Applied Probability & Statistics

Mathematics - XXXIV

Mario Lefebvre

Applied Probability and Statistics

Mario Lefebvre Applied Probability and Statistics With 58 illustrations Springer

To my parents

Preface This book is based mainly on the lecture notes that I have been using since 1993 for a course on applied probability for engineers that I teach at the lcole Polytechnique de Montrral. This course is given to electrical, computer and physics engineering students, and is normally taken during the second or third year of their curriculum. Therefore, we assume that the reader has acquired a basic knowledge of differential and integral calculus. The main objective of this textbook is to provide a reference that covers the topics that every student in pure or applied sciences, such as physics, computer science, engineering, etc., should learn in probability theory, in addition to the basic notions of stochastic processes and statistics. It is not easy to find a single work on all these topics that is both succinct and also accessible to non-

mathematicians. Because the students, who for the most part have never taken a course on prob- ability theory, must do a lot of exercises in order to master the material presented, I included a very large number of problems in the book, some of which are solved in detail. Most of the exercises proposed after each chapter are problems written es- pecially for examinations over the years. They are not, in general, routine problems, like the ones found in numerous textbooks. The exercises that can be done after a given section is read are listed in Appendix C. The reader will also find, in Appendix D, the answers to all the multiple choice questions proposed in the manual. (Of course, the student is recommended to first try to solve an exercise before looking at the answer.) Appendix E provides the answers to selected supplementary exercises included in Chapters 6 and 7. The book contains a few biographical notes on nearly all the mathematicians mentioned in the text. The reader interested in learning more about these great math- ematicians can consult the various books or Web sites dedicated to the biographies of scientists. Most of the figures in this book were realized with the help of a software program that enables one to draw curves or diagrams. When the figures involved mathematical functions, such as the exponential function, a mathematical software was used, when possible, to obtain precise curves.

Contents Preface

Contents xi

List of Tables

List of Figures

1 Introduction 1.1 The Beginnings of Probability We often hear that the theory of probability started in the seventeenth century, when a French nobleman, the Chevalier de Mr, proposed the following problem in 1654 to his friend Pascal: Why is one more likely to obtain a "6" in four throws of a die than to obtain a double "6" in 24 throws of two dice? This problem is known as de Mir's paradox. We use the word paradox, because, based on the fact that there are 6 possible results when we roll a die and 36 possible results when we roll two dice, some people thought that the two events above should have the same probability. Indeed, notice that the number of throws, divided by the number of possible results, is equal to 2/3 in both cases (4/6 = 24/36 : 2/3). Nowadays, we can easily compute the probability of each event. We find that the probability of obtaining at least one "6" in four rolls of a (fair or non-biased) die is 1 - (5/6) 4 = 671/1296 _ 0.5177, while the probability of getting at least a double "6" in throwing two dice 24 times is 1 - (35/36) 24 -- 0.4914. We can deduce that the Chevalier de Mr must have spent a lot of time throwing dice to discover such a small difference! According to some historians, this problem was not proposed by de Mr. William Feller, who wrote two books on probability theory which are considered as true classics, mentions that the problem was in fact first treated by Cardano I in the preceding century. Gerolamo Cardano, or Jerome Cardan as he is called in En- glish, was a

colorful character who, in addition to being a mathematician, was also a doctor and an astrologer. He was an inveterate gambler, which caused him many problems. He also analyzed dice games and a card game similar to poker. Further- more, he made astrological predictions. It is said that he had predicted the day of his death. Since he was in good health when the day in question arrived, he would I Gerolamo Cardano, 1501-1576, was born and died in Italy. He became interested in the do- main of probability to gain an advantage over his opponents in card and dice games. He also worked on algebraic equations. He gave the resolution method for third- and fourth-degree equations. The formula for the solution of third-degree equations had been previously ob- tained by the Italian mathematician Tartaglia.

2 1 Introduction certainly have committed suicide so as not to lose face! Note that another version of the story says that he managed to die of hunger on that day. Remark. In order to pay homage to the great mathematicians who left their mark on the history of probability, we have included some biographical notes on almost every one whose name appears in the book, for example, Poisson, who gave his name to a random variable and an important stochastic process. One thing is certain, Pascal 2 exchanged correspondence with Fermat 3 concerning the above-mentioned problem (see reference [5], p. 128, where an excerpt of a letter is reproduced). They also exchanged letters about other games of chance, including one known as the problem of points, which contributed greatly to the development of the domain of probability. The first complete treatise on the calculus of probabilities was written by Huy- gens 4 in the seventeenth century. It is however James (or Jacob) Bernoulli (see p. 70), in posthumous works published in 1713, who really founded the calculus of probabilities. Afterward, Laplace (see p. 85), with his book Thorie analytique des probabilitds written between the years 1812 and 1820, developed the theory of prob- ability in a more rigorous way. Finally, in the twentieth century, Kolmogorov (see p. 227) gave the domain of probability its modern formulation. 1.2 Examples of Applications If the conditions under which an experiment is carried out determine the result of this experiment, then it is said to be deterministic. For example, suppose that we observe an object moving in the sky along a decreasing exponential trajectory. Suppose also that we can control this object from a distance. Let x be the height of the object with respect to the ground at time 0. Then, if there are no perturbations, we can use the following model to determine the value of x(t), the height of the object

at time t: 2 Blaise Pascal, 1623-1662, was born and died in France. His father, ltienne, was his pro- fessor. He invented a calculating machine to help his father with his work as a tax collector. Along with Fermat, he is one of the founders of the theory of probability. He was also in- terested in geometry and physics. From 1654 on, he became deeply religious and published books on philosophy and theology. 3 Pierre de Fermat, 1601-1665, was born and died in France. He is especially known for his work on number theory, in particular his famous "last theorem:' He has been a precursor in the domains of probability, differential calculus and analytic geometry. He was also a lawyer and, in addition to doing his research in mathematics, he was a councillor for the parliament at Toulouse. 4 Christiaan Huygens, 1629-1695, was born and died in the Netherlands. He studied law and mathematics at the university of Leiden. Descartes showed interest in his mathematical progress. He worked, in particular, in astronomy, mechanics and optics. In 1655, using instruments he had made himself, he discovered Titan, the largest satellite of Saturn. The theory of pendulum motion is also due to him.

1.2 Examples of Applications Figure 1.1. Example of a system. dx(t) = -ax(t) dt + bu(t) dt, (1.1) x(O) = x, where a (> 0) and b are constants and the variable u(t) is called the control. Whittle (see [24]) and the author used the equations above in research papers as a rudimentary model for the displacement of an airplane that is preparing to land. When we cannot predict the result of an experiment repeated under the same con- ditions, we say that it is a random experiment. We can then make a list of all possible outcomes and try to compute the probability of each of these possible outcomes. In this textbook, we are only interested in random experiments. Probability is used in practically every pure or applied science. Examples in en- gineering, particularly in electrical engineering, where we must resort to the calculus of probabilities are the following. a) Many systems may be represented by a number of components placed in series or in parallel. For example, consider the system described by the diagram in Fig. 1.1. In reliability, we must be able to calculate the probability that such a system will function during a certain period of time, or at a given instant. In the latter case, we must then know, for each component, the probability that it functions at this instant and take into account the fact that the components perhaps do not operate independently from one another. In the case of reliability during a given period of time, we must know the distribution of the lifetime of each component. b) In communication,

we must often take into account the "noise" present in a sys- tem. For example, suppose that a system transmits either a 0 or a 1, and that there is a risk p that the number transmitted will not be received correctly (see Fig. 1.2). We may be interested in computing the probability that a 0 has been transmitted, given that a 0 has been received, or that a transmission error has occurred, etc. c) In automatic control, if we take the random perturbations into account, then the model (1.1) above becomes dx(t) = -ax(t) dt + bu(t) dt + dW(t), (1.2) where W(t) is called a Brownian motion (or a Wiener process). We also say that dW(t)/dt is a Gaussian white noise. Equation (1.2) is an example of a stochastic differential equation. d) In computer science, probability is used, in particular, to help us make decisions in expert systems. We also make use of probability in simulation and in artificial intelligence, as well as in the field of queueing theory.

Introduction Transmission Reception 1-p 0 0 ------------- p 1 P _----- 0 1-p 1 Figure 1.2. Example of a communication system. e) In physics, the domain known as statistical physics requires some knowledge of the theory of probability, and so does that of quantum mechanics. In this last case, we must be familiar with diffusion processes, such as the Wiener process. 1.3 Relative Frequencies To obtain the probability of one of the possible outcomes when we perform a random experiment, we repeat this experiment a large number of times. Let Nk (n) fk(n) := , (1.3) n where Nk (n) is the number of times that the possible outcome k has occurred during n repetitions of the experiment. The quantity f (n) is called the relative frequency of outcome k. If there are K possible outcomes, which we denote by 1,2 ..... K, then we may write that O< fk(n) < 1 fork: 1,2 ..... K (1.4) and that K Z fk(n) : 1. k=l Indeed we have, of course: Nk(n) {0, 1 ..... n}, so that fk(n) [0, 1], and (1.5) x-._X Nl(n)+...+Nx(n) n)(n)= -- -- 1. n n k=l Moreover, if A is a set that contains two possible outcomes, j and k, then fa(n) = fj(n) + fk(n), (1.6) because j and k cannot occur on the same repetition of the experiment.

1.3 Relative Frequencies 5 For instance, if we consider the random experiment that consists in observing the outcome of the roll of a die, then there are K = 6 possible results: 1, 2 ..... 6. Let A be the set {1, 6}. Because we cannot obtain both "1" and "6" on the same roll, we may write that fA(n) = f(n) + f6(n). Finally, the probability of the outcome k is obtained by taking the limit of fk(n) as the number n of repetitions tends to infinity: P[{k}] := lim f(n). (1.7)

2 Elementary Probabilities 2.1 Basic Concepts Definition 2.1.1. An experiment that can be repeated under the same conditions and whose outcome cannot be predicted with certainty is called a random experiment. Example 2.1.1 A box contains 10 brand A transistors and 10 brand B transistors. We consider the following four random experiments: El: three transistors are taken, at random and with replacement, and the number of brand A transistors (among the three selected) is counted. Remark. Sampling with (respectively without) replacement means that the object that has been selected is (resp. is not) replaced in the box before taking the next one. Therefore, in the case of sampling with (resp. without) replacement, the same object can (resp. cannot) be selected more than once. E2: three transistors are taken, at random and with replacement, and the brand of each transistor is noted. E3: transistors are taken one at a time, at random and with replacement, until a brand A transistor has been obtained; the number of brand B transistors taken before ob- taining a brand A transistor is counted. E4: a transistor is taken at random and its lifetime is measured (in hours). Definition 2.12. The set S of all possible outcomes of a random experiment is called the sample space of this experiment. Each possible outcome is also called an ele- mentary event. When a repetition of a random experiment is performed, one and only one of the elementary events occurs. That is, the elementary events are incompatible (or mutually exclusive) and exhaustive. Example 2.1.1 (continued) Corresponding to the random experiments above, we have the following sample spaces:

8 2 Elementary Probabilities S1 = {0, 1, 2, 3}. $2 = {AAA, AAB, ABA, BAA, ABB, BAB, BBA, BBB}. $3 = {0, 1 .... }. 84 = [0, o). The number of elementary events in a sample space may be finite ($1 and $2), denumerably infinite ($3), or non-denumerably infinite ($4). Remark. A set is called denumerably (or countably) infinite if we can establish a one-to-one relationship between its elements and the positivc integers. A sample space that is finite or denumerably infinite is said to be discrete, whereas if S is non-denumerably infinite, we say that it is continuous. Remark. In this chapter, we will not consider the case when the sample space S would be the union of a finite or denumerably infinite set of points and a non-denumerably infinite set of points. For instance, let E be the following random experiment: first a coin is tossed; if we get "tails," then the point 0 is chosen, otherwise a point is taken at random in the interval [1, 2]. In this case, we would have S = {0} t [1,2]. An example like this will be called a "mixed" type in Chapter 3. Definition 2.1.3.

An event is a subset of the sample space S. Thus, the empty set 0 and the sample space S itself are events. Remarks. i) The empty set 0 is called the impossible (or null) event and S is the certain event. ii) There are 2 n events that can be defined with n elementary events. iii) We generally use capital letters like A, B, etc., to denote events. Example 2.1.1 (continued) Events defined with respect to the sample spaces asso- ciated with the random experiments above are the following: A: exactly one brand A transistor is obtained; that is, A = {1 }. A2: one brand A transistor and two brand B transistors are obtained; that is, A2 = {ABB, BAB, BBA}. A3: five or six brand B transistors are picked before a (first) brand A transistor is obtained; that is, A3 = {5, 6}. A4: the selected transistor lasts more than 200 hours; that is, A4 = (200, o3). Operations with Sets (see Fig. 2.1, p. 9) Union: A U B denotes the set of outcomes that belong to A or to B (or to both). Similarly, in general, for n events we write: n A1 CI A2 CI . . . CI An = U Ai. i=1

2.1 Basic Concepts 9 Union S Union A B Intersection Intersection A B Incompatibility S Incompatibility S A S Complement Inclusion Figure 2.1. Set theory: Venn diagrams. Intersection: A fq B denotes the set of outcomes that belong to both A and B. In general, we have: n A1 fq A2 fq . . . fq An = [Ai. i=1 If two events are incompatible, then we write: A fq B = 0. Complement: The complement of an event A is the set of outcomes that do not belong to A; it is denoted by A c. Inclusion: If all the outcomes that belong to event A also belong to event B, then we say that A is included in B and we write: A C B. Equality: Two events are said to be equal if they contain the same outcomes; we then write: A = B. We can easily show the following relationships: 1) A U B = B U A and A fq B = B fq A (Commutativity). 2) A U (B U C) = (A U B) U C and A fq (B fq C) = (A fq B) fq C (Associativity).

10 2 Elementary Probabilities 3) AU(B7C) = (AUB) f(AC)andAf(BC) = (A f B) (A 7 C) (Distributivity). 4) (A f) B) c = A c B c and (A B) c = A c f) B c (De Morgan's laws). 2.2 Probability Definition 2.2.1. Let E be a random experiment and S a sample space associated with E. To each event A in S we assign a real number noted P[A], called the prob- ability of A, so that the following properties are satisfied: Axiom I: P[A] > 0 � A C S; Axiom Ih P[S] = 1 : Axiom III: If A , A2 .... is a sequence of incompatible events, then (2.1) Remarks. i) We deduce from Axiom III that if A 7/B = 0, then P[A B] = P[A] + P[B]. (2.2) Indeed, we only have to take A = A, A2 : B and A, = 0 for k

> 3, because P[0] -- 0 (see Proposition 2.2.1). ii) The function P is called a probability measure; it is a function from S into the interval [0, 1]. With the help of the three axioms above, we easily show the following proposi- tion. Proposition 22.1. We have: 1) P[A c] : 1 - P[A]. 2) P[A] < 1. 3) P[01 = 0. 4) PIOAk]=P[Ak]-ZP[AjAk]+'"+(-I'n+Ip[Ak]. k=l k=l j<k k=l 5) If A C B, then P[A] < P[B].

A B AnB 2.2 Probability 11 Figure 2.2. Figures for the proof of parts 4) and 5) of Proposition 2.2.1. Proof. 1) Since A c U A = S and A c A A ---- 13, we deduce from Axioms II and III that P[A cU A] I 1 and P[A cU A] P[A c]+ P[A] :: P[A �]= 1- P[A]. 2) P[A] = 1 - P[A c] < 1 because P[A c] > 0 (by Axiom I). 3) We have:S c =13and P[S c]= 1- P[S]= 1-1=0. 4) For n = 2, let us write A1 = A and A2 ----- B; we then have (see Fig. 2.2): P[AUB]= P[AAB c]+ P[AB]+ P[A cB], because the three events are incompatible. Moreover, P[AB e]=P[A]-P[AB] and P[A cAB]=P[B]-P[AB]. So, we have: P[A U B] = P[A] + P[B] - P[A B]. (2.3) Next, let D := B U C; then, we may write that P[A U B U C] = P[A U D] = P[A] + P[D] - P[A D] = P[A] + P[B] + P[C] - P[B C] - P[A A (B U C)] = P[A] + P[B] + P[C] - P[B C] - P[(A B) U (A C)]. Hence, we obtain the following formula: P[A U B U C] = P[A] + P[B] + P[C] - P[A B] - P[A C] - P[B C] + P[A B C]. (2.4) To prove the formula in the general case, we proceed by induction (that is, we assume that the formula is true for the case of n events and we try to show that it is then also valid for n 1 events). 5) If A C B, we may write (see Fig. 2.2) that P[A] = P[B] - P[A c B] =: P[A] < P[B]. []

12 2 Elementary Probabilities Discrete Sample Spaces If S is a discrete sample space, we may write that S = {el, e2 .... }, where ek is a possible outcome (or an elementary event). Let A C S; then the proba- bility of event A can be obtained by making use of the following formula: P[A] = Z P[Iek}]. (2.5) k:ekEA If the number n of elementary events ek is finite and if the ek's are equiprobable (or equally likely), so that P[{ek}] = 1/n Yk, then we may write that P[A] = n(a)/n, (2.6) where n(A) is the number of elementary events in A. Example 2.2.1 i) In the case of the sample space $I in Example 2.1.1, the four elementary events are not equiprobable. If we denote the probability P[{0}] by p, then we may write that P[{I}]=P[{2}]=3p and P[{3}l=p. Moreover, Hence, we find that 3 P[{k}] = I p = 1/8. k=0 P[A]= P[{1}]----3/8 (# 1/4). On the other hand, the ek's are equally likely in the case of $2 and we may write,

directly, that P[A2] : P[{ABB, BAB, BBA}] ---- 3/8. Note that the probabilities P[AI] and P[A2] must be equal because the events A and A2 correspond to the same outcome of the three draws in the random experiments El and E2. Thus, to calculate the probability of getting exactly one brand A transistor in three draws at random and with replacement from a box containing 10 brand A and 10 brand B transistors, it is simpler to consider the sample space $2 for which the outcomes are equally likely. ii) Since the sample space $3 is denumerably infinite, the ek's cannot be equiprobable. We can show that P[A3] : P[{5, 6}] : P[{5}] + P[{6}] = (10/20)5(10/20) + (1/2)6(1/2). The second equality above is obtained by incompatibility of the events {5} and {6}, while the third equality results from the notion of independence (which will be dis- cussed in Section 2.5).

2.3 Combinatorial Analysis 13 Continuous Sample Spaces When the number of possible outcomes is non-denumerably infinite, each et has a zero probability of occurring. We must then give a formula that enables us to calcu- late the probability that the outcome of the random experiment will be located in any interval [a, b]. Example 2.2.2 Suppose that P[(a, b]] : e -a/lO0 - e -b/lO0, where $0 < a < b < x$, is the probability that the lifetime of the transistor taken at random in the random experiment E4, in Example 2.1.1, belongs to the interval (a, b]. Then, we may write that P[A4] : P[(200,)] : e -2 - e - -- 0.1353. 2.3 Combinatorial Analysis 1) Tree diagrams. In order to more easily make the list of all possible outcomes when we perform a random experiment, we sometimes use a tree diagram. For ex- ample, suppose that we perform the experiment E2 in Section 2.1. The eight possible outcomes are obtained by following each branch or path in the diagram of Fig. 2.3. 2) Principle of multiplication. Suppose that we perform k consecutive random ex- periments Ei. Suppose also that the outcome of a given random experiment does not influence the subsequent experiments. Let ni be the number of possible outcomes in the sample space Si of the experiment El, for i = 1, 2 ..... k. Then there are nl �n2 � ... �nk possible results in all. Indeed, this corresponds to the number of ways of choosing an element from a set that contains n 1 elements, then an element from another set that contains n2 .... and finally an element from a set that contains nt. lStdraw 2nddraw 3rddraw Figure 2.3. Example of a tree diagram.

14 2 Elementary Probabilities 2 a T - Figure 2.4. Tree diagram drawn

compactly. For example, if we want to buy a computer system made up of three components: the computer itself, a screen and a printer, and if we can choose between three brands of computers, two brands of screens and four brands of printers, then we can buy 3 x 2 x 4 = 24 different systems. Similarly, if a cafeteria is offering for lunch a choice of two soups, three main courses and four desserts, then the person working at the cash register can come across 3 x 4 x 5 - 1 = 59 different trays (since a customer does not have to buy a dessert, for instance, but he or she must buy something). Remark. It is important that the outcome of a given experiment does not influ- ence the subsequent experiments. For example, consider the following random experiment: a die is thrown, then a coin; if "tails" is obtained, then the die is thrown a second time. In this case, the sample space of this experiment is S = { 1H ..... 6H, 1T 1 ..... 6T6} (see Fig. 2.4). So, there are 6 + 36 = 42 elementary events in S, and not 6 � 2 � 6 = 72. The first six elementary events are equiprobable among themselves if the die is non-biased, and so are the last 36. In fact, using the notion of independence (see Section 2.5), we immediately find that the first six ele- mentary events have a probability of 1/12, and the others have a probability of 1/72 (if the coin also is non-biased). 3) Permutations. Suppose that we possess n distinct objects and that we pick k at random. If the order in which the objects are chosen matters and if the objects are taken with replacement, then the number of different permutations (or arrangements) of these objects that we can obtain is given by n X n x ... X n nk; (2.7) in the case when the objects are taken without replacement, the number of different permutations is n x (n-l) x... x [n - (k -1)] -- -- n := (n), := Pk n, (2.8) (n - k) ! where k 6 {(0,)1 ..... n}. Example 2.3.1 Suppose that we want to create access codes, to a computer system, consisting of five letters taken among the 26 letters of the alphabet. Then, we can

2.3 Combinatorial Analysis 15 create 265 different codes if the same letter can be used more than once. However, if repetitions are not allowed, the number of different codes is given by 26 x 25 x 24 x 23 x 22 = 7,893,600. Remarks. i) If k = n, we have: P = n !. ii) Stirling's formula: we can show that n! (27rn)l/2nne - n. That is, the ratio n!/(2rn)l/2nne -n tends to 1 when n tends to infinity. 4) Combinations. Suppose now that we take k objects, at random and without re- placement, among n distinct objects and that the order in which the objects are cho- sen does not matter. Then, the number of different combinations (or samples) that we can obtain is given, for k 6 {(0,) 1 ..... n}, by n x (n-1)

x...x[n-(k-1)] n! () -- "-- := C;. (2.9) k! k!(n - k)! Remarks. i) Each combination of k objects enables us to form k! different permuta- tions. ii) We can easily check that C = C_ k . iii) We can show that if the objects are taken with replacement, then the number of different combinations is C +k- 1. In this case, there are no restrictions on k. iv) We have:C=0ifk >nork <0. Example 2.3.2 Suppose that we have ten diskettes in a box. If we pick three, at random and without replacement, then there are C10 - 10! -- 120 3!(10 - 3)! different groups of three diskettes that we can obtain. In practice, it is not easy to take each diskette in a really random way, so that the first diskette that we pick had 1 chance in 10 of being chosen, the second one 1 chance in 9, and the third one 1 chance in 8. To do so, we can use a random numbers generator. For instance, if we have access to a computer that generates random numbers between 0 and 1, we can generate three numbers (or more, if necessary) and take first diskette no. 2 if the first number generated is located in the subinterval [1/10, 2/10), etc. We can also use tables of random numbers. 5) Permutations of non-distinguishable objects. Suppose finally that among the n objects that we have, there are ni of type i, where i ---- 1, 2,..., k and n l q- n2 q-

16 2 Elementary Probabilities � � � + nk = n. Then, the number of different permutations of the n objects taken all at once is given by n! (2.10) nl !n2! ...nk!" The numbers obtained for varying values ofni in this expression are called the multi- nomial coefficients. Example 2.3.3 With the letters a, a, b, b, b and c, we can form 6 -- - 60 2!3!1! different code words (or access words). We will complete this section with other examples of combinatorial analysis problems. Example 2.3.4 If a letter (or two) may be repeated (only once) in Example 2.3.1, then there are P526 , C26C45! , C26C 3 5! � 4 1. - 3 22!2 ! =7,893,600+3,588,000+234,000=11,715,600 different codes. The first term above corresponds to the case when there are no repe- titions, the second term to the case when there is exactly one repetition and the third one to the case when there are exactly two repetitions. Moreover, we may write that P526 = C265 "5!" Example 2.3.5 With the letters in the word essays, the number of distinct words containing four letters that we can form is given by C 34! (with2s's) ' C 34! (with3s's) =24+36+12=72. 4! (with ls) + 22! + 3! Example 2.3.6 If we have one one-dollar coin, two quarters and three nickels, then we can pay 2 � 3 � 4 - 1 = 23 different sums exactly. Note that we can employ the principle of multiplication here, because whatever the number of coins of a certain type that we use to pay, the sum will always be different. The problem

would be more complicated if we also had a dime, for instance. Example 2.3.7 If there are n persons in a class, then the probability that at least two of these n persons have the same birthday is given (disregarding leap years) by p365 1 - n for n = 2, 3 ..... 365. 365 n

2.3 Combinatorial Analysis 17 Note that the term Pn 365 in the numerator corresponds to the number of permutations of n objects taken at random and without replacement among 365 distinct objects, while 365 n is the number of permutations with replacement of n objects taken among 365 distinct objects. We might think that this formula is equivalent to that obtained by computing the number of combinations with and without replacement, that is, Cn 365 1 for n = 2, 3 ..... 365. Cn365+n- 1 However, this last formula is wrong, because the combinations with replacement are not equally likely, so that we cannot simply divide the number of favorable cases by the total number of cases, like we did with the permutations. Example 2.3.8 We have 20 components of type I, 5 of which are defective, and 30 components of type II, 15 of which are defective. a) We want to build a system made up of ten components of type I and five compo- nents of type II placed in series. What is the probability that the system functions if the components are taken at random? b) How many distinct systems made up of four components placed in series, with at least two components of type I, can be constructed if the order of the components is taken into account? Solution. a) The total number of distinct systems that can be built is given by C12o � � C5 �. The number of distinct systems that function is Co � � c2 � Then, by equiprobability, we can write that the required probability is C�5 � C5 0.00034. � - b) The total number of distinct systems that we can construct is P45�; among those, there are P43� that contain no components of type I, and 20 � 4 � P33� that contain exactly one. Then, the required number is Pa so- Pa 3�- 80 � pO. Remark. If we assume that we cannot distinguish between two components of the same type, then the number of distinct systems that we can construct is given by 6 (with 2 components of type I) + 4 (with 3 components of type I) + 1 (with 4 components of type I) = 11.

18 2 Elementary Probabilities 2.4 Conditional Probability Let A and B be two events defined with respect to a sample space S associated with a random experiment E. Suppose that we perform the experiment E and that the event B occurs. Then B becomes the new sample space, for this trial, and in order that A too has occurred, A N B must now have occurred. Notation. The expression

P[AIB] denotes the probability of the event A, given that the event B has occurred. Definition 2.4.1. We write: P[A f3 B] PIAIB] -- P[B] ifPlB] > 0. (2.11) Remarks. i) The conditional probabilities satisfy the three axioms in the definition of probability (see p. 10). In fact, every probability is a conditional probability since, for any A C S, PIA S] PIA] PIAIS]- -- -- PIAI. PlS] 1 It follows that Proposition 2.2. l is still valid for conditional probabilities. For exam- ple, we may write that P[ACl B] : 1 - P[AIB]. However, in general, PIAIB"] 1 - PIAIBI. Moreover, PIA LJ BIG] = PIAIC] P[BIC] - P[A BICI, if P[C] > 0, etc. ii) IfA B = 0, then PIAIB] = O. iii) If B C A, then P[AIB] = I. Furthermore, if A C B, then P[A] P[AIB] -- P[B]" iv) We may have: P[AIB] < P[A], P[AIB] > P[A],or P[AIB] = P[A].That is, there is no relationship between conditional probabilities and the corresponding marginal probabilities. From Definition 2.4. l, we deduce at once the following proposition.

2.4 Conditional Probability Proposition 2.4.1. (Multiplication rule) We have: P[ANB]= P[AIB] $\diamond$ P[B] ifP[B] >O, P[ANB]=P[BIA]$\diamond$ P[A] ifP[A]>O. In general, P[A1 fq A2... fq An] = P[A1] $\diamond$ P[A21A1] $\diamond$ P[A31A1 fq A2] $\diamond$ -" $\diamond$ P[AnlA1 fq A2 fq... fq An-l] (2.13) if P[A1 fq A2 fq ... fq An-l] > O. 19 (2.12a) (2.12b) Example 2.4.1 Two components are taken at random and without replacement in a box containing ten brand A and ten brand B components. What is the probability of getting a) two brand A components? b) two components of the same brand? c) two components of different brands? Solution. Let Ak = a brand A component is obtained on the kth draw. a) We want P[A1 fq A2] = P[A2IA1]P[A1] = 9/19 x 10/20 = 9/38. Note that, according to the multiplication rule, we may also write that P[A1 fq A2] = P[AllAz]P[A2]. However, this formula does not enable us to directly find the probability P [A fq A2]. b) By symmetry (since there are as many brand A components as brand B compo- nents), we deduce from the preceding result that the required probability is given by 9/38 + 9/38 = 9/19. c) We now deduce from b) that the required probability is 1 - 9/19 = 10/19. Proposition 2.4.2. Let A and B be two events such that P[A] $\diamond$ P[B] > O. Then, P[BIA]P[A] P[AIB] -- (2.14) P[B] Proof. We only have to make use of the definition of P[AI B] and the multiplication rule. [] Remark. This result is sometimes called Bayes'formula. 1 The Reverend Thomas Bayes, 1702-176l, was born and died in England. He was first educated by tutors. After his ordination, he worked with his father, who was also a pastor. His works on probability theory were

published in a posthumous scientific article in 1764. He wanted to find a method by which "we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times ?' Nowadays, a statistical school of thought is called Bayesian.

20 2 Elementary Probabilities S B3 B 4 Figure 2.5. Example of a partition of a sample space with n = 4. Definition 2.4.2. Let B , B2 ..... Bn be events such that i) Bi fq Bj = �i # j; ii) I,_): Bk = S. We say that the events Bk form a partition of the sample space S. Remarks. i) In probability, it is natural to impose in the preceding definition the additional condition: iii) P[Bk] > 0 fork = 1,2 ..... n. ii) A and A c constitute a partition of S. Similarly, A fq B, A fq B", A'" fq B and A c fq B c constitute a partition of S. Now, let A C S. We can split up A into n disjoint and exhaustive parts as follows (see Fig. 2.5): A = (A fq B) L) (A fq B2) L)... L) (A fq B,,). Hence, using Axiom III in the definition of P[A], we obtain the following proposi- tion. Proposition 2.4.3. (Total probability rule) Let B1, B2 ..... Bn be a partition of S. Then, for any A C S we have." P[A] = P[A f3 Bk] = � P[AIBklP[BkI if P[Bkl > OVk. k=l k=l (2.15) Remark. The second equality above follows from the multiplication rule. Finally, making use of Proposition 2.4.2 and the total probability rule, we obtain Bayes' rule. Proposition 2.4.4. (Bayes' rule) Let A C S and Bt, B2 ..... Bn be a partition of S such that P[Bk] > O for k = 1, 2 ..... n. Then, P[AIBj]P[Bj] P[BjlA] = =I p[AIBk]P[Bk] for j= 1,2,...,n. (2.16)

2.5 Independence 21 Example 2.4.2 We consider a communication system that transmits either a 0 or a 1. Because of the "noise," the signal transmitted is sometimes received incorrectly. We define the events Ei = i is transmitted and Ri = i is received for i = 0 and 1. We assume that P[RoIEo] = 0.7, P[R1 [El] = 0.8 and that a 0 is transmitted 60% of the time. a) Calculate P[Eo[R1]. b) Find the probability of a transmission error. Solution. a) We have: P[Eo[R1] = P[RI[Eo]P[Eo] P[RI[Eo]P[Eo] + P[RI[E1]P[E1] (! - 0.7)(0.6) = 0.36. (1 - 0.7) (0.6) + (0.8)(0.4) b) P[Transmission error] = P[Eo f R1] + P[E1 f"l R0] = P[R1 [Eo]P[Eo] + P[Ro[E1]P[E1] = (1 - 0.7)(0.6) (1 - 0.8)(0.4) = 0.26. Remark. The events E0 and E1 constitute a partition of S, and so do R0 and R1. 2.5 Independence Definition 2..1. Let A and B be two events. We say that A and B are independent if P[A f B] : P[A]P[B]. (2.17) Remarks. i) Two

independent events may or may not be incompatible. However, if two events A are B are incompatible and independent, then P[A] or P[B] (or both) must be equal to zero. Indeed, we then have: P[A B] in-Sc' P[O] = 0 and P[A B] ind. P[A]P[B] => P[A] = 0 or P[B] = O. ii) We also define conditional independence as follows: A and B are said to be conditionally independent with respect to C if

22 2 Elementary Probabilities P[A fq BIC] = P[AICIP[BIC]. (2.18) Equation (2.18) may be satisfied, whether the events A are B are independent or not. Moreover, two independent events may be conditionally dependent with respect to an event C. For instance, suppose that we throw a (non-biased) coin twice, indepen- dently. Let us define the following events: A = "heads" is obtained on the first throw, B = "heads" is obtained on the second throw, C = exactly one "heads" is obtained in all. Then A and B (as well as A and B c) are independent, but (0 =) P[A (q BIC] PIAICIPIBIC] > O. In fact, we have: PIAIC] - PIAfqCI P[AfqB"I ind. (1/2)(1/2) PICI PICI (1/2) 2 + (1/2) 2 and P[BIC] = PIAIC],by symmetry. = 1/2 Proposition 2.5.1. Two events, A and B, such that PI A] x PI B] > O, are indepen- dent if and only if PIAIBI= PIA] or PIBIAI= P[B]. (2.19) Proof. We have: P[A fq B] = P[AIB]P[B]. Then, if A and B are independent, we obtain that P[A]PIB] = P[AIB]P[B] := PIAIB] = PIA]. Similarly, since P[A fq B] = P[BIA]P[A], we also have: P[BIA] = PIB]. Conversely, if P [A I B] = P [A], then P[A fq B] P[A] -- := P[A]P[B] = P[A fq B]. P[B] In a similar way, we show that if P[BI A] = P[B], then P[A fq B] = PIA] P[B]. [] Proposition 2.5.2. If A and B are independent, then so are A " and B. Proof. Let us consider the case when P[A] � P[B] > 0. We have: p[ACIB] = 1 - P[AIB] ind. 1 - P[A] = p[AC]. [] Remark. Likewise, we show that A and B c, as well as A c and B c, are independent events.

2.5 Independence 23 Definition 2.5.2. The events A1, A2 ..... An are said to be pairwise independent if P[Ai 71 A j] : P[Ai]P[Aj] �i 7A j, (2.20) where i, j = 1, 2 ..... n. Remark. There are C conditions to check. Definition 2.5.3. The events At, A2 ..... An are said to be (globally) independent if for any k < n, whatever the events Ai(1) ..... Ai(k), we may write that P[Ai() 71Ai(2) 71... 71 Ai(k)] = P[Aio)]P[Ai(2)]... P[Ai(k)], (2.21) where i(j) 7A i(m) if j 7A m. Remarks. i) This means that the events can be taken 2, 3, ... , n at a time and the probability of the intersection is always equal to the product of the marginal proba- bilities, that is, of the probabilities of the individual events. ii) This

time, the number of conditions that we must check is C = C - C - C = 2 n - n - 1, k=2 k=0 by Newton's binomial theorem. 2 Example 2.5.1 We consider a system made up of three components that operate independently from one another. We suppose that the system functions if at least two of its components are working. A system of this type may be represented by the diagram in Fig. 2.6, p. 24. Let the events be as follows: F = the system is functioning at time t and F/ = the component i is functioning at time t, for i = 1,2 and 3. Then, if we assume that P[F/] = 0.9 for all i, we can write (by symmetry) that P[F] = 3 x P[F 71 F2 71 F] + P[F 71 F2 71 F3] ind. 3 X (0.9)(0.9)(0.1) + (0.9) 3 = 0.9720. 2 Sir Isaac Newton, 1643-1727, was born and died in England. He is famous for his contri- butions to the fields of mechanics, optics and astronomy. He and Leibniz (and, according to some, Fermat) invented differential calculus. Actually there was a great controversy be- tween the two scientists about that, each one claiming to be the inventor of infinitesimal calculus. He also worked on alchemy and wrote theological books.

24 2 Elementary Probabilities Figure 2.6. Graphical representation of a two-out-of-three system. Example 2.5.2 A box contains 10 brand A diodes and 20 brand B diodes. Ten diodes are taken, at random and with replacement, and the brand of each one is noted. Com- pute the probability of getting at least one brand A diode. Solution. Let the events be as follows: and F = at least one brand A diode is obtained Ak = the kth diode selected is a brand A diode. We want P[F] = PIA U A2 U... U Ao] = 1 - PIA' i fq A fq... fq A'i0]. Now, because the diodes are taken with replacement, the events A are (globally) independent. Moreover, P[Ak] = 10/30 = 1/3 for any k. Then, we have: PIF] = 1 - (2/3) � - - 0.9827. Example 2.5.3 (see [14]) In a certain factory, 96% of the computers manufactured conform to the norms. Each computer produced is subjected to two independent quality control operations. We assume that each operation classifies as "good" 98% of the computers that effectively conform to the norms, and 6% of the computers that actually do not conform to the norms. Calculate the probability that a computer sold effectively conforms to the norms. Solution. Let the events be as follows: A = the computer has been classified as "good" twice and B = the computer conforms to the norms. We want

2.5 Independence P[AIB]P[B] P[BIA] = P[AIB]P[B] + P[AIBC]p[B c] ind. (0.98) 2 (0.96) = -- 0.9998. (0.98)2(0.96) (0.06)2(0.04) 25 Remarks. i) We

know that Bs := B and B2 := B c constitute a partition of S. ii) Let Ak = the computer has been classified as "good" by the kth control operation; then, we may write that A = As A2. Furthermore, As and A2 are conditionally independent with respect to B (and Be), but are not independent. Example 2.5.4 ("The Monty Hall problem") A man takes part in a television game show. At the end, he is presented with three doors and is asked to choose one among them. The grand prize is hidden, at random, behind one of the doors, while there is nothing behind the other two doors. The game show host knows where the grand prize has been hidden. Suppose that the man has chosen door no. 1 and that the host tells him that he did well in not choosing door no. 3, because there was nothing behind it. He then offers the man the opportunity to change his choice and, therefore, to select door no. 2 instead. What is the probability that the man will win the grand prize if he decides to stick with door no. 1 ? Solution. Let Ak = the grand prize is hidden behind door no. k, for k = 1, 2, 3, and let F = the game show host has eliminated door no. 3. Assume, in all logic, that if the man has chosen the right door, then the host will eliminate door no. 3 with probability 1/2. In this case, 1 1 1 1 P[F] = P[FIAs]P[As]+P[FIA2]P[A2]+P[FIA3]P[A3] = -+1-+0 = and then P[FIAs]P[As] 1/6 P[AsIF] = -- -- -- 1/3. P[F] 1/2 Therefore, the man has a probability of 2/3 of winning the grand prize if he decides to switch doors. In general, if there are n doors and if the host eliminates n - 2 among them (he naturally cannot eliminate the door chosen by the participant), then the probability that the grand prize is hidden behind the only remaining door, among the n - 1 doors in play, is equal to (n - 1)In. Example 2.5.5 (The liars problem) A says that B told him that C has lied. If the three persons tell the truth and lie with probability p (0, 1), independently from one another, what is the probability that C has indeed lied? Solution. Let the events be as follows: F = A says that B told him that C has lied and

26 2 Elementary Probabilities lied, for I = A, B, C. We want Now, we have: P[FcIF] = P[FIFc]P[Fc] P[FIFc]P[Fc] + P[FIF]P[F] PIFIFc] = p[FCA F(B] -t- P[FA 7) FB] ind. p2 -t- (1 -- p)2 and Hence, P[FIF] = P[F FB] + PIFA f F'I ind. 2p(1 - p). PIFcIFI = [p2 + (1 -- p)2l(1 -- p) lp 2 + (1 -- p)21(1 -- p) + 12p(1 -- P)IP Note that ifp = l/2,then P[FclF] = 1/2, which is logical. p2 -t- (1 -- p)2 3p 2+(1 --p)2" 2.6 Exercises, Problems, and Multiple Choice Questions Solved Exercises Exercise no. 1 (2.3) 3 License plates are made up of three letters followed by four digits. We assume that the letters I and O are never used and

that no license plates end with 0000. a) How many distinct license plates can there be? b) What is the answer in a) if, in addition, no plates bear either three identical letters or four identical digits? Solution a) The number of possible distinct plates is (24 x 24 x 24) x (104- 1) : 138,226, 176. b) In this case, the total number of possible plates is given by (243- 24) x (104- 10) = (13,800) (9990)= 137, 862,000. Exercise no. 2 (2.3) How many distinct code words made up of four letters can be formed by using (without replacement) the letters of the word ESSAY? 3 The material that must have been read to be able to solve each exercise is indicated in the parentheses.

2.6 Exercises, Problems, and Multiple Choice Questions 27 Solution If we take a single s, then we can form 4! = 24 distinct code words. If we use the two s's, then the number of distinct code words that can be formed is given by (3 4! = 3- = 36. Hence, the total number is 60 code words. 2] 2!1!1! Exercise no. 3 (2.3) A case holds two boxes containing ten objects each. a) Suppose that, among the 20 objects, there are exactly five that are defective. What is the probability that these five defective objects are all in the same box? b) If each box is made up of two rows of five objects each and if the 20 objects (that are distinguishable) are placed at random into the two boxes, how many distinct arrangements of the ten objects inside a box are there? Solution a) P [All defective objects in the same box] 15! 10!10! : 2 x ()(55) -- 2 x -- -- _ 0.0325. (0o) 5!10! 20! 20 b) The number of distinct arrangements is given by ()10! = Y6. ❖ Exercise no. 4 (2.3) A certain computer language uses the 26 letters of the alphabet and 10 special characters ($, #, etc.). We use these 36 characters to generate (at random) access codes, to a computer, made up of four characters. a) Let E be the random experiment that consists in counting the number of special characters in a code taken at random. i) Write the sample space S for this experiment. ii) Compute the probability of each elementary event. b) We consider a code taken at random. We define the events A = the given code contains at least one letter and B = the given code contains exactly one letter. Are the events A and B c incompatible? Justify your answer. Solution a) i) S = {0, 1,2, 3, 4}. (26 4 ii) P [{0}] = \36] " 0.2721; P [{1}] = - \36] " 0.4186;

28 2 Elementary Probabilities Figure 2.7. System in Exercise no. 5. P[{2}1= \36] \36] -0'2415; (:) (10"3 (26) P[{3}]= \36] -0'0619: (10 4 P [{4}] = \36] - 0.0060. b) P[A71B c] = P[{0}L{I}L{2}1 > 0, which implies that A and B c

are not incompatible. Exercise no. 5 (2.5) We have 10 brand A components, denoted by A ..... A0, and 20 brand B components, denoted by BI .... , B20. We want to build a system made up of two subsystems: three brand B components placed in series and two brand A components placed in parallel. Moreover, the two subsystems are placed in parallel (see Fig. 2.7). a) How many distinct systems can be built? Remark. The disposition of the components inside the subsystems does not matter. b) Assume that two brand A components are defective, while five brand B compo- nents are defective. What is the probability that the system functions if the compo- nents are taken at random and operate independently from one another? Solution a) There are (20) = 1140 possibilities for the first subsystem and (2 ◆) = 45 pos- sibilities for the second one. Then, by the principle of multiplication, there are 1140 x 45 = 51,300 distinct systems that can be built. b) We have: 15 x 14 x 13 91 P [Subsystem B functions] -- 20 x 19 x 18 -- 228 and

2.6 Exercises, Problems, and Multiple Choice Questions Figure 2.8. Figure for Exercise no. 6. 29 2xl 88 P [Subsystem A functions] = 1 10 x 9 90 (137" (1) = P [System functions] = 1 - \228] - - 0.9866. Exercise no. 6 (2.5) A network is constituted of four terminals and four links, as shown in the diagram in Fig. 2.8. Thus, there are two paths that connect any pair of terminals. Moreover, a terminal that transmits information to another terminal does so by sending this information in both directions, independently. The transmission is considered to be successful if the information is received along one path or the other (or both). Finally, we assume that the links fail, independently from one another, with probability 0.1. a) What is the probability that the information transmitted by terminal A to terminal B will be successful? b) What is the probability that the information transmitted by a given terminal to another terminal will be successful? c) In how many different ways can the four terminals be disposed inside the net- work if we suppose that two networks are identical when each terminal has the same neighbors in both networks? Solution Let F = the transmission is successful, and FAB= the information transmitted in the direction AB is well received, etc. a) P [F] = P [FAB t FACDB] ind. p [FAB] - ' P [FACDB] -- P [FAB] P [FACDB] in_d. (0.9) + (0.9) 3 -- (0.9) 4 = 0.9729. b) Let Dk = the minimum number of links between the two terminals is equal to k. We have: k = 1 or 2. Then, P[F] = P[F I DI]P[DI]+P[F I D2]P[D2] a_)_ (0.9729)()+P[FABDUFACD]()

30 2 Elementary Probabilities ind. (0.9729) () + [(0.9)2 + (0.9)2 _ (0.9)4] () = 0.9699. c) First we fix the position of one of the four terminals, and next we move the other three terminals. There are then () = 3 different networks (because the order of the neighbors does not matter): AB AB AC CD DC DB Exercise no. 7 (2.5) A system is made up of three components that operate independently from one another. For the system to function, at least two of its components must function. We suppose that the reliability of component no. I is equal to 0.95, that of component no. 2 to 0.9, and that of component no. 3 to 0.8. a) What is the probability that the system functions'? b) Given that the system functions, what is the probability that exactly two compo- nents function? c) Given that the system does not function, what is the probability that no components function? d) Given that component no. I functions, what is the probability that the system functions? a) Solution Let F = the system functions, and F i = component i functions, for i = I, 2, 3. P IF] = P [F n F2 n F] + P [F n F'n F3] + P [Fn F2 n F3] +PIF n F2 n F31 "d' (0.95) (0.9) (0.2) + (0.95) (0. I) (0.8) + (0.05) (0.9) (0.8) + (0.95) (0.9) (0.8) = 0.283 + 0.684 = 0.967. b) Let G = exactly two components function. P IF I G] P {GI) I (0.283) P{G[F]= -- -- P [F] 0.967 c) Let H = no components function. P[H [F c]= P [Fc [H]P[HI ___) 1(0.05)(0.1)(0.2) P IF c] I - 0.967 d) P [F [F] ind. p [F 2 F3] = P [F21 -t- P [F31 - P [F2 N F31 ind. = 0.9 + 0.8 -- (0.9) (0.8) = 0.98. 0.2927. =0,03.

2.6 Exercises, Problems, and Multiple Choice Questions Unsolved Problems 31 Problem no. 1 Two dice are thrown simultaneously. If a sum equal to six or ten is obtained, then a coin is tossed. a) How many elementary events (of the form (die i, die 2) or (die i, die 2, coin)) are there in the sample space S? b) Assume that the dice and the coin are non-biased (or well-balanced). Given that the coin has been tossed, what is the probability that a "i" has been rolled with the second die? Problem no. 2 LetAandBbeeventsforwhichP[ArB]= P [A c r B] = P [A r B c] = p (see Fig. 2.9). Calculate a) P [A c B']; b) P [A c kJ BC]. Problem no. 3 In a certain lottery, five balls are picked, at random and without replacement, among 25 balls numbered from 1 to 25. We win the grand prize if the five balls that we have chosen are selected in the same order as they appear on our ticket. a) What is the probability of winning the grand prize? b) What is the probability of not winning the grand prize because of a single ball? Problem no. 4 We have nine electronic components, including two defective ones. Four compo- nents are taken at random to construct a system in

series. a) What is the probability that the system does not function? b) If a fifth component (taken among the remaining five) is placed in parallel with the first four, what is the probability that the system will function? Problem no. 5 Let P[A [B] = ◆, P[B c] = ◆ and P[ArB c] = . Calculate a) P[A];b) p[ACT) B]. Problem no. 6 Let A and B be independent events such that P [A] < P [B] and A B Figure 2.9. Figure for Problem no. 2.

32 2 Elementary Probabilities P[A I B]+ P[B I A]= 1. Sl Transmission $2 0.05 s3 0.02 a) Calculate P [A] if P [A 71 B] = 2" b) Calculate P [B] if P [A U B] : 3"19 Problem no. 7 A communication system transmits three signals: sl, s2 and s3, with equal prob- ability. The reception is sometimes erroneous, because of the "noise." It was found, experimentally, that the probability Pij of receiving the signal s.i, given that the sig- nal si has been transmitted, is given by the following table: Reception)0"115 $2 5"3 0.1 0.90 (0.08 0.901 a) Calculate the probability that the signal s has been transmitted, given that the signal s2 has been received. b) If we assume that the transmissions are independent, what is the probability of receiving two consecutive s3 signals? Problem no. 8 A box contains five brand A, three brand B and two brand C transistors. The transistors are all distinguishable. In how many ways can we pick, at random and without replacement, four transistors if we want to get at least one of each brand? Remark. The order in which the transistors are selected does not matter. Problem no. 9 According to the data collected, 40% of all human beings have type A blood, 10% have type B, 45% type O and 5% type AB, Moreover, we think that 90% of people who have type O blood are incorrectly classified, whereas 3% of people with type B, 10% with type AB and 2% with type A blood are classified as having type O blood. a) What is the probability that a person classified as having type O blood really has this type of blood? b) If we assume that each person is classified independently from the others, what is the probability that two given persons classified with type O blood do not have this type of blood? Problem no. 10 A diskette is taken at random from a box containing 10 brand A, 15 brand B and 25 brand C diskettes. We assume that, in general, 95% of the brand A diskettes are perfect. This percentage is 97% for the brand B diskettes and 99% for those of brand C.

2.6 Exercises, Problems, and Multiple Choice Questions 33 Figure 2.10. System in Problem no. 12. a) Given that the diskette taken at random is not perfect, what is the probability that it is of brand A? b) If we repeat the random

experiment described above ten times, with replacement (so that the trials are independent), what is the probability of obtaining at least one imperfect diskette? Problem no. 11 Four calls were directed at random to a telephone central during a one-hour pe- riod. What is the probability that one call was received during each quarter of an hour of the hour considered? Problem no. 12 A system is made up of subsystems A and B placed in parallel. The subsystem A contains three components, A1, A2 and A3, placed in series, while the components B1 and B2 placed in series constitute the subsystem B (see Fig. 2.10). We suppose that the probability that the component A i functions at the end of one year is equal to 0.9, for i = 1, 2, 3, independently of the other components. We also suppose that the components B and B2 operate independently of each other. Moreover, we assume that the probability that the component Bk, k = 1,2, functions at the end of one year is equal to 0.95 if the subsystem A functions at this moment, and to 0.80 otherwise. Let us define the following events: F = the system functions at the end of one year and Fx = the subsystem X functions at the end of one year, for X = A, B. Calculate a) e [FB]; b) e [F]; c) e [FB] F]; d) e [F O F]. Problem no. 13 A person buying a certain make of car can choose one or several of the following options: A = an automatic transmission, C = air-conditioning, and M = a V6 engine. Based on the data collected so far, 90% of the customers choose at least one of the three options, 75% choose at least two and 45% choose the three options. Moreover,

34 2 Elementary Probabilities the three options taken individually are equally popular. Finally, all possible groups of two options are also equally popular. We define the event A = the customer chooses option A; similarly for C and M. a) How many different versions of the car in question can be bought if at least one of the three options is chosen? b) Are the events A, C and M incompatible? Justify your answer. c) Are the events A, C and M (globally) independent? Justify. d) Calculate the probability that a person buying this car chooses exactly one option. e) Calculate the probability that a person buying this car has chosen the three options, given that he/she has chosen at least one. Hint. Use a Venn diagram. Problem no. 14 An electricity distribution network is made up of three stations: A, B and C. We know, by experience, that during a heat wave the probability that an overload occurs at station A is equal to 0.4; this probability is equal to 0.3 for station B and to 0.2 for station C. Moreover, we assume that an overload at station A causes a breakdown of the entire network in 2% of the cases. An overload at station B brings about a

breakdown of the entire network 3% of the time, and an overload at station C causes a breakdown of the network 4% of the time. We define the events: D = a breakdown of the network occurred during a certain heat wave and A (respectively B, C) = an overload occurred at station A (resp. B, C). Remark. We assume that a network breakdown can only occur when there is an overload at one of the three stations, and that two overloads cannot happen exactly at the same moment (so that the events A, B and C are incompatible for a given breakdown). Calculate a) P[D];b)P[A] D];c) P[D] BtC]. d) Suppose that the breakdowns are independent of one another. Compute the proba- bility that two consecutive breakdowns have been caused by an overload at the same station. Problem no. 15 A respirator used during surgical operations breaks down with probability p (> 0). A breakdown of the respirator entails that the patient will die, unless a monitor detects the breakdown and warns the surgeon. In this case, there is an 80% chance that the surgeon will be able to save the patient. The monitor fails with probability 0.005, independently of the respirator. We define the events: F ----- a patient dies because of a breakdown of the respirator, R -= the respirator fails

and 2.6 Exercises, Problems, and Multiple Choice Questions 35 M = the monitor breaks down. a) The respirator is made up of three components placed in parallel that operate independently from one another. If the reliability of each component is equal to 0.9, what is the probability p that the respirator will fail? b) The monitor is constituted of two brand A and two brand B components. The four components are placed in series. We have 10 brand A and 20 brand B components to build monitors. How many different devices can we construct if we assume that the order of the four components placed in series matters and that the 10 brand A com- ponents and the 20 brand B components are not distinguishable among themselves? c) What is the answer in b) if we assume that the 30 components that we have are numbered from 1 to 30? d) Are the events R and M incompatible? Justify your answer. Answer the following questions assuming that p ----- 0.01. Calculate e) P [F]; f) P [M I F]; g) P [F I R t2 M]. h) We consider two patients that died because of respirator failures. What is the probability that the monitor worked in one case but failed in the other case? Remark. We assume that the monitor breakdowns are independent. Problem no. 16 We divide the length of telephone calls into three categories: those lasting less than one minute (I), those lasting between one and three minutes (II), and the ones lasting more than three minutes (III). Moreover,

we suppose that 60% of all calls are personal and the rest are business calls. Finally, we suppose that 10%, 20% and 70% of the personal calls are of category I, II and III, respectively. In the case of the business calls, these percentages are equal to 20%, 40% and 40%, respectively. a) Two calls are made during a certain period. If we record the total number of per- sonal calls and the number of personal calls of each category, how many elementary events (like 0000, for example) are there in the sample space S? b) We consider three independent business calls. What is the probability that there is one of each category? c) What is the probability that a given call lasts less than one minute? Problem no. 17 You are waiting for a friend who entered a shop. At the time of paying, he goes at random to one of the three cash registers: A, B or C. The probability that the service is slow at A is equal to 0.7; this probability is equal to 0.6 at B and to 0.5 at C. When he arrives, your friend tells you that the service has been slow at the cash register. a) Calculate the probability that your friend went to cash register C. b) Calculate the probability that, on two other independent visits to this shop, your friend did not go to cash register C, given that the service has been slow on each occasion.

36 2 Elementary Probabilities Problem no. 18 In a certain city, 60% of the calls to the emergency number require the help of the police department, 40% that of the ambulance service, and 20% ask for the fire department. Moreover, 40%, 20% and 5% of the calls require only the help of the police department, the ambulance service and the fire department, respectively. Finally, 5% of the calls require the help of the three services and 10% do not ask for any of these three services. a) Calculate the probability that a given call requires the help of the police department and of the ambulance service, but not of the fire department. Hint. Use a Venn diagram. b) Let the events be as follows: A = a certain call requires the help of the police department and B = a certain call requires the help of the ambulance service. Are the events A U B and A"U B'" incompatible (or mutually exclusive)? Justify your answer by a numerical computation. c) We consider ten calls to the emergency number. Among these calls, there are ex- actly seven that required the help of the police department. If four calls are taken at random (and without replacement) among the ten calls considered, what is the prob- ability that exactly two of these four calls required the help of the police department? Problem no. 19 Computers A and B exchange information within a network. The probability that the information

sent from A to B is incorrectly transmitted is equal to 0.01, while this probability is equal to 0.005 in the case when the information is sent from B to A. We consider 10 messages transmitted from A to B and 15 messages from B to A. We assume that all transmissions are independent. a) What is the probability that a message taken at random among the 25 messages considered has been incorrectly transmitted? b) Two messages are taken at random and without replacement among the 25 mes- sages considered. Given that both messages have been incorrectly transmitted, what is the probability that they were transmitted by the same computer? c) Suppose that exactly two of the 25 messages considered have been incorrectly transmitted. What is the probability that the two erroneous messages were transmit- ted by the same computer? Problem no. 20 An engineer responsible for the control of the quality in the company where she works receives a batch of 200 parts used in the computers that they build. She decides to pick 10, at random and without replacement, and to test them. Let E be the random experiment that consists in counting the number of defective parts among

2.6 Exercises, Problems, and Multiple Choice Questions 37 the 10 parts tested. Answer the following questions by assuming that there are in fact two defective parts among the 200 parts received. a) Write the sample space S. b) Calculate the probability of all the elementary events. c) We define the event A_k = there are exactly k defective parts among the 10 parts tested, for k = 0, 1 ..... Calculate $P[A_0 \cup (A^C \vee A_2)]$. Problem no. 21 A small company has two telephone lines: a local number (line 1) and an 800 number for people outside the city (line 2). We suppose that the probability that line 1 is busy during working hours is equal to 0.1; this probability is equal to 0.05 in the case of line 2. Moreover, we suppose that the probability that a customer who calls for the first time (and finds the line busy) is lost is equal to 0.3 when the call is local and to 0.4 when the call comes from outside the city. Finally, we estimate that 60% of the calls are local. a) Calculate the probability that the next customer who calls the company for the first time will be lost. b) Given that a customer has been lost after having called only once, what is the probability that this customer has called the 800 number? c) Among the next 10 customers who will call for the first time to the local number, what is the probability that at least two of them will be lost? Remark. We assume that the customers are independent of one another. Problem no. 22 a) We consider four viruses (V1, V2, V3 and V4) that can infect the computers of the computer network of a

certain institution. Let Fi = virus Vi has infected the network, for i = 1,2, 3, 4. We assume that the events F1 and F2 are independent, as well as the events F3 and F4. Moreover, the events F1 and F3 are incompatible, and so are the events FI and F4, F2 and F3, and F2 and F4. Finally, we suppose that P[Fi] = P[F2] = 0.05 and that P[F3] = P[F4] = 0.01. i) Calculate the probability that the network will be infected by any of the viruses. ii) Can we say that the event FI is included in the event F27 Justify your answer. iii) Are the four intersections of three distinct events equally likely7 Justify your answer. b) We consider five other viruses that can infect the computers of another computer network. How many conditions must we check (at most) to determine whether the five viruses are (globally) independent? Justify your answer. Problem no. 23 A sample of five empty CDs is taken at random from a batch of 100 CDs. We consider the random experiment E that consists in counting the number of CDs that conform to the norms. Assume that the batch of 100 CDs contains in fact exactly four defective CDs.

38 2 Elementary Probabilities a) Write the sample space for this random experiment if i) the CDs are taken one at a time and with replacement; ii) the five CDs are taken all at once. b) Calculate the probability that the sample of five CDs contains at least one defective in cases i) and ii) above. c) Let the events be as follows: A = there is at least one CD that conforms to the norms among the five taken at random and B = there is at least one defective CD among the five taken at random. In case i) above, are the events A' and B' incompatible? Are they independent? Justify your answers with numerical calculations. Problem no. 24 A system is made up of three components placed in parallel that operate simul- taneously. The probability that component no. 1 functions is equal to 0.9. In the case of component no. 2, this probability is equal to 0.95 if component no. 1 functions, and to 0.8 otherwise. Finally, the probability that component no. 3 functions is equal to (I.99 if components nos. 1 and 2 function, to (I.75 if components nos. 1 and 2 do not function, and to (I.8 otherwise. a) Calculate the probability that the system functions. b) Given that the system functions, what is the probability that component no. 2 functions? c) Calculate the probability that component no. 3 functions. Multiple Choice Questions Question no. 1 A box contains 100 objects, namely 40 brand A and 60 brand B objects. We take 50 objects at random and without replacement and we count the number of brand A objects obtained. Write the sample space corresponding to this random experiment. a){0 ..... 40} b){0 ..... 50} c){10 .....

40} d){10 ..... 50} e) none of these answers Question no. 2 A certain store has 5 brand A, 10 brand B and 20 brand C items in stock. How many different orders of these three brands can it satisfy? a) 35 b) 999 c) 1000 d) 1385 e) 6545

2.6 Exercises, Problems, and Multiple Choice Questions 39 Question no. 3 We have 20 components, of which two are defective. Three components are taken at random to construct a machine made up first of two components placed in parallel, and then a component placed in series. We suppose that all the components operate at the same time, and independently from one another. Calculate the probability that the machine functions. a) 0.8895 b) 0.8947 c) 0.9 d) 0.9942 e) 1 Question no. 4 Let A and B be independent events such that $P[A] = 1/4$ and $P[B] = 1/2$. More- over, let C be an event such that $P[CI Ac f)$ B] = 1/2 and $P[CIA c f)$ B c] = 1. Finally, A and C are incompatible. Calculate $P[A c f)$ B c IC]. a) 0 b) 1/3 c) 1/2 d) 2/3 e) 1 Question no. 5 A class is made up of five female students and 45 male students. Among the five girls, four are second-year students, while 30 out of the 45 boys are second- year students. Two students are taken at random, and without replacement, among the 50 students in this class. Knowing that in both cases the person chosen was a second-year student, what is the probability that a male and a female students were chosen? a) 8/289 b) 30/289 c) 60/289 d) 229/289 e) 259/289 Question no, tl Let E be the following random experiment: two dice are tossed; if the two num- bers obtained are equal, then the two dice are tossed again (only once). How many elementary events are there in the sample space S? a) 72 b) 216 c) 246 d) 252 e) 1296 Question no. 7 In a class, there are 20 students. What is the probability that exactly two of them share the same birthday? a) 0.3232 b) 0.4114 c) 0.5886 d) 0.6768 e) none of these answers Question no. 8 A number x is taken at random in the interval [0, 3] so that the probability that the number chosen is in the interval [k, k 1] is equal to (k 1)p, for k = 0, 1,2. Calculate P[A], where A := {x [0, 3] :Ix - -I > 1}. a) 1/12 b) 1/4 c) 1/3 d) 1/2 e) 2/3 Question no, 9 In a certain factory, 80% of the parts fabricated conform to the norms. Every part fabricated in this factory is subjected to three independent quality control operations. We suppose that each of these operations classifies as non-defective 95% of the parts that effectively conform to the norms, and 10% of the parts that are in fact defective. Calculate the probability that a part that has been sold conforms to the norms. a) 0.80 b) 0.9250 c) 0.95 d) 0.9744 e) 0.9997

40 2 Elementary Probabilities Figure 2.11. System in Multiple choice question no. 10. Question no. 10 We consider a system constituted of five identical components that operate inde- pendently from one another and that are connected as in the diagram in Fig. 2.11. Knowing that exactly three of the five components work, what is the probability that the system functions? a) 0.6 b) 0.7 c) 0.8 d) 0.9 e) we cannot compute it Question no. 11 The license plates in a certain country bear six characters taken at random among the 26 letters of the alphabet and the ten digits $\{0, 1 ..... 9\}$. What is the probability that a given license plate bears at least one digit? a) 0.0546 b) 0.1419 c) 0.3275 d) 0.8581 e) 0.9995 Question no. 12 A student gets up at time X and goes to bed at time Y, where $0 < X < Y < 24$. Let $S = \{(x, y) : 0 < x < y\_ < 24\}$. Write in mathematical form the event A = the student is up at least three hours more than he lies in bed. a)$\{(x,y)S:[y-x[>3\}$ b) $\{(x, y) S : [y - x[< 13.5\}$ c)$\{(x,y)6S:y-x > 13.5\}$ d)$\{(x,y)6S:y-x < 15\}$ e)$\{(x,y)6S:y-x > 15\}$ Question no. 13 Let A, B and C be events such that A C B, A and C are incompatible, $P[(B \cup C) c] = 1/10$, $P[B C] = 3/10$, $P[A c B] = 1/2$ and $P[C] = 13/20$. Calculate $P[A]$. a) 0 b) 0.05 c) 0.35 d) 0.5 e) 0.9 Question no. 14 Two men and two women sit at a circular table. In how many different ways can they be seated if the two men must be opposite each other? Remark. Two ways are equivalent if each person has the same neighbors to his/her left and to his/her right. a) 2 b) 3 c) 4 d) 6 e) 24

2.6 Exercises, Problems, and Multiple Choice Questions 41 Question no. 15 How many distinct license plates made up of three letters and three digits can there be if the three letters appear either at the beginning or at the end of the plate? a) 3F'l�F'263 b) 2(C0C326) c) 103263 d) 2(103263) e) C36 103263 Question no. 16 Let A, B and C be events such that $P[A] = P[B] = 1/2$, and A and B are independent. Calculate $P[A$ a)0 b) 1/18 c) 1/6 d) 1/3 e) 1/2 Question no. 17 The events A and B form a partition of the sample space S. Calculate $P[AI B] + P[AIBC]$. a) 0 b) 1/4 c) 1/2 d) 1 e) none of these answers Question no. 18 In a certain computer store, a study revealed that the three most popular options when people buy a new computer are the following: A = a faster processor, B = a larger hard disk and C = a wider screen. According to the data collected so far, 40% of the customers have chosen at least one of these three options, and 10% chose only option C. Moreover, all the customers who chose option A also chose option B, and vice versa. Finally, options A and B taken together have been twice as popular as option C. In the following

questions, only the options A, B and C are considered. A) How many different computers have been sold up to now by this computer store? a) 2 b) 3 c) 4 d) 6 e) 8 B) How many different computers can be sold to future customers? a) 4 b) 5 c) 6 d) 7 e) 8 C) What proportion of the customers have chosen the three options so far? a) 0 b) 0.05 c) 0.1 d) 0.15 e) 0.2 D) Up to now, 30% of the customers have been women. Moreover, 80% of these women have chosen none of the options considered. What is the probability that a man taken at random has chosen at least one of the three options? a) 2/5 b) 17/35 c) 18/35 d) 3/5 e) none of these answers E) Under the same hypotheses as in D), what is the probability that a customer that has chosen none of the three options is a woman? a) 0.2 b) 0.3 c) 0.4 d) 0.6 e) 0.8 F) We consider five persons who each bought a computer. Assuming that the cus- tomers are independent, what is the probability that at least two of them chose none of the three options? a) 0.0870 b) 0.2304 c) 0.7696 d) 0.9130 e) 0.9898

42 2 Elementary Probabilities Question no. 19 A transistor is taken at random and its lifetime is measured. The sample space for this random experiment is given by S = [0, cx). We consider the events A = [0, 1], B = [0, 2] and C = [1, cx). Give the intervals that correspond to the following events: A) D = A U (B O C); a) [0, I1 b) [0, 21 c) [I, 21 d) [2, cx) e) [I, cx) B) F =[AO(B cUC')I c. a) [0,1] b)[0,2] c) 11,2] d) 12, c) e) ll,c) Question no. 20 Two dice are thrown simultaneously. One of the dice has four red faces and two white faces, whereas the other one has two red faces and lbur white faces. Let R = at least one red face is obtained, and B = at least one white face is obtained. Calculate, assuming that all faces have an equal probability of coming up (namely, one chance in six) A) PIR]; a) I/3 b) 4/9 c) 5/9 d) 7/9 e) 8/9 B) PIR B]. a) I/3 b) 4/9 c) 5/9 d) 2/3 e) 7/9 Question no. 21 The combination of a padlock is made up of three digits (taken in the set {0, I ..... 9}). How many possible combinations are there if A) each digit cannot be chosen more than once? a) 360 b) 450 c) 495 d) 720 e) 990 B) each digit cannot be chosen more than twice? a) 495 b) 720 c) 900 d) 950 e) 990 Question no. 22 An engineer subscribes to two independent electronic mail services. The prob- ability that service no. I does not function on a given day is equal to 1/20, while the probability of a breakdown of service no. 2 is equal to only 1/100. Moreover, when service no. 1 functions, the probability that a message sent is received by its addressee is equal to 0.995. This probability is equal to 0.99 in the case of service no. 2. To be safer, the engineer wants to send an important message

through both services. A) What is the probability that the addressee receives this message? a) 0.94525 b) 0.9627 c) 0.9801 d) 0.9989 e) 0.9999 B) Given that the addressee has received the message, what is the probability that only the message that the engineer wants to send through service no. 1 arrived at its destination? a) 0.01883 b) 0.03627 c) 0.04569 d) 0.05372 e) 0.07255

2.6 Exercises, Problems, and Multiple Choice Questions 43 Question no. 23 A factory recycles defective car parts. Among the last 50 parts that it recycled, five are still defective. A company buys 20 of these 50 parts. A) What is the probability that the company receives at least one defective part? a) 0.1216 b) 0.2587 c) 0.2702 d) 0.8784 e) 0.9327 B) Given that the company has received at least one defective part, what is the prob- ability that it received exactly two? a) 0.3246 b) 0.3641 c) 0.3903 d) 0.6097 e) 0.6754 Question no. 24 Ten candidates are interviewed to fill two posts in a company. In how many ways can the company fill these posts, if A) both posts are identical? a) 45 b) 50 c) 55 d) 90 e) 100 B) one post is permanent and the other temporary? a) 50 b) 90 c) 100 d) ll0 e) 200 Question no. 25 A) A system is made up of three subsystems placed in series. Each subsystem com- prises two components placed in parallel (see system A in Fig. 2.12, p. 43). We sup- pose that all the components operate independently from one another and all have a 90% probability of working at a given time. Calculate the reliability of the system at that moment. a) 0.2710 b) 0.7290 c) 0.9266 d) 0.9703 e) 0.9991 B) Under the same hypotheses as in A), what is the reliability of a system made up of two subsystems placed in parallel, if each subsystem comprises three components placed in series (see system B in Fig. 2.12)? a) 0.9266 b) 0.9703 c) 0.9797 d) 0.9980 e) 0.9991 Question no. 26 Among the 100 passengers on board a plane landing at an airport, 80 arrive at their destination, whereas the others are in transit. Moreover, among those arriving at destination, 70% started their journey inside the country. This percentage is equal to 40% in the case of the passengers in transit. Finally, we assume that the probability that a given passenger travels for business is equal to 10%, independently of what A B Figure 2.12. Systems for Multiple choice question no. 25.

44 2 Elementary Probabilities precedes. A passenger traveling on board this plane is taken at random. What is the probability that the passenger A) arrives at destination or started his/her journey outside the country? a) 0.12 b) 0.24 c) 0.375 d) 0.56 e) 0.875 B) travels for business or started his/her journey inside

the country? a) 0.064 b) 0.424 c) 0.676 d) 0.74 e) 0.964 Question no. 27 Customers arrive at a store, at random and independently from one another, from 9:00 a.m. A) Suppose that there have been two customers between 9:00 a.m. and 10:00 a.m. i) Calculate the probability that the second customer arrived before 9:30, given that the first one arrived at 9:15. a) 1/4 b) 1/3 c) 7/15 d) 29/60 e) 1/2 ii) What is the probability that the two customers arrived inside a 15-minute interval (that can be located anywhere inside the hour considered)? a) 1/4 b) 1/3 c) 7/16 d) 1/2 e) 9/16 B) Suppose now that there have been four customers between 10:00 a.m. and 11:00 i) What is the probability that exactly three of these four customers arrived be- tween 10:00 and 10:30'? a) 1/16 b) 1/8 c) 1/4 d) 1/2 e) 3/4 ii) What is the probability that there has been one customer in each quarter of an hour of the hour considered? a) 0.00391 b) 0.00586 c) 0.04167 d) 0.06250 e) 0.09375 Question no. 28 A maker of lasers estimates that 20% of the lasers the company produces are of excellent quality, 75% are of acceptable quality, and 5% are defective. Moreover, we assume that the quality of a given laser does not depend on the other lasers. To improve the reliability of the lasers the company sells (those they think are non- defective), they submit their products to a quality control procedure. Suppose that this procedure classifies as excellent acceptable excellent acceptable acceptable defective 90% of the lasers that are of excellent quality, 10% of the lasers that are of excellent quality, 10% of the lasers that are of acceptable quality, 85% of the lasers that are of acceptable quality, 5% of the lasers that are defective, 95% of the lasers that are defective.

2.6 Exercises, Problems, and Multiple Choice Questions 45 A) Answer the following questions based on the maker's estimations. i) What is the probability that a batch of ten lasers contains at least one laser of excellent quality or at least one defective laser? a) 0.9137 b) 0.9237 c) 0.9337 d) 0.9437 e) 0.9537 ii) What is the probability that a batch of ten lasers contains at least one laser of excellent quality and at least one defective laser? a) 0.3502 b) 0.3602 c) 0.3702 d) 0.3802 e) 0.3902 B) i) Calculate the probability that a given laser has been incorrectly classified by the procedure. a) 0.135 b) 0.15 c) 0.235 d) 0.25 e) 0.3 ii) Calculate the probability that a laser sold is defective. a) 0 b) 0.0017 c) 0.0027 d) 0.0037 e) 0.0047 Question no. 29 We consider the 107 telephone numbers made up of seven digits. For security reasons, the numbers starting with 0 or with 911, or ending with 911,

are not used. We define the events A = a number starts with 0, B = a number starts with 911, and C = a number ends with 911. A) Which pairs of events are incompatible? a) A and B only b) A and C only c) B and C only d) A and B, and B and C only e) the three pairs of events B) Calculate the number of telephone numbers that are not used. That is, calculate the number of elementary events in D := A ∪ B ∪ C. a) 810,190 b) 998,190 c) 1,018,990 d) 1,431,190 e) 2,100,010 C) Let F := A (B C). How many elementary events are there in F? Suggestion. Use a Venn diagram. a) 1,000,000 b) 1,000,010 c) 1,010,000 d) 1,100,000 e) 1,110,000 Question no. 30 In an interview for a job as a computer analyst, first at most five questions are asked to every candidate. As soon as the candidate has answered four questions cor- rectly, he (or she) is hired. If the candidate does not succeed in this first stage of the interview, he is asked a last question. In this case, he is hired if and only if he answers this question correctly. Suppose that each time the candidate answers a question, he has a probability of 1/2 of responding correctly, independently from a question to another. A) What is the probability that the candidate is hired in the first stage of the inter- view? a) 1/16 b) 3/32 c) 1/8 d) 5/32 e) 3/16

46 2 Elementary Probabilities B) Let p be the answer in A). What is the probability that the candidate is hired? a) 21 P2 b) � +- c) �+p d) l-p e) l P2 C) Suppose now that, in fact, the candidate has a probability of 1/2 of responding correctly to the first question, but that his self-confidence increases after a right an- swer, so that the probability of responding correctly after a right answer is equal to 3/4. Conversely, the probability that the candidate responds correctly after a wrong answer is equal to only 1/4. What is then the probability that the candidate responds correctly to exactly one of the first two questions? a) 1/4 b) 3/8 c) 1/2 d) 5/8 e) 3/4 Question no. 31 A box contains ten components, of which two are defective. The components are taken one at a time, at random and without replacement, and are tested until the two defective components have been identified. A) What is the probability that the two defective components are identified as early as after the second test'? a) 1/100 b) 1/95 c) 1/90 d) 1/50 e) 1/45 B) What is the probability that at least one defective component has been identified after three tests (if necessary)? a) 0.4667 b) 0.4880 c) 0.5120 d) 0.5333 e) 0.5600 C) Calculate the probability that more than eight components have to be tested in order to identify the two defective ones. a) 4/15 b) 14/45 c) 16/45 d) 6/15 e) 4/9 Question no. 32 Let A, B and C be three events. Write in mathematical form the event F = exactly one

of the events A, B, C does not occur. a) (A N B) U (B N C) U (A N C) b) (A N B N C") U (A N B' N C) U (A''' N B N C) c) (A c N B c N C) U (A c N B N C") U (A N B' N C c) d) A c U B c U C''' e) (A U B U C) c Question no. 33 Let A and B be two events such that P[A] -- 1/2 and P[B] = 3/10. Which of the following statements can, under certain conditions, be true? i) P[A U B] = 4/5; ii) P[A U B]-- 13/20; iii) P[A U B] -- 2/5. a) i) only b) i) and ii) only c) i) and iii) only d) ii) and iii) only e) all Question no. 34 Each of the tools O1 ..... 05 has a particular pocket in a certain tool belt. After use, you must replace each tool in its proper pocket. Because of lack of time, you put at random (equiprobability) a tool in each pocket.

2.6 Exercises, Problems, and Multiple Choice Questions 47 A) What is the probability that the second tool that you put into the tool belt is actually in its proper pocket? a) 1/20 b) 1/12 c) 1/10 d) 1/5 e) 1/4 B) Calculate the probability that each of the tools O1, O2 and 03 are put into the right pockets? a) 1/125 b) 1/120 c) 1/60 d) 1/40 e) 1/20 Question no. 35 A company buys electrical components in batches of ten components. Upon re- ception of each batch, two components are taken at random and without replacement and are then tested. The company accepts the batch if and only if the two components tested are non-defective. Based on previous data, we estimate that the probability that a batch of ten components contains no defectives is equal to 7/10, the probability that it contains exactly one defective is 1/5, and the probability that it contains exactly two defectives is 1/10. We then find that the probability that a batch is accepted is approximately equal to 0.9222. Use, if needed, this result to calculate the probability that A) a batch contains no defectives and is accepted; a) 0.3 b) 0.6222 c) 0.7 d) 0.7222 e) 0.9222 B) a batch contains exactly two defectives or is accepted; a) 0.92 b) 0.93 c) 0.94 d) 0.95 e) 0.96 C) a batch contains exactly one defective, given that it has been rejected; a) 0.5141 b) 0.5541 c) 0.6141 d) 0.6541 e) 0.7141 D) three consecutive (independent) batches are rejected. a) 0.00047 b) 0.0014 c) 0.0778 d) 0.2157 e) 0.2333 Question no. 36 Let A, B and C be events such that P[A] = P[B] = P[C] = 1/4, A C B, and A and C are incompatible. A) Calculate P[A tA B tA C]. a) 1/4 b) 3/8 c) 1/2 d) 5/8 e) 3/4 B) Calculate P[A c N (B tA C)]. a) 1/4 b) 3/8 c) 1/2 d) 5/8 e) 3/4 Question no. 37 In physics, many systems, for example Einstein's and Debye's models for simple solids, are made up of subsystems that are independent of one another. Consider a system of this type, made up of three particles. Each particle, independently from the

other two particles, can have energy level l, 2 or 3. The energy level of the system is defined by the vector (nl, n2, n3), where ni is the number of particles having energy level i, for i : 1, 2, 3. A) How many distinct energy levels can this system have? a) 6 b) 7 c) 8 d) 9 e) 10

48 2 Elementary Probabilities B) In A), how many distinct arrangements of the energy levels of the particles corre- spond to the energy level (2, 1,0) of the system or to level (1, 1, 1) ? a) 9 b) 10 c) 12 d) 14 e) 15 C) i) Let Pi be the probability that a particle has energy level el, for i = 1,2, 3. Suppose that P1 = 1/4, P2 = 1/2 and p = 1/4. What is the probability that the three particles all have a different energy level? a) 1/16 b) 1/8 c) 3/16 d) 1/4 e) 3/8 ii) Given that the three particles have the same energy level, what is the probabil- ity that this energy level is e or e2? a) 2/3 b) 7/10 c) 4/5 d) 5/6 e) 9/10 Question no. 38 A) A number is taken at random in the set {1,2 ..... 8}, so that P[{k}] = 1/8 for k = 1,2 ..... 8. We define the events A={1,2,3,4}, B={2,3,4,5} and C= {4,5,6,7}. Which pairs of events are independent? a) (A, B) only b) (B, C) only c) (A, B) and (B, C) only d) no pairs e) all the pairs B) Suppose that in A), we have: Pl{k}] = c if k : 1,3, 5,7 and P[{k}] : c2 if k : 2, 4, 6, 8. For what values of c and c2 are the events D : { 1,2} and F : {2, 3} independent? a) c =C2 -,- 1/16 b) c =c2=3/16 C) Cl = 1/16;c2=3/16 d) Q =3/16;c2= 1/16 e) no values ofc andc2 Question no. 39 The probabilities of the events A, B and C are P[A] = 1/2, P[B] = 1/3 and P[C] = 1/4. Moreover, A and C are incompatible, B and C are incompatible, and P[A (3 B} = PIAC (3 B]. A) Calculate P[A c N B"]. a) 1/12 b) 1/6 c) 1/4 d) 1/3 e) 1/2 B) Calculate P[AI B U C"]. a) 1/4 b) 1/3 c) 1/2 d) 7/12 e) 2/3 Question no. 40 A box contains five brand A components, five components of brand B and five of brand C. Five components are taken at random and without replacement. A) What is the probability that the five components taken at random are of the same brand? 1 a) l b) c) 1 d) e)

2.6 Exercises, Problems, and Multiple Choice Questions 49 B) What is the probability that the five components are of the same brand, given that at least four of the five components are of the same brand? 1 a) 1- b) 1 c) d)l e) Question no. 41 We have three boxes containing only brand A and brand B components. The first box comprises 10 brand A and 10 brand B components, the second box contains 10 brand A and 20 brand B components, and the third one has 20 brand A and 10 brand B components. Let E be the random

experiment that consists in taking first, independently, a component at random in each of the first two boxes (starting with the first box). If the two components are of the same brand, then the random experiment is over; otherwise a component is taken at random from the third box. The brand of each component taken at random in this random experiment is noted. We consider the events Ak = a brand A component is obtained on the kth draw, for k = 1, 2, 3, Bk = a brand B component is obtained on the kth draw, for k = 1, 2, 3, F = three components have to be taken at random to complete the random experi- ment. A) How many elementary events are there in the sample space S associated with E if we take the order into account? a) 5 b) 6 c) 7 d) 8 e) 9 B) Which groups of events, among the following, are incompatible? I) A1, B2 II) A1, B1 III) A1, A2, A3 IV) A1, A2, B3 V) A1, B2, B3 a) II, III and IV only b) I and IV only c) II and III only d) I, IV and V only e) II only C) Calculate P[B3]. a) 1/18 b) 1/9 c) 1/6 d) 1/3 e) 1/2 D) Calculate P[A1 ((A2 U A3)IF]. a) 1/9 b) 2/9 c) 1/3 d) 4/9 e) 5/9 Question no. 42 A point is taken at random in the interval [0, 1], so that the probability of the event F1 = the point chosen is in the interval I = [a, b] is given by P[FI] = b - a for any I C [0, 1]. We consider the events FA, FB, FC and FD, where A=[0,�], B=[�,], C=[,] and D=[0,1]. A) Which are the only pairs of (distinct) events that are incompatible (or mutually exclusive)? a) no pairs b) (FA, FB) c) (FA, Fc) d) (FA, F) and (FA, Fc) e) (FA, F), (FA, Fo), (F, Fo) and (Fc, Fo)

50 2 Elementary Probabilities B) Which are the only pairs of (distinct) events that are independent? a) no pairs b) (FA, Fn) c) (FA, Fc) d) (FA, Fn) and (FA, Fc) e) (FA, Fn), (FA, FD), (Fn, FD) and (Fc, Fo) Question no. 43 A particle is at the origin at the initial time and is then moving on the positive integers as follows: at each time unit, a coin for which the probability of getting "tails" is equal to 1/3 (and P[{Heads}] = 2/3) is thrown (independently). If"tails" is obtained, then the particle moves one integer to the right, whereas if "heads" is obtained, the particle moves two integers to the right. A) What is the probability that the particle does not visit the point 3? a) 2/9 b) 8/27 c) 10/27 d) 4/9 e) 14/27 B) In how many different ways can the particle go from 0 to 10 without visiting the point 2? a) 21 b) 22 c) 23 d) 24 e) 25 Question no. 44 Two defective fuses have been mixed by mistake with three good fuses. The fuses are taken one at a time, at random and without replacement, and tested until the two defective fuses can be identified. A) What is the probability that the first two fuses tested are two defective or two good fuses? a) 0.25 b) 0.3 c) 0.35 d)

0.4 e) 0.5 B) What is the probability that the last fuse that needs to be tested to identify the two defective fuses is a good one? a) 0.25 b) 0.3 c) 0.35 d) 0.4 e) 0.5 Question no. 45 In a certain course, 45% of the students are in electrical engineering, 40% are in computer engineering, and 15% are in physics engineering. Furthermore, 60% of the electrical engineering students registered in this course are fourth-term students. In the case of the computer engineering students this percentage is equal to 70%, while it is equal to 90% in the case of the physics engineering students. A) Two students are taken, at random and with replacement, among the students in the course considered. What is the probability that they are both fourth-term stu- dents? a) 0.4492 b) 0.4692 c) 0.4892 d) 0.5092 e) 0.5292 B) Let PA be the answer in A). Knowing that the two students taken at random in A) are fourth-term students, what is the probability that they are both in electrical engineering? 0.0729 0.27 0.5625 a) -- b) 0.0729 PA C) -- d) 0.5625 PA e) -- PA PA PA

2.6 Exercises, Problems, and Multiple Choice Questions 51 Question no. 46 The students in a certain department of a university must choose exactly two courses among three optional courses: A, B and C. We find that the number of students registered for course A is 150% greater than the number of students reg- istered for course C. Moreover, the number of students registered for course B is 50% greater than the number of students registered for course C. A student in this department is taken at random. Let FA = the student taken at random has chosen course A, etc. A) Calculate the probability P [(FA N FB) U (FA FC) t.) (FB FC)]. a) 0.6 b) 0.7 c) 0.8 d) 0.9 e) l B) What is the probability that the student has chosen courses A and B? a) 0.4 b) 0.5 c) 0.6 d) 0.7 e) 0.8 Question no. 47 The probability that a certain person spends an amount of time (in minutes) com- prised in the interval [a, b] surfing the Internet, on a weekday, is given by b-a for10<a <b<30. 20 We assume that the days are independent. We consider a time period of five week- days, namely from Monday to Friday. A) What is the probability that the shortest period of time spent surfing the Internet, during that week, was that on Monday, and the longest period that on Friday? a) 1/120 b) 1/20 c) 1/5 d) 1/4 e) 2/5 B) What is the probability that the length of the surfing sessions has increased each day? a) 1/720 b) 1/120 c) 1/60 d) 1/40 e) 1/20 Question no. 48 We consider the system shown in Fig. 2.13. Each component functions with prob- ability 1/2, independently from the other three. O - Figure 2.13. System for Multiple choice question no. 48.

A) What is the probability that the system functions? a) 7/16 b) 1/2 c) 9/16 d) 5/8 e) 11/16 B) Let p be the answer in A). Given that the system functions, what is the probability that component no. 1 functions? a) b) Question no. 49 A hydroelectric station is subject to receive three types of shocks. The probability of a shock of type I is equal to 30%, that of a shock of type II is equal to 50%, and that of a shock of type III is equal to 20%. Moreover, the various types of shocks are incompatible. The probability that a shock of type I causes a power failure is equal to 40%. In the case of type II and type III shocks, this probability is equal to 20% and to 90%, respectively. A) What is the probability that the station receives a type II shock before a type III shock, from a given time instant? a) 3/7 b) 1/2 c) 4/7 d) 5/7 e) 6/7 B) Given that a power failure (caused by a shock) has occurred, what is the proba- bility that a type II shock has caused this failure? a) 1/7 b) 1/6 c) 1/5 d) 1/4 e) 1/3 Question no. 50 Manufactured objects are classified into three categories: compliant with the norms (I), containing one or more minor defects only (II), or containing one or more major defects (III). A) Four objects will be taken, at random and with replacement, among those that will be manufactured tomorrow. In how many different ways will it be possible to classify them if i) the order in which the objects will be taken matters? a) 64 b) 81 c) 100 d) 121 e) 196 ii) the order in which the objects will be taken does not matter? a) 12 b) 13 c) 14 d) 15 e) 16 B) Suppose that in a batch of 100 objects, there are 70 objects that comply with the norms and 23 that have one or more minor defects only. i) A sample of three objects is taken, at random and without replacement. What is the probability that the sample contains an object of each category? a) 0.0697 b) 0.0797 c) 0,0897 d) 0,0997 e) 0.1097 ii) Four objects are taken, at random and without replacement. What is the prob- ability that at least one object of each category is obtained? a) 0.0952 b) 0.1052 c) 0.1152 d) 0.1252 e) 0.1352

2.6 Exercises, Problems, and Multiple Choice Questions 53 Question no. 51 Two tests based on the DNA of an individual have been developed in order to identify the culprits in criminal cases. The tests work by comparing two samples: one taken on the victim and the other on the suspect. In the case of test A, the proba- bility of concluding that the suspect is the person wanted, given that he is effectively the culprit, is equal to 99.5%. Moreover, the probability of arriving at the same con- clusion, given that he is not the person wanted, is equal to 1.5%. In the case of test B, these probabilities are equal to 99.7% and

to 2%, respectively. Finally, we assume that the two tests are conditionally independent with respect to the event C = the suspect is the culprit, and with respect to C c. A) A suspect in a particular case has been arrested. Suppose that the probability that he is the culprit is equal to 80%. i) Given that test A clears the suspect, what is the probability that he is in fact guilty? a) 0.0179 b) 0.0189 c) 0.0199 d) 0.0209 e) 0.0219 ii) Given that the two tests concluded that the suspect is guilty, what is the prob- ability that he is really guilty? a) 0.999914 b) 0.999924 c) 0.999934 d) 0.999944 e) 0.999954 iii) If the suspect is the criminal wanted, what is the probability that one test concludes that he is guilty and the other that he is innocent? a) 0.00497 b) 0.00597 c) 0.00697 d) 0.00797 e) 0.00897 B) Suppose that ten suspects have been arrested and that one of them is the culprit. What is the probability that only the culprit is found guilty by test A? Remark. We suppose that the tests are independent from one person to another. a) 0.4685 b) 0.5685 c) 0.6685 d) 0.7685 e) 0.8685

3 Random Variables 3.1 Introduction Definition 3.1.1. Let S be a sample space associated with a random experiment E. A function X that associates a real number $X(s) = x$ with each outcome s S is called a random variable. The set of all possible values of X will be denoted by Sx (see Fig. 3.1). Example 3.1.1 i) Let Ei be the following random experiment: a transistor is taken at random in a box containing 10 brand A and 20 brand B transistors, and its brand is noted. Then, S = {A, B}. Let X be the random variable that takes the value 1 if the transistor is of brand A and the value 0 otherwise. We have: Sx = {0, 1 }. Here, the random variable X is in fact the indicator function of the event F = a brand A transistor is picked. ii) Let Eli be the random experiment that consists in taking a number at random in the interval [0, 10). Let X be the random variable that takes the value of the number obtained and Y be the random variable that is equal to the integer part of the number obtained. We have: S= [0,10), Sx=-S and Sy ={0,1 ..... 9}. In this case, the random variable X is the identity function. Note that if the number is taken at random in the interval [0, 10], then we can write that S x(s) = x x Figure 3.1. Notion of random variable.

56 3 Random Variables Sy = {0, 1 ..... 10}. However, the probability that Y takes on the value 10 is equal to zero, because the number taken at random in the interval [0, 10] should be the number 10 (see Defini- tion 3.1.2 below), and S = [0, 10] is a continuous sample space. Remark. Since a random variable is a real-valued function defined on S, the set Sx cannot have more

elements than S. Therefore, if S is a discrete sample space, then Sx is discrete. However, if S is continuous, then Sx can be either discrete or continuous (as in Example 3.1.1 ii). Note that if Z is the random variable that is equal to twice the number taken at random in the interval [0, 10) in the preceding example, that is, if we define Z = 2X, then we have: Sz = [0, 20). However, S (= Sx) and Sz are both non-denumerably infinite. Definition 3.1.2. An event with respect to Sx is a subset of Sx. Moreover, let A C S and B C Sx. We say that A and B are equivalent events if A = {s � S: X (s) � B}. (3.1) Then, we write that P[BI= P[A]. Remark. We could denote the probability P[B] of the event B by Px[B], or PIX � BI. Example 3.1.2 A box contains two brand MI, one brand M2 and one brand M3 components. Two components are taken at random and without replacement and the brand of the components is noted. Then, if' the order in which the components are chosen does not matter, we can write that S = {MIMI, MM2, MM3, M2M3}. Let X be the number of brand M components obtained. We have: Sx = {0, 1,2}. Now, let B = { 1 }. The event equivalent to B is A ---- {M M2, MM3}. It follows (by symmetry) that P[X � {11]--= P[X = 1] = P[B] = P[A] = 2 x (2)(2/4)(1/3)=2/3. Remarks. i) If we identify the two brand M1 components as M and MI2, then the sample space S can be rewritten as follows:

3.2 The Distribution Function 57 S = {MllMI2, MllM2, MlzM2, MllM3, MlzM3, M2M3}. In this case, we obtain: A = {MllM2, MlzM2, MllM3, MlzM3} and, given that the elementary events are now equiprobable, we find directly that P[X = 1] = P[B] = P[A] = 4/6 = 2/3. ii) We can also define the events F = exactly one brand M1 component is obtained, and F1 = a brand M1 component is obtained on the first draw. Then, by the total probability rule, we can write that 2 2 2 2 2 P[F] = P[F I F1]P[F1] + P[F I F']P[F'] = � + . = -, or 2 2 P[F]-- (1)(1) _ 2x2_ 2 () 6 3" 3.2 The Distribution Function Definition 3.2.1. The distribution function of the random variable X is defined,for any real number x, by Fx(x) = P[X < x]. (3.2) Remark. More precisely, we may write that Fx(x) = P [{s S: X(s) < x}]. (3.2a) Properties. i) 0 5 Fx(x) _< 1 (since Fx(x) is a probability). ii) limx Fx(x) = 1 (since the event {X < z} is certain). iii) limx_ Fx(x) = 0 (because the event {X < -z} is possible). iv) The function Fx is non-decreasing; that is, if x0 < Xl, then Fx(xo) <_ Fx(Xl) (this follows from the fact that the event {X <_ x0} is included in {X < xl }). v) We can show that the function Fx is right-continuous; that is, Fx(x) = Fx(x+), where Fx(x +) := limFx(x +3). $o Proposition 3,2.1, We have: P[a < X < b] = Fx(b) -

Fx(a).

58 3 Random Variables Proof. We only have to notice that the event {X < b} can be divided into two disjoint (and exhaustive) parts: {X <b}= {X <a}U{a < X <b}. Since the two events in the right-hand member of the equation are incompatible, it follows that Fx(b) = Fx(a) + P[a < X <_ b]. [] Corollary 3.2.1. We have: P[X = x] = Fx(x) - Fx(x-), where Fx(x-) :---- lim Fx(x - 8). ,$0 Proof. We choose a = x - 8 and b = x in the preceding proposition, and we take the limit as 8 decreases to O. [] Remark. We deduce from the corollary that if Fx is a continuous function, then the probability P[X = x] is equal to zero for any real number x. It also follows that P[a < X < b] = Pla < X < b] = P[a < X < b] = Pla < X < b] when the distribution function of X is continuous. Types of Random Variables Definition 3.2.2. Let X be a random variable that can take, at most, a denumerably infinite number of values. That is, Sx is of the form Sx = {x, x2 .... }. Then, we say that X is a discrete random variable. Example 3.2.1 Suppose that we observe the lifetime of an electric light bulb that works. Suppose also that the probability that the lifetime T (in hours) of a light bulb of this type takes on a value in the interval [a, b) is given by P[a _< T < b] = e - a/l�� - e -b/l��, where 0 < a < b _< Let X be the number of complete periods of 100 hours that the light bulb lasts. Then, Sx = {0, 1,2 .... }. Therefore, X is a discrete random variable. Now, since X only takes on non-negative integer values, we have: Fx(x)=-- P[X <_x]=O Vx <0.

3.2 The Distribution Function 59 Remark. We may write that Fx(x) : P[X E {xk Sx: xk < x}]. Since there is not a single element of Sx that is inferior to zero, we have indeed: Fx(x) = 0 for all x < 0. Next, for any real number in the interval [0, 1), we can write that Fx(x) = P[X e {0}1 --= P[X = 01 = P[T < 1001 = 1 - e -1. Remark. Since we assume that the light bulb is working, we have: P[T < 100] ---- P[0 < T < 100] ---- P[0 _< T < 100], because P [T ---- 0] ---- 0. Similarly, for any real number in the interval [1, 2), we obtain that Fx(x) = P[X=O]+P[X= 1]= 1-e -I+P[100<T <200] = 1-e -+(e --e -z)= 1-e -2 In general, for any x in the interval [n, n + 1), where n {0, 1 .... }, we find that Fx(x) --: P[X < x] = P[T < (n + 1)100] = 1 - e -n-. For example, Fx(2.5) = Fx(2) = P[X {0, 1,2}] = P[T < 300] = 1 - e -3 = 1 - e -2-1 . Thus, we can write that 0 ifx <0, Fx(x) = 1 -- e -int(x)-I ifx >_ 0, where int(x) denotes the integer part ofx. We say that Fx is a staircase function (see Fig. 3.2). Definition 3.2.3. Let X be a random variable that can take a non-denumerably in- finite number

of values. If Fx is a continuous function, then we say that X is a continuous random variable. Example 3.2.2 In the preceding example, we can write that ST : (0, cx) and that 0 ift <0, FT(t)=P[O< T <t]=P[0<T <t]= 1-e -t/◆◆ift >0, since P[T : 0] = 0 (see above) and P[T = t] = 0 (because the sample space S = (0, cx) associated with the random experiment that consists in observing the lifetime of a light bulb that works is continuous). The function FT being continuous (see Fig. 3.3), T is a continuous random variable. Remark. FT (t) is differentiable everywhere, except at t = 0.

60 3 Random Variables F x (x) 1 0.2, 0 0 o o 1 2 3 x Figure 3.2. Distribution function of the random variable X in Example 3.2.1. Definition 3.2.4. Let X be a random variable that can take a non-denumerably in- finite number of values. If Fx is not a continuous function, then we say that X is a random wtriable of mixed type, Remark. A random variable of mixed type is in fact a function that is defined at the same time on a finite or denumerably infinite set and on one or more intervals, Moreover, we can write that Fx(x) = pF(x) + (1 - p)F2(x), where 0 < p < I is the probability that X is discrete and F (x) (respectively F2(x)) is the distribution function of a discrete random variable X (resp. a continuous ran- dom variable X2), In fact, we can also write that X =X [{X is discrete} and X2=X [{X is continuous}. Example 3.2.3 In Example 3.2,1, suppose that the light bulb considered is taken at random and that the probability that it works is equal to 0.9. Let Y be the lifetime 0.8 0.6 0.4 0.2 100 200 300 400 500 t Figure 33, Distribution function of the random variable T in Example 3.2.2.

3.2 The Distribution Function 0.8 0.6 0.4 0.2- 0 100 200 300 400 500 Y Figure 3.4. Distribution function of the random variable Y in Example 3.2.3. 61 of the light bulb and let us define the event F = the light bulb works. Then, we can write, by the total probability rule, that Fy(y) =-- P[Y < y] = P[Y < y] F]P[F] + P[Y < y] FC]p[F c] = P[T < y](0.9) + P[O < y](O.1) 'v'y > O. Remark. We can write that Y] F =-- T and Y] F c =-- O. Furthermore, the probability P[0 < y] is equal to 1 if y > 0. In fact, the event {0 < y} is deterministic, so that its probability is either 0 or 1. Using the preceding example, we find that 0 ify <0, Fr(y) = 0.1 ify = 0, 1 -0.9e -y/l◆◆ ify > 0. Since Sr = [0, x) and Fr is a discontinuous function (see Fig. 3.4), Y is of mixed type. Remarks. i) We could in fact include the point y = 0 in the second interval. ii) Here, we could write that Fy(y) = O.1Fg(y) + 0.9Fr(y), where Fu(u) = " {0ifu <0, lifu >_0. Actually,

U is the constant 0 (a degenerate discrete random variable). iii) Suppose, in addition, that we replace a light bulb that is still working after 200 hours, for preventive maintenance. Let W be the time that a light bulb is used. Then, we find that (see Fig. 3.5) 0 ifw <0, Fw(w) = 1 - 0.9e -w/�� if0 < w < 200, 1 if w > 200.

62 3 Random Variables 1: 0.8 0.6 0.4 0.2 50 100 150 200 250 300 350 w Figure 3.5. Distribution function of the random variable W in Example 3.2.3. In this case, we can write that Fw(w) = pFx(w) + (1 - p)Fx2(w), where p := P[W is discrete] =0.1 + 0.9PIT >_ 2001 = 0.1 +0,9e -2 (= 1 - p = 0.9(1 - e- 2)), 0 ifw < 0, Fx,(w) = ;[if() < w < 200, 1 if w > 200 and 0 if w < 0, FX(W) = I-e -"'/u) , if0< w < 200, 1 if w > 200. Thus, we have: 0 with probability 0. l/p, XI = 200 with probability 0.9e-Z/p and (see Section 3.3) d [(I / IOO) e- "'/tt. --Fx2(tO) = l-e 2 if0 < w < 200, iX2(to) dw " 0 elsewhere. Finally, we can also generalize the notion of distribution function as follows. Definition 3.2.5. Let X be a random variable and A an event. The conditional distribution function of X, given A, is defined by P[{X < x} (A] Fx(x I A) = ifP[A] > 0. (3.3) P[A]

3.2 The Distribution Function 63 Example 3.2.4 In Example 3.2.2, let A = the light bulb lasts less than to hours. Then, we have: P[{T _< t} (3 {T < to}] Fr(t I A) = Fr(t I T <to)= PIT < to] FT{t_____) if t < to, = Fr to) ift >_ to (since T is a continuous random variable). Using Example 3.2.2, we obtain that FT(t I T <to)= 0 ift <0, 1 - e -t/lO0 if0<t <to, 1 -- e -toll�� ! ift >to. Remark. Without having to do any calculations, we could write, for instance, PIT <2001 T < 100]= 1. However, in the case of the probability P[T < 100 I T < 200], we must make use of the formula above. Similarly, let B = the light bulb lasts more than to hours. Then, Fz(t l B) = Fz(t I T > to) = 0 = P[to < T < t] PIT > to] P[{T < t} f3 {T > to}l P[T > to] ift < to, [0 ift > to Fr(t) - Fr(to) I -- FT(tO) ift <to, ift >to. Remark. In this case, we could write, directly, that P[T <_ 100l T > 2001 = 0, while the probability P[T <_ 200 I T > 100] must be calculated.

64 3 Random Variables 3.3 The Probability Mass and Density Functions Definition 3.3.1, Let X be a discrete random variable. Then, we have: Sx = {xl , x2, ... }. The function px, defined by px(xk) = P[X = xk] (3.4) for k = 1,2 ..... is called the probability mass function of X. Remarks. i) The function Px has the following properties: a) px(xk) >_ 0 Yk (since it is a probability); b) =

$p_X(x_k) = 1$ (because $P[X \in S_X] = P[S] = 1$). ii) If we assume that $x_1 < x_2 < \cdots$, then we can write that $p_X(x_1) = F_X(x_1)$ (3.5a) and $P_X(x_k) = F_X(x_k) - F_X(x_{k_-})$ for $k = 2, 3, \ldots$ (3.5b) We also have: $p_X(x) = F_X(x) - F_X(x-)$ for any real number x. (3.6) iii) Using Heaviside's function $u(x)$, defined by $\begin{bmatrix} 0 & \text{if } x < 0, \\ 1 & \text{if } x > 0, \end{bmatrix}$ (3.7) we can express the distribution function of a discrete random variable X as follows: $= \sum p_X(x_k) u(x - x_k)$. (3.8) $F_X(x)$ $k=1$ This formula is equivalent to $F_X(x) = \sum p_X(x_k)$. (3.9) Example 3.3.1 We deduce from the function $F_X(x)$ in Example 3.2.1 that $p_X(n) = F_X(n) - F_X(n-) = e^{-n} - e^{-n-1}$ for $n = 0, 1 \ldots$ (see Fig. 3.6, p. 65). To be more complete, we can add that $p_X(x) = 0$ if x is not a positive integer or zero. Remarks. i) When S_X is finite, the function P_X is often presented as a table. For example, if

3.3 The Probability Mass and Density Functions $P_X(x)$ 65 0 0 1 2 3 4 X Figure 3.6. Probability mass function of the random variable X in Example 3.3.1. then we write: $p_X(x) = \begin{cases} 1/4 & \text{if } x = -1, \\ 1/2 & \text{if } x = 0, \\ 1/4 & \text{if } x = 1, \end{cases}$ p_X(x) -1 0 1 1/4 1/2 1/4 Note that, in this particular case, we can also write the function P_X as a single for- mula as follows: $p_X(x) = \frac{x+1}{4}$ for $x = -1, 0, 1$, which is not always possible. Moreover, the function F_X can likewise be presented in the form of a table: x -1 0 1 $F_X(x)$ 1/4 3/4 1 However, we must be able to interpret these tables correctly. For example, we deduce from these two tables that $P_X(1/2) = 0$, while $F_X(1/2) = 3/4$. ii) Since $p_X(x)$ is a probability, we can easily generalize Definition 3.3.1 by con- sidering the conditional probability mass function of X, given an event A, defined by $p_X(x \mid A) = \frac{P[\{X = x\} \cap A]}{P[A]}$ if $P[A] > 0$. (3.10) For example, let $A := \{X > 0\}$. Then we find, using the random variable X of part i) above, that Indeed, we have: $p_X(0 \mid X > 0) = \frac{P[\{X = 0\} \cap \{X > 0\}]}{P[X > 0]} = \frac{P[X = 0]}{P[X > 0]} = \frac{1/2}{1/2 + 1/4} = 2/3$

66 3 Random Variables and $p_X(1 \mid X > 0) = 1 - p_X(0 \mid X > 0) = 1/3$. Note that the relationship $p_X(0 \mid X > 0) = 2 p_X(1 \mid X > 0)$ follows from the fact that $p_X(0) = 2 p_X(1)$. Now, let X be a continuous random variable and x be a point where the function F_X is differentiable. Then, we can write that $P[x < X < x + \epsilon] = F_X(x + \epsilon) - F_X(x) \simeq F_X'(x + \epsilon) - F_X(x) \simeq \epsilon f_X(x)$, where $\epsilon > 0$ is small and f_X is the derivative of F_X. Definition 3.3.2. Let X be a continuous random variable. The function f_X, defined $f_X(x) = \frac{d}{dx} F_X(x)$ (3.11) (if the derivative exists), is called the (probability) density function of X. Remark. The function $f_X(x)$ is approximately equal to the probability that the ran- dom variable X takes a value in a small interval of length around the point x (from $x - (\epsilon/2)$ to $x + (\epsilon/2)$,

for instance), divided by the length of this interval; fx(x) is not the probability that X takes on the value x (this probability is equal to zero, because X is a continuous random variable). Properties. i) fx(x) 0 (by the preceding remark, or since Fx is a non-decreasing function). ii) Integrating Equation (3.11) from - x to x, we obtain: fx(t)dt = -Fx(t)dt = Fx(t)lx_ = Fx(x) -0= Fx(x). (3.12) It follows that fS fx(x)dx = Fx(x) = (3.13) 1. Actually, any non-negative piecewise continuous function that possesses the preced- ing property is a density function. We also deduce from Equation (3.12) that Pla < X < bl = Fx(b) - Fx(a) = fx(x)dx. (3.14) Thus, the probability that X takes on a value in the interval (a, b] is given by the area under the curve fx(x) from a to b. Remarks. i) If, instead of Equation (3.13), we have:

3.3 The Probability Mass and Density Functions 0.01 0.008 0.006 � 0.004 0.002 - O 0 100 200 300 400 t Figure 3.7. Probability density function of the random variable T in Example 3.3.2. 67 fx(x)dx = c, where 0 < c < 1, we say that the random variable X is defective or improper. In the general case when the integral is equal to c > 0, we only have to define gx (x) = fx (x)/c to obtain a valid or proper random variable. Similarly, a discrete random variable can be defective. ii) If Sx is an interval or a set of intervals, rather than the entire real line, we write that the function fx is equal to zero outside Sx. Example 3.3.2 Since the distribution function of the random variable T in Example 3.2.1 is differentiable Yt -� 0 (see Example 3.2.2), the density function of T is given by {le -t/lO0 ift > 0, fT" (t) = 0 elsewhere (Fig. 3.7). Notice that the function fT" (t) is continuous everywhere, except at t = 0. Now, as in the case of the distribution function, we can generalize the definition of the density function. Definition 3.3.3. Let X be a continuous random variable and A an event. We define the conditional (probability) density function of X, given A, by d fx(x I A) = xFx(x I A) (3.15) (when the derivative exists). Example 3.3.3 The conditional (probability) density function of the random variable T in Example 3.2.1, given the event A = {T < to}, is given (see Example 3.2.4) by d fT(t I T < to) = -FT(t I T < to)

68 3 Random Variables fr(t) ift < to, le = to) = 1 0 if0 < t < to, if t > to 0 elsewhere. Remark. It is not always necessary to calculate the function Fx(x I A) to obtain fx (x t A). For example, let l/2if-I <x< 1, fx(x) = 0 elsewhere. Then, we find that fx (x I X < 0) can be obtained as follows: fx (x) fx(x Ix <0)- if-I <x <0 PlX < O| = /2 =1 if-I <x<O, 0 elsewhere. Finally, we would also

like to generalize the notion of density function to the discrete case. To this end, we consider the Dirac delta function: (x)= { 0 ifx�0, ec ifx = 0 (3.16) (with f_ (x)dx = 1). We have the following relationship between the functions u(x) and (x): u(x) = (t)dt. (3.17) 0(3 Using this relationship and Equation (3.8), we can define the density function of a discrete random variable by fx(x) = px(xk)3(x - xk) Vx R. k=l (3.18) Example 3.3.4 Let Then, we have: Fx (x) = 0 ifx <0, 0.2ifO <x < 1, 0.7ifl <x<2, 1 ifx > 2. and x 0 1 2 2; px(x) 0.2 0.5 0.3 1 fx(x)=O.23(x-O)+O.53(x-l)+O.33(x-2) Vx6R.

3.3 The Probability Mass and Density Functions fx (x) 69 2(3 0 1 2 3 4 Figure 3.8. Probability density function of the random variable in Example 3.3.5. Definition 3.3.4. Let X be a continuous random variable with Sx = [a, b]. The number Xp for which Fx(xp) = p, (3.19) where 0 < p < 1, is called the quantile of order p of X . In particular, X1 /2 x m (or) is called the median of X. Remarks. i) The number Xp is unique. However, if X is not a continuous random vari- able defined on [a, b], then we must modify the foregoing definition. For example, we define the median Xm, in the general case, as follows: P[X < Xm] >_ 1/2 and P[X >_ Xm] >_ 1/2. (3.20) If X is not a continuous random variable defined on a single interval [a, b], then the median is not necessarily unique. ii) If 100p = k, an integer, Xp is also called the kth percentile of X. The terms decile (if k is a multiple of 10) and quartile are used as well. Example 3.3.5 Let cx if 0 _< x < 2, fx(x)= c(4-x) if2 < x < 4, 0 elsewhere, where c > 0 (see Fig. 3.8). Calculate the constant c, the distribution function, the median and the 99th percentile of X. Solution. We must have: 1 = fx(x) dx = cx dx + c(4-x)dx = 4c = c = 1/4.

7O 3 Random Variables 0.8 0.6 0.4- 0.2 0 1 2 3 4 5 X Figure 3,9. Distribution function of the random variable in Example 3.3.5. Or we can use the fact that the graph of the function fx is a triangle of area equal to 4c, so that c = 1/4. Then, we can write that Fx (x) = 0 if x < 0, f) fo x t X 2 Odt+ -dt =-- if0 <x <2, o 4 8 - fo 2 [, if2 < x < 4, X) 2 t 1 (4 dt+ (4-t)dt = 1 1 ifx >4 (see Fig. 3.9). Making use of the function Fx, we find that Xm = 2 (because Fx(2) = 1/2) and, solving the equation (4 - X0.99) 2 1 -- 0.99 (with 2 < xo.99 < 4), that xo.99 ' 3.71. 3A Important Discrete Random Variables Bernoulli distribution. Suppose that we perform a random experiment E. We con- sider a particular event A. Let X be the indicator variable of A; that is, 1 Jacob (or Jacques) Bernoulli, 1654-1705, was born and died in Switzerland, He was a member of a famous

family of mathematicians and physicists of Belgian origin that in- cluded his brother Johann and his nephew Daniel (Johann's son). He first studied philosophy and theology. He taught mathematics to his younger brother Johann, but later they quarreled and ended their relationship. He contributed greatly to algebra, infinitesimal cal- culus, mechanics and geometry. He solved what is now known as the Bernoulli differential equation. His important book on probability theory was published eight years after his death, In this book, we also find the Bernoulli numbers.

3.4 Important Discrete Random Variables 71 1 if A occurs, (3.21) X = 0 ifA c occurs. We say that X has a Bernoulli distribution with parameter p, where p := P[A] is called the probability of a success. We have: x 0 1 E px(x) 1 - p p 1 This table can be replaced by px(x)=pXql-X forx=0andl, (3.22) where q := 1 - p is the probability of a failure. Remark. A parameter is a constant (generally unknown, in practice) that appears in a probability mass function or a density function, and that can take any value in or in a subset of , like [0, 1], (0, x), etc. We will see in Section 6.1 how to estimate unknown parameters. Bernoulli trials. Suppose now that we repeat the random experiment E n times. We say that the repetitions El, E2 ..... En constitute a sequence of Bernoulli trials if a) the repetitions are independent; b) the probability of a success is the same for each repetition. For example, suppose that a woman must cross n intersections with traffic lights when she goes to work. Let Ek be the random experiment that consists in observing if the woman must wait for the green light when she arrives at the kth intersection, for k = 1,2 ..... n. Then the trials E ..... En are Bernoulli trials if and only if we may assume that the probability that the woman can cross a given intersection without having to stop is the same for each intersection and does not depend on what happened when she arrived at the previous intersections. In practice, unless the traffic lights are distant enough from one another, the assumption of independence of the trials is not (exactly) satisfied. Similarly, for the assumption b) to be satisfied, the length of the green light in the cycle of the traffic lights must be the same for each intersection, which is not true in general. However, to be able to estimate the probability that the woman has to stop at exactly five intersections in ten, for instance, it is almost necessary to make such simplifying assumptions. Binomial distribution. Let X be the number of successes in n Bernoulli trials. We say that X has a binomial distribution with parameters n and p, where p is the probability of a success. We have: Sx = {0, 1 ..... n}. We write: X B(n, p).

Proposition 3.4.1. The probability mass function of X $B(n, p)$ is given by $p_X(k) = \binom{n}{k} p^k q^{n-k}$ for $k = 0, 1 \ldots n$. (3.23)

Proof. Let E be the random experiment that consists in observing, for each of the n Bernoulli trials, whether the event A occurred or not. Then, we have: $X = n$ $X = 0$ S {3A...A, Aca c^n times n times There are 2^n elementary events in all. Let s be one of the elementary events for which $X = k$. In particular, there is AA...A A"AC...A". times (n-k) times By independence, we can write that PI [s]] = $pkq,,-t$. Furthermore, there are () dif- ferent elementary events in S with exactly k successes and $(n - k)$ failures. Equation (3.23) then follows by equiprobability and incompatibility. [] Properties. i) The function P_X defined in (3.23) is a proper probability mass func- tion, because it is non-negative and $\sum p_X(k) = (17 + q)" = 1" = 1$, k=0 by Newton's binomial theorem (see the example in Fig. 3.10). ii) We can show that the function $P_X(k)$ reaches its maximum at $k_{max} = int[(n + 1)p]$. Moreover, if $(n + 1)p$ is an integer, then the maximum is also attained at $k_{max} - 1$. iii) The distribution function of X, $F_X(x)$, is obtained by summing of the function $p_X(k)$, for k from 0 to int(x): PX (x) 1/16 o 1 0 1 T) 2 3 4 X Figure 3.10. Probability mass function of a binomial random variable for which $n = 4$ and $p = 1/2$.

int(x) $F_X(x) = \sum p_X(k)$. k=0 We can use a pocket calculator to evaluate this function. There are also tables of the functions p_X and F_X. Table 3.1 (p. 74) gives the value of the function $F_X(x)$ for a few values of n 6 {2, 3 ..... 20} and of p 6 [0.05, 0.50]. Note that for values of p greater than 0.50, we can use the following relationship: $P[B(n,p) < k] = 1 - P[B(n,1-p) < n-k- 1]$ for k 6 {0,1 ..... n}. Remark. Suppose that n transistors are taken at random from a box containing m brand A and $N - m$ brand B transistors. Then, if the transistors are taken with re- placement, the number X of brand A transistors obtained has a binomial distribution with parameters n and $p = m/N$. However, if the transistors are taken without re- placement, then the trials are not independent. We can show that, in this case, we have: m N-m $p_X(k)$ -- for k = max{0, n - N + m} ..... min{n, m}. We say that X has a bypergeometric distribution with parameters N, n and m. We write: X Hyp(N, n, m). Geometric distribution. Let X be the number of Bernoulli trials needed to obtain a first success. Then, $S_X = \{ 1, 2 .... \}$. We say that X has a geometric distribution with parameter p. We write: X Geom(p) (or X Geo(p)). Proposition 3.4.2. Let X Geom(p). We have: $p_X(k) = q^{k-1} p$ for k = 1, 2 .....

(3.24) Proof. Let E be the random experiment in which the results of the Bernoulli trials are observed until event A occurs. We have: X=I X=2 X=3 S = { A , ACA, ACACA .... }. Then, we may write that px(k) := P[X = k] = P[{,ACAC... ACA}] ind. qk-lp. (k- l) times Indeed, in order that X = k, there must first be k - 1 consecutive failures, and then a success. The above result is then obtained by independence. [] Properties. i) The function Px is non-negative and

74 3 Random Variables Table 3.1. Distribution function of the binomial distribution. P n x 0.05 0.10 0.20 0.25 0.40 0.50 2 0 0.9025 0.8100 0.6400 0.5625 0.3600 0.2500 I 0.9975 0.9900 0.9600 0.9375 0.8400 0.7500 3 0 0.8574 0.7290 0.5120 0.4219 0.2160 0.1250 I 0.9927 0.9720 0.8960 0.8438 0.6480 0.5000 2 0.9999 0.9990 0.9920 0.9844 0.9360 0.8750 4 0 0.8145 0.6561 0.4096 0.3164 0.1296 0.0625 I 0.9860 0.9477 0.8192 0.7383 0.4752 0.3125 2 0.9995 0.9963 0.9728 0.9493 0.8208 0.6875 3 I .(X)(X) 0.9999 0.9984 0.9961 0.9744 0.9375 5 0 0.7738 0.5905 0.3277 0.2373 0.0778 0.0313 I 0.9774 0.9185 0.7373 0.6328 0.3370 0.1875 2 0.9988 0.9914 0.9421 0.8965 0.6826 0.5(XX) 3 I .(XXX) 0.9995 0.9933 0.9844 0.9130 0.8125 4 I .(XXX) I .(X)(X) 0.9997 0.9990 0.9898 0.9688 I 0 0 0.5987 0.3487 0.1074 0.0563 0.(X)60 0.(X) 10 I 0.9139 0.7361 0.3758 0.2440 0.0464 0.0107 2 0.9885 0.9298 0.6778 0.5256 0.1673 0.0547 3 0.9990 0.9872 0.8791 0.7759 0.3823 0.1719 4 0.9999 0.9984 0.9672 0.9219 0.6331 0.3770 5 I .(X)(X) 0.9999 0.9936 0.9803 0.8338 0.6230 6 I .()(XX) 0.9991 0.9965 0.9452 0.828 I 7 0.9999 0.9996 0.9877 0.9453 8 1.0000 I .(X)00 0.9983 0.9893 9 0.9999 0.9990 15 0 0.4633 0.2059 0.0352 0.0134 0.0005 0.0000 I 0.8290 0.5490 0.1671 0.0802 0.0052 0.0005 2 0.9638 0.8159 0.3980 0.2361 0.0271 0.0037 3 0.9945 0.9444 0.6482 0.4613 0.0905 0.0176 4 0.9994 0.9873 0.8358 0.6865 0.2173 0.0592 5 0.9999 0.9977 0.9389 0.8516 0.4032 0.1509 6 1.0000 0.9997 0.9819 0.9434 0.6098 0.3036 7 1.0000 0.9958 0.9827 0.7869 0.5000 8 0.9992 0.9958 0.9050 0.6964 9 0.9999 0.9992 0.9662 0.849 I I 0 1.0000 0.9999 0.9907 0.9408 I I 1.0000 0.9981 0.9824 12 0.9997 0.9963 13 1.0000 0.9995 14 1.0000

3.4 Important Discrete Random Variables Table 3.1. Continued. P x 0.05 0.10 0.20 0.25 0.40 0.50 0 0.3585 0.1216 0.0115 0.0032 0.0000 1 0.7358 0.3917 0.0692 0.0243 0.0005 0.0000 2 0.9245 0.6769 0.2061 0.0913 0.0036 0.0002 3 0.9841 0.8670 0.4114 0.2252 0.0160 0.0013 4 0.9974 0.9568 0.6296 0.4148

0.0510 0.0059 5 0.9997 0.9887 0.8042 0.6172 0.1256 0.0207 6 1.0000 0.9976 0.9133 0.7858 0.2500 0.0577 7 0.9996 0.9679 0.8982 0.4159 0.1316 8 0.9999 0.9900 0.9591 0.5956 0.2517 9 1.0000 0.9974 0.9861 0.7553 0.4119 10 0.9994 0.9961 0.8725 0.5881 11 0.9999 0.9991 0.9435 0.7483 12 1.0000 0.9998 0.9790 0.8684 13 1.0000 0.9935 0.9423 14 0.9984 0.9793 15 0.9997 0.9941 16 1.0000 0.9987 17 0.9998 18 1.0000 75 Zpx(k)=pZqk-I=p{l+q+q2+...}q<l P =1 =1 1 --q Hence, Px is a proper probability mass function (see Fig. 3.11). ii) The distribution function of X is given, for any integer n, by PX (x) 1/8 Figure 3.11. Probability mass function of a geometric random variable for which p = 1/2.

76 3 Random Variables 1 -- qn = 1 -- qn. (3.25) Fx(n)= px(k)= p{l +q +'"+qn-1} = P l -q k=l It follows that P[X > n] = qn for n = 1, 2 ..... Proposition 3.4.3. Let X be a geometric random variable. Then, it possesses the following memoryless property: P[X >k+j[X > j]= P[X >k] Yk, j6{1,2 .... }. (3.26) Proof. We have: P[X > k +j [X > j]= PI{X > k+j}V{X > j}] PIX > j] P[X > k + j] qk+j ---- -- q = P[X > k]. P[X > j] qJ Remark. We also say that the geometric distribution has the "no aging" or "lack of aging" property. In fact the geometric distribution is the only discrete distribution that possesses this property. Remarks. i) The geometric distribution is sometimes defined as being the number X of Bernoulli trials befire the first success. We then have: Sx = {0, 1 .... } and the function px becomes px(k)=qkp fork=0,1 ..... If we denote this variable by Y, then we can write that Y = X - 1. The choice of the random variable used when we propose a stochastic model for a given variable depends of course on the possible values of this variable. For example, if we look for a stochastic model for the number of persons involved in a car accident, then it is preferable to choose the random variable X for which S = { 1,2 .... }. On the other hand, if the variable X denotes the number of persons injured in a car accident, we should choose the random variable X such that Sx = {0, 1 .... }. ii) We can generalize the definition of the geometric random variable as follows: let X be the number of Bernoulli trials necessary to obtain r successes. Then, we have: Sx = {r, r 1 .... } and we find (the last trial being necessarily a success) that Px (k) = i) p q for k = r, r + 1 ..... (3.27) We say that X has a Pascal distribution (see p. 2), or a negative binomial distri- bution, with parameters r and p. We write: X NB(r, p).

3.4 Important Discrete Random Variables 77 Poisson distribution 2 Poisson's

theorem. Let X B(n, p). Suppose that n -- xz and p $ 0 in such a way that n x p remains equal to . Then, we find that e-C k px(k) -+ -- fork =0, 1 ..... (3.28) k! Proof. Let Pk := px(k). For k = 0, we have: (Po = qn =(1-p)n = 1- Next, we consider the ratio Pk+l _ (k-1)pk+lq n-k-I _ (n- k)p 1- pk (n)pkqn-k (k + 1)q k + 1 " Since the two factors inside the square brackets tend to 1 as n tends to infinity, we can write that lim pk+l_ n-- pk k + 1" (3.29) Using this result recurrently, we obtain Equation (3.28). Indeed, we have: lim Pl n -- -- - px(1) e -, n P0 0 + 1 2 lim P2 lim Pl n-- 1 + 1 n-- 2 ' so that etc. [] Definition 3.4.1. lf X is a discrete random variable with Sx = {0, 1 .... } and for which e-C k px(k) -- -- fork = 0, 1 ..... (3.30) k! then we say that X has a Poisson distribution with parameter t > O. We write: X Poi(). 2 Sim6on Denis Poisson, 1781-1840, was born and died in France. He first studied medicine and, from 1798, mathematics at the tcole Polytechnique de Paris, where he taught from 1802 to 1808. His professors at the lcole Polytechnique were, among others, Laplace and Lagrange. In mathematics, his main results were his papers on definite integrals and Fourier series. The Poisson distribution appeared in his important book on probability theory pub- lished in 1837. He also published works on mechanics, electricity, magnetism and astron- omy. His name is associated with various mathematical results.

78 3 Random Variables PX (x) 0.1 2 3 """ X Figure 3.12. Probability mass function of a Poisson random variable for which a = 1. Properties. i) The function Px above satisfies the two conditions required to be a proper probability mass function: it is non-negative and V "cx e - [] otk px(k) = = =o =o = 1 (see Fig. 3.12). ii) As in the case of the binomial random variables, we can use a pocket calculator or a table to evaluate the distribution function or the probability mass function of X. Table 3.2 (p. 79) gives the value of the function Fx(x) tbr a few values of the parameter e. iii) We can show that px(x) attains its maximum at x = 0 ife < 1 and at x = int(e) ife > 1. Finally, ife is an integer, then the maximum is attained at both x = e and Poisson's approximation to the binomial distribution. We deduce from Poisson's theorem that if n is sufficiently large (> 20) and p small enough (< 0.05), then we can approximate the distribution of a binomial random variable with parameters n and p by that of a Poisson random variable with parameter o =np. The approxima- tion is generally better for the function Fx than for Px (see Example 3.4.1). Remark. Before proceeding with Poisson's approximation, we must check whether the parameter p of the binomial distribution is smaller than or equal to 1/2. For

example, suppose that X B(n = 20, p = 0.9). We have: P[X = 15] -- 0.0319. Now, if we replace X by a random variable Y that has a Poisson distribution with parameter ot = 20 x 0.9 = 18, we obtain: P[X = 15] -- P[Y = 15] _ 0.0786,

3.4 Important Discrete Random Variables Table 3.2. Distribution function of the Poisson distribution.

x	0.5	1	1.5	2	5	10	15	20
0	0.6065	0.3679	0.2231	0.1353	0.0067	0.0000		
1	0.9098	0.7358	0.5578	0.4060	0.0404	0.0005		
2	0.9856	0.9197	0.8088	0.6767	0.1247	0.0028	0.0000	
3	0.9982	0.9810	0.9344	0.8571	0.2650	0.0103	0.0002	
4	0.9998	0.9963	0.9814	0.9473	0.4405	0.0293	0.0009	0.0000
5	1.0000	0.9994	0.9955	0.9834	0.6160	0.0671	0.0028	0.0001
6		0.9999	0.9991	0.9955	0.7622	0.1301	0.0076	0.0003
7		1.0000	0.9998	0.9989	0.8666	0.2202	0.0180	0.0008
8			1.0000	0.9998	0.9319	0.3328	0.0374	0.0021
9				1.0000	0.9682	0.4579	0.0699	0.0050
10					0.9863	0.5830	0.1185	0.0108
11					0.9945	0.6968	0.1848	0.0214
12					0.9980	0.7916	0.2676	0.0390
13					0.9993	0.8645	0.3632	0.0661
14					0.9998	0.9165	0.4657	0.1049
15					0.9999	0.9513	0.5681	0.1565
16					1.0000	0.9730	0.6641	0.2211
17						0.9857	0.7489	0.2970
18						0.9928	0.8195	0.3814
19						0.9965	0.8752	0.4703
20						0.9984	0.9170	0.5591
21						0.9993	0.9469	0.6437
22						0.9997	0.9673	0.7206
23						0.9999	0.9805	0.7875
24						1.0000	0.9888	0.8432
25							0.9938	0.8878
26							0.9967	0.9221
27							0.9983	0.9475
28							0.9991	0.9657
29							0.9996	0.9782
30							0.9998	0.9865
31							0.9999	0.9919
32							1.0000	0.9953

which is not a good approximation. On the other hand, let X* be the number of fail- ures among the 20 Bernoulli trials. In this case, we can approximate the distribution of X* by that of Z Poi(ot = 20 x 0.1 = 2) and write that P[X = 15] = P[X* = 5] -- P[Z = 5] -- 0.0361, which is a much better approximation. Similarly, we have: P[X < 15] = P[X* > 5] -- 0.0432 and P[Z > 5] -- 0.0527, while PlY < 15] -- 0.2867.

80 3 Random Variables Example 3.4.1 Suppose that the number X of errors in a computer program submit- ted for the first time to a mainframe computer, by each student in a certain group, is a random variable that has a Poisson distribution with parameter o = 3. Suppose also that all the computer programs submitted are independent of one another. a) What is the probability that a given program contains no errors? b) What is the probability that, among 20 programs submitted, there are at least three that contain no errors? c) Calculate approximately the probability in b) with the help of a Poisson distribu- tion. d) What is the probability that the mainframe computer receives less than five

pro- grams, from a given time instant, before it receives a first program that contains no errors? Solution. a) We seek P[X = 0] ---- e -3 0.05. Remark. The Poisson distribution as a model Ibr X is actually an approximation, since the number of errors in a program cannot tend to infinity. b) Let Y be the number of programs, among the 20 submitted, that contain no errors. a) Then Y has a binomial distribution with parameters n = 20 and p = P[X = 0] 0.05. We want 2 P[Y > 3] = 1 - P[Y < 2] _ 1 - ())(0.05)k(0.95) 2�-k k =0 " 18] = 1 - [(0.95) 2o + 20(0.05)(0.95) 9 + 190(0.05)'(0.95) " 1 - (0.3585 +0.3774+0.1887) = 1 - 0.9246 ----- 0.0754. c) We can write that P[Y > 3] P[W > 3], where W Poi(c =np - 1) ----- 1- -! -- 1--e - 1+1+ k =0 1 - (0.3679 + 0.3679 + 0.1839) = 1 - 0.9197 = 0.0803. Remark. If we replace p by 0.01, then we find that P[Y >_ 3] 0.0010 and P[W >_ 3] 0.0011. On the other hand, here if p = 0.1, we obtain that P[Y > 3] 0.3231 and P[W >__ 3] 0.3233, which is even better (in error percentage). In general, for an n fixed, the approximation should improve when the parameter p decreases. In fact, if we calculate P[Y --- 0] and the approximations obtained with a Poisson distribution for n = 20 and p = 1/100, 1/20 and 1/10, we obtain the following pairs:

3.4 Important Discrete Random Variables 81 (0.8179, 0.8187), (0.3585, 0.3679), (0.1216, 0.1353). We observe that the approximation deteriorates as p increases, the error percentages being equal to about 0.1%, 2.6% and 11.3%, respectively. But when we add many terms, the errors can cancel one another. For example, comparing the numerical val- ues of the corresponding terms in b) and c), we notice that the probability P [W = k] sometimes overestimates and sometimes underestimates the probability P[Y = k], so that the errors partly cancel. That is why it is generally easier to approximate the distribution function Fx than the probability mass function Px, as mentioned above. d) Let U be the number bf programs required to obtain a first program that contains no errors. We assume that there are at least five programs remaining to be received from the time instant considered. We seek P[U <6] =P[V <6], whereVGeom(p=P[X=0]0.05) = P[V < 5] -- 1 - (0.95) 5 -- 0.2262. Remark. The random variable U itself does not have a geometric distribution, since the group of students considered is not infinite. In fact, U is a defective random variable, because, if there are 20 students in the group (and if nobody has submitted his/her program yet), we can write that Example 3.4.2 In a certain lottery, six balls are picked at random and without re- placement among 49 balls numbered from 1 to 49. We win a prize if the combination that we have

chosen contains at least three "good" numbers. A player decides to buy one ticket per drawing until he wins a prize. What is the probability that he must buy less than ten tickets? Solution. Let M be the number of good numbers in the chosen combination, and let X be the number of tickets that the player will have to buy to win a prize for the first time. We have: $P[M > 3] = 1-P[M<2]$ 2 (k6)(&) =I-E (469) k=O (1 . 6, 096,454) + (6. 962, 598) + (15 . 123,410) = 1- 13,983,816 13,723, 192 -- 1 __ 1 - 0.9814 = 0.0186. 13,983,816

82 3 Random Variables Next, by independence of the drawings, we may write that X has a geometric distribution with p _ 0.0186. We want P[X < 10| = P[X < 9] " 1 - (1 -0.0186) 9 " 0.1555. Remark. The random variable M has in fact a hypergeometric distribution with pa- rameters N = 49, n = 6 and m = 6. If the sampling fraction n/N is small (< 0.1), then we can approximate the distribution of a hypergeometric random variable with parameters N, n and m by that of a binomial random variable with parameters n and p = m/N. In our case, we have: n/N = 6/49 > 0.1. The approximation to the probability P[M > 31 would give [(6) >3] "1-0"9724=0"0276. P[M>3I--P B 6, _ 3 Important Continuous Random Variables Uniform distribution. Suppose that a number X is taken at random in the interval [a, hi. In this case, the probability that X is in an interval of length (small enough) around a point x, where t < x < b, must be the same for any x. It follows that the probability density function of X must be a constant. Using the condition l, fx(x)dx = 1, we deduce that we must have: 1 fx(x) -- fora < x < b. (3.31) b-a Definition 3.5.1. Let X be a continuous random variable defined on the interval la, b] and whose probability density function is given by Equation (3.31). We say that X has a uniform distribution on the interval [a, bl atld we write that X U[a, b] (see Fig. 3.13, p. 83). Properties. i) The distribution function of X is given by (see Fig. 3.14, p. 83) Fx (x) = 0 ifx <a, fa x 1 dt -- x - a ira < x < b, (3.32) b-a b -a 1 ifx >b. ii) Let It, d] C [a, b], Then, we have: dmc P[c < X < d] -- b-a

3.5 Important Continuous Random Variables fx (x) 83 (0, 1/(b-a)) Figure 3.13. Probability density function of a random variable having a uniform distribution on the interval [a, b]. Thus, the probability that X is in a given subinterval depends only on the length of this subinterval. Example 3.5.1 Suppose that the number of power failures in the interval [0, t], in a certain region, has a Poisson distribution with parameter)t, where) > 0. Then, given that exactly one power failure occurred in the interval [a, b] C [0, t], we can show that the

moment at which this failure occurred is uniformly distributed on the interval [a, b]. We can also show that the waiting time between two power failures has an exponential distribution with parameter) (see Definition 3.5.2). Exponential distribution Definition 3..2. Let X be a continuous random variable defined on the interval [0, x). If the density function of X is of the form fx(x) =)e -zx for x > O, (3.33) then we say that X has an exponential distribution with parameter) > O. We write: X Exp()) (see Fig. 3.15, p. 84). F x [x) I) Figure 3.14. Distribution function of a random variable uniformly distributed on the interval [a, b].

84 3 Random Variables 1.5 1 0.5 X Figure 3.15. Density function of an exponential random variable with parameter ,k = 2. Properties. i) The function fx above is a proper probability density function, since it is non-negative and),,-x.,- -e -x" 1. ii) The distribution function of X is given by Fx(x)--- .e -xtdt= l-e -xx forx >_0 (3.34) and Fx(x) = 0 for x < 0 (see Fig. 3.16). We then obtain that PIX > x] = e -xx forx > 0, (3.35) iii) Let Y := int(X)+l. Then, Sr = {1,2 .... } and we can show (see Section 3.6) that Y Geom(p = 1 - e-X), iv) The exponential random variables are the only continuous random variables that possess the following memoryless property: 1 0.8 0.6 0.4 0.2 1 2 3 X Figure 3.16. Distribution function of an exponential random variable with parameter k = 2.

3.5 Important Continuous Random Variables 85 P[X > t + s l X > t] = P[X > s] Vs, t >0. (3.36) The demonstration of this property of the exponential distribution is similar to that of Proposition 3.4.3, p. 76, for the geometric distribution. This property is used in reliability and in queueing theory. v) The exponential distribution can be considered on the entire real line by defining)e -lxl forx]R, fx(x) = where) is a positive parameter. This random variable, called the double exponential distribution, is also known as the Laplace distribution. 3 Example 3.5.2 Suppose that the lifetime X (in years) of a machine has an exponential distribution with parameter k = 1/3. What is the probability that a three-year-old machine will still work at the end of three additional years? Solution. We want P[X >6IX >3]=P[X >6-3]=e -1 0.3679. Note that the answer would be the same if the machine were brand-new, or six years old, etc. If we apply this property to the lifetime of a car, for instance, it is clear that the assumption that we make is not entirely realistic. Indeed, the probability that a new car lasts at least three years is surely larger than the

probability that a ten-year-old car lasts at least three additional years. However, the assumption of an exponential lifetime can be acceptable for a shorter period of time. For example, the probability that a one-year-old car lasts at least two additional years should be almost equal to the probability that a brand-new car lasts at least two years. Gamma distribution Definition 3.5.3. The gamma function, denoted by F(.), is defined (for c > O) by F(c0 = x-l e -x dx. (3.37) Doing the above integral by parts, we find that F(oe) = (oe - 1)F(oe - 1) ifoe > 1. (3.38) 3 Pierre Simon (Marquis de) Laplace, 1749-1827, was born and died in France. In addition to being a mathematician and an astronomer, he was also a minister and a senator. He was made count by Napol6on and marquis by Louis XVIII. He participated in the organization of the lcole Polytechnique de Paris. His main works were on astronomy and on the calculus of probabilities: the TraitO de mcanique cOleste, published in five volumes, from 1799, and the Thdorie analytique des probabilitks, the first edition of which appeared in 1812. Many mathematical formulas bear his name.

86 3 Random Variables 1 2 3 4 5 x Figure 3.17. Probability density functions of various random variables having a gamma dis- tribution with k = I. It follows that ifoe = n � {2, 3 .... }, then F(o) = F(n) = (n - l)F(n - 1) = (n - l)(n - 2)F(n -2) ..... (n- l)(n-2),-, l,F(l)=(n- 1)!. 1 =(n- 1)!. In addition, we can show that F(1/2) = ,/-. Definition 3.5.4. Let X be a continuous non-negative random variable. If its proba- bility density function is of the form (kx)- fx(x) = for x >_ O, (3.39) F(oe) then we say that X has a gamma distribution with parameters c > 0 and) > O, We write: X Remarks. i) The parameter oe is a shape parameter, while . is a scale parameter (see Fig. 3.17), Because the shape of the density function fx changes a lot when the parameter o takes on various values, the gamma distribution is a very useful model for the applications. ii) The gamma distribution generalizes the exponential distribution since if we take = 1 in the function iX, we obtain: fx (x) = ke-)'x for x > 0. The gamma distribution is also used in reliability and it appears in the stochastic process known as the Poisson process. iii) If oe = n/2 and . = 1/2, the gamma distribution is also called the chi-square distribution with n degrees of freedom, where n is a positive integer. This distribu- tion is used in statistics, particularly in Pearson's goodness-of-fit test, which will be seen in Chapter 6. iv) Finally, if o is a positive integer, the gamma distribution is sometimes called the Erlang distribution, 4 which is very important in the

theory of queues, 4 Agner Krarup Erlang, 1878-1929, was born and died in Denmark. He was first educated by his father, who was a schoolmaster. He studied mathematics and natural sciences at the

3.5 Important Continuous Random Variables 87 Properties. i) Since F() is positive (for > 0), the function fx is non-negative. Moreover, we have: XO�-- 1 e-.x dx fx(x)dx -- f0 y=__zx 1 y_le_Y dy_ F(ot) _ 1. Thus, fx is a proper probability density function. ii) There is no simple formula that gives the distribution function of X in the general case. However, if oe is an integer, then we can show, integrating by parts repeatedly, that Fx (x) = 1 - � e -zx (') for x > 0. (3.40) k=0 That is, we have: P[X < x] = 1- P[Y < - I] = P[Y > ot], whereYPoi(.x). (3.41) Remark. We can also write that P[X < x] = P[Y > ot], because X is a continuous random variable. However, since Y is a discrete random variable, we have: P[X < x] P[Y > ot]. Note, additionally, that Equation (3.41) is exact, and not only an approximation. We will see in Chapter 5 that the time required for n events of a Poisson process to occur has a gamma distribution, whereas the number of events that occur in a given interval has a Poisson distribution, whence the above relationship between the two distributions. Example 3.5.3 Suppose that the daily consumption of electricity (in millions of kilowatt-hours) in a certain region is a random variable X having a gamma distri- bution with parameters = 3 and) = 1/2. Suppose also that the production capacity is equal to 12 millions of kilowatt-hours. Calculate the probability that the demand exceeds the capacity on a given day. Solution. We can write that P[X > 12] = PlY < 3], where Y Poi(12/2 = 6) =e 6 1+6+ --0.0620. If we make the assumption (which is probably not correct in practice) that the probability that the demand exceeds the capacity on a given day does not depend University of Copenhagen and taught in schools for several years. After meeting the chief engineer for the Copenhagen telephone company, he joined this company in 1908. He then started to apply his knowledge of probability theory to the resolution of problems related to telephone calls. He was also interested in mathematical tables.

88 3 Random Variables on what happened previously, that is, the assumption that the days are independent, then the probability that the demand exceeds the capacity on two consecutive days is given by approximately (0.0620) 2 - 0.0038. Finally, let N be the number of days in a given week on which the demand exceeds the production capacity. Under the pre- vious assumptions, we

can write that N has a binomial distribution with parameters n = 7 and p _ 0.0620. Remark. Another continuous random variable that generalizes the exponential dis- tribution and that is useful in reliability is the Weibull distribution 5 with parameters /4 > 0, y R and , > 0, whose probability density function is .[,(x)= - exp - forx > V- (3.42) Note that the exponential distribution is the particular case when/4 = l, = 0 and 3 = 1/). The parameter/4 is a shape parameter, is a position parameter and , is a scale parameter. Gaussian distribution Definition 3--. Let X be a random variable whose set of possible values is" Sx = (-,). [['the probability density fimction ([" X is (['the firm 1 { ("-')2} fir-<x<, (3.43) .l(x) - exp 2 then we say that X has a normal or Gaussian distribution w#h parameters and 2, where > O. We write: X N(iz, a2). Remark. This probability distribution is also sometimes called the place-Gauss distributkn, 6 or the place distribution (name used for the double exponential dis- tribution as well; see p. 85). Properties. i) Since a is positive, it is clear that the function fx is non-negative. To show that its integral on the entire real line is equal to 1, we consider / 2, where 5 E. H. Wallodi Weibull, 1887-1979, was born in Sweden and died in France. His family originated from Germany. He started his career in the Swedish Coast Guard. After com-pleting his doctorate, he worked as an inventor and consulting engineer for many companies in Sweden and in Germany. He published many papers on the strength of materials, fatigue and reliability, in addition to his papers, from 1939, on the distribution that bears his name. The Weibull distribution is used in many applications. 6 Carl Friedrich Gauss, 1777-1855, was born and died in Germany. Considered as one of the greatest geniuses of all time, he carried out many works on astronomy and physics, in addition to his impoant mathematical discoveries. In the field of mathematics, he was interested, in particular, in algebra and geometry He also discovered the method of least squares, which he used to make astronomical predictions. It is in fact for this method that he introduced the law of errors, that now bears his name, as a model for the eors in astronomical obseations

3.5 Important Continuous Random Variables 89 0.3 0.2 0.1 -4 -2 0 2 4 6 Figure 3.18. Probability density functions of Gaussian random variables with # = 0. f { / f 1 (x -- #)2 1 exp(--t2/2) dt, I :-- -r exp dx -- by making the change of viable t = (x - #)/a. Then, using polar coordinates, we can write that 1 1 ff_exp[-(t+s)/z],t,s: 1 = r exp(-r/2) dOdr. 2 2 We easily find that 12 = 1. Then, knowing that I 0 (because fx(x) 0), we deduce that I = 1. ii) The parameter # is

a position parameter, while a is a scale parameter. Fuaher- more, all Gaussian distributions have the same general shape, namely that of a bell (see Fig. 3.18). iii) The function fx is symmetrical with respect to #; that is, ix (x + u) = ix (-x + Moreover, fx (x) attains its maximum at x = # and has two inflection points: at x=#+r. iv) If # = 0 and r = 1, then we say that X has a standard Gaussian distribution. Its probability density function is given by 1 exp(_z2/2) for-cx < z < cx (3.44) (z) := and its distribution function is denoted by Now, if X N(#, r2), we find that its distribution function can be expressed in terms of � as follows: Fx(x)= � (--). (3.45) Thus, any probability involving a Gaussian distribution can be obtained by using the N(0, 1) distribution. To evaluate the function d, we can use a statistical software

90 3 Random Variables package or a table. Table 3.3 (p. 91) gives the value of the function qS(z) for z = 0.00... (0.01) ... 3.99 (by symmetry, we have: qs(-z) = 1 - qS(z)). From z = 3.90, we can write that qs(z) _ 1.0000. We find, with more precision, that � (4) 0.99997, (5) 0.9999997 and (6) 0.999999999. Alternatively, we can make use of the following relationship: � (z) = 1 -- Q(z), (3.46) where Q(z) is given, approximately, by Q(z) - - exp(-z-/2) for,-. > O. (3.47) (rr -- l)c.+ (z:. 2 +2rr) /2 Moreover, we have: Q(O) = 1/2 and Q(-z) = 1 - Q(c.). Finally, we also find tables of the function Q- (p) (the inverse function of Q) lbrO < p < 1 (see Table 3.4, p. 92). For instance, Q- (0.05) 1.645. Remark. The Gaussian distribution is often used as a model in situations in which we know, for example, that the random variable X that we consider cannot be neg- ative. This model can nevertheless be a good approximation to the true (generally unknown) model, as long as l* and a are such that Fx(x) - 0 ifx < O. In fact, the Gaussian distribution is useful mainly because of the central limit theorem, which will be seen in Chapter 4. Example 3.5,4 Suppose that the lifetime X of a microprocessor has a Gaussian dis- tribution with parameters l* and a 2. Calculate the probability P[I* - ka < X < #+ka],wherek{l,2 .... }. Solution. We have: P[# - ka <_ X < l* + kcr] co___n. FX(# + ka) - Fx(# - ka) = (k) - (-k) = 2(k) - 1, which is independent of # and a. For example, ifk = l, we find that (1) -- 0.8413, so that the probability that we seek is approximately equal to 0.683. Remark. If X is a continuous random variable such that Sx = (0, oo), and if Y := In X has a normal distribution with parameters # and a 2, then we say that X has a lognormal distribution (with parameters # and a2). We can show (see Theorem 3.6. l, p. 94) that the probability density function of X is given by 1 { (lnx--#) 2 } fx(x) -- V/-cr x

exp - forx > 0. (3.48) This random variable is important in many fields, notably in reliability.

3.5 Important Continuous Random Variables Table 3.3. Values of the function qb(z). 91 z +0.00 +0.01 +0.02 +0.03 +0.04 +0.05 +0.06 +0.07 +0.08 +0.09 0.0' 0.5000 0.5040 0.5080 0.5120 0.5160 0.5199 0.5239 0.5279 0.5319 0.5359 0.1 0.5398 0.5438 0.5478 0.5517 0.5557 0.5596 0.5636 0.5675 0.5714 0.5753 0.2! 0.5793 0.5832 0.5871 0.5910 0.5948 0.5987 0.6026 0.6064 0.6103 0.6141 0.3 0.6179 0.6217 0.6255 0.6293 0.6331 0.6368 0.6406 0.6443 0.6480 0.6517 0.4 0.6554 0.6591 0.6628 0.6664 0.6700 0.6736 0.6772 0.6808 0.6844 0.6879 0.5 0.6915 0.6950 0.6985 0.7019 0.7054 0.7088 0.7123 0.7157 0.7190 0.7224 0.6 0.7257 0.7291 0.7324 0.7357 0.7389 0.7422 0.7454 0.7486 0.7517 0.7549 0.7 0.7580 0.7611 0.7642 0.7673 0.7704 0.7734 0.7764 0.7794 0.7823 0.7852 0.8 0.7881 0.7910 0.7939 0.7967 0.7995 0.8023 0.8051 0.8078 0.8106 0.8133 0.9 0.8159 0.8186 0.8212 0.8238 0.8264 0.8289 0.8315 0.8340 0.8365 0.8389 1.0 0.8413 0.8438 0.8461 0.8485 0.8508 0.8531 0.8554 0.8577 0.8599 0.8621 1.1 0.8643 0.8665 0.8686 0.8708 0.8729 0.8749 0.8770 0.8790 0.8810 0.8830 1.2 0.8849 0.8869 0.8888 0.8907 0.8925 0.8944 0.8962 0.8980 0.8997 0.9015 1.3 0.9032 0.9049 0.9066 0.9082 0.9099 0.9115 0.9131 0.9147 0.9162 0.9177 1.4 0.9192 0.9207 0.9222 0.9236 0.9251 0.9265 0.9279 0.9292 0.9306 0.9319 1.5 0.9332 0.9345 0.9357 0.9370 0.9382 0.9394 0.9406 0.9418 0.9429 0.9441 1.6 0.9452 0.9463 0.9474 0.9484 0.9495 0.9505 0.9515 0.9525 0.9535 0.9545 1.7 09554 0.9564 0.9573 0.9582 0.9591 0.9599 0.9608 0.9616 0.9625 0.9633 1.8 0.9641 0.9649 0.9656 0.9664 0.9671 0.9678 0.9686 0.9693 0.9699 0.9706 1.9 0.9713 0.9719 0.9726 0.9732 0.9738 0.9744 0.9750 0.9756 0.9761 0.9767 2.0 0.9772 0.9778 0.9783 0.9788 0.9793 0.9798 0.9803 0.9808 0.9812 0.9817 2.1 0.9821 0.9826 0.9830 0.9834 0.9838 0.9842 0.9846 0.9850 0.9854 0.9857 2.2 0.9861 0.9864 0.9868 0.9871 0.9875 0.9878 0.9881 0.9884 0.9887 0.9890 2.3 0.9893 0.9896 0.9898 0.9901 0.9904 0.9906 0.9909 0.9911 0.9913 0.9916 2.4 0.9918 0.9920 0.9922 0.9925 0.9927 0.9929 0.9931 0.9932 0.9934 0.9936 2.5 0.9938 0.9940 0.9941 0.9943 0.9945 0.9946 0.9948 0.9949 0.9951 0.9952 2.6 0.9953 0.9955 0.9956 0.9957 0.9959 0.9960 0.9961 0.9962 0.9963 0.9964 2.7 0.9965 0.9966 0.9967 0.9968 0.9969 0.9970 0.9971 0.9972 0.9973 0.9974 2.8 0.9974 0.9975 0.9976 0.9977 0.9977 0.9978 0.9979 0.9979 0.9980 0.9981 2.9 0.9981 0.9982 0.9982 0.9983

0.9984 0.9984 0.9985 0.9985 0.9986 0.9986 3.0 0.9987 0.9987 0.9987
0.9988 0.9988 0.9989 0.9989 0.9989 0.9990 0.9990 3.1 0.9990 0.9991
0.9991 0.9991 0.9992 0.9992 0.9992 0.9992 0.9993 0.9993 3.2 0.9993
0.9993 0.9994 0.9994 0.9994 0.9994 0.9994 0.9995 0.9995 0.9995 3.3
0.9995 0.9995 0.9995 0.9996 0.9996 0.9996 0.9996 0.9996 0.9996 0.9997
3.4 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997
0.9998 3.5 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998
0.9998 0.9998 3.6 0.9998 0.9998 0.9999 0.9999 0.9999 0.9999 0.9999
0.9999 0.9999 0.9999 3.7 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999
0.9999 0.9999 0.9999 0.9999 3.8 0.9999 0.9999 0.9999 0.9999 0.9999
0.9999 0.9999 0.9999 0.9999 0.9999 3.9 1.0000 1.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000

92 3 Random Variables Table 3.4. Values of the function Q- 1 (p) for some values of p. p 0.10 0.05 0.01 0.005 0.001 0.0001 0.00001 Q- 1 (p) 1.282 1.645 2.326 2.576 3.090 3.719 4.265 3.6 Transformations Let X be a random variable and let Y := g(X), where g is a real-valued function defined on the real line. Then Y is also a random variable, because the composition of two functions is again a function. We want to obtain the functions Fr and fr (or Pr). For example, the distribution of the random variable V representing the speed of a particle may be known, but we might be interested in determining the distribution of its energy E, which is given by E = m V2/2, where rn is the mass of the particle. I) Discrete case. a) If X is a discrete random variable, then Y is a discrete random variable as well. Let xk l), xk2) .... be the values of X for which glxj)l=y for j= 1,2 ..... Then, we can write that PY(Y,t) : px(x,j)) for k = 1,2 ..... (3.49) j=l For example, let Sx = {-2, 0, 2} and x -2 0 2 px(x) 1/4 1/4 1/2 If Y := X 2, then the values -2 and 2 of X correspond to the same value, 4, of Y, and we find that y 0 4 E i Pr(Y) 1/4 3/4 1 In the particular case when the function g is bijective, we simply have: Pr(Yk) = px(xk), where yk = g(xk) Vk. (3.50) For instance, if Z := -X 3 in the above example, then to the values -2, 0 and 2 of X correspond respectively the values 8, 0 and -8 of Z, and we have: z -8 0 8 E pz(z) 1/2 1/4 1/4 1

3.6 Transformations 93 Example 3.6.1 Let X Geom(1/2) and { if X is even, Y := if X is odd. We have: py(1) = Z px(2k) = Z(1/2)2k-I(1/2) = 1/4 + (1/4) 2 + ... 1/3. k=I k=I It follows that pr(O) = 1 - (1/3) = 2/3. Therefore, we have: and y 0 1 E pr(y) 2/3 1/3 1 y 0 1 Fr (y) 2/3 1 Now, let Z := X - 1. Then, we have:

$Sz = \{0, 1 .... \}$ and $pz(k)=px(k+l)=(1/2)$ $(k+I-(1/2)=(1/2)$ $k+I$ for $k=0,1 ..... $ b) If X is a continuous random variable, but g is a staircase function, that is, a func- tion that takes at most a denumerably infinite number of values, then Y is a discrete random variable. We have: $PI'(Yk) := P[Y = Yk] = P[X A]$, where A is the event equivalent to $\{ Y = yk \}$ in Sx. (3.51) Example 3.6.2 Let X Exp(.) and $Y := int(X)+ 1$. We already mentioned in Section 3.5 that Y Geom($p = 1 - e-Z$). Indeed, in this case, the event equivalent to $\{Y=k\}$ is $\{k- 1 <X <k\}$ for $k6\{1,2 .... \}$, so that $pr(k) := PlY = k] = P[int(X) = k - 1] = P[k - 1 < X < k]$ f $[k\_ =e)(1-e-))=q-p, = .e-)Xdx=\_e-)x k -),(k- k-I$ where $p := 1 - e -)'$. II) Continuous case. If X is a continuous random variable and g is a continuous function, then Y is a continuous random variable as well. To obtain the distribution function of Y, we must find the probability of the event A C Sx which is equivalent to $\{Y< y\}$. Next, we differentiate $Fr(y)$ to obtain $fr(Y)$. We can also make use of the following theorem.

94 3 Random Variables Theorem 3.6.1. Suppose that the equation $y = g(x)$ has n (real) solutions: xl, ... , xn. Then, the probability density function of Y is given by $fv (Y) = Z fx (xk) Idxk/dyl$. (3.52) Remarks. i) If $n = 1$, so that the function g is bijective, then we have: $.fv(y) = .l'x(g- (y))ldg - (y) /dyl$, (3.52a) where $g 1$ is the im,erse function of g. For example, if $y = g(x) = e r$, then $x = g (y)=lny$. i i) Moreover, if $g (x) = x 2$, then we can write, in general, that $1 fr(y) = y lJx(-,/y) + fx(,/y)l$. (3.52b) However, if X is always positive, lbr instance, then $fx (- Y) = 0$. Proof (of Theorem 3.6.1). Consider the case when the function g is bijective and increasing. Then, we have: $Ply< Y<y+dyl=Plx<X<x+dxl$, where $g(x) = y J'(y)ldyl = J'x(x)ldxl ./"(y) = fx(x)ldx/dyl$. The proof in the general case is similar. Example 3.6.3 Let X N(0, 1) and $Y := X 2$. Then, we have: $Fr(y) = PIX 2 < .v]=- Pl-y 1/2 < X < y/2l = qb(y I/2)-qb(-y /2) := J'r(v) = y-1/2lqb(yl/2) q(-y/2)l =$ for $y \_ > 0$. In this example, the event $\{\_y/2 < X < y/2\}$ is equivalent to $\{Y < y\}$, and the equation $y = x 2$ has two solutions: $xl = \_y/2$ and $x2 = y/2$. Moreover, the random variable Y has a G($\diamond$, $\diamond$) distribution or a chi-square distribution with 1 degree of freedom. Example 3.6.4 Let X N(#, cr 2) and $Y := (X - #)/or$. In this case, the transfor- mation is bijective. We have: $f(Y) = fx(# + ry)ld(# + ry)/dyl - -- exp(-y2/2) I1 = q(y)$. That is, Y has a standard Gaussian distribution. In general, if $Z := aX b$, then Z has a Gaussian distribution with parameters $a# + b$ and $a2r$.

Example 3.6.5 Let X U(0, 1] and Y := -2 In X. Then, we find that Y has an exponential distribution with parameter) = 1/2. This kind of result is used in simu- lation. Remark. We must always give the set of possible values of the new random variable Y, that is, St. In the case when the function g is strictly increasing (respectively decreasing), we can show that if Sx = [a, b], then Sr = [g(a), g(b)] (resp. [g(b), g(a)]). 3.7 Mathematical Expectation and Variance Definition 3.7.1. The mathematical expectation, or the mean, of a random vari- able X is denoted by E[X] and is defined by E[X] =-- #x = Z XkPx(Xk) ifX is discrete, k=l xfx(x) dx if X is continuous. (3.53) Remarks. i) In general, we assume that the integral converges absolutely. That is, we assume that the integral of the absolute value of xfx (x) converges; similarly in the discrete case. Otherwise, the expected value of X does not exist. However, we must distinguish between the case when E[X] = x or -x, and that when E[X] = x - x. In the first case, we can simply say that the mathematical expectation of X is infinite (in absolute value), whereas in the second case it does not exist. ii) The mathematical expectation of the random variable X is a weighted average of its possible values, the weight associated with a given value being equal to its probability (in the discrete case) or to the value of the probability density function of X at this point (in the continuous case). iii) The expected value of X corresponds to the center of gravity of the distribution. iv) In the case when X is a random variable of mixed type, its expected value can be obtained by combining the two formulas in the definition of E[X]. For example, let Sx = {-1} U [0, 1].If P[X = -1] = 1/4 and fx(x)=3/4 forO<x < 1, then we can write that 1 flx3_d x = 1 3 1 E[X] = (-1). + J0 4 - + 8 -- 8" We can also make use of the Dirac delta function in the density function of X and employ the formula for continuous X as follows:

96 3 Random Variables -143(0) ifx = --l, fx(x) = � ifx [0, 1], 0 elsewhere f_ 4 f0 - = 1 1 I 3d x E[X] = xfx(x)dx = =(-1) + x 4 8 Finally, another way of expressing the density function of X is with the help of the indicator function IA, defined by lA(x) = otherwise. Then, we have: 1 3 fx(x) = 3(x + 1) + -llo, ii(x) Vx eR. v) The notion of mathematical expectation can be generalized by defining E[X I A], the conditional (mathematical) expectation of X, given an event A. We only have to replace fx(x) by fx(x [A) or px(x) by px(x [A) in the definition of E[X]. Proposition 3.7.1. The mathematical operator E is linear. That is. ['ck is a constant and gk is a real-wlued function for k = 1,2 ..... n. then

k=l k=l Proof. This result follows directly from the fact that the integral is a linear operator, and so is the summation. [] We deduce from the definition of E[X] and the preceding proposition that E[c] = c and E[aX + b] = aE[X] + b for any constants a, b and c. Remark. A (real) constant c can be considered as a degenerate random variable C whose probability mass function is 1 ifx =c, pc(x) = 0 otherwise, or whose probability density function is fc(x)=3(x-c) for allx 6R. Proposition 3.7.2. a) Let X be a non-negative continuous random variable. Then, we have: E[XI = P[X > xldx = [l - Fx(x)ldx. (3.55) b) If X is a discrete random variable whose possible values are included in the set {0, 1, 2 .... }, then 0(3 = P[X > E[X] k]. (3.56) k=0

3.7 Mathematical Expectation and Variance 97 Proof. a) We have: E[X] = xfx(x) dx = fx(x) dtdx = x)dxdt = P[X > tldt. b) The proof is similar to the previous one. Remark. We can also prove the result in a) as follows. By integrating by parts, we find that I1 - Fx(x)ldx = x[1 - Fx(x)]l + xfx(x)dx = lim x[1 - Fx(x)] + E[X]. Now, by l'Hospital's rule (H.R.), we have (assuming that E[X 2] < lim x[1 - Fx(x)] = lim 1-Fx(x) H.._R. lim x2fx(x)= O, 1/X xoo since X2fx(X) must converge (toward 0) when x tends to infinity (otherwise the integral ofx2fx (x) from 0 to infinity would diverge). Example 3.7.1 Let X "- Poi(o0. We have: Example 3.7.2 Let X "- Exp(). Making use of the preceding proposition, we obtain (see Equation (3.35)): fo �� fo �� e-XX 1 E[X] = P[X > xldx = e -zx dx -- = We can demonstrate the following proposition or use it as a definition. Proposition 3.7.3. Let X be a random variable and let Y := g(X). Then, we have." oo Z g(x)px(x) ifX is discrete, E[Y] = = � g(x)fx(x) dx is continuous. ifx oo

98 3 Random Variables Remarks. i) Let X be a continuous random variable and g be the indicator function of the event {X � A}; that is, II ifX�A, g(X) = 0 otherwise. We can write that E[g(X)] = 1. PIg(X) = 11 = PIX � AI = fx(x)dx = g(x)fx(x)dx, which corresponds to the above formula. Note that here g(X) has a Bernoulli distri- bution with parameter p = P[X � A[. ii) Let BI ..... B, be a partition of a sample space S (for which P[Bil > 0 for i = 1 ..... n). We can write that E[g(X)] = E[g(X)] BiIP[Bi]. i=I For example, if X is a random variable of discrete type, then px(Xk) := P[X = Xk [= P[X = Xk] Bi IP[Bi I. i=I It follows that Elg(X)I = Zg(xk)px(xt)= Z Zg(xt)PIX--xk l BilelBil k=I k=I i=I = g(x-)P[X ----xk I Oil P[Bi[= E[g(X) lOi]P[Oi]. i=I i=I Example 3.7.3 Let X U[0, 21. Then the probability density function of Y := X 2 is given by fr(y) =

$f_X(x)$ dXdy = fx(,f-fi) -y =1 1_ _ 1 if0<y <4, 2 2,,/ 4,,/ - - because the transformation $y = x^2$ is bijective here, since $S_x = [0, 2]$. It follows that fo 1 1 y3/2 4 E[Y]= y-@ dy-- 4 -- 7" Now, we calculate (directly) fo 2 1 dx x220 4 $E[X^2] = xZz = 6 = 7'$ Thus, Proposition 3.7.3 enables us to obtain the mean of the random variable Y much more easily.

3.7 Mathematical Expectation and Variance The Variance of a Random Variable 99 Definition 3.7.2. The variance of a random variable X is defined by VAR[X] = ax 2 ----- $E[(X - E[X])^2]$ (> 0). (3.57) Remarks. i) The variance is a measure of variability of the random variable about its mean. It is the moment of inertia of the distribution. Moreover, the variance of a random variable cannot be negative. In fact, the variance VAR[X] is strictly positive, unless X is a constant. If X = c, then E[X] = E[c] ---- c and VAR[X] = VAR[c] = E[(c- c)²] = 0. ii) We also define the standard deviation of X by STD[X] =--cr x = $(VAN[X])^{1/2}$. (3,58) The units of measure of X and of its standard deviation are the same. iii) We generalize the notion of variance by considering the conditional variance of X, given an event A, defined by VAR[X I A] = $E[(X - E[X I A])^2 I A]$. (3.59) We can show (see Proposition 3.7.5, p. 100) that VAR[X I A] = $E[X^2 I A] - (E[X I A])^2$. In fact, to calculate the conditional variance of a random variable X, given an event A, it is often easier to first find the probability density (or mass) function of Y := X I A, and then to use the definition of VAR[Y]. For example, if X U[- 1, 1] and if A = {X < 0}, we find that Y := X I A U[-1,0] and it is then easy to calculate VAR[Y] (see Example 3.7.5, p. 100). iv) We can calculate the variance of X as follows: VAN[X] = $E[(X - E[X])^2 I B_i]P[B_i]$, i=1 where the events BI ..... Bn form a partition of a sample space S and P[Bi] > 0 'i. However, in general n VAR[X] - ZVAR[X I Bil]P[Bi], i=1 because $E[(X - E[X])^2 I B_i] 7 A VAR[X I B_i]$. Proposition 3.7,4. We have: VAR[aX + b] = a²VAR[X].

100 3 Random Variables Proof. By definition, VAR[aX + b] ---- E[((aX + b) - (aE[X] + b))²] = a²E[(X - E[X])²] = a²VAR[X]. [] Proposition 3.7.5. We have: VAR[X] = $E[X^2] - (E[X])^2$. Proof. We deduce from the definition of VAR[X] that VAR[XI = E[X z] - 2E[XE[X]] E[(E[X]) z] = E[X z]- 2E[XIE[X] (E[X]) z = E[X z]- (E[X]) z (since the fact that EIX] is a constant implies that E[g(E[X])I = g(E[X]) for any function g). [] Example 3.7.4 Let X be a random variable having a Bernoulli distribution with parameter p. We have: E[X hI= lnp+Onq = p for anyn >_ 1. It follows that VAR[X] = $E[X^2] - (E[X])^2$ = p -

$p2 = p(1 - p) = pq$. (3.60) Example 3.7.5 Let X --- U[a, b]. Then, by symmetry, we can write that the mean of X is given by (a + b)/2. In addition, we have: b 1 b 3 - a 3 b 2 + ab + a 2 EIX z] = x z dx -- -- -- b - a 3(b - a) 3 Making use of Proposition 3.7.5, we find that (b - a) 2 VAR[X] -- -- (3.61) 12 Table 3.5 gives the mean and the variance of the various probability distributions defined in Sections 3.4 and 3.5. Definition 3.7.3. a) The kth moment (or order k moment) of X about the origin is defined by #k = E[Xk]for k = 0, 1,2, ... b) The kth central moment or kth moment about the mean of X is given by #k = E[(X - E[X])k]for k = O, 1 ..... Particular cases. The mean is the first moment of X about the origin, while the variance is the second central moment of X. Moreover, # = #0 = 1 and # : 0. Two other quantities, in addition to the mean and the variance, that characterize the distribution of a random variable are the skewness and the kurtosis coefficients.

3.7 Mathematical Expectation and Variance Table 3.5. Means and variances of the main probability distributions. 101 Distribution Parameters Mean Variance Bernoulli Binomial Hypergeometric Geometric Pascal Poisson Uniform Exponential Laplace Gamma Weibull Gaussian Lognormal P n and p N, n and m P r and p [a, b] oe and)v F, and fl # and r 2 # and r 2 P np m n. 1/p r/p (a + b)/2 1/. 0 ,/. F +F(I +-I) e#+�c 2 pq npq m N-n q/p2 rq/p 2 (b - a)2/12 1/. 2 2/)v 2 oe/), 2 a 2 [F(I +2 -I)- F2fI cr 2 e2#+a2(ea2 -- 1) Definition 3.7,4. a) The skewness, ill, of a random variable X is the quantity (with- out units of measure)/z3/cr 3 . b) The kurtosis of X is given by/32 -- #4/o "4. Remark. In the case of the Gaussian distribution, we find that/3i = 0 (since it is symmetrical) and/32 = 3. To complete this section, we give two results that enable us to calculate upper bounds, in terms of the mean and the variance of the random variable X, for the probabilities of events defined with respect to Sx. We can of course use these results to obtain lower bounds for the probabilities of the complements of the events in question. Theorem 3.7.1. (Markov's inequality) 7 Let X be a random variable that can take only non-negative values. Then, we have: E[X] P[X_>a]_<-- 'a >0, (3,62) a 7 Andrei Andreyevich Markov, 1856-1922, was born and died in Russia. He was a profes- sor at St. Petersburg University. His first works were on number theory and mathematical analysis. He proved the central limit theorem under quite general conditions. His study of what is now called Markov chains initiated the theory of stochastic processes. He was also interested in poetry.

Proof. Suppose that X is a discrete random variable. Then, we can write that x k <a Xk>_a Xk>_a > Z apx(xk) =aP[X > a]. xk >a The proof is similar in the case when X is a random variable of continuous type, or of mixed type. [] The following inequality can in fact be considered as a corollary to Markov's inequality. Theorem 3.7.2. (Bienaym-Chebyshev's inequality) 8'9 Let X be a random vari- able whose mean El X] = t and whose variance VAR[X] = 0 .2 exis't. Then for any a > O, we have." ,) 0.- PIIX - ul >- al < -2" (3.63) Proof. Since (X - u) 2 is a non-negative random variable, we deduce from Markov's inequality that P[(X -)2 >_ a2l <_ EI(X - u)2l VARIXI :=> PIIX - ul al -- a 2 a 2 Example 3.7.6 Let X Exp(,k). We have already found that E[X] = l/,k. We can also show that VARI X1 = l/,k 2. It follows that .-2 1 P[]X-,-] _> k] _< k2 --k2X2 fork=l,2 ..... Suppose that . = 1 and k = 2. Then, we have: P[ ]X - 1] > 2] _< 1/4. Now, since we assumed that the distribution of X is known, we can calculate directly 8 lr6n6e-Jules Bienaym6, 1796-1878, was born and died in France. He studied at the lcole Polytechnique de Paris and taught in a military academy for two years. Next, he became Inspector General of Finances. In 1848, he was named professor of probability at the Sor- bonne. A friend of Chebyshev, he translated his works from Russian into French. The work of Laplace, in particular the Thorie analytique des probabilit4s, had a great influence on his scientific thinking. 9 Pafnuty Lvovich Chebyshev, 1821-1894, was born and died in Russia. Son of a Russian nobleman, he was a professor at St. Petersburg University. He is considered the founder of the great Russian school of mathematics of the twentieth century. In addition to his contributions to number theory and to the theory of approximation of functions, he was also very interested in mechanics. By using the inequality that bears his name, he gave a simple proof of the law aflarge numbers. Moreover, he worked intensively on the central limit theorem.

3.8 Transforms 103 P[IX - 11 > 21 in=c. P[X > 31 + P[X < -11 = P[X > 31 = e -3 _ 0.0498. We notice that the value of the upper bound obtained with Bienaym6-Chebyshev's inequality is about five times greater than the exact value. However, we must re- member that the bound in question is valid for any random variable whose mean and variance are both equal to 1. 3.8 Transforms Definition 3.8.1. The characteristic function of the random variable X is defined by (bx (co) = E[eJ�X], (3.64) where j := and co E]R. Continuous case. In this case, we can write that X (co) = eJ�x fx (x) dx. (3.65) Thus, q5 x (co) is the Fourier transform m of the density function fx (x).

It follows that $f_X(x) = \int e^{-j\omega x} \phi_X(\omega)\, d\omega$. (3.66) Hence, we can state that the characteristic function of a random variable character- izes entirely the distribution of this variable. Discrete case. By definition, we have: $\phi_X(\omega) = \sum e^{j\omega x_k} p_X(x)$. k=1 In the particular case when the x's are integers, we can write that $\phi_X(\omega) = \sum e^{j\omega n} p_X(n)$" (3.67) (3.68) 10 Jean Baptiste Joseph (Baron) Fourier, 1768-1830, was born and died in France. He first studied for the priesthood, but he did not take his vows. Later, he taught at the CollÈge de France and then at the ÉcoIe Polytechnique. He took part in the French Revolution and be- came friends with Napoléon. In 1822, he published his main work, the ThÉorie analytique de la chaleur, in which he made wide use of the series that now bear his name. However, he is not the inventor of these series.

104 3 Random Variables Thus, the right-hand member of the preceding equation can be considered as the complex Fourier series of the function $\phi_X(\omega)$ (which is periodic with period 2π because e $2\pi n j = 1$ $\forall n \in \mathbb{Z}$). Then, we have: $\frac{1}{2\pi} \int_{0}^{2\pi} p_X(n) = e^{-j\omega n} \phi_X(\omega)\, d\omega$ for n =0,1,2 ..... (3.69) Proposition 3.8.1. Let X be a random variable for which all the moments about the origin exist and are finite. Then the order n moment of X, about the origin, is given by El $X^n] = (-j)^n \phi_X^{(n)}(\omega) \big|_{\omega=0}$" (3.70) Proof. We simply have to use the series expansion of $e^{j\omega X}$: $\phi_X(\omega) = E[e^{j\omega X}] = E[1 + j\omega X + (j\omega X)^2/2 + ...]$. Then, if we assume that we can interchange the summation and the mathematical expectation, we obtain: $\phi_X(\omega) = 1 + j\omega E[X] + (j\omega)^2 E[X^2] + ...$ Differentiating n times the two members of the above equation, and evaluating at $\omega = 0$, we obtain Equation (3.70). For instance, we have: $\phi_X(0) = 1$ EIX $^0] = x(0)$ Thus, the formula is valid for n = 0. Note that $\phi_X(0) = 1$ for any random viable X, which is also a consequence of the definition $\phi_X(\omega) = E[e^{j\omega x}]$. Similarly, for n = 1, we have: $\phi_X'(\omega) \big|_{\omega=0}$ $jZ - = jE[X] + (2\omega)E[X^2] + . (3\omega^2)E[X^3] + ...$ $jE[X]$ d =0 etc. [] Example 3.8.1 Let X Poi(c). We have: , $e^{j\omega k} e^{-c} \frac{c^k}{k!}$ el = e^{-c} k=0 k=0 It follows that $\_ e^- \exp(e^{j\omega})c)$.

3.8 Transforms 105 $E[X] = -j \frac{d}{d\omega} - \exp(e^{j\omega})c)\big|_{\omega=0} = -je^- \exp (e^{j\omega} \omega t)e^{j\omega}$ $\omega tj \big|_{\omega=0} = $ ' which corresponds to the result obtained in Example 3.7.1. Likewise, we find that $E[X^2] = c^2 q$-. Then, making use of Proposition 3.7.5, we obtain: VAR$[X] = E[X^2] - (E[X])^2 = 2 q_ _ ()^2 = $. So, in the case of the Poisson distribution, the parameter is both the mean and the variance of the distribution. A similar result is obtained with the exponential distribution, for

which the mean and the standard deviation are both equal to 1/.. Example 3.8,2 Let X N(#, o'2). Then, 1 e j�x exp dx {[x-(Jr-�o'2+#)12-2jr-�o'2#+(r-�o'2)2ld x -- v/-o" exp - -2- = exp < N(# + joo" 2, o'2) < = exp Hence, we can write that (jobt--o2o'2) x P[-- (E[X n] = (-j)n-7 exp ja# - g.oZo" 2 We find, in particular, that E[X] = -j exp (jo# - �o2o'2) (j# - oo'2) o=o = # and E[X2] = -exp (jr.o# - r.o2o'2) [(j# - r.oo'2)2 - o'2] w=o = #2 + o'2, so that VAR[X] = o'2. Note that since we have used the symbols #x and o'x 2 to denote the mean and the variance of the random variable X, in general, then, to be coherent, in the case of a Gaussian distribution with parameters # and o -2 the constants # and o -2 had to represent the mean and the variance of this Gaussian distribution, respectively. Remark. We have written above that the integral in the computation of the character- istic function of X N(#, o'z) can be expressed in terms of the probability that a

106 3 Random Variables random variable Y N(# + jcoo. 2, o "2) takes a value in the interval (-x, x). Of course, a Gaussian distribution cannot have a complex mean, since we only consider real variables. However, from a purely mathematical point of view, the fact that the constant # + jcoo. 2 is complex does not change anything in the computation of the integral. Now, when X is a discrete random variable whose possible values are positive integers or zero, we work with its probability generating function. Definition 3.8.2. Let X be a random variable taking its values in {0, l .... }. The (probability) generating function of X is defined by Gx(z) = E[z x] = z px(k). k:0 (3.71) Remarks. i) The generating function of X is in fact the z-transform (with a different sign in the exponent) of the sequence Po, Pl ..... where Pk := px(k) Yk. ii) Since po + pl + P2 a t- .... 1, the radius of convergence of the power series is greater than or equal to 1. iii) We can write that qSx(co) = Gx(eJ�). It follows that d ' ,,=o " ' E[XI = -.j-cox(ea") = -jG'x(ea�')e.l'j = G�(I). (3.72) Likewise, we find that E[X 2] = G(I) + G�(I). Hence, we have: VARIX] = G%(I)+ G�(I)-[G�(I)I 2. (3.73) For example, let X be a discrete random variable whose probability mass function is given by the following table: We have: Then, x 0 1 3 E px(x) 1/2 1/3 1/6 1 o 1 1 3 1 z 3 z 1 x(z)=z z 2 1 := 5 = + = which corresponds to the result obtained E[X] = 0 + 1 by using the definition of E[X]: 1 1 5

3.8 Transforms 107 iv) Proceeding as in Proposition 3.8.1, we show that 1 d k Gx(z) z o" px (k) -- k! dz Indeed, Gx(z) := E[z x] = px(O) + zpx(1) + zepx(2) +

... , which implies that Gx (0) = px (0), G(0) = px(1) + 2zpx(2) +... [z=0 = px(1), G:(0) = 2px(2) + 6zpx(3) +... [z=0 = 2px(2), etc. We can also express px (k) in terms of a complex integral as follows: 1 flz Gx(z) dz, wherer < 1 px(k) = j l =r zk4_-' - � (3.74) Example 3.8.3 Suppose that pz Gx(z)-- 1-qz' where[z[< 1/q. We have: Px (0) = Gx (0) = 0. Moreover, since we can write that Gx(z) = pz(1 + qz + (qz) + ...), we find that 1 = q-ip ..... px(k) = -. p(k!q -I) fork = 1 2, That is, the function Gx (z) above is the generating function of the geometric distri- bution with parameter p. We also deduce from the function Gx that 1 1-p E[X]=- and VAR[X]---- p p2. Finally, another function that enables us to obtain the moments of a random vari- able is the moment generating function. Definition 3.8.3. Let X be a random variable. The moment generating function of X is defined by Mx(s) = E[e-SX], (3.75) if the mathematical expectation exists.

108 3 Random Variables Remarks. i) If X is a non-negative continuous random variable, then Mx(s) is the Laplace transform of the function fx (x). ii) We have: qx (o) = Mx(-jw). It follows that dn qSx(o) w 0 dn dn dw n = = ff' n Mx(- jw) w=0 ---- (- J)n-- n Mx(w) w=o" Then, we deduce from Proposition 3.8. I that E[X n] ---- (-I) n-dn (s) s=O" ds n Mx (3.76) In fact, if we define the moment generating function of X by E[e sx] (as many authors do), then the term (- I)" disappears in the previous formula. Example 3.8.4 If X ExpO.), then we find that Mx(s) = e-SX;ke - dx -- -- ifs > -.. .+s " -(s+.)x --e ,k +s o This formula enables us to check that E[X] = I/,k and VAR[X] = I/. 2. Note that the function Mx(s) does not exist (that is, the integral diverges) ifs < -., whereas the characteristic function of X, qx (o), exists for any real value o. 3.9 Reliability Definition 3.9.1. Let T be a non-negative continuous random variable that repre- sents the lifetime of a system (or a machine, etc.). The reliability function R(t) is defined by R(t) = PIT > t] (= I - Fr(t)). (3.77) The mean lifetime of the system, E|T], can be obtained as follows (see Propo- sition 3.7.2, p. 96): E[T] = R(t) dt. (3.78) Remark. The symbol MTTF (mean time to failure) is sometimes used for the quantity E[T]. We can also define the symbols MTBF and MTTR, that is, the mean time between failures and the mean time to repair, respectively. We have: MTBF = MTTF + MTTR. Definition 3.9.2. The failure rate function of the system is denoted by r(t) and is given by r(t) = fr(t [T > t). (3.79)

3.9 Reliability 109 Remark. The function r(t), multiplied by dt, is

approximately equal to the prob- ability that a machine, that is still working at time t, breaks down in the interval (t, t + dt], because fr(t IT > t)dt -- P[t < T < t + dt] T > t]. Now, the probability that the lifetime of the system is smaller than or equal to s, given that the system is still functioning at time t, is P[T <s I T >t]:=FT(s 0 ifs <t, T > t) -= FT(S) -- FT(t) ifs > t. 1 -- FT(t) It follows that d fT(s) fT(s I T > t) = -FT(S I T > t) -- ifs > t. (3.80) 1 FT(t) Thus, since fT(t) = - -R'(t), we can also write (taking the limit as s decreases to t) that fT(t) R'(t) r(t) -- -- (3.81) 1 - FT(t) R(t) Example 3.9.1 If X Exp(.), then the failure rate function is constant: .e-)t r(t) -- --). e-,kt Proposition 3.9.1. There is a bijective relationship between the functions R(t) and r(t): R(t) = exp - r(s)ds . (3.82) Proof. We have: fot fot R'(s) r(s)ds = - R(s) ds = -lnR(t) + In R(0). Now, R(0) = l, because T is a non-negative continuous random variable. Hence, Equation (3.82) follows at once. [] Remark. We then deduce from Example 3.9.1 that the exponential distribution is the only one for which the failure rate function is constant.

! 10 3 Random Variables Reliability of Systems in Series and in Parallel a) Systems in series. If a system is made up of n subsystems placed in series and operating independently from one another, then the reliability function of the system is given by R(t)= Rk(t)=exp - [rl(s)+.--+r,,(s)]ds , k=l (3.83) where Rk(t) is the reliability function of the subsystem k, for k = 1 ..... n. This result follows from the fact that T > t if and only if Tk > t for all values of k, where Tk is the lifetime of the subsystem k. Example 3.9.2 If T has an Exp()k) distribution Vk, then R(t) = exp[-() + --- +),,)t]. Thus, it is as if there were a single component, whose lifetime has an exponential distribution with parameter) := b) Systems in parallel. In the case when the system is made up of n subsystems placed in parallel and operating independently from one another, we have: T _< t ◆# Tk _< t 'k. It follows that the reliability function of the system is given by R(t) = 1 - I-Ill - Rk(t)]. k=l (3.84) Remarks. i) We assume above that all the components begin to function at time t = 0. This is called active redundancy. Suppose now that there are n components placed in parallel, but only component no. 1 functions at first. When it fails, component no. 2 takes over, and so on. This type of redundancy is called standby (or passive) redundam3'. If the components operate independently from one another, we can show (see Chapter 4) that E[Tl = EIT] k=l and VARIT] ind. -= VARITk]. k=l In the particular case when T has an Exp(.) distribution Yk, we can also show (see Chapter 4, again) that T has a gamma distribution with

parameters n and .. It follows that n-I R(t) :=.PIT > t] = P[Poi()t) _< n - 1] = Ze-Ztot)/k! k=0 fort > 0.

3.10 Exercises, Problems, and Multiple Choice Questions 111 ii) Suppose next that a system is made up of n components, but that k (working) components are sufficient for the system to function, where 0 < k < n. Then, if all the components have the same reliability function R (t) and are all independent, the reliability function of the system is given by R(t) = [R(t)]i[1 - R(t)] n-i = 1 - , [R(t)]i[1 - R(t)] n-i. i=k i=0 iii) Finally, we can of course consider systems made up of a number of subsystems placed in series and others placed in parallel. For instance, if there are two subsys- tems in parallel followed by another subsystem placed in series, then we can write that R(t) = {1 - [1 - Rl(t)][1 - R2(t)]}R3(t). 3.10 Exercises, Problems, and Multiple Choice Questions Solved Exercises Exercise no. 1 (3.3) Let X be the execution time (in minutes) of a program submitted to a central computer. When a program contains at least one "fatal error," it is not executed (exe- cution time is equal to zero). Moreover, there is a one-minute time limit imposed on the execution time (the execution will stop at the end of one minute, even if it is not finished). Finally, we suppose that the distribution function of X, which is a random variable of mixed type, is 0 ifx <0, 0.2 if x = 0, Fx (x) = (3x + c)/5 if0 < x < 1, 1 ifx> 1. Calculate a) the constant c; b) the probability P [{X = 0} U {X = 1}]; c) the function fx (x I 0 < X < 1). Solution a) Since Fx is right-continuous, we can write that 3.0+c P[X <0] =0.2-- -- :=c= 1. 5 b) P [{X = 0/u {X = l/] : P [X : 0] + P [X = 1] --a) 0'2+[1 3"1+1 35 :0.2+0.2:0.4.

3 Random Variables d fx(x 10< X <1)=rFx(x 10< X < 1) a--[Fx (x) - Fx(0)] dx P[0<X< 1] a) 3/5 0.8 --0.2 if0 <x < 1 It follows that fx(xlO<X< 1)= lif0<x< 1, 0 elsewhere. Exercise no. 2 (3.3) Let Fx(x) = 0 ifx < 0, l/4if0 <x < 1, 1/2 if 1 <x<4, c ifx > 4. a) Find the constant c. Justify your answer. b) Calculate P[I < X < 5]. c) Find px(l) + px(2). d) Obtain fx(x I X < 1). Solution a) Since limx- Fx (x) = 1, we deduce that c = 1. b) P[l <X <5] Fx(5-) Fx(I-) t al t 3 _ = -- =c-- =l--=. c) We calculate 1 1 1 px(I)=Fx(l)-Fx(l-)-- 2 4-- 4 and 1 1 px (2) = Fx (2) - Fx (2)-- --0. 2 2 I Thus, Px (1) + Px (2) = ,. d) We have: x 0 1 4 Px (x) 4 2 X Px (x [X < 1) o Z I I 1 1 fx(xlXs1)=:(x)+z(x-) VxeR. Z Z

3.10 Exercises, Problems, and Multiple Choice Questions 113 Exercise no. 3 (3.6) A well-balanced (or non-biased) die is tossed until a "6" is obtained. a)

Given that the first three tosses produced a "1 ," what is the probability that the die will have to be tossed more than five times in all? b) Let X be the number of tosses required to obtain a first "6." We define Y = min {X, 5}. That is, Y is the minimum between X and the number 5. Calculate Pr (Y). Solution a) We seek P IX > 5 I X > 3], where X Geom(p = 1/6). We have: P[X >5 IX >3]=P[X>2]= = 3--" b) We can write that P[Y=yl=P[X=y] ify = 1,2,3,4 and Thus, we have: P[Y=5]=P[X>5]= P[X>4]= . y 1 2 3 4 5 PY(Y) (_)()(_)2 () (_)3 ()(_)4 Exercise no. 4 (3.7) Let fx(x) = { 2xe-X2 ifx > 0, 0 elsewhere. a) We define Y = 2 In X. Calculate fr (y). b) Let Z := 1/X. Calculate E [Z]. Solution a) deY/2 fv {y) = fx {x) = exp {y --e y } =2(eY/2) e-(eY/)2 eY/2 =eYe -ey ify R, since 21nO = -oc and 21nee =

114 3 Random Variables b) E [Z] = E = 2xe-X2dx = 2,f- e- dx = 2.v/ (1/2)I/2 fo� "v/'- (11/2)i/2 e 2(2/ dx Exercise no. 5 (3.7) Let X be the time (in hours) a computer terminal is used during one workday. a) Suppose that X has (approximately) a Gaussian N(4, 2) distribution. Find the number xo such that, for 90% of the days, X is greater than xo. Hint. We have: Q- (0.10) - 1.28. b) Suppose that X U[0, 8]. Calculate i) the probability density function of Y := e x and ii) the mathematical expectation of y2. Solution a) We want b) i) We calculate dlnv l +] l ife 0 <y<e 8 4* fY(Y)= {(8Y)o-' elsewhere.ifl-<y-<eS' ii) We have: f8 E[y2] = E[e2X] = e2X ldx do 8 e2 x 8 e 16- 1 = T6- o - Or f Y I e 16- 1 = e8 dy = -- i E[y2] y2 y2 e

3.10 Exercises, Problems, and Multiple Choice Questions 115 Exercise no. 6 (3.7) Four calls are received at random, according to a uniform distribution, at a central dispatch location during a one-hour period. What is the mathematical expectation of the number of calls received during the first half hour of the hour considered, knowing that at least one of the four calls was received during these first 30 minutes? Solution Let X be the number of calls received during the first half hour. Then, we can write that X B(n = 4, p = 1/2). We seek E [X I X > 1]. We have: P[X > 1]= I- P[X =O]= I- -- 16 It follows that E[X I X >_ I] 4 16 4 k=l k P [X k l X >- 1] = k=l 1 16 k = 1] {1 � 4+2 � 6+3 � 4+4 � 1} = -- _2.13. k=l 15 Exercise no. 7 (3.7) Let X be the time (in minutes) required by a technician to assemble a certain part. a) We define Y = e x . Calculate the probability density function of Y if X U(0, 1). b) Suppose that X Exp(1). Calculate P IX 2 > 4 I X 2 > 1]. c) Suppose that X N(1, crx2). i) Calculate P [IX -- II < 1] if crx2 = 0.25. ii) Find crx2 if P [X < 2] = 0.99865. Hint. Q(1) -

1.59 ⬦ 10 -1, Q(2) - 2.28 ⬦ 10 -2 , Q(3) - 1.35 ⬦ 10 -3 and Q (4) - 3.17 x 10 -5 . Solution dv dlny =-1 ifl <y <e (=0 a) fr(Y) = fx(x) __ = 1 dy y elsewhere). P[X 2 >4IX 2 > 1] =P[X >2IX > 1], becauseX>0 b) = P [X > 2-- 1] = e -l(l = e -l -- 0.3679. c) i)P[lS-ll < 1] =P[-1 <X-1 < 1]=P --- < Zo < where Z0 N(0, 1) 0.5

116 3 Random Variables = qb (2) -- qb (--2) = 2qb (2) -- 1 = 2[1 -- Q (2)1 - 1 - 2(0.9772) - 1 = 0.9544. ii) P[X <2]=qb(-x)2--1 =0.99865= Q(-)=l.35x_ ,.- 10 .3 1 1 1 ox 9 Exercise no. 8 (3.7) Let 6x fx(x) -- forx > 0 (=0elsewhere). (1 "hi- X) 4 a) Calculate the probability density function of Y := I/X. b) Use part a) to compute E IX]. c) Calculate fx (x I X < 1). d) Calculate the median of X. Solution d(l/y) _ 6) ,- -2 -- 6y if v > O. a) fy(y)=fx(l/y) T (lnt-V-I) 4 . (l+y) 4 " b) We have: [] , 6x (,+x,-3? EIX]a--)--E[YI=E .... 4dx=6 =2. x (+x) --- fo' 6x dx=6x(l+x)-3 i) fo'-3 c) P[X< 1]= (1 + x)- + 2(l+x)-3dx 2 (1 +x) -2 n 1 1 1 o = -2--y + 2 -2 4 2 2 + 1 2 0 ifx <0, := fx (x I X < 1) = 12x (1 + x)-4 if 0 < x < 1, 0 ifx > 1. d) We deduce from part c) that Fx (1) = 1/2. It follows that x0.5 = Xm = 1. Exercise no. 9 (3.7) Let X be the service time (in minutes) at a counter. We suppose that E [X] = 3. a) Give a lower bound (> 0) for P [X < 6]. Justify your answer. b) Suppose that X G(ot = 3, . = 1). Calculate P [X > 5 I X >_ 3]. c) Suppose that X N(3, 1). Calculate i) the 90th percentile of Y := 2X - 4; ii) P[W=21 W 2=4],whereW:=X-3.

3.10 Exercises, Problems, and Multiple Choice Questions Hint. We have: Q (1.2815) 0.1 and Q (2.3263) 0.01. a) We have, by Markov's inequality: Solution E [X] 1 1 1 P[X>_6] <_ ---- P[X <6]_> 1 ..... 6 2 2 2 P [X > 5] b) P[X >5[X_>3]-- P [X >_ 3] We have: and { 52] P[X<_5]=P[Poi(1.5)>_3]=l-e -5 1 +5 + . { 32] P[X_<3]= P[Poi(3)_>3]= 1-e -3 1+3+. . 117 It follows that cont. 18.5e -5 -2 18.5 P[X >5[X_>3] -- ----e --0.2946. 8.5e -3 8.5 c) i)YN(2x3-4,4x 1)=N(2,4) (Yo.9 - 2) (72) P[Y<Y⬦9]=0"9"⬦* -2- =0.9.⬦* Q yo.__ =0.1 .⬦* yo.9 - 2 1.2815 .⬦, yo.9 - 4.563. 2 ii) W N(0, 1) P[W = 2 [W 2 4] = P [W = 2 [{W = 2} U {W = -2}] = 1/2, by symmetry. Exercise no. 10 (3.7) We suppose that the length Tj (in minutes) of the telephone calls received during the day at a certain emergency number has a U(0, 4] distribution, whereas the length TN (in minutes) of those received during the night at the same emergency number has (approximately) an Exp(1) distribution. Moreover, 80% of the calls are received during the day (and 20% during the night). a) Calculate the probability that a given call lasts more than one minute.

b) What is the average length of the calls received during the night and lasting more than one minute? c) Calculate P[Tj = 1 [{Tj = 1} U {Ty = 2}].

118 3 Random Variables Solution a) Let T be the length of a given call. We can write that P [T > 1] = P [Tj > l] (0.8) -1- P [TN > l] (0.2) = (3/4) (0.8) + e -) (0.2) = 0.6 + (0.2) e - " 0.6736. b) We seek E I TN I TN > 11= I + E I TN] , by the memoryless property of Tu = 1 + 1 = 2 minutes. c) By symmetry (since .l;rj (1) = .l;rj (2) = 1/4), we can write that 1 P ITj = 11Tj = } u lTj =2}1= -. 2 Exercise no. 11 (3.7) The input to a communication channel is a certain voltage v and the output is a voltage X = v + B, where B has a standard Gaussian distribution. We use this communication channel to send binary information as follows: to send a 0, we input a voltage v = -2, while to send a 1, we input a voltage v ---= +2. The receiver decides that a 0 has been sent if the voltage X he receives is negative, and that a 1 has been sent if X is positive or zero. a) Calculate P[X = v]. b) Let Y := e x. Obtain the probability density function of Y. c) Calculate i) VAR[X] and ii) VAR[X] B = 0]. d) Find the probability that the receiver makes a detection error, if the probability of sending a 0 is equal to p. Solution a) P[X = v] = P[v + B = v] = P[B -= 0] = 0, because B is a continuous random variable. b) Y:=e x,whereX= v+BN(v,l)=y >0and fy(y)=fx(lny) dlny 1 [ ] 1 +. dy = exp - (lny-v) 2 . fory6 c) i) VAR[X] = VAR[v + B] = VARIB] = I. ii) VAR[X I B = 0] = VAR[v + 0 I B = 0] = VAR[v] = 0. d) Let F = the receiver makes a detection error, and Ek = a k is sent, for k = 0, I. We have: P[F] = P[F I Eo]P[Eo] + P[F I E]P[E]

3.10 Exercises, Problems, and Multiple Choice Questions 119 = P[-2+B >_0]p+P[2+B <0](l-p) = P[N(0, 1) > 2]p + P[N(0, 1) < -2](1 - p) Tab. 3.3 = Q(2)p + qb(-2)(1 - p) = 1 - qb(2) - 1 -- 0.9772 = 0.0228. Remark. For this last question, we could consider the constant v as a random variable V that takes on the value -2 (respectively +2) with probability p (resp. 1 - p) and write that P[F]:P[X>_O[V:-2]p+P[X <0[V:+2](1-p). However, this formula will only be given in Chapter 4 (on random vectors). Exercise no. 12 (3.8) Let X be a discrete random variable whose set of possible values is Sx : {0, 1,2 .... } and for which Mx (s) : exp {a (e -s - 1)}. Calculate a) Px (0); b) the standard deviation of 2X. Solution a) We have: 1t/= We find (see Equation (3.74) and also Example 3.8.1) that X Poi(), so that Px (0) = e -. b) Since X Poi(a), we may write that STD [2X] = (VAR [2X]) /2 : (4a) /2 : 2. Exercise no. 13 (3.8) Let x 0 1 3 px (x) 1/8 3/8 1/2 Use the generating function of X to calculate its

standard deviation. We have: Then, Solution o 1 3 1 1 3 z 3 Gx(z) := E[zX]= z g + zlg + z32 -- 8 + gz + . VAR[X] = G(1) + G%(1) - (G�(1)) 2 = (3Z)lz:l+ (+Z2) z=l-- (+) 2

120 3 Random Variables = 3 4- - -- _ 1.359. It follows that STD[XI -- 1.166. Exercise no. 14 (3.8) The probability mass function of the discrete random variable X is given by the following table: .r - 1 0 1/2 1 Px (x) 1/8 1/8 1/4 1/2 a) Calculate P[X b) Let Y := X 2. Calculate i) p,(y) and ii) the standard deviation of Y. c) Calculate the characteristic function of X at (o = rr. Simplify your answer as much as possible. d) Can we calculate the generating function of X? Justify your answer. Solution a) We have: P [X = 3/4] = 0, because 3/4 Sx. Moreover, Fx (3/4)= P[X <_ 3/4] = P[X <_ 1/2] = 1/2. Thus, P [X = 3/4] + Fx (3/4) = 1/2. b) i) We have: 1 1 I 5 PY (-') 8 4 g 1 ii) We calculate 1 1 5 11 E[Y]=0+x+I x8 16 Then, and] 1 1 5 41 e ,2 =0+ IV � 2+1 � g = 6 41 fll'] 2 43 '-J 0.4098. VAR[Y] -- 64 \16] -- 256 and STD[YI- 1 c) qbx(co) "= E[e j')x] =e -j') 1 1 1 � 1 � 4- I � 4- e jw/2 � -4 + eS�" -2 �bx(Tr)=- e -j +1+ 2e j/2 +4e j = {-1+1+2j-4} j 2 4

3.10 Exercises, Problems, and Multiple Choice Questions 121 d) We cannot calculate Gx (z), because this function is only defined when the ele- ments of Sx are all positive integers or zero, which is not the case here. Remark. We can calculate E[zX]. However, we cannot call this mathematical expec- tation the generating function of the random variable X. Exercise no. 15 (3.8) Let X N(0, 1). We define Y = e x2 and Z = 2X- 1. a) Obtain fy (y). b) Calculate E [1 / Y]. c) i) Obtain the characteristic function of Z. ii) Use bz (co) to calculate E [Z2]. Hint. We have: bx (co) = e -�2/2. Solution a) We have: y : e x2 <: x : � 1V' , where y _> 1. It follows that 1 1 C2 exp { (lw/)2 }] = 2. 2y 2 -- C2zr ClnY ,y .,=__exp - -- y-3/2(lny)-l/2 fory>_1. f_ 1 _x2/2dx f_x e_3X2/2d x b)E[1/Y] = E[e -x2] = e-X2x/e = f_': { x2 } IN(O,) = 1 exp - dx = P 2(1/3) = 3 -1/2 0.5774. c) i) bZ(co) := E[e y�z] = E[e jc�(2X-l)] =- e-J�)E[e j2�)X] = e- J�)qx(2co) = e-JO)e-(2o))2/2 = e-J o)-2o)2. ii) We can write that " d {e-JCo- 2o2(_j_4co)} co= 0 E[Z2] =-;bz (co) It�=0 -- dco = -e -jc�-zc�z [(-j - 4co) 2 - 4] co=o = -e�[(-J)2 - 4] = 5.

122 3 Random Variables Unsolved Problems Problem no. 1 Let 0 ifx < -1, 1/3 if-I _<x <0, Fx(x)= 3/4 if0_<x <2, 1 if2 _< x. Calculate a) Px (0) + Px (1); b) Fx (0] X > -1). Problem no. 2 Let 1/2 if()<x < 1, fx(x)= l/(2x) ifl _<x <e.

Calculate a) P IX < 1 I x < 21; b) ./'x (x I 1 < X < e). Problem no. 3 Let X be a random variable having a B(n = 2, p = 1/4) distribution. Calculate a) P[X >_ 1 I X _< ll:b) glX I X Problem no. 4 Let X Exp(X = 2). a) Calculate P [X 2 < 9]. b) Use Bienaym6-Chebyshev's to compute P [X - ◆ > 2]. Compare with the exact value. inequality Problem no. 5 Let X G(o = 3, X = 1/2). Calculate a) P [X < 5[: b) the standard deviation of Y := 2X -4. Problem no. 6 A machine produces parts whose diameter X (in centimeters) has approximately a Gaussian N(tt, c 2 = (0.01)2) distribution. a) What should the value oftt be so that no more than 1% of the parts have a diameter greater than 3 cm? b) Suppose that tt = 3. Calculate the probability density function of Y := IX - 3]. Problem no. 7 Let X be a discrete random variable such that px (0) = e -;, where X > 0, and px(x)-- -- forx ..... -2,-1, 1,2 ..... 2(Ixl!) Calculate a) E [IXI]; b) EIX I X > 0] if) = 1. Problem no. 8 Let I,,x/2 ifx <0, fx (x) = e_X/2 - ifx >0.

3.10 Exercises, Problems, and Multiple Choice Questions Calculate a) E [X2]; b) the characteristic function Problem no. 9 Let X U[-1, 1]. Calculate a) the kurtosis of the random variable X; b) VAR[X IX >0]. Problem no. 10 Let 123 Fx (x) = 0 ifx < -1, 1/8 if-l_ <x < 1, 3/8 ifl <x <2, 1 if2 > x. a) Calculate the probability density function of X. b) Let A := {X > 1 }. Calculate Fx (x I A). Problem no. 11 Let Fx (x) = a) Calculate P [- 1 < X < 1.5]. 0 ifx < -1, 1/2 ifx = -1, (x + 4)/6 if -1 <x <2, 1 if2 >x. b) Let A :---- {X > 1}. Calculate fx (x I A). Problem no. 12 The number of power failures that occur in a certain region during a one-year period has a Poisson distribution with parameter ct = 3. The length (in hours) of a power failure has an exponential distribution with parameter . = 1/2. a) Calculate the probability that a power failure that has already lasted two hours will end within the next 30 minutes. b) Assuming independence of the power failures from year to year, calculate the probability that, during the next ten years, there will be at least one year during which exactly one power failure will occur. Problem no. 13 Let X N (0, 1). We have: bx (co) = e -◆2/2. Calculate VAR IX2]. Problem no. 14 The time T (in months) that elapses between two car accidents at a certain in- tersection has an exponential distribution with parameter . = 1/4. The number of persons involved in an accident has a geometric distribution with parameter p = 1/3. a) Calculate the probability that a given accident involves at least four persons, know- ing that more than two persons are involved. b) Let Ak = the time elapsed between the (k - 1)st and kth accidents at this inter- section is greater than one year, for k = 2, 3,.... We assume that the events Ak are

124 3 Random Variables (globally) independent. What is the probability that, among the events A2 ..... A21, exactly two will occur? Problem no. 15 Let X be the lifetime (in months) of an electronic component. a) Suppose that X has (approximately) a Gaussian N(20, 16) distribution. Calculate P [iX[< 24]. Hint. We have: Q (1) O. 159. b) Suppose that X G(ot : 2, k); that is, fx (x) :)2xe-kX ifx > 0 (= 0 elsewhere). Calculate i) the probability density function of Y :: l/X and ii) the mathematical expectation of Y. Problem no. 16 A box contains three transistors, denoted by A, B and C. Transistor A was made by a machine that produces 3% detectives, transistor B by a machine that produces 5% detectives and transistor C by a machine that produces 7% defectives. Ten tran- sistors are taken, at random and with replacement. Let N be the number of detective transistors obtained. a) Calculate P [N : 0]. b) Does N have a binomial distribution'? If so, give its parameters. Otherwise,justify. c) Calculate the mathematical expectation of N, given that N is less than or equal to Problem no. 17 Let { kx(l-2x) if0 < x < 1/2, fx (x) = elsewhere, where k > 0. a) Find the constant k. b) Calculate the distribution function of X. c) Calculate the variance of X. d) Let Y :: I/X. Calculate i) the probability density function of Y and ii) its math- ematical expectation. Problem no. 18 a) Let fx(x) = K e - x2 ifx ◆, where K > O. i) Find the constant K. ii) Calculate fx(x 1X > 0).

3.10 Exercises, Problems, and Multiple Choice Questions 125 b) Let [2e -2(y- if y > 1, fr (y) = 0 elsewhere. i) Calculate the characteristic function of Y. ii) Use the function calculated in i) to obtain the mean of Y. Problem no. 19 A crate holds two boxes that contain ten items each. All the items were made by the same machine. Suppose that the probability that an item made by this machine is defective is equal to 0.05, independently from one item to another. a) What is the probability that there are less than three defective items among the 20 items considered, knowing that there is at least one defective among these 20 items? b) Use a Poisson distribution to compute (approximately) the probability in a). Problem no. 20 Let X be the waiting time (in minutes) before being served at a certain counter. We suppose that the distribution function of X is given by 0 ifx <0, ifx =0, Fx(x) = 3 1-ifx>0. a) Calculate P [X = 0 I X < 1] b) Calculate the probability density function of X. c) Let Y := X I {X > 0}. Calculate i) fr(y) and ii) E [Y]. Problem no. 21 Let X be a discrete random variable whose probability mass function is given by the following table: x 1 4 9 1 5 1 px (x) g a) Calculate the variance of /-. b) Let Z := X I {X > 1}. Calculate the probability density function of Z. c) i) Calculate the moment

generating function of X. ii) Use the function computed in i) to obtain E [X2]. Problem no. 22 A discrete random variable X has the following probability mass function: x -2 -1 0 1 1 1 1 1 PX (X) g 8 2 a) Calculate Fx (0) + Fx (2).

126 3 Random Variables b) Let Y := -2X. Obtain the standard deviation of Y. c) Let Z := X 2. Find fz (z). d) Calculate E . e) Calculate E IX I X < 0]. Problem no. 23 A small company has five telephone lines. We suppose that the probability that a given line is free at 11:00 a.m. is equal to 0.5 for each line, independently from one another. Moreover, this probability is the same for each workday, that is, from Monday through Friday. a) We consider a period of four weeks with five workdays each. Calculate the proba- bility that over this time period there are at least two days during which not a single telephone line is free at 11:00 a.m. b) Use a Poisson distribution to compute (approximately) the probability in a). c) If a man calls every day, exactly at 11:00 a.m., what is the average number of telephone calls that he will have to make to reach the company, knowing that on his first trial all the lines were busy'? d) During a given five-workday week, the lines were all busy at 11:00 a.m. on exactly two days. What is the probability that these two days were the Monday and Tuesday of that week? Problem no. 24 Let X be the length (in minutes) of a telephone call received by a computer consultant. a) Suppose first that X U[I, 51. Calculate P [X 2 < 9]. b) Suppose next that X G(a = 3, , = 1). Calculate P IX > 4]. c) Suppose finally that X N(3, 1) (approximately). If ten independent calls are received by the consultant during a certain period of time, what is the probability that exactly one of these calls lasts more than three minutes? d) If the characteristic function of X is given by k qSx (o) - 1 - 4jo' what is the value of the constant k? Problem no. 25 Let Px (x) = Ix (x + 1)] -1 ifx = 1, 2 ..... = 0 otherwise. a) Calculate P [2 < X < 4].

3.10 Exercises, Problems, and Multiple Choice Questions 127 b) Let Find Fr (y). c) Calculate E [X [X < 4]. d) Calculate E [X2]. Problem no. 26 1ifX<4, Y:= 0if X_ > 4. A point X is taken in the interval [-1, 1] in such a way that it is twice as likely that X is positive. We suppose that 1/3 if-1 <x<0, fx(x)= 2/3 if0<x < 1, 0 elsewhere. Calculate a) the probability density function fy (y), where Y := X2; b) the variance VAR [-2X + 4]; c) the characteristic function qx (co); d) the conditional probability density function fx (x I X < 0). Problem no. 27 A computer generates random numbers according to an exponential distribution with parameter) ---- 1/3. The numbers generated are independent

of one another. a) What is the probability that each of the first three numbers generated is greater than 9? b) Let N be the number of random numbers greater than 9 among the first 20 numbers generated. Calculate the variance of N. c) Use a Poisson distribution to calculate (approximately) the probability that exactly one of the first ten numbers generated is greater than 9. d) Knowing that the first number generated is greater than 9, what is the probability that it is greater than 12? Problem no. 28 A machine is made up of two types of parts: A and B. The number X of parts of type A that fail during a given period has a Poisson distribution with parameter a = 4. In the case of the type B parts, the number Y of failures has the following probability mass function: pg(k)= fork=0,1,2 ..... The failures of the type A and type B parts are independent. a) Calculate the probability that there is one failure (in all) during the period consid- ered. b) Calculate E [X I X < 3].

128 3 Random Variables c) We consider ten independent periods, each having the same length as the period above. What is the probability that during at least two of these periods there are no failures of the type B parts? Problem no. 29 The time T that a technician spends performing the periodical maintenance of a certain machine is divided into two parts: a fixed period of 5 minutes made up of customary checks, followed by a random period X (in minutes) which depends on the repair work that the technician has to do. a) Suppose that X U[0, 20]. Calculate the probability density function of T. b) Let X G(ot = 5,) = 1/2). Calculate E [T2]. c) Suppose that X N(10, 4). i) Calculate the probability P lIT - 151 > 6]. ii) Compare the answer in i) with the bound given by Bienaym6-Chebyshev's inequality. Hint. We have: Q (1.5) - 0.0668 and Q (3) - 0.00135. Problem no. 30 Let x/kifO <x <k, fx (x) = 0 elsewhere, where k > 0 is a constant. a) Find the constant k. b) Calculate the distribution function of X. c) Calculate the moment generating function of X. d) Let Y := I/X. Calculate VARI/-I. Problem no. 31 An airline company has realized that, on average, 4% of the people making reser- vations on a certain flight do not show up for the flight. Consequently, the policy of this company is to accept 75 reservations for a flight that can only hold 73 passen- gers. What is the probability that there is a seat available for each passenger who arrives for the flight? (Make the necessary assumptions.) Problem no. 32 An exam grade, out of I00 points, can be considered as having a Gaussian N(60, 100) distribution. a) What is the average grade in this class? b) What is the proportion of students whose grades are between 55% and 75%? c) Where should the

threshold for an A be placed, if we want 10% of the students to get an A? Problem no. 33 In information theory, Hartley's formula expresses the quantity I of information of a message as a function of the signal-to-noise ratio S/N as follows:

3.10 Exercises, Problems, and Multiple Choice Questions 129 $2I = 1 + \frac{S}{N}$ Let us write that $X = 2I$. a) Suppose that $S = 1$ and that N has a uniform distribution on the interval (0, 1). Calculate i) the probability density function of X; ii) the expected value of X. b) Suppose that $N = 1$ and that S has a uniform distribution on the interval [1,2]. Calculate the moment generating function of X. Problem no. 34 Let X be the number of different versions of a program that a computer science student has to write to obtain a program that works. We suppose that X has a geo- metric distribution with parameter $p = 1/4$. We consider a class with 20 students, who must each write a program individually. a) Calculate $P[4 < X < 6 \mid X > 2]$. b) Let N be the number of students, among the 20, who have to write exactly two versions of their programs to obtain a version that works. Calculate $P[N < 1]$. c) Use Poisson's approximation to calculate the probability in b). Problem no. 35 In information theory, the differential entropy of a continuous random variable X is defined by $H_\blacklozenge = -E[\ln f_\blacklozenge(X)]$. Moreover, the relative entropy of the continuous random variables X and Y is given by $H(f_\blacklozenge; f_r) := f_\blacklozenge(x) \ln(f_\blacklozenge(x)/f_r(x))\, dx$. a) Calculate Hx if X Exp()). b) Let $Y := e^x$, where X U[O, 1]. Find i) the probability density function of the random variable Y; ii) Hr. c) Calculate $H(f_x; f_r)$ if X U[O, 1] and Y Exp(1). Problem no. 36 a) Let X N(0, 1). We define $Y = f_X f$. i) Calculate $P[Y < 2]$. Hint. We have: Q(1) 0.159 and Q(2) - 0.0228. ii) Calculate VAR[Y]. b) Let Z G(c = 2, = 1). Obtain $E[Z^3]$ by making use of the characteristic function of Z. Hint. We have: $bz(co) = (1 - jco)^{-2}$. Problem no. 37 Let X be the number of unsuccessful interviews that a graduating student must have before getting a first job. According to the data collected, we accept that $E[X] = 2.5$.

130 3 Random Variables a) Calculate the maximum value of $P[X > 5]$. b) If we assume that $X = Y - 1$, where Y has a geometric distribution, what value of p must we choose? Justify your answer. c) If we have found that VAR[X] -- 2.5, is it better to use a geometric distribution as in b) or a Poisson distribution as a model for X? Justify your answer. d) Suppose that X has a Poisson distribution. Calculate $P[X = 0 \mid X < 1]$. Problem no. 38 I,et X be the lifetime

(in years) of a certain machine. a) Calculate $P[X^2 > 4]$ if $X \sim U[0, 10]$. b) If the value of x that maximizes the probability density function of X is $x = 4$ years, is it better to use an exponential distribution or a gamma distribution as a model for X? Justify your answer. C) Suppose that $X \sim N(4, 1)$ (approximately). i) Calculate $P[3 < X < 5]$. ii) Find the number xo such that $P[2X < xo] = 0.9$. Hint. We have: Q-(0.1) 1.2815 and Q-(0.9) -1.2815. Problem no. 39 We have a computer program that enables us to test whether a software program contains bugs or not. We suppose that, each time the program is used, the probability that it detects the presence of a given bug is equal to 0.95, independently from time to time. Furthermore, we suppose that the number N of bugs that a given software program contains has a Poisson distribution with parameter o : 0.5, and that the bugs are independent. a) Calculate $E[N \mid N \leq 2]$. b) What is the probability that we must use the program less than three times to detect the presence of a given bug'? c) Calculate the probability that the program does not detect the presence of any bugs i) in a software program that has in fact k bugs, where k $\in$ {0, 1,2 .... }; ii) in any software program. Problem no. 40 Let x2/9 if $0 < x < 3$, fx(x)= 0 elsewhere. We define $Y = int(X)$, where int denotes the integer part. Calculate a) the probability density function of Y; b) the generating function of Y; c) VAR[1/X]; d) Fx(xIX < 1).

3.10 Exercises, Problems, and Multiple Choice Questions 131 Problem no. 41 a) The random variable X has a Gaussian $N(0, 1)$ distribution. We define $Y = X^2$. i) Calculate the characteristic function of Y. ii) Give the distribution of Y, as well as its parameters. b) Let Z be a random variable that has a G(c, .) distribution. We define $W = a Z + b$, where a $\neq$ 0 and b are constants. i) Calculate the characteristic function of W. ii) For what values of a and b does the random variable W have a gamma distri- bution? /-/int. We have: Cz(co = (1 - Problem no. 42 Let 0 if x <0, c if $0 < x < 1$, 2c(2-x) if $1 < x < 2$, if x >2, fx(x) = 0 where c is a positive constant. Calculate a) the constant c; b) the distribution function of X; c) the probability density function of X, given that $X < 1$; d) the median of X; e) the 95th percentile of X. Problem no. 43 A computer system user works at home and connects to the network by modem. We suppose that the probability that the user manages to connect to the network is equal to 0.95, independently from one attempt to another. a) Let N be the number of successful connection attempts among 50 attempts. Cal- culate $P[N = 50 \mid N > 48]$. b) Use a Poisson distribution to calculate (approximately) the probability $P[N = 49]$, where N is defined in a). c) Let M be the number of attempts

needed to manage to connect to the network. Calculate the variance of M, knowing that the first attempt was unsuccessful. Problem no. 44 a) Let X U(0, 1]. We define Y = in X. Calculate i) the distribution function of Y; ii) the moment generating function of Y. b) Let Z G(c = 2, . = 1/2). Use the characteristic function of Z to obtain VAR[Z2]. Hint. We have: qz (co) = (1 - 2jco)-2.

132 3 Random Variables Problem no. 45 The users of a certain computer network can work at home with the help of a modem. Let T be the duration (in minutes) of a given user's work session. We suppose that 0 ift <0, FT(t) = t/100 if0 < t < 90, 1 if t > 90. a) What is the probability that a given user reaches the maximum duration of a work session? Justify your answer. b) Let S := T] {T < 60}. That is, S is the duration of a work session, given that it is less than 60 minutes. Calculate i) the probability density function J,(s); ii) E[S2]. c) Let 0ifT < 30, U:= 1 if 30< T <60, 2 if T >_ 60. Calculate i) Pu (u) and ii) the generating function of U . Problem no. 46 Let X be a discrete random variable whose set of possible values is Sx = {-2, 0, 2}, and let A = {X _< 0}. We suppose that Fx(x l A) = 0 ifx < -2, 3/5 if -2<x <0, 1 i fx > 0. Calculate a) P[X 2 __< 1 I A]; b) E[X A]; c) fx(x I A); d) Fx(x I A"); e) px(2) if P[X = 01 = 2P[X = 21. Problem no. 47 Let p be the probability that an individual in a population has a certain virus. To detect the presence of the virus, we analyze blood samples taken from each of the N members of this population. We assume that the N members of the population are independent. a) i) Calculate the mean and the variance of the number of persons who have the virus, given that the first two blood tests performed were positive. ii) Use Poisson's approximation to calculate the probability that at least three persons have the virus, if N = 1000 and p = 0.01. b) Suppose now that instead of proceeding as above, we rather decide to combine the samples taken from k persons before analyzing them. If the test is negative, then the k persons do not have the virus. On the other hand, if the test is positive, then the k persons must be tested again individually. Calculate the average number of tests required with this method, if we suppose that N = 500, p = 0.01 and k = 10.

3.10 Exercises, Problems, and Multiple Choice Questions 133 Multiple Choice Questions Question no. 1 We suppose that the probability that a telephone call lasts more than five minutes is equal to 0.1, independently from one call to another. A) Calculate the probability that, among 20 calls taken at

random, there are more than 18 calls that do not last more than five minutes. a) 0 b) 0.3917 c) 0.6083 d) 0.6769 e) 1 B) Calculate approximately the probability in A) using a Poisson distribution. a) 0 b) 0.3233 c) 0.4060 d) 0.6767 e) 1 C) Calculate the probability that it takes less than five calls to obtain a first call that lasts more than five minutes. a) 0.0001 b) 0.3439 c) 0.4095 d) 0.6561 e) 0.9999 D) What is the probability that, among five calls taken at random, the longest lasts less than five minutes? a) 0.00001 b) 0.1 c) 0.4095 d) 0.5905 e) 0.9 Question no. 2 Let X be the time (in days) required to repair a machine. We suppose that the average repair time is equal to four days and the standard deviation to two days. A) What is, at most (and with as much accuracy as possible), the probability that the repair time is smaller than one day or greater than seven days? a) 4/9 b) 2/3 c) 1 d) we cannot calculate it e) none of these answers B) Suppose that X has a uniform distribution on the interval [a, hi. Find the con- stant a. a) 0 b) 2 c) 4 d) we cannot calculate it e) none of these answers C) Suppose that X has a gamma distribution with parameters c : 4 and $= 1$. Calculate P [X < 4]. a) 0.2381 b) 0.3528 c) 0.5665 d) 0.5768 e) 0.6288 D) Suppose that X has a Gaussian distribution. Find the number xo such that P [IX - 41 < xo] = 0.99. Hint. We have: Q-1 (0.01) - 2.326 and Q I (0.005) - 2.576. a) 2.326 b) 2.576 Question no. 3 Let C) 4.652 d) 5.152 e) 10.304 -e xx ifx < 0, 2 -- Fx(x) = 1 ◆e -)x ifx > O, where L > 0. Calculate Fx(1 I X > 0). a) ◆e - b) e - c) l-e - d) l-◆e - e) 2 - e -x

134 3 Random Variables Question no. 4 Let Calculate Fx (0) + Fx (1/2). x -1 0 1 px(x) 1/8 3/8 1/2 a) 3/8 b) 1/2 c) 3/4 d) 7/8 e) l Question no. 5 A computer generates indcpcndcnt observations of a random variable that has a Poisson distribution with parameter $= 1$. Let N be the number of observations, among the first 100 observations, that are greater than 1. Calculate VAR[N]. a) 200(1 - 2e-)e - b) 100(1 - e-l)e -1 c) 100(1 - 2e-)e - d) 200(1 - e-)e- e) none of these answers Question no. 6 Let X be a random variable that has an exponential distribution with parameter 2.. What is the value of 2. if the 90th percentile of X is equal to 1 ? a) 0.1 b) 0.9 c) 1 d) lnl0 e) ln90 Question no. 7 Suppose that X is a continuous non-negative random variable such that E[X] = 1. We define p = P[X > 1]. Let p be the value of p if X has a uniform distribution, and P2 be the value of p if X = y2, where Y N(0, 1). Calculate p + P2. a) 0.318 b) 0.5 c) 0.659 d) 0.818 e) 1 Question no. 8 Let x -1 0 1 px(x) 1/2 1/4 1/4 Calculate STD[X2]. a) 0.1875 b) 0.4330 c) 0.6875 d) 0.8292 e) 1.1456 Question no. 9 We define Y : IX I, where X N(0, 1). What is, according

to Markov's inequal- ity, the largest value that the probability $P[Y >_{-} 2/x/--]$ can take? a) 1/2 b) 0.5642 c) 0.7071 d) 0.8862 e) 1 Question no. 10 Let $t\varphi_X(\omega) = e^{-\omega^2}$. We define $Y : 2X - 1$. Calculate a) $e^{-\omega^2}$ b) $2e^{-?} - 1$ c) e^{-4wz} d) $e^{-jw-2w2}$ e) $e^{-jw-4wz}$ Question no. 11 Let $X N(0, 1)$. We define

3.10 Exercises, Problems, and Multiple Choice Questions 135 0ifX <0, Y= lif0_ <X<l, 2ifX> 1. Calculate the generating function of Y. a) $0.341z + 0.159z^2$ b) $1 + 0.341z + 0.159z^2$ c) ⍰ $+0.341z + 0.159z^2$ d) ⍰ $+ 0.841z + Z^2$ e) $1 + 0.841z + Z^2$ Question no. 12 A certain school bought 20 computers so that its students can connect to the Internet. The school distributed access codes to the 200 students registered in the computer science course. We estimate that each student having an access code has a 0.2 probability of wishing to connect to the Internet at noon on a given day, indepen- dently from one student to another and from day to day. A) Use Poisson's approximation to calculate the probability p that all the computers are occupied, at noon, on a given day. a) 0.9990 b) 0.9992 c) 0.9994 d) 0.9996 e) 0.9998 Hint. If M -- Poi(40), then pM(20) -- 0.00019 and FM(20) -- 0.00037. Answer the following questions by assuming that the answer in A) is p = 0.9. B) What is the probability that all the computers are occupied, at noon, on at least two days among the five days of a given school week? a) 0.00046 b) 0.00856 c) 0.99144 d) 0.99954 e) 0.99990 C) What is the probability that the first day, from the beginning of the term, on which all the computers are occupied, at noon, is the second or third day of this term? a) 0.009 b) 0.0099 c) 0.099 d) 0.1539 e) 0.171 Question no. 13 Let X be the utilization time of a computer during an eight-hour workday. We propose four models for the distribution of X: X1 " U[0, 8], X2 G(2, ⍰)and X4 = 8e -X2. A) Among the above models, which are the only ones that can be the true model? c) X2, X3 d) X1, X2, X3 e) all -- Exp(1/4). Calculate rn := E[X I X > 1] and v := a) X1 b) X1, X4 B) Suppose that X VAR[X I X > 1]. a) rn = 4and v = 4 d) rn=5andv= 16 b) m=4andv=5 c) m=4andv= 16 e) m=5andv= 17 C) Suppose that X "- G(2, ⍰).Calculate P[X > 3.5]. a) 0.3041 b) 0.4779 c) 0.5578 d) 0.7440 e) 0.8088 Question no. 14 An information source generates letters taken at random among the letters a, b, c, d and e. We suppose that $P[\{a\}] = 1/2$, $P[\{b\}] = 1/4$, $P[\{c\}] = 1/8$ and

136 3 Random Variables $P[\{d\}] = P[\{e\}] = 1/16$. Moreover, a data compression system transforms the let- ters into binary strings as follows: a = 1, b = 01, c = 001, d = 0001 and e = 0000. Let X be the length of a binary

string emitted by the data compression system. Calculate A) Fx(2-) + Fx(5-); a) 1/2 b) 3/4 B) E[XIX > 1]; a) 11/8 b) 15/8 c) 9/4 C) VAR[X2]. a) 71/64 b) 5041/4096 Question no. 15 Let c) 1 d) 3/2 e) 7/4 d) 5/2 c) 11/4 FX (x) = e) 11/4 d) 1615/64 e) 42 0 if x< I, I/4 ifx = I, x2/4 if I < x < 2, I ifx >2. Calculate P[X = I] + P[X < 2]. a) I/4 b) I/2 c) 3/4 d) I e) 5/4 Question no. 16 Let X be a continuous random variable for which fx(xlX < I)=2x if0<x < I. Which of the following functions can be the function fx (x)? 1)2xif0<x < I 2)xif0<x < 3)x/4 if0 < x < 2,� 4)4x if0 < x < I/� 5) xif0<x < landl/2ifl <x <_2 a) 1 and 2 only c) all, except 4 e) none of these answers b) 1,2 and 3 only d) all Question no. 17 We suppose that the number Xof particles emitted by a radioactive source during a one-hour period has a Poisson distribution with parameter ot = 1/2, independently from hour to hour. Let N be the number of hours during which no particles are emitted, among the 24 hours of a given day. What distribution does N have? a) Poi(e -'/2) b) Poi(12) c) B(24, �) d) B(24, �e -'/2) e) B(24, e -/2) Question no. 18 Boxes contain 20 objects each. We examine the contents of the boxes until we find one that contains no defective objects. Let N be the number of boxes that we

3.10 Exercises, Problems, and Multiple Choice Questions 137 must examine to end the random experiment. What distribution does N have if the probability that an object is defective is equal to 1/10, independently from one object to another? a) Geom(1/10) b) Geom((9/10) 2�) c) B(20, 1/10) d) B(20, (9/10) 2�) e) none of these answers Question no. 19 Calculate P[X 2 < 9], where X is a random variable that has a gamma distribu- tion with parameters a = 2 and , = 1. a) e -3 b) 4e -3 c) 1 - e -3 d) 1 - 4e -3 e) 1 - e -3 Question no. 20 The lifetime (in months) of brand A components has (approximately) a Gaussian N(50, 100) distribution and that of brand/3 components a Gaussian N(60, 100) dis- tribution. A component is taken at random from a box containing 10 brand A and 20 brand B components. Calculate the probability that the component selected lasts at least 50 months. Hint. We have: Q(1) - 0.159 and Q(2) - 0.0228. a) 0.6705 b) 0.7273 c) 0.7386 d) 0.8181 e) none of these answers Question no. 21 We define Y = 1/X 2, where X is a random variable having a uniform distribu- tion on the interval [l, 3]. Calculate fr (y). a) �y-3/2 if 1/9 _< y _< 1 b) �y-3/2 if 1/9 _< y _< 1 c) � if 1 _< y _< 3 d)-ifl/9_<y< 1 e) 2if-1 _<y_< 1 Question no. 22 Let fx(x) = xe -x ifx >0. Calculate VAR[N]. a) 2 b) 8/3 c) 20/3 d) 28/3 e) none of these answers 1 1 2 1 4 GN(Z) = - q--Z q--Z � Calculate E[X-2]. a) 1/6 b) 1/4 c) 6 d) e) - Question no. 23 Let X be a

random variable having a gamma distribution with parameters a = 30 and) = 20. According to Markov's inequality, what is the minimum value of $P[X < 2]$? a) 1/6 b) l/4 c) 1/2 d) 3/4 e) 5/6 Question no. 24 A discrete random variable N has the following generating function:

138 3 Random Variables Question no. 25 Let e J�- 1 Ox (o) = k , a) where k is a constant. Calculate E[2X]. a) l/ln2 b) 2/ln2 c) ln2 d) 21n2 Question no. 26 Let e) none of these answers 0 ifx < 1, (x- 1)/2 ifl <x <2, Fx(x) = x2/8 if2 < x < 2x/, 1 ifx > 2x/. (3 <X<2)forxE[2] A) Calculate Fx x .... a) x-I b) x -3 c)x-I d) 2(x- 1) e) 2x-3 B) Calculate E[X]. a) 0.97 b) 1.47 c) 1.97 d) 2.00 e) 2.47 C) Find the 90th percentile of X. a) 0.9 b) 1.8 c) 2.0 d) 2.8 e) none of these answers D) Calculate VAR [X X<]. a) 1/96 b) 1/48 c) 1/12 d) 1/4 e) 1/2 E) Let Y := I/X. Obtain Fr(y) for � < y < 1. 1 1 _ v a) 23 2yl b)(� 1) c)--$ d) �(Y-l) e) none of these answers F) Let W := int(X). That is, W is the integer part of X. Calculate the generating function Gw(z) of W. a) �(1 + z) b) �(1 + z + z 2) c) �(z q- z 2) d) �(z q- z 2 q- z 3) e) �(l +z+z 2+z 3) Question no. 27 A certain city is supplied with electricity through a single high tension line. The number X of failures, during a one-year period, has a Poisson distribution with pa- rameter 3. Moreover, the duration Y (in hours) of a given failure has an exponential distribution with mean 2. Finally, we assume that all the failures are independent. A) Let N be the number of years, over a 25-year period, during which there has been exactly one failure. Calculate the variance of N. a) 1.2 b) 2.0 c) 3.0 d) 3.2 e) 4.0

3.10 Exercises, Problems, and Multiple Choice Questions 139 B) Let M be the number of failures required, from now, to obtain a second failure that lasts more than two hours. Calculate E[M]. a) l/e b) 2/e c) e d) 2e e) none of these answers C) Calculate E[XIX > 1]. a) 3 b) 3.16 c) 3.32 d) 3.48 e) 3.64 D) Let Z := Y I {Y < 1}. Calculate the moment generating function Mz(s) of Z at s = -1/2. a) 1/[2(1 - e-l/2)] b) 1/(1 - e -1/2) c) 1 - e -1/2 d) 2(1 - e -1/2) e) none of these answers E) Suppose that there have been exactly three failures during a given year. We can show that the total duration T (in hours) of these failures has a G(c = 3, = 1/2) distribution. Calculate $P[T < 6.5]$. a) 0.37 b) 0.41 c) 0.59 d) 0.63 e) 0.84 F) We consider 30 failures. We can show that the total duration D (in hours) of these failures has approximately a Gaussian N(60, 120) distribution. Find the number d such that $P[D < d] = 0.1$. Hint. Wehave: Q(1) _ 0.159 and Q-I(0.1) _ 1.2815. a) 46 b) 74 c) 129 d) 154 e) 214 Question no. 28

A) Let T be the time (in minutes) that a customer waits before being served at a counter when there is at least one customer in the queue when he/she arrives. We suppose that $P[T < t] = t/10$ for $0 < t < 10$. (,) i) Let $X = 1$ if the customer waits more than two minutes and $X = 0$ otherwise. Calculate the function Px if the queue is not empty when the customer arrives. 1/10 if x=0, a) $px(x) = 9/10$ ifx 1. 3/10 ifx = 0, c) $px(x) = 7/10$ ifx 1. 7/10 ifx = 0, e) $px(x) = 3/10$ ifx 1. 2/10 ifx = 0, b) $px(x) = 8/10$ ifx 1. 1/2 ifx = 0, d) $px(x) = 1/2$ ifx=1. ii) Calculate $f_T(t \mid T > 2)$ for $2 < t < 10$. a) t/8 b) t/7 c) 1/10 d) 1/8 e) 1/7 B) Suppose now that the probability that the queue is empty when the customer arrives is equal to 1/5. Let Y be the waiting time of the customer. Calculate, making use of the assumption (,) in A), the distribution function of Y for $0 < y < 10$. a) 0.08 y b) 0.1 y c) 0.2 + 0.08 y d) 0.1 q- 0.09 y e) 1

140 3 Random Variables Question no. 29 A box contains ten transistors, of which two are defective. Two transistors are taken at random and with replacement. Let X be the number of defective transistors among the two picked. We repeat this random experiment until the two transistors picked are two defective transistors. Let N be the number of repetitions required to obtain a first repetition for which the two transistors picked are defective. A) Give the distribution of X, as well as its parameter(s). Likewise for N. a) X B(2, ◆); N Geom(0.04) b) X B(10, ◆); N B(2, ◆) c) X B(2, ◆); N B(10, ◆) d) X B(10, ◆); N Poi(2) e) X B(2, ◆); N Geom(0.2) B) Calculate $P[N > 4 \mid N > 1]$. a) 0.0784 b) 0. I 153 c) 0.8847 d) 0.9216 e) 0.9600 C) Give the distribution of N, as well as its parameter(s), if the two transistors are taken without replacement (but the transistors are put back into the box after the two draws if they are not both defective). a) Geom() b)Geom() c)B(10,) d)B(10, ◆) e)Poi(2) Question no. 30 Let X be a random variable having a G(ot,.) distribution. A) Calculate $P[X > I]$ if or = 2 and . = 3. a) e -3 b) 4e -3 c) 8.5e -3 d) I - 4e -3 e) I - e -3 B) Suppose that ot = I and . = 2. Calculate $E[X 2 \mid X 2 > I]$. a) l b) 3/2 c) 2 d) 9/4 e) 5/2 C) Let Z := X 2. Find fz(z) for z > 0, if or = 2 and . = I. a) ◆e -',/ b) 4e -',/ c) 'e -',/ d) z2e -2z e) z2e -z- Question no. 31 Let X be a standard Gaussian random variable. A) Find the number x0 for which P [X 2 > x0] = 0.05. Hint. We have: Q- (0.05) - 1.645 and Q- (0.025) - 1.960. a) 1.645 b) 1.960 c) 3.290 d) 3.842 e) 3.920 B) Calculate $E[X2(X + I)]$. a) 0 b) l c) 2 d) 3 e) 5 C) Find the number(s) a for which the characteristic function of Y := aX + 1 is given by exp(jco - 2092). a)-I b) l c) q-1 d) 2 e)-1-2 Hint. We have: $q_x(\omega) = e^{-w2/2}$. D) Calculate the generating function of W, that is,

$G_W(z) := E[z^W]$, where

a) ◆ b) ◆z c) ◆(1 +z) Question no. 32 3.10 Exercises, Problems, and Multiple Choice Questions 1ifX>O, W:---- 0ifX <0. d) 1+z e) 1+z+z 2 141 Let X be the utilization time (in months) of a laser printer until its first breakdown. A) Suppose that X -- Exp(1/3). Calculate E[[X - 3[]. a) 3e -1 b) 3(1-e -1) c) 6e -1 d) 3 e) 6(1-e -1) B) Suppose that X -- G(t = 2,) = 1). We define Y = 1/(X + 1). Find f,(1/2). a) ◆ e- 1 b) e- 1 c) 2e- 1 d) 4e- 1 e) 8e- 1 C) Suppose that the time Z (in months) required for the 25th breakdown to occur has approximately a Gaussian N(75,225) distribution. Calculate P[Z < 90 [Z > 60]. Hint. We have: Q(1) - 0.159. a) 0.1891 b) 0.3180 c) 0.6820 d) 0.8109 e) 0.8410 Question no. 33 A) Let N be a random variable having a geometric distribution with parameter p. Calculate FN (n [N > 1) for n = 2, 3 ..... a) 1 - qn-1 b) 1 - qn c) 1 - qn+l d) 1 - pn-1 e) 1 - pn B) Let X -- U[0, 1]. i) Obtain the characteristic function of Y := ln(X + 1). Hint. We have: q;(w) -- eJ-. jco 2 j◆ - 1 2 j◆+l -- 1 2 j◆ 2 j◆ -- 1 2 j◆ -- 1 a)-- b) c) . d)-- e)-- jw jw+ 1 jw+ 1 2jw jwln2 ii) Let 1 ifX < 0.3, W:= 2if X>0.3. Calculate the generating function of V := 2W. a) z(0.3 + 0.7z) b) 2z(0.3 + 0.7z) c) 2z(0.3 + 1.4z) d) [z(0.3 + 0.7z)] 2 e) z2(0.3 q- 0.7z 2) Question no. 34 A number N is taken at random in the set {0, 1, 2 ..... 9}. Calculate pN(5) q- FN(5.5). a) 0.6 b) 0.65 c) 0.7 d) 0.75 e) 0.8 Question no. 35 Let fx(x)=2x if0<x < 1. Find fx(1/4[X < 1/2). a)0.5 b) 1 c) 1.5 d) 2 e)2.5

142 3 Random Variables Question no. 36 Let l/2if0<x < 1, fx(x)= kx ifl <x <2, where k is a positive constant. Calculate the 60th percentile of X. a) 1.16 b) 1.26 c) 1.36 d) 1.46 e) 1.56 Question no. 37 Suppose that X B(n = 5, p = 0.2). Calculate P[X = I IX < 1]. a) 1/2 b) 5/9 c) 2/3 d) 7/9 e) 8/9 Question no. 38 We take independent observations of a random variable X having a Poisson dis- tribution with parameter c = 2 until we obtain "0". Let N be the number of observa- tions required to end the random experiment. Calculate VAR[N - 1]. a) 43.21 b) 44.21 c) 45.21 d) 46.21 e) 47.21 Question no. 39 We consider a random variable X having an exponential distribution with param- eter . = 1. Calculate E[X [1 < X < 2]. a) 1.27 b) 1.32 c) 1.37 d) 1.42 e) 1.47 Question no. 40 Suppose that X G(c = 3, . = 1). We define Y = x/. Find fr(2). a) 32e 4 b) 64e -4 c) 32e -2 d) 64e -2 e) 32e- Question no. 41 Find the number xo for which P[X < -xo] + P[X > x0] = 1/2, where X is a standard Gaussian random variable. a) Q-(I/4) b) 1 - Q-(I/4) c) ◆Q-(I/4) d) 2Q-(1/4) e) 1 - 2Q-(1/4)

Question no. 42 Let X be a random variable having a Bernoulli distribution with parameter p = 3/4. Calculate bx (1). a) �e j b) �eJ c) �(1 + e J) d) �(3 + e j) e) �(1 + 3e J) Question no. 43 Suppose that the lifetime of a system is a random variable T whose probability density function is given by f, (t)= 1/2 ifl <t < 3. Calculate the failure rate r(t) at time t = 2. a) 1/4 b) 1/2 c) 1 d) 3/2 e) 2

3.10 Exercises, Problems, and Multiple Choice Questions 143 Question no. 44 We consider a system made up of two components placed in parallel and a third component placed in series. What is the reliability of the system at time t = 3 if the three components function independently and all have the reliability function R(t) = e -t for t > 0? a) 0.0038 b) 0.0048 Question no. 45 Let c) 0.0058 d) 0.0068 e) 0.0078 FX(x) = 0 ifx < -1, (x+l)/4if-1 <x < 1, 1/2 ifl <x<2, 3/4 if2<x <3, 1 ifx > 3. A) Calculate Fx (0] X < 2). a) 1/8 b) 1/4 c) 1/3 d) 1/2 e) 3/4 B) Calculate E[X]. a) 0 b) 3/4 c) 5/4 d) 7/4 e) 9/4 Question no. 46 An exam is made up of five multiple choice questions. For each question, five answers are proposed. Every correct answer gives two points, while 1/2 point is deducted for each wrong answer, up to a maximum of - 1. Finally, the exam is over 7 (that is, the maximum grade is 7) and the minimum grade is 0. A student answers every question at random. Let X be the grade of this student. A) Calculate E[X]. a) 1.327 b) 1.427 c) 1.527 d) 1.627 e) 1.727 B) What is the mathematical expectation of X, given that the first three answers are correct? a) 5.67 b) 5.72 c) 5.77 d) 5.82 e) 5.87 Question no. 47 We have a random numbers generator for a random variable X having a uniform distribution on the interval (0, 1). A) We want to use a transformation Y = g(X) to obtain observations of a random variable having the following probability density function: fy(y)=l/y2 ify > 1. What transformation g(X) can we use? a) l/X 2 b) el/X--e+l c) 1--1nX d) l/X e) 1-21nX B) We generate independent observations of X until we have obtained two observa- tions greater than 0.9. What is the probability that we have to generate exactly three observations ? a) 0.009 b) 0.018 c) 0.027 d) 0.036 e) 0.045

144 3 Random Variables Question no. 48 Let fx (X) = 2x/2x/x2e -x2/2 forx >0. A) Use a linear interpolation, based on the fact that Q(1) _ 0.159 and Q(2) - 0.02275, to calculate approximately the median of X. a) 1.48 b) 1.52 c) 1.56 d) 1.60 e) 1.64 B) We find that E[I/X] = 2x/ . Calculate VAR[I/X]. a) 0.3234 b) 0.3334 c) 0.3434 d) 0.3534 e) 0.3634 Question no. 49 A) Let X N(I, l).

Calculate P[X 2 - 2X > 0]. a) Q(I) b) 2Q(I) c)l - Q(I) d) 1 - 2Q(I) e) 2(1 - Q(I)) B) With the help of Bienaym6-Chebyshev's inequality, calculate an upper bound for Q(,/). a) 1/4 b) 1/3 c) 1/2 d) 2/3 e) 3/4 Question no. 50 A) Let X be a random variable having a Poisson distribution with parameter t = 1/3. Calculate E[X !]. 2e-1/3 e) -e -1/3 a) 1/3 b) 2/3 c) e -/3 d) 3 B) The characteristic function of a random variable X having an exponential distribution with parameter . is given by qbx(w) = ./(. - JW)" Use this formula to calculate E[-Y(Y 1)2/3], if Y has a gamma distribution with parameters o = 1 and . = 1. a)-11/3 b)-7/9 c)-3/4 d) 1/3 e) 3 Question no. 51 We suppose that the lifetime X of every component of the system in Figure 3.19, has a probability density function fx (x) = xe -x2/2 for x > 0. Moreover, we assume that the components operate independently from one another. A) What is the failure rate of a component at time x = 2? a) e -2 b) e 2 c) 2e 2 d) 1/2 e) 2 B) What is the value of the reliability function of the system at time x = 2? a) e -2 - 2e -6 b) 2e -2 - e -4 c) 2e -6 - e -8 d) e -2 e) e -4 + e -6 Figure 3.19. System in Multiple choice question no. 51.

3.10 Exercises, Problems, and Multiple Choice Questions Question no. 52 Let X be a random variable whose distribution function is 0 ifx <0, Fx(x) = �(x + 1) if0 < x < 1, 1 ifx > 1. [0 < X < �]+P[X= 1]. Calculate P a) 1/2 b) 5/8 c) 11/16 d) 3/4 e) 7/8 Question no. 53 A continuous random variable X has the probability density function fx(x)=le-lXl for- <x <. 2 Calculate P[-1 < X < 1] + P[X = 2]. a) e -1 b) 2e -1 c) 1 - e -1 d) e -1 + �e -2 e) 1 - e -1 + �e -2 Question no. 54 by 145 We consider a random variable X whose probability density function is defined 1 +xif-1 <x <0, fx(x) : 1 - x if0 < x < 1, 0 elsewhere. A) Find Fx (3/4). a) 3/4 b) 13/16 c) 7/8 d) 5/6 e) 31/32 B) Calculate E[XIX > 0]. a) 1/6 b) 1/4 c) 1/3 d) 1/2 e) 3/4 C)Let Y := IXI. Find f, (y) for0 <_ y _< 1. a) l b) --y c) �+y d) 2y e) 2-2y Question no. 55 Let X be a random variable having a uniform distribution on the interval [0, 1]. We define Y = max {X, �1. A) Find Fy(y) for � < y < 1. What type of random variable is Y? a) �; discrete b) y; continuous c) 2y - 1; continuous d) y; mixed e) 2y - 1; mixed B) Calculate E[Y]. a) 9/16 b) 5/8 c) 11/16 d) 3/4 e) 13/16 Question no. 56 We want to find a model for the number N of cars crossing a certain intersection during a one-minute period. According to the data collected, the most frequent value

146 3 Random Variables of N is five cars. It also happens, although not

frequently, that the traffic is much heavier. Which of the following models seems to be the most appropriate? Oeom() d) pN(n) = ()n() for n =0, 1 .... e) pN(n) = 0 for n = 1 ..... 10 Question no. 57 We look for a model, as realistic as possible, for a continuous random variable X that represents the lifetime of a machine, and whose mean and variance arc equal to 1 and 3, respectively. Which of the following distributions can be acceptable? 1) uniform 2) exponential 3) gamma 4) Gaussian 5) the square of a Gaussian N(1,3) distribution a) all except 4) b) 2), 3) and 5) only c) 3) and 5) only d) 3) only e) 4) only Question no. 58 A fabrication process for producing one-inch diameter screws is such that the diameter (in inches) of a screw can be considered as being the value taken by a random variable X having (approximately) a Gaussian N(1,625 � 10 -4) distribution, independently from one screw to another. A) What is the probability that a screw taken at random has a diameter inaccurate by more than 0.05 inch'? Hint. We have: Q(0.1) - 0.4602 and Q(0.2) - 0.4207. a) 0.8414 b) 0.8514 c) 0.8614 d) 0.8714 e) 0.8814 B) Let p be the answer in A) and q := 1 - p, We take six screws at random (and without replacement). What is the probability that less than two of these screws have a diameter inaccurate by more than 0.05 inch? a) p6 q_ pSq b) p6 q_ 6pSq c) p6 q_ 5pSq d) q6 q_ pq5 e) qe, + 6pq5 Question no. 59 The average lifetime of a certain type of tire is equal to three years, with a stan- dard deviation of 0.3 year. What can be said, with as much accuracy as possible, concerning the probability p that a tire of this type lasts more than 54 months or less than 18 months? a) p _< 1/25 b) p > 1/25 c) p > 1/25 d) p _< 24/25 e) p = 12/25 Question no. 60 The characteristic function of a discrete random variable X is given by 1 (l+eJ�+2e2J�) x() = � Calculate P[X = 1]. a) 0 b) 1/4 c) 1/2 d) 3/8 e) 5/8

3.10 Exercises, Problems, and Multiple Choice Questions 147 Question no. 61 A certain system is made up of two components placed in parallel and operat- ing independently. The lifetime of the first component has a uniform distribution on the interval (0, 1), while the lifetime of the second component has an exponen- tial distribution with parameter 2. Calculate the reliability function of the system at time 1/2. a) 0.1839 b) 0.3161 c) 0.6839 d) 0.8161 e) 0.8679 Question no. 62 The electric current X (in amperes) that passes through a resistance of r ohms is a random variable having the probability density function �if4<x<6, fx(x) = 0 elsewhere. Furthermore, the power Y (in watts) is given by Y = rX 2. A) Find the value of the distribution function Fx (5 I X

>_ 4.5). a) 1/6 b) 1/4 c) 1/3 d) 1/2 e) 5/6 B) Calculate VAR[X I X > 4.5]. a) 1/16 b) 1/12 c) 1/8 d) 3/16 e) 1/4 C) An observation of the random variable Y is generated. In which interval, among the following, is the observation most likely to be found if r = 25? a) (400, 401) b) (500, 501) c) (700, 701) d) (899, 900) e) the four preceding intervals are equally likely D) Calculate VAR[Y] if r = 30. a) 30,040 b) 30,050 c) 30,060 d) 30,070 e) 30,080 Question no. 63 Each day a student spends a random time T (in minutes) in front of her computer to do some work. We estimate that E[T] = 40 and VAR[T] = 1600. A) Let d be a duration that at least 90% of the work sessions do not exceed. Accord- ing to Bienaym6-Chebyshev's inequality, what is the minimum value of d? a) 50.98 b) 72.53 c) 102.45 d) 166.49 e) 227.68 B) Suppose that T has an exponential distribution. i) If the student has already spent 30 minutes in front of her computer, what is the probability that the work session will last at least 60 more minutes? a) 0.1245 b) 0.2231 c) 0.2548 d) 0.3679 e) 0.6321 ii) How many work sessions does it take, on average, to obtain a first session that lasts more than 90 minutes? a) 9.19 b) 9.29 c) 9.39 d) 9.49 e) 9.59

148 3 Random Variables Question no. 64 A person subscribes to a service that allows him to connect to the Internet for 30 hours per month. We suppose that the time X (in hours) that this person spends surfing the Internet during a given day has a gamma distribution with parameters ot = 3 and) = 4, independently from day to day. A) What is the probability that the person spends more than 90 minutes on the Inter- net during a given day? a) 0.022 b) 0.032 c) 0.042 d) 0.052 e) 0.062 B) Suppose that the answer in A), for other values of or and), is equal to 0.05. What is the probability that, during a given week, the person spends more than 90 minutes on the Internet on at least three days? a) 0.0018 b) 0.0028 c) 0.0038 d) 0.0048 e) 0.0058 C) We can show that the total time (in hours) during which the person wants to con- nect to the Internet, over a 30-day month, has approximately a Gaussian distribution with parameters/ = 22.5 and cr 2 = 5.625. What is the probability that the person is disconnected because he reached the time limit imposed? a) 0(-3.16) b) 0(3.16) c) 0(3.16) + � d) � - 0(3.16) e) 2 exp{-(3.16)2/2} Question no. 65 Let fx(x)=(n+l)x n for0 <x< 1, where n 6 {1,2 .... }. A) Calculate the 50th percentile (that is, the median) of X if n = 3. a) 0.8209 b) 0.8309 c) 0.8409 d) 0.8509 e) 0.8609 B) We define Y = X '. Find fv(l). a) 1 b) n + 1 c) (n + 1) n d) n/(n + 1) e) (n + l)/n C) Calculate the characteristic function bx (o) of X at o = 2, if n = 1. a) � [e2J(l - 2j) - 1] b) � [e2J(2j - 1) - 1] c) � [e2J(l - 2j) + 1] d) � [e2J(2j + 1)- 1] e) �

[e2J(2j + 1)+ 1] D) Calculate the failure rate at time x = 1/2 of a machine whose lifetime X has the above probability density function fx with n = 2. a) 5/7 b) 6/7 c) 1 d) 7/6 e) 7/5 Question no. 66 Let X be a random variable of mixed type whose distribution function is 0 ifx < -1, Fx(x) 1 - x 2 = --if-1 <x <b, 2 - 1 ifx >b.

3.10 Exercises, Problems, and Multiple Choice Questions A) What are the possible values of the constant b? a)-l<b_ < 1 b)-I <b<l c)-I <b_ <0 d)0<b<l e)0<b<l B) Calculate P[X = (b - 1)/2] + P[X > b]. b 2 b b2 d)!+b b 2 3b a)0 b) � +-c)-+2 8 2 8 e)+-+-U C) Calculate E[X I - 1 < X < b]. b3+l b3-1 2 /b3+l b3+l b- 1 b)--- c)- d) e)--- a)-- 3 3 \b 2- lJ b 2+2 Question no. 67 149 A point X is taken at random in the interval [0, 1]. Let D be the distance between X and the endpoint of the interval closer to X. That is, X if0<X< 1 D= 1-Xif� <X_<I. A) Find fx(1/4 I X2 < 1/4). a) 0 b) 1/4 c) 1/2 d) l e) 2 B) Calculate E[X 1/2 - x-l/2]. a) 2/3 b) 4/3 c) 5/3 d) 8/3 e) ec C) Find FD(d) for0 < d < 1/2. a) b) 2d c) d� d) 4d 2 e) 8d 3 D) What is the 25th percentile of D? a) 1/32 b) 1/16 c) 1/8 d) 1/6 e) 1/4 Question no. 68 A machine is made up of four components that operate independently from one another. The machine functions if at least three of its components function. The state of the components is checked at the end of each workday and the components that failed are repaired. If the machine fails during the course of the day, it will remain down until the end of that day. We estimate that a given component has a 90% prob- ability of functioning during an entire workday, whether it is new or not, and this independently from day to day. A) What is the probability that the machine functions during an entire workday? a) 0.9077 b) 0.9177 c) 0.9277 d) 0.9377 e) 0.9477 B) What is the probability that the machine functions during exactly four entire work- days over a given (seven-day) week? a) 0.0040 b) 0.0080 c) 0.0120 d) 0.0160 e) 0.0200 C) Calculate the probability that the tenth day of a certain month is the third day during which a given component failed. a) 0.0132 b) 0.0142 c) 0.0152 d) 0.0162 e) 0.0172

150 3 Random Variables D) Suppose that the number N of failures of the machine that last more than a half day, over a six-month period, has (approximately) a Poisson distribution with param- eter c = 3. Calculate VAR[N I N > 1]. a) 2.33 b) 2.66 c) 3 d) 3.33 e) 3.66 Question no. 69 A system has n components connected in parallel and functioning independently from one another. The components function one at a time, starting with component no. 1,

so that the system functions as long as there is at least one component that is not down (standby redundancy). We suppose that the lifetime T_k of component k has an exponential distribution with parameter $\lambda = 1$ for $k = 1, 2, \dots, n$. Let T be the total lifetime of the system. A) Let $S_k := \ln T_k$. Calculate the moment generating function of S_k, defined by $M_S(s) = E[e^{-ss}]$, at $s = -2$. a) $\ln 2$ b) $\ln 4$ c) $1/2$ d) 2 e) 4 B) If $n = 2$, we can show that T has a gamma distribution with parameters $\alpha = 2$ and $\lambda = 1$. That is, $f_T(t) = te^{-t}$ for $t \geq 0$. Find the failure rate $r(t)$ of the system for $t > 0$. t 2 a) t b) $1 +$ c) d) e) t C) If $n = 25$, we can show that T has approximately a Gaussian distribution with parameters $\mu = 25$ and $\sigma^2 = $ --- 25. Calculate $P[T > 30 \mid T > 25]$. a) $2Q(1)$ b) $Q(1)$ c) $Q(1)/2$ d) $1 - Q(1)$ e) $1 - 2Q(1)$ Question no. 70 Let X be a random variable whose probability density function is $]ax + b$ if $c < x < d$, $f_X(x) / 0$ elsewhere, where a, b, c and d are constants. A) What are all possible values of the constants a and b, if $c = -1$ and $d = 1$? 1 a) $a > 0; b = 0$ b) $a = \diamond; b = \diamond$ c) $a = -\diamond; b = $ B) What is the value of a that maximizes the variance of X, if $c = -1$ and $d = 1$? a) 0 b) $\diamond$ c) $-\diamond$ d) $a\diamond$ C) Let $Y := aX + b$. Find $f_Y(y)$ if $a > 0$, $c = 0$ and $d = 1$. $2y$ $2y$ a) $2y$ if $y \in [0, 1]$ b) - if $y \in [0, a]$ c) -- if y $[0, \_\_ a$ l+a e) Y d) y if y $[l 3'] $ --if y $[-l -,-1 +]$ D) Calculate the characteristic function of X, if $a = 2$, $c = 0$ and $d = 1$. a) 72 b) j, $(J - l)$ c) 2 'o-(J - l) d) $(g52 ejW \_ f2 (eJW\_ 1)$ e) $(jg2 eJW \_ J-2 eJw$

Question no. 71 A particular case of the Weibull distribution, which is often used in reliability, is the one when the probability density function of the random variable T is given by $f_T(t) = \{3t^2 e^{-t^3}$ if $t > 0$, elsewhere. A) Calculate the failure rate $r(t)$ of a system whose lifetime T has the above probability density function. __ $3t^2e^{-t^3}$ a) $3te^{-t^3}$ b) $3t^2$ c) -- d) $3t$ e) $3t^2$ $\frac{1}{1 - e^{-t^3}}$ $\frac{1}{1 - e^{-t^3}}$ B) Find the median t_m of T, as well as the mode of T, that is, the value of t that maximizes the probability density function $f_T(t)$. a) $t_m = (\ln 2)^{1/3}$; mode $= (2/3)^{1/3}$ b) $t_m = (\ln 2)^{1/3}$; mode $= (2/3)^{2/3}$ c) $t_m = (\ln 2)^{2/3}$; mode $= (2/3)^{1/3}$ d) $t_m = (\ln 2)^{1/3}$; mode $= (\ln 2)^{1/3}$ e) $t_m = (2/3)^{1/3}$; mode $= (2/3)^{1/3}$ C) Calculate $E[T^{-3/2}]$. a) $\pi/2$ b) π c) $-/2$ d) e) D) A generalization of the probability density function given above is the function $f_T(t) = \beta t^{\beta-1} e^{-t^\beta}$ if $t > 0$, elsewhere, where β is a positive constant. We can show that $E[T] = \Gamma(1 + \frac{1}{\beta})$. For what value(s) of β is the probability density function $f_T(t)$ symmetric with respect to the mean of T? a) 1 only b) 2 only c) 1,3,5 .... d) 2,4,6 .... e) none Question no. 72 We suppose that the number of calls received at a service counter during a one-hour period has a Poisson distribution, and that 30 calls are received per hour, on average.

Moreover, we suppose that the number of calls received during a given hour is independent of the number of calls received during the previous hours. A) What is the probability that exactly 30 calls are received during a given hour, and then at least 30 calls during the next hour? Hint. We have: Fx(30) -- 0.5484 and px(30) 0.0726 if X has a Poisson distribu- tion with parameter 30. a) 0.0281 b) 0.0381 c) 0.0481 d) 0.0581 e) 0.0681 B) What is the probability that at most 30 calls are received during each of exactly 20 of the 24 hours of a given day? a) 0.0017 b) 0.0027 c) 0.0037 d) 0.0047 e) 0.0057

152 3 Random Variables C) What is the average number of hours required to obtain a first hour during which at most 30 calls are received, given that during at least one of the first three hours considered at most 30 calls were received? a) 1.32 b) 1.42 c) 1.52 d) 1.62 e) 1.72 Question no. 73 A box holds l0 brand A components and 20 brand B components. The lifetime XA (in years) of the brand A components has an exponential distribution with pa- rameter , = 1/2, while the lifetime XB (in years) of those of brand B has a gamma distribution with parameter o = 3 and , = 2, so that 2e -2x ifx > 0, fx (x) = ' 0 elsewhere. A) A component is taken at random from the box. Let X be its lifetime. i) Calculate E[X]. 19 5 7 11 23 a) i b) , c) d) -- e) 1- ii) What is the probability that the component lasts more than one year? a) 0.5533 b) 0.6033 c) 0.6533 d) 0.7033 e) 0.7533 B) We can show that the total lifetime T of the 30 components has approximately a Gaussian N(50, 55) distribution. Let Z := (T - 50)/-. What is the value of z0 for which P [Z 2 < z0] 0.95? Hint. We have: Q(1.645) 0.05 and Q(1.960) 0.025. a) 1.28 b) 1.40 c) 1.92 d) 2.71 e) 3.84 Question no. 74 The distribution function of the random variable X is given by 0 ifx <c, l Fx(x) = 1 - --: ifx > C, x z where c is a constant. A) What is the set of possible values of the constant c? a)(-,) b)(0,) c)[l,) d){1,2 .... } e)(-,-1]t_J[1,x) B) Let Y :---- I/X. Find Fy(l/4) ifc ---- 2. a) 1/16 b) 1/8 c) 1/4 d) 3/8 e) 1/2 C) Calculate E[X] if c = 2. a) 3/2 b) 2 c) 5/2 d) 3 e) 7/2 Question no. 75 We say that a continuous random variable X has a beta distribution with param- eters o > 0 and/3 > 0 if

3.10 Exercises, Problems, and Multiple Choice Questions 153 { F(c+fl) x-l(1-x) -1 for0<x < 1, fx(x) = F(c)F(fl) 0 elsewhere. This distribution generalizes the U(0, 1) distribution and is notably useful in reliabil- ity. A) Calculate Fx (1/2), ifc = 1 and fl = 2. a) 9/16 b) 5/8 c) 3/4 d) 7/8 e) 15/16 B) Findfx(� X < �),ifc=3and/3=l. a) 3/4 b) 1 c) 5/4 d) 3/2 e) 2 C) Let Y := - In X. Obtain the

characteristic function of Y, if $c = fl = 1$. 1 1 jo) jo) 1 a) -- b) -- c) -- d) -- e) 1 - jo) 1 + jo) 1 - jo) 1 + jo) (1 - jo)) 2 D) The lifetime X of a certain machine has a beta distribution with parameters $a = fl = 3$. Calculate the failure rate of this machine at time $x = 1/2$. a) 7/2 b) 15/4 c) 4 d) 17/4 e) 9/2 Question no. 76 In a certain region, there are 1000 pylons (supporting an electric line), numbered from 1 to 1000. We estimate that when a thunderstorm occurs, the probability that a given pylon is hit by lightning is equal to 10 -4, independently from one pylon to another. A) What is the probability that at least two pylons are hit by lightning during a thunderstorm? a) 0.0027 b) 0.0037 c) 0.0047 d) 0.0057 e) 0.0067 B) Use a Poisson distribution to calculate (approximately) the probability that exactly two pylons are hit by lightning during a thunderstorm. a) 0.004524 b) 0.005524 c) 0.006524 d) 0.007524 e) 0.008524 C) What is the variance of the number of thunderstorms until pylon no. 1 or pylon no. 2 is hit by lightning for the first time, from a given time instant? a) 24,997,199.94 b) 24,997,299.94 c) 24,997,399.94 d) 24,997,499.94 e) 24,997,599.94 D) Suppose that the nu.mber of thunderstorms that occur over a one-year period is a random variable having a Poisson distribution with parameter $c = 50$. What is the probability that a given pylon is hit by lightning at least once during a certain year? a) 0.0030 b) 0.0035 c) 0.0040 d) 0.0045 e) 0.0050 Question no. 77 We are interested in the temperature (in degrees Celsius) in a certain city over the months of July, September and December.

154 3 Random Variables A) First, we suppose that in July, when the temperature exceeds 30 �C, the number of degrees above 30 that the temperature reaches has an exponential distribution with mean 3. What is the probability that the temperature is higher than 35 degrees, knowing that it is higher than 31 degrees? a) 0.2236 b) 0.2336 c) 0.2436 d) 0.2536 e) 0.2636 B) Next, we suppose that, during the month of September, the temperature has a Gaussian distribution with parameters/ = 15 and c 2 = 25. What is the temperature that is exceeded only 1% of the time (over a long period) in this city in September? Hint. We have: Q-(0.02) _ 2.054, Q-(0.01) _ 2.326 and Q-(0.005) _ 2.576. a) 25.63 b) 26.63 c) 27.63 d) 28.63 e) 29.63 C) Finally, we seek a model for the temperature T in December, as well as a model for T when T > 0. We consider the following distributions: 1) U 10, 10]" 2) U [5 - 5,,/3, 5 + 5,,/] ; 3) Exp(l/5); 4) G(5, 1); 5)N(5, 25). Among these models, which one seems the most appropriate for i) T if E[T] = 5 and STD[T] = 5? ii)

U:=T [{T >0}if, instead, wehave:E[T IT >0]=STD[T IT >0]=5? a) i)2;ii) 1 b) i)2;ii)3 c) i)3;ii)5 d) i)5;ii)3 e) i)5;ii)4 Question no. 78 Let X be a continuous random variable taking its values in the interval [0, x). We say that X has a Pareto distribution with parameter 0 > 0 if its probability density function is of the form ifx _>0, fx(x)= (1 +x) + 0 elsewhere, In economics, the Pareto distribution is used to represent the (poor) distribution of wealth. Suppose that, in a certain country, the wealth X of an individual (in thousands of dollars) has a Pareto distribution with parameter 0 ----- 1.2. A) Calculate fx(2 [1 < X < 3), a) 0.3754 b) 0.3954 c) 0.4154 d) 0.4354 e) 0.4554 B) What is the median wealth in this country? a) $781.80 b) $1781.80 c) $2781.80 d) $3781.80 e) $4781.80 C) We find that about 11.65% of the population has a personal fortune of at least $5000, which is the average wealth in this population. What percentage of the total wealth of this country do these 11.65% of the population own? a) 75.5% b) 77.5% c) 79.5% d) 81.5% e) 83.5%

3.10 Exercises, Problems, and Multiple Choice Questions 155 Question no. 79 Let X be a discrete random variable with probability mass function given by 1 Px(k)= fork=O, 1 ..... 0 otherwise. A) Calculate the probability that X takes on a value which is a multiple of the number 3, that is, a value in the set {3, 6, 9 .... }. a) 1/14 b) 1/12 c) 1/10 d) 1/8 e) 1/6 B) We generate (independent) random numbers according to the distribution of the random variable X. Let M be the number of random numbers that are greater than 1, among the first ten numbers generated. i) Calculate P[M = 2]. a) 0.2416 b) 0.2616 c) 0.2816 d) 0.3016 e) 0.3216 ii) Suppose that we approximate the probability P[M = k] by P[N = k], for k = 0, 1 ..... 10, where N is a random variable that has a Poisson distribution with parameter ot = 2.5. What can we assert concerning P[N = k] with respect to P[M = k] for any value k taken from the set {0, 1 ..... 10}? a) PIN = k] is strictly less than P[M = k]; b) P[N = k] can only be less than or equal to P[M = k]; c) PIN = k] is strictly greater than P[M = k]; d) P[N = k] can only be greater than or equal to P[M = k]; e) we cannot assert anything. Question no. 80 An electroluminescent diode emits light at a wavelength X (in microns) that is a continuous random variable taking its values in the interval (0, A) What is the distribution of X, if the quantile of order p of X is given by the formula 1 Xp = (1 +p), where0<p < 1? a) U(0, 1/2) b) U(0,1) c) U(1/2, 1) d) U(1/2, 3/2) e) U(1, 3/2) B) Suppose that the probability density function of X is strictly decreasing and is bounded (above) by a positive constant k. Which one, among the models below, seems the most appropriate? a) 6(1/2,

1/2) b) G(1, 1/2) c) 6(3/2, 1/2) d) 6(3/2, 1) e) 6(2, 1) C) Suppose that X has a Gaussian distribution, with parameters /x = 1 and 2 = 1/16, to which its negative part and the point 0 have been subtracted, so that c {--8(x 1) 2} forx > 0, fx(x) = - exp - where c is a constant such that fx is a valid probability density function.

156 3 Random Variables i) Find the constant c. a) l/qb(1) b) 2/qb(2) c) 4/qb(2) d) 2/qb(4) e) 4/qb(4) ii) Suppose that the diode is considered defective if the value taken by the random variable X is such that IX - 1 [> 1/4. Calculate, in terms of c, the probability that a diode is defective. Hint. We have: qb(1) _ 0.8413 and qb(2) _ 0.9772. a) 0.17c b) 0.24c c) 1 - 0.68c d) 1 - 0.24c e) 1 - 0.17c Question no. 81 Let X be a continuous random variable having the probability density function fx(x)= exp -2 ifx >0, 0 elsewhere. We say that X has a Rayleigh distribution with parameter 0 > 0. We can show that E[X] = 0 q%-/2 and VAR[X] = 02 [2 - (zr/2)]. A) Let Y :---- In X, where X has a Rayleigh distribution with parameter 0 = 1. i) Find J'(l). a) 0.1637 b) 0.1837 c) 0.2037 d) 0.2237 e) 0.2437 ii) Calculate the mathematical expectation E[e jc◆Y] at w = -2j. a) 1/2 b) l c) 3/2 d) 2 e) 5/2 B) We define Z ---- I/X. Calculate the mathematical expectation of Z, if0 ---- 1 as in A). a)/l/zr b) c) d) e)x C) What is the value of the reliability function of a system, whose lifetime has a Rayleigh distribution with parameter 0 ---- 2, at the moment that corresponds to its average lifetime? a) e -r/8 b) e -r/a c) e -r/2 d) e -r e) e -2r

4 Random Vectors 4.1 Introduction Definition 4.1.1. A function X that assigns a vector (X l (s) ..... Xn (s)) of real num- bers to each outcome s in a sample space S associated with a random experiment E is called a random vector of dimension n. Each component of the random vector is a random variable. The set of all possible values of X, denoted by Sx, is a subset of n (see Fig. 4.1). Example 4.1.1 Let E be the random experiment that consists in observing the length of a program submitted for execution to a computer, as well as the execution time. Then, the elements of the sample space S are of the form s = (n, t), where n can be the number of lines in the program and t the execution time in seconds. Now, let X := (X1, X2), where X(s) = n and X2(s) = t. The function X, which here S , .X(s) t sx X l Sx .s x, i -..,.xos) Figure 4.1. Notion of random vector.

158 4 Random Vectors is the identity function, is a random vector of dimension

2. In this example, XI is a discrete random variable, whereas X2 is (theoretically) continuous. In general, the components Xk of a random vector are random variables of the same type. Next, let E* be the random experiment that consists in observing the execution time only. Then, if s : t, where t is in seconds, we can nevertheless define a ran- dom vector X = (Xl(s) ..... Xn(s)). For instance, we can have: Xl(S) = s (= t), Xz(s) : s 2, X.(s) : min{s, 10}, X4(s) = --S 3, etc. Thus, there is no relation- ship between the dimension of the elements s of S and the dimension of the random vector X. Remarks. i) In reality, the variable X2 above is a random variable of mixed type because the execution time can be equal to zero (when there is a fatal error in the program, for example), with a strictly positive probability, or it can take on any real positive value. In practice, the execution time is rounded off, so that X2 becomes a random variable of discrete type. ii) For simplicity, we will not consider in this book the case when one (or many) random variable(s) in the random vector is (are) of mixed type. 4.2 Random Vectors of Dimension 2 I) Discrete Case Definition 4,2.1. Let Z := (X, Y) be a random vector of dimension 2. If Sz is a finite or denumerablv in.finite set of points in the plane, that is, if we can write Sz in the form Sz={(xj,yk),j: 1,2 .... ;k= 1,2 .... }, then we say that Z is a discrete random vector (of dimension 2). Remark. We may also write that SZ : SX�y : {(xj, Yk): xj E Sx, Yk Definition 4,22. Let (X, Y) be a discrete random vector. The joint probability mass function of(X, Y) is defined by px.v(xj, Yk) = P[{X : xj} {Y = Yk}] :- P[X = xj, Y : Yk] (4.1) forj, k : 1,2 ..... Properties. i) pX, y(xj, Yk) >-- 0 V(xj, Yk) (because it is a probability). ii) Since the set of points (x j, Yk) constitutes a partition of Sz, we have: Z Z px, r(xj, Yk) : 1. (4.2) j:l k:l

4.2 Random Vectors of Dimension 2 159 Let A be an event with respect to Sz. That is, A C Sz. The probability of A is given by the formula P[A] = Z px,r(x), Yk). (4.3) (xj,Yk)EA Now, making use of the total probability rule, we can write that pX(Xj) = P[X = x j] = Z P [{X : xj } { Y ---- y }] = Z Px, " (x j, y). (4.4) k=l k=l Likewise, we have: p,(y) = p[y = y] = Z Pt{X = xj} f) {Y = Yk}l = Z px,'(xj, Yk). (4.5) j=l j=l Definition 42=3. The function obtained by summing the function Px," over all pos- sible values of either of the two random variables is called a marginal probability mass function. Example 4.2.1 A box contains six transistors, including one of brand A and one of brand B. Two transistors are taken at random and without replacement. Let X (re- spectively Y) be the number of brand A (resp. B) transistors among the two

picked. Then, the joint probability mass function Px,r of the random vector Z := (X, Y) can be given in the form of a table in two dimensions as follows: y\x 0 1 2 0 16/36 8/36 1/36 1 8/36 2/36 0 2 1/36 0 0 For instance, we find that px# (O, 0) = 16/36, because there are 4 chances in 6 that the first transistor picked is neither a brand A nor a brand B transistor, and the two draws are independent. Moreover, there is 1 chance in 6 that the first transistor picked is of brand A; it follows that px,,(1, 0) = 2 x (1/6)(4/6) = 8/36 (by symmetry), etc. We can check that Px,'(j, k) > 0 for j, k = 0, 1,2, and that the sum of all the fractions in the table is equal to 1. Note that the random variables X and Y both have a binomial distribution with parameters n = 2 and p = 1/6. Next, let, for example, A = {X + Y > 1}. We have then: P[A] = 1 - P[A c] = 1 - px,,(O, 0) = 1 - 16/36 = 20/36 = 5/9,

160 4 Random Vectors because there is a single element of Sz in A c, namely (0, 0). Similarly, we find that 2 P[X = Y] = Zpx,r(j,j) = 16/36+2/36+0 = 1/2. j=0 Finally, in such a case the marginal probability mass functions are obtained by adding the elements of the lines and of the columns in the table. For example, adding the three elements of each column in the table, we deduce that j 0 1 2 E Px(j) 25/36 10/36 1/36 1 Furthermore, by symmetry, we can write that pr(k) = px(k) for k = O, 1, 2. II) Continuous Case Definition 4.2.4. Let Z := (X, Y) be a random vector of dimension 2. If Sz is a non-denumerably infinite subset of the plane, we say that Z is a continuous random vector. (We assume that X and Y are two continuous random variables.) Definition 4.2_g. Let Z := (X, Y) be a continuous random vector, and let A C Sz. The probability of the event A is given by the formula PIal=_PIZeal=fffx, r(x,y)axay. (4.6) A where fx, r(x, y) is called the joint (probability) density function of the pair (x, r'). For example, let fx'r(x'Y)={e- -vifx>O'y>O'elsewhere. Then, we have (see Fig. 4.2, p. 161): P[Y > X] = e -x-y dxdy = e -x-y dydx JO do 2 = e -x-x2 dx = e 1/4 e -(x+�)2 dx :el/4v/P[N(-�,�) > O] -- 0.55. Properties. i) We have: fx, r(x, y) P[x < X < x + dx, y < Y < y + dy] dxdy (4.7)

4.2 Random Vectors of Dimension 2 161 Y >x2 / y=x 2 Figure 4.2. Example of computation of probability in two dimensions. (the equality being obtained by taking the limit as dx and dy decrease to 0). It follows that the function fx, Y (x, y) is non-negative. ii) Since X and Y take on real values, we can write that f_f_fx, y(x,y)dxdy=l. (4.8) Definition 4.2.6. Let (X, Y) be a random vector. The joint distribution function of the pair (X, Y) is defined by Fx, r(x, y) = P[X

$\leq x, Y \leq y]$. (4.9) If (X, Y) is a discrete random vector, then $F_{X,Y}(x, y) = p_{X,Y}(x_j, y)$, (4.10) $x_j \leq x$ $y_k \leq y$ while when (X, Y) is continuous, $F_{X,Y}(x, y) = f_{X,Y}(u, v)\,dv\,du$. (4.11) Remark. We deduce from Equation (4.11) that $\partial^2 F_{X,Y}(x, y)$ $f_{X,Y}(x, y) = \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$ -- (4.12) $\partial x \partial y$ for all pairs (x, y) where the second mixed partial derivative exists. Properties. The function $F_{X,Y}$ has properties similar to those of the function F_X. i) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$ (impossible events); ii) $F_{X,Y}(\infty,\infty) = 1$ (certain event); iii) $F_{X,Y}(X_1, Y_1) \leq F_{X,Y}(x_2, Y_2)$ if $X_1 \leq x_2$ and $Y_1 \leq Y_2$ (because, in this case, $\{X \leq X_1, Y \leq Y_1\} \subset \{X \leq x_2, Y \leq Y_2\}$; see Fig. 4.3, p. 162).

162 4 Random Vectors Y (x,y) Figure 423. Non-decreasingness of the joint distribution function. iv) We can show that $\lim F_{X,Y}(x + \delta, y) = \lim F_{X,Y}(x, y + \delta) = F_{X,Y}(x, y)$. Remark. We can also check, graphically, that $P[a < X \leq b, c < Y \leq d] = F_{X,Y}(b, d) - F_{X,Y}(b, c) - F_{X,Y}(a, d) + F_{X,Y}(a,$ (4.13) Now, we have: $F_X(x) =$ -- $P[X \leq x] = P[X \leq x, Y < \infty] = F_{X,Y}(x, \infty)$. (4.14) Likewise, we obtain: $F_Y(y) = F_{X,Y}(\infty, y)$. Definition 4.2.7. The function obtained from $F_{X,Y}(x, y)$ by replacing y (respectively x) by infinity is called the marginal distribution function of X (resp. Y). Finally, we can write (see Remark i) below) that d $\frac{d}{dx} f_X(x) = $ "$_{-\infty}^r$ $f_X(x) =_{-\infty}^r f_{X,Y}(x,] =$ -- $f_{X,Y}(u, v)\,dv\,du = f_{X,Y}(x, v)\,dv$. That is, $f_X(x) = f_{X,Y}(x, y)\,dy$. (4.15) Remarks. i) In Equation (4.15), we must actually integrate over all possible values of Y, given x, a fixed value of X. Similarly, the function $f_Y(Y)$ is obtained by integrating $f_{X,Y}(x, y)$ with respect to x, from $-\infty$ to

112 4.2 Random Vectors of Dimension 2 I 112 $y=x$ Y 1 $x<y$ an.d $+y <1$ --$y=x$ $x+y=1$) X Figure 4.4. Figures for Example 4.2.2. 163 ii) Let $u_2(y)$ $I(y) := \int_{u_1(y)} g(x, y)\,dx$. Then, under certain conditions, we have the following formula (due to Leibniz): 1 $\dfrac{dI(y)}{dy} = \int_{u_1(y)}^{u_2(y)} \dfrac{\partial g(x, y)}{\partial y}\,dx + g(u_2(y), y)\dfrac{du_2(y)}{dy} - g(u_1(y), y)\dfrac{du_1(y)}{dy}$ --- Definition 4.2.8. The functions f_X and f_Y obtained from $f_{X,Y}(x, y)$ as above are called the marginal (probability) density functions of X and Y, respectively. Example 4.2.2 Let 2 if $0 < x < y < 1$, $f_{X,Y}(x, y) = 0$ elsewhere (see Fig. 4.4). Since $f_{X,Y}(x, y) $ ---- 2 and the area of the triangle on which the function is positive is equal to $1/2$, the double integral of $f_{X,Y}(x, y)$ over $S_X \times S_Y$ is equal to 1, as required. We have: $f_X(x) = $ fl $f_Y(y) = \int_0^y$ $2\,dy = 2(1-x)$ if $0 < x < 1$ (and $f_X(x) = 0$ elsewhere), $2\,dx = 2y$ if $0 < y < 1$ (and $f_Y(y) = $ --- --- 0 elsewhere). † Gottfried Wilhelm von Leibniz, 1646-1716, was born and died in Germany. He was at once mathematician, philosopher and theologian.

With Newton, he is the inventor of infinites- imal calculus. He introduced the notation for the derivative and the integral. Among his great achievements in mathematics, we count the development of the binary system and his work on determinants. He also contributed to mechanics, more precisely to dynamics. He corresponded with more than 600 European scientists. In the field of religion, he set himself the objective of reuniting the various branches of the Christian church.

164 4 Random Vectors Moreover, we can write (see Fig. 4.4, p. 163) that P[X + Y < 1]= 2dydx = 2(1 - 2x)dx = l/2. JO JO Remark. In this case, we could have computed the area of the checkered region in Fig. 4.4 and multiplied it by 2 instead. Finally, we also find that Fx,r(x, y) = 0 ifx <0ory <0, 2xy-x 2ifO<x <y < 1, 2x-x 2 if0<x < landy> 1, y2 if0<y < landx > y, 1 ifx > 1 andy> 1. Indeed, for any point (x0, Y0) in the region Sx � r, the area of the checkered region in Fig. 4.5, is given by I x o + (Yo - xo)xo. Thus, we do have: Fx.r(x,y)=2� (x2+(y-x)x) =2xy-x 2 if0<x<y< 1. The value of the function Fx, r (x, y) can then be obtained by reasoning for any point (x,y) 6 N 2. First, if 0 < x < 1 buty > l, we only have to replace y by 1 in the above formula, Next, if 0 < y < 1 but x > y, we replace x by its maximum value, namely y, in the above formula, The other two cases are elementary. We will complete this section with the notion of independence applied to random variables. Definition 4.2.9. Let (X, Y) be a random vector. We say that X and Y are indepen- dent random variables if Fx,r(x, y) = Fx(x)Fr(y) �(x, y). (4.16) 112 "-- x<y my=x) 1/2 1 X Figure 4.5. Joint distribution function in Example 4.2.2.

4.2 Random Vectors of Dimension 2 165 Equation (4.16) is equivalent to px, y(xj, Yk) px(xj)pY(yk) if(X, Y) is a discrete random vector, fx,y(x, y) = fx(x)fy(y) if(X, Y) is a continuous random vector. Remark. We can say, more generally, that X and Y are independent if P[X A, Y B] = P[X A]P[Y B], where A (respectively B) is an arbitrary event that involves only X (resp. Y). Defini- tion 4.2.9 then becomes a proposition that we can easily prove. Proposition 4.2.1. If X and Y are two independent random variables, then so are g(X) = X* and h(Y) = Y*. Proof. We have: FX*,y*(X*, y*) = P[X* <_ x*, Y* <_ y*] = P[X A, Y B], where {X 6 A} is the event in Sx that is equivalent to {X* < x*}, and {Y 6 B} C Sr is equivalent to {Y* < y*}. Then, given that X and Y are independent, we can write that Fx,,r,(x*, y*) = P[X AIP[Y B] = P[X* <_ x*IP[Y* <_ y*]. [] Example 4.2.3 The two random variables in Example

4.2.1 are not independent, because, in particular, px, r(2, 2) = 0 ⬥ (1/36) (1/36) = px(2)pr(2). Remark. In such a case, if there is one or more O's in the table, then the random variables are not independent. We can also notice that we must have: $0 < X\,Y < 2$, which directly implies that X and Y are not independent since, for instance, X = 2 = Y = 0 (because Y >_ 0). Example 4.2.4 Consider once again the function fx, r in Example 4.2.2. We already found that fx(x) = 2(1 - x), 0 < x < 1, and fr(Y) = 2y, 0 < y < 1. Since fx(O)fr(O) = 2 ⬥ 0 = 0 :/: 2 = fx,r(O, 0), X and Y are not independent random variables. Remark. Given that there is a relationship between x and y in the definition of Sxxr, namely x < y, X and Y could not be independent. We can show the following proposition.

166 4 Random Vectors Proposition 4.2.2. Let (X, Y) be a continuous random vector. Suppose that Sxx r is of the form Sxxr = {(x, y): a < x < b, c < y < d}, where a, b, c and d are constants. Then, X and Y are independent random variables if and only if (iff) we can write that fx, r(x, y) =-- g(x)h(y), (4.17) where g(x) > O iff a < x < b and h(y) > O iff c < y < d. Remark. An analogous result exists in the discrete case. 4.3 Conditionals Conditional Distribution, Probability Mass and Density Functions We discussed the notion of conditional distribution function in Section 3.2. If Y is a discrete random variable, then we can write, directly, that the conditional distribu- tion function of X, given that Y = Yk, is PlX < x, Y = y:] xj<_x Px,r(xj, yk) FxIY(X [yk) = = if pr(yk) > O. PlY = Y,] Pr(Y,) (4.18) Remark. Some authors use the notation Fx(x [Y = yk), as Fx(x [A) in Chapter 3, rather than Fxlr(x [yk). Similarly, if (X, Y) is a discrete random vector, we define the conditional prob- ability mass function of X, given that Y = yk, by Px, r (X j, yk) PXIY(Xj l Yk) -- ifpr(yk) > O. (4.19) Pr(Yk) Remark. This function has the same properties as the (marginal) probability mass functions. Furthermore, the conditional probability mass function of Y, given that X = x j, is defined in a similar way. If Y is a continuous random variable, then P[Y = y] = 0 Yy. We must therefore consider the event {y < Y <_ y dy} in (4.18) and take the limit as dy decreases to zero. If X also is a continuous random variable, we obtain the proposition that follows. Proposition 4.3.1. Let (X, Y) be a continuous random vector. The conditional dis- tribution function of X, given that Y = y, is given by jx fx, v(u, y) du FxIr(x l Y) = iffr(y) > O. (4.20) fv(Y)

4.3 Conditionals 167 Corollary 4.3.1. Let (X, Y) be a continuous random

vector. The conditional (prob- ability) density function of X, given that $Y = y$, is obtained by differentiating the function $Fxiy(x \mid y)$ above with respect to x: $fx,r(x, y)$ $fxIr(x \mid y)$ -- if $fr(y) > 0$, (4.21) $fY(Y)$ for any point x where $Fxlg(x \mid y)$ is differentiable. Proposition 4.3.2. Let (X, Y) be a random vector. X and Y are independent random variables if and only if $Fxly(x \mid y) = Fx(x)$ ◆ (x, y) (or $Frlx(y \mid x) = Fr(y)$). (4.22) If (X, Y) is a discrete random vector, this condition is equivalent to $PXIy(Xj \mid Yk) = px(Xj)$ V $(X j, y)$, (4.23) while Equation (4.22) is equivalent to $fxlr(x \mid y) = fx(x)$ ◆ (x, y) (4.24) when (X, Y) is a continuous random vector. Proof. This follows directly from Equations (4.18)-(4.21), and from the definition of independence. [] Example 4.3.1 Consider once again the function fx,r in Example 4.2.2: $fx,y(x,y)=2$ if$0<x <y< 1$ (=0elsewhere). The conditional probability density function of X, given that Y $= y$, is $fx, r(x, y)$ 2 1 $fxly(x \mid y)$ -- -- -- if$0 < x < y$ (5 1). $fr(Y)$ 2y y That is, $X \mid \{Y = y\}$ has a U$[0, y]$ distribution, where y E $[0, 1]$. Conditional Expectation As we mentioned in Section 3.7, the concept of mathematical expectation can be generalized by considering the conditional expectation of a random variable. Definition 4.3.1. Let (X, Y) be a pair of random variables. The conditional expec- tation of X, given that $Y = y$, is given by -xjpxIr(Jcj $\mid$ y) if(X, Y) is discrete, E[X] Y----y] = j=l (4.25) fS $xfxly(x \mid y)$if(X, Y) dx is continuous.

168 4 Random Vectors Now, let E[X $\mid$ Y] be a function of the random variable Y that takes on the value E[X $\mid$ Y = y] when $Y = y$. Then, E[X $\mid$ Y] also is a random variable. We could denote this random variable by g(Y), so that $g(y) = $ E[X $\mid$ Y = y], as in Section 3.6 (p. 92). The following result is important in probability. Proposition 4.3.3. Let (X, Y) be a pair of random variables. We have: E[X] : E [E[X $\mid$Y]]. (4.26) Proof. Consider the case when (X, Y) is a continuous random vector. We have: E[E[X I rll = E[Xlr = y]fr(y)dy = xfxlr(x [y)dx fr(y)dy. Now, fxlr(x $\mid$ Y)fr(Y) = fx,r(x, y). Thus, we can write (changing the order of integration) that 1 E[E[X $\mid$ YI]: fx.r(x, y)dy dx = xfx(x)dx : E[X]. The proof in the other cases is analogous. [] Remarks. i) The previous proposition is also true when we replace X by a function h(X). That is, E[E[h(X) I Y]] = E[h(X)]. (4.27) ii) In the case when Y is a discrete random variable, Equation (4.26) becomes oo E[X] = E[E[X $\mid$ Y]] = Z E[X $\mid$ Y = yk]PY(Yk). (4.28) kl iii) We can also write that VAR[X] = E[X 2] - (E[X]) 2 = E[E[X 2 $\mid$ Y]I - (E[E[X I y]])2. (4.29) Finally, let A be an event and W be the indicator variable of the event $\{X E A\}$. That is, {lifXEA, W = 0 otherwise.

(4.30) Since E[W] = P[X A], we deduce from Proposition 4.3.3 the following result, which is the equivalent of the total probability rule for random variables. Corollary 4.3.2. a) Let Y be a discrete random variable. Then, we can write that $P[X A] = Z P[X a [Y : Yk]P_Y(Y_k)$. (4.31) k=l

4.3 Conditionals 169 b) If Y is continuous, Equation (4.31) becomes $P[X A] = P[X A I Y = y]f_r(y)dy$. 0(3 (4.32) Remark. We also define the conditional variance of X, given the random variable Y, by $VAR[X I Y] ---- E[(X -- E[X I y])2 I Y] = E[X2 I Y] - (E[X I y])2$. (4.33) We can show that $VAR[X] = E[VAR[X I Y]] + VAR[E[X I Y]]$. (4.34) Example 4.3.2 Consider one last time the function $f_{x,r}$ in Example 4.2.2: $f_{x,r}(x,y)=2$ if0<x <y5 1 (=0elsewhere). Making use of Example 4.3.1, we can write that the conditional expectation of X, given that Y ---- y, is $E[X | Y = y]= y/2$, because X I {Y = Y} " U[0, y]. Moreover, we have: fo 'y $E[E[X I Y]] = E[Y/2] = $ � 2ydy = 1/3, which is indeed equal to $e[xl = x .2(1 -x)d = 1/3$. 3 1o Example 4.3.3 Suppose that the number X of customers who visit a car dealer during a one-day period has a Poisson distribution with parameter o = 20. Furthermore, suppose that the probability that a customer buys a car (on a given visit) is equal to 1/10, independently from one person to another. Let Y be the number of cars sold during one day. We can write that $Y I X -- B(n = X, p = 0.1)$. It follows that $E[Y] = E[E[Y I X]] = E[0.1X] = 0.1$ � 20 = 2. Moreover, since $VAR[Y I X] = X (0.1)(0.9)$, we deduce from Equation (4.34) that $VAR[Y] = E[(O.Og)X] + VAR[(0.1)X] = (0.09)(20) + (0.01)(20) = 2$. Finally, we can calculate P[Y> 1] = 1-P[Y=O]= 1-P[Y=O X=klP[X=k] k=O = 1 -- P[B(n = k, p = 0.1) = O]px(k) k=O

170 4 Random Vectors = I -- e -20 Z 18k = I -- e -2 -- 0.8647. k! k=0 Example 4,3.4 A number X is taken at random in the interval [0, 1], then a number Y is taken at random in the interval [0, X]. Thus, we can write that X U[0, I] and Y I X U[0, X]. Then, we obtain that $E[Y] = EIE[Y I XI] = E[X/2] = (I/2)(1/2) = I/4$. Next, in this case we have: $VAR[Y I X] = X 2$, so that l f0 l x 2 VAR[X] VAR[r] = El X2/121 + VAR [X/2I = � 1 dx + 1 1 1 1 7 = 12 3 4 12 144 Finally, we compute PlY > 1/21= PlY > 1/21X=x]fx(x)dx = PlY > 1/21X:x].ldx I l X 2 1 lnx I = .ldx=x-- -- /2 x 2 /2 1 ln2 0.1534. 2 2 Remark. We have: 1 1 $f_{x, r}(x,y)=f_{r|x}(y|x)f_x(x)=-.l =-$ if0_<y<x < 1. X X 4.4 Random Vectors of Dimension n > 2 All the definitions that we have seen in the case of pairs of random variables can easily be generalized to the case of random

vectors of dimension n, where n > 2. For instance, the joint probability density function of a continuous random vector X = (XI ..... Xn) is a non-negative function fx(xl ..... Xn) defined on a non- denumerably infinite subset Sx of R n, such that f f ... f, ... xn)dXl ...dxn = l. X

4.4 Random Vectors of Dimension n > 2 171 The marginal probability density function of Xk is obtained by integrating the function fx with respect to all variables, except Xk, for k = 1 ..... n: fxk (Xk) .... fx(xl ..... Xn) dxl ... dXk-ldXk+l ... dxn. (n-- 1) times Likewise, we have: fX,X2 (Xl, X2) .... fX(Xl ..... Xn) dx3.., dxn, (n--2) times etc. That is, we integrate the function fx with respect to all the variables that we want to remove from the random vector X. Finally, we say that the random variables X1 ..... Xn are (globally) indepen-dent if and only if P[X1 E A1 ..... Xn An] = -I P[XI AI], k=l (4.35) where A is an event that involves only X, for k = 1 ..... n. IfX = (X 1 ..... Xn) is a continuous random vector, this condition implies that fX(X1 ..... Xn) I fXk (Xk). k=l Moreover, as in the case of (random) events, some variables in a random vector may be independent and others not, as in the following example. Example 4.4.1 Let Ik(x +y)e -z if0 < x < 1,0 < y < 2, z > 0, fw(x, y, z) I 0 elsewhere be the joint probability density function of the continuous random vector W = (X, Y, Z). Using the fact that the triple integral of fw over Sw must be equal to 1, we find that the constant k is equal to 1/3. In addition, we can calculate, for instance, the probability f� [fO1 fx2 (x] 5 " 0.3066. P[X < Y,Z > 1]= +y) dydx e -zdz= e -1 6 Remark. Here, if we integrate with respect to the variable x before the variable y, then we must divide the double integral into two parts:

172 4 Random Vectors fIfolfo yl f2f01 1 P[X < Y, Z > 1] = -(x+y)dxdy+ -(x+y)dxdy e-Zdz. Furthermore, we can assert that the random variable Z is in fact independent of the variables X and Y (or, equivalently, of the random vector (X, Y)). Example 4.4.2 Let S be a sample space associated with a random experiment E, and let B1 ..... B4 be a partition of S. Suppose that we perform m independent repetitions of the experiment E and that the probability Pk :---- P[Bk] is constant from one repetition to another. Then, if Xk denotes the number of times that the event Bk has occurred among the m repetitions, for k = I ..... 4, we can write that P[XI = x, X 2 = x2, X3 = x3] := pX(Xl'X2'X3) = Xl!X2!X3!(m)1 --x2--x3)! Pl P2 P3P4 where x, x2, x3 {0. I ..... m } and x +x2 +x3 < m. This three-dimensional prob- ability distribution, called the

quadrinomial distribution with parameters p .... , P4, can easily be generalized to the case when the partition of S is made up of n events. In this case, we say that the random vector (X1 ..... X,,) has a multinomial distri- bution with parameters p ..... Pn (where $0 < Pk < I$ Yk and p + ... + Pn = I). It is actually a distribution of dimension n - I (because the sum of the Xk's is fixed). The binomial distribution is thus the particular case when n is equal to 2. Remark. The term m! Xl !XZ!X3!(m - Xl - x2 - x3)! in the function px(x, x2, x3) above is what we called the multinomial coefficients in Chapter 2. In the case of the binomial distribution, the corresponding term is ex- pressed in terms of the combinations of k objects taken (without replacement) among n distinct objects. Note, however, that if we have n objects, of which k are of type I and n - k of type II, then the number of permutations of the n objects taken all at once is given by n! k!(n - k)! -- C. Thus, the number of permutations of n objects of two types, of which k are of type I, is the same as the number of combinations of k objects taken (without replacement) among n distinct objects. An n-dimensional probability distribution that is very important for the applica- tions is the multinormal distribution, which will be seen in Section 4.7 (p. 179). 4.5 Transformations of Random Vectors In the case when X is a discrete random vector, we can proceed as in Section 3.6 to find the joint probability mass function of the new random vector Y := g(X). That

4.5 Transformations of Random Vectors 173 is, we apply the transformation g to each vector in Sx and we add the probabilities of the vectors (xl,... , xn) that correspond to the same vector (yl .... , Ym). Note that the dimension of the random vector Y is not necessarily the same as that of X. For example, consider the random vector (X, X2) in Example 4.2.1, p. 159. Let Y = (Y, Y2) := (X + X2, X - X2). Then, we find that the joint probability mass function of (Y1, Y2) is given by the following table: y2\y 0 1 2 -2 0 0 1/36 - 1 0 8/36 0 0 16/36 0 2/36 1 0 8/36 0 2 0 0 1/36 Indeed, in this case the six pairs in Sx �x2 are transformed as follows: (0, 0) -- (0, 0), (0, 1) -- (1, -1), (0, 2) -- (2, -2), (1,0) -- (1, 1), (1, 1) -- (2, 0), (2, 0) -- (2, 2). Thus, here the transformation g is bijective. Now, suppose that X is a continuous random vector. To obtain the distribu- tion function of Z := g(X ..... Xn), we can find the event A = {(x ..... x,): g(x ..... x,,) _< z} C Sx which is equivalent to the event {Z < z}. Then, we write: Fz(z)= f f ... fafX(x, ... xn) dXl'"dxn. (4.36) Similarly, if Zk := gk(X1 ..... X) for k = 1 ..... n, then to obtain the joint distribution function of the new random vector (Zl ..... Zn), we can find the event in Sx that is equivalent to

{Z1 < Zl ..... Z,, < z,,} and integrate the function fx over this event. We can also show the following proposition. Proposition 4.5.1. Let X and Y be two continuous random variables, and let V := gl(X, Y) and W := g2(X, Y). Suppose that 1) the system v = gl(x, y), w = g2(x, y) has a unique solution x = h(v, w), y = hz(v, w); 2) the functions g and g2 have continuous partial derivatives �(x, y), and the Jaco- bian of the transformation: Ogl/OX Ogl/Oy J(x,y):= Og2/Ox Og2/Oy

174 4 Random Vectors is different from zero 'v'(x, y). Then, we can write that fv, w(v, w) = fx, y(x, y) I J(x, y) [-1, where x = hi(v, to) and y = h2 (v, to). (4.37) Remarks. i) In the particular case when v = x + y and to = x - y, we find that x = (v + w)/2 and y = (v - w)/2. Moreover, in the case of such a linear transformation, the partial derivatives of the functions gl and g2 are constants. Consequently, they are continuous for any pair (x, y) R 2. ii) The Jacobian of the inverse transformation is given by Ohl/OV Ohl/Oto J*(v, to) := Oh2/Ov Oh2/to " We can show that [J*(v, w)[= [J(hl (v, to), h2(v, w))[-1 . Then, we can also write that fv.w(v, w) = J'x.y(hl (v, to), hz(v, w))[J*(v, w)[, (4.38) provided that J*(v, w) is not identical to zero. iii) When the system v = gl (x, y), w = g2(x, y) does not have a unique solution, we could apply the proposition on disjoint parts of the plane (where the solution is unique) and add all the terms obtained, as in Section 3.6. iv) The result in Proposition 4.5. I can easily be generalized to the case ofn functions of n continuous random variables X l ..... v) In the case when gk is a linear transformation 'v'k, we can write that Z := (gl(X) ..... gn(X)) = XA, where A is an n � n matrix. If A is invertible, we find that fz(zl ..... zn) = fx(xl ..... xn) I detA 1-1, (4.39) where (Xl ..... Xn) = (Zl ..... zn)A -1 . Note that we consider, in what precedes, that X (and thus Z) is a row vector. Example 4.5.1 Let X ExpO.) and Y ExpO.) be two independent random vari- ables, and let V := X + Y. To obtain the probability density function of V, we can make use of Proposition 4.5.1. We define an auxiliary variable: W = X. Then, the system v=x+y, to X has the unique solution x = w, y = v - w. Moreover, the partial derivatives of the functions gl and g2 are continuous 'v'(x, y) 6 2, and the Jacobian 11 J(x, y) = 1 0 : -1

4.5 Transformations of Random Vectors 175 is different from zero �(x, y). Then, we can write that fg,w(v, w) = fx,y(w, v - w) I -1 I -lind" fx(w)fy(v - w) := fv, w(v, w) =)e-ZW)e -z(v-w) =)2e-ZV, if w >_ 0 and v _> w. Finally, we

have: fv(v) =)2e-X dw =)2ve-X if v > 0. Remarks. i) We have, in fact: V := X + Y ii) Since the transformation considered is linear, we could have used (4.39) to obtain fv, w(v, w). iii) We must not forget to give the set of possible values of the new random vector (V, W), that is, Sv � w. For instance, if SxxY = {(x, y): O < x < y < 1} and if we consider the transformation then we find (see Fig. 4.6) that Sv�w = {(v, w): -v <w <0,0<v < lorv-2<w <0,1 <v <2}. Indeed, we simply have to consider each of the three line segments that define Sx� and find out what they become in the (v, w)-plane. For example, the line segment defined by x = 0, 0 < y < 1 becomes the line segment v= y,w=-y e* v=-w in the (v, w)-plane. 1/2 ' for0<v<l W 1 2 I I) W = -V 1/2 1 X Figure 4.6. Example of transformation of a random vector.

176 4 Random Vectors x+y<z " x+y=z Figure 4.7. Distribution function of the sum of two random variables. Remark. Let X and Y be two independent random variables, and let Z := X Y. Then, if X and Y are continuous, we can write (see Fig, 4.7) that FZ(Z) = P[X + Y <_ z] = fx(u) fr(v)dvdu. It follows that fz(z) = fx(u)fr(z - u)du. (4.40) Thus, we can state that the probability density function of the sum of two independent continuous random variables, X and Y, is the convolution product of the probability density functions of X and Y. We can use this result to calculate the probability density function of V in Example 4.5.1. We have: fv(v) = fx(u)fr(v - u)du =)e-U)e-(v-U) du =X 2 e -xv du = X2ve -xv ifv >0. We will see yet another way of obtaining this same result (and its generalization) in Section 4.9 (p. 187). 4.6 Covariance and Correlation Definition 4.6.1. Let (X, Y) be a random vector and Z := g(X, Y). The mathemat- ical expectation of Z is defined by E[Z] = Z Z g(xj, yk)px, y(xj, Yk) j=l k=l /f (X, Y) is discrete, f_ f_ g(x, y)fx, y(x, y) dxdy if(X, Y) is continuous. (4.41)

4.6 Covariance and Correlation 177 Remarks. i) The above definition of the expected value of Z is actually a theorem that we could prove. ii) We find, in particular, that E[aX + bY] = aE[X] + bE[Y] (4.42) for all constants a and b and for all random variables X and Y, and that E[gl (X)g2(Y)] = E[gl (X)]E[g2(Y)] (4.43) if X and Y are independent random variables. iii) The mathematical expectation E[XY] is called the correlation of X and Y. If E[XY] = 0, we say that X and Y are orthogonal. Definition 4.6.2. Let (X, Y) be a random vector. The covariance of X and Y is de- fined by COV[X, Y] crs, r = E[(X - E[X])(Y - E[Y])]. (4.44) Remarks. i) We have: COV[X, X] = VAR[X].

Thus, the covariance generalizes the variance. However, the covariance COV[X, Y] can be negative. ii) We easily show that COV[X, Y] = E[XY] - E[X]E[Y]. (4.45) It follows that if X and Y are independent, then COV[X, Y] = 0. Definition 4.6.3. Let (X, Y) be a random vector. The correlation coefficient of X and Y is defined by COV[X, Y] CORR[X, Y] ---- Px,Y = STD[X]STD[Y] (4.46) Remarks. i) To avoid any confusion between the correlation and the correlation coef- ficient, we will generally use the symbol Px,Y, rather than the operator CORR[X, Y]. ii) The correlation coefficient is a unitless (or dimensionless) measure of the linear relationship between X and Y. We can show that -1 <_ Px, Y < 1. Moreover, if Y = aX + b, then Px, Y = 1 if a > 0 and Px,Y = -- 1 if a < 0. iii) If X and Y are independent, then Px, Y = 0. However, the converse is not always true. If Px, r = 0, but X and Y are not independent, we say that they are simply uncorrelated. For example, let X U[-1, 1] and Y := X 2. We have then: E[XY] - E[X]E[Y] E[X 3] - 0. (1/3) x/(0 - 0) Px, Y = STD[X]STD[Y] x/l/3. STD[Y] STD[Y] because all odd moments of X are equal to zero, and the standard deviation STD[Y] is strictly positive (actually, STD[Y] = 2/V/). However, X and Y are certainly not

178 4 Random Vectors independent, since the variable Y is expressed in terms of X: Y = g(X) = X 2. In particular, we can write that X = 0 = Y = 0. Likewise, we can show that if X N(0, 1), then Px,x 2 = O. Note that Px,y = 0 is the only case that does not allow us to conclude about the independence (or not) of X and Y, since, if Px,y O, then we can state that X and Y are not independent. iv) An important particular case is the one when X and Y both have a Gaussian distribution and Px,r = 0. Indeed, we can then show (see Section 4.7, p. 179) that this implies that X and Y are independent. Thus, in the case when X and Y have Gaussian distributions, the random variables are independent if and only 'Px,r = O. In practice, we compute the point estimate (see Chapter 6, p. 253) rx, of Px. . If rx,y, called the sample correlation coefficient, is very small (in absolute value), and if the assumption of normality of the random variables seems reasonable, then we can assume the independence of X and Y when we analyze data, for example. We will see in Chapter 6 how to test, in particular, the hypothesis that a given random variable has (approximately) a Gaussian distribution. Example 4.6.1 Let I I if-y <x < y,O< y < 1, fx, r(x, y) = 0 elsewhere. We lind (see Fig. 4.8) that fx(x)= ldy= l-lxl if-I <x < 1, fr(y) = ldx=2y if0<y < 1. y Since fx(x) is symmetric with respect to 0 (and Sx = (- 1, 1), which is finite), we can write that E[X] = 0. Moreover, we have: E[XY] =

xy . 1 dxdy = O. Figure 4.8. Figures for Example 4.6.1.

4.7 Multinormal Distribution 179 Then, given that IE[Y]I < o (because Sr = (0, 1)) and STD[X]STD[Y] > 0 (because X and Y are not degenerate random variables), we can write that Px,r = O. However, X and Y are not independent, since fx(x)fy(y) = (1 - Ixl)2y 1 = fx, y(x, y) for almost all pairs (x, y). In fact, we could have stated from the outset that X and Y are not independent random variables, since there is the relationship -y < x < y between x and y in Sx � r. Remark. To obtain the correlation coefficient of the random variables X and Y, we must theoretically calculate five mathematical expectations: E[X], E[Y], E[X2], E[Y] and E[XY]. However, as can be seen in the above example, it is preferable to start by calculating the mean of XY and that of X or of Y. Indeed, if the mean E[XY] is equal to zero, as well as E[X] or E[Y], then it is not necessary to calcu- late the means of the squared variables (in the case when all expected values are real numbers and the random variables are not degenerate). 4.7 Multinormal Distribution Definition 4.7.1. Let (X, Y) be a continuous random vector. We say that the random variables X and Y have a bivariate normal distribution (or binormal distribu- tion) tf their joint probability density function is of the form { fx, r(x, y) = C exp 2(1 - p2) -2p(x-x)(Y-')]}x, (4.47) for - < x < , - < y < , where x , Y , X > O, y > O, -1 < p < 1,and (see Ng. 4.9, p. 180). We write: (X, g) N(x, r; , ;). Remarks. i) We can show that the pameter in the function fx, r above is (indeed) the correlation coecient of X and ii) We find, integrating the function fx, r(x, y) with respect to y, that X has a Gaus- sian distribution with pameters gx and . Likewise, we have: g N(gr,). We then easily obtain that X I{g = ,} N (,x + (x/r)(, -,r),(1-)). (4.49) We see that if = 0, then fxlr(x I Y) fx(x), from which we deduce the fol- lowing impoant result (see Section 4.6, p. 176): in the case of bivariate normal

180 4 Random Vectors Figure 4.9. Joint probability density function of a random vector having a bivariate normal distribution with kt X = kty = O, ax = ay = I and p = 0. distributions, if the correlation coefficient of X and Y is equal to zero, then X and Y are independent random variables. iii) We also deduce from (4.49) that E[X] Y = y] is of the form ay +b, where a and b are constants. We will see in Section 4.8 (p. 182) that this implies that, in the case of the bivariate normal distribution, the best linear estimator and the best non-linear estimator of X in terms of Y, namely E[X I Y], are identical. Example 4.7.1 Suppose that the lifetime X (in kilometers) and the thickness Y (in millimeters)

of the disc brakes of a certain company have a bivariate normal distri- bution with parameters #x = 20,000, # r = 10, cx = 1000, cr = 1 and p = 0.85. The engineer responsible for quality control wants to determine the lifetime of the brakes by measuring the thickness of the discs. If the thickness of a disc is equal to 9 mm, what is the probability that the brake will last at least 19,500 km? Solution. We seek P[X > 19, 500 I Y = 9] = P[Z > 19, 500], where Z N(19,150;277,500) Tab. 3.3 -- Q(0.66) = 1 - qb(0.66) _ 0.25. Note that the probability of the event {Y ---- 9} is zero, because Y is a continu- ous random variable. Therefore, we must not try to compute the above probability by making use of the definition of the conditional probability P[A I B] given in Chapter 2. Definition 4.7.2. We say that the random variables X i, X2 ..... Xn have a mul- tinormal distribution if they are a linear combination of the independent N(O, 1) random variables Z I ..... Zm, that is, if m Xk = #k -t- ckjZj for k = 1 ..... n, (4.50) where Iz k is a real constant for all values of k.

4.7 Multinormal Distribution 181 Now, we can show that the joint probability density function of the vector X = (X1 ..... Xn) is completely determined by the vector of means m and the covariance matrix K, where tn := (#l ..... #n) and VAR[X1] COV[X1,X2] ... COV[X1, Xn] 1 K := COV[X2, X1] VAR[X2] COV[X2, Xn] . (4.51) COV(, Xl] COVing, X2] VA[n] Remark. The matrix K is symmetric, because COV[X, Y] = COV[Y, X]. Moreover, it is non-negative definite. That is, CickCOV[Xi,Xk] 0 Ci,Ck . i=1 k=l In fact, we can also show that if K is a non-singul matrix, then fx(x) = (2)-/2(det K)-/2 exp [(- 1/2)(x - m)K - (x r - mr)] (4.52) for xk 6 k, where x := (Xl ..... Xn) and T denotes the transpose of the vector (or the matrix). Notation. We write: X N(m, K). The joint chacteristic function of the vector X, that is, X(Wl ..... Wn) := E [exp[j(wlXl +'" + wX)]], (4.53) is given by exp j Nioi -- ffikOiOk . i=1 i=1 k=l We may write that X(Wl ..... wn) = exp [jmw T -(1/2)wKw T] , (4.54) where w := (w ..... w). Remark. To show this result, we can use the fact (see Section 4.9, p. 187) that Z := X + ... + Xn N(z,a), where n Z : Oii i=1 d i k ik, i=1 k=l and that ifX N(, a2), then E[e jX] = x(1) = exp(j - (1/2)a2).

182 4 Random Vectors Properties. i) If the random vector X has a multinormal distribution and if Oik : 0 �i � k, then the random variables X, ... , Xn are independent. ii) If the random variables X ..... Xn have a multinormal distribution, then any set Y ..... Yk of random variables formed by linear combinations of the Xi's also has a multinormal distribution. More precisely, if

$Y : XA r$, where A is a $k \diamond n$ matrix of rank $k < n$ (that is, the k rows of A are linearly independent), then Y N(mA , AKAr). iii) IfX N(m, K), where K is of rank n, then there exists a matrix D such that $X = YD + m$, where Y N(0, In) (4.55) (In being the identity matrix of order n). Example 4.7.2 Let $X = (X .....X,)$ be a random vector having a multinormal dis- tribution N(0, In). Then, the mathematical expectation of the square of the distance of the vector X from the origin is given by EIX + X +... + X,]] = F.IXI $\diamond$ I = n because all the random variables Xk have a standard Gaussian distribution. In addi- tion, they are (globally) independent, since (Tij -= 0 $\diamond$i j. Then, given that Xk 2 has a chi-square distribution with 1 degree of freedom (see Example 3.6.3, p. 94), we can show (see Section 4.9, p. 187) that the squared distance := + +... + also has a chi-square distribution, with n degrees of freedom. That is, D 2 (G o = 3' " = . It follows that if n is even, then P[D > n] = P[D 2 > n 2]= PlY < n/2], where Y Poi($\diamond$n z) (see Section 3.5, p. 87). For example, ifn = 4, we have: P[D > 4] = P[Poi(8) < 2] = 9e -8 -- 0.0030. 4.8 Estimation of a Random Variable We will consider in Chapter 6 the problem of estimating the unknown parameters of a probability density (or mass) function. For example, we will see how to estimate the parameter k of a random variable X having an Exp(k) distribution, based on observations of this variable. Here, we are interested in a different type of estimation problem: let Y be a random variable. Suppose that we seek to estimate Y by a function g(X) of the

4.8 Estimation of a Random Variable

183 random variable X. We will consider three cases for the type of function g(X) that is admissible. 1) Case When g(X) Is Constant If we suppose first that g(X) is a constant, we easily find that the constant a that minimizes the mean square error e(a) := E[(Y- a) e] (4.56) is fi = E[Y], so that the minimum error emin is equal to VAR[Y]. Indeed, we have: E[(Y--a) 2] = E[Y 2]-2aE[Y]+a 2. Differentiating with respect to a and setting the derivative equal to 0, we obtain that a = ElY] does minimize the error e(a): d ae(a) = -2E[Y] + 2a = 0 $\diamond$, a = E[Y] and d e dae(a) = 2 > O. Thus, emin = E[(Y - E[r']) e] -- VAR[Y]. 2) Linear Case We now suppose that we look for a function g of the form g(X) = aX b to estimate Y. Proposition 4.8.1. The constants a and b that minimize the mean square error e(a, b) := E[(Y - (aX b)) 2] (4.57) are given by STD[Y] fi=Px,'-- and /=E[Y]-fiE[X]. STD[X] Moreover, we find that the minimum error emi n is emin = VAR[Y](1 - pzx,y). (Idea of the) Proof. Let Z := Y - aX. Then, e(a, b) := E[(Y-(aX + b)) 2] = E[(Z- b)2]. Making use of the result in Case 1, we can write that

) = E[Z] = E[Y] - aE[X]. (4.58) (4.59)

184 4 Random Vectors We then find fi by solving the equation d --EI(Y - E[Y] - a(X - E[X])) 2] ---- 0. da Finally, replacing a and b by fi and/ in e(a, b), and computing the mathematical expectation, we obtain Equation (4.59). [] 3) General Case Finally, in the general case when g(X) is any random variable, we can show the following proposition. Proposition 4.8.2. The (general)function g(X) that minimizes the mean square er- ror e(g(X)) := E[(Y - g(X)) 2] (4.60) is given by g(X) = ElY I X]. Particular cases, i) If Y = h(X), then we have: g(X) = ElY I X] = EIh(X) I X] = h(X), and it follows that the minimum error emin is equal to 0. ii) If X and Y are independent random variables, then EIY I X] = EIY]. In this case, we have: emin = VARIY1. Remarks. i) The function fiX / is the best linear estimator of Y in terms of X, while g(X) = EIY I X] is the best non-linear estimator of Y in terms of X. Note however that the best "non-linear" estimator of Y in terms of X may be linear, as in the next remark. Therefore, EIY I X] is in fact (simply) the best estimator of Y in terms of X. ii) If X and Y have Gaussian distributions, then the two estimators of Y in terms of X coincide. Indeed, we then have (see Section 4.7, p. 179): E[Y I X = x] = lZy + Px.Y (x - #x) := a*x + b*, which is linear in x. Furthermore, VARIY I X = x] = cry2(l - p2x.y), which corresponds to emin in Equation (4.59). iii) We find, differentiating the mean square error e(a, b) defined in (4.57) with re- spect to a and setting Oe/Oa equal to 0, that E[(Y - fX - [)X] = O. This equation is known as the orthogonality condition. This condition can be stated as follows: the best estimator of the form aX b of Y is such that the estimation error Y - (aX b) and the observation X are orthogonal.

4.9 Linear Combinations 185 Example 4.8.1 If X N(0, 1) and Y := X 2 , then the best non-linear estimator of Y in terms of X is directly g(X) = X 2, while the best linear estimator of Y in terms of X is given by hX with (see Section 4.6, p. 178) and / = E[Y] = E[X 2] = 1. Thus, the best linear estimator of Y in terms of X is simply the constant 1. 4.9 Linear Combinations An important particular case of transformation of random vectors is the one when Z := g(X1 ..... Xn) is a linear combination of the random variables X1 .... , Xn. That is, we can write that Z := ao q-alX1 q- .." q-anXn, (4.61) where the a's are real constants Yk. Then, making use of the (generalized) definition of E[g(X)] (see Section 4.6, p. 176), we easily show that n E[Z] = ao + -a,E[X,]. k=l (4.62) Next, let Z := alX1 + a2X2. We have: VAR[Z] = E[Z 2] - (E[Z]) 2 = E[a21X21

+ 2ala2X1X2 +aX] - (alE[X1] + a2E[X2]) 2 2 2 2 : -al�{E[XI�] - (E[Xk]) 2} -k 2ala2{E[X1X2] - E[X1]E[X2]}. k=l That is, VAR[Z] : al2VAR[XI] + a22VAR[X2] + 2alazCOV[X1, X2]. Using mathematical induction, we obtain the following proposition. (4.63) Proposition 4.9.1. Let X1 ..... Xn be random variables, and let Z be a linear com- bination of the XI 's: Z = ao - Zk:ln alXl. Then, we can write that n n _ VAR[Z] = - a2VAR[X,] + 2- aialCOV[Xi, Xl]. k=l i=1 k=I i<k (4.64)

186 4 Random Vectors Remarks. i) The constant a0 does not affect the variance of Z. ii) We can also write Equation (4.64) in the following form: VAR[Z] = Z Z aiakCOV[Xi, Xk]. i=1 k=l Particular cases, i) If the random variables are independent, Equation (4.64) be- comes VAR[Zl -- a2VAR[Xkl. (4.65) k=l ii) Suppose that the random variables Xk are independent and identically distributed (that is, they have the same distribution function). We write that the Xk's are i.i.d. variables. Then, if Sn := Xl -t- """ -t- Xn, (4.66) we can write that E[Sn] = n E[X l] and VAR[Sn] = nVAR[X l]. Remark. In summary, if the mathematical expectations of X and Y exist, we can write that E[X + gl = E[XI + EIYI. If, in addition, the random variables X and Y are independent, then E[XYI = E[XIEIYI. Likewise, VAR[X + Y] ind. VARIXI + VAR[Y] and VAR[XY] = EIX2y 21 - (E[XY]) 2 ind. EIX2]E[y2I _ (E[X]E[y])2. Note that STD[X + Y] � STD[X] + STD[Y] (in general), even if X and Y are independent, because the square root of a sum is not equal to the sum of the square roots. Reproductive Properties To obtain the probability density function of the sum of two independent continuous random variables, X and Y, we can consider the characteristic function of their sum Z: qz(og) := E[e jwz] = E[e jw(x+Y)] in2' E[eJWX]E[e jwY] = x(o))y(o)). (4.67)

4.9 Linear Combinations 187 Hence, we deduce the result that we have already mentioned in Section 4.5 (p. 176): the probability density function of Z is given by the convolution product of the density functions of X and Y. That is, fz(z) = F-l{q3x(co)q3r(co)} = fx(x) * fr(Y), (4.68) where F -1 denotes the inverse Fourier transform. The result (4.67) can be directly generalized to the case when we add n inde- pendent random variables. Making use of this generalization, we easily show the following reproductive properties. a) If Z is a linear combination of independent Gaussian distributions, then Z has a Gaussian distribution whose mean and variance are given by (4.62) and (4.65), re- spectively (actually, if the random variables are not independent, Z still has

a Gaus- sian distribution, whose variance is given by Equation (4.64)). b) The sum of n independent exponential distributions with the same parameter), (thus, the sum of n independent gamma distributions with parameters 1 and),) has a gamma distribution with parameters o = n and),. Indeed, the characteristic function of X Exp00 is given by 4)x (o) = ; -jco Then, for instance, if X I Exp(X) and X2 Exp(X) are independent, we can write that)Xl _X2 (0)) ind.)Xl (O)))X2 (0)) = , Now, if Y G(o = 2,),), we have:),2 4r (o) - (,k - jog) 2" Thus, by uniqueness, we may conclude that X1 + X2 G(a ---- 2, ;). More generally, the sum of n independent gamma distributions with parameters ak, for k = 1, 2,... , n, and ; has a gamma distribution with parameters a := = ak and ;. In particular, since a gamma G(, �) distribution is a chi- square distribution with m degrees of freedom (see Section 3.5, p. 86), the sum of n independent chi-square distributions with m degrees of freedom, for k = 1 .... , n, also has a chi-square distribution, with ml + � � � + mn degrees of freedom. c) The sum of independent Poisson distributions with parameters a, for k = 1 ..... n, also has a Poisson distribution, whose parameter is given by a :---- al + d) If X1 B(nl, p) and X2 B(n2, p) are independent, then X1 + X2 B(nl + n2, p), which actually follows directly from the definition of a binomial distribution. Remark. Since the generating function Gx and the moment generating function Mx can be expressed in terms of the characteristic function bx, Equation (4.67) above, and its generalization, are valid for Gx and Mx as well:

188 4 Random Vectors GXI+X2(Z) ind.= GXl (z)Gxz(Z) and Mxl+x2(s) ind.= MXl (s)Mx2(s). Example 4.9.1 If X N(0, 1) and Y N(- 1, 2) are independent random variables, then the random variable Z := X - 2Y has a Gaussian distribution with parameters Ixz = 0- 2 x (-1) = 2 and crz2 = I + 4 x (2) = 9. We must not forget to square the coefficients of the random variables when we calculate the variance. Note that otherwise here we would obtain a negative variance! 4.10 The Laws of Large Numbers We can show the two limit theorems that follow. Theorem 4.10.1 (Weak law of large numbers). Let X, X2 .... be an infinite se- quence of i.i.d, random variables such that E[XI] = Ix E . Let Sn be the sum of the first n variables in the sequence: Sn := -=1 Xk. Then,for any constant c > O, we have: --- <c = 1. (4.69) Remark. In statistics, we say that Ix is the mean of the population, while Sn/n is the mean of a random sample (of size n) of the population. We deduce from this theorem that the sample mean converges toward the population mean. In practice, if the

mean Ix is unknown, we can therefore estimate it by using the mean of a sample of the population. The larger the size of the sample is, the more accurate the approximation of IX by the numerical value taken by Sn/n should be. Theorem 4.10.2 (Strong law of large numbers). Let X l, X2 .... be an infinite se- quence of i.i.d, random variables such that E[X] = Ix and VAR[X] < and let Sn := n=l Xk. Then, P[lim Sn] --=Ix = 1. (4.70) Remarks. i) We must talk about measure theory to explain the difference between the two laws. The weak law of large numbers was first proved by Khintchin 2 in 1929. 2 Aleksandr Yakovlevich Khintchin, 1894-1959, was born and died in Russia. His father was an engineer. His first mathematical works were on number theory and probability. He was a professor at Moscow University. This is where he contributed in a very important way to the development of the theory of stochastic processes. Next, he became interested in statistical mechanics and in information theory. In addition to his mathematical research, he had a passion for theater and poetry.

4.11 The Central Limit Theorem 189 ii) We say that Sn/n converges in probability (respectively almost everywhere) to/z in the case of the weak (resp. strong) law of large numbers. iii) If we apply the strong law of large numbers to relative frequencies (see Chapter 1, p. 4), we find that the relative frequency of an event A converges to the probability of A with probability 1. This follows from the fact that the relative frequency of an event A is actually the mean of a sequence of independent Bernoulli random variables with parameter p, where p is at once the probability of A and the mean of the Bernoulli distributions. Example 4,10,1 Let X, X2 .... be independent random variables, all having an exponential distribution with parameter . = 1. We define the indicator variable l1 if X > 1, I = 0 otherwise for all k. Since I has a Bernoulli distribution with parameter p = P[X > 1] = e- 5, we deduce from the strong law of large numbers that n Ik lim '--=E[l]=p=e -s, k=l with probability 1. 4.11 The Central Limit Theorem Theorem 4.11.1 (Central limit theorem). Let X ..... Xn be n i.i.d, random vari- ables with finite mean I and finite variance cr 2 (cr > 0). Let Sn be the sum of the n random variables and Sn - nlz Zn .-- -- (4.71) n l /2ty Then, the distribution function of Zn tends toward that of a Gaussian N(O, 1) distri- bution. Remarks. i) We can also write the following result: Sn " N(n/z, na 2) if n is large. (4.72) Similarly, S?l -- N(/z, crYC) if n is large. (4.73) n ii) In general, if n > 30 we can use the Gaussian distribution to approximate the exact distribution of Zn. However, if the

distribution of the X's is symmetric, then fewer variables are needed in the sum Sn to obtain a good approximation (and vice versa).

190 4 Random Vectors iii) To prove the central limit theorem as stated above, we can use characteristic func- tions. We show that the characteristic function of Zn tends toward that of a standard Gaussian distribution, namely e - c�2/2.Then, since there is a bijective relationship between probability density functions and characteristic functions, we obtain the re- sult we seek. The theorem was first proved by J. W. Lindeberg in 1922. It had been demonstrated previously, among others by Lyapunov, 3 but under more stringent re- strictions. The theorem is sometimes called the Lindeberg-Feller 4 central limit the- orem, or the L6vy 5 central limit theorem. We can nowadays prove it in a few steps by making use of advanced methods (see reference [13]). iv) We can generalize the central limit theorem to the following case: let XI ..... Xn be n independent random variables. Then, under some conditions, S_{n}/n has approx- imately (when n is large enough) a Gaussian distribution with parameters #=- EIXk] and o 2 VARIXk]. (4.74) /'/ 2k= I k=l So, the random variables X do not need to be identically distributed. Note that the expressions for the mean and the variance of S_{n}/n follow at once from Equations (4.62) and (4.65) of Section 4.9. What is difficult to prove is the fact that the distri- bution of S_{n}/n is approximately Gaussian. Example 4.11.1 Let X ..... X,, be independent random variables distributed as the discrete random variable X, whose probability mass function is given by the 3 Aleksandr Mikhailovich Lyapunov, 1857-1918, was born and died in Russia. He was a mathematician and mechanical engineer. He was a school friend of Markov and a student of Chebyshev. He worked, in particular, in the field of differential equations, his methods enabling us to determine the stability of systems of ordinary differential equations, and on probability theory. Moreover, he invented important approximation methods. His doctoral thesis on the stability of motion, which was of great scientific value, was published in English in 1966. He met a violent death (he shot himself when his wile died) soon after the October Revolution of 1917. 4 William Feller, 1906-1970, was born in Croatia and died in the United States. After having been educated by tutors, he obtained university degrees in Croatia and in Germany. He taught at many renowned universities in the United States. Although he made important contributions to the field of diffusion processes in general, and to the theory of Brownian motion in particular, he is most famous for his book An Introduction

to Probability Theory and its Applications, published in two volumes, which is a real classic. The first volume is cited by almost all authors of books on probability theory. The second volume is of a higher mathematical level. 5 Paul Pierre Ldvy, 1886-1971, was born and died in France. Many members of his family were mathematicians. He studied at the lcole Polytechnique de Paris, where he gradu- ated first in his class. Next, he obtained his doctorate from the lcole des Mines de Paris. His thesis was on functional analysis. He taught at the lcole des Mines and at the lcole Polytechnique. In addition to functional analysis, he was interested in probability theory, as well as in partial differential equations and in geometry. He is considered as one of the founders of the rigorous theory of probability, because he published many important papers and books on this subject.

px (x) 4.11 The Central Limit Theorem 191)x '1 2 Figure 4.10. Probability mass function of the random variable in Example 4.11.1. following table: x -1 0 2 E Px(X) 1/2 1/8 3/8 1 (see Fig. 4.10). The exact distribution of Sn can be obtained by performing what is sometimes called the "convolution summation" of Xl, � � � , Xn. For instance, the distribution of $2 is the following: x -2 -1 0 1 2 4 E Ps2(x) 1/4 1/8 1/64 3/8 3/32 9/64 1 Indeed, there is a single pair in Sx �x2 that corresponds to -2, namely (-1, -1), and we have: ind. P[XI -= -1, X2 = -1] -=- P[XI = -1]P[X2 = -1] = (1/2) 2 = 1/4, etc. Figures 4.11 and 4.12 (p. 192) present the probability mass function of Sn for n = 16 and n = 32. We see that when n is equal to 32, the curve looks a lot like that of a Gaussian distribution. 0.07 : 0.06 0.05 0.04 0.03 0.02 0.01 : 0 ! -10 0 10 20 Figure 4.11. Probability mass function of the random variable obtained by adding 16 inde- pendent copies of the random variable in Example 4.11.1.

192 4 Random Vectors 0.05 0.04 0.03 0.02 0.01 0 -20 -10 0 10 20 30 40 Figure 4.12. Probability mass function of the random variable obtained by adding 32 inde- pendent copies of the random variable in Example 4. I I.I. Example 4.11.2 A computer, in adding numbers, rounds each number to the nearest integer. Suppose that the rounding errors are independent and have a uniform distri- bution on the interval (-I, I)" If 1500 numbers are added, what is the probability that the total error, in absolute value, exceeds 15? Solution. Let ET be the total error caused by rounding the 1500 numbers. We can write that ET = E + E2 +. � � + E5oo, where E is the error committed in rounding the kth number Yk. Since E, U(-�, I) Yk and the E,'s are independent, we

have: Er N(1500(0), 1500(1/12)) (appr.). Indeed, we have: EIEI = 0 and VARIEI = - (-I) 12 Yk. We seek Tab. 3.3 PII Er I> 151-2[1-q(1.34)1 -- 2(1 -0.91)=0.18. Remark. If we had made the error of believing that Er U(-750,750), we would have obtained: 3O Pll Er I> 15] = 1 - P[-15 _< Er <_ 15] = 1 -- --0.98, 1500 which is very far from 0.18, a vialue which here can be considered as the exact (but rounded) answer. Actually, the sum of two independent uniform random variables does not have a uniform distribution. Furthermore, the sum of 10 independent uni- form distributions already has approximately a Gaussian distribution. Thus, the sum of 1500 independent uniform distributions is practically a Gaussian random variable. Application: Approximation of a Binomial Distribution by a Gaussian Distribution Let X B(n, p). Since we can represent X as the sum of n i.i.d. Bernoulli random variables, we can use the central limit theorem to approximate the distribution of X.

4.11 The Central Limit Theorem 193 Indeed, we can write that X = Yk=ln Xk, where Xk is equal to 1 if the kth trial is a success and to 0 otherwise. That is, the binomial distribution counts the number of 1 's obtained in n Bernoulli trials. De Moivre-Laplace Approximation. 6 Let X B(n, p). If n is large enough and p sufficiently close to 1/2, then we can write that px(k) - fz(k), (4.75) where Z N(np, npq), since E[X] =np and VAR[X] = npq. Note that px(k) is a probability, whereas fz (k) is not. Validity condition: the approximation should be good if min{np, nq } is greater than or equal to 5. Thus, if p is equal to 1/2, then n = 10 is sufficient to obtain a good approximation, whereas if p = 1/100, then n must be at least equal to 500. Actually, if p is very small, or close to 1, we should use the Poisson approximation instead. Remark. We can also use the distribution function of Z to approximate Fx (k) (or px(k)). In fact, we generally get more accuracy when we seek to approximate the distribution function, as we already mentioned in Chapter 3 (when we saw Poisson's approximation). In that case, we suggest making a continuity correction. Continuity corrections. To improve the approximation, we generally use the fol- lowing corrections (where k, a and b 6 {0, 1 ..... n}): a) P[X=kl= P[k - �l < X < k + �l _ P[k - � < Z<k+�l(necessaryifwe want to make use of the distribution function of Z to approximate Px (k)); b) P[a < X < b] = P[a - � < X < b + �] -- P[a - � < Z < b + �]. This correction stems from the fact that we replace a discrete random variable by a continuous random variable (see Fig. 4.13, p. 194). Remarks. i) If we do not have a closed interval [a, b], we

can simply transform the given interval into an interval of this type before making the continuity correction. For instance, we have: + 1 : 1- 6 Abraham de Moivre, 1667-1754, was born in France and died in England. A mathematician of French origin, he emigrated to England in 1685 and made his career there. He was a friend of Newton. His most important works were on analytic geometry and on probability, of which he is one of the pioneers. In 1718, he published his book entitled The Doctrine of Chance, in which the definition of statistical independence can be found. The formula at- tributed to Stirling appeared in a book that he published in 1730. He later used this formula to prove the Gaussian approximation to the binomial distribution. An important formula in complex analysis is also due to him. It is said that, like Cardano, he predicted the day on which he would die.

194 4 Random Vectors 114 Figure 4.13. Approximation of a binomial distribution by a Gaussian distribution. ii) When we want to get an approximation for a probability like $P[X \leq b]$ or $P[X \geq a]$, adding the minimum value that X can take, namely 0, or the maximum value, namely n, generally does not change much the numerical value obtained (if the validity condition is respected). iii) We could also approximate the distribution of a Poisson random variable, with parameter oe large enough, by that of a Gaussian random variable with parameters = O "2 = Ot (see Section 4.9, p. 187). Example 4.11.3 If 20% of the diodes manufactured by a certain machine are de- fective, what is the probability that in a batch of 100 (independent) diodes taken at random (and without replacement) among those produced by this machine, there are exactly 15 defectives? Solution. Let X be the number of defective diodes, among the 100 examined. Then, X has a binomial distribution with parameters n = 100 and p = 0.2. We seek $P[X = 15] = P[14.5 < X < 15.5]$ -- $P[14.5 < Z < 15.5]$, where Z N(20, 16) Tab. 3.3 = q(1.375) - q(1.125) $\simeq$ 0.9155 - 0.8697 = 0.0458. Remarks. i) We obtain practically the same answer by making use of Equation (4.75): l exp{ 1 (15- 20)2] fz(15) -- . 2]-($\simeq$ 0.0457. In fact, it is clearly more efficient to use Equation (4.75) when we seek to approxi- mate the value of the probability mass function of X. ii) The answer obtained by using the binomial distribution is (about) 0.0481. Moreover, if we make use of the Poisson approximation instead, we find that $P[X = 15] \simeq$ 0.0516. The probability of a "success," p = 0.2, is too large here to obtain a good approximation with a Poisson distribution.

4.12 Exercises, Problems, and Multiple Choice Questions 195 iii) In the preceding example, the validity condition is respected, since we have: min{np, nq} = min{20, 80} = 20 > 5. Therefore, we were expecting to get a good approximation with a Gaussian distribution. Note, however, that the percentage of er- ror obtained by using the Gaussian approximation is about 5%, while it is about 7.3% with the Poisson approximation. Thus, here the quality of the two approximations is in fact comparable. 4.12 Exercises, Problems, and Multiple Choice Questions Solved Exercises Exercise no. 1 (4.4) Three companies, X, Y and Z, have a probability equal to 0.4, 0.3 and 0.3, re-spectively, of securing an order for a certain product. If three independent orders are placed, what is the probability that each company receives exactly one order? Solution Let W := (X, Y, Z), where X is the number of orders received by the company X, etc. Then, W has a multinomial (trinomial) distribution with parameters Px = 0.4, py = 0.3 and Pz = 0.3. We seek P[X= I,Y= I,Z= I]-- -- 3! 1!1!1! (0.4) 1 (0.3) 1 (0.3) 1 : 6 (0.036) = 0.216. Exercise no. 2 (4.6) Let X be a random variable such that E [X n] = ● for n = 1, 2 ..... We define Y = X 2. Calculate Px,r. Solution We calculate PX,Y : E [XY] - E [X] E [Y] STD [X] STD [Y] STD [X] STD [X 2] because and 1 1 STD[X 2] : E[X 4]- E[X 2] : _ : --

196 4 Random Vectors Exercise no. 3 (4.6) We define Z = X + Y, where X and Y are two independent random variables uniformly distributed on the interval [a, b]. a) Find the probability density function of Z, if a = 0 and b = 1. b) Calculate the correlation coefficient of X and Z, if a = - 1 and b = 1. Hint. We have: VAR [X] = 1/3 ifa = -1 and b = 1. Solution a) We can write, by independence, that fz (z) = fx (u) fr (z - u) du = fr (z - u) du. Now, we have: Sz = [0, 2] and It follows that 1 if0<z-u < 1 ●z- 1 <u <z, fr (z - u) --- 0 elsewhere. fz (z) = : 1 du = z if0<z<l, fz ldu=2-zifl <z<2. -I Equivalently, if Z = X + Y and W = X, then we find that fz.w(z,w)=l if0<w< 1, 0<z<2, w<z<w+l. Next, we integrate fz,w (z, w) with respect to w to obtain fz (z). b) We have: PX.Z : because E [XZ] - E [X] E [Z] (}) - (o)(o) a -- -- -- _ 0.7071, 2 and E [XZ] = E[X 2] + E [XYI ind. VAR [X] + (E [X]) 2 .q.- E [X] E [Y] 1 = VAR[X] + (0) 2 -}- 0 ● 0 = VAR [X] = - 3 1 I i'i'd" 1 () STD[Z]=(VAR[Z]): = (2VAR[X]): = .

4.12 Exercises, Problems, and Multiple Choice Questions 197 Exercise no. 4 (4.6) Let fx'Y(x'Y): {2-Xo-YifO<x < l'O< y < Calculate a) Fx,r (x, y); b) fx

(x); c) COV[X, Y]; d) P IX + Y < 1]. Solution a) We calculate f0f0 y Fx,r (x, y) = (2 - u - v) dvdu x2y xy 2 =2xy if0<x < 1,0<y < 1. 2 2 Hence, we deduce that Fx,r (x, y) = 0 ifx <0ory <0, 2xy -- ◆x2y -- ◆xY 2 if0 < x < 1,0 < y < 1, _y_ ◆22 ifx > 1,0 < y < l, -x--◆x 2 if0<x < 1,y> 1, 1 ifx > landy> 1. b) We have: 1 3 fx (x) : (2-x-y)dy= - x if0<x<l. c) First, we calculate E[XY] = xy(2- x - y) dxdy= y x x My = --. 12 Since, by symmetry, E[Y] = E IX], we can write that cov[x,r]=E[xr]-E[x][r]=- - 3 1 -0.0069. 144 d) We can write (see Fig. 4.14, p. 198) that P[X+Y < 1]= (2-x-y) dydx

198 4 Random Vectors 1/2 + y =1 Figure 4.14. Figure for part d) of Exercise no. 4. 2(1 -x)--x (1 --x) - 2x + dx = -. 3 (1 - x)2 / " dx 2 Exercise no. 5 (4.9) Let X Poi(1) and Y Poi(2) be two independent random variables. Calculate Solution We have: P 3<(X+Y)<5 =P 4<X+Y<-- = P[4<Poi(3)<6] { 343536 } = e -3 . + . + . _ 0.3193. Exercise no. 6 (4.11) Let X ..... Xso be independent random variables having a Gaussian N(0, 1) K 50 X 2 distribution. We define Y = ,,,= . a) Calculate the mean of Y. b) Use the central limit theorem to calculate P [Y < 60]. Hint. We have: E[Y 2] = 2600. Solution a) We have: E[X] = VAR [X] + (E [Xk]) 2 = 1 q- 0 2 = 1

4.12 Exercises, Problems, and Multiple Choice Questions 199 50 50 E I'1 : Z Elxl = Z l= 50. k=l k=l b) We can write that V N(50, r 2) (approximately), where r 2 = E[Y 2] - (E [y])2 = 2600 - (50) 2 = 100. Then, 60 501 Tab. 3.3 P[Y <60]--P N(0,1) < 10] =qb(1) -- 0.8413. Exercise no. 7 (4.11) Let (X1 ..... Xn) be a random vector having a multinormal distribution. We sup- pose that E [Xg] = 0 and VAR [Xg] = 1 for k = 1 ..... n, and that COV[Xj, Xg] = 0 if j - k. We define Y = X1 + X2 and Z = X3 - X4. a) What is the characteristic function of Y? Hint. If X N(#, a2), then qx (co)= exp (j#w - ◆o-2co2). b) Does the random vector (Y, Z) have a bivariate normal distribution? If so, give its five parameters; if not, justify. c) Suppose that n = 50. We take an observation of each random variable X ..... Xs0. Use the central limit theorem to calculate (approximately) the probability that exactly 25 observations are positive. Hint. We have: qb (0.14) -- 0.5557. Solution a) COV[X1, X2] = 0 B e X1 and X2 are independent. Then, we can write that Y N(0 + 0, 1 + 1), so that by (co) : exp (j (O)m - (1/2)2o) 2) : \ b) Since Y and Z are linear combinations of independent N(0, 1) random variables, the random vector (Y, Z) has a bivariate normal distribution. Its parameters are: #Y = #z = 0, rr 2 = rz 2 = 2 and PY, z = 0 (because Y and Z are independent random variables). c) Let N

be the number of positive observations. We can write that N B(n = 50, p = 1/2). We want P IN = 25] C_LT fz (25), where Z has a Gaussian N(25, 12.5) distribution. Therefore, 1 { (25 - 25) 2 } 1 P IN = 25] -- -- 4'--x/]- exp -(i-2. 0.1128. Remark. Or: P [N = 25] -- 2qb (0.14) -- 1 -- 0.1114.

200 4 Random Vectors Exercise no. 8 (4.1 I) The average lifetime of certain electronic components is equal to six months and the standard deviation of the lifetime is equal to two months. a) We consider two components of this type. Let T1 and T2 be their respective lifetimes. We suppose that T1 and T2 have (approximately) a Gaussian distribu- tion and that their correlation coefficient is equal to 1/2. Calculate the probability P [T2 > 8.5 I T1 = 4]. b) Suppose now that the components are independent, but that the distribution of their lifetime is unknown (we only know the mean and the standard deviation of the lifetime). We consider a machine made up of n components of this type placed in parallel. Furthermore, only one component is active at a time (standby redundancy). Use the central limit theorem to find the smallest value of n for which the probability that the machine functions during at least 15 years is greater than 90%. Hint. We have: Q- (0.1) 1.2815 and Q-1 (0.9) - 1.2815. Solution a) We can write that T2I{T 1 =4} N IZZ+pT2.T, -1 (4--//.1) 022 1 _p2 =N(5,3) , Tz.TI " Tab. 3.3 Then, P[T2 > 8.5]T1 =4] = P[N(5,3) > 8.5] Q(2.02) 1 -0.9783 0.022. b) Let S,, := T1 � � � T,,. If n is large enough, then by the central limit theorem we can write that S,, N (n � 6, n � 4). We seek nmin such that P [Sn > 180] > 0.90 � P [N(6n,4n) > 180] > 0.90 (18_0 - _6n 180- 6n = Q 2,v/. ,] > 0.90 = 2-- < - 1.2815. We consider the equation 6n - 2.563,,/ - 180 = 0. We find that -- -5.27 or 5.69. It follows that n -- 32.4. Thus, we must take n = 33. Exercise no. 9 (4.11) Suppose that X1 ..... Xn are independent random variables that are all uni- formly distributed on the interval [0, I]. a) Calculate P[] X1 - X2 l< 1/2]. b) i) Find the characteristic function of Y := al X1 + azX2, where al and a2 are real constants (5 0). ii) Making use of part i), can we assert that Y has a uniform distribution? If so, on what interval. If not, justify your answer. eJb_eJ)a Hint. We have: 4)x(w) -- jo4b-) ' if X U[a, b]. c) Use the central limit theorem to find the smallest value of n for which P[XI+'"+Xn <]< I0 -5. Hint. We have: Q-1(I0-5) - 4.2649.

4.12 Exercises, Problems, and Multiple Choice Questions x2 , , / x2-x 1=1/2 Figure 4.15. Figure for part a) of Exercise no. 9. 201 Solution a) We have: fx l'X2 (X1, X2) ind. 1 X 1 = 1 for 0 _< x 1 _< 1,0 < x2 < 1. Then, we can write

(see Fig. 4.15) that $= X1$ (al) $X 2$ (a2) $=$ jwal ja2 (e ja' - 1)(e ja: - 1) --ala2 2 ii) Since � (w) Cx (w) for any unifo random variable X, the random variable Y does not have a unifo distribution. c) XI+'"+Xn N nx,nx P X+'"+Xn< P N , < =P N(O, 1)< g It follows that n 24.25, whence we deduce that nmin = 25.

202 4 Random Vectors Unsolved Problems Problem no. 1 Find fx (x) if fx, Y (x, y) Problem no. 2 Let fx,r (x, Y) :- 4 Calculate P[2X 2 > Y]. Problem no. 3 Suppose that and 1 PxIv (x [y) = . ifx 2 + y2 _< 2 (= 0 elsewhere). if0 < x < 2, 0 < y < 2 (=0 elsewhere). ifx :-0, 1,2andy : 1,2, 1 Pv (Y) = if y = 1 or 2. Find Pz (z), where Z := XY, assuming that X and Y are independent random variables, Problem no. 6 Calculate VAR [XY] if X -- N(-1, 1) and Y N(2, 4) are two independent random variables. Problem no. 7 Let X1 N(0, 0 -2) and X 2 N(0, 0 "2) be two independent random variables, and let Y := X1 + X2. X Px (x) and y -1 0 1 1 1 1 pv (y) Find Px,v (x, y). Problem no. 4 Let X1 N(/, a 2)_ and X2 " N(/, a 2) be two independent random variables. We define Y : X + X2 and Y2 ----- X + 2X2. Find the joint probability density function f,2 (Y, y2) � Problem no. 5 Let

4.12 Exercises, Problems, and Multiple Choice Questions 203 a) Find the best estimator of Y, based on Xl. b) Find the best estimator of Xl, based on Y. Problem no. 8 Calculate P [2X - Y < 8], where X N(0, 1) and Y N(1, 5) are two inde- pendent random variables. Problem no. 9 Let X, Y and Z be independent random variables such that VAR [X] = 1, VAR [Y] = 4 and VAR [Z] = 11. We define W = 3X +2Y - Z. Calculate STD[W]. Problem no. 10 Suppose that X is a random variable such that E [X] = VAR [X] = 1. Use the p r-,36 central limit theorem to calculate (approximately) the probability t/k= Xk < 42], where X1 ..... X36 are independent random variables having the same distribution function as X. Problem no. 11 Let X B(n = 25, p = 1/2). Use the Gaussian approximation to the binomial distribution to calculate P [X < 12]. Problem no. 12 Let Jxifl_<x <y<2, fx,y(x,y) ! 0 elsewhere. a) Find the marginal density functions fx and ft. b) Calculate the probability P [X 2 < Y]. Problem no. 13 Let X ..... Xs0 be random variables having a geometric distribution with pa- rameter p = 1/4, and let S := X +. �. + Xs0. a) Calculate VAR [S] if Pxj,xk = 1/2 Yj 7 k. b) Suppose now that the Xk's are globally independent. Use the central limit theorem to calculate P [S < 201]. Remark. Do not use a continuity correction. Problem no. 14 Let l fx,r(x,Y)= (4xy)-7 ifO<y <x < 1, 0 elsewhere. Calculate a) fx (x) and fr (Y); b) E [Y I X]; c) P[X > �, Y > �1.

Problem no. 15 The input to a communication channel is a random variable X having a standard Gaussian distribution. The output Y is given by $Y = X + N$, where N, the noise, has

a Gaussian distribution with zero mean and variance a_N^2. Furthermore, X and N are independent. a) Calculate $P[Y = 0 \mid X = 0]$. b) Calculate the correlation coefficient, $P_{x,}'$, of X and Y. c) If $a_N^2 = 8$, we find that $P_{x, Y} = 1/3$. Moreover, the vector (X, Y) has a bivariate normal distribution. What is the best estimator of Y in terms of X? Problem no. 16 The joint probability mass function of the discrete random vector (X, Y) is given by the following table: y\x -1 0 1 0 1/9 2/9 1/9 1 1/9 2/9 1/9 2 0 1/9 0 a) Find $Pr_{1x}(Y \mid 0)$. b) Calculate $F_{x. r}(2, 0)$. c) i) Calculate the correlation coefficient of X and Y. ii) Are X and Y independent random variables? Justify your answer. d) Let $Z := X + Y$. Find $P_Z(z)$. Problem no. 17 A computer generates random numbers X according to a Gaussian distribution with parameters 0 and a^2. Let { {XifX <0, XifX>0, and Z:= 0 otherwise. Y := 0 otherwise a) Does the sum Y Z have a Gaussian distribution? Justify your answer. b) Does the pair (Y, Z) have a bivariate normal distribution? Justify. c) Calculate $E[Y \mid X \geq 1]$. Problem no. 18 In a quality control operation, we examine the paint of n brand-new cars taken at random among those produced by a certain company. Let $X_k ---- 1$ if the paint of the kth car examined has at least one flaw, and $X_k = 0$ otherwise, for k = 1,2 ..... n. We assume that the random variables X_k are independent and that the probability that the paint of a brand-new car is flawless is equal to 0.75. Therefore, X_k has a Bernoulli distribution with parameter p = 0.25 for all k. a) Calculate $\lim_n P[1-=1X/ - 1 >_]$. b) Let $Y := X_1 + X_2$ and $Z := X_1 - X_2$. Calculate COV[Y, Z]. n 10], if n 40. C) Use the central limit theorem to calculate $P[-k=l X_k == $ Hint. We have: qb (0.18) -- 0.5714.

Problem no. 19 Let (X, F) be a discrete random vector whose probability mass function is given by the following table: y \ x 0 1 2 -1 1/3 0 1/3 0 0 1/6 0 1 0 1/6 0 a) Are X and Y independent random variables? Justify your answer. b) Calculate $E[X \mid Y = -1]$. c) Let $U := X + Y$ and $V := X - Y$. Find $p_{U, V}(u, v)$. Problem no. 20 Let (X, Y) be a random vector having a bivariate normal distribution with parameters $\#x = 0$, $\#r = 1$, $a_x^2 = 1$, $a_{rz}^2 = 5$ and $P_{x,r}$. 1 a) Calculate VAR [2X - Y], if P_x, r = 2" b) Calculate $P[2X - Y > 2]$, if $p_{x,r} = O$. Hint. We have: Q (1) -

1.59 x 10 -1 and Q (2) - 2.28 x 10 -2. c) Find the best estimator of X in terms of Y, if Px,r = - 1. Problem no. 21 Let e y if0 <x < y < fx, r (x, y) = 0 elsewhere. Calculate a) the marginal probability density functions of X and Y; b) the covariance of X and Y; c) the best estimator of X in terms of Y; d) P IX = Y]. Hint. We have: f ya-l e-Ydy Problem no. 22 The joint probability mass function of the discrete random vector (X, Y) is given by the following table: y\x -1 0 1 0 1/9 1/9 1/9 2 1/9 1/9 1/6 4 1/9 1/18 1/9 a) Calculate i) Px (x); ii) P [X + Y < 2]; iii) Pxlr (x] 2); iv) Pz (z), where Z := XY; v) COV[X, Y]. b) Are X and Y independent random variables? Justify your answer.

206 4 Random Vectors Problem no. 23 A source transmits a signal X uniformly distributed on the interval (0, 2zr). Be- cause of the presence of additive Gaussian noise, the signal received, Y, is a random variable whose conditional probability density function is given by 1 [l(y-x)2} forxE(O, 2zQ yE]. fvlx (Y Ix) -- 4- exp 2 r ' a) Calculate i) fx,v (x, y); ii) the best non-linear estimator of Y in terms of X; iii) E [Y]; iv) Fx,v (zr, x). b) We consider 48 signals Xk, where k = 1,2 ..... 48, transmitted by the source. What is (approximately)the probability that the average signal M48 := & y-28__ Xk 3 9 ii) equalto is i) smaller than Hint. We have: Q (1) - 0.159 and Q (2) - 0.0228. Problem no. 24 Let X and Y be two independent random variables having, respectively, a bino- mial distribution with parameters n = 2 and p ---- 1/2, and a uniform distribution on the interval [0, 1]. Calculate a) E [X2y2]; b) P[X+ Y < 11; c) the joint distribution function of the pair (X, Y) at the point (2' 2); d) the correlation coefficient of the pair (X, Z), where Z := X + 1. Problem no. 25 Let] l/2if0<x <2, 0<y < 1, fx,v(x, y) 0 elsewhere. a) Find fx(x) and fv(Y). b) Calculate the correlation coefficient of X and Y. c) Calculate Fx, v(-l, 1) + Fx,v(l, 2). d) Let U := X + Y and V := X- Y; i) find ft, v (u, v) and draw the region where ft, v (u, v) > 0; ii) calculate E[UV]. Problem no. 26 a) Let X, Y and Z be random variables such that VAR[X] = l, VAR[Y] = 4, VAR[Z] = 9, COV[X, Y] = 1/2, COV[X, Z] = 0 and COV[Y, Z] = -1/2. Calcu- late VAR[X- ◆Y + ◆Z]. b) We take 40 independent observations of the random variable X. Let N be the num- ber of observations that are greater than 1. Use the central limit theorem to calculate (approximately) the probability P[N < 5], if X N(0, 1). Hint. We have: Q(0.8) - 0.212, Q(1) - 0.159 and Q(1.2) - 0.115.

4.12 Exercises, Problems, and Multiple Choice Questions 207 Problem no. 27 The joint conditional probability mass function PxIY (x] y) of the random

vector (X, Y) is given by the following table: y\x 0 1 2 -1 1/4 1/4 1/2 0 2/3 0 1/3 So, we have: 1/4 = PSlY(O I -1), etc. Furthermore, we suppose that Pr(Y) = 1/2 if y = -1 or0. a) Are the random variables X and Y independent? Justify your answer. Calculate b) the function px, r(x, y); c) the function Px (x); d) COV[X, Y]. Problem no. 28 Let [ke -xz-y if-c < x < c, y > O, fx, Y (x, y) = 0 elsewhere, where k > 0 is a constant. a) Find the constant k. Hint. We have: f e -(x2 +ax) dx = e a2 /4 Q(a / v/). b) Calculate P[X > Y]. Hint. We have: Q(0.7071) 0.24. c) Calculate VAR[2X - Y]. d) Let Z := XY. Find the joint probability density function of the random vector (Z, Y). Problem no. 29 Let 7r 7r fx,l,(x,y)=kcos(x+y) for0 <x < -, 0 < y < -. a) Calculate i) the constant k; ii) the joint distribution function of the random vector (X, Y); iii) the marginal probability density functions of X and Y; iv) the probability P[X < Y]. b) Are the random variables X and Y orthogonal? Justify.

208 4 Random Vectors Multiple Choice Questions Question no. 1 Let 0 ifx <0ory <0, xy/2ifO <x <2and0 < y < 1, Fx.r(x,y) = x/2 if0 < x < 2 and y > 1, y ifx >2and0<y < 1, 1 ifx >2andy > 1 be the joint distribution function of the pair (X, Y) of continuous random variables. A) Calculate Fx(3) + Fr(1/2). a) 3/2 b)2 c)Y+� d)+y e) �+l B) Are the random variables X and Y orthogonal and independent? a) not orthogonal and dependent b) not orthogonal, but independent c) orthogonal and dependent d) orthogonal and independent e) not orthogonal and we cannot tell whether they are independent C) Calculate P[X > Y]. a) 0 b) 1/4 c) 1/2 d) 3/4 e) 1 D) Calculate VAR[X Y]. a) 0 b) 1/36 c) 7/36 d) 4/9 e) 1/2 E) Let V := X 2 and W := 2Y. Calculate fv.w(v, w). a) 1/(84') b)+1/(44') c) 1/(44') d)44' e) 84' Question no. 2 Let X, X2,... be independent and identically distributed random variables such that E[X] = 0 and VAR[X] = 3. A) Suppose that the Xi's have a Gaussian distribution. Let Y := X Xz and Z := X - Xz. Are the random variables Y and Z orthogonal and independent? a) not orthogonal and dependent b) not orthogonal, but independent c) orthogonal and dependent d) orthogonal and independent e) not orthogonal and we cannot tell whether they are independent B) Under the same assumptions as in A), what is the best estimator of Y in terms of Z? a) 0 b) Y c) Z d) - e) we cannot calculate it C) Suppose that the Xi's have a uniform distribution on the interval [-3, 3]. What is the approximate distribution of X 2 + .-- + X52o ? a) U[-150, 150] b) U[0,450] c) N(0, 1) d) N(3,) e) N(150,360)

4.12 Exercises, Problems, and Multiple Choice Questions 209 Question no. 3

Let (X, Y) be a continuous random vector with joint distribution function Fx, y(x,y)=(1-e-X)(1-e -y) ifx >0, y >0. Calculate P[X > 1]. a) e -2 b) e -1 c) (1 - e-l) 2 d) 1 - e -1 e) 1 - e -2 Question no. 4 A discrete random vector (X, Y) has the joint probability mass function given by the following table: y\x 0 1 2 0 1/8 Pl 1/4 1 1/8 1/8 p2 What values can the constants Pl and P2 take if X and Y are independent? a) Pl = 1/4, P2 = 1/8 b) Pl = 3/8, P2 = 0 c) Pl = 1/8, P2 -: 1/4 d) P l = 3/16, P2 ---- 3/16 e) none of these answers Question no. 5 Let X "- N(0, 1). Calculate the correlation coefficient of X and Y := X 4. Hint. We have: <Px (w) = e -◆2/2 . a)-I b) 0 c) 1/2 d) l/x/ e) l Question no. 6 A random vector (X, Y) has a bivariate normal distribution with parameters /x = 0,/r = 0, crx2 = 1, cr 2 = 4 and p = 0. Calculate P[XY < 0]. a) 1/16 b) 1/8 c) 1/4 d) 1/2 e) 3/4 Question no. 7 Let X and Y be two independent random variables, both having a Poisson distri- bution with parameter c = 1.5. Calculate P[X + Y > 1.5]. a) 0.05 b) 0.20 c) 0.58 d) 0.80 e) 0.95 Question no. 8 If 10% of the diodes manufactured by a certain machine are defective, what is the probability that in a lot of 100 diodes taken at random (and without replacement) among those produced by this machine, there are less than 15 defectives? a) Q(1.67) b) Q(1.5) c) Q(0.05) d) 1 - Q(1.67) e) 1 - Q(1.5) Question no. 9 The joint probability mass function of a discrete random vector (X, Y) is given by () e-1 forx :0, 1,2;y= 0, 1,2, px, r(x, y) = 4y! .... Calculate P[X = 2Y]. a) 0 b) ◆e -1 C) ◆e -1 d) e -1 e) 1

210 4 Random Vectors Question no. 10 Let (X, Y) be a continuous random vector for which the conditional probability density function of X, given that Y = y, is given by fxiy(x] y) = 1-e-X/Y forx > 0,0 < y < 1. Y Calculate E[X], if E[Y] = 1/2. a) 1/4 b) l/2 c) 2 d) y e) o Question no. 11 A point (X, Y, Z) is taken at random inside the unit sphere, so that 3 J'x.r.Z(x, y, z) =-- ifx 2+y2+z 2 < 1. 4Jr - Calculate the probability that the distance between the chosen point and the origin is smaller than 1/2. a) 1/64 b) 1/8 c) 1/4 d) Jr/8 e) 1/2 Question no. 12 Let fx.r(x,y)=l if0<x< 1,0<y< 1 be the joint probability density function of the continuous random vector (X, Y). We define U = X 2 and V = y2. Find fu.v(u, v) for0 < u < 1,0 < v < I. a) 1 b) x2y 2 c) ◆(uv) -1 d) ◆(up) 1/2 e)] (uv) 1/2 Question no. 13 Let (X, Y) be a random vector having a bivariate normal distribution with pa- rameters #x = 0, #r = 0, Cx = 1, cr = l and p = 0. Calculate the best estimator of X 2 in terms of Y. a) 0 b) X 2 c) Y d) Y2 e) none of these answers Question no. 14 Let X ..... X0o be a set of independent random variables such that E[Xk] = 0 and VAR[Xk] = l for k = 1, 2 ..... 100.

Use the central limit theorem to calculate the probability P[I __o0 Xk l> 10].
a) Q(1) b) 2Q(I) c) 1 - Q(1) d) 1 - 2Q(1) e) 2(1 - Q(1)) Question no. 15
Suppose that X B(n = 100, p). What is the value of p (smaller than or equal to
1/2), if P[X = 100p] -- 0.13307 a) 1/10 b) 1/5 c) 3/10 d) 2/5 e) 1/2 Question
no. 16 Let X, X2 and X3 be N(0, 1) random variables such that Px,x2 = O,
Px,x3 = 0 and Px2, x3 = -- 1.

4.12 Exercises, Problems, and Multiple Choice Questions 211 a) li = O, 12 =
0 d) li = 1,12 = 1 Question no. 17 A) Let Y := XI + X2. Calculate PY, X3. a)-
1/2 b)-l/,,/ c) 0 d) 1/,,/ e) 1/2 B) Calculate ml := E[X3 I X2] and m2 := E[X I
X2]. a)ml =0, m2 = 1 b) ml = -X2, m2 = 1 c) ml = -X2, m2 = X d) ml = X2,m2
= 1 e) ml = X2,m2 = X C) Let Xll ..... Xn be n independent random variables
distributed like XI. We define 1 ifX1 >0, / = 0 otherwise, for k = 1 ..... n.
According to the strong law of large numbers, what is, with prob- ability 1, the
value of the limits ll := lim _I and 12:= lim ' b) li = 1/2, 12 = 1/4 c) li = 1/2,
12 = 1/2 e) ll = cx), 12 = oo Calculate P[X 2 + y2 > 2 In 4] if X and Y are two
independent random variables having a Gaussian N(0, 1) distribution. a) l/8 b)
1/4 c) 1/2 d) 3/4 e) 7/8 Question no. 18 Find the characteristic function of Z :=
X + 2Y, where X and Y are indepen- dent random variables having an
exponential distribution with parameter a and with parameter , respectively.
Hint. If X Exp()), then cx(w) =)/ () - a) (7:7-) + 2(_-) b) ' 2/3 � (2e--:v) d)
(_-)(_jw) 2 e) 2(_--) (_j) Question no. 19 Let XI, X2, ... be independent
random variables that are all distributed as X U(-�, �), and let Sn := =i X.
Use the central limit theorem to calculate P[S2500 _< 125]. a) Q(1) b) 2Q(1)
c) 1 - 2Q(1) d) 2(1 - Q(1)) e) (1 - Q(1)) 2 Question no. 20 A number X is
taken at random in the interval (0, 1], and then a number Y is taken at random
in the interval (0, X]. We can show that 1 fx,r(x,y)=- if0<y<x <1. X Calculate
P[Y < 1/2]. a) -T--In 2 b) I2 c) In 2 d) l qn_____2 e) none of these answers

212 4 Random Vectors Question no. 21 Let FX, Y(X, Y) = 1 Calculate Fx(l/2)
+ Fx(3/2). a) 1/2 b) 3/2 c) �Y d) l +�y Question no. 22 0 ifx <0ory <0,
xyifO<x < 1,0 < y_< 1, x if0<x<l,y> 1, y ifx> 1,0<y< 1, ifx> l,y> 1. e) -y
Calculate ELY], if Y] X N(X, 1) and X Exp(2). a) 1/2 b) 2 c) X/2 d) X e) 2X
Question no. 23 Suppose that X has a uniform distribution on the interval (-1,
1), We define Y = X 2. Are the random variables X and Y orthogonal and
independent? a) not orthogonal and not independent b) not orthogonal, but
independent c) orthogonal and correlated d) orthogonal and independent e)

orthogonal, uncorrelated but not independent Question no. 24 We consider a random vector (X, Y) having a bivariate normal distribution with parameters #x = 1, #y = 2, x 2 = 1, y2 = 4 and p = 1/2. Calculate E[XY]. a) l b) 2 c) 3 d) 4 e) 5 Question no. 25 Use the central limit theorem to calculate P[(XI + - - - + Xn) 2 > c], where the random variables XI ..... Xn are independent and all have an exponential distribu- tion with parameter 2, and c is a positive constant, a) n-ze7 e) 1 - 2Q(-------) Question no. 26 Let px, y(j,k) = k! for j,k =0, 1,2 ..... Calculate P[X = Y]. a)0 b) �e -I c) �e -1/2 d)e -I e) e -1/2 Question no. 27 Calculate E[X Y] if we suppose that Y has a uniform distribution on the interval (0, 1) and that

a) 2/9 b) 1/3 Question no. 28 Let 4.12 Exercises, Problems, and Multiple Choice Questions 2x fxig(xly)=- if0<x <y. c) 2/3 d) cx e) does not exist 213 1 fx,y(x,y)=- if0<x <2,0<y <2. 4 Find fv,w(v, w), where V := X - 2 and W := -Y. a) � if-2< v <0,-2 <w <0 b) � if0<v<2,-2<w<0 c) �vw if-Z<v<0,-Z<w<0 d)-4 I- if-Z<v<0,0<w<2 e) none of these answers Question no. 29 Suppose that px(x)= 1/2 ifx=lor2. We consider n independent random variables, XI ..... Xn, such that Pxi (x) =- Px (x) for i = 1, 2 ..... n. Calculate the generating function Gsn (z) of the random variable Sn := XI + "" + Xn. a) �(z +z 2) b)[�(z + z2)] I/n c)[�(z + z2)] n d) n(z + z 2) e) (z + z 2) Question no. 30 Suppose that X N(0, 1), Y N(1,4) and Px,Y = -1/2. Calculate the best estimator of X 2 in terms of Y. a)] b) L - c) q- I-- - d) (Y-I)216 e) q- (Y-I)2 16 Question no. 31 Calculate P[X 2 + y2 < 1], if 2ifO <x <y_ < 1, fx,y(x, y) = 0 elsewhere. a):r/16 b):r/8 c):r/4 d) 1/2 Question no. 32 Let Calculate Fx,�(25, 5). a) 0 b) 0.03125 c) 0.05 d) 0.95 Question no. 33 e) 3/4 forx=O, 1 ..... 20;y= 1,2 ..... e) 0.96875 The joint probability mass function pxA,(x, y) of a discrete random vector is given by

214 4 Random Vectors 1/5 if (x, y) = (0, 0), (2,1), px,v(x,y)= 1/lOif(x,y)=(O,- l),(O,l),(1,O),(l,1),(2,-1),(2,0), 0 otherwise. Calculate EIX I Y = 0]. a) 0 b) 3/10 c) 2/5 d) 3/4 e) 4/5 Question no. 34 Suppose that X U(0, 1) and Y U(0, 2) are independent random variables. Find the function .f'v.v(v, y), where V := X + Y. a) � ify<v<y+l,O<y<2 b)_ if0<v<3,0<y<2 c) � ify<v<3,0<y<2 d) � if0<v< 1,0<y<2 e) � ifO<v<y+l,O<y<2 Question no. 35 Let X U(-2, 2). We define Y = X 2 and Z = X 3. Are the random variables Y and Z correlated'? independent? orthogonal? a) orthogonal and independent b) orthogonal and uncorrelated, but not independent c) not orthogonal, but independent d) not

orthogonal, uncorrelated, but not independent e) orthogonal, uncorrelated, but we cannot conclude regarding independence Question no. 36 The random variables X, X2 and X3 have a Gaussian N(0, 2) distribution. We suppose that COV[X, X2] = 0 and COVIX, X3] = -2. Calculate the variance VARIX + X2 + X3]. a) 2 b) 4 c) 6 d) 8 e) 10 Question no. 37 We define S0 = o Xk, where the Xk's are independent random variables, all having a Poisson distribution with parameter ot = 1. i) Calculate PISo > 2]. ii) Use the central limit theorem to calculate P[So > 2]. a) i) 0.9972; ii) q(0.8) b) i) 0.9972; ii) q(2.53) c) i) 0.9995; ii) q(0.7) d) i) 0.9995; ii) qb(0.8) e) i) 0.9995; ii) qb(2.53) Question no. 38 Let 3 fx, v(x, y) = - if0 < y < 1, X 2 < y. 4 A) Find fx(x) and fv(Y). a) fx(x) = �(1 - x 2) if l x l< 1; iv(Y) = y/2 if0 < y < 1 b) fx(x)= -](1 -x 2) if Ix l< 1; fv(Y)= l if0 < y < 1

4.12 Exercises, Problems, and Multiple Choice Questions 215 c) fx(x) = 2 if[x 1< 1; f,(y) = yI/2 if0 < y < 1 d) fx(x) = 2 if[x 1< 1; f,(y) = - ifx 2 < y < 1 e) fx(x) = 2 if[x 1< 1; f,(y) = 1 if0 < y < 1 B) Calculate PlY >_ X]. a) 1/8 b) 1/4 c) 1/2 d) 3/4 e) 7/8 C) Calculate Fx,,(�, 3)" a) 0 b) 1/27 c) 1/9 d) 1/3 e) l D) Are the random variables X and Y independent and orthogonal? a) correlated, therefore dependent, and not orthogonal b) dependent, but orthogonal c) independent, but not orthogonal d) independent and orthogonal e) uncorrelated, but dependent, and not orthogonal E) Find the best linear estimator of Y in terms of X. a) 0 b) 1/2 c) 3/5 d) X e) I+x2 2 F) We take ten independent observations of the pair (X, Y). i) What is the probability p that they are all located in the first quadrant? ii) What is the approximate value of p, according to the central limit theorem? I e-5 c) i) (�)10;ii) v' e-5 a) i) (�)0;ii) l e-2. b) i) (�)10;ii) I e-5 d) i) (�)0. ii) l e-2. e) i) (�)10. ii) Question no. 39 Let Ix+xy if0 <x < 1,-1 < y < 1, fx, r(x,y) / 0 elsewhere. A) Calculate f x (�). a) � b) l c)- d)2 e) �(1 +y) B) Calculate PIXY < 0]. a) � b) � c) � 6) 3 2. e) � C) Are the random variables X and Y orthogonal? correlated? independent? a) orthogonal and independent b) orthogonal and correlated c) not orthogonal, but independent d) not orthogonal, but correlated e) not orthogonal and uncorrelated, but not independent D) Let Z := 1IX. What is the best linear estimator of X in terms of Z? a) l/Z b) Z c)-6Z+6 d) 2/3 e) 2 E) Let Y, Y2 .... be independent random variables having the same distribution function as Y. We define Sn n

216 4 Random Vectors i) By the weak law of large numbers, toward what

value does Sn tend as n -- I d) x e) none of these answers a)-x b) 0 c).x ii) According to the central limit theorem, what is (approximately) the characteristic function of Sn? Hint. If X N(#, r2), then (px(w) = exp{j#co - �r2co2}. _ I co 2} b) exp{�njw- nw 2} c) exp{.jco- co 2} a) exp{ jw d) exp{ jw - co 2 } e) exp{- �co 2 } Question no. 411 A point X I is taken in the interval [0, 1] according to a uniform distribution. Let xl be the value taken by XI. Next, a point X2 is taken uniformly in the interval [xl, 11. We consider the random vector (XI, X2). A) What is the joint probability density function of (X i, X2)? a) 1 ifO _< x I _< 1,0 < x2 < 1 b) � if0 _< Xl -x2 _< 2 c) if0<xl <x2 < 1 d) if0<xl <x2 < 1 e) I if() <Xl <x < 1 (21n2- l)(l+Xl) B) Calculate E[X2 [XI = 3 b) c) d) 1+xf e) a) 2 2 C) Calculate E[X]. 5 b) II c) 2 d) 13 e) 7 a) 1 3 Question no. 41 A) Let X be a random variable having a uniform distribution on the interval [- 1, 1]. We define Y : X 3. Calculate VAR[X - Y]. a) 0.0762 b) 0.1905 c) 0.2762 d) 0.4762 e) 0.8762 B) Let XI ..... X36 be independent random variables such that 1 fxk(x)=- ill <x <e, X for k = 1 ..... 36. Use the central limit theorem to calculate the probability 36 P[I-Ik:l Xk > elg]. a) Q(0.38) b) Q(0.48) c) Q(0.58) d) Q(0.68) e) Q(0.78) Question no. 42 Let 1(@) 0 ifx<-lory<0, Fx,,(x,y) = (1 -e -y) if-1 <_ x < 1 andy > 0, 1-e Y ifx> landy>0 be the joint distribution function of the continuous random vector (X, Y).

4.12 Exercises, Problems, and Multiple Choice Questions 217 A) Calculate Fx (1/2). a) 1/8 b) 1/4 c) 1/2 d) 3/4 e) 7/8 B) Calculate P[X + Y > 0]. a) 0.8161 b) 0.8261 c) 0.8361 d) 0.8461 e) 0.8561 C) Are the random variables X and Y independent (or simply uncorrelated)? orthog- onal? a) independent and orthogonal b) uncorrelated, but not independent, and orthogonal c) correlated, therefore not independent, but orthogonal d) independent, but not orthogonal e) uncorrelated, but not independent, and not orthogonal Question no. 43 Let X1 ..... X36 be independent random variables having an exponential distri- bution with parameter 2. A) We define Y = 2X1 + X2. Calculate the covariance of Y and X1. a) 0 b) 1/4 c) 1/2 d) 3/4 e) l B) According to the central limit theorem, what is the approximate distribution of s := Xl +... + x36? Hint. We have: I'(x) = f yx-l e-y dy. a) N(0, 1) b) N(18, 9) c) N(18, 18) d) N(18, 36) e) N(18, 45) Question no. 44 Let (X, Y) be a random vector with joint probability density function Jx+yifO<x < 1, 0<y < 1, fx, l(x,y) I 0 elsewhere. A) Calculate Fx#(�, �)+ Fx, l(�,)- a) 1/16 b) 1/8 c) 1/4 d) 3/8 e) 1/2 B) Calculate P[-} < �]. a) 5/24 b) 1/4 c) 7/24 d) 1/3 e) 3/8 C) Calculate E[X [Y = �]. a) 11/18 b) 2/3

c) 13/18 d) 7/9 e) 5/6 D) We define the random variables $U = X + Y$ and $V = XY$. We can show that $U 2$ if $0 < u < 1$, $/2- 2v$ if $0 < v < 1$, $f(u) = 2u -- U 2$ if $1 < u < 2$, and $f_V(v) = | 0$ elsewhere. 0 elsewhere Calculate COV[U, V]. a)-1/18 b) 0 c) 1/18 d) 1/12 e) 1/9

218 4 Random Vectors E) Let V, V2 ..... V36 be independent random variables having the same probability density function as the random variable $V = XY$. i) Let $W := V - 2V2 + 3V3$. Find the characteristic function $4,w(w)$ of W in terms of that of V. a) 24'v (w) b) 4'v (w) + 4'v (-2w) + 4'v (3w) c) q}v (r..o) n t- q}2 (r..o) n t- q}3 V (r..o) d) qv()q2((o)q() e) qSv(w)qSv(-2w)qSv(3w) 36 ii) Use the central limit theorem to calculate $P[-= Vk > 11]$. a) 1 - Q(0.707) b) Q(0.707) c) 1 - Q(I/2) d) Q(I/2) e) 1/2 Question no. 45 We consider the random variables $V'.=aX+(I-a)Y$ and $W:=(l-a)X+aY$, where X $N(0, 1)$ and Y $N(- 1, 1)$ arc independent random variables, and where a is a constant taken from the interval $[0, l]$. A) For what value(s) of a are the random variables V and W orthogonal? a) 0only b) 1/2 only c) lonly d)()and lonly e) 0,1/2and lonly B) For what value(s) of a are the random variables V and W completely correlated (that is, $P.w = 1$)? a) 0 only b) 1/2 only c) 1 only d) 0 and 1 only e) 0, 1/2 and 1 only C) What is the best linear estimator of V in terms W, if $a = l/4$? 3 b)_3 3 3 3 3( ) a)- a a+sW c).gW d) W+ e)-+ 5 W+ D) Find the characteristic function of W, if $a = 1/2$. 1 2 2. Hint. If X $N(/z, or2)$, then $q_x(w) = \exp(j/zw - cr w)$. e) $\exp(- \blacklozenge w 2)$ E) Suppose that X ..... X30 and Y ..... Y30 are independent random variables distributed like X and like Y, respectively. We define $D_i = X_i - Y_i$ for $i = 1$ ..... 30. Let M be the number of Di's that will take on a positive value. Use a Gaussian distribution to calculate (approximately) $P[M = 28]$. Hint. We have: qb(0.7071) -- 0.7602 and qb(0.8071) -- 0.7902. a) 0.00045 b) 0.0045 c) 0.0145 d) 0.0245 e) 0.0445 Question no. 46 The joint probability density function of the continuous random vector (X, Y) is given by $]\blacklozenge x2(2+3y 2)$ if $-1 < x < 1, 0 < y < 1$, fx,r(x, y) ! 0 elsewhere.

We can show that $E[X] =0$, 4.12 Exercises, Problems, and Multiple Choice Questions 219 19 $E[X2] = 3 7$ and $ElY 2] = --$. 5' $E['] = 1-5$ 45 A) Are the random variables X and Y independent, or simply uncorrelated, and orthogonal? a) independent and orthogonal b) uncorrelated, but not independent, and orthogonal c) independent, but not orthogonal d) uncorrelated, but not independent, and not orthogonal e) not independent and not orthogonal B)

Calculate P[X > O, Y < �]. 1 a)6 b) � c) d) � e) C) Let Z := XY. What is the best estimator of Z in terms of X? a) 0 b) X c) �X d)2X e) XY D) Let W := 2X+3Y, and let Wl ..... W25 be independent random variables having the same distribution function as W. According to the central limit theorem, what is the approximate distribution of the sum W1 + ... + W25? a) N(41.75, 78.4375) b) N(41.75, 79.4375) c) N(42.75, 77.4375) d) N(43.75, 79.4375) e) N(43.75, 78.4375) Question no. 47 The joint probability density function of the continuous random vector (X, Y) is given by k if0<x<y< 1, fx,y(x, y) = 0 elsewhere. A) Find the constant k. a) 1/16 b) 1/8 c) 1/4 d) 1/2 e) 1 B) Calculate fx (1/4). a) k/2 b) k c) 2k d) 4k e) 8k C) Find the best (non-linear) estimator of Y in terms of X. a) 1-x3/2 b) -x3/2 c) d) - e) X 3(1--XU2) 3 3(1-XU2) Question no. 48 Suppose that X ..... X30 are independent random variables having a U(- 1, 1) distribution. We define S = '2�__1 X. A) Calculate the mean of S. a) 5 b) 10 c) 15 d) 20 e) 30 B) Use the central limit theorem to calculate (approximately) P [S > 12.5]. Hint. We have: E[S 2] = 924/9. a) Q(1.13) b) Q(1.23) c) Q(1.33) d) Q(1.43) e) Q(1.53)

5 Stochastic Processes 5.1 Introduction Definition 5.1.1. Let E be a random experiment and let S be a sample space associ- ated with E. A stochastic process (or random process) is a set of random variables {X(t, s), t E T}. Thus, with each result s S, we associate a function X (t, s). The domain oft is a set T of real numbers. Remarks. i) As in the case of random variables, we often suppress the argument s, and we denote the stochastic process (sp) by {X(t), t T}. In fact, many authors simply denote the stochastic process by X(t), without specifying the set T. This notation is generally sufficiently clear, because in most cases that are of interest to us the set T is the interval [0, x). However, there can be some confusion between the stochastic process itself and the value taken by the process at a (fixed) time t. Consequently, we will use the more complete notation. ii) The graph of X(t, s) as a function of t, for a fixed s, is called a realization or a trajectory (or a sample path) of the stochastic process. iii) If T is a denumerably infinite set, then we say that {X(t), t T} is a discrete- time stochastic process. Generally, we take T = {0, 1 .... }. We then have a sequence of random variables that we denote by X0, X1 ..... If T is an interval, or a set of intervals, {X(t), t T} is called a continuous- time stochastic process. Actually, we sometimes encounter the case when the set T is finite. However, this case is much less important in practice. iv) Moreover, we say that {X(t), t T} is a discrete-state (respectively continuous-

state) stochastic process if the set of possible values of the X(t)'s, called the state space, is finite or denumerably infinite (resp. non-denumerably infinite). Now, if t is fixed, then X(t, s) =-- X(t) is a random variable. We define the following functions. 1) First-order distribution function: $F(x; t) = P[X(t) < x] = P[\{s\ S: X(t, s) < x/\}]$. (5.1)

222 5 Stochastic Processes Remark. Note that the function $F(x; t)$ is in fact the distribution function of the random variable X(t), so that $F(x; t) = F_{x(t)}(x)$. Therefore, it is not a new concept. 2) First-order density function (for continuous-state sp's): $f(x; t) = x F(x; t)$ '�x where $F(x; t)$ is differentiable. (5.2) 3) Second-order distribution and density functions: $F(x, x_2; t, t_2) = P[X(t) < x, X(t_2) < x_2]$ and 02 $.f(x. x_2; t_l, t_2) -- --F(X_l, x_2: t_l, t_2)$ 'Ox Ox2 'v'(x, x2) where the second mixed partial derivative exists. 4) nth-order distribution function: (5.3) (5.4) $F(x_l ...... :_n; t_l ..... t_{,,}) = P[X(t_l) < x_l ..... X(t_n) < x_{nl}$. (5.5) Example 5.1.1 A Bernoulli process is a sequence X, X2 .... of random variables associated with Bernoulli trials. It is therefore a discrete-time and discrete-state stochastic process, since $T = \{ 1,2 .... \}$ and $S_{xk} = \{0, 1 \}$ 'v'k. Then, by indepen- dence, we may write, for instance, that $p(x, y; t_l = 1, t_2 = 2) = P[X_I : x, X_2 : y] = p^X(1 - p)^{l-x} p^Y(1 - p)^{l-y} = p^{X+Y}(1 _ p)^{2-x-y}$, where x and y 6 {0, 1 }. Furthermore, if we say that an arrival corresponds to a success, then the number N of trials between two arrivals (including the trial on which a success occurred) has a geometric distribution. 5.2 Characteristics of Stochastic Processes In order to establish the properties of a stochastic process, we must theoretically know its nth-order distribution function. However, we can also (partially) character- ize a stochastic process by its moments, with the help of the following functions. 1) Mean of a stochastic process: $m_x(t) = E[X(t)I = \int x f(x; t) dx$ (continuous case). (5.6)

5.2 Characteristics of Stochastic Processes 223 2) Autocorrelation function: ffff $R_x(t_l, t_2) = E[X(t_l)X(t_2)] = \int X_l X_2 f(x_l, x_2; t_l, t_2) dx_l dX_2$. (5.7) Remark. The value of $R_x (t, t) = E[X^2(t)]$ is sometimes called the average power of the stochastic process $\{X(t), t\ T\}$. 3) Autocovariance function: $C_x(t_l, t_2) = R_x(t_l, t_2) - m_x(t_l)m_x(t_2)$. (5.8) In particular, the variance of X (t) is given by $VAR[X(t)] = C_x (t, t)$. (5.9) 4) Correlation coefficient: $P_X(t_l, t_2) = C_x (t_l, t_2) [C_x (t_l, t_l)C_x (t_2, t_2)]^{1/2}$" (5.10) We have: $-1 <_ p_x(t_l, t_2) <_ 1$ and $p_x(t, t) = 1$. Remark. The functions $R_x (t_l, t_2)$, $C_x (q, t_2)$ and $P_x (t_l, t_2)$ are non-negative definite. Indeed, we have, for example: $a_i a_k R_x(t_i, t_k) = E a_i X(t_i > 0$ Vai G .

i=1 k=l i=1 Definition 5.2.1. If mx(t) is a constant and Rx(tl, t2) = Rx(t2 - q) Vtl, t2 T, we say that the stochastic process {X(t), t T} is wide-sense stationary (WSS). Remarks. i) We could replace the function Rx by the function Cx in the definition, since if mx (t) = c, then the functions Rx and Cx only differ by the constant c 2. ii) It would be more rigorous to write in the definition that we must have Rx (tl, t2) = R] (t2 - t), where R] is a function of a single variable. However, the formulation used above is classic. iii) The concept of stationarity will be seen in detail in Section 5.6 (p. 235). Definition 5.2.2. If the increments X(t4) - X(t3) and X(t2) - X(q) of the sp {X(t), t T} are independent (respectively uncorrelated) Vt < t2 <_ t3 < t4, we say that {X (t), t T} is a stochastic process with independent increments (resp. uncorrelated increments).

224 5 Stochastic Processes Definition 5.2.3. If the distribution of X (t2 + r) - X (tl q- r) and that of X (t2) - X (tl) are identical Yr, we say that {X(t), t E T} is a stochastic process with stationary increments. Remark. If two random variables are identically distributed, this means that they have the same distribution function. However, this does not mean that they are equal. Suppose now that we consider two stochastic processes: {X(t), t E T} and {Y(t), t 6 T2}. We then define the following functions. 1) Cross-correlation function: Rx.y(tl, t2) = ElX(tl)Y(t2)] = Ry, x(t2, tl). (5.11) 2) Cross-covariance function: Cx.y(tl, t2) = Rx,y(tl, t2) - mx(tl)mr(t2). (5.12) Definition 5.2.4. The stochastic processes {X(t), t TI} and { Y (t), t T2} are said to be a) orthogonal if Rx. r (t l, tz) = 0 Ytl, t2; b) uncorrelated if Cx, r (t l, t2) = 0 Ytl, tz. Definition 5.2.5. The stochastic processes {X(t), t TI} and {Y(t), t E T2} are called independent if the random variables X (tl) ..... X (tn) and Y (rl) ..... Y (rm) are (globally) independent Yn, m e { 1,2 .... }. Example 5.2.1 Let {X(t), t > 0} be the random process defined by X (t) = y2t fort >0, where Y is a random variable having a uniform distribution on the interval (0, 1). a) Calculate i) the first-order density function of the process; ii) the mean E[X(t)] of the process for t > 0; iii) the autocovariance function Cx(t, t s) of the process for s, t > 0. b) Is the process WSS? Solution. a) i) We can write (see Section 3.6, p. 94) that f(x; t) =- fx(t)(x) = fy(y) dx , where y = 1 1 = 1 - 2v/ if x6 (0, t).

5.3 Markov Chains 225 ii) We have: fo I t fort O. E[X(t)] = y2t. 1 dy = - We can also use part i) and write that t ex(tl = xdx = dx -- = fort >0. iii) Since X(t)X(t + s) = y2t y2 (t + s) = y4 t(t + s), we first calculate E[X(t)X(t + s)] =

$E[y4t(t+s)]$ -- $t(t)s.+5$ for $s, t > 0$. It follows that $Cx(t, t s) = E[X(t)X(t s)]$ - $E[X(t)]E[X(t s)]$ ii) $t(t+s)$ $t(t+s)$ $4t(t+s)$ .... for $s, t > 0$. 5 9 45 b) Since the mean of $X(t)$ depends on t (see part ii) of a)), the stochastic process $\{X(t), t > 0\}$ is not WSS. Note that the condition $Cx(t, t+s) = Cx(s)$ is not fulfilled either. Therefore, even if the mean $E[X(t)]$ were constant, the process would not be WSS. 5.3 Markov Chains The most important discrete-time stochastic process is known as a Markov chain. Definition 5.3.1. A stochastic process $\{Xn, n = O, 1 .... \}$ whose state space is either finite or denumerably infinite is called a Markov chain if $P[Xn+I = j I Xn = i, Xn-I = in-I ..... XO = i0]$ (=1) $P[Xn+I = j I Xn = i]$ (2) = Pij (5.13) for all states io ..... in-l, i, j and for all n $>\_$ O. Remarks. i) The equation (1) is called Markov's property. It means that the future, given the past and the present, is independent of the past and thus depends only on the present. ii) If the equation (2) is satisfied, we actually have a stationary (or time-homogeneous) Markov chain, because $P[Xn+t = j I Xn = i]$ does not depend on n. In the general

226 5 Stochastic Processes case, we can denote this probability by Pij (n). We can also write Pij as follows: pi,j. In fact, it is sometimes necessary to separate the states i and j by a comma to avoid any confusion (between the states 1 and 11, for instance). iii) We generally use the set N � := $\{0, 1 .... \}$ as the state space. Now, we deduce from the definition of the Pij's that Pij > 0 '�i, j (5.14) and E PiJ = 1 j=0 Vi. (5.15) Definition 5.3.2. The matrix P defined by r P() Pol � � � "] (5.6) is called the one-step transition probability matrix (or simply the transition ma- trix) of the Markov chain. Remark. The matrix P is called a stochastic matrix. If, in addition, we have: E PiJ = 1 i=0 V j, (5.17) the matrix P is said to be doubly stochastic. Definition 5.3.3. The set $\{Pi, i = O, 1 .... \}$, where Pi is defined by Pi = $P[Xo = i]$, (5.18) is called the initial distribution of the Markov chain. Finally, let Pij := $P[Xm+n = j [Xm = i]$, where m, n, i, j > 0. (5.19) That is, Pij denotes the probability of moving from state i to state j in n steps. Then, if p(n) denotes the matrix of the Pij s, we can show that $p(n) = pn$. (5.20) It follows that $p(n+m) = pnpm$. (5.21)

5.3 Markov Chains 227 The various equations obtained from Equation (5.21) are known as the Chapman- Kolmogorov 1 equations. Particular cases 1) A Markov chain whose state space is the set Z of all integers is called a random walk if Pi,i+l = P = 1 - Pi,i-1 for i = 0, �1, �2 .... (5.22) for a certain p 6 (0, 1). So, we have (with q := 1 - p): p= q 0 p q 0 p " If the state space is $\{0, 1 .....$

N} and Poo = PNN = 1, the states 0 and N are said to be absorbing. If P01 = 1, the state 0 is called reflecting. 2) Let { Yn, n = 0, 1 .... } be a set of i.i.d, random variables whose possible values are integers. Let ak := P[Yn = k] Then, if we define Xn=-Yk forn=0,1 ..... k=0 we find that {Xn, n = 0, 1 .... } is a Markov chain and that Pij = aj-i. We then say that the chain is homogeneous with respect to the state variable. Example 5.3.1 Suppose that a particle moves on the set of all integers and that, when it is in state i, the probability that it jumps from that state to state j depends also on the state it visited before i. More precisely, suppose that 1 Andrei Nikolaevich Kolmogorov, 1903-1987, was born and died in Russia. He was a great mathematician who, before getting his Ph.D., had already published 18 scientific papers, many of which were written during his undergraduate studies. His work on Markov processes in continuous time and with continuous state space is the basis of the theory of diffusion processes. His book on theoretical probability, published in 1933, marks the be- ginning of modern probability theory. He also contributed in an important way to many other domains of mathematics, notably to the theory of dynamical systems.

228 5 Stochastic Processes P[Xn+ : j I Xn = i, Xn- : k, Xn-2 : in-2 ..... Xo : i0] : e[Xn+l = J I xn : i, Xn- 1 = k] for all states io ..... in-2, i, j, k 6 Z and for n = 1, 2,.... If the state space of the stochastic process {Xn, n = 0, 1 .... }, where Xn denotes the position of the particle after n transitions, is the set Z of all integers, then the process is not a Markov chain. However, it is possible to satisfy Markov's property by defining a state of the process as the vector (k, i) consisting of the last two states visited by the particle (after at least one transition). For example, a generalization of the random walk is obtained by writing that the probability of going from (il, jl) to (i2, j2) is given by Pl if j2 = jl + 1,i2 = j andil = jl - 1, q ifj2:j-l,i2=jandi :j- 1, P(il,jl),(i2,j2) : P2 if j2 : j + l, i2 = j and il = jl + l, q2 ifj2=j- l,i2=jandi =j +1, 0 otherwise, where p + q + p2 + q2 = I. Similarly, if we must keep track of the last three states visited by the particle, then the states of the Markov chain will be triplets, etc. Thus, even if Markov's property is not satisfied with a given state space, it may be satisfied with another state space. 5.4 The Poisson Process One of the most important stochastic processes in applied sciences is the Poisson process. This process counts the number of events that occur in an interval. It is therefore a particular counting process, which we define first. Definition 5.4.1. A stochastic process {N(t), t >_ 0} is called a counting process if N (t) denotes

the total number of events that occurred up to time t. Properties. i) $N(t) \geq 0$ $\forall$ t $\geq O$. ii) $N(t)$ is a random variable whose possible values are 0, 1 ..... iii) $N(t)$ is non-decreasing. iv) If $t < t2$, then $N(t2) - N(t)$ is the number of events that occurred in the interval $(t, t2]$, Definition 5.4.2. A counting process $\{N(t), t \geq 0\}$ is called a Poisson process with rate ($> O$) if i) $N(0) = 0$; ii) $\{N(t), t \geq 0\}$ has independent increments; iii) $N(t + z) - N(z)$ Poi$Ot)$ ◆, $t > O$.

5.4 The Poisson Process 229 Remarks. i) The condition iii) implies that a Poisson process has stationary incre- ments. Moreover, we can write that $N(t) = N(t O) - N(O) = N(t)$ Poi$(.t)$ $\forall$ t $> O$. ii) We can in fact replace the condition iii) by the following condition: the sp $\{N(t), t \geq 0\}$ has stationary increments and where o(6) is such that $P[N(6) = 1] = .6 + 0(6)$, $P[N(6) = O] = 1 - .6 + o(6)$, (5.23) $0(6)$ lim -- = 0. (5.24) $\$0$ 6 This means that, in a short interval, the probability of two or more events occurring is negligible. Now, we have: $E[N(t)] = .t$ (because $N(t)$ Poi$(.t)$), 1 (5.25) $E[N2(t)] = VAR[N(t)] + (E[N(t)])$ 2 = .t + .2t2. Hence, if tl $< t2$, we can write (making use of the fact that the increments of a Poisson process are independent) that $RN(tl, t2) := E[N(tl)N(t2)] = E[N(tl)(N(t2) - N(tl))] + E[N2(tl)]$ in. $E[N(tl)]E[N(t2) - N(tl)] + E[N2(q)] =)tl)(t2 -- t) + ()tl +)2t12) =)2tlt2 +)tl$. Since $RN(tl, t2) = .RN(t2, tl)$, it follows that $[)tl q-)2tl t2 RN(tl, t2) =)t2 q-)2tlt2$ if tl $\leq$ t2, (5.26) if tl $>$ t2. Finally, we deduce from Equation (5.26) that $CN(tl, t2) ---- RN(tl, t2) -- ()tl)()t2) =)$ min$\{tl, t2\}$. (5.27) Properties. i) Let T1 be the arrival time of the first event, and Tn be the interarrival time between the (n - 1)st and the nth event, for n = 2, 3 ..... We can show that the Tk's are independent random variables for k $\geq$ 1 and that Tk Exp$(.)$ $\forall$k. Indeed, we have: $P[T1 > t] = P[N(t) = 0] = e^{-zt}$.

230 5 Stochastic Processes x(t) Figure 5.1. Example of trajectory of a telegraph signal process. It follows that the time required for n events to occur, from any time instant, has a gamma distribution with parameters ot = n and) (see Section 4.9, p. 187). If we denote the arrival times of the events by S, $\$2$ ..... we have: $S,, = Y'= Tk$ for n=l,2 ..... ii) Given that a single event occurred in the interval $(0, t]$, the time T, at which it happened has a uniform distribution on the interval $(0, t]$, Remark. The Poisson process is a special case of what is known as birth and death processes, which themselves are particular continuous-time Markov chains. It is also a particular renewal process. Now, from a Poisson process, we can define a stochastic process called the tele-

graph signal process, as follows: I ifN(t) is even, (5.28) X(t)= -I ifN(t) isodd (see Fig. 5. I), We have: ()t)2k P[X(t) = I I = PIN(t) = 2k] = e -t -- e-tcosh()t), (5,29) =o =o (2k) ! where "cosh" denotes the hyperbolic cosine. Then, we find that P[X(t) = -I] = I - e-Ztcosh()t) = e-Ztsinh()t). (5.30) It follows that E[X (t)] = I . [e-Ztcosh()t)] + (-I) - [e-Ztsinh()t)] = e -2)t. (5,31) Finally, we also find that Rx(t, t s) = e -2Isl. (5.32) Indeed, if s > 0, we can write that

5.4 The Poisson Process 231 Rx(t, t q- s) := E[X(t)X(t q- s)] = E[(--1)N(t)(--1) N(t+s)] = E[(--1)2N(t)(--1) N(t+s)-N(t)] .= E[(--1) N(s)] = E[X(s)] = e -2)s. Remark. {X (t), t > 0} is actually called a semi-random telegraph signal process. A random telegraph signal process is defined by Y(t) = Z. X(t), (5.33) where Z is a random variable independent of X(t) for all t and such that P[Z = 1] = P[Z = -1] = 1/2. We have: E[Y(t)] ind. E[Z]E[X (t)] = O . E[X (t)] = 0 and Rr(t, t + s) = E[ZX(t) ◆ ZX(t + s)] ind. E[Z2]Rx(t, t + s) = 1 e -2" s . Thus, {Y(t), t >_ 0} is a WSS stochastic process. Note that Z 2 is actually equal to the constant 1. Example 5.4.1 Telephone calls arrive at an exchange according to a Poisson process, at the (average) rate of one call per minute. a) What is the probability that (at least) five minutes have elapsed since i) the last call arrived? ii) the penultimate call arrived? b) Calculate the probability that no calls arrive during the next two minutes, knowing that at least one call was received over the last two minutes. Solution. Let N(t) be the number of calls in the interval [0, t]. We have: N(t) Poi(1 ◆ t), where t is in minutes. a) i) We seek P[N(5)=0]=P[Y=0], whereYPoi(5) = e -5 -- 0.0067. ii) We now seek P[N(5) < 1] = P[Y <_ 1] = e-5(5 + 1) _ 0.0404. b) Since the Poisson process has independent increments, we simply have to calcu- late P[N(2) = O] = e -2 -- 0.1353.

232 5 Stochastic Processes 5.5 The Wiener Process The most important continuous-time and continuous-state stochastic process is the Wiener 2 process, also known as Brownian motion. It is a particular case of the pro- cesses called Gaussian, which we define first. Definition 5.5.1. A stochastic process {X(t), t >_ 0} is called a Gaussian process if the random variables X (tl) ..... X (tn) have a multinormal distribution for all n and ◆tl ..... tn . Remark. Under some conditions, continuous-time and continuous-state stochastic processes are called diffusion processes. Consider a particle moving at random on the set of all integers. More precisely, consider the symmetric random walk, that is, the (discrete-time) Markov chain for which

$p_{i,i+1} = p_{i,i-1} = 1/2$ for $i = 0, \pm 1, \pm 2$ ..... (5.34) Suppose that we modify the process as follows: instead of moving one unit to the left or to the right at each time unit, the particle moves unit to the left or to the right at each 3 time unit, where and 3 are very small positive quantities. So, we speed up the process: the particle moves very often, but covers very little distance at each displacement. Let X(t) be the position of the particle at time t. Suppose that it is at the origin at time t = 0. Then, the position of the particle after the first n displacements is given by X (n3) = (2N - n), (5.35) where N is the number of steps to the right made in the course of these n displace- ments. Since N has a binomial distribution with parameters n and p = I/2, we can write that E[X(n3)] = (2(n/2) - n) = 0, 1 (5.36) VAR[X (n3)] = 42VAR[N] = 42(n/4) = n 2. Suppose now that we let 3 decrease to 0 in the random walk. We then obtain a continuous-time process. Since VAR[X(t)] = t2/3 (5.37) if t = n3, we must express in terms of 3. Let 2 = 0.23. (5.38) 2 Norbert Wiener, 1894-1964, was born in the United States and died in Sweden. He obtained his Ph.D. in philosophy from Harvard University at the age of 18. His research subject was mathematical logic. After a stay in Europe to study mathematics, he started working at the Massachusetts Institute of Technology, where he did some research on Brownian motion. He contributed, in particular, to communication theory and to control. In fact, he is the inventor of cybernetics, which is the "science of communication and control in the animal and the machine."

5.5 The Wiener Process 233 Then, when decreases to zero, the process obtained is also a continuous-state stochastic process and we have: E[X(t)] = 0, 1 (5.39) VAR[X(t)] cr2t as 4, 0. In fact, the variance of X(t) is equal to c2t for any positive value of . The stochastic process {W(t), t > 0} defined by W(t) = lim X(t) (5.40) $0 is called the Wiener process or, as mentioned above, a Brownian motion, named in honor of the English botanist Robert Brown, 3 who observed in 1827 the movement of microscopic particles suspended in a fluid. This motion seems completely irregu- lar, because of the molecular collisions that occur continuously. The process is also known as the Wiener-Einstein process or the Wiener-Bachelier process. Using the Gaussian approximation to the binomial distribution, we can write that P[W(t) _< w] = q . (5.41) That is, W(t) N(0, cNt). Remark. If we start with a non-symmetric random walk, then we obtain (under some conditions) a Wiener process with non-zero drift coefficient , so that W(t) N(/xt, o2t). We now give a more formal definition of a Wiener process. Definition S.82. A stochastic process {W(t), t

$> 0\}$ is called a Wiener process if i) $W(0) = 0$; ii) $\{W(t), t > 0\}$ is a process with independent and stationary increments; iii) $W(t)$ N(O, cr2t) Yt > O. Remarks. i) When $c = 1$, the process is called the standard Brownian motion. Since the variance of $W(t)/cr$ is equal to t, we can always transform an arbitrary Wiener process into a standard Brownian motion. ii) We can generalize the definition by letting $W(0)$ take on any real value w0 and then by replacing condition iii) by $W(t)$ N(w0, c2t) 'it > 0. Note that the process $\{W^*(t), t > 0\}$ defined by $W^*(t) = W(t) - wo$ then satisfies the preceding definition (if w0 is a constant). Actually, we could consider the case when $W(0)$ is a random variable W0 independent of W (t) Yt > O. iii) We can show that $W(t)$ is a continuous function of t with probability 1 (see Fig. 5.2, p. 234). 3 Robert Brown, 1773-1858, was born in Scotland and died in England. He is considered the greatest British botanist of the nineteenth century. He took part in an expedition to Aus- tralia, from which he brought many unknown species to Europe. In 1810, he was elected to the Fellowship of the Royal Society.

234 5 Stochastic Processes w(t) Figure 5.2. Example of trajectory of a Wiener process. To obtain the nth-order density function of $\{W(t), t > 0\}$, we use the fact that the Wiener process has independent and stationary increments. Then, if t < t2 < � � � < tn, we can write that f(Wl ..... Wn; tl ..... tn) = f(Wl; tl) IEI f(wk - U/k-l; tk - tk-l), (5.42) k=2 where f(w; t) is the probability density function of a Gaussian N(0, 0-2t) distribu- tion. Finally, if t < r, we have: Cw(t, r) -= COV[W(t), W(r)] = COV[W(t), W(t) + W(r) - W(t)] = COV[W(t), W(t)] + COV[W(t), W(r) - W(t)] in. COV[W(t), W(t)] = VAR[W(t)] = 0-2t, where we used the fact that COV[X, Y Z] = E[X(Y Z)] - E[X]E[Y Z] = {E[XY] - E[X]E[Y]} + {E[XZ] - E[X]E[Z]} = COV[X, Y] + COV[X, Z]. In the general case, we obtain: Cw(t, r) = 0 -2 min{t, r}. (5.43) Example 5.5.1 What is the distribution of $W(t) + W(r)$, where t < r? Solution. We can write that $W(t) + W(r) = 2W(t) + [W(r) - W(t)] := Y + Z$, where Y and Z are independent random variables having Gaussian distributions with zero means and variances given by

5.6 Stationarity VAR[Y] = 4VAR[W(t)] = 4ret and VAR[Z] = re(r -- t), 235 because the Wiener process has stationary increments. It follows that $W(t) + W(r)$ N(0, r2(3t + r)). Remark. It is important not to forget that the random variables $Z := W(r) -- W(t)$ and $W(r - t)$ have the same distribution, but are not identical. For example, suppose that t = 1 and r = 2. If W(2- 1) = 0, then this

does not imply that $Z = W(2) - W(1) = 0$, since $P[W(2) = 0] = 0$. Remark. Let $\{W(t), t > 0\}$ be a standard Brownian motion, and let f be a func- tion whose derivative exists and is continuous in the interval $[a, b]$. We define the stochastic integral f(t) dW(t). = f(b)W(b) -- f(a)W(a) -- W(t) df(t). (5.44) Note that this formula is the one that we obtain by integrating by parts. Moreover, if $f(t) = t\,2$, for example, then we have: W(t) df(t) = 2t W(t) dr. The process $\{dW(t), t > 0\}$ is called a (Gaussian) white noise. In engineering, people often write: dW(t)/dt = (= (t)). We can show that the integral defined in (5.44) has a Gaussian distribution with zero mean and variance b $f2(t)$ dt. 5.6 Stationarity Definition 5.6.1. A stochastic process $\{X(t), t\,T\}$ is said to be stationary, or strict-sense stationary (SSS), if F(Xl ..... Xn; tl ..... tn) = F(Xl ..... Xn; tl q- r ..... tn q- r) (5.45) 'r, n and tl ..... tn. If the process starts at $t = O$, for instance, then $t + r$ must be non-negative ◆k in Equation (5.45). Likewise, the sp's $\{X(t), t\,6\,TI\}$ and $\{Y(t), t\,6\,T2\}$ are jointly stationary if the joint distribution function of the processes remains unchanged if we add the constant r to every value oft in the function.

236 5 Stochastic Processes We deduce from the definition that if $\{X(t), t\,E\,T\}$ is a continuous SSS process, then f(x;t) = f(x;t+r) Yt, r, f(xl,x2;tl,t2) = f(xl,x2;tl +r, t2+r) Ytl,t2, r, so that f(x; t) = f(x) Yt, 1 (5.46) f(xl,x2;tl,t2)= f(xl,x2;t2-tl) Ytl,t2. These equations lead to the definition of WSS processes, introduced in Section 5.2. Definition 5.ti.2. A stochastic process $\{X(t), t\,T\}$ is said to be wide-sense sta- tionary (WSS) if E[X(t)] = m (a constant), 1 (5.47) E[X(t)X(t + r)] = Rx(r) (= Rx(-r)). We deduce from (5.47) that E[X2(t)] = Rx(O), (5.48) so that the average power of a WSS stochastic process is independent of t. Further- more, we can write that and Cx(t, t + r) = Cx(r) = Rx(r) - m 2 (5.49) px(t, t H- r) = px(r) = Cx(r)/Cx(O). (5.50) Finally, the sp's $\{X(t), t\,6\,T\}$ and $\{Y(t), t\,6\,T2\}$ are said to be jointly WSS if $\{X(t), t\,6\,T\}$ and $\{Y(t), t\,6\,T2\}$ are WSS and Rx,y(t + r, t) = E[X(t + r)Y(t)] = Rx,y(r) (5 Rx, y(-r), in general). (5.51) Similarly, we will have: Cx, y(t + r, t) = Cx, y(r) = Rx, y(r) - mxmy. (5.52) Remark. An SSS stochastic process is also WSS, but the converse is not always true. Proposition 5.ti.1. If a Gaussian process $\{X(t), t\,T\}$ is WSS, then it is also SSS. Proof. This follows from the fact that a Gaussian process is completely determined by its mean and its autocovariance function. Indeed, the nth-order characteristic

5.6 Stationarity 237 function of the process $\{X(t), t\,6\,T\}$ is given by exp jmx -

ok - $C_x(t_i - t_k) \sigma_i \sigma_k$. (5.53) $k=l$ $i=1$ $k=l$ Note that this function is invariant under any change of origin. [] Example 5.6.1 The Poisson process is not stationary. Indeed, we have: $E[N(t)] = \lambda t$. Since the mean of $N(t)$ is not constant, $\{N(t), t > 0\}$ is not WSS. Consequently, $\{N(t), t > 0\}$ is not stationary. However, the random telegraph signal $\{Y(t), t >_{-} 0\}$ is WSS because (see p. 231) $E[Y(t)] = 0$ and $R_Y(t, t + r) = e^{-2zr}$, where we assume that $r > 0$. Actually, we can show that the random telegraph signal is (strict-sense) stationary. However, the semi-random telegraph signal is not stationary, not even WSS. Example 5.6.2 The Wiener process has a mean equal to zero. However, its autoco- variance function is given by $C_W(l_1, l_2) = \sigma^2 \min\{t_1, t_2\}$. (5.54) Since this function does not depend only on the difference between t_2 and t_1, $\{W(t), t >_{-} 0\}$ is not a stationary process. In fact, we can show that $\min\{t_1, t_2\} = (/1 q- l_2 - |t_1 - t_2|)$. Remark. Given that a Gaussian process is completely determined by its mean and its autocovariancefunction, and that the Brownian motion is Gaussian, we can also define a Wiener process as a Gaussian process, with zero mean and autocovariance function given by Equation (5.54). In addition, the process must start at the origin. Example 5.6.3 Let $\{Z(t), t > 0\}$ be the process defined by $Z(0) = 0$ and $Z(t) := tW(1/t)$ if $t > 0$, where $\{W(t), t > 0\}$ is a Brownian motion. We have: $E[Z(t)] = tE[W(1/t)] = t \cdot 0 = 0$

238 5 Stochastic Processes and $C_Z(t_1, t_2) = E[Z(t_1)Z(t_2)] - 0^2 = E[t_1 W(1/t_1) t_2 W(1/t_2)] = t_1 t_2 C_W(1/t_1, 1/t_2) = t_1 t_2 \sigma^2 \min\{t^{-1}, t^{-2}\} = \sigma^2 \min\{t_1, t_2\}$. Since $\{Z(t), t > 0\}$ is a Gaussian process (because it is a linear transformation of a Gaussian process) that starts at the origin and whose mean and autocovariance function are identical to those of $\{W(t), t > 0\}$, we can state that it is a Brownian motion. 5.7 Ergodicity To estimate the mean $m_X(t)$ of a stochastic process $\{X(t), t \in T\}$ at time t, we must take a large number of observations $X(t, s_k)$ of the process. We write: $m_X(t) -- - X(t, s_k)$. (5.55) n $k=l$ That is, we estimate the mean of the sp at time t by the mean of a random sample taken at this time instant. Suppose, however, that we only have a single observation, $X(t, s)$, of $X(t)$. Can we then use the temporal mean $X(t, s) dt$ (5.56) $\langle x(t) \rangle_s := s$ to estimate $m_X(t)$? A necessary condition is that $m_X(t)$ be a constant. Remark. In (5.56), we can also integrate over the interval $[0, 2S]$ (for example if $T = [0, \infty]$). Definition 5.7.1. A stochastic process $\{X(t), t \in T\}$ is said to be ergodic if, with probability 1, every characteristic of $\{X(t), t \in T\}$ can be obtained from a single realization $X(t, s)$ of $X(t)$. There exist many types

of ergodicity. Here, we will limit ourselves to the case when the sp $\{X(t), t\ T\}$ is mean ergodic. Definition 5.7.2. A stochastic process $\{ X (t), t\ T \}$, whose mean is a constant m, is said to be mean ergodic/f, with probability 1, we have: lira $(X(t))s = m$. (5.57) Now, $(X(t))s$ is a random variable whose mean is given by if_, $E[X (t, s)]dt = m\ dt = m$. (5.58) $F[<X(t))s]= s\ s$

5.7 Ergodicity 239 Hence, we deduce that the sp $\{X(t), t\ E\ T\}$, for which mx(t) =-- m, is mean ergodic if and only if the variance of the random variable $(X(t))s$ decreases to 0 as S tends to infinity. Remark. For example, if X has a Gaussian N(#, r 2) distribution, then the mean of a random sample of size n of X has a Gaussian N(#, o'Z/n) distribution. Since VAR[)] . 0 as n tends to infinity, we can conclude that limn- = . We can show the following proposition. Proposition 5.7.1. Let $\{X(t), t\ T\}$ be a WSS process. Then, $\{X(t), t\ T\}$ is mean ergodic if and only VAR[(X(t))s] = 2s Cx(r) 1 - dr $ 0 as S . (5.59) Remarks. i) In the general case when the stochastic process $\{X(t), t\ E\ T\}$ is not necessarily WSS, the criterion (5.59) becomes lim VAR[(X(t))s] = lim Cx(tl, t2)dtldt2 = O. (5.60) S-- ec S-- ec ' S S ii) There exist other (mean) ergodicity criteria. For instance, a WSS process $\{X(t), t\ 6\ T\}$ is mean ergodic if Cx(O) < cxz and lim Cx(r) = 0. (5.61) Example 5.7.1 Consider once again the random telegraph signal, denoted by $\{Y(t), t > 0\}$ in Section 5.4. We already saw that Rr(r) = e -2)ll. Since $\{Y(t), t > 0\}$ is a process with zero mean, we can write that It follows that Cy(r) = e -2)Irl. e_2)r r VAR[(Y(t))s] = - 1 - dr. Given that the expression between the square brackets is bounded by 0 and 1 (when r varies from 0 to 2S), we can replace this term by 1, and we then find that 1 -- e -4)S VAR[(Y(t))s] < 2)S It is now easy to see that the variance of (Y(t))s decreases to 0 as S tends to infinity. Thus, by the previous proposition, the random telegraph signal is mean ergodic. Finally, we could have used the criterion (5.61) above. Indeed, we obtain: Cx(O) = 1 < cxz and lim Cx(r) = lim e -2)lrl ---- 0. Note, however, that this criterion is a sufficient, but not a necessary condition.

240 5 Stochastic Processes 5.8 Exercises, Problems, and Multiple Choice Questions Solved Exercises Exercise no. 1 (5.3) We suppose that the probability that a certain machine functions without failure today is equal to 0.7 if the machine functioned without failure yesterday and the day before yesterday (state 0); 0.5 if the machine functioned without failure yesterday, but not the day before yes- terday (state 1); 0.4 if the machine functioned without

failure the day before yesterday, but not yes- terday (state 2); 0.2 if the machine did not function without failure, neither yesterday nor the day before yesterday (state 3). a) Find the one-step transition probability matrix of the Markov chain associated with the functioning state of the machine. b) Calculate p0 2.1, that is, the probability of moving from state 0 to state 1 in two steps. c) Calculate the average number of days without failure of the machine over the next two days, given that the Markov chain is presently in state 0. Solution a) We have, for example: $p_{0,2}$: P[At least one failure today { state 0] = 1 - 0.7 = 0.3, etc. We find that 0.7 0 0.3 0 0.5 0 0.5 0 0.2 0 0.8 b) We seek p_2 ol = Po,oPo.1 + po, 1Pl + $p_{0,z}$P2l + po,3P3l = 0 +0 + (0.3)(0.4) +0 = 0.12. c) Let N be the number of days without failure over the two-day period. We have: 2 if the Markov chain moves from 0 to 0 to 0, 1 if the Markov chain moves from 0 to 2 to 1 or from 0 to 0 to 2, 0 if the Markov chain moves from 0 to 2 to 3. It follows that

5.8 Exercises, Problems, and Multiple Choice Questions E [N] = 2 x Po,oPo,o + 1 x (Po,2P2,1 + Po,oPo,2) +0 x Po,2P2,3 = 2 (0.7) 2 + (0.3) (0.4) + (0.7) (0.3) + 0 = 1.31. 241 Exercise no. 2 (5.4) Let N (t) be the number of failures of a computer system in the interval [0, t]. We suppose that {N(t), t > 0} is a Poisson process with rate . = 1 per week. a) Calculate the probability that i) the system functions without failure during two consecutive weeks; ii) the system has exactly two failures during a given week, knowing that it has functioned without failure during the previous two weeks; iii) less than two weeks elapse before the third failure occurs. b) Let Z (t) := e -N(t) for t _> 0. Is the stochastic process {Z(t), t > 0} wide-sense stationary? Justify your answer. Hint. We have: E[e -sx] = exp lot (e -s - 1)], if X Poi(ot). Solution a) i) We seek P IN (2) = 0] = P [Poi (2) = 0] = e -2 -- 0.1353. ii) Since the increments of a Poisson process are independent, we look for 12 P IN (1) = 2] = P [Poi(1) = 2] = e -1- 0.1839. 2! iii) Let $3 be the time when the third failure occurs. We have: $3 < 2 ❓:> N (2-) > 3. We then calculate P[S3 <2]=P[N(2) >3]= 1-P[N(2) <2]= 1-P[Poi(2)<2] I e 2 2. 1 = 1-- e -2+e-22+ - = 1--5e -2 - 0.3233. Or: $3 G(o = 3, ;k = 1), so that fo 2x2e-x 2 f02 P[S3 <2]= dx=-- e -x + xe -x dx 0 02 -x 2 = --2e -2- xe-X] + e -x dx = --4e -2 - e 10 = 1- 5e -2. b) We have: N(t) Poi(t), so that E [Z (t)] = exp [t (e -1 - 1)]. Since E [Z (t)] is not a constant, the stochastic process {Z(t), t > 0} is not WSS.

242 5 Stochastic Processes Unsolved Problems Problem no. 1 Let {N(t), t _>

0} be a Poisson process with rate . We define the stochastic process {X(t), 0 < t < 1} by X(t):N(t)-tN(1) for0<t < 1. a) Calculate the mean of X (t). b) Calculate the autocovariance function of the process {X (t), 0 < t < 1}. Hint. We have: RN (tl, t2) =)2tlt2 +) min{tl, t2}. c) Is the process {X(t), 0 < t < 1} WSS? Justify. Problem no. 2 Let {W(t), t _> 0} be a standard Brownian motion. We define X (t) = W 2 (t) for t > 0. a) Is the stochastic process {X (t), t > 0} a Wiener process? Justify. b) Is {X (t), t > 0} i) WSS? ii) mean ergodic? Justify your answers. Problem no. 3 The stochastic process {X(t), t > 0} is defined by X(t)=e -rt for t >0. We suppose that Y has a uniform distribution on the interval (0, 1). a) Calculate E [X (t)]. b) Calculate Cx (t, t + s), where s, t > 0. c) Is the stochastic process {X(t), t > 0} WSS? Justify your answer. d) Find the first-order density function of the process {X (t), t > 0}. Problem no. 4 We suppose that customers arrive at a counter in a bank according to a Poisson process with rate) = 10 per hour. Let N (t) be the number of customers in the interval [0, t]. a) What is the probability that no customers arrive during a 15-minute period? b) Knowing that eight customers arrived during a given hour, what is the probability that at most two customers arrive during the following hour? c) Knowing that a customer arrived during a 15-minute period, what is the probability that he arrived during the first five minutes of the period considered? d) Let X (t) := N (t). Is the stochastic process {X(t), t > 0} WSS? Justify your answer. Hint. We have: RN (tl, t2) = 10 min{q, t2} + 100qt2.

5.8 Exercises, Problems, and Multiple Choice Questions 243 Problem no. 5 Let {W(t), t process {X(t), t >_ 0} by X(t)=e -tw(e 2t) for t >0, where ot is a positive constant. a) Calculate the mean of X (t). b) Calculate the autocovariance function of the process {X (t), t > 0}. Hint. We have: Cw (t, r) ---- rain{t, r}. c) Is the stochastic process {X(t), t >_ 0} i) a Brownian motion? ii) stationary? iii) mean ergodic? Justify your answers. Problem no. 6 Let {N(t), t > 0} be a Poisson process with rate . > 0. We define the stochastic process {X (t), t > 0} by N(t + 32) - N (t) X(t)= for t >_0, where 3 > 0 is a constant. a) Is the process {X (t), t >_ 0} a Poisson process? Justify. b) Calculate the mean of X (t). c) Calculate the autocovariance function of the process {X(t), t > 0} for tl = 1, t2:2and3: 1. Hint. We have: RN (tl, t2) :) min{tl, t2} + .2tlt2. d) We consider the process {Y(t), t >_ 0} defined from {X(t), t > 0} by taking the limit as 3 decreases toward zero. We can show that E [Y (t)] ---- 0 and that -t2:0, Cy (tl,t2) = 0 otherwise. Is the process {Y(t), t

≥ 0} i) WSS? ii) mean ergodic? Justify your answers. e) Let Zn :: N (n) for n : 0,1, 2 ..... i) The stochastic process {Zn, n : 0, 1 .... } is a Markov chain. Justify this as- sertion. ii) Calculate Pi,j for i, j E {0, 1, 2 .... }. Problem no. 7 Let {W(t), t $\geq$ 0} be a standard Brownian motion. We define the stochastic process {X(t), t $\geq$ 0} by X(t) : W(t) W(t) fort $\geq$ 0. a) Calculate the mean of X (t). b) Calculate COV[X(t), X(t r)] for r > O. Hint. We have: COV[W(tl), W(t2)] = min{tl, t2}. > 0} be a standard Brownian motion. We define the stochastic

244 5 Stochastic Processes c) Is the stochastic process {X(t), t $\geq$ 0} i) Gaussian? ii) stationary? iii) a Brownian motion? Justify your answers. d) Calculate the correlation coefficient of W(t) and W(t 2) for t > 0. Problem no. 8 Let {Xn, n = 0, 1 .... } be a Markov chain whose state space is the set {0, 1 } and whose one-step transition probability matrix P is given by p l-p whereO $\leq$ p $\leq$ I. a) Suppose that p = 1 and that X0 = O. Calculate E[X2]. b) Suppose that p= I/2 and that P[Xo = 0]= PIXo = I1= I/2. We define the continuous-time stochastic process {Y(t), t > O} by Y(t) = tXit for t > O, where [t] denotes the integer part of t. i) Calculate Cr(t, t + I). ii) Is the stochastic process {Y(t), t > O} WSS? Justify your answer. iii) Calculate limn--.ec P[Xn = 0]. Multiple Choice Questions Question no. 1 A man plays independent repetitions of the following game: at each repetition, he throws a dart onto a circular target. Suppose that the distance D (in centimeters) between the impact point of the dart and the center of the target has a U[0, 30] dis- tribution. If D $\leq$ 5, the player wins $1 ; if 5 < D < 25, the player neither wins nor loses anything; if D $\geq$ 25, the player loses $I. The player's initial fortune is equal to $I and he will stop playing when either he is ruined or his fortune reaches $3. Let Xn be the fortune of the player after n repetitions. Then, the stochastic process {Xn, n = 0, 1 .... } is a Markov chain. A) Find the one-step transition probability matrix of the chain. a) 1/62/3 I/6 0 I/3 I/3 I/3 0 0 1/62/3 I/6 b) 0 I/3 I/3 I/3 0 1/62/3 I/6 0 0 0 dt /62/3/6 0 /62/3/6 0 0 I/6 2/3 I/6 e) | 0 I/6 2/3 I/6 0 0 0 I 0 I 0 B) Calculate E[X]. a) I/3 b) I c) II/18 d) 4/3 e) 29/18 /3 /3 /3 o c) 1/31/31/3 o o /3 /3 /3 o /3 /3 /3

5.8 Exercises, Problems, and Multiple Choice Questions 245 Suppose now that the man never stops playing, so that the state space of the Markov chain is the set {0, + 1, +2 .... }. Suppose also that the duration T (in sec- onds) of a repetition of the game has an exponential distribution with mean 30. Then, the

stochastic process $\{N(t), t > 0\}$, where $N(t)$ denotes the number of repetitions played in the interval $[0, t]$, is a Poisson process with rate $) = 2$ per minute. C) Calculate the probability that the player has completed at least three repetitions in less than two minutes. a) 5e -4 b) 13e -4 c) 1 -5e -4 d) 1 - 13e -4 e) 1 - e -4 D) Calculate (approximately) the probability $P[N(25) < 50]$. a)"0 b) 1/4 c) 1/2 d) 3/4 e)" 1 Question no. 2 Let $\{X_n, n = 1, 2 .... \}$ be a Bernoulli process. That is, the random variables XI, X2 .... are independent and all have a Bernoulli distribution with parameter p. Calculate the autocorrelation function Rx(n, m) of the process, if p = 1/2 and n,m E $\{1,2 .... \}$. a) 1/4 for all n, m b) 1/2 for all n, m c) 1/4 ifn -J: m; 1/2 ifn ---- rn d) 1/2 ifn -J: m; 1/4 ifn ---- rn e) none of these answers Question no. 3 The one-step transition probability matrix P of a Markov chain whose state space is $\{0, 1 \}$ is given by 0 " Calculate E[X2], if $P[X_o = O] = 1/3$. a) 1/12 b) 1/4 c) 1/3 d) 1/2 e) 11/12 Question no. 4 Let $N(t)$ be the number of telephone calls received at an exchange in the interval $[0, t]$. We suppose that $\{N(t), t > 0\}$ is a Poisson process with rate $) = 10$ per hour. Calculate the probability that no calls are received during each of two consecutive 15-minute periods. a) e -5 b) e -2"5 c) 2e -2'5 d) e -0"25 e) 2e -0"25 Question no. 5 Calculate the variance of $W(t) - 2W(z)$ for $t < z$, where $\{W(t), t > 0\}$ is a standard Brownian motion. a) t + 2z b) t + 4z c) 3t + 4 d) 5t + 4 e) 9t + 2 Question no. 6 Let $\{X(t), t >\_ 0\}$ be the stochastic process defined by $X(t) = tU + l$, where U is a random variable having a U(0, 1) distribution. Find the first-order den- sity function of the process.

246 5 Stochastic Processes 1 lif0<x< t a) lif0<x < 1 b) lif0 <t < l c) 2if0<t <x d) 7 l ifl <x<t+l e) 7 Question no. 7 The customers of a newspaper salesman arrive according to a Poisson process with rate . = 2 per minute. Calculate the probability that at least one customer arrives in the interval (to, t0 + 2], given that there has been exactly one customer in the interval (to - l, t0 + 1]. a) e-2/2 b) e 2 c) (1 - e-2)/2 d) 1 - e -2 e) 1 - (e-2/2) Question no. 8 Let $\{W(t), t >\_ 0\}$ be a standard Brownian motion. We define $X(t) = IW(t)l$ for $t >\_ 0$. Is the stochastic process $\{X(t), t >\_ 0\}$ Gaussian and stationary'? a) Gaussian and SSS b) Gaussian, but not stationary c) not Gaussian, but SSS d) not Gaussian and not stationary e) not Gaussian, but WSS (only) Question no. 9 Let $\{X(t), t >\_ 0\}$ be a stochastic process whose autocorrelation function is Rx(tl, t2) = e -It-t-I + 1 and autocovariance function is Cx(t, t2) = e -Iq-t'-I. Is the stochastic process $\{X(t), t >\_ 0\}$ WSS and mean ergodic? a) WSS and ergodic b) not stationary, but ergodic c) WSS, but not ergodic d) not stationary and not

ergodic e) we cannot conclude Question no. 10 Let {X(t), t > 0} be the stochastic process defined by X(t)=W(t+I)-W(I) for t_> 0, where {W(t), t >_ 0} is a standard Brownian motion. A) Calculate the autocovariance function of the process {X (t), t >_ 0}. a) min{tl, t2} - I b) min{t, t2} c) min{t, t2} + I d) 2min{q,t2} e) none of these answers B) Is the process {X(t), t >_ 0} Gaussian and a standard Brownian motion? a) Gaussian and a standard Brownian motion b) Gaussian and a non-standard Brownian motion c) Gaussian, but not a Brownian motion d) not Gaussian and not a standard Brownian motion e) not Gaussian and not a non-standard Brownian motion C) Is the process {X(t), t >_ 0} stationary and mean ergodic? a) WSS only, and ergodic b) SSS and ergodic c) SSS and not ergodic d) not stationary, but ergodic e) not stationary and not ergodic

5.8 Exercises, Problems, and Multiple Choice Questions 247 Question no. 11 Let {Yk, k = 1, 2 .... } be a set of independent random variables having a Bernoulli distribution with parameter p = 1/3. We define Xn = -]=l Yk for n = 1, 2 ..... Then, the stochastic process {Xn, n = 1, 2 .... } is a Markov chain. Calculate p03,2. a) 1/27 b) 2/27 c) 1/9 d) 2/9 e) 4/9 Question no. 12 We consider a standard Brownian motion { W (t), t_> 0}. Calculate VAR[W(4) - 2W(1)]. a) 0 b) 2 c) 4 d) 6 e) 8 Question no. 13 Let {X(t), t_> 0} be a WSS stochastic process, with zero mean and with an autocorrelation function given by Rx () = e-lrl. Let Y(t) :: tX2(1/t) for t > 0. Is the stochastic process {Y(t), t > 0} WSS and mean ergodic? a) not WSS and not ergodic b) not WSS, but ergodic c) WSS, but not ergodic d) WSS and ergodic e) not WSS and, consequently, we cannot conclude about the ergodicity Question no. 14 Let {Xn, n : 0, 1 .... } be a random walk for which 2 1 pi,i+l : and pi,i-1 ---- for i {0, �1, �2 .... }. Calculate E[X2 I Xo --- 0]. a) 0 b) 1/9 c) 1/3 d) 4/9 e) 2/3 Question no. 15 We consider a stochastic process {X(t), t > 0} for which 1 1 Cx(tl, t2) -- and Rx(t, t2) -- It2 -- tl 1 It2 -- tl 1 Is the process stationary and mean ergodic? a) not WSS and not ergodic b) not WSS, but ergodic c) SSS, but not ergodic d) WSS and ergodic e) WSS (only), but not ergodic 4.

248 5 Stochastic Processes Question no. 16 We define the stochastic process {X(t), t > 0} by X(t) = t/Y for t > 0, where Y has a uniform distribution on the interval (0, 2). Find f(x; t) forx > t/2. t t t t 1 a) - b) c) d) e) Question no. 17 The stochastic process {X(t), t > 0} is defined by X(t)=N(t+I)-N(I) fort >0, where {N(t), t > 0} is a Poisson process with rate)` > 0. Calculate Cx(s, t) for

0<s<t. Hint. We have: Cu(s, t) = Ks,if0 < s < t. a))`(s -- 1 ) b) Ks c) )`(t - s) d)),st e) Question no. 18 Let {W(t), t >_ OI be a standard Brownian motion. We detine X(t)=-W(t) fort >0. Is the stochastic process {X(t), t > 0} Gaussian? Is it a Brownian motion? a) not Gaussian and not a Brownian motion b) not Gaussian, but a Brownian motion c) Gaussian, but not a Brownian motion d) Gaussian and a standard Brownian motion e) Gaussian and a non-standard Brownian motion Question no. 19 Let {Xn, n = 0, I, 2 .... } be a Bernoulli process. That is, the random variables X,, are independent and all have a Bernoulli distribution with parameter p. Calculate the particular case p(0, 1; nl = 0, n2 = 1) of the second-order probability mass function of the process. a) 0 b) p2 c) p(l - p) d) 2p(l - p) e) (1 - p)2 Question no. 20 We define Y(t) = / 0 if N(t) is odd, 1 if N(t) is even, where {N(t),t _> 0} is a Poisson process with rate), = 1. We can show that PlY(t) = 1] = (1 + e-2t)/2 for t >_ 0. Next, let Xn := Y(n) for n = 0, 1 ..... Then, {Xn, n = 0, 1 .... } is a Markov chain. Calculate its one-step transition proba- bility matrix. [1/2 -] [10 [] I lq-e 22 l-e-2 12 e) I-e22 lq-e-2 12 I/2. c) d) l-e -2 I+e 2 l-e 2 a) 1/2 1/2J b) 1 1+ e-2 2 2 2 2

5.8 Exercises, Problems, and Multiple Choice Questions 249 Question no. 21 Let {X(t), t > 0} be a Gaussian process such that X(0) = 0, E[X(t)] = Izt for t > 0, where/z 0, and Rx(t,t+r)=2t+Iz2t(t+r) fort, r >0. Is the stochastic process {Y(t), t > 0}, where Y(t) := X (t)-/zt, a Brownian motion? Is it stationary? a) not a Brownian motion and not stationary b) not a Brownian motion, but stationary c) a standard Brownian motion, but not stationary d) a non-standard Brownian motion, but not stationary e) a non-standard Brownian motion and stationary Question no. 22 Suppose that {X(t), t > 0} is a Gaussian process, with zero mean, whose auto- covariance function is given by Cx(t,t+r)=e -t fort, r >_0. We define Y(t) = X2(t). Is the stochastic process {Y(t), t > 0} stationary? Is it mean ergodic? a) not stationary and not ergodic b) not stationary, but ergodic c) WSS only, but not ergodic d) SSS, but not ergodic e) WSS only, and ergodic Question no. 23 Let {Xn, n = 0, 1 .... } be a Markov chain whose state space i.s the set {0, 1 } and whose one-step transition probability matrix is given by A) Calculate Cx(t = 0, t2 = 1), if P[Xo = 0] = P[Xo = 1] = 1/2. a)-1/4 b)-l/8 c) 0 d) 1/4 e) 1/2 B) Calculate limn_ P[X = 0 I X0 = 0]. a) 0 b) 1/2 c) 1 d) does not exist e) none of these answers Question no. 24 The failures of a certain machine happen according to a Poisson process with rate) = 1 per week. A) What is the probability that the machine has at least one failure during each of the first two weeks considered? a) e -2 b) 1 -

2e - c) (1 - e-) 2 d) 2e -I e) 1 - e -2 B) Suppose that exactly five failures have occurred during the first four weeks con- sidered. Let M be the number of failures during the fourth of these four weeks. Calculate ElM I M > 0]. a) 1.24 b) 1.34 c) 1.44 d) 1.54 e) 1.64

250 5 Stochastic Processes Question no. 25 A) Let X, X 2 .... be an infinite sequence of independent random variables having a Poisson distribution with parameter ot = 1. We define Yn =) Xk forn= 1,2 ..... k=l 4 Then, {Yn, n = 1,2 .... } is a Markov chain. Calculate Pl.3, that is, the probability of moving from state 1 to state 3 in four steps. a) 0.1065 b) 0.1165 c) 0.1265 d) 0.1365 e) 0.1465 B) Let {N(t), t > 0} be a Poisson process with rate . > 0. We define M (t) = N(). Mz(t) = N(2t), M3(t) = N(t + 2) - N(2). Which of the {Mk (t), t > 0} stochastic processes is (are) also a Poisson process? a) M2(t) only b) M3(t) only c) Mz(t) and M3(t) only d) none e) all Question no. 26 The power failures in a certain region occur according to a Poisson process with rate . = l/5 per week. Moreover, the duration X (in hours) of a given power failure has an exponential distribution with parameter -2 = l/2. Finally, we assume that the durations of the various power failures are independent random variables. A) What is the probability that the longest, among the first three power failures ob- served, lasts more than four hours? a) 0.3435 b) 0.3535 c) 0.3635 d) 0.3735 e) 0.3835 B) Suppose that there has been exactly one power failure during the first week con- sidered. What is the probability that the failure had still not been repaired at the end of the week in question? a) 0.0019 b) 0.0119 c) 0.0219 d) 0.0319 e) 0.0419 Question no. 27 The flow of a certain river can be in one of the following three states: 0: low flow, 1: average flow, 2: high flow. We suppose that the stochastic process {Xn, n = 0, 1 .... }, where Xn represents the state of the river flow on the nth day, is a Markov chain. Furthermore, we estimate that the probability that the flow moves from state i to state j in one day is given by the formula 1 - - - li - jlzri, Pi,j -- 2 where 0 < :'ri < 1, for i, j = 0, 1,2.

5.8 Exercises, Problems, and Multiple Choice Questions 251 A) Calculate the probability that the river flow moves from state 0 to state 1 in one day. a) 1/6 b) 1/4 c) 1/3 d) 5/12 e) 1/2 B) Calculate the probability that the river flow moves from state 0 to state 2 in two days. a) 1/6 b) 1/4 c) 1/3 d) 5/12 e) 1/2 Question no. 28 A machine is made up of two components that operate independently. The life- time Ti (in days) of component i has an exponential

distribution with parameter for/ = 1,2. A) Suppose that the two components are placed in parallel, and that ,kl = 2 = In 2. When the machine breaks down, the two components are replaced by new ones at the beginning of the following day. Let Xn be the number of components that function at the end of n days. Then, the stochastic process {Xn, n = 0, 1 .... } is a Markov chain. Calculate its one-step transition probability matrix. a) | 1/2 0 1/2 b) 1/2 1/2 0 c) 1/2 1/2 11/4 1/4 1/2 1/4 1/4 1/2 1/2 1/4 1/4..] d) 0 1/2 1/2 e) 1/2 1/2 0 1/4 1/4 1/2 1/4 1/2 1/4 B) Suppose that the two components are placed in series and that as soon as a com- ponent fails, it is replaced by a new one. Let N(t) be the number of replacements in the interval [0, t]. We can show that the stochastic process {N(t), t > 0} is a Poisson process. Give its rate Hint. If To is the time between two replacements, then we can write that To min{T1, T2}. a) �(1 +2) b))l +2 c) �1 2 d))1)2 e) min{,kl,,k2} Question no. 29 Let {Xn, n = O, 1 .... } be a Markov chain whose state space is the set {0, 1, 2, 3, 4}, and whose one-step transition probability matrix is i 0 0 O0 0.5 0.2 0.3 0 0 P= 0 0 10 0 0 01 0 1 00 A) Calculate the probability that the process moves from state 1 to state 2 in four steps. a) 0.2824 b) 0.2924 c) 0.3024 d) 0.3124 e) 0.3224

252 5 Stochastic Processes B) Suppose that X0 = 1. Let N be the number of times that state 1 will be visited, including the initial state. Calculate E[N]. a) 1.25 b) 1.5 c) 1.75 d) 2 e) 2.5 Question no. 30 We define the stochastic process {X(t), 0 < t < 1} by X(t)=N(t 2)-t2N(1) for0<t < 1, where {N(t), t > 0} is a Poisson process with rate . > 0. A) Calculate the autocorrelation function Rx(tl, t2) of {X(t), 0 < t < 1 } at tl = 1/4 and t2 = 1/2. Hint. We have: R N (tl, t2) =)2tlt2 +) min{tl, t2}. a) ./64 b) ./32 c) 3./64 d) ./16 e) 5./64 B) Calculate P[X(t) > 0 I N(1) = 1] for0 < t < 1. a) 1/2 b) t c) l-t d) t e) l-t

6 Estimation and Testing 6.1 Point Estimation Definition 6.1.1. A random sample of size n of a random variable X is a set XI ..... Xn of independent random variables whose distribution is identical to that of X. That is, Xg has the same distribution function as X, for k = 1 ..... n. The random variable X is also called the population and each Xg is an observation of X. Definition 6.1.2. A statistic is a function, that does not depend on any unknown parameter, of the observations in a random sample. The distribution of a statistic is called a sampling distribution. Example 6.1.1 The sample mean: := - (6.1) n k=l is a statistic. Similarly, the sample variance: S 2 := . (Xk 2)2 (6.2) n-1 k=l is a statistic. However, the quantity (Xk_ /)2 n 1 k=l is a statistic only if/, := E[X]

is known. If the population X is Gaussian, then the sampling distribution of is also Gaus- sian. Finally, if X does not have a Gaussian distribution, but n is large enough, then the sampling distribution of is approximately Gaussian, by the central limit theo- rem.

254 6 Estimation and Testing Definition 6.1.3. A point estimate of an unknown parameter, O, of a population is a real number that corresponds to this parameter. Definition 6.1.4. An estimator of an unknown parameter, O, of a population is a statistic T = g(Xi ..... Xn) that corresponds to this parameter. Remarks. i) An estimator T is a random variable, since it is a function of a random sample. ii) The quantity 0 can, actually, be a vector: 0 = (01 ..... Ore). Example 6.1.2 The sample mean, ,, is an estimator of the (unknown) population mean #. Likewise, S 2 is an estimator ofo. 2 := VAR[X]. Moreover, numerical values obtained with a particular random sample, that is, � and s 2, are point estimates of # and o.2, respectively. Properties Definition 6.1.5. An estimator T of an unknown parameter 0 is said to be unbiased if E[T [= O. Furthermore, the bias $'T is defined by Bias[Tl = E[T[- O. (6.3) Remark. The estimator T (= T(n)) is said to be asymptotically unbiased if lim E[T] = O. Example 6.1.3 The sample mean is an unbiased estimator of #, because Likewise, S 2 is an unbiased estimator ofo. 2. Indeed, we can show that $2 :: !-(Xk--)2 _ 1 ['-X2_n.21k � k= n- 1 n-- 1 k,=l Moreover (see Chapter 4, p. 185), VAR[,] ind.� VAR[Xk]n 2 i.d. n 21 no.2 : --.�2n k:l Then, making use of the formula (6.4)

6.1 Point Estimation 255 E[X 2] = VAR[X] + (E[X]) 2, we find that E[S 2] = 0 "2. Indeed, we have: Definition 6.1.6. The mean square error of the estimator T of an unknown popula- tion parameter 0 is given by MSE[T] = E[(T - 0)2]. (6.5) Remarks. i) Expanding the squared expression in the above definition, we find that MSE[T] = VAR[T] + (Bias[T]) 2. (6.6) Thus, if T is an unbiased estimator of 0, then MSE[T] = VAR[T]. ii) Let T1 and T2 be two estimators of 0. We say that T1 is relatively more efficient (or simply better) than T2 if MSE[T] < MSE[T2]. (6.7) Example 6.1.4 Let X1 ..... Xn be a random sample ofa Gaussian population X with mean # and variance 0.2, where # and 0 .2 are unknown parameters. Then, making use of Equation (6.4), we can write that O-2 MSE[] = VAR[] = --. (6.8) n Moreover, we can show that MSE[S2] _ 2 0.4. (6.9) n-1 Actually, we find that the (biased) estimator of 0 .2 obtained by dividing by n rather than by n - 1 in (6.2) is relatively more efficient than S 2.

Definition 6.1."I. Let T be an estimator of O based on a random sample of size n. We say that T is a consistent estimator of O if T converges in probability to O, that is, if lim P[IT-OI <]= 1 'v' >0. (6.10) Example 6.1.5 The sample mean, , is a consistent estimator of #, because, by Bienaym6-Chebyshev's inequality, we can write that VAR[] 0 .2 P[[- #[< $\diamond$] _> 1 62 -- 1 - --n62 " 1 as n -+ x '$\diamond$e > 0. Note that we can also use the weak law of large numbers to show that " is a consis- tent estimator of #.

256 6 Estimation and Testing The Method of Maximum Likelihood Definition 6.1.8. Let X ..... Xn be a random sample of a population X whose prob- ability density (or mass)function f x (x; 0) depends on an unknown parameter 0. The likelihood function of the sample is L(O) = fi fx(Xk; 0). k=l (6.11) Definition 6.1.9. The maximum likelihood estimator, OM L, of the parameter 0 ts the value of O that maximizes the likelihood function. Remark. We can have: 0 = (0 ..... Om). Example 6.1.6 a) Let X ..... X. be a random sample of X Exp(X). Then, we have: L(X) : -I Xe-XX = Xne-Xnf ifXk > 0 Vk. k=l Since) maximizes LOQ if and only if) maximizes In LOQ, we can consider In LOQ = n lnX - Xn2. Finally, we find, by differentiating In LOQ (with respect to)) and setting this deriva- tive equal to zero, that XtL = 1/2. We can check that it is a maximum. Indeed, we have: d 2 n dX InLOQ=- <0. b) If X Poi(a), we have: _i e-ax e-nat n'f(L(ot) = -- n if Xk {0, 1, .. } k= Xk ! l-lk= Xk ! " = lnL(eO=-- ne+n21ne--lnXk! k=l d n2 = --lnL() =-n+-- =00 =2. da Since d 2 n2 da 2 In L(a) -- a2 (at least for n large enough), we can write that Yk

6.1 Point Estimation 257 Remarks. i) Method of moments. We set, in the case when X Exp(X), 1 E[X]=2$\diamond$>- =2. It follows that the estimator of the parameter X by the method of moments, ,M, is also given by 1/.. Similarly, when X Poi(), proceeding as above, we find at once that 4 = (= 4L). In general, we set - Xm E[X m] = k for m = 1, 2 ..... k=l (6.12) We use j (helpful) equations to estimate the j unknown parameters of the function fx(x; 0). For example, if X U[-0, 0], then we use the equation obtained with m ---- 2 (because the equation for m = 1 does not enable us to estimate 0): F/ " k=l That is, (20) 2 n 2 [-- X 2-11/2 12 k=l n k=l Actually, some authors instead use the equation VAR[X] = 3 2 "0- VAR[X] ---- (Xk -- .)2 n-- k=l to estimate the parameter 0, which yields a different estimator: k=l However, since E[X] = 0, if n is large enough, then should be approximately equal to zero, so that the two estimators will produce almost equal point estimates. ii) Let X U[0, 0]. Then,

1 L(O)=-- ifXk [O,O]�k. O n We find that OML = max{X1 ..... Xn}, while OM = 2R. For example, if we have the following particular observations: 1, 2 and 9, then OM ---- 9 and 04 = 2 � 4 = 8 (which is impossible).

258 6 Estimation and Testing 6.2 Estimation by Confidence Intervals Definition 6.2.1. An interval [LC, U C] is called a (two-sided) confidence interval at 100(1 - ot)%forO if P[LC(X1 ..... Xn) < 0 < UC(XI ..... Xn)] = 1 - a. (6.13) Remarks. i) 0 is a parameter, whereas LC and UC are random variables. ii) We call LC and UC the lower confidence and upper confidence limits, respec- tively. Moreover, 1 - is called the confidence coefficient. iii) One-sided confidence intervals with a lower bound and an upper bound are given respectively by [LC,) and (-, UC], where now P[LC < O] = P[O < UC1 = 1 - . (6.14) a) Confidence interval for/z; r known Let X, ..., Xn be a random sample ofa Gaussian population X whose mean is unknown, but whose variance r 2 is known. Then, we can write that It follows that " N(#, cr2/n) := -- N(0, 1). (6.15) where z/2 is defined by (see Fig. 6.1) � (z/2) = 1 - - = 1 - , (6.16) or Q(zc/2) 2 =' (6.17) -Z/2 Z ct/2 " Figure 6.1. Definition of the quantity z/2.

6.2 Estimation by Confidence Intervals Table 6.1. Values of zc for various values of a. 259 0.25 0.10 0.05 0.025 0.01 0.005 0.001 0.0005 0.674 1.282 1.645 1.960 2.326 2.576 3.090 3.291 That is, z, is the 100(1 - oe)th percentile of the standard Gaussian distribution (if 100(1 - oe) is an integer). We deduce from Equation (6.16) that P [f(- z,/2(cr/x/') < lz < f(+ z,/2(cr/x/')] = 1 - oe. (6.18) Therefore, the interval [f(- z,/2(cr / x/'), f(+ z,/2(cr / x/')] (6.19) is a 100(1 - 00% confidence interval for Remarks. i) One-sided 100(1 - 00% confidence intervals for/z are given by [J - z,(cr/x/'), x) and (-x, J z,(cr/x/')]. (6.20) ii) The confidence intervals given by expressions (6.19) and (6.20) are still valid (ap- proximately), even if the population X is not Gaussian, as long as n is large enough (by the central limit theorem). iii) The values of z, can be obtained using a statistical software package or found in a table. The most useful values are given in Table 6.1. Furthermore, by symmetry, we have: z- = -z. Finally, there are also formulas that give good approximations to z,. Example 6.2.1 Let xl ..... x0 be a particular random sample of a population X N(/z, cr 2 = 0.04). Suppose that � = 10.5. Then, a confidence interval for/z with confidence coefficient equal to 1 - oe = 0.95 is given by 10.5 -t- 1.960(0.2/x/q-d) == [10.38, 10.62] (approximately). Remark. We cannot write that P [/z 6 [10.38, 10.62]] _ 0.95, because /z is not a random variable. Once

the data have been collected, we calculate a deterministic interval. Then, the parameter/z is or is not in this interval. There is no probability left. b) Confidence interval for/; tr unknown If the variance cr 2 of the population X is unknown, but n is sufficiently large, then we can still use expressions (6.19) and (6.20). We only have to replace cr by its estimator S. However, if n is small (< 30), these expressions are no longer valid. In this case, X must have a Gaussian distribution. We consider T .-- S/x/-d -- {[(n - 1)(S/cr)2]/(n - 1)} 1/2' (6.21)

26O 6 Estimation and Testing 0.4 0.3 0.2 0.1 0 finity (N(0,1)) ribution) 4 -2 0 2 4 Figure 6.2. Examples of Student's tk distributions. We can show that " and S are independent random variables and that (n -- 1) $2 2 (7 2 Xn-I" (6.22) That is, (n - 1)($2/(7 2) has a chi-square distribution with n - 1 degrees offreedom. Since the numerator on the right-hand term of Equation (6.2) has a Gaussian N(0, 1) distribution, then, by definition, T has a Student's t distribution with n - I degrees of freedom. We write: T tn-1 (see Fig. 6.2). Proceeding as in Case a) above, we find that [2 -- tc/2,n-1 (S/q/'), 2 + tc/2,n-1 (S/q/)] (6.23) is a 100(1 -oe)% confidence interval for p, where tc/Z.n-1 is defined by (see Fig. 6.3) PIT < tc/Z.n-1] = I - - ifT tn-1. (6.24) Remarks. i) The formulas for the one-sided confidence intervals are easily deduced from (6.23). We obtain: [2-t,n_(S/v),ec) and (-ec, 2+t,n_(S/v)]. (6.25) fT (t) tou2, n-1 Figure 6.3. Definition of the quantity tct/2,n_ 1.

6.2 Estimation by Confidence Intervals Table 62. Values of t0.025, n and t0.05,n for various values of n. 261 n 1 2 3 4 5 6 7 8 t0.025,n 12.706 4.303 3.182 2.776 2.571 2.447 2.365 2.306 t0.05,n 6.314 2.920 2.353 2.132 2.015 1.943 1.895 1.860 n 9 10 15 20 25 30 40 x .025,n 2.262 2.228 2.131 2.086 2.060 2.042 2.021 1.960 N.05,n 1.833 1.812 1.753 1.725 1.708 1.697 1.684 1.645 ii) The values of t,,n are obtained using a statistical software package or found in a table. The values of t0.025,n and of t0.05,n for many values of n are given in Table 6.2. Furthermore, by symmetry, as in the case of the standard Gaussian distribution, we have: tl_a, n = -ta, n. iii) Student is the pseudonym of W. S. Gosset. 1 The probability density function of a Student distribution with k degrees of freedom is given by I[(k + 1)/2] 1 fT(t) = 4,i(k/2) [1 + (t2/k)] (k+l)/2 for-x < t < x. (6.26) Its mean is equal to zero (for k > 1) and its variance to k/(k - 2) (for k > 2). Moreover, the Student distribution with 1 degree of freedom is a particular case of the Cauchy 2 distribution, whose

probability density function is given by 1 fx(x)= [(-) 2]rfl 1+ for-x < x < x, (6.27) where 6 and fl > 0 are parameters. The tl distribution corresponds to the case when = 0 and fl = 1. We can show that this Cauchy distribution can also be obtained from two independent N(0, cr 2) random variables, X and Y, by defining Z = XWY. Note that the mathematical expectation of this random variable does not exist, because 1 William Sealy Gosset, 1876-1937, was born and died in England. He studied chemistry and mathematics at Oxford University. Next, he worked as a chemist for the Guinness brewery, in Ireland, where he invented a statistical test for the control of the quality of the beer. This test uses the distribution that bears his name. He published many papers on statistics. His research was motivated by practical problems. 2 (Baron) Augustin Louis Cauchy, 1789-1857, was born and died in France. He is considered the father of mathematical analysis and the inventor of the theory of functions of a complex variable. He studied at the]cole Polytechnique de Paris and at the]cole des Ponts et Chauss6es. He was professor at the]cole Polytechnique and at the Collge de France. His political ideas and his religious convictions caused him a lot of trouble. He wrote 789 scientific papers on mathematics, mechanics, optics, etc. His contributions to mathematical physics are very important. Many formulas or mathematical results bear his name.

262 6 Estimation and Testing 1 E[X] = x dx 7r(x + 1) and this improper integral does not converge. However, if we consider the Cauchy principal value of the integral, defined by lim x 2 dx, ' ,. 7r(x + 1) then we can write that E[X] = 0. Finally, given that E[X 2] = 0, the variance of X is infinite (or does not exist). Example 6.2.2 If cr is unknown in Example 6.2.1 and if the standard deviation of the sample is equal to 0.25, then the confidence interval becomes 10.5 + 2.262(0.25//') = [10.32, 10.681 (approximately). Remark. Since t.n > z for any n E { 1,2 .... }, the confidence interval obtained when the population standard deviation is unknown should generally be wider than the corresponding confidence interval calculated with cr known, which is logical. However, it can happen in practice that the estimator S of cr considerably underes- timates the true value of the population standard deviation, so that the confidence interval in the case when cr is unknown is narrower. For instance, if we had obtained a value of s equal to 0.15 rather than 0.25 above, then the confidence interval would have been [10.39, 10.61] (approximately), which is a slightly narrower interval than the one calculated in Example 6.2.1. 6.3

(Pearson's) Chi-Square Goodness-of-Fit Test Let X be a random variable whose probability density (or mass) function fx(x; O) is unknown. We want to test the null hypothesis H0: fx = fo (6.28) against the alternative hypothesis H: fx gk fo, (6.29) where f0 is a given function. The procedure, proposed by K. Pearson, 3 is the follow- ing: 3 Karl Pearson, 1857-1936, was born and died in England. He studied at Cambridge Univer- sity and then made his career in London. His most important works were on applications of statistics to biological problems. His contributions to statistics, in particular his goodness-of-fit test published in 1900, are contained in his papers on the theory of evolution. He also contributed to regression analysis. The term "standard deviation" is due to him. He and Fisher totally disagreed about the statistical techniques they were using. Moreover, he studied, in Germany, medieval German literature. He was even offered a post in this field at Cambridge University.

fx(X) 6.3 (Pearson's) Chi-Square Goodness-of-Fit Test 263 2 Figure 6.4. Definition of the quantities X2/2,n -ce 1 and X1 1" i) we divide the set Sx of possible values of X into k disjoint and exhaustive classes (or intervals); ii) we take a random sample of size n from the population X; iii) we calculate k D2 := Z (nj - m j) 2 j=l mj (6.30) where nj is the number of observations in the jth class (the observed frequency) and mj denotes the expected frequency under Ho, that is, the number of observations that the jth class should count, on average, if the hypothesis H0 is true; iv) we can show that if H0 is true and if the number n of observations is large enough, then D 2 X2_r_]. That is, D 2 has (approximately) a chi-square distribution with k - r - 1 degrees of freedom, where r is the number of unknown parameters of the function f0 that we must estimate. We reject H0 at the significance level ot if and only if where (see Fig. 6.4) 2 (6.31) 02 > Xce,k-r-1 , P[X<x,n]=l-ot ifXXn . (6.32) Remarks. i) In general, it is preferable that mj >_ 5 Yj. If this is not the case, we can combine (adjacent) classes and therefore reduce k. ii) The k intervals (or classes) are not necessarily of equal width. Likewise, it is not necessary that the corresponding probabilities be equal. iii) Before drawing the random sample, we can state that if the null hypothesis H0 is true, then the number Nj of observations that will fall in the jth class is a random variable having a binomial distribution with parameters n and p j, where pj is the

264 6 Estimation and Testing Table 6.3. Values of X,n for various values ofc

and n. n X02.025,n 2 X0.05,n X(.95,n 2 X0.975,n 1 2 3 4 5 6 7 8 9 5.02 7.38 9.35 11.14 12.83 14.45 16.01 17.53 19.02 3.84 5.99 7.81 9.49 11.07 12.59 14.07 15.51 16.92 0 + 0.10 0.35 0.71 1.15 1.64 2.17 2.73 3.33 0 + 0.05 0.22 0.48 0.83 1.24 1.69 2.18 2.70 10 15 20 25 30 40 50 100 20.48 27.49 34.17 40.65 46.98 59.34 71.42 129.56 18.31 25.00 31.41 37.65 43.77 55.76 67.50 124.34 3.94 7.26 10.85 14.61 18.49 26.51 34.76 77.93 3.25 6.27 9.59 13.12 16.79 24.43 32.36 74.22 2 X0.025,n Xg.O5,n X02.95,n X02.975,n probability that the variable X takes a value in the jth class (if Ho is true). We have: mj - npj. Furthermore, we know that the binomial distribution tends to a Gaussian distribution. Finally, the square of a standard Gaussian distribution is a chi-square distribution with one degree of freedom. iv) The values of X2,n are obtained using a statistical software package or a table. There are also formulas that give excellent approximations, in particular: 2 Xc,n - n Zc + 1 - - n (Wilson-Hilferty's approximation) (6.33) and X2 1 [Zc, + 2n/Z- 1] 2 (Fisher's approximation). (6.34) The values of X,n for various values of ot and n are given in Table 6.3, p. 264. Contrary to z, and tc.n, we cannot write that X-,n = -X,n2 (because the chi-square distribution is non-negative). Example 6.3.1 We want to test the hypothesis that the lifetime X (in months) of an electronic component has an exponential distribution. We collected the following observations: j [0, 10) [10, 20) [20, 30) [30, 40) [40, nj 35 20 18 8 19 100 Moreover, the mean of the observations is equal to 25. Thus, we want to test H0: X Exp())

6.3 (Pearson's) Chi-Square Goodness-of-Fit Test 265 against HI: X does not have an exponential distribution. Suppose that = 0.05. First, we must estimate the unknown parameter ,. We al- ready saw, in Section 6.1, p. 256, that the maximum likelihood estimator of, is given by 1/). Then, we set fx(x) = le-x/25 forx > 0. 25 Next, we find that P[a < X < b] = e -a/25 - e -1'/25 for0<a < b, whence we deduce the following table: j [0, 10) [10, 20) [20, 30) [30, 40) [40, x) E nj 35 20 18 8 19 n = 100 () pj 0.33 0.22 0.15 0.10 0.20 1 (npj =) mj 33 22 15 10 20 100 Now, given that mj > 5 Y j, we calculate D2 (35 - 33) 2 (19 - 20) 2 -- +...+ _ 1.35. 33 20 Remark. We could use d 2 to distinguish between the test statistic D 2 and the value taken by this statistic. 2 Since X0.05,5-1-1 - 7.81, we accept the exponential model, with , = 1/25, at significance level = 0.05. Remarks. i) If we wanted to test, for instance, the hypothesis that the random variable X has a Gaussian distribution, then we would have to add the interval (-, 0) (in which there are of course no observations). ii) The parameter

is actually defined by = P[Error of type I] = P[Reject H0 [H0 is true]. (6.35) Here, the notation for conditional probability given in Chapter 2 has been used. How- ever, we will never have to calculate the probability of the event {H0 is true}. It is merely a statement that we assume to be true in the calculation of the type I error riskl We also define the parameter/3 by the formula /3 = P[Error of type II] = P[Accept H0 [H0 is false]. (6.36) The quantity 1 -/3 is called the power of the test. A test generally more powerful than that of Pearson is the Kolmogorov-Smirnov test. To test normality, the Shapiro-Wilk test (for n < 50) is considered by many people to be the best.

266 6 Estimation and Testing Note that, contrary to , the value of fl is not unique in a given problem, but rather depends on the particular alternative hypothesis Hi that we assume to be true. There are actually an infinite number of particular alternative hypotheses Hi, whereas H0 is unique. For example, above if the hypothesis Hi is true, then the random variable X can have any distribution, except the exponential distribution. 6.4 Tests of Hypotheses on the Parameters I) Comparisons between a parameter and a constant a) Test of a theoretical mean/z; ◆y known Let Xl ..... X,, be a random sample of size n of a random variable X with un- known mean #, but known variance cr . We want to test against H: It -◆/zo. (6.37) Remark. A hypothesis of the form Ho:/z ----- /zo is called simple, whereas HI:/z :/zo is a multiple or composite hypothesis. Moreover, the test in (6.37) is said to be two- tailed. A right-tailed or upper- tailed (respectively left-tailed or lower-tailed) test is given by Ho: /z =/zo against Hi: /z > /zo (resp./z < /zo). (6.38) In this case, we may write the null hypothesis Ho as follows: H0:/z _</zo (resp./z >_/zo). (6.39) To perform the test in (6.37), we use the statistic Zo .- (6.40) /v' " When H0 is true, has (at least approximately if n is large enough) a Gaussian distribution with parameters/zo and a2/n. It follows that Zo N(0, 1). (6.41) We reject H0 at significance level if and only if IZol > z/z, (6.42) where z/e is such that (z/e) : 1 - a/2 (see Section 6.2, p. 258). Remarks. i) We have, indeed: P[Reject H0 I H0 is true] = P[1Z0l > z/e I/z :/z0] : 2(a/2) : a. (6.43)

6.4 Tests of Hypotheses on the Parameters 267 ii) In the case of the one-tailed tests, we reject H0 if and only if Z0>z if Hi:#>#0, (6.44) Z0 < --z if HI: # < #0. Example 6.4.1 Let xl ..... x25 be a particular random sample of a Gaussian popula- tion with variance cr 2 = 4. Suppose that 2 = 2.8. To test H0:#=2 against HI:##2, we calculate 2.8 -2 --2. zo- 2/v We then choose or, the

significance level of the test. Let ot = 0.02; then z,/2 = zo.ol -- 2.326. Since 121 _< 2.326, we do not reject the null hypothesis. On the other hand, if we choose oe = 0.05, we obtain: z,/2 = zo.o25 -- 1.960. Since 121 > 1.960, we reject the hypothesis Ho: # ---- 2 at significance level oe ---- 0.05. Remarks. i) Statistical software programs generally give the value of oe that corre- sponds to the observed statistic z0. This value is called the P-value. In the preceding example, we find that z/2 = 2 if and only if oe _ 0.0455. Thus, if we choose oe smaller than 0.0455, then we cannot reject H0. In general, it is difficult to give an exact formula for the P-value. However, in the case of the hypothesis test on a mean with cr known, it is easy to show that 2[1 - (Izol)] ifH: # #o, P = I - qb(zo) ifH: # > #o, qb (zo) if HI : # < #o. (6.45) ii) In the case of the tests (especially the one-tailed tests) on the parameters of a random variable, the null hypothesis is usually the one that we think is false, or at least that we put in doubt. Consequently, we try to reject it. For this reason, we prefer to say that we cannot reject H0, rather than saying that we accept Ho. Type H error If the hypothesis HI is true, that is, if # = #I = #0 A, where A 0, then J N(#I, cr2/n) and it follows that Therefore, (6.46) fl (= fl(A)) = P[Accept Ho I Ho is false] = P[-z,/2 <_ Zo <_ z,/2 I # = #o + A]

268 6 Estimation and Testing = P - Zc/ - -- < N(O,1) < Zc/ (6.47) Remark. In the case of the one-tailed tests, we find that � z, if HI: # > #0, � z, if H: < 0. (6.48) Example 6.4.1 (continued) If the true value of the population mean is # = #l = 2.6 = 2 0.6, and if a = 0.05, we calculate /3(0.6) _ (1.96- (0.6)(5/2)) - (-1.96- (0.6)(5/2)) Tab. 3.3 = (0.46) - (-3.46) -- 0.6772 - (1 - 0.9997) = 0.6769. Thus, the power of the test, when # = 2.6, is given by 1 -/3(0.6) " 0.3231, which is weak. If we wish to increase this power, we can take more observations. Indeed, the value of/3 decreases as n increases (see (6.47)). For instance, if we take n = 100, then we obtain that/3(0.6) _ 0.1492, so that 1 - /3(0.6) -- 0.8508. In practice, we would like 1 -/3(A) to be greater than or equal to 0.8 for a difference A that we deem significant between #0 and the true population mean. Sample size We can make use of Equation (6.47) to calculate the value ofn required to get a certain/3, given a and A. Indeed, if A > 0 we have: so that This last equation implies that (6.49) /3- (z/2 A/-ff). (6.50) z- -- z/2 -- -- (6.51) O" Since zl- = -z, we obtain the following formula: (z/ + z) n -- (6.52) A2 ' where A = # - #o.

Remarks. i) If A < 0, then 6.4 Tests of Hypotheses on the Parameters 269

$(Zc_{\alpha}/2 A/) - 1$ (6.53) and we obtain the same formula as above for n. ii) In the case of the one-tailed tests, we find the following exact result: (z, + zt)2r 2 n -- (6.54) A 2 Example 6.4.2 Suppose that we test Ho:#=2 against Hi:#5 2 with ot = 0.05. Then, if # = 2.5 and cr 2 = 1, to obtain a value of/3 smaller than or equal to 0.10, we must take (zo.o25 + zo.02 $\diamond$ 1 (1.960 + 1.282) 2 n 42. (2.5 - 2) 2 0.25 Remark. Here, we have: _ q(-1.960 - (0.5),/) -- q(-5.20) _ 0. b) Test of a theoretical mean ; a unknown Let X ..... Xn be a random sample of a random variable X having a Gaussian distribution with mean # and variance cr 2 unknown. To test Ho: # = #o, (6.55) we use the statistic To .-- -- (6.56) s / ,f' where S is the sample standard deviation. We' can write (see p. 260) that To tn-1. (6.57) We reject Ho at significance level ot if and only if ITol > tot/2,n-1 if Hi: # 5 & #o, TO > tt,n-1 if ill: # > #0, TO < -ta,n-1 if HI:# < #0. (6.58)

270 6 Estimation and Testing Remark. If the sample size is large enough (> 30), we can use the test in Case a) above; we simply replace a by its estimator S. Moreover, in this case, the assumption of a Gaussian distribution for X is not essential. Example 6.4.3 Let 5, 15, 14, 12, 4, 8, 9, 12, 10, 17, 12, 11 be a particular random sample of a Gaussian population with mean/ and variance a 2 unknown. We choose = 0.05 and we want to test Ho:z= 15 against H:z$\diamond$ 15. We first find that 2 = 10.75 and s -- 3.84. Next, we calculate 10.75 - 15 - __ - -3.83. 3.84//-i- Finally, since to.o25.11 2.201, we have: Ito] > to.o25.11 and we can reject the null hypothesis Ho at level = 0.05. Remark. Actually, we must set the hypotheses Ho and H prior to taking the random sample. Type H error and sample size By definition, in the two-tailed case, fl = P[Accept Ho I Ho is false] = P[17b] < t/2,n-I I/ =/o -t- A], (6.59) where A $\diamond$ 0. Now, if z = zo + A, we find that -- , (6.6O) To.-- S/V/- x/W2/(n_ l) where Y N(Ax/-d/a, 1) and W 2 Xn2_ are independent random variables. We say that To has a non-central t distribution with n - 1 degrees of freedom and non- centrality parameter A x/-d/a. Consequently, it is not easy to calculate the exact value of ft. There exist approximation formulas for ft. For example, we can show that (t'n--[-V/- + (t2a,n_,/2n) fl() -- qb ,,_ 1 ifH:/ >/o, (6.61) where ft -- ft 0 .-- -- (6.62)

6.4 Tests of Hypotheses on the Parameters 271 Since 0. is unknown, we must replace it by its point estimate s (if the random sample has already been taken) to calculate , or we can express the difference # - #0 in terms of 0.. There also exist curves, called operating characteristic curves, that give the value of/3 in

terms of a parameter like above for a fixed c and many values of n. More-over, we can use these curves to (approximately) calculate the size n of the random sample that we must take to obtain a given value of/3. Finally, we also find tables that give the value of n that corresponds to/3, when c and are fixed. In this book, as far as the computation of/3 or n is concerned, we will generally simply consider the case when the size of the random sample (taken or that must be taken) is large. The problem is then reduced to the preceding one where 0. is known and we can use the formulas already obtained. We must however replace 0. by s (if the data have been gathered). c) Test of a theoretical variance 2 Let X1, . . . , X n be a random sample of a Gaussian random variable X, with mean # and variance 0.2 unknown. To test the null hypothesis Ho: 0.2 = 0.o 2, we use the statistic W02 .-- (n- 1)S 2 0.02 (6.63) When Ho is true, W02 has a chi-square distribution with n - 1 degrees of freedom. We reject Ho at significance level a if and only if 2 2 if HI: 0 .2 0.0 2 , Wg > Xot/Z,n_ 1 or W02 < Xl_(ot/Z),n_ 1 Wd2> 2 Xc,n_ 1 if Hl: 0.e > 0.02, (6.64) W,2 < . 2 if nl: 0.2 < 0.02. 0 ," 1 -or,n- 1 Remark. As in the case of the test of a theoretical mean # with 0. known, we could consider the problem of testing H0:0.2 = 0.g when this time (we assume that) the population mean is known. However, in practice this case is much less important. Example 6.4A (See [4].) Let 10.2, 9.9, 9.8, 10.1, 10.1, 9.9, 10.0, 9.7, 10.1, 10.3 be a particular random sample of a random variable X that represents the weight of soap boxes. We assume that X has (approximately) a Gaussian N(#, 0.2) distribution and we want to test Ho: 0.2 = 0.04 against H1 : 0.2 > 0.04 at level c = 0.05. We find that 2 = 10.01 and Z]◆l x/2 = 1002.31. It follows that

272 6 Estimation and Testing 10 s2 = (xi - .f)2 10- 1 i=1 ---- x/2-10◆2 " 9 ' so that 0.31 w " = 7.75. 0.04 Finally, since 2 Xo.o5,9 " 16.92, we cannot reject the null hypothesis H0 at significance level a = 0.05. Remark. Here, since s 2 " 0.034 is smaller than 0.04, we certainly could not reject Ho with a small a. Type II error and sample size Suppose that the alternative hypothesis is H: r 2 < ro2 and that the true value of r 2 is r 2 < %2. We have: = p[(n-l) $2 2 o? - l- (.n-I O" = PIx.-2 >_)Xl2_.n_l], (6.65) where 0-2 ◆ 0 (6.66) Similarly, we find that 2 pIz2_ <_ ZZ,,._] ifH: r 2 > r02, Z 2 72 PIZz-(a/2).,-I < Xn2-1 < Xa/2.n-ll if HI: # a02. (6.67) Now, if n is large, we may write that Xn 2 N(n, 2n). (6.68) Hence, we can approximately calculate fl(Z) with the help of the preceding formulas. To determine the size of the random sample needed to obtain a fl smaller than or equal to a given value, we can use the following

formulas: 3 1 Io-0z + o'IZ/5]2 (one-tailed cases), 3 _r,oz,/+,,zl(two_tailedcase), +2L -o J (6.69) where r 2 is the true variance of the population�

6.4 Tests of Hypotheses on the Parameters 273 Example 6.4.4 (continued) Suppose that we gathered 51 observations rather than 10 in the preceding example. If 0.i = 0.06, then ;k = 2/3 and we calculate 8(, = 2/3) = P[;(52o < (2/3);(o2.o5,5ol Tab. 3.3 Tab/.,,6.3_ P[N(50, 100) _< 45] = qb(-0.5) _ 0.31. If we want to be able to detect, with a probability of at least 0.90, that 0.2 is greater than 0.04, when the population variance is in fact equal to 0.06, we must take 3 ! F0.2 Z0.05 "-b Z0.I0] 2 n -- + 2 L ,/g-hZ - 0.2 J - 103.83. Thus, we must take at least 104 observations. Remark. We can show that if n is large enough, then we may write that S N(0.,). (6.70) It follows that S - 0.o N(0, 1). (6.71) z0 .- 0.0/,/ - Therefore, when n is large, we can also reject Ho if and only if]Zo] > zooC if HI: 0 .2 5 0.g, Zo > z if HI: 0.2 > 0.o 2, (6.72) Z0 < -z, if HI : 0.2 < 0.g. II) Comparisons between two parameters a) Test of two theoretical means; variances known Let X1 ..... Xm be a random sample of a random variable X with unknown mean/zx and known variance 0.x 2 , and Y1 ..... Yn be a random sample of a random variable Y with unknown mean /zr and known variance 0.r 2. We assume that the random variables Xi and Yj are independent �i, j. To test H0:/zx -/zr = A, (6.73) we use the statistic z� - v/- + Since J and I ? are independent, we have:)-I?--N(/zX-/ZY'0.A+ -)m (6.74) (6.75)

274 6 Estimation and Testing (at least approximately if m and n are large enough). It follows that Zo N(0, 1). (6.76) We reject H0 at significance level oe if and only if I IZol > z,/2 if H,:/zx -/zr -� A, / Zo > z, ifH:/zx /zr > A. (6.77) Remark. In the case of the tests comparing two parameters, we can limit ourselves to the two alternative hypotheses above, since we want to determine whether the two parameters differ by a constant A (H0) or not (H), or whether the difference between one parameter and the other is smaller than or equal to A (Ho) or greater than A (H). Example 6.4.5 Let x ..... xm be a particular random sample of X N(/zx, 2) and Y ..... Y5 be a particular random sample of Y N(/zr, 3). We assume that the samples are independent and we want to test Ho: /zx =/zy against H: /zy > /zx. We simply have to take A = 0 and reverse the role of X and Y in the above formulas. We choose = 0.05. Suppose that 2 = 4 and . = 5. Then we calculate 5-4 1.58. zo- 2-- 3 nt- 0 Since zo.o5 - 1.645, we

cannot reject the null hypothesis $H_0: \mu_X = \mu_Y$ at level = 0.05. Remark. If we had obtained $z_0 = 1.68$, for instance, then we could have rejected $H0$ and accepted $H: \mu_X < \mu_Y$. However, we could not reject H_0 with $z_0 = 1.68$ if the alternative hypothesis were rather $H: \mu_X \neq \mu_Y$, since $|1.68| = 1.68$ is not greater than $z_{0.025} = 1.960$. Type II error and sample size Proceeding as in the case of the test of a theoretical mean with σ known, we find that $\beta(\delta) = \{ \Phi(z_{c/2} - \delta) - \Phi(-z_c/-\delta)\, \text{if}\, H: \mu_X - \mu_Y > A'\Phi(z_c - \delta)\, \text{if}\, H: \mu_X \; r \; A$, (6.78) where $X - Y - t \; \delta$.-- (6.79)

6.4 Tests of Hypotheses on the Parameters 275 Similarly, we find that to obtain a given β, we must draw random samples of size $(z/2 + + m = n$ " (6.80) $(\#X -- \#r -- A)$ 2 when the alternative hypothesis is $Hi: \#x -- gr \neq A$. In the case of the one-tailed test, we replace /2 by in the preceding foula, and the result is now exact, rather than approximate. Remark. In theory, it is not necessary to take $m = n (= c)$. We can choose m, for instance, and then e value of n that coesponds to a given is $2c - m$. We can proceed in that manner, in paicular, when it is more expensive to draw a random sample from one random variable than from the other. b) Test of two theoretical means; variances unknown Let $X_1 ..... X_m$ and $Y_1 ..... Y_n$ be random samples of a random variable X and of a random variable Y, respectively. We assume that all the parameters are unknown and that X and Y have independent Gaussian distributions. We want to test $H0: \#x - \#" = A$. (6.81) Case 1: $\sigma_X^2 = \sigma_Y^2$ We first assume that the variances are unknown, but equal. In practice, we per- form a test of equality of variances (which will be seen further on) to make sure that this hypothesis is reasonable. We define 2 $(m -- 1)S2x + (n - 1)S2y$ (6.82) $Sp = m+n-2$ ' where s_X^2 and s_Y^2 are the sample variances, and we consider the statistic T_0 .-- (6.83) $S \; pm- +$ We may write that $H_0 \; T_0 \; t_{m+n-2}$. It follows that we reject $H0$ at significance level c if and only if $|T0| > t_{/2,m+n-2}$ if $H: \#x -- \#" \; A$, $T_0 > t_{,m+n-2}$ if $H1 \#x \#, > A$. (6.84) (6.85)

276 6 Estimation and Testing Case 2: $\sigma_X \neq \sigma_Y$ If we cannot assume that $\sigma_X = \sigma_Y$, then we must use the statistic 2-- T_0^* .-- , (6.86) x which, when H_0 is true, has approximately a t distribution with v degrees of freedom, where $(a + b)^2 \; v$.-- (6.87) a 2 b 2 -"S- 1 -'1- n- 1 with $S_2 \; S_2 \neq x$ and b " $Y \; a$.= -- .= --. to.l m n We then replace T_0 by $T0^*$ and $m + n - 2$ by v in the preceding case. Remarks i) We also find in several books the formula $(a + b)^2 \; v -- 2$. (6.89) a 2 b 2 m+l + n+l ii) We can show that $\min\{m - 1, n-1\} < v < m+n-2$. iii) When the sample sizes are large enough, we can use the standard Gaussian dis-

tribution as an approximate distribution (under H0) for 7b or To*. c) Test of two means with paired observations Suppose that we gather pairs (X, Y) ..... (Xn, Yn) of observations of the ran- dom variables X N(/x, ax 2) and Y N(/y, ay2), and that the conditions under which the observations are taken vary (or may vary) from one pair to the other. In this case, the observations Xk and Yk are not independent. To test Ho:/x -/ Y = A, we define the variable D = X - Y and we calculate the differences D := X - Y for k = 1 ..... n. The problem of testing H0 is then reduced to that of testing a theoretical mean/) with a) unknown, and we can use the formulas in Case I) b), p. 269. Example 6.4.6 (See [11].) In a test of hardness of materials, a steel ball is pressed into each material and the diameter of the indentation is measured. This diameter is related to the hardness. We have two types of steel balls, which we use with ten different materials. We want to determine whether the two types of steel produce equivalent results. We collected the following data:

6.4 Tests of Hypotheses on the Parameters 277 Material 1 2 3 4 5 6 7 8 9 10 Steell 75 46 57 43 58 32 61 56 34 65 Steel2 52 41 43 47 32 49 52 44 57 60 Let X be the indentation diameter obtained with steel 1 and Y be the indentation diameter with steel 2. We define D = X - Y and we want to test H0: /ZO = 0 against H1: /ZO - 0. Let = 0.05. We calculate the differences dk for k = 1 ..... 10: Material 1 2 3 4 5 6 7 8 9 10 dk 23 5 14 --4 26 --17 9 12 --23 5 We find that d = 5 and SD 15.85. We calculate next to- so/v'- 15.8 - 1. Since t0.025,10-1 " 2.262, we do not reject Ho at level oe = 0.05. d) Test of the equality of two variances Let X N(/zx, ax 2) and Y N(/zr, a2) be two independent random variables, and let S and S, be the variances of a random sample of size m of X and of a random sample of size n of Y, respectively. We assume that all the parameters are unknown. To test the null hypothesis H0: ox 2 = o 2, (6.90) we use the statistic S 2 . x F0 .= -- (6.91) We can show that if H0 is true, then F0 has a (Fisher) 4 F distribution with m - 1 and n - 1 degrees of freedom. We write: Fo Fm-l,n-. We reject H0 at significance level a if and only if 4 Sir Ronald Fisher, 1890-1962, was born in England and died in Australia. He studied as- tronomy at Cambridge University, but he worked as a biologist in a research center, where he made contributions to genetics and statistics. He is considered a founder of the mod- ern theory of statistics because of his numerous contributions to this field. In particular, he invented the analysis of variance and the concept of maximum likelihood, in addition to giving the exact sampling distributions of many statistics. Like Gosset's

methods, the sta- tistical methods that Fisher developed are valid for small samples, whereas Pearson worked with large samples. From 1933, he became a professor, first in London and then in Cam- bridge.

278 6 Estimation and Testing Table 6.4. Values of Fo.o25,n, n and Fo.os,n,n for several values of n. n 1 2 3 4 5 6 7 8 Fo.o25,n ,n 647.8 39.00 15.44 9.60 7.15 5.82 4.99 4.43 Fo.o5,n,n 161.4 19.00 9.28 6.39 5.05 4.28 3.79 3.44 n 9 10 15 20 30 60 120 Fo.o25,n, n 4.03 3.72 2.86 2.46 2.07 1.67 1.43 1.00 F0.05, n, n 3.18 2.98 2.40 2.12 1.84 1.53 1.35 1.00 {F0> Fe/2,m-l,n-I or Fo < F (e/2),m-l,n-l if Hl: tTi t7 , FO > Fc,,,,- ,n-l if H: o" > �'r" (6.92) Remarks. i) The critical values F,m-,n- are defined in the same manner as the other critical values (z,, t,,,,, and 2 (6.93) ii) We find For,m-I,n-I with the help of a statistical software package or in a statistical table. Several values of Fo.0z.s.m,n and Fo.o.s,m,,are given in Table 6.4 278, in the case when m = n. The values of Fo.oz.s,n.,, and Fo.05.,,, for n and n2 E {1,2 ..... 12} can be found in Appendix B, p. 342. Moreover, we have the following relationship: 1 Fl-,nl,n2 -- -- (6.94) Fe,n2,n t Finally, for values of n and n2 large enough, we can use the approximation formula Fo,n,2 -- exp - -- + z, + . (6.95) nl iii) As in the case of the test of a theoretical variance 0 "2, an approximate test, valid when the sample sizes are large, can be constructed based on the Gaussian distribu- tion. Indeed, if m and n are large enough, then we may write that Sx N 0.x, 2mJ and Sy N 0.,, , (6.96) so that Sx - Sy Ho Zo := N(0, 1), (6.97)

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 279 2 is defined in (6.82). We then reject H0: ax 2 = ar 2 at significance level cz if where Sp and only if [Z0[> ;a/2 if Hi: ax 2 :flay 2, Z0 > z if Hi ax 2 > ar 2. (6.98) iv) The probability density function of a random variable having a Fisher distribution with rn and n degrees of freedom is given by F[(m + n)/2]mm/2n n/2 X (m/2)-t fx(x) = F(m/2)F(n/2) (mx -1- n) (m+n)/2 for x > 0. (6.99) We find that n 2n2(m + n - 2) E[X]-- ifn >2 and VAR[X]= ifn >4. n - 2 m(n - 2)2(n - 4) (6.100) Example 6.4.7 We consider two manufacturing processes for bottles. Let X N(/xx, ax 2) be the capacity of the bottles made with the manufacturing process cur- rently in use, and let Y N(txr, 2 r) be the capacity of the bottles made with a new manufacturing process. All the parameters are unknown, but we think that/xx We want to test at level c = 0.05 Ho:a2x=a2r against Hl:a2x >a2r. We take at random 20 bottles made with the current process and 25 fabricated with the new process. We find that s2x _

0.0144 and s2g -- 0.0064. We calculate f0 := -- s2x 0.0144 = 2.25. s2}, -- 0.0064 Now, we find in a statistical table that F0.05,19,24 " 2.04 (2.02 according to (6.95)). Consequently, we can reject H0 at level c = 0.05. 6 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises Solved Exercises Exercise no. 1 (6.1) Let 1 x'O fx(x;O) = -xe- / forx > 0, where 0 > 0 is an unknown parameter. We have: E[X] = 20 and VAR[X] = 202. a) Find the maximum likelihood estimator of 0. b) Calculate the mean square error of the estimator OML obtained in a). c) What is the approximate distribution of OML? Justify your answer.

280 6 Estimation and Testing Solution a) We calculate L(O) -I-2Xke x/� 0-2n 12Ixk ex p Xk/O k=l k=l k=l == lnL (0) = -2nln0 + lnX: - k=l k=l odd ln L (O) = 02n 0 '- -- Xk21"1 k=l k=l 2 2 In L (0) = 2n 2 n Remark. Note that -- - Zk=l Xk < 0 0 < 2. We have indeed: 0ML = �2 < 2. b) E[OMLI a= E[2/21 = �E[X] = = 0 MSE[OMLI = VAR [OML] = VAR[2/2]- VAR[X] 202 02 4n 4n 2n c) By the central limit theorem, we can write that 2 has approximately a Gaussian - b) 02 N(E IX], �VAR [X]) distribution. Then, OML �X . N(0, N). Exercise no. 2 (6.3) An engineer responsible for the control of the quality in a company that man- ufactures modems picks, at the end of each workday, two modems at random (and without replacement) among those manufactured on that day. Let X be the number of defective modems among the two examined. We suppose that X B(n = 2, p). a) We can show that the maximum likelihood estimator of p, based on m observa- tions, is given by PML = 2/2, where 2 is the mean of a random sample (of size m) of X. Calculate the mean square error of PML. b) We gathered the following data over a 30-day period: j 0 1 2 nj 20 8 2 where j is the number of defective modems in the sample of size 2 and nj is the number of days on which we observed j defectives in the sample. Test the hypothesis that X B (n = 2, p = 1/5). Use oe = 0.05. 2 2 Hint. We have: X0.05,1 - 3.84, X0.05,2 - 5.99 and 2 X0.o5,3 - 7.81.

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 281 Solution a) We first calculate 1 - 1 2p E [pMI.] =- E[J/2] ----- - E[X] ----- -E [X] -- 2 -- p It follows that Bias[pML] = 0, and then MSE [PMI.] = VAR [PMI.] =-- VAR[J/2] -- 1 VAR [X] p (1 - p) 4 rn 2m b) Since P0 = (4/5) 2 = 16/25 and P2 = (1/5) 2 = 1/25, we can construct the following table: j 0 1 2 nj 20 8 2 pj 16/25 8/25 1/25 30pj = mj 19.2 9.6 1.2 30 1 30 Given that m2 = 1.2 < 5, we combine the last two classes. We obtain: j 0 1 or 2 nj 20 10 30 pj

16/25 9/25 1 m] 19.2 10.8 30 Next, we calculate D2 (20- 19.2) 2 (10- 10.8) 2 -- + -- 0.0926. 19.2 10.8 2 Since D 2 _ 0.0926 is smaller than Xo.o5,2-o-1 - 3.84, we can accept the model X B(n = 2, p = 1/5), with = 0.05. Exercise no. 3 (6.3) Let X be the number of flaws in the paint of new cars. We suppose that X Poi00. We constructed the following table with the help of a random sample of 100 CarS Number of flaws 0 1 2 3 Number of cars 40 36 20 4 a) Calculate the maximum likelihood estimator of the parameter), based on the above data. b) Test the fit of the model X Poi0 = 1) to the data. Use = 0.05. Hint. We have: 2 2 X0.05, 2 " 5.99 and Xo.o5,3 - 7.81.

282 6 Estimation and Testing Solution a) We have (see Section 6. l, p. 256): Next, we calculate Y=(O�40+I �36+2�20+3�4)-- 1 = 0.88. 100 Thus, ML : 0.88. b) We have: X H Poi () = 1) = Pk := PX (k) = e-I F Yk > 0. We then complete the table as follows: k 0 1 2 3 + nk 40 36 20 4 100 = n (--) p 0.368 0.368 0.184 0.080 1 (100p =) m 36.8 36.8 18.4 8 100 Since m > 5 Yk, next we calculate D2 (40 - 36.8) 2 (36 - 36.8) 2 (20 - 18.4) 2 (4 - 8) 2 _ 36.8 36.8 18.4 8 _ 2.44. 2 " 7.81 we accept the proposed model at significance Given that D 2 < X0.o5,4_0_i , level o = 0.05. Exercise no. 4 (6.4) We believe that the courier service of company A is faster than that of company B. Independent observations of the delivery times (in days) of the two courier ser- vices yielded: nA : 5, 2A = 1.2; n,= 10, 2,= 1.5. Moreover, we assume that the standard deviations of the delivery times are o A = O" B = 0.2. a) What additional assumptions must we make to be able to perform the comparison? Set also the hypotheses H0 and H. b) What is the smallest value of ot for which we can reject H0 if z0 = -2.74?

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 283 Solution a) Let XA (respectively XB) be the delivery time with company A (resp. B). We assume that XA and XB have a Gaussian distribution, so that XA N(/ZA, aA 2 = 0.04) and XB N(/ZB, aB 2 = 0.04). Furthermore, we assume that XA and XB are independent random variables. Finally, we set: n0: 13, A --- 13, B and H1: b) We reject H0 if and only if z0 < -zc �" -2.74 < -zc. We have: Tab. 3.3 (2.74) 0.9969. Therefore, we can write that Otmi n ' (1 -- 0.9969) + = 0.0031 + Exercise no. 5 (6.4) A manufacturer of synthetic rubber asserts that the average hardness of its rubber is equal to 64.3 degrees Shore. Previous experiments show that the standard devia- tion of the hardness is equal to two degrees Shore. We believe that the manufacturer's

assertion either underestimates or overestimates the average rubber hardness. There- fore, we perform a test of hypotheses. If the true mean (/) is indeed 64.3 degrees, the probability of reaching this conclusion must be equal to 0.95. Moreover, if the difference between/ and 64.3 is -4- 1 degree, the test performed should conclude, with a probability of at least 0.9, that/z is not equal to 64.3. We assume that the rubber hardness has (approximately) a Gaussian distribution. a) How many observations must we take? b) If $2 = 65$ and n is the value computed in a), what is the conclusion of the test? c) Calculate the probability of concluding that the average rubber hardness is equal to 64.3, if the true average hardness is 65 degrees Shore and if we use the sample size computed in a). d) How many observations must we draw if we want to reject H0 with a type I error risk of 0.05 when $2 = 65$?

284 6 Estimation and Testing Solution a) We have: ot = 0.05 and fl (l# - 64.31 = 1) = 1 - 0.9 = 0.1. Therefore, we must take (Z0.025 q'- Z0.10) 2 (2) 2 Tab. 6.1 n "" (1.960 + 1.282) 2 4 (# _ 64.3)2 -- - "" 42.04. Thus, we must draw n = 43 (or 42) observations. b) We want to test Ho: # = 64.3 against Hi: #-� 64.3 at significance level a = 0.05. We reject Ho if and only if 2 2 < Cl := 64.3 - zo.o25 -- 64.3 - 0.6 = 63.7 /43 or 2 2 > C2 : 64.3 + 1.960 "" 64.3 + 0.6 = 64.9. Since $2 = 65 > 64.9$, we reject the null hypothesis Ho at level ot = 0.05. Remark. lf2 = 65 with n = 42 (rather than 43), then C2 "" 64.9 still and the conclu- sion is the same. c) We want to calculate/4 (/z = 65). We can write that 19) 65] fl (= 65) P [63.7 2 64.9 I = =P[63.7N(65,)564.9] P [-4.26 5 N(0,1) 5 -0.33] Tab. 3.3 P IN(0, 1) 5 -0.33] 0.37. Remark. We could have used Equation (6.47) (p. 268) instead. d) We seek n (minimal) such that 2 65 > 64.3 + z0.025- 1: q/ > 5.6 � n > 31.3. Thus, we must take at least 32 observations. Exercise no. 6 (6.4) We study the thermal efficiency X (in %) of Diesel engines made by a major car maker. We assume that X has (approximately) a Gaussian distribution with standard deviation equal to 2. Tests performed on 25 engines yielded the following results: $2 = 31.4$ and s = 1.6.

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 285 a) Test the hypothesis H0:/z = 32.3 against Hi:/z -fi 32.3. Use a -- 0.05. b) Compute the probability of making a type II error in a) if the average thermal efficiency is in fact 31.3%. c) Based on the data collected, can we conclude that the standard deviation of X is actually smaller than 2%7 Perform a test of hypotheses at significance level a = 0.01. Hint. We have: 2

X0.99,24 10.86. Solution a) We want to test Ho:/z = 32.3 against HI: /z �
32.3 at significance level a = 0.05. We reject Ho if and only if 2 � < C1 :=
32.3 - zo.o25 -- 31.516 V25 or 2 - > C2 := 32.3 + 1.960 -- 33.084. Since � =
31.4 < 31.516, we reject Ho at level oe = 0.05. b) We seek /3 (/z = 31.3) _ e
[31.516 _< 2 _< 33.084 I/z = 31.3] _P[31.516<_N(31.3, zAS) <33.084]= P
[0.54 _< N(0,1) < 4.46] Tab. 3.3 _ 1 - 0.7054 = 0.2946. c) We want to test Ho:
o -2 = 4 against at level ot = 0.01. We reject Ho if and only if Since Hi:o- 2 <4
(25- 1)S 2 < 4 1-O.Ol,25-1 - 10.86. S 2 = (1.6) 2 = 2.56 :=> 6s 2 = 15.36 _>
10.86, we cannot reject Ho at level ot = 0.01.

286 6 Estimation and Testing Unsolved Problems Problem no. 1 Let X ..... Xn
be a random sample of the random variable X whose probability density
function is fx (x) = 202. if 0 < x < 20 (= 0 elsewhere). We have: VAR[X] =
202/9. a) We propose the following estimator of the parameter 0: T = ,'.
Calculate the mean square error of T. b) Calculate the maximum likelihood
estimator of 0. Problem no. 2 Let 1.1,1.3, 1.9, 2.1,2.5 and 3.4 be a particular
random sample of X N(/, 0"2). Calculate a) a one-sided 95% confidence
interval with an upper bound for b) a one-sided 90% confidence interval with
a lower bound for/, if we suppose that 0 .2:0.5. Problem no. 3 We collected the
following observations of a random variable X: j 1 2 3 4 nj 24 16 6 4 Test the
hypothesis H0: X Geom(p = I/2). Use a = 0.05. Problem no. 4 Let X ..... Xn be
a random sample of size n of a discrete random variable X whose probability
mass function is given by px (k; O) = O (l -O) k- fork= 1,2 ..... where 0 6 (0, 1)
is an unknown parameter. a) Find the maximum likelihood estimator of the
parameter 0. b) We collected the following observations of the random
variable X: 1 2 3 20 15 15 Test the hypothesis Ho:Px(k)=(1/2) k fork= 1,2 ....
Use a = 0. 1 0. 2 2.706, ,g.0,2 - 4.605 and 2 Hint. We have: X0.0, - ,0.10,3 -
6.251. against H" px (k) 7 (1/2) :.

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary
Exercises 287 Problem no. 5 Let X1 ..... Xn be a random sample of size n of a
random variable X having a Gaussian N(0, 4) distribution, where 0 is an
unknown parameter. a) Suppose that n : 25 and I Xk = 2.5. Find a constant c
such that the interval (-x, c] is a 95% confidence interval for 0. b) We propose
the following two estimators of the unknown parameter 0: T1 = --Xk and T2 :
d Xk, n k=I k=I where d is a positive constant. i) Calculate the mean square
error of T2. ii) Find for what values of d the estimator T2 is relatively more

efficient than TI, ifn=4and0= 1. Problem no. 6 Let [ifx 1, fx (x; 0) = 0 elsewhere, where 0 > 0. a) Find 0ML, that is, the maximum likelihood estimator of the parameter b) Calculate the mean square error of Hint. We have: E IX] = 0 + 1 and VAR IX] = 0 2. c) Let T := c) be an estimator of 0, where X* is the mean of a random sample of size n of X. Find the value of the constant c that minimizes the mean square error of T. Problem no. 7 Let X1 ..... Xn be a random sample of size n of a random variable X whose probability density function is 1 -x/O fx(x;O)=Re - forx_>0. We can show that the maximum likelihood estimator of the unknown parameter 0 (> 0) is given by OMC = 2. a) Calculate the mean square error of OMC. b) A particular random sample of size n = 80 of X has been drawn. The mean and the variance of the sample are equal to 2 and 5, respectively. Calculate an approximate 95% confidence interval for the parameter 0. c) Give an approximate 95% confidence interval for 0, if in b) the sample variance is unknown. Justify your answer.

288 6 Estimation and Testing Problem no. 8 Let X be a random variable of continuous type whose probability density func- tion is given by l.-x3e -x/ ifx > 0, fx (x;/) = 0 ifx _< 0, where/ > 0 is a parameter. a) Given a random sample of size n, X ..... X,,, find the maximum likelihood esti- mator of/. b) Consider a random sample of size four (n = 4): X, X2, X3, X4, and the following two estimators for/: i) the estimator obtained in a) (with n = 4); ii)/ := 0(X + 2X2 + 3X3 + 4X4). Compare these two estimators with regard to their bias and their mean square error, and determine which one is preferable. Hint. For the random variable X considered, we have: E [X] = 4/ and VAR [X] = 4/ 2. Problem no. 9 Let (1 - 0)/3 ifx = 0, px(x; 0) = 1/3 ifx = 1, (1 + 0)/3 ifx = 2, where - 1 < 0 < 1 is an unknown parameter. A random sample of size n = 30 of X has enabled us to construct the following table: i 0 1 2 ni 10 12 8 a) We propose the following estimator of the parameter 0: := (" - 1). Calculate the mean square error of . Hint. We have: E[X] = 1 + 20 and VAR[X] = - -0 . b) Test the hypothesis 1 H0: px(x; O) = px(x; 0 = O) = -z forx = 0, l, 2. Use ot = 0.05. Problem no. 10 A particular random sample of nine cigarettes of a certain brand contains an av- erage of 4.2 mg of nicotine per cigarette. The cigarette maker claims that the content X of nicotine in its cigarettes is not, on average, greater than 3.5 mg. We assume that X N(#; 1.96). a) Can we put the cigarette maker's assertion in doubt? Use t = 0.05.

6.5 Exemises, Problems, and Multiple Choice Questions, Supplementary Exercises 289 b) What is the probability of concluding, in a), that the cigarette maker is wrong if # = 3.8? c) What is the smallest number of cigarettes that we should examine if we want the probability computed in b) to be greater than or equal to 90%? d) To better protect the consumer, is it preferable to perform the test in a) with oe = 0.05 or oe = 0.01 ? Justify your answer. Problem no. 11 We want to compare the compressive forces (in kilograms per square centimeter) of two types of concrete. We want to show that the second type of concrete has a greater compressive force, on average, than the first one. Measurements made with two specimens are given below: Concrete 1 295 319 304 302 Concrete2 318 316 312 318 a) Define the variables and specify the necessary basic assumptions. b) Compare, at significance level ot = 0.05, the average compressive forces, assuming that o'12 o'. c) Show, using a test of hypotheses, that the variances are actually not significantly different at level ot = 0.05. d) Perform again the test of comparison of the population means in b) with o'12 = o'22. Problem no. 12 Let X -- N(#, o'2). A particular random sample of size n = 25 of X yielded $\diamond$ = 0.2 and s = 0.5. a) We want to test H0: # = 0 against Hi: # $\diamond$ 0 at significance level ot = 0.05. Calculate the value of the statistic used to perform the test. b) We also want to test H0:o'2 = 0.2 against Hi: o'2 > 0.2 at significance level ot = 0.05. What is the value of the percentile used to make the decision (that is, the critical value)? Problem no. 13 We are interested in the lifetime X (in tens of thousands of kilometers) of tires of a certain brand. We wish to test the hypothesis H0: #x = 50 against HI: #x < 50 at significance level ot = 0.05. We assume that X -- N(#x, 25) (approximately). a) What is the value of/3 if n = 9 and #x = 45? b) What is the smallest value ofn for which/3 < 0.10 if#x = 45? Problem no. 14 Let X -- N(#, o'12) and X2 " N(#2, o'22) be two independent random variables. Particular random samples of X and X2 yielded: n = 9, $\diamond$1 = 5, sl = 2; n 2 = 10, -2 = 3, $2 = 1.

290 6 Estimation and Testing a) Is the variance r 2 significantly greater than r227 Use ot = 0.05. b) If we assume that the theoretical variances are equal, can we then state that the theoretical means are significantly different? Use t = 0.05. Problem no. 15 We assume that all the students taking a certain course are equally gifted. We also assume that the mark obtained by a student on the final exam has (approximately) a Gaussian distribution. The results obtained on this exam, by the students of two groups with different teachers, are the following: Group n 2 s 1 49 10.5 4.1 2 35 9.3 3.8 Let Xi be the mark that a

student taking the course with teacher i will have on the final exam, for $i = 1, 2$. We assume that $X_1 \sim N(\#_1, r_1^2)$ and $X_2 \sim N(\#_2, r_{22})$ are independent random variables. a) Can we conclude that the difference between the variations in the marks is signif- icant if we use a significance level ot $= 0.10$? Specify the null hypothesis Ho and the alternative hypothesis H. Hint. Use the fact that $F_{0.05,8,3} > $ b) Can we postulate that the teacher has a significant effect on the results of his students? Use $t = 0.05$ and specify H0 and H. c) What is the smallest number of observations that we need to be able to detect, with a probability of at least 90%, an actual one-mark difference between the means and #2? d) What is, approximately, the value of/4 for the test performed in b) if# - #2 ---- 2? Problem no. 16 The ohmic resistance of a certain electronic component must be equal, on aver- age, to 400 ohms. A random sample of 16 components, drawn from a large batch of components, yielded the following observations: 392 396 386 389 388 387 403 397 401 391 400 402 394 406 406 400 We assume that the ohmic resistance has approximately a Gaussian distribution. a) Can we affirm, at significance level $t = 0.05$, that the batch meets the norm of 400 ohms? b) Calculate the probability of making a type II error with the test performed in a) (that is, with the rejection constants for 2 computed in a)) if the ohmic resistance has actually a Gaussian $N(405, 49)$ distribution. c) Under the same assumptions as in b), how many observations must we take if the probability of making a type II error must be smaller than or equal to 0.05?

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 291 Problem no. 17 We use two identical machines to manufacture a certain part. We wish to deter- mine whether the two machines have the same variability with respect to an impor- tant characteristic of this part. Particular random samples, drawn from the production of each machine, yielded the following results: Machine A 140 135 140 138 135 138 140 Machine B 135 138 136 140 138 135 139 We have: $X_A = 138$, S_A -- 2.24, $\diamondsuit_B = 137.29$ and s_B -- 1.98. a) Specify the assumptions that we must make to be able to perform the test. b) Perform the appropriate statistical test, using a significance level equal to 0.10. What is the conclusion? c) Test the hypothesis that the 14 observations collected come from a uniform dis- tribution over the interval [135, 140]. Use the significance level $a = 0.05$ and the following intervals: [135, 137.5) and [137.5, 140]. Problem no. 18 We are interested in the breaking strength of metal rods made by two companies. We have a batch of

rods manufactured by each company. We draw at random ten rods from each batch and we measure the strength needed to break each one of them. We obtain the following results: Batch 1 53.2 53.9 53.1 50.9 42.2 52.9 55.8 41.9 50.0 49.0 Batch2 55.0 64.5 53.0 57.8 56.1 58.0 55.8 50.8 54.0 52.2 We assume that the breaking strength of the metal rods has approximately a Gaussian distribution. a) Test the hypothesis that the average breaking strength is the same in the two batches, at significance level a = 0.05. Hint. We have: to.o25,18 " 2.101. b) Test, assuming that the 20 rods actually come from the same company, the hy- pothesis that i) the average breaking strength of a metal rod is equal to 55 and ii) the variance of the breaking strength of a metal rod is 20. Use = 0.05. Hint. We have: t0.025,19 " 2.093, Xo2.o25,19 _ 32.85 and X0.975 ,2 19 -- 8.91. Problem no. 19 Measurements of percentages of elongation have been made on ten steel parts. Five of these parts were treated with method A (aluminum only) and the other five with method B (aluminum plus calcium), yielding the following results: MethodA(%) 28 29 31 33 30 MethodB(%) 34 27 30 36 33 a) Test the equality of variances. Use a = 0.10. b) Can we conclude that the two methods give, on average, the same results? Use = 0.05.

292 6 Estimation and Testing Multiple Choice Questions Question no. 1 Let X ..... Xn be a random sample of a random variable X having a Bernoulli distribution with parameter p. A) Find the maximum likelihood estimator PML of the parameter p. a) I/(n') b) n" c) 1/" d) /n e) B) Give the exact and the approximate distribution of n'. a) B(I, p) and Poi(np) b) B(I, p) and N(np, npq) c) B(I, p) and N(0, I) d) B(n, p) and N(p, pq/n) e) B(n, p) and N(np, npq) C) Give a formula for an approximate 100(I - 00% confidence interval for p. a) p+zc/2[P-2] I/2 b)" +Zc/2[] I/2 c)" +za/212'12)] I/2 d) p + za/2 P(In-P) e)) + za/2 i��-Yr)n Question no. 2 Let Xl ..... Xn be a random sample of a random variable X having a uniform distribution on the interval [0, 0]. We propose the following estimator of the un- known parameter 0: = 2'. Calculate the mean square error of . Hint. We have: VARIXI = 02/12. 02 02 02 a) 0 b) c) T d) - e) �- Question no. 3 Let X be a random variable having a Gaussian distribution with unknown pa- rameters. A particular random sample xi ..... x25 of X has yielded the following results: 25 25 2 1550. z-,xk= 175 and z-,xk = k=l k=l Calculate a two-sided 95% confidence interval for E[X]. a) 7 + 0.7360 to.o25.24 b) 7 + 0.7360 to.o25.25 c) 7 + 0.7360 zo.25 d) 7 + 0.5417 to.o25.24 e) 7 + 0.5417 zo.025 Question no. 4 We can show that a theoretical

confidence interval at approximately 95% for the parameter ot of a random variable having a Poisson distribution is given by + 1.96 STD[], where " is the mean of a random sample of size n of X. A par- ticular random sample of size n = 1000 has yielded a sample mean of 0.4. Calculate (approximately) the confidence interval for or, based on this particular sample. a) 0.4 + 0.02 b) 0.4 + 0.03 c) 0.4 + 0.04 d) 0.4 + 0.05 e) we cannot calculate it Question no. 5 We want to test the hypothesis that a random variable X has a Gaussian N(0, 1) distribution. A (particular) random sample of size n = 100 of X has enabled us to construct the following table:

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 293 j (-x, -0.674) [-0.674,0) [0, 0.674) [0.674, x) nj 20 25 25 30 Calculate the statistic D 2 used to carry out the test. Hint. We have: Q(0.674) - 0.25. a) 0 b) 4/25 c) 1 d) 2 e) 4 Question no. 6 Let 1 fx(x;O)=-- if-0 <x <0, 20 where 0 > 0 is an unknown parameter. We have: E[X 2] = 02/3 and E[X 4] = 04/5. A) TO estimate the unknown parameter/ := 02, we propose the following estimator: /:: -- Xk, n k=l where Xl ..... Xn is a random sample of size n of X. Calculate the mean square error of /32 4/32 5@ a) - b)4 c) d) e) B) We collected the following observations of the random variable X: j (-2, -1) [-1, 0) [0, 1) [1, 2) nj 27 28 32 33 We want to test the hypothesis 1 H0: fx(x;O)=fx(x;O=2)=- if -2<x <2. 4 Calculate the statistic D 2 used to carry out the test and give the number d of degrees of freedom of D 2 (under H0). a) D 2 " 0.8667; d = 2 b) D 2 " 0.8667; d = 3 c) D 2 -- 0.8739; d = 2 d) D 2 -- 0.8739; d = 3 e) none of these answers Question no. 7 Let Xl ..... Xn be a random sample of a random variable X having an exponen- tial distribution with parameter ,. We define 0 = /. Find the maximum likelihood estimator of the parameter 0. a) 1/"- b) 1/ c) 1/ 2 d) e) Question no. 8 The following data have been collected by tossing a die 90 times:

294 6 Estimation and Testing j 1 2 3 4 5 6 nj 13 12 16 18 15 16 Let X be the number obtained by tossing the die. We want to test the hypothesis H0: P[X is even] = P[X is odd]. Calculate the statistic D 2 used to perform the test and give the number d of degrees of freedom of D 2 under the null hypothesis Ho. a) D 2=0;d=0 b) D 2=0.05,;d= 1 c) D 2=0.05,;d=2 d) D 2= 1.6;d----4 e) D 2= 1.6;d=5 Question no. 9 A particular 95% confidence interval for the mean # of a population X , -t- 3 N(#, 9) calculated from a random sample of size n = 99, is given by 10 zo.o2.s . A 100th observation, xoo, is taken. Calculate the new

confidence interval, if xo0 = 11. a) 10 4- zo.o25(0.3) b) 10.01 4- zo.o25(0.3) c) 10 4- z0.o25(0.299) d) 10.01 4- zo.o2.s(0.299) e) none of these answers Question no. 10 We want to test the hypothesis that the following data come from a B(3, �) dis- tribution: j 0 1 2 nj 6 18 16 Calculate the statistic D 2 used to make the decision and give the number d of degrees of freedom of D 2 under the null hypothesis. a) D 2 1.87; d = 2 b) D 2 1.87; d = 3 c) D 2 3.87; d = 3 d) D 2 5.87; d = 2 e) none of these answers Question no. 11 Suppose that X has a uniform distribution on the interval [0, 0]. To estimate the unknown parameter 0, we propose the estimator = 2, where 2 is the mean of a random sample of size n of X. Calculate the mean square error of g. 0 2 0 2 0 2 0 2 0 2 0 2 0 2 a) - b) - + - c) d) T e) ; + Question no. 12 Let X be a random variable having a Poisson distribution with unknown parame- ter 0 > 0. We consider a random sample of size n (> 30) of X. Use the central limit theorem to obtain an approximate 100(1 - o0% confidence interval for 0. a) 2 4- zot/201/2 b) 2 4- zot/2(O/n)1/2 c) .J 4- zot/2(f(/n) d)) 4- zo/2(f(/n 1/2) e) f(4- zo/2(f(/n) 1/2 Question no. 13 Suppose that X N(0, 0), where 0 is an unknown parameter. Find the estimator of 0 by the method of maximum likelihood.

interval for/. a) 2 + za/2() d) 2 2 -+- zot/2(12l/n 1/2) Question no. 15 Let 6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 295 n 1 n 1 n a) j2 b) b Y-k=l X 2 c) Y-k=l X 2 d) Y-k=l(Xk -- j)2 e) _1E n k=l (Xk -- 2)2 Question no. 14 Let X N(/x,/xa). Obtain a formula for an approximate 100(1 -o0% confidence b) 2 :t: zo/2(l/n 1/2) c) zt: zo/2(12l/n) e) + zot/z(ll/n 1/2) 1 fx(x; O) = -de -Ixll� forx 6 N, where 0 is a positive parameter. We can show that VAR[X] = 202. A) Find the maximum likelihood estimator of 0. 1 n 1 n 1 n 1 n a)0 b)Y-k=lXk c)Y-k=lIXkl d)(Y-k=lX2) 1/2 e) (k=l X) 1/2 B) We consider the following estimator of the pameter fl := 02: k=l Calculate the bias of/. a) 0 b) fl c) nn_--fl d) 5-_llfl e) �fl Question no. 16 Two makers of batteries for laptop computers claim that with their batteries the computers can be used for at least three hours before the batteries have to be recharged. Batteries from these two makers have been tested with ten different com- puters. The data are the following: Computer 1 2 3 4 5 6 7 8 9 10 2.9 2.8 2.9 3.2 3.0 3.1 2.7 2.9 2.7 2.9 3.1 3.2 3.3 3.0 2.9 2.9 3.1 3.2 2.8 3.2 where Ti is the functioning time (in hours) obtained with a brand i battery, for i = 1,2. We assume that has (approximately) a Gaussian distribution. A) We believe that the standard deviation of the random variable TI is greater than 0.2. Calculate the statistic w02 needed to perform the test. a) 0.636 b) 1.145 c)

5.725 d) 10.305 e) 11.450 B) Next, we want to test the hypothesis that there is not a significant difference between the mean of TI and that of T2. Calculate the appropriate statistic. a)-2.21 b)-2.06 c)-1.96 d) 2.21 e) 7.00

296 6 Estimation and Testing C) We find that the standard deviation of the random sample of T2 is (about) 0.1636. How many observations should we take (approximately) in order that fl -- 0.10, if #r2 ---- 2.9, when we perform the test of H0:#r2 ---- 3 against H" #r2 < 3 at significance level = 0.05? a) 23 b) 25 c) 28 d) 31 e) 34 Question no. 17 A company has invented a device that is supposed to reduce fuel consumption by at least 10%. A) Let R be the reduction in fuel consumption obtained with the device in question. We suppose that R has a gamma distribution with parameters and .. i) If = 1, we find that the maximum likelihood estimator of the parameter 0 := 1/. is 0t/. = /, where / is the mean of a random sample of size n of R. Calculate the mean square error of OMI.. Hint. We have: EIR] = /. and VAR[RI = /.2. 0 e) 02 1 b) 0 c) d) - a) -N ii) Estimate the parameter by the method of moments, if = .. n n 1 n R 2 a)/ b) l// c) [(� Yk= Rk2) -- l] - d)[�Yk=Rk2] -' e)Yk= k B) The device has been installed on ten cars of the same make and their fuel con-sumption has been measured over a 200-km distance. The results are given in the table below: Car 1 2 3 4 5 6 7 8 9 10 X 17.5 18.3 19.1 16.4 18.9 17.8 20.2 19.4 17.6 18.4 where X is the fuel consumption with the device installed. We assume that X has (approximately) a Gaussian distribution. Knowing that the average fuel consumption of this make of car, without the device, is equal to 10 liters/100 km, we wish to test at significance level = 0.025 Ho:#x= 18 against H:#x > 18. i) Give the value of the statistic and of the percentile used to make the decision. a) 1.04; 2.262 b) 1.14; 2.262 c) 3.29; 2.262 d) -0.95; 1.960 e) 0.95; 1.960 ii) Use a linear interpolation, based on the fact that qb(1) _ 0.8413 and qb(1.5) _ 0.9332, to compute approximately the value of fl if the true value of the mean of X is#x = 19andifcrx= 1. a) 0.092 b) 0.102 c) 0.112 d) 0.122 e) 0.132 iii) What is the smallest number of observations that must be taken to be able to detect, with a probability of at least 0.95, that #x is greater than 18 if we suppose that the true value of #x is 19 and that crx = 1 ? a) 12 b) 13 c) 14 d) 15 e) 16

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 297 Question no. 18 Let X1 ..... Xn be a random sample of a random variable X whose probability density function is defined by fx(x; 0) = Ox 0-1

if $0 < x < 1$, where $0 > 0$ is an unknown parameter. A) Find the estimator of 0 by the method of moments. a) 2 b) l-2 c) 2(1-2) d)(1-2)/2 e) 2/(1-2) B) In order to check whether the model fx(x; O) proposed above is adequate for a certain random variable X, we perform a (Pearson's) goodness-of-fit test. We gather 162 observations of X and we group them into three intervals. We obtain the follow- ing table: Interval (0, ⬥) [3') [' 1) Number of observations 14 62 86 Moreover, we find that the value of the maximum likelihood estimator of the param- eter 0 calculated from these 162 observations is OMI = 2. Calculate the statistic D 2 used to carry out the test and give the number d of degrees of freedom associated with D 2 . a) D 22.25;d= 1 b) D 23.25;d= 1 c) D 24.25;d= 1 d) D 22.25;d=2 e) D 23.25;d=2 Question no. 19 A maker of electronic chips asserts that the average size # (in micrometers) of certain components of its chips does not exceed 2. In order to check this assertion, we decide to perform a test of hypotheses. The test used should be such that there is a 0.05 probability of rejecting the maker's assertion if # is equal to 2. This probability should be 0.9 if # is equal to 2.5. We assume that the size of the components in question has (approximately) a Gaussian N(#, 0.16) distribution. A) Give the hypotheses that we want to test. a) Ho: # = 2; Hi:/z = 2.5 b) Ho: # = 2; Hi:/z 5& 2 c) H0: # = 2; Hi: # < 2 d) H0: # = 2; Hi: # > 2 e) Ho: # = 2.5; Hi:/z > 2.5 B) What is the smallest size of the random sample that must be drawn? a) 2 b) 6 c) 15 d) 18 e) 35 C) If/z = 2.4 and n = 10, what is the probability of rejecting the maker's assertion? a) qb(1.1) b) qb(1.2) c) qb(1.3) d) qb(1.4) e) qb(1.5) Question no. 20 The lifetime X of the transistors produced by a certain company has a Gaussian distribution with a standard deviation of 30 days. The company aims at an average lifetime of 210 days. In order to estimate the mean # of X, the engineer responsible

298 6 Estimation and Testing for the control of the quality examines 36 transistors taken at random and obtains a sample mean of 225 days. A 95% confidence interval for p is then given by [215.2, 234.8]. A) We want to reduce by half the width of the confidence interval for p by increasing the size of the sample. What value of n must be used? a) 54 b) 72 c) 90 d) 108 e) 144 B) We now want to reduce by half the width of the confidence interval for , while keeping the sample size equal to n = 36. What value of ot must then be chosen? a) ⬥Q(0.98) b) Q(0.98) c) 2Q(0.98) d) ⬥Q(1.96) e) 2Q(1.96) C) Next, the engineer wants to test, at level 0.025, whether the data indicate that the average lifetime is greater than 210 days. i) Give the hypotheses that we want to test. a)

H0:/ = 210; H:/ <210 b) Ho:/ = 210; H:/ # 210 c) H0:/=210:H:/>210 d) H0:/=210;H:l >225 e) H0:/ = 225; H:/ < 210 ii) With the test used in i), what is the probability of concluding that the average lifetime is not greater than 210 days, when it is actually 220 days? a) Q(0.04) b) Q(-0.04) c) Q(0.14) d) Q(-0.14) e) Q(0.24) D) Suppose that the population variance is unknown. What is the value of the statistic used to test the hypothesis that the mean of X is greater than 210 days, if the sum of the squares of the 36 observations is equal to 1,875,000? a) 2.02 b) 2.12 c) 2.22 d) 2.32 e) 2.42 Question no. 21 Let X be a random variable having the following probability density function: I 20xe-◆X2 ifx > 0, fx(x; O) 0 ifx <0, where 0 > 0 is an unknown parameter. A) Find the maximum likelihood estimator of the unknown parameter 0. a) .- b) (..)2 C) (..)-2 d) '=' X2 e) n B) We collected 30 observations of the random variable X, with which the following frequency table was constructed: 12 8 10 We want to test, at significance level ot = 0.05, the hypothesis H0: fx(x; O) : fx(x; O : 1) : 2xe -x2 against Hi: fx(x) : 2xe -x2

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 299 for x > 0. Give the value of the statistic used to perform the test, as well as the number d of degrees of freedom associated with this statistic (under H0). a) 0.20; d = 1 b) 0.20; d = 2 c) 0.40; d = 1 d) 0.40; d = 2 e) 0.60; d = 1 Question no. 22 A tire manufacturer claims that the average braking distance for a car equipped with its brand Z tires is not greater than the average braking distance obtained with the more expensive brand W tires. We believe that this statement is false and we decide to test it by measuring the braking distance (in meters) from 100 km/h to a complete stop. The data are the following: Brand Z tires 40 41 39 44 45 Brand W tires 38 40 38 41 45 Let Z be the braking distance obtained with the brand Z tires, and W the braking distance obtained with those of brand W. We assume that Z N(#z, az 2) and W N(#w, av) (approximately). A) Suppose that the measurements have been made with five different makes of cars, the first time using the brand Z tires, and the second time using the brand W tires. Give the value of the statistic used to test, at level x = 0.05, the hypothesis H0: #z = #w. Give also the value of the percentile c to which the statistic in question is compared. a)0.81;c 1.860 b) 0.81;c 2.306 c) 2.75;c 1.860 d) 2.75; c _ 2.132 e) 2.75; c _ 2.776 B) Suppose that the ten braking distance measurements have been made with the same car, on ten different days. Calculate, assuming that az = aw = 2.5, the value of the appropriate statistic to test at level u = 0.05 the hypothesis H0 : #z = #w, and

give the value of the percentile c to which this statistic is compared. a) 0.81;c _ 1.645 b)0.81;c -- 1.860 c) 0.89; c -- 1.645 d) 0.89; c -- 1.860 e) 0.89; c - 2.306 C) Under the same assumptions as in B), what is the probability of detecting that the average braking distance #z is greater than the mean #w when, in fact, #z - #w = 1.25? a) Q(0.791) b) Q(0.854) c) Q(2.436) d) qb(0.791) e) qb(0.854) Question no. 23 Suppose that X N(-0, 0), where 0 > 0 is an unknown parameter. A) We propose the following estimator of the parameter 0: := -){. Calculate the mean square error of . 0 c) 0 d) ◆2 a) 02 b) 02 + - e) 02 0 B) Use the estimator = -) to obtain a formula giving an approximate 100(1 -) % confidence interval for 0, based on a random sample of size n = 100, if we assume that) < 0.

300 6 Estimation and Testing a) -," zlz-- Z (--2)1/2" b) -," 4- zc,/2(o) c) -," 4- c,/2- (-2)/2rg 01/2 01/2 d) -" 4- Zo/2T6- e) -Y(4- z, lo Question no. 24 The speed X of the microchips of a certain company is supposed to be 2 GHz. We assume that X has (approximately) a Gaussian N(#, a2) distribution. A) We take a random sample of size n = 9 of X and we compute the average speed ◆ of the microchips. If the sample standard deviation s is equal to 0.2, for what values of◆ can we conclude that the average speed of the microchips is smaller than 2 GHz? Use = 0.025. a)Y< 1.85 b)Y> 1.85 c)Y< 1.95 d) J> 1.95 e)[◆-2[<0.15 B) Suppose that the average speed of the microchips is actually equal to 2.1 GHz, and that a = 0.25. What is the probability of rejecting the hypothesis H0: # = 2 (accepting H : # < 2 instead) at level a = 0.05, if we draw a random sample of size n= 16? a) (2.2) b) (3.2) c) Q(2.2) d) Q(3.2) e) 2 (2.2) - 1 C) What is the value of the statistic used to test the hypothesis Ho: a2 = 0.02 against H: r 2 > 0.02 if a random sample of size n = 10 yielded 10 10 xk =20.17 and xk 2=40.775? k=l k=l What is the conclusion of the test ifa ---- 0.05? a) 4.6; we reject H0 b) 4.6; we accept H0 c) 9.2; we reject Ho d) 9.2; we accept H0 e) 18.4; we reject H0 Question no. 25 Suppose that X is a random variable having a Weibull distribution with parame- ters 15 > 0, y 6 R anda = l. That is, fx(x; 15,) = 15(x -)/- exp{-(x -)/} forx > . A) Let ?/ = 0. Give the value of the estimator of the unknown parameter 15 by the method of moments, if a random sample of X yielded ◆ = 6. Hint. Set y = x in the calculation of the mathematical expectation of X and use the fact that f0 = F(). xa-le-X dx a) 1/5 b) 1/4 c) 1/3 d) 1/2 e) 1 B) Let 15 = 2. Estimate the parameter y by the method of maximum likelihood with the help of a single observation, X, of X. 1 a) X1-◆ b) X1-4 c) X1 d) Xl+4 e) Xl+

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 301 C) Test, at significance level oe = 0.05, the hypothesis H0 that fx(x;,/)=fx(x;=l,/=O)=e -x forx > 0, if we collected the following data: j [0, 1) [1, 2) [2, 3) [3, nj 80 40 20 10 Give the value of the statistic D 2 and the conclusion of the test. a) 5.94; we accept H0 b) 6.94; we accept H0 c) 6.94; we reject H0 d) 7.94; we accept H0 e) 7.94; we reject H0 Question no. 26 A car maker, in its publicity, claims that the power X (in horsepower) of the engines of its cars of brand A is 200 hp. Tests made with 16 independent vehicles gave an average power of 196 hp and a sample standard deviation s of 5 hp. We assume that X has approximately a Gaussian N(#, r 2) distribution. A) We can state, with a risk oe of type I error equal to 0.05, that the average power of the engines is in fact smaller than 200 hp. Give the values of the statistic and the percentile that were both used to make the decision. a)-3.2;-2.131 b)-3.2;-1.753 c)-3.2;-1.645 d) 3.2; 1.753 e) 3.2; 2.131 B) Based on the data collected, can we conclude, if ot = 0.05, that r x > 4? Give the value of the statistic used and the conclusion of the test. a) 4.69; we accept H0 b) 18.75; we accept H0 c) 18.75; we reject H0 d) 23.44; we accept H0 e) 23.44; we reject H0 C) What is the smallest number of observations that should be taken in B) to be able to state, with a probability of at least 0.9, that rx > 4 if, in fact, rx = 6? a) 23 b) 24 c) 25 d) 26 e) 27 Question no. 27 The random variable X has the probability density function 3 x2 fx(x; 0) = 2 if-0 < x < 0, where 0 (> 0) is an unknown parameter. A) Find the estimator of the parameter 0 by the method of moments. n n 2 \1/2 a)0 b)) c)(35-))) 1/2 d) 3 k=l X2 k e) (n k=l Xk) B) We constructed the frequency table below from a random sample of size 80 of X: j [-2,-1) [--1, O) [0, 1) [1,2] nj 32 4 5 39

302 6 Estimation and Testing Give the value of the statistic D 2 used to test H0:fx(x;0)= fx(x;O=2)=]-6 x2 if -2<x <2. How many degrees of freedom does the statistic D have (under H0)? a) D =27/35;d=2 b) D =-32/35;d=2 c) D =32/35;d=3 d) D 2=37/35;d=3 e) D =-37/35;d=4 Question no. 28 We wish to compare the fuel consumption (in liters per 100 km) of two makes of cars, A and B. We believe that the cars of brand B use less fuel, on average, than those of brand A. Let XA (respectively XB) be the fuel consumption of the cars of brand A (resp. B). We assume that X A N(/A, aA 2) and XB N(/B, a 2) are independent random variables. Random samples of XA and XB have been taken. The data are the following: X A 10.5 10.9 9.8 10.2 11.9 9.9 10.7 10.1 10.6 XB 9.7 10.0 10.3 9.4 9.6 9.8 10.2 10.0 A) We must first test the equality

of variances. We find that if c = 0.05, then we conclude that aA 2 = a 2 . What is the value of the statistic used to test, at significance level c = 0.05, the hypothesis that the average fuel consumption of the brand B cars is smaller than that of the brand A cars? What is the alternative hypothesis H ? a) 2.46; H: p.A > P-B b) 2.56; H: p.A > P-B C) 2.46; H: p.A d) 2.56; H: A < B e) 2.66; H: A > B B) If we choose c = 0.10 instead, then we must conclude that aA 2 : a 2 . What is, in this case, the value of the statistic used to test, at level c = 0.10, the hypothesis in A)? What is the conclusion of the test? a) 2.56; we accept H0 b) 2.66; we accept H0 c) 2.56; we reject H0 d) 2.66; we reject H0 e) 2.56; we cannot conclude C) Suppose that we exclude the observation 11.9 from the random sample of XA, because this observation is doubtful. What is then the value of the statistic that we use to test the hypothesis Ho: aA 2 = a 2 against H: aA 2 :fi a 2 ? What is the conclusion of the test if or = 0.05? a) 1.68; we accept Ho b) 1.78; we accept Ho c) 2.68; we accept Ho d) 2.78; we reject Ho e) 3.68; we reject Ho Supplementary Exercises Question no. 1 Obtain the maximum likelihood estimator of the parameter 0 in the probability density function fx(x; O) = e -◆lxl forx Z

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 303 Question no. 2 Suppose that the random variable X has the following probability density func- tion: fx(x; , 6) = Le -x/ forx _> 0, where > 0 and/3 > 0 are unknown constants. Find the estimator of the parameter by the method of maximum likelihood. Question no. 3 Suppose that X1 N(#Xl, 1) and X2 N(#x2, rx22) are independent random variables. Random samples from X1 and X2 yielded the following table: n . s X 1 10 0.1 0.9 X 2 100 -0.2 1.1 a) Test the hypothesis H0: Lx I = /Lx 2 against HI: Lx I > tZX 2 at significance level = 0.05. b) What is (approximately) the probability of making a type II error in a) if, in fact, #X 1 -- #X 2 = 0.5? Question no. 4 A random variable having a generalized Pareto distribution is such that fx(x;a, fl)----a(/3-2)(l+ax) -◆+1 forx >0, where > 0 and/3 > 2 are unknown parameters. Estimate and/3 by the method of moments, if we assume that/3 > 4. Question no. 5 The tensile strength X (in kN/m 2) of a certain synthetic fiber has a Gaussian N(#, 100) distribution. The company that makes this fiber claims that the average tensile strength is greater than or equal to 240 kN/m 2. In order to check the assertion, we perform a statistical test. A random sample of size n = 18 yielded 2 = 237 kN/m 2 . a) What is the alternative hypothesis H 1 for this test? b) What is the value of the statistic Z0 used to perform the test? c) Calculate the value of/3 if

= 235, n = 18 and = 0.05. d) What is the smallest number of observations that must be taken if the probability of rejecting the company's assertion must be of at least 0.975 when # = 236 and = 0.05? Question no. 6 In order to compare heat loss from steel pipes and from glass pipes, we consider seven pairs of pipes (made up of a steel pipe and a glass pipe). The pipes in a given pair have the same diameter, but diameters differ from one pair to another. Water is

304 6 Estimation and Testing passed through the pipes at the same initial temperature and the heat loss (in �C) is measured over a 50-meter distance. The results obtained are the following: Pair number 1 2 3 4 5 6 7 Steel pipe 4.6 3.7 4.2 1.9 4.8 6.1 4.7 Glass pipe 2.5 4.2 2.0 1.8 2.7 3.2 3.0 We assume that the heat loss has approximately a Gaussian distribution. a) What is the value of the statistic used to test the hypothesis that the heat loss is, on average, greater in steel pipes than in glass pipes? b) What is the number of degrees of freedom associated with the test statistic in a)? Question no. 7 Let $fx(x; 0) =$ 1/0 for $0 < x < 0$. We want to test H0:0 = 1 against H: $0 > 1$. To do so, we take a single observation, X, of X and we reject H0 if and only if $X > 0.9$. a) What is the type I error risk, or, of the test'? b) What is the value of if0 = 1.5? Question no. 8 The precision of the manufacturing process of bolts is under statistical control as long as the standard deviation of the diameter of the bolts does not exceed 1.5 mm. We want to test, at significance level ot = 0.05, the hypothesis that the manu- facturing process is under control. We assume that the diameter X of the bolts has approximately a Gaussian N(z, cr 2) distribution. a) What is the conclusion of the test if the sum of 20 measurements of the diameter is equal to 109.8 and the sum of the squares of the measurements is 655.6? 2 Hint. We have: X0.05. 9 30.144. b) What is the type II error risk of the test performed in a) if the standard deviation of X is actually equal to 2.2 mm? c) Determine the (smallest) size of the random sample needed to obtain a type II error risk of 0.05 when the standard deviation of X is equal to 2.2 mm. Question no. 9 The strength level of material A (respectively B) has approximately a Gaussian distribution: X A N(/A, ffA 2) (resp. XB N(zB, cry)). To test the hypothesis H0: /A A = /A B against H:/A A :) /A B, we took two independent random samples: XA 2.08 3.77 2.46 2.20 2.58 3.35 3.52 Xn 3.13 3.07 3.66 3.51 3.22 We compute -A = 2.85, s -- 0.464; �n = 3.32, a) Perform the test with a type I error risk ot of 0.10. b) Perform the test at significance level ot = 0.05. s -- 0.065.

6.5 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 305 Remark. In a) and b), we must first test the equality of variances with the same type I error risk (10% and 5%, respectively). Question no. 10 (See [1113 The average percentage of waste produced by a certain manufacturing operation is supposed to be smaller than or equal to 7%. We took several days at random and we calculated the percentages of waste: 6.5151, 7.4970, 7.4601, 6.3723, 8.3257, 9.8199, 9.5618. We assume that the percentage X of waste has (approximately) a Gaussian $N(\#, cr\ 2)$ distribution. a) Based on the data, can we conclude that # is significantly greater than 7%? Use a = 0.01. Hint. We have: t0.01,6 " 3.143. b) If it is important to detect a ratio (# - 0.07)/or of 0.5 with a probability of at least 95%, what is the smallest number of observations that must be taken? c) For a ratio (# - 0.07)/or of 2, what is the power of the test performed in a)? Question no. 11 (See [11].) The diameter of a steel ball has been measured by 12 individuals, using two types of pairs of calipers. The following data, where for instance 5 represents 0.265 and 4 represents 0.264, have been collected: Individual 1 2 3 4 5 6 7 8 9 10 11 12 Calipersl 5 5 6 7 7 5 7 7 5 8 8 5 Calipers2 4 5 4 6 7 8 4 5 5 7 8 9 a) Is there a significant difference between the means of the populations from which the two samples come? Use a = 0.05 and specify the basic assumptions that we must make to perform the test. Hint. We have: t0.025,11 -- 2.201. b) Is the standard deviation of the first population significantly greater than 0.001 ? Use t = 0.10. 2 _ 17.28. Hint. We have:

7 Simple Linear Regression 7.1 Introduction: The Model We saw in Chapter 6 how to test the hypothesis that the data we collected come from a given distribution, such as a Gaussian distribution. When the random variable of interest is a function of a deterministic variable (that we can control), regression can help us find the form of the relationship between the random and deterministic variables. In this chapter, we will treat in detail the basic problem, namely that of finding the best linear relationship between the random variable and the deterministic variable. We will also briefly consider the case when the relationship between the two variables is assumed to be non-linear. Let Y be a random variable and x a deterministic variable (that is, non-random). We have a random sample (xl, YI) ..... (Xn, Yn) and we want to find a mathematical relationship that expresses Y in terms of x. The variable x is called the independent variable and Y is called the dependent or response variable. In the case of simple linear regression, the model that we propose is

of the form (7.1) where e is an error term. We assume that each observation Y_i of Y satisfies the equation $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, (7.2) where $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1$ ..... n, and that the random variables ϵ_i are indepen- dent. Note that we take for granted that the variance of ϵ_i is the same for all values of i. We can thus write that $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$ for $i = 1$ ..... n, (7.3) because β_0 and β_1 are parameters (that is, constants).

308 7 Simple Linear Regression Remark. The model is known as the simple linear regression model, because there is a single independent variable, and the model is linear with respect to the parameters. The model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 x_{2i} + \epsilon_i$ (7.4) is linear with respect to the parameters and is a polynomial of degree two, while $Y_i = \beta_0 e^{x_i} + \epsilon_i$ (7.5) is a non-linear model with respect to the parameters. The best estimators of the parameters β_0 and β_1, that is, the minimum variance unbiased estimators of β_0 and β_1, are obtained using the method of least squares. We define the sum $SS = \sum_{i=1}^{n} \epsilon_i^2 = \sum (Y_i - \beta_0 - \beta_1 x_i)^2$. (7.6) The estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 by the method of least squares are the values of β_0 and β_1 that minimize the sum SS. We set and $\frac{\partial SS}{\partial \beta_0} = -2 \sum (Y_i - \beta_0 - \beta_1 X_i) = 0$, $\frac{\partial SS}{\partial \beta_1} = -2 \sum x_i (Y_i - \beta_0 - \beta_1 x_i)$. The solution of these two equations, called the normal equations, is (7.7) $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i Y_i - n \bar{x} \bar{Y}}{\sum x_i^2 - n \bar{x}^2}$ (7.8) Remarks. i) In the case when we assume that the errors ϵ_i are independent and all have a Gaussian $N(0, \sigma^2)$ distribution, as above, we find that the estimators given by the method of least squares and those obtained with the method of maximum likelihood are the same. ii) The ϵ_i's are called the (theoretical) residuals. The quantity ϵ_i represents the dif- ference between the observation Y_i and the value calculated from the proposed model $Y_i = \beta_0 + \beta_1 X_i$. To predict the value of Y when $x = x_i$, we then use the equation and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$. (7.9)

7.1 Introduction: The Model 309 iii) The prediction equation should (theoretically) be used only for values of x in the interval $[x_{(1)}, x_{(n)}]$, where $x_{(1)} := \min\{x_1 ..... x_n\}$ and $x_{(n)} := \max\{x_1 ..... x_n\}$. (7.11) iv) Even if we know that $Y = 0$ when $x = 0$, we do not set $\beta_0 = 0$. We generally obtain better results with β_0 in the model. However, if we indeed wish to propose the regression model through the origin: $Y = \beta_1 x + \epsilon$, (7.12) where $\epsilon \sim N(0, \sigma^2)$, we easily find that the estimator of the parameter β_1 obtained from the method of least squares, based on a random sample (X_1, Y_1) ..... (x_n, Y_n), is (7.13) Example 7.1.1 (See [17].) We want to determine how the tensile strength of a certain alloy depends

on the percentage of zinc it contains. We have the following data: %of zinc 4.7 4.8 4.9 5.0 5.1 Tensile strength 1.2 1.4 1.5 1.5 1.7 We obtain the graph in Fig. 7.1. We consider the simple linear regression model: $Y = \beta_0 + \beta_1 x + \epsilon$, where x is the percentage of zinc and Y is the tensile strength. Remark. It is important to correctly identify the random variable and the determinis- tic variable in the problem. Tensile strength 1.7- 1.6. 1.5- 1.4- 1.3" 1.2. 5.1,1.68) � , . 4) I I I I I 4.7 4.8 4.9 5.0 5.1 % of zinc Figure 7.1. Graph in Example 7.1.1.

310 7 Simple Linear Regression We find that $\bar{x}$ = 4.9, Then, we have: and 5 5 1.46, Zxi2 = 120.15 and Z xiYi = 35.88. i=1 i=1 Y-,= xiYi - 5$\bar{x}$. = 35.88 - 5(4.9)(1.46) = 1.1 y-.= xi 2 - 522 120.15 - 5(4.9) 2 /o = . -/$\bar{x}$ = 1.46 - (1.1) (4.9) = -3.93. Thus, the prediction equation is given by = -3.93 + l.lx. 7.2 Tests of Hypotheses Since the estimators /o and / are linear combinations of independent Gaussian random variables, they also have a Gaussian distribution. Moreover, we can show that/0 and/ are unbiased estimators of/3o and , and that VARIol = o.2 and VARIetal where O-2 (7.14) SSs ' SS x ..: Z(Xi_ ,)2 : Z X- n.2 : Z x'2t (-,i:l xi) 2 n i:l i=l i:l (7.15) Remark. The estimators/o and are generally not independent. In fact, we can show that o.2. COV[/o,/l] -- (7.16) SSx" Since two Gaussian random variables are independent if and only if their covariance (or their correlation coefficient) is equal to zero (see p. 180), we can state that/0 and /3 are independent random variables if and only if$\bar{x}$ = 0. However, let / :/0 +/$\bar{x}$- (7.17) We find that COV[/,/,] = 0, so that/6 and/, are independent estimators. Next, we define the sum of the squares of the errors (or of the residuals) SSE = Z(Yi -]i)2. (7.18) i=1

7.2 Tests of Hypotheses 311 We can show that E[SSE] = (n - 2)a 2. (7.19) It follows that an unbiased estimator of the variance cr 2 is given by (the mean square of the errors) SSE r = MSE := . (7.20) n-2 a) To test the null hypothesis Ho: rio =/300, (7.21) we use the statistic to - rioo no TO := " tn-. (7.22) MSE ;q-x We then reject Ho at significance level c if and only if IZol > t/2,n-2 ifH: r0 � fl00, TO > tc,n-2 if HI:riO > floo, TO < -t,n-2 if H:riO < riO0. (7.23) b) Similarly, to test Ho: ri = rico, we use the statistic / -rico /40 TO .-- " tn-2 /MSe/SSx and we reject Ho at significance level c if and only if ITol > tc/2,n-2 if ill: ril rilO, TO > tc,n-2 if HI: ril > ril0, TO < -tc,n-2 if ill: fl < fl0. (7.24) (7.25) (7.26) Particular Case: Test of the (Global) Significance of Regression The test of Ho: fll =0 against Hi: fl � 0 (7.27) is very important, because if

we cannot reject the null hypothesis Ho, then we must conclude that there is no linear relationship between x and Y (or, at least, we cannot conclude that there is a significant linear relationship between x and Y). To perform

312 7 Simple Linear Regression the test, we can proceed as above. In practice, we perform instead, equivalently, an analysis of the variance. Note first that when we choose fl0 = 0, we have: T- MSe/SSr" (7.28) We can show that = ss, := }2(9, - f)2. (7.29) i=1 We say that the sum of squares SSR is the sum of squares due to regression. It can be shown that the sums SSR and SSF. are independent random variables having chi-square distributions with 1 and n - 2 degrees of freedom, respectively. It follows that F0 := T02 = SSI , FI,,,-2. (7.30) MSF. Remarks. i) Actually, by definition, the square of a random variable having a Student distribution with n degrees of freedom is a random variable having a Fisher distribu- tion with 1 and n degrees of freedom. Therefore. we could have written directly that T -. Fl,n-2, when the null hypothesis H0 is true. ii) Since SSR has a single degree of freedom, the mean square MS8 and the sum of squares SSR are equal. Finally, we can show that E[SS] = c* 2 + flSSx ' c* 2. (7.31) Since E[MSt] = c, 2, the larger the value of the statistic Fo, the smaller the proba- bility that the null hypothesis Ho is true. We reject Ho at significance level c if and only if Fo > Fa,.n-2. (7.32) The procedure to test the global significance of regression can be summarized using an analysis of variance table (see Table 7. I, p. 313), in which the total sum of squares SSr is defined by SST = Y(Yi -]7.)2 = Yi 2 - n 2. (7.33) i=1 i=1 In practice, we first calculate the sum of squares SSr. Next, we have: SS =-- = where SSxr := (xi -�)(Yi - ') = xiYi -n�'. (7.34) i=1 i=1 We can then complete the analysis of variance table.

7.3 Confidence Intervals and Ellipses Table 7.1. Analysis of variance. Source Sum Degrees Mean of variation of squares of freedom squares F 0 MSR Regression SSR 1 MSR MSE Error SSE n - 2 MSE Total SS T n - 1 313 Example 7.2.1 We found in Example 7.1.1 that = -3.93 + 1.1x. We now want to test Ho:fl =0 against Hl:fl 5 0. We have: 5 SSr = Zy- 5y 2 = 10.79- 5(1.46) 2 = 0.132 i=I and ssR = iSSxv = (1.1) ('-xiYi- 5Y) = (1.1)(35.88- 35.77)=0.121. i=I Thus, SSE = SST - SSR -- 0.011 and MSR 0.121 fo -- MSe -- 0.011/3 --33. We find in the table giving the values of Fo.os,nl,,2, Appendix A, p. 343, that Fo.o5,1,3 -- 10.1. Since 33 > 10.1, we can reject Ho at significance level o = 0.05. Note that in fact we have: Fo.o5,1,3 = to2.o25,3 � Therefore, we can

also use Table 6.2, p. 261, to obtain the critical values Fo.os, I,n-2. 7.3 Confidence Intervals and Ellipses Using the following results: 6/S_x tn_2 and]1 tn-2, (7.35) y2 6 / +S where 6 = = /SSe/(n - 2), we can calculate confidence intervals for the parameters fl0 and ill. We find that the two-sided confidence intervals at 100(1 -o)%

314 7 Simple Linear Regression are given (in compact form) respectively by rio + t/,_e + -- and f l [31 � ta/2,n-2`8 SSx" 2 (7.36) s& (7.37) Similarly, we can show that a 100(1 - o0% confidence interval for the average value of the random variable Y when x = " lxll ), x(n)], that is, for fl0 + fll ', is Oq-l q-ta/2"n-2`8 -q- SS----- (7.38) We see that the width of the interval is a minimum when = . Now, the preceding formula is valid when we wish to compute a confidence interval for fl0 + fil, based on the n observations collected. In the case when we wish to compute a confidence interval for a new observation (that is, for a prediction) of Y at x : instead, the term a 2 is added to the variance and the formula becomes / 1 (_))2 0 -t- 1 � t/Z,n-2`8 1 + - + -- (7.39) n SS- Remark. If we want a confidence interval for the mean of m new observations of Y at 1 x = s e, we only have to replace the term I (added in front of -) by in the preceding formula. Finally, we can show that the interior of the ellipse defined by n(flo --/)2 + 2n(flo -- 0)(ill -- ill) + (ill -- fi,)2 Z xi2 = 2`82 Fm2'n-2 i=1 (7.40) is a 100(1 - o)% confidence region for the pair (rio, ill). Example 7.3.1 A two-sided 95% confidence interval for a new observation of Y when x = = 4.9, in Example 7.1.1, is given by / 1 -3.93 + 1.1 (4.9) -t- t0.025,3`8 1 + + 0 _ 1.46 -t- 0.21, because we have: Tab. 6.2 t0.025,3 " 3.182. (see Example 7.2.1, p. 313) and

7.5 The Analysis of Residuals 315 7.4 The Coefficient of Determination Definition 7.4.1. The coefficient of determination, R 2, is obtained by dividing the sum of squares SSR by the sum of squares SST : R2 - SSR. (7.41) SSr The coefficient R 2, also called the squared correlation coefficient, enables us to measure the fit of the model to the data. If all the data points fall on the regression line, then we have: R 2 : 1. In general, R 2 takes on a value in the interval [0, 1] and gives us the percentage of the total variation SST that is explained by the regression model. It follows that 1 - R 2 -- SSE (7.42) SST is the proportion of SS that is not explained by the model. Remarks. i) The expression correlation coefficient is in fact not correct, because x is not a random variable. Indeed, a correlation coefficient is only defined for two

random variables, X and Y. ii) The quantity R : .v is called the fit index. It is often used to measure the quality of the regression model. However, it is important to know that if the value of R is large, this does not necessarily mean that the (simple) linear regression model is the right one. Similarly, if the value of R is small, we should not conclude that the model must be rejected. The index fit is actually a measure of the improvement to the fit obtained by including the term fllx in the model, in comparison to the regression model Y : flo + e. Example 7.4.1 Using the results of Example 7.2.1, we find that R2 _ 0.121 0.917 0.132 with the data in Example 7.1.1. 7.5 The Analysis of Residuals We already called the quantities e i = Yi -- flO -- fllXi the (theoretical) residuals. We define the sample residuals by ei :Yi --i. (7.43) The analysis of the residuals, which is generally carried out with the help of a statis- tical software program, enables us to check, in particular, the validity of the assump- tion made at the beginning of this chapter, namely that the error terms ei all have a

316 7 Simple Linear Regression Gaussian distribution (with zero mean and common variance (72). If this assumption is true, then the quantities ei Zi : -, (7.44) (7 called the standardized residuals, should be particular observations of a random variable Z having approximately a standard Gaussian distribution. Since P[-2< Z <2]--0.95 and P[-3 < Z < 3] -- 0.997, (7.45) there should not be more than (about) 5% of the zi's such that Izil 2, and almost no zi's such that Izil > 3. Remark. If the size of the random sample is small, then we must use the Student distribution. That is, we must compare the zi's to the values taken by a random variable having a tn-2 distribution. Example 7.5.1 The regression equation obtained with the data provided in Example 7.1.1 is . = -3.93 + l.lx, and we have: 6" ---- 0x/0--.011/3. Making use of these results, we can build the following table: xi Yi i ei zi 4.7 1.2 1.24 -0.04 -0.67 4.8 1.4 1.35 0.05 0.83 4.9 1.5 1.46 0.04 0.67 5.0 1.5 1.57 -0.07 -1.17 5.1 1.7 1.68 0.02 0.33 We see that all the standardized residuals are small. Therefore, the normality assump- tion of the random variables i seems reasonable. Remark. Since n = 5, we should compare the residuals to the values taken by a t3 distribution. Now, we have: P[-3.182 < T < 3.182] " 0.05 if T t3. Therefore, the residuals could have been located in the interval [-3.182, 3.182]. However, the size of the particular random sample is a little small here to carry out an analysis of residuals. The analysis of the residuals also enables us to check the other basic assumptions made in simple linear regression: a) the model is of the form Y = flo + flx; b)

the errors ε_i are independent and all have the same variance (σ^2.

7.5 The Analysis of Residuals 317 e i x X X X X X X X X X X X X X X X X X X X X X Xi ε r Yi Figure 7.2. Residuals forming a uniform band. To do so, we plot the graphs of the residuals ei against the x i's and the i 'S. In all cases, the points in these graphs should form a uniform band, as in Fig. 7.2. Figure 7.3, p. 318, shows cases when we can conclude that i) the model is not linear in x (graphs (a) and (b)); ii) the variance of the errors is not constant (graphs (c) and (d)); iii) both at the same time (graphs (e) and (f)). Remark. If there exists at least one value of the independent variable x for which we have at least two observations of the dependent variable Y (and if there are at least three different values of x in all), then we can perform the following test of the fit of the model to the data: let Y/,1 .... , Yi,ni be ni observations of the random variable Y when x = xi, for i = 1 ..... k. We define the sum of the squares due to pure error: SSpE = Z Z (Yi'j -- lpi)2' (7.46) i=1 j=l where 1 ni /'li .= (7.47) and we set SSE = SSpE + SSL, (7.48) where SSL is the sum of squares due to the lack of fit of the model. Next, we consider the statistic SS/(k - 2) F0 := SSpe/(n - k)' (7.49) k where n := Y-4=1 ni is the total number of observations in the random sample. We reject, at significance level oe, the hypothesis that the model is adequate for the data when F0 > F,k-Z,,,-k.

318 e 7 Simple Linear Regression x x x x x x x x x x x .... x x x (a) = Xi ε r i e i x x x x x x ___ x____ x .... _x__x____ X X X X X X X X X X X (b) X i or i e i x x x x x x xX x x x (c) X i or i e i x x x x x x x x x x x x x x x x x x x (d) X i or i e i X X X x X X _X. X X X X -x .... X X X X X X i or i x x x x x x x x x x x x x x x (e) (f) Figure 7.3. Residuals showing at least one assumption not satisfied. 7.6 Curvilinear Regression Sometimes, a non-linear regression model can be reduced to a linear regression model by transforming the Y variable or the x variable (or both at the same time). For example, if we propose the model (without an error term) Y =$\beta_0 e/x$, (7.50) then it suffices to take the (natural) logarithm on each side of the equation, and next to set Y' = In Y and β_0' = In β_0. We thus obtain the model

Similarly, the model becomes if we define yt = _ In Y. 7.6 Curvilinear Regression 319 yt = ; ___ lX. (7.51) 1 r = (7.52) exp(flo + 1 X) Y'=fl0+fl1X (7.53) Remarks. i) If we assume in (7.51) that Y[= β_0 + $\beta_1$xi + ε_i, (7.54) where ε_i N(0, 0 "2) for i = 1 ..... n, then this implies that the original model is Y/ = fl0

e/hxi+i. (7.55) An error term of the form e ei may not be realistic in many situations. On the other hand, assuming that the error term has a lognormal distribution can be interesting in some cases. ii) The non-linear models that can be linearized are called intrinsically linear. iii) We realize that the relationship between x and Y is not linear by looking at the graph of the Yi's against the xi 's, or at the graph of the ei's against the i 'S. Example 7.6.1 We have the following data: x 1 2 3 7 Y 13.0 21.9 29.8 30.4 11.8 24.7 24.1 35.7 where x is the time (in days) elapsed after the setting of a cement and Y is the tensile strength (in kilograms per square centimeter) of the cement. We propose the curvilinear regression model Y = oe- x. a) Transform the variables to obtain a linear model. b) Estimate the parameters fl0 and fll by the method of least squares. c) Calculate a 95% prediction interval for Y when x = 5 and 10. Remark. When there is more than one value of Y for a given value of x, we should specify in the problem whether all the observations are independent (actually, we should always specify in the problem that the observations are independent). Here, we assume that we indeed have eight values taken by independent observations.

320 7 Simple Linear Regression Solution. a) We have: Y =/3oe -x � In Y = In/30 -/3x. Therefore, we simply have to set Y' = In Y,/3 = In/30 and/3t = _/3 to obtain y' =/3 +/3'1 x. Remark. We could also have set x' = -x, rather than/3' = -/3. b) We must first transform the data. We obtain: x 1 2 3 7 " Yt 2.565 3.086 3.395 3.414 2.468 3.207 3.182 3.575 Next, we calculate fi and fi'l using the formulas in Section 7.1. We find that fi) " 2.69 and fi'l " 0.13. It follows that 14.74 and fil = -i -0.13. c) The 95% prediction interval for Y', when x = , is given by / 1 (_.)2 2.447 We find that � = 3.25, SSx = 41.5 and (with the transformed data) SSr - 1.114. It follows that SSE = SSr - (fitl)2 SSx " 0.413, and then Thus, we obtain (O'l 3) 1/2 6" " " 0.262. =5=�,3.340.70 and = 10=�,3.990.96. Finally, in terms of the variable Y, the prediction intervals become Yl{x=5} 6114.0, 56.8] and YI {x= 10} 6120.7,141.2]. Remarks. i) Looking at the data, we see that Y increases when x increases. Therefore, the fact that fi is negative is logical.

7.7 Correlation 321 ii) Since/1 is a real parameter, the model considered is equivalent to the model Y = floe fllx. iii) If we had proposed the model Y = /0 /1 In x, for instance, then it would have been sufficient to set x' = In x to obtain the simple linear regression model. Moreover, if we assume that where ei N(0,

(r 2) for i = 1 ..... n, then we have: Yi =/3o +/31 In xi + el. That is, if we only transform the independent variable x, then the error term ei in the model is unchanged. 7.7 Correlation When x is not a deterministic variable, but rather a random variable X, we as- sume that (X1, Y1) ..... (X, Y. is a random sample of size n of the random vector If we assume further that X has a Gaussian N(#x, rx) distribution and Y (x, Y). rr), then we can show that the expected value and the variance of N(#y, Y, given that X = x, are given by (see p. 179) E[Y [X = x] = #y -+- tox,y--(x - #x) (7.56) x and VAR[Y I X = x] = Cry(1 - Px,y). (7.57) It follows that we can write that E[Y [X = x] = 1o llx, (7.58) where o'y O'y 0 := Y -- PX,Y--#X and /31 := Px,Y--. (7.59) o- X o- X We can show that the maximum likelihood estimators of the parameters/0 and/1 are given by 1 = i=1 (Xi -2)(I - f') = i=1 Yi(Xi - 2) (7.60) zin=l (Xi -- 2) 2 zin=l (Xi - 2) 2 and 0 = I? - ◆12. (7.61) Note that these estimators are of the same form as those obtained by the method of least squares in Section 7.1.

322 7 Simple Linear Regression Now, the estimator fix, r of the theoretical correlation coefficient Px, r is the sam- ple correlation coefficient Rx,r, defined by n X i=(i - 2)(Yi - f') SSxy RX'y = n X n "-- (7.62) [Zi=I (i -- ,)2i=l(Yi -- f')2] l/2 Remark. We find that the square of the sample correlation coefficient, rx.y, of the observations in a particular random sample of a pair (X, Y) is equal to the coefficient of determination: r 2 ---- R 2. However, the coefficient of determination was defined X.Y for a deterministic variable x (and a random variable Y). As we already mentioned, it is not rigorous to state that the coefficient of determination is the square of the sample correlation coefficient of the variables x and Y. Test of Hypotheses To test the hypothesis Ho: Px.Y = 0 against H: Px.Y # O, (7.63) we use the statistic To:= /n 2 Rx, y no -- tn-2. (7.64) /1 -- R 2 X,Y We reject Ho at significance level oe if and only if Remarks. i) Since SSr = SSr and fi = SSxr/SSx, we can write that (SSx'] '/2 RX'y = / \-'T J (7.65) Therefore, testing Ho: Px,Y = 0 is equivalent, from a mathematical point of view, to testing H0:/4 = 0. ii) To test the more general hypothesis Ho: Px, Y = P0, we can use the statistic v/_{-3 [(I+Rx.Y)_In(I+P◆]I Z0.-- In 1--Rx.y \1--oo/3' (7.66) which has approximately a standard Gaussian distribution if the hypothesis H0 is true and if the size n of the random sample is large enough. We reject H0 at significance level ot if and only if [Z0[> zu/2. Example 7.7.1 The following data are the maximum and minimum temperatures (in degrees Fahrenheit) registered during a winter week in an American city:

7.7 Correlation 323 Sun. Mon. Tue. Wed. Thu. Fri. Sat. Maximum 6 11 14 12 5 --2 --9 Minimum --22 --17 --15 --9 --24 --29 --35 Let Y be the maximum temperature and X be the minimum temperature. We compute Y _ --21.57 and -- 5.29. Moreover, 7 SSx = x -- 7. 2 463.71; i=1 7 SSr = y2 i -- 7 2 411.43; i=1 7 SSx, y -- xiYi -- 7Y5; -- 414.14. i=1 It follows that --SSx'YO.89 and /o=-/Y_24.55, SSx so that we estimate/*'lx := E[Y [X = x] by /2qx -- 24.55 + 0.89x. Remark. We could also compute *Xly - --26.89 + 1.01y. The sample correlation coefficient rx, is given by SSx, 0.9482. rx,, -- Since R 2 = r 2 0.8990, we say that the model explains about 89.9% of the x,Y variation in the data. Finally, to test Ho: Px, =0 against H" Px, 0, we calculate the statistic to := rx,, 6.67. /1 --r 2 x,Y Tab. 6.2 Given that t0.025,5 " 2.571, we can reject Ho at significance level e = 0.05 and conclude that the correlation coefficient ,ox, is not equal to zero.

324 7 Simple Linear Regression 7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises Solved Exercises Exercise no. 1 (7.2) Researchers think that there is a relationship between the strength Y (in tens of kilograms) of a certain metal part and the time x (in minutes) allowed for its cooling after its fabrication. They measured the strength of ten parts cooled over different time periods and obtained the results that are summarized below: 10 I(1 I(1 2:40.9, .:74.4, Zx/:17,077, Zy/:55,504, ZxiYi:30,436. i=l i:1 i=1 We propose the simple linear regression model: Y =/40 +/4ix + e, where e N (0, 0"2). a) Estimate the parameters/o and/ by the method of least squares. b) Estimate the parameter c) Test the hypothesis Ho:/I = 0 against H:/ :/: 0. Use a = 0.05. a) We have: IO /1 = i:I xiYi -- 102. El'__ / , X/2 --1022 and Solution 30, 436 - (10) (40.9) (74.4) 6.4 17,077 - (10) (40.9) 2 348.9 /o = . -/2 -- 74.4 - (0.0183) (40.9) _ 73.6498. b) We calculate and IO SSr = Zyi - 10. 2 = 55,504 - (10)(74.4) 2 = 150.4 i=1 SSR = x2 i _ 1022 a) / (348.9) -- 0.117. Then, we can write that SSr - SSR) I/2 3" = gO----- -- 4.33. 0.0183

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 325 c) We reject H0:/31 = 0 if and only if SSR b) 0.117 62 -- (4.33)2 -- 0.0062 > Fo.os, 1,10-2. p. 343 SinceF0.05,1,8 -- 5.32, we cannot reject H0 if c = 0.05. Exercise no. 2 (7.4) We study the tensile strength Y of a certain alloy as a function of the percentages Xl and x2 of two of its components. We took ten specimens of this alloy, that were produced under different conditions, and we obtained the following data: Y 213 220 216 225

235 218 239 243 233 240 Xl 13 15 14 18 19 20 22 17 16 18 x2 2.1 2.3 2.2 2.5 3.2 2.4 3.4 4.1 4.0 4.3 = -10 X 2 = 3028, Y.O 1X2i = 30.5 and Remark. We have: y.]� 1Xli 172, /-i=l li ..01Xi =99.65. a) We first propose the simple linear regression models: Y =/30 +/3ixi + e, where e N (0, or2), for i = 1,2. Test, at significance level a = 0.05, the null hypothesis Ho: /3i = 0 against Hi:/3i 0, for i = 1,2. b) Based on the quantities R 2 and 6 "2, which variable, Xl or x2, seems to better explain the tensile strength of the alloy? Remark. The smaller the value of 62 is, the more adequate the model seems to be for the data. c) We also consider the model Y = exp (/31Xl q- E), where e N (0, O'2). Use the data collected to estimate the parameter/31 by the method of least squares. Solution a) We first calculate SS, = 1105.6. In the case of the simple linear regression model with the variable Xl, we have: SSR = 21 Xi -- 10� : 69.6/12. \i=1 We find that/1 ' 2.221. It follows that SSR - 343.33 and SSE - 762.27 = 2 _ 762.27__ _ 95.3 = fo _ __343'33 8 95.3

326 7 Simple Linear Regression p. 343 Werereject Ho:/3 = 0at significancelevelot = 0.05 if and only if fo > Fo.o5,,o-2 5.32. Thus, we cannot reject Ho. When we consider the model with the variable x2, we obtain that SSR - 942.11, so that SSE - 163.49 : 3 -2 _ 20.4 := fo _ __942"11 _46.1. 20.4 Therefore, here we can reject Ho:/32 = 0 at significance level ot = 0.05, since 46.1 > 5.32. b) In the case of the variable x, we have: 6 .2 a) -- 95.3 and R2 SSR a) = -- 0.31, SSr while for the variable x2, we obtain that c3 "2 20.4 and R2 a) 942.11 -- -- _ 0.85. 1105.6 Given that R 2 is larger and that the value of 2 is smaller with the variable x2, it is x2 that better explains the tensile strength. c) We have: lnY=flx + e. We set 10 SS (1) = (lnyi - lXli) 2 . i=1 Then, we have: d 1o -�= I Xli lnyi d-SS (/1) = 0 ,� 2 Zi=I (lnyi - lXii) (-Xli) = 0 1 -- Z01 X21i We find that 10 10 Z Xli ln yi 934.5 and '-X21i = 3028. i=1 i=1 It follows that 934.5 " 0.31. 3028

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 327 Unsolved Problems Problem no. 1 The following (independent) data have been collected: 1 2 3 4 5 6 7 8 9 10 x 12 12 10 10 10 3 2 20 20 30 Y 20 40 30 80 50 50 90 30 40 40 We find that 10 10 10 xi 129, Zyi=470' Zx/2 i=1 i=1 i=1 10 -xiYi = 5250, SST = 4410. i=1 = 2301, 10 Z y/2 = 26, 500, i=1 We consider the model Y = flo+fllX +�, where � N(0, a2). a) What percentage of the total variation in the values of Y can be explained by a linear relationship between the variables Y and x? b) Can we reject the hypothesis

$H_0: \beta_1 -= 0$ (against $H_i: \beta \# 0$) if we choose $= 0.05$? Problem no. 2 On a Monday, a pisciculturist scatters a large quantity of baby fish into a fish pond. Each subsequent Monday, he catches at random 500 fish, marks them with a number (different every week), and then puts them back into the water. After a few hours, he throws a net, counts the total number of fish that he caught and the number of those marked with the number of the current week, before putting the fish back into the water. He obtains the following chronological table giving the number of marked fish and the total number of fish caught: Week 1 2 3 4 5 6 7 8 No. marked 17 19 18 23 24 26 27 29 No. caught 1004 1011 1008 1015 1003 1017 1003 1013 Remark. The precision of the calculations is important in this problem. a) Let P be the proportion of marked fish in the fish pond. The breeder considers the following model, from the fourth week: $P = $ ore t for t [4, 8]. Estimate the parameters and β by the method of least squares.

328 7 Simple Linear Regression b) By extrapolating theproposed model, what is the expected number of fish in the fish pond after 40 weeks? c) Assuming that the fish remain edible and that they gain 7% in weight every week, from the fourth week until maturity after 50 weeks, should the pisciculturist harvest the fish after eight weeks or wait for maturity? Remark. The value of the fish is proportional to the weight times the number of fish. Problem no. 3 (See [22].) An industrial engineer proposed the model $T = y/4 -k$ for the time T (in seconds) needed to accomplish a simple manual task, as a function of the number k of times the task has been practiced, where/4 and y are parameters that depend on the task and on the individual. We have the following data: T 22.4 2 ! .3 ! 9.7 15.6 ! 5.2 ! 3.9 13.7 k 0 1 2 3 4 5 6 a) Transform the proposed model into a linear model and estimate the parameters and g by the method of least squares. b) We suppose that $T = y/4-ke$, where $e \; N (0, \sigma^2)$. i) Estimate the parameter σ. ii) Test the hypothesis $H_0:/4 = 1$ against $H_i: /4 \; 1$ with a type I error risk oe equal to 0.05. iii) Calculate an approximate 95% confidence interval for g. Problem no. 4 We study the production Y (in cubic meters per second) of a manufacturing pro- cess, as a function of the temperature x (in degrees Celsius). For temperatures in- creasing by 100 degrees from 100 to 600 $\diamond$C, the production increased from 49 to 68. We have: 6 6 6 $\sum z y_i = 369$, $\sum y^2 = 22{,}947$ and $\sum x_i Y_i = 135{,}400$. i=1 i=1 i=1 a) We propose the simple linear regression model: $Y =/0 +/4_1 x + e$, where $e \; N (0, \sigma^2)$. i) Estimate the parameters/0 and/41 by the method of least squares. ii) Calculate the percentage of variation explained by the model.

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 329 iii) Test, at significance level $\alpha = 0.01$, the hypothesis $H0:,\beta_1 = 0$ against $1 \neq 0$. Hint. We have: $t_{0.005,4} -- 4.604$. b) We also propose the following curvilinear regression model: $Y=a_0+a_1/+e$, where $e \sim N(0, a_2)$. We then find that $/6=1$ Yi " 6847.934. i) Calculate the point estimates of the parameters α_0 and α_q by the method of least squares. ii) Calculate the percentage of variation explained by the model. Problem no. 5 A trucking company wishes to determine the relationship between the age of a truck and the number of days per year it spends being repaired. To do so, it took six tracks at random and obtained the following data: x 8 1 6 3 5 2 Y 9 16 1 4 0 10 where x is the age of the truck (in years) and Y is the number of days it spent being repaired over a one-year period. a) Without doing any calculations, explain why, among the three models that follow, model (2) seems the most appropriate: (1) (2) (3) where $e \sim N(0, \sigma^2)$. b) Based only on the coefficient of determination R^2, which model, between models (1) and (3) above, is the better one? $Y = \beta_0 + \beta_1 x + e$, $Y = \beta_0 + \beta_1 x^2 - 6$, $Y = \beta_0 + \beta_1 e^{-x} + e$, Multiple Choice Questions Question no. 1 We study the relationship between the number x of years elapsed since receiving their first degree and the annual salary Y (in thousands of dollars) of engineers. The values taken by the observations in a random sample of size $n = 10$ yielded the following table: x 1 2 3 4 5 Y 28 30 40 55 45 31 35 36 40 60

330 7 Simple Linear Regression We propose the model $Y = \beta_0 +\beta_1 x^2 + e$, where $e \sim N(0, \sigma^2)$. We then have: $SS_r = 1016$ and $\hat{\sigma}^2 -- 33.5$. A) Estimate the parameter β_0 by the method of least squares. a) 0 b) 1 c) 6.1 d) 21.7 e) 29 B) Calculate the percentage of variation explained by the proposed model. a) 54.2% b) 73.2% c) 73.6% d) 85.6% e) 85.8% C) Estimate the annual salary of an engineer having ten years experience, according to the proposed model. a) 100,000 b) 104,000 c) 105,000 d) 121,700 e) 129,000 Question no. 2 The real speed of three makes of cars, when their speedometers indicate 100 km/h, has been measured for four cars of each make: Make 1 Make 2 Make 3 105 103 108 101 101 97 99 104 102 96 98 103 We propose the model $V = \beta_0 + \beta_1 x + $, where V is the real speed, x is the speed indicated by the speedometer and $e \sim N(0, \sigma^2)$. Estimate the parameter β_1 by the method of least squares. a) 0 b) 1 c) 1.1 d) e) indeterminate Question no. 3 We are interested in the relationship between two random variables, X and Y. We collected five pairs (x_i, Y_i) of particular observations of (X, Y). The results are summarized below: 5 5 5 5 5

Zxi =5, Zyi: 15, Zx7:55, Zy/2=51, Zxiyi=51. i:1 i=1 i:1 i:1 i=1 A) Calculate the sample correlation coefficient, rx, y. a) 0 b) 0.194 c) 0.4 d) 0.775 e) 0.8 B) Suppose that X N(#x, rx22) and Y N(#r', rr22). Estimate the mean of Y, given that X = 0. a)-3 b) 0 c) 1.2 d) 3 e) 4.8 C) If, with other data, we obtained R 2 = 0.04 and/ > 0, what is then the value of the statistic used to test the hypothesis H0: p = 0 against H: p - 0 ? a)-0.577 b)-0.069 c) 0.069 d) 0.354 e) 0.577

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 331 Question no. 4 We consider the following data: x 1 1.5 2 2.5 3 Y 13 16 18 18 19 We propose the curvilinear regression model Y = fl0 +/1 lnx + s, where s N(0, a2). We calculate /5=llnX i-3.1135, Z=lYi=84, Z=lYilnxi - -56.3309, -/5___ 11n 2 xi -- 2.6914, -=1Y] = 1434. A) Estimate the parameter fll by the method of least squares. a) 2.80 b) 5.35 c) 11.20 d) 13.47 e) 17.12 B) Calculate the value of the statistic F0 used to test the null hypothesis H0: fll = 0. a) 4.11 b) 16.92 c) 50.77 d) 72.30 e) 75.20 C) What is the percentage of variation explained by the model? a) 5.58% b) 12.62% c) 80.0% d) 88.22% e) 94.42% Supplementary Exercises Question no. 1 Preliminary calculations made with a series of particular observations (xi, Yi), i = 1 ..... 100, of a random variable Y and a deterministic variable x yielded lOO z]O01Yi = .di=l i i=l xi = 500, 1000, 100 x 2 = 24, 775, =1 (Yi -)2 = 39,600, rx,r = 0.9, where rx,r is the sample "correlation coefficient" defined by]001 (Xi -- -)(Yi -- f) rx, Y 2. 2 -, 100. y)2] 1/2" []001(Xi--)2_,i=ltYi- A first look at the graph of these data points suggest a simple linear regression model: Y = flo +/1X n t- 6, where e N (0, a2).

332 7 Simple Linear Regression a) Calculate y] (xi - ◆) (Yi -). b) Estimate all the parameters in the model. c) Perform an analysis of the variance and test the null hypothesis H0:/31 = 0 at significance level ot = 0.05. Hint. We have: F0.05,1,98 -- 3.94. Question no. 2 We have the following table showing the progression of an epidemic over time: t 10 12 15 20 23 25 27 30 N 40 35 30 25 20 20 15 10 where t denotes the number of weeks elapsed since the outbreak of the epidemic and N is the number (in tens) of persons infected. We wish to forecast the value of N after 40 weeks. We propose the following model: N=flO+/31t+e, where e " N (0, or2). We find that n = 8, Y8i= ti = 162, Y= t/2 = 3652, Y=I ni = 195, Y/8=l n/2 = 5475, y/8=l tini = 3435. a) Estimate all the parameters in the model and calculate the table of analysis of variance. b) Give the forecasted value of N after 40 weeks. c) Reminder. We define a

100(1 - or) % one-sided confidence interval with an upper bound for a parameter 0 by the determination of a statistic UC such that P[t9 < UC]= 1-or. Calculate a 99% one-sided confidence interval with an upper bound for the value of N after 40 weeks. Hint. We have: t0.0,6 -- 3.143. Question no. 3 We are interested in the relationship between the speed x (in kilometers per hour) reached in 10 seconds by 15 luxury cars and the braking distance Y (in meters) until a complete stop. The following data have been collected: x 104.2 106.1 105.6 106.3 101.7 104.4 102.0 103.8 Y 39.8 40.4 39.9 40.8 33.7 39.5 33.0 37.0 x 104.0 101.5 101.9 100.6 104.9 106.2 103.1 Y 37.0 33.2 33.9 29.9 39.5 40.6 35.1

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises a) Estimate the parameters in the model Y =/30 +/31 x + , where -- N(0, a2). 333 b) Calculate the table of analysis of variance and test the null hypothesis H0:/31 = 0 at significance level t = 0.05. c) Calculate 95% confidence intervals for the coefficients fl0 and ill. d) Perform an analysis of the residuals to check whether the basic assumptions are satisfied. e) Calculate a 95% prediction interval for Y when x = 100,103 and 106. Hint. We have: t0.025,13 " 2.160. Question no. 4 The data that follow show the effect of time on the content of hydrogen for two (independent) steel specimens stored at 20 �C: t 1 2 6 17 30 H 7.7 7.5 6.1 5.7 4.2 8.4 8.1 6.8 5.3 4.5 where t denotes the time (in hours) and H is the content of hydrogen (in parts per million). We propose the model H=o+/311nt+e, where e " N (0, r2). a) Estimate all the parameters in the model. b) Calculate the table of analysis of variance and test the null hypothesis H0:/3 = 0 at significance level o = 0.05. Question no. 5 We seek to establish a relationship between the diameter x (in centimeters) of the filament of an electric light bulb and its lifetime Y (in hours). We have the following data: x 0.15 0.20 0.25 0.30 0.40 0.50 0.60 0.60 Y 120 165 204 238 296 373 410 403 x 0.60 0.70 0.70 0.80 0.80 0.80 0.90 1.00 Y 420 462 455 520 518 525 580 600 a) Estimate the parameters of the simple linear regression model. b) Calculate the analysis of variance table and test the null hypothesis H0:/31 = 0 at significance level ot = 0.05. c) Perform an analysis of the residuals to check whether the basic assumptions are satisfied.

334 7 Simple Linear Regression Question no. 6 The law of perfect gases relating the pressure P (in kilograms per square cen- timeter) and the volume v (in cubic centimeters): Pv � = c can be written, after transformation, in the

form Y ---- c - Fx, where $Y := \ln P$, $x := \ln v$ and $c := \ln c$. Particular observations of v and P yielded (v, P) ----(50,64.7), (60,51.3), (70,40.5), (90, 25.9), (100,7.8). Remark. Useful calculations: -'1 xi 21.35984, z I(X) i= Yi -- 17.11712,]x/2 291.57317, E] y/2 61.40147. 1 .,ciYi -- 72.26249, a) Estimate the parameters a and F by the method of least squares. b) Calculate the analysis of variance table of the model $Y = a - gx$, as well as the percentage of variation explained. c) Calculate a confidence interval at 95% for the average value of P when v = 80. Question no. 7 The regression model through the origin is given by Y=flx+, whereN(0, cr2). a) Find the estimator of the parameter fl by the method of least squares, if we have n observations of the pair (x, Y). b) We suppose that a recently built structure sinks into the ground according to the model $Z = 10 - 10e-Cx$, where x is the age (in months) of the structure and Z is the sinking (in centimeters). Find a transformation that reduces the above model to the regression model through the origin. Question no. 8 To evaluate a measurement process, an operator who knows the measuring in- strument well obtains two measurements, X and Y, of 15 parts in a batch. A mea- surement process is said to have a good repeatability if the correlation coefficient, rx, y, of the random sample of the pair (X, Y) is greater than 0.80. The data are the following:

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 335 X 34 56 6 50 33 43 49 17 54 24 35 46 10 51 25 Y 45 44 19 55 17 32 37 5 54 18 26 43 16 55 11 Remark. Useful calculations: Y = 35.53, = 31.8, sx -- 16.20 and s- _ 17.01. a) Does the measurement process have a good repeatability? b) Calculate the covariance sx,v := 5= 1 (xi - Y) (Yi -- 3) of the particular obser- vations of X and Y. c) We consider the pairs (X1, Y1) ..... (X15, Y15), (Y1, X1) ..... (Yl, They constitute a random sample of size n = 30 of a pair of random variables that we denote by (U, V). i) Explain why the standard deviation Su of the observations of U and that of the observations of V are equal. ii) We assume that U and V have a Gaussian distribution, so that E[V I U = u] = 13o + 131 u. Estimate, using the 30 particular observations of (U, V), the parameters 130 and 131 by the method of maximum likelihood. Question no. 9 During the 1840s, the physicist J. Forbes measured in 17 places, in the Alps and in Scotland, the barometric pressure Y (in inches of mercury) and the boiling temperature x (in degrees Fahrenheit) of water. At the time, barometers were fragile and it was much easier to determine the boiling temperature of water in mountainous regions than to use a barometer. From this

temperature, the barometric pressure could be determined. The data collected by Forbes are presented below: x 194.5 194.3 197.0 198.4 199.4 199.9 Y 20.79 20.79 22.40 22.67 23.15 23.35 x 200.9 201.1 201.4 201.3 203.6 204.6 Y 23.89 23.99 24.02 24.01 25.14 26.57 x 209.5 208.6 210.7 211.9 212.2 Y 28.49 27.76 29.04 29.88 30.06 a) Draw the graph of the data points (xi, Yi) and that of the data points (xi, In Yi). Which graph looks more like a straight line? b) Perform a regression analysis according to the simple linear regression model. That is, estimate the parameters in the model, calculate the analysis of variance table and perform an analysis of the residuals.

336 7 Simple Linear Regression c) Perform a regression analysis of the model In Y = '0 + q x + , where N(0, r2). d) Which model is superior, based on the percentage of variation explained by the model and on the behavior of the residuals? Question no. 10 Let (x, y) ..... (xm, yo) be ten particular observations of a pair of random variables (X, Y) We have: K-'l◆ x. = z]OI Yi = 55, SX = Sy = 3.03 and ◆ /4=1 t z]O I xi y i = 380. a) Calculate the sample correlation coefficient, rx, r. b) Calculate -.]◆ 1 x/2 - 1022. Question no. 11 An experiment similar to that of Forbes, described in no. 9, was realized by J. Hooker in the Himalayas. His data are the following: x Y x Y x Y 210.8 29.211 210.2 28.559 208.4 27.972 202.5 24.697 200.6 23.726 200.1 23.369 199.5 23.030 197.0 21.892 196.4 21.928 196.3 21.654 195.6 21.605 a) Perform a regression 193.6 21.212 193.4 20.480 191.4 19.758 191.1 19.490 190.6 19.386 189.5 18.869 188.8 18.356 188.5 18.507 186.0 17.221 185.7 17.267 analysis of the model 185.6 17.062 184.6 16.881 184.1 16.959 184.1 16.817 183.2 16.385 182.4 16.235 181.9 16.106 181.9 15.928 181.0 15.919 180.6 15.376 In Y = go +)/1 x n t- 5, where N(0, r2). b) Compare the results to those in part c) of exercise no. 9. c) Perform a regression analysis of the model in a), using the 48 data of Hooker and Forbes together. Question no. 12 In a simple linear regression problem, we find that the parameter fll is not signif- icantly different from zero. a) We then propose the model Y =/30 + e, where e N (0, r2). Estimate/30 by the method of least squares.

7.8 Exercises, Problems, and Multiple Choice Questions, Supplementary Exercises 337 b) We also propose the model Y = /50 +/51x 2 + 6:. Find the value/o of/50 that minimizes the function SS(flo,/1) := in=l(Yi -/50 -/lXi2) 2. Question no. 13 The voltage U (in volts) of a capacitor, with initial charge uo, is, after t seconds, given by the equation U = uo e-t. An experiment yielded the

following results: t 0 1 2 3 4 5 6 7 8 9 10 U 100 75 55 40 30 20 15 10 10 5 5 a) Make a transformation to obtain a linear model with respect to the parameters. b) Fit the new model by the method of least squares. c) Calculate 95% confidence intervals for u0 and ft. Question no. 14 We propose the model $Y = flo +/51X + 6:$, where $6: N (0, a2)$, relating the variables x and Y. We have the following results: $--=1Yi = 10$, $--=1Y/ = 50$ and the sum of squares SSn is equal to 25. a) Calculate the coefficient of determination, R 2. b) Calculate 9a. Question no. 15 Boyle's law relates the pressure P of a gas to the volume v it occupies, according to the equation Pv a =fl, where a and/5 are two constants. Fit the model using the following data: v 0.2 1.0 0.8 1.0 1.6 0.4 P 0.6 1.0 1.5 2.0 2.5 3.5 Question no. 16 We consider the model $Y =/5ox o$. a) Make a transformation to linearize the model. Express the new variables y1 and x I, as well as the new parameters/56 and/511, in terms of the old ones. b) An analysis of variance table yielded a statistic fo equal to 12. Can we then reject, at significance level a = 0.01, the hypothesis $Ho:/5I 1 = 0$, if we have ten particular observations of (x, Y)? Give the value of the percentile used to perform the test. Hint. We have: to.oo5,8 -- 3.355.

338 7 Simple Linear Regression Question no. 17 (See [2].) A screw manufacturer wishes to inform its customers about the relationship be- tween the nominal length x and the actual length Y of its screws (in inches). The following values, taken by independent observations of Y, have been collected: x 1/4 1/4 1/4 1/2 1/2 1/2 Y 0.262 0.262 0.245 0.496 0.512 0.490 x 3/4 3/4 3/4 1 1 1 Y 0.743 0.744 0.751 0.976 1.010 1.004 x 5/4 5/4 5/4 3/2 3/2 3/2 Y 1.265 1.254 1.252 1.498 1.518 1.504 x 7/4 7/4 7/4 2 2 2 Y 1.738 1.759 1.750 2.005 1.992 1.992 We propose the simple linear regressionmodel $Y = lo +ill x +$, where $N(O, a2)$. a) Estimate all the parameters in the model. b) Obtain a 95% confidence interval for the average value of Y when x = 1, using only the three particular observations of Y for this value of x. c) Obtain a 95% confidence interval for the average value of Y when x = I, using all the data available. d) Calculate the "correlation coefficient" (see no. 1 above) of the pairs of data (x, y).

Appendix A: Mathematical Formulas Logarithms lnab=lna+lnb; lna/b=lna-lnb; lna b=blna Geometric Series If [r[< 1, we find that Z ark _ a . Z ak r k ar k=0 1 r k=o (1 __ -)2 ' = ar_ ; Zark_ a(1-r n+l) Z ark 1 r 1 - r k:! k=O Limits L'Hospital's rule. Suppose that limxxo f(x) = limxxo g(x) = 0, or that limxxo

f(x) = limxxo g(x) = ◆xz. Then, under some conditions, we can write that f(x) f'(x) lim -- lim -o g(x) xxo g'(x)" If the functions if(x) and gt(x) satisfy the same conditions as f(x) and g(x), we can repeat the process. Moreover, the constant x0 may be equal to Derivatives Derivative of a quotient: d f(x) dx g(x) g(x)f'(x) -- f(x)g'(x) g2(x) if g(x) : 0.

340 Appendix A: Mathematical Formulas Remark. This formula can also be obtained by differentiating the product f(x)h(x), where h(x) := 1/g(x). Chain rule. Ifu = g(x), then we have: d d du -x f(u) = -u f(u)" -x = f'(u)g'(x) = f'[g(x)]g'(x). For example, if u = x 2 , we calculate d-exp(u 2) exp(uZ)2u 2x 4x " exp(x4). Integrals Integration by parts: f udv = uv - f vdu. Integration by substitution. If the inverse function g-1 exists, then we can write that f(x) dx = f[g(y)]g'(y) dy, where a = g(c) e* c = g-(a) andb = g(d) e* d = g-(b). For example, dx " = 2ye y dy = 2yeYl- 2 e y dy = 2e 2+2. In this example, y = g-(x) = x /2 andx = g(y) = y2.

Appendix B: Quantiles of the Sampling Distributions Gaussian distribution ot 0.25 O. 10 0.05 0.025 0.01 0.005 0.001 0.0005 z 0.674 1.282 1.645 1.960 2.326 2.576 3.090 3.291 Student distribution n 1 2 3 4 5 6 7 8 to.o25,n 12.706 4.303 3.182 2.776 2.571 2.447 2.365 2.306 to.os,n 6.314 2.920 2.353 2.132 2.015 1.943 1.895 1.860 n 9 10 15 20 25 30 40 to.o25,n 2.262 2.228 2.131 2.086 2.060 2.042 2.021 1.960 to.os,n 1.833 1.812 1.753 1.725 1.708 1.697 1.684 1.645 Chi-square distribution n 1 2 3 4 5 6 7 8 9 2 5.02 7.38 9.35 11.14 12.83 14.45 16.01 17.53 19.02 XO.O25,n 2 3.84 5.99 7.81 9.49 11.07 12.59 14.07 15.51 16.92 XO.O5,n 2 0 + 0.10 0.35 0.71 1.15 1.64 2.17 2.73 3.33 XO.95,n 2 0 + 0.05 0.22 0.48 0.83 1.24 1.69 2.18 2.70 XO.975,n

342 Appendix B: Quantiles of the Sampling Distributions Cbi-square distribution (continued) n 10 15 20 25 30 40 50 100 2 20.48 27.49 34.17 40.65 46.98 59.34 71.42 129.56 X0.025,n 2 Xo.os,n 18.31 25.00 31.41 37.65 43.77 55.76 67.50 124.34 2 3.94 7.26 10.85 14.61 18.49 26.51 34.76 77.93 XO.95,n 2 3.25 6.27 9.59 13.12 16.79 24.43 32.36 74.22 XO.975,n Fisher distribution n 1 2 3 4 5 6 7 8 Fo.o25a. 647.8 39.00 15.44 9.60 7.15 5.82 4.99 4.43 Fo.05.,., 161.4 19.00 9.28 6.39 5.05 4.28 3.79 3.44 n 9 10 15 20 30 60 120 c Fo.o25.nm 4.03 3.72 2.86 2.46 2.07 1.67 1.43 1.00 Fo.oS.n., 3.18 2.98 2.40 2.12 1.84 1.53 1.35 1.00 Values of F0.o25,nl.n2 n2\nl 1 2 3 4 5 6 7 8 9 10 11 12 648 800 864 900 922 937 948 957 963 969 973 977 2 38.5 39.0 39.2

39.2 39.3 39.3 39.4 39.4 39.4 39.4 39.4 39.4 3 17.4 16.0 15.4 15.1 14.9 14.7 14.6 14.5 14.5 14.4 14.4 14.3 4 12.2 10.6 9.98 9.60 9.36 9.20 9.07 8.98 8.90 8.84 8.79 8.75 5 10.0 8.43 7.76 7.39 7.15 6.98 6.85 6.76 6.68 6.62 6.57 6.52 6 8.81 7.26 6.60 6.23 5.99 5.82 5.70 5.60 5.52 5.46 5.41 5.37 7 8.07 6.54 5.89 5.52 5.29 5.12 4.99 4.90 4.82 4.76 4.71 4.67 8 7.57 6.06 5.42 5.05 4.82 4.65 4.53 4.43 4.36 4.30 4.24 4.20 9 7.21 5.71 5.08 4.72 4.48 4.32 4.20 4.10 4.03 3.96 3.91 3.87 10 6.94 5.46 4.83 4.47 4.24 4.07 3.95 3.85 3.78 3.72 3.66 3.62 11 6.72 5.26 4.63 4.28 4.04 3.88 3.76 3.66 3.59 3.53 3.47 3.43 12 6.55 5.10 4.47 4.12 3.89 3.73 3.61 3.51 3.44 3.37 3.32 3.28

Appendix B: Quantiles of the Sampling Distributions Values of fo.o5,ni,n2 343 n2\nl 1 2 3 4 5 6 7 8 9 10 11 12 1 161 200 216 225 230 234 237 239 241 242 243 244 2 18.5 19.0 19.2 19.2 19.3 19.3 19.4 19.4 19.4 19.4 19.4 19.4 3 10.1 9.55 9.28 9.12 9.01 8.94 8.89 8.85 8.81 8.79 8.76 8.74 4 7.71 6.94 6.59 6.39 6.26 6.16 6.09 6.04 6.00 5.96 5.94 5.91 5 6.61 5.79 5.41 5.19 5.05 4.95 4.88 4.82 4.77 4.74 4.70 4.68 6 5.99 5.14 4.76 4.53 4.39 4.28 4.21 4.15 4.10 4.06 4.03 4.00 7 5.59 4.74 4.35 4.12 3.97 3.87 3.79 3.73 3.68 3.64 3.60 3.57 8 5.32 4.46 4.07 3.84 3.69 3.58 3.50 3.44 3.39 3.35 3.31 3.28 9 5.12 4.26 3.86 3.63 3.48 3.37 3.29 3.23 3.18 3.14 3.10 3.07 10 4.96 4.10 3.71 3.48 3.33 3.22 3.14 3.07 3.02 2.98 2.94 2.91 11 4.84 3.98 3.59 3.36 3.20 3.09 3.01 2.95 2.90 2.85 2.82 2.79 12 4.75 3.89 3.49 3.26 3.11 3.00 2.91 2.85 2.80 2.75 2.72 2.69

Appendix C: Classification of the Exercises The list of exercises (among the multiple choice questions, problems or supplemen- tary exercises) that can be done after having read each section is given below. The symbol 1 m (respectively 1 s), for example, denotes the multiple choice question (resp. supplementary exercise) no. 1 of the chapter in question. Chapter 2 2.1: lm, 6m, 12m, 19m, 32m 2.2: 2; 8m, 13m, 36m, 46m 2.3: 3, 8, 11, 18, 20; 2m, 7m, 1 lm, 14m, 15m, 21m, 24m, 29m, 34m 2.4: 1,5, 24; 17m, 23m, 31m, 39m, 40m, 44m, 48m, 50m 2.5: all the others Chapter 3 3.1: -- 3.2: 3m, 15m, 52m 3.3: 1,2, 10, 11,42; 4m, 16m, 28m, 34m, 35m, 36m, 53m 3.4: 12, 19, 31,34; lm, 12m, 17m, 18m, 29m, 37m, 56m, 79m 3.5: 14, 27, 38; 6m, 19m, 20m, 41m, 58m, 64m, 80m 3.6: 6, 35; 21m, 40m, 47m 3.7: 3, 4, 5, 7, 9, 15, 16, 17, 20, 22, 23, 25, 28, 29, 32, 37, 39, 43, 46, 47; 2m, 5m, 22m, 23m, 30m, 32m, 38m, 39m, 45m, 46m, 48m, 49m, 54m, 55m, 57m, 59m, 66m, 67m, 68m, 72m, 73m, 74m, 76m, 77m, 78m 3.8: 8, 13, 18, 21,24, 26, 30, 33, 36, 40, 41,44, 45; 7m, 8m, 9m, 10m, llm, 13m, 14m, 24m, 25m, 26m, 27m, 31m, 33m, 42m, 50m,

60m, 62m, 63m, 70m 3.9: 43m, 44m, 51m, 61m, 65m, 69m, 71m, 75m, 81m Chapter 4 4.1: -- 4.2: 1,2, 12; 3m, 4m, 9m, 17m, 20m, 21m, 26m, 31m, 32m 4.3: 3, 14; 10m, 22m, 33m, 40m

346 Appendix C: Classification of the Exercises 4.4: l I m 4.5: 4, 5, 19; 12m, 28m, 34m 4.6: 6, 16, 22, 24, 25, 27, 29; I m, 5m, 23m, 27m, 35m, 42m 4.7: 15, 17; 6m,24m 4.8: 7,20,21; 13m, 30m,47m 4.9: 8, 9, 28; 7m, 18m, 29m, 36m 4.10:16m 4.11: 10, 11, 12, 13, 18, 23, 26; 2m, 8m, 14m, 15m, 19m, 25m, 37m, 38m, 39m, 41 m, 43m, 44m, 45m, 46m, 48m Chapter 5 7.1: 2m 7.2: I s 7.3: 2s, 17s 7A: 1; 14s 7.5: 3s, 5s 7o6: 2, 3, 4, 5; lm, 4m; 4s, 6s, 7s, 9s, 1 Is, 12s, 13s, 15s, 16s 7,7: 3m; 8s, 10s Chapter 7 5.1: -- 5.2: 3; 2m, 6m, 16m, 19m 5.3: 8; 3m, l lm, 14m,23m,27m,29m 5A: 1,4; I m, 4m, 7m. 17m, 20m. 24m, 25m, 26m, 28m, 30m 5.5: 7; 5m, 12m, 18m 5.6: 8m, 21 m 5.7: 2,5,6; 9m, 10m, 13m, 15m, 22m Chapter 6 6.1: 1,6,8; 2m, Tm, llm, 13m, 15m 6.: 2,5,7; lm,3m,4m,gm, 12m, 14m, 23m; Is, 2s,4s 6.3: 3, 4, 9; 5m, 6m, 8m, 10m, 18m, 21 m, 25m, 27m 6A: 10-19; 16m, 17m, 19m,20m,22m,24m,26m,28m; 3s,5s-I Is

Appendix D: Answers to the Multiple Choice Questions Chapter 2 la;2d;3b;4d;5c;6c;7a;8c;9e;lOc;lld;12c;13b;14a;15d;16c;17d; 18 c ,e ,b ,b ,c ,d ; 19 b,e; 20 d,c; 21 d,e; 22 d,a; 23 e,c; 24 a,b; 25 d,a; 26 b,c; 27 b,c,c,e; 28 d,a,a,c; 29 a,c,b; 30 e,b,a; 31 e,d,c; 32 b; 33 b; 34 d,c; 35 c,e,a,a; 36 c,a; 37 e,a,c,e; 38 b,d; 39 d,e; 40 c,b; 41 b,a,c,d; 42 c,e; 43 e,a; 44 d,d; 45 b,a; 46 e,c; 47 b,b; 48 d,d; 49 e,a; 50 b,d,a,e; 51 c,b,d,e. Chapter 3 1 b,c,b,d; 2 a,e,c,d; 3 c; 4 e; 5 a; 6 d; 7 d; 8 b; 9 c; 10 e; 11 c; 12 e,d,c; 13 b,d,b; 14 d,e,d; 15 e; 16 c; 17 e; 18 b; 19 d; 20 b; 21 a; 22 d; 23 b; 24 e; 25 a; 26 e,c,e,b,a,c; 27 d,d,b,a,d,a; 28 b,d,c; 29 a,d,a; 30 b,e,a; 31 d,b,e,c; 32 c,d,d; 33 a,b,e; 34 c; 35 d; 36 b; 37 b; 38 e; 39 d; 40 a; 41 a; 42 e; 43 c; 44 b; 45 d,c; 46 a,b; 47 d,b; 48 c,e; 49 b,a; 50 e,a; 51 e,c; 52 b; 53 c; 54 e,c,e; 55 d,b; 56 a; 57 d; 58 a,e; 59 a; 60 b; 61 c; 62 c,d,a,e; 63 d,b,d; 64 e,c,b; 65 c,e,a,b; 66 c,b,d; 67 e,d,b,c; 68 e,a,e,b; 69 d,a,a; 70 d,a,d,c; 71 e,a,d,e; 72 b,b,c; 73 b,c,e; 74 c,a,c; 75 c,d,a,b; 76 c,a,d,e; 77 e,b,d; 78 d,a,d; 79 a,c,e; 80 c,b,e,e; 81 b,d,c,b. Chapter 4 1 a,b,d,c,a; 2 d,a,e; 3 b; 4 c; 5 b; 6 d; 7 d; 8 e; 9 c; 10 b; 11 b; 12 d; 13 e; 14 b; 15 a; 16 b,c,c; 17 b; 18 c; 19 c; 20 d; 21 b; 22 a; 23 e; 24 c; 25 a; 26 c; 27 a; 28 a; 29 c; 30 e; 31 c; 32 e; 33 d; 34 a; 35 b; 36 a; 37 b; 38 a,e,b,b,c,e; 39 b,a,c,d,d,b; 40 d,d,b; 41 a,c; 42 d,a,a; 43 c,e; 44 e,a,a,d,e,a; 45 d,b,e,a,c; 46 a,c,d,e; 47 d,c,a; 48 b,e. Chapter 5 1 d,e,d,c; 2 c; 3 e; 4 a; 5 d; 6 e; 7 e; 8 d; 9 a; 10 b,a,e; 11 d; 12 c; 13 a; 14 e; 15 d; 16 d; 17 b; 18 d; 19 c; 20 d; 21 d; 22 a; 23 a,d; 24 c,e;

25 e,c; 26 b,b; 27 c,b; 28 e,b; 29 c,a; 30 c,d.

348 Appendix D: Answers to the Multiple Choice Questions Chapter 6 1 e,e,c; 2 e; 3 a; 4 c; 5 d; 6 e,b; 7 a; 8 b; 9 b; 10 e; 11 b; 12 e; 13 c; 14 e; 15 c,d; 16 c,b,a; 17 e,c,a,d,b; 18 e,a; 19 d,b,e; 20 e,c,c,a,d; 21 e,b; 22 d,c,b; 23 c,c; 24 a,d,b; 25 c,b,e; 26 b,d,e; 27 e,c; 28 b,d,a. Chapter 7 1 e,c,e; 2 e; 3 d,c,d; 4 b,c,e.

Appendix E: Answers to Selected Supplementary Exercises Chapter 6 5. a) H1 : # < 240; b) - 1.27; c) 0.32; d) 82. 6. a) Test with paired observations; to -- 3.23; b) 6. 7. a) 0.1; b) 0.6. 2 do not reject Ho: o-2 > (1.5)2; b) 0.21; c) 40. 8. a) Wo 2 -- 23.5 < Xo.o5,19 := we 9. a) fo --7.14=weacceptHl:o-a 2 7e 2. � o-e, td -- --1.67 := we do not reject Ho; b) we do not reject Ho: o-a 2 = o-2; to -- -- 1.46 := we do not reject Ho. 10. a) to - 1.81 < to.ol,6 := we do not reject Ho: # = 0.07; b) the formula with known o- gives 63 or 64 (exact answer = 66); c) the formula with known o- gives 0.9985 (more precise answer = 0.95). 11. a) Xi = diameter measured with calipers i; we assume that Xi N(#i, o'/2); we perform a test with paired observations; to -- 0.43 := we do not reject Ho: #1 = #2; 2 . b) Wo 2 16.24 < Xo.lo,11 := we do not reject Ho: o'2 = (0.001)2 Chapter 7 2. a) /o - 52.38; /1 - -1.383; 62 _ 1.89; fo - 375.93; b) (40) _ -2.94 < 0 0; c) [0, 3.43] (approximately). 4. a)]o - 8.31;/1 - -1.08; 6 "2 0.166; b)fo - 113.55 > 5.32; we reject Ho. 6. a) -- 14.76; 2.65; b) fo -- 13.4; R 2 -- 0.81; c) [12.3, 44.3] (approximately). 8. a) rx,y 0.83 =: yes; b) 230; c) i) the observations of U and of V are the same, only the order of the observations is different; ii)/o -- 6.4;/1 -- 0.81. 10. a) 0.94; b) 82.6. 12. a) 1; b) 17" -/1 X/2 Ft " i=l 14. a) 0.83; b) 1.5. 16. a) Y' = lnY; x' = lnx; = lno; tl = o; b) 12 > Fo.oA,8 = to2.oo5,8 -- 11.26; we reject Ho.

Bibliography 1. Barnes, J. Wesley, Statistical Analysis for Engineers: A Computer-Based Approach, Prentice-Hall, Englewood Cliffs, New Jersey, 1988. 2. Bowker, Albert H., and Lieberman, Gerald J., Engineering Statistics, Prentice-Hall, En- glewood Cliffs, New Jersey, 1959. 3. Breiman, Leo, Probability and Stochastic Processes: With a View Toward Applications, Houghton Mifflin, Boston, 1969. 4. Cartier, Jacques; Parent, R6gis; and Picard, Jean-Marc, Inference Statistique, ditions Sciences et Culture, Montreal, 1976. 5. Chung, Kai Lai, Elementary Probability Theory with Stochastic Processes, Springer- Verlag, New York, 1975. 6. C16ment, Bernard, Analyse Statistique,

Tome I and II, Lectures notes, cole Polytech- nique de Montr6al, 1988. 7. Dougherty, Edward R., Probability and Statistics for the Engineering, Computing, and Physical Sciences, Prentice-Hall, Englewood Cliffs, New Jersey, 1990. 8. Feller, William, An Introduction to Probability Theory and Its Applications, Volume I, 3rd Edition, Wiley, New York, 1968. 9. Feller, William, An Introduction to Probability Theory and Its Applications, Volume II, 2nd Edition, Wiley, New York, 1971. 10. Hastings, Kevin J., Probability and Statistics, Addison-Wesley, Reading, Massachusetts, 1997. 11. Hines, William W., and Montgomery, Douglas C., Probability and Statistics in Engineer- ing and Management Science, 3rd Edition, Wiley, New York, 1990. 12. Hogg, Robert V., and Craig, Allen T., Introduction to Mathematical Statistics, 3rd Edition, Macmillan, New York, 1970. 13. Kallenberg, Olav, Foundations of Modern Probability, 2nd Edition, Springer-Verlag, New York, 2002. 14. Kre, Paul, Introduction aux Math6matiques et g leurs Applications Fondamentales, Dunod, Paris, 1969. 15. Leon-Garcia, Alberto, Probability and Random Processes for Electrical Engineering, 2nd Edition, Addison-Wesley, Reading, Massachusetts, 1994. 16. Lindgren, Bernard W., Statistical Theory, 3rd Edition, Macmillan, New York, 1976. 17. Maksoudian, Y. Leon, Probability and Statistics with Applications, International, Scran- ton, Pennsylvania, 1969.

352 Bibliography 18. Miller, Irwin; Freund, John E.; and Johnson, Richard A., Probability and Statistics for Engineers, 4th Edition, Prentice-Hall, Englewood Cliffs, New Jersey, 1990. 19. Papoulis, Athanasios, Probability, Random Variables, and Stochastic Processes, 3rd Edi- tion, McGraw-Hill, New York, 1991. 20. Parzen, Emmanuel, Stochastic Processes, Holden-Day, San Francisco, 1962. 21. Roberts, Richard A., An Introduction to Applied Probability, Addison-Wesley, Reading, Massachusetts, 1992. 22. Ross, Sheldon M., Introduction to Probability and Statistics for Engineers and Scientists, Wiley, New York, 1987. 23. Ross, Sheldon M., Introduction to Probability Models, 7th Edition, Academic Press, San Diego, 2000. 24. Whittle, Peter, Optimization Over Time, Volume I, Wiley, Chichester, 1982.