



Milovan Krnjajić

INTRODUCTION TO PROBABILITY AND STOCHASTIC PROCESSES

Exercises and Solutions

**Singidunum University
Belgrade**

Milovan Krnjajić

**INTRODUCTION TO
PROBABILITY AND STOCHASTIC
PROCESSES**

Exercises and Solutions

First edition

Belgrade, 2020.

INTRODUCTION TO PROBABILITY AND STOCHASTIC PROCESSES - Exercises and Solutions

Author:

Milovan Krnjajić

Reviewers:

Prof. Milan Milosavljević,
Singidunum University, Belgrade, Serbia

Dr. Stanislav Mišković,
Principal Data Scientist, San Jose CA, USA

Publisher:

SINGIDUNUM UNIVERSITY
32 Danijelova Street, Bgrade
www.singidunum.ac.rs

For publisher:

Milovan Stanišić, PhD

Prepress:

Milovan Krnjajić
Miloš Višnjic

Design:

Aleksandar Mihajlović, MA

Year:

2020

Circulation:

10

Printed by:

Caligraph, Belgrade

ISBN: 978-86-7912-739-6

Copyright © 2020 Singidunum University, Belgrade

All rights reserved. No part of this work covered by the copyright herein may be reproduced, transmitted, stored or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, without the prior written permission of the publisher.

Contents

Preface	V
Exercises in Probability I	1
Exercises in Probability II	11
Exercises in Probability III	23
Exercises in Stochastic Processes	33
Solutions to Exercises in Probability I	41
Solutions to Exercises in Probability II	59
Solutions to Exercises in Probability III	85
Solutions to Exercises in Stochastic Processes	113
Bibliography	125

Preface

The prologue of a well known book on probability¹ cites Edward Davenant (a XVII century English clergyman and academic) who reportedly said that he “would have a man *knockt* in the head who should write anything in *Mathematiques* that had been written before”. The question “what’s new here then?” immediately arises for anyone who attempts to write on a standard academic subject. Given the abundance of study material available online, this question remains especially relevant today.

This solution manual is based on the set of problems the author collected during many years of his professional work on statistical modelling and analysis, both in industry and academia. Some exercises in particular come from the material he used while teaching the undergraduate courses in Probability Theory and Stochastic Processes at the National University of Ireland, Galway.

Most of the problems found in this monograph are standard in the sense that they (or their versions) can be found in a myriad of textbooks and course syllabi with similar exercises and solutions. Although the text doesn’t claim to formulate new problems, the manual provides a collection of highly detailed solutions and answers. This serves as a nod to the students who are struggling to cope with the subject matter (and who doesn’t at times?) who are eager to understand all the steps of reasoning and derivations employed in the solutions. In addition, the text offers a balanced mix of problems: some of which are solved using basic logic and computation, while others are theoretical, and require rigorous proofs.

The problems in the chapter ‘Probability I’ are concerned with the algebra of events, and require minimum mathematical skills. The section ‘Probability II’ deals mostly with the countable sample spaces and discrete random variables. ‘Probability III’ focuses on uncountable sample spaces and continuous random variables and may be slightly more mathematically demanding. The problems in ‘Stochastic Processes’ cover the basics of random processes with discrete state space, typically taught in undergraduate introductory courses.

Milovan Krnjajić

Belgrade, October 2020.

¹Billingsley, P. (1995) Probability and Measure

PART

I

Exercises in Probability I

Exercises in Probability I

Exercise 1.1 Let A and B be two independent events.

- (a) Prove that A and B^c are independent events, where B^c is the complement of event B .
- (b) Are A^c and B^c also independent?

Exercise 1.2 If A and B are two mutually exclusive events, when are they also independent?

Exercise 1.3 What are the underlying probability spaces for the following experiments:

- (a) three tosses of a biased coin.
- (b) a box contains two red and two black balls; two balls are drawn without replacement.
- (c) a biased coin is tossed repeatedly until a head shows up.

Exercise 1.4 Let Ω be a sample space partitioned into sets A and A^c . Assume that for an event B on the sample space Ω the conditional probabilities $P(B|A)$ and $P(B|A^c)$ are defined. Find $P(A|B)$ and $P(A^c|B)$.

Exercise 1.5 Let A and B be two events defined on the same sample space. Using axioms of probability show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. *Hint:* Draw a Venn diagram and express $A \cup B$ as a union of three disjoint subsets.

Exercise 1.6 A fair coin is tossed repeatedly. Show that with probability one, a head appears sooner or later. Likewise, show that any given finite sequence of heads and tails will occur eventually with probability one.

Exercise 1.7 Prove Boole's inequality:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

Exercise 1.8 Show that $P(A \cup B \cup C) = 1 - P(A^c|B^c \cap C^c)P(B^c|C^c)P(C^c)$.

Exercise 1.9 Calculate $P(A|B)$ when (a) $A \cap B = \emptyset$, (b) $A \subset B$, and (c) $B \subset A$.

Exercise 1.10 Let $P(A) = 0.43$ and $P(B) = 0.87$, where A and B are events. Show that $P(A \cap B) \geq 0.30$.

Exercise 1.11 An urn contains five red and seven black balls. A ball is randomly selected from the urn, its color noted and then two balls of the same color are put in the urn. This procedure is repeated four times. What is the probability of observing the following sequence of colors: "red, red, black, black" (in the given order)?

Exercise 1.12 A pair of fair dice are to be rolled repeatedly and the sum S of the two numbers is observed for each roll. What is the probability of observing $S = 7$ before $S = 3$?

Exercise 1.13 Which of the following two events is more likely?

- (a) Four rolls of a fair die show at least one six.
- (b) Twenty four rolls of two fair dice show double six at least once.

Exercise 1.14 A student claims that two events A and B are *mutually exclusive* and have probabilities $P(A) = 0.7$ and $P(B) = 0.6$. Is this possible? Explain.

Exercise 1.15 When playing a state lottery a player chooses 8 out of numbers 1 to 40. Lottery commission then performs an experiment that selects eight out of these forty numbers. We assume that all of $\binom{40}{8}$ combinations are equally likely. What is the probability that a player has

- (a) all 8 of the winning numbers?
- (b) at least 7 of the winning numbers?
- (c) at least 6 of the winning numbers?

Exercise 1.16 From a group of 6 freshmen, 8 sophomores, 8 juniors and 6 seniors a committee of size 4 is randomly selected. Find the probability that the committee will consist of

- (a) one student from each class

(b) two sophomore students and two juniors

(c) only sophomores or juniors

Exercise 1.17 A bank has good and bad loans. The chance of a bad loan defaulting in any year is 40%, while the chance of a good loan defaulting is only 5%. Currently the bank assesses that 80% of its loans are good. What proportion of loans will default in the next year?

If a loan defaults, what is the probability that the bank had correctly classified it as a bad loan?

Exercise 1.18 A box contains three coins of which one is a regular coin, one double-headed and one double-tailed. A coin is selected randomly and flipped. What is the probability that it is a double-headed coin if it shows a head?

Exercise 1.19 Two players alter rolling dice. To win, each player needs to make a specific score. If neither rolls the required score, the play continues. Player A wins each roll with probability a and player B wins with probability b . What is the probability that A wins (a) if he makes the first roll; (b) what if he makes the second roll?

Exercise 1.20 Let the sample space $\Omega = 1, 2, 3, 4$, and consider events $A = 1, 2$, $B = 1, 3$, $C = 2, 3$. Also, let $P(n) = 1/4$, where $n = 1, 2, 3, 4$. Are events A, B, C jointly independent? Does $P(A, B, C) = P(A)P(B)P(C)$ imply pairwise independence of (A, B) , (A, C) , (B, C) ?

Exercise 1.21 Bob and Mike play tennis and the game came to deuce. Bob wins any point with probability p_1 , independently of any other point. What is the probability that Bob wins the game?

Exercise 1.22 Suppose three fair coins are flipped. Then, it must be that at least two are alike. Also, it's even chance whether the third is a head or a tail. So, the probability that all three are the same is $1/2$. Or, is it?

Exercise 1.23 A box contains two marbles for which it is known that each is either white or black. Show that one is black and the other is white!

Exercise 1.24 Each of three boxes contains two marbles with following distribution of colors: BB , BW , WW . Select a box at random and remove a marble without looking at it. Note that it's equally likely for the other marble to be black or white. Therefore, the chance that the selected box contains identical marbles is $1/2$. Do you agree or disagree, and why?

Exercise 1.25 You've met a relative after many years and learn that she has two children. Moreover, at least one of them is a boy. What is the probability that both children are boys?

Exercise 1.26 A box contains twenty coins but five are counterfeit. A thief steals five randomly selected coins. After that the owner picks randomly one coin from the remaining fifteen.

(a) What is the probability that this coin is counterfeit?

(b) Given that the coin picked by the owner is counterfeit, what is the probability that the thief stole at least one counterfeit coin?

Exercise 1.27 Show that the probability that exactly one of the events A and B occurs is $P(A \cup B) = P(A) + P(B) - 2P(A \cap B)$.

Exercise 1.28 Three cards are selected without replacement from a regular deck of 52 cards. Find the probability of not obtaining a spade.

Exercise 1.29 Let 90% of a sphere be colored red with rest 10% black. Show that regardless of the distribution of colors, there is always a way to inscribe a cube with all red vertices.

Exercise 1.30 A laboratory blood test is 95 percent correct in detecting a certain disease when it is, in fact, present. In other words, the *sensitivity* of the test is 95%. However, the test also gives a "false positive" with probability 0.02 for the healthy persons tested. That is, if a healthy person is tested, then with probability 0.02 the test will come out positive. This is also phrased as – the test has 98% *specificity*, or the test comes out negative 98% of time when a person is healthy. If one percent of the population actually has the disease (the prevalence of the

disease is one percent), what is the probability that a randomly chosen person has the disease given that the test came out positive?

Exercise 1.31 Alice goes on vacation and asks her relative Bob to take care of her flowers in her absence. The flowers are very delicate and have 15% chance of withering away. If not watered that chance is 80%. Alice is 90% certain that Bob will take good care of the flowers.

- (a) What is the probability that upon her return Alice will find flowers alive?
- (b) If her flowers have withered away, what is the probability that Bob forget to fulfill his promise?

Exercise 1.32 Assume that there are n possible outcomes in a trial and let p_i be the probability of the i -th outcome. Note that $\sum_{i=1}^n p_i = 1$. If two independent trials are observed, find the probability that the outcome in the second trial is larger than in the first.

Exercise 1.33 Mr. Smart Alek has come up with a gambling system he thinks would win him money. When he bets on red, he places \$10 only after he has seen 10 consecutive spins that landed on black. He is convinced that the probability of 11 consecutive spins resulting in black is very small. What's your opinion about this system?

Exercise 1.34 Let A, B, C be events on the same sample space. Show the following

- (a) the conditional independence of A and B given C does not imply independence of A and B
- (b) independence of A and B does not imply conditional independence of A and B given C
- (c) For which events C are both (a) and (b) satisfied?

Exercise 1.35 Let a coin have probability of 0.8 of landing on heads when flipped. Alice observes the result (either heads or tails) and goes to tell Bob. However, Alice's attention span is fairly short and there is 0.4 chance that she would forget the result by the time she communicates it to Bob. If Alice forgets the result, instead

of admitting it, she'd tell with equal probability that it was either heads or tails. If Alice remembers the result, she'd faithfully convey it.

- (a) What is the probability that Bob will be told that the result was heads?
- (b) What is the probability that Bob is given the correct result?
- (c) Given that Bob is told that the coin landed on heads, what is the probability that it was in fact so?

Exercise 1.36 Marbles are randomly removed from a box that initially contained 20 red and 10 blue marbles.

- (a) What is the probability that all of the red balls are removed before all of the blue ones have been removed?

Now, assume that the box contains 20 red 10 blue and 8 green marbles.

- (b) What is the probability that all of the red marbles are removed before all the blue ones have been removed?
- (c) What is the probability that the marbles are drawn in the order blue, red, green?
- (d) What is the probability that the group of blue marbles is the first of the three groups to be removed?

Exercise 1.37 Let's play a game with an ordinary deck of 52 cards. Assume that a dealer randomly draws two cards without replacement. Find the probability that both cards are aces conditioned on the additional information from the dealer in each case below.

- (a) Dealer says that one of the cards is the ace of hearts.
- (b) The first selected card is an ace.
- (c) The second card is an ace.
- (d) At least one card is an ace

Exercise 1.38 Let a drug D have 99% success rate in cases where the bad side effects do not arise. The bad side effects do occur in 5% of cases where the drug D is administered. If side effects are present the success rate for the drug falls to 30%. What is the unconditional probability of success if this drug is used?

Exercise 1.39 An exam multiple choice question offers n choices. A student either knows to answer the question with probability p or guesses randomly with probability $1 - p$. Given that the answer is correct, what is the probability that the student did not guess it, that is, he knew the answer?

Exercise 1.40 Let the sequence $\{A_n, n \geq 1\}$ be either increasing or decreasing sequence of events. Show that P is continuous function, that is,

$$\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n).$$

Exercise 1.41 Assume that you are looking for a top-floor apartment. An ad in the newspaper says that there are two vacant apartments in a building with 7 floors with 8 apartments per floor. You decide to go out and check the apartments only if there is at least an even chance that the top apartments are available for rent. Find the probability that there is a vacant apartment on the top floor? Will you go to check this building?

Exercise 1.42 A group of five people simultaneously enter an elevator at the ground floor. There are 10 upper floors. The persons choose their exit floors independently of each other.

(a) Specify an appropriate sample space and determine the probability that they are all going to different floors when each person randomly chooses one of the 10 floors as the exit floor.

(b) How does the answer change when each person chooses with probability $1/2$ the 10th floor as the exit floor and the other floors remain equally likely as the exit floor with a probability of $1/18$ each.

Exercise 1.43 The Manhattan distance of a point (x, y) in the plane to the origin $(0, 0)$ is defined as $|x| + |y|$. You choose at random a point in the unit square $\{(x, y) : 0 \leq x, y \leq 1\}$. What is the probability that the Manhattan distance of this point to the point coordinate origin, $(0, 0)$, is no more than d for $0 \leq d \leq 2$?

Exercise 1.44 In the movie *Deer Hunter* there is a greasy scene of two prisoners of war being forced to play Russian roulette. In a somewhat different yet equally

bizarre scenario, assume that three persons, A , B and C are “playing” the Russian roulette in which they take turns pulling the trigger of a six-cylinder revolver loaded with one bullet. Before each turn, the magazine is spun to randomly select a new cylinder to fire until the deadly shot. Three players take the order $A, B, C, A, B, C, \dots$. Determine for each of the three players the probability that the player will be the one to trigger the deadly round.

Exercise 1.45 Three players A , B and C , play a poker game. They need to choose the dealer and they do it by each of them drawing a card from a freshly shuffled deck until someone gets an ace. They draw a card in order $A, B, C, A, B, C, \dots$. The first player receiving an ace starts the game as dealer. Does everyone have an equal chance to become dealer?

Exercise 1.46 An oil explorer performs a test to find out whether oil is likely to be in a certain area. The probability that the test correctly indicates the presence of oil is 90% if oil is indeed present in the test area and the probability of a false positive is 15% if no oil is present in the test area, or in other words, the test is 85% negative in case the oil is really not present. The explorer’s prior belief regarding the presence of oil in the test area can be quantified as prior probability of 40%. Use Bayes’ rule in odds form to update the value of the probability of oil being present in the test area given that the test gives a positive signal.

Exercise 1.47 Assume that two numbers from 1 to 10 are generated randomly and independently of each other.

(a) What is the probability that both numbers are even if the information is provided that at least one is?

(b) What is the probability that both numbers are even given the information that the number two is among the two randomly generated numbers?

Exercise 1.48 Assume that five people are sitting at a bar. Two of them order rum and the other three order vodka. The waiter forgot who ordered what and puts the drinks in a random order for the five persons.

(a) Specifying an appropriate sample space find the probability that each person gets the correct drink.

(b) Solve the problem using conditional probabilities.

Exercise 1.49 Assume that $P(A|B) > P(A)$. Is it then true that $P(B|A) > P(B)$?

Exercise 1.50 Let X and Y be the scores on two fair dice taking values in the set $\{1, 2, \dots, 6\}$. Let $A_1 = \{X+Y = 9\}$, $A_2 = \{X \in \{1, 2, 3\}\}$, and $A_3 = \{Y \in \{3, 4, 6\}\}$. Show that $P(A_1 \cup A_2 \cup A_3) = P(A_1)P(A_2)P(A_3)$. Are these three events independent?

PART II

Exercises in Probability II

Exercises in Probability II

Exercise 2.1 A box contains twenty balls numbered from one to twenty from which three balls are randomly selected without replacement. If we bet that at least one of drawn balls has a number ≥ 17 , what is the probability of winning the bet?

Exercise 2.2 Let $P(X = i) = k\lambda^n/n!, n = 0, 1, 2, \dots$ where λ is a positive constant, be the probability mass function of a random variable X . Find (a) $P(X = 0)$; (b) $P(X > 2)$.

Exercise 2.3 Let X and Y be independent Poisson random variables with parameters λ, μ , respectively. Prove that the distribution of $X + Y$ is Poisson with parameter $\lambda + \mu$.

Exercise 2.4 Consider an experiment in which a coin is flipped *once* and lands on heads with probability p , and tails $q = 1 - p$. Let X and Y be the numbers of heads and tails, respectively. It's not surprising that X and Y are not independent because $P(X = 1, Y = 1) = 0$, but $P(X = 1)P(Y = 1) = p(1 - p)$.

Assume that the coin is tossed N times where N is a random number following the Poisson distribution with parameter λ . It is a remarkable fact that now X and Y are independent random variables. Prove that it is indeed so.

Exercise 2.5 Let X and Y be independent random variables, each taking the values -1 or 1 with probability $1/2$, and let $Z = XY$.

(a) Show that X, Y, Z are pairwise independent.

(b) Are X, Y, Z jointly independent?

Exercise 2.6 Let X and Y be independent positive integer-valued random variables having the same probability mass function $f(x) = 2^{-x}, x = 1, 2, \dots$. Find:

(a) $P(\min(X, Y) < x)$, (b) $P(Y > X)$, (c) $P(X = Y)$, (d) $P(X > kY)$, for a given integer k , (e) $P(X \text{ divides } Y)$, (f) $P(X = rY)$, for a given positive rational r .

Exercise 2.7 Let X have the probability mass function f and $Y = g(X)$ where $g : R \rightarrow R$. Prove that

$$E(g(X)) = \sum_x g(x)f(x)$$

provided that the sum is absolutely convergent.

Exercise 2.8 Prove that $E(aX + bY) = aE(X) + bE(Y)$, where $a, b \in R$, and X and Y are discrete random variables.

Exercise 2.9 Let X and Y be independent Bernoulli random variables with parameter $p = 0.5$. Show that $X + Y$ and $|X - Y|$ are uncorrelated yet not independent.

Exercise 2.10 Show that if X and Y are independent then

$$E(XY) = E(X)E(Y).$$

X and Y are called *uncorrelated* if $E(XY) = E(X)E(Y)$. Note that the independent variables are uncorrelated, however, the converse is *not* true.

Exercise 2.11 Let X and Y be random variables. Show that

(a) $\text{var}(aX) = a^2\text{var}(X)$ for $a \in R$,

(b) $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$ if X and Y are uncorrelated.

Exercise 2.12 Let $G = (V, E)$ be a finite graph. For any subset W of vertices and any edge e in E , define the indicator function

$$I_W(e) = \begin{cases} 1 & \text{if } e \text{ connects } W \text{ and } W^c \\ 0 & \text{otherwise.} \end{cases}$$

Set $N_W = \sum_{e \in E} I_W(e)$, that is, N_W is the number of edges which connect vertices in W with those in W^c . Using the *probabilistic method* show that there exists $W \subseteq V$ such that $N_W \geq |E|/2$.

Exercise 2.13 Using the *probabilistic method* again, prove the claim in the following problem. Assume that there is a sequence of 17 integrated circuits (IC-s) arranged in a circle, and that it's known that five are faulty. Show that there is a subsequence of seven IC-s in which at least three are faulty.

Exercise 2.14 Let there be a random number of coins on the table, where N has the Poisson distribution with parameter λ . Let each coin be flipped once with heads appearing with probability p . Prove that the total number of heads also has the Poisson distribution but with parameter $p\lambda$.

Exercise 2.15 Let X and Y be independent discrete random variables and let $g, h : R \rightarrow R$.

- (a) Show that $g(X)$ and $h(Y)$ are independent.
- (b) A more general statement is that X and Y are independent if and only if $f_{X,Y}(x, y)$ can be factorized as the product $g(x)h(y)$ of a function g of x alone and a function h of y alone. Prove this.

Exercise 2.16 Let X and Y be independent Poisson variables with parameters λ and μ , respectively.

- (a) Show that $X + Y$ is Poisson with parameter $\lambda + \mu$.
- (b) Show that the conditional distribution of X , given $X + Y = n$, is binomial.

Exercise 2.17 If the values of random variable X are only positive integers, show that

$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

Exercise 2.18 There are n students at the graduation ceremony. At the end everyone tosses their academic graduation cap in the air. Assume that each cap has a unique student ID on it. The caps are then collected in one pile and students randomly select a cap. What is the probability that not a single student gets the correct cap?

Exercise 2.19 Let X and Y be discrete random variables with the joint probability mass function

$$f_{X,Y}(x, y) = k[(x + y - 1)(x + y)(x + y + 1)]^{-1}, \quad \text{where } x, y = 1, 2, 3, \dots$$

- (a) Find the marginal distributions of X and Y .

(b) Determine the normalizing constant k .

(c) What is the covariance of X and Y ?

Exercise 2.20 Let X and Y be discrete random variables with mean zero, variance one and covariance ρ . Show that

$$E(\max\{X^2, Y^2\}) \leq 1 + \sqrt{1 - \rho^2}.$$

Exercise 2.21 Let $f(X, Y)$ be the joint mass function of discrete random variables X and Y .

(a) Show that $E(\log f_X(X)) \geq E(\log f_Y(X))$

(b) The following expression is the *mutual information* between X and Y :

$$I = E_{X,Y} \left[\log \left(\frac{f(X, Y)}{f_X(X)f_Y(Y)} \right) \right]$$

Prove that $I \geq 0$ with equality if and only if X and Y are independent.

Exercise 2.22 Prove that the conditional expectation $\psi(X) = E(Y|X)$ satisfies

$$E(E(Y|X)) = E(Y).$$

Here, $\psi(X) = E(Y|X)$, emphasizes that $E(Y|X)$ is a function of the random variable X . Note also that the first expectation operator in $E(E(Y|X))$ is with respect to the distribution of X , whereas the second, $E(Y|X)$, is (obviously) over the conditional distribution of $Y|X$.

Exercise 2.23 Prove that the conditional expectation $\psi(X) = E(Y|X)$ satisfies

$$E(E(Y|X)g(X)) = E(Yg(X))$$

for any function g for which both expectations exist. Note that the expectation in $E(Yg(X))$ is with respect to the joint distribution of X, Y .

Exercise 2.24 An investor takes a position in N stocks, where N has the Poisson distribution with parameter λ . Each investment (bet!) will be profitable with prob-

ability p independently of other investments. Let K be the number of successful investments. Find $E(K|N)$, $E(K)$, and $E(N|K)$.

Exercise 2.25 Let X and Y be two discrete random variables not necessarily independent. Prove that $P(X + Y = z) = \sum_x f(z, z - x)$.

Exercise 2.26 A manufacturing process in a company yields 1% defective parts independently of each other. The parts are sold in packages of ten parts per package and the company will replace a package free of charge or give money back if more than one part is defective. What proportion of packages sold must the company replace?

Exercise 2.27 Assume that in a jury trial it takes 8 out of 12 jurors to convict the defendant. In other words, for the defendant to be convicted at least 8 out of 12 jurors must cast a guilty vote. Assume that the jurors make their decisions independently. If each juror makes the right decision with probability p , what is the probability that the jury will do a good job, that is, render a correct decision.

Exercise 2.28 Let X denote a binomial random variable with parameters (n, p) , where $0 < p < 1$. Then, as k assumes values from 0 to n , the probability $P(X = k)$ first increases monotonically, reaches a maximum and then decreases monotonically. The largest value is reached when k is the largest integer $\leq (n + 1)p$. Prove these claims.

Exercise 2.29 Let $X \sim \text{Binomial}(n, p)$, that is, random variable X has a binomial distribution with parameters n, p . Find the expectation and the variance of X .

Exercise 2.30 Find the expectation and the variance of the Poisson distribution with parameter λ

Exercise 2.31 Poisson distribution can be used to approximate a binomial random variable. Show that both distributions have approximately equal expectation and variance for the case when n is large and p is small so that np is moderate.

Exercise 2.32 Let X be a random variable with expectation μ and variance σ^2 . Define $Y = (X - \mu)/\sigma$. Find $E(Y)$ and $V(Y)$.

Exercise 2.33 Let $X_1, X_2, \dots, X_n$ be random variables that are independent and with the identical distribution (usually abbreviated, iid random variables). Assume that $E(X_i) = \mu$ and let $Y_i = X_i/(X_1 + X_2 + \dots + X_n)$, for $i = 1, 2, \dots, n$. Find $E(Y_i)$.

Exercise 2.34 For two random variable X and Y show that

$$(E(XY))^2 \leq E(X^2)E(Y^2).$$

Exercise 2.35 For random variables X and Y , define the correlation coefficient, ρ_{XY} , as

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

where σ_X^2 and σ_Y^2 are variances of X and Y , respectively. Show that

$$-1 \leq \rho_{XY} \leq 1$$

Exercise 2.36 Let X and Y be random variables and $Y = aX + b$.

- (a) Find the covariance of X and Y
- (b) Find the correlation coefficient between X and Y .

Exercise 2.37 Let $X_1, X_2, \dots, X_n$ be n random variables. Show that

$$\text{Var} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j X_i X_j \text{Cov}(X_i, X_j).$$

Exercise 2.38 Let X denote a geometric random variable that counts the number of Bernoulli trials until the first success occurs (note that the first successful trial is included in the count). Then, $P(X = n) = (1 - p)^{n-1}p$, $n = 1, 2, \dots$ where p is the probability of success.

- (a) Show that the success will eventually occur for any $p > 0$.
- (b) Find the expectation and the variance of the geometric random variable.

Exercise 2.39 Show that for any distribution $p(x)$ with mean μ and variance σ^2

$$\sigma^2 = \sum_{x \in D} x^2 p(x) - \mu^2,$$

where D is the domain of the random variable X .

Exercise 2.40 Not all distributions have a finite mean or variance.

(a) Let a probability mass function $p(x)$ be defined as $p(x) = c/x^2, x = \pm 1, \pm 2, \dots$. Show that the distribution has no well defined mean.

(b) Probability mass function is $p(x) = 2c/x^2, x \geq 1$. Prove that the mean is infinite.

(c) Let $p(x) = c/x^3, x = \pm 1, \pm 2, \dots$. Show that the mean is finite whereas the variance is infinite.

Exercise 2.41 Recall the definition of variance, $V(Y) = E((Y - E(X))^2)$. Now, consider the following definition of conditional variance:

$$V(Y|X = x) = E((Y - E(Y|X = x))^2|X = x).$$

Prove that $V(Y) = E(V(Y|X)) + V(E(Y|X))$.

Exercise 2.42 Consider a radioactive source that emits particles which are detected and counted by the Geiger counter. The number of particles that are emitted in a given time interval is Poisson distributed with expected value λ . It's assumed that each emitted particle is detected and recorded by the counter with probability p , independently of the other particles. Let X denote the random variable that counts the detected particles in a given time interval and let Y be the number of undetected particles in the time interval.

(a) Find the probability mass functions of X and Y ?

(b) Are X and Y independent?

Exercise 2.43 Assume you are in a casino that is offering a new game which it advertises as giving a player the fantastic chances of wins.

The details of the game are as follows. You may simultaneously roll one red die and three blue dice. The stake is \$1. If none of the blue dice matches the red die, you lose your stake; otherwise, you get anyway paid $k + 1$ dollars if exactly k of the blue dice match the red die. In the case that exactly one blue die matches the red die, you get paid an additional 0.50 if the other two blue dice match.

(a) What is the expected payoff of the game?

(b) Would you like to play this game?

Exercise 2.44 Imagine that two players, A and B , are invited to playing the following game of chance. The first player to win a total of ten games is the overall winner. Assume that each of the two players has an even chance to win any given game. Then, without any warning, the game is suddenly interrupted and must be ended. This happens at a moment that the player A has won a games and the player B has won b games with $a < 10$ and $b < 10$. What is the probability that the player A would have been the overall winner in a case the competition continued interrupted?

Exercise 2.45 Assume that a player practices skills for a certain game. On a given day, the player achieves scores taking values from the range 101 to 110. with probability 0.1, independent of other days. The player wants to improve his score so he decides to play on three different days and to declare as a representative score the minimum X of the scores X_1, X_2, X_3 , obtained on these three different days.

(a) Find the probability mass function of X .

(b) By how much has the player's expected score improved after playing on three days?

Exercise 2.46 Each day, a diligent policeman stops several traffic offenders. On any given day, the number of traffic offenders he stops can be any number from one to six with even chance. The policeman makes his interventions independently of what he has done in the previous days. Let X be the number of traffic offenders the policeman stops in 10 days. Find the mean and the variance of X .

Exercise 2.47 An office worker drives to work 5 days a week for a full year (50 weeks), and get a traffic ticket with probability $p = 0.02$ on any specific day, independently of other days. Let X be the total number of tickets that person gets in a year.

(a) What is the probability that the number of tickets you get is exactly equal to the expected value of X ?

(b) Calculate the probability in (a) using a Poisson approximation.

(c) Any one of the tickets is \$10 or \$20 or \$50 with respective probabilities 0.5, 0.3, and 0.2, independently of other tickets. Find the mean and the variance of the amount of money the office worker pays in traffic tickets during the year.

(d) Suppose that the probability, p , of getting a ticket is unknown, but the person got 5 tickets during the year. Let the estimate of p be given by the sample mean $\hat{p} = 5/250 = 0.02$.

What is the range of possible values of p assuming that the difference between p and the sample mean $\hat{p}$ is within 5 times the standard deviation of the sample mean?

Exercise 2.48 The joint probability mass function of (X, Y) is given by

$$p_{XY}(x, y) = \begin{cases} k(2x + y) & x = 1, 2; y = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(a) Find the value of k .

(b) Find the marginal pmf's of X and Y .

(c) Are X and Y independent?

Exercise 2.49 Let the joint pmf of (X, Y) be specified as:

$$p_{XY}(x, y) = \begin{cases} kx^2y & x = 1, 2; y = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}$$

- (a) What is the value of k ?
- (b) Find the marginal probability mass functions of X and Y .
- (c) Are X and Y independent?

Exercise 2.50 Let $X_1 + X_2 + \cdots + X_n$ be a sequence of random variables representing the results of independent trials. All random variables have the same mean and the same (finite) variance, that is, $E(X_i) = \mu$, and $V(X_i) = \sigma^2$, for all $i = 1, 2, \dots, n$. Let $S_n = X_1 + X_2 + \cdots + X_n$. Then for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0.$$

This is a simple version of the weak law of large numbers (WLLN).

PART

III

Exercises in Probability III

Exercises in Probability III

Exercise 3.1 Assume that X is a continuous random variable whose probability density function is given by $f(x) = c(2x - x^2)$ for $0 < x < 2$, and zero otherwise.

(a) Find the value of c . (b) Find $P(X > 1)$.

Exercise 3.2 The length of time in days that a machine functions before breaking down is a random variable with the density function given by $f(x) = \lambda \exp(-x/100)$, where $x \geq 0$, and is zero otherwise.

(a) What is the probability that the machine will function between 50 and 150 days without a problem.

(b) What is the probability that it will work well for less than 100 hours?

Exercise 3.3 Let $f(x)$ be a density function of a random variable X , given by $f(x) = 1$ for $0 < x < 1$ and zero otherwise. Find $E(\exp(X))$.

Exercise 3.4 For a non-negative continuous random variable X , prove that

$$E(X) = \int_0^{\infty} P(X > x) dx.$$

Exercise 3.5 For a non-negative continuous random variable X , with density function $f_X(x)$, and function $g(x) \geq 0$, prove that

$$E(g(X)) = \int_0^{\infty} g(x) f_X(x) dx.$$

Exercise 3.6 Let X have uniform distribution defined on interval $0 \leq x \leq 1$ and define a new random variable, Y , as $Y = a + (b - a)X$. Find the density function of Y .

Exercise 3.7 Let random variable X have the exponential distribution with parameter λ , where $\lambda > 0$. Then, its CDF is

$$F(x) = 1 - \exp(-\lambda x), x \geq 0.$$

This distribution arises in practice in cases of waiting time between unpredictable events such as arrivals of airplanes, phone calls at a call-center, earthquakes, emission of radioactive particles, visitors at a bar, etc. Its discrete analogue is the geometric distribution.

- (a) Considering a sequence of Bernoulli trials occurring at diminishingly small epochs of time derive the continuous distribution of the first success.
- (b) Find the expectation of X .

Exercise 3.8 Let $X \sim \text{Exp}(-\lambda)$, that is, X is a random variable having the exponential distribution with parameter λ . Find $E(X)$ and $V(X)$.

Exercise 3.9 Let $X \sim N(0, 1)$ and $Y = e^X$. Find the density function of Y .

Exercise 3.10 2. Let X and Y be two independent random variables with the same distribution and density functions F and f , respectively.

- (a) Prove that $V = \max\{X, Y\}$ has distribution function $P(V \leq x) = F(x)^2$ and density function $f_V(x) = 2f(x)F(x)$, $x \in R$.
- (b) Find the density function of $U = \min\{X, Y\}$

Exercise 3.11 Let X be a continuous random variable with finite variance. Show that the minimum of $E(X - a)^2$ is achieved at $a = E(X)$.

Exercise 3.12 Let X be a random variable with uniform density on $[a, b]$. Find its mean and variance.

Exercise 3.13 Consider a continuous r.v. X with probability density function $f_X(x) = 0, x < 0$. Show that for any $a > 0$

$$P(X \geq a) \leq EX/a.$$

This is Markov inequality.

Exercise 3.14 Let X be a continuous random variable with $EX = \mu$ and $\text{Var}(X) = \sigma^2$. Prove that, for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \sigma^2/a^2.$$

This is known as the Chebyshev inequality.

Exercise 3.15 A random variable X has *memoryless property* if

$$P(X \leq x + t | X > t) = P(X \leq x), \quad x, t > 0$$

Show that if X is non-negative and continuous and also memoryless, then X must be an exponential random variable.

Exercise 3.16 Let X and Y be independent random variables with finite variances, and let $U = X + Y$ and $V = XY$. What is the condition for U and V to be uncorrelated?

Exercise 3.17 Let $X_k : 1 \leq k \leq n$ be independent and identically distributed with finite variance, and define $\bar{X} = (\sum_{k=1}^n X_k)/n$. Show that $\text{Cov}(\bar{X}, X_r - \bar{X}) = 0$.

Exercise 3.18 A unit stick is broken uniformly at random and then the larger piece is broken again uniformly at random. Show that the probability the three pieces may be used to form a triangle is $2 \log 2 - 1$.

Exercise 3.19 Consider a similar stick-breaking problem to the previous one in which the stick is broken into three pieces at two randomly chosen points. Find the probability that no piece is longer than half the length of the stick.

Exercise 3.20 Let X and Y be defined by $X = \cos \Theta$, $Y = \sin \Theta$,

where $\Theta \sim \text{Uniform}(0, 2\pi)$

(a) Show that X and Y are uncorrelated.

(b) Show that X and Y are not independent.

Exercise 3.21 Let X and Y be continuous random variables. Show that

$E(E(Y|X)) = E(Y)$, where the outer expectation is with respect to the distribution of X .

Exercise 3.22 Show that if X and Y are independent random variables then

$$(g(X)h(Y)) = E(g(X))E(h(Y)).$$

Exercise 3.23 Let X and Y have the joint density $f(x, y) = cxy$, for $0 < x < y < 1$. Find c .

Exercise 3.24 Let random variable X have a Pareto distribution with parameter 4, and let f_X be its probability density function

$$f_X(t) = \begin{cases} kt^{-4} & (t > 1) \\ 0 & (t \leq 1) \end{cases}$$

where k is some positive constant.

(a) Find the constant k .

(b) Find the d.f. F_X .

(c) Find μ_X, σ_X

(d) Find $P(|X - \mu_X| \leq 1.5\sigma_X)$.

(e) Use Chebyshev's inequality to estimate $P(|X - \mu_X| \leq 1.5\sigma_X)$.

Exercise 3.25 Prove that the following result for the total expectation:

$$E(X) = \int_{-\infty}^{\infty} E(X|Y = y)f_Y(y)dy.$$

Exercise 3.26 Assume that a light bulb has exponentially distributed lifetime Y . However, due to the variations in the manufacturing process the company has been experiencing some quality control problems. On any given day, the parameter λ of the probability density function (PDF) of Y is actually a random variable, uniformly distributed in the interval $[1, 1.5]$. A light bulb is tested and its lifetime recorded. What can be said about the underlying parameter λ ?

Exercise 3.27 Assume that we are interested in a detection of a transmitted binary signal S . The signal is transmitted according to the following probabilities: $P(S = 1) = p$ and $P(S = -1) = 1 - p$. The received signal is $Y = N + S$, where N is normal noise, with zero mean and unit variance, independent of the signal S .

What is the probability that $S = 1$, as a function of the observed value y of Y ?

Exercise 3.28 Let X have the PDF

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|},$$

where $\lambda > 0$.

(a) Check whether $f_X(x)$ is a proper probability density function by verifying that it satisfies the normalization condition.

(b) Find the mean and variance of X .

Exercise 3.29 A player throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of the player's hit from the center.

(a) Find the PDF, the mean, and the variance of X .

(b) The target h has an inner circle of radius t . If $X \leq t$, the player gets a score of $S = 1/X$, otherwise his score is $S = 0$. Find the CDF of S .

Exercise 3.30 Let the joint pdf of a bivariate r.v. (X, Y) be specified as

$$f_{XY}(x, y) = \begin{cases} k(x+y) & 0 < x < 2, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(a) Find the value of k .

(b) Find the marginal PDF-s of X and Y .

(c) Are X and Y independent?

Exercise 3.31 The joint pdf of a bivariate r.v. (X, Y) is given by

$$f_{XY}(x, y) = \begin{cases} \frac{1}{y} e^{-x/y} e^{-y} & x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that $f_{XY}(x, y)$ is a proper probability density function.

(b) Find $P(X > 1 \mid Y = y)$.

Exercise 3.32 Sterling formula is

$$\lim_{x \rightarrow \infty} x! = \sqrt{2\pi x} \exp(-x) x^x.$$

Prove the formula using the approximation of Poisson distribution by the Gaussian distribution.

Exercise 3.33 Let $Z \sim N(0, 1)$ so that $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})$, where $-\infty < z < \infty$, and let $X = \mu + \sigma Z$.

(a) Find the PDF of X .

(b) Find the PDF of X^5 .

(c) Let $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$. Express the probability $P(a < X < b)$ in terms of Φ , where a and b are real constants.

(d) Let $Y = aX + b$ where a and b are real constants. Express $E(Y)$ and $V(Y)$ in terms of μ, σ, a and b .

(e) The moment generating function of a random variable T is $G_T(t) = e^{t^2}$. Find the mean and the variance of T .

Exercise 3.34 Let random variable X have a standard normal distribution. Use first principles to answer the following questions.

(a) What is the probability density of the random variable $Y = X^2$?

(b) The random variable X has a standard normal distribution. What is the probability density of the random variable $Y = \sqrt{X}$?

10E – 16

Exercise 3.35 The random variable X has the probability density function $f(x) = (8/\pi) \sqrt{x(1-x)}$, for $0 < x < 1$ and $f(x) = 0$ otherwise. What is the probability density function of the random variable $Y = 2X + 1$?

Exercise 3.36 Show that a Poisson distribution can be approximated by a normal distribution.

Exercise 3.37 You bought a bulb which came from supplier A or B with probabilities p_A and p_B , respectively, with $p_A + p_B = 1$. Supplier A, B manufacture the

bulbs with exponentially distributed lifetimes with expected value $1/\mu_A, 1/\mu_B$. Assume that the bulb has already lasted s time units. What is the probability that the bulb will last for another t time units?

Exercise 3.38 Let the continuous random variables X and Y have the joint density function $f(x, y) = 6(x - y)$, for $0 < y < x < 1$ and $f(x, y) = 0$ otherwise. Find the correlation coefficient of X and Y .

Exercise 3.39 Let S_n be a binomial random variable with parameters $(n = 36, p = 0.5)$.

- (a) Find the exact value of $P(S_n \leq 21)$;
- (b) Calculate $P(S_n \leq 21)$ using the approximation based on the Central Limit Theorem.
- (c) Calculate the same probability using DeMoivre-Laplace approximation to the binomial distribution.

Exercise 3.40 Let's consider what can be learned about whether an event occurred based on observations of a random variable. Recall the form of Bayes theorem using events: $P(A|B) = P(B|A)P(A)/P(B)$. Note that it is necessary that $P(B) > 0$.

Now, assume that Y is a continuous random variable, and we observe $Y = y$. How can we derive $P(A|Y = y)$ when the probability of a continuous variable taking a specific value is zero?

Exercise 3.41 Example 4.5. A linear Function of a Normal Random Variable is Normal. Suppose that X is a normal random variable with mean μ and variance σ^2 , and let $Y = aX + b$, where a and b are scalars, with $a \neq 0$.

Exercise 3.42 Consider arriving at a bus stop at time $t = 0$. Two buses A and B are in operation. The arrival time X of bus A is a uniform random variable on $[0, t_A]$, whereas the arrival time Y of bus B is a uniform random variable on $[0, t_B]$, where $t_A \leq t_B$. The arrival times, that is X and Y , are independent.

What is the probability that the bus A arrives first?

Exercise 3.43 Let X and Y have a joint density $f(x, y) = 1/\pi$ for $(x, y) \in \mathcal{C}$, where $\mathcal{C}$ is a circle with unit radius.

Are X and Y independent random variables?

Exercise 3.44 Consider independent standard normal random variables W, X, Y, Z . Find $P(W + X > Y + Z + 1)$.

Exercise 3.45 The radius of a circle is uniformly distributed on $(0, 1)$. What is the probability density of the area of the circle?

Exercise 3.46 Use direct calculations to answer the following questions. (a) The random variable X has a standard normal distribution. What is the probability density of the random variable $Y = X^2$?

(b) The random variable X has a standard normal distribution. What is the probability density of the random variable $Y = \sqrt{|X|}$?

Exercise 3.47 Let X and Y have joint density $f(x, y) = 1/\pi$, for $x^2 + y^2 \leq 1$.

(a) Find $E(X^2 + Y^2)$ and $E(\sqrt{X^2 + Y^2})$;

(b) Find $E(X \wedge Y)$;

(c) Find $E(X^2/(X^2 + Y^2))$.

(d) Are X and Y independent?

Exercise 3.48 Let X and Y have joint density $f(x, y) = 1/\pi$, for $x^2 + y^2 \leq 1$.

(a) Find EX , EY , and $E(XY)$.

(b) Are X and Y independent? Are they correlated?

(c) Now, let $U = |X|$, and $V = |Y|$. Are U, V correlated? Are they independent?

Exercise 3.49 Let the random variables U and V have joint density

$$f(u, v) = e^{-v}, \quad \text{for } 0 < u < v < \infty$$

Find (a) the marginals $f_U(u), f_V(v)$ (b) the conditional densities $f_{U|V}(u | v)$ and $f_{V|U}(v | u)$; (c) What is the density of $Y = V - U$?

Exercise 3.50 Let X and Y be two independent exponential random variables with parameters λ and μ , respectively. Find $P(X < Y)$.

PART IV

Exercises in Stochastic Processes

Exercises in Stochastic Processes

In the chapter on Stochastic Process we include more exercises on probability generating functions and moment generating functions as they are important and useful devices for solving problems that involve finding properties of sums of random variables. Using generating functions, for example, allows for working with *random* number of summands in a sum of random variables. In general, it's easier to work with generating functions than to work directly with density functions. In particular, generating functions are helpful in solving random walk problems where methods of counting sample paths would not work – as the number of steps of the random walk may not be given in advance and is a random variable itself.

Exercise 4.1 Let $\{X_1, X_2, \dots\}$ be independent and identically distributed rv-s having a common distribution $p(x)$. Specify the sum S_N to have N summands, where N is a random variable independent of X_i -s: $S_N = X_1 + X_2 + \dots + X_N$.

- (a) Using conditioning, find the expected value of S_N .
- (b) Find the expectation of the sum using generating functions.

Exercise 4.2 Let X be a Poisson random variable with parameter λ .

- (a) Find the moment generating function, $M_X(t)$, of X .
- (b) Using $M_X(t)$ find the mean and variance of X .

Exercise 4.3 Let X be an exponential rv with parameter λ . (a) Find the $MGF_X(t)$; (b) Using the MGF find the expectation and variance of X .

Exercise 4.4 State, giving reasons, which of the following matrices correspond to a transition matrix of a Markov chain:

1.

$$\begin{pmatrix} 0.5 & 0.1 & 0.3 \\ 0.2 & 0.2 & 0.7 \\ 0.3 & 0.6 & 0.1 \end{pmatrix}$$

2.

$$\begin{pmatrix} 0.8 & 0 & 0.2 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.6 & 0.3 \end{pmatrix}$$

3.

$$\begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.1 & 0.8 & 0.1 \end{pmatrix}$$

4.

$$\begin{pmatrix} 1/2 & 1/2 & -1/2 & 1/2 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1/2 & 1/2 & 1/2 & 1/2 \end{pmatrix}$$

Exercise 4.5 We want to describe the weather in London using three states, (R) rainy, (C) cloudy and (S) sunny. Furthermore (and quite unrealistically), assume that the states last 24 hours and may change every morning. State of the weather each morning is assumed to be governed by a Markov chain with the transition matrix

$$P = \begin{matrix} & \begin{matrix} R & C & S \end{matrix} \\ \begin{matrix} R \\ C \\ S \end{matrix} & \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.5 & 0 \\ 0.4 & 0.3 & 0.3 \end{pmatrix} \end{matrix}$$

- (a) If it is cloudy on Monday, what is the probability that it is raining for the rest of the week, *i.e.* Tuesday through to Saturday?
- (b) If Monday is sunny, what is the probability that it rains on Wednesday?
- (c) If it rains on Monday, what is the probability of a cloudy Tuesday?
If you are now told that it was sunny on Wednesday, how does this affect the answer to the previous question?
Do these results contradict the Markov property?
- (d) Solve the stationary equations to obtain the stationary probabilities and hence determine the proportion of time the weather is in each of the three states.

- (e) What is the expected number of days between two sunny days?
- (f) What is the probability distribution for the length of a continuous period of cloudy weather? What is the average length of this period?

Exercise 4.6 Let $\{X_n\}_{n \in N_0}$ be a Markov chain with the transition matrix

$$P = \begin{array}{c} \begin{array}{ccc} & \text{A} & \text{B} & \text{C} \\ \text{A} & 1/4 & 1/2 & 1/4 \\ \text{B} & 0 & 1/3 & 2/3 \\ \text{C} & 1/3 & 1/3 & 1/3 \end{array} \end{array}$$

- (a) Classify all the states of the Markov chain.
- (b) Find all stationary distributions. How many stationary distributions are there?
- (c) The chain starts from state A. What is the expected number of steps before it returns to A?
- (d) How many times, on average, does the chain visit state B between two consecutive visits to A?
- (e) Each time the chain visits state A, the amount of \$1 is added to an (initially empty) account, \$2 for state B, and nothing for state C. Estimate the amount of money in the account after 10,000 transitions. Assume that the law of large numbers for Markov chains provides an adequate approximation.

Exercise 4.7 Consider a discrete time branching process in which each particle produces a family in the next generation. The family sizes are independent random variables, X_i , identically distributed according to a distribution with the probability generating function $G_X(s)$. Suppose that Z_n is the size of the n th generation with $G_n(s)$ as its probability generating function, and that there is initially exactly one particle. For the branching process described above, do the following:

- (a) Express the size, Z_{n+1} , of the $(n+1)$ th generation as the sum of Z_n terms corresponding to the family sizes produced in the previous generation.
- (b) Show that for $n = 0, 1, \dots$

$$G_{n+1}(s) = G_n(G_X(s)).$$

(c) Hence, or otherwise, show that

$$E[Z_{n+1}] = \mu E[Z_n]$$

where μ is the mean of the distribution of the family size, i.e. $\mu = E[X]$.

(d) Recall that initially (in the 0th generation), there is only a single particle. Show that

$$E[Z_n] = \mu^n.$$

Exercise 4.8 Assume that the offspring distribution (i.e. the distribution of family sizes) from the previous problem is

$$P(X = k) = \left(\frac{2}{3}\right)^k \frac{1}{3}, \quad k = 0, 1, 2, \dots$$

and do the following:

(a) Show, by induction, that

$$G_n(s) = \frac{1 - 2^n - 2(1 - 2^{n-1})s}{1 - 2^{n+1} - 2(1 - 2^n)s}.$$

(b) Calculate the mean of Z_n .

(c) Calculate the probability of extinction by the n th generation.

(d) Find the probability of ultimate extinction.

Exercise 4.9 Let P denote a Markov matrix. Show that P^n is also a Markov matrix for any positive integer n

Exercise 4.10 A professor of Probability Theory gives exams that can be difficult, medium or easy. If an exam was difficult the next one will have an equal chance to be either medium or easy. If the given exam is medium or easy, then there is 0.5 probability that the next test will be of equal difficulty and a 0.25 probability for each of the other two levels of probability. Specify the transition matrix for this Markov chain and find the steady state probabilities.

Exercise 4.11 Let a Markov chain have two states and the following transition probability matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(a) Find the stationary distribution of the chain.

(b) Find the limiting distribution, $\lim_{n \rightarrow \infty} P^n$.

Exercise 4.12 Show that in a finite state Markov chain not all states can be transient.

Exercise 4.13 Let P be a transition probability matrix representing a two-state Markov chain:

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \quad \text{where } 0 < a < 1, 0 < b < 1.$$

(a) Show that the n -step transition probability matrix P^n is given by

$$P^n = \frac{1}{a+b} \left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$$

(b) Find P^n when $n \rightarrow \infty$.

Exercise 4.14 5.47. Consider a Markov chain with two states and transition probability matrix

$$P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

(a) Find the stationary distribution $\hat{p}$ of the chain.

(b) Find $\lim_{n \rightarrow \infty} P^n$.

(c) Find $\lim_{n \rightarrow \infty} P^n$ by first evaluating P^n .

Exercise 4.15 Let T_n denote the time of the n th event of a Poisson process $X(t)$ with rate λ . Suppose that one event has occurred in the interval $(0, t)$. Show that the conditional distribution of arrival time T_1 is uniform over $(0, t)$

Exercise 4.16 Consider a Poisson process $X(t)$ with rate λ . Show that the time intervals between successive events (or interarrival times) in this process are iid exponential r.v.'s with parameter λ .

Exercise 4.17 Prove that if the number of states is n , and if state j is accessible from state i , then it is accessible in n or fewer steps

Exercise 4.18 Recal the following definition and conventions.

For any states i and j define f_{ij}^n to be the probability that, starting in i , the first transition into j occurs at time n . Formally,

$$\begin{aligned} f_{i1}^0 &= 0 \\ f_{ij}^n &= P \{X_n = j, X_k \neq j, k = 1, \dots, n-1 \mid X_0 = i\} \end{aligned}$$

Let

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^n$$

Then f_{ij} denotes the probability of ever making a transition into state j , given that the process starts in i . (Note that for $i \neq j$, f_{ij} is positive if, and only if, j is accessible from i . State j is said to be recurrent if $f_{jj} = 1$, and transient otherwise.

Show that

$$P_{ij}^n = \sum_{k=0}^n f_{ij}^k P_{ij}^{n-k}$$

Exercise 4.19 A particle moves among n locations that are arranged in a circle (with the neighbors of location n being $n-1$ and 1). At each step, it moves one position either in the clockwise position with probability p or in the counterclockwise position with probability $1-p$. Find the transition probabilities of the reverse chain.

Exercise 4.20 If $f_{ij} < 1$ and $f_{jj} < 1$, show that: $\sum_{n=1}^{\infty} P_n^n < \infty$.

PART V

Solutions to Exercises in Probability I

Solutions to Exercises in Probability I

Solution 1.1 (a) To prove independence of A and B^c we need to show that $P(A, B^c) = P(A)P(B^c)$. Note that $P(A, B) = P(A \text{ and } B)$ is the same as $P(A \cap B)$.

Now, $P(A \cap B^c) = P(A \setminus (A \cap B)) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$. Drawing a simple Venn diagram would clarify the first equality.

(b) Having proved that A and B^c are independent, we use the same steps to prove that A^c and B^c are independent.

$$P(A^c \cap B^c) = P(A^c \setminus (A^c \cap B)) = P(A^c) - P(A^c \cap B) = P(A^c) - P(A^c)P(B) = P(A^c)P(B^c).$$

Solution 1.2 In general, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. For mutually exclusive events $P(A \cup B) = P(A) + P(B)$, and also $P(A \cap B) = P(A)P(B) = 0$, implying that $P(A)$ or $P(B)$ is zero.

Note: Only dependent events can be mutually exclusive (assuming strictly positive probability of each event).

Solution 1.3 The sampling space $\mathcal{F}$ can be taken to be the set of all the subsets of outcomes Ω , with the probability of any member of $\mathcal{F}$ equal to the sum of probabilities of the elements of the member subset.

(a) $\Omega = \{H, T\}^3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$, the set of triples of heads and tails. Assuming independence, the probability of n heads in any triple is $p^n(1 - p)^{3-n}$, where p is the probability of a head at each throw.

(b) Here, $\Omega = \{R, B\}^2 = \{RR, RB, BR, BB\}$. Probabilities of drawing two balls without replacement are: $P(RR) = P(BB) = (2/4)(1/3)$, and $P(RB) = P(BR) = (2/4)(1/3)$.

Solution 1.4 $P(B|A) = P(B \text{ and } A)/P(A) = P(B \cap A)/P(A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$, where we used the Law of Total Probability in the last step.

Solution 1.5 $P(A \cup B) = P((A \cap B^c) \cup (A \cap B) \cup (B \cap A^c))$
 $= P(A \cap B^c) + P(A \cap B) + P(B \cap A^c)$

Now, $P(A) = P(A \cap B^c) + P(A \cap B)$ and $P(B) = P(B \cap A^c) + P(B \cap C)$, and thus $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Solution 1.6 Probability that no head turns up is

$$P(\text{no heads ever}) = \lim_{n \rightarrow \infty} P(\text{no heads in first } n \text{ tosses}) = \lim_{n \rightarrow \infty} 2^{-n} = 0.$$

Thus, the probability that a head will turn up eventually is $1 - P(\text{no heads ever}) = 0$.

Solution 1.7 Let $B = \bigcup_{i=1}^n A_i$. Then,

$$P\left(\bigcup_{i=1}^n A_i\right) = P(B) = \sum_{\omega \in B} P(\omega) \quad (1)$$

Likewise,

$$\sum_{i=1}^n P(A_i) = \sum_{i=1}^n \sum_{\omega \in A_i} P(\omega) \quad (2)$$

Note that each ω from B appears in (1) exactly once, whereas it appears at least once in (2). Therefore, (1) is less or at most equal to (2).

Solution 1.8 Using De Morgan laws:

$P(A \cup B \cup C) = P((A^c \cap B^c \cap C^c)^c) = 1 - P(A^c \cap B^c \cap C^c)$, which is by the multiplication rule equal to $1 - P(A^c|B^c \cap C^c)P(B^c|C^c)P(C^c)$.

Solution 1.9 (a) $P(A|B) = P(\emptyset) = 0$; (b) $P(A|B) = P(A)/P(B)$; (c) $P(A|B) = 1$.

Solution 1.10 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Thus,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1 = 0.43 + 0.87 - 1 = 0.30.$$

Solution 1.11 The probability of the sequence "red, red, black, black" (in the given order) is $P = (5/12)(6/13)(7/14)(8/15)$.

Solution 1.12 There are six pairs with sum equal 7, $(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)$. Likewise, there are two pairs with the sum equal 3. Therefore, probability of observing $S = 7$ before observing $S = 3$ is $6/(6 + 2)$.

Solution 1.13 Let A be the event that at least one six shows in four rolls of a single die. Then, A^c denotes the event that no sixes turned up. Thus, $P(A) = 1 - P(A^c) = 1 - (5/6)^4 \approx 0.518$. Likewise, let B denote the event that at least a double six shows up in twenty four rolls of two dice. Then, B^c is the event that no double sixes show up. Thus, $P(A) = 1 - P(A^c) = 1 - (35/36)^{24} \approx 0.491$, meaning that the first gamble is to be preferred.

This is **de Méré's** problem. Chevalier de Méré (1607-1684) posed the above problem to Blaise Pascal who gave the correct answer. Further communication between Pascal and Pierre de Fermat paved the way for probability theory to emerge as a new mathematical branch and academic discipline.

Solution 1.14 If the events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$. Therefore, the given probabilities are incorrect as their sum is > 1 .

Solution 1.15 (a) $1/\binom{40}{8}$;

(b) $\binom{8}{7}\binom{32}{1}/\binom{40}{8}$;

(c) add the values from (a) and (b) to $\binom{8}{6}\binom{32}{2}/\binom{40}{8}$.

Solution 1.16 (a) $(6 \times 8 \times 8 \times 6)/\binom{28}{4}$;

(b) $\binom{8}{2}\binom{8}{2}/\binom{28}{4}$;

(c) $\binom{16}{4}/\binom{28}{4}$.

Solution 1.17 The proportion of loans defaulting in a year is $0.8 \times 0.05 + 0.2 \times 0.8 = 12\%$. Let D denote default and B denote the bad loan. We need to find $P(B|D)$. Using Bayes Theorem,

$$P(B|D) = \frac{P(D|B)P(B)}{P(D|B)P(B) + P(D|B^c)P(B^c)} = \frac{0.4 \times 0.2}{0.4 \times 0.2 + 0.05 \times 0.8} = \frac{2}{3}.$$

Solution 1.18 Let H be the event that the flipped coin shows a head, and D that the coin is double headed. Note that the sample space, Ω , consists of the sides of the coins (obverse and revers), where is $|\Omega| = 6$. Then, $P(H) = 3/6 = 1/2$, whereas $P(H \cap D) = 1/3$. From here, $P(D|H) = P(H \cap D)/P(H) = (1/3)/(1/2) = 2/3$.

A naive (and wrong) argument would be as follows: When the flipped coin shows a head, it's obviously either regular or double-headed. Since the coin was picked at random it means that the probability of it being regular is equal to that of being double headed, hence $P(D|H) = 1/2$.

Solution 1.19 Let p_1 be the probability of A 's winning conditioned on the event that he plays first, and p_2 when A plays second. Then, $p_1 = a + (1 - a)p_2$ and $p_2 = (1 - b)p_1$. Solving the two equations yields

$$p_1 = \frac{1}{1 - (1 - a)(1 - b)} \text{ and } p_2 = \frac{a(1 - b)}{1 - (1 - a)(1 - b)}$$

Solution 1.20 Pairwise probabilities are:

$$P(A \cap B) = P(1) = 1/4 = P(A)P(B) = (1/2)(1/2);$$

$$P(A \cap C) = P(2) = 1/4 = P(A)P(C) = (1/2)(1/2);$$

$$P(B \cap C) = P(3) = 1/4 = P(B)P(C) = (1/2)(1/2).$$

Therefore, A, B, C are pairwise independent. However, the joint probability $P(A \cap B \cap C) = P(\emptyset) = 0$, which differs from $P(A)P(B)P(C) = 1/8$ meaning that A, B, C are not jointly independent. In other words, pairwise independence does not imply joint independence.

Note: The textbooks define **mutual independence** of n events to hold for any subset of $k \leq n$ events, where $k = 2, 3, \dots, n$. Thus, so defined mutual independence implies pairwise independence, but the reverse is not true.

Solution 1.21 According to tennis rules, when at deuce, a player (say, Bob) wins as soon as he has won two more points in total than his opponent. Note that Bob can win only after an even number $(2n + 2)$ of points have been played. Bob wins $n + 2$ while Mike wins n . Let B_{2n+2} denote the event that Bob wins at the point $2n + 2$. Note that if the win is achieved at point $2n + 2$ it means that there have been n deuces. At each of n deuces there are two ways to proceed through the game – either Bob wins a point, gaining the advantage, and then loses the next point, or Mike wins a point and then loses one. Then, at the last deuce, Bob wins two points and thus the game. Therefore, there are 2^n outcomes in the event

B_{2n+2} . Note that the last two plays are known as just two Bob's wins, so they do not contribute to the number of outcomes. However, the two last wins affect the probability of each outcome, which is $p^{n+2}(1-p)^n$. Adding probabilities of each outcome we get the probability B_{2n+2} , that Bob will win the game after n deuces:

$$P(B_{2n+2}) = 2^n p^{n+2} (1-p)^n$$

It's important to note that starting from a deuce, after an even number of steps (points played) the game is either won or it's a deuce again. After an odd number of steps either player has the advantage.

Therefore, denoting by $P(B_{2n})$ the probability that Bob wins the game at the point $2n$, we obtain the overall probability of Bob's win as:

$$\begin{aligned} P_B &= P(B_2) + P(B_4) + P(B_6) + \dots \\ &= p^2 + 2p^3(1-p) + 2^2p^4(1-p)^2 + \dots \\ &= p^2(1 + 2p(1-p) + (2p(1-p))^2 + \dots) = p^2/(1 - 2p(1-p)) \end{aligned}$$

Solution 1.22 Probability that all three coins show the same side is $P(\text{all same}) = P(HHH) + P(TTT) = 1/8 + 1/8 = 1/4$. This is the correct answer.

So, what's wrong with the reasoning given in the problem?

The problem is with confusing two sample spaces. Initial sample space does not identify the third coin, it's determined by two others that are flipped. The chance that the third coin is a head or tail is a conditional probability.

Let A denote the event that there are at least two heads. Then, $P(\text{third coin is } H|A) = 1/4$, $P(\text{third coin is } T|A) = 3/4$. Likewise, let B denote the event that there are at least two tails. Then, $P(\text{third coin is } T|B) = 1/4$, $P(\text{third coin is } H|B) = 3/4$.

As we see, in neither case is the chance even that the third coin is a head or a tail. This problem is known as *Galton's paradox (1894)*.

Solution 1.23 This is the reasoning of Lewis Carroll (who of course, knew the correct answer, but made this up as a joke). Assume that the colors are equally likely. Then,

$$P(BB) = P(WW) = 1/4, P(BW) = 1/2.$$

Carroll proposes to add a black marble to the box, shake it, and randomly pick a marble. What is now the probability of the marble being black?

Since now we have possible triplets (BBB) , (BBW) , (BWW) , by conditioning on these three cases we have

$$P(B) = 1 \times P(BBB) + 2/3 \times P(BBW) + 1/3 \times P(BWW) = (1)(1/4) + (2/3)(1/2) + (1/3)(1/2) = 2/3.$$

However, since the probability of drawing a black marble is $2/3$, then there must be two black and one white marble in the box. In other words, before adding the black marble, the box must have contained one black and one white marble!

This obviously can't be right and it's not, because in these two experiments there are two different sample spaces. It's incorrect to use the probabilities assigned to pairs of marbles from the first sample space (based on the box containing two marbles) when calculating unconditional probability of drawing a black marble from the box with three marbles.

Solution 1.24 This is the Bertrand's box paradox. The reasoning given in the text of the problem is incorrect because it mixes up unconditional probability of a marble being black or white with the conditional probability when the box is randomly picked. The correct answer is:

$P(B|b) = P(W|w) = 2/3$, where B denotes the event that both marbles are black, W that both are white, whereas b, w denote a single observed marble, either black or white, once the box is selected. In order to compute the answer we use Bayes Theorem:

$$P(B|b) = P(B)P(b|B)/P(b) = (1/3)(1)/((1/3)(1) + (1/3)(1/2) + (1/3)(0)) = 2/3.$$

Solution 1.25 One can reason that the second child is a boy or a girl with an equal chance, that is, that the probability of both being boys is 0.5.

However, this is wrong. It is indeed true that a probability of a child being a boy is 0.5, but what's asked for is the conditional probability that both children are boys given that we know that one is. In other words, we need to find the conditional probability that the other child is a boy given that we know that one is a boy.

Let B denote the event that both children are boys and A the event that at least one is a boy. Then,

$$P(B|A) = P(B \cap A)/P(A) = P(B)/P(A) = P(\{b, b\})/P(\{b, g\}, \{g, b\}, \{b, b\}) = (1/4)/(3/4) = 1/3.$$

Note that the notation $\{b, g\}$ means, for example, that the older child is a boy and the younger is a girl.

Solution 1.26 (a) Note that any coin has an equal chance to be the one picked by the owner regardless of the fact that five had been stolen. This means that the probability is $5/20 = 1/4$.

(b) Let's assume that one coin is first removed and then the thief steals five coins. This does not change the probabilities of interest. Let C denote the event that the removed coin was counterfit and let E denote the event that the thief stole at least one counterfit coin. Thus $P(E|C)$ is the probability that the thief stole at least one counterfit coin given that the removed coin was counterfit. Likewise, $P(E^c|C)$ is the probability that all the stolen coins were counterfit.

Then, $P(E|C) = 1 - P(E^c|C) = 1 - (15/19)(14/18)(13/17)(12/16)(11/15) \approx 0.742$.

Solution 1.27 The event that exactly one of A or B occurs corresponds to the symmetric difference of two events. Thus, $P(A \triangle B) = P(A \cup B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$

Solution 1.28 A "solution": The total number of possible selections of three cards from a deck of 52 is $\binom{52}{3}$. The first card can't be a spade, so the number of choices at step one is 49. After we remove the first card, there are $3 \times 13 - 1 = 38$ non-spade cards left. After removing the second card, there are 37 non-spade cards left. Thus the probability is $(39 \times 38 \times 37)/\binom{52}{3}$.

The problem is that there is a double counting in the numerator. For example, 2 of clubs, 3 of diamonds and 4 of clubs should not be distinguished from 3 of diamonds, 4 of clubs and 2 of clubs. We counted one favorable event six times!

The correct answer is $(39 \times 38 \times 37)/3! \binom{52}{3} = \binom{39}{3}/\binom{52}{3}$.

Solution 1.29 Let A_i be the event that the vertex i of the cube is black, where $i = 1, \dots, 8$. Then, $P(A_i) = 0.1$. The probability that at least one vertex is black is $P(\cup_{i=1}^8 A_i) \leq \sum_{i=1}^8 P(A_i) = 0.8$. Thus, there is a probability of at least 0.2 that the cube will have only red vertices.

Solution 1.30 Let $P(+|D) = 0.95$ be the test *sensitivity*, that is, the probability of the test being positive when the disease is present. Let $P(-|D^c) = 0.98$ be the test *specificity*, that is, the probability of the test being negative when the disease is not present. Let $P(D) = 0.01$ be the prevalence of the disease, that is, the fraction of population having the disease (then the fraction of healthy population is $P(D^c) = 0.99$).

Applying Bayes theorem:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{P(+|D)P(D)}{P(+|D)P(D) + (1 - P(-|D^c))P(D^c)}$$

$$P(D|+) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.02 \times 0.99} = 0.324$$

Solution 1.31 (a) Let A denote the event that flowers are well and alive when Alice gets back and let W be the event that Bob waters flowers. Then, by the law of total probability, $P(A) = P(A|W)P(W) + P(A|W^c)P(W^c) = (0.85)(0.9) + (0.2)(0.1) = 0.785$.

(b) Probability that Bob didn't water the flowers once Alice finds them withered away is: $P(W^c|A^c) = P(A^c|W^c)P(W^c)/P(A^c) = (0.80)(0.1)/(0.215) \approx 0.372$.

Solution 1.32 Let A denote the event that the outcome of the first trial is larger, B that the outcome is larger in the second trial, and E that the outcomes are equal in both trials. Obviously, $P(A) + P(B) + P(E) = 1$. By symmetry, $P(A) = P(B)$ so the answer is

$$P(A) = P(B) = (1 - P(E))/2 = (1 - \sum_{i=1}^n p_i^2)/2 \quad (1)$$

Here is an alternative way to solve the problem.

$$\begin{aligned} P(B) &= \sum_{i=1}^n \sum_{j>i}^n P(\text{outcomes of the first and second trials are } i, j, \text{ respectively}) \\ &= \sum_{i=1}^n \sum_{j>i} p_i p_j \end{aligned}$$

The above expression for $P(B)$ is the same as (1) as shown below:

$$\begin{aligned} 1 &= \sum_{i=1}^n p_i \sum_{j=1}^n p_j = \sum_i \sum_j p_i p_j \\ &= \sum_i p_i^2 + \sum_i \sum_{j \neq i} p_i p_j = \sum_i p_i^2 + 2 \sum_i \sum_{j>i} p_i p_j \end{aligned}$$

Solution 1.33 Assuming that the results of roulette spins are independent events, the conditional probability of the next outcome is unchanged by the previous spins, for example, that 10 spins landed on black. Mr. Smart Alek will have no advantage with his system over placing his bets on red without any knowledge of the previous results. The only advantage may be that Mr. Smart Alek will have to wait decent amount of time between any 10 consecutive black results, and thus loose less money than otherwise.

Solution 1.34 (a) Consider flipping of two fair coins and let A denote the event that the first coin shows H , B the event that the second coin shows H and C the event that both coins show the same side. Then A and B are independent, however, they are *not* independent given C .

(b) Now, roll two dice and let A denote the event that the smaller score is 3, B the event that the larger is 6, and let C be the event that the smaller score is no more than 3, and the larger is 4 or more. Then A and B are *not* independent, yet they are conditionally independent.

(c) The cases are equivalent if $P(C) = 1$.

Solution 1.35 Let H denote the event that coin lands heads, T_H that Bob is told the coin landed heads, F the event that Alice forgets the results, and C that Bob got the correct information.

(a) Using the Law of Total Probability,

$$P(T_H) = P(T_H|F)P(F) + P(T_H|F^c)P(F^c) = 0.5 \times 0.4 + P(H)(0.6) = 0.68.$$

(b) Likewise, $P(C) = P(C|F)P(F) + P(C|F^c)P(F^c) = 0.5 \times 0.4 + 1 \times 0.6 = 0.8$.

(c) Also, $P(H|T_H) = P(H \cap T_H)/P(T_H)$, and using the Law of Total Probability again,

$$\begin{aligned} P(H \cap T_H) &= P(H \cap T_H|F)P(F) + P(H \cap T_H|F^c)P(F^c) \\ &= P(H|F)P(H|T_H \cap F)P(F) + P(H)P(F^c) \\ &= 0.8 \times 0.5 \times 0.4 + 0.8 \times 0.6 = 0.64 \end{aligned}$$

Thus, $P(H|T_H) = P(H \cap T_H)/P(T_H) = 0.64/0.68 \approx 0.941$.

Solution 1.36 (a) Note that the probability that all red marbles are removed before the blue ones is the same as the probability that the last marble is blue. Each of 30 marbles has the same probability of being the last in the box, which means that the probability that such a marble is blue is $10/30$.

(b) Note that now only the green marbles remain, so that we look for the probability that the last of red or blue marbles (30 of them) is actually blue. So the desired probability that all red marbles be removed before the blue ones is again $10/30$.

(c) Let B_1, R_2, G_3 denote, respectively, the events that the all blue marbles are removed first, all red marbles are removed second and the last eight remaining marbles are green. Then, by Multiplication Rule,

$$P(B_1, R_2, G_3) = P(G_3)P(R_2|G_3)P(B_1|R_2 \cap G_3) = (8/38)(20/30)(1) = 8/57.$$

Here, $P(G_3) = 8/38$, as it is the probability that the very last marble is green;

$P(R_2|G_3) = 20/30$, as it is the probability that out of red or blue marbles it is the red marble which is removed before the last marble which is green. Since there are 20 red marbles and 10 blue, this probability is $20/30$.

The probability $P(B_1|R_2 \cap G_3)$ is obviously 1.

$$(d) P(B_1) = P(B_1, G_2, R_3) + P(B_1, R_2, G_3) = (20/38)(8/18)(8/57) \approx 0.374.$$

Solution 1.37 Let B denote the event that both cards are aces.

(a) Let A denote the event that one card is ace of spades. Then,

$$P(B|A) = P(B \cap A)/P(A) = (3/\binom{52}{2})/(51/\binom{52}{2}) = 3/51$$

(b) $P(B | \text{the first card is an ace}) = 3/51$, as the second card can be any of the remaining 51, with three aces being favorable outcomes.

(c) There is no difference whether the first or the second card is an ace, so the probability is the same as in (b). We can arrive at the same result by working out the details without reference to (b). First, let A denote the event that the second card is an ace, and C denote the event that the first card is not an ace but the second is. Then, we have,

$$\begin{aligned} P(B|A) &= P(B \cap A)/P(A) = P(B)/P(A) \\ &= P(B)/(P(B) + P(C)) \\ &= \frac{(4/52)(3/51)}{((4/52)(3/51) + (48/52)(4/51))} \end{aligned}$$

(d) Let A denote that at least one of the two cards is an ace. Note that $P(A) = 1 - (48/52)(47/51)$. Then, $P(B|A) = P(B)/P(A) = 1/33$.

Solution 1.38 Let A be the event that the drug is successful and B be the event that no side effects arise. Then, $P(A|B \cap D) = 99/100$; $P(B|D) = 95/100$; $P(A|B^c \cap D) = 30/100$; and $P(B^c|D) = 5/100$.

Then, $P(A|D) = P(A|B \cap D) \times P(B|D) + P(A|B^c \cap D) \times P(B^c|D) \approx 0.95$.

Solution 1.39 Let A denote the event that the answer is correct and K that the student knew the answer. We need to compute $P(K|A)$. We'll use Bayes theorem and hence have to determine unconditional probability of correct answer: $P(A) = P(A|K)P(K) + P(A|K^c)P(K^c) = (1)p + (1/n)(1-p)$. Hence, $P(K|A) = P(A|K)P(K)/P(A) = np/((n-1)p + 1)$.

It's worth noting that for larger n , the student is more likely that he answered correctly based on his knowledge. This justifies giving multiple choice tests.

Solution 1.40 Let's first work with an increasing sequence of events $\{A_n, n \geq 1\}$ such that $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$. Define the events B_n so that $B_1 = A_1$ and

$$B_n = A_n \cap \left(\bigcup_{i=1}^{n-1} A_i \right)^c = A_n \cap A_{n-1}^c.$$

Note that B_n consists of all points in A_n which are not in any of the previous

$A_i, i < n$. Events B_i are mutually exclusive (why?) so that $\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$ and $\cup_{i=1}^n B_i = \cup_{i=1}^n A_i$ for all $n \geq 1$. Now,

$$\begin{aligned} P(\lim_{i \rightarrow \infty} A_i) &= P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{\infty} B_i) = \sum_i P(B_i) \\ &= \lim_{n \rightarrow \infty} \sum_i^n P(B_i) = \lim_{n \rightarrow \infty} P(\bigcup_i^n B_i) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

Similarly for a decreasing sequence.

Solution 1.41 Assume that the 56 apartments are numbered as 1, 2, . . . , 56. Take as sample space the set of all unordered pairs i, j of two distinct numbers from 1, 2, . . . , 56, where each pair corresponds to two vacant apartments. The sample space has $\binom{56}{2} = 1,540$ equally likely elements. The number of elements for which there is one apartment vacant on the top floor is $\binom{52}{1} \binom{48}{1} = 384$ and the number of elements for which there are two apartments vacant on the top floor is $\binom{8}{2} = 28$. Hence the desired probability is $(384 + 28)/1,540 = 0.2675$.

An alternative solution can be based on considering an ordered sample space. Take as ordered sample the set of all 56! permutations of the apartments in the building, where the first two elements in the permutation refer to the vacant apartments. The number of permutations with no vacancy on the top floor is $48 \times 47 \times 54!$. Hence the probability that there is no vacancy on the top floor is $48 \times 47 \times 54! / 56! = 0.7325$, that is, the probability of finding an apartment on the top floor is 0.2675. It seems that you'll not be checking that building given that the probability is less than 50%.

Third way to solve the problem is using conditional probabilities. Let E_1 be the event that the first vacancy is not on the top floor and E_2 be the event that the second vacancy is not on the top floor. The desired probability is $1 - P(E_1|E_2) = 1 - P(E_1)P(E_2|E_1)$, which is equal to $1 - (48/56) \times (47/55) = 0.2675$.

Solution 1.42 (a) A same sample space can be used to answer both questions. Specify an ordered sample space whose outcomes are given by all possible 5-tuples $(i_1, i_2, \dots, i_5)$, where i_k is the exit floor for person k for $k = 1, \dots, 5$. The sample space has 105 outcomes. Let us first consider the case that the floors are equally likely to be chosen as exit floor. Then the probability $1/10^5$ is assigned to each

element of the sample outcomes. The number of outcomes for which the five persons are all going to different floors is $10 \times 9 \times 8 \times 7 \times 6$. Hence the probability that they are all going to different floors is $10 \times 9 \times 8 \times 7 \times 6 / 10^5 = 0.3024$.

(b) For the case that the 10 floors are not equally likely to be chosen as the exit floor, let each outcome $(i_1, i_2, i_3, i_4, i_5)$ have exactly k of the components that are equal to 10. Then the probability $(1/2)^k (1/18)^{5-k}$ is assigned to each outcome $(i_1, i_2, i_3, i_4, i_5)$.

The number of outcomes for which the five persons are all going to different floors and floor 10 is not chosen as the exit floor is equal to $9 \times 8 \times 7 \times 6 \times 5 = 15,120$, while the number of outcomes for which the five persons are all going to different floors and exactly one person chooses floor 10 as the exit floor is equal to $\binom{5}{1} \times 9 \times 8 \times 7 \times 6 = 15,120$. Then the probability that the persons are all going to different floors is

$$15,120 \times \left(\frac{1}{18}\right)^5 + 15,120 \times \frac{1}{2} \left(\frac{1}{18}\right)^4 = 0.080.$$

Solution 1.43 The sample space of this experiment is the set $\{(x, y) : 0 \leq x, y \leq 1\}$.

The probability $P(A)$ assigned to each subset A of the unit square is the area of the set A . For fixed a with $0 < a < 2$, let A be the subset of points (x, y) in the unit square that satisfy $x + y \leq a$. It helps to draw a picture in order to note that the area of the set A is given by $a^2/2$ for $0 < a \leq 1$ and by $1 - (2 - a)^2/2$ for $(1 < a \leq 2)$. Based on the symmetry argument, it is clear that the probability of a Manhattan distance of no more than a is also equal to $a^2/2$ for $0 < a \leq 1$ and by $1 - (2 - a)^2/2$ for $(1 < a \leq 2)$.

Solution 1.44 Let A_i be the event of the deadly shot occurring at the i th trial. Then $P(A_i) = \left(\frac{5}{6}\right)^{i-1} \frac{1}{6}$ for $i = 1, 2, \dots$. The events A_i are mutually exclusive. The probability that player A will shoot himself dead is

$$P\left(\bigcup_{k=0}^{\infty} A_{1+3k}\right) = \frac{1}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^{3k} = \frac{1/6}{1 - (5/6)^3} = 0.3956$$

The probability that desperado B will shoot himself dead is

$$P\left(\bigcup_{k=0}^{\infty} A_{2+3k}\right) = \frac{1}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^{1+3k} = \frac{1}{6} \times \frac{5}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^{3k} = (5/6) \times 0.3956 = 0.3297$$

The probability that desperado C will shoot himself dead is

$$P\left(\bigcup_{k=0}^{\infty} A_{3+3k}\right) = \frac{1}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^{2+3k} = \frac{1}{6} \times \left(\frac{5}{6}\right)^2 \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^{3k} = (5/6)^2 \times 0.3956 = 0.2747$$

Solution 1.45 Let p_k be the probability that the first ace appears at the k th card. Invoking the multiplication probability rule,

$$p_k = P(A_1 A_2 \dots A_n) = P(A_1) P(A_2 | A_1) \dots P(A_n | A_1 \dots A_{n-1}),$$

it follows that $p_1 = 4/52$, $p_2 = (48/52) \times (4/51)$, and

$$p_k = (48/52) \times (47/51) \times \dots \times \frac{48 - k + 2}{52 - k + 2} \times \frac{4}{52 - k + 1}, k = 3, \dots, 49$$

Thus, the three players do not have the same chance to become the dealer. For $X = (A, B, \text{ and } C)$, let P_X be the probability that player X becomes the dealer. Then $P_A > P_B > P_C$, because the probability p_k is decreasing in k . The probabilities can be calculated as $P_A = \sum_{k=0}^{16} p_{1+3k} = 0.36$, $P_B = \sum_{k=0}^{15} p_{2+3k} = 0.3328$, $P_C = \sum_{k=0}^{15} p_{3+3k} = 0.3072$.

Solution 1.46 Bayes rule in odds form: “*Posterior odds = Prior odds $\times$ Likelihood ratio*”.

Using little algebra, this can be expressed as

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \times \frac{P(E|H)}{P(E|H^c)}$$

where H is a hypothesis, H^c is its complement, E is evidence (data), $P(H)/P(H^c)$ are prior odds, $P(H|E)/P(H^c|E)$ are posterior odds, and $P(E|H)$ is the likelihood under hypothesis H .

Now, let the hypothesis H be the event that oil is present in the test area and

the evidence E be the event that the test gives a positive signal. This implies

$$\begin{aligned} P(H|E)/P(H^c|E) &= P(H)/P(H^c) \times P(E|H)/P(E|H^c) \\ &= (0.4/0.6) \times (0.9/0.15) = 4.0. \end{aligned}$$

Thus, the odds of finding oil are four times larger after test comes out positive. The corresponding posterior probability is 80%.

Solution 1.47 (a) Let the hypothesis H denote the event that both numbers are even and the evidence E denote the event that there is an even number among the two numbers. Then the posterior odds ratio is

$$\begin{aligned} P(H|E)/P(H^c|E) &= P(H)/P(H^c) \times P(H|E)/P(H^c|E) \\ &= (25/100)/(75/100) \times (1/(50/100)) = 2/3. \end{aligned}$$

Odds $2/3$ correspond to the probability of $2/(2+3) = 0.4$.

(b) Now, let the hypothesis H be the event that both numbers are even and the evidence E be the event that there the number two is among the two numbers. Then,

$$\begin{aligned} P(H|E)/P(H^c|E) &= P(H)/P(H^c) \times P(H|E)/P(H^c|E) \\ &= (25/100)/(75/100) \times (9/19)/(10/100) = 30/19. \end{aligned}$$

The odds $30/19$ correspond to the probability of $30/49$.

Solution 1.48 (a) Consider an ordered sample space with outcomes including all possible $5!$ orderings of the five people, where the first two people in the ordering get rum from the waiter and the other three get vodka. The number of possible orderings in which the first two people have ordered rum and the other people have ordered vodka is $2! \times 3!$. Hence the desired probability is $(2! \times 3!)/5! = 0.1$. Alternatively, we can take an unordered sample space whose outcomes are given by all possible choices of two people from the five people, where these two people get rum. This leads to the same probability of $1/\binom{5}{2} = 0.1$.

(b) In order to use conditional probabilities, let's assume that the first two drinks served are glasses of rum, and let A_1 be the event that the first rum is given to a

person having ordered rum and A_2 be the event that the second rum is also given correctly, to the other person who ordered rum. Then, the probability that the rum was served correctly is $P(A_1 \cup A_2) = P(A_1 A_2)$ and it can be evaluated as $P(A_1 A_2) = P(A_1)P(A_2|A_1) = (2/5)(1/4) = 0.1$

Solution 1.49 Given that $P(A|B) > P(A)$, we have $P(A \cup B)/P(B) > P(A)$ and thus $P(A \cup B) > P(A)P(B)$

However, $P(A \cup B) = P(B|A)P(A)$. Since $P(A \cup B) > P(A)P(B)$ we have that $P(B|A)P(A) > P(A)P(B)$, so $P(B|A) > P(B)$.

There fore if event A is “informative” about event B in the sense that the conditional probability $P(A|B)$ increases over $P(A)$ then the reverse is also true – event B is “informative” about event A .

Solution 1.50 Note that

$$A_1 = \{(3, 6), (4, 5), (6, 3), (5, 4)\}$$

$$A_2 = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (2, 6), (3, 1), (3, 2), \dots, (3, 6)\}$$

$$A_3 = \{(1, 3), (2, 3), \dots, (6, 3), (1, 4), (2, 4), \dots, (6, 4), (1, 6), (2, 6), \dots, (6, 6)\}$$

Thus, $P(A_1) = 4/36 = 1/9$, $P(A_2) = P(A_3) = 18/36 = 1/2$.

Moreover, $P(A_1 \cup A_2 \cup A_3) = P(X = 3, Y = 6) = P(\{(3, 6)\}) = 1/36$.

Now, $P(A_1)P(A_2)P(A_3) = (1/9)(1/2)(1/2) = (1/36) = P(A_1 \cup A_2 \cup A_3)$, implying that the events A_1, A_2, A_3 are jointly independent.

Checking the pairwise independence we see that

$$P(A_2 \cup A_3) = P(X \in A_2, X \in A_3) = P(X = 3) = 1/6,$$

whereas $P(A_2)P(A_3) = 1/4$. Therefore the events A_1, A_2, A_3 are not pairwise independent and therefore they are not (fully) mutually independent (or, we just say they are not independent).

PART VI

Solutions to Exercises in Probability II

Solutions to Exercises in Probability II

Solution 2.1 Let X denote the largest number among the three selected balls, that is, X is a random variable with values in $\{3, 4, \dots, 20\}$. Also, assume that any triplet of randomly selected balls is equally likely. Obviously, there are $\binom{20}{3}$ such triplets. Then,

$$P(X = i) = \binom{i-1}{2} / \binom{20}{3}, \quad i = 3, \dots, 20$$

The above equality follows because the event $\{X = i\}$ corresponds to the number of selections containing ball i and the number of selections of two other balls being numbered from 1 to $i - 1$.

The relevant probabilities are: $P(X = 20) = \binom{19}{2} / \binom{20}{3} = 0.15$, $P(X = 19) \approx 0.134$, $P(X = 18) \approx 0.119$, and $P(X = 17) \approx 0.105$.

The event $\{X \geq 17\}$ is the union of disjoint events $\{X = i\}$, where $i = 17, 18, 19, 20$ so its overall probability is the sum of individual probabilities. Hence, the probability of winning the bet is $P(X \geq 17) = 0.508$.

Solution 2.2 First we need to compute the normalizing constant k in order to ensure that $\sum_{n=0}^{\infty} P(X = n) = 1$. Thus, $k \sum_{n=0}^{\infty} \lambda^n / n! = 1$, yielding $k = \exp(-\lambda)$ since $\sum_{n=0}^{\infty} \lambda^n / n! = \exp(\lambda)$. From here,

(a) $P(X = 0) = \exp(-\lambda);$

(b) $P(X > 0) = 1 - P(X \leq 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$
 $= 1 - \exp(-\lambda) - \lambda \exp(-\lambda) - \lambda^2 \exp(-\lambda) / 2.$

Solution 2.3

$$\begin{aligned} P(X + Y = n) &= \sum_{k=0}^n P(X = n - k, Y = k) = \sum_{k=0}^n P(X = n - k) P(Y = k) \\ &= \sum_{k=0}^n \frac{\exp(-\lambda)}{(n-k)!} \frac{\exp(-\mu)}{k!} = \frac{\exp(-\lambda-\mu)}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^{n-k} \mu^k \\ &= \frac{\exp(-\lambda-\mu)}{n!} (\lambda + \mu)^n. \end{aligned}$$

The last step follows from applying the binomial theorem to the summation.

The last expression is the probability mass function of the Poisson random variable with parameter $\lambda + \mu$.

Solution 2.4

$$\begin{aligned} P(X = x, Y = y) &= P(X = x, Y = y | N = x + y) P(N = x + y) \\ &= \binom{x+y}{x} p^x q^y \lambda^{x+y} \exp(-\lambda) / (x + y)! \\ &= (\lambda p)^x (\lambda q)^y \exp(-\lambda) / (x! y!). \end{aligned}$$

An analogous result holds for Y so that the independence between X and Y follows immediately: $P(X = x, Y = y) = P(X = x)P(Y = y)$.

Solution 2.5 To show pairwise independence compute the following probabilities:

$P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = 1/4$; $P(X = 1, Z = 1) = P(X = 1)P(Z = 1) = 1/4$, and $P(Y = 1, Z = 1) = P(Y = 1)P(Z = 1) = 1/4$. With similar equations holding for value -1 instead of 1 it is clear that (X, Y) are pairwise independent, as are (X, Z) and (Y, Z) .

Joint independence does not hold because: $P(X = 1, Y = 1, Z = -1) = 0$, whereas $P(X = 1)P(Y = 1)P(Z = -1) = 1/8$.

Solution 2.6 (a) If $x \geq 1$ then $P(\min(X, Y) \leq x) = 1 - P(X > x, Y > x)$

$$= 1 - P(X > x)P(Y > x) = 1 - 4^{-x} = P(Y > X) + P(X = Y).$$

(b) Note that $P(Y > X) = P(Y < X)$ by symmetry and

$$P(Y > X) + P(Y < X) + P(X = Y) = 1.$$

Now, $P(X = Y) = \sum_x P(Y = x, X = x) = \sum_x 4^{-x} = 1/3$. Thus, $P(Y > X) = 1/3$.

(c) The answer is $1/3$, according to (b).

(d)

$$\begin{aligned} P(X \geq kY) &= \sum_{y=1}^{\infty} P(X \geq ky, Y = y) \\ &= \sum_{y=1}^{\infty} P(X \geq ky, Y = y) = \sum_{y=1}^{\infty} P(X \geq ky)P(Y = y) \\ &= \sum_{y=1}^{\infty} \sum_{x=0}^{\infty} 2^{-ky-x} 2^{-y} = 2/(2^{k+1} - 1) \end{aligned}$$

Note that $\sum_{x \geq ky} 2^{-x} 2^{-y} = \sum_{x=0}^{\infty} 2^{-ky-x} 2^{-y}$.

(e)

$$\begin{aligned} P(X \text{ divides } Y) &= \sum_{y=1}^{\infty} P(Y = kX) \\ &= \sum_{y=1}^{\infty} \sum_{x=1}^{\infty} P(Y = y, X = x) \\ &= \sum_{k=1}^{\infty} \sum_{y=1}^{\infty} 2^{-ky} 2^{-x} = \sum_{k=1}^{\infty} 1/(2^{k+1} - 1). \end{aligned}$$

(f) Let $r = m/n$ where m and n are relatively prime (that is, not having common integer factors). Then

$$\begin{aligned} P(X = rY) &= \sum_{y=1}^{\infty} P(X = km, Y = kn) \\ &= \sum_{y=1}^{\infty} \sum_{x=1}^{\infty} P(Y = y, X = x) \\ &= \sum_{k=1}^{\infty} 2^{-km} 2^{-kn} = 1/(2^{m+n} - 1) \end{aligned}$$

Note that $nX = mY$, so when x ranges over km , Y ranges over kn .

Solution 2.7 A word on notation: as written, the sum $\sum_x$, where $x \in R$, appears to be uncountable, however, all except countably many x are zero.

Also, it is required that the sum be absolutely convergent. That is because we want $E[X]$ to be unaffected by reordering of its terms.

$$\begin{aligned} E[g(X)] &= \sum_y y P(g(X) = y) = \sum_y y \sum_{\{x: g(x)=y\}} P(X = x) \\ &= \sum_y y \sum_{\{x: g(x)=y\}} f(x) = \sum_x g(x) f(x). \end{aligned}$$

This relation is known as the Law of unconscious statistician. The important point is that in order to compute the expectation of $Y = g(X)$ it is *not* necessary to know the distribution of Y . Working with the distributions (probability mass function) of X suffices.

Solution 2.8 Let $A_x = \{\omega : X(\omega) = x\} = \{X = x\}$, that is A_x is a set of all ω such that $X(\omega) = x$. Likewise, $A_y = \{\omega : X(\omega) = y\} = \{X = y\}$. Then, we can represent

$aX + bY$ as the sum:

$$aX + bY = \sum_{x,y} (ax + by) I_{A_x \cap B_y},$$

where I_A is an indicator function. Now, use the fact that $E(I_A) = P(A)$ to get

$$E(aX + bY) = E\left(\sum_{x,y} (ax + by) I_{A_x \cap B_y}\right) = \sum_{x,y} (ax + by) P(A_x \cap B_y)$$

Note that

$$\sum_y P(A_x \cap B_y) = P(A_x \cap (\bigcup_y B_y)) = P(A_x \cap \Omega) = P(A_x),$$

and similarly $\sum_x P(A_x \cap B_y) = P(B_y)$. Now,

$$\begin{aligned} E(aX + bY) &= \sum_x ax \sum_y P(A_x \cap B_y) + \sum_y by \sum_x P(A_x \cap B_y) \\ &= a \sum_x x P(A_x) + b \sum_y P(B_y) = aE(X) + bE(Y). \end{aligned}$$

Solution 2.9 To show that $X + Y$ and $|X - Y|$ are uncorrelated we find the expectation of their product $E(XY)$ and compare with the product of their expectations $E(X)E(Y)$. If $E(XY) = E(X)E(Y)$ then X and Y are uncorrelated. In other words, X and Y are uncorrelated if the covariance $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 0$.

Note that $E(X + Y) = E(X) + E(Y) = 2p = 1$. Now find $E(|X - Y|)$

$$\begin{aligned} E|X - Y| &= E\left(\sum_{x,y} |x - y| I_{\{X=x, Y=y\}}\right) = \sum_{x,y} |x - y| P(X = x, Y = y) \\ &= P(X = 1, Y = 0) + P(X = 0, Y = 1) = 1/4 + 1/4 = 1/2. \end{aligned}$$

Therefore $E(X + Y)E(|X - Y|) = (1)(1/2) = 1/2$. Furthermore,

$$E((X + Y)|X - Y|) = E(X|X - Y|) + E(Y|X - Y|)$$

Note that because of the symmetry, $E(X|X - Y|) = E(Y|X - Y|)$, so it suffices

to find $E(X|X - Y|)$ only.

$$E(X|X - Y|) = E\left(\sum_{x,y} x|x - y|I_{A_x \cap B_y}\right) = P(X = 1, Y = 0) = 1/4,$$

so that $E((X + Y)|X - Y|) = 1/4 + 1/4 = 1/2$. Therefore $E((X + Y)|X - Y|) = 1/2 = E(X + Y)E(|X - Y|)$, that is, X and Y are uncorrelated.

At the same time, $P(X + Y = 0, |X - Y| = 0) = 1/4$ and $P(X + Y = 0)P(|X - Y| = 0) = (1/4)(1/2) = 1/8$, implying that $X + Y$ and $|X - Y|$ are not independent.

Solution 2.10 Let $A_x = \{X = x\}$ and $B_y = \{Y = y\}$. Then

$$XY = \sum_{x,y} xyI_{A_x \cap B_y}$$

Expectation of XY is

$$\sum_{x,y} xyP(A_x \cap B_y) = \sum_{x,y} xyP(A_x)P(B_y) = \sum_x xP(A_x) \sum_y yP(A_y) = E(X)E(Y).$$

Solution 2.11 (a) Using the linearity property, $E(aX) = aE(X)$,

$$\begin{aligned} \text{var}(aX) &= E((aX - E(aX))^2) = E(a^2(X - E(X))^2) \\ &= a^2E((X - E(X))^2) = a^2\text{var}(X) \end{aligned}$$

(b) Note that $\text{var}(X) = E(X - EX)^2 = E(X^2) - (EX)^2$ (why?). Then, using the fact that X and Y are uncorrelated, that is, $\text{cov}(X, Y) = 0 = E(XY) - E(X)E(Y)$,

$$\begin{aligned} \text{var}(X + Y) &= E((X + Y - E(X + Y))^2) \\ &= E(X^2 + 2XY + Y^2 - (EX)^2 - (EY)^2 - 2EXEY) \\ &= \text{var}(X) + \text{var}(Y) + 2(XY - EXEY) = \text{var}(X) + \text{var}(Y). \end{aligned}$$

Solution 2.12 To solve the graph problem we use the *probabilistic* method – so, let's color randomly the vertices so that each vertex can be colored red or blue with probability $1/2$.

Let R be the *random* set of vertices colored red and N_R the size of R . Then

$E(N_R) = |E|/2$ because the vertices are colored red with probability $1/2$. However, this means that there must be a value of $N_R > |E|/2$, which proves the claim of the problem.

Note that we've used the fact that if $EX > x$ then $P(X > x) > 0$. In other words, if the mean of a random variable is greater than some value x , then $P(X > x)$ is (strictly) positive. Moreover, knowing that $P(X > x)$ is (strictly) positive, we can also say that *there exists* a value x of the random variable X , such that $X > x$.

Solution 2.13 Let's label the IC-s from 1 to 17, and let I_i be the indicator function that IC i is faulty. Let F_k be the number of faulty IC-s among the subsequence $k + 1, k + 2, \dots, k + 7$, where these numbers are used modulo 17.

Now, select *randomly* an IC and let its label be K . Each IC is equally likely to be selected. Then,

$$E(F_K) = \sum_{k=1}^{17} (1/17)I_k + (1/17)I_{k+1} + \dots + I_{k+7} = (7/17) \sum_{i=1}^{17} I_i = 35/17.$$

Therefore, we showed that $E(F_K) > 2$, and given that F_K is integer valued, it means that $P(F_K \geq 3) > 0$. In other words, it's true that there are at least three faulty IC-s among any subsequence of seven of them.

Solution 2.14 Let H denote the total number of heads. Then,

$$\begin{aligned} P(H = h) &= \sum_{n=h}^{\infty} P(H = h|N = n)P(N = n) = \sum_{n=h}^{\infty} p^h(1-p)^{n-h}\lambda^n \exp(-\lambda)/n! \\ &= \frac{(\lambda p)^h \exp(-\lambda)}{h!} \sum_{n=h}^{\infty} [(1-p)\lambda]^{n-h} \exp(-\lambda(1-p))/(n-h)! \\ &= \frac{(\lambda p)^h \exp(-\lambda)}{h!} \sum_{k=0}^{\infty} [(1-p)\lambda]^k \exp(-\lambda(1-p))/(k)! \end{aligned}$$

Note that in the 1st step we substituted k for $n - h$ to obtained a clear form of the Poisson distribution with parameter $\lambda(1-p)$. Since the sum is equal to one, the result follows – the distribution of the total number of heads is Poisson with parameter (λp) .

To summarize this result:

If $N \sim \text{Poisson}(\lambda)$ and $H_i \sim \text{Bernoulli}(p)$,
then $H_1 + H_2 \dots + H_N \sim \text{Poisson}(\lambda p)$.

Solution 2.15 (a) Assume that all $a, b \in R$. Then,

$$\begin{aligned} P(g(X) = a, h(Y) = b) &= \sum_{\mathcal{F}} P(X = x, Y = y), \text{ where } \mathcal{F} = \{x, y : g(x) = a, h(y) = b\} \\ &= \sum_{\mathcal{F}} P(X = x)P(Y = y) \\ &= \sum_{\{x: g(x)=a\}} P(X = x) \sum_{\{y: h(y)=b\}} P(Y = y) \\ &= P(g(X) = a)P(h(Y) = b) \end{aligned}$$

(b) Assume that $f_{X,Y}(x, y) = g(x)h(y)$. Then,

$$\begin{aligned} f_X(x) &= \sum_y f_{X,Y}(x, y) = g(x) \sum_y h(y), \text{ and} \\ f_Y(y) &= \sum_x f_{X,Y}(x, y) = h(y) \sum_x g(x). \end{aligned}$$

Note that

$$\sum_x f_X(x) = \sum_y f_Y(y) = \sum_x g(x) \sum_y h(y) = 1,$$

yielding

$$f_X(x)f_Y(y) = (g(x) \sum_y h(y))(h(y) \sum_x g(x)) = g(x)h(y) = f_{X,Y}(x, y).$$

Solution 2.16 (a) Consider the sum $X + Y = n$. Then,

$$\begin{aligned} P(X + Y = n) &= \sum_{k=0}^n P(X = n - k, Y = k) = \sum_{k=0}^n P(X = n - k)P(Y = k) \\ &= \sum_{k=0}^n \{\exp(-\lambda)\lambda^{n-k}/(n - k)!\} \{\exp(-\mu)\mu^k/k!\} \\ &= \exp(\lambda - \mu)/n! \sum_{k=0}^n \binom{n}{k} \lambda^{n-k} \mu^k = \exp(\lambda - \mu)(\lambda + \mu)^n/n!, \end{aligned}$$

with the binomial theorem used in the last step.

(b) Proving that the conditional distribution is binomial, we have,

$$\begin{aligned} P(X = k | X + Y = n) &= P(X = k, X + Y = n) / P(X + Y = n) \\ &= P(X = k)P(Y = n - k) / P(X + Y = n) \\ &= \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu}\right)^k \left(\frac{\mu}{\lambda + \mu}\right)^{n-k} \end{aligned}$$

The last expression for the binomial distribution follows by writing explicit expressions for all involved probabilities (all have Poisson distributions) and noting that $P(X + Y = n)$ is Poisson with parameter $\lambda + \mu$.

Solution 2.17 We have

$$\begin{aligned} \sum_{k=1}^{\infty} kP(X = k) &= \sum_{k=1}^{\infty} \sum_{n=1}^k P(X = k) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(X = k) \\ &= \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} P(X > n) \end{aligned}$$

Solution 2.18 To solve this problem we use a form of the *inclusion–exclusion* principle or formula:

$$\begin{aligned} P\left(\bigcup_{k=1}^n E_k\right) &= \sum_{1 \leq i \leq n} P(E_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(E_{i_1} \cap E_{i_2}) \\ &\quad + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \dots \\ &\quad + (-1)^{n-1} P(E_1 \cap E_2 \dots \cap E_n). \end{aligned}$$

Now, let E_i denote the event that the i -th cap is correctly located. Then, the probability that at least one cap is correctly located is

$$\begin{aligned} P\left(\bigcup_{k=1}^n E_k\right) &= n \frac{1}{n} - \binom{n}{2} \frac{1}{n(n-1)} + \binom{n}{3} \frac{1}{n(n-1)(n-2)} \dots (-1)^{n-1} \frac{1}{n!} \\ &= 1 - 1/2! + 1/3! - \dots + (-1)^{n-1}/n! \end{aligned}$$

Thus, the probability that *no* cap found its rightful owner is

$$1 - P\left(\bigcup_{k=1}^n E_k\right) = 1/2! - 1/3! - \dots + (-1)^{n-1}/n!$$

Note that this probability is $1/e \approx 0.3678794$ when n tends to infinity.

Solution 2.19 (a) We need to find $f_X(x)$ and $f_Y(y)$. Now,

$$\begin{aligned}
 f_X(x) &= P(X = x) = \sum_{y=1}^{\infty} P(X = x, Y = y) \\
 &= \frac{k}{2} \sum_{y=1}^{\infty} \frac{1}{(x+y-1)(x+y)} - \frac{1}{(x+y+1)(x+y)} \\
 &= \frac{k}{2} \sum_{y=1}^{\infty} \frac{1}{x+y-1} - \frac{2}{x+y} + \frac{1}{x+y+1}
 \end{aligned}$$

where we applied a simple identity, $\frac{2}{(a-1)(a+1)} = \frac{1}{a-1} - \frac{1}{a+1}$, to the expression that defines $f_{X,Y}(x, y)$. and a similar identity $\frac{1}{a(a+1)} = \frac{1}{a} - \frac{1}{a+1}$, to the fraction terms under the first summation above.

Now, note that the terms in the final summation form a telescoping sum for $y = 1, 2, \dots$, with the total reducing to $\frac{1}{x(x+1)}$.

Therefore, $f_X(x) = \frac{k}{2x(x+1)}$. Also, $f_X(x) = \frac{k}{2x(x+1)}$, by symmetry.

(b) To determine k note that $\sum_{x=1}^{\infty} = 1$, so it follows that $k = 2$.

(c) Note that the sum, $E(X) = \sum_{x=1}^{\infty} 1/(x+1)$, diverges. Since $E(X) = \infty$, the covariance is undefined, or we say that it does not exist.

Solution 2.20 It is helpful to use this identity: $\max\{a, b\} = (a + b)/2 + |a - b|/2$. Then,

$$\begin{aligned}
 E(\max\{X^2, Y^2\}) &= E((X^2 + Y^2)/2 + |X^2 - Y^2|/2) \\
 &= 1 + E(|X^2 - Y^2|)/2 = 1 + E(|(X - Y)(X + Y)|)/2,
 \end{aligned}$$

since $\text{var}(X) = \text{var}(Y) = E(X^2) = E(Y^2) = 1$. Furthermore,

$$\begin{aligned}
 E(|(X - Y)(X + Y)|) &= \sum_{x,y} f(x, y) |(x - y)(x + y)| \\
 &= \sum_{x,y} |\sqrt{f(x, y)}(x - y) \sqrt{f(x, y)}(x + y)| \\
 &\leq \sqrt{\sum_{x,y} f(x, y)(x - y)^2} \sqrt{\sum_{x,y} f(x, y)(x + y)^2} \\
 &= \sqrt{E[(X - Y)^2]E[(X + Y)^2]}
 \end{aligned}$$

after applying the Cauchy–Schwarz–Bunyakovski inequality.

Therefore, noting that $\rho = E(XY)$ when $EX = EY = 0$, we get

$$\begin{aligned} E(\max\{X^2, Y^2\}) &\leq 1 + \frac{1}{2}\sqrt{E[(X - Y)^2]E[(X + Y)^2]} \\ &= 1 + \frac{1}{2}\sqrt{(2 - 2\rho)(2 + 2\rho)} \\ &= 1 + \sqrt{1 - \rho^2} \end{aligned}$$

Solution 2.21 (a) Note that $\log y \leq y - 1$ with equality if and only if $y = 1$. Then,

$$E(\log f_Y(X)) - E(\log f_X(X)) = E\left(\frac{f_Y(X)}{f_X(X)}\right) \leq E\left(\frac{f_Y(X)}{f_X(X)} - 1\right)$$

Note that the expectation is with respect to X , that is,

$$E_X\left[\frac{f_Y(X)}{f_X(X)}\right] = \sum_x f_X(x) \frac{f_Y(x)}{f_X(x)} = 1,$$

where we assumed that X and Y have the same domain. Therefore $E(\log f_Y(X)) \leq E(\log f_X(X))$, with equality if and only if $f_X = f_Y$.

(b) We use the following well known inequality (*prove it!*), $\log z \geq 1 - 1/z$, with equality if and only if $z = 1$. Then,

$$\begin{aligned} I &= E_{X,Y}\left[\log\left(\frac{f(X,Y)}{f_X(X)f_Y(Y)}\right)\right] = \sum_{\mathcal{A}} f(x,y) \log\left(\frac{f(x,y)}{f_X(x)f_Y(y)}\right) \\ &\geq \sum_{\mathcal{A}} f(x,y) \left(1 - \frac{f_X(x)f_Y(y)}{f(x,y)}\right) = \sum_{\mathcal{A}} f(x,y) - \sum_{\mathcal{A}} f_X(x)f_Y(y) \\ &= 1 - \sum_{\mathcal{A}} f_X(x)f_Y(y) \geq 0. \end{aligned}$$

Therefore, $I \geq 0$. The equality holds if and only if X and Y are independent.

Note that $\mathcal{A} = \{(x, y) : f(x, y) > 0\}$, and $\sum_{\mathcal{A}} f_X(x)f_Y(y) \leq 1$.

Solution 2.22 In the following, we use the *law of unconscious statistician*, $E(g(X)) = \sum_x f(x)g(x)$,

$$\begin{aligned} E(E(Y|X)) &= E_X(E_{Y|X}(Y|X)) = \sum_x f_X(x)E(Y|X) \\ &= \sum_{x,y} f_X(x)f_{Y|X}(y) = \sum_{x,y} f_{X,Y}(x,y)y \\ &= \sum_y y \sum_x f_{X,Y}(x,y) = \sum_y y f_Y(y) = E(Y). \end{aligned}$$

The equality $E(E(Y|X)) = E(X)$ is sometimes called the *law of double expectation*. The usefulness of this result stems also from the fact that $E(Y)$ can be calculated as

$$E(Y) = \sum_x E(Y|X = x)P(X = x).$$

Solution 2.23 Note that the *law of double expectation* is obtained as the special case of the expression here, by setting $g(X) = 1$ for all x .

$$\begin{aligned} E(E(Y|X)g(X)) &= \sum_x f_X(x)g(x)E(Y|X = x) \\ &= \sum_x f_X(x)g(x) \sum_y y f_{Y|X}(y|x) = \sum_x g(x) \sum_y y f_X(x) f_{Y|X}(y|x) \\ &= \sum_x g(x) \sum_y y f_{X,Y}(x, y) = \sum_{x,y} g(x) y f_{X,Y}(x, y) = E(Yg(X)). \end{aligned}$$

Solution 2.24 We have the following distributions:

$$f_n(N) = \frac{\lambda^n}{n!} \exp(-\lambda), \quad f(K|N(k|n)) = \binom{n}{k} p^k (1-p)^{n-k}$$

Now,

$$E(K|N = n) = \sum_k k f_{K|N}(k|n) = pn$$

Thus, $E(K|N) = pN$ and $E(K) = E(E(K|N)) = E(pN) = pE(N) = p\lambda$.

In order to derive $E(K|N)$ we need the conditional mass function $f_{N|K}(N|K)$. Note that $n \geq k$. We use Bayes theorem:

$$\begin{aligned} f_{N|K}(n|k) &= P(N = n|K = k) \\ &= \frac{P(N=n)P(K=k|N=n)}{P(K=k)} = \frac{P(N=n)P(K=k|N=n)}{\sum_{j=k}^{\infty} P(K=k|N=j)P(N=j)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} (\lambda^n \exp(-\lambda)/n!)}{\sum_{j=k}^{\infty} \binom{j}{k} p^k (1-p)^{j-k} (\lambda^j / j!) \exp(-\lambda)} = \frac{\binom{n}{k} ((1-p)\lambda)^n / n!}{\sum_{j=k}^{\infty} \binom{j}{k} ((1-p)\lambda)^j / j!}, \text{ let } (1-p)\lambda = \alpha \\ &= \frac{\alpha^n / (n-k)!}{\sum_{j=k}^{\infty} \alpha^j / (j-k)!} = \frac{\alpha^{n-k} / (n-k)!}{\sum_{j=k}^{\infty} \alpha^{j-k} / j!} = \frac{\alpha^{n-k} / (n-k)!}{\sum_{i=0}^{\infty} \alpha^i / i!} \\ &= \frac{\alpha^{n-k}}{(n-k)!} \exp(-\alpha). \end{aligned}$$

Thus, we got

$$f_{N|K}(n|k) = \alpha^{n-k} \exp(-\alpha) / (n-k)!.$$

which is a shifted Poisson distribution with the expectation equal to $k + (1 - p)\lambda$,
The expectation can be derived from

$$E(N|K = k) = \sum_{n \geq k}^{\infty} n f_{N|K}(n|k)$$

and this is left as an exercise for the student.

Solution 2.25 The relevant set to consider consists of all x, y pairs which sum to z .
The set can be expressed as

$$\{X + Y = z\} = \bigcup_x (\{X = x\} \cap \{Y = z - x\}).$$

Note that the union consists of disjoint subsets as it is partitioned by single values x . We can say that at most countably many terms have non-zero probability.
Then,

$$P(X + Y = z) = \sum_x P(X = x, Y = z - x) = \sum_x f_{X,Y}(x, z - x)$$

Note that if X and Y are independent we have

$$\begin{aligned} P(X + Y = z) &= f_{X+Y}(z) = \sum_x f_{X,Y}(x, z - x) \\ &= \sum_x f_X(x) f_Y(z - x) \\ &= \sum_y f_X(z - y) f_Y(y) \end{aligned}$$

The mass function of $X + Y$ is the convolution of the mass functions of X and Y .

Solution 2.26 Let X be the number of defective parts in a package. Note that the probability of a defective part is 0.01.

Obviously, X is a binomial variable $X \sim \text{Binomial}(n, p)$, where $n = 10, p = 0.01$.

The company has to replace a package if there are 2 or more defects in it. There-

fore, the probability that the package will need to be replaced is

$$1 - P(X = 0) - P(X = 1) = 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \approx 0.04.$$

Thus, about 0.4% of the packages need to be replaced.

Solution 2.27 There is not enough information to answer the question.

In case the defendant is innocent, the probability of the jury making a correct decision (thus freeing defendant) is

$$\sum_{i=5}^{12} \binom{12}{i} p^i (1-p)^{12-i}.$$

If the defendant is guilty the probability of the jury making a correct decision (to convict) is

$$\sum_{i=5}^{12} \binom{12}{i} p^i (1-p)^{12-i}.$$

If we assume that the defendant is guilty with probability θ then we can condition on whether he is guilty or not we get the probability of a correct decision as follows.

$$(1 - \theta) \sum_{i=5}^{12} \binom{12}{i} p^i (1-p)^{12-i} + \theta \sum_{i=5}^{12} \binom{12}{i} p^i (1-p)^{12-i}.$$

Solution 2.28 Let's consider the ratio $P(X = k)/P(X = k - 1)$ and find out for which values of k is the ratio greater or less than 1.

$$\frac{P(X=k)}{P(X=k-1)} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)}$$

Thus, $P(X = k) \geq P(X = k - 1)$ if and only if $(n - k + 1)p \geq k(1 - p)$, which is the same as $k \leq (n + 1)p$.

Solution 2.29 A binomial random variable $X \sim \text{Binom}(n, p)$ can be written as the sum of n independent Bernoulli variables, B_i , where $B_i = 1$ with probability p :

$$X = B_1 + B_2 + \dots + B_n, \quad i = 1, 2, \dots, n.$$

Then, $E(X) = E(B_1 + B_2 + \dots + B_n) = E(B_1) + E(B_2) + \dots + E(B_n) = np$.

Likewise, $V(X) = V(B_1 + \dots + B_n) = V(B_1) + \dots + V(B_n) = np(1 - p)$.

Note that the expectation of a sum of random variables is the sum of the corresponding expectations regardless of the random variables being independent or not. However, the variance of the sum is equal to the sum of the variances only if the variables are pairwise uncorrelated. Since in our case they are independent, it is sufficient for the result to hold.

Solution 2.30 We start from the definition to find the expectation of a Poisson random variable.

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \lambda^k \exp(-\lambda)/k! = \lambda \sum_{k=1}^{\infty} \lambda^{k-1} \exp(-\lambda)/(k-1)! \\ &= \lambda \exp(-\lambda) \sum_{k=0}^{\infty} \lambda^k/k! = \lambda. \end{aligned}$$

Note the changes in indexes, and also the last step where we used the fact that $\exp(\lambda) = \sum_{k=0}^{\infty} \lambda^k/k!$

To find variance, we first need $E(X^2)$ since $V(X) = E(X^2) - (E(X))^2$.

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \lambda^k \exp(-\lambda)/k! = \lambda \sum_{k=1}^{\infty} k \lambda^{k-1} \exp(-\lambda)/(k-1)! \\ &= \lambda \sum_{k=0}^{\infty} (k+1) \lambda^k \exp(-\lambda)/k! \\ &= \lambda \left\{ \sum_{k=0}^{\infty} k \lambda^k \exp(-\lambda)/k! + \sum_{k=0}^{\infty} \lambda^k \exp(-\lambda)/k! \right\} \\ &= \lambda(\lambda + 1). \end{aligned}$$

Therefore, $V(X) = E(X^2) - (E(X))^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda$.

Solution 2.31 The expectation and variance of a binomial random variable X are: $E(X) = np$ and $V(X) = np(1 - p)$. Assume that n is large, p small and np moderate. Now, for a binomial X and $\lambda = np$, we have

$$\begin{aligned} P(Y = k) &= \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n!}{(n-k)!k!} (\lambda/n)^k (1 - \lambda/n)^{n-k} \\ &= \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} \frac{\lambda^k (1 - \lambda/n)^n}{k! (1 - \lambda/n)^k} \end{aligned}$$

Based on n large and p being small, we use the following approximations:

$$(1 - \lambda/n)^n \rightarrow \exp(-\lambda); \quad (1 - \lambda/n)^k \rightarrow 1;$$

$$n(n-1)(n-2)\dots(n-k+1)/n^k \rightarrow 1;$$

to arrive at $P(X = k) \approx \exp(-\lambda)\lambda^k/k! = \exp(-np)(np)^k/k!$

Therefore, the number of successes occurring is approximately $\text{Poisson}(np)$.

In practice, to approximate a binomial random variables with $(n \geq 100)$ and for small n , say, $n \leq 0.01$, we can use a Poisson with $\lambda = np \leq 20$.

Solution 2.32 We have $Y = (X - \mu)/\sigma$, so that

$$E(Y) = E((X - \mu)/\sigma) = E(X - \mu)/\sigma = (E(X) - \mu)/\sigma = (\mu - \mu)/\sigma = 0.$$

Likewise,

$$V(Y) = V(X - \mu)/\sigma^2 = V(X)/\sigma^2 = \sigma^2/\sigma^2 = 1.$$

Solution 2.33 According to their definition, Y_i are independent and identically distributed. Let $E(Y_i) = \mu_i$. Consider the sum $Y = Y_1 + Y_2 + \dots + Y_n$. Obviously, $Y = 1$. Then, $E(Y) = 1 = E(Y_1) + E(Y_2) + \dots + E(Y_n) = n\mu_i$. Therefore $E(Y_i) = 1/n$ for all $i = 1, 2, \dots, n$.

Solution 2.34 We need to prove that $(E(XY))^2 \leq E(X^2)E(Y^2)$, which is a variant of Cauchy-Schwarz inequality. Consider the expression $E(X - aY)^2$ where $a \in R$. Then, $E(X - aY)^2 \geq 0$ for all a . Furthermore, $E(X^2) - 2aE(XY) + a^2E(Y^2) \geq 0$.

Now, a value of a that minimizes the above expression is $a = E(XY)/E(Y^2)$. In order to do it, we find the first derivative with respect to a and set to zero to obtain the result. (Why is this a minimum and not a maximum?) The inequality becomes $E(X^2) - 2(E(XY))^2/E(Y^2) + (E(XY))^2/E(Y^2) \geq 0$ which is the same as $(E(XY))^2 \leq E(X^2)E(Y^2)$.

Solution 2.35 Cauchy-Schwarz inequality of the previous problem states that

$$(E(XY))^2 \leq E(X^2)E(Y^2).$$

This implies

$$(E((X - \mu_X)(Y - \mu_Y)))^2 \leq E((X - \mu_X)^2)E((Y - \mu_Y)^2)$$

or $\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2$. Therefore $\rho_{XY}^2 = \sigma_{XY}^2 / \sigma_X^2 \sigma_Y^2 \leq 1$.

Then, it follows that $-1 \leq \rho_{XY} \leq 1$.

Solution 2.36 (a) We have that $Y = aX + b$. Thus, $E(XY) = E(X(aX + b)) = aE(X^2) + bE(X)$. Also, $E(Y) = aE(X) + b$. Furthermore, the covariance is,

$$\begin{aligned} \text{Cov}(X, Y) &= \sigma_{XY} = E(XY) - E(X)E(Y) \\ &= aE(X^2) + bE(X) - E(X)(aE(X) + b) \\ &= a(E(X^2) - (E(X))^2) = a\sigma_X^2 \end{aligned}$$

(b) We first find σ_Y by noting that $V(Y) = V(aX + b) = a^2V(X)$. Thus, $\sigma_Y^2 = a^2\sigma_X^2$, or $\sigma_Y = |a|\sigma_X$. Then, the correlation coefficient of X and Y is

$$\rho_{XY} = \sigma_{XY} / (\sigma_X \sigma_Y) = a\sigma_X^2 / (\sigma_X |a| \sigma_X) = \frac{a}{|a|}.$$

Thus, $\rho_{XY} = a/|a|$, implying $\rho_{XY} = 1$, for $a > 0$, and $\rho_{XY} = -1$, for $a < 0$.

Solution 2.37 If $X_1, X_2, \dots, X_n$ are pairwise independent then

$$\text{Var} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i).$$

Assume now that X_i -s are not independent and let $Y = \sum_{i=1}^n a_i X_i$. Then,

$$\begin{aligned} \text{Var}(Y) &= E((Y - EY)^2) = E \left\{ \left(\sum_{i=1}^n a_i (X_i - EX_i) \right)^2 \right\} \\ &= E \left\{ \sum_{i=1}^n \sum_{j=1}^n a_i a_j (X_i - EX_i)(X_j - EX_j) \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E[(X_i - EX_i)(X_j - EX_j)] \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j) \end{aligned}$$

Now, if X_i, X_j are pairwise uncorrelated, then $\text{Cov}(X_i, X_j) = \text{Var}(X_i)$ for $i = j$ and $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$. Therefore, for pairwise uncorrelated X_i, X_j , we

obtain the same result as for independent X_i :

$$\text{Var} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i).$$

Solution 2.38 (a) The number of possible trials has no upper bound. Thus, the probability of the event E that the success will occur eventually is

$$P(E) = \sum_{n=1}^{\infty} P(X = n) = p \sum_{n=1}^{\infty} (1-p)^{n-1}.$$

Since the geometric series sums to one it follows that $P(E) = p$, that is, there is a positive probability that the success will eventually occur.

(b) Let $q = 1 - p$. Then,

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} nq^{n-1}p = p \sum_{n=1}^{\infty} \frac{d}{dq}(q^n) = p \frac{d}{dq} \left(\sum_{n=1}^{\infty} q^n \right) \\ &= p \frac{d}{dq} (1/(1-q)) = p/(1-q)^2 = 1/p. \end{aligned}$$

Since $V = E(X^2) - (E(X))^2$ we first find $E(X^2)$ and also let $q = 1 - p$. Then,

$$\begin{aligned} E(X^2) &= \sum_{n=1}^{\infty} n^2 q^{n-1} p = p \sum_{n=1}^{\infty} \frac{d}{dq}(nq^n) = p \frac{d}{dq} \left(\sum_{n=1}^{\infty} nq^n \right) \\ &= p \frac{d}{dq} \left[\frac{q}{1-q} \left(\sum_{n=1}^{\infty} nq^{n-1} p \right) \right] = p \frac{d}{dq} \left[\frac{q}{1-q} E(X) \right] = p \frac{d}{dq} \left[\frac{q}{(1-q)^2} \right] \\ &= 2/p^2 - 1/p. \end{aligned}$$

Solution 2.39 By definition $V(X) = \sigma^2 = E((X - E(X))^2)$, hence

$$\begin{aligned} \sigma^2 &= \sum_{x \in D} p(x)((x - \mu)^2) = \sum_{x \in D} p(x)(x^2 - 2x\mu + \mu^2) \\ &= \sum_{x \in D} p(x)x^2 - 2 \sum_{x \in D} p(x)x\mu + \sum_{x \in D} p(x)\mu^2 \\ &= \sum_{x \in D} p(x)x^2 - 2\mu \sum_{x \in D} p(x)x + \mu^2 \sum_{x \in D} p(x) \\ &= \sum_{x \in D} p(x)x^2 - 2\mu^2 + \mu^2 \\ &= \sum_{x \in D} p(x)x^2 - \mu^2 \end{aligned}$$

Solution 2.40 (a) Let $p(x) = c/x^2, x = \pm 1, \pm 2, \dots$. To determine c we use the fact that the sum of all probabilities is one. Thus, $\sum_{x=1}^{\infty} 1/x^2 = \pi^2/6$, and $\sum_{x=-1}^{-\infty} 1/x^2 = \pi^2/6$, so that $c = 3/\pi^2$. However,

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xp(x) + \sum_{x=-1}^{-\infty} xp(x) \\ &= 3/\pi^2 \sum_{x=1}^{\infty} 1/x + 3/\pi^2 \sum_{x=-1}^{-\infty} 1/x \\ &= \infty - \infty \end{aligned}$$

Therefore, the expectation is not well defined – it does not exist.

(b) Let $p(x) = 2c/x^2, x \geq 1$. Then, as in (a) we have $E(X) = 2c \sum_{x=1}^{\infty} 1/x = \infty$, so the expectation is infinite.

(c) Here, $p(x) = c/x^3, x \geq 1$. The expectation is

$$E(X) = \sum_{x=1}^{\infty} xp(x) = c \sum_{x=1}^{\infty} 1/x^2 = c\pi^2/6$$

Thus, the mean is finite. However,

$$E(X^2) = \sum_{x=1}^{\infty} x^2 p(x) = c \sum_{x=1}^{\infty} 1/x = \infty,$$

so the variance is infinite.

Solution 2.41 Starting from the definition of the conditional variance,

$$V(Y|X = x) = E((Y - E(Y|X = x))^2|X = x),$$

we have

$$\begin{aligned} V(Y) &= E((Y - EY)^2) = E(E((Y - E(Y|X))^2)|X) \\ &= E(E((Y - E(Y|X) + E(Y|X) - EY)^2)|X) \\ &= E(E((Y - E(Y|X))^2)|X) + E(E((E(Y|X) - EY)^2)|X) \\ &\quad + 2E[E((Y - E(Y|X))(E(Y|X) - EY))|X] \end{aligned}$$

Note that the last term in the sequence of above equalities is zero because

$E((Y - E(Y|X))|X) = EY - E(E(Y|X)) = EY - EY = 0$. Hence,

$$\begin{aligned} V(Y) &= E(E((Y - E(Y|X))^2|X)) + E((E(Y|X) - EY)^2|X) \\ &= E(V(Y|X)) + V(E(Y|X)). \end{aligned}$$

Solution 2.42 Let the random variable N be the number of particles emitted in the given time interval. Note that $P(X = j, Y = k) = P(X = j, Y = k, N = j + k)$ and $P(AB) = P(A | B)P(B)$. Then, it follows that

$$\begin{aligned} P(X = j, Y = k) &= P(X = j, Y = k | N = j + k)P(N = j + k) \\ &= \binom{j+k}{j} p^j (1-p)^k e^{-\lambda} \frac{\lambda^{j+k}}{(j+k)!} \end{aligned}$$

$$P(X = j, Y = k) = e^{-\lambda} \frac{(\lambda p)^j}{j!} \times e^{-\lambda(1-p)} \frac{(\lambda(1-p))^k}{k!} \quad \text{for all } j, k \geq 0$$

Note further that $P(X = j) = \sum_{k=0}^{\infty} P(X = j, Y = k)$ and $P(Y = k) = \sum_{j=0}^{\infty} P(X = j, Y = k)$. Therefore, it follows that

$$P(X = j) = e^{-\lambda p} \frac{(\lambda p)^j}{j!} \quad \text{and} \quad P(Y = k) = e^{-\lambda(1-p)} \frac{(\lambda(1-p))^k}{k!}$$

for all $j, k \geq 0$.

This means that X and Y are Poisson random variables with parameters λp and $\lambda(1-p)$. Given that $P(X = j, Y = k) = P(X = j)P(Y = k)$ for all j, k , we have shown that X and Y are independent.

Solution 2.43 (a) Let the random variable X be the payoff of the game. Then,

$$\begin{aligned} P(X = 0) &= \frac{6 \times 5 \times 5 \times 5 \times 5}{6^4} = \frac{125}{216} \\ P(X = 2) &= \frac{6 \times \binom{3}{1} \times 1 \times 5 \times 4}{6^4} = \frac{60}{216} \\ P(X = 2.5) &= \frac{6 \times \binom{3}{1} \times 1 \times 5 \times 1}{6^4} = \frac{15}{216} \\ P(X = 3) &= \frac{6 \times \binom{3}{2} \times 1 \times 1 \times 5}{6^4} = \frac{15}{216} \\ P(X = 4) &= \frac{6 \times 1 \times 1 \times 1 \times 1}{6^4} = \frac{1}{216} \end{aligned}$$

Alternatively, using conditional probabilities, the probability mass function of

X can be calculated as $P(X = 0) = (\frac{5}{6})^3$, $P(X = 2) = \binom{3}{1} \frac{1}{6} (\frac{5}{6})^2 \times \frac{4}{5}$, $P(X = 2.5) = \binom{3}{1} \frac{1}{6} (\frac{5}{6})^2 \times \frac{1}{5}$, $P(X = 3) = \binom{3}{2} (\frac{1}{6})^2 \frac{5}{6}$, and $P(X = 4) = (\frac{1}{6})^3$ Hence

$$E(X) = 0 \times \frac{125}{216} + 2 \times \frac{60}{216} + 2.5 \times \frac{15}{216} + 3 \times \frac{15}{216} + 4 \times \frac{1}{216} = 0.956$$

We see that the expected payoff is $E(X) - 1 = -0.044 < 0$.

(b) Given that the payoff is negative, it means that the game is disadvantageous to the player. The house has a large edge of 4.4%, so the game should not be played unless one enjoys “free” drinks over simple reasoning.

Solution 2.44 To solve the problem requires some pondering until we realize that the desired probability that the player A will be the overall winner is the probability that A wins at least $10 - a$ games of the additional $19 - a - b$ games. Take as sample space all possible sequences of “ones” and “zeros” to a length of $19 - a - b$, where A ’s win is recorded as a “one” and a loss as a “zero”. Assign to each element of the sample space the same probability of $1/(19 - a - b)$. Let A_k be the event that the player A wins exactly k of these games. The set A_k contains $\binom{19-a-b}{k}$ elements. The events A_k are disjoint and so the resulting probability is the sum of individual probabilities:

$$\sum_{k=10-a}^{19-a-b} P(A_k) = \sum_{k=10-a}^{19-a-b} \binom{19-a-b}{k} (1/2)^{19-a-b}.$$

Solution 2.45 (a) The set of the possible values of the random variable X is $\{101, 102, \dots, 110\}$. Then, the probability mass function is given by

$$p_X(k) = \begin{cases} P(X > k-1) - P(X > k), & \text{if } k = 101, 102, \dots, 110, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $P(X > 100) = 1$ and that for $k = 101, 102, \dots, 110$, we have

$$\begin{aligned} P(X > k) &= P(X_1 > k, X_2 > k, X_3 > k) \\ &= P(X_1 > k)P(X_2 > k)P(X_3 > k) = (110 - k)^3/10^3. \end{aligned}$$

Then, it follows that

$$p_X(k) = \begin{cases} \frac{(111-k)^3 - (110-k)^3}{10^3}, & \text{if } k = 101, 102, \dots, 110, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Since X_i are uniformly distributed over the integers in the set $\{101, 102, \dots, 110\}$, we have $E(X_i) = (101 + 110)/2 = 105.5$.

$$E(X) = \sum_{k=-\infty}^{\infty} kp_X(k) = \sum_{k=101}^{110} kp_X(k) = \sum_{k=101}^{110} \frac{(111-k)^3 - (110-k)^3}{10^3}$$

Evaluating the above expression yields $E(X) = 103.025$. Therefore, the expected improvement is $105.5 - 103.025 = 2.475$.

Solution 2.46 Let X_i be the number of traffic offenders the policeman stops on day i . Then, X_i are independent random variables that with an equal chance assume values in the set $\{1, 2, 3, 4, 5, 6\}$. We are interested in $X = \sum_{i=1}^{10} X_i$, where

$$E(X) = E\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} E(X_i) = 35.$$

where we used the value for $E(X_i) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$.

Using the fact that the X_i -s are independent, the variance is computed as:

$$V(X) = V\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} V(X_i) \approx 29.17.$$

Here, we used the result that $V(X_i) = E(X_i^2) - (E(X_i))^2 = (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - 3.5^2 = 2.917$.

Solution 2.47 (a) We note that the X has the binomial distribution with parameters of the corresponding pmf, $p = 0.02$ and $n = 250$. The mean is $E[X] = np = 250 \times 0.02 = 5$. Then the probability that the number of tickets one gets is exactly equal to the expected value of X is

$$P(X = 5) = \binom{250}{5} 0.02^5 0.98^{245} = 0.1773.$$

(b) In order to use Poisson approximation, we set $\lambda = np = 250 \times 0.02 = 5$. Therefore, the approximate probability is

$$P_{Poisson}(X = 5) = \exp(-\lambda)\lambda^5/5! = 0.1755.$$

The two probabilities, binomial and Poisson, are very close to each other.

(c) Let Y denote the total money paid for traffic tickets in a year. Then, $E(Y) = 50 \sum_{i=1}^5 E(Y_i)$, where Y_i is the amount paid on day i .

We have the following probabilities of paying the traffic tickets:

$$P(Y_i = y_i) = \begin{cases} 0.98, & \text{if } y = 0 \\ 0.01, & \text{if } y = 10 \\ 0.006, & \text{if } y = 20 \\ 0.004, & \text{if } y = 50. \end{cases}$$

Then, the mean is $E(Y_i) = (0.01)(10) + (0.006)(20) + (0.004)(50) = 0.42$.

The variance is, $V(Y_i) = E(Y_i^2) - E(Y_i)^2 = (0.01)(10^2) + (0.006)(20^2) + (0.004)(50^2) - (0.42)^2 = 13.22$.

Then, the mean of Y is $E(Y) = 250E(Y_i) = 105$. Now, use the independence of Y_i to calculate $V(Y) = 250V(Y_i) = 3,305$.

(d) The standard deviation of the sample mean is

$$\sigma = \sqrt{p(1-p)/n} = \sqrt{p(1-p)/250}.$$

Then, the possible values of p within 5 standard deviations from $\hat{p}$ are those that satisfy $0 < p < 1$ and

$$|p - \hat{p}| \leq 5\sigma, \text{ or } (p - \hat{p})^2 \leq 25p(1-p)/250.$$

When simplified, the inequality is $275p^2 - 35p + 0.1 \leq 0$, which is satisfied for $0.0025 \leq p \leq 0.1245$.

Solution 2.48 (a) It will be helpful to recall the properties of the joint probability

mass function of a bivariate (X, Y) :

$$\begin{aligned} 0 &\leq p_{XY}(x, y) \leq 1; \\ \sum_x \sum_y p_{XY}(x, y) &= 1; \\ P[(X, Y) \in \mathcal{A}] &= \sum_{(x, y) \in R_{\mathcal{A}}} p_{XY}(x, y), \end{aligned}$$

where $R_{\mathcal{A}}$ is the range space of X and Y corresponding to $\mathcal{A}$.

Then, we have

$$\sum_x \sum_y p_{XY}(x, y) = \sum_{x=1}^2 \sum_{y=1}^2 k(2x + y) = k(18) = 1.$$

Therefore, $k = 1/18$.

(b) Noting the properties of joint pmf-s, the marginal pmf of X is

$$\begin{aligned} p_X(x) &= \sum_y p_{XY}(x, y) = \sum_{y=1}^2 \frac{1}{18} (2x + y) \\ &= \frac{1}{18} (2x + 1) + \frac{1}{18} (2x + 2) = \frac{1}{18} (4x + 3), \quad x = 1, 2. \end{aligned}$$

Likewise, the marginal pmf of Y is

$$\begin{aligned} p_Y(y) &= \sum_x p_{XY}(x, y) = \sum_{x=1}^2 \frac{1}{18} (2x + y) \\ &= \frac{1}{18} (2 + y) + \frac{1}{18} (4 + y) = \frac{1}{18} (2y + 6), \quad y = 1, 2. \end{aligned}$$

(c) Now, $p_X(x_i) p_Y(y_j) \neq p_{XY}(x_i, y_j)$; implying that X and Y are *not* independent.

Solution 2.49 (a) Let's recall again the properties of joint pmf-s:

$$\begin{aligned} 0 &\leq p_{XY}(x, y) \leq 1; \\ \sum_x \sum_y p_{XY}(x, y) &= 1; \\ P[(X, Y) \in \mathcal{A}] &= \sum_{(x, y) \in R_{\mathcal{A}}} p_{XY}(x, y), \end{aligned}$$

where $R_{\mathcal{A}}$ is the range space of X and Y corresponding to $\mathcal{A}$. Now, we have

$$\begin{aligned}\sum_x \sum_y p_{XY}(x, y_j) &= \sum_{x_i=1}^2 \sum_{y_j=1}^3 kx_i^2 y_j \\ &= k(1 + 2 + 3 + 4 + 8 + 12) = k(30) = 1\end{aligned}$$

Thus, $k = \frac{1}{30}$. (b) The marginal pmf of X is

$$p_X(x) = \sum_y p_{XY}(x, y) = \sum_{y=1}^3 x^2 y / 30 = \frac{1}{5} x^2 \quad x_i = 1, 2$$

Then, the marginal pmf of Y is

$$p_Y(y) = \sum_x p_{XY}(x, y) = \sum_{x=1}^2 x^2 y / 30 = \frac{1}{6} y, \quad y = 1, 2, 3.$$

c) Now, $p_x(x) p_y(y) = x^2 y / 30 = p_{xy}(x, y)$ From here, we conclude that X and Y are independent.

Solution 2.50 Random variables $X_1, X_2, \dots, X_n$ are independent and have the same distributions. That means that the variance of the sum of individual variances is the sum of individual variances σ^2 .

Let $S_n = X_1 + X_2 + \dots + X_n$. Then

$$E(S_n) = \mu \quad \text{and} \quad V(S_n) = n\sigma^2.$$

Furthermore,

$$V\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n} \quad \text{and also} \quad E\left(\frac{S_n}{n}\right) = \mu$$

Now we apply the Chebyshev's Inequality, so for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Therefore, for a fixed ϵ

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - \mu \right| \geq \epsilon \right) = 0.$$

An equivalent expression of the previous result is:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - \mu \right| < \epsilon \right) = 1.$$

PART VII

Solutions to Exercises in Probability III

Solutions to Exercises in Probability III

Solution 3.1 (a) Given that f is a probability density function, it has to satisfy

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Then, $c \int_0^2 (2x - x^2)dx = 1$, or $1 = c(x^2 - x^3/3)|_0^2$, or $c = 3/4$.

$$(b) P(X > 1) = \int_1^{\infty} f(x)dx = (3/4) \int_1^2 (2x - x^2)dx = (3/4)((x^2 - x^3/3)|_1^2) = 2/3.$$

Solution 3.2 (a) We first find λ . The density function satisfies $\int_{-\infty}^{\infty} f(x)dx = 1$ or $\lambda \int_{-\infty}^{\infty} \exp(-x/100)dx = 1$. This yields $1 = -100\lambda \exp(-x/100)|_0^{\infty} = 100\lambda$. Thus, $\lambda = 1/100$

The probability that the machine would function between 50 and 150 days is

$$\begin{aligned} P(50 < X < 150) &= \int_{50}^{150} (1/100) \exp(-x/100)dx = -\exp(-x/100) \Big|_{50}^{150} \\ &= \exp(-1/2) - \exp(-3/2) \approx 0.384. \end{aligned}$$

(b) Analogously to (a) we have

$$\begin{aligned} P(X < 100) &= \int_0^{100} (1/100) \exp(-x/100)dx = -\exp(-x/100) \Big|_0^{100} \\ &= 1 - \exp(-1) \approx 0.633. \end{aligned}$$

Thus, the machine will break down about 63.3% of time before working 100 hours with no problem.

Solution 3.3 Let $Y = \exp(X)$ Then, the cumulative distribution function of Y is

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(\exp(X) < y) = P(X < \log(y)) \\ &= \int_0^{\log(y)} f(t)dt = \log(y). \end{aligned}$$

The density function of Y is $f_Y(y) = d/dx(F_Y(y)) = 1/y$. Note that y has different limits from x , that is, $1 < y < e$. Now, $E(\exp(X)) = E(Y) = \int_{-\infty}^{\infty} t f_Y(t)dt =$

$$\int_1^e dt = e - 1.$$

An alternative (and more general) way to derive $E(\exp(X))$ is to use the continuous version of the law of *unconscious statistician*,

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Then, $E(\exp(X)) = \int_0^1 \exp(x) dx = e - 1$.

Solution 3.4 Let $f_X(x)$ be the probability density function of the continuous random variable X . Then,

$$\int_0^{\infty} P(X > x) dx = \int_0^{\infty} \int_x^{\infty} f_X(t) dt dx,$$

where $P(X > y) = \int_y^{\infty} f_X(t) dt$. To finish the proof, we interchange the bounds of integration:

$$\int_0^{\infty} P(X > x) dx = \int_0^{\infty} \left(\int_0^t dx \right) f_X(t) dt = \int_0^{\infty} t f_X(t) dt = E(X).$$

Solution 3.5 Note that a function of a random variable is also a random variable, and since $g \geq 0$, we can start from the result of the previous exercise, $E(g(X)) = \int_0^{\infty} P(g(X) > t) dt$.

Then, we have $P(g(X) > t) = \int_{\mathcal{A}} f_X(x) dx$, where $\mathcal{A} = \{x : g(x) > t\}$. Furthermore,

$$E(g(X)) = \int_0^{\infty} \int_{\mathcal{A}} f_X(x) dx dt = \int_{\mathcal{A}} \left(\int_0^{g(x)} dt \right) f_X(x) dx = \int_{\mathcal{A}} g(x) f_X(x) dx.$$

Solution 3.6 We have that $X \sim \text{Uniform}(0, 1)$ and $Y = a + (b - a)X$. Hence, the cumulative distribution function (CDF) of Y is

$$F_Y(y) = P(Y \leq y) = P(a + (b - a)X \leq y) = P(X \leq \frac{y - a}{b - a}) = F_X\left(\frac{y - a}{b - a}\right).$$

Note that the range of Y is $a \leq y \leq b$. Differentiating F_X , we get $f_Y(y) = 1/(b - a)$.

Solution 3.7 (a) Assume that the Bernoulli trials are performed at times $\epsilon, 2\epsilon, 3\epsilon$ and let T be the waiting time for the first success. Then $P(T > k\epsilon) = (1 - p)^k$, where for a fixed time, t , the number of trials (time epochs) is approximately $k = t/\epsilon$. Now, let $\epsilon \rightarrow 0$ and $p \rightarrow 0$ in such a way that p/ϵ tends to a positive constant λ or $p \approx \lambda\epsilon$. Then,

$$P(T > t) = P(T > k\epsilon) = (1 - p)^k \approx (1 - \lambda\epsilon)^{t/\epsilon} \rightarrow \exp(-\lambda t).$$

(b) Note that $P(X > x) = 1 - F(x)$, and so

$$EX = \int_0^{\infty} P(X > x) dx = \int_0^{\infty} (1 - F(x)) dx = \int_0^{\infty} \exp(-\lambda x) dx = 1/\lambda.$$

Solution 3.8 The density of the distribution of the exponential random variable is $f_X(x) = \lambda \exp(-\lambda x)$.

Thus, the mean of X is $E(X) = \int_0^{\infty} x \lambda \exp(-\lambda x) dx$.

We integrate by parts ($u = x, dv = \lambda \exp(-\lambda x) dx$) to obtain

$$E(X) = -x \exp(-\lambda x) \Big|_0^{\infty} + \exp(-\lambda x) dx = 1/\lambda.$$

Next, we find $E(X^2) = \int_0^{\infty} x^2 \lambda \exp(-\lambda x) dx$. Integrating by parts now requires substitution $u = x^2, dv = \lambda \exp(-\lambda x)$, yielding

$$E(X^2) = -x^2 \exp(-\lambda x) \Big|_0^{\infty} + 2 \int_0^{\infty} x \exp(-\lambda x) dx = 1/\lambda = 2/\lambda^2.$$

Therefore, $V(X) = E(X^2) - (EX)^2 = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2$.

Solution 3.9 Let Φ denote the $N(0, 1)$ distribution function. Then,

$P(Y \leq y) = P(X \leq \log y) = P(\exp(X) \leq y) = \Phi(\log y)$, and

$$f_Y(y) = \frac{1}{y} f_X(\log y) = \frac{1}{y\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log y)^2\right), \quad 0 < y < \infty.$$

Solution 3.10 (a) $\text{Max}\{X, Y\} \leq x$ if and only if $X \leq x$ and $Y \leq x$. Then, by independence,

$$P(\text{max}\{X, Y\} \leq x) = P(X \leq x, Y \leq x) = P(X \leq x)P(Y \leq x) = F(x)^2.$$

Now, differentiating the expression for CDF we obtain the density function of $G = \text{max}\{X, Y\}$, $f_G(x) = 2F(x)f(x)$.

(a) Similarly, $H = \text{min}\{X, Y\} > x$ if and only if $X > x$ and $Y > x$. Therefore,

$$P(H \leq x) = 1 - P(\text{min}\{X, Y\} > x) = 1 - P(X > x)P(Y > x) = 1 - (1 - F(x))^2.$$

Again, by differentiating the CDF, we obtain $f_H(x) = 2f(x)(1 - F(x))$

Solution 3.11 Expand the expression $E((X - a)^2)$ as

$$\begin{aligned} E((X - a)^2) &= E([(X - EX) + (EX - a)]^2) \\ &= E((X - EX)^2) + 2(X - EX)(EX - a) + E((EX - a)^2) \\ &= V(X) + (EX - a)^2 \end{aligned}$$

It's clear that the minimum is achieved when the second term is zero, that is, $EX = a$.

Solution 3.12 If $X \sim \text{Uniform}(a, b)$, it means that $f_X(x) = 1/(b - a)$ for $a \leq x \leq b$, otherwise f_X is zero. Thus,

$$E(X) = \int_a^b 1/(b - a)x dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = (a + b)/2.$$

$$\text{Furthermore, } E(X^2) = \int_a^b 1/(b - a)x^2 dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = (b^2 + ab + a^2)/3.$$

$$\text{Therefore, } V(X) = E(X^2) - E(X)^2 = (b^2 + ab + a^2)/3 - (b + a)^2/4 = (b - a)^2/12.$$

Solution 3.13 We have that $P(X > a) = \int_a^\infty f_X(x) dx$. Given that $f_X(x) = 0, x < 0$, the expectation is, $EX = \int_0^\infty x f_X(x) dx$. Note the lower bounds of integration to

conclude that

$$EX \geq \int_a^\infty x f_X(x) dx \geq \int_a^\infty a f_X(x) dx = aP(X > a).$$

Therefore, $P(X > a) \leq EX/a$.

Solution 3.14 To find the probability $P(|X - \mu| \geq a)$ we find the probability mass under the tails of the density for values beyond $\pm a$. Thus,

$$P(|X - \mu| \geq a) = \int_{-\infty}^{\mu-a} f_X(x) dx + \int_{\mu+a}^{\infty} f_X(x) dx = \int_{\mathcal{A}} f_X(x) dx,$$

where $\mathcal{A} = \{x : |x - \mu| \geq a\}$. Furthermore,

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx \geq \int_{\mathcal{A}} (x - \mu)^2 f_X(x) dx \geq \int_{\mathcal{A}} a^2 f_X(x) dx = a^2 P(|X - \mu| \geq a)$$

Therefore, $P(|X - \mu| \geq a) \leq \sigma^2/a^2$. Note that the Chebyshev inequality is remarkable because it holds for any distribution with finite mean and variance. However, depending on the ratio σ^2/a^2 , the inequality may be very weak.

Solution 3.15 The memoryless property, $P(X \leq x + t | X > t)P(X \leq x)$, $x, t > 0$, is equivalent to $P(X \leq x + t, X > t)/P(X > t) = P(X \leq x)$ or

$P(t < X \leq x + t) = P(X > t)P(0 < X \leq x)$ since X is non-negative. If we express the previous equality using the CDF, $F_X(x)$ of X , and writing only F , just dropping the subscript, we get

$$F(x + t) - F(t) = [F(x) - F(0)][1 - F(t)]$$

Note that $F(0) = 0$, rearrange the equation and divide by t to get:

$$\frac{F(x + t) - F(t)}{t} = \frac{F(t)(1 - F(x))}{t}$$

Taking the limit of both sides as $t \rightarrow 0$ we obtain $F'(x) = F'(0)(1 - F(x))$

Here, $F'(x)$ denotes the derivative of F . Now, let $G(x) = 1 - F(x)$, Then, we have $G'(x) = -G'(0)G(x)$. Note that $G'(0) = -c_1$ is a constant, so we have $G'(x)/G(x) = -c_1$ and from here, $G(x) = c_2 \exp(-c_1 x)$, as the solution for this well known differential equation. Furthermore, $c_2 = 1$ and $-c_1$ we denote by λ to get $G(x) = \exp(-\lambda x)$ and finally:

$$F(x) = 1 - G(x) = 1 - \exp(-\lambda x),$$

which is the CDF of the exponential random variable with density $\lambda \exp(-\lambda x)$

Solution 3.16 For U and V to be uncorrelated it suffices that $E(UV) = (EU)(EV)$, or $E((X + Y)XY) = E(X + Y)E(XY)$.

Thus, $E(X^2Y) + E(XY^2) = EXE(XY) + EXE(XY)$. Then, by independence of X and Y ,

$$\begin{aligned} E(X^2)EY + (EX)E(Y^2) &= (EX)E(XY) + (EX)E(XY) \\ (EY)E(X^2) + (EX)E(Y^2) &= (EX)^2(EY) + (EY)^2(EX) \\ (EY)(E(X^2) - (EX)^2) &= (EX)(E(Y^2) - (EY)^2). \end{aligned}$$

In other words, U and V are uncorelated if $(EY)(VX) = (EX)(VY)$.

Solution 3.17 Note that $E(\bar{X}) = \mu$, so $E(X_r - \bar{X}) = 0$. Then,

$$\begin{aligned} E\{\bar{X}(X_r - \bar{X})\} &= \frac{1}{n}E\left(\sum_s X_r X_s\right) - E(\bar{X}^2) \\ &= \frac{1}{n}(\sigma^2 + n\mu^2) - (\text{var}(\bar{X}) + E(\bar{X})^2) \\ &= \frac{1}{n}(\sigma^2 + n\mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = 0 \end{aligned}$$

Solution 3.18 Let X be the length of the shorter piece of the first break, so $f_X(x) = 2$ for $x \in [0, \frac{1}{2}]$. The other two pieces have lengths $U(1 - X)$ and $(1 - U)(1 - X)$ where U is uniformly distributed on $[0, 1]$.

A triangle can be formed from these lengths if and only if

$$U(1 - X) \leq X + (1 - U)(1 - X), \quad (1 - U)(1 - X) \leq X + U(1 - X)$$

where $X < 1 - X$. Conditional on $X = x$, the above occurs if and only if $1 - 2x \leq 2U(1 - x) \leq 1$, which has probability

$$\mathbb{P}\left(\frac{1 - 2x}{2(1 - x)} \leq U \leq \frac{1}{2(1 - x)}\right) = \frac{1}{2(1 - x)} - \frac{1 - 2x}{2(1 - x)} = \frac{x}{1 - x}$$

The unconditional probability is

$$\int_0^{\frac{1}{2}} \frac{x}{1 - x} f_X(x) dx = \int_0^{\frac{1}{2}} \frac{2x}{1 - x} dx = 2 \log 2 - 1$$

Solution 3.19 Let X and Y be the two random points at which the stick is broken where X is the point that is closest to the left end point of the stick. Assume that the stick is of a unit length. Then, all three pieces are smaller than a half, only if $X \leq 0.5$, $Y - X \leq 0.5$ and $1 - Y \leq 0.5$.

In other words, (X, Y) should satisfy $0 \leq X \leq 0.5$ and $0.5 \leq Y \leq 0.5 + X$. Then the joint density function $f(x, y)$ of (X, Y) is given by $f(x, y) = 2$ for $0 < x < y < 1$ and 0 otherwise.

In order to convince yourself of this, note that

$$X = \min(U_1, U_2), \text{ and } Y = \max(U_1, U_2),$$

where U_1 and U_2 are independent and uniformly distributed on $(0, 1)$.

Furthermore, for any $0 < x < y < 1$ and $dx > 0, dy > 0$ sufficiently small,

$$\begin{aligned} & P(x \leq X \leq x + dx, y \leq Y \leq y + dy) \\ &= P(x \leq U_1 \leq x + dx, y \leq U_2 \leq y + dy) \\ &+ P(x \leq U_2 \leq x + dx, y \leq U_1 \leq y + dy). \end{aligned}$$

By the independence of U_1 and U_2 , this gives $P(x \leq X \leq x + dx, y \leq Y \leq y + dy) = 2dxdy$, showing that $f(x, y) = 2$ for $0 < x < y < 1$. It now follows that $P(\text{no piece is longer than half the length of the stick})$

$$= \int_0^{0.5} dx \int_{0.5}^{0.5+x} 2dy = 2 \int_0^{0.5} xdx = 0.25.$$

Solution 3.20 We have

$$f_{\theta}(\theta) = \begin{cases} 1/(2\pi), & 0 < \theta < 2\pi \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{2\pi} \cos \theta f_{\theta}(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \cos \theta d\theta = 0$$

Similarly for $E(Y)$,

$$E(Y) = \frac{1}{2\pi} \int_0^{2\pi} \sin \theta d\theta = 0$$

$$E(XY) = \frac{1}{2\pi} \int_0^{2\pi} \cos \theta \sin \theta d\theta = \frac{1}{4\pi} \int_0^{2\pi} \sin 2\theta d\theta = 0 = E(X)E(Y)$$

Therefore, X and Y are uncorrelated.

(b)

$$E(X^2) = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \theta d\theta = \frac{1}{4\pi} \int_0^{2\pi} (1 + \cos 2\theta) d\theta = \frac{1}{2}$$

$$E(Y^2) = \frac{1}{2\pi} \int_0^{2\pi} \sin^2 \theta d\theta = \frac{1}{4\pi} \int_0^{2\pi} (1 - \cos 2\theta) d\theta = \frac{1}{2}$$

$$E(X^2Y^2) = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \theta \sin^2 \theta d\theta = \frac{1}{16\pi} \int_0^{2\pi} (1 - \cos 4\theta) d\theta = \frac{1}{8}$$

Hence

$$E(X^2Y^2) = \frac{1}{8} \neq \frac{1}{4} = E(X^2)E(Y^2)$$

If X and Y are independent then $E(g(X)h(Y)) = E(g(X))E(h(Y))$. In our case this means that we would have $E(X^2Y^2) = E(X^2)E(Y^2)$, however, this is not true, therefore, X and Y are not independent.

Solution 3.21 In the derivation below we use the following statements:

$$f_{Y|X}(y|x) = f_{XY}(x, y)/f_X(x), \text{ where } f_X(x) > 0; \text{ and } E(Y|x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

Then, we have

$$\begin{aligned}
 E[E(Y | X)] &= \int_{-\infty}^{\alpha} E(Y | x) f_X(x) dx = \int_{\alpha}^{\infty} \left[\int_{\alpha}^w y f_{Y|X}(y | x) dy \right] f_X(x) dx \\
 &= \int_{-\infty}^{\infty} \int_{-x}^{\infty} y \frac{f_{XY}(x, y)}{f_X(x)} f_X(x) dx dy = \int_{-x}^{\infty} y \left[\int_{-\infty}^{\infty} f_{XY}(x, y) dx \right] dy \\
 &= \int_{-\alpha}^{\alpha} y f_Y(y) dy = E[Y],
 \end{aligned}$$

which proves the result.

Solution 3.22 Given that X and Y are independent, so are $g(X)$ and $h(Y)$, and we have

$$\begin{aligned}
 E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y) f_{XY}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y) f_X(x) f_Y(y) dx dy \\
 &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \int_{-\infty}^{\infty} h(y) f_Y(y) dy \\
 &= E[g(X)] E[h(Y)]
 \end{aligned}$$

Solution 3.23 We have $f(x, y) = cxy$, for $0 < x < y < 1$. Knowing that $\iint_A f(x, y) dx dy = 1$, where A is the domain where $f(x, y) > 0$, we compute:

$$\int_0^1 \int_x^1 cxy dx dy = \int_0^1 cx \left(\int_x^1 y dy \right) dx = \int_0^1 cx(1 - x^2)/2 dx = c/8 = 1.$$

Thus, $c = 8$.

Solution 3.24 We have,

$$f_X(t) = \begin{cases} kt^{-4} & (t > 1) \\ 0 & (t \leq 1), \end{cases}$$

where k is some positive constant.

(a) Total probability, or the area under the probability curve must be one, so to find k , we integrate the p.d.f.

$$1 = \int_{-\infty}^{\infty} f_X(t) dt = k \int_1^{\infty} t^{-4} dt = k/3$$

Thus, $k = 3$.

(b) The distribution function F_X is given by:

$$\begin{aligned} F_X(t) &= P(X \leq t) = \int_{-\infty}^t f_X(u) du \\ &= \begin{cases} 3 \int_1^t u^{-4} du, & t > 1 \\ 0, & t \leq 1 \end{cases} \\ &= \begin{cases} 1 - (1/t)^3, & t > 1 \\ 0, & t \leq 1 \end{cases} \end{aligned}$$

(c) First find the mean μ_X

$$\mu_X = \int_{-\infty}^{\infty} t f_X(t) dt = 3 \int_1^{\infty} t^{-3} dt = 3/2$$

and $E(X^2)$

$$E(X^2) = \int_{-\infty}^{\infty} t^2 f_X(t) dt = 3 \int_1^{\infty} t^{-2} dt = 3$$

$$\text{Var}(X) = \sigma_X^2 = E(X^2) - E(X)^2 = 3/4$$

(d) Now, we compute the desired probability noting that $\sigma = \sqrt{3}/2$. Thus,

$$\begin{aligned} P(|X - \mu_X| \leq 1.5\sigma_X) &= P(|X - 3/2| \leq 3\sqrt{3}/4) \\ &= P\left(\frac{6 - 3\sqrt{3}}{4} < X < \frac{6 + 3\sqrt{3}}{4}\right) \\ &= P\left(1 < X < \frac{6 + 3\sqrt{3}}{4}\right) = 3 \int_1^{\frac{6+3\sqrt{3}}{4}} t^{-4} dt = 0.954 \end{aligned}$$

Note the lower integration limit of 1 according to the domain of X , which was necessary to use since $(6 - 3\sqrt{3})/4 = 0.954 < 1$.

(e) According to Chebyshev's inequality

$$P(|X - \mu_X| \leq 1.5\sigma_X) \geq 1 - \left(\frac{1}{1.5}\right)^2 \approx 0.556$$

Although consistent with the correct value found in (d), the estimate computed

based on Chebyshev's inequality may appear as very inaccurate. Yet, Chebyshev provided a way to estimate the probability based on so little available information, which is nothing short of remarkable.

Solution 3.25

$$\begin{aligned}
 & \int_{-\infty}^{\infty} E[X | Y = y] f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx \right] f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x | y) f_Y(y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right] dx \\
 &= \int_{-\infty}^{\infty} x f_X(x) dx = E[X].
 \end{aligned}$$

Solution 3.26 We can model the parameter λ in terms of a uniform random variable Z with PDF

$$f_{\Lambda}(\lambda) = 2, \quad \text{for } 1 \leq \lambda \leq 1.5.$$

The conditional distribution, PDF $f_{\Lambda|Y}(\lambda | y)$, captures the available information about Λ .

Note the continuous form of Bayes' rule:

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}.$$

Then, we can express the conditional PDF for $\Lambda|Y$ as:

$$f_{\Lambda|Y}(\lambda | y) = \frac{f_{\Lambda}(\lambda) f_{Y|\Lambda}(y | \lambda)}{\int_{-\infty}^{\infty} f_{\Lambda}(t) f_{Y|\Lambda}(y | t) dt} = \frac{2\lambda e^{-\lambda y}}{\int_1^{3/2} 2te^{-ty} dt}$$

for $1 \leq \lambda \leq 3/2$. Calculation of the integral is left as an exercise.

Solution 3.27 We have $Y = S + N$. Note that conditional on $S = s$, the random variable Y has a normal distribution with mean s and unit variance.

Now, applying the Bayes rule in continuous form, we obtain

$$P(S = 1 | Y = y) = \frac{p_S(1)f_{Y|S}(y | 1)}{f_Y(y)} = \frac{\frac{p}{\sqrt{2\pi}}e^{-(y-1)^2/2}}{\frac{p}{\sqrt{2\pi}}e^{-(y-1)^2/2} + \frac{1-p}{\sqrt{2\pi}}e^{-(y+1)^2/2}}$$

After some simplifications, we get

$$P(S = 1 | Y = y) = \frac{pe^y}{pe^y + (1-p)e^{-y}}$$

It should be noted that the probability $P(S = 1 | Y = y)$ tends to zero as y decreases to $-\infty$ and approaches 1 as y increases to ∞ . Also, $P(S = 1 | Y = y)$, is monotonically increasing in between, which agrees with intuition.

Solution 3.28 (a) We check if the PDF integrates to 1:

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{\lambda}{2}e^{-\lambda|x|}dx = 2 \cdot \frac{1}{2} \int_0^{\infty} \lambda e^{-\lambda x}dx = 2 \cdot \frac{1}{2} = 1$$

where we have used the fact $\int_0^{\infty} \lambda e^{-\lambda x}dx = 1$, which is the normalization property of the exponential PDF. By symmetry of the PDF, we have $E[X] = 0$. Furthermore,

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \frac{\lambda}{2}e^{-\lambda|x|}dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x}dx = \frac{2}{\lambda^2}$$

where we have used the fact that the second moment of the exponential PDF is $2/\lambda^2$. Therefore,

$$\text{var}(X) = E[X^2] - (E[X])^2 = 2/\lambda^2$$

The expression, $\frac{\lambda}{2}e^{-\lambda|x|}$, represents the PDF of the Laplace distribution.

Solution 3.29 (a) We first calculate the CDF of X . For $x \in [0, r]$, we have

$$F_X(x) = P(X \leq x) = \frac{\pi x^2}{\pi r^2} = \left(\frac{x}{r}\right)^2$$

For $x < 0$, we have $F_X(x) = 0$, and for $x > r$, we have $F_X(x) = 1$. By differentiat-

ing, we obtain the PDF

$$f_X(x) = \begin{cases} \frac{2x}{r^2}, & \text{if } 0 \leq x \leq r \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$E[X] = \int_0^r \frac{2x^2}{r^2} dx = \frac{2r}{3}$$

Also

$$E[X^2] = \int_0^r \frac{2x^3}{r^2} dx = \frac{r^2}{2}$$

so

$$\text{var}(X) = E[X^2] - (E[X])^2 = \frac{r^2}{2} - \frac{4r^2}{9} = \frac{r^2}{18}$$

(b) The player gets a positive score in the range $[1/t, \infty)$ if and only if $X \leq t$, and otherwise he gets a score of 0. Thus, for $s < 0$, the CDF of S is $F_S(s) = 0$. For $0 \leq s < 1/t$, we have $F_S(s) = P(S \leq s) = P(\text{player's hit is outside the inner circle}) = 1 - P(X \leq t) = 1 - \frac{t^2}{r^2}$. For $1/t < s$, the CDF of S is given by

$$F_S(s) = P(S \leq s) = P(X \leq t)P(S \leq s | X \leq t) + P(X > t)P(S \leq s | X > t).$$

We have $P(X \leq t) = t^2/r^2$, $P(X > t) = 1 - t^2/r^2$ and since $S = 0$ when $X > t$ we have $P(S \leq s | X > t) = 1$. Furthermore,

$$\begin{aligned} P(S \leq s | X \leq t) &= P(1/X \leq s | X \leq t) \\ &= \frac{P(1/s \leq X \leq t)}{P(X \leq t)} = \frac{(\pi t^2 - \pi(1/s)^2)/(\pi r^2)}{\pi t^2/(\pi r^2)} = 1 - \frac{1}{s^2 t^2} \end{aligned}$$

Using the above results, we obtain

$$P(S \leq s) = \frac{t^2}{r^2} \left(1 - \frac{1}{s^2 t^2} \right) + 1 - \frac{t^2}{r^2} = 1 - \frac{1}{s^2 r^2}$$

Summarizing the results of the preceding calculations, the CDF of S is

$$F_S(s) = \begin{cases} 0, & \text{if } s < 0 \\ 1 - t^2/r^2, & \text{if } 0 \leq s < 1/t \\ 1 - 1/s^2r^2, & \text{if } 1/t \leq s \end{cases}$$

Solution 3.30 Integrating the bivariate PDF

$$\begin{aligned} \int_{-\infty}^{\pi} \int_{-\infty}^{\infty} f_{Xr}(x, y) dx dy &= k \int_0^2 \int_0^2 (x + y) dx dy \\ &= k \int_0^2 \left(\frac{x^2}{2} + xy \right) \Big|_{x=0}^{x=2} dy \\ &= k \int_0^2 (2 + 2y) dy = 8k = 1 \end{aligned}$$

Thus $k = \frac{1}{8}$. (b) By Eq. (3.30), the marginal pdf of X is

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \frac{1}{8} \int_0^2 (x + y) dy \\ &= \frac{1}{8} \left(xy + \frac{y^2}{2} \right) \Big|_{y=0}^{y=2} = \begin{cases} \frac{1}{4}(x + 1) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

since $f_{Xr}(x, y)$ is symmetric with respect to x and y , the marginal pdf of Y is

$$f_r(y) = \begin{cases} \frac{1}{4}(y + 1) & 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

(c) since $f_{Xr}(x, y) \neq f_X(x)f_r(y)$, X and Y are not independent.

Solution 3.31 (a) We have

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Xr}(x, y) dx dy &= \int_0^{\infty} \int_0^{\infty} \frac{1}{y} e^{-x/y} e^{-y} dx dy \\ &= \int_0^{\infty} \frac{1}{y} e^{-y} \left(\int_0^{\infty} e^{-x/y} dx \right) dy \\ &= \int_0^{\infty} \frac{1}{y} e^{-y} \left[-ye^{-x/y} \Big|_{x=0}^{x=\infty} \right] dy \\ &= \int_0^{\infty} e^{-y} dy = 1 \end{aligned}$$

(b) We find the marginal pdf on Y .

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \frac{1}{y} e^{-y} \int_0^{\infty} e^{-x/y} dx \\ &= \frac{1}{y} e^{-y} \left[-y e^{-x/y} \Big|_{x=0}^{x=\infty} \right] = e^{-y} \end{aligned}$$

Then, the conditional pdf of X is

$$f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \begin{cases} \frac{1}{y} e^{-x/y} & x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} P(X > 1 | Y = y) &= \int_1^{\infty} f_{X|Y}(x, y) dx = \int_1^{\infty} \frac{1}{y} e^{-x/y} dx \\ &= -e^{-x/y} \Big|_{x=1}^{x=\infty} = e^{-1/y} \end{aligned}$$

Solution 3.32 Let X be a random variable with the Poisson distribution having the PMF:

$$P(X = x) = \exp(-\lambda) \lambda^x / x!$$

Note that $E(X) = V(X) = \mu$, that is the expectation and variance of the Poisson distribution are the same and equal to the parameter of the distribution.

Next, assuming that x is large, we approximate the histogram of X with a Gaussian distribution with density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

by setting x to μ and also equating σ^2 with μ .

The expression for f_X reduces to:

$$f_X(x)|_{x=\mu} = \frac{1}{\sqrt{2\pi}}.$$

This value is then equated with $P(X = x)|_{\lambda=x}$ to get:

$$\exp(-x) x^x / x! = \frac{1}{\sqrt{2\pi}},$$

and from here the Sterling formula follows directly:

$$x! \approx \sqrt{2\pi} \exp(-x)x^x.$$

Solution 3.33 (a) We start from

$$P(X < x) = P(\mu + \sigma Z < x) = P(Z < \frac{x - \mu}{\sigma})$$

Next, in terms of CDF, and PDF, this implies $F_X(x) = F_Z((x - \mu)/\sigma)$ and by differentiating, $f_X(x) = (1/\sigma)f_Z((x - \mu)/\sigma)$.

(b) Let $Y = X^5$. Then,

$$f_Y(y) = f_X(x) \left| \frac{dy}{dx} \right|^{-1} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y^{1/5} - \mu)^2}{2\sigma^2}\right) \frac{1}{5y^{4/5}}.$$

(c) We have that $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$. Then,

$$\begin{aligned} P(a < X < b) &= P(X < b) - P(X < a) \\ &= P((x - \mu)/\sigma < (b - \mu)/\sigma) - P((x - \mu)/\sigma < (a - \mu)/\sigma) \\ &= P(Z < (b - \mu)/\sigma) - P(Z < (a - \mu)/\sigma) \\ &= \Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma). \end{aligned}$$

(d) Here, $Y = aX + b$, so $EY = aE(X) + b = a\mu + b$.

Furthermore, $V(Y) = V(aX + b) = V(aX) = a^2V(X) = a^2\sigma^2$.

(e) Moment generating function is given as $G_Z(t) = e^{t^2}$.

Then, $G'_Z(0) = 2te^{t^2}|_{t=0} = 0$. Thus, $EZ = 0$.

Moreover, $G''_Z(0) = (4t^2e^{t^2} + 2te^{t^2})|_{t=0} = 2 = E(Z^2)$.

Hence, $V(Z) = E(Z^2) - (EZ)^2 = 2$.

Solution 3.34 (a) The probability distribution function of Y is given by

$$\begin{aligned} P(Y \leq y) &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{y}} e^{-\frac{1}{2}x^2} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{y}} e^{-\frac{1}{2}x^2} dx \end{aligned}$$

Now we differentiate $P(Y \leq y)$ to get the probability density function of Y :

$$\begin{aligned} g(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}} e^{-\frac{1}{2}y} \quad \text{for } y \geq 0 \end{aligned}$$

Note that $g(y) = 0$ otherwise. This shows that Y has a gamma density with shape parameter $\frac{1}{2}$ and scale parameter $\frac{1}{2}$.

(b) Noting that $P(Y \leq y) = P(|X| \leq y^2)$ for $y \geq 0$, it follows that

$$P(Y \leq y) = P(-y^2 \leq X \leq y^2) = \Phi(y^2) - \Phi(-y^2) \quad \text{for } y \geq 0$$

Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

Differentiation of the distribution function $P(Y \leq y)$ yields the density function of Y :

$$2y\phi(y^2) + 2y\phi(-y^2) = \frac{4y}{\sqrt{2\pi}} e^{-\frac{1}{2}y^4} \quad \text{for } y > 0.$$

Solution 3.35 The probability distribution function of Y is given by

$$\begin{aligned} P(Y \leq y) &= P(2X - 1 \leq y) = P\left(X \leq \frac{1}{2}(y + 1)\right) \\ &= \int_0^{\frac{1}{2}(y+1)} \frac{8}{\pi} \sqrt{x(1-x)} dx, \quad -1 \leq y \leq 1 \end{aligned}$$

The derivative of $P(Y \leq y)$ is

$$(8/\pi) \sqrt{\frac{1}{2}(y+1) \left(1 - \frac{1}{2}(y+1)\right)} \times \frac{1}{2} = \frac{2}{\pi} \sqrt{1-y^2}, \quad \text{for } -1 < y < 1.$$

Hence the probability density function of Y is given by $g(y) = \frac{2}{\pi} \sqrt{1-y^2}$, for $-1 < y < 1$, and $g(y) = 0$ otherwise.

Solution 3.36 We use moment generating functions for the proof. Now, let

$$X_n \sim \text{Poisson}(n), \quad \text{for } n = 1, 2, \dots$$

Note here, that instead of the customary λ , we use n to indicate only the integer values of the parameter of the distribution. That's because we are interested in large values of λ , where large integers are precise enough to describe the mean and also to tie the increasing parameter with the limiting sequence of random variables we use below.

Then, the PMF of X_n is

$$f_{X_n}(x) = \frac{n^x e^{-n}}{x!} \quad x = 0, 1, 2, \dots$$

The moment generating function of X_n is

$$M_{X_n}(t) = E[e^{tX_n}] = e^{n(e^t-1)}, \text{ for } -\infty < t < \infty.$$

Note that taking the limits of X directly is not helpful since

$$\lim_{n \rightarrow \infty} M_{X_n}(t) = \lim_{n \rightarrow \infty} e^{n(e^t-1)} = \infty$$

Instead, we work with a standardized version of the Poisson random variable $\frac{X_n - n}{\sqrt{n}}$ and derive its limiting moment generating function

$$\begin{aligned} \lim_{n \rightarrow \infty} M_{(X_n - n)/\sqrt{n}}(t) &= \lim_{n \rightarrow \infty} E \left[\exp \left(t \cdot \frac{X_n - n}{\sqrt{n}} \right) \right] \\ &= \lim_{n \rightarrow \infty} \exp(-t\sqrt{n}) E \left[\exp \left(\frac{tX_n}{\sqrt{n}} \right) \right] \\ &= \lim_{n \rightarrow \infty} \exp(-t\sqrt{n}) \exp \left(n \left(e^{t/\sqrt{n}} - 1 \right) \right) \\ &= \lim_{n \rightarrow \infty} \exp \left(-t\sqrt{n} + n \left(tn^{-1/2} + t^2 n^{-1}/2 + t^3 n^{-3/2}/6 + \dots \right) \right) \\ &= \lim_{n \rightarrow \infty} \exp \left(t^2/2 + t^3 n^{-1/2}/6 + \dots \right) \\ &= \exp(t^2/2) \quad -\infty < t < \infty. \end{aligned}$$

Now, the result is the moment generating function of a standard normal random variable. However, this result also means that the original X_n has a limiting distribution that is normal with mean n and variance n .

Solution 3.37 Let the X denote the original lifetime of the bulb. To calculate the

desired probability

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)}$$

we need the probability distribution of X . We do not know which manufacturer the bulb is from, but can condition X on the events that it came from either. Let A_1, A_2 be the events that the bulb was manufactured by A or B , respectively. Then, applying the law of conditional probability,

$$\begin{aligned} P(X > x) &= P(X > x \mid A_1) P(A_1) + P(X > x \mid A_2) P(A_2) \\ &= p_1 e^{-\mu_1 x} + p_2 e^{-\mu_2 x}, \quad \text{for } x \geq 0. \end{aligned}$$

This gives

$$P(X > s + t \mid X > s) = \frac{p_1 e^{-\mu_1(s+t)} + p_2 e^{-\mu_2(s+t)}}{p_1 e^{-\mu_1 s} + p_2 e^{-\mu_2 s}}$$

Solution 3.38 The marginal density of X is given by

$$f_X(x) = 6 \int_0^x (x - y) dy = 3x^2 \quad \text{for } 0 < x < 1,$$

and $f_X(x) = 0$ otherwise. Likewise, the marginal density of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = 6 \int_y^1 (x - y) dx = 3y^2 - 6y + 3, \quad \text{for } 0 < y < 1,$$

and $f_Y(y) = 0$ otherwise. An easy calculation finds that

$$E(X) = 3/4, \quad \sigma(X) = \sqrt{3}/4, \quad E(Y) = 1/4, \quad \sigma(Y) = \sqrt{3/80}$$

and

$$E(XY) = 6 \int_0^1 dx \int_0^x xy(x - y) dy = \int_0^1 x^4 dx = 1/5$$

Finally,

$$\rho(X, Y) = \frac{1/5 - (3/4) \times (1/4)}{(1/4)\sqrt{3} \times \sqrt{3/80}} = 0.149.$$

Solution 3.39 (a) The exact value of the probability, $P(S_n \leq a)$, we obtain from:

$$P(S_n \leq a) = \sum_{k=0}^a \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{21} \binom{36}{k} (1/2)^{36} = 0.8785$$

since $n = 36, p = 0.5$.

(b) Using the CLT:

$$P(S_n \leq a) = \Phi\left(\frac{a - E(S_n)}{\sigma_{S_n}}\right) = \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right) = \Phi(1) = 0.8413.$$

Here, $\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-z^2/2) dz$, with Z having the standard normal distribution.

(c) Using deMoivre-Laplace approximation, we have:

$$P(k \leq S_n \leq l) \approx \Phi\left(\frac{l + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

In our case,

$$P(S_n \leq 21) \approx \Phi\left(\frac{l + 1/2 - np}{\sqrt{np(1-p)}}\right) = \Phi(1.17) = 0.879,$$

since the second term, $\Phi(-\infty)$, is zero.

Note that deMoivre-Laplace approximation allows for the computation of probability for a single value of the approximated binomial variable. For example, $P(S_n = 19) \approx \Phi((19.5 - 18)/3) - \Phi((18.5 - 18)/3) = 0.124$, what is close to the exactly calculated probability $\binom{36}{19} (1/2)^{36} = 0.1251$.

Solution 3.40 We start by considering a well defined event in the sense that it has non-zero probability. Let $B = \{y \leq Y \leq y + \epsilon\}$, where $\epsilon > 0$. Then,

$$P(B) = P(\{y \leq Y \leq y + \epsilon\}) = \int_y^{y+\epsilon} f_Y(t) dt$$

Thus, $P(B)$ is non-zero as long as $f_Y > 0$. Now, we can write the Bayes rule in

the following form:

$$P(A \mid y \leq Y \leq y + \epsilon) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(y \leq Y \leq y + \epsilon \mid A)P(A)}{P(y \leq Y \leq y + \epsilon)}$$

Furthermore,

$$P(y \leq Y \leq y + \epsilon \mid A) = \int_y^{y+\epsilon} f_{Y|A}(t) dt$$

and

$$P(y \leq Y \leq y + \epsilon) = \int_y^{y+\epsilon} f_Y(t) dt$$

Now, taking the limit, as $\epsilon \rightarrow 0$

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} P(A \mid y \leq Y \leq y + \epsilon) &= P(A \mid Y = y) \\ \lim_{\epsilon \rightarrow 0} \int_y^{y+\epsilon} f_{Y|A}(t) dt &= f_{Y|A}(y)\epsilon \\ \lim_{\epsilon \rightarrow 0} \int_y^{y+\epsilon} f_Y(t) dt &= f_Y(y)\epsilon \end{aligned}$$

we get the following version of the Bayes rule:

$$P(A \mid Y = y) = \frac{f_{Y|A}(y)P(A)}{f_Y(y)}.$$

Solution 3.41 In order to find the distribution of $Y = g(X) = aX + b$ we can work with the result that $f_Y(y) = f_X(g^{-1}(x))|dx/dy|$. Thus, we have

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

and also $g(x) = ax + b$, $|dx/dy| = 1/|a|$. Therefore,

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\left(\frac{y-b}{a} - \mu\right)^2 / 2\sigma^2\right\} \\ &= \frac{1}{\sqrt{2\pi}|a|\sigma} \exp\left\{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right\} \end{aligned}$$

We recognize this as a normal PDF with mean $a\mu + b$ and variance $a^2\sigma^2$, that is,

Y is a normal random variable.

An alternative way to determine the distribution of Y is to note that the normal distribution of X is characterized by two parameters, $\mu = E(X)$ and $\sigma^2 = V(X)$.

Here, we have $E(Y) = E(aX + b) = a\mu + b$, and $V(Y) = V(aX + b) = a^2\sigma^2$. Thus, the distribution of Y has exactly the same parameters as obtained above, that is,

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

Solution 3.42 Bus arriving times have a joint density $f(x, y) = f_X(x)f_Y(y)$, which factorizes since X and Y are independent.

Note that $f_X(x) = 1/t_A$ for $x \leq t_A$, and $f_Y(y) = 1/t_B$ for $y \leq t_B$.

Let bus A arrive at time x , $0 \leq x \leq t_A$. Then, A will arrive first in case that bus B arrives at a time between x and t_B . This happens with probability $(t_B - x)/t_B = 1 - x/t_B$.

Let E be event that A arrives first. Then, $P(E|X = x) = 1 - x/t_B$. However,

$$P(E) = \int_0^{t_A} P(E|X = x)f_X(x)dx = \int_0^{t_A} (1 - x/t_B)(1/t_A)dx = 1 - t_A/(2t_B).$$

Solution 3.43 To check independence we find the marginal densities of f_X and f_Y noting that $x^2 + y^2 \leq 1$. Then,

$$f_X = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1/\pi dy = 2\sqrt{1-x^2}/\pi$$

. Likewise,

$$f_Y = 2\sqrt{1-y^2}/\pi.$$

It's obvious that $f(x, y) = 1/\pi \neq f_X(x)f_Y(y)$, that is, X and Y are not independent.

Solution 3.44 This problem may hint at convolutions for sums of random variables, however, the solution to it is quite simple. Note that all involved random vari-

ables have identical distribtuion (standard normal) and are also independent.

We want to find, $P(W + X > Y + Z + 1) = P(W + X - Y - Z > 1)$.

Then, $T = W + X - Y - Z$ is also a normal random variable (as a sum of normal rv-s) and thus has the mean, $ET = E(W + X - Y - Z) = 0$. Also, $V(T) = V(W + X - Y - Z) = 4$.

Now, normalize T such that $Q = (T - ET)/\sigma_T = T/2$.

Q has a normal distribution with mean zero and $V(Q) = (1/2)^2 V(T) = 1$.

In other words Q is a standard normal rv.

Hence, $P(T > 1) = P(2Q > 1) = P(Q > 1/2) = 1 - \Psi(1/2) = 0.3085$.

Solution 3.45 Let Y denote the area of the circle, where $Y = \pi X^2$, where X is uniformly distributted on $[0, 1]$, with density $f_X(x) = 1$, for $(0 < x < 1)$ and $f_X(x) = 0$ otherwise. The range for Y is $0 < y < \pi$, so that

$$F_Y(y) = P(Y \leq y) = P(X \leq \sqrt{y/\pi}) = \int_0^{\sqrt{y/\pi}} dx = \sqrt{y/\pi},$$

where X is uniformly distributed on $(0, 1)$. The density function of X is given by $f(x) = 1$ for $0 < x < 1$ and $f(x) = 0$ otherwise.

Now, differentiate the $F_Y(y)$ to get the density function, $f_Y(y)$ of the area of the circle:

$$f_Y(y) = \frac{1}{2\sqrt{\pi y}}, \text{ for } 0 \leq y \leq \pi, \text{ otherwise } 0.$$

Solution 3.46 (a) The probability distribution function of Y is given by

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{y}} e^{-\frac{1}{2}x^2} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{y}} e^{-\frac{1}{2}x^2} dx \end{aligned}$$

We then differente $F_Y(y)$ to get the probability density function of Y

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}} e^{-\frac{1}{2}y}, \quad \text{for } y \geq 0,$$

and $g(y) = 0$ otherwise. We recognize that Y has a gamma density with shape parameter $\frac{1}{2}$ and scale parameter $\frac{1}{2}$.

(b) Note first that, given $Y = \sqrt{|X|}$, we have that $P(Y \leq y) = P(|X| \leq y^2)$, for $y \geq 0$. Then it follows that

$$F_Y(y) = P(Y \leq y) = P(-y^2 \leq X \leq y^2) = \Phi(y^2) - \Phi(-y^2), \quad \text{for } y \geq 0.$$

Again, differentiate $F_Y(y)$ to get the density function of Y

$$f_Y(y) = 2y\phi(y^2) + 2y\phi(-y^2) = \frac{4y}{\sqrt{2\pi}}e^{-\frac{1}{2}y^4} \quad \text{for } y > 0$$

and is equal to zero otherwise. Note that $\phi(x) = d\Phi/dx = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$.

Solution 3.47 (a) We need to evaluate

$$E(X^2 + Y^2) = \iint_A f(x, y) dx dy = 1\pi \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dx dy$$

Note that the area over which to evaluate the integrals is that of a unit circle with the center in the origin of the coordinate system.

One of the ways in which we can proceed is to use the substitution $x = r \sin \theta$, $y = r \cos \theta$, implying $x^2 + y^2 = r^2 dr d\theta$. Therefore,

$$E(X^2 + Y^2) = 1/\pi \int_0^{2\pi} \int_0^1 r^2 dr d\theta = 2/3.$$

(b) Note that by symmetry, the answer is the same in each octant. Therefore

$$\frac{1}{8}E(X \wedge Y) = \iint_{0 < x < y} y f(x, y) dx dy = \frac{1}{\pi} \int_0^{\pi/4} \int_0^1 r^2 dr d\theta = \frac{1}{3\pi}(1 - 1/\sqrt{2})$$

Thus, $E(X \wedge Y) = \frac{8}{3\pi}(1 - 1/\sqrt{2})$

(c) Let $A = E(X^2/(X^2 + Y^2))$ and $B = E(Y^2/(X^2 + Y^2))$. Again, note the

symmetry here, as $A = B$ and also $A + B = 1$. Thus, $E(X^2/(X^2 + Y^2)) = E(Y^2/(X^2 + Y^2)) = 1/2$.

Solution 3.48 (a) We need to evaluate $EX, EY, E(XY)$. Again, the area over which to evaluate the integrals is that of a unit circle with the center in the origin. Intuitively, we should get the result $EX = EY = 0$ (why?). Let's compute EX

$$E(X) = \iint_A f(x, y) dx dy = 1/\pi \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} x dx dy = 2/\pi \int_{-1}^1 x \sqrt{1-x^2} dx.$$

Note that $x\sqrt{1-x^2}$ is an odd function, hence this integral is zero. Similarly, we obtain $EY = 0$ and $E(XY) = 0$.

(b) The result, $E(XY) = 0$, tells us that X and Y are uncorrelated. However, they are not independent. We can see that by comparing the joint distribution $f(x, y) = 1/\pi$ on $(0 < x, y < 1)$ with the marginals which we compute as follows.

$$f_X(x) = 1/\pi \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = 2/\pi \sqrt{1-x^2}$$

Likewise, $f_Y(y) = 2/\pi \sqrt{1-y^2}$. We see that $f_X(x)f_Y(y) \neq f(x, y) = 1/\pi$. Hence X and Y are not independent.

(c) Now, $EU = E(|X|) = E(V)$,

$$EU = E(|X|) = 4/\pi \int_0^1 \int_0^{\pi/2} r^2 \cos \theta dr d\theta = 4/(3\pi).$$

Also,

$$E(UV) = E(|XY|) = 4/\pi \int_0^1 \int_0^{\pi/2} r^3 \cos \theta \sin \theta dr d\theta = 1/(2\pi).$$

Thus, $\text{cov}(U, V) = 1/(2\pi(-(4/(3\pi))^2) < 0$, that is, U and V are negatively correlated.

Solution 3.49 (a) To compute $f_U(u)$,

$$f_U(u) = \int_u^\infty f(u, v) dv = e^{-u}$$

Thus, U is exponential rv with parameter 1. Then

$$f_{V|U} = \frac{f(u, v)}{f_U(u)} = e^{u-v}, \quad v > u$$

so the conditional density of V given $U = u$ is exponential on (u, ∞) , with parameter 1. Again

$$f_V(v) = \int_0^v f(u, v) du = v e^{-v}$$

This is a gamma density. Then

$$f_{U|V} = \frac{f(u, v)}{f_V(v)} = \frac{1}{v}, \quad 0 < u < v$$

so the conditional density of U given $V = v$ is uniform on $(0, v)$

Note that

$$P(X \leq x) = \int_{-\infty}^{\infty} P(X \leq x | Y = y) f_Y(y) dy.$$

Therefore,

$$P(V - U \leq y) = \int_0^\infty P(V - U \leq y | U = u) f_U(u) du$$

Now

$$\begin{aligned} P(V - U \leq y | U = u) &= \int_u^{y+u} f_{V|U}(v | u) dv \\ &= e^{+u} (e^{-u} - e^{-y-u}) = 1 - e^{-y} \end{aligned}$$

Finally,

$$\begin{aligned} P(V - U \leq y) &= \int_0^\infty (1 - e^{-y}) e^{-u} du \\ &= 1 - e^{-y} \end{aligned}$$

since this is the distribution function of an exponential random variable, it follows that $Y = V - U$ has density e^{-y} .

Since $(P(V - U \leq y))$ does not depend on u , it is also true that U and $V - U$ are independent.

Solution 3.50 Let $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$. Then,

$$\begin{aligned} P(X < Y) &= \int_0^\infty P(X < Y \mid Y = y) f_Y(y) dy \\ &= \int_0^\infty (1 - e^{-\lambda y}) \mu e^{-\mu y} dy \\ &= 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu}. \end{aligned}$$

Another way to get the same result would be to use a double integral, however, we know (by now) the CDF of an exponential random variable and are free to use it directly.

PART
VIII

**Solutions to Exercises
in Stochastic Processes**

Solutions to Exercises in Stochastic Processes

Solution 4.1 (a) We have,

$$\begin{aligned} E(T) &= \sum_n E(S_N | N = n) p_N(n) = \sum_n E(S_n) p_N(n) \\ &= \sum_n n E(X_1) p_N(n) = E(X_1) \sum_n n p_N(n) \\ &= E(X_1) E(N). \end{aligned}$$

(b) Let's find the moment generating function (MGF) of T :

$$\begin{aligned} M_T(t) &= E(\exp(tT)) = E[E(\exp(tT) | N)] = \\ &= E[E(e^{t(X_1 + \dots + X_N)})] = E[E(e^{tX_1}) \dots E(e^{tX_N})] \\ &= E[(e^{tX})^N] = E(M_X(t)^N) = G_N(M_X(t)). \end{aligned}$$

We use the result, $M_T(t) = G_N(M_X(t))$, to find ET :

$$ET = M'_T(0) = G'_N(M_X(0)) M'_X(0) = E(N) E(X),$$

since $G'_N(1) = E(N)$, $M_X(0) = 1$, and $M'_X(0) = E(X)$.

Note that for the sake of notational simplicity we used $E(X)$ instead of $E(X_1)$, as all $E(X_i)$ are equal since X_i have identical distributions.

Solution 4.2 (a)

$$\begin{aligned} M(t) &= E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \lambda^k / k! \\ &= e^{-\lambda} \sum_{k=0}^{\infty} (e^t \lambda)^k / k! = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}. \end{aligned}$$

$$(b) E(X) = M'(t)|_{t=0} = \lambda e^t e^{\lambda(e^t - 1)} = \lambda.$$

$$E(X^2) = M''(t)|_{t=0} = (\lambda e^t)^2 e^{\lambda(e^t - 1)} + \lambda e^t e^{\lambda(e^t - 1)} = \lambda^2 + \lambda.$$

Therefore, $V(X) = E(X^2) - (E(X))^2 = \lambda$.

Solution 4.3 (a)

$$M(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda-t)x} dx = \lambda/(\lambda - t).$$

Note that $M(t)$ is defined only for $t < \lambda$.

(b) Now, $M'(t) = \lambda/(\lambda - t)^2$ and $M''(t) = 2\lambda/(\lambda - t)^3$.

Hence, $E(X) = M'(0) = 1/\lambda$; $E(X^2) = M''(0) = 2/\lambda^2$, and $V(X) = 1/\lambda^2$.

Solution 4.4 (a) No. The second row sum is greater than one.

(b) Yes. Probabilities are non-negative, rows sum to one.

(c) No. Not a quadratic matrix.

(d) No. Negative entries.

Solution 4.5 (a) $(0.5)(0.6)^4 = 0.00648$; **(b)** $(0.3)^2 + (0.3)(0.5) + (0.4)(0.2) = 0.32$;

(c) Probability of cloudy Tuesday is 0.2.

$P(X_1 = C | X_0 = R, X_2 = S) = P(X_2 = S, X_1 = C | X_0 = R) / P(X_2 = S | X_0 = R)$.
Therefore, $\{X_2 = S\} \cap \{X_1 = C\} = \emptyset$ implying that the desired probability is zero. This does not contradict Markov property, because the future is informative here, the past is not.

(d) Solve the following system of equations, using also the fact that $w_1 + w_2 + w_3 = 1$.

$$0.6w_1 + 0.5w_2 + 0.4w_3 = w_1$$

$$0.2w_1 + 0.5w_2 + 0.3w_3 = w_2$$

$$0.2w_1 + 0w_2 + 0.3w_3 = w_3$$

The solution is $w = (7/13, 4/13, 2/13)$.

(e) Expectation of the number of days between two sunny days is $\mu_{RR} - 1 = 13/2 - 1 = 11/2$.

(f) Let p be the probability of staying cloudy (event L). Then, $P(L = l) = p^{l-1}(1-p)$. Thus, $E(L) = 1/p = 2$.

Solution 4.6 (a) The set of states of the chain is irreducible and regular.

(b) The stationary distribution $\pi = (\pi_1, \pi_2, \pi_3)$ satisfies $\pi P = \pi$. Thus,

$$0.6\pi_1 + 0.5\pi_2 + 0.4\pi_3 = \pi_1$$

$$0.2\pi_1 + 0.5\pi_2 + 0.3\pi_3 = \pi_2$$

$$0.2\pi_1 + 0\pi_2 + 0.3\pi_3 = \pi_3$$

Solving the system (taking into account that $\pi_1 + \pi_2 + \pi_3 = 1$) give the unique stationary distribution $\pi = (4/19, 6/19, 9/19)$.

(c) The number of steps between two returns to A is $E(\tau_A(1)) = 1/\pi_A$. Therefore, $E(\tau_A(1)) = 19/4$

(d) The number of visits to B between two consecutive visits to A is $E(\sum_k^{\tau_A(1)} 1\{x_k = B\}) = \pi_B/\pi_A = 6/4$.

(e) $\lim_{n \rightarrow \infty} 1/n \sum_{k=0}^{n-1} f(X_k) = \sum_{A,B,C} f_i \pi$. Here, $f(A) = 1, f(B) = 2, f(C) = 0$ so
 $M = \sum_{k=0}^{10000} f(X_k) \approx 10001/10000 \sum_{k=0}^{10000} \approx \sum_{k=0}^{10000} (4/19(1) + 6/19(2)) \approx \8422 .

Solution 4.7 (a) Let $Z_{n+1} = X_1 + X_2 + \dots + X_{Z_n}$.

(b) We want to find the probability generating function, $G_{n+1} = E(S^{Z_{n+1}})$. Thus,

$$\begin{aligned} G_{n+1}(s) &= E_{Z_{n+1}}(s^{Z_{n+1}}) = E_{Z_n}(E_{Z_{n+1}}(s^{Z_{n+1}}|Z_n)) \\ &= E_{Z_n}(E(s^{X_1+X_2+\dots+X_{Z_n}}|Z_n)) \\ &= E([E(s^X)]^{Z_n}) = G_n(G_X(s)). \end{aligned}$$

Note that $E(X_i)$ are the same for all i , hence we use X instead of any X_1 or any X_i to simplify the notation.

(c) We need to show that $E(X_{n+1}) = \mu E(Z_n)$, where $\mu = E(X_i), i = 1, 2, \dots$. We start, from the expression for $E(Z_{n+1}) = E(X_1 + \dots + X_{Z_n})$. Then by the law of double expectation, we have

$$\begin{aligned} E_{Z_{n+1}} &= E_{Z_n}(E(X_1 + \dots + X_{Z_n}|Z_n)) = E_{Z_n}(E(X_1) + \dots + E(X_{Z_n})) \\ &= E_{Z_n}(Z_n E(X_1)) = E(X_1)E(Z_n) = \mu E(Z_n). \end{aligned}$$

(d) Need to show that $E(Z_n) = \mu^n$, where $\mu = E(X_i), i = 1, 2, \dots$. Note that

$Z_0 = 1$, $Z_1 = X_1$, $Z_2 = X_1 + X + 2 + \dots + X_{Z_1}$ Then,

$$\begin{aligned} E(Z_1) &= E(X_1) = \mu; \\ E(Z_2) &= E(X_1 + \dots + EX_{Z_1}) \\ &= E(E(X_1 + \dots + EX_{Z_1} | Z_1)) \\ &= E(Z_1 E(X_1)) = \mu E(Z_1) = \mu^2 \end{aligned}$$

Similarly, $E(Z_3) = \mu^3$ and a simple induction argument proves that indeed, $E(Z_n) = \mu^n$.

Solution 4.8 We have the distribution, $P(X = k) = \left(\frac{2}{3}\right)^k \frac{1}{3}$, $k = 0, 1, 2, \dots$, and want to show that

$$G_n(s) = \frac{1 - 2^n - 2(1 - 2^{n-1})s}{1 - 2^{n+1} - 2(1 - 2^n)s}.$$

For $n = 0$, $G_0(s)$ is the PGF of the distribution of the size of the zeroth generation, which is a constant equal 1. Thus we have, $G_0(s) = \sum_{k \geq 0} P(X_0 = k) = s$, because the only non-zero term is $sP(X_0 = 1) = s$.

We also find the PGF of X , which is the distribution for the first generation:

$$G_X(s) = \sum_{k \geq 0} s^k (2/3)^k (1/3) = (1/3) \sum_{k \geq 0} (2s/3)^k = (1/3)/(1 - 2s/3) = 1/(3 - 2s).$$

Now, check that the expression for $G_n(s)$ satisfies the cases $n = 0$ and $n = 1$:

$$G_n(s) = \frac{1 - 2^n - 2(1 - 2^{n-1})s}{1 - 2^{n+1} - 2(1 - 2^n)s} \Big|_{n=0} = \frac{1 - 2^0 - 2(1 - 2^{0-1})s}{1 - 2^{0+1} - 2(1 - 2^0)s} = s.$$

$$G_n(s) = \frac{1 - 2^n - 2(1 - 2^{n-1})s}{1 - 2^{n+1} - 2(1 - 2^n)s} \Big|_{n=1} = \frac{1 - 2^1 - 2(1 - 2^{1-1})s}{1 - 2^{1+1} - 2(1 - 2^1)s} = 1/(3 - 2s).$$

The induction step: assuming that the expression $G_n(s)$ holds for n , we show

that it also holds for $n + 1$.

$$\begin{aligned} G_{n+1}(s) &= G_n(G(s)) = G_n(G_X(s)) = G_n(1/(3 - 2s)) \\ &= \frac{1 - 2^n - 2(1 - 2^{n-1})/(3 - 2s)}{1 - 2^{n+1} - 2(1 - 2^n)/(3 - 2s)} \\ &= \frac{1 - 2^{n+1} - 2(1 - 2^{n-1})s}{1 - 2^{n+1} - 2(1 - 2^n)s} \end{aligned}$$

Thus, the expression is correct also for $n + 1$ and thereby for any n .

(b) $E(Z_n) = \mu^n = (EX)^n = (G'_X(1))^n$ We know that

$$G'_X(1) = (1/(3 - 2s'))|_{s=1} = 2/(3 - 2s)^2|_{s=1} = 2. \text{ Thus, } E(Z_n) = 2^n.$$

(c) Probability of extinction by n -th generation is

$$P(Z_n = 0) = G_n(s)|_{s=0} = (1 - 2^n)/(1 - 2^{n+1}), n = 1, 2, \dots$$

(d) Probability of ultimate extinction is

$$P(\lim_n Z_n) = \lim_n (1 - 2^n)/(1 - 2^{n+1}) = 1/2.$$

Solution 4.9 Let the following matrix be Markov.

$$P = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{12} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Note the property of a Markov matrix: $\sum_{j=1}^m p_{ij} = 1$, for all $i = 1, 2, \dots, m$, which simply says that the Markov chain is guaranteed to move to some state starting in any state. Now,

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{12} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Using compact notation, we have $Pa = a$, where $a^T = [1 \ 1 \ \dots \ 1]$. Multiplying on

the left both sides of the above equation by P , we obtain

$$P^2 a = P a = a$$

which indicates that P^2 is also a Markov matrix. If we repeat the multiplication by P it's obvious that $P^n a = a$, in other words, P^n is also Markov.

Solution 4.10 This Markov chain has three states, D , M , and E , corresponding to the levels of exam difficulty.

$$P = \begin{matrix} & \begin{matrix} D & M & E \end{matrix} \\ \begin{matrix} D \\ M \\ E \end{matrix} & \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{matrix}$$

The chain has a single recurrent class, which is aperiodic. Based on the fact that in the steady state we have $w = wP$, the balance equations can be read directly from the matrix

$$\begin{aligned} w_1 &= 0w_1 + (1/4)w_2 + (1/4)w_3 \\ w_2 &= (1/2)w_1 + (1/2)w_2 + (1/4)w_3 \\ w_3 &= (1/2)w_1 + (1/4)w_2 + (1/2)w_3 \end{aligned}$$

Note that the system is underspecified in the sense that one equation follows from the other two, so we use the additional piece of information, which is that all steady state probabilities sum to 1, $w_1 + w_2 + w_3 = 1$.

The solution is $w = (1/5, 2/5, 2/5)$.

Solution 4.11 By the definition of the stationarity $w = wP$ where w is the stationary distribution. Then, for $w = (w_1, w_2)$ and given P we have $(w_1, w_2) = (w_2, w_1)$, that is $w_1 = w_2$, and from $w_1 + w_2 = 1$, it follows that $w_1 = w_2 = 1/2$.

(b) Find $\lim_{n \rightarrow \infty} P^n$. Multiplying P by itself we see that the result alternates between

$$P^n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \text{ for } n = 1, 3, 5, \dots$$

and

$$P^n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ for } n = 2, 4, 6, \dots$$

Therefore, the limiting distribution does not exist.

Solution 4.12 Assume that all states $1, 2, 3, \dots, n$ are transient. By definition of transience, after a finite number of steps the first state will not be visited any more. The same holds for the second state and then for the n -th state. Then it follows that after a finite amount of time no more state will be visited. This is a clear contradiction as the chain must be in some state. Hence a finite state Markov chain must have some recurrent states.

Solution 4.13 (a) We want to decompose the stochastic matrix P so that $P = E\Lambda E^{-1}$, where E is a matrix whose columns are (right) eigenvectors of P . Note that $P^n = E\Lambda^n E^{-1}$.

We first find the eigenvalues in the usual way. Note that we know that one eigenvalue is 1 because $P\mathbf{1} = P$ given that rows of P sum to one.

We find eigenvalues from the customary characteristic equation of P that follows from $(P - \lambda I)v = 0$, where v is an eigenvector. Then, the expression is zero if the corresponding determinant is zero.

$$\begin{aligned} |\lambda I - P| &= \begin{vmatrix} \lambda - (1-a) & -a \\ -b & \lambda - (1-b) \end{vmatrix} \\ &= (\lambda - 1)(\lambda - 1 + a + b) = 0 \end{aligned}$$

Thus, the eigenvalues of P are $\lambda_1 = 1$ and $\lambda_2 = 1 - a - b$.

The next step is to calculate eigenvectors, and when all the dust settles down (after doing straightforward yet somewhat tedious algebra) we get:

$$P^n = \frac{1}{a+b} \left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$$

(b) It's easy to see that for $0 < a < 1, 0 < b < 1$, then $0 < 1 - a < 1$ and $|1 - a - b| < 1$. Therefore, $\lim_{n \rightarrow \infty} (1 - a - b)^n = 0$

$$\lim_{n \rightarrow \infty} P^n = \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix}$$

Solution 4.14 (a) Recall that the stationary distribution, $\hat{p}$, satisfies the equation

$\hat{p}P = \hat{p}$. Thus, we have

$$\begin{bmatrix} p_1 & p_2 \end{bmatrix} \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \end{bmatrix}$$

which yields

$$\begin{aligned} \frac{3}{4}p_1 + \frac{1}{2}p_2 &= p_1 \\ \frac{1}{4}p_1 + \frac{1}{2}p_2 &= p_2 \end{aligned}$$

Each of these equations is equivalent to $p_1 = 2p_2$. since $p_1 + p_2 = 1$, we obtain

$$\hat{\mathbf{p}} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

(b) since the Markov chain is regular we know that $\lim_{n \rightarrow \infty} P^n = \hat{P}$. Note that the rows of $\hat{P}$ are identical and equal to the stationary distribution.

$$\lim_{n \rightarrow \infty} P^n = \lim_{n \rightarrow \infty} \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}^n = \begin{bmatrix} \hat{p} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

(c) Setting $a = \frac{1}{4}$ and $b = \frac{1}{2}$ in the expression for the n -th power of P , we get

$$P^n = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix} - \left(\frac{1}{4}\right)^n \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

since $\lim_{n \rightarrow \infty} \left(\frac{1}{4}\right)^n = 0$, we obtain

$$\lim_{n \rightarrow \infty} P^n = \lim_{n \rightarrow \infty} \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}^n = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Solution 4.15 For $\tau \leq t$

$$\begin{aligned} P[T_1 \leq \tau \mid X(t) = 1] &= \frac{P[T_1 \leq \tau, X(t) = 1]}{P[X(t) = 1]} \\ &= \frac{P[X(\tau) = 1, X(t) - X(\tau) = 0]}{P[X(t) = 1]} \\ &= \frac{P[X(\tau) = 1]P[X(t) - X(\tau) = 0]}{P[X(t) = 1]} \\ &= \frac{\lambda\tau e^{-\lambda\tau} e^{-\lambda(t-\tau)}}{\lambda t e^{-\lambda t}} = \tau/t \end{aligned}$$

which implies that T_1 is uniform over $(0, t)$.

Solution 4.16 Let $Z_1, Z_2, \dots$ be the r.v.'s representing the lengths of interarrival times in the Poisson process $X(t)$. First, notice that $\{Z_1 > t\}$ takes place if and only if no event of the Poisson process occur in the interval $(0, t)$, and thus, knowing that $p_0(t) = \exp(-\lambda t)$, or

$$\begin{aligned} P(Z_1 > t) &= P\{X(t) = 0\} = e^{-\lambda t} \\ F_{Z_1}(t) &= P(Z_1 \leq t) = 1 - e^{-\lambda t} \end{aligned}$$

Hence Z_1 is an exponential r.v. with parameter λ . Let $f_1(t)$ be the pdf of Z_1 . Then

we have

$$\begin{aligned}
 P(Z_2 > t) &= \int P(Z_2 > t \mid Z_1 = \tau) f_1(\tau) d\tau \\
 &= \int P[X(t + \tau) - X(\tau) = 0] f_1(\tau) d\tau \\
 &= e^{-\lambda t} \int f_1(\tau) d\tau = e^{-\lambda t}
 \end{aligned}$$

which indicates that Z_2 is also an exponential r.v. with parameter λ and is independent of Z_1 . Repeating the same argument, we conclude that $Z_1, Z_2, \dots$ are iid exponential r.v.'s with parameter λ .

Solution 4.17 By Pigeon hole principle which says that if n pigeons return to their $m(< n)$ home (through hole), then at least one home contains more than one pigeon.

Consider any path of states $i_0 = i, i_1, \dots, i_n = j$ such that $P_{i_k, i_{k+1}} > 0$. Call this a path from i to j . If j can be reached from i , then there must be a path from i to j . Let $i_0 \dots, i_n$ be such a path. If all of values $i_0, \dots, i_n$ are not distinct, then there must be a subpath from i to j having fewer elements (for instance, if $i, 1, 2, 4, 1, 3, j$ is a path, then so is $i, 1, 3, j$). Hence, if a path exists, there must be one with all distinct states.

Solution 4.18 Let Y denote the first passage time to the state j starting the state i at time 0.

$$\begin{aligned}
 P_{ij}^n &= P(X_n = j \mid X_0 = i) \\
 &= \sum_{k=0}^n P(X_n = j, Y = k \mid X_0 = i) \\
 &= \sum_{k=0}^n P(X_n = j \mid Y = k, X_0 = i) P(Y = k \mid X_0 = i) \\
 &= \sum_{k=0}^n P(X_n = j \mid X_k = j) P(Y = k \mid X_0 = i) = \sum_{k=0}^n P_{jj}^{n-k} f_{ij}^k
 \end{aligned}$$

Solution 4.19 (a) The reverse time chain has transition matrix

$$\overleftarrow{P}_{i,j} \equiv \frac{\pi_j P_{j,i}}{\pi_i}$$

To find it, we need to first find the stationary vector π . By symmetry (or by noting that the chain is doubly stochastic), $\pi_j = 1/n, j = 1, \dots, n$. Hence,

$$P_{ij}^* = \pi_j P_{ji} / \pi_i = P_{ji} = \begin{cases} p & \text{if } j = i - 1 \\ 1 - p & \text{if } j = i + 1 \end{cases}$$

Solution 4.20 (a)

$$\begin{aligned}
 \sum_{n=1}^{\infty} P_{ij}^n &= E[\text{number of visits to } j \mid X_0 = i] \\
 &= E[\text{number of visits to } j \mid \text{ever visit } j, X_0 = i] f_{ij} \\
 &= (1 + E[\text{number of visits to } j \mid X_0 = j]) f_{ij} \\
 &= \frac{f_{ij}}{1 - f_{jj}} < \infty
 \end{aligned}$$

since $(1 + \text{number of visits to } j \mid X_0 = j)$ is a geometric rv variable with mean $\frac{1}{1-f_{jj}}$

BIBLIOGRAPHY

Bibliography

Ash, R. (2008) *Basic Probability Theory*. Dover Publications Inc, New York.

Bertsekas D. and Tsitsiklis J. (2008) *Introduction to Probability*. Athena Scientific. Belmont, Massachusetts.

Coleman R. (1974) *Stochastic Processes*. George Allen & Unwin Ltd. London, UK.

Grimmet G. and Stirzaker D. (2001) *Probability and Random Processes*. Oxford University Press Inc. New York.

Hogg R. and Craig A. (1995) *Introduction to Mathematical Statistics*. Prentice Hall, Inc. New Jersey.

Hwei P. Hsu (1997) *Probability, Random Variables, & Random Processes*. McGraw Hill Inc. New York.

Ross, S. (1994) *A first Course in Probability*. Prentice Hall Inc. Englewood Cliffs. New Jersey.

Ross, S. (1996) *Stochastic Processes*. John Wiley & Sons Inc. New York

Stirzaker D. (2005) *Probability and Random Variables*. Cambridge University Press. Cambridge, UK.

**CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд**

519.21(075.8)(076)(0.034.2)

KRNJAJIĆ, Milovan, 1957-

Introduction to probability and stochastic processes [Elektronski izvor] : exercises and solutions / Milovan Krnjajić. - 1st ed. - Belgrade : Singidunum University, 2020 (Belgrade : Caligraph). - 1 elektronski optički disk (CD-ROM) ; 12 cm

Sistemske zahteve: Nisu navedeni. - Naslov sa nasl. strane dokumenta. - Tiraž 10. - Bibliografija.

ISBN 978-86-7912-739-6

a) Стохастички процеси -- Задаци б) Теорија вероватноће -- Задаци

COBISS.SR-ID 28842249

© 2020

All rights reserved. No part of this work covered by the copyright herein may be reproduced, transmitted, stored or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, without the prior written permission of the publisher.



Milovan Krnjajić

INTRODUCTION TO PROBABILITY AND STOCHASTIC PROCESSES

Exercises and Solutions

This solution manual is based on the set of problems the author collected during many years of his professional work on statistical modelling and analysis, both in industry and academia. Some exercises in particular come from the material he used while teaching the undergraduate courses in Probability Theory and Stochastic Processes at the National University of Ireland, Galway.

Most of the problems found in this monograph are standard in the sense that they (or their versions) can be found in a myriad of textbooks and course syllabi with similar exercises and solutions. Although the text doesn't claim to formulate new problems, the manual provides a collection of highly detailed solutions and answers. This serves as a nod to the students who are struggling to cope with the subject matter (and who doesn't at times?) who are eager to understand all the steps of reasoning and derivations employed in the solutions. In addition, the text offers a balanced mix of problems: some of which are solved using basic logic and computation, while others are theoretical, and require rigorous proofs.

