Student’s Solution Manual to Accompany

Probability: The Science of Uncertainty with Applications to Investments, Insurance, and Engineering

Michael A. Bean
## Contents

Introduction...1  
Chapter One Solutions...3  
Chapter Two Solutions...6  
Chapter Three Solutions...15  
Chapter Four Solutions...23  
Section 4.1.13 Exercises...23  
Section 4.2.4 Exercises...38  
Section 4.3.3 Exercises...38  
Chapter Five Solutions...44  
Chapter Six Solutions...59  
Chapter Seven Solutions...77  
Chapter Eight Solutions...101  
Chapter Nine Solutions...129  
Chapter Ten Solutions...162
Introduction

This manual contains complete solutions to approximately one quarter of the exercises in the book *Probability: The Science of Uncertainty with Applications to Investments, Insurance, and Engineering*. It is an ideal companion to the textbook and is recommended for students who are preparing to take an examination in probability set by a professional society such as the Society of Actuaries, the Casualty Actuarial Society, or the Canadian Institute of Actuaries. This manual will also be a valuable resource for students interested in seeing worked-out solutions to problems that go beyond the examples given in the textbook.

■ How To Use This Manual

Mathematics is a subject that can only be learned through practice. Hence, before consulting the solutions in this manual students should have made a serious attempt to do the textbook exercises on their own. Although there are many ways in which this manual can be used as a supplement to the textbook, we believe that students will learn the most by using the manual in the following way:

1. Read the assigned chapter or section of the textbook thoroughly before attempting any of the exercises.

2. Re-read each of the examples in the assigned reading paying close attention to the key points and steps required to obtain the final answer.

3. Without consulting the solutions that accompany the examples, recreate solutions for each of the examples in the assigned reading and verify that the answers obtained agree with those given in the textbook.

4. Without consulting the solutions in this manual, attempt to solve each of the exercises accompanying the assigned reading. If necessary, read the sections and examples of the textbook that pertain to the given exercises once again. Do not immediately consult the solution manual when confronted with a problem that is difficult to solve.
5. If a particular problem still seems intractable, then read the first few sentences of the solution from this manual and try to solve the rest of the problem on your own.

6. After completing these steps, carefully read over the sections of this manual that pertain to the assigned exercises, making note of all key solution points and any alternative approaches that you may not have considered.

We trust that you will find this manual helpful in your study of probability. Comments on the contents or structure of this manual can be sent to the publisher at the address printed in the front of the textbook.
Chapter One Solutions

2. a. The terms Bayesian and frequentist refer to interpretations of probability. The frequentist (also called objectivist) interpretation of probability is a perspective in which probabilities are considered to be long run relative frequencies. The Bayesian (also called subjectivist) interpretation of probability is a perspective in which probabilities are considered to be measures of belief that can change over time to reflect new information. See sections 1.6 and 1.9 in the textbook.

b. The insurance principle is the basis of actuarial science, whereas the principle of no arbitrage is the basis of financial engineering. The insurance principle asserts that for any group of homogeneous and independent risks, the average loss per individual becomes more certain as the size of the group increases. The principle of no arbitrage asserts that any two securities that provide the same future cash flow and have the same level of risk must sell for the same price. The insurance principle can be used to determine the pure cost of insurance for a large group of independent, homogeneous risks. The principle of no arbitrage can be used to determine the (theoretically correct) price of a security relative to the prices of other securities in an active market.

c. Both moral hazard and adverse selection arise in insurance from an insurer's inability to access perfect information about the insured person. Adverse selection arises from an inability to distinguish completely the good risks from the bad. Moral hazard arises from the behavioral changes that insurance protection induces after it is purchased. Adverse selection can be minimized through the use of a good risk classification scheme (and to a lesser extent, policy design). Moral hazard is primarily mitigated through policy design (by including, for example, deductibles and coinsurance provisions which require policyholders to share in the losses).

d. Actuarial science is the subject that is concerned with analyzing the adverse financial consequences of large, unpredictable losses and with designing mechanisms to cushion
the harmful financial effects of such losses. Financial engineering is the subject that is concerned with analyzing risk in financial markets and with designing products and techniques to manage that risk. Actuarial science is based on applications of the insurance principle, whereas financial engineering is based on applications of the principle of no arbitrage and the principle of optimality. Historically, actuarial science developed to address contingencies in a company's liabilities, whereas financial engineering developed to address contingencies in the company's assets.

10. This question and the one following it are designed to give students an appreciation of the differences between the frequentist and Bayesian interpretations of probability. They are also designed to illustrate some of the strengths and weaknesses of each interpretation.

a. Recall that a frequentist considers a probability to be a constant long-run relative frequency, whereas a Bayesian considers a probability to be a measure of belief that can change over time to reflect new information. Looking at the data in the problem, a frequentist would note that the average number of accidents per year over the five year period is 72 \( (90+70+75+60+65)/5 \) and would estimate the probability of an accident in the coming year to be 7.2\%. A Bayesian, on the other hand, might notice the decline in accident frequency over time and choose to alter his/her opinion of the accident frequency to take this new information into account. As a result, a Bayesian might estimate the probability of an accident in the coming year to be around 6\%. A frequentist might also notice the apparent decline in accident frequency over time but, in the absence of further data, would interpret the relatively high first year accident frequency of 90 to be simply an "above average" observation that could have occurred in any year. The frequentist would interpret the first year observation in this way because the frequentist considers the accident probability to be an inherent constant that does not change with the arrival of new data.

b. One would expect the accident frequency to decrease as drivers gain experience. Hence, it is not realistic to assume a constant accident frequency over time. A Bayesian can easily incorporate this anticipated decrease into future estimates for the accident frequency because a Bayesian considers probability to be a measure of belief that can change over time to reflect new information and personal opinion. A frequentist could explain the change in accident frequency by arguing that the observed data do not come from the same experiment and hence should not be combined to determine an estimate for the future accident frequency. Indeed, a frequentist could argue that the data come from five distinct experiments, where each experiment is defined by the number of
years of driving experience, and that the only way to obtain meaningful estimates of the accident frequency is to consider several groups of 1000 newly licensed 18-year old drivers over a 5 year period. The estimates of accident frequency determined in this way will differ with the number of years of driving experience, as intuition suggests they should.
Chapter Two Solutions

6. a. The given distribution function is a step function with jumps at \( x = -1 \) and \( x = \frac{2}{3} \).

Hence, the only values of \( x \) for which \( p_X[x] > 0 \) are \( x = -1 \) and \( x = \frac{2}{3} \). To see why this is so, consider, for example, the point \( x = 0 \). From the definition of \( F_X \), \( F_X[0] = \frac{1}{3} \) and \( F_X[-1] = \frac{1}{3} \). However,

\[
F_X[0] = \Pr[X \leq 0] = \Pr[X \leq -1] + \Pr[-1 < X < 0] + \Pr[X = 0] = F_X[-1] + \Pr[-1 < X < 0] + p_X[0].
\]

Hence, \( \Pr[-1 < X < 0] + p_X[0] = 0 \). So, since probabilities cannot be negative, we have \( p_X[0] = 0 \) as claimed. One can show in a similar way that \( p_X[x] = 0 \) for \( x \neq -1, \frac{2}{3} \).

Now

\[
p_X \left[ \frac{2}{3} \right] = \Pr \left[ X = \frac{2}{3} \right] = \Pr \left[ X \leq \frac{2}{3} \right] - \Pr \left[ X < \frac{2}{3} \right] = F_X \left[ \frac{2}{3} \right] - F_X \left[ \frac{2}{3} \right].
\]

Further, since \( p_X[x] = 0 \) for \( -1 < x < \frac{2}{3} \), as demonstrated earlier, we have

\[
\Pr \left[ X < \frac{2}{3} \right] = \Pr[X \leq -1] + \Pr \left[ 1 < X < \frac{2}{3} \right] = \Pr[X \leq -1] = F_X[-1].
\]

Hence,

\[
p_X \left[ \frac{2}{3} \right] = F_X \left[ \frac{2}{3} \right] - F_X[-1] = 1 - \frac{1}{3} = \frac{2}{3}.
\]
In particular, the value of \( p_X\left(\frac{2}{3}\right) \) is the size of the jump on the graph of \( F_X \) at the point \( x = \frac{2}{3} \). In a similar manner, we have \( p_X[-1] = \frac{1}{3} \).

Consequently, the probability mass function of \( X \) is given by

\[
p_X[-1] = \frac{1}{3}, \quad p_X\left(\frac{2}{3}\right) = \frac{2}{3}, \quad \text{and} \quad p_X[x] = 0 \quad \text{for all other} \quad x.
\]

b. From the probability mass function determined in part a, the expected value is given by

\[
E[X] = (-1)\left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{1}{9}.
\]

c. Using the observations given in part a, we have

\[
\Pr\left[ X < \frac{2}{3} \right] = \Pr[X \leq -1] + \Pr\left[ 1 < X < \frac{2}{3} \right] = F_X[-1] + 0 = \frac{1}{3}.
\]

d. The graphs of \( p_X \) and \( F_X \) can be created using Mathematica or similar computer software. The graphs given here have the same scale to facilitate comparisons.
12. Let \( X_1 \) and \( X_2 \) be as defined in the question.

a. If only the suit of a card is observed, then the sample space is 
\[ S = \{\text{HH, HD, DH, DD}\}, \]
where H represents hearts, D represents diamonds, and where order is respected. If the two hearts are distinguishable from each other and the two diamonds are distinguishable from each other, then the sample space is 
\[ S = \{H_1 H_2, \ H_2 H_1, \ H_1 D_1, \ H_1 D_2, \ H_2 D_1, \ H_2 D_2, \ D_1 H_1, \ D_1 H_2, \ D_2 H_1, \ D_2 H_2, \ D_1 D_2, \ D_2 D_1\} \]
where the subscripts identify the given heart or diamond uniquely and order is respected. Since we are only interested in the suit of a card in this question, we can take the sample space to be \( S = \{\text{HH, HD, DH, DD}\} \). Note, however, that the principle of equal likelihood does not apply to \( S \) because we have suppressed information on the individual outcomes. See section 3.1 for a discussion of how the choice of sample space can affect the validity of the principle of equal likelihood.

b. The random variables \( X_1 \) and \( X_2 \) are not independent because knowledge of the value assumed by one of the variables changes the probability distribution of the other. For example, if it is known that \( X_1 \) assumes the value 1 (i.e., the card on the left is a heart),
then it is less likely that \( X_2 = 1 \) because only one of the three remaining cards is a heart, the other two cards being diamonds. However, the random variables \( X_1 \) and \( X_2 \) are identically distributed because in the absence of knowledge of the left card, the probability that the right card is a heart is \( \frac{1}{2} \) and in the absence of knowledge of the right card, the probability that the left card is a heart is \( \frac{1}{2} \).

c. Using the relative frequency interpretation of probability, the probability mass function for the random vector \((X_1, X_2)\) is given by

\[
\begin{align*}
p_{X_1, X_2}[1, 1] &= \frac{1}{6}, & p_{X_1, X_2}[1, 0] &= \frac{1}{3}, & p_{X_1, X_2}[0, 1] &= \frac{1}{3}, & p_{X_1, X_2}[0, 0] &= \frac{1}{6}.
\end{align*}
\]

Consider, for example, the point \((x_1, x_2) = (1, 1)\). Suppose that the given experiment is repeated \(n\) times, where \(n\) is a large number. Then according to the relative frequency interpretation of probability, approximately \(\frac{n}{2}\) of the ordered pairs \((x_1, x_2)\) are of the form \((1, \ast)\). Suppose that exactly \(n^*\) are of this form. Then according to the relative frequency interpretation of probability, approximately \(\frac{1}{3}\) of these \(n^*\) ordered pairs are of the form \((1, 1)\). Consequently, of the original \(n\) observations, approximately \(\frac{1}{3} n^* \approx \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)n = \frac{n}{6}\) are of the form \((1, 1)\). Therefore, by the relative frequency interpretation of probability again, \(p_{X_1, X_2}[1, 1] = \frac{1}{6}\) as claimed. The values of \(p_{X_1, X_2}\) at the points \((1, 0)\), \((0, 1)\), and \((0, 0)\) are determined using a similar argument. See also the discussion in section 2.2 of the textbook.

d. The contingency table for \(X_1\) and \(X_2\) is as follows:
15. a. From the general relationship

\[ F_X(x) = \int_{-\infty}^{x} f_X(s) \, ds \]

(see section 2.3 in the textbook) and the given form of the density function, it follows that

\[ F_X(x) = 1 - \frac{1}{x^2} \quad \text{for } x \geq 1, \quad F_X(x) = 0 \quad \text{for } x < 1. \]

Indeed, for \( x \geq 1 \),

\[ F_X(x) = \int_1^{x} \frac{2}{s^3} \, ds = \left[ -\frac{1}{s^2} \right]_1^{x} = 1 - \frac{1}{x^2}. \]

b. From the given formula for \( f_X \) and the formula for the expectation of a continuous random variable (section 2.3 in the textbook), it follows that

\[ E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx = \int_1^{\infty} x \frac{2}{x^3} \, dx = -2 x^{-1} \bigg|_{1}^{\infty} = 2. \]
Note that we obtain the same answer using the formula $E[X] = \int_0^{\infty} S_X[x] \, dx$. Indeed, since $S_X[x] = 1/x^2$ for $x \geq 1$ and $S_X[x] = 1$ for $x < 1$, we have

$$E[X] = 1 + \int_1^{\infty} \frac{1}{x^2} \, dx = 2.$$  

c. From the formula for $F_X$ determined in part b, we have

$$\Pr[X > 4] = 1 - \Pr[X \leq 4] = 1 - F_X[4] = 1 - \left(1 - \frac{1}{16}\right) = \frac{1}{16}.$$  

d. The graphs of $f_X$ and $F_X$ can be created using Mathematica or similar computer software. Note that only the values for $x \geq 1$ have been plotted since $f_X$ and $F_X$ are both zero for $x < 1$. 

![Graph of $f_X$ and $F_X$]
Note that the slope of the graph of $F_X$ at $x = a$ is $f_X(a)$. In particular, the slope of $F_X$ at $x = 1$ is 2. (Here, we are implicitly considering the slope at $x = 1$ to be the slope determined by considering points to the right of $x = 1$.)

20. Since $X$ and $Y$ are independent, we can complete the contingency table using the multiplicative relationship $p_{X,Y}[x, y] = p_X[x] p_Y[y]$. We also need to use the facts that $\sum p_X[x] = 1$ and $\sum p_Y[y] = 1$. The procedure for completing the table is as follows:

i. We are given that $p_X[2] = .4$. Hence, $p_X[1] = .6$.

ii. Using the value of $p_X[1]$ obtained in i, the given values for $p_{X,Y}[1, 1]$ and $p_{X,Y}[1, 4]$, and the multiplicative relationship for $p_{X,Y}[x, y]$, we get $p_Y[1] = .4$ and $p_Y[4] = .2$.


iv. From the marginal distributions for $X$ and $Y$, we can complete the rest of the contingency table using the multiplicative relationship for $p_{X,Y}$.

The completed contingency table for $X$ and $Y$ is as follows:
23. Let $V$ be the value of a $1$ investment in the given security two days after the date of initial investment.

a. A $1$ investment that gains 50% and then loses 40% will only be worth

\[(\$1.00)(1.50)(0.60) = \$0.90.\]

Since there is an equal chance of a gain or loss on any given day, this suggests that it is not beneficial to hold the security for 2 days.

b. There are four possible outcomes: gains on both days, losses on both days, a gain followed by a loss, or a loss followed by a gain. Since gains and losses on different days are independent, it follows that

<table>
<thead>
<tr>
<th>$X$</th>
<th>2</th>
<th>12</th>
<th>04</th>
<th>08</th>
<th>04</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.16</td>
<td>.12</td>
<td>.04</td>
<td>.08</td>
<td>.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Y$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.24</td>
<td>.18</td>
<td>.06</td>
<td>.12</td>
</tr>
</tbody>
</table>

\[.4 \quad .3 \quad .1 \quad .2 \quad 1\]
\[
V = (1.50)^2 \text{ with probability } \frac{1}{4},
\]
\[
V = (1.50)(0.60) \text{ with probability } \frac{1}{2},
\]
\[
V = (0.60)^2 \text{ with probability } \frac{1}{4}.
\]

Hence, the probability mass function for \( V \) is given by
\[
p_V[0.36] = \frac{1}{4}, \quad p_V[0.90] = \frac{1}{2}, \quad p_V[2.25] = \frac{1}{4}, \quad p_V[v] = 0 \text{ otherwise.}
\]

c. From the answer to part b, we have
\[
\Pr[V > 1] = \frac{1}{4}
\]

and
\[
E[V] = (2.25) \left( \frac{1}{4} \right) + (0.90) \left( \frac{1}{2} \right) + (0.36) \left( \frac{1}{4} \right) = 1.1025.
\]

Hence, there is only a 25% chance of our coming out ahead after two days. This agrees with our observation in part a. The fact that \( E[V] > 1 \) may lead one to believe that the investment is a good one. However, this is only true in certain circumstances, as explained in the next part of the question.

d. From part c, \( E[V] = 1.1025 > 1. \) Since \( E[V] \) represents the average accumulation per investment for a large number of independent investments of the type described, it follows that the investment opportunity is a good one if we can make a large number of independent investments of this type. This is so even though only one quarter of the investments will be profitable (\( \Pr[V > 1] = .25 \) from part c) because the gains, when they occur, more than make up for the losses on the other three quarters of the investments. Note the importance of the assumption that investment returns on the individual investments are independent: If all of the investments had the same 2-day gain-loss pattern, then it would not be advantageous to invest for the reasons given in part a.
2. We are given that \( \Pr[E] = .3 \), \( \Pr[F] = .5 \), \( \Pr[E \mid F] = .4 \), and we are required to calculate \( \Pr[E \cap F] \), \( \Pr[E \cup F] \), and \( \Pr[F \mid E] \). From the definition of conditional probability and the given information, we have

\[
\Pr[E \cap F] = \Pr[E \mid F] \Pr[F] = (.4)(.5) = .2.
\]

Hence,

\[
\Pr[F \mid E] = \frac{\Pr[E \cap F]}{\Pr[E]} = \frac{.2}{.3} = \frac{2}{3}
\]

and

\[
\Pr[E \cup F] = \Pr[E] + \Pr[F] - \Pr[E \cap F] = .3 + .5 - .2 = .6.
\]

8. In this question, we are required to determine estimates for \( \Pr[E \cup F] \) and \( \Pr[E \cap F] \) when given the values of \( \Pr[E] \) and \( \Pr[F] \) only.

From Boole’s inequality (exercise 6), we know that \( \Pr[E \cup F] \geq \Pr[E] + \Pr[F] \). We also know from basic properties of probabilities that \( \Pr[E \cup F] \leq \Pr[E] \), \( \Pr[E \cup F] \leq \Pr[F] \), and \( \Pr[E \cup F] \leq 1 \). Consequently, \( \Pr[E \cup F] \) satisfies the inequality \( \max[\Pr[E], \Pr[F]] \leq \Pr[E \cup F] \leq \min[1, \Pr[E] + \Pr[F]] \).

This is the strongest statement one can make about \( \Pr[E \cup F] \) without having information about the nature of \( E \cap F \). The case \( E \subset F \) illustrates that the lower bound is best possible and the case \( E \cap F = \emptyset \) illustrates that the upper bound is best possible.

We can also derive sharp estimates for \( \Pr[E \cap F] \) using Bonferroni’s inequality and
properties of probabilities. Indeed, from Bonferroni’s inequality (exercise 7), we know that \( \Pr[E \cap F] \geq \Pr[E] + \Pr[F] - 1 \). We also know that \( \Pr[E \cap F] \leq \Pr[E] \), \( \Pr[E \cap F] \leq \Pr[F] \), and \( \Pr[E \cap F] \geq 0 \). Consequently, 
\[
\max[0, \Pr[E] + \Pr[F] - 1] \leq \Pr[E \cap F] \leq \min[\Pr[E], \Pr[F]].
\]

This is the strongest statement one can make about \( \Pr[E \cap F] \) without having information about \( E \cup F \). The case \( E \subset F \) illustrates that the upper bound is best possible and the case \( E \cup F = S \) illustrates that the lower bound is best possible.

We are now ready to answer the question.

a. Suppose that \( \Pr[E] = .7 \) and \( \Pr[F] = .4 \). Then the strongest statements we can make about \( \Pr[E \cup F] \) and \( \Pr[E \cap F] \) are
\[
.7 \leq \Pr[E \cup F] \leq 1
\]
and
\[
.1 \leq \Pr[E \cap F] \leq .4.
\]

b. Suppose that \( \Pr[E] = .6 \) and \( \Pr[F] = .2 \). Then the strongest statements we can make about \( \Pr[E \cup F] \) and \( \Pr[E \cap F] \) are
\[
.6 \leq \Pr[E \cup F] \leq .8
\]
and
\[
0 \leq \Pr[E \cap F] \leq .2.
\]

c. Since probabilities are always less than or equal to 1, Boole’s inequality provides nontrivial information for \( \Pr[E \cup F] \) if and only if \( \Pr[E] + \Pr[F] < 1 \).

d. Since probabilities are always greater than or equal to 0, Bonferroni’s inequality provides nontrivial information for \( \Pr[E \cap F] \) if and only if \( \Pr[E] + \Pr[F] > 1 \).

e. Since it is not possible for both the statements \( \Pr[E] + \Pr[F] < 1 \) and \( \Pr[E] + \Pr[F] > 1 \) to be true, it follows from parts c and d that it is not possible for both Boole’s inequality and Bonferroni’s inequality to provide nontrivial information. Only one of these inequalities can provide nontrivial information for a given pair of events. This was illustrated numerically in parts a and b.
11. Let $E$, $B$, $M$ be the following events:

- $E$: Employee owns units of the equity fund.
- $B$: Employee owns units of the bond fund.
- $M$: Employee owns units of the money market fund.

We are given the following information: $\Pr(E) = .15$, $\Pr(B) = .28$, $\Pr(M) = .30$, $\Pr(E \cap B) = .08$, $\Pr(E \cap M) = .10$, $\Pr(B \cap M) = .15$, $\Pr(E \cap B \cap M) = .05$. From this information, it is straightforward to construct a Venn diagram using *Mathematica* or similar computer software.

![Venn Diagram](image)

The answers to parts a through f follow directly from this Venn diagram.

a. The percentage of eligible employees currently participating in the pension plan is $\Pr(E \cup B \cup M) = .45$.

b. The percentage of eligible employees currently not participating is $\Pr((E \cup B \cup M)^c) = .55$. 
c. The percentage of participating employees who direct their contributions to a single fund can be calculated by dividing the fraction of all eligible employees who direct their contributions to a single fund by the fraction of all eligible employees who are participating. The fraction of eligible employees who direct their contributions to a single fund is equal to the fraction of eligible employees whose contributions go to the equity fund only plus the fraction whose contributions go to the bond fund only plus the fraction whose contributions go to the money market fund only. From the Venn diagram, it follows that the fraction of eligible employees whose contributions go to a single fund is \(0.02 + 0.10 + 0.10 = 0.22\). Consequently, the percentage of participating employees who direct their contributions to a single fund is
\[
\frac{0.02 + 0.10 + 0.10}{0.45} = \frac{0.22}{0.45} = 0.55.
\]
The desired probability can also be described in probability notation as follows:
\[
\Pr[(E \setminus (B \cup M)) \cup (B \setminus (E \cup M)) \cup (M \setminus (E \cup M)) \cap (E \cup B \cup M)].
\]
From this expression, it should be clear that working with the Venn diagram is the best approach to take to determine the desired probability!

d. The percentage of participating employees who direct their contributions to at least two different funds is simply the complement of the probability determined in part c. Hence, the desired probability is
\[
1 - \frac{22}{45} = \frac{23}{45}.
\]

e. The fraction of participants with bond shares who also own stock shares is
\[
\Pr[E \mid B] = \frac{\Pr[E \cap B]}{\Pr[B]} = \frac{0.08}{0.28} = \frac{2}{7}.
\]
f. The fraction of participants with money market shares who also own stock shares is
\[
\Pr[E \mid M] = \frac{\Pr[E \cap M]}{\Pr[M]} = \frac{0.10}{0.30} = \frac{1}{3}.
\]

14. Let C, I, P be the following events:
   C: Worker belongs to a company pension plan.
Let $C, I, P$ be the following events:

- $C$: Worker belongs to a company pension plan.
- $I$: Worker has an IRA.
- $P$: Worker has private savings in excess of $5000.

We are given the following information:

- $\Pr[C] = .25$, $\Pr[I] = .20$, $\Pr[P] = .30$,
- $\Pr[C \cap I \cap P] = .05$, $\Pr[(C \cup I \cup P)^c] = .55$, $\Pr[I \mid C] = .60$, $\Pr[I \mid P] = \frac{1}{3}$.

To construct the Venn diagram for this problem, let $x = \Pr[(C \cap P) \setminus I]$.

The value of $x$ can be determined from the condition that all the probabilities in this Venn diagram must sum to 1. That is,

$$(.10 - x) + .10 + .05 + x + .05 + (.20 - x) + 0 + .55 = 1.$$ 

Solving for $x$, we obtain $x = .05$. With this information and the Venn diagram just constructed, we can provide answers to parts a through d.

a. The fraction of people with an IRA who also have private retirement savings in excess of $5000$ is, by Bayes' theorem, equal to

$$\Pr[P \mid I] = \frac{\Pr[I \mid P] \Pr[P]}{\Pr[I]} = \frac{\left(\frac{1}{3}\right)(.30)}{.20} = \frac{1}{2}.$$
b. The fraction of people with an IRA that also belong to a company pension plan is, by Bayes' theorem, equal to
\[ \Pr[C \mid I] = \frac{\Pr[I \mid C] \Pr[C]}{\Pr[I]} = \frac{.60 \times .25}{.20} = .75. \]

c. The fraction of people who belong to a company pension plan that have no other retirement savings besides social security is
\[ \Pr[C \setminus (I \cup P) \mid C] = \frac{.10 - x \times .25}{.25} = \frac{.25 - x}{.25} = \frac{5 - 5x}{5} = \frac{1 - x}{1}. \]

d. The fraction of people with private savings in excess of $5000 that do not participate in a company pension plan is
\[ \Pr[C^c \mid P] = \frac{\Pr[C^c \cap P]}{\Pr[P]} = \frac{(20 - x) + .05}{.30} = \frac{20 - x}{.30} = \frac{2}{3}. \]

17. Let G, B, A, C be the following events:

- **G**: Policyholder classified as a good risk.
- **B**: Policyholder classified as a bad risk.
- **A**: Policyholder classified as an average risk.
- **C**: Policyholder files an accident claim.

We are given the following information: \( \Pr[G] = .30, \Pr[B] = .20, \Pr[A] = .50, \Pr[C \mid G] = .05, \Pr[C \mid B] = .40, \Pr[C \mid A] = .10. \)

a. The probability that a randomly chosen customer files an accident claim in the coming year is, by the law of total probability,
\[ \Pr[C] = \Pr[C \mid G] \Pr[G] + \Pr[C \mid B] \Pr[B] + \Pr[C \mid A] \Pr[A] = (.05) (.30) + (.40) (.20) + (.10) (.50) = .015 + .08 + .05 = .145. \]

b. Using Bayes' theorem and the answer to part a, the desired probabilities are
Chapter Three Solutions

21

\[ \Pr[G \mid C] = \frac{\Pr[C \mid G] \Pr[G]}{\Pr[C]} = \frac{(0.05)(0.30)}{0.145} = \frac{15}{145} = \frac{3}{29}, \]

and

\[ \Pr[B \mid C] = \frac{\Pr[C \mid B] \Pr[B]}{\Pr[C]} = \frac{(0.40)(0.20)}{0.145} = \frac{80}{145} = \frac{16}{29}, \]

respectively.

c. Let \( x \) be the required value for \( \Pr[A] \). Then \( \Pr[B] = 0.70 - x \), \( \Pr[G] = 0.30 \), and by the law of total probability,

\[
\Pr[C] = \Pr[C \mid G] \Pr[G] + \Pr[C \mid B] \Pr[B] + \Pr[C \mid A] \Pr[A] = (.05)(.30) + (.40)(.70 - x) + (.10) x = .295 - .30x.
\]

Hence \( \Pr[C] \leq 0.10 \) if and only if \( .295 - .30 x \leq .10 \), i.e., if and only if \( x \geq .65 \). Consequently, for the company's requirement to be met, at least 65% of the company's customers must be classified as average risks.

23. Let \( P \) and \( Q \) represent the following events:

- \( P \): Student passes test.
- \( Q \): Student is qualified.

We are given the following information: \( \Pr[Q^c] = 0.20 \), \( \Pr[P \mid Q] = 0.85 \), \( \Pr[P^c \mid Q^c] = 0.80 \).

a. \( \Pr[Q^c \mid P] \) represents the fraction of students who pass the test who are unqualified. \( \Pr[Q \mid P^c] \) represents the fraction of students who fail the test who are qualified. These quantities represent errors in the testing procedure and should be as small as possible.

b. If the college is primarily concerned with screening unqualified applicants, then minimizing \( \Pr[Q^c \mid P] \) is more important than minimizing \( \Pr[Q \mid P^c] \).

c. If the college is primarily concerned with reducing the number of qualified students
who are denied admission because they failed the test, then minimizing $\Pr[Q \mid P^c]$ is more important than minimizing $\Pr[Q^c \mid P]$. This implicitly assumes that every applicant who passes the test is granted admission, everyone who fails is denied admission, and that the college has the capacity to accommodate whatever number of candidates pass the test.

d. By the law of total probability and the property $\Pr[P \mid Q^c] = 1 - \Pr[P^c \mid Q^c]$, we have $\Pr[P] = \Pr[P \mid Q] \Pr[Q] + \Pr[P \mid Q^c] \Pr[Q^c] = (.85)(1 - .20) + (1 - .80)(.20) = .72$.

Hence, by Bayes’ theorem and the properties $\Pr[P \mid Q^c] = 1 - \Pr[P^c \mid Q^c]$ and $\Pr[P^c \mid Q] = 1 - \Pr[P \mid Q]$, we have

$$\Pr[Q^c \mid P] = \frac{\Pr[P \mid Q^c] \Pr[Q^c]}{\Pr[P]} = \frac{(1 - .80)(.20)}{.72} = \frac{1}{18} \approx .06$$

and

$$\Pr[Q \mid P^c] = \frac{\Pr[P^c \mid Q] \Pr[Q]}{\Pr[P^c]} = \frac{(1 - .85)(1 - .20)}{1 - .72} = \frac{3}{7} \approx .43.$$  

Consequently, if the goal of the test is to limit the number of unqualified applicants who gain admission then the test is fairly good because $\Pr[Q^c \mid P]$ is small. However, if the goal is to limit the number of qualified applicants who are denied admission then the test is not very good because $\Pr[Q \mid P^c]$ is quite large.

e. The files of students who drop out (i.e., are unqualified) are likely to be kept separate from the files of students who continue. If the files are organized in this way, then it is relatively simple to determine the fraction of drop-outs who failed the initial test and the fraction of continuing students who passed the initial test. Since the files are likely to be organized in this way, it is more likely that the values of $\Pr[P \mid Q]$ and $\Pr[P^c \mid Q^c]$ will be observed than the values of $\Pr[Q^c \mid P]$ and $\Pr[Q \mid P^c]$. In fact, if the test results are used for any sort of screening, then it is not even possible to observe the value of $\Pr[Q \mid P^c]$ because students who fail the test and as a result are denied admission will never get a chance to prove that they are qualified.
4. The probability masses associated with a mixed distribution occur at the points of discontinuity of the distribution function. From the definition of $F$, it is clear that $F$ has two jump discontinuities: one of size $\frac{1}{6}$ at $x = 0$ and the other of size $\frac{1}{3}$ at $x = 2$. Between the points $x = 0$ and $x = 2$ there is a continuous distribution of probability given by the density

$$f(x) = F'(x) = \frac{1}{4} \quad \text{for} \ 0 < x < 2.$$ 

Note that the amount of probability between 0 and 2 is $\int_0^2 \frac{1}{4} \, dx = \frac{1}{2}$. When this number is added to the probability masses at $x = 0$ and $x = 2$, we obtain 1, as we should. This describes the form of the generalized density for $X$.

To express $F_X$ as a weighted sum of a continuous distribution function and a discrete distribution function, we first extract a continuous distribution function $F_C$ from $F_X$ by removing the jumps from the graph of $F_X$ and scaling appropriately. The jumps can be removed by subtracting appropriate multiples of the unit step functions at $x = 0$ and $x = 2$. The appropriate scaling factor is determined by the requirement that $\lim_{x \to \infty} F_C(x) = 1$. Following these steps, we obtain

$$F_C(x) = 2 \left( F_X(x) - \frac{1}{6} u_0(x) - \frac{1}{3} u_2(x) \right),$$

where $u_a$ is the unit step function at the point $x = a$, i.e.,
$u_a[x] = 0$ for $x < a,$
$u_a[x] = 1$ for $x \geq a.$

Rearranging this equation for $F_C,$ we obtain

$$F_X[x] = \frac{1}{2} F_C[x] + \frac{1}{6} u_0[x] + \frac{1}{3} u_2[x].$$

Hence,

$$F_X[x] = \frac{1}{2} F_C[x] + \frac{1}{2} F_D[x],$$

where $F_D[x] = \frac{1}{3} u_0[x] + \frac{2}{3} u_2[x].$ Note that $F_D$ is the distribution function for the discrete distribution with probability masses of $\frac{1}{3}$ at $x = 0$ and $\frac{2}{3}$ at $x = 2.$ Note also that

$F_C[x] = \frac{1}{2} x$ for $0 \leq x < 2.$ Consequently, we have shown that $F_X$ can be written as a weighted sum of a continuous distribution function and a discrete distribution function as required.

7. a. From the definition of the given distribution function, there are five different values for $F_{X_1}, X_2$ and these values are assumed on five distinct regions. Hence, the graph of $F_{X_1, X_2}$ can be represented in two dimensions using five degrees of shading. This two-dimensional graph can be created using Mathematica or similar computer software.
Note that in this graph the more lightly shaded regions are the regions on which the value of $F_{X_1,X_2}$ is greater. Note further that this graph only displays the portion of $F_{X_1,X_2}$ inside the square $[-1, 3] \times [-1, 3]$. However, from the given graph, the nature of $F_{X_1,X_2}$ outside this square should be readily apparent.

b. The two-dimensional graph created in part a suggests that a three-dimensional representation of the given $F_{X_1,X_2}$ will consist of several blocks with rectangular faces. The required three-dimensional graph can be created using Mathematica or similar computer software.
Note that the view point for this picture is in the third octant (i.e., the octant with \( x_1 < 0, \ x_2 < 0, \) and \( z > 0 \)) rather than the more customary first octant (i.e., the octant with \( x_1 > 0, \ x_2 > 0, \) and \( z > 0 \)). Since \( F_{X_1,X_2} \) is increasing in both \( x_1 \) and \( x_2 \), we get a better visual representation for the graph of \( F_{X_1,X_2} \) by doing this. Choosing a view point in the third octant also facilitates comparisons with the two-dimensional graph generated in part a and makes the determination of a formula for \( p_{X_1,X_2} \) in part c simpler.

c. It is relatively straightforward to determine the probability mass function for a discrete univariate random variable from its distribution function. Indeed, the locations and sizes of the probability masses are simply the locations and sizes of the jumps in the graph of the distribution function. However, determining the probability mass function for a discrete bivariate random variable from its distribution function is not quite so simple.

It is still true that the presence of a probability mass results in a jump on the graph of
Chapter Four Solutions

\[ F_{X_1, X_2} \] at the location of the probability mass. The demonstration of this fact in the bivariate case is similar to its demonstration in the univariate case (see Example 1, section 4.1.2 of the textbook). However, it is no longer true that every jump on the graph of \( F_{X_1, X_2} \) arises in this way: Looking at the graph generated in part b, we can see that this graph has jumps along each of the lines \( x_1 = a \) for \( a > 0 \). If each of these jumps corresponded to a different probability mass, the set of points with non-zero probability mass would be infinite. But then the function \( F_{X_1, X_2} \) itself would have an infinite number of values rather than the five that it actually does.

If we think a little bit harder about what actually happens to the distribution function at a point where a probability mass is located, we soon realize that at such points a new "block" is created. (Consider the graph in part b.) From this realization, it follows that for the specified \( F_{X_1, X_2} \) the only possible locations for probability masses are the points \((0, 0), (0, 1), (1, 0), (1, 1)\) (see the graph in part b). After a little reflection, it becomes apparent that there are non-zero probability masses at each of these points.

The size of the probability mass at \((0, 0)\) is relatively straightforward to determine. From the graph of \( F_{X_1, X_2} \), it is \( \frac{1}{8} \). The size of the probability mass at \((0, 1)\) can be determined by moving along the line \( x_1 = 0 \) and calculating the size of the jump that occurs when the point \((0, 1)\) is reached. Following this procedure, we find that \[ p_{X_1, X_2}(0, 1) = \frac{3}{8} - \frac{1}{8} = \frac{1}{4}. \] Using a similar approach, we get \[ p_{X_1, X_2}(1, 0) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8}. \] The size of the probability mass at \((1, 1)\) is then determined by the requirement that the probability masses must sum to 1. Hence, \[ p_{X_1, X_2}(1, 1) = \frac{1}{2}. \]

To summarize, the probability mass function for the specified distribution is

\[
\begin{align*}
p_{X_1, X_2}(0, 0) &= \frac{1}{8}, & p_{X_1, X_2}(0, 1) &= \frac{1}{4}, & p_{X_1, X_2}(1, 0) &= \frac{1}{8}, & p_{X_1, X_2}(1, 1) &= \frac{1}{2}.
\end{align*}
\]

It is straightforward to check that the distribution function corresponding to this probability mass function has the form specified. Hence, by the uniqueness of probability mass functions, this must be the correct definition for \( p_{X_1, X_2} \).

d. The graph of the probability mass function specified in part c can be created using Mathematica or similar computer software.
Note that the view point for this graph is in the third octant to facilitate comparisons with part b.

e. The distribution functions of bivariate and univariate distributions have the following similarities:

i. Both are non-decreasing.

ii. Both have values between 0 and 1.

iii. Both having limiting values of 0 and 1 at "extreme" locations.

However, they also have some important differences:

i. Not every jump on the graph of a bivariate distribution function corresponds to a
probability mass (see the discussion in the answer to part c).

ii. The function $F_{X_1, X_2}$ need not tend to 1 along every line to infinity. Consider, for example the line $x_1 = \frac{1}{2}$ or the line $x_2 = \frac{1}{2}$ on the graph constructed in part b.

iii. It is not generally possible to determine the value of $p_{X_1, X_2}$ by looking at the differences in height between two neighboring planes. Instead, one must consider the relationships among the heights of all neighboring planes at the point where a probability mass is located. As an illustration, consider the point $(1, 1)$ for the $F_{X_1, X_2}$ of this problem (see the answer to part c for details).

9. a. A graph of the region of nonzero probability can be created using Mathematica or similar computer software.
A graph of the density function in three-dimensional space can be created using Mathematica or similar computer software.
b. Recall that from a graphical perspective, conditional densities are scaled cross-sections (see section 4.1.9). From the graph of the bivariate density $f_{X_1, X_2}$ created in part a, the cross-sections parallel to the respective axes define rectangular regions. We can see this more clearly from graphs that highlight the cross-sections:
From these two graphs, we can make the following observations:

i. The cross-section defined by $X_1 = x_1$ outlines a rectangle with base length $2 - x_1$ and height $\frac{1}{2}$. Hence for each $x_1$, the distribution of $X_2 \mid X_1 = x_1$ is uniform on $(0, 2 - x_1)$, that is,
   
   \[ f_{X_2 \mid X_1 = x_1}[x_2] = \frac{1}{2 - x_1} \quad \text{for } x_2 \in (0, 2 - x_1). \]

ii. The cross-section defined by $X_2 = x_2$ outlines a rectangle with base length $2 - x_2$ and height $\frac{1}{2}$. Hence for each $x_2$, the distribution of $X_1 \mid X_2 = x_2$ is uniform on $(0, 2 - x_2)$, that is,
From the formulas for \( f_{X_2|X_1=x_1} \) and \( f_{X_1|X_2=x_2} \) just determined, it is clear that knowledge of the value assumed by one of the random variables affects the distribution of probability for the other. Consequently, \( X_1 \) and \( X_2 \) are not independent.

c. Recall that from a graphical perspective, marginal densities are projections. From the graphs created in part b, we can make the following observations:

i. The cross-section defined by \( X_1 = x_1 \) outlines a rectangle with base length \( 2 - x_1 \) and height \( \frac{1}{2} \). Hence the amount of probability projected onto the \( x_1 \)-axis at the point \( x_1 \) is \( \frac{1}{2} (2 - x_1) \) (the area of this rectangle). Consequently, the marginal density of \( X_1 \) is given by \( f_{X_1}[x_1] = \frac{1}{2} (2 - x_1) \) for \( x_1 \in (0, 2) \).

ii. The cross-section defined by \( X_2 = x_2 \) outlines a rectangle with base length \( 2 - x_2 \) and height \( \frac{1}{2} \). Hence the amount of probability projected onto the \( x_2 \)-axis at the point \( x_2 \) is \( \frac{1}{2} (2 - x_2) \) (the area of this rectangle). Consequently, the marginal density of \( X_2 \) is given by \( f_{X_2}[x_2] = \frac{1}{2} (2 - x_2) \) for \( x_2 \in (0, 2) \).

It is straightforward to verify that these formulas are correct using the algebraic definitions given in section 4.1.8. Indeed,

\[
f_{X_1}[x_1] = \int_{-\infty}^{\infty} f_{X_1, X_2}[x_1, x_2] \, dx_2 = 
\int_{-\infty}^{0} 0 \, dx_2 + \int_{0}^{2-x_1} \frac{1}{2} \, dx_2 + \int_{2-x_1}^{\infty} 0 \, dx_2 = \frac{1}{2} (2 - x_1) \quad \text{for} \ x_1 \in (0, 2)
\]

and similarly,

\[
f_{X_2}[x_2] = \frac{1}{2} (2 - x_2) \quad \text{for} \ x_2 \in (0, 2).
\]

d. The formulas for the conditional densities determined in part b can be verified using the algebraic definition of conditional density given in section 4.1.9 and the formulas for the marginal densities determined in part c. Indeed,
\[ f_{X_1 \mid X_2} [x_1] = \frac{f_{X_1, X_2} [x_1, x_2]}{f_{X_2} [x_2]} = \frac{\frac{1}{2}}{2 - x_2} = \frac{1}{2 - x_2} \text{ for } x_1 \in (0, 2 - x_2) \text{ and } x_2 \in (0, 2). \]

Note that for fixed \( x_2 \in (0, 2) \), \( f_{X_1, X_2} [x_1, x_2] = 0 \) for \( x_1 < 0 \) or \( x_1 > 2 - x_2 \). Similarly,

\[ f_{X_2 \mid X_1} [x_2] = \frac{f_{X_1, X_2} [x_1, x_2]}{f_{X_1} [x_1]} = \frac{\frac{1}{2}}{2 - x_1} = \frac{1}{2 - x_1} \text{ for } x_2 \in (0, 2 - x_1) \text{ and } x_1 \in (0, 2). \]

Note that these formulas only hold for the specified values of \( x_1 \) and \( x_2 \).

e. Graphs of the marginal and conditional densities are as follows:
These graphs are consistent with the graphical interpretations of $f_{X_1}$, $f_{X_2}$, $f_{X_1|X_2}$, and $f_{X_2|X_1}$ considered in part b and part c.

f. Recall that probabilities associated with bivariate distributions can be interpreted as volumes of particular regions under the two-dimensional surface defined by the density function. Since the density function in this exercise assumes the constant value $\frac{1}{2}$ on the region of nonzero probability, it follows that the probability $\Pr[X_1 > 2 X_2]$ is equal to $\frac{1}{2}$ times the area of the region of nonzero probability defined by $X_1 > 2 X_2$. The latter region is illustrated in the following graph:
From basic geometry, the shaded region in this graph has area

\[
\left( \frac{1}{2} \right) \left( \frac{4}{3} \right) \left( \frac{2}{3} \right) + \left( \frac{1}{2} \right) \left( 2 - \frac{4}{3} \right) \left( \frac{2}{3} \right) = \frac{4}{9} + \frac{2}{9} = \frac{2}{3}. 
\]

Consequently, the desired probability is

\[ \Pr[X_1 > 2X_2] = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}. \]

11. a. By the distributional form of the law of total probability,
Using integration by parts twice, we have

\[
\int_0^\infty 4\lambda^2 e^{-\lambda (x+2)} d\lambda = 4\lambda^2 \frac{e^{-\lambda (x+2)}}{-(x+2)} \bigg|_{\lambda=0}^\infty - \int_0^\infty 8\lambda e^{-\lambda (x+2)} \frac{d\lambda}{-(x+2)} = \\
0 + \frac{8}{x+2} \int_0^\infty \lambda e^{-\lambda (x+2)} d\lambda = \frac{8}{x+2} \left( \frac{\lambda e^{-\lambda (x+2)}}{-(x+2)} \bigg|_{\lambda=0}^\infty - \int_0^\infty 1 \times \frac{e^{-\lambda (x+2)}}{-(x+2)} d\lambda \right) = \\
\frac{8}{x+2} \left( 0 + \frac{1}{x+2} \int_0^\infty e^{-\lambda (x+2)} d\lambda \right) = \frac{8}{(x+2)^3}.
\]

So

\[ f_X[x] = \frac{8}{(x+2)^3} \quad \text{for } x > 0. \]

Using this formula, we have

\[
Pr[X > 2] = \int_2^\infty 8 (x+2)^{-3} dx = -4 (x+2)^{-2} \bigg|_2^\infty = \frac{1}{4}.
\]

b. If a claim of size two is received, then the insurer's belief about the true value of $\lambda$ going forward is captured by the distribution of $\Lambda | X = 2$. Using the distributional form of Bayes' theorem we have

\[ f_{\Lambda|X=2}[\lambda] = \frac{f_{X|\Lambda=2}[2] f_{\Lambda}[\lambda]}{f_X[2]} . \]

From the given information,

\[ f_{X|\Lambda=2}[2] = \lambda e^{-\lambda x} |_{x=2} = \lambda e^{-2\lambda}, \]

\[ f_{\Lambda}[\lambda] = 4\lambda e^{-2\lambda}. \]

Further from part a,

\[ f_X[2] = 8 (x+2)^{-3} |_{x=2} = \frac{1}{8}. \]
Consequently, the density of $\Lambda \mid X = 2$ is given by

$$f_{\Lambda \mid X=2}[\lambda] = \frac{\left(\lambda e^{-2.1}\right) \left(4 \lambda e^{-2.1}\right)}{1/8} = 32 \lambda^2 e^{-4.1} \quad \text{for } \lambda > 0.$$ 

This density encapsulates the insurer's belief about $\lambda$ going forward.

**Section 4.2.4 Exercises**

3. The expectation of a function of a mixed random variable can be calculated by considering sums over the discrete part and integrals over the continuous part (see section 4.2.1 for details). Hence

$$E[|X+1|] = \left(\frac{1}{x+1}\right)_{x=-2} \times \frac{1}{4} + \left(\frac{1}{x+1}\right)_{x=2} \times \frac{1}{4} + \int_{-1}^{0} \left(\frac{1+x}{2}\right) dx + \int_{0}^{1} \left(\frac{1-x}{2}\right) dx =$$

$$= \frac{1}{4} + \frac{3}{4} + \frac{1}{2} \int_{-1}^{0} (x+1)^2 dx + \frac{1}{2} \int_{0}^{1} (1-x^2) dx = 1 + \frac{1}{6} + \frac{1}{3} = \frac{3}{2}.$$ 

Note that $|x+1| = x+1$ for $x \geq -1$.

**Section 4.3.3 Exercises**

3. One of the important properties of the moment generating function for a random variable $X$ is that it characterizes the distribution of $X$, i.e., there is one and only one moment generating function associated with each probability distribution (see section 4.3.1). Hence, if we can construct a probability distribution whose moment generating function is the one given in this exercise, then that distribution must be the distribution of $X$. 


The presence in \( M_X \) of \( \frac{1}{1-t} \), which is the moment generating function for an exponential distribution with parameter \( \lambda = 1 \) (see Example 6 of section 4.3.1 and Example 2 of section 4.2.1), suggests that the distribution of \( X \) contains an "exponential component". At the same time, the presence of the term \( \frac{1}{4} \) suggests that there is a probability mass of size \( \frac{1}{4} \) at \( x = 0 \). Taken together, these observations suggest that \( X \) has a mixed distribution with a discrete probability mass at \( x = 0 \) and a continuous exponential part on the interval \( x > 0 \). As an initial guess, consider the distribution for a random variable \( Y \) with probability mass \( \frac{1}{4} \) at \( y = 0 \) and continuous distribution on \( y > 0 \) given by

\[
f_Y(y) = \frac{3}{4} e^{-y}, \quad y > 0.
\]

From the definition of moment generating function and the formula for calculating the expectation of a mixed random variable (see sections 4.3.1 and 4.2.1), we have

\[
M_Y(t) = E_Y[e^{tY}] = e^0 \cdot \frac{1}{4} + \int_0^\infty e^{ty} \cdot \frac{3}{4} e^{-y} \, dy = \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{1-t}, \quad t > 1
\]

which is identical to the moment generating function given. Hence, by the uniqueness of moment generating functions, it follows that the mass-density function for \( X \) is given by

\[
p_X[0] = \frac{1}{4}, \quad f_X(x) = -e^{-x}, \quad x > 0.
\]

From this, it follows that the distribution function of \( X \) is

\[
F_X(x) = \Pr[X \leq x] = \begin{cases} 
\frac{1}{4} + \frac{3}{4} (1 - e^{-x}) & \text{for } x \geq 0, \\
0 & \text{otherwise}.
\end{cases}
\]

Hence

\[
F_X(x) = \begin{cases} 
1 - \frac{3}{4} e^{-x} & \text{for } x \geq 0, \\
0 & \text{otherwise}.
\end{cases}
\]

The graphs of \( f_X \) and \( F_X \) can be created using Mathematica or similar computer software.
6. From the formulas given in section 4.3.1 (and derived in exercise 5), the mean, variance, and skewness can be determined from either the moment generating function or the cumulant generating function. For the distributions of this question, we will use whichever approach is simpler from a computational viewpoint. This means considering the cumulant generating function for parts a, d, and e, and the moment generating
From the formulas given in section 4.3.1 (and derived in exercise 5), the mean, variance, and skewness can be determined from either the moment generating function or the cumulant generating function. For the distributions of this question, we will use whichever approach is simpler from a computational viewpoint. This means considering the cumulant generating function for parts a, d, and e, and the moment generating function for parts b and c. Note however, that in each part, the mean, variance, and skewness can be determined using both approaches.

a. Since \( M_X[t] = (1 - t)^{-1} \), we have
\[
\psi_X[t] = -\log[1 - t].
\]
Differentiating \( \psi_X \) successively, we have
\[
\psi_X'[t] = (1 - t)^{-2},
\]
\[
\psi_X''[t] = 2 (1 - t)^{-3}.
\]
Hence
\[
\mu_X = \psi_X'[0] = 1,
\]
\[
\sigma_X^2 = \psi_X''[0] = 1,
\]
\[
\gamma_X = \frac{\psi_X^{(3)}[0]}{\psi_X''[0]^{3/2}} = \frac{2}{1} = 2.
\]

b. Since \( M_X[t] = \frac{1}{2} e^t + \frac{1}{2} e^{-2t} \), we have
\[
M_X'[t] = \frac{1}{2} e^t - e^{-2t},
\]
\[
M_X''[t] = \frac{1}{2} e^t + 2 e^{-2t},
\]
\[
M_X^{(3)}[t] = \frac{1}{2} e^t - 4 e^{-2t}.
\]
Hence
\[
E[X] = M_X'[0] = \frac{-1}{2},
\]
\[ E[X^2] = M_X^{(2)}[0] = \frac{5}{2}, \]
\[ E[X^3] = M_X^{(3)}[0] = \frac{-7}{2}. \]

Consequently,
\[ \mu_X = \frac{-1}{2}, \]
\[ \sigma_X^2 = E[X^2] - E[X]^2 = \frac{5}{2} - \left(\frac{-1}{2}\right)^2 = \frac{9}{4}, \]
and
\[ \gamma_X = \frac{E[X^3] - 3 E[X^2] E[X] + 2 E[X]^3}{\sigma_X^3} = \frac{\left(\frac{-7}{2}\right) - 3 \left(\frac{5}{2}\right) \left(\frac{-1}{2}\right) + 2 \left(\frac{-1}{2}\right)^3}{\left(\frac{9}{4}\right)^{3/2}} = 0. \]

Note that it would be much more complicated to use the cumulant generating function here since the expression for \( \psi_X^{(k)} \) cannot be simplified to any great extent.

9. Recall that for non-negative random variables \( X \), the expected value can be calculated using the following formula:
\[ E[X] = \int_0^\infty S_X[x] \, d\, x \]
(see section 4.3.2). Hence for the random variable \( X \) with survival function
\[ S_X[x] = (1 + x)^{-2}, \quad x \geq 0 \]
we have
\[ E[X] = \int_0^\infty (1 + x)^{-2} \, d\, x = -(1 + x)^{-1} \bigg|_0^\infty = 1. \]

The variance of this particular Pareto distribution is actually infinite. To see this, note that
\[ f_X(x) = -S'_X(x) = \frac{2}{(-x)} \]

and

\[ E[X^2] = \int_0^\infty x^2 f_X(x) \, dx = \int_0^\infty 2x^2 (1 + x)^{-3} \, dx. \]

However, from the relationship \( 2x^2 \geq \frac{1}{2} (1 + x)^2 \) which holds for \( x \geq 1 \), we have

\[ \int_1^\infty 2x^2 (1 + x)^{-3} \, dx \geq \frac{1}{2} \int_1^\infty (1 + x)^{-1} \, dx = \frac{1}{2} \log(1 + x) \big|_1^\infty = \infty. \]

Consequently \( E[X^2] = \infty \) and so \( \text{Var}(X) = \infty \) as well.
Chapter Five Solutions

1. Recall that the probability mass function of the binomial distribution is given by

\[ p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n \]

and the mean and variance are

\[ \mu_X = np, \]
\[ \sigma_X^2 = np(1-p). \]

Using the general formulas

\[ E[(X - \mu_X)^3] = E[X^3] - 3E[X^2]E[X] + 2E[X]^3, \]
\[ M_X^{(3)}(0) = E[X^3] \]

and the formula for the moment generating function of a binomial random variable, one can show that the skewness of a binomial random variable is given by

\[ \gamma_X = \frac{1 - 2p}{(np(1-p))^{1/2}}. \]

From these formulas, one can give the qualitative descriptions requested in part a.

a. Suppose first that n is fixed. Then as p increases from 0 to 1, the distribution of probability moves from being concentrated around the point \( x = 0 \) to being concentrated around the point \( x = n \). In the limiting cases \( p = 0 \) and \( p = 1 \), the distribution reduces to a point mass. From the formula for \( \gamma_X \), it follows that the distribution is positively
skewed when \( p < \frac{1}{2} \), negatively skewed when \( p > \frac{1}{2} \), and has zero skew when \( p = \frac{1}{2} \).

From the formula for \( p_X \), it follows that the distribution obtained by replacing \( p \) with \( 1 - p \) is the reflection of the given binomial distribution in the line \( x = \frac{n}{2} \) and is itself a binomial distribution. Indeed,

\[
p_X\left[ \frac{n}{2} - k \right] = \frac{n!}{\left( \frac{n}{2} - k \right)! \left( \frac{n}{2} + k \right)!} p_{\frac{n}{2} - k} (1 - p)^{n/2 + k}.
\]

Hence a binomial distribution is symmetric if and only if \( p = \frac{1}{2} \).

From the formula for \( \sigma_X^2 \), it follows that the binomial distribution with the greatest variance for a given \( n \) is the one with \( p = \frac{1}{2} \) and the variance equals 0 when \( p = 0 \) or \( p = 1 \) (the cases in which the distribution reduces to a point mass). It also follows that the variance of a binomial distribution is invariant with respect to interchanging \( p \) and \( 1 - p \), which makes sense since the distributions with \( p \) and \( 1 - p \) interchanged are mirror images of one another, as noted earlier. From the formula for \( \mu_X \), it is clear that the mean of a binomial distribution is directly proportional to \( p \).

Now suppose that \( p \) is fixed and is not equal to 0 or 1. Then as \( n \) increases, \( \gamma_X \) approaches 0. Indeed,

\[
\gamma_X = \frac{1 - 2p}{(np(1-p))^{1/2}} \to 0 \quad \text{as} \quad n \to \infty.
\]

Hence for any fixed \( p \), the distribution becomes more symmetric as \( n \to \infty \). Moreover by considering graphs of \( p_X \) with the same \( p \) and various \( n \) it is apparent that the distribution becomes more "bell-shaped" as \( n \to \infty \). (This can be proved directly from the formula for \( p_X \), but students are not expected to furnish such a proof at this point in the book.) For fixed \( p \), the mean and variance are both directly proportional to \( n \).

Hence although the distribution becomes more bell-shaped as \( n \to \infty \), it is also true that the distribution's variance increases without bound as \( n \to \infty \).

b. From the formula for \( p_X \), it follows that

\[
\frac{p_X[x + 1]}{p_X[x]} = \left( \frac{n - x}{x + 1} \right) \left( \frac{p}{1 - p} \right) = \left( \frac{n + 1}{x + 1} - 1 \right) \left( \frac{p}{1 - p} \right)
\]
for \(x = 0, 1, \ldots, n - 1\). Hence, the ratio \(p_X[x + 1]/p_X[x]\) is decreasing for all \(x\).

Consequently to show that \(p_X\) first increases and then decreases it suffices to show that \(p_X[1]/p_X[0] > 1\) and \(p_X[n]/p_X[n - 1] < 1\). Now

\[
\frac{p_X[1]}{p_X[0]} = \frac{n \cdot p}{1 - p}
\]

and

\[
\frac{p_X[n]}{p_X[n - 1]} = \left(\frac{1}{n}\right) \left(\frac{p}{1 - p}\right)
\]

Hence

\[
\frac{p_X[1]}{p_X[0]} > 1 \Leftrightarrow n \cdot p > 1 - p \Leftrightarrow p > \frac{1}{n + 1}
\]

and

\[
\frac{p_X[n]}{p_X[n - 1]} < 1 \Leftrightarrow p < 1 - \frac{1}{n + 1}.
\]

Therefore, if \(n\) and \(p\) are such that \(1/(n + 1) < p < 1 - 1/(n + 1)\), then the graph of \(p_X\) first increases and then decreases. On the other hand, if \(p \leq 1/(n + 1)\) then \(p_X[1]/p_X[0] \leq 1\) in which case \(p_X[x + 1]/p_X[x] \leq 1\) for all \(x = 0, 1, \ldots, n - 1\) and the graph of \(p_X\) is always decreasing, whereas if \(p \geq n/(n + 1)\) then \(p_X[n]/p_X[n - 1] \geq 1\) in which case \(p_X[x + 1]/p_X[x] \geq 1\) for all \(x = 0, 1, \ldots, n - 1\) and the graph of \(p_X\) is always increasing.

c. From the answer to part b, the ratio \(p_X[x + 1]/p_X[x]\) is decreasing. Hence to determine the modes we need only determine the integers \(m\) for which

\[
\frac{p_X[m + 1]}{p_X[m]} \leq 1 \quad \text{and} \quad \frac{p_X[m]}{p_X[m - 1]} \geq 1.
\]

Now
\[
\frac{p_X[x+1]}{p_X[x]} \leq 1 \iff \left(\frac{n-x}{x+1}\right)\left(\frac{p}{1-p}\right) \leq 1 \iff x \geq (n+1)p - 1
\]

and
\[
\frac{p_X[x]}{p_X[x-1]} \geq 1 \iff \left(\frac{n-x+1}{x}\right)\left(\frac{p}{1-p}\right) \geq 1 \iff x \leq (n+1)p.
\]

Consequently the modes are the integers \(m\) such that
\[(n+1)p - 1 \leq m \leq (n+1)p.\]

Therefore when \((n+1)p\) is an integer there are two modes, one at \((n+1)p\) and one at \((n+1)p - 1\). Otherwise, the distribution has only one mode.

**Comment on Exercises 7 and 8:** The objective of exercises 7 and 8 is to introduce students to the important topic of distribution fitting. In many practical problems, one is given a set of data rather than an explicit formula for a distribution function and one must make inferences on the basis of this data alone. One approach to take in such situations is to fit the data to a distribution with known characteristics and use the fitted distribution as the probability model. This approach is known as *parametric* modeling and is illustrated in its simplest form in exercises 7 and 8. There are two questions that arise naturally in parametric modeling:

i. Which family of distributions (e.g., Poisson, binomial, negative binomial, etc) models the data best?

ii. Which distribution in the selected family (i.e., which values of the parameters in the selected family) models the data best?

Students will not yet have the background to answer either of these questions adequately (unless they have already taken a course in statistical estimation, which usually follows a course in probability). Exercises 7 and 8 provide an introduction to these important statistical questions in a relatively simple setting.

7. Let \(N\) be the number of claims associated with a randomly chosen policy.

   a. Let \(\hat{p}_N\) denote the relative frequency function for the given data set. (In statistics, this
a. Let \( p_{\hat{N}} \) denote the relative frequency function for the given data set. (In statistics, this is also referred to as the frequency function for the empirical distribution.) Then \( p_{\hat{N}} \) is given by

\[
\begin{align*}
\hat{p}_{\hat{N}}[0] &= .122, \\
\hat{p}_{\hat{N}}[1] &= .188, \\
\hat{p}_{\hat{N}}[2] &= .188, \\
\hat{p}_{\hat{N}}[3] &= .156, \\
\hat{p}_{\hat{N}}[4] &= .117, \\
\hat{p}_{\hat{N}}[5] &= .082, \\
\hat{p}_{\hat{N}}[6] &= .055, \\
\hat{p}_{\hat{N}}[7] &= .035, \\
\hat{p}_{\hat{N}}[8] &= .022, \\
\hat{p}_{\hat{N}}[9] &= .013, \\
\hat{p}_{\hat{N}}[10] &= .022. 
\end{align*}
\]

b. A bar chart for \( \hat{p}_{\hat{N}} \) can be created using Mathematica or similar computer software.

![Bar chart](image)

Note that the distribution is discrete and positively skewed. Based on the distributions studied in chapter 5, this suggests that possible models include the Poisson, negative binomial, or binomial with parameter \( p \) small.

c. From part a, the implied mean is

\[
(0)(.122) + (1)(.188) + (2)(.188) + ... + (9)(.013) + (10)(.022) = 2.998
\]

and the implied second moment is

\[
(0^2)(.122) + (1^2)(.188) + (2^2)(.188) + ... + (9^2)(.013) + (10^2)(.022) = 14.622.
\]

Hence the implied variance is

\[
14.622 - (2.998)^2 = 5.633996.
\]
To one decimal place, the implied mean and variance are 3.0 and 5.6 respectively. From these calculations it appears that $E[N] < \text{Var}(N)$. Since negative binomial distributions have this property (i.e., mean less than variance) but Poisson and binomial distributions do not, a negative binomial model would appear to be the best among these three choices.

d. The simplest way to fit a distribution to data is to match moments. Since the negative binomial distribution has two parameters, this means equating the first and second moments. We are instructed to fit the distribution by equating means and variances. Since the mean, variance and second moment are related by the general formula $\text{Var}(X) = E[X^2] - E[X]^2$, this approach will give us the same result, ignoring rounding errors. Note that moment matching is the simplest but by no means the only way to fit a distribution to data. Students will encounter other techniques when they take a course in statistical estimation.

From section 5.3, we know that the mean and variance of a negative binomial distribution with parameters $r$ and $p$ are

$$E[N] = \frac{r(1 - p)}{p}$$

and

$$\text{Var}(N) = \frac{r(1 - p)}{p^2}$$

respectively. In part c we determined the implied mean and variance for the given data set to be 3.0 and 5.6 to one decimal place respectively. Equating these respective values we have

$$\frac{r(1 - p)}{p} = 3.0$$

and

$$\frac{r(1 - p)}{p^2} = 5.6.$$

This system is solved most easily by dividing the first equation by the second and then
substituting the resulting value of \( p \) into the first equation to determine \( r \). When we do this, we obtain

\[
r = \frac{45}{13}, \quad p = \frac{15}{28}.
\]

Let \( p_N \) denote the probability mass function for the negative binomial distribution with parameters \( r = 45/13 \) and \( p = 15/28 \). Then

\[
p_N[n] = \frac{\Gamma\left(n + \frac{45}{13}\right)}{\Gamma\left(\frac{45}{13}\right) \Gamma[n + 1]} \left(\frac{15}{28}\right)^{\frac{45}{13}} \left(\frac{13}{28}\right)^n \quad \text{for } n = 0, 1, 2, \ldots
\]

Note that the more general form of the negative binomial probability mass function must be used here since the estimated value of \( r \) is not an integer. Approximate numerical values of \( p_N[n] \) for \( n = 0, 1, 2, \ldots, 10 \) can be easily determined using *Mathematica* or similar computer software:

<table>
<thead>
<tr>
<th>( n )</th>
<th>( p_N[n] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.115264</td>
</tr>
<tr>
<td>1</td>
<td>0.185245</td>
</tr>
<tr>
<td>2</td>
<td>0.191861</td>
</tr>
<tr>
<td>3</td>
<td>0.162168</td>
</tr>
<tr>
<td>4</td>
<td>0.121626</td>
</tr>
<tr>
<td>5</td>
<td>0.0842695</td>
</tr>
<tr>
<td>6</td>
<td>0.0551765</td>
</tr>
<tr>
<td>7</td>
<td>0.034626</td>
</tr>
<tr>
<td>8</td>
<td>0.021023</td>
</tr>
<tr>
<td>9</td>
<td>0.0124302</td>
</tr>
<tr>
<td>10</td>
<td>0.00719178</td>
</tr>
</tbody>
</table>

Comparing these numbers to the relative frequency function determined in part a, it
Comparing these numbers to the relative frequency function determined in part a, it appears that the negative binomial distribution with \( r = 45/13 \) and \( p = 15/28 \) is a reasonable fit.

e. We computed the implied mean and variance in part c by assuming that all probability at values greater than or equal to 10 is concentrated at the value 10. The effect of this assumption is to underestimate the true mean and variance of the distribution. To compensate for this, we could round up the implied mean and variance before equating them to the negative binomial distribution mean and variance formulas. For example, if we estimate the implied mean and variance as 3 and 6 respectively, then the parameter values for the corresponding negative binomial distribution are \( r = 3 \) and \( p = .5 \) and the probability mass function for this distribution is

\[
p_N[n] = \binom{n + 2}{2} \left( \frac{1}{2} \right)^{n+3} \quad \text{for } n = 0, 1, 2, \ldots
\]

Approximate numerical values for this particular set of \( p_N[n] \) for \( n = 0, 1, 2, \ldots, 10 \) can be determined using Mathematica or similar software.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( p_N[n] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>1</td>
<td>0.1875</td>
</tr>
<tr>
<td>2</td>
<td>0.1875</td>
</tr>
<tr>
<td>3</td>
<td>0.15625</td>
</tr>
<tr>
<td>4</td>
<td>0.117188</td>
</tr>
<tr>
<td>5</td>
<td>0.0820313</td>
</tr>
<tr>
<td>6</td>
<td>0.0546875</td>
</tr>
<tr>
<td>7</td>
<td>0.0351563</td>
</tr>
<tr>
<td>8</td>
<td>0.0219727</td>
</tr>
<tr>
<td>9</td>
<td>0.0134277</td>
</tr>
<tr>
<td>10</td>
<td>0.00805664</td>
</tr>
</tbody>
</table>

Comparing these values to the corresponding values for the negative binomial...
distribution with $r = 45/13$ and $p = 15/28$, we see that the fit when $r = 3$ and $p = .5$
appears to be slightly better.

f. The desired probability is $\Pr[N > 2] = 1 - \Pr[N = 0] - \Pr[N = 1] - \Pr[N = 2]$. 
According to the model constructed in part d, i.e.,
$N \sim \text{NegativeBinomial}(45/13, 15/28),
\Pr[N = 0] \approx .115264, \Pr[N = 1] \approx .185245, \Pr[N = 2] \approx .191861.$

Hence
$\Pr[N > 2] = 1 - .115264 - .185245 - .191861 = .50763.$

If a negative binomial model with parameters $r = 3$ and $p = .5$ is used instead (see the
answer to part e), then
$\Pr[N = n] = \binom{n + 2}{n} \left(\frac{1}{2}\right)^{n+3} \text{ for } n = 0, 1, 2, ...$

and the required probability is
$\Pr[N > 2] = 1 - \left(\frac{1}{2}\right)^3 - 3 \left(\frac{1}{2}\right)^4 - 6 \left(\frac{1}{2}\right)^5 = \frac{1}{2}.$

12. In each part of this question, one must first recognize the given moment generating
function as the moment generating function of a particular special distribution. Then
using the uniqueness property of the moment generating function and properties of the
identified special distribution, it is straightforward to determine $E[X], \text{Var}(X), \Pr[X > 1]$
and $\Pr[X = 2]$.

a. $X \sim \text{Binomial}(10, .25)$. Hence
$E[X] = (10)(.25) = 2.5,$
$\text{Var} (X) = (10)(.25)(.75) = 1.875,$
$\Pr[X > 1] = 1 - \Pr[X = 0] - \Pr[X = 1] =
1 - \binom{10}{0}(.75)^{10} - \binom{10}{1}(.25)(.75)^9 = 1 - (3.25)(.75)^9 \approx .75597477,$
Pr[X = 2] = \binom{10}{2} (.25)^2 (.75)^8 \approx .28156757.

b. \( X \sim \text{NegativeBinomial}(3, .25) \). Hence
\[
E[X] = \frac{3 (.75)}{.25} = 9,
\]
\[
\text{Var}(X) = \frac{3 (.75)}{(2.25)^2} = 36,
\]
\[
\Pr[X > 1] = 1 - \Pr[X = 0] - \Pr[X = 1] =
1 - \left(\frac{2}{2}\right) (.25)^3 - \left(\frac{3}{2}\right) (.25)^3 (.75) = 1 - (3.25) (.25)^3 = .94921875,
\]
\[
\Pr[X = 2] = \left(\frac{4}{2}\right) (.25)^3 (.75)^2 = \left(\frac{3}{8}\right).
\]

c. \( X \sim \text{Poisson}(2) \). Hence
\[
E[X] = 2,
\]
\[
\text{Var}(X) = 2,
\]
\[
\Pr[X > 1] = 1 - \Pr[X = 0] - \Pr[X = 1] = 1 - \frac{2^0 e^{-2}}{0!} - \frac{2^1 e^{-2}}{1!} = 1 - 3 e^{-2} \approx .59399415,
\]
\[
\Pr[X = 2] = \frac{2^2 e^{-2}}{2!} = 2 e^{-2} \approx .27067057.
\]

16. Let \( X \) be the number of delinquencies in the first month and let \( \Lambda \) be the expected number of delinquencies per month. Then from the given information, a reasonable model for \( X \) is \( (X \mid \Lambda = \lambda) \sim \text{Poisson}(\lambda) \) where \( \Lambda \) has the density
\[
f_\Lambda(\lambda) = (0.02)^2 \lambda e^{-0.02\lambda} \text{ for } \lambda > 0.
\]

By the law of total probability,
\[
\Pr[X = x] = \int_0^\infty \frac{\lambda^x e^{-\lambda}}{x!} f_\lambda[\lambda] \, d\lambda = \int_0^\infty (0.02)^2 \frac{\lambda^{x+1} e^{-1.02 \lambda}}{x!} \, d\lambda.
\]

Using integration by parts we have
\[
\int_0^\infty \lambda^{x+1} e^{-1.02 \lambda} \, d\lambda = \\
\lambda^{x+1} e^{-1.02 \lambda} \bigg|_0^\infty - \int_0^\infty (x + 1) \lambda^x \frac{e^{-1.02 \lambda}}{1.02} \, d\lambda = \frac{x + 1}{1.02} \int_0^\infty \lambda^x e^{-1.02 \lambda} \, d\lambda.
\]

By repeated application of this formula we obtain
\[
\int_0^\infty \lambda^{x+1} e^{-1.02 \lambda} \, d\lambda = \frac{(x + 1)!}{(1.02)^{x+1}} \int_0^\infty e^{-1.02 \lambda} \, d\lambda = \frac{(x + 1)!}{(1.02)^{x+2}}.
\]

Hence
\[
\Pr[X = x] = \int_0^\infty (0.02)^2 \frac{\lambda^{x+1} e^{-1.02 \lambda}}{x!} \, d\lambda = \\
\left(\frac{(0.02)^2}{x!}\right) \frac{(x + 1)!}{(1.02)^{x+2}} = (x + 1) \left(\frac{0.02}{1.02}\right)^2 \left(\frac{1}{1.02}\right)^x = (x + 1) \left(\frac{1}{51}\right)^2 \left(\frac{50}{51}\right)^x.
\]

Therefore, the probability that there are fewer than 50 delinquencies in the first month is
\[
\Pr[X < 50] = \sum_{x=0}^{49} (x + 1) \left(\frac{1}{51}\right)^2 \left(\frac{50}{51}\right)^x.
\]

We could evaluate this sum using a computer. However there is a more elegant way, which we now describe.

Consider the quantity defined by
\[
g[r] = \sum_{j=0}^n r^j.
\]

Note that from the formula for the sum of a finite geometric series,
\[ g[r] = \frac{1 - r^{n+1}}{1 - r}. \]

Hence
\[ g'[r] = \sum_{j=1}^{n} j r^{j-1} \]

and also
\[ g'[r] = \frac{-(n+1) r^n (1 - r) - (1 - r^{n+1})(-1)}{(1 - r)^2} = -\frac{r^n (n + 1)}{(1 - r)} + \frac{1 - r^{n+1}}{(1 - r)^2}. \]

Equating these two expressions for \( g'[r] \) we obtain
\[ \sum_{j=1}^{n} j r^{j-1} = \frac{1 - r^{n+1}}{(1 - r)^2} - \frac{r^n (n + 1)}{1 - r}. \]

Putting \( r = \frac{50}{51} \) and \( n = 50 \) into this equation we have
\[ \sum_{j=1}^{50} j \left( \frac{50}{51} \right)^{j-1} = 1 - \left( \frac{50}{51} \right)^{51} - \frac{(51) \left( \frac{50}{51} \right)^{50}}{(1 \ 51)} \]
that is,
\[ \sum_{j=1}^{50} j \left( \frac{50}{51} \right)^{j-1} = 51^2 \left( 1 - \left( \frac{50}{51} \right)^{51} - \left( \frac{50}{51} \right)^{50} \right). \]

By changing the index of summation we also have
\[ \sum_{j=1}^{50} j \left( \frac{50}{51} \right)^{j-1} = \sum_{x=0}^{49} (x + 1) \left( \frac{50}{51} \right)^x. \]

Consequently,
\[
\sum_{x=0}^{49} (x + 1) \left( \frac{50}{51} \right)^x = 51^2 \left( 1 - \frac{50}{51} \right)^{51} - \left( \frac{50}{51} \right)^{50}.
\]

It follows that the desired probability is
\[
\Pr[X < 50] = \sum_{x=0}^{49} (x + 1) \left( \frac{1}{51} \right)^2 \left( \frac{50}{51} \right)^x = 1 - \left( \frac{50}{51} \right)^{51} - \left( \frac{50}{51} \right)^{50} \approx .26422909.
\]

Comment: The alert reader may have noticed that \( \Lambda \sim \text{Gamma}(2, 0.02) \). Using the fact that
\( (X \mid \Lambda = \lambda) \sim \text{Poisson} (\lambda) \) and \( \Lambda \sim \text{Gamma} (r, \alpha) \Rightarrow \)
\( X \sim \text{NegativeBinomial} (r, \alpha / (\alpha + 1)) \),

it then follows that \( X \sim \text{NegativeBinomial} \left( 2, \frac{1}{51} \right) \). Hence the probability mass function of \( X \) is
\[
p_X(x) = \binom{x + 1}{1} \left( \frac{1}{51} \right)^2 \left( \frac{50}{51} \right)^x \text{ for } x = 0, 1, \ldots
\]

which is precisely the formula derived earlier.

20. Let \( X_1 \) be the number of claims submitted in a month for a group known to be a low utilizer, let \( X_2 \) be the number of claims submitted in a month for a group known to be a high utilizer, and let \( N \) be the number of claims submitted in a month for the group under consideration. Let \( C \) be defined as follows:
\[
C = \begin{cases} 
1 & \text{if given group is low utilizer,} \\
2 & \text{if given group is high utilizer.}
\end{cases}
\]

Then from the given information, \( \Pr[C = 1] = .70, \Pr[C = 2] = .30, X_1 \sim \text{Poisson}(20) \) and \( X_2 \sim \text{Poisson}(50) \). We are interested in the probability that the given group submits fewer than 20 claims in the first month. By the law of total probability, this is
\[
\Pr[N < 20] = \Pr[N < 20 \mid C = 1] \Pr[C = 1] + \Pr[N < 20 \mid C = 2] \Pr[C = 2] = \\
\Pr[X_1 < 20] \Pr[C = 1] + \Pr[X_2 < 20] \Pr[C = 2] = \\
(.70) \Pr[X_1 < 20] + (.30) \Pr[X_2 < 20].
\]
Since $X_1 \sim \text{Poisson}(20)$ and $X_2 \sim \text{Poisson}(50)$, we have
\[
\Pr[X_1 < 20] = \sum_{x=0}^{19} \frac{20^x e^{-20}}{x!}
\]
and
\[
\Pr[X_2 < 20] = \sum_{x=0}^{19} \frac{50^x e^{-50}}{x!}.
\]
These sums can be evaluated numerically using Mathematica or similar computer software. We find that
\[
\sum_{x=0}^{19} \frac{20^x e^{-20}}{x!} \approx 0.470257
\]
and
\[
\sum_{x=0}^{19} \frac{50^x e^{-50}}{x!} \approx 4.79136 \times 10^{-7}.
\]
Consequently, the desired probability is
\[
\Pr[N < 20] = (.70) \Pr[X_1 < 20] + (.30) \Pr[X_2 < 20] \approx (.70) (.470257) + (.30) (4.79136 \times 10^{-7}) \approx .32918 = 33\%.
\]

24. Let $X$ be the number of policyholders that file at least one claim during the first year and let $P$ be the probability that a given policyholder files at least one claim. Then assuming claims are independent, $(X \mid P = p) \sim \text{Binomial}(100, p)$. We are given that the density of $P$ is
\[
f_P(p) = 3(1 - p)^2, \quad 0 < p < 1.
\]
Hence by Bayes' theorem,
\[ f_{P|X=x}[p] = \frac{f_{X|P=x}[x] f_P[p]}{f_X[x]} = \frac{\binom{100}{x} p^x (1 - p)^{100-x} \cdot 3 (1 - p)^2}{f_X[x]} \propto p^x (1 - p)^{102-x} \]

where the terms not containing \( p \) have been omitted from the proportionality. The proportionality constant can be determined in principle from the condition

\[ \int_0^1 f_{P|X=x}[p] \, dp = 1. \]

Since there are no claims filed in the first year, the event of interest is \( X = 0 \). In this case, the proportionality constant is relatively simple to determine. From the formula for \( f_{P|X=x} \) we have

\[ f_{P|X=0}[p] = \frac{(1 - p)^{102}}{\int_0^1 (1 - p)^{102} \, dp} = \frac{(1 - p)^{102}}{-(1 - p)^{103}} \bigg|_0^1 = 103 (1 - p)^{102}. \]

Therefore the desired probability is

\[ \Pr[P > .10 \mid X = 0] = \int_{.10}^1 f_{P|X=0}[p] \, dp = \int_{.10}^1 103 (1 - p)^{102} \, dp = -(1 - p)^{103} \bigg|_{.10}^1 = (.90)^{103} \approx 1.9363 \times 10^{-5}. \]

Note that \((.90)^{103}\) is extremely small. Hence it is very unlikely that \( P \) will exceed 10% if no claims are observed during the first year.
Chapter Six Solutions

1. Recall that the probability density function of the gamma distribution is given by

\[ f_X(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma[r]}, \quad x > 0 \]

and the mean, variance, and skewness are

\[ \mu_X = \frac{r}{\lambda}, \]
\[ \sigma_X^2 = \frac{r}{\lambda^2}, \]
\[ \gamma_X = \frac{2}{\lambda^{3/2}}. \]

From these formulas, one can give the qualitative descriptions requested in part a.

a. The distribution is positively skewed for all \( r \) and \( \lambda \). The distributions with the greatest skew are the ones for which \( r \) is small. As \( r \) increases with \( \lambda \) held fixed, the distribution of probability moves to the right, becomes more spread out, and becomes more symmetric. On the other hand, as \( \lambda \) increases with \( r \) held fixed, the distribution of probability moves to the left and becomes less spread out, but the skewness does not change. These characteristics are clear from the formulas for the mean, variance, and skewness.

Now consider what happens when \( r \to 0 \) or \( \lambda \to 0 \). As \( r \to 0 \) with \( \lambda \) held fixed, the distribution becomes more concentrated around \( x = 0 \). In the limiting case \( r = 0 \), the distribution reduces to a point mass at 0. Indeed, for any \( x \neq 0 \) and any \( \lambda \geq 0 \),
\[ f_X(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma[r]} \rightarrow 0 \text{ as } r \to 0 \]

and for \( r < 1, f_X(x) \to \infty \text{ as } x \to 0^+ \). As \( \lambda \to 0 \) with \( r \) held fixed, the distribution of probability moves to the right. From the interpretation of a gamma random variable as a waiting time, the limiting distribution in the case \( \lambda = 0 \) could be considered a point mass at infinity. However, strictly speaking, the distribution is not defined at \( \lambda = 0 \).

b. Different values of the parameter \( r \) result in density curves of a different shape. For example, when \( r < 1 \) the density function is unbounded and becomes infinite at \( x = 0 \), when \( r = 1 \) the density function is strictly decreasing with a maximum at \( x = 0 \), and when \( r > 1 \) the density function increases and then decreases and attains its maximum at a point \( x > 0 \). (See figures 6.1 and 6.2b in the textbook.) From these descriptions, it is clear that the "shape" of the graph is not the same for all values of \( r \). Hence it is appropriate to consider \( r \) to be a "shape parameter".

The easiest way to see why \( \lambda \) can be considered a scale parameter is to plot a few graphs of gamma densities with different \( \lambda \) values and the same \( r \) value. Consider for example plots of the densities for Gamma(2, 1) and Gamma(2, 2). This can be done using Mathematica or similar computer software.
Without referring to the axis scale, the two graphs appear to be the same. However, if the graphs are plotted using the same scale, we see that they are in fact quite different:

This suggests that changes in $\lambda$ amount to changes in scale.

We can also see this by looking directly at the formula for the density function:
\[ f_X(x) = \frac{\lambda (\lambda x)^{r-1} e^{-\lambda x}}{\Gamma(r)}. \]

Consider the change of scale given by the substitution \( u = \lambda x \). Since probability densities measure probability per unit length, they are affected by the choice of unit length. This means that the substitution \( u = \lambda x \) will also affect the scale of the graph in the vertical direction. Indeed, the change in vertical scale will be given by \( v = y/\lambda \), where \((x, y)\) represents two-dimensional coordinates before the change of scale and \((u, v)\) represents two-dimensional coordinates after the change of scale. Hence applying this substitution to the formula for \( f_X \) we obtain

\[ f[u] = \frac{u^{r-1} e^{-u}}{\Gamma(r)}, \]

where \( f \) is the density function in the new coordinate system (i.e., after the change of scale). Note that the value of \( r \) has not changed. This shows that changes in \( \lambda \) are related to changes in scale. For this reason, it is appropriate to consider \( \lambda \) a "scale parameter".

c. The derivative of the density function is

\[ f'_X(x) = \frac{\lambda^r}{\Gamma(r)} \left\{ (r-1)x^{r-2} e^{-\lambda x} + x^{r-1} (-\lambda) e^{-\lambda x} \right\} = \frac{\lambda^r x^{r-2} e^{-\lambda x} \Gamma(r)}{\Gamma(r)} [1 - x]. \]

Hence \( f_X \) is increasing for \( r - 1 - \lambda x > 0 \), i.e., for \( x < (r - 1)/\lambda \) and \( f_X \) is decreasing for \( r - 1 - \lambda x < 0 \), i.e., for \( x > (r - 1)/\lambda \). Consequently, \( f_X \) is decreasing for all \( x > 0 \) if and only if \( (r - 1)/\lambda \leq 0 \), i.e., if and only if \( r \leq 1 \).

5. a. Exponential(100) (time measured in hours) or Exponential\(\left(\frac{5}{3}\right)\) (time measured in minutes).

b. Beta(6, 96). If a sample of size \( n \) is drawn from a population with replacement and the sample contains \( x \) defectives, then the fraction of defectives in the entire population is Beta\(\left(x + 1, n - x + 1\right)\) (see section 6.3.3). Note that the sampling was probably done with replacement, but if the population is large relative to the sample size, the difference between sampling with and without replacement is small.
c. Lognormal($\mu$, $\sigma$). There is insufficient information in the statement of the question to specify $\mu$ and $\sigma$. Note that this is an approximate model. The precise value of the security after one year is $X_1 X_2 \cdots X_n$ where $X_j$ is the accumulation of a $\$1$ investment in the $j$-th day and $n$ is the number of trading days. By assumption, the $X_j$ are independent, identically distributed, and positive random variables. Since $n$ is reasonably large, the multiplicative form of the central limit theorem applies. Hence $X_1 X_2 \cdots X_n \approx \text{Lognormal}(\mu, \sigma)$ where $\mu, \sigma$ are the mean and standard deviation of $\log[X_1 X_2 \cdots X_n]$.

d. Exponential(3), time measured in months. Since the failure rate is constant, there is no aging. Consequently, the distribution is exponential.

9. Let $T_l, T_r$ be the total service times for the left and right machines respectively and let $T_l^*, T_r^*$ be the corresponding remaining service times. Let $T$ be the waiting time until the first machine becomes available when both machines are in use. Suppose that $T_l, T_r, T_l^*, T_r^*$, and $T$ are all measured in seconds. Then $T = \min(T_l^*, T_r^*)$.

We are not explicitly told what models to use for $T_l$ and $T_r$. In the interest of simplicity, let's assume that both $T_l$ and $T_r$ have exponential distributions. Since the exponential distribution has the memoryless property it follows from this assumption that $T_l^*$ and $T_r^*$ are exponentially distributed with $T_l^* \sim T_l$ and $T_r^* \sim T_r$. Note that in this context the memoryless property means that knowledge of the time that a machine has already spent servicing a customer has no effect on the distribution of the remaining service time. This is not an unreasonable assumption to make in this context as anyone who has stood behind a customer performing multiple transactions can attest! Since the average service times are 30 seconds and 20 seconds for the left and right machines respectively, it follows that $T_l \sim \text{Exponential}\left(\frac{1}{30}\right)$, $T_r \sim \text{Exponential}\left(\frac{1}{20}\right)$ and also $T_l^* \sim \text{Exponential}\left(\frac{1}{30}\right), T_r^* \sim \text{Exponential}\left(\frac{1}{20}\right)$.

In section 6.1.1, it was shown that if $T_1 \sim \text{Exponential}(\lambda_1), T_2 \sim \text{Exponential}(\lambda_2)$ and $T_1$, $T_2$ are independent then $\min(T_1, T_2) \sim \text{Exponential}(\lambda_1 + \lambda_2)$. Since $T = \min(T_l^*, T_r^*)$, it
follows that $T$ has an exponential distribution with parameter $\lambda = \frac{1}{30} + \frac{1}{20} = \frac{1}{12}$, i.e., $T \sim \text{Exponential}(\frac{1}{12})$. This fact will be used to answer parts a through e.

a. Since $T \sim \text{Exponential}(\frac{1}{12})$, we have $E[T] = 12$. Hence the person at the front of the line should expect to wait 12 seconds.

b. The desired probability is
\[
\Pr[T > 15] = e^{-15/12} = e^{-5/4} \approx .2865.
\]

c. From part a, the expected waiting time for a person at the front of the line is 12 seconds. Hence we should expect the line to move every 12 seconds. It follows that the person who is currently third in line should expect to wait 36 seconds. This result can also be derived more formally using the approach outlined in part d.

d. Let $T_j$ be the time that the $j$-th person in line must wait for service after making it to the front of the line and let $T_j^*$ be the amount of time that the $j$-th person must wait in total. Then
\[
T_j^* = T_1 + T_2 + \cdots + T_j.
\]

From earlier comments, $T_j \sim \text{Exponential}(\frac{1}{12})$ for all $j$. Moreover, since machine service times are independent and exponentially distributed (i.e., "memoryless"), the $T_j$ are also independent. Hence from section 6.1.2 we have
\[
T_j^* \sim \text{Gamma}\left(j, \frac{1}{12}\right).
\]

Consequently the expected waiting time for the person currently third in line is
\[
E[T_3^*] = \frac{3}{1/12} = 36 \text{ seconds}
\]

(the answer obtained in part c) and the probability that this person must wait more than 30 seconds is
\[
\Pr[T_3^* > 30] = \sum_{n=0}^{\infty} \frac{\left(\frac{1}{12}\right)^n (30)^n}{n!} e^{-30/12} = e^{-30/12} \left\{ 1 + \frac{5}{2} + \left(\frac{1}{2}\right)^2 \left(\frac{5}{2}\right)^2 \right\} = \frac{53}{8} e^{-5/2} \approx .5438.
\]
e. To answer the question of this part, we need only consider the machine on the left. The desired probability is
\[ \Pr[T_l > 60] = e^{-60/30} = e^{-2} \approx .1353. \]

14. Let \( X_j \) be the dollar increase on the \( j \)-th trading day. By assumption the \( X_j \) are independent and identically distributed with probability distribution given by
\[
X_j = \begin{cases} 
2 & \text{with probability } .50, \\
-1 & \text{with probability } .50.
\end{cases}
\]

Since the current price of the stock is $100, its price \( n \) trading days hence is
\[ S_n = 100 + X_1 + X_2 + \cdots + X_n. \]
We are interested in determining \( \Pr[S_{50} > 145] \).

Let \( I_j \) be an indicator of a price increase on the \( j \)-th trading day. Then
\[ I_j \sim \text{Binomial}[1, .50] \]
and
\[ X_j = 3I_j - 1. \]

Hence
\[ S_n = 100 + 3(I_1 + \cdots + I_n) - n = 100 - n + 3Y \]
where \( Y = I_1 + \cdots + I_n \sim \text{Binomial}[n, .50] \). Consequently,
\[
\Pr[S_n > 145] = \Pr[100 - n + 3Y > 145] = \Pr\left[Y > 15 + \frac{n}{3}\right] = \sum_{k=k^*}^n \binom{n}{k}(.50)^n
\]
where \( k^* = 15 + \left\lfloor \frac{n}{3} \right\rfloor + 1 = 16 + \left\lfloor \frac{2}{3} \right\rfloor \). Here \([x]\) denotes the integer part of \( x \), i.e., the greatest integer less than or equal to \( x \). For \( n = 50 \) we have \( k^* = 32 \). Hence
\[
\Pr[S_{50} > 145] = \sum_{k=32}^{50} \binom{50}{k}(.50)^{50}.
\]

The latter sum can be determined numerically using Mathematica or similar computer software. When we do this we find that
An alternative approach to determining $\Pr[S_{50} > 145]$ is to use a normal approximation for $S_n$. From the definition of $X_j$ we have

$$E[X_j] = (2)(.50) + (-1)(.50) = 0.50,$$

$$\text{Var}(X_j) = E[X_j^2] - E[X_j]^2 = [(2)^2(.50) + (-1)^2(.50)] - (0.50)^2 = 2.25.$$ 

Hence

$$E[S_n] = 100 + \sum_{j=1}^{n} E[X_j] = 100 + \frac{n}{2},$$

$$\text{Var}(S_n) = \sum_{j=1}^{n} \text{Var}(X_j) = 2.25n,$$

where the formula for the variance follows from the independence of the $X_j$. It follows that for $n$ sufficiently large,

$$S_n \approx \text{Normal} \left( 100 + \frac{n}{2}, 1.5\sqrt{n} \right).$$

Using this approximation and correcting for continuity we have

$$\Pr[S_{50} > 145] =$$

$$\Pr[S_{50} \geq 145.5] = \Pr \left[ \frac{S_{50} - 125}{1.5\sqrt{50}} \geq \frac{145.5 - 125}{1.5\sqrt{50}} \right] \approx \Pr[Z \geq 1.9328] = 1 - \Phi[1.9328]$$

where $Z \sim \text{Normal}(0, 1)$ and $\Phi$ is the distribution function of $Z$. From the table in Appendix E and using linear interpolation we have

$$\Phi[1.9328] \approx (.72)\Phi[1.93] + (.28)\Phi[1.94] = (.72)(.9732) + (.28)(.9738) = .973368.$$ 

Consequently,

$$\Pr[S_{50} > 145] \approx 1 - \Phi[1.9328] \approx 1 - .973368 \approx .02663.$$
Note that a correction for continuity was appropriate in this case because the values of $S_n$ are all integers. If the daily price movements (i.e., the values of $X_j$) had not been whole dollar amounts then it would not have been appropriate to correct for continuity in the approximation of $S_n$.

It is instructive to compare the value calculated for $\Pr[S_{50} > 145]$ under a normal approximation for $S_{50}$ to the exact value determined earlier. Recall that the exact value of $\Pr[S_{50} > 145]$ was determined to be $.0324543$ and the value of $\Pr[S_{50} > 145]$ under a normal approximation was determined to be $.02663$. To the nearest percentage point, both values are about 3%. If this degree of precision in the answer is sufficient then it is reasonable to use the normal approximation. However, if greater precision is required then the desired probability must be calculated exactly.

18. Let $P$ be the fraction of the company's policies for which a claim is filed. From section 6.3.3, we know that if a sample of size $n$ is drawn with replacement from a population whose members are one of two types and the sample contains $x$ items of a particular type, then the fraction of items of this type in the entire population has the distribution $\text{Beta}(x + 1, n - x + 1)$. In this exercise, the sample size is $n = 100$ and $x = 5$. We are not told whether the sampling is done with or without replacement. However, since the number of policies is likely to be very large relative to the size of the sample (which we know to be 100), we may assume that the sampling is done with replacement. Hence an appropriate model for $P$ is $P \sim \text{Beta}(6, 96)$ and the desired probability is

$$\Pr[P > .10] = 1 - \Pr[P \leq .10] = 1 - \frac{1}{B[6, 96]} \int_0^{10} x^5 (1 - x)^{95} \, dx.$$ 

We could calculate the latter integral using successive applications of integration by parts; however this would require five iterations! Alternatively, we can use the formula $\Pr[\text{Beta}(r, s) \leq x] = \Pr[\text{Binomial}(r + s - 1, x) \geq r]$.

Using this result we have

$$\Pr[P \leq .10] =$$

$$\Pr[\text{Beta}(6, 96) \leq .10] = \Pr[\text{Binomial}(101, .10) \geq 6] = 1 - \sum_{x=0}^{5} \binom{101}{x} (.10)^x (.90)^{101-x}.$$
Hence
\[ \Pr[P > .10] = \sum_{x=0}^{5} \binom{101}{x} (0.1)^x (0.9)^{101-x}. \]

The latter sum can be evaluated numerically using Mathematica or similar computer software. When we do this we find that the desired probability is
\[ \Pr[P > .10] \approx 0.0541903. \]

22. Let \( X \) be the number of heads obtained in 1000 tosses of the selected coin and let \( I \) be an indicator of the fairness of the coin, i.e.,
\[ I = \begin{cases} 
1 & \text{if selected coin is fair}, \\
0 & \text{if selected coin is biased}. 
\end{cases} \]

Since the gambler concludes that the coin is biased if \( X \geq 525 \) and concludes that it is fair otherwise, the probability that the gambler reaches a false conclusion is, by the law of total probability,
\[ \Pr[X \geq 525 \mid I = 1] \Pr[I = 1] + \Pr[X < 525 \mid I = 0] \Pr[I = 0]. \]

Consider first the quantity \( \Pr[X \geq 525 \mid I = 1] \). This is the probability of reaching a false conclusion when the coin being tossed is known to be fair. Note that the distribution of \( X \mid I = 1 \) is binomial with parameters \( n = 1000 \) and \( p = .50 \). (The total number of tosses is 1000 and since the coin is fair, the probability of heads on a single toss of the coin is .50.) Hence
\[ \Pr[X \geq 525 \mid I = 1] = \sum_{x=525}^{1000} \binom{1000}{x} (0.5)^{1000}. \]

We can evaluate this sum using Mathematica or similar computer software. When we do this we find that
\[ \Pr[X \geq 525 \mid I = 1] \approx 0.0606071. \]

Alternatively, we can evaluate the probability using a normal approximation with continuity correction. When we do this, we obtain
Pr\[X \geq 525 \mid I = 1\] = Pr\[X \geq 524.5 \mid I = 1\] = 
\[
\frac{Pr\left[\frac{X - (1000)(.50)}{\sqrt{(1000)(.50)(.50)}} \geq \frac{524.5 - (1000)(.50)}{\sqrt{(1000)(.50)(.50)}} \mid I = 1\right]}{Pr[Z \geq 1.5495]}.
\]

From Appendix E of the textbook and using linear interpolation as appropriate we have 
\[\Phi[1.5495] \approx (.05) \Phi[1.54] + (.95) \Phi[1.55] = (.05)(.9382) + (.95)(.9394) = .93934.\]

Hence
\[Pr[X \geq 525 \mid I = 1] \approx Pr[Z \geq 1.5495] = 1 - \Phi[1.5495] = .06066,\]

which is close to the value .0606071 calculated directly.

Now consider the quantity Pr\[X < 525 \mid I = 0\]. This is the probability of reaching a false conclusion when the coin being tossed is known to be the biased one. Since the probability of heads for the coin known to be biased is 55% by assumption, the distribution of \(X \mid I = 0\) is binomial with parameters \(n = 1000\) and \(p = .55\). Hence
\[Pr[X < 525 \mid I = 0] = \sum_{x=0}^{524} \binom{1000}{x} (.55)^x (.45)^{1000-x}.\]

Once again, we can evaluate this probability using Mathematica or similar computer software. When we do this we find that
\[Pr[X < 525 \mid I = 0] \approx .0526817.\]

Alternatively, we can use a normal approximation with continuity correction:
\[Pr[X < 525 \mid I = 0] = Pr[X \leq 524.5 \mid I = 0] = \frac{Pr\left[\frac{X - (1000)(.55)}{\sqrt{(1000)(.55)(.45)}} \leq \frac{524.5 - (1000)(.55)}{\sqrt{(1000)(.55)(.45)}} \mid I = 0\right]}{\Phi[-1.6209]} = 1 - \Phi[1.6209].\]

From Appendix E of the textbook and using linear interpolation as appropriate we have 
\[\Phi[1.6209] \approx (.91) \Phi[1.62] + (.09) \Phi[1.63] \approx (.91)(.9474) + (.09)(.9484) = .94749.\]

Hence
\[Pr[X < 525 \mid I = 0] \approx 1 - \Phi[1.6209] \approx .05251,\]
which is close to the value .0526817 calculated directly.

The only remaining probabilities to consider are \( \Pr[I = 0] \) and \( \Pr[I = 1] \). Since the gambler has one coin of each type and selects the coin to flip at random, we must have

\[
\Pr[I = 0] = \frac{1}{2} \quad \text{and} \quad \Pr[I = 1] = \frac{1}{2}.
\]

Putting this together, we find that the probability of reaching a false conclusion is

\[
\Pr[X \geq 525 \mid I = 1] \Pr[I = 1] + \Pr[X < 525 \mid I = 0] \Pr[I = 0] =
\]

\[
(.0606071)\left(\frac{1}{2}\right) + (.0526817)\left(\frac{1}{2}\right) = .0566444.
\]

Note that we have used the numerical values computed directly from the binomial sums by Mathematica when determining the final answer. However, the answer obtained using the normal approximation is similar.

30. Let \( X \) be the insurer's payment in dollars for a randomly selected policy and let \( I \) be an indicator of a claim for this policy. Then according to the assumptions,

\[
I = \begin{cases} 
1 & \text{with probability } .25, \\
0 & \text{with probability } .75,
\end{cases}
\]

and

\( (X \mid I = 1) \sim \text{Pareto}(3, 100) \).

Hence

\[
S_{X|I=1}[x] = \left(\frac{100}{100 + x}\right)^3 \quad \text{for } x > 0.
\]

a. The desired probability is \( \Pr[X > 50] \). By the law of total probability we have

\[
\Pr[X > 50] = \Pr[X > 50 \mid I = 1] \Pr[I = 1] + \Pr[X > 50 \mid I = 0] \Pr[I = 0].
\]

Clearly \( \Pr[X > 50 \mid I = 0] = 0 \) since no payment is made if no claim is submitted. From the formula for \( S_{X|I=1} \) stated earlier we also have
Consequently,

\[ \text{Pr}[X > 50 | I = 1] = S_{X|I=1}(50) = \left( \frac{100}{100 + 50} \right)^3 = \left( \frac{2}{3} \right)^3. \]

\[ \text{Pr}[X > 50] = \left( \frac{2}{3} \right)^3 (.25) + (0) (.75) = \frac{2}{27}. \]

b. The desired probability is \( \text{Pr}[X \leq 10] \). Arguing as in part a we have

\[ S_{X|I=1}(10) - \text{Pr}[I = 1] + 0 \cdot \text{Pr}[I = 0] = \left( \frac{100}{100 + 10} \right)^3 (.25) = \left( \frac{10}{11} \right)^3 \left( \frac{1}{4} \right) \approx .18782870. \]

Hence

\[ \text{Pr}[X \leq 10] = 1 - \text{Pr}[X > 10] \approx 1 - .18782870 = .81217130. \]

c. Applying the law of total probability as in parts a and b we have for \( x \geq 0 \),

\[ S_X[x] = \text{Pr}[X > x] = \text{Pr}[X > x | I = 1] \text{Pr}[I = 1] + \text{Pr}[X > x | I = 0] \text{Pr}[I = 0] = S_{X|I=1}(x) \cdot \text{Pr}[I = 1] + 0 \cdot \text{Pr}[I = 0] = \left( \frac{100}{100 + x} \right)^3 (.25). \]

Since the payment on a given policy cannot be negative we must also have

\[ S_X[x] = \text{Pr}[X > x] = 1 \quad \text{for } x < 0. \]

Consequently, the survival function of \( X \) is given by

\[ S_X[x] = \left( \frac{100}{100 + x} \right)^3 (.25) \quad \text{for } x \geq 0, \]

\[ S_X[x] = 1 \quad \text{for } x < 0. \]

It follows that the distribution function \( F_X \) is given by

\[ F_X[x] = 1 - \left( \frac{100}{100 + x} \right)^3 (.25) \quad \text{for } x \geq 0, \]

\[ F_X[x] = 0 \quad \text{for } x < 0. \]
Note that
\[
\Pr[X = 0] = \Pr[X = 0 | I = 1] \Pr[I = 1] + \Pr[X = 0 | I = 0] \Pr[I = 0] = 0 \cdot \Pr[I = 1] + 1 \cdot \Pr[I = 0] = .75.
\]

This also follows from the formula for \( F_X \). Hence we see that \( X \) has a mixed distribution with a probability mass of size .75 at \( x = 0 \) (representing the event that no claim is submitted) and a continuous distribution of probability on \( x > 0 \).

d. Recall that for nonnegative random variables \( X \) we have
\[
E[\frac{X}{X}] = \int_0^\infty \frac{100}{100 + x} \left( .25 \right) \left( 100 + x \right)^{-2} \, dx = \frac{\left( .25 \right) \left( 100 \right)^3}{2} = 12.5.
\]

To determine the variance of \( X \) we need to consider the density function \( f_X \). From part c, it follows that the continuous part of the distribution has density function
\[
f_X(x) = -S'_X(x) = \left( .25 \right) \left( 1 + \frac{x}{100} \right)^{-4} \text{ for } x > 0.
\]

The discrete part consists of a probability mass of size .75 at \( x = 0 \). Hence
\[
E[X^2] = 0^2 \cdot \Pr[X = 0] + \int_0^\infty x^2 f_X(x) \, dx = 0^2 \cdot (.75) + \left( .25 \right) \int_0^\infty x^2 \left( 1 + \frac{x}{100} \right)^{-4} \, dx.
\]

The integral \( \int_0^\infty x^2 \left( 1 + \frac{x}{100} \right)^{-4} \, dx \) can be determined by recursively applying integration by parts. Alternatively, one could recognize this integral as the second moment of a Pareto distribution with parameter \( s = 3, \beta = 100 \), and use the formula for the second moment stated in section 6.1.3. Taking the latter approach we have
\[
\int_0^\infty x^2 \left( 1 + \frac{x}{100} \right)^{-4} \, dx = \frac{100^2 \cdot 2}{(3-1)(3-2)} = 100^2.
\]
Consequently, the second moment of $X$ is

$$E[X^2] = (.25) \int_0^\infty x^2 \cdot \frac{3}{100} \left(1 + \frac{x}{100}\right)^{-4} \, dx = \left(\frac{1}{4}\right) 100^2 = 2500.$$

It follows that

$$\text{Var}(X) = E[X^2] - E[X]^2 = 2500 - (12.5)^2 = 2343.75.$$

32. In this exercise, students determine the probability mass function for a continuous mixture of binomial random variables with beta mixing weights. An application of this is considered in exercise 35.

Suppose that $(N \mid P = p) \sim \text{Binomial}(m, p)$ and $P \sim \text{Beta}(r, s)$. Then from sections 5.1 and 6.3.3 we have

$$p_{N\mid P = p}[n] = \binom{m}{n} p^n (1 - p)^{m-n} \quad \text{for } n = 0, 1, \ldots, m$$

and

$$f_P[p] = \frac{p^{r-1} (1 - p)^{s-1}}{B[r, s]} \quad \text{for } 0 < p < 1.$$

Hence by the law of total probability,

$$p_N[n] = \Pr[N = n] = \int_0^1 \Pr[N = n \mid P = p] \, f_P[p] \, dp =$$

$$= \int_0^1 p_{N\mid P = p}[n] \, f_P[p] \, dp = \int_0^1 \left\{ \binom{m}{n} p^n (1 - p)^{m-n} \right\} \left(\frac{p^{r-1} (1 - p)^{s-1}}{B[r, s]}\right) \, dp =$$

$$= \frac{\binom{m}{n}}{B[r, s]} \int_0^1 p^{n+r-1} (1 - p)^{m+s-n-1} \, dp.$$

From section 6.3.3 we have
\[
\int_0^1 p^{r-1} (1 - p)^{m-n+s-1} dp = B[r, m-n+s].
\]

From Appendix C of the textbook we have

\[
B[r, s] = \frac{\Gamma[r] \Gamma[s]}{\Gamma[r+s]}
\]

and

\[
B[n+r, m-n+s] = \frac{\Gamma[n+r] \Gamma[m-n+s]}{\Gamma[m+r+s]}. \tag{1}
\]

Consequently,

\[
p_N[n] = \binom{m}{n} \frac{\Gamma[r+s]}{\Gamma[r] \Gamma[s]} \frac{\Gamma[n+r] \Gamma[m-n+s]}{\Gamma[m+r+s]}
\]

as required.

For positive integer values of the argument, the gamma function reduces to a factorial. In particular,

\[\Gamma[x] = (x-1)! \quad \text{for } x = 1, 2, 3, \ldots.\]

Hence if \(r\) and \(s\) are positive integers, the formula for \(p_N\) becomes

\[
p_N[n] = \binom{m}{n} \frac{(r+s-1)! (n+r-1)! (m-n+s-1)!}{(r-1)! (s-1)! (m+r+s-1)!},\]

which is the formula that we were required to derive.

35. Let \(N\) be the number of companies that use their line of credit in the coming month.

From the information given in the question, an appropriate model for \(N\) is as follows: \(N \mid P = p \sim \text{Binomial}(10, p), P \sim \text{Beta}(2, 3)\). From exercise 32, the probability mass function for the unconditional distribution of \(N\) is given by
**Chapter Six Solutions**

\[ p_N[n] = \binom{10}{n} \frac{\Gamma[5] \Gamma[n + 2] \Gamma[13 - n]}{\Gamma[2] \Gamma[3] \Gamma[15]} = \binom{10}{n} \frac{4! (n + 1)! (12 - n)!}{1! \times 2! \times 14!} \]

for \( n = 0, 1, \ldots, 10 \). Hence the desired probability is

\[ \Pr[N \leq 2] = p_N[0] + p_N[1] + p_N[2] = \binom{10}{0} \frac{4! \times 1! \times 12!}{1! \times 2! \times 14!} + \binom{10}{1} \frac{4! \times 2! \times 11!}{1! \times 2! \times 14!} + \binom{10}{2} \frac{4! \times 3! \times 10!}{1! \times 2! \times 14!} \approx 0.31068931. \]

42. Let \( X \) be the size of a random claim. We are given that \( (X \mid \Lambda = \lambda) \sim \text{Exponential}(\lambda) \) and \( \Lambda \sim \text{Gamma}(2, 100) \). We are also given that the size of the first claim is \( x_1 = 200 \). Since \( \text{Exponential}(\lambda) \) is a special case of \( \text{Gamma}(r, \lambda) \) with \( r = 1 \), we can use the result of exercise 41 to determine the distribution of \( \Lambda \mid X = x_1 \). In exercise 41, it was shown that if \( (X \mid \Lambda = \lambda) \sim \text{Gamma}(r, \lambda) \) and \( \Lambda \sim \text{Gamma}(s, \beta) \) then \( (\Lambda \mid X = x) \sim \text{Gamma}(r + s, x + \beta) \). In this exercise, we have \( r = 1, s = 2, \beta = 100, \) and \( x = 200 \). Hence

\( (\Lambda \mid X = x_1) \sim \text{Gamma}(3, 300) \).

It follows that an appropriate model for \( \Lambda \) going forward is \( \text{Gamma}(3, 300) \).

46. a. Let \( X \) be the eye pressure measurement for a randomly selected person and let \( I \) be an indicator for glaucoma such that

\[ I = \begin{cases} 
1 & \text{if eye is diseased,} \\
0 & \text{if eye is healthy.} 
\end{cases} \]

We are given that \( (X \mid I = 1) \sim \text{Normal}(25, 1), (X \mid I = 0) \sim \text{Normal}(20, 1) \), and \( \Pr[I = 1] = .10 \). We are interested in determining \( \Pr[I = 1 \mid X = x] \). Using Bayes' theorem we have

\[ p_{I \mid X = i}[I] = \frac{f_{X \mid I = i}[x] p_I[i]}{f_X[x]} \quad \text{for} \quad i = 0, 1. \]

Since \( X \mid I = i \) has a normal distribution we have
The desired probability is

\[
\Pr[I = 1 \mid X = x] = \frac{p_{I|X=x}[1]}{p_{I|X=x}} = \frac{e^{-(x-25)^2/2}}{e^{-(x-25)^2/2} + 9 e^{-(x-20)^2/2}}.
\]

Simplifying this expression we have

\[
\Pr[I = 1 \mid X = x] = \frac{1}{1 + 9 e^{(225-10x)/2}}.
\]

b. We are interested in the values of \(x\) for which \(p_{I|X=x}[1] > \frac{1}{2}\). From part a,

\[
p_{I|X=x}[1] = \frac{1}{1 + 9 e^{(225-10x)/2}}.
\]

Hence

\[
p_{I|X=x}[1] > \frac{1}{2} \iff 1 + 9 e^{(225-10x)/2} < 2 \iff \\
9 e^{(225-10x)/2} < 1 - 9 \iff 2(25 - 10x) < \log \left[ \frac{1}{9} \right] \iff \\
x > - \left\{ \frac{225}{2} - \log \left[ \frac{1}{9} \right] \right\} \iff x > 22.5 + \frac{1}{5} \log[9] \approx 22.9394.
\]
Chapter Seven Solutions

4. Consider the transformation

\[ Y = \begin{cases} \sqrt{X} & \text{for } X \geq 0, \\ 0 & \text{for } X < 0. \end{cases} \]

If \( X \) assumes both positive and negative values, as is the case when \( X \sim \text{Normal}(0, 1) \), then this transformation is not one-to-one. In this situation, \( Y \) will have a probability mass at 0 equal to \( \Pr[X \leq 0] \) and a distribution of probability on the interval \( y > 0 \).

The best way to approach the problem of determining a formula for \( F_Y \) is to consider a graph of the transformation with the event \( Y \leq y \) highlighted:

From this graph we see that for \( y \geq 0 \),
\[ F_Y[y] = \Pr[Y \leq y] = \Pr[\sqrt{X} \leq y] = \Pr[X \leq y^2] = F_X[y^2] \]

and for \( y < 0, F_Y[y] = 0 \). Hence the distribution function of \( Y \) under the given transformation is in general

\[
F_Y[y] = \begin{cases} 
F_X[y^2] & \text{for } y \geq 0, \\
0 & \text{for } y < 0.
\end{cases}
\]

Suppose that \( X \sim \text{Normal}(0, 1) \). Then from section 6.2.1,

\[
F_X[x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt \quad \text{for } x \in \mathbb{R}.
\]

Hence under the given transformation,

\[
F_Y[y] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y^2} e^{-t^2/2} \, dt \quad \text{for } y \geq 0,
\]

\[
F_Y[y] = 0 \quad \text{for } y < 0.
\]

It follows that \( Y \) has a mixed distribution with a probability mass of size \( \frac{1}{2} \) at \( y = 0 \) and a continuous distribution of probability on \( y > 0 \) given by

\[
f_Y[y] = \sqrt{\frac{2}{\pi}} \, y \, e^{-y^2/2}.
\]

The latter expression is determined using the fundamental theorem of calculus and the Chain rule to calculate the derivative of \( F_Y \). In particular,

\[
\frac{d}{dy} \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y^2} e^{-t^2/2} \, dt \right) = 2 \, y \cdot \left( \frac{1}{\sqrt{2\pi}} \, e^{-t^2/2} \right) \bigg|_{t=y^2} = \sqrt{\frac{2}{\pi}} \, y \, e^{-y^2/2}.
\]

Therefore, the (generalized) density function for \( Y \) when \( X \sim \text{Normal}(0, 1) \) is given by
Comment: The probability mass at 0 was inadvertently omitted from the answer given in the "Answers to Selected Exercises" section of the textbook. The complete answer is the one given here.

6. When \( X \) is a continuous random variable the stated properties of the limited expected value function are straightforward to prove using the formulas

\[
\begin{align*}
E[X; m] &= \int_{-\infty}^{m} x f_X(x) \, dx + m S_X[m], \\
E[X; m] &= E[X] - \int_m^\infty (x - m) f_X(x) \, dx
\end{align*}
\]

derived in section 7.2 of the textbook. Hence to begin, let's assume that the distribution of \( X \) is continuous.

Under this assumption, the survival function of \( X \) is continuous. Hence from the formula

\[
E[X; m] = \int_{-\infty}^{m} x f_X(x) \, dx + m S_X[m]
\]

the quantity \( E[X; m] \) is continuous as a function of \( m \). Differentiating this formula with respect to \( m \) and using the fundamental theorem of calculus as appropriate we have

\[
\frac{d}{dm} E[X; m] = \frac{d}{dm} \int_{-\infty}^{m} x f_X(x) \, dx + \frac{d}{dm} m S_X[m] = m f_X[m] + S_X[m] + m S_X'[m] = m f_X[m] + S_X[m] + m (-f_X[m]) = S_X[m]
\]

at all points \( m \) for which \( S_X \) is differentiable. Consequently,

\[
\frac{d}{dm} E[X; m] \geq 0 \quad \text{for all } m
\]
(since survival functions are always nonnegative) and so \( E[X; m] \) is increasing (or more precisely nondecreasing) in \( m \). Differentiating the formula \( \frac{d}{dm} E[X; m] = S_X[m] \) just derived we obtain
\[
\frac{d^2}{dm^2} E[X; m] = S_X'[m] = -f_X[m].
\]
From this it follows that
\[
\frac{d^2}{dm^2} E[X; m] \leq 0 \quad \text{for all } m
\]
(since density functions are always nonnegative) and so \( E[X; m] \) is concave as a function of \( m \). Now
\[
E[X; m] \leq E[X],
\]
a fact that follows directly from the formula
\[
E[X; m] = E[X] - \int_m^\infty (x - m) f_X[x] \, dx.
\]
Hence
\[
\lim_{m \to \infty} E[X; m] \leq E[X].
\]
On the other hand,
\[
\lim_{m \to \infty} E[X; m] \geq E[X]
\]
which can be demonstrated by letting \( m \to \infty \) in the formula
\[
E[X; m] = \int_{-\infty}^m x f_X[x] \, dx + m S_X[m]
\]
noting that \( m S_X[m] \geq 0 \) and \( \int_{-\infty}^\infty x f_X[x] \, dx = E[X] \). Consequently,
\[
\lim_{m \to \infty} E[X; m] = E[X].
\]
Hence we have shown in the case where \( X \) has a continuous distribution that

i. \( E[X; m] \) is increasing, continuous, and concave as a function of \( m \).

ii. \( E[X; m] \to E[X] \) as \( m \to \infty \).

iii. \( \frac{d}{dm} E[X; m] = S_X[m] \).

These are the three properties of \( E[X; m] \) that we were required to demonstrate.

The demonstration of these three properties when \( X \) does not have a continuous distribution is more delicate. Essentially the same line of reasoning can be followed in the general case by interpreting the probability densities to be generalized densities and the integrals to be generalized integrals in the sense described in the appendix to chapter 4. However, the validity of some of the intermediate steps must be carefully established. This requires demonstrating that the fundamental theorem of calculus and the formula for integration by parts hold for generalized integrals.

When \( X \) is discrete, it is also possible to demonstrate the three properties of \( E[X; m] \) by determining an explicit formula for \( E[X; m] \). To illustrate the key steps with this approach, consider a discrete distribution with exactly three probability masses. Suppose that these probability masses are located at the points \( a_1, a_2, a_3 \) where \( a_1 < a_2 < a_3 \) and have respective sizes \( k_1, k_2, k_3 \). Then from the definition of limited expected value, we have

\[
E[X; m] = \begin{cases} 
  m & \text{for } m < a_1, \\
  k_1 a_1 + (k_2 + k_3) m & \text{for } a_1 \leq m < a_2, \\
  (k_1 a_1 + k_2 a_2) + k_3 m & \text{for } a_2 \leq m < a_3, \\
  E[X] & \text{for } m \geq a_3.
\end{cases}
\]

It follows from this that \( E[X; m] \), when considered as a function of \( m \), is piecewise linear with slopes that decrease from 1 to 0. Moreover, the slope of \( E[X; m] \) at \( x = m \) is the sum of the probability masses to the right of \( m \), i.e., \( \frac{d}{dm} E[X; m] = S_X[m] \), and the ultimate value of \( E[X; m] \) is \( E[X] \).

A graph of \( E[X; m] \) illustrating these properties can be created using Mathematica or similar computer software.
From this graph, it is clear that when $X$ is a discrete random variable with three probability masses, the function $E[X; m]$ has the properties stated in the question. The demonstration of these properties when $X$ has more than three probability masses is similar.

It is worth noting that although the graph of $E[X; m]$ has some of the characteristics of the graph of a distribution function (nondecreasing, limiting value), $E[X; m]$ is not a distribution function. Moreover, the graph of $E[X; m]$ is generally quite different from the graph of the corresponding $F_X$. We can illustrate this concretely by creating the graph of the distribution function that corresponds to the graph of $E[X; m]$ just given.
Note that the sizes of the respective jumps at \(a_1, a_2, a_3\) are \(k_1, k_2, k_3\).

This completes the required demonstration.

10. From section 7.4, the actuarial present value of $1 payable at uncertain future time \(T\) is 
\[
E[e^{-rT}] \quad \text{where} \quad r \quad \text{is the constant per annum continuously compounded discount rate.}
\]
If \(T_1, \ldots, T_n\) are the future lifetimes of the individuals insured under a joint life policy then the time of payment of the benefit on this policy is \(T = \min(T_1, \ldots, T_n)\) (see exercise 9 for the definition of a joint life policy). From the solution to exercise 9, it follows that if the \(T_j\) are independent and exponentially distributed with \(T_j \sim \text{Exponential}(\lambda_j)\) then 
\(T \sim \text{Exponential}(\lambda_1 + \cdots + \lambda_n)\).

Now the actuarial present value of $1 when the future lifetime random variable \(T\) has an exponential distribution with parameter \(\lambda\) is in general \(\lambda/(r + \lambda)\) where \(r\) is the constant per annum continuously compounded discount rate. Indeed,
\[
E[e^{-rT}] = \int_0^\infty e^{-rt} f_T(t) \, dt = \int_0^\infty e^{-rt} \cdot \lambda e^{-\lambda t} \, dt = \lambda \int_0^\infty e^{-(r+\lambda)t} \, dt = \frac{\lambda}{r+\lambda}.
\]

Hence if \(T = \min(T_1, \ldots, T_n)\) where the \(T_j\) are independent and exponentially distributed with \(T_j \sim \text{Exponential}(\lambda_j)\) then the actuarial present value of $1 payable at future time \(T\)
is
\[ \frac{\sum_{j=1}^{n} \lambda_j}{r + \sum_{j=1}^{n} \lambda_j}. \]

Consequently, the actuarial present value for $1 of benefit on a joint life policy for which the insured lives have future lifetimes that are independent and exponentially distributed is
\[ \frac{\sum_{j=1}^{n} \lambda_j}{r + \sum_{j=1}^{n} \lambda_j}, \]

where $r$ is the constant per annum continuously compounded discount rate and $\lambda_1, \ldots, \lambda_n$ are the parameters of the exponential distributions for the individual lives.

14. Let $X$ be the amount of a random claim measured in thousands of dollars and let $Y$ be the reimbursement corresponding to this claim. From the given information,
\[ Y = \min (.70 (X - 0.5)^+, \ 5) \]

where for any random variable $W$,
\[ W^+ = \begin{cases} 0 & \text{if } W \leq 0, \\ W & \text{if } W > 0. \end{cases} \]

We are also given that $X \sim \text{Pareto}(3, \ 2)$. Hence the survival function of $X$ is given by
\[ S_X(x) = \left( \frac{2}{2 + x} \right)^3 \text{ for } x \geq 0. \]

a. To determine a formula for $F_Y$, we consider separately the cases $y < 0$, $y = 0$, $0 < y < 5$, and $y \geq 5$. Since the reimbursement amount $Y$ lies between 0 and 5 (according to the terms of the contract), the cases $y < 0$ and $y \geq 5$ are trivial. In particular, $F_Y[y] = 0$ for $y < 0$ and $F_Y[y] = 1$ for $y \geq 5$. Hence we need only consider the cases $y = 0$ and $0 < y < 5$ in detail.

Consider first the case $y = 0$. Note that from the definition of $Y$, the event $Y \leq 0$ is
Consider first the case $y = 0$. Note that from the definition of $Y$, the event $Y \leq 0$ is equivalent to the event $X \leq 0.5$. Hence $F_Y[0] = \Pr[Y \leq 0] = \Pr[X \leq 0.5] = F_X[0.5]$.

Since $S_X[x] = (2/2 + x)^3$ for $x \geq 0$, it follows that

$$F_Y[0] = F_X[0.5] = 1 - S_X[0.5] = 1 - \left(\frac{2}{2.5}\right)^3 = 1 - \left(\frac{4}{5}\right)^3 = .488.$$  

Now consider the case $0 < y < 5$. From the definition of $Y$, the event $Y \leq y$ is equivalent to the event $X \leq 0.5 + \frac{y}{.70}$. Hence for $0 < y < 5$,

$$F_Y[y] = \Pr[Y \leq y] = \Pr[X \leq 0.5 + \frac{y}{.70}] = F_X\left[0.5 + \frac{y}{.70}\right] = 1 - S_X\left[0.5 + \frac{y}{.70}\right].$$

Since $S_X[x] = (2/2 + x)^3$ for $x \geq 0$, it follows that

$$F_Y[y] = 1 - \left(\frac{2}{2.5 + \frac{y}{.70}}\right)^3 = 1 - \left(\frac{1.4}{1.75 + y}\right)^3$$

for $0 < y < 5$.

Combining the formulas for $F_Y$ in the various cases, we obtain

$$F_Y[y] = \begin{cases} 
0 & \text{for } y < 0, \\
1 - \left(\frac{1.4}{1.75 + y}\right)^3 & \text{for } 0 \leq y < 5, \\
1 & \text{for } y \geq 5.
\end{cases}$$

Note that $Y$ has probability masses at $y = 0$ and $y = 5$ and a continuous distribution of probability on $(0, 5)$. The probability mass at $y = 0$ represents the event that the submitted claim is less than or equal to the deductible. The probability mass at $y = 5$ represents the event that the submitted claim is such that the cap has been reached.

b. The expected reimbursement is $E[Y]$. The simplest way to determine $E[Y]$ when $X \sim \text{Pareto}(3, 2)$ is to express $Y$ as a linear combination of functions of the form $\min(X, \cdot \cdot \cdot)$ and then use the formula for the limited expected value function of a Pareto random variable determined in exercise 5.
From the definition of the contract terms,
\[ Y = \min (.70 \ (X - 0.5)^+, \ 5). \]

Using properties of the minimum function and the fact that
\((X - d)^+ = X - \min(X, d)\)
we have
\[ Y = \min (.70 \ (X - 0.5)^+, \ 5) =
    .70 \ \min\left( (X - 0.5)^+, \ \frac{50}{7} \right) =
    .70 \ \min\left( X - \min(X, 0.5), \ \frac{50}{7} \right) =
    .70 \ \left\{ \min\left( X, \ \frac{50}{7} + \min(X, 0.5) \right) - \min(X, 0.5) \right\}. \]

Considering separately the cases \(X < \frac{50}{7} + \frac{1}{2}\) and \(X \geq \frac{50}{7} + \frac{1}{2}\) we also have
\[ \min\left( X, \ \frac{50}{7} + \min(X, 0.5) \right) = \min\left( X, \ \frac{50}{7} + \frac{1}{2} \right). \]

Hence
\[ Y = .70 \ \left\{ \min\left( X, \ \frac{50}{7} + \frac{1}{2} \right) - \min\left( X, \ \frac{1}{2} \right) \right\}. \]

Taking the expectation, it follows that
\[ E[Y] = .70 \ \left\{ E\left[ X; \ \frac{50}{7} + \frac{1}{2} \right] - E\left[ X; \ \frac{1}{2} \right] \right\}. \]

Now in the solution to exercise 5, students will have shown that if \(X \sim \text{Pareto}(s, \ \beta)\) with \(s \neq 1\) then
\[ E[X; m] = \frac{\beta}{s - 1} \ \left\{ 1 - \left( \frac{\beta}{\beta + m} \right)^{s-1} \right\}. \]

Hence for \(X \sim \text{Pareto}(3, \ 2)\) we have
\[ E[X; m] = 1 - \left( \frac{2}{2 + m} \right)^2 \text{ for } m \geq 0. \]

In particular,
\[ E \left[ X; \frac{50}{7} + \frac{1}{2} \right] = 1 - \left( \frac{2}{2 + \frac{50}{7} + \frac{1}{2}} \right)^2 = 1 - \left( \frac{28}{135} \right)^2 \]

and
\[ E \left[ X; \frac{1}{2} \right] = 1 - \left( \frac{2}{2 + \frac{1}{2}} \right)^2 = 1 - \left( \frac{4}{5} \right)^2. \]

Since the distribution of submitted claims (in thousands of dollars) is assumed to be Pareto(3, 2), it follows that the expected reimbursement (in thousands of dollars) is
\[ E[Y] = \begin{align*}
\begin{array}{ll}
0.7 \left( E \left[ X; \frac{50}{7} + \frac{1}{2} \right] - E \left[ X; \frac{1}{2} \right] \right) & = 0.7 \left( \left( 1 - \left( \frac{28}{135} \right)^2 \right) - \left( 1 - \left( \frac{4}{5} \right)^2 \right) \right) \\
\end{array}
\end{align*} \approx 0.41788752.

Consequently, the expected reimbursement is $417.89.

**Comment:** The answer given in the "Answers to Selected Exercises" section at the back of the textbook is incorrect in some printings.

17. Consider a contract in which the insurer pays 80% of the first $2000 of eligible claim expenses in excess of a $200 deductible and 50% of any remaining claim expenses, with a maximum possible payment of $10,000. Let \( X \) be the size of a random loss and let \( Y \) be the insurer's payment toward this loss. Then
\[
\begin{align*}
Y &= 0 & \text{if } X < 200, \\
Y &= 0.8 (X - 200) & \text{if } 200 \leq X < 2200, \\
Y &= 0.8 (2000) + 0.5 (X - 2200) & \text{if } 2200 \leq X < x^*, \\
Y &= 10,000 & \text{if } X \geq x^*,
\end{align*}
\]
where \( x^* \) is such that
(.80)(2000) + (.50)(x^* - 2200) = 10,000, \\
i.e., x^* = 19,000.

a. The easiest way to decompose the contract into a portfolio of single layer contracts is to consider a graph of the insurer’s payout $Y$ as a function of the loss $X$:

By limiting the plot range, we can get a better sense of the true character of the graph for losses $X$ with $X \leq 500$: 
In this "close-up" graph, the dotted line has slope 1 and is included to give the reader a sense of the relative size of the slope of the graph of $Y$ on the interval $200 < X < 2200$.

Put

$Y_1 = \min (Y, 1600)$

and

$Y_2 = Y - Y_1$.

From the graphs just given, it should be clear that $Y_1$ and $Y_2$ represent the payouts on single layer contracts and $Y_1 + Y_2 = Y$. This can be confirmed by considering graphs of $Y_1$ and $Y_2$ as functions of $X$ and recalling the shape of the graph of the insurer's payout on a single layer combination contract:
By limiting the plot range of the graph of $Y_1$ we obtain a graph that is immediately recognizable as the payout on a single layer contract:
It follows from these graphs that \( Y_1 \) represents the insurer's payout on a single layer contract with deductible \( d = \$200 \), coinsurance fraction \( a = 80 \% \) and maximum reimbursement \( m = \$1600 \), and \( Y_2 \) represents the insurer's payout on a single layer contract with deductible \( d = \$2200 \), coinsurance fraction \( a = 50 \% \) and maximum reimbursement \( m = \$8400 \). Hence the double layer combination contract given in the statement of the question can be considered a portfolio of the single layer contracts \( C_1, C_2 \) with respective parameters
\[
d = 200, \ a = 0.80, \ m = 1600
\]
and
\[
d = 2200, \ a = 0.50, \ m = 8400.
\]

Comment: The technique to decompose contracts with more than two layers is similar.

b. Let \( C_1 \) be the single layer contract with parameters \( d_1 = 200, \ a_1 = 0.80, \ m_1 = 1600 \) and let \( C_2 \) be the single layer contract with parameters \( d_2 = 2200, \ a_2 = 0.50, \ m_2 = 8400 \). In part a, it was shown that the double layer combination contract of this question is equivalent to a portfolio of the single layer contracts \( C_1 \) and \( C_2 \). Hence
\[
h_w[X] = h_{w}^{C_1}[X] + h_{w}^{C_2}[X]
\]
where \( h_{w}^{C_1}[X] \) is the writer's payout on contract \( C_1 \) and \( h_{w}[X] \) is the writer's payout on the
given double layer contract. We are interested in $E[h_w[X]]$ when $X$ has an exponential
distribution with mean 4000, i.e., $X \sim \text{Exponential}(1/4000)$.

From the general formula in section 7.3 for the expected payout on a single layer
contract we have

$$E[h_w^C[X]] = \alpha_1 \left\{ E \left[ X; d_1 + \frac{m_1}{\alpha_1} \right] - E[X; d_1] \right\}.$$

Moreover, for $X \sim \text{Exponential}(\lambda)$ we have

$$E[X; m] = \frac{1}{\lambda} \left( 1 - e^{-\lambda m} \right)$$

(see Example 2 of section 7.2 of the textbook). This follows directly from the formula
$E[X; m] = \int_0^m S_X[x] \, dx$ for nonnegative $X$ using the fact that $S_X[x] = e^{-\lambda x}$ for
$X \sim \text{Exponential}(\lambda)$. Consequently,

$$E[h_w^C[X]] = \alpha_1 \left\{ E \left[ X; d_1 + \frac{m_1}{\alpha_1} \right] - E[X; d_1] \right\} =$$

$$(.80) \left\{ E \left[ X; 200 + \frac{1600}{80} \right] - E[X; 200] \right\} = (.80) \left[ E[X; 2200] - E[X; 200] \right] =$$

$$(.80) \left[ 4000 \left( 1 - e^{-2200/4000} \right) - 4000 \left( 1 - e^{-200/4000} \right) \right] = 3200 \left( e^{-0.05} - e^{-0.55} \right)$$

and

$$E[h_w^C[X]] = \alpha_2 \left\{ E \left[ X; d_2 + \frac{m_2}{\alpha_2} \right] - E[X; d_2] \right\} =$$

$$(.50) \left\{ E \left[ X; 2200 + \frac{8400}{50} \right] - E[X; 2200] \right\} = (.50) \left[ E[X; 19000] - E[X; 2200] \right] =$$

$$(.50) \left[ 4000 \left( 1 - e^{-19000/4000} \right) - 4000 \left( 1 - e^{-2200/4000} \right) \right] = 2000 \left( e^{-0.55} - e^{-4.75} \right).$$

Therefore, the expected payout to the policyholder (i.e., the policyholder's expected
reimbursement) when $X$ has an exponential distribution with mean $4000$ is

$$E[h_w[X]] = E[h_w^C[X]] + E[h_w^G[X]] = 3200 \left( e^{-0.05} - e^{-0.55} \right) + 2000 \left( e^{-0.55} - e^{-4.75} \right) =$$

$$3200 e^{-0.05} - 1200 e^{-0.55} - 2000 e^{-4.75} \approx 2334.29.$$
19. Let $X$ be the total allowable expenses in thousands of dollars incurred during the year for a given policyholder and let $Y$ be the amount reimbursed. To answer parts a and b of this question, we need to determine a formula for $E[Y]$ in which the deductible and the out-of-pocket maximum are unspecified.

Let $d$ be the deductible in thousands of dollars and let $M$ be the out-of-pocket maximum in thousands of dollars. (We use an upper case $M$ here rather than a lower case $m$ to avoid possible confusion with the notation of section 7.3 of the textbook, where $m$ is used to denote the maximum amount payable by the insurer. Note that $M$ is a fixed constant in this context, not a random variable, despite our convention of reserving upper case letters for random quantities.) Then the formula for $Y$ in terms of $d$ and $M$ is

\[
Y = 0 \quad \text{if } 0 \leq X \leq d, \\
Y = (.80)(X - d) \quad \text{if } d < X < \beta, \\
Y = X - M \quad \text{if } X \geq \beta,
\]

where $\beta$ is the value of $X$ at which the out-of-pocket maximum is breached. The value of $\beta$ can be determined explicitly by solving the equation
\[
d + (.20)(X - d) = M.
\]

The left hand side of this equation is the amount paid by the policyholder under the assumption that total expenses are greater than the deductible but less than the amount at which the out-of-pocket maximum is triggered. The right hand side of this equation is the out-of-pocket maximum. Note that the deductible is considered an out-of-pocket expense. Solving this equation we obtain
\[
X = d + \frac{M - d}{.20}.
\]

Hence
\[
\beta = d + \frac{M - d}{.20}.
\]

Substituting this into the formula for $Y$ we obtain
\[
Y = 0 \quad \text{if } 0 \leq X \leq d,
\]
Accordingly, and applying the assumption, $X \sim \text{Exponential}(1)$. Hence

\[
E[Y] = \int_0^d 0 \cdot f_X(x) \, dx + \int_d^{d+(M-d)/20} (0.80) (x-d) f_X(x) \, dx + \int_{d+(M-d)/20}^\infty (x-M) f_X(x) \, dx =
\]

\[
0 + \int_d^{5M-4d} (0.80) (x-d) e^{-x} \, dx + \int_{5M-4d}^\infty (x-M) e^{-x} \, dx.
\]

Applying integration by parts we have

\[
\int_d^{5M-4d} (0.80) (x-d) e^{-x} \, dx = (0.80) (x-d) \frac{e^{-x}}{(-1)} \bigg|_d^{5M-4d},
\]

\[
= -0.80 (5M - 5d) \left(-e^{-5M+4d}\right) - 0 + (0.80) \frac{e^{-x}}{(-1)} \bigg|_d^{5M-4d} =
\]

\[
-4(M-d) e^{-5M+4d} + \{0.80(-e^{-5M+4d}) - 0.80(-e^{-d})\} =
\]

\[
0.80 e^{-d} - 4(M-d) + 0.80 e^{-5M+4d},
\]

and similarly

\[
\int_{5M-4d}^\infty (x-M) e^{-x} \, dx =
\]

\[
(x-M) \frac{e^{-x}}{(-1)} \bigg|_{5M-4d}^\infty \frac{e^{-x}}{(-1)} = 0 - (4M-4d) \frac{e^{-x}}{(-1)} \bigg|_{5M-4d}^\infty + (4M-4d) e^{-5M+4d} =
\]

\[
4(M-d) e^{-5M+4d} + e^{-5M+4d} = (4(M-d) + 1) e^{-5M+4d}.
\]

Consequently,

\[
E[Y] = 0 + \left(0.80 e^{-d} - 4(M-d) + 0.80 e^{-5M+4d}\right) + (4(M-d) + 1) e^{-5M+4d} =
\]

\[
0.80 e^{-d} + 0.20 e^{-5M+4d}.
\]

According to the given information, the insurer would like the average reimbursement per member to be no greater than $200. Therefore the values of $M$ and $d$ must be such that
.80 e^{-d} + .20 e^{-5M+4d} \leq .20.

a. Suppose that the deductible remains at $500, i.e., d = 0.5. Then the out-of-pocket maximum must be selected such that

\[ .80 e^{-0.5} + .20 e^{-5M+4(0.5)} \leq .20. \]

Since \( .80 e^{-0.5} \approx .485 > .20 \), there is no value of \( M \) for which this inequality holds. It follows that there is no level of the out-of-pocket maximum that would result in the insurer’s average reimbursement being at most $200 if the deductible remains at $500.

b. Suppose that the maximum out-of-pocket expense remains at $1500, i.e., \( M = 1.5 \). Then the deductible must be selected such that

\[ .80 e^{-d} + .20 e^{-5(1.5)+4d} \leq .20, \]

that is,

\[ 4 + e^{-7.5d} - e^d \leq 0. \]

The latter inequality is a quintic in \( e^d \) and must be solved numerically. We do this by considering the function \( g \) defined as \( g(x) = e^{-7.5}x^3 - x + 4 \). As a first step to determining the roots of \( g \), it is useful to consider a graph of \( g \) for \( x \geq 0 \):

![Graph of g(x)](image)

This graph suggests that \( g(x) > 0 \) for all \( x > 0 \). By considering the derivative \( g' \) one can
show using standard Calculus techniques that this is in fact the case. It follows that
\(4 + e^{-7.5} 5^d - e^d > 0\)
for all \(d \geq 0\). Therefore, there is no level of the deductible that would result in the insurer's average reimbursement being at most $200 if the out-of-pocket maximum remains at $1500.

24. Let \(P\) be the required annual premium payable continuously under the requirement that the present value of premium payments exceeds the present value of benefit payments 95% of the time. Let \(T\) be the future lifetime of the insured and let \(r\) be the per annum continuously compounded rate of interest. By assumption, \(T\) is uniformly distributed on \((0, 20)\) and \(r = 6\%\). From the discussion in section 7.4 of the textbook, the present value of premium payments is
\[
P \cdot \frac{1 - e^{-rT}}{r}
\]
and the present value of benefits is
\[
10,000 e^{-rT}.
\]
Hence the requirement on the premium \(P\) is
\[
\text{Pr}\left[P \frac{1 - e^{-rT}}{r} > 10,000 e^{-rT}\right] = .95.
\]
Now
\[
P \cdot \frac{1 - e^{-rT}}{r} > 10,000 e^{-rT} \iff P - P e^{-rT} > 10,000 r e^{-rT} \iff
\]
\[
(P + 10,000 r) e^{-rT} < P \iff e^{-rT} < \frac{P}{P + 10,000 r} \iff T > -\frac{1}{r} \log \left[\frac{P}{P + 10,000 r}\right].
\]
Hence
\[
\Pr \left[ P \frac{1 - e^{-rT}}{r} > 10,000 \ e^{-rT} \right] = \\
\Pr \left[ T > \frac{1}{r} \log \left( \frac{P}{P + 10,000 \ r} \right) \right] = \frac{20 - (\frac{1}{r} \log \left[ P/(P + 10,000 \ r) \right])}{20}
\]

(provided that \( \frac{1}{r} \log \left[ P/(P + 10,000 \ r) \right] \leq 20 \)). It follows from the requirement

\[
\Pr \left[ P \frac{1 - e^{-rT}}{r} > 10,000 \ e^{-rT} \right] = .95
\]

that

\[
\frac{20 + \frac{1}{r} \log \left[ P/(P + 10,000 \ r) \right]}{20} = .95.
\]

Substituting \( r = .06 \) into this equation and rearranging as appropriate, we find that

\[
\log \left( \frac{P}{P + 600} \right) = -.06
\]

or equivalently

\[
\frac{P}{P + 600} = e^{-0.06}.
\]

Solving this equation for \( P \) we have

\[
P = \frac{600 \ e^{-0.06}}{1 - e^{-0.06}} \approx 9,703.00.
\]

Therefore the annual premium payable continuously that should be charged if the insurer wishes the present value of premiums to exceed the present value of benefits 95% of the time is $9,703. This is significantly greater than the annual premium determined by equating actuarial present values ($836.57). In fact the annual premium payable continuously is greater than the single premium that one must pay today under an analogous 95% criterion ($9,417.65)!
To understand these apparently contradictory results, note that the $9,703 figure just calculated is a per annum amount that is payable continuously. In the absence of life contingencies (i.e., assuming survival to the end of the year), $9,703 payable continuously throughout the year is equivalent to

$$\$9,703 \left( \frac{1 - e^{-0.06}}{0.06} \right) = \$9,417.65$$

payable at the beginning of the year or

$$\$9,417.65 \times e^{0.06} = \$10,000$$

payable at the end of the year. Hence the 95% criterion in the current exercise amounts to setting the premium under the assumption that death occurs at the end of the first year of the policy. The reason that this happens is that the future lifetime of the policyholder is uniformly distributed on (0, 20) and as a result the probability of survival beyond the first year is exactly 95%. If the future lifetime were modeled using any other distribution, the premium could not be set in such a simple way.

26. Let $T$ be the lifetime of the system measured in hours and let $T_A, T_B, T_C, T_D$ be the lifetimes of the system components $A, B, C, D$ respectively. Then from the configuration given in Figure 7.13 of the textbook, $T = \max (\min (T_A, T_B), \max (T_C, T_D))$.

Put $Y_1 = \min(T_A, T_B)$ and $Y_2 = \max(T_C, T_D)$. Then since $T_A, T_B, T_C, T_D$ are independent, the distributions of $Y_1$ and $Y_2$ are given by

$$S_{Y_1}[t] = S_{T_A}[t] S_{T_B}[t],$$

$$F_{Y_2}[t] = F_{T_C}[t] F_{T_D}[t]$$

(see section 7.5 of the textbook) and so the distribution of $T$ is given by

$$F_T[t] = F_{Y_1}[t] F_{Y_2}[t] = (1 - S_{T_A}[t] S_{T_B}[t]) \left( F_{T_C}[t] F_{T_D}[t] \right).$$

Now by assumption, $T_A \sim \text{Exponential}(\frac{1}{4})$, $T_B \sim \text{Exponential}(\frac{1}{3})$, $T_C \sim \text{Exponential}(\frac{1}{2})$, $T_D \sim \text{Exponential}(1)$. Hence, the distribution function of $T$ is given by
\[ F_T[t] = (1 - S_{T_a}[t] S_{T_b}[t]) F_{T_a}[t] F_{T_b}[t] = (1 - e^{-t/4}) (1 - e^{-t/2}) (1 - e^{-t}) = (1 - e^{-7t/12}) (1 - e^{-t/2}) (1 - e^{-t}) \]

for \( t \geq 0 \).

Since \( T \) (being a lifetime random variable) is nonnegative, the expected value of \( T \) can be determined using the formula \( E[T] = \int_0^\infty S_T[t] \, dt \). Hence

\[ E[T] = \int_0^\infty S_T[t] \, dt = \int_0^\infty (1 - F_T[t]) \, dt = \int_0^\infty \left\{1 - (1 - e^{-7t/12}) (1 - e^{-t/2}) (1 - e^{-t})\right\} \, dt. \]

However,

\[ (1 - e^{-7t/12}) (1 - e^{-t/2}) (1 - e^{-t}) = (1 - e^{-7t/12}) (1 - e^{-t/2} - e^{-t} + e^{-3t/2}) = 1 - e^{-7t/2} - e^{-t} + e^{-3t/2} - e^{-7t/12} + e^{-13t/12} + e^{-19t/12} - e^{-25t/12}. \]

Consequently,

\[ E[T] = \int_0^\infty \left\{e^{-7t/2} + e^{-t} - e^{-3t/2} + e^{-7t/12} - e^{-13t/12} - e^{-19t/12} + e^{-25t/12}\right\} \, dt. \]

Using the fact that \( \int_0^\infty e^{-t/\lambda} \, dt = 1/\lambda \) we obtain

\[ E[T] = 2 + 1 - \frac{2}{3} + \frac{12}{7} - \frac{12}{13} - \frac{12}{19} + \frac{25}{25} = \frac{385,519}{129,675} \approx 2.97296318. \]

Hence the expected life of the system is about 3 hours.

30. Suppose that \( X = \tan[\Theta] \) and \( \Theta \) is uniformly distributed on \((-\pi, \pi)\). The distribution function for \( X \) can be determined by considering a graph of the transformation \( X = \tan[\Theta] \) with the event \( X \leq x \) highlighted:
Note that in this graph it is assumed that $x > 0$.

Suppose that $x > 0$. Then the event $X \leq x$ is equivalent to the event

$$-\pi < \Theta \leq \arctan[x] - \pi \quad \text{or} \quad -\frac{\pi}{2} < \Theta \leq \arctan[x] \quad \text{or} \quad \frac{\pi}{2} < \Theta < \pi$$

where $\arctan[x]$ is the branch of the inverse tangent function with values in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Hence

$$\Pr[X \leq x] = \Pr[-\pi < \Theta \leq \arctan[x] - \pi] + \Pr[-\frac{\pi}{2} < \Theta \leq \arctan[x]] + \Pr[\frac{\pi}{2} < \Theta < \pi]$$

and so since $\Theta$ is uniformly distributed on $(-\pi, \pi)$,

$$\Pr[X \leq x] = \frac{(\arctan[x] - \pi) - (-\pi)}{2\pi} + \frac{\arctan[x] - (-\frac{\pi}{2})}{2\pi} + \frac{\pi - \frac{\pi}{2}}{2\pi} = \frac{\arctan[x]}{\pi} + \frac{1}{2}.$$ 

On the other hand, suppose that $x \leq 0$. Then the event $X \leq x$ is equivalent to the event

$$-\frac{\pi}{2} < \Theta \leq \arctan[x] \quad \text{or} \quad \frac{\pi}{2} < \Theta < \arctan[x] + \pi.$$

Hence
\[
\Pr[X \leq x] = \Pr\left[ -\frac{\pi}{2} < \theta \leq \arctan(x) \right] + \Pr\left[ \frac{\pi}{2} < \theta < \arctan(x) + \pi \right] = \\
\arctan[x] - \left( -\frac{\pi}{2} \right) \quad \arctan[x] + \pi - \left( -\frac{\pi}{2} \right) = \frac{\arctan[x]}{2\pi} + \frac{1}{2}
\]

which is the same as the formula obtained in the case \( x > 0 \).

Consequently, the distribution function of \( X \) is given by
\[
F_X(x) = \frac{1}{2} + \frac{\arctan[x]}{\pi} \quad \text{for} \ x \in \mathbb{R}
\]

and so the density function of \( X \) is
\[
f_X(x) = F_X'(x) = \frac{1}{\pi} \cdot \frac{d}{dx} \arctan(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} \quad \text{for} \ x \in \mathbb{R}
\]

as required.
3. Let $X, Y$ be random variables with joint density function

$$f_{X,Y}(x, y) = \begin{cases} 
1 & \text{for } x + 2y \leq 2, \ x \geq 0, \ y \geq 0, \\
0 & \text{otherwise}, 
\end{cases}$$

and put $S = X + Y$ and $D = X - Y$. We are required to determine the distributions of $S$ and $D$.

Note that $X$ and $Y$ are not independent. This follows immediately from an examination of the region of nonzero probability for $f_{X,Y}$.
Since $f_{X,Y}$ is constant on the region of nonzero probability, probabilities associated with $X$ and $Y$ can be determined by calculating areas in the $x$-$y$ plane and then multiplying by the constant value of $f_{X,Y}$, which is 1 in this case. This is the approach we take to determine the distribution functions of $S$ and $D$.

**Distribution of $S$:** From the graph of the region of nonzero probability for $f_{X,Y}$ the calculation of $F_S(s) = \Pr[S \leq s]$ can be separated into four cases:

i. $s \leq 0$,

ii. $0 < s \leq 1$, 

iii. $1 < s \leq 2$,
iv. $s > 2$.

This corresponds to a partition of $s$ values according to the shapes of the regions
$\mathcal{R}_0 = \{(x, y) : x + y \leq s, \ x + 2y \leq 2, \ x \geq 0, \ y \geq 0\}$.

Since $f_{X,Y}$ is 0 outside the region $\{(x, y) : x + 2y \leq 2, \ x \geq 0, \ y \geq 0\}$, it is clear that
$F_S[s] = 0$ for all $s \leq 0$ and $F_S[s] = 1$ for all $s > 2$. Hence we need only consider the cases
$0 < s \leq 1, 1 < s \leq 2$ in detail.

Consider first the case $0 < s \leq 1$. In this case, the region of integration for calculating
$F_S[s] = \Pr[S \leq s] = \Pr[X + Y \leq s]$ is triangular with base and height equal to $s$:
Since $f_{X,Y}$ is constant and equal to 1 on this region, it follows that

$$F_3[s] = \frac{1}{2} s^2$$

for $0 < s \leq 1$.

Now consider the case $1 < s \leq 2$. In this case the region of integration for calculating $\Pr[X + Y \leq s]$ is a quadrilateral that can be decomposed into a rectangle and two right-angled triangles:

Using basic geometry, the area of the shaded region is
\[ \frac{1}{2} (2s - 2) (1 - (2 - s)) + (2 - s) (2s - 2) + \frac{1}{2} (s - (2s - 2)) (2 - s) = (s - 1) (s - 1) + 2 (2 - s) (s - 1) + \frac{1}{2} (2 - s) (2 - s) = -\frac{1}{2} s^2 + 2s - 1. \]

Hence
\[ F_S[s] = -\frac{1}{2} s^2 + 2s - 1 \quad \text{for} \ 1 < s \leq 2. \]

Consequently, the distribution function of \( S \) is given by
\[
\begin{align*}
F_S[s] &= 0 \quad \text{for} \ s \leq 0, \\
F_S[s] &= \frac{1}{2} s^2 \quad \text{for} \ 0 < s \leq 1, \\
F_S[s] &= -\frac{1}{2} s^2 + 2s - 1 \quad \text{for} \ 1 < s \leq 2, \\
F_S[s] &= 1 \quad \text{for} \ s > 2.
\end{align*}
\]

The graph of \( F_S \) is as follows:

From this graph and the preceding formula for \( F_S \) it is clear that \( F_S \) has the properties of a distribution function. Hence there is no obvious mistake in the formula derived for \( F_S \).
Distribution of $D$: From the graph of the region of nonzero probability for $f_{X,Y}$ the calculation of $F_D[d] = \Pr[D \leq d]$ can be separated into four cases:

i. $d \leq -1$,

ii. $-1 < d \leq 0$,

iii. $0 < d \leq 2$,

iv. $d > 2$.

This corresponds to a partition of $d$ values according to the shapes of the regions

$\mathcal{R}_d = \{ (x, y) : x - y \leq d, x + 2y \leq 2, x \geq 0, y \geq 0 \}$.

Since $f_{X,Y}$ is 0 outside the region $\{ (x, y) : x + 2y \leq 2, x \geq 0, y \geq 0 \}$, it is clear that $F_D[d] = 0$ for all $d \leq -1$ and $F_D[d] = 1$ for all $d > 2$. Hence we need only consider the cases $-1 < d \leq 0, 0 < d \leq 2$ in detail.

Consider first the case $-1 < d \leq 0$. In this case, the region of integration for calculating $F_D[d] = \Pr[D \leq d] = \Pr[X - Y \leq d]$ is triangular:
Using basic geometry, the area of the shaded region is

\[ \frac{1}{2} (1 - (-d)) \left( \frac{2}{3} (d + 1) \right) = \frac{1}{3} (d + 1)^2. \]

Since \( f_{X,Y} \) is constant and equal to 1 on this region, it follows that

\[ F_D[d] = \frac{1}{3} (d + 1)^2 \quad \text{for} \quad -1 < d \leq 0. \]

Now consider the case \( 0 < d \leq 2 \). In this case the region of integration for calculating \( \Pr[X - Y \leq d] \) is a quadrilateral:
The area of the shaded region can be determined by subtracting the area of the triangle bounded by the points \((d, 0), (2, 0), \left(\frac{2}{3} d + \frac{2}{3}, \frac{2}{3} - \frac{1}{3} d\right)\) from the area of the triangle bounded by the points \((0, 0), (2, 0), (0, 1)\). Taking this approach, the shaded area is

\[
\frac{1}{2} (2) (1) - \frac{1}{2} (2 - d) \left(\frac{2}{3} - \frac{d}{3}\right) = 1 - \frac{1}{6} (2 - d)^2.
\]

Since \(f_{X,Y}\) is constant and equal to 1 on the shaded region, it follows that

\[
F_D[d] = 1 - \frac{1}{6} (2 - d)^2 \quad \text{for} \quad 0 < d \leq 2.
\]

Consequently, the distribution function of \(D\) is given by
\[ F_D[d] = \begin{cases} 0 & \text{for } d \leq -1, \\ \frac{1}{3} (d + 1)^2 & \text{for } -1 < d \leq 0, \\ 1 - \frac{1}{6} (2 - d)^2 & \text{for } 0 < d \leq 2, \\ 1 & \text{for } d > 2. \end{cases} \]

The graph of \( F_D \) is as follows:

From this graph and the preceding formula for \( F_D \) it is clear that \( F_D \) has the properties of a distribution function. Hence there is no obvious mistake in the formula derived for \( F_D \).

**Comment on Exercise 10:** The solution to exercise 10 makes use of the following formula:

\[ \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (x + u)^2} \, dx = \frac{1}{2} e^{-|u|} \quad \text{for } u \in \mathbb{R}. \]

To facilitate understanding of the solution to exercise 10 we will derive this formula here. The derivation is also interesting in its own right as it demonstrates an integration technique that is often useful.

Put
\[ g[u] = \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(x^2 + (u/x)^2\right)} \, dx. \]

From this definition, it should be clear that \( g[u] = g[-u] \) for all \( u \). Hence, it suffices to prove that
\[ g[u] = \frac{1}{2} e^{-u} \quad \text{for } u \geq 0. \]

We will show that \( g'[u] = -g[u] \) for \( u > 0 \)

and then use standard solution techniques for differential equations to deduce that \( g[u] = \frac{1}{2} e^{-u} \) for \( u \geq 0 \).

Differentiating \( g \) (by interchanging the order of differentiation and integration as appropriate) we obtain
\[
\frac{d}{du} \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(x^2 + (u/x)^2\right)} \, dx = \int_{x=0}^{\infty} \frac{d}{du} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(x^2 + (u/x)^2\right)} \right\} \, dx = \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} \left(-\frac{1}{2} \cdot \frac{1}{x^2} \right) \left(2 \frac{1}{x^2} \right) \, dx = -u \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(x^2 + (u/x)^2\right)} \frac{1}{x^2} \, dx.
\]

Applying the substitution \( x = u/y \) with fixed \( u > 0 \) to the latter integral we have
\[
-u \int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left((u/y)^2 + u^2\right)} \frac{1}{x^2} \, dx = -u \int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left[(u/y)^2 + u^2\right]} \left(\frac{y}{u}\right)^2 \left(-\frac{u}{y^2}\right) \, dy = - \int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left[(u/y)^2 + u^2\right]} \, dy = -g[u].
\]

Hence
\[ g'[u] = -g[u] \quad \text{for } u > 0 \]
as claimed.

Now the equation \( g'[u] = -g[u] \) is a separable differential equation and can be solved
fairly easily. Written in terms of the variables \( x \) and \( y \) where \( y = g[u] \) the equation has the form
\[
\frac{dy}{dx} = -y.
\]

Separating the variables in the standard way, we have
\[
\frac{dy}{y} = -dx
\]
from which we obtain, on integration of both sides,
\[
\log y = -x + C
\]
where \( C \) is an arbitrary constant. Solving for \( y \), we obtain the solution
\[
y = y_0 e^{-x}
\]
where \( y_0 \) is the value of \( y \) when \( x = 0 \). In terms of the earlier notation we have
\[
g[u] = g[0] e^{-u} \quad \text{for } u > 0.
\]

To complete the derivation we need only determine the value of \( g[0] \). From the definition of \( g \) we have
\[
g[0] = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.
\]
The integrand of the latter integral is the density function for a standard normal random variable, i.e., it is the density function for Normal(0, 1). Since the standard normal density is symmetric about 0 and probability densities integrate to 1 it follows that
\[
g[0] = \frac{1}{2}.
\]
Consequently,
\[
g[u] = \frac{1}{2} e^{-u} \quad \text{for } u \geq 0.
\]
Since $g[u] = g[-u]$ for all $u$, it follows that
\[ g[u] = \frac{1}{2} e^{-|u|} \quad \text{for all } u. \]

Therefore
\[ \int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 + u^2)} \, dx = \frac{1}{2} e^{-|u|} \quad \text{for } u \in \mathbb{R}, \]

which is the formula we set out to prove.

10. Suppose that $X$ and $Y$ are independent random variables with $X \sim \text{Exponential}(1)$ and $Y \sim \text{Normal}(0, 1)$. We are asked to determine the density function for $X^{1/2} Y$.

Note that for $X \sim \text{Exponential}(1), X^{1/2} \sim \text{Weibull}(1, 2)$ (see section 6.2.1 of the textbook). Hence the distribution of $X^{1/2} Y$ is identical to the distribution of $V Y$ where $V \sim \text{Weibull}(1, 2)$.

In exercise 11, the student is asked to determine the distribution of $X Y$ under the assumption that $X \sim \text{Normal}(0, 1), Y \sim \text{Weibull}(\sqrt{2}, 2)$, and $X, Y$ are independent. (The distribution types of $X$ and $Y$ are not explicitly given in the statement of exercise, but they are easy to deduce from the forms of the given density functions.) From the observations just made, this is essentially the same problem as the current exercise (the only difference being the parameters of the Weibull distribution). Hence the density function obtained in the current exercise should have the same basic form as the one given in exercise 11.

Put $W = X^{1/2} Y$ where $X \sim \text{Exponential}(1), Y \sim \text{Normal}(0, 1)$, and $X, Y$ are independent. We will determine the distribution of $W$ directly without making use of the observation that $X^{1/2} \sim \text{Weibull}(1, 2)$. Since $X$ cannot assume negative values, $X^{1/2}$ is well defined and it makes sense to ask for the distribution of $W$.

Note that the distribution of $W$ is symmetric about $w = 0$. In particular,
\[ \Pr[W \leq w] = \Pr[W \geq -w] \quad \text{for all } w. \]
The demonstration of this fact is more subtle than it may at first appear. Suppose that we have determined a formula for the distribution function of \( W = X Y \). Then since \( Y \sim \text{Normal}(0, 1) \Rightarrow -Y \sim \text{Normal}(0, 1) \) and since \(-Y\) and \(X\) are independent whenever \(Y\) and \(X\) are independent, we can follow exactly the same steps to determine the distribution function of \( -W = X^{1/2}(-Y) \) and the result will be the same distribution function. Consequently, when \( X \sim \text{Exponential}(1), Y \sim \text{Normal}(0, 1) \) and \(X, Y\) are independent, the distributions of \( X^{1/2} Y \) and \(-X^{1/2} Y\) are identical. (Note that this argument does not contradict the comments made in section 4.1.11 of the textbook where it is pointed out that substitution of equivalent random variables is not in general valid.) Hence for any \( w \),

\[
\Pr[W \leq w] = \Pr[X^{1/2} Y \leq w] = \Pr[-X^{1/2} Y \geq -w] = \Pr[W \geq -w].
\]

It follows that the distribution of \( W \) is symmetric. Consequently, it suffices to determine the density of \( W \) in the case \( w > 0 \).

Suppose henceforth that \( w > 0 \). We determine \( \Pr[W > w] \) by integrating the joint density of \( X \) and \( Y \) over the region defined by the event \( X^{1/2} Y > w \). From the given information, the joint density of \( X \) and \( Y \) is

\[
f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x} e^{-y^2/2} \quad \text{for } x > 0 \text{ and } y \in \mathbb{R}.
\]

A graph of the event \( X^{1/2} Y > w \) can be created using Mathematica or similar computer software.
Carrying out the integration of $f_{X,Y}$ over the shaded region, we have
\[
\Pr[W \geq w] = \Pr\left[Y \geq \frac{w}{X^{1/2}}\right] = \int_{y \geq \frac{w}{x^{1/2}}, x > 0, y > 0} \frac{1}{\sqrt{2\pi}} e^{-x} e^{-y^2/2} \, dx \, dy = \\
\int_{y=0}^{\infty} \int_{x=(w/y)^2}^{\infty} \left(\frac{1}{\sqrt{2\pi}} e^{-y^2/2}\right) e^{-x} \, dx \, dy = \int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \left(\int_{x=(w/y)^2}^{\infty} e^{-x} \, dx\right) \, dy = \\
\int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{-(w/y)^2} \, dy = \int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{w}{\sqrt{2}}\right)^2} \, dy = 1 - \frac{1}{2} \left(\frac{w}{\sqrt{2}}\right)^2.
\]

Now from the comment preceding the solution to this exercise,
\[
\int_{x=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(x^2 + (w/x)^2\right)} \, dx = \frac{1}{2} e^{-|u|}, \quad u \in \mathbb{R}.
\]

Substituting $u = w \sqrt{2}$ and using the assumption $w > 0$ we have
\[
\int_{y=0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{w}{\sqrt{2}}\right)^2} \, dy = \frac{1}{2} e^{-w \sqrt{2}}.
\]

Hence
\[
\Pr[W \geq w] = \frac{1}{2} e^{-w \sqrt{2}} \quad \text{for } w > 0.
\]

Since $W$ is continuous, $\Pr[W = w] = 0$ for all $w$. Consequently,
\[
S_w[w] = \Pr[W > w] = \Pr[W \geq w] - \Pr[W = w] = \frac{1}{2} e^{-w \sqrt{2}} - 0 = \frac{1}{2} e^{-w \sqrt{2}} \quad \text{for } w > 0
\]
and
\[
f_w[w] = -S_w'[w] = - \frac{d}{dw} \left(\frac{1}{2} e^{-w \sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-w \sqrt{2}} \quad \text{for } w > 0.
\]

Since the distribution of $W$ is symmetric, it follows that the density function of $W$ is in general
Chapter 15.
Hence it follows that $f_{W}(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$ for $w \in \mathbb{R}$.

This is the density function that we were required to derive.

15. We are asked to show that if $X_1, \ldots, X_n$ are independent Cauchy random variables then the arithmetic average $X_1\ldots + X_n/n$ also has a Cauchy distribution.

At first glance this may appear to be a violation of the law of large numbers since the distribution of $X_1\ldots + X_n/n$ is the same for all $n$, and hence does not become more concentrated as $n$ increases. However, recall that one of the conditions specified in the statement of the law of large numbers is that the mean of the distribution exist and be finite (see section 8.4.2 of the textbook and take note of the section of the proof where this condition is used). From section 7.6 of the textbook, we know that the mean of the Cauchy distribution does not exist. Hence the law of large numbers is not applicable to random variables having a Cauchy distribution.

To prove the required result, it suffices to show that a weighted sum of any two independent Cauchy random variables is Cauchy. Indeed, if $k Y_1 + (1 - k) Y_2$ is Cauchy whenever $Y_1$ and $Y_2$ are independent Cauchy random variables and $k > 0$ then since

\[
X_1 + \cdots + X_n/n = \frac{X_1 + \cdots + X_{n-1} + 1}{n} - \frac{X_n}{n},
\]

it follows by induction that $(X_1 + \cdots + X_n)/n$ has a Cauchy distribution whenever $X_1, \ldots, X_n$ are independent Cauchy random variables. Consequently if $X_1, \ldots, X_n$ are independent Cauchy random variables and if $k Y_1 + (1 - k) Y_2$ is Cauchy whenever $Y_1$ and $Y_2$ are independent Cauchy random variables and $k > 0$, then by induction, $(X_1 + \cdots + X_n)/n$ has a Cauchy distribution.

Hence suppose that $Y_1$ and $Y_2$ are independent Cauchy random variables, $k$ is a positive real number, and consider the weighted sum $S = k Y_1 + (1 - k) Y_2$. Since $Y_1$ and $Y_2$ are independent, the density function of $S$ can be determined using the convolution formula given in section 8.1.3 of the textbook:

\[
f_{S}(s) = \int_{-\infty}^{\infty} f_{k Y_1}(x) f_{Y_2}(s - x) \, dx.
\]

Since $Y_1$ and $Y_2$ are Cauchy random variables, the densities of $k Y_1$ and $(1 - k) Y_2$ are as
Since $Y_1$ and $Y_2$ are Cauchy random variables, the densities of $kY_1$ and $H_1-kL_2$ are as follows:

$$ f_{kY_1}[x] = \frac{k}{\pi} \cdot \frac{1}{x^2 + k^2}, $$

$$ f_{(1-k)Y_2}[x] = \frac{1-k}{\pi} \cdot \frac{1}{x^2 + (1-k)^2}. $$

Hence

$$ f_{kY_1}[x] f_{(1-k)Y_2}[s-x] = \frac{k (1-k)}{\pi^2} \cdot \frac{1}{x^2 + k^2} \cdot \frac{1}{(x-s)^2 + (1-k)^2}. $$

Using the method of partial fractions, the latter expression can be decomposed into a sum of functions that are more easily integrated. Applying this method we have

$$ \frac{1}{x^2 + k^2} \cdot \frac{1}{(x-s)^2 + (1-k)^2} = \frac{A + B x}{x^2 + k^2} + \frac{C + D x}{(x-s)^2 + (1-k)^2} $$

where

$$ A = \frac{s^2 + 1 - 2k}{(s^2 + 1 - 2k)^2 + 4s^2 k^2}, $$

$$ B = \frac{2s}{(s^2 + 1 - 2k)^2 + 4s^2 k^2}, $$

$$ C = \frac{3s^2 - 1 + 2k}{(s^2 + 1 - 2k)^2 + 4s^2 k^2}, $$

$$ D = -B. $$

Consequently,
\[ f_3[s] = \int_{-\infty}^{\infty} f_k y_s y_{(1-k)} y_{(s-x)} \, dx = \frac{k(1-k)}{\pi^2} \left\{ \int_{-\infty}^{\infty} \frac{1}{x^2 + k^2} \, dx + \int_{-\infty}^{\infty} \frac{1}{(x-s)^2 + (1-k)^2} \, dx \right\}. \]

Applying the substitution \( x = k \, w \) to the first integral in the latter expression we have
\[
\int_{-\infty}^{\infty} \frac{A + B x}{x^2 + k^2} \, dx = \frac{1}{k} \int_{-\infty}^{\infty} \frac{A + B \, k \, w}{w^2 + 1} \, dw.
\]

and applying the substitution \( x = s + (1-k) \, w \) to the second integral in this expression we have
\[
\int_{-\infty}^{\infty} \frac{C + D x}{(x-s)^2 + (1-k)^2} \, dx = \frac{1}{1-k} \int_{-\infty}^{\infty} \frac{C + D ((1-k) \, w + s)}{w^2 + 1} \, dw.
\]

Substituting these expressions into the previous formula for \( f_3 \) and simplifying we have
\[
f_3[s] = \frac{k(1-k)}{\pi^2} \left\{ \int_{-\infty}^{\infty} \frac{A + B x}{x^2 + k^2} \, dx + \int_{-\infty}^{\infty} \frac{C + D x}{(x-s)^2 + (1-k)^2} \, dx \right\} = \frac{1}{\pi^2} \left\{ (1-k) \int_{-\infty}^{\infty} \frac{A + B \, k \, w}{w^2 + 1} \, dw + k \int_{-\infty}^{\infty} \frac{C + D ((1-k) \, w + s)}{w^2 + 1} \, dw \right\} = \frac{1}{\pi^2} \left\{ \int_{-\infty}^{\infty} \frac{A (1-k) + C k + D k \, s}{w^2 + 1} \, dw \right\} + 0 = \frac{1}{\pi^2} \left\{ A (1-k) + C k - B k \, s \right\} \left[ \arctan(w) \right]_{w=-\infty}^{w=\infty} = \frac{1}{\pi^2} \left\{ A (1-k) + C k - B k \, s \right\} \left( \frac{\pi}{2} - \left( -\frac{\pi}{2} \right) \right) = \frac{1}{\pi} \left\{ A (1-k) + C k - B k \, s \right\}.
\]

(Note the use of the relationship \( D = -B \).

Substituting the values of \( A, B, C \) previously specified into the latter expression we have
\[
\frac{1}{\pi} \left\{ A (1-k) + C k - B k \, s \right\} = \]
\[
\frac{1}{\pi} \frac{(s^2 + 1 - 2k)(1 - k) + (4s^2 - (s^2 + 1 - 2k))k - (2s)k}{(s^2 + 1 - 2k)^2 + 4s^2k^2} = \frac{1}{\pi} \frac{s^2 + (1 - 2k)^2}{(s^2 + 1 - 2k)^2 + 4s^2k^2} = \frac{1}{\pi} \frac{s^2 + (1 - 2k)^2}{(s^2 + 1)(s^2 + 1 - 2k)^2} = \frac{1}{\pi} \frac{1}{s^2 + 1}.
\]

Consequently,

\[f_S(s) = \frac{1}{\pi} \frac{1}{s^2 + 1} \quad \text{for } s \in \mathbb{R},\]

which we recognize as the density function of a Cauchy distribution.

Hence we have shown that if \( Y_1 \) and \( Y_2 \) are independent Cauchy random variables and \( k > 0 \) then \( kY_1 + (1 - k)Y_2 \) has a Cauchy distribution. Applying this result repeatedly in the manner discussed earlier it follows that for any independent Cauchy random variables \( X_1, \ldots, X_n \), the arithmetic average \((X_1 + \cdots + X_n)/n\) has a Cauchy distribution. This is the result we were required to prove.

17. Let \( X \) and \( Y \) be the returns on securities \( S_1 \) and \( S_2 \) respectively. We are given that \( \mu_X = 10\% \), \( \sigma_X = 5\% \), \( \mu_Y = 20\% \), \( \sigma_Y = 15\% \), and \( \rho_{XY} = .30 \). The equation in the risk-reward plane for the possible portfolios consisting of \( S_1 \) and \( S_2 \) is

\[\sigma_R^2 = A(\mu_R - \mu_0)^2 + \sigma_0^2\]

where

\[A = \frac{(\sigma_X - \sigma_Y)^2 + 2(1 - \rho_{XY})\sigma_X\sigma_Y}{(\mu_X - \mu_Y)^2} = \frac{(5\% - 15\%)^2 + 2(1 - .30)(5\%)(15\%) }{(10\% - 20\%)^2} = 2.05,\]
\[
\mu_0 = \frac{\mu_X \sigma_Y^2 - (\mu_X + \mu_Y) \rho_{XY} \sigma_X \sigma_Y + \mu_Y \sigma_X^2}{(\sigma_X - \sigma_Y)^2 + 2 (1 - \rho_{XY}) \sigma_X \sigma_Y} = \frac{(10 \%) (15 \%)^2 - (10 \% + 20 \%) (0.30) (15 \%) + (20 \%) (5 \%)^2}{(5 \% - 15 \%)^2 + 2 (1 - .30) (5 \%) (15 \%)} \approx 10.12195 \%,
\]

and
\[
\sigma_0^2 = \frac{\sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)}{(\sigma_X - \sigma_Y)^2 + 2 (1 - \rho_{XY}) \sigma_X \sigma_Y} = \frac{(5 \%)^2 (15 \%)^2 (1 - (.30)^2)}{(5 \% - 15 \%)^2 + 2 (1 - .30) (5 \%) (15 \%)} \approx (4.99695029 \%)^2.
\]

Note that in decimal form, \( \mu_0 = 0.1012195 \) and \( \sigma_0 = 0.0499695029 \). It doesn't matter which form (decimal or percentage) is used in the equations that follow as long as we are consistent and the answers are interpreted appropriately. Note that \( A \) is the same regardless of the form used for \( \mu_0 \) and \( \sigma_0 \).

a. Let \( \alpha \) be the fraction invested in \( S_1 \) that minimizes the variance of the portfolio. Then \( \mu_0 = \alpha \mu_X + (1 - \alpha) \mu_Y \).

From the preceding comments, \( \mu_0 \approx 10.12195 \% \), \( \mu_X = 10 \% \), \( \mu_Y = 20 \% \).

Hence the required fraction is
\[
\alpha = \frac{\mu_0 - \mu_Y}{\mu_X - \mu_Y} = \frac{10.12195 \% - 20 \%}{10 \% - 20 \%} \approx .9878.
\]

So to minimize the variance, about 98.78% should be invested in \( S_1 \).

b. The expected return on the minimum variance portfolio is
This exercise is based on a story told by the Nobel Prize winning economist Paul Samuelson. It is designed to illustrate the difference between risk pooling and risk sharing and expose a common fallacy associated with the law of large numbers. See also section 8.4.3 of the textbook and Samuelson's 1963 paper "Risk and Uncertainty: A Fallacy of Large Numbers" published in volume 98 of the journal Scientia.

a. Let $J_k$ be John's winnings on the $k$-th bet and let $S_J$ be John's accumulated winnings on 100 bets. Then from the given information, the $J_k$ are independent and identically distributed with distribution given by

$$J_k = \begin{cases} 2000 & \text{with probability .50,} \\ -1000 & \text{with probability .50,} \end{cases}$$

and $S_J = J_1 + \cdots + J_k$. The riskiness of a bet or group of bets can be analyzed by considering the variance. From the definition of $J_k$ we have

$$\text{Var } (J_k) = E[J_k^2] - E[J_k]^2 = \left(2000^2 \cdot .50 + (-1000)^2 \cdot .50\right) - \left((2000) \cdot .50 + (-1000) \cdot .50\right)^2 = 12,500,000 - 250,000 = 12,250,000 = (1500)^2.$$

Since the $J_k$ are independent and identically distributed we also have

$$\text{Var } (S_J) = \text{Var } (J_1) + \cdots + \text{Var } (J_{100}) = 100 \cdot \text{Var } (J_k) = \left(15,000\right)^2.$$

Hence by agreeing to 100 bets, John actually increases his risk (by a factor of 100 when risk is measured by the variance). Consequently, if John is not willing to take on the risk of loss associated with a single bet, he should definitely not take on the risk associated with 100 such bets. It follows that John's reasoning is not sound.

b. Let $M_k$ be Matt's winnings on the $k$-th bet and let $S_M$ be Matt's accumulated winnings on 1000 bets. Then from the given information

$$M_k = \begin{cases} 2 & \text{with probability .50,} \\ -1 & \text{with probability .50,} \end{cases}$$

and $S_M = M_1 + \cdots + M_{1000}$. Hence
\[
\text{Var} (M_k) = E\left[M_k^2\right] - E[M_k]^2 = \\
\left\{(2)^2 (.50) + (-1)^2 (.50)\right\} - \left\{(2)(.50) + (-1)(.50)\right\}^2 = \frac{5}{2} - \left(\frac{1}{2}\right)^2 = 2.25
\]

and

\[
\text{Var} (S_M) = \text{Var} (M_1) + \cdots + \text{Var} (M_{1000}) = 1000 \text{Var} (M_k) = 1000 (2.25) = 2250.
\]

Moreover,

\[E[S_M] = 1000 E[M_k] = 500.\]

Now recall from part a that the variance and expected value on a single bet of the type proposed by Paul are

\[
\text{Var} (J_k) = 2,250,000
\]

and

\[E[J_k] = 500.\]

Hence by breaking up Paul’s proposed bet into 1000 pieces, Matt has \textit{reduced} the risk as measured by the variance by a factor of 1000 without changing the expected value. So unlike John, Matt’s reasoning is sound.

c. In order for Matt and John to take advantage of Paul’s offer they need to share the risk of loss. One way of doing this is to share accumulated winnings on the 100 bets with 100 like-minded people. Let \(S\) denote the share of the accumulated winnings for one of these 100 people. Then using the notation of part a,

\[
S = \frac{1}{100} (J_1 + \cdots + J_{100}).
\]

Since the \(J_k\) are independent and identically distributed with

\[E[J_k] = 500,\]

\[\text{Var} (J_k) = 2,250.00,\]

we have
\[ E[S] = \frac{1}{100} (E[J_1] + \cdots + E[J_{100}]) = E[J_k] = 500 \]

and

\[ \text{Var}(S) = \left( \frac{1}{100} \right)^2 (\text{Var}(J_1) + \cdots + \text{Var}(J_{100})) = \frac{1}{100} \text{Var}(J_k) = 22,500. \]

Clearly the risk of loss for one of the 100 people sharing the accumulated winnings is less than the risk faced on a single bet with no sharing of winnings even though the expected gain in both situations is the same. It follows that if John and Matt wish to take advantage of Paul’s tempting offer, they should find 98 other people interested in a bet of this type and share accumulated gains or losses with these people.

If John and Matt still find the risk of loss to be too high, they can reduce the risk further by increasing the number of people with whom gains and losses are shared. Note however that if the number of bets remains at 100, then increasing the number of people sharing in the winnings will reduce the expected gain for each member.

*Comment:* A concrete application of the risk reduction technique presented in part c is presented in the solution to exercise 22, part c.

22. Let \( L_j \) be the loss incurred and \( kE[L_j] \) the amount of premium collected by insurer \( j \) where \( k > 1 \) and \( k \) is the same for both insurers. Under the proposed coinsurance arrangement, premiums and losses are shared equally. Hence each insurer pays \( L \) where \( L = \frac{L_1 + L_2}{2} \) and receives \( k(E[L_1] + E[L_2])/2 \). The losses \( L_j \) are assumed to be independent and identically distributed.

a. Since \( L = \frac{L_1 + L_2}{2} \) and the \( L_j \) are independent and identically distributed, we have

\[ E[L] = \frac{1}{2} (E[L_1] + E[L_2]) = \frac{1}{2} (2E[L_j]) = E[L_j] \]

and

\[ \text{Var}(L) = \frac{1}{4} (\text{Var}(L_1) + \text{Var}(L_2)) = \frac{1}{4} (2\text{Var}(L_j)) = \frac{1}{2} \text{Var}(L_j). \]
Hence
\[ \mu_L = \mu_{L_j}, \]
\[ \sigma_L = \frac{1}{\sqrt{2}} \sigma_{L_j}. \]

Consequently,
\[ \frac{\mu_L}{\sigma_L} = \sqrt{2} \left( \frac{\mu_{L_j}}{\sigma_{L_j}} \right). \]

b. In exercise 21, it is argued that the ratio \( \mu/\sigma \) can be considered a rudimentary measure of solvency with larger values of \( \mu/\sigma \) indicating a higher probability of solvency. From part a of the present exercise, we have
\[ \frac{\mu_L}{\sigma_L} = \sqrt{2} \left( \frac{\mu_{L_j}}{\sigma_{L_j}} \right) > \frac{\mu_{L_j}}{\sigma_{L_j}}. \]

Consequently, according to the risk measure \( \mu/\sigma \) the insurers are more secure with coinsurance.

We can also reach this conclusion by calculating insolvency probabilities directly. Since the insurers share the collected premiums equally, the probability that aggregate claims exceed aggregate premiums when insurer \( j \) participates in the coinsurance arrangement is
\[ \Pr[L > k \left( \frac{E[L_1] + E[L_2]}{2} \right)] = \Pr[L > k \mu_{L_j}]. \]

Under a normal approximation, the latter probability is
\[ \Pr\left[ L > k \mu_{L_j} \right] = \Pr\left[ \frac{L - \mu_L}{\sigma_L} > \frac{k \mu_{L_j} - \mu_L}{\sigma_L} \right] = \Pr\left[ \frac{L - \mu_L}{\sigma_L} > \frac{(k-1) \mu_{L_j}}{\sigma_{L_j} / \sqrt{2}} \right] \approx \Pr[Z > \sqrt{2} \left( k-1 \right) \frac{\mu_{L_j}}{\sigma_{L_j}}] = 1 - \Phi \left[ \sqrt{2} \left( k-1 \right) \frac{\mu_{L_j}}{\sigma_{L_j}} \right]. \]
Hence the probability of insolvency for insurer \( j \) when the coinsurance arrangement is in place is approximately

\[
1 - \Phi \left( \sqrt{2} \left( k - 1 \right) \frac{\mu_{L_j}}{\sigma_{L_j}} \right).
\]

From exercise 21, the probability of insolvency for insurer \( j \) when the coinsurance arrangement is not in place is approximately

\[
1 - \Phi \left( k - 1 \frac{\mu_{L_j}}{\sigma_{L_j}} \right).
\]

Since \( \Phi \) is increasing and \( k > 1 \) we have

\[
1 - \Phi \left( \sqrt{2} \left( k - 1 \right) \frac{\mu_{L_j}}{\sigma_{L_j}} \right) < 1 - \Phi \left( k - 1 \frac{\mu_{L_j}}{\sigma_{L_j}} \right).
\]

Consequently, the probability of insolvency is lower when the coinsurance arrangement is in place. Hence the insurers are more secure with coinsurance, as argued earlier.

c. In the course of doing business, insurance companies naturally accumulate pools of risk, some of which can be large. Pooling actually increases risk. To understand why, consider \( n \) independent, identically distributed losses and let \( S = X_1 + \cdots + X_n \) be the aggregate loss. Using standard properties of the variance we have

\[
\text{Var}(S) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) = n \text{ Var}(X_j)
\]

from which it follows that

\[
\text{Var}(S) \to \infty \quad \text{as} \quad n \to \infty.
\]

This shows that the uncertainty in the payout associated with a pool of risks increases without bound as the number of policies in the pool increases, i.e., pooling increases risk.

Now the basic principle of insurance is that risk can be reduced through sharing. The mathematical justification of this principle follows from the law of large numbers (see sections 1.4 and 8.4.3 of the textbook). One way that an insurer can share the risk associated with a pool of risks is to enter into a coinsurance agreement with other
insurers. This is the approach considered in part b of this exercise. Another way that an insurer can share the risk is to sell shares to the public.

A share in a company represents an ownership position in the company and entitles the holder of the share to a portion of the future profits generated by the company. Since future profits are unpredictable, the value of a company's shares will fluctuate over time, generally increasing during periods of rising profits and decreasing during periods of falling profits. When an insurance company experiences a large unexpected loss, profits are generally impacted in a negative way and the value of the company declines. For a company with shares outstanding this decline in value will be reflected in a lower trading price for the company's shares. In effect, the unexpected loss will be shifted from the company to the shareholders.

Now if the number of shareholders is large then the amount of the loss shifted to an individual shareholder (through a reduction in share price) will be relatively small and limited to the value of the individual's shares. Moreover the risk of loss for an individual shareholder will be much less than the risk associated with the entire pool accumulated by the company. Indeed, if $S$ is the loss on the pool at the company level and this loss is spread equally over the $m$ shares outstanding then the risk of loss on an individual share as given by the variance is

$$\text{Var} \left( \frac{S}{m} \right) = \frac{1}{m^2} \text{Var} (S) \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$ 

Consequently, the risk to individual shareholders is relatively small and approaches 0 as the number of shareholders becomes arbitrarily large.

It follows from these observations that selling shares to the public is an effective way for an insurance company to disperse the risk associated with large risk pools. Selling shares also makes it possible for individuals of relatively modest means to share in the often generous profits associated with underwriting large risk pools without having to assume an unacceptably high level of risk.

24. Let $X_j$ be the payout on the $j$-th policy during the coming year and let $S$ be the aggregate payout during the coming year on the portfolio of $m$ policies. Assume that all death claim payments are made at the end of the year. Then from the given information, the $X_j$ are independent and identically distributed with distribution given by
\[ X_j = \begin{cases} 25,000 \, e^{-0.05} & \text{with probability .01,} \\ 0 & \text{with probability .99,} \end{cases} \]

and \( S = X_1 + \cdots + X_m \). Note that \( 25,000 \, e^{-0.05} \) is the present value of 25,000 discounted one year at a continuously compounded rate of 5% per annum. Let \( P \) be the per-policy premium that must be charged for the insurer to be 99% confident that total premiums exceed total claims. Then \( P \) is determined by the condition
\[ \Pr[S \leq m \, P] = .99. \]

Now if \( m \) is sufficiently large, then the distribution of \( S \) is approximately normal with mean
\[ E[S] = m \, E[X_j] \]
and variance
\[ \operatorname{Var}(S) = m \, \operatorname{Var}(X_j). \]

From the information on the distribution of \( X_j \) stated earlier we have
\[ E[X_j] = (25,000 \, e^{-0.05}) \cdot (.01) \approx 237.81 \]
and
\[ \operatorname{Var}(X_j) = E[X_j^2] - (E[X_j])^2 = (25,000 \, e^{-0.05})^2 \cdot (.01) - (25,000 \, e^{-0.05} \cdot .01)^2 = (25,000)^2 \, e^{-10} \cdot .01 \cdot .99 \approx (2366.15)^2. \]

Hence for \( m \) sufficiently large, the distribution of \( S \) is approximately normal with
\[ E[S] \approx 237.81 \, m \]
and
\[ \operatorname{Var}(S) \approx (2366.15)^2 \, m. \]

Assuming this to be the case, the condition
\[ \Pr[S \leq m \, P] = .99 \]

is approximately equivalent to the condition
where $Z \sim \text{Normal}(0, 1)$. In terms of the standard normal distribution function, the latter condition is

$$\Phi \left[ \frac{P - 237.81}{2366.15 \sqrt{m}} \right] = .99.$$  

Now from Appendix E of the textbook, we have $\Phi[2.326] \approx .99$. Hence the condition

$$\Phi \left[ \frac{P - 237.81}{2366.15 \sqrt{m}} \right] = .99$$  

is equivalent to

$$\frac{P - 237.81}{2366.15 \sqrt{m}} \approx 2.326.$$  

Solving the latter equation for $P$ we obtain

$$P \approx 237.81 + \frac{5503.67}{\sqrt{m}}.$$  

Therefore to be 99% confident that total premiums exceed total claims, the insurer must charge each of the $m$ policyholders the dollar amount

$$237.81 + \frac{5503.67}{\sqrt{m}}.$$  

Note that the $237.81$ figure here is the expected loss per policy and the $\frac{5503.67}{\sqrt{m}}$ figure is the additional per policy amount that must be collected to ensure with 99% confidence that claims are covered in the aggregate. The additional per-policy amount is not insignificant (unless $m$ is more than 1 million). The reason for this is that claim payments, while unlikely (1% probability of occurrence), are quite large ($25,000) when they occur.
27. Let \( X_j \) be the loss on the \( j \)-th contract measured in thousands of dollars and let \( S \) be the aggregate loss in thousands of dollars. By assumption, the \( X_j \) are independent and identically distributed with \( X_j \sim \text{Exponential}(1/5) \) and \( S = X_1 + \cdots + X_{100} \). Let \( P \) be the premium in thousands of dollars collected on each contract. By assumption, \( P = 5.05 \).

We are required to determine \( \Pr[S > 100P] \) using a normal power approximation for \( S \). In order to apply the normal approximation formula to \( S \) we need to determine values for \( \mu_S, \sigma_S, \) and \( \gamma_S \). From sections 4.2.1, 4.2.2, and 4.2.3 of the textbook, the mean, variance, and skewness of a sum \( S \) of independent random variables \( X_1, \ldots, X_n \) are respectively

\[
\mu_S = \mu_{X_1} + \cdots + \mu_{X_n},
\]

\[
\sigma_S^2 = \sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2,
\]

\[
\gamma_S = \frac{\gamma_{X_1} \sigma_{X_1}^3 + \cdots + \gamma_{X_n} \sigma_{X_n}^3}{\sigma_S^3}.
\]

When the \( X_j \) are identically distributed, as is the case in this question, these formulas reduce to

\[
\mu_S = n \mu_X,
\]

\[
\sigma_S^2 = n \sigma_X^2,
\]

\[
\gamma_S = \frac{n \gamma_X \sigma_X^3}{\sigma_S^3} = \frac{n \gamma_X \sigma_X^3}{n^{3/2} \sigma_X^3} = \frac{\gamma_X}{n^{1/2}}
\]

where \( X \sim X_j \). Now for \( X \sim \text{Exponential}(\lambda) \) we have

\[
\mu_X = \frac{1}{\lambda},
\]

\[
\sigma_X^2 = \frac{1}{\lambda^2},
\]
\( \gamma_X = 2 \)

(see section 6.1.1 of the textbook). Consequently, the mean, variance, and skewness for the aggregate loss random variable \( S \) of this question are given by

\[
\begin{align*}
\mu_S &= 100 \mu_X = 100 \left( \frac{1}{\lambda} \right) = 100 \left( \frac{1}{1/5} \right) = 500, \\
\sigma_S^2 &= 100 \sigma_X^2 = 100 \left( \frac{1}{\lambda^2} \right) = 100 \cdot \frac{1}{(1/5)^2} = 2500, \\
\gamma_S &= \frac{\gamma_X}{100^{1/2}} = \frac{2}{10} = 0.2.
\end{align*}
\]

Using these values of \( \mu_S, \sigma_S, \) and \( \gamma_S \) in the normal power approximation formula given in section 8.6 we find that

\[
\begin{align*}
\Pr[S \leq 100 P] &= \Pr[S \leq 505] = F_S[505] \approx \Phi \left[ \frac{3}{\gamma_S} + \sqrt{\frac{9}{\gamma_S} + 1 + \frac{6}{\gamma_S} \cdot \frac{505 - 500}{\sigma_S}} \right] \\
&= \Phi \left[ -(3)(5) + \sqrt{(9)(25) + 1 + (6)(5) \cdot \frac{505 - 500}{50}} \right] \\
&= \Phi[15 + \sqrt{229}] \approx \Phi[1.3274595] \approx \Phi[1.327].
\end{align*}
\]

From the table for the standard normal distribution function in Appendix E of the textbook we have

\( \Phi[1.327] \approx 0.73 \, \Phi[.13] + .27 \Phi[.14] = 0.73 \cdot 0.5517 + 0.27 \cdot 0.5557 = 0.55278. \)

Hence

\( \Pr[S \leq 100 P] \approx 0.55278. \)

Consequently, the probability that aggregate losses exceed total premiums collected is

\( \Pr[S > 100 P] = 1 - \Pr[S \leq 100 P] \approx 1 - 0.55278 \approx 0.44722 \)

using a normal power approximation.
Comment: It is actually possible to specify the exact distribution of \( S \) in this question. Indeed, since \( S \) is a sum of independent, identically distributed exponential random variables, \( S \) must have a gamma distribution. In particular, \( S \sim \text{Gamma}(100, 1/5) \). Using this exact distribution and the fact that the survival function of a gamma distribution whose parameter \( r \) is a positive integer has the form

\[
S_T[t] = \sum_{n=0}^{r-1} \frac{(\lambda t)^n}{n!} e^{-\lambda t}
\]

(see section 6.1.2 of the textbook), it is possible to calculate the exact value of the desired probability. Indeed,

\[
\Pr[S > 505] = \sum_{n=0}^{99} \frac{((1/5) (505))^n}{n!} e^{-(1/5) (505)} = \sum_{n=0}^{99} \frac{101^n}{n!} e^{-101}
\]

The numerical value of this probability can be determined using Mathematica or similar computer software. When we do this we find that

\[
\sum_{n=0}^{99} \frac{101^n}{n!} e^{-101} \approx 0.447104.
\]

Comparing this value to the value \( .44722 \) of \( \Pr[S > 505] \) that was previously determined using a normal power approximation, we see that the normal power approximation is quite close. The advantage of using a normal power approximation instead of the exact distribution to calculate probabilities (even in problems such as the one just considered where it is relatively straightforward to determine the exact distribution) is that the normal power approximation only requires the use of a hand-held calculator and a table of values for the standard normal distribution function whereas the exact distribution may require more sophisticated computer software such as Mathematica.
Chapter Nine Solutions

1. For a given throw of the dart, let \( X \) be the distance from the fixed origin to the place where the dart lands and let \( Y \) be the number of the target at which the dart is aimed. Suppose that the target number is selected by the roll of a die.

   a. \( E[X | Y] \) is the average distance from the origin of the darts thrown at target \( Y \). (If the target shooter is fairly accurate, this should be close to the distance from the origin to the location of target \( Y \).)

   \( \text{Var}(X | Y) \) is the variability in the distance from the origin of the darts thrown at target \( Y \). (Note that this is not necessarily a good measure of the error in the target shooter's aim: If the darts all land far from the target but close to each other then \( \text{Var}(X | Y) \) will be small even though the error associated with each throw is large.)

   \( E[E[X | Y]] \) is the average of the average distances weighted according to the number of darts aimed at each target.

   \( E[\text{Var}(X | Y)] \) is the average variability in hitting the selected target (whatever it happens to be) weighted according to the number of darts aimed at each target.

   \( \text{Var}(E[X | Y]) \) is the variability in the average dart distances for the targets weighted according to the number of darts aimed at each target. (If the target shooter is fairly accurate then this is a measure of the proximity of the target locations.)

   b. The formula \( E[X] = E[E[X | Y]] \) has the following interpretation: The average distance from the origin of all darts tossed can be determined by calculating the average distance from the origin for darts thrown at each specific target and then averaging these values according to the frequency with which each target is selected.

   The formula \( \text{Var}(X) = E[\text{Var}(X | Y)] + \text{Var}(E[X | Y]) \) has the following interpretation:
The variability in the distance from the origin of all darts tossed is equal to the sum of the average variability in hitting the selected target and the variability in the average distances for each target. (If the target shooter is fairly accurate then this formula asserts that the variability in dart locations for all darts is determined by the variability around each target and the spread of the target locations.)

3. Let $C$ be a risk classification variable defined as follows:

$$C = 1 \quad \text{if client has no credit history,}$$
$$C = 2 \quad \text{if client has history of late payments,}$$
$$C = 3 \quad \text{if client has declared personal bankruptcy.}$$

Let $L_j$ be the loss on a randomly selected account for which it is known that $C = j$. We are given that

$$L_1 \sim \text{Exponential}(1/100),$$
$$L_2 \sim \text{Exponential}(1/500),$$
$$L_3 \sim \text{Exponential}(1/1000),$$

and the distribution of $C$ is

$$C = 1 \quad \text{with probability .20,}$$
$$C = 2 \quad \text{with probability .50,}$$
$$C = 3 \quad \text{with probability .30.}$$

Let $L$ denote the loss on a randomly selected account for which the risk class is not known. Parts a through d are concerned with the distribution of $L$.

a. The expected loss on a randomly selected account can be determined using the formula for unconditional expectation given in section 9.3 of the textbook,

$$E[L] = E_C[E[L \mid C]].$$

From the given information,

$$E[L \mid C] = E[L_1] \quad \text{with probability .20,}$$
1. From the given information, 
\[ E[L \mid C] = E[L_2] \] with probability .50, 

Moreover, \( E[L_1] = 100, E[L_2] = 500, E[L_3] = 1000 \). Hence 
\[
E[E[L \mid C]] = E[L_1](.20) + E[L_2](.50) + E[L_3](.30) = (100)(.20) + (500)(.50) + (1000)(.30) = 570.
\]

Consequently, the expected loss on a randomly selected account is 
\[ E[L] = E[E[L \mid C]] = 570. \]

b. The variance of the loss on a randomly selected account can be determined using the formula for unconditional variance given in section 9.3 of the textbook, 
\[ \text{Var}(L) = E[\text{Var}(L \mid C)] + \text{Var}(E[L \mid C]). \]

From the given information,
\[ \text{Var}(L \mid C) = \text{Var}(L_1) \] with probability .20, 
\[ \text{Var}(L \mid C) = \text{Var}(L_2) \] with probability .50, 
\[ \text{Var}(L \mid C) = \text{Var}(L_3) \] with probability .30, 

and
\[ E[L \mid C] = E[L_1] \] with probability .20, 
\[ E[L \mid C] = E[L_2] \] with probability .50, 

Moreover, since \( L_1 \sim \text{Exponential}(1/100), L_2 \sim \text{Exponential}(1/500), \) and \( L_3 \sim \text{Exponential}(1/1000) \) we have 
\[ E[L_1] = 100, \ E[L_2] = 500, \ E[L_3] = 1000 \]

and 
\[ \text{Var}(L_1) = 100^2, \ \text{Var}(L_2) = 500^2, \ \text{Var}(L_3) = 1000^2. \]
Hence

\[
E[\text{Var}(L \mid C)] = \text{Var}(L_1)(.20) + \text{Var}(L_2)(.50) + \text{Var}(L_3)(.30) = (100)^2(.20) + (500)^2(.50) + (1000)^2(.30) = 427,000,
\]

\[
E[\text{Var}(L \mid C)]^2 = E[L_1]^2(.20) + E[L_2]^2(.50) + E[L_3]^2(.30) = (100)^2(.20) + (500)^2(.50) + (1000)^2(.30) = 427,000,
\]

and

\[
E[E[L \mid C]] = E[L_1](.20) + E[L_2](.50) + E[L_3](.30) = (100)(.20) + (500)(.50) + (1000)(.30) = 570.
\]

Consequently,

\[
\text{Var}(E[L \mid C]) = E[\text{Var}(L \mid C)]^2 - (E[E[L \mid C]])^2 = 427,000 - (570)^2 = 102,100
\]

and so

\[
\text{Var}(L) = E[\text{Var}(L \mid C)] + \text{Var}(E[L \mid C]) = 427,000 + 102,100 = 529,100.
\]

Therefore the variance of the loss on a randomly selected account is 529,100 squared dollars. This is equivalent to a standard deviation of about $727.39.

**Important Comment:** Note that in this particular question it just so happens that the values of \(E[\text{Var}(L \mid C)]\) and \(E[E[L \mid C]]^2\) are the same. However, the reader should keep in mind that the values of \(E[\text{Var}(L \mid C)]\), \(E[E[L \mid C]]^2\) are generally different. The reason they are the same in this question is that the exponential distribution has the property that its variance is equal to the square of its mean. Indeed for \(X \sim \text{Exponential}(\lambda)\), \(E[X] = 1/\lambda\) and \(\text{Var}(X) = 1/\lambda^2 = E[X]^2\).

c. The desired probability can be determined using the law of total probability. Indeed, for any \(l > 0\) we have

\[
\text{Pr}[L > l] = \text{Pr}[L > l \mid C = 1] \text{Pr}[C = 1] + \text{Pr}[L > l \mid C = 2] \text{Pr}[C = 2] + \text{Pr}[L > l \mid C = 3] \text{Pr}[C = 3] = \text{Pr}[L_1 > l] \text{Pr}[C = 1] + \text{Pr}[L_2 > l] \text{Pr}[C = 2] + \text{Pr}[L_3 > l] \text{Pr}[C = 3].
\]

From the given information,

\[
L_1 \sim \text{Exponential}(1/100),
\]
\( L_2 \sim \text{Exponential}(1/500), \)
\( L_3 \sim \text{Exponential}(1/1000) \)

and
\[
\Pr[C = 1] = .20, \quad \Pr[C = 2] = .50, \quad \Pr[C = 3] = .30.
\]

Hence
\[
\Pr[L > l] = e^{-l/100} (.20) + e^{-l/500} (.50) + e^{-l/1000} (.30) \quad \text{for } l > 0.
\]

Consequently, the probability that the company loses more than $500 on a randomly selected account is
\[
\Pr[L > 500] = e^{-5} (.20) + e^{-1} (.50) + e^{-1/2} (.30) \approx .36724651 \approx 37 \%.
\]

d. From part a, the expected loss on a randomly selected account is $570 and from part b the standard deviation of the loss is $727.39. Hence if the finance company wishes to recoup its expected loss and have a safety margin of 1 standard deviation for each customer, it should charge each client an administration fee of $1297.39 (= $570 + $727.39). From a practical viewpoint, it is probably not possible to charge a fee of this size upfront since people with a poor or nonexistent credit history are unlikely to have that much cash lying around. However, this fee could be recouped by increasing the interest rate on the loan and assessing periodic "hidden" fees that are less visible to the borrower.

Comment: As a check on our calculations in parts a and b, we can use the survival function determined in part c to calculate \( E[L] \) and \( \text{Var}(L) \) directly, i.e., without making use of the formulas for unconditional mean and variance given in section 9.3 of the textbook. From part c, the survival function of \( L \) is
\[
S_L[l] = .20 \, e^{-l/100} + .50 \, e^{-l/500} + .30 \, e^{-l/1000} \quad \text{for } l > 0.
\]

Hence \( L \) is a mixture of \( \text{Exponential}(1/100) \), \( \text{Exponential}(1/500) \), and \( \text{Exponential}(1/1000) \) with respective mixing weights .20, .50, and .30. It follows from section 9.4 that the moment generating function of \( L \) is
\[
M_L[t] = .20 \frac{1}{1 - 100 \, t} + .50 \frac{1}{1 - 500 \, t} + .30 \frac{1}{1 - 1000 \, t} \quad \text{for } t < \frac{1}{1000}.
\]
(Recall from section 6.1.1 that the moment generating function of \( X \sim \text{Exponential}(\lambda) \) is \( M_X(t) = \lambda/(\lambda - t) \) for \( t < \lambda \). Consequently, using the fact that \( E[L^k] = M'_k[0] \) for \( k = 1, 2, \ldots \) (see section 4.3.1 of the textbook) we have
\[
E[L] = M'_1[0] = \left[ (.20) (1 - 100 t)^{-2} (100) + (.50) (1 - 500 t)^{-2} (500) + (.30) (1 - 1000 t)^{-2} (1000) \right] |_{t=0} = 570
\]
and
\[
E[L^2] = M'_2[0] = \left[ (.20) (2) (1 - 100 t)^{-3} (100)^2 + (.50) (2) (1 - 500 t)^{-3} (500)^2 + (.30) (2) (1 - 1000 t)^{-3} (1000)^2 \right] |_{t=0} = 854,000.
\]
Therefore
\[
E[L] = 570
\]
and
\[
\text{Var}(L) = E[L^2] - E[L]^2 = 854,000 - (570)^2 = 529,100,
\]
which are identical to the values given in parts a and b, as they should be.

5. Let \( L \) be the size of the uncertain loss and let \( K \) be the indemnified amount.

a. Suppose that the insurer caps the indemnified amount at \( m \). Then
\[
K = \min(L, m) = \begin{cases} L & \text{if } L \leq m, \\ m & \text{if } L > m. \end{cases}
\]
The distribution function for \( K \) can be determined by applying the law of total probability with the conditions \( L \leq m \) and \( L > m \). Indeed, for any \( k \) we have
\[
\Pr[K \leq k] = \Pr[K \leq k \mid L \leq m] \Pr[L \leq m] + \Pr[K \leq k \mid L > m] \Pr[L > m] = \\
\Pr[L \leq k \mid L \leq m] \Pr[L \leq m] + \Pr[m \leq k \mid L > m] \Pr[L > m] = \\
\frac{\Pr[L \leq m]}{\Pr[L \leq \min(m, k)] + \Pr[m \leq k \text{ and } L > m]} \Pr[L > m].
\]
Hence for \( k < m \) we have
\[
\Pr[K \leq k] = \Pr[L \leq \min(m, k)] + \Pr[m \leq k \text{ and } L > m] = \Pr[L \leq k] + 0 = \Pr[L \leq k]
\]
and for \( k \geq m \) we have
\[
\Pr[K \leq k] = \Pr[L \leq \min(m, k)] + \Pr[m \leq k \text{ and } L > m] = \Pr[L \leq m] + \Pr[L > m] = 1.
\]
Consequently, the distribution function of \( K \) is given by
\[
F_K[k] = \begin{cases} 
F_L[k] & \text{for } k < m, \\
1 & \text{for } k \geq m.
\end{cases}
\]

b. Suppose that the indemnified amount is subject to a deductible \( d \) but no cap. Then
\[
K = \begin{cases} 
0 & \text{if } L \leq d, \\
L - d & \text{if } L > d.
\end{cases}
\]
The distribution function for \( K \) can be determined by applying the law of total probability with the conditions \( L \leq d \) and \( L > d \). Indeed, for any \( k \) we have
\[
\Pr[K \leq k] = \Pr[K \leq k \mid L \leq d] \Pr[L \leq d] + \Pr[K \leq k \mid L > d] \Pr[L > d] = \\
\frac{\Pr[0 \leq k \mid L \leq d] \Pr[L \leq d] + \Pr[L - d \leq k \mid L > d] \Pr[L > d]}{\Pr[L \leq d] + \Pr[L > d]} = \\
\frac{\Pr[L \leq d]}{\Pr[L \leq d] + \Pr[L > d]} \Pr[L > d] + \frac{\Pr[d < L \leq d + k]}{\Pr[L > d]} \Pr[L > d] = \\
\Pr[0 \leq k \text{ and } L \leq d] + \Pr[d < L \leq d + k].
\]
Hence for \( k \geq 0 \) we have
\[
\Pr[K \leq k] = \Pr[0 \leq k \text{ and } L \leq d] + \Pr[d < L \leq d + k] = \\
\Pr[L \leq d] + \Pr[d < L \leq d + k] = \Pr[L \leq d + k] = F_L[d + k]
\]
and for \( k < 0 \) we trivially have
\[
\Pr[K \leq k] = \Pr[0 \leq k \text{ and } L \leq d] + \Pr[d < L \leq d + k] = 0.
\]
Consequently, the distribution function of \( K \) is given by
\[
F_K[k] = \begin{cases} 
F_L[d + k] & \text{for } k \geq 0, \\
0 & \text{for } k < 0.
\end{cases}
\]

c. Suppose that the indemnified amount is subject to both a deductible \( d \) and a cap \( m \). Then
\( K = 0 \) if \( L \leq d \),
\( K = L - d \) if \( d < L \leq d + m \),
\( K = m \) if \( L > d + m \).

The distribution function for \( K \) can be determined by applying the law of total probability with the conditions \( L \leq d, d < L \leq d + m, L > d + m \). Indeed for any \( k \) we have
\[
\Pr[K \leq k] = \Pr[K \leq k \mid L \leq d] \Pr[L \leq d] + \\
\Pr[K \leq k \mid d < L \leq d + m] \Pr[d < L \leq d + m] + \Pr[K \leq k \mid L > d + m] \Pr[L > d + m] = \\
\Pr[0 \leq k \mid L \leq d] \Pr[L \leq d] + \Pr[L - d \leq k \mid d < L \leq d + m] \Pr[d < L \leq d + m] + \\
\Pr[m \leq k \mid L > d + m] \Pr[L > d + m] = \\
\Pr[0 \leq k \text{ and } L \leq d] + \Pr[L - d \leq k \text{ and } d < L \leq d + m] + \Pr[m \leq k \text{ and } L > d + m] = \\
\Pr[0 \leq k \text{ and } L \leq d] + \Pr[d < L \leq \min (d + k, d + m)] + \Pr[m \leq k \text{ and } L > d + m].
\]

Hence for \( 0 \leq k < m \) we have
\[
\Pr[K \leq k] = \\
\Pr[0 \leq k \text{ and } L \leq d] + \Pr[d < L \leq \min (d + k, d + m)] + \Pr[m \leq k \text{ and } L > d + m] = \\
\Pr[L \leq d] + \Pr[d < L \leq d + k] + 0 = \Pr[L \leq d + k],
\]
for \( k \geq m \) we have
\[
\Pr[K \leq k] = \\
\Pr[0 \leq k \text{ and } L \leq d] + \Pr[d < L \leq \min (d + k, d + m)] + \Pr[m \leq k \text{ and } L > d + m] = \\
\Pr[L \leq d] + \Pr[d < L \leq d + m] + \Pr[L > d + m] = 1,
\]
and for \( k < 0 \) we have
\[
\Pr[K \leq k] = \Pr[0 \leq k \text{ and } L \leq d] + \\
\Pr[d < L \leq \min (d + k, d + m)] + \Pr[m \leq k \text{ and } L > d + m] = 0 + 0 + 0 = 0.
\]

Consequently, the distribution function of \( K \) is given by
\[
F_X[k] = 0 \quad \text{for } k < 0, \\
F_X[k] = F_X[d + k] \quad \text{for } 0 \leq k < m, \\
F_X[k] = 1 \quad \text{for } k \geq m.
\]

7. Suppose that \( X = BI \), where \( I \sim \text{Bernoulli}(p) \) and \( B \) is a nonnegative random variable which is independent of \( I \).
a. The distribution function of $X$ can be determined by conditioning on $I$ and using the law of total probability as follows:

$$F_X(x) = \Pr[X \leq x] = \Pr[X \leq x \mid I = 0] \Pr[I = 0] + \Pr[X \leq x \mid I = 1] \Pr[I = 1] =$$

$$\Pr[B I \leq x \mid I = 0] \Pr[I = 0] + \Pr[B I \leq x \mid I = 1] \Pr[I = 1] =$$

$$\Pr[0 \leq x] \cdot (1 - p) + \Pr[B \leq x] \cdot p.$$

Note that $\Pr[B I \leq x \mid I = 1] = \Pr[B \leq x]$ since $B$ is independent of $I$. Hence for $x \geq 0$

$$F_X(x) = \Pr[0 \leq x] \cdot (1 - p) + \Pr[B \leq x] \cdot p = (1 - p) + p F_B(x),$$

and for $x < 0$

$$F_X(x) = \Pr[0 \leq x] \cdot (1 - p) + \Pr[B \leq x] \cdot p = 0 \cdot (1 - p) + 0 \cdot p = 0.$$

(Note that $\Pr[B \leq x] = 0$ for $x < 0$ since $B$ is assumed to be nonnegative.) Consequently, the distribution of $X$ is given by

$$F_X(x) = \begin{cases} (1 - p) + p F_B(x) & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases}$$

as required. Since $S_X(x) = 1 - F_X(x)$ for all $x$, it follows that the survival function of $X$ is

$$S_X(x) = \begin{cases} 1 - ((1 - p) + p F_B(x)) & \text{for } x \geq 0, \\ 1 - 0 & \text{for } x < 0, \end{cases}$$

$$= \begin{cases} p (1 - F_B(x)) & \text{for } x \geq 0, \\ 1 & \text{for } x < 0, \end{cases} = \begin{cases} p S_X(x) & \text{for } x \geq 0, \\ 1 & \text{for } x < 0, \end{cases}$$

as required.

In general, $X$ has a mixed distribution with a probability mass at $x = 0$. When $B$ is continuous and $p < 1$, $X$ has a probability mass at $x = 0$ and a continuous distribution of probability on the interval $x > 0$. Sample graphs of $F_X$ and $S_X$ when $B$ is continuous can be created using Mathematica or similar computer software.
b. The graph of the generalized density $f_X$ corresponding to the graphs of $F_X$ and $S_X$ given in part a is as follows:
Note that the area under the continuous part of the density is equal to $p$. The probability mass at $x = 0$ is equal to the probability that $B = 0$ or $I = 0$. Since $B$ is assumed to be continuous, $\Pr[B = 0] = 0$ and so $\Pr[X = 0] = \Pr[I = 0] = 1 - p$.

c. Applying the formula for unconditional expectation given in section 9.3 we have

$$E[X] = E[I E[X | I]] = E[X | I = 0] \Pr[I = 0] + E[X | I = 1] \Pr[I = 1] =$$
$$E[I B | I = 0] \Pr[I = 0] + E[I B | I = 1] \Pr[I = 1] =$$
$$E[0] \Pr[I = 0] + E[B] \Pr[I = 1] = 0 \cdot (1 - p) + \mu_B p = p \mu_B$$

as required. Note that

$$E[I B | I = 1] = E[B | I = 1] = E[B]$$

since $B$ is independent of $I$.

To determine the formula for $\text{Var}(X)$ we apply the formula for unconditional variance which in this context has the form

$$\text{Var}(X) = E[I \text{Var}(X | I)] + \text{Var}_I(E[X | I]).$$

Arguing as before we have
\[ E[\text{Var}(X | I)] = \text{Var}(X | I = 0) \Pr[I = 0] + \text{Var}(X | I = 1) \Pr[I = 1] = \\
\text{Var}(B | I = 0) \Pr[I = 0] + \text{Var}(B | I = 1) \Pr[I = 1] = \\
\text{Var}(0) \Pr[I = 0] + \text{Var}(B) \Pr[I = 1] = 0 \cdot (1 - p) + \sigma_B^2 p = p \sigma_B^2. \]

Since
\[ E[I | X] = \begin{cases} 
0 & \text{if } I = 0, \\
\mu_B & \text{if } I = 1,
\end{cases} \]
we also have
\[ \text{Var}_I(E[X | I]) = \mu_B^2 \Pr[I = 1] \Pr[I = 0] = \mu_B^2 p (1 - p). \]

Hence the unconditional variance of \( X \) is given by
\[ \text{Var}(X) = E[I \text{Var}(X | I)] + \text{Var}_I(E[X | I]) = p \sigma_B^2 + \mu_B^2 p (1 - p) \]
as required.

d. The moment generating function of \( X \) is defined by
\[ M_X[t] = E[e^{tX}]. \]

Using the formula for unconditional expectation given in section 9.3 of the textbook we have
\[ E[e^{tX}] = E[E[e^{tX} | I]] = E[e^{tX} | I = 0] \Pr[I = 0] + E[e^{tX} | I = 1] \Pr[I = 1] = \\
E[e^{tB} | I = 0] \Pr[I = 0] + E[e^{tB} | I = 1] \Pr[I = 1] = \\
E[e^0] \Pr[I = 0] + E[e^{tB}] \Pr[I = 1] = 1 \cdot (1 - p) + M_B[t] p. \]

Consequently,
\[ M_X[t] = (1 - p) + p M_B[t] \]
for all \( t \) where \( M_B \) is defined. Differentiating this equation \( k \) times we obtain
\[ M_X^{(k)}[t] = p M_B^{(k)}[t]. \]

Hence using the fact that \( E[X^k] = M_X^{(k)}[0] \) for any random variable \( X \) we have
\[ E[X^k] = p E[B^k] \quad \text{for } k = 1, 2, \ldots. \]
as required.

e. The third central moment can be calculated in terms of the moments about 0 as follows:

\[ E[(X - \mu_X)^3] = E[X^3] - 3 E[X^2] E[X] + 2 E[X]^3 \]

(see exercise 5 part a of section 4.3 or simply expand the trinomial \((X - \mu_X)^3\) and use the linearity property of expectation). From part d of the current exercise,

\[ E[X^k] = p E[B^k] \quad \text{for } k = 1, 2, \ldots. \]

Hence substituting this expression for \(E[X^k]\) into the preceding formula for \(E[(X - \mu_X)^3]\) we obtain

\[
\]

\[
\]

as required.

A formula for \(E[(X - \mu_X)^3]\) in terms of the statistics \(\mu_B, \sigma_B,\) and \(\gamma_B\) can also be given.

Using the general relationships


\[ E[(B - \mu_B)^3] = \gamma_B \sigma_B^3, \]

\[ \gamma_B^2 = E[B^2] - E[B]^2, \]

we have

\[ E[B^3] = \gamma_B \sigma_B^3 + \mu_B^3 \]

and

\[
\]

\[
\gamma_B \sigma_B^3 + 3 (\gamma_B^2 + \mu_B^3) \mu_B - 2 \mu_B^3 = \gamma_B \sigma_B^3 + 3 \sigma_B^2 \mu_B^3 + \mu_B^3.
\]

Substituting these formulas for \(E[B^2]\) and \(E[B^3]\) into the formula for \(E[(X - \mu_X)^3]\)
derived earlier we obtain


\[ p \left( \gamma_B \sigma_B^3 + 3 \, \sigma_B \mu_B + \mu_B^3 \right) - 3 \, p^2 \left( \sigma_B^2 + \mu_B^2 \right) \mu_B + 2 \, p^3 \, \mu_B^3 = \]

\[ p \, \gamma_B \sigma_B^3 + \left( 3 \, p - 3 \, p^2 \right) \sigma_B^2 \mu_B + \left( p - 3 \, p^2 + 2 \, p^3 \right) \mu_B^3 = \]

\[ p \, \gamma_B \sigma_B^3 + 3 \, p \, (1 - p) \, \sigma_B^2 \mu_B + p \, (1 - p) \, (1 - 2 \, p) \, \mu_B^3 \]

as required.

Comment: The assumption that \( B \) and \( I \) are independent was inadvertently omitted in the statement of the question in section 9.6. However, this assumption is necessary for the formulas just derived to hold in general. To see that this is so consider the case \( B = I \) (perfect dependence). Then \( X = I^2 \) and so the distribution of \( X \) is Bernoulli\((p)\),

\[
X = \begin{cases} 
1 & \text{with probability } p, \\
0 & \text{with probability } 1 - p.
\end{cases}
\]

Hence

\[
F_X[x] = \begin{cases} 
0 & \text{for } x < 0, \\
1 - p & \text{for } 0 \leq x < 1, \\
1 & \text{for } x \geq 1,
\end{cases}
\]

which is clearly different from the formula one gets in part a of this exercise. Indeed, substituting the distribution function \( F_B \) when \( B = I \) into the formula in part a we obtain

\[
F_X[x] = \begin{cases} 
0 & \text{for } x < 0, \\
1 - p^3 & \text{for } 0 \leq x < 1, \\
1 & \text{for } x \geq 1.
\end{cases}
\]

Consequently for the formulas in parts a through e to hold in general, we require that \( B \) and \( I \) be independent.

12. Let \( X_j \) be the payout on a randomly selected policy of type \( j \), let \( I_j \) be an indicator of a claim occurrence on a randomly selected policy of type \( j \), and let \( L_j \) be the claim size on
Let \( X_j \) be the payout on a randomly selected policy of type \( j \), let \( I_j \) be an indicator of a claim occurrence on a randomly selected policy of type \( j \), and let \( L_j \) be the claim size on a policy of type \( j \) for which a claim is known to have occurred. Then from the given information,

\[
I_1 = \begin{cases} 
1 & \text{with probability 25 \%}, \\
0 & \text{with probability 75 \%},
\end{cases}
\]

\[
I_2 = \begin{cases} 
1 & \text{with probability 40 \%}, \\
0 & \text{with probability 60 \%},
\end{cases}
\]

\[
X_1 = \begin{cases} 
L_1 & \text{if } I_1 = 1, \\
0 & \text{if } I_1 = 0,
\end{cases}
\]

\[
X_2 = \begin{cases} 
L_2 & \text{if } I_2 = 1, \\
0 & \text{if } I_2 = 0,
\end{cases}
\]

and the probability mass functions of \( L_1 \) and \( L_2 \) are respectively

\[
p_{L_1}[1000] = .20, \quad p_{L_1}[5000] = .50, \quad p_{L_1}[10000] = .30
\]

and

\[
p_{L_2}[1000] = .70, \quad p_{L_2}[5000] = .20, \quad p_{L_2}[10000] = .10.
\]

Let \( Y \) be the payout on a randomly selected policy whose type is not known and let \( C \) be an indicator of type defined as

\[
C = \begin{cases} 
1 & \text{if policyholder is of type 1}, \\
2 & \text{if policyholder is of type 2}.
\end{cases}
\]

Then from the given information,

\[
C = \begin{cases} 
1 & \text{with probability 30 \%}, \\
2 & \text{with probability 70 \%},
\end{cases}
\]

and

\[
Y = \begin{cases} 
X_1 & \text{if } C = 1, \\
X_2 & \text{if } C = 2.
\end{cases}
\]

With this notation, the solutions to parts a through d can be presented in an organized way.

a. The desired quantities are \( E[X_1] \) and \( \text{Var}(X_1) \). Using the formulas for unconditional
expectation and variance given in section 9.3 of the textbook we have

\[ E[X_1] = E_I [E[X_1 | I_1]] = E[X_1 | I_1 = 1] \Pr[I_1 = 1] + E[X_1 | I_1 = 0] \Pr[I_1 = 0] = E[L_1] \Pr[I_1 = 1] + E[0] \Pr[I_1 = 0] = E[L_1] (.25) + (0) (.75) = (.25) E[L_1] \]

and

\[ \text{Var}(X_1) = E_I [\text{Var}(X_1 | I_1)] + \text{Var}_I (E[X_1 | I_1]) = \text{Var}(X_1 | I_1 = 1) \Pr[I_1 = 1] + \text{Var}(X_1 | I_1 = 0) \Pr[I_1 = 0] + \text{Var}_I (E[X_1 | I_1]) = \text{Var}(L_1)(.25) + (0)(.75) + \text{Var}_I (E[X_1 | I_1]) = (.25) \text{Var}(L_1) + \text{Var}_I (E[X_1 | I_1]). \]

Now

\[ E[X_1 | I_1] = \begin{cases} E[L_1] & \text{if } I_1 = 1, \\ E[0] & \text{if } I_1 = 0; \end{cases} = \begin{cases} E[L_1] & \text{if } I_1 = 1, \\ 0 & \text{if } I_1 = 0; \end{cases} = E[L_1] \cdot I_1. \]

Hence

\[ \text{Var}_I (E[X_1 | I_1]) = E[L_1]^2 \text{Var}(I_1) = E[L_1]^2 \Pr[I_1 = 1] \Pr[I_1 = 0] = E[L_1]^2 (.25) (.75) = .1875 E[L_1]^2 \]

and so

\[ \text{Var}(X_1) = (.25) \text{Var}(L_1) + \text{Var}_I (E[X_1 | I_1]) = (.25) \text{Var}(L_1) + (.1875) E[L_1]^2. \]

(Note that the formula \( \text{Var}(I_1) = \Pr[I_1 = 1] \Pr[I_1 = 0] \) follows directly from the formula for the variance of a binomial random variable with \( n = 1 \) and \( p = \Pr[I_1 = 1] \).)

From the distribution of \( L_1 \) we have

\[ E[L_1] = (1000)(.20) + (5000)(.50) + (10000)(.30) = 5700 \]

and

\[ E[L_1^2] = (1000)^2 (.20) + (5000)^2 (.50) + (10000)^2 (.30) = 42,700,000. \]

Hence

\[ \text{Var}(L_1) = E[L_1^2] - E[L_1]^2 = 42,700,000 - (5700)^2 = 10,210,000. \]

Consequently, the mean and variance of \( X_1 \) are
\[ E[X_1] = (.25) E[L_1] = (.25)(5700) = 1425 \]

and

\[
\text{Var}(X_1) = \\
(\text{.25)} \text{Var}(L_1) + (.1875) E[L_1]^2 = (.25)(10,210,000) + (.1875)(5700)^2 = 8,644,375 .
\]

These are the numerical values of the desired quantities.

b. The desired quantities are \( E[X_2] \) and \( \text{Var}(X_2) \). Arguing as in part a we have

\[
E[X_2] = E[L_2] \Pr[I_2 = 1]
\]

and

\[
\text{Var}(X_2) = \text{Var}(L_2) \Pr[I_2 = 1] + \text{Var}_{L_2}(E[X_2 | I_2]) = \\
\text{Var}(L_2) \Pr[I_2 = 1] + \text{Var}(E[L_2 | I_2]) = \text{Var}(L_2) \Pr[I_2 = 1] + E[L_2]^2 \text{Var}(I_2) = \\
\text{Var}(L_2) \Pr[I_2 = 1] + E[L_2]^2 \Pr[I_2 = 1] \Pr[I_2 = 0].
\]

Hence

\[
E[X_2] = (.40) E[L_2]
\]

and

\[
\text{Var}(X_2) = (.40) \text{Var}(L_2) + (.40)(.60) E[L_2]^2 = (.40) \text{Var}(L_2) + (.24) E[L_2]^2.
\]

Now from the distribution of \( L_2 \) we have

\[
E[L_2] = (1000)(.70) + (5000)(.20) + (10000)(.10) = 2700,
\]

\[
E[L_2^2] = (1000)^2 (.70) + (5000)^2 (.20) + (10000)^2 (.10) = 15,700,000,
\]

and

\[
\text{Var}(L_2) = E[L_2^2] - E[L_2]^2 = 15,700,000 - (2700)^2 = 8,410,000.
\]

Consequently, the mean and variance of \( X_2 \) are

\[
E[X_2] = (.40) E[L_2] = (.40)(2700) = 1080
\]

and

\[
\text{Var}(X_2) = (.40) \text{Var}(L_2) + (.24) E[L_2]^2 = (.40)(8,410,000) + (.24)(2700)^2 = 5,113,600.
\]
These are the numerical values of the desired quantities.

c. The desired quantities are $E[Y]$ and $\text{Var}(Y)$. Using the formulas for unconditional expectation and variance given in section 9.3 of the textbook we have

$$E[Y] = E[C][E[Y \mid C]] =$$

$$E[Y \mid C = 1] \Pr[C = 1] + E[Y \mid C = 2] \Pr[C = 2] = E[X_1] \Pr[C = 1] + E[X_2] \Pr[C = 2]$$

and

$$\text{Var}(Y) = E[C][\text{Var}(Y \mid C)] + \text{Var}_C(E[Y \mid C]) =$$

$$\text{Var}(Y \mid C = 1) \Pr[C = 1] + \text{Var}(Y \mid C = 2) \Pr[C = 2] + \text{Var}_C(E[Y \mid C]) =$$

$$\text{Var}(X_1) \Pr[C = 1] + \text{Var}(X_2) \Pr[C = 2] + \text{Var}_C(E[Y \mid C]).$$

Now

$$E[Y \mid C] = \begin{cases} 
E[X_1] & \text{if } C = 1, \\
E[X_2] & \text{if } C = 2. 
\end{cases}$$

From parts a and b we have

$$E[X_1] = 1425$$

and

$$E[X_2] = 1080.$$

Further, from the information given in the statement of the question,

$$\Pr[C = 1] = .30, \quad \Pr[C = 2] = .70.$$ 

Hence

$$E[Y \mid C] = \begin{cases} 
1425 & \text{with probability .30}, \\
1080 & \text{with probability .70}, 
\end{cases}$$

and so

$$\text{Var}(E[Y \mid C]) = E[E[Y \mid C]^2] - (E[E[Y \mid C]])^2 =$$

$$\left\{(1425)^2 (.30) + (1080)^2 (.70)\right\} - \left\{(1425) (.30) + (1080) (.70)\right\}^2 =$$

$$1,425,667.50 - 1,400,672.25 = 24,995.25.$$ 

It follows that
\[
\text{Var}(Y) = \text{Var}(X_1) \Pr[C = 1] + \text{Var}(X_2) \Pr[C = 2] + \text{Var}(E[Y \mid C]) = \\
\text{Var}(X_1)(.30) + \text{Var}(X_2)(.70) + 24,995.25.
\]

Now from parts a and b we also have
\[
\text{Var}(X_1) = 8,644.375
\]
and
\[
\text{Var}(X_2) = 5,113,600.
\]

Hence substituting these values and the earlier values for \(E[X_1], E[X_2]\) into the formulas for \(E[Y]\) and \(\text{Var}(Y)\) we have
\[
E[Y] = E[X_1] \Pr[C = 1] + E[X_2] \Pr[C = 2] = (1425)(.30) + (1080)(.70) = 1183.50
\]
and
\[
\text{Var}(Y) = \text{Var}(X_1)(.30) + \text{Var}(X_2)(.70) + 24,995.25 = \\
(8,644.375)(.30) + (5,113,600)(.70) + 24,995.25 = 6,197,827.75.
\]

These are the numerical values of the desired quantities.

d. The desired probability is \(\Pr[Y \geq 5000]\). This can be determined by successive applications of the law of total probability.

To make the calculations more transparent, we will first determine a general formula for \(\Pr[Y \geq y]\) in the case \(y > 0\). Conditioning on the risk class \(C\) we have
\[
\Pr[Y \geq y] = \Pr[Y \geq y \mid C = 1] \Pr[C = 1] + \Pr[Y \geq y \mid C = 2] \Pr[C = 2] = \\
\Pr[X_1 \geq y](.30) + \Pr[X_2 \geq y](.70)
\]
and conditioning on the event of a claim occurrence we have
\[
\Pr[X_j \geq y] = \Pr[X_j \geq y \mid I_j = 1] \Pr[I_j = 1] + \Pr[X_j \geq y \mid I_j = 0] \Pr[I_j = 0] = \\
\Pr[L_j \geq y] \Pr[I_j = 1] + \Pr[0 \geq y] \Pr[I_j = 0].
\]

Using the fact that \(\Pr[0 \geq y] = 0\) for \(y > 0\) it follows that
\[
\Pr[Y \geq y] = \Pr[X_1 \geq y](.30) + \Pr[X_2 \geq y](.70) = \\
\Pr[L_1 \geq y] \Pr[I_1 = 1](.30) + \Pr[L_2 \geq y] \Pr[I_2 = 1](.70) = \\
\Pr[L_1 \geq y](.25)(.30) + \Pr[L_2 \geq y](.40)(.70) = \Pr[L_1 \geq y](.075) + \Pr[L_2 \geq y](.28)
for $y > 0$. Now from the distributions of $L_1$ and $L_2$ we have

$$
Pr[L_1 \geq 5000] = Pr[L_1 = 5000] + Pr[L_1 = 10000] = .50 + .30 = .80
$$

and

$$
$$

Consequently, the probability that the payout on a randomly selected policy is at least $5000 is

$$
Pr[Y \geq 5000] = 
Pr[L_1 \geq 5000] (.075) + Pr[L_2 \geq 5000] (.28) = (.80) (.075) + (.30) (.28) = .1440.
$$

17. Let $X_1, \ldots, X_{500}$ be the insurer's payouts on the 500 policies for which the claim incidence probability is 10%, let $Y_1, \ldots, Y_{2000}$ be the insurer's payouts on the 2000 policies for which the claim incidence probability is 5%, and let $S$ be the insurer's total payout on all policies. Suppose that the $X_i, Y_j$ and $S$ are all measured in $100,000$ lots. Then

$$
S = X_1 + \cdots + X_{500} + Y_1 + \cdots + Y_{2000}.
$$

Our objective is to determine the per-policy premium such that aggregate premiums exceed aggregate payments 95% of the time.

Suppose that the $X_i$ and $Y_j$ are mutually independent. Then the distributions of the sums $X_1 + \cdots + X_{500}$, $Y_1 + \cdots + Y_{2000}$ are approximately normal and so since the sum of independent normal random variables has a normal distribution, it follows that the distribution of $S$ is approximately normal. To use a normal approximation for $S$ we need only determine $E[S]$ and $\text{Var}(S)$.

Since the $X_i$ and $Y_j$ are mutually independent and $X_i \sim X$, $Y_j \sim Y$ for $X, Y$ of the type specified in the statement of the question, it follows that

$$
E[S] = 500 E[X] + 2000 E[Y],
$$

$$
\text{Var}(S) = 500 \text{Var}(X) + 2000 \text{Var}(Y).
$$

From the given information,
\( X = I B \)

where \( I \sim \text{Bernoulli}(0.10) \) and \( B \) is a truncated exponential random variable with parameters \( \lambda = 1 \) and \( m = 2.5 \), and

\( Y = I B \)

where \( I \sim \text{Bernoulli}(0.05) \) and \( B \) is a truncated exponential random variable with parameters \( \lambda = 2 \) and \( m = 5 \). In exercise 7 part c it is shown that if \( X = I B \) with \( I \sim \text{Bernoulli}(p) \) and \( B \) nonnegative then

\[
E[X] = p \mu_B,
\]

\[
\text{Var}(X) = p (1 - p) \mu_B^2 + p \sigma_B^2.
\]

From Example 3, section 4.2.1 and Example 2, section 4.2.2 of the textbook, the mean and variance of a truncated exponential random variable \( B \) with parameters \( \lambda \) and \( m \) are

\[
E[B] = \frac{1}{\lambda} \left( 1 - e^{-\lambda m} \right),
\]

\[
\text{Var}(B) = \frac{1}{\lambda^2} \left( 1 - 2 \lambda m e^{-\lambda m} - e^{-2\lambda m} \right).
\]

Hence if \( X = I B \) with \( I \sim \text{Bernoulli}(p) \) and \( B \) is a truncated exponential with parameters \( \lambda \) and \( m \), then

\[
E[X] = p \mu_B = \frac{p}{\lambda} \left( 1 - e^{-\lambda m} \right),
\]

\[
\text{Var}(X) = p (1 - p) \mu_B^2 + p \sigma_B^2 = p (1 - p) \left( 1 - e^{-\lambda m} \right)^2 + \frac{p}{\lambda^2} \left( 1 - 2 \lambda m e^{-\lambda m} - e^{-2\lambda m} \right).
\]

It follows that the mean and variance of the \( X_j \) (i.e., the case \( p = 0.10, \lambda = 1, m = 2.5 \)) are

\[
E[X_j] = \frac{0.10}{1} \left( 1 - e^{-(1)(2.5)} \right) = 0.10 \left( 1 - e^{-2.5} \right),
\]
\[ \text{Var} (X_j) = (.10) (.90) \frac{1}{12} \left( 1 - e^{-(1)(2.5)} \right)^2 + \frac{10}{12} \left( 1 - (2)(1)(2.5) e^{-(1)(2.5)} - e^{-(2)(1)(2.5)} \right) = .09 \left( 1 - e^{-2.5} \right)^2 + .10 \left( 1 - 5 \ e^{-2.5} - e^{-5} \right) = .19 - .68 \ e^{-2.5} - .01 \ e^{-5} \]

and the mean and variance of the \( Y_j \) (i.e., the case \( p = .05, \lambda = 2, m = 5 \)) are

\[ E[ Y_j ] = \frac{.05}{2} \left( 1 - e^{-(2)(5)} \right) = .025 \left( 1 - e^{-10} \right), \]

\[ \text{Var} ( Y_j ) = (.05) (.95) \frac{1}{2^2} \left( 1 - e^{-(2)(5)} \right)^2 + \frac{.05}{2^2} \left( 1 - (2) (2) (5) e^{-(2)(5)} - e^{-(2)(2)(5)} \right) = \]

\[ .011875 \left( 1 - e^{-10} \right)^2 + .0125 \left( 1 - 20 \ e^{-10} - e^{-20} \right) = .024375 - .27375 e^{-10} - .000625 e^{-20}. \]

Consequently, the mean and variance of \( S \) are

\[ E[S] = 500 \ E[X] + 2000 \ E[Y] = 500 \ (.10 \left( 1 - e^{-2.5} \right)) + 2000 \ (.025 \left( 1 - e^{-10} \right)) = 100 - 50 \ e^{-2.5} - 50 \ e^{-10} \approx 95.89348 \]

and

\[ \text{Var} (S) = 500 \ \text{Var} (X) + 2000 \ \text{Var} (Y) = 500 \ (.19 - .68 \ e^{-2.5} - .01 \ e^{-5}) + 2000 \ (.024375 - .27375 e^{-10} - .000625 e^{-20}) = 143.75 - 340 \ e^{-2.5} - 5 \ e^{-5} - 547.50 \ e^{-10} - 1.25 \ e^{-20} \approx 115.7825543 \approx (10.76023)^2. \]

a. Suppose that each policyholder is charged the same premium. Let \( P \) be the per-policy premium in $100,000 lots. Then we require

\[ \Pr[S < 2500 \ P] = .95. \]

Using a normal approximation for \( S \) we have

\[ \Pr[S < 2500 \ P] = \Pr \left[ \frac{S - E[S]}{\sqrt{\text{Var}(S)}} < \frac{2500 \ P - 95.89348}{10.76023} \right] \approx \Pr \left[ Z < \frac{2500 \ P - 95.89348}{10.76023} \right]. \]

Now \( \Phi(1.645) \approx .95. \) Hence the required premium \( P \) is given by
\[
\frac{2500 \cdot P - 95.89348}{10.76023} \approx 1.645,
\]

that is,
\[
P \approx 0.0454376.
\]

Consequently, if all policyholders are charged the same amount, the required premium is $4543.76 per policy.

b. Suppose that the insurer charges a policyholder with expected loss \( \mu \) the premium \( P = (1 + \vartheta) \mu \) where \( \vartheta \) is the same for all policyholders. Then the premium for the 500 policyholders with claim incidence probability 10% is
\[
P = (1 + \vartheta) \cdot E[X] = (1 + \vartheta) \cdot 0.10 \cdot (1 - e^{-2.5}) \approx 0.09179150 \cdot (1 + \vartheta)
\]

and the premium for the 2000 policyholders with claim incidence probability 5% is
\[
P = (1 + \vartheta) \cdot E[Y] = (1 + \vartheta) \cdot 0.025 \cdot (1 - e^{-10}) = 0.02499887 \cdot (1 + \vartheta).
\]

The requirement that aggregate premiums exceed aggregate payments 95% of the time takes the form
\[
\Pr[S < (500 \cdot 0.09179150 + 2000 \cdot 0.02499887) (1 + \vartheta)] = 0.95,
\]

i.e.,
\[
\Pr[S < 95.89348 (1 + \vartheta)] = 0.95.
\]

Using a normal approximation for \( S \) we have
\[
\Pr[S < 95.89348 (1 + \vartheta)] = \Pr\left[ \frac{S - E[S]}{\sqrt{\text{Var}(S)}} < \frac{95.89348 (1 + \vartheta) - 95.89348}{10.76023} \right] \approx \Phi \left[ \frac{95.89348 (1 + \vartheta) - 95.89348}{10.76023} \right],
\]

Since \( \Phi[1.645] \approx 0.95 \), the requirement on \( \vartheta \) is
\[
\frac{95.89348 (1 + \vartheta) - 95.89348}{10.76023} \approx 1.645,
\]

i.e.,
\[ \vartheta \approx .18458584. \]

Consequently, the required premium for policies with claim incidence probability 10\% is given by
\[ P = (1 + \vartheta) E[X] = (1.18458584) (.09179150) = .1087349, \]
i.e., \$10,873.49 per policy, and the required premium for policies with claim incidence probability 5\% is given by
\[ P = (1 + \vartheta) E[Y] = (1.18458584) (.02499887) = .02961331, \]
i.e., \$2961.33 per policy.

**Comment:** The total premium collected under the criterion in part a is
\[ (2500) (\$4543.76) \approx \$11,359,400. \]

The total premium collected under the criterion in part b is
\[ (500) (\$10,873.49) + (2000) (\$2,961.33) = \$11,359,405. \]

Ignoring rounding error, we see that the total premium collected is the same. The advantage of the criterion in part b is that it charges a premium that distinguishes between the risks. This can help to avoid a problem with adverse selection.

20. Let \( S \) be the total amount of claims in hundreds of dollars. Then
\[ S = W_1 + \cdots + W_N \]
where \( W_j \) is the size of the \( j \)-th claim in hundreds of dollars and \( N \) is the number of claims. We are given that \( N \sim \text{Binomial}(4, .25) \) and the distribution of \( W_j \) is
\[ \Pr[W_j = 1] = .3, \quad \Pr[W_j = 2] = .45, \quad \Pr[W_j = 3] = .25. \]

We are also given that the \( W_j \) are mutually independent and independent of \( N \). We are asked to determine \( \Pr[S > 5] \).

By the law of total probability,
\[
\Pr[S > 5] = \\
\sum_{n=0}^{\infty} \Pr[S > 5 \mid N = n] \Pr[N = n] = \sum_{n=0}^{\infty} \Pr[W_1 + \cdots + W_N > 5 \mid N = n] \Pr[N = n].
\]

Since the \(W_j\) are independent of \(N\),
\[
\Pr[W_1 + \cdots + W_N > 5 \mid N = n] = \Pr[W_1 + \cdots + W_n > 5].
\]

Moreover, since \(N \sim \text{Binomial}(4, .25)\),
\[
\Pr[N = n] = \binom{4}{n} (.25)^n (.75)^{4-n} \quad \text{for } n = 0, 1, 2, 3, 4.
\]

Hence
\[
\Pr[S > 5] = \sum_{n=0}^{4} \Pr[W_1 + \cdots + W_n > 5] \binom{4}{n} (.25)^n (.75)^{4-n}.
\]

Now since the only possible values of \(W_j\) are 1, 2, or 3, the aggregate claim \(S\) cannot exceed 5 if \(N \leq 1\). In particular,
\[
\Pr[W^0 > 5] = 0,
\]
\[
\Pr[W > 5] = 0,
\]
where \(W^0\) represents \(W_1 + \cdots + W_n\) when \(n = 0\). Hence
\[
\Pr[S > 5] = \Pr[W_1 + W_2 > 5] \binom{4}{2} (.25)^2 (.75)^2 + \\
\Pr[W_1 + W_2 + W_3 > 5] \binom{4}{3} (.25)^3 (.75) + \Pr[W_1 + W_2 + W_3 + W_4 > 5] (.25)^4.
\]

Consider \(\Pr[W_1 + W_2 > 5]\). Since the possible values of \(W_j\) are 1, 2, or 3, the only way that the aggregate claim \(S\) can exceed 5 is if both \(W_1\) and \(W_2\) are equal to 3, i.e.,
\[
\Pr[W_1 + W_2 > 5] = \Pr[W_1 = 3 \text{ and } W_2 = 3].
\]

Since the \(W_j\) are independent, it follows that
\[
\Pr[W_1 + W_2 > 5] = \Pr[W_1 = 3] \Pr[W_2 = 3] = (.25)^2 = .0625.
\]
Now consider $\Pr[W_1 + W_2 + W_3 > 5]$. Put $X = W_1 + W_2 + W_3$. For simplicity we will determine $\Pr[X \leq 5]$ and then deduce the value for $\Pr[X > 5]$. The values of the triple $(W_1, W_2, W_3)$ for which $X \leq 5$ are as follows:

$(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 1), (1, 2, 2), (1, 3, 1), (2, 1, 1), (2, 1, 2), (2, 2, 1), (3, 1, 1)$.

Since the $W_j$ are independent, it follows that

$$\Pr[X \leq 5] = \Pr[W_1 = 1] \Pr[W_2 = 1] \Pr[W_3 = 1] + \Pr[W_1 = 1] \Pr[W_2 = 2] \Pr[W_3 = 1] + \cdots + \Pr[W_1 = 3] \Pr[W_2 = 1] \Pr[W_3 = 1] = (.3)^3 + 3 (.3)^2 (.45) + 3 (.3)^3 (.25) + 3 (.3) (.45)^2 = .39825.$$  

Hence

$$\Pr[W_1 + W_2 + W_3 > 5] = \Pr[X > 5] = 1 - \Pr[X \leq 5] = 1 - .39825 = .60175.$$  

Finally consider $\Pr[W_1 + W_2 + W_3 + W_4 > 5]$. Put $Y = W_1 + W_2 + W_3 + W_4$. As in the previous case, it is simpler to determine $\Pr[Y \leq 5]$ and deduce the value of $\Pr[Y > 5]$. The values of the quadruple $(W_1, W_2, W_3, W_4)$ for which $Y \leq 5$ are as follows:

$(1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 2, 1), (1, 2, 1, 1), (2, 1, 1, 1)$.

Hence arguing as before,

$$\Pr[Y \leq 5] = \Pr[W_1 = 1] \Pr[W_2 = 1] \Pr[W_3 = 1] \Pr[W_4 = 1] + \Pr[W_1 = 1] \Pr[W_2 = 2] \Pr[W_3 = 1] \Pr[W_4 = 1] + \cdots + \Pr[W_1 = 3] \Pr[W_2 = 1] \Pr[W_3 = 1] \Pr[W_4 = 1] = (.3)^4 + 4 (.3)^3 (.45) = .0567.$$  

It follows that

$$\Pr[W_1 + W_2 + W_3 + W_4 > 5] = \Pr[Y > 5] = 1 - \Pr[Y \leq 5] = 1 - .0567 = .9433.$$  

Substituting these results into the earlier formula for $\Pr[S > 5]$ we obtain

$$\Pr[S > 5] = \Pr[W_1 + W_2 > 5] \binom{4}{2} (.25)^2 (.75)^2 + \Pr[W_1 + W_2 + W_3 > 5] \binom{4}{3} (.25)^3 (.75) + \Pr[W_1 + W_2 + W_3 + W_4 > 5] (.25)^4 = (.0625) (6) (.25)^2 (.75)^2 + (.60175) (4) (.25)^3 (.75) + (.9433) (.25)^4 = .04507539 \approx 4.5\%.$$
Consequently, the probability that aggregate claims on the given policy exceed $500 is about 4.5%.

24. Let $Y_1, Y_2, Y_3, Y_4$ be the aggregate claims in thousands of dollars in the seasons spring, summer, fall, winter respectively and let $C$ denote the aggregate claims for the entire year. By assumption, the $Y_j$ have compound Poisson distributions with intensity parameters

$$\lambda_1 = 100, \quad \lambda_2 = 75, \quad \lambda_3 = 90, \quad \lambda_4 = 200$$

and claim amount random variables

Exponential (1), Exponential (2), Exponential (1), Exponential (.25).

Moreover, the $Y_j$ are assumed to be independent. Hence from section 9.5.2 of the textbook or exercise 22, $C$ has a compound Poisson distribution with intensity parameter

$$\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 100 + 75 + 90 + 200 = 465$$

and claim amount $W$ a mixture of Exponential(1), Exponential(2), Exponential(1), and Exponential(.25) with mixing weights

$$\lambda_1/\lambda = 100/465 = 20/93, \quad \lambda_2/\lambda = 75/465 = 15/93, \quad \lambda_3/\lambda = 90/465 = 18/93, \quad \lambda_4/\lambda = 200/465 = 40/93.$$

a. From section 9.5.2 of the textbook or exercise 21 part b, the formulas for the expected value, variance, and skewness of a compound Poisson random variable are as follows:

$$E[C] = \lambda E[W],$$

$$\text{Var} (C) = \lambda E[W^2],$$

$$\gamma_C = \frac{1}{\lambda^{1/2}} \cdot \frac{E[W^3]}{E[W^2]^{3/2}}.$$

In this exercise, $W$ is a mixture of Exponential(1), Exponential(2), Exponential(1), Exponential(0.25) with mixing weights

$$\frac{20}{93}, \frac{15}{93}, \frac{18}{93}, \frac{40}{93}$$

and $\lambda = 465$. Hence using the general formula for the moment generating function of a mixture and the formula for the moment generating function of an exponential random variable we have
\[
M_W[t] = \frac{20}{38} \cdot \frac{1}{1-t} + \frac{15}{15} \cdot \frac{2}{2-t} + \frac{18}{40} \cdot \frac{1}{1-t} + \frac{40}{93} \cdot \frac{0.25}{2 - t} + \frac{0.25}{93} \cdot \frac{1}{1-4t}.
\]

It follows that
\[
E[W^k] = M_W^{(k)}[0] = \left. \frac{38}{93} \cdot \frac{d^k}{dt^k}\left(\frac{1}{1-t}\right)\right|_{t=0} + \left. \frac{15}{93} \cdot \frac{d^k}{dt^k}\left(\frac{2}{2-t}\right)\right|_{t=0} + \left. \frac{40}{93} \cdot \frac{d^k}{dt^k}\left(\frac{1}{1-4t}\right)\right|_{t=0} = \frac{38}{93} \cdot \frac{k!}{1} + \frac{15}{93} \cdot \frac{k!}{2} + \frac{40}{93} \cdot \frac{k!}{4}. \]

In particular,
\[
E[W] = \frac{411}{186} = \frac{137}{62},
\]
\[
E[W^2] = \frac{2727}{186} = \frac{909}{62},
\]
\[
E[W^3] = \frac{62,397}{372} = \frac{20,799}{124}.
\]

Consequently, the mean, variance, and skewness of the aggregate claim random variable is
\[
E[C] = \lambda E[W] = 465 \left(\frac{137}{62}\right) = \frac{2055}{2},
\]
\[
\text{Var}(C) = \lambda E[W^2] = 465 \left(\frac{909}{62}\right) = \frac{13,635}{2},
\]
\[
\gamma_C = \frac{1}{\lambda^{1/2}} \cdot \frac{E[W^3]}{E[W^2]^{3/2}} = \frac{1}{\sqrt{465}} \cdot \frac{20,799}{\left(\frac{909}{62}\right)^{3/2}} = \frac{2311}{303 \sqrt{3030}} \approx .13855940.
\]

b. The desired probability is \(\Pr[C > 1200]\). (Note that \(C\) is measured in thousands of
The desired probability is \( \Pr@C_1200D \).
(Note that \( C \) is measured in thousands of dollars.)
From part a, we have
\[
E[C] = 1027.50, \\
\text{Var} (C) = 6817.50, \\
\gamma_C = .13855940.
\]
Hence using a normal approximation we have
\[
\Pr[C > 1200] = \Pr\left[ \frac{C - E[C]}{\sqrt{\text{Var} (C)}} > \frac{1200 - 1027.50}{\sqrt{6817.50}} \right] \approx \Pr[Z > 2.0892] = 1 - \Phi[2.0892]
\]
where \( Z \sim \text{Normal}(0, 1) \) and \( \Phi \) is the standard normal distribution function. From the table in Appendix E of the textbook and using linear interpolation we have
\[
\Phi[2.0892] \approx (.08)\Phi[2.08] + (.92)\Phi[2.09] \approx (.08)(.9812) + (.92)(.9817) = .98166.
\]
Consequently, under a normal approximation the desired probability is
\[
\Pr[C > 1200] \approx 1 - \Phi[2.0892] \approx 1 - .98166 = 1.834\%.
\]
We can also approximate the desired probability using a normal power approximation. From section 8.6 of the textbook we have the approximation
\[
F_C[c] \approx \Phi\left[ \frac{-3}{\gamma_C} + \frac{9}{\gamma_C^2} + 1 + \frac{6}{\gamma_C \sigma_C} \cdot \frac{c - \mu_C}{\sigma_C} \right].
\]
Substituting the values of \( \mu_C, \sigma_C, \) and \( \gamma_C \) specified earlier into this formula we have
\[
F_C[c] \approx \Phi\left[ \frac{-3}{.13855940} + \frac{9}{(.13855940)^2} + 1 + \frac{6}{(.13855940)} \cdot \frac{c - 1027.50}{\sqrt{6817.50}} \right] \approx \Phi\left[ -21.65136 + \sqrt{469.7816 + 43.3027 \left( \frac{c - 1027.50}{82.5682} \right)} \right].
\]
Hence the desired probability is
\[
\Pr[C > 1200] = 1 - F_C[1200] \approx 1 - \Phi[2.0182].
\]
From Appendix E of the textbook we have
\[ \Phi[2.0182] \approx (.18) \Phi[2.01] + (.82) \Phi[2.02] \approx (.18) (.9778) + (.82) (.9783) = .97821. \]

Consequently, under a normal power approximation the desired probability is
\[ \Pr[C > 1200] \approx 1 - \Phi[2.0182] \approx 1 - .97821 = 2.179\%. \]

Both approximations give a numerical value of \( \Pr[C > 1200] \) that is fairly small. However, the value of 2.179% determined by the normal power approximation is probably closer to the true value because it incorporates the impact of the skewness of \( C \) which is not entirely negligible in this case. Since the normal distribution is symmetric (and hence has skewness of 0), it has a tendency to underestimate probability in the right tail of distributions with a positive skew.

30. Let \( N \) be the number of claims received on this policy and let \( X \) be the aggregate amount of claims for this policy. Then from the given information, the distribution of \( N \) is given by
\[ p_N[0] = .50, \quad p_N[1] = .30, \quad p_N[2] = .20 \]

and
\[
\begin{align*}
X &= 0 & \text{if } N = 0, \\
X &= W_1 & \text{if } N = 1, \\
X &= W_1 + W_2 & \text{if } N = 2,
\end{align*}
\]

where \( W_j \sim \text{Exponential}(1/2) \) and the \( W_j \) are independent. From section 6.1.2 of the textbook, we know that the sum of independent exponential random variables with common parameter \( \lambda \) has a gamma distribution. Hence \( W_1 + W_2 \sim \text{Gamma}(2, \frac{1}{2}) \).

Consequently, \( X \) is a mixture of a certain point mass at 0, the distribution \( \text{Exponential}(\frac{1}{2}) \), and the distribution \( \text{Gamma}(2, \frac{1}{2}) \) with respective mixing weights .50, .30, .20.

a. From the observation that \( X \) is a mixture of the type just described it follows that the distribution function of \( X \) is given by
\[ F_X[x] = .50 F_{X_1}[x] + .30 F_{X_2}[x] + .20 F_{X_3}[x] \]

where

\[ F_{X_1}[x] = \begin{cases} 
1 & \text{for } x \geq 0, \\
0 & \text{for } x < 0, 
\end{cases} \]

\[ F_{X_2}[x] = \begin{cases} 
1 - e^{-x/2} & \text{for } x \geq 0, \\
0 & \text{for } x < 0, 
\end{cases} \]

and

\[ F_{X_3}[x] = \begin{cases} 
1 - e^{-x/2} \left( 1 + \frac{x}{2} \right) & \text{for } x \geq 0, \\
0 & \text{for } x < 0. 
\end{cases} \]

(See section 6.1.2 of the textbook for the formula for the survival function of a gamma distribution whose parameter \( r \) is a positive integer.) Simplifying we obtain

\[ F_X[x] = \begin{cases} 
1 - e^{-x/2} \left( .50 + .20 \frac{x}{2} \right) & \text{for } x \geq 0, \\
0 & \text{for } x < 0; \end{cases} \]

\[ = \begin{cases} 
1 - .50 e^{-x/2} \left( 1 + \frac{x}{2} \right) & \text{for } x \geq 0, \\
0 & \text{for } x < 0. \end{cases} \]

b. To determine the unconditional mean and variance of \( X \) we condition on \( N \) and use the formulas

\[ E[X] = E_N[E[X \mid N]]. \]

\[ \text{Var}(X) = E_N[\text{Var}(X \mid N)] + \text{Var}_N(E[X \mid N]). \]

From the earlier observations, we have

\[ E[X \mid N] = E[0] \quad \text{if } N = 0, \]

\[ E[X \mid N] = E[W_1] \quad \text{if } N = 1, \]

\[ E[X \mid N] = E[W_1 + W_2] \text{ if } N = 2, \]

and
\[ \text{Var}(X \mid N) = \text{Var}(0) \quad \text{if } N = 0, \]
\[ \text{Var}(X \mid N) = \text{Var}(W_1) \quad \text{if } N = 1, \]
\[ \text{Var}(X \mid N) = \text{Var}(W_1 + W_2) \quad \text{if } N = 2. \]

Since \( W_j \sim \text{Exponential}(1/2) \) and the \( W_j \) are independent, it follows from the formulas for the mean and variance of an exponential distribution (section 6.1.1 of the textbook) and the distribution of \( N \) specified earlier that
\[ E[X \mid N] = \begin{cases} 0 & \text{with probability } .50, \\ 2 & \text{with probability } .30, \\ 4 & \text{with probability } .20, \end{cases} \]
and
\[ \text{Var}(X \mid N) = \begin{cases} 0 & \text{with probability } .50, \\ 4 & \text{with probability } .30, \\ 8 & \text{with probability } .20. \end{cases} \]

Consequently,
\[ E_N[E[X \mid N]] = (0)(.50) + (2)(.30) + (4)(.20) = 1.4, \]
\[ E_N[\text{Var}(X \mid N)] = (0)(.50) + (4)(.30) + (8)(.20) = 2.8, \]
and
\[ \text{Var}_N(E[X \mid N]) = E_N[\text{Var}(X \mid N)] = 2.8. \]

Therefore, the unconditional mean and variance of \( X \) are
\[ E[X] = E_N[E[X \mid N]] = 1.4, \]
\[ \text{Var}(X) = E_N[\text{Var}(X \mid N)] + \text{Var}_N(E[X \mid N]) = 2.8 + 2.44 = 5.24. \]

c. From the solution to part a, we have
\[ S_X[x] = .50 e^{-x/2} \left(1 + \frac{x}{5}\right) \quad \text{for } x \geq 0. \]

Hence the probability that aggregate claims exceed three is
\[ S_X[3] = .50 e^{-3/2} \left(1 + \frac{3}{5}\right) = .80 e^{-3/2} \approx .1785. \]
1. Let $\mu_S, \sigma_S, \mu_B, \sigma_B$ be the expected return and standard deviation of return for the stock and bond funds respectively, and let $\rho$ be the correlation of the returns for the two funds. We are given that $\mu_S = 15\%$, $\sigma_S = 20\%$, $\mu_B = 5\%$, $\sigma_B = 10\%$, $\rho = 0.25$.

a. From section 10.1 of the textbook, the equation of the curve in the risk-reward plane representing the possible portfolios is

$$\sigma^2 = A (\mu - \mu_0)^2 + \sigma_0^2$$

where

$$A = \frac{(\sigma_S - \sigma_B)^2 + 2 (1 - \rho) \sigma_S \sigma_B}{(\mu_S - \mu_B)^2},$$

$$\mu_0 = \frac{\mu_S \sigma_B^2 - (\mu_S + \mu_B) \rho \sigma_S \sigma_B + \mu_B \sigma_S^2}{(\sigma_S - \sigma_B)^2 + 2 (1 - \rho) \sigma_S \sigma_B},$$

$$\sigma_0^2 = \frac{\sigma_S^2 \sigma_B^2 (1 - \rho^2)}{(\sigma_S - \sigma_B)^2 + 2 (1 - \rho) \sigma_S \sigma_B}.$$

Substituting the given values of $\mu_S, \sigma_S, \mu_B, \sigma_B, \rho$ into these expressions we have

$$A = \frac{(20\% - 10\%)^2 + 2 (1 - .25) (20\%) (10\%)}{(15\% - 5\%)^2} = 4,$$
\[ \mu_0 = \frac{(15\%)^2 - (15\% + 5\%) (0.25) (20\%) (10\%) + (5\%) (20\%)^2}{(20\% - 10\%)^2 + 2 (1 - .25) (20\%) (10\%)} = 6.25\%, \]

\[ \sigma_0^2 = \frac{(20\%)^2 (10\%)^2 (1 - (.25)^2)}{(20\% - 10\%)^2 + 2 (1 - .25) (20\%) (10\%)} = 93.75\%(%)^2 = (9.68\%)^2. \]

Consequently, the equation in the risk-reward plane is
\[ \sigma^2 = 4 (\mu - 6.25\%)^2 + (9.68\%)^2. \]

b. From part a, the portfolio with the least risk has risk-reward coordinates
\[ \sigma_0 = 9.68\%, \]
\[ \mu_0 = 6.25\%. \]

c. The graph is straightforward to create using Mathematica or similar computer software.

d. For an investor with utility functional \( U = \mu - k \sigma^2 \), the optimal allocation in the stock fund is
\[ \mu^* - \mu_B \]
\[ \mu_S - \mu_B \]
where \( \mu^* \) is the return on the optimal portfolio and given by
\[ \mu^* = \mu_0 + \frac{1}{2kA} \]

(see section 10.1 of the textbook). In this exercise, \( k = 1/100 \). Hence the return on the optimal portfolio is
\[ \mu^* = \mu_0 + \frac{1}{2kA} = 6.25\% + \frac{1}{2 \left( \frac{1}{100\%} \right)}(4) = 18.75\% \]

and the fraction invested in the stock fund is
\[ \mu^* - \mu_B = \frac{18.75 \% - 5 \%}{15 \% - 5 \%} = 1.375. \]

The interpretation of this is as follows: For every $1000 of investable assets, one should sell short $375 of bonds and invest $1375 in stocks. If it is not possible to sell short the bond fund, then the next best thing for this particular investor is to invest 100% of the portfolio in the stock fund.

3. Let \( x_1, x_2, x_3 \) be the weights of \( S_1, S_2, S_3 \) in the optimal risky portfolio and \( \lambda \) be the reward-to-variability ratio for the optimal risky portfolio when the risk-free rate is \( r_f \).

From section 10.3 of the textbook, \( x_1, x_2, x_3 \), and \( \lambda \) satisfy the following system:

\[
\begin{align*}
\sigma_1^2 (\lambda x_1) + \sigma_{12} (\lambda x_2) + \sigma_{13} (\lambda x_3) &= \mu_1 - r_f, \\
\sigma_{21} (\lambda x_1) + \sigma_2^2 (\lambda x_2) + \sigma_{23} (\lambda x_3) &= \mu_2 - r_f, \\
\sigma_{31} (\lambda x_1) + \sigma_{32} (\lambda x_2) + \sigma_3^2 (\lambda x_3) &= \mu_3 - r_f.
\end{align*}
\]

This system has a unique solution in \( \lambda x_1, \lambda x_2, \lambda x_3 \) provided that the variance-covariance matrix is nondegenerate, which we will assume. Hence

\[
\begin{align*}
\lambda x_1 &= c_{11} + c_{12} r_f, \\
\lambda x_2 &= c_{21} + c_{22} r_f, \\
\lambda x_3 &= c_{31} + c_{32} r_f,
\end{align*}
\]

where the \( c_{ij} \) are independent of \( r_f \). We are given that when \( r_f = 5 \% \),

\[
\begin{align*}
x_1 &= \frac{74}{93}, \quad x_2 = \frac{-60}{93}, \quad x_3 = \frac{79}{93}, \quad \lambda = 46.5.
\end{align*}
\]

Substituting these values into the preceding system we obtain

\[
\begin{align*}
(46.5) \left( \frac{74}{93} \right) &= c_{11} + c_{12} (5), \\
(46.5) \left( \frac{-60}{93} \right) &= c_{21} + c_{22} (5).
\end{align*}
\]
(46.5) \[ \frac{79}{93} = c_{31} + c_{32} \]

where the units of \( c_{i1} \) and \( c_{i2} \) are as appropriate. We are also given that when \( r_f = 10 \% \),

\[ x_1 = \frac{12}{19}, \quad x_2 = \frac{-20}{19}, \quad x_3 = \frac{27}{19}, \quad \lambda = 19. \]

Hence arguing as before we obtain

\[ (19) \left( \frac{12}{19} \right) = c_{11} + c_{12} \]
\[ (19) \left( \frac{-20}{19} \right) = c_{21} + c_{22} \]
\[ (19) \left( \frac{27}{19} \right) = c_{31} + c_{32} \]

The six equations in the \( c_{ij} \) are most easily solved by considering them as three sets of two equations:

\[ 37 = c_{11} + 5 c_{12}, \]
\[ 12 = c_{11} + 10 c_{12}; \]
\[ -30 = c_{21} + 5 c_{22}, \]
\[ -20 = c_{21} + 10 c_{22}; \]
\[ 39.5 = c_{31} + 5 c_{32}, \]
\[ 27 = c_{31} + 10 c_{32}. \]

Solving these equations we obtain

\[ c_{11} = 62, \quad c_{12} = -5, \]
\[ c_{21} = -40, \quad c_{22} = 2, \]
\[ c_{31} = 52, \quad c_{32} = -2.5. \]

Hence the coordinates of the optimal risky portfolio are given by
\[ \lambda x_1 = 62 - 5 r_f, \]
\[ \lambda x_2 = -40 + 2 r_f, \]
\[ \lambda x_3 = 52 - 2.5 r_f. \]

In particular, for \( r_f = 15 \% \) we have
\[ \lambda x_1 = -13, \quad \lambda x_2 = -10, \quad \lambda x_3 = 14.5 \]
and so
\[ x_1 = \frac{26}{17}, \quad x_2 = \frac{20}{17}, \quad x_3 = \frac{-29}{17}. \]

8. We are given that the expected return and standard deviation on the market portfolio are \( \mu_M = 15 \) and \( \sigma_M = 15 \) and the expected return and standard deviation on the global minimum variance portfolio are \( \mu_0 = 5 \) and \( \sigma_0 = 5 \). Note that in this context the global minimum variance portfolio refers to the portfolio among those with no position in the risk-free asset whose variance is the smallest. Since both the market portfolio and the global minimum variance portfolio must lie on the minimum variance set of risky portfolios ("risky" meaning that there is no position in the risk-free asset), we can determine the equation in the risk-reward plane for the minimum variance set. Indeed, from sections 10.1, 10.2, and 10.3 of the textbook, the minimum variance set has equation
\[ \sigma^2 = A (\mu - \mu_0)^2 + \sigma_0^2 \]
where \((\sigma_0, \mu_0)\) are the risk-reward coordinates for the global minimum variance portfolio. Substituting the values of \((\sigma, \mu)\) for the market portfolio into this equation and solving for \(A\), we find that \(A = 2\). Hence the minimum variance set for the risky portfolios is defined by
\[ \sigma^2 = 2 (\mu - 5)^2 + 25. \]

We are given that the rate of return on risk-free investments is 3\% and short positions in the risk-free asset are not allowed. Consequently, the set of efficient portfolios consists of two separate pieces:
i. The line segment between the points (0, 3 %) and T where T is the point of tangency for the curve \( \sigma^2 = 2 (\mu - 5)^2 + 25 \) and the tangent line through (0, 3 %).

ii. The portion of the curve \( \sigma^2 = 2 (\mu - 5)^2 + 25 \) to the right of the point T.

Note that by the definition of a zero-beta companion portfolio, T is the efficient portfolio whose zero beta companion has expected return 3%.

Because short positions are not allowed in the risk-free asset, the market portfolio will lie on the portion of the curve \( \sigma^2 = 2 (\mu - 5)^2 + 25 \) that is to the right of T. We don't actually need to make this observation to determine the answers to parts a through d. It will be a consequence of our calculations. However, it is helpful to keep this fact in mind as we proceed through the rest of the solution.

a. Let \((\sigma_{Z(M)}, \mu_{Z(M)})\) be the risk-reward coordinates for the zero beta companion of the market portfolio M. From the definition of a zero beta companion, \(\mu_{Z(M)}\) is the \(\mu\)-intercept of the tangent line to the hyperbola \(\sigma^2 = 2 (\mu - 5)^2 + 25\) through the point \((\sigma_M, \mu_M)\). The value of \(\mu_{Z(M)}\) can be determined from the equation:

\[
\left( \frac{d\mu}{d\sigma} \right)_{(\sigma, \mu)=(\sigma_M, \mu_M)} = \frac{\mu_M - \mu_{Z(M)}}{\sigma_M - 0}.
\]

Note that the left side of this equation is the slope of the hyperbola at the point \((\sigma_M, \mu_M)\) and the right side of this equation is the slope of the line through the points \((0, \mu_{Z(M)})\) and \((\sigma_M, \mu_M)\). A formula for the derivative \(d\mu/d\sigma\) can be determined by implicitly differentiating the equation \(\sigma^2 = 2 (\mu - 5)^2 + 25\):

\[
\frac{d\mu}{d\sigma} = \frac{\sigma}{2 (\mu - 5)}.
\]

Substituting \((\sigma_M, \mu_M) = (15, 15)\) into the equation for \(\mu_{Z(M)}\) and solving we obtain \(\mu_{Z(M)} = 3.75\).

Since the zero beta companion of M must lie on the minimum variance set, we can determine the standard deviation of the zero beta companion by substituting this value for \(\mu_{Z(M)}\) into the equation \(\sigma^2 = 2 (\mu - 5)^2 + 25\). When we do this, we find that
\[ \sigma^2_Z(M) = 2(\mu_Z(p) - 5)^2 + 25 = 2(3.75 - 5)^2 + 25 = 28.125. \]

Hence the expected return and standard deviation for the zero beta companion of the market portfolio are
\[ \mu_Z(M) = 3.75, \]
\[ \sigma_Z(M) = \sqrt{28.125} \approx 5.30. \]

b. From section 10.3 of the textbook, we know that every portfolio on the minimum variance set can be constructed using only the market portfolio and its zero beta companion. Indeed, for any portfolio \( Q \) on the minimum variance set, the weights \( x_1, x_2 \) in \( M, Z(M) \) are given by
\[ x_1 = \frac{z_1}{z_1 + z_2}, \quad x_2 = \frac{z_2}{z_1 + z_2} \]
where
\[ \begin{pmatrix} \sigma^2_M & 0 \\ 0 & \sigma^2_Z(M) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \mu_M - r \\ \mu_Z(M) - r \end{pmatrix} \]
and where \( r \) is the expected return on the zero beta companion of \( Q \). Since \( T \) is a portfolio on the minimum variance set, it follows that \( T \) can be constructed using only \( M \) and \( Z(M) \). Since the expected return on the zero beta companion of \( T \) is 3% and \( (\sigma_Z(M), \mu_Z(M)) = (5.30, 3.75) \) from part a, it follows that the weights \( x_1, x_2 \) in \( M, Z(M) \) can be determined by solving
\[ \begin{pmatrix} 15^2 & 0 \\ 0 & 28.125 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 15 - 3 \\ 3.75 - 3 \end{pmatrix}. \]

When we do this, we find that
\[ z_1 = \frac{12}{225}, \quad z_2 = \frac{0.75}{28.125} \]
and so
Chapter Ten Solutions

Consequently, we can replicate portfolio $T$ by investing $\frac{2}{3}$ of our assets in the market portfolio and the remaining $\frac{1}{3}$ in the zero beta companion of the market portfolio.

c. Let $R_T, R_M, R_{Z(M)}$ denote the returns on the portfolio $T$, the market portfolio $M$, and the zero beta companion of the market portfolio respectively. Then from part b,

$$R_T = x_1 R_M + x_2 R_{Z(M)} = \frac{2}{3} R_M + \frac{1}{3} R_{Z(M)}.$$

Hence using the fact that $\mu_{Z(M)} = 3.75$ from part a,

$$E[R_T] = \frac{2}{3} E[R_M] + \frac{1}{3} E[R_{Z(M)}] = \frac{2}{3} (15) + \frac{1}{3} (3.75) = 11.25.$$

Now the CAPM for the efficient portfolios lying to the right of the tangency portfolio $T$ (i.e., the efficient portfolios with expected return greater than $E[R_T]$) is given by

$$E[R] = E[R_{Z(M)}] + \beta (E[R_M] - E[R_{Z(M)}]),$$

with the restriction on $\beta$ to be determined shortly. (See section 10.4 of the textbook.) This version of the CAPM is used since the efficient portfolios with expected return greater than $E[R_T]$ do not have a position (either long or short) in the risk-free asset. Since $T$ does not have a position (either long or short) in the risk-free asset, it follows that $E[R_T]$ and $\beta_T$ satisfy the preceding CAPM equation, i.e.,

$$E[R_T] = E[R_{Z(M)}] + \beta_T (E[R_M] - E[R_{Z(M)}]).$$

Substituting the given and calculated values into this equation, we have

$$11.25 = 3.75 + \beta_T (15 - 3.75)$$

from which it follows that

$$\beta_T = \frac{2}{3}. $$
Since the CAPM equation is increasing in $\beta$ and since all efficient portfolios with expected return less than $E[R_T]$ lie on the straight line through $(0, 3\%)$, $(\sigma_T, \mu_T)$ (and hence contain a long position in the risk-free asset), it follows that the preceding CAPM equation only holds for $\beta \geq \beta_T$, i.e., for $\beta \geq \frac{2}{3}$.

d. From part c, the security market line for efficient portfolios with expected return greater than $E[R_T]$ is
$$E[R] = E[R_{Z(M)}] + \beta (E[R_M] - E[R_{Z(M)}]).$$

As noted in part c, this equation holds for $\beta \geq \beta_T$. Substituting the given and calculated values into this equation, we see that the security market line for $\beta \geq \frac{2}{3}$ is given by
$$E[R] = 3.75 + 11.25 \beta.$$

The security market line for efficient portfolios with $\beta < \frac{2}{3}$ is simply the straight line in the $\beta-\mu$ plane through the points representing $T$ and the risk-free asset. The reason for this is that all of the efficient portfolios with expected return less than $E[R_T]$ lie on the straight line in the $\sigma-\mu$ plane through the points $(0, 3\%), (\sigma_T, \mu_T)$, and so $T$ plays the role of the market portfolio in this case. The equation of the line in the $\beta-\mu$ plane through the points representing $T$ and the risk-free asset is in general
$$E[R] = r_f + \beta \frac{E[R_T] - r_f}{\beta_T}.$$

Substituting the values of $E[R_T]$ and $\beta_T$ determined in part c and the given value of the risk-free rate we have
$$E[R] = 3 + 12.375 \beta$$

for $\beta \leq \frac{2}{3}$.

Consequently, the security market line for efficient portfolios is given by
$$E[R] = \begin{cases} 
3 + 12.375 \beta & \text{for } \beta \leq \frac{2}{3}, \\
3.75 + 11.25 \beta & \text{for } \beta \geq \frac{2}{3}.
\end{cases}$$
The security market line for *individual risky assets* is given by the general form of the CAPM for all values of $\beta$:

$$E[R] = E[R_{Z(M)}] + \beta (E[R_M] - E[R_{Z(M)})].$$

Substituting the values of $E[R_M]$ and $E[R_{Z(M)}]$ into this equation we obtain

$$E[R] = 3.75 + 11.25 \beta \quad \text{for all values of } \beta.$$

This equation provides information on the "required return" for a risky asset to be a candidate for investment. Note that the "required return" is higher for an individual risky asset of risk level $\beta$ (with $\beta < \frac{2}{3}$) than it is for an efficient portfolio of risk level $\beta$.

This makes perfect sense since an efficient portfolio with $\beta < \frac{2}{3}$ has a return component (the risk-free component) that is completely certain.