Yates Probability 3rd Edition solutions

Probability for Electrical and Computer Engineering (Louisiana State University)

Probability and Stochastic Processes

A Friendly Introduction for Electrical and Computer Engineers

Third Edition

INSTRUCTOR'S SOLUTION MANUAL

Roy D. Yates, David J. Goodman, David Famolari

September 8, 2014

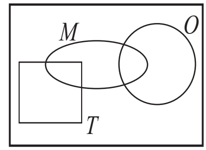
Comments on this Solutions Manual

- This solution manual is mostly complete. Please send error reports, suggestions, and comments to ***ryates@winlab.rutgers.edu***.
- To make solution sets for your class, use the *Solution Set Constructor* at the instructors site ***www.winlab.rutgers.edu/probsolns***.
- Send email to ***ryates@winlab.rutgers.edu*** for access to the instructors site.
- Matlab functions written as solutions to homework problems can be found in the archive `matsoln3e.zip` (available to instructors). Other Matlab functions used in the text or in these homework solutions can be found in the archive `matcode3e.zip`. The .m files in `matcode3e` are available for download from the Wiley website. Two other documents of interest are also available for download:
  - A manual `probmatlab3e.pdf` describing the `matcode3e .m` functions
  - The quiz solutions manual `quizsol.pdf`.
- This manual uses a page size matched to the screen of an iPad tablet. If you do print on paper and you have good eyesight, you may wish to print two pages per sheet in landscape mode. On the other hand, a "Fit to Paper" printing option will create "Large Print" output.

Problem Solutions – Chapter 1

Problem 1.1.1 Solution

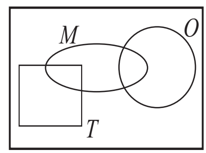
Based on the Venn diagram on the right, the complete Gerlandas pizza menu is



- Regular without toppings
- Regular with mushrooms
- Regular with onions
- Regular with mushrooms and onions
- Tuscan without toppings
- Tuscan with mushrooms

Problem 1.1.2 Solution

Based on the Venn diagram on the right, the answers are mostly fairly straightforward. The only trickiness is that a pizza is either Tuscan (T) or Neapolitan (N) so $\{N, T\}$ is a partition but they are not depicted as a partition. Specifically, the event N is the region of the Venn diagram outside of the "square block" of event T. If this is clear, the questions are easy.



(a) Since $N = T^c$, $N \cap M \neq \phi$. Thus N and M are not mutually exclusive.

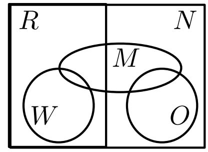
(b) Every pizza is either Neapolitan (N), or Tuscan (T). Hence $N \cup T = S$ so that N and T are collectively exhaustive. Thus its also (trivially) true that $N \cup T \cup M = S$. That is, R, T and M are also collectively exhaustive.

(c) From the Venn diagram, T and O are mutually exclusive. In words, this means that Tuscan pizzas never have onions or pizzas with onions are never Tuscan. As an aside, "Tuscan" is a fake pizza designation; one shouldn't conclude that people from Tuscany actually dislike onions.

(d) From the Venn diagram, $M \cap T$ and O are mutually exclusive. Thus Gerlanda's doesn't make Tuscan pizza with mushrooms and onions.

(e) Yes. In terms of the Venn diagram, these pizzas are in the set $(T \cup M \cup O)^c$.

Problem 1.1.3 Solution



At Ricardo's, the pizza crust is either Roman (R) or Neapolitan (N). To draw the Venn diagram on the right, we make the following observations:

- The set $\{R, N\}$ is a partition so we can draw the Venn diagram with this partition.
- Only Roman pizzas can be white. Hence $W \subset R$.
- Only a Neapolitan pizza can have onions. Hence $O \subset N$.
- Both Neapolitan and Roman pizzas can have mushrooms so that event M straddles the $\{R, N\}$ partition.
- The Neapolitan pizza can have both mushrooms and onions so $M \cap O$ cannot be empty.
- The problem statement does not preclude putting mushrooms on a white Roman pizza. Hence the intersection $W \cap M$ should not be empty.

Problem 1.2.1 Solution

(a) An outcome specifies whether the connection speed is high (h), medium (m), or low (l) speed, and whether the signal is a mouse click (c) or a tweet (t). The sample space is

$$S = \{ht, hc, mt, mc, lt, lc\}. \tag{1}$$

(b) The event that the wi-fi connection is medium speed is $A_1 = \{mt, mc\}$.

(c) The event that a signal is a mouse click is $A_2 = \{hc, mc, lc\}$.

(d) The event that a connection is either high speed or low speed is $A_3 = \{ht, hc, lt, lc\}$.

(e) Since $A_1 \cap A_2 = \{mc\}$ and is not empty, A_1, A_2, and A_3 are not mutually exclusive.

(f) Since

$$A_1 \cup A_2 \cup A_3 = \{ht, hc, mt, mc, lt, lc\} = S, \tag{2}$$

the collection A_1, A_2, A_3 is collectively exhaustive.

Problem 1.2.2 Solution

(a) The sample space of the experiment is

$$S = \{aaa, aaf, afa, faa, ffa, faf, aff, fff\}. \tag{1}$$

(b) The event that the circuit from Z fails is

$$Z_F = \{aaf, aff, faf, fff\}. \tag{2}$$

The event that the circuit from X is acceptable is

$$X_A = \{aaa, aaf, afa, aff\}. \tag{3}$$

(c) Since $Z_F \cap X_A = \{aaf, aff\} \neq \phi$, Z_F and X_A are not mutually exclusive.

(d) Since $Z_F \cup X_A = \{aaa, aaf, afa, aff, faf, fff\} \neq S$, Z_F and X_A are not collectively exhaustive.

(e) The event that more than one circuit is acceptable is

$$C = \{aaa, aaf, afa, faa\}. \tag{4}$$

The event that at least two circuits fail is

$$D = \{ffa, faf, aff, fff\}. \tag{5}$$

(f) Inspection shows that $C \cap D = \phi$ so C and D are mutually exclusive.

(g) Since $C \cup D = S$, C and D are collectively exhaustive.

Problem 1.2.3 Solution

The sample space is

$$S = \{A\clubsuit, \ldots, K\clubsuit, A\diamondsuit, \ldots, K\diamondsuit, A\heartsuit, \ldots, K\heartsuit, A\spadesuit, \ldots, K\spadesuit\}. \tag{1}$$

The event H that the first card is a heart is the set

$$H = \{A\heartsuit, \ldots, K\heartsuit\}. \tag{2}$$

The event H has 13 outcomes, corresponding to the 13 hearts in a deck.

Problem 1.2.4 Solution

The sample space is

$$S = \left\{ \begin{array}{l} 1/1 \ldots 1/31, 2/1 \ldots 2/29, 3/1 \ldots 3/31, 4/1 \ldots 4/30, \\ 5/1 \ldots 5/31, 6/1 \ldots 6/30, 7/1 \ldots 7/31, 8/1 \ldots 8/31, \\ 9/1 \ldots 9/31, 10/1 \ldots 10/31, 11/1 \ldots 11/30, 12/1 \ldots 12/31 \end{array} \right\}. \tag{1}$$

The event H defined by the event of a July birthday is given by the following et with 31 sample outcomes:

$$H = \{7/1, 7/2, \ldots, 7/31\}. \tag{2}$$

Problem 1.2.5 Solution

Of course, there are many answers to this problem. Here are four partitions.

1. We can divide students into engineers or non-engineers. Let A_1 equal the set of engineering students and A_2 the non-engineers. The pair $\{A_1, A_2\}$ is a partition.

2. To separate students by GPA, let B_i denote the subset of students with GPAs G satisfying $i-1 \le G < i$. At Rutgers, $\{B_1, B_2, \ldots, B_5\}$ is a partition. Note that B_5 is the set of all students with perfect 4.0 GPAs. Of course, other schools use different scales for GPA.

3. We can also divide the students by age. Let C_i denote the subset of students of age i in years. At most universities, $\{C_{10}, C_{11}, \ldots, C_{100}\}$ would be an event space. Since a university may have prodigies either under 10 or over 100, we note that $\{C_0, C_1, \ldots\}$ is always a partition.

4. Lastly, we can categorize students by attendance. Let D_0 denote the number of students who have missed zero lectures and let D_1 denote all other students. Although it is likely that D_0 is an empty set, $\{D_0, D_1\}$ is a well defined partition.

Problem 1.2.6 Solution

Let R_1 and R_2 denote the measured resistances. The pair (R_1, R_2) is an outcome of the experiment. Some partitions include

1. If we need to check that neither resistance is too high, a partition is

$$A_1 = \{R_1 < 100, R_2 < 100\}, \qquad A_2 = \{R_1 \geq 100\} \cup \{R_2 \geq 100\}. \tag{1}$$

2. If we need to check whether the first resistance exceeds the second resistance, a partition is

$$B_1 = \{R_1 > R_2\} \qquad B_2 = \{R_1 \leq R_2\}. \tag{2}$$

3. If we need to check whether each resistance doesn't fall below a minimum value (in this case 50 ohms for R_1 and 100 ohms for R_2), a partition is C_1, C_2, C_3, C_4 where

$$C_1 = \{R_1 < 50, R_2 < 100\}, \qquad C_2 = \{R_1 < 50, R_2 \geq 100\}, \tag{3}$$
$$C_3 = \{R_1 \geq 50, R_2 < 100\}, \qquad C_4 = \{R_1 \geq 50, R_2 \geq 100\}. \tag{4}$$

4. If we want to check whether the resistors in parallel are within an acceptable range of 90 to 110 ohms, a partition is

$$D_1 = \left\{(1/R_1 + 1/R_2)^{-1} < 90\right\}, \tag{5}$$
$$D_2 = \left\{90 \leq (1/R_1 + 1/R_2)^{-1} \leq 110\right\}, \tag{6}$$
$$D_2 = \left\{110 < (1/R_1 + 1/R_2)^{-1}\right\}. \tag{7}$$

Problem 1.3.1 Solution

(a) A and B mutually exclusive and collectively exhaustive imply $P[A]+P[B] = 1$. Since $P[A] = 3\,P[B]$, we have $P[B] = 1/4$.

(b) Since $P[A \cup B] = P[A]$, we see that $B \subseteq A$. This implies $P[A \cap B] = P[B]$. Since $P[A \cap B] = 0$, then $P[B] = 0$.

(c) Since it's always true that $P[A \cup B] = P[A] + P[B] - P[AB]$, we have that

$$P[A] + P[B] - P[AB] = P[A] - P[B]. \tag{1}$$

This implies $2\,P[B] = P[AB]$. However, since $AB \subset B$, we can conclude that $2\,P[B] = P[AB] \leq P[B]$. This implies $P[B] = 0$.

Problem 1.3.2 Solution

The roll of the red and white dice can be assumed to be independent. For each die, all rolls in $\{1, 2, \ldots, 6\}$ have probability $1/6$.

(a) Thus

$$P[R_3W_2] = P[R_3]\,P[W_2] = \frac{1}{36}.$$

(b) In fact, each pair of possible rolls R_iW_j has probability $1/36$. To find $P[S_5]$, we add up the probabilities of all pairs that sum to 5:

$$P[S_5] = P[R_1W_4] + P[R_2W_3] + P[R_3W_2] + P[R_4W_1] = 4/36 = 1/9.$$

Problem 1.3.3 Solution

An outcome is a pair (i, j) where i is the value of the first die and j is the value of the second die. The sample space is the set

$$S = \{(1,1), (1,2), \ldots, (6,5), (6,6)\}\,. \tag{1}$$

with 36 outcomes, each with probability $1/36$ Note that the event that the absolute value of the difference of the two rolls equals 3 is

$$D_3 = \{(1,4), (2,5), (3,6), (4,1), (5,2), (6,3)\}\,. \tag{2}$$

Since there are 6 outcomes in D_3, $P[D_3] = 6/36 = 1/6$.

Problem 1.3.4 Solution

(a) FALSE. Since $P[A] = 1 - P[A^c] = 2\,P[A^c]$ implies $P[A^c] = 1/3$.

(b) FALSE. Suppose $A = B$ and $P[A] = 1/2$. In that case,

$$P[AB] = P[A] = 1/2 > 1/4 = P[A]\,P[B]. \tag{1}$$

(c) TRUE. Since $AB \subseteq A$, $P[AB] \leq P[A]$, This implies

$$P[AB] \leq P[A] < P[B]. \tag{2}$$

(d) FALSE: For a counterexample, let $A = \phi$ and $P[B] > 0$ so that $A = A \cap B = \phi$ and $P[A] = P[A \cap B] = 0$ but $0 = P[A] < P[B]$.

Problem 1.3.5 Solution

The sample space of the experiment is

$$S = \{LF, BF, LW, BW\}. \tag{1}$$

From the problem statement, we know that $P[LF] = 0.5$, $P[BF] = 0.2$ and $P[BW] = 0.2$. This implies $P[LW] = 1 - 0.5 - 0.2 - 0.2 = 0.1$. The questions can be answered using Theorem 1.5.

(a) The probability that a program is slow is

$$P[W] = P[LW] + P[BW] = 0.1 + 0.2 = 0.3. \tag{2}$$

(b) The probability that a program is big is

$$P[B] = P[BF] + P[BW] = 0.2 + 0.2 = 0.4. \tag{3}$$

(c) The probability that a program is slow or big is

$$P[W \cup B] = P[W] + P[B] - P[BW] = 0.3 + 0.4 - 0.2 = 0.5. \tag{4}$$

Problem 1.3.6 Solution

A sample outcome indicates whether the cell phone is handheld (H) or mobile (M) and whether the speed is fast (F) or slow (W). The sample space is

$$S = \{HF, HW, MF, MW\}. \tag{1}$$

The problem statement tells us that $\mathrm{P}[HF] = 0.2$, $\mathrm{P}[MW] = 0.1$ and $\mathrm{P}[F] = 0.5$. We can use these facts to find the probabilities of the other outcomes. In particular,

$$\mathrm{P}[F] = \mathrm{P}[HF] + \mathrm{P}[MF]. \tag{2}$$

This implies

$$\mathrm{P}[MF] = \mathrm{P}[F] - \mathrm{P}[HF] = 0.5 - 0.2 = 0.3. \tag{3}$$

Also, since the probabilities must sum to 1,

$$\begin{aligned} \mathrm{P}[HW] &= 1 - \mathrm{P}[HF] - \mathrm{P}[MF] - \mathrm{P}[MW] \\ &= 1 - 0.2 - 0.3 - 0.1 = 0.4. \end{aligned} \tag{4}$$

Now that we have found the probabilities of the outcomes, finding any other probability is easy.

(a) The probability a cell phone is slow is

$$\mathrm{P}[W] = \mathrm{P}[HW] + \mathrm{P}[MW] = 0.4 + 0.1 = 0.5. \tag{5}$$

(b) The probability that a cell hpone is mobile and fast is $\mathrm{P}[MF] = 0.3$.

(c) The probability that a cell phone is handheld is

$$\mathrm{P}[H] = \mathrm{P}[HF] + \mathrm{P}[HW] = 0.2 + 0.4 = 0.6. \tag{6}$$

Problem 1.3.7 Solution

A reasonable probability model that is consistent with the notion of a shuffled deck is that each card in the deck is equally likely to be the first card. Let H_i denote the event that the first card drawn is the ith heart where the first heart is the ace, the second heart is the deuce and so on. In that case, $P[H_i] = 1/52$ for $1 \leq i \leq 13$. The event H that the first card is a heart can be written as the mutually exclusive union

$$H = H_1 \cup H_2 \cup \cdots \cup H_{13}. \tag{1}$$

Using Theorem 1.1, we have

$$P[H] = \sum_{i=1}^{13} P[H_i] = 13/52. \tag{2}$$

This is the answer you would expect since 13 out of 52 cards are hearts. The point to keep in mind is that this is not just the common sense answer but is the result of a probability model for a shuffled deck and the axioms of probability.

Problem 1.3.8 Solution

Let s_i denote the outcome that the down face has i dots. The sample space is $S = \{s_1, \ldots, s_6\}$. The probability of each sample outcome is $P[s_i] = 1/6$. From Theorem 1.1, the probability of the event E that the roll is even is

$$P[E] = P[s_2] + P[s_4] + P[s_6] = 3/6. \tag{1}$$

Problem 1.3.9 Solution

Let s_i equal the outcome of the student's quiz. The sample space is then composed of all the possible grades that she can receive.

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}. \tag{1}$$

Since each of the 11 possible outcomes is equally likely, the probability of receiving a grade of i, for each $i = 0, 1, \ldots, 10$ is $P[s_i] = 1/11$. The probability that the student gets an A is the probability that she gets a score of 9 or higher. That is

$$P[\text{Grade of A}] = P[9] + P[10] = 1/11 + 1/11 = 2/11. \tag{2}$$

The probability of failing requires the student to get a grade less than 4.

$$\begin{aligned} P[\text{Failing}] &= P[3] + P[2] + P[1] + P[0] \\ &= 1/11 + 1/11 + 1/11 + 1/11 = 4/11. \end{aligned} \tag{3}$$

Problem 1.3.10 Solution

Each statement is a consequence of part 4 of Theorem 1.4.

(a) Since $A \subset A \cup B$, $\mathrm{P}[A] \leq \mathrm{P}[A \cup B]$.

(b) Since $B \subset A \cup B$, $\mathrm{P}[B] \leq \mathrm{P}[A \cup B]$.

(c) Since $A \cap B \subset A$, $\mathrm{P}[A \cap B] \leq \mathrm{P}[A]$.

(d) Since $A \cap B \subset B$, $\mathrm{P}[A \cap B] \leq \mathrm{P}[B]$.

Problem 1.3.11 Solution

Specifically, we will use Theorem 1.4(c) which states that for any events A and B,

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] - \mathrm{P}[A \cap B]. \tag{1}$$

To prove the union bound by induction, we first prove the theorem for the case of $n = 2$ events. In this case, by Theorem 1.4(c),

$$\mathrm{P}[A_1 \cup A_2] = \mathrm{P}[A_1] + \mathrm{P}[A_2] - \mathrm{P}[A_1 \cap A_2]. \tag{2}$$

By the first axiom of probability, $\mathrm{P}[A_1 \cap A_2] \geq 0$. Thus,

$$\mathrm{P}[A_1 \cup A_2] \leq \mathrm{P}[A_1] + \mathrm{P}[A_2]. \tag{3}$$

which proves the union bound for the case $n = 2$.Now we make our induction hypothesis that the union-bound holds for any collection of $n-1$ subsets. In this case, given subsets $A_1, \ldots, A_n$, we define

$$A = A_1 \cup A_2 \cup \cdots \cup A_{n-1}, \qquad B = A_n. \tag{4}$$

By our induction hypothesis,

$$\mathrm{P}[A] = \mathrm{P}[A_1 \cup A_2 \cup \cdots \cup A_{n-1}] \leq \mathrm{P}[A_1] + \cdots + \mathrm{P}[A_{n-1}]. \tag{5}$$

This permits us to write

$$\begin{aligned}
\mathrm{P}[A_1 \cup \cdots \cup A_n] &= \mathrm{P}[A \cup B] \\
&\leq \mathrm{P}[A] + \mathrm{P}[B] \qquad \text{(by the union bound for } n = 2\text{)} \\
&= \mathrm{P}[A_1 \cup \cdots \cup A_{n-1}] + \mathrm{P}[A_n] \\
&\leq \mathrm{P}[A_1] + \cdots \mathrm{P}[A_{n-1}] + \mathrm{P}[A_n]
\end{aligned} \tag{6}$$

which completes the inductive proof.

Problem 1.3.12 Solution

It is tempting to use the following proof:

> Since S and ϕ are mutually exclusive, and since $S = S \cup \phi$,
>
> $$1 = P[S \cup \phi] = P[S] + P[\phi].$$
>
> Since $P[S] = 1$, we must have $P[\phi] = 0$.

The above "proof" used the property that for mutually exclusive sets A_1 and A_2,

$$P[A_1 \cup A_2] = P[A_1] + P[A_2]. \tag{1}$$

The problem is that this property is a consequence of the three axioms, and thus must be proven. For a proof that uses just the three axioms, let A_1 be an arbitrary set and for $n = 2, 3, \ldots$, let $A_n = \phi$. Since $A_1 = \cup_{i=1}^{\infty} A_i$, we can use Axiom 3 to write

$$P[A_1] = P[\cup_{i=1}^{\infty} A_i] = P[A_1] + P[A_2] + \sum_{i=3}^{\infty} P[A_i]. \tag{2}$$

By subtracting $P[A_1]$ from both sides, the fact that $A_2 = \phi$ permits us to write

$$P[\phi] + \sum_{n=3}^{\infty} P[A_i] = 0. \tag{3}$$

By Axiom 1, $P[A_i] \geq 0$ for all i. Thus, $\sum_{n=3}^{\infty} P[A_i] \geq 0$. This implies $P[\phi] \leq 0$. Since Axiom 1 requires $P[\phi] \geq 0$, we must have $P[\phi] = 0$.

Problem 1.3.13 Solution

Following the hint, we define the set of events $\{A_i | i = 1, 2, \ldots\}$ such that $i = 1, \ldots, m$, $A_i = B_i$ and for $i > m$, $A_i = \phi$. By construction, $\cup_{i=1}^{m} B_i = \cup_{i=1}^{\infty} A_i$. Axiom 3 then implies

$$P[\cup_{i=1}^{m} B_i] = P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]. \tag{1}$$

For $i > m$, $\text{P}[A_i] = \text{P}[\phi] = 0$, yielding the claim $\text{P}[\cup_{i=1}^m B_i] = \sum_{i=1}^m \text{P}[A_i] = \sum_{i=1}^m \text{P}[B_i]$.

Note that the fact that $\text{P}[\phi] = 0$ follows from Axioms 1 and 2. This problem is more challenging if you just use Axiom 3. We start by observing

$$\text{P}\left[\cup_{i=1}^m B_i\right] = \sum_{i=1}^{m-1} \text{P}\left[B_i\right] + \sum_{i=m}^{\infty} \text{P}\left[A_i\right]. \tag{2}$$

Now, we use Axiom 3 again on the countably infinite sequence $A_m, A_{m+1}, \ldots$ to write

$$\sum_{i=m}^{\infty} \text{P}\left[A_i\right] = \text{P}\left[A_m \cup A_{m+1} \cup \cdots\right] = \text{P}\left[B_m\right]. \tag{3}$$

Thus, we have used just Axiom 3 to prove Theorem 1.3:

$$\text{P}\left[\cup_{i=1}^m B_i\right] = \sum_{i=1}^{m} \text{P}\left[B_i\right]. \tag{4}$$

Problem 1.3.14 Solution

Each claim in Theorem 1.4 requires a proof from which we can check which axioms are used. However, the problem is somewhat hard because there may still be a simpler proof that uses fewer axioms. Still, the proof of each part will need Theorem 1.3 which we now prove.

For the mutually exclusive events $B_1, \ldots, B_m$, let $A_i = B_i$ for $i = 1, \ldots, m$ and let $A_i = \phi$ for $i > m$. In that case, by Axiom 3,

$$\begin{aligned}
\text{P}\left[B_1 \cup B_2 \cup \cdots \cup B_m\right] &= \text{P}\left[A_1 \cup A_2 \cup \cdots\right] \\
&= \sum_{i=1}^{m-1} \text{P}\left[A_i\right] + \sum_{i=m}^{\infty} \text{P}\left[A_i\right] \\
&= \sum_{i=1}^{m-1} \text{P}\left[B_i\right] + \sum_{i=m}^{\infty} \text{P}\left[A_i\right]. \qquad (1)
\end{aligned}$$

Now, we use Axiom 3 again on $A_m, A_{m+1}, \ldots$ to write

$$\sum_{i=m}^{\infty} \mathrm{P}\left[A_i\right] = \mathrm{P}\left[A_m \cup A_{m+1} \cup \cdots\right] = \mathrm{P}\left[B_m\right]. \tag{2}$$

Thus, we have used just Axiom 3 to prove Theorem 1.3:

$$\mathrm{P}\left[B_1 \cup B_2 \cup \cdots \cup B_m\right] = \sum_{i=1}^{m} \mathrm{P}\left[B_i\right]. \tag{3}$$

(a) To show $\mathrm{P}[\phi] = 0$, let $B_1 = S$ and let $B_2 = \phi$. Thus by Theorem 1.3,

$$\mathrm{P}\left[S\right] = \mathrm{P}\left[B_1 \cup B_2\right] = \mathrm{P}\left[B_1\right] + \mathrm{P}\left[B_2\right] = \mathrm{P}\left[S\right] + \mathrm{P}\left[\phi\right]. \tag{4}$$

Thus, $\mathrm{P}[\phi] = 0$. Note that this proof uses only Theorem 1.3 which uses only Axiom 3.

(b) Using Theorem 1.3 with $B_1 = A$ and $B_2 = A^c$, we have

$$\mathrm{P}\left[S\right] = \mathrm{P}\left[A \cup A^c\right] = \mathrm{P}\left[A\right] + \mathrm{P}\left[A^c\right]. \tag{5}$$

Since, Axiom 2 says $\mathrm{P}[S] = 1$, $\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$. This proof uses Axioms 2 and 3.

(c) By Theorem 1.8, we can write both A and B as unions of mutually exclusive events:

$$A = (AB) \cup (AB^c), \qquad B = (AB) \cup (A^cB). \tag{6}$$

Now we apply Theorem 1.3 to write

$$\mathrm{P}\left[A\right] = \mathrm{P}\left[AB\right] + \mathrm{P}\left[AB^c\right], \qquad \mathrm{P}\left[B\right] = \mathrm{P}\left[AB\right] + \mathrm{P}\left[A^cB\right]. \tag{7}$$

We can rewrite these facts as

$$\mathrm{P}[AB^c] = \mathrm{P}[A] - \mathrm{P}[AB], \qquad \mathrm{P}[A^cB] = \mathrm{P}[B] - \mathrm{P}[AB]. \tag{8}$$

Note that so far we have used only Axiom 3. Finally, we observe that $A \cup B$ can be written as the union of mutually exclusive events

$$A \cup B = (AB) \cup (AB^c) \cup (A^cB). \tag{9}$$

Once again, using Theorem 1.3, we have

$$\mathrm{P}[A \cup B] = \mathrm{P}[AB] + \mathrm{P}[AB^c] + \mathrm{P}[A^cB] \tag{10}$$

Substituting the results of Equation (8) into Equation (10) yields

$$\mathrm{P}\left[A \cup B\right] = \mathrm{P}\left[AB\right] + \mathrm{P}\left[A\right] - \mathrm{P}\left[AB\right] + \mathrm{P}\left[B\right] - \mathrm{P}\left[AB\right], \tag{11}$$

which completes the proof. Note that this claim required only Axiom 3.

(d) Observe that since $A \subset B$, we can write B as the mutually exclusive union $B = A \cup (A^cB)$. By Theorem 1.3 (which uses Axiom 3),

$$\mathrm{P}\left[B\right] = \mathrm{P}\left[A\right] + \mathrm{P}\left[A^cB\right]. \tag{12}$$

By Axiom 1, $\mathrm{P}[A^cB] \geq 0$, hich implies $\mathrm{P}[A] \leq \mathrm{P}[B]$. This proof uses Axioms 1 and 3.

Problem 1.4.1 Solution

Each question requests a conditional probability.

(a) Note that the probability a call is brief is

$$\mathrm{P}\left[B\right] = \mathrm{P}\left[H_0B\right] + \mathrm{P}\left[H_1B\right] + \mathrm{P}\left[H_2B\right] = 0.6. \tag{1}$$

The probability a brief call will have no handoffs is

$$\mathrm{P}\left[H_0|B\right] = \frac{\mathrm{P}\left[H_0B\right]}{\mathrm{P}\left[B\right]} = \frac{0.4}{0.6} = \frac{2}{3}. \tag{2}$$

(b) The probability of one handoff is $\mathrm{P}[H_1] = \mathrm{P}[H_1B] + \mathrm{P}[H_1L] = 0.2$. The probability that a call with one handoff will be long is

$$\mathrm{P}\left[L|H_1\right] = \frac{\mathrm{P}\left[H_1L\right]}{\mathrm{P}\left[H_1\right]} = \frac{0.1}{0.2} = \frac{1}{2}. \tag{3}$$

(c) The probability a call is long is $\mathrm{P}[L] = 1 - \mathrm{P}[B] = 0.4$. The probability that a long call will have one or more handoffs is

$$\begin{aligned}\mathrm{P}\left[H_1 \cup H_2|L\right] &= \frac{\mathrm{P}\left[H_1L \cup H_2L\right]}{\mathrm{P}\left[L\right]} \\ &= \frac{\mathrm{P}\left[H_1L\right] + \mathrm{P}\left[H_2L\right]}{\mathrm{P}\left[L\right]} = \frac{0.1+0.2}{0.4} = \frac{3}{4}.\end{aligned} \tag{4}$$

Problem 1.4.2 Solution

Let s_i denote the outcome that the roll is i. So, for $1 \leq i \leq 6$, $R_i = \{s_i\}$. Similarly, $G_j = \{s_{j+1}, \ldots, s_6\}$.

(a) Since $G_1 = \{s_2, s_3, s_4, s_5, s_6\}$ and all outcomes have probability 1/6, $P[G_1] = 5/6$. The event $R_3G_1 = \{s_3\}$ and $P[R_3G_1] = 1/6$ so that

$$P[R_3|G_1] = \frac{P[R_3G_1]}{P[G_1]} = \frac{1}{5}. \tag{1}$$

(b) The conditional probability that 6 is rolled given that the roll is greater than 3 is

$$P[R_6|G_3] = \frac{P[R_6G_3]}{P[G_3]} = \frac{P[s_6]}{P[s_4, s_5, s_6]} = \frac{1/6}{3/6}. \tag{2}$$

(c) The event E that the roll is even is $E = \{s_2, s_4, s_6\}$ and has probability 3/6. The joint probability of G_3 and E is

$$P[G_3E] = P[s_4, s_6] = 1/3. \tag{3}$$

The conditional probabilities of G_3 given E is

$$P[G_3|E] = \frac{P[G_3E]}{P[E]} = \frac{1/3}{1/2} = \frac{2}{3}. \tag{4}$$

(d) The conditional probability that the roll is even given that it's greater than 3 is

$$P[E|G_3] = \frac{P[EG_3]}{P[G_3]} = \frac{1/3}{1/2} = \frac{2}{3}. \tag{5}$$

Problem 1.4.3 Solution

Since the 2 of clubs is an even numbered card, $C_2 \subset E$ so that $P[C_2E] = P[C_2] = 1/3$. Since $P[E] = 2/3$,

$$P[C_2|E] = \frac{P[C_2E]}{P[E]} = \frac{1/3}{2/3} = 1/2. \tag{1}$$

The probability that an even numbered card is picked given that the 2 is picked is

$$P[E|C_2] = \frac{P[C_2E]}{P[C_2]} = \frac{1/3}{1/3} = 1. \tag{2}$$

Problem 1.4.4 Solution

Let A_i and B_i denote the events that the ith phone sold is an Apricot or a Banana respectively. Our goal is to find $\mathrm{P}[B_1B_2]$, but since it is not clear where to start, we should plan on filling in the table

	A_2	B_2
A_1		
B_1		

This table has four unknowns: $\mathrm{P}[A_1A_2]$, $\mathrm{P}[A_1B_2]$, $\mathrm{P}[B_1A_2]$, and $\mathrm{P}[B_1B_2]$. We start knowing that

$$\mathrm{P}[A_1A_2] + \mathrm{P}[A_1B_2] + \mathrm{P}[B_1A_2] + \mathrm{P}[B_1B_2] = 1. \tag{1}$$

We still need three more equations to solve for the four unknowns. From "sales of Apricots and Bananas are equally likely," we know that $\mathrm{P}[A_i] = \mathrm{P}[B_i] = 1/2$ for $i = 1, 2$. This implies

$$\mathrm{P}[A_1] = \mathrm{P}[A_1A_2] + \mathrm{P}[A_1B_2] = 1/2, \tag{2}$$
$$\mathrm{P}[A_2] = \mathrm{P}[A_1A_2] + \mathrm{P}[B_1A_2] = 1/2. \tag{3}$$

The final equation comes from "given that the first phone sold is a Banana, the second phone is twice as likely to be a Banana," which implies $\mathrm{P}[B_2|B_1] = 2\,\mathrm{P}[A_2|B_1]$. Using Bayes' theorem, we have

$$\frac{\mathrm{P}[B_1B_2]}{\mathrm{P}[B_1]} = 2\frac{\mathrm{P}[B_1A_2]}{\mathrm{P}[B_1]} \qquad \Longrightarrow \qquad \mathrm{P}[B_1A_2] = \frac{1}{2}\,\mathrm{P}[B_1B_2]. \tag{4}$$

Replacing $\mathrm{P}[B_1A_2]$ with $\mathrm{P}[B_1B_2]/2$ in the the first three equations yields

$$\mathrm{P}[A_1A_2] + \mathrm{P}[A_1B_2] + \frac{3}{2}\,\mathrm{P}[B_1B_2] = 1, \tag{5}$$
$$\mathrm{P}[A_1A_2] + \mathrm{P}[A_1B_2] = 1/2, \tag{6}$$
$$\mathrm{P}[A_1A_2] + \frac{1}{2}\,\mathrm{P}[B_1B_2] = 1/2. \tag{7}$$

Subtracting (6) from (5) yields $(3/2)\,\mathrm{P}[B_1B_2] = 1/2$, or $\mathrm{P}[B_1B_2] = 1/3$, which is the answer we are looking for.

At this point, if you are curious, we can solve for the rest of the probability table. From (4), we have $\mathrm{P}[B_1A_2] = 1/6$ and from (7) we obtain $\mathrm{P}[A_1A_2] = 1/3$. It then follows from (6) that $\mathrm{P}[A_1B_2] = 1/6$. The probability table is

$$\begin{array}{c|cc} & A_2 & B_2 \\ \hline A_1 & 1/3 & 1/6 \\ B_1 & 1/6 & 1/3 \end{array}.$$

Problem 1.4.5 Solution

The first generation consists of two plants each with genotype yg or gy. They are crossed to produce the following second generation genotypes, $S = \{yy, yg, gy, gg\}$. Each genotype is just as likely as any other so the probability of each genotype is consequently 1/4. A pea plant has yellow seeds if it possesses at least one dominant y gene. The set of pea plants with yellow seeds is

$$Y = \{yy, yg, gy\}. \tag{1}$$

So the probability of a pea plant with yellow seeds is

$$\mathrm{P}[Y] = \mathrm{P}[yy] + \mathrm{P}[yg] + \mathrm{P}[gy] = 3/4. \tag{2}$$

Problem 1.4.6 Solution

Define D as the event that a pea plant has two dominant y genes. To find the conditional probability of D given the event Y, corresponding to a plant having yellow seeds, we look to evaluate

$$\mathrm{P}[D|Y] = \frac{\mathrm{P}[DY]}{\mathrm{P}[Y]}. \tag{1}$$

Note that $\mathrm{P}[DY]$ is just the probability of the genotype yy. From Problem 1.4.5, we found that with respect to the color of the peas, the genotypes yy, yg, gy, and gg were all equally likely. This implies

$$\mathrm{P}[DY] = \mathrm{P}[yy] = 1/4 \qquad \mathrm{P}[Y] = \mathrm{P}[yy, gy, yg] = 3/4. \tag{2}$$

Thus, the conditional probability can be expressed as

$$\mathrm{P}[D|Y] = \frac{\mathrm{P}[DY]}{\mathrm{P}[Y]} = \frac{1/4}{3/4} = 1/3. \tag{3}$$

Problem 1.4.7 Solution

The sample outcomes can be written ijk where the first card drawn is i, the second is j and the third is k. The sample space is

$$S = \{234, 243, 324, 342, 423, 432\}. \tag{1}$$

and each of the six outcomes has probability 1/6. The events E_1, E_2, E_3, O_1, O_2, O_3 are

$$E_1 = \{234, 243, 423, 432\}, \qquad O_1 = \{324, 342\}, \tag{2}$$

$$E_2 = \{243, 324, 342, 423\}, \qquad O_2 = \{234, 432\}, \tag{3}$$

$$E_3 = \{234, 324, 342, 432\}, \qquad O_3 = \{243, 423\}. \tag{4}$$

(a) The conditional probability the second card is even given that the first card is even is

$$\mathrm{P}[E_2|E_1] = \frac{\mathrm{P}[E_2E_1]}{\mathrm{P}[E_1]} = \frac{\mathrm{P}[243, 423]}{\mathrm{P}[234, 243, 423, 432]} = \frac{2/6}{4/6} = 1/2. \tag{5}$$

(b) The conditional probability the first card is even given that the second card is even is

$$\mathrm{P}[E_1|E_2] = \frac{\mathrm{P}[E_1E_2]}{\mathrm{P}[E_2]} = \frac{\mathrm{P}[243, 423]}{\mathrm{P}[243, 324, 342, 423]} = \frac{2/6}{4/6} = 1/2. \tag{6}$$

(c) The probability the first two cards are even given the third card is even is

$$\mathrm{P}[E_1E_2|E_3] = \frac{\mathrm{P}[E_1E_2E_3]}{\mathrm{P}[E_3]} = 0. \tag{7}$$

(d) The conditional probabilities the second card is even given that the first card is odd is

$$\mathrm{P}[E_2|O_1] = \frac{\mathrm{P}[O_1E_2]}{\mathrm{P}[O_1]} = \frac{\mathrm{P}[O_1]}{\mathrm{P}[O_1]} = 1. \tag{8}$$

(e) The conditional probability the second card is odd given that the first card is odd is

$$\mathrm{P}[O_2|O_1] = \frac{\mathrm{P}[O_1O_2]}{\mathrm{P}[O_1]} = 0. \tag{9}$$

Problem 1.4.8 Solution

The problem statement yields the obvious facts that $P[L] = 0.16$ and $P[H] = 0.10$. The words "10% of the ticks that had either Lyme disease or HGE carried both diseases" can be written as

$$P[LH|L \cup H] = 0.10. \tag{1}$$

(a) Since $LH \subset L \cup H$,

$$P[LH|L \cup H] = \frac{P[LH \cap (L \cup H)]}{P[L \cup H]} = \frac{P[LH]}{P[L \cup H]} = 0.10. \tag{2}$$

Thus,

$$P[LH] = 0.10\, P[L \cup H] = 0.10\,(P[L] + P[H] - P[LH]). \tag{3}$$

Since $P[L] = 0.16$ and $P[H] = 0.10$,

$$P[LH] = \frac{0.10\,(0.16 + 0.10)}{1.1} = 0.0236. \tag{4}$$

(b) The conditional probability that a tick has HGE given that it has Lyme disease is

$$P[H|L] = \frac{P[LH]}{P[L]} = \frac{0.0236}{0.16} = 0.1475. \tag{5}$$

Problem 1.5.1 Solution

From the table we look to add all the mutually exclusive events to find each probability.

(a) The probability that a caller makes no hand-offs is

$$P[H_0] = P[LH_0] + P[BH_0] = 0.1 + 0.4 = 0.5. \tag{1}$$

(b) The probability that a call is brief is

$$P[B] = P[BH_0] + P[BH_1] + P[BH_2] = 0.4 + 0.1 + 0.1 = 0.6. \tag{2}$$

(c) The probability that a call is long or makes at least two hand-offs is

$$\begin{aligned} P[L \cup H_2] &= P[LH_0] + P[LH_1] + P[LH_2] + P[BH_2] \\ &= 0.1 + 0.1 + 0.2 + 0.1 = 0.5. \end{aligned} \tag{3}$$

Problem 1.5.2 Solution

(a) From the given probability distribution of billed minutes, M, the probability that a call is billed for more than 3 minutes is

$$\begin{aligned} P[L] &= 1 - P[3 \text{ or fewer billed minutes}] \\ &= 1 - P[B_1] - P[B_2] - P[B_3] \\ &= 1 - \alpha - \alpha(1-\alpha) - \alpha(1-\alpha)^2 \\ &= (1-\alpha)^3 = 0.57. \end{aligned} \quad (1)$$

(b) The probability that a call will billed for 9 minutes or less is

$$P[9 \text{ minutes or less}] = \sum_{i=1}^{9} \alpha(1-\alpha)^{i-1} = 1 - (0.57)^3. \quad (2)$$

Problem 1.5.3 Solution

(a) For convenience, let $p_i = P[FH_i]$ and $q_i = P[VH_i]$. Using this shorthand, the six unknowns $p_0, p_1, p_2, q_0, q_1, q_2$ fill the table as

$$\begin{array}{c|ccc} & H_0 & H_1 & H_2 \\ \hline F & p_0 & p_1 & p_2 \\ V & q_0 & q_1 & q_2 \end{array}. \quad (1)$$

However, we are given a number of facts:

$$p_0 + q_0 = 1/3, \qquad p_1 + q_1 = 1/3, \quad (2)$$

$$p_2 + q_2 = 1/3, \qquad p_0 + p_1 + p_2 = 5/12. \quad (3)$$

Other facts, such as $q_0+q_1+q_2 = 7/12$, can be derived from these facts. Thus, we have four equations and six unknowns, choosing p_0 and p_1 will specify the other unknowns. Unfortunately, arbitrary choices for either p_0 or p_1 will lead to negative values for the other probabilities. In terms of p_0 and p_1, the other unknowns are

$$q_0 = 1/3 - p_0, \qquad p_2 = 5/12 - (p_0 + p_1), \tag{4}$$
$$q_1 = 1/3 - p_1, \qquad q_2 = p_0 + p_1 - 1/12. \tag{5}$$

Because the probabilities must be nonnegative, we see that

$$0 \le p_0 \le 1/3, \tag{6}$$
$$0 \le p_1 \le 1/3, \tag{7}$$
$$1/12 \le p_0 + p_1 \le 5/12. \tag{8}$$

Although there are an infinite number of solutions, three possible solutions are:

$$p_0 = 1/3, \qquad p_1 = 1/12, \qquad p_2 = 0, \tag{9}$$
$$q_0 = 0, \qquad q_1 = 1/4, \qquad q_2 = 1/3. \tag{10}$$

and

$$p_0 = 1/4, \qquad p_1 = 1/12, \qquad p_2 = 1/12, \tag{11}$$
$$q_0 = 1/12, \qquad q_1 = 3/12, \qquad q_2 = 3/12. \tag{12}$$

and

$$p_0 = 0, \qquad p_1 = 1/12, \qquad p_2 = 1/3, \tag{13}$$
$$q_0 = 1/3, \qquad q_1 = 3/12, \qquad q_2 = 0. \tag{14}$$

(b) In terms of the p_i, q_i notation, the new facts are $p_0 = 1/4$ and $q_1 = 1/6$. These extra facts uniquely specify the probabilities. In this case,

$$p_0 = 1/4, \qquad p_1 = 1/6, \qquad p_2 = 0, \tag{15}$$
$$q_0 = 1/12, \qquad q_1 = 1/6, \qquad q_2 = 1/3. \tag{16}$$

Problem 1.6.1 Solution

This problems asks whether A and B can be independent events yet satisfy $A = B$? By definition, events A and B are independent if and only if $\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B]$. We can see that if $A = B$, that is they are the same set, then

$$\mathrm{P}[AB] = \mathrm{P}[AA] = \mathrm{P}[A] = \mathrm{P}[B]. \tag{1}$$

Thus, for A and B to be the same set and also independent,

$$\mathrm{P}[A] = \mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B] = (\mathrm{P}[A])^2. \tag{2}$$

There are two ways that this requirement can be satisfied:

- $\mathrm{P}[A] = 1$ implying $A = B = S$.
- $\mathrm{P}[A] = 0$ implying $A = B = \phi$.

Problem 1.6.2 Solution

From the problem statement, we learn three facts:

$$\begin{aligned}
\mathrm{P}[AB] &= 0 && \text{(since A and B are mutually exclusive)} && (1)\\
\mathrm{P}[AB] &= \mathrm{P}[A]\,\mathrm{P}[B] && \text{(since A and B are independent)} && (2)\\
\mathrm{P}[A] &= \mathrm{P}[B] && \text{(since A and B are equiprobable)} && (3)
\end{aligned}$$

Applying these facts in the given order, we see that

$$0 = \mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B] = (\mathrm{P}[A])^2. \tag{4}$$

It follows that $\mathrm{P}[A] = 0$.

Problem 1.6.3 Solution

Let A_i and B_i denote the events that the ith phone sold is an Apricot or a Banana respectively. The works "each phone sold is twice as likely to be an Apricot than a Banana" tells us that

$$\mathrm{P}[A_i] = 2\,\mathrm{P}[B_i]. \tag{1}$$

However, since each phone sold is either an Apricot or a Banana, A_i and B_i are a partition and

$$P[A_i] + P[B_i] = 1. \tag{2}$$

Combining these equations, we have $P[A_i] = 2/3$ and $P[B_i] = 1/3$. The probability that two phones sold are the same is

$$P[A_1A_2 \cup B_1B_2] = P[A_1A_2] + P[B_1B_2]. \tag{3}$$

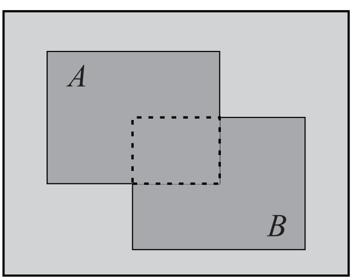
Since "each phone sale is independent,"

$$P[A_1A_2] = P[A_1]P[A_2] = \frac{4}{9}, \qquad P[B_1B_2] = P[B_1]P[B_2] = \frac{1}{9}. \tag{4}$$

Thus the probability that two phones sold are the same is

$$P[A_1A_2 \cup B_1B_2] = P[A_1A_2] + P[B_1B_2] = \frac{4}{9} + \frac{1}{9} = \frac{5}{9}. \tag{5}$$

Problem 1.6.4 Solution



In the Venn diagram, assume the sample space has area 1 corresponding to probability 1. As drawn, both A and B have area 1/4 so that $P[A] = P[B] = 1/4$. Moreover, the intersection AB has area 1/16 and covers 1/4 of A and 1/4 of B. That is, A and B are independent since

$$P[AB] = P[A]P[B]. \tag{1}$$

Problem 1.6.5 Solution

(a) Since A and B are mutually exclusive, $P[A \cap B] = 0$. Since $P[A \cap B] = 0$,

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = 3/8. \tag{1}$$

A Venn diagram should convince you that $A \subset B^c$ so that $A \cap B^c = A$. This implies

$$P[A \cap B^c] = P[A] = 1/4. \tag{2}$$

It also follows that $P[A \cup B^c] = P[B^c] = 1 - 1/8 = 7/8$.

(b) Events A and B are dependent since $P[AB] \neq P[A] P[B]$.

Problem 1.6.6 Solution

(a) Since C and D are independent,

$$P[C \cap D] = P[C] P[D] = 15/64. \tag{1}$$

The next few items are a little trickier. From Venn diagrams, we see

$$P[C \cap D^c] = P[C] - P[C \cap D] = 5/8 - 15/64 = 25/64. \tag{2}$$

It follows that

$$P[C \cup D^c] = P[C] + P[D^c] - P[C \cap D^c] \tag{3}$$

$$= 5/8 + (1 - 3/8) - 25/64 = 55/64. \tag{4}$$

Using DeMorgan's law, we have

$$P[C^c \cap D^c] = P[(C \cup D)^c] = 1 - P[C \cup D] = 15/64. \tag{5}$$

(b) Since $P[C^c D^c] = P[C^c] P[D^c]$, C^c and D^c are independent.

Problem 1.6.7 Solution

(a) Since $A \cap B = \emptyset$, $P[A \cap B] = 0$. To find $P[B]$, we can write

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \tag{1}$$

or

$$5/8 = 3/8 + P[B] - 0. \tag{2}$$

Thus, $P[B] = 1/4$. Since A is a subset of B^c, $P[A \cap B^c] = P[A] = 3/8$. Furthermore, since A is a subset of B^c, $P[A \cup B^c] = P[B^c] = 3/4$.

(b) The events A and B are dependent because

$$\mathrm{P}[AB] = 0 \neq 3/32 = \mathrm{P}[A]\,\mathrm{P}[B]. \tag{3}$$

Problem 1.6.8 Solution

(a) Since C and D are independent $\mathrm{P}[CD] = \mathrm{P}[C]\,\mathrm{P}[D]$. So

$$\mathrm{P}[D] = \frac{\mathrm{P}[CD]}{\mathrm{P}[C]} = \frac{1/3}{1/2} = 2/3. \tag{1}$$

In addition, $\mathrm{P}[C \cap D^c] = \mathrm{P}[C] - \mathrm{P}[C \cap D] = 1/2 - 1/3 = 1/6$. To find $\mathrm{P}[C^c \cap D^c]$, we first observe that

$$\mathrm{P}[C \cup D] = \mathrm{P}[C] + \mathrm{P}[D] - \mathrm{P}[C \cap D] = 1/2 + 2/3 - 1/3 = 5/6. \tag{2}$$

By De Morgan's Law, $C^c \cap D^c = (C \cup D)^c$. This implies

$$\mathrm{P}[C^c \cap D^c] = \mathrm{P}[(C \cup D)^c] = 1 - \mathrm{P}[C \cup D] = 1/6. \tag{3}$$

Note that a second way to find $\mathrm{P}[C^c \cap D^c]$ is to use the fact that if C and D are independent, then C^c and D^c are independent. Thus

$$\mathrm{P}[C^c \cap D^c] = \mathrm{P}[C^c]\,\mathrm{P}[D^c] = (1 - \mathrm{P}[C])(1 - \mathrm{P}[D]) = 1/6. \tag{4}$$

Finally, since C and D are independent events, $\mathrm{P}[C|D] = \mathrm{P}[C] = 1/2$.

(b) Note that we found $\mathrm{P}[C \cup D] = 5/6$. We can also use the earlier results to show

$$\begin{aligned} \mathrm{P}[C \cup D^c] &= \mathrm{P}[C] + \mathrm{P}[D] - \mathrm{P}[C \cap D^c] && (5)\\ &= 1/2 + (1 - 2/3) - 1/6 = 2/3. && (6) \end{aligned}$$

(c) By Definition 1.6, events C and D^c are independent because

$$\mathrm{P}[C \cap D^c] = 1/6 = (1/2)(1/3) = \mathrm{P}[C]\,\mathrm{P}[D^c]. \tag{7}$$

Problem 1.6.9 Solution

For a sample space $S = \{1, 2, 3, 4\}$ with equiprobable outcomes, consider the events

$$A_1 = \{1, 2\} \qquad A_2 = \{2, 3\} \qquad A_3 = \{3, 1\}. \tag{1}$$

Each event A_i has probability 1/2. Moreover, each pair of events is independent since

$$\mathrm{P}[A_1A_2] = \mathrm{P}[A_2A_3] = \mathrm{P}[A_3A_1] = 1/4. \tag{2}$$

However, the three events A_1, A_2, A_3 are not independent since

$$\mathrm{P}[A_1A_2A_3] = 0 \neq \mathrm{P}[A_1]\,\mathrm{P}[A_2]\,\mathrm{P}[A_3]. \tag{3}$$

Problem 1.6.10 Solution

There are 16 distinct equally likely outcomes for the second generation of pea plants based on a first generation of $\{rwyg, rwgy, wryg, wrgy\}$. These are:

$$\begin{array}{cccc} rryy & rryg & rrgy & rrgg \\ rwyy & rwyg & rwgy & rwgg \\ wryy & wryg & wrgy & wrgg \\ wwyy & wwyg & wwgy & wwgg \end{array}$$

A plant has yellow seeds, that is event Y occurs, if a plant has at least one dominant y gene. Except for the four outcomes with a pair of recessive g genes, the remaining 12 outcomes have yellow seeds. From the above, we see that

$$\mathrm{P}[Y] = 12/16 = 3/4 \tag{1}$$

and

$$\mathrm{P}[R] = 12/16 = 3/4. \tag{2}$$

To find the conditional probabilities $\mathrm{P}[R|Y]$ and $\mathrm{P}[Y|R]$, we first must find $\mathrm{P}[RY]$. Note that RY, the event that a plant has rounded yellow seeds, is the set of outcomes

$$RY = \{rryy, rryg, rrgy, rwyy, rwyg, rwgy, wryy, wryg, wrgy\}. \tag{3}$$

Since $\mathrm{P}[RY] = 9/16$,

$$\mathrm{P}[Y|R] = \frac{\mathrm{P}[RY]}{\mathrm{P}[R]} = \frac{9/16}{3/4} = 3/4 \tag{4}$$

and

$$\mathrm{P}[R|Y] = \frac{\mathrm{P}[RY]}{\mathrm{P}[Y]} = \frac{9/16}{3/4} = 3/4. \tag{5}$$

Thus $\mathrm{P}[R|Y] = \mathrm{P}[R]$ and $\mathrm{P}[Y|R] = \mathrm{P}[Y]$ and R and Y are independent events. There are four visibly different pea plants, corresponding to whether the peas are round (R) or not (R^c), or yellow (Y) or not (Y^c). These four visible events have probabilities

$$\mathrm{P}[RY] = 9/16 \qquad \mathrm{P}[RY^c] = 3/16, \tag{6}$$
$$\mathrm{P}[R^cY] = 3/16 \qquad \mathrm{P}[R^cY^c] = 1/16. \tag{7}$$

Problem 1.6.11 Solution

(a) For any events A and B, we can write the law of total probability in the form of

$$\mathrm{P}[A] = \mathrm{P}[AB] + \mathrm{P}[AB^c]. \tag{1}$$

Since A and B are independent, $\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B]$. This implies

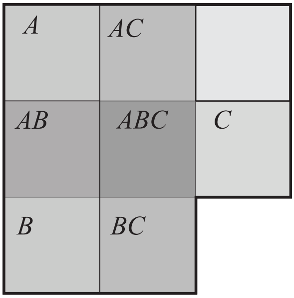
$$\mathrm{P}[AB^c] = \mathrm{P}[A] - \mathrm{P}[A]\,\mathrm{P}[B] = \mathrm{P}[A]\,(1 - \mathrm{P}[B]) = \mathrm{P}[A]\,\mathrm{P}[B^c]. \tag{2}$$

Thus A and B^c are independent.

(b) Proving that A^c and B are independent is not really necessary. Since A and B are arbitrary labels, it is really the same claim as in part (a). That is, simply reversing the labels of A and B proves the claim. Alternatively, one can construct exactly the same proof as in part (a) with the labels A and B reversed.

(c) To prove that A^c and B^c are independent, we apply the result of part (a) to the sets A and B^c. Since we know from part (a) that A and B^c are independent, part (b) says that A^c and B^c are independent.

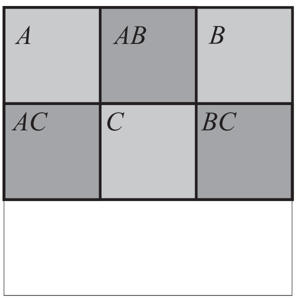
Problem 1.6.12 Solution



In the Venn diagram at right, assume the sample space has area 1 corresponding to probability 1. As drawn, A, B, and C each have area 1/2 and thus probability 1/2. Moreover, the three way intersection ABC has probability 1/8. Thus A, B, and C are mutually independent since

$$P[ABC] = P[A] P[B] P[C]. \tag{1}$$

Problem 1.6.13 Solution



In the Venn diagram at right, assume the sample space has area 1 corresponding to probability 1. As drawn, A, B, and C each have area 1/3 and thus probability 1/3. The three way intersection ABC has zero probability, implying A, B, and C are not mutually independent since

$$P[ABC] = 0 \neq P[A] P[B] P[C]. \tag{1}$$

However, AB, BC, and AC each has area 1/9. As a result, each pair of events is independent since

$$P[AB] = P[A] P[B], \quad P[BC] = P[B] P[C], \quad P[AC] = P[A] P[C]. \tag{2}$$

Problem 1.7.1 Solution

We can generate the 200×1 vector $\mathbf{T}$, denoted `T` in MATLAB, via the command

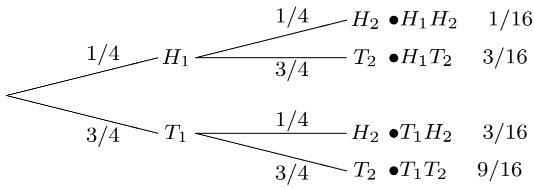
```
T=50+ceil(50*rand(200,1))
```

Keep in mind that `50*rand(200,1)` produces a 200×1 vector of random numbers, each in the interval $(0, 50)$. Applying the ceiling function converts these random numbers to rndom integers in the set $\{1, 2, \ldots, 50\}$. Finally, we add 50 to produce random numbers between 51 and 100.

Problem Solutions – Chapter 2

Problem 2.1.1 Solution

A sequential sample space for this experiment is



(a) From the tree, we observe

$$P[H_2] = P[H_1H_2] + P[T_1H_2] = 1/4. \tag{1}$$

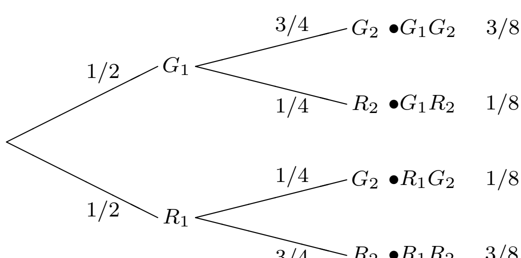
This implies

$$P[H_1|H_2] = \frac{P[H_1H_2]}{P[H_2]} = \frac{1/16}{1/4} = 1/4. \tag{2}$$

(b) The probability that the first flip is heads and the second flip is tails is $P[H_1T_2] = 3/16$.

Problem 2.1.2 Solution

The tree with adjusted probabilities is



From the tree, the probability the second light is green is

$$P[G_2] = P[G_1G_2] + P[R_1G_2] = 3/8 + 1/8 = 1/2. \tag{1}$$

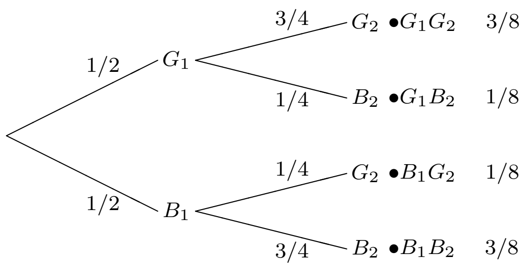
The conditional probability that the first light was green given the second light was green is

$$P[G_1|G_2] = \frac{P[G_1G_2]}{P[G_2]} = \frac{P[G_2|G_1]\,P[G_1]}{P[G_2]} = 3/4. \tag{2}$$

Finally, from the tree diagram, we can directly read that $P[G_2|G_1] = 3/4$.

Problem 2.1.3 Solution

Let G_i and B_i denote events indicating whether free throw i was good (G_i) or bad (B_i). The tree for the free throw experiment is

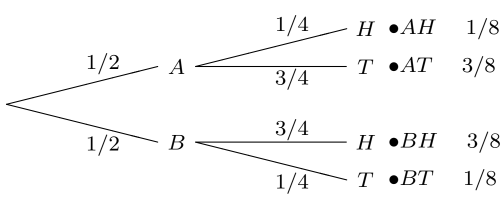


The game goes into overtime if exactly one free throw is made. This event has probability

$$P[O] = P[G_1B_2] + P[B_1G_2] = 1/8 + 1/8 = 1/4. \tag{1}$$

Problem 2.1.4 Solution

The tree for this experiment is



The probability that you guess correctly is

$$\mathrm{P}[C] = \mathrm{P}[AT] + \mathrm{P}[BH] = 3/8 + 3/8 = 3/4. \tag{1}$$

Problem 2.1.5 Solution

The $\mathrm{P}[-|H]$ is the probability that a person who has HIV tests negative for the disease. This is referred to as a false-negative result. The case where a person who does not have HIV but tests positive for the disease, is called a false-positive result and has probability $\mathrm{P}[+|H^c]$. Since the test is correct 99% of the time,

$$\mathrm{P}[-|H] = \mathrm{P}[+|H^c] = 0.01. \tag{1}$$

Now the probability that a person who has tested positive for HIV actually has the disease is

$$\mathrm{P}[H|+] = \frac{\mathrm{P}[+,H]}{\mathrm{P}[+]} = \frac{\mathrm{P}[+,H]}{\mathrm{P}[+,H] + \mathrm{P}[+,H^c]}. \tag{2}$$

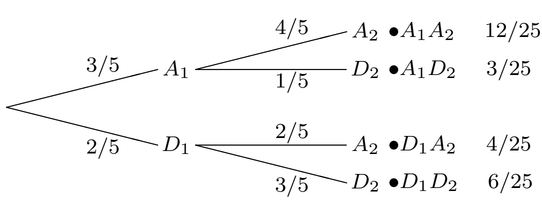
We can use Bayes' formula to evaluate these joint probabilities.

$$\begin{aligned}
\mathrm{P}[H|+] &= \frac{\mathrm{P}[+|H]\,\mathrm{P}[H]}{\mathrm{P}[+|H]\,\mathrm{P}[H] + \mathrm{P}[+|H^c]\,\mathrm{P}[H^c]} \\
&= \frac{(0.99)(0.0002)}{(0.99)(0.0002) + (0.01)(0.9998)} \\
&= 0.0194.
\end{aligned} \tag{3}$$

Thus, even though the test is correct 99% of the time, the probability that a random person who tests positive actually has HIV is less than 0.02. The reason this probability is so low is that the a priori probability that a person has HIV is very small.

Problem 2.1.6 Solution

Let A_i and D_i indicate whether the ith photodetector is acceptable or defective.



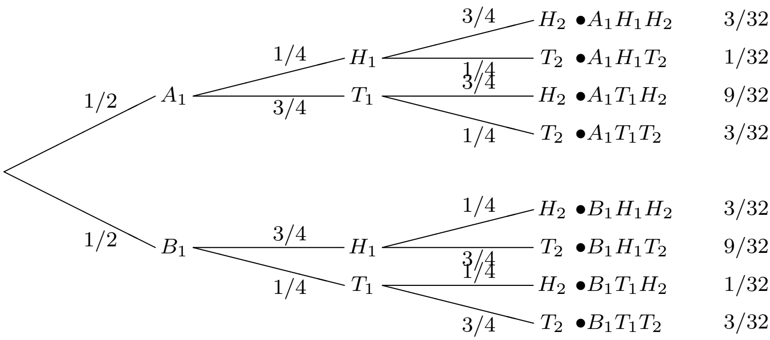
(a) We wish to find the probability $\mathrm{P}[E_1]$ that exactly one photodetector is acceptable. From the tree, we have

$$\mathrm{P}\left[E_1\right] = \mathrm{P}\left[A_1 D_2\right] + \mathrm{P}\left[D_1 A_2\right] = 3/25 + 4/25 = 7/25. \tag{1}$$

(b) The probability that both photodetectors are defective is $\mathrm{P}[D_1 D_2] = 6/25$.

Problem 2.1.7 Solution

The tree for this experiment is



The event $H_1 H_2$ that heads occurs on both flips has probability

$$\mathrm{P}\left[H_1 H_2\right] = \mathrm{P}\left[A_1 H_1 H_2\right] + \mathrm{P}\left[B_1 H_1 H_2\right] = 6/32. \tag{1}$$

The probability of H_1 is

$$\begin{aligned} \mathrm{P}\left[H_1\right] &= \mathrm{P}\left[A_1 H_1 H_2\right] + \mathrm{P}\left[A_1 H_1 T_2\right] + \mathrm{P}\left[B_1 H_1 H_2\right] + \mathrm{P}\left[B_1 H_1 T_2\right] \\ &= 1/2. \end{aligned} \tag{2}$$

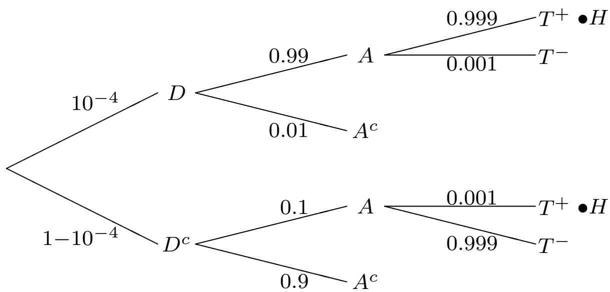
Similarly,

$$\begin{aligned} \mathrm{P}\left[H_2\right] &= \mathrm{P}\left[A_1 H_1 H_2\right] + \mathrm{P}\left[A_1 T_1 H_2\right] + \mathrm{P}\left[B_1 H_1 H_2\right] + \mathrm{P}\left[B_1 T_1 H_2\right] \\ &= 1/2. \end{aligned} \tag{3}$$

Thus $\mathrm{P}[H_1 H_2] \neq \mathrm{P}[H_1]\,\mathrm{P}[H_2]$, implying H_1 and H_2 are not independent. This result should not be surprising since if the first flip is heads, it is likely that coin B was picked first. In this case, the second flip is less likely to be heads since it becomes more likely that the second coin flipped was coin A.

Problem 2.1.8 Solution

We start with a tree diagram:



(a) Here we are asked to calculate the conditional probability $P[D|A]$. In this part, its simpler to ignore the last branches of the tree that indicate the lab test result. This yields

$$
\begin{aligned}
P[D|A] = \frac{P[DA]}{P[A]} &= \frac{P[AD]}{P[DA] + P[D^cA]} \\
&= \frac{(10^{-4})(0.99)}{(10^{-4})(0.99) + (0.1)(1 - 10^{-4})} \\
&= 9.89 \times 10^{-4}. \qquad (1)
\end{aligned}
$$

The probability of the defect D given the arrhythmia A is still quite low because the probability of the defect is so small.

(b) Since the heart surgery occurs if and only if the event T^+ occurs, H and T^+ are the same event and (from the previous part)

$$
\begin{aligned}
P[H|D] = P\left[T^+|D\right] &= \frac{P[DT^+]}{P[D]} \\
&= \frac{10^{-4}(0.99)(0.999)}{10^{-4}} = (0.99)(0.999). \qquad (2)
\end{aligned}
$$

(c) Since the heart surgery occurs if and only if the event T^+ occurs, H and T^+ are the same event and (from the previous part)

$$P[H|D^c] = P\left[T^+|D^c\right] = \frac{P[D^c T^+]}{P[D^c]}$$
$$= \frac{(1-10^{-4})(0.1)(0.001)}{1-10^{-4}} = 10^{-4}. \tag{3}$$

(d) Heart surgery occurs with probability

$$P[H] = P[H|D]\,P[D] + P[H|D^c]\,P[D^c]$$
$$= (0.99)(0.999)(10^{-4}) + (10^{-4})(1-10^{-4})$$
$$= 1.99 \times 10^{-4}. \tag{4}$$

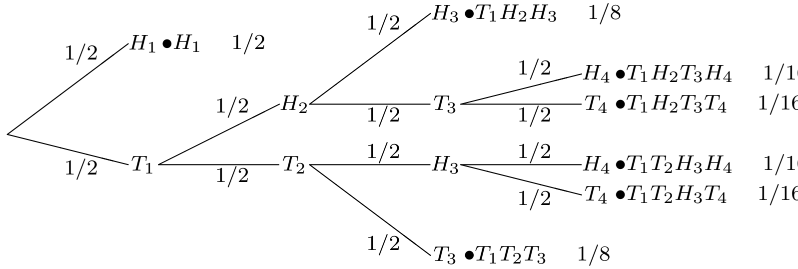
(e) Given that heart surgery was performed, the probability the child had no defect is

$$P[D^c|H] = \frac{P[D^c H]}{P[H]}$$
$$= \frac{(1-10^{-4})(0.1)(0.001)}{(0.99)(0.999)(10^{-4}) + (10^{-4})(1-10^{-4})}$$
$$= \frac{1-10^{-4}}{2-10^{-2}-10^{-3}+10^{-4}} = 0.5027. \tag{5}$$

Because the arrythmia is fairly common and the lab test is not fully reliable, roughly half of all the heart surgeries are performed on healthy infants.

Problem 2.1.9 Solution

(a) The primary difficulty in this problem is translating the words into the correct tree diagram. The tree for this problem is shown below.



(b) From the tree,

$$\begin{aligned} \mathrm{P}[H_3] &= \mathrm{P}[T_1H_2H_3] + \mathrm{P}[T_1T_2H_3H_4] + \mathrm{P}[T_1T_2H_3H_4] \\ &= 1/8 + 1/16 + 1/16 = 1/4. \end{aligned} \tag{1}$$

Similarly,

$$\begin{aligned} \mathrm{P}[T_3] &= \mathrm{P}[T_1H_2T_3H_4] + \mathrm{P}[T_1H_2T_3T_4] + \mathrm{P}[T_1T_2T_3] \\ &= 1/8 + 1/16 + 1/16 = 1/4. \end{aligned} \tag{2}$$

(c) The event that Dagwood must diet is

$$D = (T_1H_2T_3T_4) \cup (T_1T_2H_3T_4) \cup (T_1T_2T_3). \tag{3}$$

The probability that Dagwood must diet is

$$\begin{aligned} \mathrm{P}[D] &= \mathrm{P}[T_1H_2T_3T_4] + \mathrm{P}[T_1T_2H_3T_4] + \mathrm{P}[T_1T_2T_3] \\ &= 1/16 + 1/16 + 1/8 = 1/4. \end{aligned} \tag{4}$$

The conditional probability of heads on flip 1 given that Dagwood must diet is

$$\mathrm{P}[H_1|D] = \frac{\mathrm{P}[H_1D]}{\mathrm{P}[D]} = 0. \tag{5}$$

Remember, if there was heads on flip 1, then Dagwood always postpones his diet.

(d) From part (b), we found that $P[H_3] = 1/4$. To check independence, we calculate

$$\begin{aligned} P[H_2] &= P[T_1H_2H_3] + P[T_1H_2T_3] + P[T_1H_2T_4T_4] = 1/4 \\ P[H_2H_3] &= P[T_1H_2H_3] = 1/8. \end{aligned} \tag{6}$$

Now we find that

$$P[H_2H_3] = 1/8 \neq P[H_2]\,P[H_3]\,. \tag{7}$$

Hence, H_2 and H_3 are dependent events. In fact, $P[H_3|H_2] = 1/2$ while $P[H_3] = 1/4$. The reason for the dependence is that given H_2 occurred, then we know there will be a third flip which may result in H_3. That is, knowledge of H_2 tells us that the experiment didn't end after the first flip.

Problem 2.1.10 Solution

(a) We wish to know what the probability that we find no good photodiodes in n pairs of diodes. Testing each pair of diodes is an independent trial such that with probability p, both diodes of a pair are bad. From Problem 2.1.6, we can easily calculate p.

$$p = P[\text{both diodes are defective}] = P[D_1D_2] = 6/25. \tag{1}$$

The probability of Z_n, the probability of zero acceptable diodes out of n pairs of diodes is p^n because on each test of a pair of diodes, both must be defective.

$$P[Z_n] = \prod_{i=1}^{n} p = p^n = \left(\frac{6}{25}\right)^n \tag{2}$$

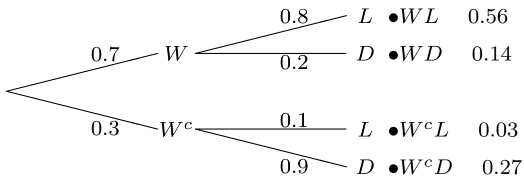
(b) Another way to phrase this question is to ask how many pairs must we test until $P[Z_n] \leq 0.01$. Since $P[Z_n] = (6/25)^n$, we require

$$\left(\frac{6}{25}\right)^n \leq 0.01 \quad \Rightarrow \quad n \geq \frac{\ln 0.01}{\ln 6/25} = 3.23. \tag{3}$$

Since n must be an integer, $n = 4$ pairs must be tested.

Problem 2.1.11 Solution

The starting point is to draw a tree of the experiment. We define the events W that the plant is watered, L that the plant lives, and D that the plant dies. The tree diagram is



It follows that

(a) $P[L] = P[WL] + P[W^cL] = 0.56 + 0.03 = 0.59$.

(b)

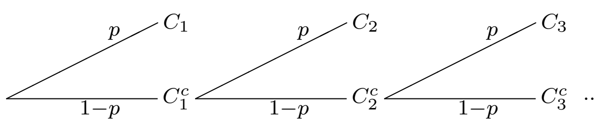
$$P[W^c|D] = \frac{P[W^cD]}{P[D]} = \frac{0.27}{0.14 + 0.27} = \frac{27}{41}. \tag{1}$$

(c) $P[D|W^c] = 0.9$.

In informal conversation, it can be confusing to distinguish between $P[D|W^c]$ and $P[W^c|D]$; however, they are simple once you draw the tree.

Problem 2.1.12 Solution

The experiment ends as soon as a fish is caught. The tree resembles



From the tree, $P[C_1] = p$ and $P[C_2] = (1-p)p$. Finally, a fish is caught on the nth cast if no fish were caught on the previous $n-1$ casts. Thus,

$$P[C_n] = (1-p)^{n-1}p. \tag{1}$$

Problem 2.2.1 Solution

Technically, a gumball machine has a finite number of gumballs, but the problem description models the drawing of gumballs as sampling from the machine without replacement. This is a reasonable model when the machine has a very large gumball capacity and we have no knowledge beforehand of how many gumballs of each color are in the machine. Under this model, the requested probability is given by the multinomial probability

$$\begin{aligned} \mathrm{P}\left[R_2 Y_2 G_2 B_2\right] &= \frac{8!}{2!\,2!\,2!\,2!}\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^2 \\ &= \frac{8!}{4^{10}} \approx 0.0385. \end{aligned} \tag{1}$$

Problem 2.2.2 Solution

In this model of a starburst package, the pieces in a package are collected by sampling without replacement from a giant collection of starburst pieces.

(a) Each piece is "berry or cherry" with probability $p = 1/2$. The probability of only berry or cherry pieces is $p^{12} = 1/4096$.

(b) Each piece is "not cherry" with probability $3/4$. The probability all 12 pieces are "not pink" is $(3/4)^{12} = 0.0317$.

(c) For $i = 1, 2, \ldots, 6$, let C_i denote the event that all 12 pieces are flavor i. Since each piece is flavor i with probability $1/4$, $\mathrm{P}[C_i] = (1/4)^{12}$. Since C_i and C_j are mutually exclusive,

$$\mathrm{P}[F_1] = \mathrm{P}[C_1 \cup C_2 \cup \cdots \cup C_4] = \sum_{i=1}^{4} \mathrm{P}[C_i] = 4\,\mathrm{P}[C_1] = (1/4)^{11}.$$

Problem 2.2.3 Solution

(a) Let B_i, L_i, O_i and C_i denote the events that the ith piece is Berry, Lemon, Orange, and Cherry respectively. Let F_1 denote the event that all three pieces draw are the same flavor. Thus,

$$F_1 = \{S_1S_2S_3, L_1L_2L_3, O_1O_2O_3, C_1C_2C_3\} \tag{1}$$

$$\mathrm{P}\left[F_1\right] = \mathrm{P}\left[S_1S_2S_3\right] + \mathrm{P}\left[L_1L_2L_3\right] + \mathrm{P}\left[O_1O_2O_3\right] + \mathrm{P}\left[C_1C_2C_3\right] \tag{2}$$

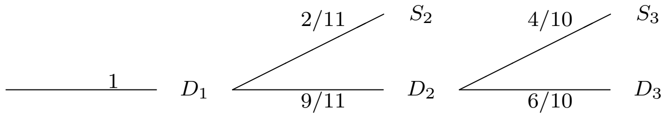
Note that

$$\mathrm{P}[L_1L_2L_3] = \frac{3}{12} \cdot \frac{2}{11} \cdot \frac{1}{10} = \frac{1}{220} \tag{3}$$

and by symmetry,

$$\mathrm{P}[F_1] = 4\,\mathrm{P}[L_1L_2L_3] = \frac{1}{55}. \tag{4}$$

(b) Let D_i denote the event that the ith piece is a different flavor from all the prior pieces. Let S_i denote the event that piece i is the same flavor as a previous piece. A tree for this experiment is



Note that:

- $\mathrm{P}[D_1] = 1$ because the first piece is "different" since there haven't been any prior pieces.
- The second piece is the same as the first piece with probability $2/11$ because in the remaining 11 pieces there are 2 pieces that are the same as the first piece. Alternatively, out of 11 pieces left, there are 3 colors each with 3 pieces (that is, 9 pieces out of 11) that are different from the first piece.
- Given the first two pieces are different, there are 2 colors, each with 3 pieces (6 pieces) out of 10 remaining pieces that are a different flavor from the first two pieces. Thus $\mathrm{P}[D_3|D_2D_1] = 6/10$.

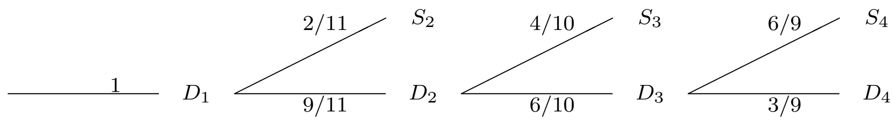
It follows that the three pieces are different with probability

$$\mathrm{P}\left[D_1D_2D_3\right] = 1\left(\frac{9}{11}\right)\left(\frac{6}{10}\right) = \frac{27}{55}. \tag{5}$$

Problem 2.2.4 Solution

(a) Since there are only three pieces of each flavor, we cannot draw four pieces of all the same flavor. Hence $P[F_1] = 0$.

(b) Let D_i denote the event that the ith piece is a different flavor from all the prior pieces. Let S_i denote the event that piece i is the same flavor as a previous piece. A tree for this experiment is relatively simple because we stop the experiment as soon as we draw a piece that is the same as a previous piece. The tree is



Note that:

- $P[D_1] = 1$ because the first piece is "different" since there haven't been any prior pieces.
- For the second piece, there are 11 pieces left and 9 of those pieces are different from the first piece drawn.
- Given the first two pieces are different, there are 2 colors, each with 3 pieces (6 pieces) out of 10 remaining pieces that are a different flavor from the first two pieces. Thus $P[D_3|D_2D_1] = 6/10$.
- Finally, given the first three pieces are different flavors, there are 3 pieces remaining that are a different flavor from the pieces previously picked.

Thus $P[D_4|D_3D_2D_1] = 3/9$. It follows that the three pieces are different with probability

$$P[D_1D_2D_3D_4] = 1\left(\frac{9}{11}\right)\left(\frac{6}{10}\right)\frac{3}{9} = \frac{9}{55}. \quad (1)$$

An alternate approach to this problem is to assume that each piece is distinguishable, say by numbering the pieces 1, 2, 3 in each flavor. In addition, we define the outcome of the experiment to be a 4-permutation of the 12 distinguishable pieces. Under this model, there are $(12)_4 = \frac{12!}{8!}$ equally likely outcomes in the sample space. The number of outcomes in which all four pieces are different is $n_4 = 12 \cdot 9 \cdot 6 \cdot 3$ since there are 12 choices for the first piece drawn, 9 choices for the second piece from the three remaining flavors, 6 choices for the third piece and three choices for the last piece. Since all outcomes are equally likely,

$$P[F_4] = \frac{n_4}{(12)_4} = \frac{12 \cdot 9 \cdot 6 \cdot 3}{12 \cdot 11 \cdot 10 \cdot 9} = \frac{9}{55} \tag{2}$$

(c) The second model of distinguishable starburst pieces makes it easier to solve this last question. In this case, let the outcome of the experiment be the $\binom{12}{4} = 495$ combinations or pieces. In this case, we are ignoring the order in which the pieces were selected. Now we count the number of combinations in which we have two pieces of each of two flavors. We can do this with the following steps:

1. Choose two of the four flavors.
2. Choose 2 out of 3 pieces of one of the two chosen flavors.
3. Choose 2 out of 3 pieces of the other of the two chosen flavors.

Let n_i equal the number of ways to execute step i. We see that

$$n_1 = \binom{4}{2} = 6, \qquad n_2 = \binom{3}{2} = 3, \qquad n_3 = \binom{3}{2} = 3. \tag{3}$$

There are $n_1 n_2 n_3 = 54$ possible ways to execute this sequence of steps. Since all combinations are equally likely,

$$P[F_2] = \frac{n_1 n_2 n_3}{\binom{12}{4}} = \frac{54}{495} = \frac{6}{55}. \tag{4}$$

Problem 2.2.5 Solution

Since there are $H = \binom{52}{7}$ equiprobable seven-card hands, each probability is just the number of hands of each type divided by H.

(a) Since there are 26 red cards, there are $\binom{26}{7}$ seven-card hands with all red cards. The probability of a seven-card hand of all red cards is

$$P[R_7] = \frac{\binom{26}{7}}{\binom{52}{7}} = \frac{26!\,45!}{52!\,19!} = 0.0049. \tag{1}$$

(b) There are 12 face cards in a 52 card deck and there are $\binom{12}{7}$ seven card hands with all face cards. The probability of drawing only face cards is

$$P[F] = \frac{\binom{12}{7}}{\binom{52}{7}} = \frac{12!\,45!}{5!\,52!} = 5.92 \times 10^{-6}. \tag{2}$$

(c) There are 6 red face cards (J, Q, K of diamonds and hearts) in a 52 card deck. Thus it is impossible to get a seven-card hand of red face cards: $P[R_7F] = 0$.

Problem 2.2.6 Solution

There are $H_5 = \binom{52}{5}$ equally likely five-card hands. Dividing the number of hands of a particular type by H will yield the probability of a hand of that type.

(a) There are $\binom{26}{5}$ five-card hands of all red cards. Thus the probability getting a five-card hand of all red cards is

$$P[R_5] = \frac{\binom{26}{5}}{\binom{52}{5}} = \frac{26!\,47!}{21!\,52!} = 0.0253. \tag{1}$$

Note that this can be rewritten as

$$P[R_5] = \frac{26}{52}\frac{25}{51}\frac{24}{50}\frac{23}{49}\frac{22}{48},$$

which shows the successive probabilities of receiving a red card.

(b) The following sequence of subexperiments will generate all possible "full house"

1. Choose a kind for three-of-a-kind.
2. Choose a kind for two-of-a-kind.

3. Choose three of the four cards of the three-of-a-kind kind.
4. Choose two of the four cards of the two-of-a-kind kind.

The number of ways of performing subexperiment i is

$$n_1 = \binom{13}{1}, \qquad n_2 = \binom{12}{1}, \qquad n_3 = \binom{4}{3}, \qquad n_4 = \binom{4}{2}. \tag{2}$$

Note that $n_2 = \binom{12}{1}$ because after choosing a three-of-a-kind, there are twelove kinds left from which to choose two-of-a-kind. is

The probability of a full house is

$$P[\text{full house}] = \frac{n_1 n_2 n_3 n_4}{\binom{52}{5}} = \frac{3,744}{2,598,960} = 0.0014. \tag{3}$$

Problem 2.2.7 Solution

There are $2^5 = 32$ different binary codes with 5 bits. The number of codes with exactly 3 zeros equals the number of ways of choosing the bits in which those zeros occur. Therefore there are $\binom{5}{3} = 10$ codes with exactly 3 zeros.

Problem 2.2.8 Solution

Since each letter can take on any one of the 4 possible letters in the alphabet, the number of 3 letter words that can be formed is $4^3 = 64$. If we allow each letter to appear only once then we have 4 choices for the first letter and 3 choices for the second and two choices for the third letter. Therefore, there are a total of $4 \cdot 3 \cdot 2 = 24$ possible codes.

Problem 2.2.9 Solution

We can break down the experiment of choosing a starting lineup into a sequence of subexperiments:

1. Choose 1 of the 10 pitchers. There are $N_1 = \binom{10}{1} = 10$ ways to do this.

2. Choose 1 of the 15 field players to be the designated hitter (DH). There are $N_2 = \binom{15}{1} = 15$ ways to do this.

3. Of the remaining 14 field players, choose 8 for the remaining field positions. There are $N_3 = \binom{14}{8}$ to do this.

4. For the 9 batters (consisting of the 8 field players and the designated hitter), choose a batting lineup. There are $N_4 = 9!$ ways to do this.

So the total number of different starting lineups when the DH is selected among the field players is

$$N = N_1 N_2 N_3 N_4 = (10)(15)\binom{14}{8} 9! = 163{,}459{,}296{,}000. \tag{1}$$

Note that this overestimates the number of combinations the manager must really consider because most field players can play only one or two positions. Although these constraints on the manager reduce the number of possible lineups, it typically makes the manager's job more difficult. As for the counting, we note that our count did not need to specify the positions played by the field players. Although this is an important consideration for the manager, it is not part of our counting of different lineups. In fact, the 8 nonpitching field players are allowed to switch positions at any time in the field. For example, the shortstop and second baseman could trade positions in the middle of an inning. Although the DH can go play the field, there are some coomplicated rules about this. Here is an excerpt from Major League Baseball Rule 6.10:

> The Designated Hitter may be used defensively, continuing to bat in the same position in the batting order, but the pitcher must then bat in the place of the substituted defensive player, unless more than one substitution is made, and the manager then must designate their spots in the batting order.

If you're curious, you can find the complete rule on the web.

Problem 2.2.10 Solution

When the DH can be chosen among all the players, including the pitchers, there are two cases:

- The DH is a field player. In this case, the number of possible lineups, N_F, is given in Problem 2.2.9. In this case, the designated hitter must be chosen from the 15 field players. We repeat the solution of Problem 2.2.9 here: We can break down the experiment of choosing a starting lineup into a sequence of subexperiments:

1. Choose 1 of the 10 pitchers. There are $N_1 = \binom{10}{1} = 10$ ways to do this.
2. Choose 1 of the 15 field players to be the designated hitter (DH). There are $N_2 = \binom{15}{1} = 15$ ways to do this.
3. Of the remaining 14 field players, choose 8 for the remaining field positions. There are $N_3 = \binom{14}{8}$ to do this.
4. For the 9 batters (consisting of the 8 field players and the designated hitter), choose a batting lineup. There are $N_4 = 9!$ ways to do this.

So the total number of different starting lineups when the DH is selected among the field players is

$$N = N_1 N_2 N_3 N_4 = (10)(15)\binom{14}{8} 9! = 163{,}459{,}296{,}000. \tag{1}$$

- The DH is a pitcher. In this case, there are 10 choices for the pitcher, 10 choices for the DH among the pitchers (including the pitcher batting for himself), $\binom{15}{8}$ choices for the field players, and 9! ways of ordering the batters into a lineup. The number of possible lineups is

$$N' = (10)(10)\binom{15}{8} 9! = 233,513,280,000. \tag{2}$$

The total number of ways of choosing a lineup is $N + N' = 396{,}972{,}576{,}000$.

Problem 2.2.11 Solution

(a) This is just the multinomial probability

$$\begin{aligned} \mathrm{P}[A] &= \binom{40}{19, 19, 2} \left(\frac{19}{40}\right)^{19} \left(\frac{19}{40}\right)^{19} \left(\frac{2}{40}\right)^{2} \\ &= \frac{40!}{19!19!2!} \left(\frac{19}{40}\right)^{19} \left(\frac{19}{40}\right)^{19} \left(\frac{2}{40}\right)^{2}. \end{aligned} \tag{1}$$

(b) Each spin is either green (with probability 19/40) or not (with probability 21/40). If we call landing on greeen a success, then G_{19} is the probability of 19 successes in 40 trials. Thus

$$P[G_{19}] = \binom{40}{19}\left(\frac{19}{40}\right)^{19}\left(\frac{21}{40}\right)^{21}. \tag{2}$$

(c) If you bet on red, the probability you win is 19/40. If you bet green, the probability that you win is 19/40. If you first make a random choice to bet red or green, (say by flipping a coin), the probability you win is still $p = 19/40$.

Problem 2.2.12 Solution

(a) We can find the number of valid starting lineups by noticing that the swingman presents three situations: (1) the swingman plays guard, (2) the swingman plays forward, and (3) the swingman doesn't play. The first situation is when the swingman can be chosen to play the guard position, and the second where the swingman can only be chosen to play the forward position. Let N_i denote the number of lineups corresponding to case i. Then we can write the total number of lineups as $N_1 + N_2 + N_3$. In the first situation, we have to choose 1 out of 3 centers, 2 out of 4 forwards, and 1 out of 4 guards so that

$$N_1 = \binom{3}{1}\binom{4}{2}\binom{4}{1} = 72. \tag{1}$$

In the second case, we need to choose 1 out of 3 centers, 1 out of 4 forwards and 2 out of 4 guards, yielding

$$N_2 = \binom{3}{1}\binom{4}{1}\binom{4}{2} = 72. \tag{2}$$

Finally, with the swingman on the bench, we choose 1 out of 3 centers, 2 out of 4 forward, and 2 out of four guards. This implies

$$N_3 = \binom{3}{1}\binom{4}{2}\binom{4}{2} = 108, \tag{3}$$

and the total number of lineups is $N_1 + N_2 + N_3 = 252$.

Problem 2.2.13 Solution

What our design must specify is the number of boxes on the ticket, and the number of specially marked boxes. Suppose each ticket has n boxes and $5+k$ specially marked boxes. Note that when $k>0$, a winning ticket will still have k unscratched boxes with the special mark. A ticket is a winner if each time a box is scratched off, the box has the special mark. Assuming the boxes are scratched off randomly, the first box scratched off has the mark with probability $(5+k)/n$ since there are $5+k$ marked boxes out of n boxes. Moreover, if the first scratched box has the mark, then there are $4+k$ marked boxes out of $n-1$ remaining boxes. Continuing this argument, the probability that a ticket is a winner is

$$p = \frac{5+k}{n}\frac{4+k}{n-1}\frac{3+k}{n-2}\frac{2+k}{n-3}\frac{1+k}{n-4} = \frac{(k+5)!(n-5)!}{k!n!}. \tag{1}$$

By careful choice of n and k, we can choose p close to 0.01. For example,

n	9	11	14	17
k	0	1	2	3
p	0.0079	0.012	0.0105	0.0090

(2)

A gamecard with $N=14$ boxes and $5+k=7$ shaded boxes would be quite reasonable.

Problem 2.3.1 Solution

(a) Since the probability of a zero is 0.8, we can express the probability of the code word 00111 as 2 occurrences of a 0 and three occurrences of a 1. Therefore

$$\mathrm{P}[00111] = (0.8)^2(0.2)^3 = 0.00512. \tag{1}$$

(b) The probability that a code word has exactly three 1's is

$$\mathrm{P}[\text{three 1's}] = \binom{5}{3}(0.8)^2(0.2)^3 = 0.0512. \tag{2}$$

Problem 2.3.2 Solution

Given that the probability that the Celtics win a single championship in any given year is 0.32, we can find the probability that they win 8 straight NBA championships.

$$P[8 \text{ straight championships}] = (0.32)^8 = 0.00011. \tag{1}$$

The probability that they win 10 titles in 11 years is

$$P[10 \text{ titles in } 11 \text{ years}] = \binom{11}{10}(.32)^{10}(.68) = 0.00084. \tag{2}$$

The probability of each of these events is less than 1 in 1000! Given that these events took place in the relatively short fifty year history of the NBA, it should seem that these probabilities should be much higher. What the model overlooks is that the sequence of 10 titles in 11 years started when Bill Russell joined the Celtics. In the years with Russell (and a strong supporting cast) the probability of a championship was much higher.

Problem 2.3.3 Solution

We know that the probability of a green and red light is 7/16, and that of a yellow light is 1/8. Since there are always 5 lights, G, Y, and R obey the multinomial probability law:

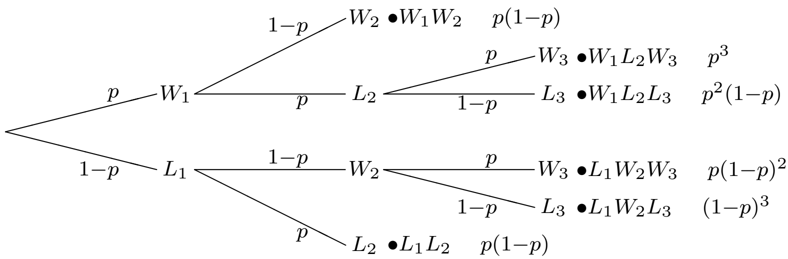
$$P[G = 2, Y = 1, R = 2] = \frac{5!}{2!1!2!}\left(\frac{7}{16}\right)^2\left(\frac{1}{8}\right)\left(\frac{7}{16}\right)^2. \tag{1}$$

The probability that the number of green lights equals the number of red lights

$$\begin{aligned} P[G = R] &= P[G = 1, R = 1, Y = 3] + P[G = 2, R = 2, Y = 1] \\ &\quad + P[G = 0, R = 0, Y = 5] \\ &= \frac{5!}{1!1!3!}\left(\frac{7}{16}\right)\left(\frac{7}{16}\right)\left(\frac{1}{8}\right)^3 + \frac{5!}{2!1!2!}\left(\frac{7}{16}\right)^2\left(\frac{7}{16}\right)^2\left(\frac{1}{8}\right) \\ &\quad + \frac{5!}{0!0!5!}\left(\frac{1}{8}\right)^5 \\ &\approx 0.1449. \end{aligned} \tag{2}$$

Problem 2.3.4 Solution

For the team with the homecourt advantage, let W_i and L_i denote whether game i was a win or a loss. Because games 1 and 3 are home games and game 2 is an away game, the tree is



The probability that the team with the home court advantage wins is

$$\begin{aligned} P[H] &= P[W_1W_2] + P[W_1L_2W_3] + P[L_1W_2W_3] \\ &= p(1-p) + p^3 + p(1-p)^2. \end{aligned} \tag{1}$$

Note that $P[H] \leq p$ for $1/2 \leq p \leq 1$. Since the team with the home court advantage would win a 1 game playoff with probability p, the home court team is less likely to win a three game series than a 1 game playoff!

Problem 2.3.5 Solution

(a) There are 3 group 1 kickers and 6 group 2 kickers. Using G_i to denote that a group i kicker was chosen, we have

$$P[G_1] = 1/3, \qquad P[G_2] = 2/3. \tag{1}$$

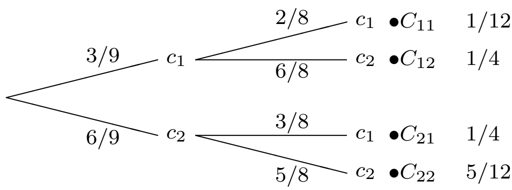
In addition, the problem statement tells us that

$$P[K|G_1] = 1/2, \qquad P[K|G_2] = 1/3. \tag{2}$$

Combining these facts using the Law of Total Probability yields

$$\begin{aligned} P[K] &= P[K|G_1]\,P[G_1] + P[K|G_2]\,P[G_2] \\ &= (1/2)(1/3) + (1/3)(2/3) = 7/18. \end{aligned} \tag{3}$$

(b) To solve this part, we need to identify the groups from which the first and second kicker were chosen. Let c_i indicate whether a kicker was chosen from group i and let C_{ij} indicate that the first kicker was chosen from group i and the second kicker from group j. The experiment to choose the kickers is described by the sample tree:



Since a kicker from group 1 makes a kick with probability 1/2 while a kicker from group 2 makes a kick with probability 1/3,

$$\mathrm{P}[K_1K_2|C_{11}] = (1/2)^2, \qquad \mathrm{P}[K_1K_2|C_{12}] = (1/2)(1/3), \tag{4}$$

$$\mathrm{P}[K_1K_2|C_{21}] = (1/3)(1/2), \qquad \mathrm{P}[K_1K_2|C_{22}] = (1/3)^2. \tag{5}$$

By the law of total probability,

$$\begin{aligned}\mathrm{P}[K_1K_2] &= \mathrm{P}[K_1K_2|C_{11}]\,\mathrm{P}[C_{11}] + \mathrm{P}[K_1K_2|C_{12}]\,\mathrm{P}[C_{12}] \\ &\quad + \mathrm{P}[K_1K_2|C_{21}]\,\mathrm{P}[C_{21}] + \mathrm{P}[K_1K_2|C_{22}]\,\mathrm{P}[C_{22}] \\ &= \frac{1}{4}\frac{1}{12} + \frac{1}{6}\frac{1}{4} + \frac{1}{6}\frac{1}{4} + \frac{1}{9}\frac{5}{12} = 15/96. \end{aligned} \tag{6}$$

It should be apparent that $\mathrm{P}[K_1] = \mathrm{P}[K]$ from part (a). Symmetry should also make it clear that $\mathrm{P}[K_1] = \mathrm{P}[K_2]$ since for any ordering of two kickers, the reverse ordering is equally likely. If this is not clear, we derive this result by calculating $\mathrm{P}[K_2|C_{ij}]$ and using the law of total probability to calculate $\mathrm{P}[K_2]$.

$$\mathrm{P}[K_2|C_{11}] = 1/2, \qquad \mathrm{P}[K_2|C_{12}] = 1/3, \tag{7}$$

$$\mathrm{P}[K_2|C_{21}] = 1/2, \qquad \mathrm{P}[K_2|C_{22}] = 1/3. \tag{8}$$

By the law of total probability,

$$\begin{aligned}
\mathrm{P}[K_2] &= \mathrm{P}[K_2|C_{11}]\,\mathrm{P}[C_{11}] + \mathrm{P}[K_2|C_{12}]\,\mathrm{P}[C_{12}] \\
&\qquad + \mathrm{P}[K_2|C_{21}]\,\mathrm{P}[C_{21}] + \mathrm{P}[K_2|C_{22}]\,\mathrm{P}[C_{22}] \\
&= \frac{1}{2}\frac{1}{12} + \frac{1}{3}\frac{1}{4} + \frac{1}{2}\frac{1}{4} + \frac{1}{3}\frac{5}{12} = \frac{7}{18}.
\end{aligned} \tag{9}$$

We observe that K_1 and K_2 are not independent since

$$\mathrm{P}[K_1K_2] = \frac{15}{96} \neq \left(\frac{7}{18}\right)^2 = \mathrm{P}[K_1]\,\mathrm{P}[K_2]\,. \tag{10}$$

Note that $15/96$ and $(7/18)^2$ are close but not exactly the same. The reason K_1 and K_2 are dependent is that if the first kicker is successful, then it is more likely that kicker is from group 1. This makes it more likely that the second kicker is from group 2 and is thus more likely to miss.

(c) Once a kicker is chosen, each of the 10 field goals is an independent trial. If the kicker is from group 1, then the success probability is $1/2$. If the kicker is from group 2, the success probability is $1/3$. Out of 10 kicks, there are 5 misses iff there are 5 successful kicks. Given the type of kicker chosen, the probability of 5 misses is

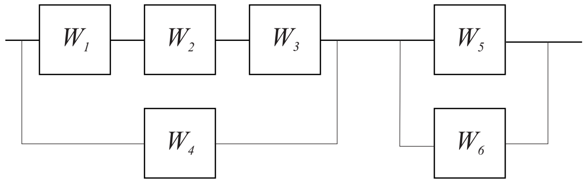
$$\mathrm{P}[M|G_1] = \binom{10}{5}(1/2)^5(1/2)^5, \qquad \mathrm{P}[M|G_2] = \binom{10}{5}(1/3)^5(2/3)^5. \tag{11}$$

We use the Law of Total Probability to find

$$\begin{aligned}
\mathrm{P}[M] &= \mathrm{P}[M|G_1]\,\mathrm{P}[G_1] + \mathrm{P}[M|G_2]\,\mathrm{P}[G_2] \\
&= \binom{10}{5}\left((1/3)(1/2)^{10} + (2/3)(1/3)^5(2/3)^5\right).
\end{aligned} \tag{12}$$

Problem 2.4.1 Solution

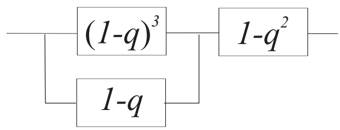
From the problem statement, we can conclude that the device components are configured in the following way.



To find the probability that the device works, we replace series devices 1, 2, and 3, and parallel devices 5 and 6 each with a single device labeled with the probability that it works. In particular,

$$\mathrm{P}[W_1W_2W_3] = (1-q)^3, \tag{1}$$
$$\mathrm{P}[W_5 \cup W_6] = 1 - \mathrm{P}[W_5^c W_6^c] = 1 - q^2. \tag{2}$$

This yields a composite device of the form



The probability $\mathrm{P}[W']$ that the two devices in parallel work is 1 minus the probability that neither works:

$$\mathrm{P}\left[W'\right] = 1 - q(1-(1-q)^3). \tag{3}$$

Finally, for the device to work, both composite device in series must work. Thus, the probability the device works is

$$\mathrm{P}[W] = [1 - q(1-(1-q)^3)][1-q^2]. \tag{4}$$

Problem 2.4.2 Solution

Suppose that the transmitted bit was a 1. We can view each repeated transmission as an independent trial. We call each repeated bit the receiver decodes as 1 a success. Using $S_{k,5}$ to denote the event of k successes in the five trials, then the probability k 1's are decoded at the receiver is

$$\mathrm{P}[S_{k,5}] = \binom{5}{k} p^k (1-p)^{5-k}, \quad k = 0, 1, \ldots, 5. \tag{1}$$

The probability a bit is decoded correctly is

$$\mathrm{P}[C] = \mathrm{P}[S_{5,5}] + \mathrm{P}[S_{4,5}] = p^5 + 5p^4(1-p) = 0.91854. \tag{2}$$

The probability a deletion occurs is

$$\mathrm{P}[D] = \mathrm{P}[S_{3,5}] + \mathrm{P}[S_{2,5}] = 10p^3(1-p)^2 + 10p^2(1-p)^3 = 0.081. \tag{3}$$

The probability of an error is

$$\mathrm{P}[E] = \mathrm{P}[S_{1,5}] + \mathrm{P}[S_{0,5}] = 5p(1-p)^4 + (1-p)^5 = 0.00046. \tag{4}$$

Note that if a 0 is transmitted, then 0 is sent five times and we call decoding a 0 a success. You should convince yourself that this a symmetric situation with the same deletion and error probabilities. Introducing deletions reduces the probability of an error by roughly a factor of 20. However, the probability of successfull decoding is also reduced.

Problem 2.4.3 Solution

Note that each digit 0 through 9 is mapped to the 4 bit binary representation of the digit. That is, 0 corresponds to 0000, 1 to 0001, up to 9 which corresponds to 1001. Of course, the 4 bit binary numbers corresponding to numbers 10 through 15 go unused, however this is unimportant to our problem. the 10 digit number results in the transmission of 40 bits. For each bit, an independent trial determines whether the bit was correct, a deletion, or an error. In Problem 2.4.2, we found the probabilities of these events to be

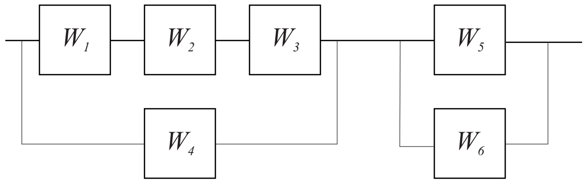
$$\mathrm{P}[C] = \gamma = 0.91854, \qquad \mathrm{P}[D] = \delta = 0.081, \qquad \mathrm{P}[E] = \epsilon = 0.00046. \tag{1}$$

Since each of the 40 bit transmissions is an independent trial, the joint probability of c correct bits, d deletions, and e erasures has the multinomial probability

$$\mathrm{P}[C = c, D = d, E = d] = \begin{cases} \frac{40!}{c!d!e!}\gamma^c\delta^d\epsilon^e & c + d + e = 40; c, d, e \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 2.4.4 Solution

From the statement of Problem 2.4.1, the configuration of device components is



By symmetry, note that the reliability of the system is the same whether we replace component 1, component 2, or component 3. Similarly, the reliability is the same whether we replace component 5 or component 6. Thus we consider the following cases:

I **Replace component 1** In this case

$$\mathrm{P}[W_1W_2W_3] = (1-\frac{q}{2})(1-q)^2, \tag{1}$$

$$\mathrm{P}[W_4] = 1-q, \tag{2}$$

$$\mathrm{P}[W_5 \cup W_6] = 1-q^2. \tag{3}$$

This implies

$$\begin{aligned}\mathrm{P}[W_1W_2W_3 \cup W_4] &= 1-(1-\mathrm{P}[W_1W_2W_3])(1-\mathrm{P}[W_4]) \\ &= 1-\frac{q^2}{2}(5-4q+q^2). \end{aligned}\tag{4}$$

In this case, the probability the system works is

$$\begin{aligned}\mathrm{P}[W_I] &= \mathrm{P}[W_1W_2W_3 \cup W_4]\,\mathrm{P}[W_5 \cup W_6] \\ &= \left[1-\frac{q^2}{2}(5-4q+q^2)\right](1-q^2). \end{aligned}\tag{5}$$

II **Replace component 4** In this case,

$$\mathrm{P}[W_1W_2W_3] = (1-q)^3, \tag{6}$$

$$\mathrm{P}[W_4] = 1-\frac{q}{2}, \tag{7}$$

$$\mathrm{P}[W_5 \cup W_6] = 1-q^2. \tag{8}$$

This implies

$$\begin{aligned} \mathrm{P}[W_1W_2W_3 \cup W_4] &= 1-(1-\mathrm{P}[W_1W_2W_3])(1-\mathrm{P}[W_4]) \\ &= 1-\frac{q}{2}+\frac{q}{2}(1-q)^3. \end{aligned} \tag{9}$$

In this case, the probability the system works is

$$\begin{aligned} \mathrm{P}[W_{II}] &= \mathrm{P}[W_1W_2W_3 \cup W_4]\,\mathrm{P}[W_5 \cup W_6] \\ &= \left[1-\frac{q}{2}+\frac{q}{2}(1-q)^3\right](1-q^2). \end{aligned} \tag{10}$$

III **Replace component 5** In this case,

$$\mathrm{P}[W_1W_2W_3] = (1-q)^3, \tag{11}$$

$$\mathrm{P}[W_4] = 1-q, \tag{12}$$

$$\mathrm{P}[W_5 \cup W_6] = 1-\frac{q^2}{2}. \tag{13}$$

This implies

$$\begin{aligned} \mathrm{P}[W_1W_2W_3 \cup W_4] &= 1-(1-\mathrm{P}[W_1W_2W_3])(1-\mathrm{P}[W_4]) \\ &= (1-q)\left[1+q(1-q)^2\right]. \end{aligned} \tag{14}$$

In this case, the probability the system works is

$$\begin{aligned} \mathrm{P}[W_{III}] &= \mathrm{P}[W_1W_2W_3 \cup W_4]\,\mathrm{P}[W_5 \cup W_6] \\ &= (1-q)\left(1-\frac{q^2}{2}\right)\left[1+q(1-q)^2\right]. \end{aligned} \tag{15}$$

From these expressions, its hard to tell which substitution creates the most reliable circuit. First, we observe that $\mathrm{P}[W_{II}] > \mathrm{P}[W_I]$ if and only if

$$1-\frac{q}{2}+\frac{q}{2}(1-q)^3 > 1-\frac{q^2}{2}(5-4q+q^2). \tag{16}$$

Some algebra will show that $\mathrm{P}[W_{II}] > \mathrm{P}[W_I]$ if and only if $q^2 < 2$, which occurs for all nontrivial (i.e., nonzero) values of q. Similar algebra will show that $\mathrm{P}[W_{II}] > \mathrm{P}[W_{III}]$ for all values of $0 \leq q \leq 1$. Thus the best policy is to replace component 4.

Problem 2.5.1 Solution

Rather than just solve the problem for 50 trials, we can write a function that generates vectors **C** and **H** for an arbitrary number of trials n. The code for this task is

```
function [C,H]=twocoin(n);
C=ceil(2*rand(n,1));
P=1-(C/4);
H=(rand(n,1)< P);
```

The first line produces the $n \times 1$ vector `C` such that `C(i)` indicates whether coin 1 or coin 2 is chosen for trial i. Next, we generate the vector `P` such that `P(i)=0.75` if `C(i)=1`; otherwise, if `C(i)=2`, then `P(i)=0.5`. As a result, `H(i)` is the simulated result of a coin flip with heads, corresponding to `H(i)=1`, occurring with probability `P(i)`.

Problem 2.5.2 Solution

Rather than just solve the problem for 100 trials, we can write a function that generates n packets for an arbitrary number of trials n. The code for this task is

```
function C=bit100(n);
% n is the number of 100 bit packets sent
B=floor(2*rand(n,100));
P=0.03-0.02*B;
E=(rand(n,100)< P);
C=sum((sum(E,2)<=5));
```

First, `B` is an $n \times 100$ matrix such that `B(i,j)` indicates whether bit `i` of packet `j` is zero or one. Next, we generate the $n \times 100$ matrix `P` such that `P(i,j)=0.03` if `B(i,j)=0`; otherwise, if `B(i,j)=1`, then `P(i,j)=0.01`. As a result, `E(i,j)` is the simulated error indicator for bit `i` of packet `j`. That is, `E(i,j)=1` if bit `i` of packet `j` is in error; otherwise `E(i,j)=0`. Next we sum across the rows of `E` to obtain the number of errors in each packet. Finally, we count the number of packets with 5 or more errors.

For $n = 100$ packets, the packet success probability is inconclusive. Experimentation will show that `C=97`, `C=98`, `C=99` and `C=100` correct packets are typica values that might be observed. By increasing n, more consistent results are obtained.

For example, repeated trials with $n = 100,000$ packets typically produces around $C = 98,400$ correct packets. Thus 0.984 is a reasonable estimate for the probability of a packet being transmitted correctly.

Problem 2.5.3 Solution

To test n 6-component devices, (such that each component works with probability q) we use the following function:

```
function N=reliable6(n,q);
% n is the number of 6 component devices
%N is the number of working devices
W=rand(n,6)>q;
D=(W(:,1)&W(:,2)&W(:,3))|W(:,4);
D=D&(W(:,5)|W(:,6));
N=sum(D);
```

The $n \times 6$ matrix `W` is a *logical* matrix such that `W(i,j)=1` if component `j` of device `i` works properly. Because `W` is a logical matrix, we can use the MATLAB logical operators `|` and `&` to implement the logic requirements for a working device. By applying these logical operators to the $n \times 1$ columns of `W`, we simulate the test of `n` circuits. Note that `D(i)=1` if device `i` works. Otherwise, `D(i)=0`. Lastly, we count the number `N` of working devices. The following code snippet produces ten sample runs, where each sample run tests `n=100` devices for $q = 0.2$.

```
>> for n=1:10, w(n)=reliable6(100,0.2); end
>> w
w =
    82   87   87   92   91   85   85   83   90   89
>>
```

As we see, the number of working devices is typically around 85 out of 100. Solving Problem 2.4.1, will show that the probability the device works is actually 0.8663.

Problem 2.5.4 Solution

The code

```
function n=countequal(x,y)
%Usage:  n=countequal(x,y)
%n(j)= # elements of x = y(j)
[MX,MY]=ndgrid(x,y);
%each column of MX = x
%each row of MY = y
n=(sum((MX==MY),1))';
```

for `countequal` is quite short (just two lines excluding comments) but needs some explanation. The key is in the operation

`[MX,MY]=ndgrid(x,y).`

The MATLAB built-in function `ndgrid` facilitates plotting a function $g(x, y)$ as a surface over the x, y plane.The x, y plane is represented by a grid of all pairs of points $x(i), y(j)$. When `x` has n elements, and `y` has m elements, `ndgrid(x,y)` creates a grid (an $n \times m$ array) of all possible pairs `[x(i) y(j)]`. This grid is represented by two separate $n \times m$ matrices: `MX` and `MY` which indicate the `x` and `y` values at each grid point. Mathematically,

`MX(i,j) = x(i),    MY(i,j)=y(j).`

Next, `C=(MX==MY)` is an $n \times m$ array such that `C(i,j)=1` if `x(i)=y(j)`; otherwise `C(i,j)=0`. That is, the jth column of `C` indicates indicates which elements of `x` equal `y(j)`. Lastly, we sum along each column j to count number of elements of `x` equal to `y(j)`. That is, we sum along column j to count the number of occurrences (in `x`) of `y(j)`.

Problem 2.5.5 Solution

For arbitrary number of trials n and failure probability q, the following functions evaluates replacing each of the six components by an ultrareliable device.

```
function N=ultrareliable6(n,q);
% n is the number of 6 component devices
%N is the number of working devices
for r=1:6,
    W=rand(n,6)>q;
    R=rand(n,1)>(q/2);
    W(:,r)=R;
    D=(W(:,1)&W(:,2)&W(:,3))|W(:,4);
    D=D&(W(:,5)|W(:,6));
    N(r)=sum(D);
end
```

This code is based on the code for the solution of Problem 2.5.3. The $n \times 6$ matrix `W` is a *logical* matrix such that `W(i,j)=1` if component `j` of device `i` works properly. Because `W` is a logical matrix, we can use the MATLAB logical operators `|` and `&` to implement the logic requirements for a working device. By applying these logical opeators to the $n \times 1$ columns of `W`, we simulate the test of `n` circuits. Note that `D(i)=1` if device `i` works. Otherwise, `D(i)=0`. Note that in the code, we first generate the matrix `W` such that each component has failure probability q. To simulate the replacement of the jth device by the ultrareliable version by replacing the jth column of `W` by the column vector `R` in which a device has failure probability $q/2$. Lastly, for each column replacement, we count the number `N` of working devices. A sample run for $n = 100$ trials and $q = 0.2$ yielded these results:

```
>> ultrareliable6(100,0.2)
ans =
    93    89    91    92    90    93
```

From the above, we see, for example, that replacing the third component with an ultrareliable component resulted in 91 working devices. The results are fairly inconclusive in that replacing devices 1, 2, or 3 should yield the same probability of device failure. If we experiment with $n = 10,000$ runs, the results are more definitive:

```
>> ultrareliable6(10000,0.2)
ans =
        8738        8762        8806        9135        8800        8796
>> >> ultrareliable6(10000,0.2)
ans =
        8771        8795        8806        9178        8886        8875
>>
```

In both cases, it is clear that replacing component 4 maximizes the device reliability. The somewhat complicated solution of Problem 2.4.4 will confirm this observation.

Problem Solutions – Chapter 3

Problem 3.2.1 Solution

(a) We wish to find the value of c that makes the PMF sum up to one.

$$P_N(n) = \begin{cases} c(1/2)^n & n = 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Therefore, $\sum_{n=0}^{2} P_N(n) = c + c/2 + c/4 = 1$, implying $c = 4/7$.

(b) The probability that $N \leq 1$ is

$$\mathrm{P}[N \leq 1] = \mathrm{P}[N = 0] + \mathrm{P}[N = 1] = 4/7 + 2/7 = 6/7. \tag{2}$$

Problem 3.2.2 Solution

(a) We must choose c to make the PMF of V sum to one.

$$\sum_{v=1}^{4} P_V(v) = c(1^2 + 2^2 + 3^2 + 4^2) = 30c = 1. \tag{1}$$

Hence $c = 1/30$.

(b) Let $U = \{u^2 | u = 1, 2, \ldots\}$ so that

$$\mathrm{P}[V \in U] = P_V(1) + P_V(4) = \frac{1}{30} + \frac{4^2}{30} = \frac{17}{30}. \tag{2}$$

(c) The probability that V is even is

$$\mathrm{P}[V \text{ is even}] = P_V(2) + P_V(4) = \frac{2^2}{30} + \frac{4^2}{30} = \frac{2}{3}. \tag{3}$$

(d) The probability that $V > 2$ is

$$\mathrm{P}[V > 2] = P_V(3) + P_V(4) = \frac{3^2}{30} + \frac{4^2}{30} = \frac{5}{6}. \tag{4}$$

Problem 3.2.3 Solution

(a) We choose c so that the PMF sums to one.

$$\sum_x P_X(x) = \frac{c}{2} + \frac{c}{4} + \frac{c}{8} = \frac{7c}{8} = 1. \tag{1}$$

Thus $c = 8/7$.

(b)

$$\mathrm{P}[X = 4] = P_X(4) = \frac{8}{7 \cdot 4} = \frac{2}{7}. \tag{2}$$

(c)

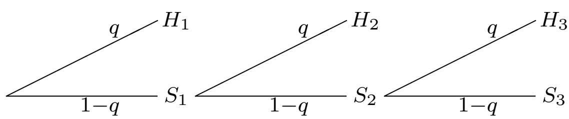
$$\mathrm{P}[X < 4] = P_X(2) = \frac{8}{7 \cdot 2} = \frac{4}{7}. \tag{3}$$

(d)

$$\mathrm{P}[3 \leq X \leq 9] = P_X(4) + P_X(8) = \frac{8}{7 \cdot 4} + \frac{8}{7 \cdot 8} = \frac{3}{7}. \tag{4}$$

Problem 3.2.4 Solution

(a) The starting point is to draw a tree diagram. On swing i, let H_i denote the event that Casey hits a home run and S_i the event that he gets a strike. The tree is



We see that Casey hits a home run as long as he does not strike out. That is,

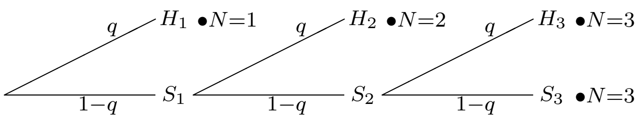
$$\mathrm{P}[H] = 1 - \mathrm{P}[S_3] = 1 - (1-q)^3 = 1 - (0.95)^3. \tag{1}$$

A second way to solve this problem is to observe that Casey hits a home run if he hits a homer on any of his three swings, $H = H_1 \cup H_2 \cup H_3$. Since $\mathrm{P}[H_i] = (1-q)^{i-1}q$ and since the events H_i are mutually exclusive,

$$\mathrm{P}[H] = \mathrm{P}[H_1] + \mathrm{P}[H_2] + \mathrm{P}[H_3] = 0.05(1 + 0.95 + 0.95^2) \tag{2}$$

This can be simplified to the first answer.

(b) Now we label the outcomes of the tree with the sample values of N:



From the tree,

$$P_N(n) = \begin{cases} q & n = 1, \\ (1-q)q & n = 2, \\ (1-q)^2 & n = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 3.2.5 Solution

(a) To find c, we apply the constraint $\sum_n P_N(n) = 1$, yielding

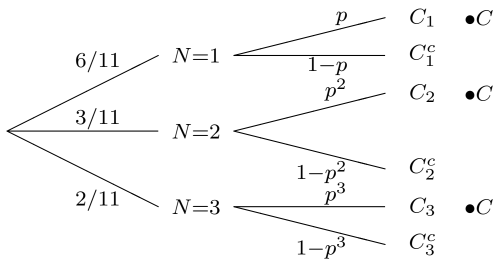
$$1 = \sum_{n=1}^{3} \frac{c}{n} = c\left(1 + \frac{1}{2} + \frac{1}{3}\right) = c\left(\frac{11}{6}\right). \tag{1}$$

Thus $c = 6/11$.

(b) The probability that N is odd is

$$\mathrm{P}[N \text{ is odd}] = P_N(1) + P_N(3) = c\left(1+\frac{1}{3}\right) = c\left(\frac{4}{3}\right) = \frac{24}{33}. \tag{2}$$

(c) We can view this as a sequential experiment: first we divide the file into N packets and then we check that all N packets are received correctly. In the second stage, we could specify how many packets are received correctly; however, it is sufficient to just specify whether the N packets are all received correctly or not. Using C_n to denote the event that n packets are transmitted and received correctly, we have

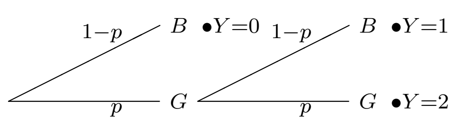


We see that

$$\begin{aligned}\mathrm{P}[C] &= \mathrm{P}[C_1] + \mathrm{P}[C_2] + \mathrm{P}[C_3] \\ &= \frac{6p}{11} + \frac{3p^2}{11} + \frac{2p^3}{11} = \frac{p(6+3p+2p^2)}{11}.\end{aligned} \tag{3}$$

Problem 3.2.6 Solution

Using B (for Bad) to denote a miss and G (for Good) to denote a successful free throw, the sample tree for the number of points scored in the 1 and 1 is

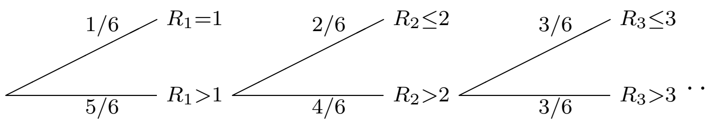


From the tree, the PMF of Y is

$$P_Y(y) = \begin{cases} 1-p & y=0, \\ p(1-p) & y=1, \\ p^2 & y=2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 3.2.7 Solution

Note that $N > 3$ if we roll three rolls satisfying $R_1 > 1$, $R_2 > 2$ and $R_3 > 3$. The tree for this experiment is



We note that

$$\mathrm{P}\left[N > 3\right] = \mathrm{P}\left[R_1 > 1, R_2 > 2, R_3 > 3\right] = \frac{5}{6} \cdot \frac{4}{6} \cdot \frac{3}{6} = \frac{5}{18}. \tag{1}$$

Problem 3.2.8 Solution

The probability that a caller fails to get through in three tries is $(1-p)^3$. To be sure that at least 95% of all callers get through, we need $(1-p)^3 \leq 0.05$. This implies $p = 0.6316$.

Problem 3.2.9 Solution

In Problem 3.2.8, each caller is willing to make 3 attempts to get through. An attempt is a failure if all n operators are busy, which occurs with probability $q = (0.8)^n$. Assuming call attempts are independent, a caller will suffer three failed attempts with probability $q^3 = (0.8)^{3n}$. The problem statement requires that $(0.8)^{3n} \leq 0.05$. This implies $n \geq 4.48$ and so we need 5 operators.

Problem 3.2.10 Solution

From the problem statement, a single is twice as likely as a double, which is twice as likely as a triple, which is twice as likely as a home-run. If p is the probability of a home run, then

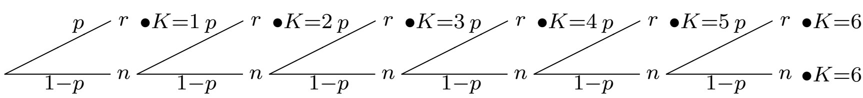
$$P_B(4) = p \qquad P_B(3) = 2p \qquad P_B(2) = 4p \qquad P_B(1) = 8p \tag{1}$$

Since a hit of any kind occurs with probability of .300, $p + 2p + 4p + 8p = 0.300$ which implies $p = 0.02$. Hence, the PMF of B is

$$P_B(b) = \begin{cases} 0.70 & b = 0, \\ 0.16 & b = 1, \\ 0.08 & b = 2, \\ 0.04 & b = 3, \\ 0.02 & b = 4, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 3.2.11 Solution

(a) In the setup of a mobile call, the phone will send the "SETUP" message up to six times. Each time the setup message is sent, we have a Bernoulli trial with success probability p. Of course, the phone stops trying as soon as there is a success. Using r to denote a successful response, and n a non-response, the sample tree is



(b) We can write the PMF of K, the number of "SETUP" messages sent as

$$P_K(k) = \begin{cases} (1-p)^{k-1}p & k = 1, 2, \ldots, 5, \\ (1-p)^5 p + (1-p)^6 = (1-p)^5 & k = 6, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Note that the expression for $P_K(6)$ is different because $K = 6$ if either there was a success or a failure on the sixth attempt. In fact, $K = 6$ whenever there were failures on the first five attempts which is why $P_K(6)$ simplifies to $(1-p)^5$.

(c) Let B denote the event that a busy signal is given after six failed setup attempts. The probability of six consecutive failures is $\mathrm{P}[B] = (1-p)^6$.

(d) To be sure that $\mathrm{P}[B] \leq 0.02$, we need $p \geq 1 - (0.02)^{1/6} = 0.479$.

Problem 3.3.1 Solution

(a) If it is indeed true that Y, the number of yellow M&M's in a package, is uniformly distributed between 5 and 15, then the PMF of Y, is

$$P_Y(y) = \begin{cases} 1/11 & y = 5, 6, 7, \dots, 15 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

(b)

$$\mathrm{P}[Y < 10] = P_Y(5) + P_Y(6) + \cdots + P_Y(9) = 5/11. \tag{2}$$

(c)

$$\mathrm{P}[Y > 12] = P_Y(13) + P_Y(14) + P_Y(15) = 3/11. \tag{3}$$

(d)

$$\mathrm{P}[8 \leq Y \leq 12] = P_Y(8) + P_Y(9) + \cdots + P_Y(12) = 5/11. \tag{4}$$

Problem 3.3.2 Solution

Since there are five colors, each M&M is red with probability $p = 1/5$. Since each bag has 25 M&Ms, R is a binomial $(n = 25, p = 1/5)$ random variable. Thus,

$$P_R(r) = \binom{25}{r} \left(\frac{1}{5}\right)^r \left(\frac{4}{5}\right)^{25-r}. \tag{1}$$

The probability of no red pieces is $P_R(0) = (4/5)^{25} = 0.0038$.

Problem 3.3.3 Solution

(a) Each paging attempt is an independent Bernoulli trial with success probability p. The number of times K that the pager receives a message is the number of successes in n Bernoulli trials and has the binomial PMF

$$P_K(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) Let R denote the event that the paging message was received at least once. The event R has probability

$$\mathrm{P}[R] = \mathrm{P}[B > 0] = 1 - \mathrm{P}[B = 0] = 1 - (1-p)^n. \tag{2}$$

To ensure that $\mathrm{P}[R] \geq 0.95$ requires that $n \geq \ln(0.05)/\ln(1-p)$. For $p = 0.8$, we must have $n \geq 1.86$. Thus, $n = 2$ pages would be necessary.

Problem 3.3.4 Solution

The roll of the two dice are denoted by the pair

$$(i, j) \in S = \{(1,1), (1,2), \ldots, (6,6)\}. \tag{1}$$

Each pair is an outcome. There are 36 pairs and each has probability 1/36. The event "doubles" is $\{(1,1), (2,2), \ldots, (6.6)\}$ has probability $p = 6/36 = 1/6$. If we define "doubles" as a successful roll, the number of rolls N until we observe doubles is a geometric (p) random variable and has expected value $\mathrm{E}[N] = 1/p = 6$.

Problem 3.3.5 Solution

Whether a hook catches a fish is an independent trial with success probability h. The the number of fish hooked, K, has the binomial PMF

$$P_K(k) = \begin{cases} \binom{m}{k} h^k (1-h)^{m-k} & k = 0, 1, \ldots, m, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 3.3.6 Solution

(a) Let X be the number of times the frisbee is thrown until the dog catches it and runs away. Each throw of the frisbee can be viewed as a Bernoulli trial in which a success occurs if the dog catches the frisbee an runs away. Thus, the experiment ends on the first success and X has the geometric PMF

$$P_X(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

(b) The child will throw the frisbee more than four times iff there are failures on the first 4 trials which has probability $(1-p)^4$. If $p = 0.2$, the probability of more than four throws is $(0.8)^4 = 0.4096$.

Problem 3.3.7 Solution

Each paging attempt is a Bernoulli trial with success probability p where a success occurs if the pager receives the paging message.

(a) The paging message is sent again and again until a success occurs. Hence the number of paging messages is $N = n$ if there are $n-1$ paging failures followed by a paging success. That is, N has the geometric PMF

$$P_N(n) = \begin{cases} (1-p)^{n-1}p & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The probability that no more three paging attempts are required is

$$\mathrm{P}[N \leq 3] = 1 - \mathrm{P}[N > 3] = 1 - \sum_{n=4}^{\infty} P_N(n) = 1 - (1-p)^3. \tag{2}$$

This answer can be obtained without calculation since $N > 3$ if the first three paging attempts fail and that event occurs with probability $(1-p)^3$. Hence, we must choose p to satisfy $1 - (1-p)^3 \geq 0.95$ or $(1-p)^3 \leq 0.05$. This implies

$$p \geq 1 - (0.05)^{1/3} \approx 0.6316. \tag{3}$$

Problem 3.3.8 Solution

The probability of more than 500,000 bits is

$$P[B > 500{,}000] = 1 - \sum_{b=1}^{500{,}000} P_B(b) \tag{1}$$

$$= 1 - p \sum_{b=1}^{500{,}000} (1-p)^{b-1}. \tag{2}$$

Math Fact B.4 implies that $(1-x)\sum_{b=1}^{500{,}000} x^{b-1} = 1 - x^{500{,}000}$. Substituting, $x = 1 - p$, we obtain:

$$\begin{aligned} P[B > 500{,}000] &= 1 - (1 - (1-p)^{500{,}000}) \\ &= (1 - 0.25 \times 10^{-5})^{500{,}000} \\ &\approx \exp(-500{,}000/400{,}000) = 0.29. \end{aligned} \tag{3}$$

Problem 3.3.9 Solution

(a) K is a Pascal $(5, p = 0.1)$ random variable and has PMF

$$P_K(k) = \binom{k-1}{4} p^5 (1-p)^{k-5} = \binom{k-1}{4} (0.1)^5 (0.9)^{k-5}. \tag{1}$$

(b) L is a Pascal $(k = 33, p = 1/2)$ random variable and so its PMF is

$$P_L(l) = \binom{l-1}{32} p^{33} (1-p)^{l-33} = \binom{l-1}{32} \left(\frac{1}{2}\right)^l. \tag{2}$$

(c) M is a geometric $(p = 0.01)$ random variable, You should know that that M has PMF

$$P_M(m) = \begin{cases} (1-p)^{m-1} p & m = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 3.3.10 Solution

Since an average of $T/5$ buses arrive in an interval of T minutes, buses arrive at the bus stop at a rate of $1/5$ buses per minute.

(a) From the definition of the Poisson PMF, the PMF of B, the number of buses in T minutes, is

$$P_B(b) = \begin{cases} (T/5)^b e^{-T/5}/b! & b = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) Choosing $T = 2$ minutes, the probability that three buses arrive in a two minute interval is

$$P_B(3) = (2/5)^3 e^{-2/5}/3! \approx 0.0072. \tag{2}$$

(c) By choosing $T = 10$ minutes, the probability of zero buses arriving in a ten minute interval is

$$P_B(0) = e^{-10/5}/0! = e^{-2} \approx 0.135. \tag{3}$$

(d) The probability that at least one bus arrives in T minutes is

$$\mathrm{P}[B \geq 1] = 1 - \mathrm{P}[B = 0] = 1 - e^{-T/5} \geq 0.99. \tag{4}$$

Rearranging yields $T \geq 5 \ln 100 \approx 23.0$ minutes.

Problem 3.3.11 Solution

(a) If each message is transmitted 8 times and the probability of a successful transmission is p, then the PMF of N, the number of successful transmissions has the binomial PMF

$$P_N(n) = \begin{cases} \binom{8}{n} p^n (1-p)^{8-n} & n = 0, 1, \ldots, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The indicator random variable I equals zero if and only if $N = 8$. Hence,

$$\mathrm{P}[I = 0] = \mathrm{P}[N = 0] = 1 - \mathrm{P}[I = 1] \tag{2}$$

Thus, the complete expression for the PMF of I is

$$P_I(i) = \begin{cases} (1-p)^8 & i = 0, \\ 1-(1-p)^8 & i = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 3.3.12 Solution

The requirement that $\sum_{x=1}^{n} P_X(x) = 1$ implies

$$n = 1: \qquad c(1)\left[\frac{1}{1}\right] = 1, \qquad c(1) = 1, \tag{1}$$

$$n = 2: \qquad c(2)\left[\frac{1}{1} + \frac{1}{2}\right] = 1, \qquad c(2) = \frac{2}{3}, \tag{2}$$

$$n = 3: \qquad c(3)\left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3}\right] = 1, \qquad c(3) = \frac{6}{11}, \tag{3}$$

$$n = 4: \qquad c(4)\left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right] = 1, \qquad c(4) = \frac{12}{25}, \tag{4}$$

$$n = 5: \qquad c(5)\left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right] = 1, \qquad c(5) = \frac{12}{25}, \tag{5}$$

$$n = 6: \qquad c(6)\left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{6}\right] = 1, \qquad c(6) = \frac{20}{49}. \tag{6}$$

As an aside, find $c(n)$ for large values of n is easy using the recursion

$$\frac{1}{c(n+1)} = \frac{1}{c(n)} + \frac{1}{n+1}. \tag{7}$$

Problem 3.3.13 Solution

(a) Each of the four m&m's is equally likely to be red or green. Hence the number of red m&m's is a binomial ($n = 4, p = 1/2$) random variable N with PMF

$$P_N(n) = \binom{4}{n}(1/2)^n(1/2)^{4-n} = \binom{4}{n}\left(\frac{1}{16}\right). \tag{1}$$

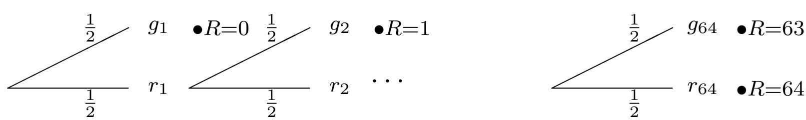
The probability of an equal number of red and green is

$$\mathrm{P}[E] = \mathrm{P}[N = 2] = P_N(2) = \binom{4}{2}\left(\frac{1}{16}\right) = \frac{3}{8}. \tag{2}$$

(b) In the bag of 64 m&m's, each m&m is green with probability 1/2 so that G is a binomial ($n = 64, p = 1/2$) random variable with PMF

$$P_G(g) = \binom{64}{g}(1/2)^g(1/2)^{64-g} = \binom{64}{g}2^{-64}. \tag{3}$$

(c) This is similar to the number of geometric number number of trials needed for the first success, except things are a little trickier because the bag may have all red m&m's. To be more clear, we will use r_i and g_i to denote the color of the ith m&m eaten. The tree is



From the tree, we see that

$$\mathrm{P}[R = 0] = 2^{-1}, \qquad \mathrm{P}[R = 1] = 2^{-2}, \cdots \qquad \mathrm{P}[R = 63] = 2^{-64}, \tag{4}$$

and $\mathrm{P}[R = 64] = 2^{-64}$. The complete PMF of R is

$$P_R(r) = \begin{cases} (1/2)^{r+1} & r = 0, 1, 2, \ldots, 63, \\ (1/2)^{64} & r = 64, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 3.3.14 Solution

(a) We can view whether each caller knows the birthdate as a Bernoulli trial. As a result, L is the number of trials needed for 6 successes. That is, L has a Pascal PMF with parameters $p = 0.75$ and $k = 6$ as defined by Definition 3.7. In particular,

$$P_L(l) = \begin{cases} \binom{l-1}{5}(0.75)^6(0.25)^{l-6} & l = 6, 7, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The probability of finding the winner on the tenth call is

$$P_L(10) = \binom{9}{5}(0.75)^6(0.25)^4 \approx 0.0876. \tag{2}$$

(c) The probability that the station will need nine or more calls to find a winner is

$$\begin{aligned} \mathrm{P}[L \geq 9] &= 1 - \mathrm{P}[L < 9] \\ &= 1 - P_L(6) - P_L(7) - P_L(8) \\ &= 1 - (0.75)^6[1 + 6(0.25) + 21(0.25)^2] \approx 0.321. \end{aligned} \tag{3}$$

Problem 3.3.15 Solution

The packets are delay sensitive and can only be retransmitted d times. For $t < d$, a packet is transmitted t times if the first $t-1$ attempts fail followed by a successful transmission on attempt t. Further, the packet is transmitted d times if there are failures on the first $d-1$ transmissions, no matter what the outcome of attempt d. So the random variable T, the number of times that a packet is transmitted, can be represented by the following PMF.

$$P_T(t) = \begin{cases} p(1-p)^{t-1} & t = 1, 2, \ldots, d-1, \\ (1-p)^{d-1} & t = d, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 3.3.16 Solution

(a) Note that $L_1 = l$ if there are l consecutively arriving planes landing in l minutes, which has probability p^l, followed by an idle minute without an arrival, which has probability $1-p$. Thus,

$$P_{L_1}(l) = \begin{cases} p^l(1-p) & l = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) As indicated, $W = 10$ if there are ten consecutive takeoffs in ten minutes. This occurs if there are no landings for 10 minutes, which occurs with probability $(1-p)^{10}$. Thus $\mathrm{P}[W = 10] = (1-p)^{10}$.

(c) The key is to recognize that each takeoff is a "success" with success probability $1-p$. Moreover, in each one minute slot, a success occurs with probability $1-p$, independent of the result of the trial in any other one minute slot. Since your plane is the tenth in line, your plane takes off when the tenth success occurs. Thus W is a Pascal $(10, 1-p)$ random variable and has PMF

$$\begin{aligned} P_W(w) &= \binom{w-1}{9}(1-p)^{10}(1-(1-p))^{w-10} \\ &= \binom{w-1}{9}(1-p)^{10}p^{w-10}. \end{aligned} \tag{2}$$

Problem 3.3.17 Solution

(a) Since each day is independent of any other day, $\mathrm{P}[W_{33}]$ is just the probability that a winning lottery ticket was bought. Similarly for $\mathrm{P}[L_{87}]$ and $\mathrm{P}[N_{99}]$ become just the probability that a losing ticket was bought and that no ticket was bought on a single day, respectively. Therefore

$$\mathrm{P}[W_{33}] = p/2, \qquad \mathrm{P}[L_{87}] = (1-p)/2, \qquad \mathrm{P}[N_{99}] = 1/2. \tag{1}$$

(b) Suppose we say a success occurs on the kth trial if on day k we buy a ticket. Otherwise, a failure occurs. The probability of success is simply 1/2. The random variable K is just the number of trials until the first success and has the geometric PMF

$$P_K(k) = \begin{cases} (1/2)(1/2)^{k-1} = (1/2)^k & k = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) The probability that you decide to buy a ticket and it is a losing ticket is $(1-p)/2$, independent of any other day. If we view buying a losing ticket as a Bernoulli success, R, the number of losing lottery tickets bought in m days, has the binomial PMF

$$P_R(r) = \begin{cases} \binom{m}{r}[(1-p)/2]^r[(1+p)/2]^{m-r} & r = 0, 1, \ldots, m, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(d) Letting D be the day on which the j-th losing ticket is bought, we can find the probability that $D = d$ by noting that $j-1$ losing tickets must have been purchased in the $d-1$ previous days. Therefore D has the Pascal PMF

$$P_D(d) = \begin{cases} \binom{d-1}{j-1}[(1-p)/2]^j[(1+p)/2]^{d-j} & d = j, j+1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 3.3.18 Solution

(a) Let S_n denote the event that the Sixers win the series in n games. Similarly, C_n is the event that the Celtics in in n games. The Sixers win the series in 3 games if they win three straight, which occurs with probability

$$\mathrm{P}[S_3] = (1/2)^3 = 1/8. \tag{1}$$

The Sixers win the series in 4 games if they win two out of the first three games and they win the fourth game so that

$$\mathrm{P}[S_4] = \binom{3}{2}(1/2)^3(1/2) = 3/16. \tag{2}$$

The Sixers win the series in five games if they win two out of the first four games and then win game five. Hence,

$$\mathrm{P}[S_5] = \binom{4}{2}(1/2)^4(1/2) = 3/16. \tag{3}$$

By symmetry, $\mathrm{P}[C_n] = \mathrm{P}[S_n]$. Further we observe that the series last n games if either the Sixers or the Celtics win the series in n games. Thus,

$$\mathrm{P}[N = n] = \mathrm{P}[S_n] + \mathrm{P}[C_n] = 2\,\mathrm{P}[S_n]. \tag{4}$$

Consequently, the total number of games, N, played in a best of 5 series between the Celtics and the Sixers can be described by the PMF

$$P_N(n) = \begin{cases} 2(1/2)^3 = 1/4 & n = 3, \\ 2\binom{3}{1}(1/2)^4 = 3/8 & n = 4, \\ 2\binom{4}{2}(1/2)^5 = 3/8 & n = 5, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

(b) For the total number of Celtic wins W, we note that if the Celtics get $w < 3$ wins, then the Sixers won the series in $3 + w$ games. Also, the Celtics win 3 games if they win the series in 3,4, or 5 games. Mathematically,

$$\mathrm{P}[W = w] = \begin{cases} \mathrm{P}[S_{3+w}] & w = 0, 1, 2, \\ \mathrm{P}[C_3] + \mathrm{P}[C_4] + \mathrm{P}[C_5] & w = 3. \end{cases} \tag{6}$$

Thus, the number of wins by the Celtics, W, has the PMF shown below.

$$P_W(w) = \begin{cases} \mathrm{P}[S_3] = 1/8 & w = 0, \\ \mathrm{P}[S_4] = 3/16 & w = 1, \\ \mathrm{P}[S_5] = 3/16 & w = 2, \\ 1/8 + 3/16 + 3/16 = 1/2 & w = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

(c) The number of Celtic losses L equals the number of Sixers' wins W_S. This implies $P_L(l) = P_{W_S}(l)$. Since either team is equally likely to win any game, by symmetry, $P_{W_S}(w) = P_W(w)$. This implies $P_L(l) = P_{W_S}(l) = P_W(l)$. The complete expression of for the PMF of L is

$$P_L(l) = P_W(l) = \begin{cases} 1/8 & l = 0, \\ 3/16 & l = 1, \\ 3/16 & l = 2, \\ 1/2 & l = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Problem 3.3.19 Solution

Since a and b are positive, let K be a binomial random variable for n trials and success probability $p = a/(a+b)$. First, we observe that the sum of over all possible values of the PMF of K is

$$\begin{aligned} \sum_{k=0}^{n} P_K(k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} \left(\frac{a}{a+b}\right)^k \left(\frac{b}{a+b}\right)^{n-k} \\ &= \frac{\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}}{(a+b)^n}. \end{aligned} \tag{1}$$

Since $\sum_{k=0}^{n} P_K(k) = 1$, we see that

$$(a+b)^n = (a+b)^n \sum_{k=0}^{n} P_K(k) = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{2}$$

Problem 3.4.1 Solution

Using the CDF given in the problem statement we find that

(a) $P[Y < 1] = 0$ and $P[Y \leq 1] = 1/4$.

(b)

$$\mathrm{P}[Y > 2] = 1 - \mathrm{P}[Y \leq 2] = 1 - 1/2 = 1/2. \quad (1)$$

$$\mathrm{P}[Y \geq 2] = 1 - \mathrm{P}[Y < 2] = 1 - 1/4 = 3/4. \quad (2)$$

(c)

$$\mathrm{P}[Y = 3] = F_Y\left(3^+\right) - F_Y\left(3^-\right) = 1/2. \quad (3)$$

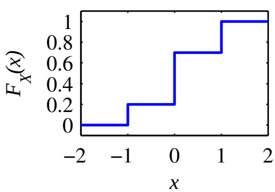
$$\mathrm{P}[Y > 3] = 1 - F_Y(3) = 0. \quad (4)$$

(d) From the staircase CDF of Problem 3.4.1, we see that Y is a discrete random variable. The jumps in the CDF occur at at the values that Y can take on. The height of each jump equals the probability of that value. The PMF of Y is

$$P_Y(y) = \begin{cases} 1/4 & y = 1, \\ 1/4 & y = 2, \\ 1/2 & y = 3, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Problem 3.4.2 Solution

(a) The given CDF is shown in the diagram below.



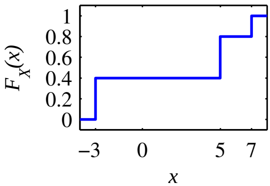
$$F_X(x) = \begin{cases} 0 & x < -1, \\ 0.2 & -1 \leq x < 0, \\ 0.7 & 0 \leq x < 1, \\ 1 & x \geq 1. \end{cases} \quad (1)$$

(b) The corresponding PMF of X is

$$P_X(x) = \begin{cases} 0.2 & x = -1, \\ 0.5 & x = 0, \\ 0.3 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Problem 3.4.3 Solution

(a) Similar to the previous problem, the graph of the CDF is shown below.



$$F_X(x) = \begin{cases} 0 & x < -3, \\ 0.4 & -3 \leq x < 5, \\ 0.8 & 5 \leq x < 7, \\ 1 & x \geq 7. \end{cases} \tag{1}$$

(b) The corresponding PMF of X is

$$P_X(x) = \begin{cases} 0.4 & x = -3 \\ 0.4 & x = 5 \\ 0.2 & x = 7 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Problem 3.4.4 Solution

Let $q = 1 - p$, so the PMF of the geometric (p) random variable K is

$$P_K(k) = \begin{cases} pq^{k-1} & k = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For any integer $k \geq 1$, the CDF obeys

$$F_K(k) = \sum_{j=1}^{k} P_K(j) = \sum_{j=1}^{k} pq^{j-1} = 1 - q^k. \tag{2}$$

Since K is integer valued, $F_K(k) = F_K(\lfloor k \rfloor)$ for all integer and non-integer values of k. (If this point is not clear, you should review Example 3.22.) Thus, the complete expression for the CDF of K is

$$F_K(k) = \begin{cases} 0 & k < 1, \\ 1 - (1-p)^{\lfloor k \rfloor} & k \geq 1. \end{cases} \tag{3}$$

Problem 3.4.5 Solution

Since mushrooms occur with probability 2/3, the number of pizzas sold before the first mushroom pizza is $N = n < 100$ if the first n pizzas do not have mushrooms followed by mushrooms on pizza $n+1$. Also, it is possible that $N = 100$ if all 100 pizzas are sold without mushrooms. the resulting PMF is

$$P_N(n) = \begin{cases} (1/3)^n(2/3) & n = 0, 1, \ldots, 99, \\ (1/3)^{100} & n = 100, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For integers $n < 100$, the CDF of N obeys

$$F_N(n) = \sum_{i=0}^{n} P_N(i) = \sum_{i=0}^{n} (1/3)^i(2/3) = 1 - (1/3)^{n+1}. \tag{2}$$

A complete expression for $F_N(n)$ must give a valid answer for every value of n, including non-integer values. We can write the CDF using the floor function $\lfloor x \rfloor$ which denote the largest integer less than or equal to X. The complete expression for the CDF is

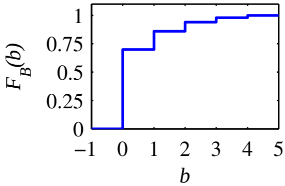
$$F_N(x) = \begin{cases} 0 & x < 0, \\ 1 - (1/3)^{\lfloor x \rfloor + 1} & 0 \le x < 100, \\ 1 & x \ge 100. \end{cases} \tag{3}$$

Problem 3.4.6 Solution

From Problem 3.2.10, the PMF of B is

$$P_B(b) = \begin{cases} 0.70 & b = 0, \\ 0.16 & b = 1, \\ 0.08 & b = 2, \\ 0.04 & b = 3, \\ 0.02 & b = 4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The corresponding CDF is



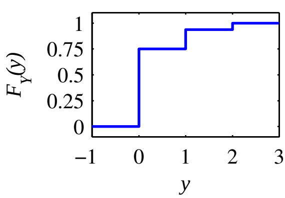
$$F_B(b) = \begin{cases} 0 & b < 0, \\ 0.70 & 0 \leq b < 1, \\ 0.86 & 1 \leq b < 2, \\ 0.94 & 2 \leq b < 3, \\ 0.98 & 3 \leq b < 4, \\ 1.0 & b \geq 4. \end{cases} \tag{2}$$

Problem 3.4.7 Solution

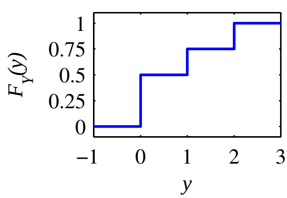
In Problem 3.2.6, we found the PMF of Y. This PMF, and its corresponding CDF are

$$P_Y(y) = \begin{cases} 1-p & y = 0, \\ p(1-p) & y = 1, \\ p^2 & y = 2, \\ 0 & \text{otherwise,} \end{cases} \qquad F_Y(y) = \begin{cases} 0 & y < 0, \\ 1-p & 0 \leq y < 1, \\ 1-p^2 & 1 \leq y < 2, \\ 1 & y \geq 2. \end{cases} \tag{1}$$

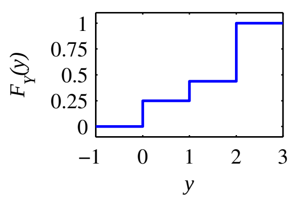
For the three values of p, the CDF resembles



$p = 1/4$



$p = 1/2$



$p = 3/4$

Problem 3.4.8 Solution

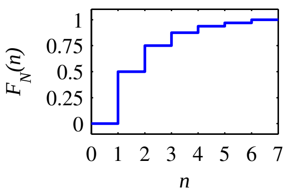
From Problem 3.2.11, the PMF of the number of call attempts is

$$P_N(n) = \begin{cases} (1-p)^{k-1}p & k = 1, 2, \ldots, 5, \\ (1-p)^5 p + (1-p)^6 = (1-p)^5 & k = 6, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For $p = 1/2$, the PMF can be simplified to

$$P_N(n) = \begin{cases} (1/2)^n & n = 1, 2, \ldots, 5, \\ (1/2)^5 & n = 6, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The corresponding CDF of N is



$$F_N(n) = \begin{cases} 0 & n < 1, \\ 1/2 & 1 \le n < 2, \\ 3/4 & 2 \le n < 3, \\ 7/8 & 3 \le n < 4, \\ 15/16 & 4 \le n < 5, \\ 31/32 & 5 \le n < 6, \\ 1 & n \ge 6. \end{cases} \tag{3}$$

Problem 3.5.1 Solution

For this problem, we just need to pay careful attention to the definitions of mode and median.

(a) The mode must satisfy $P_X(x_{\text{mod}}) \ge P_X(x)$ for all x. In the case of the uniform PMF, any integer x' between 1 and 100 is a mode of the random variable X. Hence, the set of all modes is

$$X_{\text{mod}} = \{1, 2, \ldots, 100\}. \tag{1}$$

(b) The median must satisfy $\text{P}[X < x_{\text{med}}] = \text{P}[X > x_{\text{med}}]$. Since

$$\text{P}[X \le 50] = \text{P}[X \ge 51] = 1/2. \tag{2}$$

we observe that $x_{\text{med}} = 50.5$ is a median since it satisfies

$$\text{P}[X < x_{\text{med}}] = \text{P}[X > x_{\text{med}}] = 1/2. \tag{3}$$

In fact, for any x' satisfying $50 < x' < 51$, $\text{P}[X < x'] = \text{P}[X > x'] = 1/2$. Thus,

$$X_{\text{med}} = \{x \mid 50 < x < 51\}. \tag{4}$$

Problem 3.5.2 Solution

Voice calls and data calls each cost 20 cents and 30 cents respectively. Furthermore the respective probabilities of each type of call are 0.6 and 0.4.

(a) Since each call is either a voice or data call, the cost of one call can only take the two values associated with the cost of each type of call. Therefore the PMF of X is

$$P_X(x) = \begin{cases} 0.6 & x = 20, \\ 0.4 & x = 30, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The expected cost, $E[C]$, is simply the sum of the cost of each type of call multiplied by the probability of such a call occurring.

$$E[C] = 20(0.6) + 30(0.4) = 24 \text{ cents.} \tag{2}$$

Problem 3.5.3 Solution

(a) J has the Poisson PMF

$$P_J(j) = \begin{cases} t^j e^{-t}/j! & j = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

It follows that

$$0.9 = P[J > 0] = 1 - P_J(0) = 1 - e^{-t} \qquad \Longrightarrow \ t = \ln(10) = 2.3. \tag{2}$$

(b) For $k = 0, 1, 2, \ldots$, $P_K(k) = 10^k e^{-10}/k!$. Thus

$$P[K = 10] = P_K(10) = 10^{10} e^{-10} = 0.1251. \tag{3}$$

(c) L is a Poisson $(\alpha = E[L] = 2)$ random variable. Thus its PMF is

$$P_L(l) = \begin{cases} 2^l e^{-2}/l! & l = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

It follows that

$$P[L \leq 1] = P_L(0) + P_L(1) = 3e^{-2} = 0.406. \tag{5}$$

Problem 3.5.4 Solution

Both coins come up heads with probability $p = 1/4$ and we call this a success. Thus X, the number of successes in 10 trials, is a binomial $(n = 10, p = 1/4)$ random variable with PMF

$$P_X(x) = \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}. \tag{1}$$

You should know that $\mathrm{E}[X] = np = 2.5$. Since the average return is 2.5 dollars, the probability you do "worse than average" is

$$\begin{aligned} \mathrm{P}[X < \mathrm{E}[X]] &= \mathrm{P}[X < 2.5] \\ &= \mathrm{P}[X \leq 2] \\ &= P_X(0) + P_X(1) + P_X(2) \\ &= \left(\frac{3}{4}\right)^{10} + 10\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^9 + 45\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^8 \\ &= 7\left(\frac{3}{4}\right)^9 = 0.5256. \end{aligned} \tag{2}$$

Problem 3.5.5 Solution

(a) Each packet transmission is a Bernoulli trial with success probability 0.95 and X is the number of packet failures (received in error) in 10 trials. Since the failure probability is $p = 0.05$, X has the binomial $(n = 10, p = 0.05)$ PMF

$$P_X(x) = \binom{10}{x}(0.05)^x(0.95)^{10-x}. \tag{1}$$

(b) When you send 12 thousand packets, the number of packets received in error, Y, is a binomial $(n = 12000, p = 0.05)$ random variable. The expected number received in error is $\mathrm{E}[Y] = np = 600$ per hour, or about 10 packets per minute. Keep in mind this is a reasonable figure if you are an active data user.

Problem 3.5.6 Solution

From the solution to Problem 3.4.1, the PMF of Y is

$$P_Y(y) = \begin{cases} 1/4 & y = 1, \\ 1/4 & y = 2, \\ 1/2 & y = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of Y is

$$E[Y] = \sum_y y P_Y(y) = 1(1/4) + 2(1/4) + 3(1/2) = 9/4. \tag{2}$$

Problem 3.5.7 Solution

From the solution to Problem 3.4.2, the PMF of X is

$$P_X(x) = \begin{cases} 0.2 & x = -1, \\ 0.5 & x = 0, \\ 0.3 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of X is

$$E[X] = \sum_x x P_X(x) = -1(0.2) + 0(0.5) + 1(0.3) = 0.1. \tag{2}$$

Problem 3.5.8 Solution

From the solution to Problem 3.4.3, the PMF of X is

$$P_X(x) = \begin{cases} 0.4 & x = -3, \\ 0.4 & x = 5, \\ 0.2 & x = 7, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of X is

$$E[X] = \sum_x x P_X(x) = -3(0.4) + 5(0.4) + 7(0.2) = 2.2. \tag{2}$$

Problem 3.5.9 Solution

From Definition 3.6, random variable X has PMF

$$P_X(x) = \begin{cases} \binom{4}{x}(1/2)^4 & x = 0, 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of X is

$$\begin{aligned} \mathrm{E}[X] &= \sum_{x=0}^{4} x P_X(x) \\ &= 0\binom{4}{0}\frac{1}{2^4} + 1\binom{4}{1}\frac{1}{2^4} + 2\binom{4}{2}\frac{1}{2^4} + 3\binom{4}{3}\frac{1}{2^4} + 4\binom{4}{4}\frac{1}{2^4} \\ &= [4 + 12 + 12 + 4]/2^4 = 2. \end{aligned} \tag{2}$$

Problem 3.5.10 Solution

X has PMF

$$P_X(x) = \begin{cases} 1/5 & x = 1, 2, \ldots, 5, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

and expected value $\mathrm{E}[X] = 3$.

(a) It follows that $\mathrm{P}[X = \mathrm{E}[X]] = P_X(3) = 1/5$.

(b) Also,

$$\mathrm{P}[X > \mathrm{E}[X]] = \mathrm{P}[X > 3] = P_X(4) + P_X(5) = 2/5. \tag{2}$$

Problem 3.5.11 Solution

K has expected value $\mathrm{E}[K] = 1/p = 11$ and PMF

$$P_K(k) = \begin{cases} (1-p)^{k-1}p & k = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) From these facts,

$$\begin{aligned}
\mathrm{P}[K = \mathrm{E}[K]] = P_K(11) &= (1-p)^{10}p \\
&= (10/11)^{10}(1/11) = 10^{10}/11^{11} = 0.035.
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{P}[K > \mathrm{E}[K]] &= \mathrm{P}[K > 11] \\
&= \sum_{x=12}^{\infty} P_K(x) \\
&= \sum_{x=12}^{\infty} (1-p)^{x-1}p \\
&= p[(1-p)^{11} + (1-p)^{12} + \cdots] \\
&= p(1-p)^{11}[1 + (1-p) + (1-p)^2 + \cdots] \\
&= (1-p)^{11} = (10/11)^{11} = 0.3505. \qquad (2)
\end{aligned}$$

The answer $(1-p)^{11}$ can also be found by recalling that $K > 11$ if and only if there are 11 failures before the first success, an event which has probability $(1-p)^{11}$.

(c)

$$\begin{aligned}
\mathrm{P}[K < \mathrm{E}[K]] &= 1 - \mathrm{P}[K \geq \mathrm{E}[K]] \\
&= 1 - (\mathrm{P}[K = \mathrm{E}[K]] + \mathrm{P}[K > \mathrm{E}[K]]) \\
&= 1 - \left((10/11)^{10}(1/11) + (10/11)^{11}\right) \\
&= 1 - (10/11)^{10}. \qquad (3)
\end{aligned}$$

Note that $(10/11)^{10}$ is the probability of ten straight failures. As long as this does NOT occur, then $K < 11$.

Problem 3.5.12 Solution

(a) The probability of 20 heads in a row is $p = (1/2)^{20} = 2^{-20}$. Hence the PMF of R is

$$P_R(r) = \begin{cases} 1 - 2^{-20} & r = 0, \\ 2^{-20} & r = 20 \cdot 10^6, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) Let's say a success occurs if a contestant wins the game. The success probability is $p = 2^{-20}$. Note that there are $L = l$ losing customers if we observe l failures followed by a success on trial $l + 1$. Note that $L = 0$ is possible. Thus $\mathrm{P}[L = l] = (1 - p)^l p$. The PMF of L is

$$P_L(l) = \begin{cases} (1 - p)^l p & l = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) The expected reward earned by a customer is

$$\mathrm{E}[R] = 20 \cdot 10^6 p = \frac{20 \cdot 10^6}{2^{20}}. \tag{3}$$

You might know that $2^{10} = 1024$ and so $2^{20} = (1024)^2 > 10^6$. Thus $\mathrm{E}[R] < 20$. In fact, $\mathrm{E}[R] = 19.07$. That is, the casino collects \$20 from each player but on average pays out \$19.07 to each customer.

Problem 3.5.13 Solution

The following experiments are based on a common model of packet transmissions in data networks. In these networks, each data packet contains a cylic redundancy check (CRC) code that permits the receiver to determine whether the packet was decoded correctly. In the following, we assume that a packet is corrupted with probability $\epsilon = 0.001$, independent of whether any other packet is corrupted.

(a) Let $X = 1$ if a data packet is decoded correctly; otherwise $X = 0$. Random variable X is a Bernoulli random variable with PMF

$$P_X(x) = \begin{cases} 0.001 & x = 0, \\ 0.999 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The parameter $\epsilon = 0.001$ is the probability a packet is corrupted. The expected value of X is

$$\mathrm{E}[X] = 1 - \epsilon = 0.999. \tag{2}$$

(b) Let Y denote the number of packets received in error out of 100 packets transmitted. Y has the binomial PMF

$$P_Y(y) = \begin{cases} \binom{100}{y}(0.001)^y(0.999)^{100-y} & y = 0, 1, \dots, 100, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The expected value of Y is

$$\mathrm{E}[Y] = 100\epsilon = 0.1. \tag{4}$$

(c) Let L equal the number of packets that must be received to decode 5 packets in error. L has the Pascal PMF

$$P_L(l) = \begin{cases} \binom{l-1}{4}(0.001)^5(0.999)^{l-5} & l = 5, 6, \dots, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The expected value of L is

$$\mathrm{E}[L] = \frac{5}{\epsilon} = \frac{5}{0.001} = 5000. \tag{6}$$

(d) If packet arrivals obey a Poisson model with an average arrival rate of 1000 packets per second, then the number N of packets that arrive in 5 seconds has the Poisson PMF

$$P_N(n) = \begin{cases} 5000^n e^{-5000}/n! & n = 0, 1, \dots, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The expected value of N is $\mathrm{E}[N] = 5000$.

Problem 3.5.14 Solution

(a) When K is geometric ($p = 1/3$), it has PMF

$$P_K(k) = \begin{cases} (1-p)^{k-1}p & k = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

and $\mathrm{E}[K] = 1/p = 3$. In this case,

$$\begin{aligned} \mathrm{P}[K < \mathrm{E}[K]] = \mathrm{P}[K < 3] &= P_K(1) + P_K(2) \\ &= p + (1-p)p = 5/9 = 0.555. \end{aligned} \tag{2}$$

(b) When K is binomial ($n = 6, p = 1/2$), it has PMF

$$P_K(k) = \binom{6}{k} p^k (1-p)^{6-k} = \binom{6}{k} \left(\frac{1}{2}\right)^k \tag{3}$$

and $\mathrm{E}[K] = np = 3$. In this case,

$$\begin{aligned} \mathrm{P}[K < \mathrm{E}[K]] = \mathrm{P}[K < 3] &= P_K(0) + P_K(1) + P_K(2) \\ &= \left(\frac{1}{2}\right)^6 (1 + 6 + 15) \\ &= \frac{22}{64} = \frac{11}{32} = 0.344. \end{aligned} \tag{4}$$

(c) When K is Poisson ($\alpha = 3$), it has PMF

$$\begin{aligned} P_K(k) &= \begin{cases} \alpha^k e^{-\alpha}/k! & k = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 3^k e^{-3}/k! & k = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{5}$$

and $E[K] = \alpha = 3$. In this case,

$$\begin{aligned} P[K < E[K]] = P[K < 3] &= P_K(0) + P_K(1) + P_K(2) \\ &= e^{-3}\left(1 + 3 + \frac{3^2}{3!}\right) \\ &= (11/2)e^{-3} = 0.2738. \end{aligned} \tag{6}$$

(d) When K is discrete uniform $(0, 6)$, it has PMF

$$P_K(k) = \begin{cases} 1/7 & k = 0, 1, \ldots, 6, \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

and $E[K] = (0 + 6)/2 = 3$. In this case,

$$\begin{aligned} P[K < E[K]] = P[K < 3] &= P_K(0) + P_K(1) + P_K(2) \\ &= 3/7 = 0.429. \end{aligned} \tag{8}$$

Problem 3.5.15 Solution

In this "double-or-nothing" type game, there are only two possible payoffs. The first is zero dollars, which happens when we lose 6 straight bets, and the second payoff is 64 dollars which happens unless we lose 6 straight bets. So the PMF of Y is

$$P_Y(y) = \begin{cases} (1/2)^6 = 1/64 & y = 0, \\ 1 - (1/2)^6 = 63/64 & y = 64, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

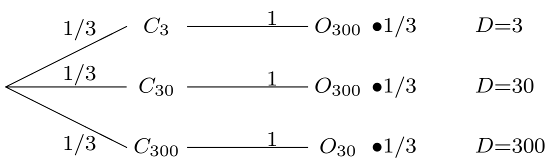
The expected amount you take home is

$$E[Y] = 0(1/64) + 64(63/64) = 63. \tag{2}$$

So, on the average, we can expect to break even, which is not a very exciting proposition.

Problem 3.5.16 Solution

(a) This problem is a little simpler than Monty Hall because the host never has a choice. Assuming you never switch, the tree is

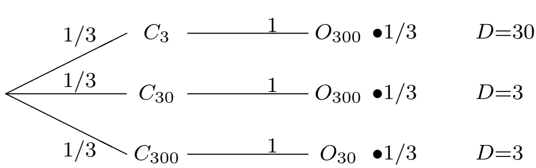


The PMF of D is

$$P_D(d) = \begin{cases} 1/3 & d = 3, 30, 300, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of D is $\mathrm{E}[D] = (3 + 30 + 300)/3 = 111$. In fact, the tree isn't really necessary for this part; because you never switch D is simply the dollars in the suitcase you initially choose.

(b) This problem is slightly less simple than the previous step since the reward D depends on what suitcase the host opens. Otherwise the tree is the same. Assuming you always switch, the tree is

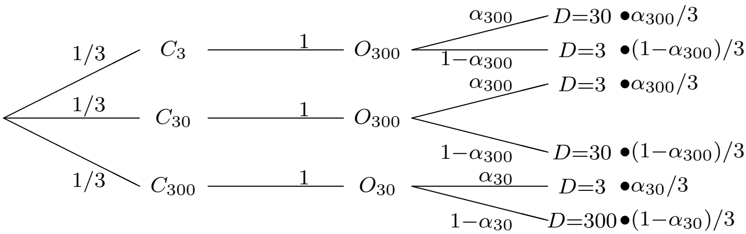


The PMF of D is

$$P_D(d) = \begin{cases} 2/3 & d = 3 \\ 1/3 & d = 30 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The expected value of D is $\mathrm{E}[D] = (2/3)3 + (1/3)30 = 12$. So switching is a terrible idea. This should be fairly intuitive since the host is leading you to switch to the suitcase with less money.

(c) This problem is a little complicated since the reward depends on your random decision. The tree is



We can solve for the PMF of D by adding the probabilities of the outcomes corresponding to each possible D. In terms of the α_i, we have

$$P_D(d) = \begin{cases} (1+\alpha_{30})/3 & d = 3, \\ 1/3 & d = 30, \\ (1-\alpha_{30})/3 & d = 300, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The expected value of D is

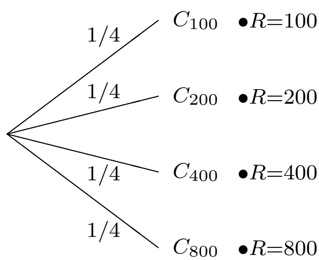
$$\mathrm{E}\left[D\right] = \left(\frac{1+\alpha_{30}}{3}\right)3 + \left(\frac{1}{3}\right)30 + \left(\frac{1-\alpha_{30}}{3}\right)300 = 111 - 99\alpha_{30}. \tag{4}$$

Note that we can choose any $\alpha_{30} \in [0,1]$. To maximize $\mathrm{E}[D]$, you should choose $\alpha_{30} = 0$. This is logical because if the host opens the suitcase with 30 dollars, then you should realize you have the suitcase with 300 dollars and you definitely should not switch!

However, adopting this rule earns you the same expected reward $\mathrm{E}[D] = 111$ that you get with the "never switch" rule! In fact, what is curious is that the expected reward is insensitive to α_{300}. You get the same average reward no matter what you do when the host opens the 300 dollar suitcase. Frankly, this is something of a surprise.

Problem 3.5.17 Solution

(a) Since you ignore the host, you pick your suitcase and open it. The tree is



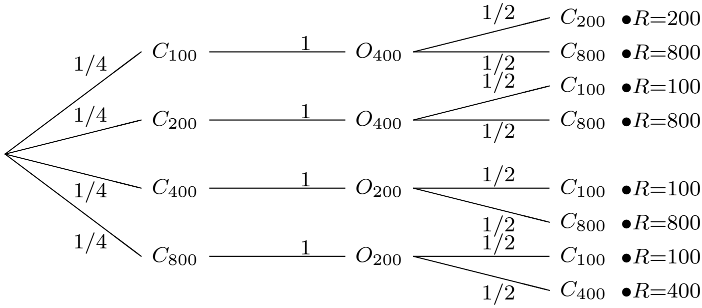
The PMF of R is just

$$P_R(r) = \begin{cases} 1/4 & r = 100, 200, 400, 800, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of R is

$$\mathrm{E}[R] = \sum_r r P_R(r) = \frac{1}{4}(100 + 200 + 400 + 800) = 375. \tag{2}$$

(b) In this case, the tree diagram is



All eight outcomes are equally likely. It follows that

r	100	200	400	800
$P_R(r)$	3/8	1/8	1/8	3/8

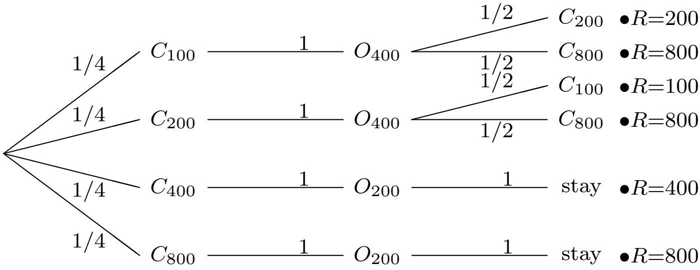
The expected value of R is

$$\mathrm{E}[R] = \frac{3}{8}(100) + \frac{1}{8}(200+400) + \frac{3}{8}(800) = 412.5. \tag{3}$$

(c) You can do better by making your decision whether to switch to one of the unopened suitcases depend on what suitcase the host opened. In particular, studying the tree from part (b), we see that if the host opens the \$200 suitcase, then your originally chosen suitcase is either the \$400 suitcase or the \$800 suitcase. That is, you learn you have already picked one of the two best suitcases and it seems likely that you would be better to not switch. On the other hand, if the host opens the \$400 suitcase, then you have learned that your original choice was either the \$100 or \$200 suitcase. In this case, switching gives you a chance to win the \$800 suitcase. In this case switching seems like a good idea. Thus, our intuition suggests that

- switch if the host opens the \$400 suitcase;
- stay with your original suitcase if the host opens the \$200 suitcase.

To verify that our intuition is correct, we construct the tree to evaluate this new switching policy:



It follows that

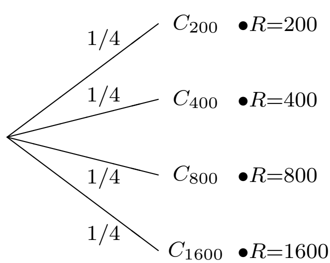
r	100	200	400	800
$P_R(r)$	1/8	1/8	1/4	1/2

The expected value of R is

$$\mathrm{E}[R] = \frac{1}{8}(100+200) + \frac{1}{4}(400) + \frac{1}{2}(800) = 537.5. \tag{4}$$

Problem 3.5.18 Solution

(a) Since you ignore the host, you pick your suitcase and open it. The tree is



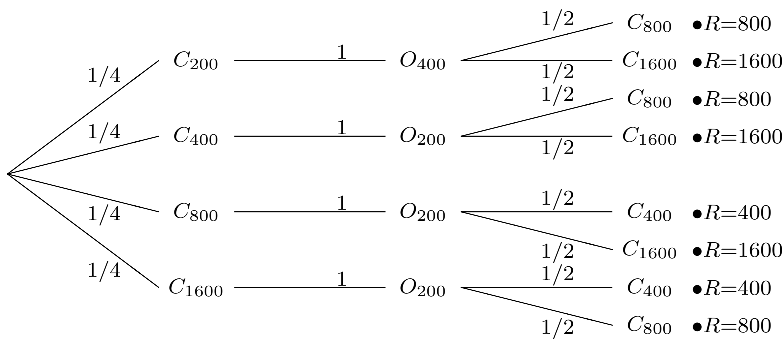
The PMF of R is just

$$P_R(r) = \begin{cases} 1/4 & r = 200, 400, 800, 1600 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of R is

$$\mathrm{E}[R] = \sum_r r P_R(r) = \frac{1}{4}(200 + 400 + 800 + 1600) = 750. \tag{2}$$

(b) In this case, the tree diagram is



All eight outcomes are equally likely and R as PMF

r	400	800	1600
$P_R(r)$	1/4	3/8	3/8

The expected value of R is

$$\mathrm{E}[R] = \frac{1}{4}400 + \frac{3}{8}800 + \frac{3}{8}1600 = 1000. \tag{3}$$

Problem 3.5.19 Solution

By the definition of the expected value,

$$\begin{aligned} \mathrm{E}[X_n] &= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} && (1) \\ &= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-1-(x-1))!} p^{x-1}(1-p)^{n-1-(x-1)}. && (2) \end{aligned}$$

With the substitution $x' = x - 1$, we have

$$\mathrm{E}\left[X_n\right] = np \underbrace{\sum_{x'=0}^{n-1} \binom{n-1}{x'} p^{x'} (1-p)^{n-x'}}_{1} = np \sum_{x'=0}^{n-1} P_{X_{n-1}}(x) = np. \qquad (3)$$

The above sum is 1 because it is the sum of a binomial random variable for $n-1$ trials over all possible values.

Problem 3.5.20 Solution

We write the sum as a double sum in the following way:

$$\sum_{i=0}^{\infty} \mathrm{P}\left[X > i\right] = \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} P_X(j). \qquad (1)$$

At this point, the key step is to reverse the order of summation. You may need to make a sketch of the feasible values for i and j to see how this reversal occurs. In this case,

$$\sum_{i=0}^{\infty} \mathrm{P}\left[X > i\right] = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} P_X(j) = \sum_{j=1}^{\infty} j P_X(j) = \mathrm{E}\left[X\right]. \qquad (2)$$

Problem 3.5.21 Solution

(a) A geometric (p) random variable has expected value $1/p$. Since R is a geometric random variable with $\mathrm{E}[R] = 100/m$, we can conclude that R is a geometric ($p = m/100$) random variable. Thus the PMF of R is

$$P_R(r) = \begin{cases} (1 - m/100)^{r-1}(m/100) & r = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

(b) From the problem statement, $\mathrm{E}[R] = 100/m$. Thus

$$\mathrm{E}\left[W\right] = \mathrm{E}\left[5mR\right] = 5m\,\mathrm{E}\left[R\right] = 5m\frac{100}{m} = 500. \qquad (2)$$

(c) She wins the money if she does work $W \geq 1000$, which has probability

$$\mathrm{P}\left[W \geq 1000\right] = \mathrm{P}\left[5mR \geq 1000\right] = \mathrm{P}\left[R \geq \frac{200}{m}\right]. \tag{3}$$

Note that for a geometric (p) random variable X and an integer x_0,

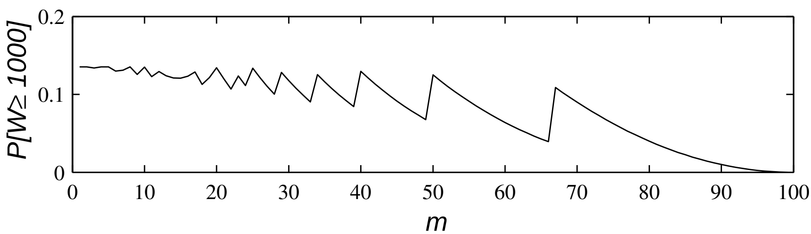
$$\begin{aligned}\mathrm{P}\left[X \geq x_0\right] &= \sum_{x=x_0}^{\infty} P_X(x) \\ &= (1-p)^{x_0-1} p \left(1 + (1-p) + (1-p)^2 + \cdots\right) \\ &= (1-p)^{x_0-1}. \end{aligned} \tag{4}$$

Thus for the geometric $(p = m/100)$ random variable R,

$$\mathrm{P}\left[R \geq \frac{200}{m}\right] = \mathrm{P}\left[R \geq \left\lceil \frac{200}{m} \right\rceil\right] = \left(1 - \frac{m}{100}\right)^{\left\lceil \frac{200}{m} \right\rceil - 1}, \tag{5}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

As a function of m, the probability of winning is an odd sawtooth function that has a peak each time $200/m$ is close to an integer. Here is a plot if you're curious:



It does happen to be true in this case that $\mathrm{P}[W \geq 1000]$ is maximized at $m = 1$. For $m = 1$,

$$\mathrm{P}\left[W \geq 1000\right] = \mathrm{P}\left[R \geq 200\right] = (0.99)^{199} = 0.1353. \tag{6}$$

Problem 3.5.22 Solution

(a) A geometric (p) random variable has expected value $1/p$. Since R is a geometric random variable with $E[R] = 200/m$, we can conclude that R is a geometric ($p = m/200$) random variable. Thus the PMF of R is

$$P_R(r) = \begin{cases} (1 - m/200)^{r-1}(m/200) & r = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) From the problem statement, $E[R] = 200/m$. Thus

$$E[W] = E[4mR] = 4m\,E[R] = 4m\frac{200}{m} = 800. \tag{2}$$

(c) She wins the money if she does work $W \geq 1000$, which has probability

$$P[W \geq 1000] = P[4mR \geq 1000] = P\left[R \geq \frac{250}{m}\right]. \tag{3}$$

Note that for a geometric (p) random variable X and an integer x_0,

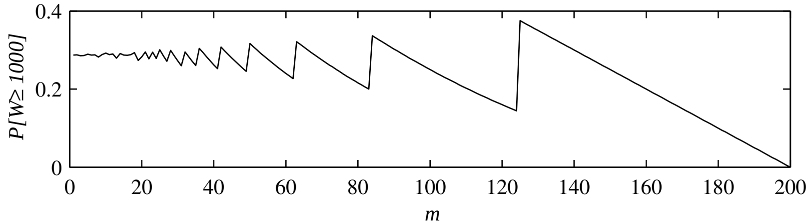
$$\begin{aligned} P[X \geq x_0] &= \sum_{x=x_0}^{\infty} P_X(x) \\ &= (1-p)^{x_0-1} p \left(1 + (1-p) + (1-p)^2 + \cdots\right) \\ &= (1-p)^{x_0-1}. \end{aligned} \tag{4}$$

Thus for the geometric ($p = m/200$) random variable R,

$$P\left[R \geq \frac{250}{m}\right] = P\left[R \geq \left\lceil \frac{250}{m} \right\rceil\right] = \left(1 - \frac{m}{200}\right)^{\left\lceil \frac{250}{m} \right\rceil - 1}. \tag{5}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

As for finding the optimal value of m, the probability of winning is a very odd sawtooth function that has a peak each time $250/m$ and whose highest peak is at $m = 125$. (This isn't obvious, unless you exercised the power of MATLAB or your graphing calculator.) Here is a plot if you're curious:



In this case, at $m = 125$,

$$\mathrm{P}\left[W \geq 1000\right] = \mathrm{P}\left[R \geq \frac{250}{125}\right] = \left(1 - \frac{125}{200}\right)^{\left\lceil \frac{250}{125} \right\rceil - 1} = \left(\frac{3}{8}\right). \tag{6}$$

Problem 3.6.1 Solution

From the solution to Problem 3.4.1, the PMF of Y is

$$P_Y(y) = \begin{cases} 1/4 & y = 1, \\ 1/4 & y = 2, \\ 1/2 & y = 3, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

(a) Since Y has range $S_Y = \{1, 2, 3\}$, the range of $U = Y^2$ is $S_U = \{1, 4, 9\}$. The PMF of U can be found by observing that

$$\mathrm{P}\left[U = u\right] = \mathrm{P}\left[Y^2 = u\right] = \mathrm{P}\left[Y = \sqrt{u}\right] + \mathrm{P}\left[Y = -\sqrt{u}\right]. \tag{2}$$

Since Y is never negative, $P_U(u) = P_Y(\sqrt{u})$. Hence,

$$P_U(1) = P_Y(1) = 1/4, \tag{3}$$
$$P_U(4) = P_Y(2) = 1/4, \tag{4}$$
$$P_U(9) = P_Y(3) = 1/2. \tag{5}$$

For all other values of u, $P_U(u) = 0$. The complete expression for the PMF of U is

$$P_U(u) = \begin{cases} 1/4 & u = 1, \\ 1/4 & u = 4, \\ 1/2 & u = 9, \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

(b) From the PMF, it is straightforward to write down the CDF

$$F_U(u) = \begin{cases} 0 & u < 1, \\ 1/4 & 1 \leq u < 4, \\ 1/2 & 4 \leq u < 9, \\ 1 & u \geq 9. \end{cases} \tag{7}$$

(c) From Definition 3.13, the expected value of U is

$$\mathrm{E}[U] = \sum_u u P_U(u) = 1(1/4) + 4(1/4) + 9(1/2) = 5.75. \tag{8}$$

From Theorem 3.10, we can calculate the expected value of U as

$$\begin{aligned} \mathrm{E}[U] = \mathrm{E}\left[Y^2\right] &= \sum_y y^2 P_Y(y) \\ &= 1^2(1/4) + 2^2(1/4) + 3^2(1/2) = 5.75. \end{aligned} \tag{9}$$

As we expect, both methods yield the same answer.

Problem 3.6.2 Solution

From the solution to Problem 3.4.2, the PMF of X is

$$P_X(x) = \begin{cases} 0.2 & x = -1, \\ 0.5 & x = 0, \\ 0.3 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The PMF of $V = |X|$ satisfies

$$P_V(v) = \mathrm{P}[|X| = v] = P_X(v) + P_X(-v). \tag{2}$$

In particular,

$$P_V(0) = P_X(0) = 0.5, \tag{3}$$
$$P_V(1) = P_X(-1) + P_X(1) = 0.5. \tag{4}$$

The complete expression for the PMF of V is

$$P_V(v) = \begin{cases} 0.5 & v = 0, \\ 0.5 & v = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

(b) From the PMF, we can construct the staircase CDF of V.

$$F_V(v) = \begin{cases} 0 & v < 0, \\ 0.5 & 0 \leq v < 1, \\ 1 & v \geq 1. \end{cases} \tag{6}$$

(c) From the PMF $P_V(v)$, the expected value of V is

$$\mathrm{E}[V] = \sum_v P_V(v) = 0(1/2) + 1(1/2) = 1/2. \tag{7}$$

You can also compute $\mathrm{E}[V]$ directly by using Theorem 3.10.

Problem 3.6.3 Solution

From the solution to Problem 3.4.3, the PMF of X is

$$P_X(x) = \begin{cases} 0.4 & x = -3, \\ 0.4 & x = 5, \\ 0.2 & x = 7, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The PMF of $W = -X$ satisfies

$$P_W(w) = \mathrm{P}[-X = w] = P_X(-w). \tag{2}$$

This implies

$$P_W(-7) = P_X(7) = 0.2 \tag{3}$$

$$P_W(-5) = P_X(5) = 0.4 \tag{4}$$

$$P_W(3) = P_X(-3) = 0.4. \tag{5}$$

The complete PMF for W is

$$P_W(w) = \begin{cases} 0.2 & w = -7, \\ 0.4 & w = -5, \\ 0.4 & w = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

(b) From the PMF, the CDF of W is

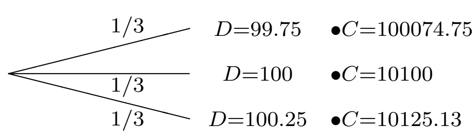
$$F_W(w) = \begin{cases} 0 & w < -7, \\ 0.2 & -7 \leq w < -5, \\ 0.6 & -5 \leq w < 3, \\ 1 & w \geq 3. \end{cases} \tag{7}$$

(c) From the PMF, W has expected value

$$\mathrm{E}[W] = \sum_w w P_W(w) = -7(0.2) + -5(0.4) + 3(0.4) = -2.2. \tag{8}$$

Problem 3.6.4 Solution

A tree for the experiment is



Thus C has three equally likely outcomes. The PMF of C is

$$P_C(c) = \begin{cases} 1/3 & c = 100{,}074.75, 10{,}100, 10{,}125.13, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 3.6.5 Solution

(a) The source continues to transmit packets until one is received correctly. Hence, the total number of times that a packet is transmitted is $X = x$ if the first $x-1$ transmissions were in error. Therefore the PMF of X is

$$P_X(x) = \begin{cases} q^{x-1}(1-q) & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The time required to send a packet is a millisecond and the time required to send an acknowledgment back to the source takes another millisecond. Thus, if X transmissions of a packet are needed to send the packet correctly, then the packet is correctly received after $T = 2X - 1$ milliseconds. Therefore, for an odd integer $t > 0$, $T = t$ iff $X = (t+1)/2$. Thus,

$$P_T(t) = P_X((t+1)/2) = \begin{cases} q^{(t-1)/2}(1-q) & t = 1, 3, 5, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 3.6.6 Solution

The cellular calling plan charges a flat rate of \$20 per month up to and including the 30th minute, and an additional 50 cents for each minute over 30 minutes. Knowing that the time you spend on the phone is a geometric random variable M with mean $1/p = 30$, the PMF of M is

$$P_M(m) = \begin{cases} (1-p)^{m-1}p & m = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The monthly cost, C obeys

$$P_C(20) = \mathrm{P}[M \leq 30] = \sum_{m=1}^{30} (1-p)^{m-1}p = 1 - (1-p)^{30}. \tag{2}$$

When $M \geq 30$, $C = 20 + (M-30)/2$ or $M = 2C - 10$. Thus,

$$P_C(c) = P_M(2c - 10) \qquad c = 20.5, 21, 21.5, \ldots. \tag{3}$$

The complete PMF of C is

$$P_C(c) = \begin{cases} 1-(1-p)^{30} & c = 20, \\ (1-p)^{2c-10-1}p & c = 20.5, 21, 21.5, \ldots. \end{cases} \tag{4}$$

Problem 3.6.7 Solution

(a) A student is properly counted with probability p, independent of any other student being counted. Hence, we have 70 Bernoulli trials and N is a binomial $(70, p)$ random variable with PMF

$$P_N(n) = \binom{70}{n} p^n (1-p)^{70-n}. \tag{1}$$

(b) There are two ways to find this. The first way is to observe that

$$\begin{aligned} \mathrm{P}[U = u] = \mathrm{P}[N = 70 - u] &= P_N(70 - u) \\ &= \binom{70}{70-u} p^{70-u}(1-p)^{70-(70-u)} \\ &= \binom{70}{u}(1-p)^u p^{70-u}. \end{aligned} \tag{2}$$

We see that U is a binomial $(70, 1-p)$. The second way is to argue this directly since U is counting overlooked students. If we call an overlooked student a "success" with probability $1-p$, then U, the number of successes in n trials, is binomial $(70, 1-p)$.

(c)

$$\begin{aligned} \mathrm{P}[U \geq 2] &= 1 - \mathrm{P}[U < 2] \\ &= 1 - (P_U(0) + P_U(1)) \\ &= 1 - (p^{70} + 70(1-p)p^{69}). \end{aligned} \tag{3}$$

(d) The binomial $(n = 70, 1-p)$ random variable U has $\mathrm{E}[U] = 70(1-p)$. Solving $70(1-p) = 2$ yields $p = 34/35$.

Problem 3.6.8 Solution

(a) Each M&M is double counted with probability p. Since there are 20 M&Ms, the number of double-counted M&Ms, R, has the binomial PMF

$$P_R(r) = \binom{20}{r} p^r (1-p)^{20-r}. \tag{1}$$

(b) The professor counts each M&M either once or twice so that $N = 20 + R$. This implies

$$\begin{aligned} P_N(n) = \mathrm{P}[20 + R = n] &= P_R(n-20) \\ &= \binom{20}{n-20} p^{n-20} (1-p)^{40-n}. \end{aligned} \tag{2}$$

Problem 3.7.1 Solution

Let W_n equal the number of winning tickets you purchase in n days. Since each day is an independent trial, W_n is a binomial $(n, p = 0.1)$ random variable. Since each ticket costs 1 dollar, you spend n dollars on tickets in n days. Since each winning ticket is cashed for 5 dollars, your profit after n days is

$$X_n = 5W_n - n. \tag{1}$$

It follows that

$$\mathrm{E}[X_n] = 5\,\mathrm{E}[W_n] - n = 5np - n = (5p-1)n = -n/2. \tag{2}$$

On average, you lose about 50 cents per day.

Problem 3.7.2 Solution

From the solution to Quiz 3.6, we found that $T = 120 - 15N$. By Theorem 3.10,

$$\begin{aligned} \mathrm{E}[T] &= \sum_{n \in S_N} (120 - 15n) P_N(n) && (1) \\ &= 0.1(120) + 0.3(120-15) + 0.3(120-30) + 0.3(120-45) = 93. && (2) \end{aligned}$$

Also from the solution to Quiz 3.6, we found that

$$P_T(t) = \begin{cases} 0.3 & t = 75, 90, 105, \\ 0.1 & t = 120, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Using Definition 3.13,

$$\mathrm{E}[T] = \sum_{t \in S_T} t P_T(t) = 0.3(75) + 0.3(90) + 0.3(105) + 0.1(120) = 93. \tag{4}$$

As expected, the two calculations give the same answer.

Problem 3.7.3 Solution

Whether a lottery ticket is a winner is a Bernoulli trial with a success probability of 0.001. If we buy one every day for 50 years for a total of $50 \cdot 365 = 18250$ tickets, then the number of winning tickets T is a binomial random variable with mean

$$\mathrm{E}[T] = 18250(0.001) = 18.25. \tag{1}$$

Since each winning ticket grosses \$1000, the revenue we collect over 50 years is $R = 1000T$ dollars. The expected revenue is

$$\mathrm{E}[R] = 1000\,\mathrm{E}[T] = 18250. \tag{2}$$

But buying a lottery ticket everyday for 50 years, at \$2.00 a pop isn't cheap and will cost us a total of $18250 \cdot 2 = \$36500$. Our net profit is then $Q = R - 36500$ and the result of our loyal 50 year patronage of the lottery system, is disappointing expected loss of

$$\mathrm{E}[Q] = \mathrm{E}[R] - 36500 = -18250. \tag{3}$$

Problem 3.7.4 Solution

Let X denote the number of points the shooter scores. If the shot is uncontested, the expected number of points scored is

$$\mathrm{E}[X] = (0.6)2 = 1.2. \tag{1}$$

If we foul the shooter, then X is a binomial random variable with mean $\mathrm{E}[X] = 2p$. If $2p > 1.2$, then we should not foul the shooter. Generally, p will exceed 0.6 since a free throw is usually even easier than an uncontested shot taken during the action of the game. Furthermore, fouling the shooter ultimately leads to the the detriment of players possibly fouling out. This suggests that fouling a player is not a good idea. The only real exception occurs when facing a player like Shaquille O'Neal whose free throw probability p is lower than his field goal percentage during a game.

Problem 3.7.5 Solution

Given the distributions of D, the waiting time in days and the resulting cost, C, we can answer the following questions.

(a) The expected waiting time is simply the expected value of D.

$$\mathrm{E}[D] = \sum_{d=1}^{4} d \cdot P_D(d) = 1(0.2) + 2(0.4) + 3(0.3) + 4(0.1) = 2.3. \qquad (1)$$

(b) The expected deviation from the waiting time is

$$\mathrm{E}[D - \mu_D] = \mathrm{E}[D] - \mathrm{E}[\mu_d] = \mu_D - \mu_D = 0. \qquad (2)$$

(c) C can be expressed as a function of D in the following manner.

$$C(D) = \begin{cases} 90 & D = 1, \\ 70 & D = 2, \\ 40 & D = 3, \\ 40 & D = 4. \end{cases} \qquad (3)$$

(d) The expected service charge is

$$\mathrm{E}[C] = 90(0.2) + 70(0.4) + 40(0.3) + 40(0.1) = 62 \text{ dollars.} \qquad (4)$$

Problem 3.7.6 Solution

FALSE: For a counterexample, suppose

$$P_X(x) = \begin{cases} 0.5 & x = 0.1, 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

In this case, $\mathrm{E}[X] = 0.5(0.1 + 0.5) = 0.3$ so that $1/\mathrm{E}[X] = 10/3$. On the other hand,

$$\mathrm{E}[1/X] = 0.5\left(\frac{1}{0.1} + \frac{1}{0.5}\right) = 0.5(10 + 2) = 6. \tag{2}$$

Problem 3.7.7 Solution

As a function of the number of minutes used, M, the monthly cost is

$$C(M) = \begin{cases} 20 & M \le 30 \\ 20 + (M - 30)/2 & M \ge 30 \end{cases} \tag{1}$$

The expected cost per month is

$$\begin{aligned} \mathrm{E}[C] &= \sum_{m=1}^{\infty} C(m) P_M(m) \\ &= \sum_{m=1}^{30} 20 P_M(m) + \sum_{m=31}^{\infty} (20 + (m - 30)/2) P_M(m) \\ &= 20 \sum_{m=1}^{\infty} P_M(m) + \frac{1}{2} \sum_{m=31}^{\infty} (m - 30) P_M(m). \end{aligned} \tag{2}$$

Since $\sum_{m=1}^{\infty} P_M(m) = 1$ and since $P_M(m) = (1-p)^{m-1} p$ for $m \ge 1$, we have

$$\mathrm{E}[C] = 20 + \frac{(1-p)^{30}}{2} \sum_{m=31}^{\infty} (m - 30)(1-p)^{m-31} p. \tag{3}$$

Making the substitution $j = m - 30$ yields

$$\mathrm{E}[C] = 20 + \frac{(1-p)^{30}}{2} \sum_{j=1}^{\infty} j(1-p)^{j-1} p = 20 + \frac{(1-p)^{30}}{2p}. \tag{4}$$

Problem 3.7.8 Solution

Since our phone use is a geometric random variable M with mean value $1/p$,

$$P_M(m) = \begin{cases} (1-p)^{m-1}p & m = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For this cellular billing plan, we are given no free minutes, but are charged half the flat fee. That is, we are going to pay 15 dollars regardless and \$1 for each minute we use the phone. Hence $C = 15 + M$ and for $c \geq 16$, $P[C = c] = P[M = c - 15]$. Thus we can construct the PMF of the cost C

$$P_C(c) = \begin{cases} (1-p)^{c-16}p & c = 16, 17, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Since $C = 15 + M$, the expected cost per month of the plan is

$$E[C] = E[15 + M] = 15 + E[M] = 15 + 1/p. \tag{3}$$

In Problem 3.7.7, we found that that the expected cost of the plan was

$$E[C] = 20 + [(1-p)^{30}]/(2p). \tag{4}$$

In comparing the expected costs of the two plans, we see that the new plan is better (i.e. cheaper) if

$$15 + 1/p \leq 20 + [(1-p)^{30}]/(2p). \tag{5}$$

A simple plot will show that the new plan is better if $p \leq p_0 \approx 0.2$.

Problem 3.7.9 Solution

We consider the cases of using standard devices or ultra reliable devices separately. In both cases, the methodology is the same. We define random variable W such that $W = 1$ if the circuit works or $W = 0$ if the circuit is defective. (In the probability literature, W is called an indicator random variable.) The PMF of W depends on whether the circuit uses standard devices or ultra reliable devices. We then define the revenue as a function $R(W)$ and we evaluate $E[R(W)]$.

The circuit with standard devices works with probability $(1-q)^{10}$ and generates revenue of k dollars if all of its 10 constituent devices work. In this case, $W = W_s$ has PMF

$$P_{W_s}(w) = \begin{cases} 1-(1-q)^{10} & w=0, \\ (1-q)^{10} & w=1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In addition, let R_s denote the profit on a circuit with standard devices. We observe that we can we can express R_s as a function $r_s(W_s)$:

$$R_s = r_s(W_s) = \begin{cases} -10 & W_s = 0, \\ k-10 & W_s = 1. \end{cases} \tag{2}$$

Thus we can express the expected profit as

$$\begin{aligned} \mathrm{E}[R_s] &= \mathrm{E}[r_s(W)] \\ &= \sum_{w=0}^{1} P_{W_s}(w)\, r_s(w) \\ &= P_{W_s}(0)\,(-10) + P_{W_s}(1)\,(k-10) \\ &= (1-(1-q)^{10})(-10) + (1-q)^{10}(k-10) = (0.9)^{10}k - 10. \end{aligned} \tag{3}$$

To examine the circuit with ultra reliable devices, let $W = W_u$ indicate whether the circuit works and let $R_u = r_u(W_u)$ denote the profit on a circuit with ultrareliable devices. W_u has PMF

$$P_{W_u}(w) = \begin{cases} 1-(1-q/2)^{10} & w=0, \\ (1-q/2)^{10} & w=1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The revenue function is

$$R_u = r_u(W_u) = \begin{cases} -30 & W_u = 0, \\ k-30 & W_u = 1. \end{cases} \tag{5}$$

Thus we can express the expected profit as

$$\begin{aligned}
\mathrm{E}\left[R_u\right] &= \mathrm{E}\left[r_u(W_u)\right] \\
&= \sum_{w=0}^{1} P_{W_u}(w)\, r_u(w) \\
&= P_{W_u}(0)\,(-30) + P_{W_u}(1)\,(k-30) \\
&= (1-(1-q/2)^{10})(-30) + (1-q/2)^{10}(k-30) \\
&= (0.95)^{10}k - 30. \qquad (6)
\end{aligned}$$

Now we can compare $\mathrm{E}[R_s]$ and $\mathrm{E}[R_u]$ to decide which circuit implementation offers the highest expected profit. The inequality $\mathrm{E}[R_u] \geq \mathrm{E}[R_s]$, holds if and only if

$$k \geq 20/[(0.95)^{10} - (0.9)^{10}] = 80.21. \qquad (7)$$

So for $k < \$80.21$ using all standard devices results in greater revenue, while for $k > \$80.21$ more revenue will be generated by implementing the circuit with all ultra-reliable devices. That is, when the price commanded for a working circuit is sufficiently high, we should build more-expensive higher-reliability circuits.

If you have read ahead to Section 7.1 and learned about conditional expected values, you might prefer the following solution. If not, you might want to come back and review this alternate approach after reading Section 7.1.

Let W denote the event that a circuit works. The circuit works and generates revenue of k dollars if all of its 10 constituent devices work. For each implementation, standard or ultra-reliable, let R denote the profit on a device. We can express the expected profit as

$$\mathrm{E}\left[R\right] = \mathrm{P}\left[W\right]\mathrm{E}\left[R|W\right] + \mathrm{P}\left[W^c\right]\mathrm{E}\left[R|W^c\right]. \qquad (8)$$

Let's first consider the case when only standard devices are used. In this case, a circuit works with probability $\mathrm{P}[W] = (1-q)^{10}$. The profit made on a working device is $k-10$ dollars while a nonworking circuit has a profit of -10 dollars. That is, $\mathrm{E}[R|W] = k-10$ and $\mathrm{E}[R|W^c] = -10$. Of course, a negative profit is actually a loss. Using R_s to denote the profit using standard circuits, the expected profit is

$$\begin{aligned}
\mathrm{E}\left[R_s\right] &= (1-q)^{10}(k-10) + (1-(1-q)^{10})(-10) \\
&= (0.9)^{10}k - 10. \qquad (9)
\end{aligned}$$

And for the ultra-reliable case, the circuit works with probability $P[W] = (1 - q/2)^{10}$. The profit per working circuit is $E[R|W] = k - 30$ dollars while the profit for a nonworking circuit is $E[R|W^c] = -30$ dollars. The expected profit is

$$\begin{aligned} E[R_u] &= (1 - q/2)^{10}(k - 30) + (1 - (1 - q/2)^{10})(-30) \\ &= (0.95)^{10}k - 30. \end{aligned} \tag{10}$$

Not surprisingly, we get the same answers for $E[R_u]$ and $E[R_s]$ as in the first solution by performing essentially the same calculations. it should be apparent that indicator random variable W in the first solution indicates the occurrence of the conditioning event W in the second solution. That is, indicators are a way to track conditioning events.

Problem 3.7.10 Solution

(a) There are $\binom{46}{6}$ equally likely winning combinations so that

$$q = \frac{1}{\binom{46}{6}} = \frac{1}{9{,}366{,}819} \approx 1.07 \times 10^{-7}. \tag{1}$$

(b) Assuming each ticket is chosen randomly, each of the $2n - 1$ other tickets is independently a winner with probability q. The number of other winning tickets K_n has the binomial PMF

$$P_{K_n}(k) = \begin{cases} \binom{2n-1}{k} q^k (1 - q)^{2n-1-k} & k = 0, 1, \ldots, 2n - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) Since there are $K_n + 1$ winning tickets in all, the value of your winning ticket is $W_n = n/(K_n + 1)$ which has mean

$$E[W_n] = n\,E\left[\frac{1}{K_n + 1}\right]. \tag{3}$$

Calculating the expected value

$$E\left[\frac{1}{K_n + 1}\right] = \sum_{k=0}^{2n-1} \left(\frac{1}{k + 1}\right) P_{K_n}(k) \tag{4}$$

is fairly complicated. The trick is to express the sum in terms of the sum of a binomial PMF.

$$\begin{aligned} \mathrm{E}\left[\frac{1}{K_n+1}\right] &= \sum_{k=0}^{2n-1} \frac{1}{k+1} \frac{(2n-1)!}{k!(2n-1-k)!} q^k (1-q)^{2n-1-k} \\ &= \frac{1}{2n} \sum_{k=0}^{2n-1} \frac{(2n)!}{(k+1)!(2n-(k+1))!} q^k (1-q)^{2n-(k+1)}. \end{aligned} \tag{5}$$

By factoring out $1/q$, we obtain

$$\begin{aligned} \mathrm{E}\left[\frac{1}{K_n+1}\right] &= \frac{1}{2nq} \sum_{k=0}^{2n-1} \binom{2n}{k+1} q^{k+1} (1-q)^{2n-(k+1)} \\ &= \frac{1}{2nq} \underbrace{\sum_{j=1}^{2n} \binom{2n}{j} q^j (1-q)^{2n-j}}_{A}. \end{aligned} \tag{6}$$

We observe that the above sum labeled A is the sum of a binomial PMF for $2n$ trials and success probability q over all possible values except $j=0$. Thus

$$A = 1 - \binom{2n}{0} q^0 (1-q)^{2n-0} = 1-(1-q)^{2n}. \tag{7}$$

This implies

$$\mathrm{E}\left[\frac{1}{K_n+1}\right] = \frac{1-(1-q)^{2n}}{2nq}. \tag{8}$$

Our expected return on a winning ticket is

$$\mathrm{E}\left[W_n\right] = n\,\mathrm{E}\left[\frac{1}{K_n+1}\right] = \frac{1-(1-q)^{2n}}{2q}. \tag{9}$$

Note that when $nq \ll 1$, we can use the approximation that $(1-q)^{2n} \approx 1-2nq$ to show that

$$\mathrm{E}\left[W_n\right] \approx \frac{1-(1-2nq)}{2q} = n, \qquad nq \ll 1. \tag{10}$$

However, in the limit as the value of the prize n approaches infinity, we have

$$\lim_{n\to\infty} \mathrm{E}\left[W_n\right] = \frac{1}{2q} \approx 4.683 \times 10^6. \tag{11}$$

That is, as the pot grows to infinity, the expected return on a winning ticket doesn't approach infinity because there is a corresponding increase in the number of other winning tickets. If it's not clear how large n must be for this effect to be seen, consider the following table:

n	10^6	10^7	10^8
$\mathrm{E}\left[W_n\right]$	9.00×10^5	4.13×10^6	4.68×10^6

(12)

When the pot is \$1 million, our expected return is \$900,000. However, we see that when the pot reaches \$100 million, our expected return is very close to $1/(2q)$, less than \$5 million!

Problem 3.7.11 Solution

(a) There are $\binom{46}{6}$ equally likely winning combinations so that

$$q = \frac{1}{\binom{46}{6}} = \frac{1}{9{,}366{,}819} \approx 1.07 \times 10^{-7}. \tag{1}$$

(b) Assuming each ticket is chosen randomly, each of the $2n-1$ other tickets is independently a winner with probability q. The number of other winning tickets K_n has the binomial PMF

$$P_{K_n}(k) = \begin{cases} \binom{2n-1}{k} q^k (1-q)^{2n-1-k} & k = 0, 1, \ldots, 2n-1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Since the pot has $n + r$ dollars, the expected amount that you win on your ticket is

$$\mathrm{E}\left[V\right] = 0(1-q) + q\,\mathrm{E}\left[\frac{n+r}{K_n+1}\right] = q(n+r)\,\mathrm{E}\left[\frac{1}{K_n+1}\right]. \tag{3}$$

Note that $\mathrm{E}[1/K_n + 1]$ was also evaluated in Problem 3.7.10. For completeness, we repeat those steps here.

$$\mathrm{E}\left[\frac{1}{K_n+1}\right] = \sum_{k=0}^{2n-1} \frac{1}{k+1} \frac{(2n-1)!}{k!(2n-1-k)!} q^k (1-q)^{2n-1-k}$$
$$= \frac{1}{2n} \sum_{k=0}^{2n-1} \frac{(2n)!}{(k+1)!(2n-(k+1))!} q^k (1-q)^{2n-(k+1)}. \tag{4}$$

By factoring out $1/q$, we obtain

$$\mathrm{E}\left[\frac{1}{K_n+1}\right] = \frac{1}{2nq} \sum_{k=0}^{2n-1} \binom{2n}{k+1} q^{k+1} (1-q)^{2n-(k+1)}$$
$$= \frac{1}{2nq} \underbrace{\sum_{j=1}^{2n} \binom{2n}{j} q^j (1-q)^{2n-j}}_{A}. \tag{5}$$

We observe that the above sum labeled A is the sum of a binomial PMF for $2n$ trials and success probability q over all possible values except $j = 0$. Thus $A = 1 - \binom{2n}{0} q^0 (1-q)^{2n-0}$, which implies

$$\mathrm{E}\left[\frac{1}{K_n+1}\right] = \frac{A}{2nq} = \frac{1-(1-q)^{2n}}{2nq}. \tag{6}$$

The expected value of your ticket is

$$\mathrm{E}[V] = \frac{q(n+r)[1-(1-q)^{2n}]}{2nq}$$
$$= \frac{1}{2}\left(1 + \frac{r}{n}\right)[1-(1-q)^{2n}]. \tag{7}$$

Each ticket tends to be more valuable when the carryover pot r is large and the number of new tickets sold, $2n$, is small. For any fixed number n, corresponding to $2n$ tickets sold, a sufficiently large pot r will guarantee that $\mathrm{E}[V] > 1$. For example if $n = 10^7$, (20 million tickets sold) then

$$\mathrm{E}[V] = 0.44\left(1 + \frac{r}{10^7}\right). \tag{8}$$

If the carryover pot r is 30 million dollars, then $\mathrm{E}[V] = 1.76$. This suggests that buying a one dollar ticket is a good idea. This is an unusual situation because normally a carryover pot of 30 million dollars will result in far more than 20 million tickets being sold.

(c) So that we can use the results of the previous part, suppose there were $2n-1$ tickets sold before you must make your decision. If you buy one of each possible ticket, you are guaranteed to have one winning ticket. From the other $2n-1$ tickets, there will be K_n winners. The total number of winning tickets will be K_n+1. In the previous part we found that

$$\mathrm{E}\left[\frac{1}{K_n+1}\right] = \frac{1-(1-q)^{2n}}{2nq}. \tag{9}$$

Let R denote the expected return from buying one of each possible ticket. The pot had r dollars beforehand. The $2n-1$ other tickets are sold add $n-1/2$ dollars to the pot. Furthermore, you must buy $1/q$ tickets, adding $1/(2q)$ dollars to the pot. Since the cost of the tickets is $1/q$ dollars, your expected profit

$$\begin{aligned}\mathrm{E}\left[R\right] &= \mathrm{E}\left[\frac{r+n-1/2+1/(2q)}{K_n+1}\right] - \frac{1}{q} \\ &= \frac{q(2r+2n-1)+1}{2q}\,\mathrm{E}\left[\frac{1}{K_n+1}\right] - \frac{1}{q} \\ &= \frac{[q(2r+2n-1)+1](1-(1-q)^{2n})}{4nq^2} - \frac{1}{q}. \end{aligned} \tag{10}$$

For fixed n, sufficiently large r will make $\mathrm{E}[R] > 0$. On the other hand, for fixed r, $\lim_{n\to\infty} \mathrm{E}[R] = -1/(2q)$. That is, as n approaches infinity, your expected loss will be quite large.

Problem 3.8.1 Solution

Given the following PMF

$$P_N(n) = \begin{cases} 0.2 & n = 0, \\ 0.7 & n = 1, \\ 0.1 & n = 2, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

the calculations are straightforward:

(a) $E[N] = (0.2)0 + (0.7)1 + (0.1)2 = 0.9$.

(b) $E[N^2] = (0.2)0^2 + (0.7)1^2 + (0.1)2^2 = 1.1$.

(c) $\text{Var}[N] = E[N^2] - E[N]^2 = 1.1 - (0.9)^2 = 0.29$.

(d) $\sigma_N = \sqrt{\text{Var}[N]} = \sqrt{0.29}$.

Problem 3.8.2 Solution

From the solution to Problem 3.4.1, the PMF of Y is

$$P_Y(y) = \begin{cases} 1/4 & y = 1, \\ 1/4 & y = 2, \\ 1/2 & y = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of Y is

$$E[Y] = \sum_y y P_Y(y) = 1(1/4) + 2(1/4) + 3(1/2) = 9/4. \tag{2}$$

The expected value of Y^2 is

$$E\left[Y^2\right] = \sum_y y^2 P_Y(y) = 1^2(1/4) + 2^2(1/4) + 3^2(1/2) = 23/4. \tag{3}$$

The variance of Y is

$$\text{Var}[Y] = E\left[Y^2\right] - (E[Y])^2 = 23/4 - (9/4)^2 = 11/16. \tag{4}$$

Problem 3.8.3 Solution

From the solution to Problem 3.4.2, the PMF of X is

$$P_X(x) = \begin{cases} 0.2 & x = -1, \\ 0.5 & x = 0, \\ 0.3 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of X is

$$\mathrm{E}[X] = \sum_x x P_X(x) = (-1)(0.2) + 0(0.5) + 1(0.3) = 0.1. \tag{2}$$

The expected value of X^2 is

$$\mathrm{E}\left[X^2\right] = \sum_x x^2 P_X(x) = (-1)^2(0.2) + 0^2(0.5) + 1^2(0.3) = 0.5. \tag{3}$$

The variance of X is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 0.5 - (0.1)^2 = 0.49. \tag{4}$$

Problem 3.8.4 Solution

From the solution to Problem 3.4.3, the PMF of X is

$$P_X(x) = \begin{cases} 0.4 & x = -3, \\ 0.4 & x = 5, \\ 0.2 & x = 7, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of X is

$$\mathrm{E}[X] = \sum_x x P_X(x) = -3(0.4) + 5(0.4) + 7(0.2) = 2.2. \tag{2}$$

The expected value of X^2 is

$$\mathrm{E}\left[X^2\right] = \sum_x x^2 P_X(x) = (-3)^2(0.4) + 5^2(0.4) + 7^2(0.2) = 23.4. \tag{3}$$

The variance of X is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 23.4 - (2.2)^2 = 18.56. \tag{4}$$

Problem 3.8.5 Solution

(a) The expected value of X is

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x=0}^{4} x P_X(x) \\
&= 0\binom{4}{0}\frac{1}{2^4} + 1\binom{4}{1}\frac{1}{2^4} + 2\binom{4}{2}\frac{1}{2^4} + 3\binom{4}{3}\frac{1}{2^4} + 4\binom{4}{4}\frac{1}{2^4} \\
&= [4+12+12+4]/2^4 = 2. \qquad (1)
\end{aligned}$$

The expected value of X^2 is

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= \sum_{x=0}^{4} x^2 P_X(x) \\
&= 0^2\binom{4}{0}\frac{1}{2^4} + 1^2\binom{4}{1}\frac{1}{2^4} + 2^2\binom{4}{2}\frac{1}{2^4} + 3^2\binom{4}{3}\frac{1}{2^4} + 4^2\binom{4}{4}\frac{1}{2^4} \\
&= [4+24+36+16]/2^4 = 5. \qquad (2)
\end{aligned}$$

The variance of X is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 5 - 2^2 = 1. \qquad (3)$$

Thus, X has standard deviation $\sigma_X = \sqrt{\mathrm{Var}[X]} = 1$.

(b) The probability that X is within one standard deviation of its expected value is

$$\begin{aligned}
\mathrm{P}\left[\mu_X - \sigma_X \leq X \leq \mu_X + \sigma_X\right] &= \mathrm{P}\left[2-1 \leq X \leq 2+1\right] \\
&= \mathrm{P}\left[1 \leq X \leq 3\right]. \qquad (4)
\end{aligned}$$

This calculation is easy using the PMF of X:

$$\mathrm{P}\left[1 \leq X \leq 3\right] = P_X(1) + P_X(2) + P_X(3) = 7/8. \qquad (5)$$

Problem 3.8.6 Solution

(a) The expected value of X is

$$\begin{aligned} \mathrm{E}[X] &= \sum_{x=0}^{5} x P_X(x) \\ &= 0\binom{5}{0}\frac{1}{2^5} + 1\binom{5}{1}\frac{1}{2^5} + 2\binom{5}{2}\frac{1}{2^5} \\ &\qquad + 3\binom{5}{3}\frac{1}{2^5} + 4\binom{5}{4}\frac{1}{2^5} + 5\binom{5}{5}\frac{1}{2^5} \\ &= [5+20+30+20+5]/2^5 = 5/2. \end{aligned} \tag{1}$$

The expected value of X^2 is

$$\begin{aligned} \mathrm{E}\left[X^2\right] &= \sum_{x=0}^{5} x^2 P_X(x) \\ &= 0^2\binom{5}{0}\frac{1}{2^5} + 1^2\binom{5}{1}\frac{1}{2^5} + 2^2\binom{5}{2}\frac{1}{2^5} \\ &\qquad + 3^2\binom{5}{3}\frac{1}{2^5} + 4^2\binom{5}{4}\frac{1}{2^5} + 5^2\binom{5}{5}\frac{1}{2^5} \\ &= [5+40+90+80+25]/2^5 = 240/32 = 15/2. \end{aligned} \tag{2}$$

The variance of X is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 15/2 - 25/4 = 5/4. \tag{3}$$

By taking the square root of the variance, the standard deviation of X is $\sigma_X = \sqrt{5/4} \approx 1.12$.

(b) The probability that X is within one standard deviation of its mean is

$$\begin{aligned} \mathrm{P}[\mu_X - \sigma_X \le X \le \mu_X + \sigma_X] &= \mathrm{P}[2.5 - 1.12 \le X \le 2.5 + 1.12] \\ &= \mathrm{P}[1.38 \le X \le 3.62] \\ &= \mathrm{P}[2 \le X \le 3]. \end{aligned} \tag{4}$$

By summing the PMF over the desired range, we obtain

$$\begin{aligned} \mathrm{P}\left[2 \leq X \leq 3\right] &= P_X(2) + P_X(3) \\ &= 10/32 + 10/32 = 5/8. \end{aligned} \tag{5}$$

Problem 3.8.7 Solution

For $Y = aX + b$, we wish to show that $\mathrm{Var}[Y] = a^2 \mathrm{Var}[X]$. We begin by noting that Theorem 3.12 says that $\mathrm{E}[aX + b] = a\,\mathrm{E}[X] + b$. Hence, by the definition of variance.

$$\begin{aligned} \mathrm{Var}\left[Y\right] &= \mathrm{E}\left[(aX + b - (a\,\mathrm{E}\left[X\right] + b))^2\right] \\ &= \mathrm{E}\left[a^2(X - \mathrm{E}\left[X\right])^2\right] \\ &= a^2\,\mathrm{E}\left[(X - \mathrm{E}\left[X\right])^2\right]. \end{aligned} \tag{1}$$

Since $\mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{Var}[X]$, the assertion is proved.

Problem 3.8.8 Solution

Given that

$$Y = \frac{1}{\sigma_x}(X - \mu_X), \tag{1}$$

we can use the linearity property of the expectation operator to find the mean value

$$\mathrm{E}\left[Y\right] = \frac{\mathrm{E}\left[X - \mu_X\right]}{\sigma_X} = \frac{\mathrm{E}\left[X\right] - \mathrm{E}\left[X\right]}{\sigma_X} = 0. \tag{2}$$

Using the fact that $\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X]$, the variance of Y is found to be

$$\mathrm{Var}\left[Y\right] = \frac{1}{\sigma_X^2} \mathrm{Var}\left[X\right] = 1. \tag{3}$$

Problem 3.8.9 Solution

With our measure of jitter being σ_T, and the fact that $T = 2X - 1$, we can express the jitter as a function of q by realizing that

$$\mathrm{Var}[T] = 4\,\mathrm{Var}[X] = \frac{4q}{(1-q)^2}. \tag{1}$$

Therefore, our maximum permitted jitter is

$$\sigma_T = \frac{2\sqrt{q}}{(1-q)} = 2 \text{ ms}. \tag{2}$$

Solving for q yields $q^2 - 3q + 1 = 0$. By solving this quadratic equation, we obtain

$$q = \frac{3 \pm \sqrt{5}}{2} = 3/2 \pm \sqrt{5}/2. \tag{3}$$

Since q must be a value between 0 and 1, we know that a value of $q = 3/2 - \sqrt{5}/2 \approx 0.382$ will ensure a jitter of at most 2 milliseconds.

Problem 3.8.10 Solution

The PMF of K is the Poisson PMF

$$P_K(k) = \begin{cases} \lambda^k e^{-\lambda}/k! & k = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The mean of K is

$$\mathrm{E}[K] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda. \tag{2}$$

To find $\mathrm{E}[K^2]$, we use the hint and first find

$$\mathrm{E}[K(K-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!}. \tag{3}$$

By factoring out λ^2 and substituting $j = k - 2$, we obtain

$$\mathrm{E}[K(K-1)] = \lambda^2 \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!}}_{1} = \lambda^2. \tag{4}$$

The above sum equals 1 because it is the sum of a Poisson PMF over all possible values. Since $\mathrm{E}[K] = \lambda$, the variance of K is

$$\begin{aligned} \mathrm{Var}[K] &= \mathrm{E}\left[K^2\right] - (\mathrm{E}[K])^2 \\ &= \mathrm{E}[K(K-1)] + \mathrm{E}[K] - (\mathrm{E}[K])^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned} \tag{5}$$

Problem 3.8.11 Solution

The standard deviation can be expressed as

$$\sigma_D = \sqrt{\text{Var}[D]} = \sqrt{\text{E}\left[D^2\right] - \text{E}\left[D\right]^2}, \tag{1}$$

where

$$\text{E}\left[D^2\right] = \sum_{d=1}^{4} d^2 P_D(d) = 0.2 + 1.6 + 2.7 + 1.6 = 6.1. \tag{2}$$

So finally we have

$$\sigma_D = \sqrt{6.1 - 2.3^2} = \sqrt{0.81} = 0.9. \tag{3}$$

Problem 3.9.1 Solution

For a binomial (n, p) random variable X, the solution in terms of math is

$$\text{P}\left[E_2\right] = \sum_{x=0}^{\lfloor\sqrt{n}\rfloor} P_X\left(x^2\right). \tag{1}$$

In terms of MATLAB, the efficient solution is to generate the vector of perfect squares $x =$ `[0 1 4 9 16 ...]` and then to pass that vector to the `binomialpmf.m`. In this case, the values of the binomial PMF are calculated only once. Here is the code:

```
function q=perfectbinomial(n,p);
i=0:floor(sqrt(n));
x=i.^2;
q=sum(binomialpmf(n,p,x));
```

For a binomial $(100, 0.5)$ random variable X, the probability X is a perfect square is

```
>> perfectbinomial(100,0.5)
ans =
    0.0811
```

Problem 3.9.2 Solution

```
function x=shipweight8(m);
sx=1:8;
p=[0.15*ones(1,4) 0.1*ones(1,4)];
x=finiterv(sx,p,m);
```

The random variable X given in Example 3.27 is just a finite random variable. We can generate random samples using the `finiterv` function. The code is shown on the left.

Here are some examples of sample runs of `shipweight8`:

```
>> shipweight8(12)
ans =
    8     4     1     2     3     7     1     1     2     5     6     5
>> shipweight8(12)
ans =
    4     4     2     6     2     5     2     3     5     6     1     8
```

Problem 3.9.3 Solution

Recall that in Example 3.27 that the weight in pounds X of a package and the cost $Y = g(X)$ of shipping a package were described by

$$P_X(x) = \begin{cases} 0.15 & x = 1,2,3,4, \\ 0.1 & x = 5,6,7,8, \\ 0 & \text{otherwise}, \end{cases} \qquad Y = \begin{cases} 105X - 5X^2 & 1 \le X \le 5, \\ 500 & 6 \le X \le 10. \end{cases} \tag{1}$$

```
%shipcostpmf.m
sx=(1:8)';
px=[0.15*ones(4,1); ...
        0.1*ones(4,1)];
gx=(sx<=5).* ...
    (105*sx-5*(sx.^2))...
   + ((sx>5).*500);
sy=unique(gx)'
py=finitepmf(gx,px,sy)'
```

The `shipcostpmf` script on the left calculates the PMF of Y. The vector `gx` is the mapping $g(x)$ for each $x \in S_X$. In `gx`, the element 500 appears three times, corresponding to $x = 6$, $x = 7$, and $x = 8$. The function `sy=unique(gx)` extracts the unique elements of `gx` while `finitepmf(gx,px,sy)` calculates the probability of each element of `sy`.

Here is the output:

```
>> shipcostpmf
sy =
   100   190   270   340   400   500
py =
   0.15    0.15    0.15    0.15    0.10    0.30
```

Problem 3.9.4 Solution

First we use **shipweight8** from Problem 3.9.2 to generate m samples of the package weight X. Next we convert that to m samples of the shipment cost Y. Summing these samples and dividing by m, we obtain the average cost of m samples. Here is the code:

```
function y=avgship(m);
x=shipweight8(m);
yy=cumsum([100 90 80 70 60]);
yy=[yy 500 500 500];
y=sum(yy(x))/m;
```

Each time we perform the experiment of executing the function `avgship(m)`, we generate m random samples of X, and m corresponding samples of Y. The sum $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ is random. For $m = 10$, four samples of $\overline{Y}$ are

```
>> [avgship(10) avgship(10) avgship(10) avgship(10) avgship(10)]
ans =
   365    260    313    255    350
```

For $m = 100$, the results are arguably more consistent:

```
>> [avgship(100) avgship(100) avgship(100) avgship(100)]
ans =
  328.1000  316.6000  325.9000  339.7000
```

Finally, for $m = 1000$, we obtain results reasonably close to $E[Y]$:

```
>> [avgship(1000) avgship(1000) avgship(1000) avgship(1000)]
ans =
  323.4900  321.8400  319.0500  325.5600
```

In Chapter 10, we will develop techniques to show how $\overline{Y}$ converges to $E[Y]$ as $m \to \infty$.

Problem 3.9.5 Solution

Suppose X_n is a Zipf $(n, \alpha = 1)$ random variable and thus has PMF

$$P_X(x) = \begin{cases} c(n)/x & x = 1, 2, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The problem asks us to find the smallest value of k such that $P[X_n \leq k] \geq 0.75$. That is, if the server caches the k most popular files, then with $P[X_n \leq k]$ the request is for one of the k cached files. First, we might as well solve this problem for any probability p rather than just $p = 0.75$. Thus, in math terms, we are looking for

$$k = \min\left\{k' \middle| \, P\left[X_n \leq k'\right] \geq p\right\}. \tag{2}$$

What makes the Zipf distribution hard to analyze is that there is no closed form expression for

$$c(n) = \left(\sum_{x=1}^{n} \frac{1}{x}\right)^{-1}. \tag{3}$$

Thus, we use MATLAB to grind through the calculations. The following simple program generates the Zipf distributions and returns the correct value of k.

```
function k=zipfcache(n,p);
%Usage: k=zipfcache(n,p);
%for the Zipf (n,alpha=1) distribution, returns the smallest k
%such that the first k items have total probability p
pmf=1./(1:n);
pmf=pmf/sum(pmf); %normalize to sum to 1
cdf=cumsum(pmf);
k=1+sum(cdf<=p);
```

The program `zipfcache` generalizes 0.75 to be the probability p. Although this program is sufficient, the problem asks us to find k for all values of n from 1 to 10^3!. One way to do this is to call `zipfcache` a thousand times to find k for each value of n. A better way is to use the properties of the Zipf PDF. In particular,

$$P\left[X_n \leq k'\right] = c(n) \sum_{x=1}^{k'} \frac{1}{x} = \frac{c(n)}{c(k')}. \tag{4}$$

Thus we wish to find

$$k = \min\left\{k' \middle| \frac{c(n)}{c(k')} \geq p\right\} = \min\left\{k' \middle| \frac{1}{c(k')} \geq \frac{p}{c(n)}\right\}. \tag{5}$$

Note that the definition of k implies that

$$\frac{1}{c(k')} < \frac{p}{c(n)}, \qquad k' = 1, \ldots, k-1. \tag{6}$$

Using the notation $|A|$ to denote the number of elements in the set A, we can write

$$k = 1 + \left| \left\{ k' | \frac{1}{c(k')} < \frac{p}{c(n)} \right\} \right|. \tag{7}$$

This is the basis for a very short MATLAB program:

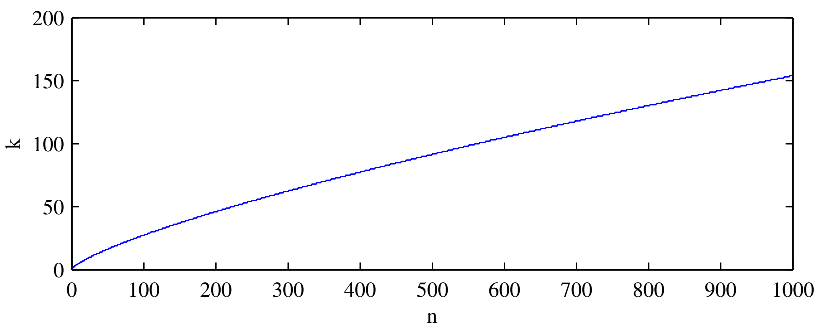
```
function k=zipfcacheall(n,p);
%Usage: k=zipfcacheall(n,p);
%returns vector k such that the first
%k(m) items have total probability >= p
%for the Zipf(m,1) distribution.
c=1./cumsum(1./(1:n));
k=1+countless(1./c,p./c);
```

Note that `zipfcacheall` uses a short MATLAB program `countless.m` that is almost the same as `count.m` introduced in Example 3.40.If `n=countless(x,y)`, then `n(i)` is the number of elements of `x` that are strictly less than `y(i)` while `count` returns the number of elements less than or equal to `y(i)`.

In any case, the commands

```
k=zipfcacheall(1000,0.75);
plot(1:1000,k);
```

is sufficient to produce this figure of k as a function of m:



We see in the figure that the number of files that must be cached grows slowly with the total number of files n.

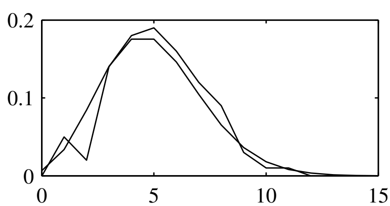
Finally, we make one last observation. It is generally desirable for MATLAB to execute operations in parallel. The program `zipfcacheall` generally will run faster than n calls to `zipfcache`. However, to do its counting all at once, `countless` generates and $n \times n$ array. When n is not too large, say $n \leq 1000$, the resulting array with $n^2 = 1{,}000{,}000$ elements fits in memory. For much larger values of n, say $n = 10^6$ (as was proposed in the original printing of this edition of the text, `countless` will cause an "out of memory" error.

Problem 3.9.6 Solution

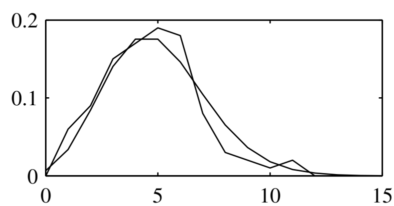
We use `poissonrv.m` to generate random samples of a Poisson ($\alpha = 5$) random variable. To compare the Poisson PMF against the output of `poissonrv`, relative frequencies are calculated using the `hist` function. The following code plots the relative frequency against the PMF.

```
function diff=poissontest(alpha,m)
x=poissonrv(alpha,m);
xr=0:ceil(3*alpha);
pxsample=hist(x,xr)/m;
pxsample=pxsample(:);
%pxsample=(countequal(x,xr)/m);
px=poissonpmf(alpha,xr);
plot(xr,pxsample,xr,px);
diff=sum((pxsample-px).^2);
```

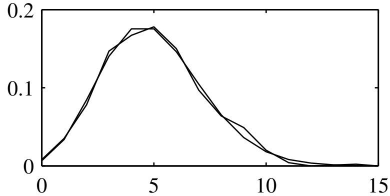
For $m = 100, 1000, 10000$, here are sample plots comparing the PMF and the relative frequency. The plots show reasonable agreement for $m = 10000$ samples.



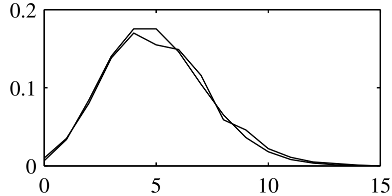
(a) $m = 100$



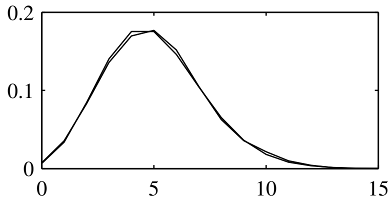
(b) $m = 100$



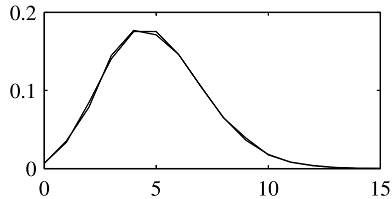
(a) $m = 1000$



(b) $m = 1000$



(a) $m = 10{,}000$



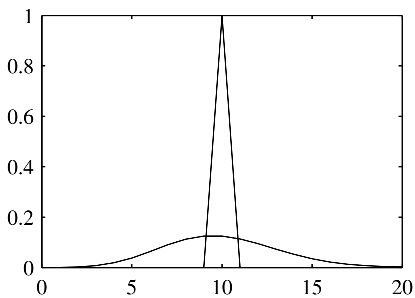
(b) $m = 10{,}000$

Problem 3.9.7 Solution

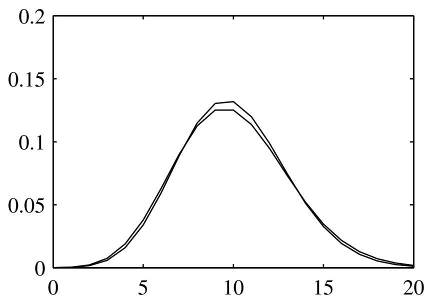
We can compare the binomial and Poisson PMFs for $(n, p) = (100, 0.1)$ using the following MATLAB code:

```
x=0:20;
p=poissonpmf(100,x);
b=binomialpmf(100,0.1,x);
plot(x,p,x,b);
```

For $(n, p) = (10, 1)$, the binomial PMF has no randomness. For $(n, p) = (100, 0.1)$, the approximation is reasonable:

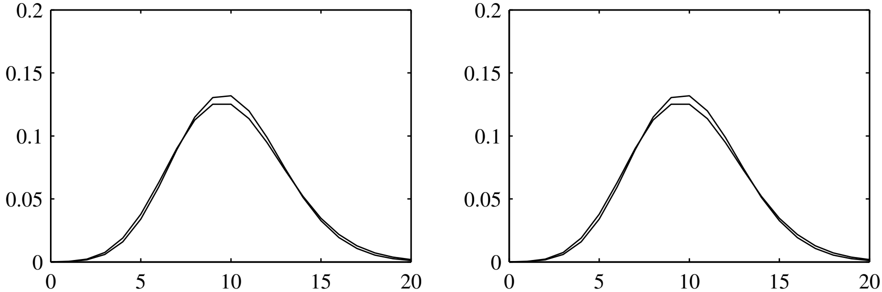


(a) $n = 10, p = 1$



(b) $n = 100, p = 0.1$

Finally, for $(n, p) = (1000, 0.01)$, and $(n, p) = (10000, 0.001)$, the approximation is very good:



(a) $n = 1000, p = 0.01$ (b) $n = 10000, p = 0.001$

Problem 3.9.8 Solution

Following the Random Sample algorithm, we generate a sample value $R =$ `rand(1)` and then we find k^* such that

$$F_K(k^* - 1) < R < F_K(k^*) . \tag{1}$$

From Problem 3.4.4, we know for integers $k \geq 1$ that geometric (p) random variable K has CDF $F_K(k) = 1 - (1 - p)^k$. Thus,

$$1 - (1 - p)^{k^* - 1} < R \leq 1 - (1 - p)^{k^*}. \tag{2}$$

Subtracting 1 from each side and then multiplying through by -1 (which reverses the inequalities), we obtain

$$(1 - p)^{k^* - 1} > 1 - R \geq (1 - p)^{k^*}. \tag{3}$$

Next we take the logarithm of each side. Since logarithms are monotonic functions, we have

$$(k^* - 1) \ln(1 - p) > \ln(1 - R) \geq k^* \ln(1 - p). \tag{4}$$

Since $0 < p < 1$, we have that $\ln(1 - p) < 0$. Thus dividing through by $\ln(1 - p)$ reverses the inequalities, yielding

$$k^* - 1 < \frac{\ln(1 - R)}{\ln(1 - p)} \leq k^*. \tag{5}$$

Since k^* is an integer, it must be the smallest integer greater than or equal to $\ln(1-R)/\ln(1-p)$. That is, following the last step of the random sample algorithm,

$$K = k^* = \left\lceil \frac{\ln(1-R)}{\ln(1-p)} \right\rceil \tag{6}$$

The MATLAB algorithm that implements this operation is quite simple:

```
function x=geometricrv(p,m)
%Usage: x=geometricrv(p,m)
%   returns m samples of a geometric (p) rv
r=rand(m,1);
x=ceil(log(1-r)/log(1-p));
```

Problem 3.9.9 Solution

For the PC version of MATLAB employed for this test, `poissonpmf(n,n)` reported `Inf` for $n = n^* = 714$. The problem with the `poissonpmf` function in Example 3.37 is that the cumulative product that calculated $n^k/k!$ can have an overflow. Following the hint, we can write an alternate `poissonpmf` function as follows:

```
function pmf=poissonpmf(alpha,x)
%Poisson (alpha) rv X,
%out=vector pmf: pmf(i)=P[X=x(i)]
x=x(:);
if (alpha==0)
    pmf=1.0*(x==0);
else
    k=(1:ceil(max(x)))';
    logfacts =cumsum(log(k));
    pb=exp([-alpha; ...
        -alpha+ (k*log(alpha))-logfacts]);
    okx=(x>=0).*(x==floor(x));
    x=okx.*x;
    pmf=okx.*pb(x+1);
end
%pmf(i)=0 for zero-prob x(i)
```

By summing logarithms, the intermediate terms are much less likely to overflow.

Problem Solutions – Chapter 4

Problem 4.2.1 Solution

The CDF of X is

$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/2 & -1 \leq x < 1, \\ 1 & x \geq 1. \end{cases} \tag{1}$$

Each question can be answered by expressing the requested probability in terms of $F_X(x)$.

(a)

$$\begin{aligned} \mathrm{P}[X > 1/2] &= 1 - \mathrm{P}[X \leq 1/2] \\ &= 1 - F_X(1/2) = 1 - 3/4 = 1/4. \end{aligned} \tag{2}$$

(b) This is a little trickier than it should be. Being careful, we can write

$$\begin{aligned} \mathrm{P}[-1/2 \leq X < 3/4] = \mathrm{P}&[-1/2 < X \leq 3/4] \\ &+ \mathrm{P}[X = -1/2] - \mathrm{P}[X = 3/4]. \end{aligned} \tag{3}$$

Since the CDF of X is a continuous function, the probability that X takes on any specific value is zero. This implies $\mathrm{P}[X = 3/4] = 0$ and $\mathrm{P}[X = -1/2] = 0$. (If this is not clear at this point, it will become clear in Section 4.7.) Thus,

$$\begin{aligned} \mathrm{P}[-1/2 \leq X < 3/4] &= \mathrm{P}[-1/2 < X \leq 3/4] \\ &= F_X(3/4) - F_X(-1/2) = 5/8. \end{aligned} \tag{4}$$

(c)

$$\begin{aligned} \mathrm{P}[|X| \leq 1/2] &= \mathrm{P}[-1/2 \leq X \leq 1/2] \\ &= \mathrm{P}[X \leq 1/2] - \mathrm{P}[X < -1/2]. \end{aligned} \tag{5}$$

Note that $P[X \leq 1/2] = F_X(1/2) = 3/4$. Since $P[X = -1/2] = 0$,

$$P[X < -1/2] = P[X \leq -1/2] = F_X(-1/2) = 1/4. \tag{6}$$

This implies

$$\begin{aligned} P[|X| \leq 1/2] &= P[X \leq 1/2] - P[X < -1/2] \\ &= 3/4 - 1/4 = 1/2. \end{aligned} \tag{7}$$

(d) Since $F_X(1) = 1$, we must have $a \leq 1$. For $a \leq 1$, we need to satisfy

$$P[X \leq a] = F_X(a) = \frac{a+1}{2} = 0.8. \tag{8}$$

Thus $a = 0.6$.

Problem 4.2.2 Solution

The CDF of V was given to be

$$F_V(v) = \begin{cases} 0 & v < -5, \\ c(v+5)^2 & -5 \leq v < 7, \\ 1 & v \geq 7. \end{cases} \tag{1}$$

(a) For V to be a continuous random variable, $F_V(v)$ must be a continuous function. This occurs if we choose c such that $F_V(v)$ doesn't have a discontinuity at $v = 7$. We meet this requirement if $c(7+5)^2 = 1$. This implies $c = 1/144$.

(b)

$$P[V > 4] = 1 - P[V \leq 4] = 1 - F_V(4) = 1 - 81/144 = 63/144. \tag{2}$$

(c)

$$P[-3 < V \leq 0] = F_V(0) - F_V(-3) = 25/144 - 4/144 = 21/144. \tag{3}$$

(d) Since $0 \leq F_V(v) \leq 1$ and since $F_V(v)$ is a nondecreasing function, it must be that $-5 \leq a \leq 7$. In this range,

$$P[V > a] = 1 - F_V(a) = 1 - (a+5)^2/144 = 2/3. \tag{4}$$

The unique solution in the range $-5 \leq a \leq 7$ is $a = 4\sqrt{3} - 5 = 1.928$.

Problem 4.2.3 Solution

(a) By definition, $\lceil nx \rceil$ is the smallest integer that is greater than or equal to nx. This implies $nx \leq \lceil nx \rceil \leq nx + 1$.

(b) By part (a),

$$\frac{nx}{n} \leq \frac{\lceil nx \rceil}{n} \leq \frac{nx+1}{n}. \tag{1}$$

That is,

$$x \leq \frac{\lceil nx \rceil}{n} \leq x + \frac{1}{n}. \tag{2}$$

This implies

$$x \leq \lim_{n\to\infty} \frac{\lceil nx \rceil}{n} \leq \lim_{n\to\infty} x + \frac{1}{n} = x. \tag{3}$$

(c) In the same way, $\lfloor nx \rfloor$ is the largest integer that is less than or equal to nx. This implies $nx - 1 \leq \lfloor nx \rfloor \leq nx$. It follows that

$$\frac{nx-1}{n} \leq \frac{\lfloor nx \rfloor}{n} \leq \frac{nx}{n}. \tag{4}$$

That is,

$$x - \frac{1}{n} \leq \frac{\lfloor nx \rfloor}{n} \leq x. \tag{5}$$

This implies

$$\lim_{n\to\infty} x - \frac{1}{n} = x \leq \lim_{n\to\infty} \frac{\lceil nx \rceil}{n} \leq x. \tag{6}$$

Problem 4.2.4 Solution

In this problem, the CDF of W is

$$F_W(w) = \begin{cases} 0 & w < -5, \\ (w+5)/8 & -5 \leq w < -3, \\ 1/4 & -3 \leq w < 3, \\ 1/4 + 3(w-3)/8 & 3 \leq w < 5, \\ 1 & w \geq 5. \end{cases} \tag{1}$$

Each question can be answered directly from this CDF.

(a)

$$P[W \leq 4] = F_W(4) = 1/4 + 3/8 = 5/8. \tag{2}$$

(b)

$$P[-2 < W \leq 2] = F_W(2) - F_W(-2) = 1/4 - 1/4 = 0. \tag{3}$$

(c)

$$P[W > 0] = 1 - P[W \leq 0] = 1 - F_W(0) = 3/4. \tag{4}$$

(d) By inspection of $F_W(w)$, we observe that $P[W \leq a] = F_W(a) = 1/2$ for a in the range $3 \leq a \leq 5$. In this range,

$$F_W(a) = 1/4 + 3(a-3)/8 = 1/2. \tag{5}$$

This implies $a = 11/3$.

Problem 4.3.1 Solution

$$f_X(x) = \begin{cases} cx & 0 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) From the above PDF we can determine the value of c by integrating the PDF and setting it equal to 1, yielding

$$\int_0^2 cx\, dx = 2c = 1. \tag{2}$$

Therefore $c = 1/2$.

(b) $P[0 \leq X \leq 1] = \int_0^1 \frac{x}{2}\, dx = 1/4$.

(c) $P[-1/2 \leq X \leq 1/2] = \int_0^{1/2} \frac{x}{2}\, dx = 1/16$.

(d) The CDF of X is found by integrating the PDF from 0 to x.

$$F_X(x) = \int_0^x f_X(x')\, dx' = \begin{cases} 0 & x < 0, \\ x^2/4 & 0 \leq x \leq 2, \\ 1 & x > 2. \end{cases} \tag{3}$$

Problem 4.3.2 Solution

From the CDF, we can find the PDF by direct differentiation. The CDF and corresponding PDF are

$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/2 & -1 \leq x \leq 1, \\ 1 & x > 1, \end{cases} \tag{1}$$

$$f_X(x) = \begin{cases} 1/2 & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 4.3.3 Solution

We find the PDF by taking the derivative of $F_U(u)$ on each piece that $F_U(u)$ is defined. The CDF and corresponding PDF of U are

$$F_U(u) = \begin{cases} 0 & u < -5, \\ (u+5)/8 & -5 \leq u < -3, \\ 1/4 & -3 \leq u < 3, \\ 1/4 + 3(u-3)/8 & 3 \leq u < 5, \\ 1 & u \geq 5, \end{cases} \tag{1}$$

$$f_U(u) = \begin{cases} 0 & u < -5, \\ 1/8 & -5 \leq u < -3, \\ 0 & -3 \leq u < 3, \\ 3/8 & 3 \leq u < 5, \\ 0 & u \geq 5. \end{cases} \tag{2}$$

Problem 4.3.4 Solution

For $x < 0$, $F_X(x) = 0$. For $x \geq 0$,

$$\begin{aligned} F_X(x) &= \int_0^x f_X(y)\, dy \\ &= \int_0^x a^2 y e^{-a^2y^2/2}\, dy = -e^{-a^2y^2/2}\Big|_0^x = 1 - e^{-a^2x^2/2}. \end{aligned} \tag{1}$$

A complete expression for the CDF of X is

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-a^2x^2/2} & x \geq 0 \end{cases} \tag{2}$$

Problem 4.3.5 Solution

For $x > 2$,

$$f_X(x) = (1/2) f_2(x) = (c_2/2)e^{-x}. \tag{1}$$

The non-negativity requirement $f_X(x) \geq 0$ for all x implies $c_2 \geq 0$. For $0 \leq x \leq 2$, non-negativity implies

$$\frac{c_1}{2} + \frac{c_2}{2}e^{-x} \geq 0, \qquad 0 \leq x \leq 2. \tag{2}$$

Since $c_2 \geq 0$, we see that this condition is satisfied if and only if

$$\frac{c_1}{2} + \frac{c_2}{2} e^{-2} \geq 0, \tag{3}$$

which simplifies to $c_1 \geq -c_2 e^{-2}$. Finally the requirement that the PDF integrates to unity yields

$$\begin{aligned} 1 &= \frac{1}{2} \int_{-\infty}^{\infty} f_1(x)\, dx + \frac{1}{2} \int_{-\infty}^{\infty} f_2(x)\, dx \\ &= \frac{1}{2} \int_0^2 c_1\, dx + \frac{1}{2} \int_0^{\infty} c_2 e^{-x}\, dx \\ &= c_1 + c_2/2. \end{aligned} \tag{4}$$

Thus $c_1 = 1 - c_2/2$ and we can represent our three constraints in terms of c_2 as

$$c_2 \geq 0, \qquad 1 - c_2/2 \geq -c_2 e^{-2}. \tag{5}$$

This can be simplified to

$$c_1 = 1 - c_2/2, \qquad 0 \leq c_2 \leq \frac{1}{1/2 - e^{-2}} = 2.742. \tag{6}$$

We note that this problem is tricky because $f_X(x)$ can be a valid PDF even if $c_1 < 0$.

Problem 4.3.6 Solution

$$f_X(x) = \begin{cases} ax^2 + bx & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

First, we note that a and b must be chosen such that the above PDF integrates to 1.

$$\int_0^1 (ax^2 + bx)\, dx = a/3 + b/2 = 1 \tag{2}$$

Hence, $b = 2 - 2a/3$ and our PDF becomes

$$f_X(x) = x(ax + 2 - 2a/3) \tag{3}$$

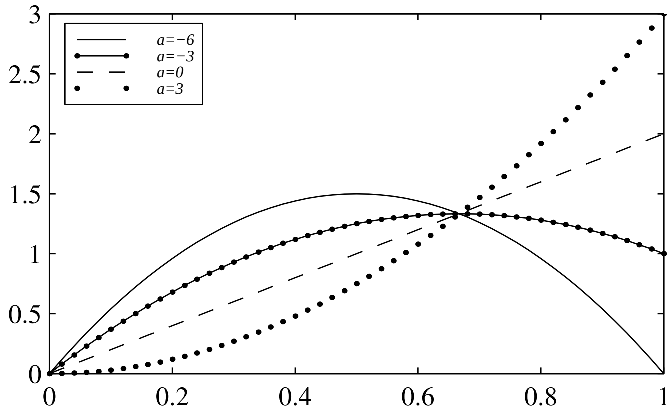
For the PDF to be non-negative for $x \in [0,1]$, we must have $ax + 2 - 2a/3 \geq 0$ for all $x \in [0,1]$. This requirement can be written as

$$a(2/3 - x) \leq 2, \qquad 0 \leq x \leq 1. \tag{4}$$

For $x = 2/3$, the requirement holds for all a. However, the problem is tricky because we must consider the cases $0 \leq x < 2/3$ and $2/3 < x \leq 1$ separately because of the sign change of the inequality. When $0 \leq x < 2/3$, we have $2/3 - x > 0$ and the requirement is most stringent at $x = 0$ where we require $2a/3 \leq 2$ or $a \leq 3$. When $2/3 < x \leq 1$, we can write the constraint as $a(x - 2/3) \geq -2$. In this case, the constraint is most stringent at $x = 1$, where we must have $a/3 \geq -2$ or $a \geq -6$. Thus a complete expression for our requirements are

$$-6 \leq a \leq 3, \qquad b = 2 - 2a/3. \tag{5}$$

As we see in the following plot, the shape of the PDF $f_X(x)$ varies greatly with the value of a.



Problem 4.4.1 Solution

$$f_X(x) = \begin{cases} 1/4 & -1 \leq x \leq 3, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We recognize that X is a uniform random variable from [-1,3].

(a) $\mathrm{E}[X] = 1$ and $\mathrm{Var}[X] = \frac{(3+1)^2}{12} = 4/3$.

(b) The new random variable Y is defined as $Y = h(X) = X^2$. Therefore

$$h(\mathrm{E}[X]) = h(1) = 1 \tag{2}$$

and

$$\mathrm{E}[h(X)] = \mathrm{E}\left[X^2\right] = \mathrm{Var}[X] + \mathrm{E}[X]^2 = 4/3 + 1 = 7/3. \tag{3}$$

(c) Finally

$$\mathrm{E}[Y] = \mathrm{E}[h(X)] = \mathrm{E}\left[X^2\right] = 7/3, \tag{4}$$

$$\mathrm{Var}[Y] = \mathrm{E}\left[X^4\right] - \mathrm{E}\left[X^2\right]^2 = \int_{-1}^{3} \frac{x^4}{4}\,dx - \frac{49}{9} = \frac{61}{5} - \frac{49}{9}. \tag{5}$$

Problem 4.4.2 Solution

(a) Since the PDF is uniform over [1,9]

$$\mathrm{E}[X] = \frac{1+9}{2} = 5, \qquad \mathrm{Var}[X] = \frac{(9-1)^2}{12} = \frac{16}{3}. \tag{1}$$

(b) Define $h(X) = 1/\sqrt{X}$ then

$$h(\mathrm{E}[X]) = 1/\sqrt{5}, \tag{2}$$

$$\mathrm{E}[h(X)] = \int_{1}^{9} \frac{x^{-1/2}}{8}\,dx = 1/2. \tag{3}$$

(c)

$$\mathrm{E}[Y] = \mathrm{E}[h(X)] = 1/2, \tag{4}$$

$$\begin{aligned}\mathrm{Var}[Y] &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 \\ &= \int_{1}^{9} \frac{x^{-1}}{8}\,dx - \mathrm{E}[X]^2 = \frac{\ln 9}{8} - 1/4.\end{aligned} \tag{5}$$

Problem 4.4.3 Solution

The CDF of X is

$$F_X(x) = \begin{cases} 0 & x < 0, \\ x/2 & 0 \le x < 2, \\ 1 & x \ge 2. \end{cases} \tag{1}$$

(a) To find $E[X]$, we first find the PDF by differentiating the above CDF.

$$f_X(x) = \begin{cases} 1/2 & 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The expected value is then

$$E[X] = \int_0^2 \frac{x}{2}\,dx = 1. \tag{3}$$

(b)

$$E\left[X^2\right] = \int_0^2 \frac{x^2}{2}\,dx = 8/3, \tag{4}$$

$$\text{Var}[X] = E\left[X^2\right] - E[X]^2 = 8/3 - 1 = 5/3. \tag{5}$$

Problem 4.4.4 Solution

We can find the expected value of X by direct integration of the given PDF.

$$f_Y(y) = \begin{cases} y/2 & 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expectation is

$$E[Y] = \int_0^2 \frac{y^2}{2}\,dy = 4/3. \tag{2}$$

To find the variance, we first find the second moment

$$E\left[Y^2\right] = \int_0^2 \frac{y^3}{2}\,dy = 2. \tag{3}$$

The variance is then $\text{Var}[Y] = E[Y^2] - E[Y]^2 = 2 - (4/3)^2 = 2/9$.

Problem 4.4.5 Solution

The CDF of Y is

$$F_Y(y) = \begin{cases} 0 & y < -1, \\ (y+1)/2 & -1 \leq y < 1, \\ 1 & y \geq 1. \end{cases} \tag{1}$$

(a) We can find the expected value of Y by first differentiating the above CDF to find the PDF

$$f_Y(y) = \begin{cases} 1/2 & -1 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It follows that

$$\mathrm{E}[Y] = \int_{-1}^{1} y/2 \, dy = 0. \tag{3}$$

(b)

$$\mathrm{E}\left[Y^2\right] = \int_{-1}^{1} \frac{y^2}{2} \, dy = 1/3, \tag{4}$$

$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2 = 1/3 - 0 = 1/3. \tag{5}$$

Problem 4.4.6 Solution

To evaluate the moments of V, we need the PDF $f_V(v)$, which we find by taking the derivative of the CDF $F_V(v)$. The CDF and corresponding PDF of V are

$$F_V(v) = \begin{cases} 0 & v < -5, \\ (v+5)^2/144 & -5 \leq v < 7, \\ 1 & v \geq 7, \end{cases} \tag{1}$$

$$f_V(v) = \begin{cases} 0 & v < -5, \\ (v+5)/72 & -5 \leq v < 7, \\ 0 & v \geq 7. \end{cases} \tag{2}$$

(a) The expected value of V is

$$\begin{aligned} \mathrm{E}\left[V\right] = \int_{-\infty}^{\infty} v f_V(v)\ dv &= \frac{1}{72}\int_{-5}^{7}(v^2+5v)\,dv \\ &= \frac{1}{72}\left(\frac{v^3}{3}+\frac{5v^2}{2}\right)\Bigg|_{-5}^{7} \\ &= \frac{1}{72}\left(\frac{343}{3}+\frac{245}{2}+\frac{125}{3}-\frac{125}{2}\right) = 3. \end{aligned} \tag{3}$$

(b) To find the variance, we first find the second moment

$$\begin{aligned} \mathrm{E}\left[V^2\right] = \int_{-\infty}^{\infty} v^2 f_V(v)\ dv &= \frac{1}{72}\int_{-5}^{7}(v^3+5v^2)\,dv \\ &= \frac{1}{72}\left(\frac{v^4}{4}+\frac{5v^3}{3}\right)\Bigg|_{-5}^{7} \\ &= 6719/432 = 15.55. \end{aligned} \tag{4}$$

The variance is $\mathrm{Var}[V] = \mathrm{E}[V^2] - (\mathrm{E}[V])^2 = 2831/432 = 6.55$.

(c) The third moment of V is

$$\begin{aligned} \mathrm{E}\left[V^3\right] &= \int_{-\infty}^{\infty} v^3 f_V(v)\ dv \\ &= \frac{1}{72}\int_{-5}^{7}(v^4+5v^3)\,dv \\ &= \frac{1}{72}\left(\frac{v^5}{5}+\frac{5v^4}{4}\right)\Bigg|_{-5}^{7} = 86.2. \end{aligned} \tag{5}$$

Problem 4.4.7 Solution

To find the moments, we first find the PDF of U by taking the derivative of $F_U(u)$.

The CDF and corresponding PDF are

$$F_U(u) = \begin{cases} 0 & u < -5, \\ (u+5)/8 & -5 \le u < -3, \\ 1/4 & -3 \le u < 3, \\ 1/4 + 3(u-3)/8 & 3 \le u < 5, \\ 1 & u \ge 5. \end{cases} \tag{1}$$

$$f_U(u) = \begin{cases} 0 & u < -5, \\ 1/8 & -5 \le u < -3, \\ 0 & -3 \le u < 3, \\ 3/8 & 3 \le u < 5, \\ 0 & u \ge 5. \end{cases} \tag{2}$$

(a) The expected value of U is

$$\begin{aligned} \mathrm{E}[U] = \int_{-\infty}^{\infty} u f_U(u)\, du &= \int_{-5}^{-3} \frac{u}{8}\, du + \int_{3}^{5} \frac{3u}{8}\, du \\ &= \left.\frac{u^2}{16}\right|_{-5}^{-3} + \left.\frac{3u^2}{16}\right|_{3}^{5} = 2. \end{aligned} \tag{3}$$

(b) The second moment of U is

$$\begin{aligned} \mathrm{E}\left[U^2\right] = \int_{-\infty}^{\infty} u^2 f_U(u)\, du &= \int_{-5}^{-3} \frac{u^2}{8}\, du + \int_{3}^{5} \frac{3u^2}{8}\, du \\ &= \left.\frac{u^3}{24}\right|_{-5}^{-3} + \left.\frac{u^3}{8}\right|_{3}^{5} = 49/3. \end{aligned} \tag{4}$$

The variance of U is $\mathrm{Var}[U] = \mathrm{E}[U^2] - (\mathrm{E}[U])^2 = 37/3$.

(c) Note that $2^U = e^{(\ln 2)U}$. This implies that

$$\int 2^u\, du = \int e^{(\ln 2)u}\, du = \frac{1}{\ln 2} e^{(\ln 2)u} = \frac{2^u}{\ln 2}. \tag{5}$$

The expected value of 2^U is then

$$\begin{aligned}
\mathrm{E}\left[2^U\right] &= \int_{-\infty}^{\infty} 2^u f_U(u)\ du \\
&= \int_{-5}^{-3} \frac{2^u}{8}\ du + \int_{3}^{5} \frac{3 \cdot 2^u}{8}\ du \\
&= \left.\frac{2^u}{8\ln 2}\right|_{-5}^{-3} + \left.\frac{3 \cdot 2^u}{8\ln 2}\right|_{3}^{5} = \frac{2307}{256\ln 2} = 13.001. \qquad (6)
\end{aligned}$$

Problem 4.4.8 Solution

The Pareto (α, μ) random variable has PDF

$$f_X(x) = \begin{cases} (\alpha/\mu)\,(x/\mu)^{-(\alpha+1)} & x \geq \mu, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

The nth moment is

$$\mathrm{E}\left[X^n\right] = \int_{\mu}^{\infty} x^n \frac{\alpha}{\mu}\left(\frac{x}{\mu}\right)^{-(\alpha+1)}\ dx = \mu^n \int_{\mu}^{\infty} \frac{\alpha}{\mu}\left(\frac{x}{\mu}\right)^{-(\alpha-n+1)}\ dx. \qquad (2)$$

With the variable substitution $y = x/\mu$, we obtain

$$\mathrm{E}\left[X^n\right] = \alpha\mu^n \int_{1}^{\infty} .y^{-(\alpha-n+1)}\, dy \qquad (3)$$

We see that $\mathrm{E}[X^n] < \infty$ if and only if $\alpha - n + 1 > 1$, or, equivalently, $n < \alpha$. In this case,

$$\begin{aligned}
\mathrm{E}\left[X^n\right] &= \left.\frac{\alpha\mu^n}{-(\alpha-n+1)+1} y^{-(\alpha-n+1)+1}\right|_{y=1}^{y=\infty} \\
&= \left.\frac{-\alpha\mu^n}{\alpha-n} y^{-(\alpha-n)}\right|_{y=1}^{y=\infty} = \frac{\alpha\mu^n}{\alpha-n}. \qquad (4)
\end{aligned}$$

Problem 4.5.1 Solution

Since Y is a continuous uniform $(a = 1, b = 5)$ random variable, we know that

$$f_Y(y) = \begin{cases} 1/4 & 1 \le y \le 5, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

and that

$$E[Y] = \frac{a+b}{2} = 3, \qquad \text{Var}[Y] = \frac{(b-a)^2}{12} = \frac{4}{3}. \tag{2}$$

With these facts, the remaining calculations are straightforward:

(a) $P[Y > E[Y]] = P[Y > 3] = \int_3^5 \frac{1}{4}\,dy = \frac{1}{2}$.

(b) $P[Y \le \text{Var}[Y]] = P[X \le 4/3] = \int_1^{4/3} \frac{1}{4}\,dy = \frac{1}{12}$.

Problem 4.5.2 Solution

Note that Y has PDF

$$f_Y(y) = \begin{cases} 1/20 & -10 \le y < 10, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Hence,

$$P[|Y| < 3] = P[-3 < Y < 3] = \int_{-3}^{3} f_Y(y)\,dy = \int_{-3}^{3} \frac{1}{20}\,dy = \frac{3}{10}. \tag{2}$$

Problem 4.5.3 Solution

The reflected power Y has an exponential $(\lambda = 1/P_0)$ PDF. From Theorem 4.8, $E[Y] = P_0$. The probability that an aircraft is correctly identified is

$$P[Y > P_0] = \int_{P_0}^{\infty} \frac{1}{P_0} e^{-y/P_0}\,dy = e^{-1}. \tag{1}$$

Fortunately, real radar systems offer better performance.

Problem 4.5.4 Solution

From Appendix A, we observe that an exponential PDF Y with parameter $\lambda > 0$ has PDF

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In addition, the mean and variance of Y are

$$\mathrm{E}[Y] = \frac{1}{\lambda}, \qquad \mathrm{Var}[Y] = \frac{1}{\lambda^2}. \tag{2}$$

(a) Since $\mathrm{Var}[Y] = 25$, we must have $\lambda = 1/5$.

(b) The expected value of Y is $\mathrm{E}[Y] = 1/\lambda = 5$.

(c)

$$P[Y > 5] = \int_5^\infty f_Y(y)\, dy = -e^{-y/5}\Big|_5^\infty = e^{-1}. \tag{3}$$

Problem 4.5.5 Solution

An exponential (λ) random variable has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

and has expected value $\mathrm{E}[Y] = 1/\lambda$. Although λ was not specified in the problem, we can still solve for the probabilities:

(a) $P[Y \geq \mathrm{E}[Y]] = \int_{1/\lambda}^\infty \lambda e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_{1/\lambda}^\infty = e^{-1}.$

(b) $P[Y \geq 2\,\mathrm{E}[Y]] = \int_{2/\lambda}^\infty \lambda e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_{2/\lambda}^\infty = e^{-2}.$

Problem 4.5.6 Solution

From Appendix A, an Erlang random variable X with parameters $\lambda > 0$ and n has PDF

$$f_X(x) = \begin{cases} \lambda^n x^{n-1} e^{-\lambda x}/(n-1)! & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In addition, the mean and variance of X are

$$E[X] = \frac{n}{\lambda}, \qquad \text{Var}[X] = \frac{n}{\lambda^2}. \tag{2}$$

(a) Since $\lambda = 1/3$ and $E[X] = n/\lambda = 15$, we must have $n = 5$.

(b) Substituting the parameters $n = 5$ and $\lambda = 1/3$ into the given PDF, we obtain

$$f_X(x) = \begin{cases} (1/3)^5 x^4 e^{-x/3}/24 & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(c) From above, we know that $\text{Var}[X] = n/\lambda^2 = 45$.

Problem 4.5.7 Solution

Since Y is an Erlang random variable with parameters $\lambda = 2$ and $n = 2$, we find in Appendix A that

$$f_Y(y) = \begin{cases} 4ye^{-2y} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) Appendix A tells us that $E[Y] = n/\lambda = 1$.

(b) Appendix A also tells us that $\text{Var}[Y] = n/\lambda^2 = 1/2$.

(c) The probability that $1/2 \leq Y < 3/2$ is

$$P[1/2 \leq Y < 3/2] = \int_{1/2}^{3/2} f_Y(y)\, dy = \int_{1/2}^{3/2} 4ye^{-2y}\, dy. \tag{2}$$

This integral is easily completed using the integration by parts formula

$$\int u\,dv = uv - \int v\,du \tag{3}$$

with

$$\begin{aligned} u &= 2y, & dv &= 2e^{-2y}, \\ du &= 2dy, & v &= -e^{-2y}. \end{aligned}$$

Making these substitutions, we obtain

$$\begin{aligned} P\left[1/2 \leq Y < 3/2\right] &= -2ye^{-2y}\Big|_{1/2}^{3/2} + \int_{1/2}^{3/2} 2e^{-2y}\,dy \\ &= 2e^{-1} - 4e^{-3} = 0.537. \end{aligned} \tag{4}$$

Problem 4.5.8 Solution

Since U is continuous uniform and zero mean, we know that U is a $(-c, c)$ continuous uniform random variable with PDF

$$f_U(u) = \begin{cases} \frac{1}{2c} & -c \leq U \leq c, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

for some constant c. We also know that $\text{Var}[U] = (2c)^2/12 = c^2/3$. Thus,

$$\begin{aligned} P\left[U^2 \leq \text{Var}[U]\right] &= P\left[U^2 \leq c^2/3\right] \\ &= P\left[-c/\sqrt{3} \leq U \leq c/\sqrt{3}\right] \\ &= \int_{-c/\sqrt{3}}^{c/\sqrt{3}} \frac{1}{2c}\,du \\ &= \frac{1}{2c}\frac{2c}{\sqrt{3}} = \frac{1}{\sqrt{3}} = 0.5774. \end{aligned} \tag{2}$$

Problem 4.5.9 Solution

Since U is a continuous uniform random variable with $\mathrm{E}[U] = 10$, we know that $u = 10$ is the midpoint of a uniform (a, b) PDF. That is, for some constant $c > 0$, U is a continuous uniform $(10 - c, 10 + c)$ random variable with PDF

$$f_U(u) = \begin{cases} 1/(2c) & 10 - c \leq u \leq 10 + c, \\ 0 & \text{otherwise.} \end{cases}$$

This implies

$$\begin{aligned} \frac{1}{4} = \mathrm{P}[U > 12] &= \int_{12}^{\infty} f_U(u)\, du \\ &= \int_{12}^{10+c} \frac{1}{2c}\, du = \frac{10 + c - 12}{2c} = \frac{1}{2} - \frac{1}{c}. \end{aligned} \tag{1}$$

This implies $1/c = 1/4$ or $c = 4$. Thus U is a uniform $(6, 14)$ random variable with PDF

$$f_U(u) = \begin{cases} 1/8 & 6 \leq u \leq 14, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\mathrm{P}[U < 9] = \int_{-\infty}^{9} f_U(u)\, du = \int_{6}^{9} \frac{1}{8}\, du = \frac{3}{8}.$$

Problem 4.5.10 Solution

(a) The PDF of a continuous uniform $(-5, 5)$ random variable is

$$f_X(x) = \begin{cases} 1/10 & -5 \leq x \leq 5, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) For $x < -5$, $F_X(x) = 0$. For $x \geq 5$, $F_X(x) = 1$. For $-5 \leq x \leq 5$, the CDF is

$$F_X(x) = \int_{-5}^{x} f_X(\tau)\, d\tau = \frac{x + 5}{10}. \tag{2}$$

The complete expression for the CDF of X is

$$F_X(x) = \begin{cases} 0 & x < -5, \\ (x+5)/10 & 5 \leq x \leq 5, \\ 1 & x > 5. \end{cases} \tag{3}$$

(c) The expected value of X is

$$\int_{-5}^{5} \frac{x}{10}\,dx = \left.\frac{x^2}{20}\right|_{-5}^{5} = 0. \tag{4}$$

Another way to obtain this answer is to use Theorem 4.6 which says the expected value of X is $\mathrm{E}[X] = (5 + -5)/2 = 0$.

(d) The fifth moment of X is

$$\int_{-5}^{5} \frac{x^5}{10}\,dx = \left.\frac{x^6}{60}\right|_{-5}^{5} = 0. \tag{5}$$

(e) The expected value of e^X is

$$\int_{-5}^{5} \frac{e^x}{10}\,dx = \left.\frac{e^x}{10}\right|_{-5}^{5} = \frac{e^5 - e^{-5}}{10} = 14.84. \tag{6}$$

Problem 4.5.11 Solution

For a uniform $(-a, a)$ random variable X,

$$\mathrm{Var}[X] = (a - (-a))^2/12 = a^2/3. \tag{1}$$

Hence $\mathrm{P}[|X| \leq \mathrm{Var}[X]] = \mathrm{P}[|X| \leq a^2/3]$. Keep in mind that

$$f_X(x) = \begin{cases} 1/(2a) & -a \leq x \leq a, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

If $a^2/3 > a$, that is $a > 3$, then we have $\mathrm{P}[|X| \leq \mathrm{Var}[X]] = 1$. Otherwise, if $a \leq 3$,

$$\mathrm{P}\left[|X| \leq \mathrm{Var}[X]\right] = \mathrm{P}\left[|X| \leq a^2/3\right] = \int_{-a^2/3}^{a^2/3} \frac{1}{2a}\,dx = a/3. \tag{3}$$

Problem 4.5.12 Solution

We know that X has a uniform PDF over $[a, b)$ and has mean $\mu_X = 7$ and variance $\text{Var}[X] = 3$. All that is left to do is determine the values of the constants a and b, to complete the model of the uniform PDF.

$$\text{E}[X] = \frac{a+b}{2} = 7, \qquad \text{Var}[X] = \frac{(b-a)^2}{12} = 3. \tag{1}$$

Since we assume $b > a$, this implies

$$a + b = 14, \qquad b - a = 6. \tag{2}$$

Solving these two equations, we arrive at

$$a = 4, \qquad b = 10. \tag{3}$$

And the resulting PDF of X is,

$$f_X(x) = \begin{cases} 1/6 & 4 \le x \le 10, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 4.5.13 Solution

Given that

$$f_X(x) = \begin{cases} (1/2)e^{-x/2} & x \ge 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

(a)

$$\text{P}[1 \le X \le 2] = \int_1^2 (1/2)e^{-x/2}\,dx = e^{-1/2} - e^{-1} = 0.2387. \tag{2}$$

(b) The CDF of X may be be expressed as

$$F_X(x) = \begin{cases} 0 & x < 0, \\ \int_0^x (1/2)e^{-x/2}\,d\tau & x \ge 0, \end{cases} = \begin{cases} 0 & x < 0, \\ 1 - e^{-x/2} & x \ge 0. \end{cases} \tag{3}$$

(c) X is an exponential random variable with parameter $a = 1/2$. By Theorem 4.8, the expected value of X is $\text{E}[X] = 1/a = 2$.

(d) By Theorem 4.8, the variance of X is $\text{Var}[X] = 1/a^2 = 4$.

Problem 4.5.14 Solution

Given the uniform PDF

$$f_U(u) = \begin{cases} 1/(b-a) & a \leq u \leq b, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The mean of U can be found by integrating

$$\mathrm{E}[U] = \int_a^b u/(b-a)\,du = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}. \tag{2}$$

Where we factored $(b^2-a^2) = (b-a)(b+a)$. The variance of U can also be found by finding $\mathrm{E}[U^2]$.

$$\mathrm{E}\left[U^2\right] = \int_a^b u^2/(b-a)\,du = \frac{(b^3-a^3)}{3(b-a)}. \tag{3}$$

Therefore the variance is

$$\mathrm{Var}[U] = \frac{(b^3-a^3)}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}. \tag{4}$$

Problem 4.5.15 Solution

Let X denote the holding time of a call. The PDF of X is

$$f_X(x) = \begin{cases} (1/\tau)e^{-x/\tau} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We will use $C_A(X)$ and $C_B(X)$ to denote the cost of a call under the two plans. From the problem statement, we note that $C_A(X) = 10X$ so that $\mathrm{E}[C_A(X)] = 10\,\mathrm{E}[X] = 10\tau$. On the other hand

$$C_B(X) = 99 + 10(X-20)^+, \tag{2}$$

where $y^+ = y$ if $y \geq 0$; otherwise $y^+ = 0$ for $y < 0$. Thus,

$$\begin{aligned} \mathrm{E}[C_B(X)] &= \mathrm{E}\left[99 + 10(X-20)^+\right] \\ &= 99 + 10\,\mathrm{E}\left[(X-20)^+\right] \\ &= 99 + 10\,\mathrm{E}\left[(X-20)^+ | X \leq 20\right] \mathrm{P}[X \leq 20] \\ &\qquad + 10\,\mathrm{E}\left[(X-20)^+ | X > 20\right] \mathrm{P}[X > 20]. \end{aligned} \tag{3}$$

Given $X \leq 20$, $(X-20)^+ = 0$. Thus $\mathrm{E}[(X-20)^+|X \leq 20] = 0$ and

$$\mathrm{E}\left[C_B(X)\right] = 99 + 10\,\mathrm{E}\left[(X-20)|X > 20\right]\mathrm{P}\left[X > 20\right]. \tag{4}$$

Finally, we observe that $\mathrm{P}[X > 20] = e^{-20/\tau}$ and that

$$\mathrm{E}\left[(X-20)|X > 20\right] = \tau \tag{5}$$

since given $X \geq 20$, $X - 20$ has a PDF identical to X by the memoryless property of the exponential random variable. Thus,

$$\mathrm{E}\left[C_B(X)\right] = 99 + 10\tau e^{-20/\tau} \tag{6}$$

Some numeric comparisons show that $\mathrm{E}[C_B(X)] \leq \mathrm{E}[C_A(X)]$ if $\tau > 12.34$ minutes. That is, the flat price for the first 20 minutes is a good deal only if your average phone call is sufficiently long.

Problem 4.5.16 Solution

The integral I_1 is

$$I_1 = \int_0^\infty \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^\infty = 1. \tag{1}$$

For $n > 1$, we have

$$I_n = \int_0^\infty \underbrace{\frac{\lambda^{n-1}x^{n-1}}{(n-1)!}}_{u} \underbrace{\lambda e^{-\lambda x}\,dt}_{dv}. \tag{2}$$

We define u and dv as shown above in order to use the integration by parts formula $\int u\,dv = uv - \int v\,du$. Since

$$du = \frac{\lambda^{n-1}x^{n-1}}{(n-2)!}\,dx, \qquad v = -e^{-\lambda x}, \tag{3}$$

we can write

$$\begin{aligned} I_n &= uv|_0^\infty - \int_0^\infty v\,du \\ &= -\frac{\lambda^{n-1}x^{n-1}}{(n-1)!}e^{-\lambda x}\Big|_0^\infty + \int_0^\infty \frac{\lambda^{n-1}x^{n-1}}{(n-2)!}e^{-\lambda x}\,dx = 0 + I_{n-1} \end{aligned} \tag{4}$$

Hence, $I_n = 1$ for all $n \geq 1$.

Problem 4.5.17 Solution

For an Erlang (n, λ) random variable X, the kth moment is

$$\begin{aligned}
\mathrm{E}\left[X^k\right] &= \int_0^\infty x^k f_X(x)\, dt \\
&= \int_0^\infty \frac{\lambda^n x^{n+k-1}}{(n-1)!} e^{-\lambda x}\, dt = \frac{(n+k-1)!}{\lambda^k (n-1)!} \underbrace{\int_0^\infty \frac{\lambda^{n+k} x^{n+k-1}}{(n+k-1)!} e^{-\lambda t}\, dt}_{1}. \qquad (1)
\end{aligned}$$

The above integral equals 1 since it is the integral of an Erlang $(n+k, \lambda)$ PDF over all possible values. Hence,

$$\mathrm{E}\left[X^k\right] = \frac{(n+k-1)!}{\lambda^k (n-1)!}. \tag{2}$$

This implies that the first and second moments are

$$\mathrm{E}[X] = \frac{n!}{(n-1)!\lambda} = \frac{n}{\lambda}, \qquad \mathrm{E}\left[X^2\right] = \frac{(n+1)!}{\lambda^2 (n-1)!} = \frac{(n+1)n}{\lambda^2}. \tag{3}$$

It follows that the variance of X is n/λ^2.

Problem 4.5.18 Solution

In this problem, we prove Theorem 4.11 which says that for $x \geq 0$, the CDF of an Erlang (n, λ) random variable X_n satisfies

$$F_{X_n}(x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}. \tag{1}$$

We do this in two steps. First, we derive a relationship between $F_{X_n}(x)$ and $F_{X_{n-1}}(x)$. Second, we use that relationship to prove the theorem by induction.

(a) By Definition 4.7, the CDF of Erlang (n, λ) random variable X_n is

$$F_{X_n}(x) = \int_{-\infty}^x f_{X_n}(t)\, dt = \int_0^x \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}\, dt. \tag{2}$$

(b) To use integration by parts, we define

$$u = \frac{t^{n-1}}{(n-1)!}, \qquad dv = \lambda^n e^{-\lambda t}\,dt, \tag{3}$$

$$du = \frac{t^{n-2}}{(n-2)!}, \qquad v = -\lambda^{n-1} e^{-\lambda t}. \tag{4}$$

Thus, using the integration by parts formula $\int u\,dv = uv - \int v\,du$, we have

$$\begin{aligned} F_{X_n}(x) = \int_0^x \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}\,dt &= -\frac{\lambda^{n-1} t^{n-1} e^{-\lambda t}}{(n-1)!}\bigg|_0^x + \int_0^x \frac{\lambda^{n-1} t^{n-2} e^{-\lambda t}}{(n-2)!}\,dt \\ &= -\frac{\lambda^{n-1} x^{n-1} e^{-\lambda x}}{(n-1)!} + F_{X_{n-1}}(x) \end{aligned} \tag{5}$$

(c) Now we do proof by induction. For $n = 1$, the Erlang (n, λ) random variable X_1 is simply an exponential random variable. Hence for $x \geq 0$, $F_{X_1}(x) = 1 - e^{-\lambda x}$. Now we suppose the claim is true for $F_{X_{n-1}}(x)$ so that

$$F_{X_{n-1}}(x) = 1 - \sum_{k=0}^{n-2} \frac{(\lambda x)^k e^{-\lambda x}}{k!}. \tag{6}$$

Using the result of part (a), we can write

$$\begin{aligned} F_{X_n}(x) &= F_{X_{n-1}}(x) - \frac{(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!} \\ &= 1 - \sum_{k=0}^{n-2} \frac{(\lambda x)^k e^{-\lambda x}}{k!} - \frac{(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}, \end{aligned} \tag{7}$$

which proves the claim.

Problem 4.5.19 Solution

For $n = 1$, we have the fact $\mathrm{E}[X] = 1/\lambda$ that is given in the problem statement. Now we assume that $\mathrm{E}[X^{n-1}] = (n-1)!/\lambda^{n-1}$. To complete the proof, we show that this implies that $\mathrm{E}[X^n] = n!/\lambda^n$. Specifically, we write

$$\mathrm{E}\left[X^n\right] = \int_0 x^n \lambda e^{-\lambda x}\,dx. \tag{1}$$

Now we use the integration by parts formula $\int u\,dv = uv - \int v\,du$ with $u = x^n$ and $dv = \lambda e^{-\lambda x}\,dx$. This implies $du = nx^{n-1}\,dx$ and $v = -e^{-\lambda x}$ so that

$$\begin{aligned} \mathrm{E}\left[X^n\right] &= -x^n e^{-\lambda x}\Big|_0^\infty + \int_0^\infty nx^{n-1}e^{-\lambda x}\,dx \\ &= 0 + \frac{n}{\lambda}\int_0^\infty x^{n-1}\lambda e^{-\lambda x}\,dx \\ &= \frac{n}{\lambda}\,\mathrm{E}\left[X^{n-1}\right]. \end{aligned} \tag{2}$$

By our induction hyothesis, $\mathrm{E}[X^{n-1}] = (n-1)!/\lambda^{n-1}$ which implies

$$\mathrm{E}\left[X^n\right] = n!/\lambda^n. \tag{3}$$

Problem 4.5.20 Solution

(a) Since $f_X(x) \geq 0$ and $x \geq r$ over the entire integral, we can write

$$\int_r^\infty x f_X(x)\,dx \geq \int_r^\infty r f_X(x)\,dx = r\,\mathrm{P}\left[X > r\right]. \tag{1}$$

(b) We can write the expected value of X in the form

$$\mathrm{E}\left[X\right] = \int_0^r x f_X(x)\,dx + \int_r^\infty x f_X(x)\,dx. \tag{2}$$

Hence,

$$r\,\mathrm{P}\left[X > r\right] \leq \int_r^\infty x f_X(x)\,dx = \mathrm{E}\left[X\right] - \int_0^r x f_X(x)\,dx. \tag{3}$$

Allowing r to approach infinity yields

$$\begin{aligned} \lim_{r\to\infty} r\,\mathrm{P}\left[X > r\right] &\leq \mathrm{E}\left[X\right] - \lim_{r\to\infty}\int_0^r x f_X(x)\,dx \\ &= \mathrm{E}\left[X\right] - \mathrm{E}\left[X\right] = 0. \end{aligned} \tag{4}$$

Since $r\,\mathrm{P}[X > r] \geq 0$ for all $r \geq 0$, we must have $\lim_{r\to\infty} r\,\mathrm{P}[X > r] = 0$.

(c) We can use the integration by parts formula $\int u\,dv = uv - \int v\,du$ by defining $u = 1 - F_X(x)$ and $dv = dx$. This yields

$$\int_0^\infty [1 - F_X(x)]\,dx = x[1 - F_X(x)]|_0^\infty + \int_0^\infty x f_X(x)\;dx. \tag{5}$$

By applying part (a), we now observe that

$$x\,[1 - F_X(x)]|_0^\infty = \lim_{r\to\infty} r[1 - F_X(r)] - 0 = \lim_{r\to\infty} r\,\mathrm{P}\,[X > r]\,. \tag{6}$$

By part (b), $\lim_{r\to\infty} r\,\mathrm{P}[X > r] = 0$ and this implies

$$x\,[1 - F_X(x)]|_0^\infty = 0. \tag{7}$$

Thus,

$$\int_0^\infty [1 - F_X(x)]\,dx = \int_0^\infty x f_X(x)\;dx = \mathrm{E}\,[X]\,. \tag{8}$$

Problem 4.6.1 Solution

Given that the peak temperature, T, is a Gaussian random variable with mean 85 and standard deviation 10 we can use the fact that $F_T(t) = \Phi((t - \mu_T)/\sigma_T)$ and Table 4.2 on page 143 to evaluate:

$$\begin{aligned}
\mathrm{P}\,[T > 100] &= 1 - \mathrm{P}\,[T \le 100] \\
&= 1 - F_T(100) \\
&= 1 - \Phi\left(\frac{100-85}{10}\right) \\
&= 1 - \Phi(1.5) = 1 - 0.933 = 0.066, \qquad (1)\\
\mathrm{P}\,[T < 60] &= \Phi\left(\frac{60-85}{10}\right) \\
&= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - .993 = 0.007, \qquad (2)\\
\mathrm{P}\,[70 \le T \le 100] &= F_T(100) - F_T(70) \\
&= \Phi(1.5) - \Phi(-1.5) = 2\Phi(1.5) - 1 = .866. \qquad (3)
\end{aligned}$$

Problem 4.6.2 Solution

The standard normal Gaussian random variable Z has mean $\mu = 0$ and variance $\sigma^2 = 1$. Making these substitutions in Definition 4.8 yields

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{1}$$

Problem 4.6.3 Solution

(a)

$$\begin{aligned} P[V > 4] = 1 - P[V \leq 4] &= 1 - P\left[\frac{V-0}{\sigma} \leq \frac{4-0}{\sigma}\right] \\ &= 1 - \Phi(4/\sigma) \\ &= 1 - \Phi(2) = 0.023. \end{aligned} \tag{1}$$

(b)

$$P[W \leq 2] = P\left[\frac{W-2}{5} \leq \frac{2-2}{5}\right] = \Phi(0) = \frac{1}{2}. \tag{2}$$

(c)

$$\begin{aligned} P[X \leq \mu + 1] &= P[X - \mu \leq 1] \\ &= P\left[\frac{X-\mu}{\sigma} \leq \frac{1}{\sigma}\right] \\ &= \Phi(1/\sigma) = \Phi(0.5) = 0.692. \end{aligned} \tag{3}$$

(d)

$$\begin{aligned} P[Y > 65] &= 1 - P[Y \leq 65] \\ &= 1 - P\left[\frac{Y-50}{10} \leq \frac{65-50}{10}\right] \\ &= 1 - \Phi(1.5) = 1 - 0.933 = 0.067. \end{aligned} \tag{4}$$

Problem 4.6.4 Solution

In each case, we are told just enough to determine μ.

(a) Since $0.933 = \Phi(1.5)$,

$$\Phi(1.5) = P[X \le 10] = P\left[\frac{X-\mu}{\sigma} \le \frac{10-\mu}{\sigma}\right] = \Phi\left(\frac{10-\mu}{\sigma}\right). \tag{1}$$

It follows that

$$\frac{10-\mu}{\sigma} = 1.5,$$

or $\mu = 10 - 1.5\sigma = -5$.

(b)

$$\begin{aligned} P[Y \le 0] &= P\left[\frac{Y-\mu}{\sigma} \le \frac{0-\mu}{\sigma}\right] \\ &= \Phi\left(\frac{-\mu}{10}\right) = 1 - \Phi\left(\frac{\mu}{10}\right) = 0.067. \end{aligned} \tag{2}$$

It follows that $\Phi(\mu/10) = 0.933$ and from the table we see that $\mu/10 = 1.5$ or $\mu = 15$.

(c) From the problem statement,

$$P[Y \le 10] = P\left[\frac{Y-\mu}{\sigma} \le \frac{10-\mu}{\sigma}\right] = \Phi\left(\frac{10-\mu}{\sigma}\right) = 0.977. \tag{3}$$

From the $\Phi(\cdot)$ table,

$$\frac{10-\mu}{\sigma} = 2 \qquad \Longrightarrow \qquad \sigma = 5 - \mu/2. \tag{4}$$

(d) Here we are not told the standard deviation σ. However, since

$$P[Y \le 5] = 1 - P[Y > 5] = 1/2 \tag{5}$$

and Y is a Gaussian (μ, σ), we know that

$$P[Y \le 5] = P\left[\frac{Y-\mu}{\sigma} \le \frac{5-\mu}{\sigma}\right] = \Phi\left(\frac{5-\mu}{\sigma}\right) = \frac{1}{2}. \tag{6}$$

Since $\Phi(0) = 1/2$, we see that

$$\Phi\left(\frac{5-\mu}{\sigma}\right) = \Phi(0) \qquad \Longrightarrow \qquad \mu = 5. \tag{7}$$

Problem 4.6.5 Solution

Your body temperature T (in degrees Fahrenheit) satisfies

$$\mathrm{P}\left[T > 100\right] = \mathrm{P}\left[\frac{T-98.6}{0.4} > \frac{100-98.6}{0.4}\right] = Q(3.5) = 2.32 \times 10^{-4}. \tag{1}$$

According to this model, if you were to record your body temperature every day for 10,000 days (over 27 years), you would expect to measure a temperature over 100 perhaps 2 or 3 times. This seems very low since a 100 degree body temperature is a mild and not uncommon fever. What this suggests is that this is a good model for when you are healthy but is not a good model for when you are sick. When you are healthy, 2×10^{-4} might be a reasonable value for the probability of an elevated temperature. However, when you are sick, you need a new model for body temperatures such that $\mathrm{P}[T > 100]$ is much higher.

Problem 4.6.6 Solution

Using σ_T to denote the (unknown) standard deviation of T, we can write

$$\begin{aligned}
\mathrm{P}\left[T < 66\right] &= \mathrm{P}\left[\frac{T-68}{\sigma_T} < \frac{66-68}{\sigma_T}\right] \\
&= \Phi\left(\frac{-2}{\sigma_T}\right) = 1 - \Phi\left(\frac{2}{\sigma_T}\right) = 0.1587.
\end{aligned} \tag{1}$$

Thus $\Phi(2/\sigma_T) = 0.8413 = \Phi(1)$. This implies $\sigma_T = 2$ and thus T has variance $\mathrm{Var}[T] = 4$.

Problem 4.6.7 Solution

X is a Gaussian random variable with zero mean but unknown variance. We do know, however, that

$$\mathrm{P}\left[|X| \leq 10\right] = 0.1. \tag{1}$$

We can find the variance $\mathrm{Var}[X]$ by expanding the above probability in terms of the $\Phi(\cdot)$ function.

$$\mathrm{P}\left[-10 \leq X \leq 10\right] = F_X(10) - F_X(-10) = 2\Phi\left(\frac{10}{\sigma_X}\right) - 1. \tag{2}$$

This implies $\Phi(10/\sigma_X) = 0.55$. Using Table 4.2 for the Gaussian CDF, we find that $10/\sigma_X = 0.15$ or $\sigma_X = 66.6$.

Problem 4.6.8 Solution

Repeating Definition 4.11,

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-u^2/2}\, du. \tag{1}$$

Making the substitution $x = u/\sqrt{2}$, we have

$$Q(z) = \frac{1}{\sqrt{\pi}} \int_{z/\sqrt{2}}^\infty e^{-x^2}\, dx = \frac{1}{2}\mathrm{erfc}\left(\frac{z}{\sqrt{2}}\right). \tag{2}$$

Problem 4.6.9 Solution

Moving to Antarctica, we find that the temperature, T is still Gaussian but with variance 225. We also know that with probability 1/2, T exceeds -75 degrees. First we would like to find the mean temperature, and we do so by looking at the second fact.

$$\mathrm{P}\left[T > -75\right] = 1 - \mathrm{P}\left[T \leq -75\right] = 1 - \Phi\left(\frac{-75 - \mu_T}{15}\right) = 1/2 \tag{1}$$

By looking at the table we find that if $\Phi(x) = 1/2$, then $x = 0$. Therefore,

$$\Phi\left(\frac{-75 - \mu_T}{15}\right) = 1/2 \tag{2}$$

implies that $(-75 - \mu_T)/15 = 0$ or $\mu_T = -75$. Now we have a Gaussian T with expected value -75 and standard deviation 15. So we are prepared to answer the following problems:

$$\mathrm{P}[T > 0] = Q\left(\frac{0-(-75)}{15}\right) = Q(5) = 2.87 \times 10^{-7}, \tag{3}$$

$$\begin{aligned} \mathrm{P}[T < -100] = F_T(-100) &= \Phi\left(\frac{-100-(-75)}{15}\right) \\ &= \Phi(-5/3) = 1 - \Phi(5/3) = 0.0478. \end{aligned} \tag{4}$$

Problem 4.6.10 Solution

In this problem, we use Theorem 4.14 and the tables for the Φ and Q functions to answer the questions. Since $\mathrm{E}[Y_{20}] = 40(20) = 800$ and $\mathrm{Var}[Y_{20}] = 100(20) = 2000$, we can write

$$\begin{aligned} \mathrm{P}[Y_{20} > 1000] &= \mathrm{P}\left[\frac{Y_{20}-800}{\sqrt{2000}} > \frac{1000-800}{\sqrt{2000}}\right] \\ &= \mathrm{P}\left[Z > \frac{200}{20\sqrt{5}}\right] = Q(4.47) = 3.91 \times 10^{-6}. \end{aligned} \tag{1}$$

The second part is a little trickier. Since $\mathrm{E}[Y_{25}] = 1000$, we know that the prof will spend around \$1000 in roughly 25 years. However, to be certain with probability 0.99 that the prof spends \$1000 will require more than 25 years. In particular, we know that

$$\begin{aligned} \mathrm{P}[Y_n > 1000] = \mathrm{P}\left[\frac{Y_n - 40n}{\sqrt{100n}} > \frac{1000-40n}{\sqrt{100n}}\right] \\ = 1 - \Phi\left(\frac{100-4n}{\sqrt{n}}\right) = 0.99. \end{aligned} \tag{2}$$

Hence, we must find n such that

$$\Phi\left(\frac{100-4n}{\sqrt{n}}\right) = 0.01. \tag{3}$$

Recall that $\Phi(x) = 0.01$ for a negative value of x. This is consistent with our earlier observation that we would need $n > 25$ corresponding to $100 - 4n < 0$. Thus, we use the identity $\Phi(x) = 1 - \Phi(-x)$ to write

$$\Phi\left(\frac{100-4n}{\sqrt{n}}\right) = 1 - \Phi\left(\frac{4n-100}{\sqrt{n}}\right) = 0.01 \tag{4}$$

Equivalently, we have

$$\Phi\left(\frac{4n - 100}{\sqrt{n}}\right) = 0.99 \tag{5}$$

From the table of the Φ function, we have that $(4n - 100)/\sqrt{n} = 2.33$, or

$$(n - 25)^2 = (0.58)^2 n = 0.3393n. \tag{6}$$

Solving this quadratic yields $n = 28.09$. Hence, only after 28 years are we 99 percent sure that the prof will have spent \$1000. Note that a second root of the quadratic yields $n = 22.25$. This root is not a valid solution to our problem. Mathematically, it is a solution of our quadratic in which we choose the negative root of $\sqrt{n}$. This would correspond to assuming the standard deviation of Y_n is negative.

Problem 4.6.11 Solution

We are given that there are 100,000,000 men in the United States and 23,000 of them are at least 7 feet tall, and the heights of U.S men are independent Gaussian random variables with mean $5'10''$.

(a) Let H denote the height in inches of a U.S male. To find σ_X, we look at the fact that the probability that $P[H \geq 84]$ is the number of men who are at least 7 feet tall divided by the total number of men (the frequency interpretation of probability). Since we measure H in inches, we have

$$P[H \geq 84] = \frac{23{,}000}{100{,}000{,}000} = \Phi\left(\frac{70 - 84}{\sigma_X}\right) = 0.00023. \tag{1}$$

Since $\Phi(-x) = 1 - \Phi(x) = Q(x)$,

$$Q(14/\sigma_X) = 2.3 \cdot 10^{-4}. \tag{2}$$

From Table 4.3, this implies $14/\sigma_X = 3.5$ or $\sigma_X = 4$.

(b) The probability that a randomly chosen man is at least 8 feet tall is

$$P[H \geq 96] = Q\left(\frac{96 - 70}{4}\right) = Q(6.5). \tag{3}$$

Unfortunately, Table 4.3 doesn't include $Q(6.5)$, although it should be apparent that the probability is very small. In fact, MATLAB will calculate $Q(6.5) = 4.0 \times 10^{-11}$.

(c) First we need to find the probability that a man is at least 7'6".

$$\mathrm{P}[H \geq 90] = Q\left(\frac{90-70}{4}\right) = Q(5) \approx 3 \cdot 10^{-7} = \beta. \tag{4}$$

Although Table 4.3 stops at $Q(4.99)$, if you're curious, the exact value is $Q(5) = 2.87 \cdot 10^{-7}$.

Now we can begin to find the probability that no man is at least 7'6". This can be modeled as 100,000,000 repetitions of a Bernoulli trial with parameter $1-\beta$. The probability that no man is at least 7'6" is

$$(1-\beta)^{100{,}000{,}000} = 9.4 \times 10^{-14}. \tag{5}$$

(d) The expected value of N is just the number of trials multiplied by the probability that a man is at least 7'6".

$$\mathrm{E}[N] = 100{,}000{,}000 \cdot \beta = 30. \tag{6}$$

Problem 4.6.12 Solution

This problem is in the wrong section since the $\mathrm{erf}(\cdot)$ function is defined later on in Section 4.8 as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\,du. \tag{1}$$

(a) Since Y is Gaussian $(0, 1/\sqrt{2})$, Y has variance $1/2$ and

$$f_Y(y) = \frac{1}{\sqrt{2\pi(1/2)}} e^{-y^2/[2(1/2)]} = \frac{1}{\sqrt{\pi}} e^{-y^2}. \tag{2}$$

For $y \geq 0$, $F_Y(y) = \int_{-\infty}^{y} f_Y(u)\,du = 1/2 + \int_0^y f_Y(u)\,du$. Substituting $f_Y(u)$ yields

$$F_Y(y) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^y e^{-u^2}\,du = \frac{1}{2} + \mathrm{erf}(y). \tag{3}$$

(b) Since Y is Gaussian $(0, 1/\sqrt{2})$, $Z = \sqrt{2}Y$ is Gaussian with expected value $\mathrm{E}[Z] = \sqrt{2}\,\mathrm{E}[Y] = 0$ and variance $\mathrm{Var}[Z] = 2\,\mathrm{Var}[Y] = 1$. Thus Z is Gaussian $(0, 1)$ and

$$\begin{aligned}\Phi(z) = F_Z(z) &= \mathrm{P}\left[\sqrt{2}Y \le z\right] \\ &= \mathrm{P}\left[Y \le \frac{z}{\sqrt{2}}\right] \\ &= F_Y\left(\frac{z}{\sqrt{2}}\right) = \frac{1}{2} + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right).\end{aligned} \tag{4}$$

Problem 4.6.13 Solution

First we note that since W has an $N[\mu, \sigma^2]$ distribution, the integral we wish to evaluate is

$$I = \int_{-\infty}^{\infty} f_W(w)\, dw = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(w-\mu)^2/2\sigma^2}\, dw. \tag{1}$$

(a) Using the substitution $x = (w-\mu)/\sigma$, we have $dx = dw/\sigma$ and

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx. \tag{2}$$

(b) When we write I^2 as the product of integrals, we use y to denote the other variable of integration so that

$$\begin{aligned}I^2 &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right)\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy.\end{aligned} \tag{3}$$

(c) By changing to polar coordinates, $x^2 + y^2 = r^2$ and $dx\, dy = r\, dr\, d\theta$ so that

$$\begin{aligned}I^2 &= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\, dr\, d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} -e^{-r^2/2}\Big|_0^{\infty}\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1.\end{aligned} \tag{4}$$

Problem 4.6.14 Solution

(a) Let $Y = X - k$. At time t, Y is a Gaussian $(0, \sqrt{t})$ random variable and since $V = Y^+$,

$$\mathrm{E}\left[V\right] = \mathrm{E}\left[Y^+\right] = \int_0^\infty y f_Y(y)\, dy = \frac{1}{\sqrt{2\pi t}} \int_0^\infty y e^{-y^2/2t}\, dy. \tag{1}$$

With the variable substitution $w = y^2/2t$, we have $dw = (y/t)\, dy$ and

$$\mathrm{E}\left[V\right] = \frac{t}{\sqrt{2\pi t}} \int_0^\infty e^{-w}\, dw = \sqrt{\frac{t}{2\pi}}. \tag{2}$$

(b)

$$\mathrm{P}\left[R > 0\right] = \mathrm{P}\left[V - d > 0\right] = \mathrm{P}\left[V > d\right]. \tag{3}$$

Since V is nonnegative, $\mathrm{P}[V > d] = 1$ for $d < 0$. Thus $d_0 \geq 0$ and for $d = d_0 \geq 0$,

$$\begin{aligned} \mathrm{P}\left[R_0 > 0\right] &= \mathrm{P}\left[(X-k)^+ > d_0\right] \\ &= \mathrm{P}\left[X - k > d_0\right] \\ &= \mathrm{P}\left[\frac{X-k}{\sqrt{t}} > \frac{d_0}{\sqrt{t}}\right] = Q\left(\frac{d_0}{\sqrt{t}}\right). \end{aligned} \tag{4}$$

Note that $Q(0) = 1/2$ and thus $d_0 = 0$.

(c) Finding d_1 is even simpler:

$$(0.01)d_1\, \mathrm{E}\left[R\right] = \mathrm{E}\left[V - d_1\right] = \mathrm{E}\left[V\right] - d_1. \tag{5}$$

Thus,

$$d_1 = \frac{\mathrm{E}\left[V\right]}{1.01} = \frac{1}{1.01}\sqrt{\frac{t}{2\pi}}. \tag{6}$$

(d) The strategy "Buy if $d \leq d_0$" reduced to "Buy if $d \leq 0$." That is, you buy the option if someone is willing to give it to you for free or if someone is willing to pay you to take it. Of course, no one will be willing to pay you to take an option. This strategy is too conservative in that the requirement $\mathrm{P}[R > 0] \geq 1/2$ is not possible at any option price $d > 0$ because, under our model, half of the time the option will end up being worthless.

On the other hand, The strategy "Buy if $d \leq d_1$" is at least plausible; it may be that someone will be willing to sell you the option for $d \leq d_1$ dollars. However, whether the strategy is reasonable depends on whether you have accurate knowledge that $\mathrm{E}[V] = \sqrt{t/(2\pi)}$? As the purchaser of the option, you are counting on the future price being sufficiently random that there is a decent chance that $X > k + (1.01)d_1$. This really depends on the variance of X growing linearly with t.

Problem 4.6.15 Solution

This problem is mostly calculus and only a little probability. The result is a famous formula in the analysis of radio systems. From the problem statement, the SNR Y is an exponential $(1/\gamma)$ random variable with PDF

$$f_Y(y) = \begin{cases} (1/\gamma)e^{-y/\gamma} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Thus, from the problem statement, the BER is

$$\begin{aligned} \overline{P}_e = \mathrm{E}\left[P_e(Y)\right] &= \int_{-\infty}^{\infty} Q(\sqrt{2y}) f_Y(y)\, dy \\ &= \int_0^{\infty} Q(\sqrt{2y}) \frac{y}{\gamma} e^{-y/\gamma}\, dy. \end{aligned} \tag{2}$$

Like most integrals with exponential factors, its a good idea to try integration by parts. Before doing so, we recall that if X is a Gaussian $(0,1)$ random variable with CDF $F_X(x)$, then

$$Q(x) = 1 - F_X(x). \tag{3}$$

It follows that $Q(x)$ has derivative

$$Q'(x) = \frac{dQ(x)}{dx} = -\frac{dF_X(x)}{dx} = -f_X(x) = -\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \tag{4}$$

To solve the integral, we use the integration by parts formula

$$\int_a^b u\,dv = uv|_a^b - \int_a^b v\,du, \tag{5}$$

where

$$u = Q(\sqrt{2y}), \qquad dv = \frac{1}{\gamma}e^{-y/\gamma}\,dy, \tag{6}$$

$$du = Q'(\sqrt{2y})\frac{1}{\sqrt{2y}} = -\frac{e^{-y}}{2\sqrt{\pi y}}, \qquad v = -e^{-y/\gamma}. \tag{7}$$

From integration by parts, it follows that

$$\begin{aligned}
\overline{P}_e &= uv|_0^\infty - \int_0^\infty v\,du \\
&= -Q(\sqrt{2y})e^{-y/\gamma}\Big|_0^\infty - \int_0^\infty \frac{1}{\sqrt{y}}e^{-y[1+(1/\gamma)]}\,dy \\
&= 0 + Q(0)e^{-0} - \frac{1}{2\sqrt{\pi}}\int_0^\infty y^{-1/2}e^{-y/\bar{\gamma}}\,dy,
\end{aligned} \tag{8}$$

where $\bar{\gamma} = \gamma/(1+\gamma)$. Next, recalling that $Q(0) = 1/2$ and making the substitution $t = y/\bar{\gamma}$, we obtain

$$\overline{P}_e = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\bar{\gamma}}{\pi}}\int_0^\infty t^{-1/2}e^{-t}\,dt. \tag{9}$$

From Math Fact B.11, we see that the remaining integral is the $\Gamma(z)$ function evaluated $z = 1/2$. Since $\Gamma(1/2) = \sqrt{\pi}$,

$$\overline{P}_e = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\bar{\gamma}}{\pi}}\Gamma(1/2) = \frac{1}{2}\left[1 - \sqrt{\bar{\gamma}}\right] = \frac{1}{2}\left[1 - \sqrt{\frac{\gamma}{1+\gamma}}\right]. \tag{10}$$

Problem 4.6.16 Solution

(a) Let $V = (k - X)^+$ so that $R = g_p(X) = d - V$. This implies

$$\mathrm{E}[R] = d - \mathrm{E}[V], \qquad \mathrm{Var}[R] = \mathrm{Var}[V]. \tag{1}$$

From the problem statement, the PDF of X is

$$f_X(x) = \begin{cases} 1/(2t) & k - t \le x \le k + t, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It follows that

$$\begin{aligned}
\mathrm{E}[V] &= \mathrm{E}\left[(k - X)^+\right] \\
&= \int_{-\infty}^{\infty} (k - x)^+ f_X(x)\, dx \\
&= \int_{k-t}^{k+t} (k - x)^+ \frac{1}{2t}\, dx \\
&= \frac{1}{2t} \int_{k-t}^{k} (k - x)\, dx = -\frac{1}{4t}(k - x)^2 \bigg|_{x=k-t}^{x=k} = \frac{t}{4}.
\end{aligned} \tag{3}$$

and

$$\begin{aligned}
\mathrm{E}\left[V^2\right] = \mathrm{E}\left[((k - X)^+)^2\right] &= \int_{-\infty}^{\infty} ((k - x)^+)^2 f_X(x)\, dx \\
&= \int_{k-t}^{k+t} ((k - x)^+)^2 \frac{1}{2t}\, dx \\
&= \frac{1}{2t} \int_{k-t}^{k} (k - x)^2\, dx \\
&= -\frac{1}{6t}(k - x)^3 \bigg|_{x=k-t}^{x=k} = \frac{t^2}{6}.
\end{aligned} \tag{4}$$

It follows that

$$\mathrm{Var}[V] = \mathrm{E}\left[V^2\right] - (\mathrm{E}[V])^2 = \frac{t^2}{6} - \frac{t^2}{16} = \frac{5t^2}{48}. \tag{5}$$

Finally, we go back to the beginning to write

$$\mathrm{E}\left[R\right] = d - \frac{t}{4}, \qquad \mathrm{Var}[R] = \mathrm{Var}[V] = \frac{5t^2}{48}. \tag{6}$$

(b) For the call let $W = (X-k)^+$ so that $R' = 2d-(V+W)$. Since the PDF $f_X(x)$ is symmetric around $x = k$, one can deduce from symmetry that V and W are identically distributed. Thus $\mathrm{E}[W] = \mathrm{E}[V] = t/4$ and $\mathrm{E}[W^2] = \mathrm{E}[V^2] = t^2/6$. Thus,

$$\mathrm{E}[V+W] = 2\,\mathrm{E}[V] = \frac{t}{2}$$

and

$$\mathrm{E}[R'] = 2d - \mathrm{E}[V+W] = 2d - \frac{t}{2} = 2\,\mathrm{E}[R].$$

Since $R' = 2d - (V+W)$, $\mathrm{Var}[R'] = \mathrm{Var}[V+W]$. This requires us to find $\mathrm{E}[(V+W)^2]$. At this point, we note that V and W are not independent. In fact, if $V > 0$, then $W = 0$ but if $W > 0$, then $V = 0$. This implies that $VW = 0$ and, of course, $\mathrm{E}[VW] = 0$. It then follows that

$$\begin{aligned} \mathrm{E}\left[(V+W)^2\right] &= \mathrm{E}\left[V^2 + 2VW + W^2\right] \\ &= \mathrm{E}\left[V^2\right] + \mathrm{E}\left[W^2\right] \\ &= 2\,\mathrm{E}\left[V^2\right] = \frac{t^2}{3}. \end{aligned} \tag{7}$$

Finally, we can write

$$\begin{aligned} \mathrm{Var}[R'] &= \mathrm{Var}[V+W] \\ &= \mathrm{E}\left[(V+W)^2\right] - \left(\mathrm{E}\left[V+W\right]\right)^2 \\ &= \frac{t^2}{3} - \left(\frac{t}{2}\right)^2 = \frac{t^2}{12}. \end{aligned} \tag{8}$$

(c) What makes the straddle attractive compared to selling either the call or the put is that the expected reward is doubled since $\mathrm{E}[R'] = 2\,\mathrm{E}[R]$ but the variance is cut in half from $\mathrm{Var}[R] = t^2/6$ to $\mathrm{Var}[R'] = t^2/12$. We can interpret this as saying that the profit is more predictable.

Problem 4.6.17 Solution

First we recall that the stock price at time t is X, a uniform $(k-t, k+t)$ random variable. The profit from the straddle is $R' = 2d - (V + W)$ where

$$V = (k - X)^+, \qquad W = (X - k)^+. \tag{1}$$

To find the CDF, we write

$$\begin{aligned} F_{R'}(r) = \mathrm{P}\left[R' \le r\right] &= \mathrm{P}\left[2d - (V + W) \le r\right] \\ &= \mathrm{P}\left[V + W \ge 2d - r\right]. \end{aligned} \tag{2}$$

Since $V + W$ is non-negative,

$$F_{R'}(r) = \mathrm{P}\left[V + W \ge 2d - r\right] = 1, \qquad r \ge 2d. \tag{3}$$

Now we focus on the case $r \le 2d$. Here we observe that $V > 0$ and $W > 0$ are mutually exclusive events. Thus, for $2d - r \ge 0$,

$$F_{R'}(r) = \mathrm{P}\left[V \ge 2d - r\right] + \mathrm{P}\left[W \ge 2d - r\right] = 2\,\mathrm{P}\left[W \ge 2d - r\right]. \tag{4}$$

since W and V are identically distributed. Since $W = (X - k)^+$ and $2d - r \ge 0$,

$$\begin{aligned} \mathrm{P}\left[W \ge 2d - r\right] &= \mathrm{P}\left[(X - k)^+ \ge 2d - r\right] \\ &= \mathrm{P}\left[X - k \ge 2d - r\right] \\ &= \begin{cases} 0 & (2d - r) > t, \\ \frac{t-(2d-r)}{2t} & (2d - r) \le t. \end{cases} \end{aligned} \tag{5}$$

We can combine the above results in the following statement:

$$F_{R'}(r) = 2\,\mathrm{P}\left[W \ge 2d - r\right] = \begin{cases} 0 & r < 2d - t, \\ \frac{t-2d+r}{t} & 2d - t \le r \le 2d, \\ 1 & r \ge 2d. \end{cases} \tag{6}$$

The PDF of R' is

$$f_{R'}(r) = \begin{cases} \frac{1}{t} & 2d - t \le r \le 2d, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

It might appear that this is a good strategy since you may expect to receive a return of $\mathrm{E}[R'] > 0$ dollars; however this is not free because you assume the risk of a significant loss. In a real investment, the PDF of the price X is not bounded and the loss can be very very large. However, in the case of this problem, the bounded PDF for X implies the loss is not so terrible. From part (a), or by examination of the PDF $f_{R'}(r)$, we see that

$$\mathrm{E}[R'] = \frac{4d - t}{2}.$$

Thus $\mathrm{E}[R'] > 0$ if and only if $d > t/4$. In the worst case of $d = t/4$, we observe that R' has a uniform PDF over $(-t/2, t/2)$ and the worst possible loss is $t/2$ dollars. Whether the risk of such a loss is worth taking for an expected return $\mathrm{E}[R']$ would depend mostly on your financial capital and your investment objectives, which were not included in the problem formulation.

Problem 4.7.1 Solution

(a) Using the given CDF

$$\mathrm{P}[X < -1] = F_X\left(-1^-\right) = 0, \tag{1}$$
$$\mathrm{P}[X \leq -1] = F_X(-1) = -1/3 + 1/3 = 0. \tag{2}$$

Where $F_X(-1^-)$ denotes the limiting value of the CDF found by approaching -1 from the left. Likewise, $F_X(-1^+)$ is interpreted to be the value of the CDF found by approaching -1 from the right. We notice that these two probabilities are the same and therefore the probability that X is exactly -1 is zero.

(b)

$$\mathrm{P}[X < 0] = F_X\left(0^-\right) = 1/3, \tag{3}$$
$$\mathrm{P}[X \leq 0] = F_X(0) = 2/3. \tag{4}$$

Here we see that there is a discrete jump at $X = 0$. Approached from the left the CDF yields a value of 1/3 but approached from the right the value is $2/3$. This means that there is a non-zero probability that $X = 0$, in fact that probability is the difference of the two values.

$$\mathrm{P}[X = 0] = \mathrm{P}[X \leq 0] - \mathrm{P}[X < 0] = 2/3 - 1/3 = 1/3. \tag{5}$$

(c)

$$\mathrm{P}[0 < X \leq 1] = F_X(1) - F_X\left(0^+\right) = 1 - 2/3 = 1/3, \tag{6}$$
$$\mathrm{P}[0 \leq X \leq 1] = F_X(1) - F_X\left(0^-\right) = 1 - 1/3 = 2/3. \tag{7}$$

The difference in the last two probabilities above is that the first was concerned with the probability that X was strictly greater then 0, and the second with the probability that X was greater than or equal to zero. Since the the second probability is a larger set (it includes the probability that $X = 0$) it should always be greater than or equal to the first probability. The two differ by the probability that $X = 0$, and this difference is non-zero only when the random variable exhibits a discrete jump in the CDF.

Problem 4.7.2 Solution

Similar to the previous problem we find

(a)

$$\mathrm{P}[X < -1] = F_X\left(-1^-\right) = 0, \qquad \mathrm{P}[X \leq -1] = F_X(-1) = 1/4. \tag{1}$$

Here we notice the discontinuity of value $1/4$ at $x = -1$.

(b)

$$\mathrm{P}[X < 0] = F_X\left(0^-\right) = 1/2, \qquad \mathrm{P}[X \leq 0] = F_X(0) = 1/2. \tag{2}$$

Since there is no discontinuity at $x = 0$, $F_X(0^-) = F_X(0^+) = F_X(0)$.

(c)

$$\mathrm{P}[X > 1] = 1 - \mathrm{P}[X \leq 1] = 1 - F_X(1) = 0, \tag{3}$$
$$\mathrm{P}[X \geq 1] = 1 - \mathrm{P}[X < 1] = 1 - F_X\left(1^-\right) = 1 - 3/4 = 1/4. \tag{4}$$

Again we notice a discontinuity of size $1/4$, here occurring at $x = 1$.

Problem 4.7.3 Solution

(a) By taking the derivative of the CDF $F_X(x)$ given in Problem 4.7.2, we obtain the PDF

$$f_X(x) = \begin{cases} \frac{\delta(x+1)}{4} + 1/4 + \frac{\delta(x-1)}{4} & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The first moment of X is

$$\begin{aligned} \mathrm{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\, dx \\ &= x/4|_{x=-1} + x^2/8\big|_{-1}^{1} + x/4|_{x=1} \\ &= -1/4 + 0 + 1/4 = 0. \end{aligned} \tag{2}$$

(c) The second moment of X is

$$\begin{aligned} \mathrm{E}\left[X^2\right] &= \int_{-\infty}^{\infty} x^2 f_X(x)\, dx \\ &= x^2/4\big|_{x=-1} + x^3/12\big|_{-1}^{1} + x^2/4\big|_{x=1} \\ &= 1/4 + 1/6 + 1/4 = 2/3. \end{aligned} \tag{3}$$

Since $\mathrm{E}[X] = 0$, $\mathrm{Var}[X] = \mathrm{E}[X^2] = 2/3$.

Problem 4.7.4 Solution

The PMF of a Bernoulli random variable with mean p is

$$P_X(x) = \begin{cases} 1-p & x = 0, \\ p & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The corresponding PDF of this discrete random variable is

$$f_X(x) = (1-p)\delta(x) + p\delta(x-1). \tag{2}$$

Problem 4.7.5 Solution

The PMF of a geometric random variable with mean $1/p$ is

$$P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The corresponding PDF is

$$\begin{aligned} f_X(x) &= p\delta(x-1) + p(1-p)\delta(x-2) + \cdots \\ &= \sum_{j=1}^{\infty} p(1-p)^{j-1}\delta(x-j). \end{aligned} \tag{2}$$

Problem 4.7.6 Solution

(a) Since the conversation time cannot be negative, we know that $F_W(w) = 0$ for $w < 0$. The conversation time W is zero iff either the phone is busy, no one answers, or if the conversation time X of a completed call is zero. Let A be the event that the call is answered. Note that the event A^c implies $W = 0$. For $w \geq 0$,

$$\begin{aligned} F_W(w) &= \mathrm{P}[A^c] + \mathrm{P}[A]\, F_{W|A}(w) \\ &= (1/2) + (1/2)F_X(w). \end{aligned} \tag{1}$$

Thus the complete CDF of W is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ 1/2 + (1/2)F_X(w) & w \geq 0 \end{cases} \tag{2}$$

(b) By taking the derivative of $F_W(w)$, the PDF of W is

$$f_W(w) = \begin{cases} 0 & w < 0 \\ (1/2)\delta(w) + (1/2)f_X(w). & w \geq 0. \end{cases} \tag{3}$$

Next, we keep in mind that since X must be nonnegative, $f_X(x) = 0$ for $x < 0$. Hence,

$$f_W(w) = (1/2)\delta(w) + (1/2)f_X(w). \tag{4}$$

(c) From the PDF $f_W(w)$, calculating the moments is straightforward.

$$E[W] = \int_{-\infty}^{\infty} w f_W(w)\,dw = \frac{1}{2}\int_{-\infty}^{\infty} w f_X(w)\,dw = E[X]/2. \tag{5}$$

The second moment is

$$E\left[W^2\right] = \int_{-\infty}^{\infty} w^2 f_W(w)\,dw = \frac{1}{2}\int_{-\infty}^{\infty} w^2 f_X(w)\,dw = E\left[X^2\right]/2. \tag{6}$$

The variance of W is

$$\begin{aligned} \text{Var}[W] &= E\left[W^2\right] - (E[W])^2 = E\left[X^2\right]/2 - (E[X]/2)^2 \\ &= \frac{1}{2}\text{Var}[X] + \frac{1}{4}(E[X])^2. \end{aligned} \tag{7}$$

Problem 4.7.7 Solution

The professor is on time 80 percent of the time and when he is late his arrival time is uniformly distributed between 0 and 300 seconds. The PDF of T, is

$$f_T(t) = \begin{cases} 0.8\delta(t-0) + \frac{0.2}{300} & 0 \le t \le 300, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The CDF can be found be integrating

$$F_T(t) = \begin{cases} 0 & t < -1, \\ 0.8 + \frac{0.2t}{300} & 0 \le t < 300, \\ 1 & t \ge 300. \end{cases} \tag{2}$$

Problem 4.7.8 Solution

Let G denote the event that the throw is good, that is, no foul occurs. The CDF of D obeys

$$F_D(y) = P[D \le y|G]\,P[G] + P[D \le y|G^c]\,P[G^c] \tag{1}$$

Given the event G,

$$P[D \le y|G] = P[X \le y - 60] = 1 - e^{-(y-60)/10} \quad (y \ge 60) \tag{2}$$

Of course, for $y < 60$, $\mathrm{P}[D \le y|G] = 0$. From the problem statement, if the throw is a foul, then $D = 0$. This implies

$$\mathrm{P}\left[D \le y|G^c\right] = u(y) \tag{3}$$

where $u(\cdot)$ denotes the unit step function. Since $\mathrm{P}[G] = 0.7$, we can write

$$\begin{aligned} F_D(y) &= \mathrm{P}\left[G\right]\mathrm{P}\left[D \le y|G\right] + \mathrm{P}\left[G^c\right]\mathrm{P}\left[D \le y|G^c\right] \\ &= \begin{cases} 0.3u(y) & y < 60, \\ 0.3 + 0.7(1 - e^{-(y-60)/10}) & y \ge 60. \end{cases} \end{aligned} \tag{4}$$

Another way to write this CDF is

$$F_D(y) = 0.3u(y) + 0.7u(y-60)(1 - e^{-(y-60)/10}) \tag{5}$$

However, when we take the derivative, either expression for the CDF will yield the PDF. However, taking the derivative of the first expression perhaps may be simpler:

$$f_D(y) = \begin{cases} 0.3\delta(y) & y < 60, \\ 0.07e^{-(y-60)/10} & y \ge 60. \end{cases} \tag{6}$$

Taking the derivative of the second expression for the CDF is a little tricky because of the product of the exponential and the step function. However, applying the usual rule for the differentation of a product does give the correct answer:

$$\begin{aligned} f_D(y) &= 0.3\delta(y) + 0.7\delta(y-60)(1 - e^{-(y-60)/10}) + 0.07u(y-60)e^{-(y-60)/10} \\ &= 0.3\delta(y) + 0.07u(y-60)e^{-(y-60)/10}. \end{aligned} \tag{7}$$

The middle term $\delta(y-60)(1 - e^{-(y-60)/10})$ dropped out because at $y = 60$, $e^{-(y-60)/10} = 1$.

Problem 4.7.9 Solution

The professor is on time and lectures the full 80 minutes with probability 0.7. In terms of math,

$$\mathrm{P}\left[T = 80\right] = 0.7. \tag{1}$$

Likewise when the professor is more than 5 minutes late, the students leave and a 0 minute lecture is observed. Since he is late 30% of the time and given that he is late, his arrival is uniformly distributed between 0 and 10 minutes, the probability that there is no lecture is

$$P[T = 0] = (0.3)(0.5) = 0.15 \tag{2}$$

The only other possible lecture durations are uniformly distributed between 75 and 80 minutes, because the students will not wait longer then 5 minutes, and that probability must add to a total of $1 - 0.7 - 0.15 = 0.15$. So the PDF of T can be written as

$$f_T(t) = \begin{cases} 0.15\delta(t) & t = 0, \\ 0.03 & 75 \leq 7 < 80, \\ 0.7\delta(t-80) & t = 80, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 4.8.1 Solution

Taking the derivative of the CDF $F_Y(y)$ in Quiz 4.2, we obtain

$$f_Y(y) = \begin{cases} 1/4 & 0 \leq y \leq 4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We see that Y is a uniform $(0,4)$ random variable. By Theorem 6.3, if X is a uniform $(0,1)$ random variable, then $Y = 4X$ is a uniform $(0,4)$ random variable. Using `rand` as MATLAB's uniform $(0,1)$ random variable, the program `quiz31rv` is essentially a one line program:

```
function y=quiz31rv(m)
%Usage y=quiz31rv(m)
%Returns the vector y holding m
%samples of the uniform (0,4) random
%variable Y of Quiz 3.1
y=4*rand(m,1);
```

Problem 4.8.2 Solution

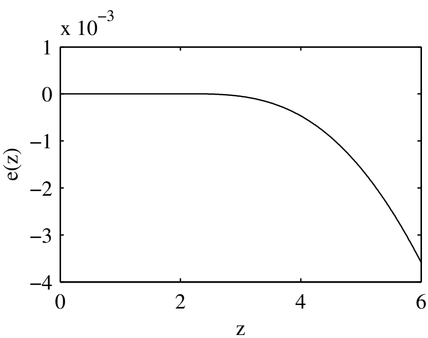
The code for $\hat{Q}(z)$ is the MATLAB function

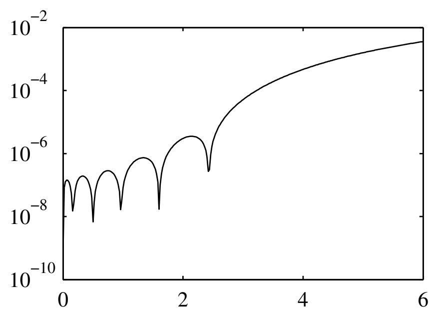
```
function p=qapprox(z);
%approximation  to the Gaussian
% (0,1) complementary CDF Q(z)
t=1./(1.0+(0.231641888.*z(:)));
a=[0.127414796; -0.142248368; 0.7107068705; ...
        -0.7265760135; 0.5307027145];
p=([t t.^2 t.^3 t.^4 t.^5]*a).*exp(-(z(:).^2)/2);
```

This code generates two plots of the relative error $e(z)$ as a function of z:

```
z=0:0.02:6;
q=1.0-phi(z(:));
qhat=qapprox(z);
e=(q-qhat)./q;
plot(z,e); figure;
semilogy(z,abs(e));
```

Here are the output figures of `qtest.m`:





The left side plot graphs $e(z)$ versus z. It appears that the $e(z) = 0$ for $z \leq 3$. In fact, $e(z)$ is nonzero over that range, but the relative error is so small that it isn't visible in comparison to $e(6) \approx -3.5 \times 10^{-3}$. To see the error for small z, the right hand graph plots $|e(z)|$ versus z in log scale where we observe very small relative errors on the order of 10^{-7}.

Problem 4.8.3 Solution

By Theorem 4.9, if X is an exponential (λ) random variable, then $K = \lceil X \rceil$ is a geometric (p) random variable with $p = 1 - e^{-\lambda}$. Thus, given p, we can write $\lambda = -\ln(1-p)$ and $\lceil X \rceil$ is a geometric (p) random variable. Here is the MATLAB function that implements this technique:

```
function k=georv(p,m);
lambda= -log(1-p);
k=ceil(exponentialrv(lambda,m));
```

To compare this technique with that use in `geometricrv.m`, we first examine the code for `exponentialrv.m`:

```
function x=exponentialrv(lambda,m)
x=-(1/lambda)*log(1-rand(m,1));
```

To analyze how $m = 1$ random sample is generated, let $R =$ `rand(1,1)`. In terms of mathematics, `exponentialrv(lambda,1)` generates the random variable

$$X = -\frac{\ln(1-R)}{\lambda} \tag{1}$$

For $\lambda = -\ln(1-p)$, we have that

$$K = \lceil X \rceil = \left\lceil \frac{\ln(1-R)}{\ln(1-p)} \right\rceil \tag{2}$$

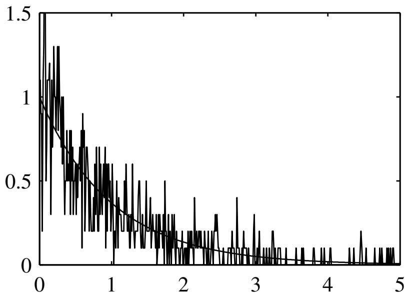
This is precisely the same function implemented by `geometricrv.m`. In short, the two methods for generating geometric (p) random samples are one in the same.

Problem 4.8.4 Solution

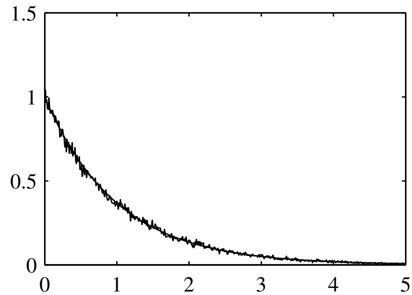
(a) To test the exponential random variables, the following code

```
function exponentialtest(lambda,n)
delta=0.01;
x=exponentialrv(lambda,n);
xr=(0:delta:(5.0/lambda))';
fxsample=(histc(x,xr)/(n*delta));
fx=exponentialpdf(lambda,xr);
plot(xr,fx,xr,fxsample);
```

generates n samples of an exponential λ random variable and plots the relative frequency $n_i/(n\Delta)$ against the corresponding exponential PDF. Note that the `histc` function generates a histogram using `xr` to define the edges of the bins. Two representative plots for $n = 1{,}000$ and $n = 100{,}000$ samples appear in the following figure:



`exponentialtest(1,1000)`



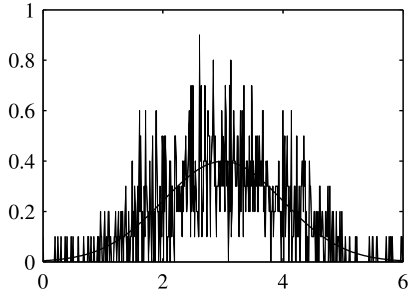
`exponentialtest(1,100000)`

For $n = 1{,}000$, the jaggedness of the relative frequency occurs because δ is sufficiently small that the number of sample of X in each bin $i\Delta < X \leq (i+1)\Delta$ is fairly small. For $n = 100{,}000$, the greater smoothness of the curve demonstrates how the relative frequency is becoming a better approximation to the actual PDF.

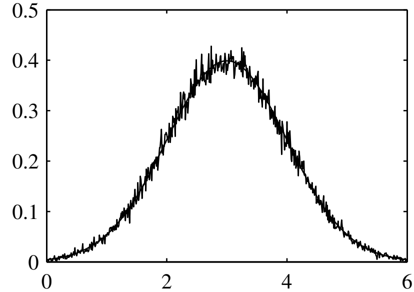
(b) Similar results hold for Gaussian random variables. The following code generates the same comparison between the Gaussian PDF and the relative frequency of n samples.

```
function gausstest(mu,sigma2,n)
delta=0.01;
x=gaussrv(mu,sigma2,n);
xr=(0:delta:(mu+(3*sqrt(sigma2))))';
fxsample=(histc(x,xr)/(n*delta));
fx=gausspdf(mu,sigma2,xr);
plot(xr,fx,xr,fxsample);
```

Here are two typical plots produced by `gaussiantest.m`:



`gausstest(3,1,1000)`



`gausstest(3,1,100000)`

Problem Solutions – Chapter 5

Problem 5.1.1 Solution

(a) The probability P$[X \leq 2, Y \leq 3]$ can be found be evaluating the joint CDF $F_{X,Y}(x,y)$ at $x = 2$ and $y = 3$. This yields

$$\mathrm{P}[X \leq 2, Y \leq 3] = F_{X,Y}(2,3) = (1 - e^{-2})(1 - e^{-3}) \tag{1}$$

(b) To find the marginal CDF of X, $F_X(x)$, we simply evaluate the joint CDF at $y = \infty$.

$$F_X(x) = F_{X,Y}(x,\infty) = \begin{cases} 1 - e^{-x} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) Likewise for the marginal CDF of Y, we evaluate the joint CDF at $X = \infty$.

$$F_Y(y) = F_{X,Y}(\infty,y) = \begin{cases} 1 - e^{-y} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 5.1.2 Solution

(a) Because the probability that any random variable is less than $-\infty$ is zero, we have

$$F_{X,Y}(x,-\infty) = \mathrm{P}[X \leq x, Y \leq -\infty] \leq \mathrm{P}[Y \leq -\infty] = 0 \tag{1}$$

(b) The probability that any random variable is less than infinity is always one.

$$F_{X,Y}(x,\infty) = \mathrm{P}[X \leq x, Y \leq \infty] = \mathrm{P}[X \leq x] = F_X(x)\,. \tag{2}$$

(c) Although P$[Y \leq \infty] = 1$, P$[X \leq -\infty] = 0$. Therefore the following is true.

$$F_{X,Y}(-\infty,\infty) = \mathrm{P}[X \leq -\infty, Y \leq \infty] \leq \mathrm{P}[X \leq -\infty] = 0. \tag{3}$$

(d) Part (d) follows the same logic as that of part (a).

$$F_{X,Y}(-\infty, y) = \mathrm{P}\left[X \leq -\infty, Y \leq y\right] \leq \mathrm{P}\left[X \leq -\infty\right] = 0. \tag{4}$$

(e) Analogous to Part (b), we find that

$$F_{X,Y}(\infty, y) = \mathrm{P}\left[X \leq \infty, Y \leq y\right] = \mathrm{P}\left[Y \leq y\right] = F_Y(y). \tag{5}$$

Problem 5.1.3 Solution

We wish to find $\mathrm{P}[x_1 \leq X \leq x_2 \cup y_1 \leq Y \leq y_2]$. We define events

$$A = \{x_1 \leq X \leq x_2\}, \qquad B = \{y_1 \leq Y \leq y_2\} \tag{1}$$

so that $\mathrm{P}[A \cup B]$ is the probability of observing an X, Y pair in the "cross" region. By Theorem 1.4(c),

$$\mathrm{P}\left[A \cup B\right] = \mathrm{P}\left[A\right] + \mathrm{P}\left[B\right] - \mathrm{P}\left[AB\right] \tag{2}$$

Keep in mind that the intersection of events A and B are all the outcomes such that both A and B occur, specifically, $AB = \{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\}$. It follows that

$$\begin{aligned} \mathrm{P}\left[A \cup B\right] = \mathrm{P}\left[x_1 \leq X \leq x_2\right] &+ \mathrm{P}\left[y_1 \leq Y \leq y_2\right] \\ &- \mathrm{P}\left[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\right]. \end{aligned} \tag{3}$$

By Theorem 5.2,

$$\begin{aligned} &\mathrm{P}\left[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\right] \\ &\quad = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1). \end{aligned} \tag{4}$$

Expressed in terms of the marginal and joint CDFs,

$$\begin{aligned} \mathrm{P}\left[A \cup B\right] = F_X(x_2) - F_X(x_1) &+ F_Y(y_2) - F_Y(y_1) \\ &- F_{X,Y}(x_2, y_2) + F_{X,Y}(x_2, y_1) \\ &+ F_{X,Y}(x_1, y_2) - F_{X,Y}(x_1, y_1). \end{aligned} \tag{5}$$

Problem 5.1.4 Solution

Its easy to show that the properties of Theorem 5.1 are satisfied. However, those properties are necessary but not sufficient to show $F(x, y)$ is a CDF. To convince ourselves that $F(x, y)$ is a valid CDF, we show that for all $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$\mathrm{P}\left[x_1 < X_1 \leq x_2, y_1 < Y \leq y_2\right] \geq 0. \tag{1}$$

In this case, for $x_1 \leq x_2$ and $y_1 \leq y_2$, Theorem 5.2 yields

$$\begin{aligned}
\mathrm{P}\left[x_1 < X \leq x_2, y_1 < Y \leq y_2\right] &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \\
&= F_X(x_2)\, F_Y(y_2) - F_X(x_1)\, F_Y(y_2) \\
&\qquad - F_X(x_2)\, F_Y(y_1) + F_X(x_1)\, F_Y(y_1) \\
&= [F_X(x_2) - F_X(x_1)][F_Y(y_2) - F_Y(y_1)] \\
&\geq 0. \qquad (2)
\end{aligned}$$

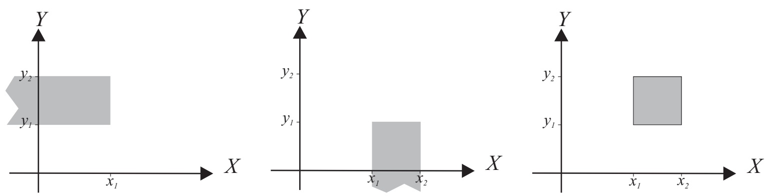
Hence, $F_X(x)F_Y(y)$ is a valid joint CDF.

Problem 5.1.5 Solution

In this problem, we prove Theorem 5.2 which states

$$\begin{aligned}
\mathrm{P}\left[x_1 < X \leq x_2, y_1 < Y \leq y_2\right] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) \\
- F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1). \qquad (1)
\end{aligned}$$

(a) The events A, B, and C are



A B C (2)

(b) In terms of the joint CDF $F_{X,Y}(x, y)$, we can write

$$\mathrm{P}[A] = F_{X,Y}(x_1, y_2) - F_{X,Y}(x_1, y_1), \tag{3}$$

$$\mathrm{P}[B] = F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_1), \tag{4}$$

$$\mathrm{P}[A \cup B \cup C] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_1). \tag{5}$$

(c) Since A, B, and C are mutually exclusive,

$$\mathrm{P}[A \cup B \cup C] = \mathrm{P}[A] + \mathrm{P}[B] + \mathrm{P}[C]. \tag{6}$$

However, since we want to express

$$\mathrm{P}[C] = \mathrm{P}[x_1 < X \leq x_2, y_1 < Y \leq y_2] \tag{7}$$

in terms of the joint CDF $F_{X,Y}(x,y)$, we write

$$\begin{aligned} \mathrm{P}[C] &= \mathrm{P}[A \cup B \cup C] - \mathrm{P}[A] - \mathrm{P}[B] \\ &= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1), \end{aligned} \tag{8}$$

which completes the proof of the theorem.

Problem 5.1.6 Solution

The given function is

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-(x+y)} & x, y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

First, we find the CDF $F_X(x)$ and $F_Y(y)$.

$$F_X(x) = F_{X,Y}(x, \infty) = \begin{cases} 1 & x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$F_Y(y) = F_{X,Y}(\infty, y) = \begin{cases} 1 & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Hence, for any $x \geq 0$ or $y \geq 0$,

$$\mathrm{P}[X > x] = 0, \qquad \mathrm{P}[Y > y] = 0. \tag{4}$$

For $x \geq 0$ and $y \geq 0$, this implies

$$\mathrm{P}[\{X > x\} \cup \{Y > y\}] \leq \mathrm{P}[X > x] + \mathrm{P}[Y > y] = 0. \tag{5}$$

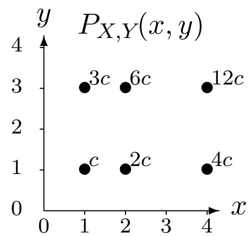
However,

$$\begin{aligned} \mathrm{P}[\{X > x\} \cup \{Y > y\}] &= 1 - \mathrm{P}[X \leq x, Y \leq y] \\ &= 1 - (1 - e^{-(x+y)}) = e^{-(x+y)}. \end{aligned} \tag{6}$$

Thus, we have the contradiction that $e^{-(x+y)} \leq 0$ for all $x, y \geq 0$. We can conclude that the given function is not a valid CDF.

Problem 5.2.1 Solution

In this problem, it is helpful to label points with nonzero probability on the X, Y plane:



(a) We must choose c so the PMF sums to one:

$$\begin{aligned}\sum_{x=1,2,4}\sum_{y=1,3} P_{X,Y}(x,y) &= c\sum_{x=1,2,4} x \sum_{y=1,3} y \\ &= c\,[1(1+3)+2(1+3)+4(1+3)] = 28c. \end{aligned} \tag{1}$$

Thus $c = 1/28$.

(b) The event $\{Y < X\}$ has probability

$$\begin{aligned}\mathrm{P}[Y < X] &= \sum_{x=1,2,4}\sum_{y<x} P_{X,Y}(x,y) \\ &= \frac{1(0)+2(1)+4(1+3)}{28} = \frac{18}{28}. \end{aligned} \tag{2}$$

(c) The event $\{Y > X\}$ has probability

$$\begin{aligned}\mathrm{P}[Y > X] &= \sum_{x=1,2,4}\sum_{y>x} P_{X,Y}(x,y) \\ &= \frac{1(3)+2(3)+4(0)}{28} = \frac{9}{28}. \end{aligned} \tag{3}$$

(d) There are two ways to solve this part. The direct way is to calculate

$$\mathrm{P}[Y = X] = \sum_{x=1,2,4} \sum_{y=x} P_{X,Y}(x,y) = \frac{1(1) + 2(0)}{28} = \frac{1}{28}. \tag{4}$$

The indirect way is to use the previous results and the observation that

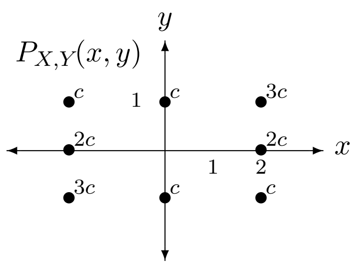
$$\begin{aligned} \mathrm{P}[Y = X] &= 1 - \mathrm{P}[Y < X] - \mathrm{P}[Y > X] \\ &= 1 - 18/28 - 9/28 = 1/28. \end{aligned} \tag{5}$$

(e)

$$\begin{aligned} \mathrm{P}[Y = 3] &= \sum_{x=1,2,4} P_{X,Y}(x,3) \\ &= \frac{(1)(3) + (2)(3) + (4)(3)}{28} = \frac{21}{28} = \frac{3}{4}. \end{aligned} \tag{6}$$

Problem 5.2.2 Solution

On the X, Y plane, the joint PMF is



(a) To find c, we sum the PMF over all possible values of X and Y. We choose c so the sum equals one.

$$\sum_x \sum_y P_{X,Y}(x,y) = \sum_{x=-2,0,2} \sum_{y=-1,0,1} c\,|x+y| = 6c + 2c + 6c = 14c. \tag{1}$$

Thus $c = 1/14$.

(b)

$$\begin{aligned}
\text{P}[Y < X] &= P_{X,Y}(0,-1) + P_{X,Y}(2,-1) + P_{X,Y}(2,0) + P_{X,Y}(2,1) \\
&= c + c + 2c + 3c = 7c = 1/2. \qquad (2)
\end{aligned}$$

(c)

$$\begin{aligned}
\text{P}[Y > X] &= P_{X,Y}(-2,-1) + P_{X,Y}(-2,0) + P_{X,Y}(-2,1) + P_{X,Y}(0,1) \\
&= 3c + 2c + c + c = 7c = 1/2. \qquad (3)
\end{aligned}$$

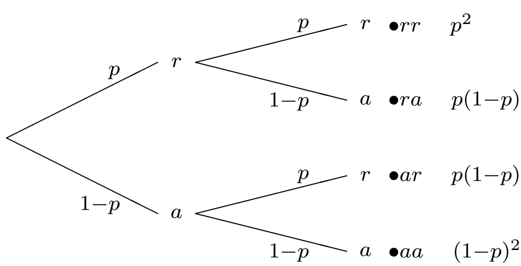
(d) From the sketch of $P_{X,Y}(x,y)$ given above, $\text{P}[X = Y] = 0$.

(e)

$$\begin{aligned}
\text{P}[X < 1] = P_{X,Y}(-2,-1) &+ P_{X,Y}(-2,0) + P_{X,Y}(-2,1) \\
&+ P_{X,Y}(0,-1) + P_{X,Y}(0,1) \\
&= 8c = 8/14. \qquad (4)
\end{aligned}$$

Problem 5.2.3 Solution

Let r (reject) and a (accept) denote the result of each test. There are four possible outcomes: rr, ra, ar, aa. The sample tree is



Now we construct a table that maps the sample outcomes to values of X and Y.

outcome	$\text{P}[\cdot]$	X	Y
rr	p^2	1	1
ra	$p(1-p)$	1	0
ar	$p(1-p)$	0	1
aa	$(1-p)^2$	0	0

(1)

This table is esentially the joint PMF $P_{X,Y}(x,y)$.

$$P_{X,Y}(x,y) = \begin{cases} p^2 & x=1, y=1, \\ p(1-p) & x=0, y=1, \\ p(1-p) & x=1, y=0, \\ (1-p)^2 & x=0, y=0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 5.2.4 Solution

The sample space is the set $S = \{hh, ht, th, tt\}$ and each sample point has probability 1/4. Each sample outcome specifies the values of X and Y as given in the following table

outcome	X	Y
hh	0	1
ht	1	0
th	1	1
tt	2	0

(1)

The joint PMF can represented by the table

$P_{X,Y}(x,y)$	$y=0$	$y=1$
$x=0$	0	1/4
$x=1$	1/4	1/4
$x=2$	1/4	0

(2)

Problem 5.2.5 Solution

As the problem statement says, reasonable arguments can be made for the labels being X and Y or x and y. As we see in the arguments below, the lowercase choice of the text is somewhat arbitrary.

- *Lowercase axis labels:* For the lowercase labels, we observe that we are depicting the masses associated with the joint PMF $P_{X,Y}(x,y)$ whose arguments are x and y. Since the PMF function is defined in terms of x and y, the axis labels should be x and y.

- *Uppercase axis labels:* On the other hand, we are depicting the possible outcomes (labeled with their respective probabilities) of the pair of random variables X and Y. The corresponding axis labels should be X and Y just as in Figure 5.2. The fact that we have labeled the possible outcomes by their probabilities is irrelevant. Further, since the expression for the PMF $P_{X,Y}(x,y)$ given in the figure could just as well have been written $P_{X,Y}(\cdot,\cdot)$, it is clear that the lowercase x and y are not what matter.

Problem 5.2.6 Solution

As the problem statement indicates, $Y = y < n$ if and only if

A: the first y tests are acceptable, and

B: test $y+1$ is a rejection.

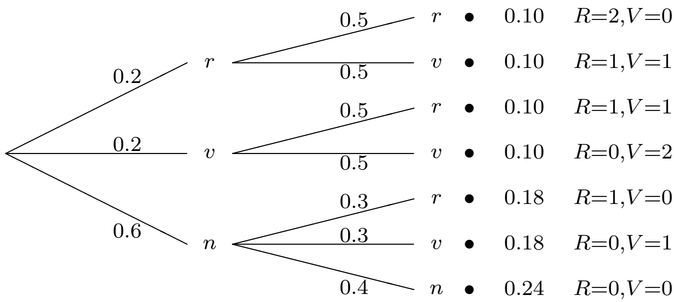
Thus $\mathrm{P}[Y = y] = \mathrm{P}[AB]$. Note that $Y \leq X$ since the number of acceptable tests before the first failure cannot exceed the number of acceptable circuits. Moreover, given the occurrence of AB, the event $X = x < n$ occurs if and only if there are $x - y$ acceptable circuits in the remaining $n - y - 1$ tests. Since events A, B and C depend on disjoint sets of tests, they are independent events. Thus, for $0 \leq y \leq x < n$,

$$\begin{aligned} P_{X,Y}(x,y) &= \mathrm{P}\left[X = x, Y = y\right] \\ &= \mathrm{P}\left[ABC\right] \\ &= \mathrm{P}\left[A\right]\mathrm{P}\left[B\right]\mathrm{P}\left[C\right] \\ &= \underbrace{p^y}_{\mathrm{P}[A]}\underbrace{(1-p)}_{\mathrm{P}[B]}\underbrace{\binom{n-y-1}{x-y}p^{x-y}(1-p)^{n-y-1-(x-y)}}_{\mathrm{P}[C]} \\ &= \binom{n-y-1}{x-y}p^{x}(1-p)^{n-x}. \end{aligned} \tag{1}$$

When all n tests are acceptable, $y = x = n$. Thus $P_{X,Y}(n,n) = p^n$.

Problem 5.2.7 Solution

(a) Using r, v, and n to denote the events that (r) Rutgers scores, (v) Villanova scores, and (n) neither scores in a particular minute, the tree is:



From the leaf probabilities, we can write down the joint PMF of R and V, taking care to note that the pair $R = 1, V = 1$ occurs for more than one outcome.

$P_{R,V}(r, v)$	$v = 0$	$v = 1$	$v = 2$
$r = 0$	0.24	0.18	0.1
$r = 1$	0.18	0.2	0
$r = 2$	0.1	0	0

(b)

$$\begin{aligned} \mathrm{P}[T] = \mathrm{P}[R = V] &= \sum_i P_{R,V}(i, i) \\ &= P_{R,V}(0, 0) + P_{R,V}(1, 1) \\ &= 0.24 + 0.2 = 0.44. \end{aligned} \tag{1}$$

(c) By summing across the rows of the table for $P_{R,V}(r, v)$, we obtain $P_R(0) = 0.52$, $P_R(1) = 0.38$, and $P_R(r)2 = 0.1$. The complete expresson for the marginal PMF is

r	0	1	2
$P_R(r)$	0.52	0.38	0.10

(d) For each pair (R, V), we have $G = R + V$. From first principles.

$$P_G(0) = P_{R,V}(0,0) = 0.24, \tag{2}$$
$$P_G(1) = P_{R,V}(1,0) + P_{R,V}(0,1) = 0.36, \tag{3}$$
$$P_G(2) = P_{R,V}(2,0) + P_{R,V}(1,1) + P_{R,V}(0,2) = 0.4. \tag{4}$$

The complete expression is

$$P_G(g) = \begin{cases} 0.24 & g = 0, \\ 0.36 & g = 1, \\ 0.40 & g = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 5.2.8 Solution

The joint PMF of X and K is $P_{K,X}(k,x) = \mathrm{P}[K = k, X = x]$, which is the probability that $K = k$ and $X = x$. This means that both events must be satisfied. The approach we use is similar to that used in finding the Pascal PMF in Example 3.13. Since X can take on only the two values 0 and 1, let's consider each in turn. When $X = 0$ that means that a rejection occurred on the last test and that the other $k-1$ rejections must have occurred in the previous $n-1$ tests. Thus,

$$P_{K,X}(k,0) = \binom{n-1}{k-1}(1-p)^{k-1}p^{n-1-(k-1)}(1-p), \qquad k = 1, \ldots, n. \tag{1}$$

When $X = 1$ the last test was acceptable and therefore we know that the $K = k \leq n-1$ tails must have occurred in the previous $n-1$ tests. In this case,

$$P_{K,X}(k,1) = \binom{n-1}{k}(1-p)^{k}p^{n-1-k}p, \qquad k = 0, \ldots, n-1. \tag{2}$$

We can combine these cases into a single complete expression for the joint PMF.

$$P_{K,X}(k,x) = \begin{cases} \binom{n-1}{k-1}(1-p)^k p^{n-k} & x = 0, k = 1, 2, \ldots, n, \\ \binom{n-1}{k}(1-p)^k p^{n-k} & x = 1, k = 0, 1, \ldots, n-1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 5.2.9 Solution

Each circuit test produces an acceptable circuit with probability p. Let K denote the number of rejected circuits that occur in n tests and X is the number of acceptable circuits before the first reject. The joint PMF, $P_{K,X}(k,x) = \mathrm{P}[K = k, X = x]$ can be found by realizing that $\{K = k, X = x\}$ occurs if and only if the following events occur:

A The first x tests must be acceptable.

B Test $x+1$ must be a rejection since otherwise we would have $x+1$ acceptable at the beginnning.

C The remaining $n-x-1$ tests must contain $k-1$ rejections.

Since the events A, B and C are independent, the joint PMF for $x + k \leq r$, $x \geq 0$ and $k \geq 0$ is

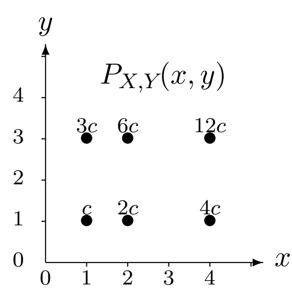
$$P_{K,X}(k,x) = \underbrace{p^x}_{\mathrm{P}[A]} \underbrace{(1-p)}_{\mathrm{P}[B]} \underbrace{\binom{n-x-1}{k-1}(1-p)^{k-1}p^{n-x-1-(k-1)}}_{\mathrm{P}[C]} \tag{1}$$

After simplifying, a complete expression for the joint PMF is

$$P_{K,X}(k,x) = \begin{cases} \binom{n-x-1}{k-1} p^{n-k}(1-p)^k & x+k \leq n, x \geq 0, k \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 5.3.1 Solution

On the X, Y plane, the joint PMF $P_{X,Y}(x,y)$ is



By choosing $c = 1/28$, the PMF sums to one.

(a) The marginal PMFs of X and Y are

$$P_X(x) = \sum_{y=1,3} P_{X,Y}(x,y) = \begin{cases} 4/28 & x = 1, \\ 8/28 & x = 2, \\ 16/28 & x = 4, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

$$P_Y(y) = \sum_{x=1,2,4} P_{X,Y}(x,y) = \begin{cases} 7/28 & y = 1, \\ 21/28 & y = 3, \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

(b) The expected values of X and Y are

$$\mathrm{E}[X] = \sum_{x=1,2,4} x P_X(x) = (4/28) + 2(8/28) + 4(16/28) = 3, \tag{3}$$

$$\mathrm{E}[Y] = \sum_{y=1,3} y P_Y(y) = 7/28 + 3(21/28) = 5/2. \tag{4}$$

(c) The second moments are

$$\begin{aligned} \mathrm{E}\left[X^2\right] &= \sum_{x=1,2,4} x P_X(x) \\ &= 1^2(4/28) + 2^2(8/28) + 4^2(16/28) = 73/7, \end{aligned} \tag{5}$$

$$\mathrm{E}\left[Y^2\right] = \sum_{y=1,3} y P_Y(y) = 1^2(7/28) + 3^2(21/28) = 7. \tag{6}$$

The variances are

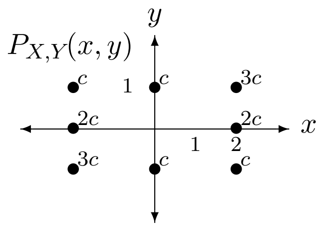
$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 10/7, \tag{7}$$

$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 3/4. \tag{8}$$

The standard deviations are $\sigma_X = \sqrt{10/7}$ and $\sigma_Y = \sqrt{3/4}$.

Problem 5.3.2 Solution

On the X, Y plane, the joint PMF is



The PMF sums to one when $c = 1/14$.

(a) The marginal PMFs of X and Y are

$$P_X(x) = \sum_{y=-1,0,1} P_{X,Y}(x,y) = \begin{cases} 6/14 & x = -2, 2, \\ 2/14 & x = 0, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

$$P_Y(y) = \sum_{x=-2,0,2} P_{X,Y}(x,y) = \begin{cases} 5/14 & y = -1, 1, \\ 4/14 & y = 0, \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

(b) The expected values of X and Y are

$$\mathrm{E}[X] = \sum_{x=-2,0,2} x P_X(x) = -2(6/14) + 2(6/14) = 0, \tag{3}$$

$$\mathrm{E}[Y] = \sum_{y=-1,0,1} y P_Y(y) = -1(5/14) + 1(5/14) = 0. \tag{4}$$

(c) Since X and Y both have zero mean, the variances are

$$\begin{aligned}
\operatorname{Var}[X] = \mathrm{E}\left[X^2\right] &= \sum_{x=-2,0,2} x^2 P_X(x) \\
&= (-2)^2(6/14) + 2^2(6/14) = 24/7, && (5)\\
\operatorname{Var}[Y] = \mathrm{E}\left[Y^2\right] &= \sum_{y=-1,0,1} y^2 P_Y(y) \\
&= (-1)^2(5/14) + 1^2(5/14) = 5/7. && (6)
\end{aligned}$$

The standard deviations are $\sigma_X = \sqrt{24/7}$ and $\sigma_Y = \sqrt{5/7}$.

Problem 5.3.3 Solution

We recognize that the given joint PMF is written as the product of two marginal PMFs $P_N(n)$ and $P_K(k)$ where

$$P_N(n) = \sum_{k=0}^{100} P_{N,K}(n,k) = \begin{cases} \frac{100^n e^{-100}}{n!} & n = 0, 1, \ldots, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$P_K(k) = \sum_{n=0}^{\infty} P_{N,K}(n,k) = \begin{cases} \binom{100}{k} p^k (1-p)^{100-k} & k = 0, 1, \ldots, 100, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 5.3.4 Solution

For integers $0 \le x \le 5$, the marginal PMF of X is

$$P_X(x) = \sum_y P_{X,Y}(x,y) = \sum_{y=0}^{x} (1/21) = \frac{x+1}{21}. \tag{1}$$

Similarly, for integers $0 \le y \le 5$, the marginal PMF of Y is

$$P_Y(y) = \sum_x P_{X,Y}(x,y) = \sum_{x=y}^{5} (1/21) = \frac{6-y}{21}. \tag{2}$$

The complete expressions for the marginal PMFs are

$$P_X(x) = \begin{cases} (x+1)/21 & x = 0, \ldots, 5, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$P_Y(y) = \begin{cases} (6-y)/21 & y = 0, \ldots, 5, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The expected values are

$$\mathrm{E}\left[X\right] = \sum_{x=0}^{5} x \frac{x+1}{21} = \frac{70}{21} = \frac{10}{3}, \tag{5}$$

$$\mathrm{E}\left[Y\right] = \sum_{y=0}^{5} y \frac{6-y}{21} = \frac{35}{21} = \frac{5}{3}. \tag{6}$$

Problem 5.3.5 Solution

The joint PMF of N, K is

$$P_{N,K}(n,k) = \begin{cases} (1-p)^{n-1} p/n & k = 1, 2, \ldots, n, \\ & n = 1, 2 \ldots, \\ o & \text{otherwise.} \end{cases} \tag{1}$$

For $n \geq 1$, the marginal PMF of N is

$$P_N(n) = \sum_{k=1}^{n} P_{N,K}(n,k) = \sum_{k=1}^{n} (1-p)^{n-1} p/n = (1-p)^{n-1} p. \tag{2}$$

The marginal PMF of K is found by summing $P_{N,K}(n,k)$ over all possible N. Note that if $K = k$, then $N \geq k$. Thus,

$$P_K(k) = \sum_{n=k}^{\infty} \frac{1}{n} (1-p)^{n-1} p. \tag{3}$$

Unfortunately, this sum cannot be simplified.

Problem 5.3.6 Solution

For $n = 0, 1, \ldots$, the marginal PMF of N is

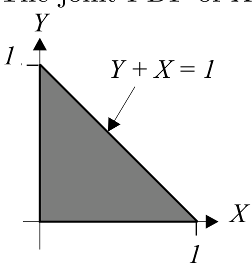
$$P_N(n) = \sum_k P_{N,K}(n,k) = \sum_{k=0}^{n} \frac{100^n e^{-100}}{(n+1)!} = \frac{100^n e^{-100}}{n!}. \tag{1}$$

For $k = 0, 1, \ldots$, the marginal PMF of K is

$$\begin{aligned} P_K(k) = \sum_{n=k}^{\infty} \frac{100^n e^{-100}}{(n+1)!} &= \frac{1}{100} \sum_{n=k}^{\infty} \frac{100^{n+1} e^{-100}}{(n+1)!} \\ &= \frac{1}{100} \sum_{n=k}^{\infty} P_N(n+1) \\ &= \mathrm{P}[N > k] / 100. \end{aligned} \tag{2}$$

Problem 5.4.1 Solution

(a) The joint PDF of X and Y is



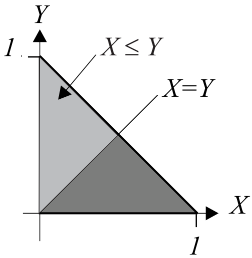
$$f_{X,Y}(x,y) = \begin{cases} c & x + y \leq 1, x, y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

To find the constant c we integrate over the region shown. This gives

$$\int_0^1 \int_0^{1-x} c\, dy\, dx = cx - \frac{cx}{2}\bigg|_0^1 = \frac{c}{2} = 1. \tag{1}$$

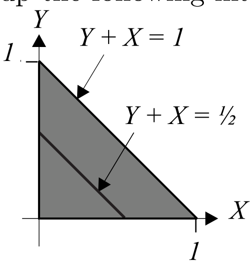
Therefore $c = 2$.

(b) To find the $\mathrm{P}[X \leq Y]$ we look to integrate over the area indicated by the graph



$$P[X \leq Y] = \int_0^{1/2} \int_x^{1-x} dy\, dx$$
$$= \int_0^{1/2} (2-4x)\, dx$$
$$= 1/2. \tag{2}$$

(c) The probability $P[X+Y \leq 1/2]$ can be seen in the figure. Here we can set up the following integrals



$$P[X+Y \leq 1/2] = \int_0^{1/2} \int_0^{1/2-x} 2\, dy\, dx$$
$$= \int_0^{1/2} (1-2x)\, dx$$
$$= 1/2 - 1/4 = 1/4. \tag{3}$$

Problem 5.4.2 Solution

We are given the joint PDF

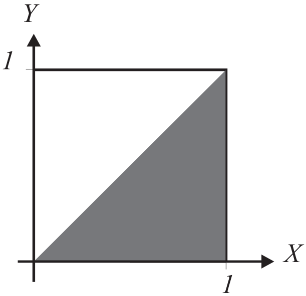
$$f_{X,Y}(x,y) = \begin{cases} cxy^2 & 0 \leq x, y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) To find the constant c integrate $f_{X,Y}(x,y)$ over the all possible values of X and Y to get

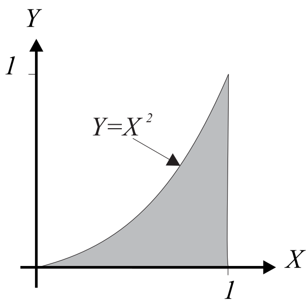
$$1 = \int_0^1 \int_0^1 cxy^2\, dx\, dy = c/6. \tag{2}$$

Therefore $c = 6$.

(b) The probability $P[X \geq Y]$ is the integral of the joint PDF $f_{X,Y}(x,y)$ over the indicated shaded region.



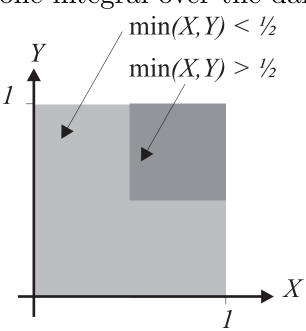
$$
\begin{aligned}
\mathrm{P}\left[X \geq Y\right] &= \int_0^1 \int_0^x 6xy^2 \, dy \, dx \\
&= \int_0^1 2x^4 \, dx \\
&= 2/5. \qquad (3)
\end{aligned}
$$



Similarly, to find $\mathrm{P}[Y \leq X^2]$ we can integrate over the region shown in the figure.

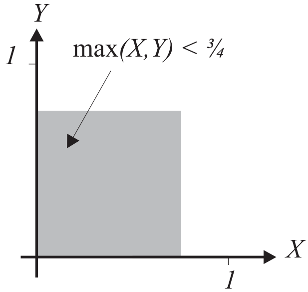
$$
\begin{aligned}
\mathrm{P}\left[Y \leq X^2\right] &= \int_0^1 \int_0^{x^2} 6xy^2 \, dy \, dx \\
&= 1/4. \qquad (4)
\end{aligned}
$$

(c) Here we can choose to either integrate $f_{X,Y}(x,y)$ over the lighter shaded region, which would require the evaluation of two integrals, or we can perform one integral over the darker region by recognizing



$$
\begin{aligned}
\mathrm{P}\left[\min(X,Y) \leq 1/2\right] &= 1 - \mathrm{P}\left[\min(X,Y) > 1/2\right] \\
&= 1 - \int_{1/2}^1 \int_{1/2}^1 6xy^2 \, dx \, dy \\
&= 1 - \int_{1/2}^1 \frac{9y^2}{4} \, dy \\
&= \frac{11}{32}. \qquad (5)
\end{aligned}
$$

(d) The probability $\mathrm{P}[\max(X,Y) \leq 3/4]$ can be found be integrating over the shaded region shown below.



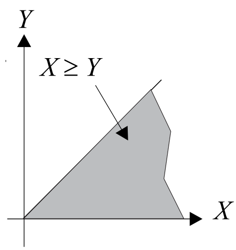
$$
\begin{aligned}
P[\max(X,Y) \leq 3/4] &= P[X \leq 3/4, Y \leq 3/4] \\
&= \int_0^{\frac{3}{4}} \int_0^{\frac{3}{4}} 6xy^2 \, dx \, dy \\
&= \left(x^2\Big|_0^{3/4}\right)\left(y^3\Big|_0^{3/4}\right) \\
&= (3/4)^5 = 0.237. \qquad (6)
\end{aligned}
$$

Problem 5.4.3 Solution

The joint PDF of X and Y is

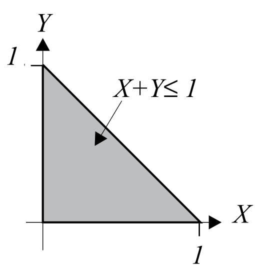
$$
f_{X,Y}(x,y) = \begin{cases} 6e^{-(2x+3y)} & x \geq 0, y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)
$$

(a) The probability that $X \geq Y$ is:



$$
\begin{aligned}
P[X \geq Y] &= \int_0^\infty \int_0^x 6e^{-(2x+3y)} \, dy \, dx \\
&= \int_0^\infty 2e^{-2x}\left(-e^{-3y}\Big|_{y=0}^{y=x}\right) dx \\
&= \int_0^\infty [2e^{-2x} - 2e^{-5x}] \, dx = 3/5. \qquad (2)
\end{aligned}
$$

The probability $P[X+Y \leq 1]$ is found by integrating over the region where $X + Y \leq 1$:



$$
\begin{aligned}
P[X+Y \leq 1] &= \int_0^1 \int_0^{1-x} 6e^{-(2x+3y)} \, dy \, dx \\
&= \int_0^1 2e^{-2x}\left[-e^{-3y}\Big|_{y=0}^{y=1-x}\right] dx \\
&= \int_0^1 2e^{-2x}\left[1 - e^{-3(1-x)}\right] dx \\
&= -e^{-2x} - 2e^{x-3}\Big|_0^1 \\
&= 1 + 2e^{-3} - 3e^{-2}. \qquad (3)
\end{aligned}
$$

(b) The event $\{\min(X,Y) \geq 1\}$ is the same as the event $\{X \geq 1, Y \geq 1\}$. Thus,

$$\mathrm{P}\left[\min(X,Y) \geq 1\right] = \int_1^\infty \int_1^\infty 6e^{-(2x+3y)}\, dy\, dx = e^{-(2+3)}. \tag{4}$$

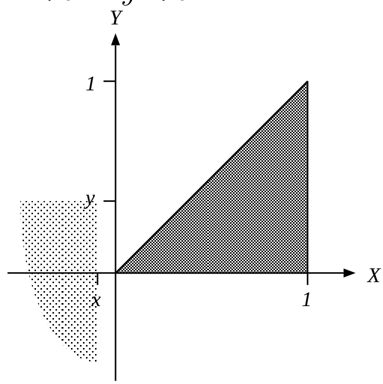
(c) The event $\{\max(X,Y) \leq 1\}$ is the same as the event $\{X \leq 1, Y \leq 1\}$ so that

$$\mathrm{P}\left[\max(X,Y) \leq 1\right] = \int_0^1 \int_0^1 6e^{-(2x+3y)}\, dy\, dx = (1-e^{-2})(1-e^{-3}). \tag{5}$$

Problem 5.4.4 Solution

The only difference between this problem and Example 5.8 is that in this problem we must integrate the joint PDF over the regions to find the probabilities. Just as in Example 5.8, there are five cases. We will use variable u and v as dummy variables for x and y.

- $x < 0$ or $y < 0$

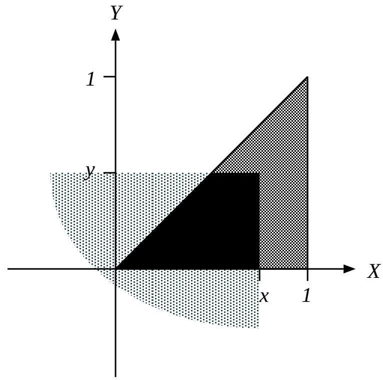


In this case, the region of integration doesn't overlap the region of nonzero probability and

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\; du\, dv \\ &= 0. \end{aligned} \tag{1}$$

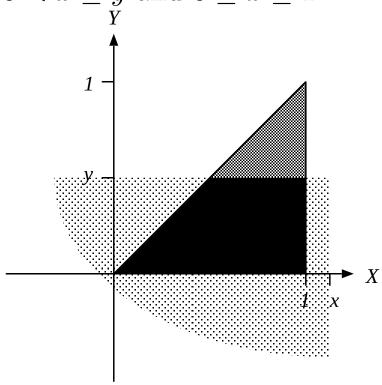
- $0 < y \leq x \leq 1$

  In this case, the region where the integral has a nonzero contribution is



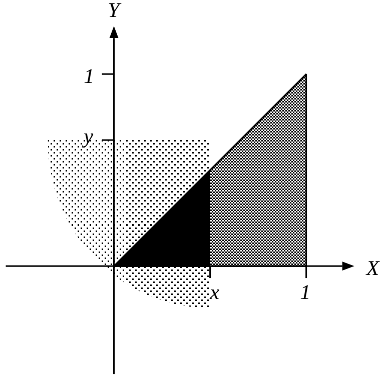
$$
\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(u,v)\;dy\,dx \\
&= \int_{0}^{y}\int_{v}^{x} 8uv\,du\,dv \\
&= \int_{0}^{y} 4(x^2 - v^2)v\,dv \\
&= 2x^2v^2 - v^4\Big|_{v=0}^{v=y} \\
&= 2x^2y^2 - y^4. \qquad (2)
\end{aligned}
$$

- $0 < x \leq y$ and $0 \leq x \leq 1$



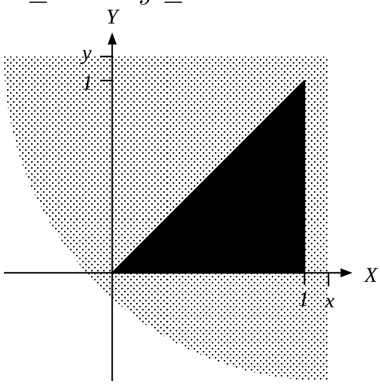
$$
\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(u,v)\;dv\,du \\
&= \int_{0}^{x}\int_{0}^{u} 8uv\,dv\,du \\
&= \int_{0}^{x} 4u^3\,du = x^4. \qquad (3)
\end{aligned}
$$

- $0 < y \leq 1$ and $x \geq 1$



$$
\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\; dv\, du \\
&= \int_{0}^{y} \int_{v}^{1} 8uv\, du\, dv \\
&= \int_{0}^{y} 4v(1-v^2)\, dv \\
&= 2y^2 - y^4. \qquad (4)
\end{aligned}
$$

- $x \geq 1$ and $y \geq 1$



In this case, the region of integration completely covers the region of nonzero probability and

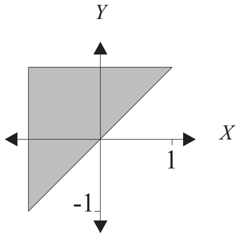
$$
\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\; du\, dv \\
&= 1. \qquad (5)
\end{aligned}
$$

The complete answer for the joint CDF is

$$
F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0, \\ 2x^2y^2 - y^4 & 0 < y \leq x \leq 1, \\ x^4 & 0 \leq x \leq y, 0 \leq x \leq 1, \\ 2y^2 - y^4 & 0 \leq y \leq 1, x \geq 1, \\ 1 & x \geq 1, y \geq 1. \end{cases} \qquad (6)
$$

Problem 5.5.1 Solution

The joint PDF (and the corresponding region of nonzero probability) are



$$f_{X,Y}(x,y) = \begin{cases} 1/2 & -1 \leq x \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a)

$$\mathrm{P}[X > 0] = \int_0^1 \int_x^1 \frac{1}{2}\, dy\, dx = \int_0^1 \frac{1-x}{2}\, dx = 1/4 \tag{2}$$

This result can be deduced by geometry. The shaded triangle of the X, Y plane corresponding to the event $X > 0$ is $1/4$ of the total shaded area.

(b) For $x > 1$ or $x < -1$, $f_X(x) = 0$. For $-1 \leq x \leq 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy = \int_x^1 \frac{1}{2}\, dy = (1-x)/2. \tag{3}$$

The complete expression for the marginal PDF is

$$f_X(x) = \begin{cases} (1-x)/2 & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(c) From the marginal PDF $f_X(x)$, the expected value of X is

$$\begin{aligned} \mathrm{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\; dx = \frac{1}{2} \int_{-1}^{1} x(1-x)\, dx \\ &= \frac{x^2}{4} - \frac{x^3}{6}\bigg|_{-1}^{1} = -\frac{1}{3}. \end{aligned} \tag{5}$$

Problem 5.5.2 Solution

(a) The integral of the PDF over all x, y must be unity. Thus

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx\, dy \\ &= \int_0^1 \int_0^1 cx\, dx\, dy = c \left(\int_0^1 x\, dx \right) \left(\int_0^1 dy \right) = \frac{c}{2}. \end{aligned} \tag{1}$$

Thus $c = 2$.

(b) The marginal PDF of X is $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$. This integral is zero for $x < 0$ or $x > 1$. For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy = \int_0^1 2x\, dy = 2x. \tag{2}$$

Thus,

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(c) To check independence, we check whether $f_{X,Y}(x, y)$ equals $f_X(x) f_Y(y)$. Since $0 \le Y \le 1$, we know that $f_Y(y) = 0$ for $y < 0$ or $y > 1$. For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx = \int_0^1 2x\, dx = x^2\Big|_0^1 = 1. \tag{4}$$

Thus Y has the uniform PDF

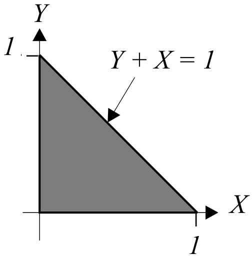
$$f_Y(y) = \begin{cases} 1 & 0 \le y \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

and we see that $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. Thus X and Y are independent.

Problem 5.5.3 Solution

X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2 & x + y \le 1, x, y \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$



Using the figure to the left we can find the marginal PDFs by integrating over the appropriate regions.

$$f_X(x) = \int_0^{1-x} 2\,dy = \begin{cases} 2(1-x) & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Likewise for $f_Y(y)$:

$$f_Y(y) = \int_0^{1-y} 2\,dx = \begin{cases} 2(1-y) & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 5.5.4 Solution

Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/(\pi r^2) & 0 \leq x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The marginal PDF of X is

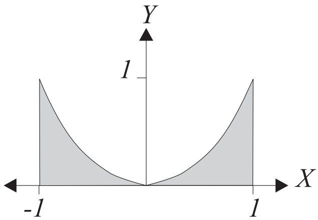
$$f_X(x) = 2\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2}\,dy = \begin{cases} \frac{2\sqrt{r^2-x^2}}{\pi r^2} & -r \leq x \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) Similarly, for random variable Y,

$$f_Y(y) = 2\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{1}{\pi r^2}\,dx = \begin{cases} \frac{2\sqrt{r^2-y^2}}{\pi r^2} & -r \leq y \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 5.5.5 Solution

The joint PDF of X and Y and the region of nonzero probability are



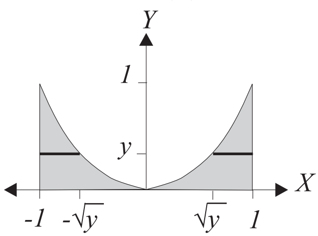
$$f_{X,Y}(x,y) = \begin{cases} 5x^2/2 & -1 \leq x \leq 1, 0 \leq y \leq x^2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We can find the appropriate marginal PDFs by integrating the joint PDF.

(a) The marginal PDF of X is

$$f_X(x) = \int_0^{x^2} \frac{5x^2}{2}\, dy = \begin{cases} 5x^4/2 & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) Note that $f_Y(y) = 0$ for $y > 1$ or $y < 0$. For $0 \leq y \leq 1$,



$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \\ &= \int_{-1}^{-\sqrt{y}} \frac{5x^2}{2}\, dx + \int_{\sqrt{y}}^{1} \frac{5x^2}{2}\, dx \\ &= 5(1 - y^{3/2})/3. \end{aligned} \tag{3}$$

The complete expression for the marginal CDF of Y is

$$f_Y(y) = \begin{cases} 5(1 - y^{3/2})/3 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 5.5.6 Solution

In this problem, the joint PDF is

$$f_{X,Y}(x,y) = \begin{cases} 2\,|xy|\,/r^4 & 0 \leq x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) Since $|xy| = |x||y|$, for $-r \le x \le r$, we can write

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \frac{2|x|}{r^4} \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} |y|\, dy. \tag{2}$$

Since $|y|$ is symmetric about the origin, we can simplify the integral to

$$f_X(x) = \frac{4|x|}{r^4} \int_0^{\sqrt{r^2-x^2}} y\, dy = \frac{2|x|}{r^4} y^2 \Big|_0^{\sqrt{r^2-x^2}} = \frac{2|x|(r^2 - x^2)}{r^4}. \tag{3}$$

Note that for $|x| > r$, $f_X(x) = 0$. Hence the complete expression for the PDF of X is

$$f_X(x) = \begin{cases} \frac{2|x|(r^2-x^2)}{r^4} & -r \le x \le r, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(b) Note that the joint PDF is symmetric in x and y so that $f_Y(y) = f_X(y)$.

Problem 5.5.7 Solution

First, we observe that Y has mean $\mu_Y = a\mu_X + b$ and variance $\text{Var}[Y] = a^2 \text{Var}[X]$. The covariance of X and Y is

$$\begin{aligned} \text{Cov}[X,Y] &= \text{E}[(X-\mu_X)(aX+b-a\mu_X-b)] \\ &= a\,\text{E}\left[(X-\mu_X)^2\right] \\ &= a\,\text{Var}[X]. \end{aligned} \tag{1}$$

The correlation coefficient is

$$\rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}} = \frac{a\,\text{Var}[X]}{\sqrt{\text{Var}[X]}\sqrt{a^2\,\text{Var}[X]}} = \frac{a}{|a|}. \tag{2}$$

When $a > 0$, $\rho_{X,Y} = 1$. When $a < 0$, $\rho_{X,Y} = -1$.

Problem 5.5.8 Solution

The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} (x+y)/3 & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) We first find the marginal PDFs of X and Y. For $0 \leq x \leq 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$
$$= \int_0^2 \frac{x+y}{3}\, dy = \frac{xy}{3} + \frac{y^2}{6}\bigg|_{y=0}^{y=2} = \frac{2x+2}{3}. \tag{2}$$

For $0 \leq y \leq 2$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$$
$$= \int_0^1 \left(\frac{x}{3} + \frac{y}{3}\right) dx = \frac{x^2}{6} + \frac{xy}{3}\bigg|_{x=0}^{x=1} = \frac{2y+1}{6}. \tag{3}$$

Complete expressions for the marginal PDFs are

$$f_X(x) = \begin{cases} \frac{2x+2}{3} & 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} \frac{2y+1}{6} & 0 \leq y \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(b) The expected value of X is

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$
$$= \int_0^1 x\frac{2x+2}{3}\, dx = \frac{2x^3}{9} + \frac{x^2}{3}\bigg|_0^1 = \frac{5}{9}. \tag{5}$$

The second moment of X is

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$
$$= \int_0^1 x^2\frac{2x+2}{3}\, dx = \frac{x^4}{6} + \frac{2x^3}{9}\bigg|_0^1 = \frac{7}{18}. \tag{6}$$

The variance of X is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 7/18 - (5/9)^2 = 13/162. \tag{7}$$

(c) The expected value of Y is

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\, dy \\ &= \int_0^2 y \frac{2y+1}{6}\, dy = \left.\frac{y^2}{12} + \frac{y^3}{9}\right|_0^2 = \frac{11}{9}. \end{aligned} \tag{8}$$

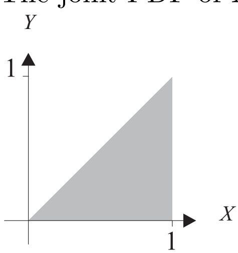
The second moment of Y is

$$\begin{aligned} E\left[Y^2\right] &= \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy \\ &= \int_0^2 y^2 \frac{2y+1}{6}\, dy = \left.\frac{y^3}{18} + \frac{y^4}{12}\right|_0^2 = \frac{16}{9}. \end{aligned} \tag{9}$$

The variance of Y is $\text{Var}[Y] = E[Y^2] - (E[Y])^2 = 23/81$.

Problem 5.5.9 Solution

(a) The joint PDF of X and Y and the region of nonzero probability are



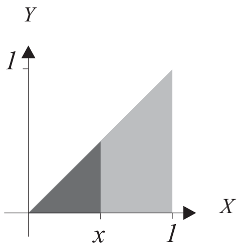
$$f_{X,Y}(x,y) = \begin{cases} cy & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) To find the value of the constant, c, we integrate the joint PDF over all x and y.

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy &= \int_0^1 \int_0^x cy\, dy\, dx = \int_0^1 \frac{cx^2}{2}\, dx \\ &= \left.\frac{cx^3}{6}\right|_0^1 = \frac{c}{6}. \end{aligned} \tag{2}$$

Thus $c = 6$.

(c) We can find the CDF $F_X(x) = P[X \le x]$ by integrating the joint PDF over the event $X \le x$. For $x < 0$, $F_X(x) = 0$. For $x > 1$, $F_X(x) = 1$. For $0 \le x \le 1$,

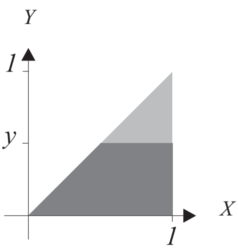


$$F_X(x) = \iint_{x' \le x} f_{X,Y}(x', y')\, dy'\, dx' = \int_0^x \int_0^{x'} 6y'\, dy'\, dx' = \int_0^x 3(x')^2\, dx' = x^3. \tag{3}$$

The complete expression for the joint CDF is

$$F_X(x) = \begin{cases} 0 & x < 0, \\ x^3 & 0 \le x \le 1, \\ 1 & x \ge 1. \end{cases} \tag{4}$$

(d) Similarly, we find the CDF of Y by integrating $f_{X,Y}(x, y)$ over the event $Y \le y$. For $y < 0$, $F_Y(y) = 0$ and for $y > 1$, $F_Y(y) = 1$. For $0 \le y \le 1$,

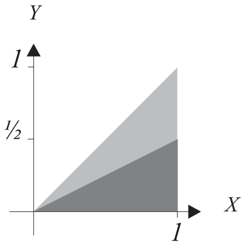


$$F_Y(y) = \iint_{y' \le y} f_{X,Y}(x', y')\, dy'\, dx' = \int_0^y \int_{y'}^1 6y'\, dx'\, dy' = \int_0^y 6y'(1 - y')\, dy' = 3(y')^2 - 2(y')^3\Big|_0^y = 3y^2 - 2y^3. \tag{5}$$

The complete expression for the CDF of Y is

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ 3y^2 - 2y^3 & 0 \le y \le 1, \\ 1 & y > 1. \end{cases} \tag{6}$$

(e) To find $P[Y \leq X/2]$, we integrate the joint PDF $f_{X,Y}(x,y)$ over the region $y \leq x/2$.



$$P[Y \leq X/2] = \int_0^1 \int_0^{x/2} 6y \, dy \, dx = \int_0^1 3y^2 \Big|_0^{x/2} dx = \int_0^1 \frac{3x^2}{4} dx = 1/4. \tag{7}$$

Problem 5.6.1 Solution

The key to this problem is understanding that "small order" and "big order" are synonyms for $W = 1$ and $W = 5$. Similarly, "vanilla", "chocolate", and "strawberry" correspond to the events $D = 20$, $D = 100$ and $D = 300$.

(a) The following table is given in the problem statement.

	vanilla	choc.	strawberry
small order	0.2	0.2	0.2
big order	0.1	0.2	0.1

This table can be translated directly into the joint PMF of W and D.

$P_{W,D}(w,d)$	$d = 20$	$d = 100$	$d = 300$
$w = 1$	0.2	0.2	0.2
$w = 5$	0.1	0.2	0.1

(1)

(b) We find the marginal PMF $P_D(d)$ by summing the columns of the joint PMF. This yields

$$P_D(d) = \begin{cases} 0.3 & d = 20, \\ 0.4 & d = 100, \\ 0.3 & d = 300, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) To check independence, we calculate the marginal PMF

$$P_W(w) = \sum_{d=20,100,300} P_{W,D}(w,d) = \begin{cases} 0.6 & w = 1, \\ 0.4 & w = 5, \end{cases} \tag{3}$$

and we check if $P_{W,D}(w,d) = P_W(w)P_D(d)$. In this case, we see that

$$P_{W,D}(1,20) = 0.2 \neq P_W(1)\,P_D(20) = (0.6)(0.3). \tag{4}$$

Hence W and D are dependent.

Problem 5.6.2 Solution

The key to this problem is understanding that "Factory Q" and "Factory R" are synonyms for $M = 60$ and $M = 180$. Similarly, "small", "medium", and "large" orders correspond to the events $B = 1$, $B = 2$ and $B = 3$.

(a) The following table given in the problem statement

	Factory Q	Factory R
small order	0.3	0.2
medium order	0.1	0.2
large order	0.1	0.1

can be translated into the following joint PMF for B and M.

$P_{B,M}(b,m)$	$m = 60$	$m = 180$
$b = 1$	0.3	0.2
$b = 2$	0.1	0.2
$b = 3$	0.1	0.1

(1)

(b) Before we find $E[B]$, it will prove helpful to find the marginal PMFs $P_B(b)$ and $P_M(m)$. These can be found from the row and column sums of the table of the joint PMF

$P_{B,M}(b,m)$	$m=60$	$m=180$	$P_B(b)$
$b=1$	0.3	0.2	0.5
$b=2$	0.1	0.2	0.3
$b=3$	0.1	0.1	0.2
$P_M(m)$	0.5	0.5	

(2)

The expected number of boxes is

$$\mathrm{E}[B] = \sum_b bP_B(b) = 1(0.5) + 2(0.3) + 3(0.2) = 1.7. \tag{3}$$

(c) From the marginal PMFs we calculated in the table of part (b), we can conclude that B and M are not independent. since $P_{B,M}(1,60) \neq P_B(1)P_M(m)60$.

Problem 5.6.3 Solution

Flip a fair coin 100 times and let X be the number of heads in the first 75 flips and Y be the number of heads in the last 25 flips. We know that X and Y are independent and can find their PMFs easily.

$$P_X(x) = \binom{75}{x}(1/2)^{75}, \qquad P_Y(y) = \binom{25}{y}(1/2)^{25}. \tag{1}$$

The joint PMF of X and N can be expressed as the product of the marginal PMFs because we know that X and Y are independent.

$$P_{X,Y}(x,y) = \binom{75}{x}\binom{25}{y}(1/2)^{100}. \tag{2}$$

Problem 5.6.4 Solution

We can solve this problem for the general case when the probability of heads is p. For the fair coin, $p = 1/2$. Viewing each flip as a Bernoulli trial in which heads is a success, the number of flips until heads is the number of trials needed for the first success which has the geometric PMF

$$P_{X_1}(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Similarly, no matter how large X_1 may be, the number of *additional* flips for the second heads is the same experiment as the number of flips needed for the first occurrence of heads. That is, $P_{X_2}(x) = P_{X_1}(x)$. Moreover, the flips needed to generate the second occurrence of heads are independent of the flips that yield the first heads. Hence, it should be apparent that X_1 and X_2 are independent and

$$\begin{aligned} P_{X_1,X_2}(x_1,x_2) &= P_{X_1}(x_1)\, P_{X_2}(x_2) \\ &= \begin{cases} (1-p)^{x_1+x_2-2}p^2 & x_1 = 1,2,\ldots; x_2 = 1,2,\ldots, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

However, if this independence is not obvious, it can be derived by examination of the sample path. When $x_1 \geq 1$ and $x_2 \geq 1$, the event $\{X_1 = x_1, X_2 = x_2\}$ occurs iff we observe the sample sequence

$$\underbrace{tt\cdots t}_{x_1-1 \text{ times}} \quad h \quad \underbrace{tt\cdots t}_{x_2-1 \text{ times}} \quad h \tag{3}$$

The above sample sequence has probability $(1-p)^{x_1-1}p(1-p)^{x_2-1}p$ which in fact equals $P_{X_1,X_2}(x_1,x_2)$ given earlier.

Problem 5.6.5 Solution

From the problem statement, X and Y have PDFs

$$f_X(x) = \begin{cases} 1/2 & 0 \leq x \leq 2, \\ 0 & \text{otherwise,} \end{cases}, \qquad f_Y(y) = \begin{cases} 1/5 & 0 \leq y \leq 5, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Since X and Y are independent, the joint PDF is

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) = \begin{cases} 1/10 & 0 \leq x \leq 2, 0 \leq y \leq 5, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 5.6.6 Solution

X_1 and X_2 are independent random variables such that X_i has PDF

$$f_{X_i}(x) \begin{cases} \lambda_i e^{-\lambda_i x} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

To calculate $P[X_2 < X_1]$, we use the joint PDF $f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$.

$$\begin{aligned}
P[X_2 < X_1] &= \iint_{x_2<x_1} f_{X_1}(x_1)\, f_{X_2}(x_2)\, dx_1\, dx_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \int_{x_2}^\infty \lambda_1 e^{-\lambda_1 x_1}\, dx_1\, dx_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} e^{-\lambda_1 x_2}\, dx_2 \\
&= \int_0^\infty \lambda_2 e^{-(\lambda_1+\lambda_2)x_2}\, dx_2 = \frac{\lambda_2}{\lambda_1+\lambda_2} \qquad (2)
\end{aligned}$$

Problem 5.6.7 Solution

(a) We find k by the requirement that the joint PDF integrate to 1. That is,

$$\begin{aligned}
1 = \int_{-\infty}^\infty \int_{-\infty}^\infty f_{X,Y}(x,y)\, dx\, dy &= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} (k+3x^2)\, dx\, dy \\
&= \left(\int_{-1/2}^{1/2} dy\right)\left(\int_{-1/2}^{1/2} (k+3x^2)\, dx\right) \\
&= kx + x^3\Big|_{x=-1/2}^{x=1/2} = k + 1/4 \qquad (1)
\end{aligned}$$

Thus k=3/4.

(b) For $-1/2 \le x \le 1/2$, the marginal PDF of X is

$$f_X(x) = \int_{-\infty}^\infty f_{X,Y}(x,y)\, dy = \int_{-1/2}^{1/2} (k+3x^2)\, dy = k + 3x^2. \qquad (2)$$

The complete expression for the PDF of X is

$$f_X(x) = \begin{cases} k+3x^2 & -1/2 \le x \le 1/2, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

(c) For $-1/2 \leq y \leq 1/2$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x)\ dx$$
$$= \int_{-1/2}^{1/2} (k + 3x^2)\, dx = kx + x^3\Big|_{x=-1/2}^{x=1/2} = k + 1/4. \tag{4}$$

Since $k = 3/4$, Y is a continuous uniform $(-1/2, 1/2)$ random variable with PDF

$$f_Y(y) = \begin{cases} 1 & -1/2 \leq y \leq 1/2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

(d) We need to check whether $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. If you solved for k in part (a), then from (b) and (c) it is obvious that this equality holds and thus X and Y are independent. If you were not able to solve for k in part (a), testing whether $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ yields the requirement $1 = k + 1/4$. With some thought, you should have gone back to check that $k = 3/4$ solves part (a). This would lead to the correct conclusion that X and Y are independent.

Problem 5.6.8 Solution

Random variables X_1 and X_2 are iiid with PDF

$$f_X(x) = \begin{cases} x/2 & 0 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) Since X_1 and X_2 are identically distributed they will share the same CDF $F_X(x)$.

$$F_X(x) = \int_0^x f_X(x')\ dx' = \begin{cases} 0 & x \leq 0, \\ x^2/4 & 0 \leq x \leq 2, \\ 1 & x \geq 2. \end{cases} \tag{2}$$

(b) Since X_1 and X_2 are independent, we can say that

$$P[X_1 \leq 1, X_2 \leq 1] = P[X_1 \leq 1]\,P[X_2 \leq 1]$$
$$= F_{X_1}(1)\,F_{X_2}(1) = [F_X(1)]^2 = \frac{1}{16}. \tag{3}$$

(c) For $W = \max(X_1, X_2)$,

$$F_W(1) = P[\max(X_1, X_2) \le 1] = P[X_1 \le 1, X_2 \le 1]. \tag{4}$$

Since X_1 and X_2 are independent,

$$F_W(1) = P[X_1 \le 1] P[X_2 \le 1] = [F_X(1)]^2 = 1/16. \tag{5}$$

(d)

$$F_W(w) = P[\max(X_1, X_2) \le w] = P[X_1 \le w, X_2 \le w]. \tag{6}$$

Since X_1 and X_2 are independent,

$$\begin{aligned} F_W(w) &= P[X_1 \le w] P[X_2 \le w] \\ &= [F_X(w)]^2 = \begin{cases} 0 & w \le 0, \\ w^4/16 & 0 \le w \le 2, \\ 1 & w \ge 2. \end{cases} \end{aligned} \tag{7}$$

Problem 5.6.9 Solution

This problem is quite straightforward. From Theorem 5.5, we can find the joint PDF of X and Y is

$$f_{X,Y}(x, y) = \frac{\partial^2 [F_X(x) F_Y(y)]}{\partial x \, \partial y} = \frac{\partial [f_X(x) F_Y(y)]}{\partial y} = f_X(x) f_Y(y). \tag{1}$$

Hence, $F_{X,Y}(x, y) = F_X(x) F_Y(y)$ implies that X and Y are independent.

If X and Y are independent, then

$$f_{X,Y}(x, y) = f_X(x) f_Y(y). \tag{2}$$

By Definition 5.3,

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) \, dv \, du \\ &= \left(\int_{-\infty}^{x} f_X(u) \, du \right) \left(\int_{-\infty}^{y} f_Y(v) \, dv \right) \\ &= F_X(x) F_X(x). \end{aligned} \tag{3}$$

Problem 5.7.1 Solution

We recall that the joint PMF of W and D is

$$\begin{array}{c|ccc} P_{W,D}(w,d) & d=20 & d=100 & d=300 \\ \hline w=1 & 0.2 & 0.2 & 0.2 \\ \hline w=5 & 0.1 & 0.2 & 0.1 \end{array} \tag{1}$$

In terms of W and D, the cost (in cents) of a shipment is $C = WD$. The expected value of C is

$$\begin{aligned} \mathrm{E}[C] &= \sum_{w,d} wdP_{W,D}(w,d) \\ &= 1(20)(0.2) + 1(100)(0.2) + 1(300)(0.2) \\ &\qquad + 5(20)(0.3) + 5(100)(0.4) + 5(300)(0.3) = 764 \text{ cents.} \end{aligned} \tag{2}$$

Problem 5.7.2 Solution

We recall that B and M have joint PMF

$$\begin{array}{c|cc} P_{B,M}(b,m) & m=60 & m=180 \\ \hline b=1 & 0.3 & 0.2 \\ b=2 & 0.1 & 0.2 \\ b=3 & 0.1 & 0.1 \end{array} \tag{1}$$

In terms of M and B, the cost (in cents) of sending a shipment is $C = BM$. The expected value of C is

$$\begin{aligned} \mathrm{E}[C] &= \sum_{b,m} bmP_{B,M}(b,m) \\ &= 1(60)(0.3) + 2(60)(0.1) + 3(60)(0.1) \\ &\qquad + 1(180)(0.2) + 2(180)(0.2) + 3(180)(0.1) = 210 \text{ cents.} \end{aligned} \tag{2}$$

Problem 5.7.3 Solution

We solve this problem using Theorem 5.9. To do so, it helps to label each pair

X, Y with the sum $W = X + Y$:

$P_{X,Y}(x,y)$	$y=1$	$y=2$	$y=3$	$y=4$
$x=5$	0.05 $_{W=6}$	0.1 $_{W=7}$	0.2 $_{W=8}$	0.05 $_{W=9}$
$x=6$	0.1 $_{W=7}$	0.1 $_{W=8}$	0.3 $_{W=9}$	0.1 $_{W=10}$

It follows that

$$\begin{aligned}\mathrm{E}\left[X+Y\right] &= \sum_{x,y}(x+y)P_{X,Y}(x,y) \\ &= 6(0.05)+7(0.2)+8(0.3)+9(0.35)+10(0.1) = 8.25. \end{aligned} \tag{1}$$

and

$$\begin{aligned}\mathrm{E}\left[(X+Y)^2\right] &= \sum_{x,y}(x+y)^2 P_{X,Y}(x,y) \\ &= 6^2(0.05)+7^2(0.2)+8^2(0.3)+9^2(0.35)+10^2(0.1) \\ &= 69.15. \end{aligned} \tag{2}$$

It follows that

$$\begin{aligned}\mathrm{Var}[X+Y] &= \mathrm{E}\left[(X+Y)^2\right] - \left(\mathrm{E}\left[X+Y\right]\right)^2 \\ &= 69.15 - (8.25)^2 = 1.0875. \end{aligned} \tag{3}$$

An alternate approach would be to find the marginals $P_X(x)$ and $P_Y(y)$ and use these to calculate $\mathrm{E}[X]$, $\mathrm{E}[Y]$, $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$. However, we would still need to find the covariance of X and Y to find the variance of $X+Y$.

Problem 5.7.4 Solution

Using the following probability model

$$P_X(k) = P_Y(k) = \begin{cases} 3/4 & k=0, \\ 1/4 & k=20, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We can calculate the requested moments.

$$\mathrm{E}[X] = 3/4 \cdot 0 + 1/4 \cdot 20 = 5. \tag{2}$$

$$\mathrm{Var}[X] = 3/4 \cdot (0-5)^2 + 1/4 \cdot (20-5)^2 = 75. \tag{3}$$

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[X] = 2\,\mathrm{E}[X] = 10. \tag{4}$$

Since X and Y are independent, Theorem 5.17 yields

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] = 2\,\mathrm{Var}[X] = 150 \tag{5}$$

Since X and Y are independent, $P_{X,Y}(x,y) = P_X(x)P_Y(y)$ and

$$\begin{aligned}\mathrm{E}\left[XY2^{XY}\right] &= \sum_{x=0,20}\sum_{y=0,20} XY2^{XY} P_{X,Y}(x,y) \\ &= (20)(20)2^{20(20)} P_X(20)\, P_Y(20) = 2.75 \times 10^{12}.\end{aligned} \tag{6}$$

Problem 5.7.5 Solution

We start by observing that

$$\mathrm{Cov}[X,Y] = \rho\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = 1.$$

This implies

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y] = 1 + 4 + 2(1) = 7.$$

Problem 5.7.6 Solution

The expected value is

$$\mathrm{E}[W] = \mathrm{E}[2X+2Y] = 2\,\mathrm{E}[X] + 2\,\mathrm{E}[Y] = 2. \tag{1}$$

The variance is

$$\begin{aligned}\mathrm{Var}[W] = \mathrm{Var}[2(X+Y)] &= 4\,\mathrm{Var}[X+Y] \\ &= 4\left(\mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y]\right) \\ &= 4\left(3^2 + 4^2 + 2(-3)\right) = 76.\end{aligned} \tag{2}$$

Problem 5.7.7 Solution

We will solve this problem when the probability of heads is p. For the fair coin, $p = 1/2$. The number X_1 of flips until the first heads and the number X_2 of additional flips for the second heads both have the geometric PMF

$$P_{X_1}(x) = P_{X_2}(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Thus, $\mathrm{E}[X_i] = 1/p$ and $\mathrm{Var}[X_i] = (1-p)/p^2$. By Theorem 5.11,

$$\mathrm{E}[Y] = \mathrm{E}[X_1] - \mathrm{E}[X_2] = 0. \tag{2}$$

Since X_1 and X_2 are independent, Theorem 5.17 says

$$\mathrm{Var}[Y] = \mathrm{Var}[X_1] + \mathrm{Var}[-X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] = \frac{2(1-p)}{p^2}. \tag{3}$$

Problem 5.7.8 Solution

(a) Since $\mathrm{E}[-X_2] = -\mathrm{E}[X_2]$, we can use Theorem 5.10 to write

$$\begin{aligned} \mathrm{E}[X_1 - X_2] = \mathrm{E}[X_1 + (-X_2)] &= \mathrm{E}[X_1] + \mathrm{E}[-X_2] \\ &= \mathrm{E}[X_1] - \mathrm{E}[X_2] \\ &= 0. \end{aligned} \tag{1}$$

(b) By Theorem 4.5(f), $\mathrm{Var}[-X_2] = (-1)^2 \mathrm{Var}[X_2] = \mathrm{Var}[X_2]$. Since X_1 and X_2 are independent, Theorem 5.17(a) says that

$$\begin{aligned} \mathrm{Var}[X_1 - X_2] &= \mathrm{Var}[X_1 + (-X_2)] \\ &= \mathrm{Var}[X_1] + \mathrm{Var}[-X_2] \\ &= 2\,\mathrm{Var}[X]. \end{aligned} \tag{2}$$

Problem 5.7.9 Solution

(a) Since X and Y have zero expected value, $\text{Cov}[X,Y] = \text{E}[XY] = 3$, $\text{E}[U] = a\,\text{E}[X] = 0$ and $\text{E}[V] = b\,\text{E}[Y] = 0$. It follows that

$$\begin{aligned}\text{Cov}[U,V] &= \text{E}[UV] \\ &= \text{E}[abXY] \\ &= ab\,\text{E}[XY] = ab\,\text{Cov}[X,Y] = 3ab. \end{aligned} \tag{1}$$

(b) We start by observing that $\text{Var}[U] = a^2\,\text{Var}[X]$ and $\text{Var}[V] = b^2\,\text{Var}[Y]$. It follows that

$$\begin{aligned}\rho_{U,V} &= \frac{\text{Cov}[U,V]}{\sqrt{\text{Var}[U]\,\text{Var}[V]}} \\ &= \frac{ab\,\text{Cov}[X,Y]}{\sqrt{a^2\,\text{Var}[X] b^2\,\text{Var}[Y]}} = \frac{ab}{\sqrt{a^2b^2}}\rho_{X,Y} = \frac{1}{2}\frac{ab}{|ab|}. \end{aligned} \tag{2}$$

Note that $ab/|ab|$ is 1 if a and b have the same sign or is -1 if they have opposite signs.

(c) Since $\text{E}[X] = 0$,

$$\begin{aligned}\text{Cov}[X,W] &= \text{E}[XW] - \text{E}[X]\,\text{E}[U] \\ &= \text{E}[XW] \\ &= \text{E}[X(aX+bY)] \\ &= a\,\text{E}\left[X^2\right] + b\,\text{E}[XY] \\ &= a\,\text{Var}[X] + b\,\text{Cov}[X,Y]. \end{aligned} \tag{3}$$

Since X and Y are identically distributed, $\text{Var}[X] = \text{Var}[Y]$ and

$$\frac{1}{2} = \rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sqrt{\text{Var}[X]\,\text{Var}[Y]}} = \frac{\text{Cov}[X,Y]}{\text{Var}[X]} = \frac{3}{\text{Var}[X]}. \tag{4}$$

This implies $\text{Var}[X] = 6$. From (3), $\text{Cov}[X,W] = 6a + 3b$. Thus X and W are uncorrelated if $6a + 3b = 0$, or $b = -2a$.

Problem 5.7.10 Solution

FALSE. To generate a counterexample, Suppose Y_1 and Y_2 have correlation coefficient ρ. In this case,

$$\operatorname{Var}[Y_1 + Y_2] = \operatorname{Var}[Y_1] + \operatorname{Var}[Y_2] + 2\operatorname{Cov}[Y_1, Y_2]. \tag{1}$$

Since Y_1 and Y_2 are identically distributed,

$$\operatorname{Var}[Y_2] = \operatorname{Var}[Y_1] \tag{2}$$

and

$$\operatorname{Cov}[Y_1, Y_2] = \rho \operatorname{Var}[Y_1]. \tag{3}$$

Thus,

$$\operatorname{Var}[Y_1 + Y_2] = 2\operatorname{Var}[Y_1] + 2\rho \operatorname{Var}[Y_1] = 2(1+\rho)\operatorname{Var}[Y_1]. \tag{4}$$

We see that $\operatorname{Var}[Y_1 + Y_2] < \operatorname{Var}[Y_1]$ if $2(1+\rho) < 1$, or $\rho < -1/2$.

Problem 5.7.11 Solution

(a) Since $E[V] = E[X] - E[Y] = 0$,

$$\begin{aligned}\operatorname{Var}[V] = E\left[V^2\right] &= E\left[X^2 - 2XY + Y^2\right] \\ &= \operatorname{Var}[X] + \operatorname{Var}[Y] - 2\operatorname{Cov}[X, Y] \\ &= \operatorname{Var}[X] + \operatorname{Var}[Y] - 2\sigma_X\sigma_Y\rho_{X,Y} \\ &= 20 - 16\rho_{X,Y}. \end{aligned} \tag{1}$$

This is minimized when $\rho_{X,Y} = 1$. The minimum possible variance is 4. On other hand, $\operatorname{Var}[V]$ is maximized when $\rho_{X,Y} = -1$; the maximum possible variance is 36.

(b) Since $E[W] = E[X] - E[2Y] = 0$,

$$\begin{aligned}\operatorname{Var}[W] = E\left[W^2\right] &= E\left[X^2 - 4XY + 4Y^2\right] \\ &= \operatorname{Var}[X] + 4\operatorname{Var}[Y] - 4\operatorname{Cov}[X, Y] \\ &= \operatorname{Var}[X] + 4\operatorname{Var}[Y] - 4\sigma_X\sigma_Y\rho_{X,Y} \\ &= 68 - 32\rho_{X,Y}. \end{aligned} \tag{2}$$

This is is minimized when $\rho_{X,Y} = 1$ and maximized when $\rho_{X,Y} = -1$. The minimum and maximum possible variances are 36 and 100.

Problem 5.7.12 Solution

(a) The first moment of X is

$$\mathrm{E}[X] = \int_0^1 \int_0^1 4x^2 y \, dy \, dx = \int_0^1 2x^2 \, dx = \frac{2}{3}. \tag{1}$$

The second moment of X is

$$\mathrm{E}\left[X^2\right] = \int_0^1 \int_0^1 4x^3 y \, dy \, dx = \int_0^1 2x^3 \, dx = \frac{1}{2}. \tag{2}$$

The variance of X is $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 1/2 - (2/3)^2 = 1/18$.

(b) The mean of Y is

$$\mathrm{E}[Y] = \int_0^1 \int_0^1 4xy^2 \, dy \, dx = \int_0^1 \frac{4x}{3} \, dx = \frac{2}{3}. \tag{3}$$

The second moment of Y is

$$\mathrm{E}\left[Y^2\right] = \int_0^1 \int_0^1 4xy^3 \, dy \, dx = \int_0^1 x \, dx = \frac{1}{2}. \tag{4}$$

The variance of Y is

$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 1/2 - (2/3)^2 = 1/18. \tag{5}$$

(c) To find the covariance, we first find the correlation

$$\mathrm{E}[XY] = \int_0^1 \int_0^1 4x^2 y^2 \, dy \, dx = \int_0^1 \frac{4x^2}{3} \, dx = \frac{4}{9}. \tag{6}$$

The covariance is thus

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \frac{4}{9} - \left(\frac{2}{3}\right)^2 = 0. \tag{7}$$

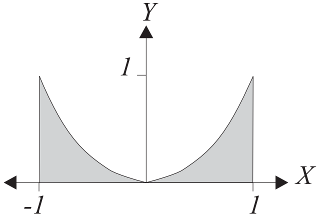
(d) $E[X+Y] = E[X] + E[Y] = 2/3 + 2/3 = 4/3$.

(e) By Theorem 5.12, the variance of $X+Y$ is

$$\text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}[X,Y] = 1/18 + 1/18 + 0 = 1/9. \tag{8}$$

Problem 5.7.13 Solution

The joint PDF of X and Y and the region of nonzero probability are



$$f_{X,Y}(x,y) = \begin{cases} 5x^2/2 & -1 \le x \le 1, 0 \le y \le x^2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The first moment of X is

$$E[X] = \int_{-1}^{1} \int_0^{x^2} x \frac{5x^2}{2}\,dy\,dx = \int_{-1}^{1} \frac{5x^5}{2}\,dx = \left.\frac{5x^6}{12}\right|_{-1}^{1} = 0. \tag{2}$$

Since $E[X] = 0$, the variance of X and the second moment are both

$$\text{Var}[X] = E\left[X^2\right] = \int_{-1}^{1} \int_0^{x^2} x^2 \frac{5x^2}{2}\,dy\,dx = \left.\frac{5x^7}{14}\right|_{-1}^{1} = \frac{10}{14}. \tag{3}$$

(b) The first and second moments of Y are

$$E[Y] = \int_{-1}^{1} \int_0^{x^2} y \frac{5x^2}{2}\,dy\,dx = \frac{5}{14}, \tag{4}$$

$$E\left[Y^2\right] = \int_{-1}^{1} \int_0 x^2 y^2 \frac{5x^2}{2}\,dy\,dx = \frac{5}{26}. \tag{5}$$

Therefore, $\text{Var}[Y] = 5/26 - (5/14)^2 = .0576$.

(c) Since $\mathrm{E}[X] = 0$, $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[XY]$. Thus,

$$\mathrm{Cov}\left[X,Y\right] = \mathrm{E}\left[XY\right] = \int_1^1 \int_0^{x^2} xy \frac{5x^2}{2}\, dy\, dx = \int_{-1}^1 \frac{5x^7}{4}\, dx = 0. \tag{6}$$

(d) The expected value of the sum $X+Y$ is

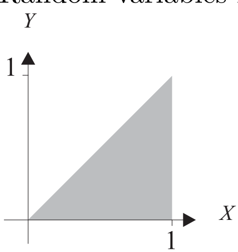
$$\mathrm{E}\left[X+Y\right] = \mathrm{E}\left[X\right] + \mathrm{E}\left[Y\right] = \frac{5}{14}. \tag{7}$$

(e) By Theorem 5.12, the variance of $X+Y$ is

$$\begin{aligned}\mathrm{Var}[X+Y] &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}\left[X,Y\right] \\ &= 5/7 + 0.0576 = 0.7719. \end{aligned} \tag{8}$$

Problem 5.7.14 Solution

Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Before finding moments, it is helpful to first find the marginal PDFs. For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy = \int_0^x 2\, dy = 2x. \tag{2}$$

Note that $f_X(x) = 0$ for $x < 0$ or $x > 1$. For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dx = \int_y^1 2\, dx = 2(1-y). \tag{3}$$

Also, for $y < 0$ or $y > 1$, $f_Y(y) = 0$. Complete expressions for the marginal PDFs are

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} f_Y(y) \qquad = \begin{cases} 2(1-y) & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(a) The first two moments of X are

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\ dx = \int_0^1 2x^2\, dx = 2/3, \tag{5}$$

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\ dx = \int_0^1 2x^3\, dx = 1/2. \tag{6}$$

The variance of X is $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 1/2 - 4/9 = 1/18$.

(b) The expected value and second moment of Y are

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \int_0^1 2y(1-y)\, dy = y^2 - \left.\frac{2y^3}{3}\right|_0^1 = \frac{1}{3}, \tag{7}$$

$$\mathrm{E}\left[Y^2\right] = \int_{-\infty}^{\infty} y^2 f_Y(y)\ dy = \int_0^1 2y^2(1-y)\, dy = \left.\frac{2y^3}{3} - \frac{y^4}{2}\right|_0^1 = \frac{1}{6}. \tag{8}$$

The variance of Y is $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = 1/6 - 1/9 = 1/18$.

(c) Before finding the covariance, we find the correlation

$$\mathrm{E}[XY] = \int_0^1 \int_0^x 2xy\, dy\, dx = \int_0^1 x^3\, dx = 1/4 \tag{9}$$

The covariance is

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 1/36. \tag{10}$$

(d) $\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y] = 2/3 + 1/3 = 1$

(e) By Theorem 5.12,

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X, Y] = 1/6. \tag{11}$$

Problem 5.7.15 Solution

Since $\mathrm{E}[Y] = \mathrm{E}[X] = \mathrm{E}[Z] = 0$, we know that

$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right], \qquad \mathrm{Var}[X] = \mathrm{E}\left[X^2\right], \qquad \mathrm{Var}[Z] = \mathrm{E}\left[Z^2\right], \tag{1}$$

and

$$\operatorname{Cov}[X,Y] = \operatorname{E}[XY] = \operatorname{E}[X(X+Z)] = \operatorname{E}\left[X^2\right] + \operatorname{E}[XZ]. \tag{2}$$

Since X and Z are independent, $\operatorname{E}[XZ] = \operatorname{E}[X]\operatorname{E}[Z] = 0$, implying

$$\operatorname{Cov}[X,Y] = \operatorname{E}\left[X^2\right]. \tag{3}$$

Independence of X and Z also implies $\operatorname{Var}[Y] = \operatorname{Var}[X] + \operatorname{Var}[Z]$, or equivalently, since the signals are all zero-mean,

$$\operatorname{E}\left[Y^2\right] = \operatorname{E}\left[X^2\right] + \operatorname{E}\left[Z^2\right]. \tag{4}$$

These facts imply that the correlation coefficient is

$$\begin{aligned}\rho_{X,Y} = \frac{\operatorname{E}\left[X^2\right]}{\sqrt{\operatorname{E}[X^2]\operatorname{E}[Y^2]}} &= \frac{\operatorname{E}\left[X^2\right]}{\sqrt{\operatorname{E}[X^2](\operatorname{E}[X^2]+\operatorname{E}[Z^2])}} \\ &= \frac{1}{\sqrt{1+\frac{\operatorname{E}[Z^2]}{\operatorname{E}[X^2]}}}. \end{aligned}\tag{5}$$

In terms of the signal to noise ratio, we have

$$\rho_{X,Y} = \frac{1}{\sqrt{1+\frac{1}{\Gamma}}}. \tag{6}$$

We see in (6) that $\rho_{X,Y} \to 1$ as $\Gamma \to \infty$.

Problem 5.8.1 Solution

Independence of X and Z implies

$$\operatorname{Var}[Y] = \operatorname{Var}[X] + \operatorname{Var}[Z] = 1^2 + 4^2 = 17. \tag{1}$$

Since $\operatorname{E}[X] = \operatorname{E}[Y] = 0$, the covariance of X and Y is

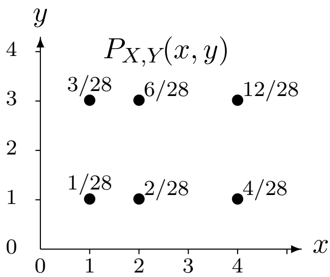
$$\operatorname{Cov}[X,Y] = \operatorname{E}[XY] = \operatorname{E}[X(X+Z)] = \operatorname{E}\left[X^2\right] + \operatorname{E}[XZ]. \tag{2}$$

Since X and Z are independent, $\operatorname{E}[XZ] = \operatorname{E}[X]\operatorname{E}[Z] = 0$. Since $\operatorname{E}[X] = 0$, $\operatorname{E}[X^2] = \operatorname{Var}[X] = 1$. Thus $\operatorname{Cov}[X,Y] = 1$. Finally, the correlation coefficient is

$$\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sqrt{\operatorname{Var}[X]}\sqrt{\operatorname{Var}[Y]}} = \frac{1}{\sqrt{17}} = 0.243. \tag{3}$$

Since $\rho_{X,Y} \neq 0$, we conclude that X and Y are dependent.

Problem 5.8.2 Solution



In Problem 5.2.1, we found the joint PMF $P_{X,Y}(x,y)$ as shown. Also the expected values and variances were

$$\mathrm{E}[X] = 3, \qquad \mathrm{Var}[X] = 10/7, \tag{1}$$

$$\mathrm{E}[Y] = 5/2, \qquad \mathrm{Var}[Y] = 3/4. \tag{2}$$

We use these results now to solve this problem.

(a) Random variable $W = Y/X$ has expected value

$$\begin{aligned}\mathrm{E}[Y/X] &= \sum_{x=1,2,4}\sum_{y=1,3} \frac{y}{x} P_{X,Y}(x,y) \\ &= \frac{1}{1}\frac{1}{28} + \frac{3}{1}\frac{3}{28} + \frac{1}{2}\frac{2}{28} + \frac{3}{2}\frac{6}{28} + \frac{1}{4}\frac{4}{28} + \frac{3}{4}\frac{12}{28} = 15/14.\end{aligned} \tag{3}$$

(b) The correlation of X and Y is

$$\begin{aligned}r_{X,Y} &= \sum_{x=1,2,4}\sum_{y=1,3} xy P_{X,Y}(x,y) \\ &= \frac{1\cdot1\cdot1}{28} + \frac{1\cdot3\cdot3}{28} + \frac{2\cdot1\cdot2}{28} + \frac{2\cdot3\cdot6}{28} + \frac{4\cdot1\cdot4}{28} + \frac{4\cdot3\cdot12}{28} \\ &= 210/28 = 15/2.\end{aligned} \tag{4}$$

Recognizing that $P_{X,Y}(x,y) = xy/28$ yields the faster calculation

$$\begin{aligned}r_{X,Y} = \mathrm{E}[XY] &= \sum_{x=1,2,4}\sum_{y=1,3} \frac{(xy)^2}{28} \\ &= \frac{1}{28}\sum_{x=1,2,4} x^2 \sum_{y=1,3} y^2 \\ &= \frac{1}{28}(1+2^2+4^2)(1^2+3^2) = \frac{210}{28} = \frac{15}{2}.\end{aligned} \tag{5}$$

(c) The covariance of X and Y is

$$\operatorname{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \frac{15}{2} - 3\frac{5}{2} = 0. \tag{6}$$

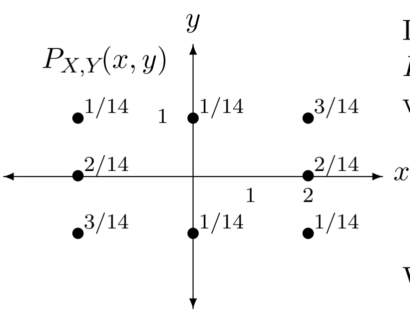
(d) Since X and Y have zero covariance, the correlation coefficient is

$$\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} = 0. \tag{7}$$

(e) Since X and Y are uncorrelated, the variance of $X+Y$ is

$$\operatorname{Var}[X+Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] = \frac{61}{28}. \tag{8}$$

Problem 5.8.3 Solution



In Problem 5.2.1, we found the joint PMF $P_{X,Y}(x,y)$ shown here. The expected values and variances were found to be

$$\mathrm{E}[X] = 0, \qquad \operatorname{Var}[X] = 24/7, \tag{1}$$
$$\mathrm{E}[Y] = 0, \qquad \operatorname{Var}[Y] = 5/7. \tag{2}$$

We need these results to solve this problem.

(a) Random variable $W = 2^{XY}$ has expected value

$$\begin{aligned}
\mathrm{E}\left[2^{XY}\right] &= \sum_{x=-2,0,2}\ \sum_{y=-1,0,1} 2^{xy} P_{X,Y}(x,y) \\
&= 2^{-2(-1)}\frac{3}{14} + 2^{-2(0)}\frac{2}{14} + 2^{-2(1)}\frac{1}{14} + 2^{0(-1)}\frac{1}{14} + 2^{0(1)}\frac{1}{14} \\
&\qquad + 2^{2(-1)}\frac{1}{14} + 2^{2(0)}\frac{2}{14} + 2^{2(1)}\frac{3}{14} \\
&= 61/28.
\end{aligned} \tag{3}$$

(b) The correlation of X and Y is

$$\begin{aligned} r_{X,Y} &= \sum_{x=-2,0,2} \sum_{y=-1,0,1} xy P_{X,Y}(x,y) \\ &= \frac{-2(-1)(3)}{14} + \frac{-2(0)(2)}{14} + \frac{-2(1)(1)}{14} \\ &\qquad + \frac{2(-1)(1)}{14} + \frac{2(0)(2)}{14} + \frac{2(1)(3)}{14} \\ &= 4/7. \end{aligned} \tag{4}$$

(c) The covariance of X and Y is

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 4/7. \tag{5}$$

(d) The correlation coefficient is

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{2}{\sqrt{30}}. \tag{6}$$

(e) By Theorem 5.16,

$$\begin{aligned} \mathrm{Var}[X+Y] &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y] \\ &= \frac{24}{7} + \frac{5}{7} + 2\frac{4}{7} = \frac{37}{7}. \end{aligned} \tag{7}$$

Problem 5.8.4 Solution

In the solution to Quiz 5.3, the joint PMF and the marginal PMFs are

$P_{H,B}(h,b)$	$b=0$	$b=2$	$b=4$	$P_H(h)$
$h=-1$	0	0.4	0.2	0.6
$h=0$	0.1	0	0.1	0.2
$h=1$	0.1	0.1	0	0.2
$P_B(b)$	0.2	0.5	0.3	

(1)

From the joint PMF, the correlation coefficient is

$$
\begin{aligned}
r_{H,B} = \mathrm{E}[HB] &= \sum_{h=-1}^{1} \sum_{b=0,2,4} hbP_{H,B}(h,b) \\
&= -1(2)(0.4) + 1(2)(0.1) + -1(4)(0.2) + 1(4)(0) \\
&= -1.4. \qquad (2)
\end{aligned}
$$

since only terms in which both h and b are nonzero make a contribution. Using the marginal PMFs, the expected values of X and Y are

$$
\mathrm{E}[H] = \sum_{h=-1}^{1} hP_H(h) = -1(0.6) + 0(0.2) + 1(0.2) = -0.2, \qquad (3)
$$

$$
\mathrm{E}[B] = \sum_{b=0,2,4} bP_B(b) = 0(0.2) + 2(0.5) + 4(0.3) = 2.2. \qquad (4)
$$

The covariance is

$$
\mathrm{Cov}[H,B] = \mathrm{E}[HB] - \mathrm{E}[H]\,\mathrm{E}[B] = -1.4 - (-0.2)(2.2) = -0.96. \qquad (5)
$$

Problem 5.8.5 Solution

X and Y are independent random variables with PDFs

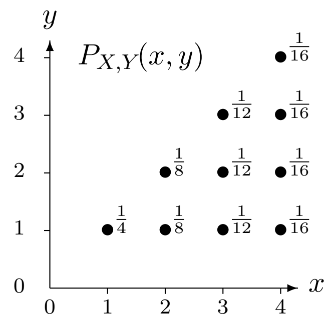
$$
f_X(x) = \begin{cases} \frac{1}{3}e^{-x/3} & x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} \frac{1}{2}e^{-y/2} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)
$$

(a) Since X and Y are exponential random variables with parameters $\lambda_X = 1/3$ and $\lambda_Y = 1/2$, Appendix A tells us that $\mathrm{E}[X] = 1/\lambda_X = 3$ and $\mathrm{E}[Y] = 1/\lambda_Y = 2$. Since X and Y are independent, the correlation is $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y] = 6$.

(b) Since X and Y are independent, $\mathrm{Cov}[X,Y] = 0$.

Problem 5.8.6 Solution

From the joint PMF



we can find the marginal PMF for X or Y by summing over the columns or rows of the joint PMF.

$$P_Y(y) = \begin{cases} 25/48 & y = 1, \\ 13/48 & y = 2, \\ 7/48 & y = 3, \\ 3/48 & y = 4, \\ 0 & \text{otherwise,} \end{cases} \qquad P_X(x) = \begin{cases} 1/4 & x = 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The expected values are

$$\mathrm{E}[Y] = \sum_{y=1}^{4} y P_Y(y) = 1\frac{25}{48} + 2\frac{13}{48} + 3\frac{7}{48} + 4\frac{3}{48} = 7/4, \tag{2}$$

$$\mathrm{E}[X] = \sum_{x=1}^{4} x P_X(x) = \frac{1}{4}(1+2+3+4) = 5/2. \tag{3}$$

(b) To find the variances, we first find the second moments.

$$\mathrm{E}\left[Y^2\right] = \sum_{y=1}^{4} y^2 P_Y(y) = 1^2 \frac{25}{48} + 2^2 \frac{13}{48} + 3^2 \frac{7}{48} + 4^2 \frac{3}{48} = 47/12, \tag{4}$$

$$\mathrm{E}\left[X^2\right] = \sum_{x=1}^{4} x^2 P_X(x) = \frac{1}{4}\left(1^2 + 2^2 + 3^2 + 4^2\right) = 15/2. \tag{5}$$

Now the variances are

$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 47/12 - (7/4)^2 = 41/48, \tag{6}$$

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 15/2 - (5/2)^2 = 5/4. \tag{7}$$

(c) To find the correlation, we evaluate the product XY over all values of X and Y. Specifically,

$$\begin{aligned} r_{X,Y} = \mathrm{E}[XY] &= \sum_{x=1}^{4} \sum_{y=1}^{x} xy P_{X,Y}(x, y) \\ &= \frac{1}{4}(1) + \frac{1}{8}(2+4) \\ &\qquad + \frac{1}{12}(3+6+9) + \frac{1}{16}(4+8+12+16) \\ &= 5. \end{aligned} \tag{8}$$

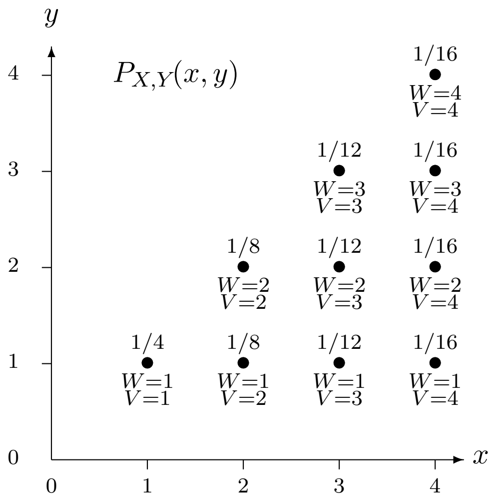
(d) The covariance of X and Y is

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 5 - (7/4)(10/4) = 10/16. \tag{9}$$

(e) The correlation coefficient is

$$\rho_{X,Y} = \frac{\mathrm{Cov}[W, V]}{\sqrt{\mathrm{Var}[W]\,\mathrm{Var}[V]}} = \frac{10/16}{\sqrt{(41/48)(5/4)}} \approx 0.605. \tag{10}$$

Problem 5.8.7 Solution



To solve this problem, we identify the values of $W = \min(X,Y)$ and $V = \max(X,Y)$ for each possible pair x, y. Here we observe that $W = Y$ and $V = X$. This is a result of the underlying experiment in that given $X = x$, each $Y \in \{1, 2, \ldots, x\}$ is equally likely. Hence $Y \leq X$. This implies $\min(X,Y) = Y$ and $\max(X,Y) = X$.

Using the results from Problem 5.8.6, we have the following answers.

(a) The expected values are

$$E[W] = E[Y] = 7/4, \qquad E[V] = E[X] = 5/2. \tag{1}$$

(b) The variances are

$$\text{Var}[W] = \text{Var}[Y] = 41/48, \qquad \text{Var}[V] = \text{Var}[X] = 5/4. \tag{2}$$

(c) The correlation is

$$r_{W,V} = E[WV] = E[XY] = r_{X,Y} = 5. \tag{3}$$

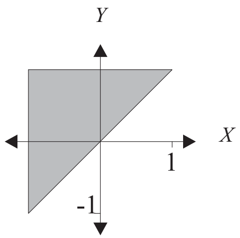
(d) The covariance of W and V is

$$\text{Cov}[W,V] = \text{Cov}[X,Y] = 10/16. \tag{4}$$

(e) The correlation coefficient is

$$\rho_{W,V} = \rho_{X,Y} = \frac{10/16}{\sqrt{(41/48)(5/4)}}. \approx 0.605. \tag{5}$$

Problem 5.8.8 Solution



Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/2 & -1 \leq x \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The region of possible pairs (x, y) is shown with the joint PDF. The rest of this problem is just calculus.

$$r_{X,Y} = \int_{-1}^{1} \int_{x}^{1} \frac{xy}{2}\, dy\, dx = \frac{1}{4} \int_{-1}^{1} x(1-x^2)\, dx = \left. \frac{x^2}{8} - \frac{x^4}{16} \right|_{-1}^{1} = 0. \tag{2}$$

$$\begin{aligned} \mathrm{E}\left[e^{X+Y}\right] &= \int_{-1}^{1} \int_{x}^{1} \frac{1}{2} e^x e^y\, dy\, dx \\ &= \frac{1}{2} \int_{-1}^{1} e^x (e^1 - e^x)\, dx \\ &= \left. \frac{1}{2} e^{1+x} - \frac{1}{4} e^{2x} \right|_{-1}^{1} = \frac{e^2}{4} + \frac{e^{-2}}{4} - \frac{1}{2}. \end{aligned} \tag{3}$$

Problem 5.8.9 Solution

(a) Since $\hat{X} = aX$, $\mathrm{E}[\hat{X}] = a\,\mathrm{E}[X]$ and

$$\begin{aligned} \mathrm{E}\left[\hat{X} - \mathrm{E}\left[\hat{X}\right]\right] &= \mathrm{E}\left[aX - a\,\mathrm{E}\left[X\right]\right] \\ &= \mathrm{E}\left[a(X - \mathrm{E}\left[X\right])\right] = a\,\mathrm{E}\left[X - \mathrm{E}\left[X\right]\right]. \end{aligned} \tag{1}$$

In the same way, since $\hat{Y} = cY$, $\mathrm{E}[\hat{Y}] = c\,\mathrm{E}[Y]$ and

$$\begin{aligned} \mathrm{E}\left[\hat{Y} - \mathrm{E}\left[\hat{Y}\right]\right] &= \mathrm{E}\left[aY - a\,\mathrm{E}\left[Y\right]\right] \\ &= \mathrm{E}\left[a(Y - \mathrm{E}\left[Y\right])\right] = a\,\mathrm{E}\left[Y - \mathrm{E}\left[Y\right]\right]. \end{aligned} \tag{2}$$

(b)

$$\begin{aligned}
\operatorname{Cov}\left[\hat{X},\hat{Y}\right] &= \mathrm{E}\left[(\hat{X}-\mathrm{E}\left[\hat{X}\right])(\hat{Y}-\mathrm{E}\left[\hat{Y}\right])\right] \\
&= \mathrm{E}\left[ac(X-\mathrm{E}\left[X\right])(Y-\mathrm{E}\left[Y\right])\right] \\
&= ac\,\mathrm{E}\left[(X-\mathrm{E}\left[X\right])(Y-\mathrm{E}\left[Y\right])\right] = ac\operatorname{Cov}\left[X,Y\right]. \qquad (3)
\end{aligned}$$

(c) Since $\hat{X} = aX$,

$$\begin{aligned}
\operatorname{Var}[\hat{X}] &= \mathrm{E}\left[(\hat{X}-\mathrm{E}\left[\hat{X}\right])^2\right] \\
&= \mathrm{E}\left[(aX-a\,\mathrm{E}\left[X\right])^2\right] \\
&= \mathrm{E}\left[a^2(X-\mathrm{E}\left[X\right])^2\right] = a^2\operatorname{Var}[X]. \qquad (4)
\end{aligned}$$

In the very same way,

$$\begin{aligned}
\operatorname{Var}[\hat{Y}] &= \mathrm{E}\left[(\hat{Y}-\mathrm{E}\left[\hat{Y}\right])^2\right] \\
&= \mathrm{E}\left[(cY-c\,\mathrm{E}\left[Y\right])^2\right] \\
&= \mathrm{E}\left[c^2(Y-\mathrm{E}\left[Y\right])^2\right] = c^2\operatorname{Var}[Y]. \qquad (5)
\end{aligned}$$

(d)

$$\begin{aligned}
\rho_{\hat{X},\hat{Y}} &= \frac{\operatorname{Cov}\left[\hat{X},\hat{Y}\right]}{\sqrt{\operatorname{Var}[\hat{X}]\operatorname{Var}[\hat{Y}]}} \\
&= \frac{ac\operatorname{Cov}\left[X,Y\right]}{\sqrt{a^2c^2\operatorname{Var}[X]\operatorname{Var}[Y]}} = \frac{\operatorname{Cov}\left[X,Y\right]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} = \rho_{X,Y}. \qquad (6)
\end{aligned}$$

Problem 5.8.10 Solution

For this problem, calculating the marginal PMF of K is not easy. However, the marginal PMF of N is easy to find. For $n = 1, 2, \ldots$,

$$P_N(n) = \sum_{k=1}^{n} \frac{(1-p)^{n-1}p}{n} = (1-p)^{n-1}p. \qquad (1)$$

That is, N has a geometric PMF. From Appendix A, we note that

$$\mathrm{E}[N] = \frac{1}{p}, \qquad \mathrm{Var}[N] = \frac{1-p}{p^2}. \tag{2}$$

We can use these facts to find the second moment of N.

$$\mathrm{E}\left[N^2\right] = \mathrm{Var}[N] + (\mathrm{E}[N])^2 = \frac{2-p}{p^2}. \tag{3}$$

Now we can calculate the moments of K.

$$\mathrm{E}[K] = \sum_{n=1}^{\infty}\sum_{k=1}^{n} k\frac{(1-p)^{n-1}p}{n} = \sum_{n=1}^{\infty}\frac{(1-p)^{n-1}p}{n}\sum_{k=1}^{n} k. \tag{4}$$

Since $\sum_{k=1}^{n} k = n(n+1)/2$,

$$\mathrm{E}[K] = \sum_{n=1}^{\infty}\frac{n+1}{2}(1-p)^{n-1}p = \mathrm{E}\left[\frac{N+1}{2}\right] = \frac{1}{2p} + \frac{1}{2}. \tag{5}$$

We now can calculate the sum of the moments.

$$\mathrm{E}[N+K] = \mathrm{E}[N] + \mathrm{E}[K] = \frac{3}{2p} + \frac{1}{2}. \tag{6}$$

The second moment of K is

$$\mathrm{E}\left[K^2\right] = \sum_{n=1}^{\infty}\sum_{k=1}^{n} k^2\frac{(1-p)^{n-1}p}{n} = \sum_{n=1}^{\infty}\frac{(1-p)^{n-1}p}{n}\sum_{k=1}^{n} k^2. \tag{7}$$

Using the identity $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$, we obtain

$$\mathrm{E}\left[K^2\right] = \sum_{n=1}^{\infty}\frac{(n+1)(2n+1)}{6}(1-p)^{n-1}p = \mathrm{E}\left[\frac{(N+1)(2N+1)}{6}\right]. \tag{8}$$

Applying the values of $\mathrm{E}[N]$ and $\mathrm{E}[N^2]$ found above, we find that

$$\mathrm{E}\left[K^2\right] = \frac{\mathrm{E}\left[N^2\right]}{3} + \frac{\mathrm{E}[N]}{2} + \frac{1}{6} = \frac{2}{3p^2} + \frac{1}{6p} + \frac{1}{6}. \tag{9}$$

Thus, we can calculate the variance of K.

$$\mathrm{Var}[K] = \mathrm{E}\left[K^2\right] - (\mathrm{E}\left[K\right])^2 = \frac{5}{12p^2} - \frac{1}{3p} + \frac{5}{12}. \tag{10}$$

To find the correlation of N and K,

$$\begin{aligned} r_{N,K} = \mathrm{E}\left[NK\right] &= \sum_{n=1}^{\infty}\sum_{k=1}^{n} nk\frac{(1-p)^{n-1}p}{n} \\ &= \sum_{n=1}^{\infty}(1-p)^{n-1}p\sum_{k=1}^{n} k. \end{aligned} \tag{11}$$

Since $\sum_{k=1}^{n} k = n(n+1)/2$,

$$r_{N,K} = \sum_{n=1}^{\infty}\frac{n(n+1)}{2}(1-p)^{n-1}p = \mathrm{E}\left[\frac{N(N+1)}{2}\right] = \frac{1}{p^2}. \tag{12}$$

Finally, the covariance is

$$\mathrm{Cov}\left[N,K\right] = r_{N,K} - \mathrm{E}\left[N\right]\mathrm{E}\left[K\right] = \frac{1}{2p^2} - \frac{1}{2p}. \tag{13}$$

Problem 5.9.1 Solution

X and Y have joint PDF

$$f_{X,Y}(x,y) = ce^{-(x^2/8)-(y^2/18)}. \tag{1}$$

The omission of any limits for the PDF indicates that it is defined over all x and y. We know that $f_{X,Y}(x,y)$ is in the form of the bivariate Gaussian distribution so we look to Definition 5.10 and attempt to find values for σ_Y, σ_X, $\mathrm{E}[X]$, $\mathrm{E}[Y]$ and ρ. First, we know that the constant is

$$c = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}. \tag{2}$$

Because the exponent of $f_{X,Y}(x,y)$ doesn't contain any cross terms we know that ρ must be zero, and we are left to solve the following for $\mathrm{E}[X]$, $\mathrm{E}[Y]$, σ_X, and σ_Y:

$$\left(\frac{x-\mathrm{E}\left[X\right]}{\sigma_X}\right)^2 = \frac{x^2}{8}, \qquad \left(\frac{y-\mathrm{E}\left[Y\right]}{\sigma_Y}\right)^2 = \frac{y^2}{18}. \tag{3}$$

From which we can conclude that

$$\mathrm{E}[X] = \mathrm{E}[Y] = 0, \tag{4}$$
$$\sigma_X = \sqrt{8}, \tag{5}$$
$$\sigma_Y = \sqrt{18}. \tag{6}$$

Putting all the pieces together, we find that $c = \frac{1}{24\pi}$. Since $\rho = 0$, we also find that X and Y are independent.

Problem 5.9.2 Solution

Because X and Y are independent and Gaussian, they are also bivariate Gaussian. By Theorem 5.21, a linear combination is Gaussian. For each part, all we need to do is calculate the expected value and variance and then write down the corresponding Gaussian PDF.

(a) For $V = X + Y$,

$$\mu_V = \mathrm{E}[X] + \mathrm{E}[Y] = 1 + 2 = 3, \tag{1}$$
$$\sigma_V^2 = \mathrm{Var}[W] = \mathrm{Var}[X] + \mathrm{Var}[Y] = 2^2 + 4^2 = 20. \tag{2}$$

The PDF of V is

$$f_V(v) = \frac{1}{\sqrt{2\pi\sigma_V^2}} e^{-(v-\mu_V)^2/2\sigma_V^2} = \frac{1}{\sqrt{40\pi}} e^{-(v-3)^2/40}. \tag{3}$$

(b) For $W = 3X + 2Y$:

$$\mathrm{E}[W] = 3\,\mathrm{E}[X] + 2\,\mathrm{E}[Y] = 3(1) + 2(2) = 7, \tag{4}$$
$$\mathrm{Var}[W] = 3^2\,\mathrm{Var}[X] + 2^2\,\mathrm{Var}[Y] = 9(4) + 4(16) = 100. \tag{5}$$

Now we can write down the PDF of W as

$$f_W(w) = \frac{1}{\sqrt{2\pi\,\mathrm{Var}[W]}} e^{-(w-\mathrm{E}[W])^2/2\,\mathrm{Var}[W]}$$
$$= \frac{1}{\sqrt{200\pi}} e^{-(w-7)^2/200}. \tag{6}$$

Problem 5.9.3 Solution

FALSE: Let $Y = X_1 + aX_2$. If $\mathrm{E}[X_2] = 0$, then $\mathrm{E}[Y] = \mathrm{E}[X_1]$ for all a. Since Y is Gaussian (by Theorem 5.21), $\mathrm{P}[Y \leq y] = 1/2$ if and only if $\mathrm{E}[Y] = \mathrm{E}[X_1] = y$. We obtain a simple counterexample when $y = \mathrm{E}[X_1] - 1$.

Note that the answer would be true if we knew that $\mathrm{E}[X_2] \neq 0$. Also note that the variance of W will depend on the correlation between X_1 and X_2, but the correlation is irrelevant in the above argument.

Problem 5.9.4 Solution

By Theorem 5.21, a sum of independent Gaussian random variables is Gaussian. Thus, $Y = X_1 + X_2$ is Gaussian. Next we observe that

$$\mathrm{E}[Y] = \mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 0.$$

Thus,

$$\begin{aligned}
\mathrm{Var}[Y] = \mathrm{Var}[X_1 + X_2] &= \mathrm{E}\left[(X_1 + X_2)^2\right] \\
&= \mathrm{E}\left[X_1^2 + 2x_1x_2 + X_2^2\right] \\
&= 1 + 2\,\mathrm{E}\left[X_1X_2\right] + 1. \qquad (1)
\end{aligned}$$

For Y to be idemtical to X_1 and X_2 requires $\mathrm{Var}[Y] = 1$. This requires $\mathrm{E}[X_1X_2] = \mathrm{Cov}[X_1, X_2] = -1/2$.

Problem 5.9.5 Solution

For the joint PDF

$$f_{X,Y}(x, y) = ce^{-(2x^2 - 4xy + 4y^2)}, \qquad (1)$$

we proceed as in Problem 5.9.1 to find values for σ_Y, σ_X, $\mathrm{E}[X]$, $\mathrm{E}[Y]$ and ρ.

(a) First, we try to solve the following equations

$$\left(\frac{x - \mathrm{E}[X]}{\sigma_X}\right)^2 = 4(1 - \rho^2)x^2, \qquad (2)$$

$$\left(\frac{y - \mathrm{E}[Y]}{\sigma_Y}\right)^2 = 8(1 - \rho^2)y^2, \qquad (3)$$

$$\frac{2\rho}{\sigma_X\sigma_Y} = 8(1 - \rho^2). \qquad (4)$$

The first two equations yield $\mathrm{E}[X] = \mathrm{E}[Y] = 0$.

(b) To find the correlation coefficient ρ, we observe that

$$\sigma_X = 1/\sqrt{4(1-\rho^2)}, \qquad \sigma_Y = 1/\sqrt{8(1-\rho^2)}. \tag{5}$$

Using σ_X and σ_Y in the third equation yields $\rho = 1/\sqrt{2}$.

(c) Since $\rho = 1/\sqrt{2}$, now we can solve for σ_X and σ_Y.

$$\sigma_X = 1/\sqrt{2}, \qquad \sigma_Y = 1/2. \tag{6}$$

(d) From here we can solve for c.

$$c = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} = \frac{2}{\pi}. \tag{7}$$

(e) X and Y are dependent because $\rho \neq 0$.

Problem 5.9.6 Solution

The event B is the set of outcomes satisfying $X^2 + Y^2 \leq 2^2$. Of ocurse, the calculation of $P[B]$ depends on the probability model for X and Y.

(a) In this instance, X and Y have the same PDF

$$f_X(x) = f_Y(x) = \begin{cases} 0.01 & -50 \leq x \leq 50, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Since X and Y are independent, their joint PDF is

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)\, f_Y(y) \\ &= \begin{cases} 10^{-4} & -50 \leq x \leq 50, -50 \leq y \leq 50, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

Because X and Y have a uniform PDF over the bullseye area, $P[B]$ is just the value of the joint PDF over the area times the area of the bullseye.

$$P[B] = P\left[X^2 + Y^2 \leq 2^2\right] = 10^{-4} \cdot \pi 2^2 = 4\pi \cdot 10^{-4} \approx 0.0013. \tag{3}$$

(b) In this case, the joint PDF of X and Y is inversely proportional to the area of the target.

$$f_{X,Y}(x,y) = \begin{cases} 1/[\pi 50^2] & x^2+y^2 \le 50^2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The probability of a bullseye is

$$\mathrm{P}[B] = \mathrm{P}\left[X^2+Y^2 \le 2^2\right] = \frac{\pi 2^2}{\pi 50^2} = \left(\frac{1}{25}\right)^2 \approx 0.0016. \tag{5}$$

(c) In this instance, X and Y have the identical Gaussian $(0,\sigma)$ PDF with $\sigma^2 = 100$; i.e.,

$$f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}. \tag{6}$$

Since X and Y are independent, their joint PDF is

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{7}$$

To find $\mathrm{P}[B]$, we write

$$\begin{aligned} \mathrm{P}[B] = \mathrm{P}\left[X^2+Y^2 \le 2^2\right] &= \iint_{x^2+y^2\le 2^2} f_{X,Y}(x,y)\, dx\, dy \\ &= \frac{1}{2\pi\sigma^2} \iint_{x^2+y^2\le 2^2} e^{-(x^2+y^2)/2\sigma^2}\, dx\, dy. \end{aligned} \tag{8}$$

This integral is easy using polar coordinates. With the substitutions $x^2+y^2 = r^2$, and $dx\,dy = r\,dr\,d\theta$,

$$\begin{aligned} \mathrm{P}[B] &= \frac{1}{2\pi\sigma^2} \int_0^2 \int_0^{2\pi} e^{-r^2/2\sigma^2} r\, dr\, d\theta \\ &= \frac{1}{\sigma^2} \int_0^2 r e^{-r^2/2\sigma^2}\, dr \\ &= -e^{-r^2/2\sigma^2}\Big|_0^2 = 1 - e^{-4/200} \approx 0.0198. \end{aligned} \tag{9}$$

Problem 5.9.7 Solution

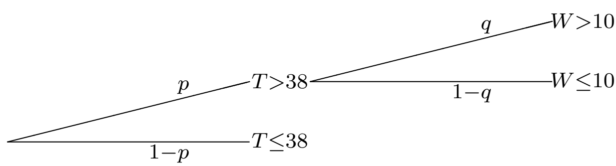
(a) The person's temperature is high with probability

$$p = \mathrm{P}[T > 38] = \mathrm{P}[T - 37 > 38 - 37] = 1 - \Phi(1) = 0.159. \tag{1}$$

Given that the temperature is high, then W is measured. Since $\rho = 0$, W and T are independent and

$$q = \mathrm{P}[W > 10] = \mathrm{P}\left[\frac{W-7}{2} > \frac{10-7}{2}\right] = 1 - \Phi(1.5) = 0.067. \tag{2}$$

The tree for this experiment is



The probability the person is ill is

$$\begin{aligned} \mathrm{P}[I] &= \mathrm{P}[T > 38, W > 10] \\ &= \mathrm{P}[T > 38]\,\mathrm{P}[W > 10] = pq = 0.0107. \end{aligned} \tag{3}$$

(b) The general form of the bivariate Gaussian PDF is

$$f_{W,T}(w,t) = \frac{\exp\left[-\dfrac{\left(\frac{w-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(w-\mu_1)(t-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{t-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}. \tag{4}$$

With $\mu_1 = \mathrm{E}[W] = 7$, $\sigma_1 = \sigma_W = 2$, $\mu_2 = \mathrm{E}[T] = 37$ and $\sigma_2 = \sigma_T = 1$ and $\rho = 1/\sqrt{2}$, we have

$$f_{W,T}(w,t) = \frac{\exp\left[-\frac{(w-7)^2}{4} - \frac{\sqrt{2}(w-7)(t-37)}{2} + (t-37)^2\right]}{2\pi\sqrt{2}}. \tag{5}$$

To find the conditional probability $\mathrm{P}[I|T=t]$, we need to find the conditional PDF of W given $T=t$. The direct way is simply to use algebra to find

$$f_{W|T}(w|t) = \frac{f_{W,T}(w,t)}{f_T(t)}. \tag{6}$$

The required algebra is essentially the same as that needed to prove Theorem 7.15. Its easier just to apply Theorem 7.15 which says that given $T=t$, the conditional distribution of W is Gaussian with

$$\mathrm{E}[W|T=t] = \mathrm{E}[W] + \rho\frac{\sigma_W}{\sigma_T}(t - \mathrm{E}[T]), \tag{7}$$

$$\mathrm{Var}[W|T=t] = \sigma_W^2(1-\rho^2). \tag{8}$$

Plugging in the various parameters gives

$$\mathrm{E}[W|T=t] = 7 + \sqrt{2}(t-37), \quad \text{and} \quad \mathrm{Var}[W|T=t] = 2. \tag{9}$$

Using this conditional mean and variance, we obtain the conditional Gaussian PDF

$$f_{W|T}(w|t) = \frac{1}{\sqrt{4\pi}}e^{-\left(w-(7+\sqrt{2}(t-37))\right)^2/4}. \tag{10}$$

Given $T=t$, the conditional probability the person is declared ill is

$$\begin{aligned}
\mathrm{P}[I|T=t] &= \mathrm{P}[W > 10|T=t] \\
&= \mathrm{P}\left[\frac{W-(7+\sqrt{2}(t-37))}{\sqrt{2}} > \frac{10-(7+\sqrt{2}(t-37))}{\sqrt{2}}\right] \\
&= \mathrm{P}\left[Z > \frac{3-\sqrt{2}(t-37)}{\sqrt{2}}\right] \\
&= Q\left(\frac{3\sqrt{2}}{2} - (t-37)\right).
\end{aligned} \tag{11}$$

Problem 5.9.8 Solution

(a) As $0.5X_1$ and $0.5X_2$ are independent Gaussian random variables, the sum X is a Gaussian random variable with

$$\begin{aligned} \mathrm{E}[X] &= 0.5\,\mathrm{E}[X_1] + 0.5\,\mathrm{E}[X_2] = 0, & (1)\\ \mathrm{Var}[X] &= \mathrm{Var}[0.5X_1] + \mathrm{Var}[0.5X_2] \\ &= 0.25\,\mathrm{Var}[X_1] + 0.25\,\mathrm{Var}[X_2] = 1. & (2) \end{aligned}$$

Thus X is a Gaussian $(0,1)$ random variable. It follows that

$$\begin{aligned} \mathrm{P}[A] = \mathrm{P}[X > 1] &= 1 - \mathrm{P}[X \leq 1] \\ &= 1 - \Phi(1) = 1 - 0.841 = 0.159. & (3) \end{aligned}$$

(b)

$$\begin{aligned} \mathrm{P}[A] = \mathrm{P}[\max(X_1, X_2) \geq 1] &= 1 - \mathrm{P}[\max(X_2, X_2) \leq 1] \\ &= 1 - \mathrm{P}[X_1 \leq 1, X_2 \leq 1]. & (4) \end{aligned}$$

Since X_1 and X_2 are iid,

$$\begin{aligned} \mathrm{P}[A] &= 1 - \mathrm{P}[X_1 \leq 1]\,\mathrm{P}[X_2 \leq 1] \\ &= 1 - (\mathrm{P}[X_1 \leq 1])^2 \\ &= 1 - [\Phi(1/\sqrt{2})]^2 \\ &= 1 - \Phi^2(0.707) \approx 1 - (0.758)^2 = 0.42. & (5) \end{aligned}$$

(c) Recall that $X = 0.5(X_1 + X_2)$. If you plot the regions

$$\begin{aligned} \{X > 1\} &= \{(X_1, X_2) | 0.5(X_1 + X_2) > 1\}, & (6)\\ \{\max(X_1, X_2) > 1\} &= \{(X_1, X_2) | \max(X_1, X_2) > 1\}, & (7) \end{aligned}$$

in the (X_1, X_2) plane, you will observe that

$$\{X > 1\} \subset \{\max(X_1, X_2) > 1\}.$$

It follows that

$$\begin{aligned} P[A] &= P[\{X > 1\} \cup \{\max(X_1, X_2) > 1\}] \\ &= P[\{\max(X_1, X_2) > 1\}] = 0.42. \end{aligned} \tag{8}$$

(d) Now

$$\begin{aligned} P[A] &= P[\max(X_1, X_2) > 1, \min(X_1, X_2) > 0] \\ &= P[\max X_1, X_2 \geq 1, X_1 > 0, X_2 > 0]. \end{aligned} \tag{9}$$

When you draw a picture of this set in the (X_1, X_2) plane, you will see that this is the first quadrant $\{X_1 > 0, X_2 > 0\}$ except that the unit square $\{0 < X_1 < 1, 0 < X_2 < 1\}$ is removed. The probability of A is

$$\begin{aligned} P[A] &= P[X_1 > 0, X_2 > 0] - P[0 < X_1 < 1, 0 < X_2 < 1] \\ &= P[X_1 > 0]\, P[X_2 > 0] - P[0 < X_1 < 1]\, P[0 < X_2 < 1] \\ &= P[X_1 > 0]^2 - P[0 < X_1 < 1]^2 \\ &= [1 - \Phi(0)]^2 - [\Phi(1/\sqrt{2}) - \Phi(0)]^2 \\ &\approx 1/4 - [\Phi(0.7) - 1/2]^2 \\ &= 0.25 - (0.258)^2 = 0.183. \end{aligned} \tag{10}$$

Problem 5.9.9 Solution

The key to this problem is knowing that a sum of independent Gaussian random variables is Gaussian.

(a) First we calculate the mean and variance of Y. Since the expectation of the sum is always the sum of the expectations,

$$E[Y] = \frac{1}{2} E[X_1] + \frac{1}{2} E[X_2] = 74. \tag{1}$$

Since X_1 and X_2 are independent, the variance of the sum is the sum of the variances:

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{Var}[X_1/2] + \mathrm{Var}[X_2/2] \\ &= \frac{1}{4} \mathrm{Var}[X_1] + \frac{1}{4} \mathrm{Var}[X_2] = 16^2/2 = 128. \end{aligned} \tag{2}$$

Thus, Y has standard deviation $\sigma_Y = 8\sqrt{2}$. Since we know Y is Gaussian,

$$\begin{aligned} P[A] = P[Y \geq 90] &= P\left[\frac{Y - 74}{8\sqrt{2}} \geq \frac{90 - 74}{8\sqrt{2}}\right] \\ &= Q(\sqrt{2}) = 1 - \Phi(\sqrt{2}). \end{aligned} \tag{3}$$

(b) With weight factor w, Y is still Gaussian, but with

$$E[Y] = w\,E[X_1] + (1-w)\,E[X_2] = 74, \tag{4}$$

$$\begin{aligned} \text{Var}[Y] &= \text{Var}[wX_1] + \text{Var}[(1-w)X_2] \\ &= w^2\,\text{Var}[X_1] + (1-w)^2\,\text{Var}[X_2] = 16^2[w^2 + (1-w)^2]. \end{aligned} \tag{5}$$

Thus, Y has standard deviation $\sigma_Y = 16\sqrt{w^2 + (1-w)^2}$. Since we know Y is Gaussian,

$$\begin{aligned} P[A] = P[Y \geq 90] &= P\left[\frac{Y - 74}{\sigma_Y} \geq \frac{90 - 74}{\sigma_Y}\right] \\ &= 1 - \Phi\left(\frac{1}{\sqrt{w^2 + (1-w)^2}}\right). \end{aligned} \tag{6}$$

Since $\Phi(x)$ is increasing in x, $1 - \Phi(x)$ is decreasing in x. To maximize $P[A]$, we want the argument of the Φ function to be as small as possible. Thus we want $w^2 + (1-w)^2$ to be as large as possible. Note that $w^2 + (1-w)^2$ is a parabola with a minimum at $w = 1/2$; it is maximized at $w = 1$ or $w = 0$.

That is, if you need to get exam scores around 74, and you need 90 to get an A, then you need to get lucky to get an A. With $w = 0$, you just need to be lucky on exam 1. With $w = 0$, you need only be lucky on exam 2. It's more likely that you are lucky on one exam rather than two.

(c) With the maximum of the two exams,

$$\begin{aligned}
\mathrm{P}[A] &= \mathrm{P}[\max(X_1, X_2) > 90] \\
&= 1 - \mathrm{P}[\max(X_1, X_2) \le 90] \\
&= 1 - \mathrm{P}[X_1 \le 90, X_2 \le 90] \\
&= 1 - \mathrm{P}[X_1 \le 90]\,\mathrm{P}[X_2 \le 90] \\
&= 1 - (\mathrm{P}[X_1 \le 90])^2 \\
&= 1 - \left[\Phi\left(\frac{90-74}{16}\right)\right]^2 = 1 - \Phi^2(1).
\end{aligned} \tag{7}$$

(d) Let N_c and N_a denote the number of A's awarded under the rules in part (c) and part (a). The expected additional number of A's is

$$\begin{aligned}
\mathrm{E}[N_c - N_a] &= 100[1 - \Phi^2(1)] - 100[1 - \Phi(\sqrt{2})] \\
&= 100[\Phi(\sqrt{2}) - \Phi^2(1)] = 21.3.
\end{aligned} \tag{8}$$

Problem 5.9.10 Solution

The given joint PDF is

$$f_{X,Y}(x,y) = de^{-(a^2x^2+bxy+c^2y^2)} \tag{1}$$

In order to be an example of the bivariate Gaussian PDF given in Definition 5.10, we must have

$$a^2 = \frac{1}{2\sigma_X^2(1-\rho^2)}, \qquad c^2 = \frac{1}{2\sigma_Y^2(1-\rho^2)},$$

$$b = \frac{-\rho}{\sigma_X\sigma_Y(1-\rho^2)}, \qquad d = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

We can solve for σ_X and σ_Y, yielding

$$\sigma_X = \frac{1}{a\sqrt{2(1-\rho^2)}}, \qquad \sigma_Y = \frac{1}{c\sqrt{2(1-\rho^2)}}. \tag{2}$$

Plugging these values into the equation for b, it follows that $b = -2ac\rho$, or, equivalently, $\rho = -b/2ac$. This implies

$$d^2 = \frac{1}{4\pi^2\sigma_X^2\sigma_Y^2(1-\rho^2)} = (1-\rho^2)a^2c^2 = a^2c^2 - b^2/4. \tag{3}$$

Since $|\rho| \leq 1$, we see that $|b| \leq 2ac$. Further, for any choice of a, b and c that meets this constraint, choosing $d = \sqrt{a^2c^2 - b^2/4}$ yields a valid PDF.

Problem 5.9.11 Solution

From Equation (5.68), we can write the bivariate Gaussian PDF as

$$f_{X,Y}(x,y) = \frac{1}{\sigma_X\sqrt{2\pi}}e^{-(x-\mu_X)^2/2\sigma_X^2}\frac{1}{\tilde{\sigma}_Y\sqrt{2\pi}}e^{-(y-\tilde{\mu}_Y(x))^2/2\tilde{\sigma}_Y^2}, \tag{1}$$

where $\tilde{\mu}_Y(x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$ and $\tilde{\sigma}_Y = \sigma_Y\sqrt{1-\rho^2}$. However, the definitions of $\tilde{\mu}_Y(x)$ and $\tilde{\sigma}_Y$ are not particularly important for this exercise. When we integrate the joint PDF over all x and y, we obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy \\
&\quad= \int_{-\infty}^{\infty}\frac{1}{\sigma_X\sqrt{2\pi}}e^{-(x-\mu_X)^2/2\sigma_X^2}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\tilde{\sigma}_Y\sqrt{2\pi}}e^{-(y-\tilde{\mu}_Y(x))^2/2\tilde{\sigma}_Y^2}\,dy}_{1}\,dx \\
&\quad= \int_{-\infty}^{\infty}\frac{1}{\sigma_X\sqrt{2\pi}}e^{-(x-\mu_X)^2/2\sigma_X^2}\,dx. \qquad (2)
\end{aligned}$$

The marked integral equals 1 because for each value of x, it is the integral of a Gaussian PDF of one variable over all possible values. In fact, it is the integral of the conditional PDF $f_{Y|X}(y|x)$ over all possible y. To complete the proof, we see that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\frac{1}{\sigma_X\sqrt{2\pi}}e^{-(x-\mu_X)^2/2\sigma_X^2}\,dx = 1. \tag{3}$$

since the remaining integral is the integral of the marginal Gaussian PDF $f_X(x)$ over all possible x.

Problem 5.9.12 Solution

(a)

$$\begin{aligned}
F_{Y_1}(y) &= \mathrm{P}\left[Y_1 \leq y\right] \\
&= \mathrm{P}\left[Y_1 \leq y | X_2 > 0\right] \mathrm{P}\left[X_2 > 0\right] + \mathrm{P}\left[Y_1 \leq y | X_2 \leq 0\right] \mathrm{P}\left[X_2 \leq 0\right] \\
&= \frac{1}{2} \mathrm{P}\left[Y_1 \leq y | X_2 > 0\right] + \frac{1}{2} \mathrm{P}\left[Y_1 \leq y | X_2 \leq 0\right] \\
&= \frac{1}{2} \mathrm{P}\left[X_1 \leq y | X_2 > 0\right] + \frac{1}{2} \mathrm{P}\left[-X_1 \leq y | X_2 \leq 0\right] \\
&= \frac{1}{2} \mathrm{P}\left[X_1 \leq y\right] + \frac{1}{2} \mathrm{P}\left[-X_1 \leq y\right] \\
&= \frac{1}{2} \mathrm{P}\left[X_1 \leq y\right] + \frac{1}{2} \mathrm{P}\left[X_1 \geq -y\right] \\
&= \frac{1}{2}\Phi(y) + \frac{1}{2}(1 - \Phi(-y)) = \Phi(y). \qquad (1)
\end{aligned}$$

(b) Since $F_{Y_1}(y_1)y = \Phi(y)$, Y_1 is Gaussian $(0, 1)$. Since X_1 and X_2 are iid and Y_2 is the identical to Y_1 except that X_1 and X_2 have their roles reversed, symmetry implies that Y_2 is identical to Y_1.

(c) Y_1 and Y_2 are NOT bivariate Gaussian. Since $\operatorname{sgn}(xy) = \operatorname{sgn}(x)\operatorname{sgn}(y)$, we see that

$$\begin{aligned}
\operatorname{sgn}(Y_1) &= \operatorname{sgn}(X_1)\operatorname{sgn}(\operatorname{sgn}(X_2)) = \operatorname{sgn}(X_1)\operatorname{sgn}(X_2), \\
\operatorname{sgn}(Y_2) &= \operatorname{sgn}(X_2)\operatorname{sgn}(\operatorname{sgn}(X_1)) = \operatorname{sgn}(X_2)\operatorname{sgn}(X_1). \qquad (2)
\end{aligned}$$

That is, Y_1 and Y_2 always have the same sign. Thus $\mathrm{P}[Y_1 < 0, Y_2 > 0] = 0$, but this is would be impossible if Y_1 and Y_2 were bivariate Gaussian.

Problem 5.10.1 Solution

The repair of each laptop can be viewed as an independent trial with four possible outcomes corresponding to the four types of needed repairs.

(a) Since the four types of repairs are mutually exclusive choices and since 4 laptops are returned for repair, the joint distribution of $N_1, \ldots, N_4$ is the multinomial PMF

$$\begin{aligned} P_{N_1,\ldots,N_4}(n_1, \ldots, n_4) \\ &= \binom{4}{n_1, n_2, n_3, n_4} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4} \\ &= \binom{4}{n_1, n_2, n_3, n_4} \left(\frac{8}{15}\right)^{n_1} \left(\frac{4}{15}\right)^{n_2} \left(\frac{2}{15}\right)^{n_3} \left(\frac{1}{15}\right)^{n_4}. \end{aligned} \tag{1}$$

(b) Let L_2 denote the event that exactly two laptops need LCD repairs. Thus $\mathrm{P}[L_2] = P_{N_1}(2)$. Since each laptop requires an LCD repair with probability $p_1 = 8/15$, the number of LCD repairs, N_1, is a binomial $(4, 8/15)$ random variable with PMF

$$P_{N_1}(n_1) = \binom{4}{n_1} (8/15)^{n_1} (7/15)^{4-n_1}. \tag{2}$$

The probability that two laptops need LCD repairs is

$$P_{N_1}(2) = \binom{4}{2} (8/15)^2 (7/15)^2 = 0.3717. \tag{3}$$

(c) A repair is type (2) with probability $p_2 = 4/15$. A repair is type (3) with probability $p_3 = 2/15$; otherwise a repair is type "other" with probability $p_o = 9/15$. Define X as the number of "other" repairs needed. The joint PMF of X, N_2, N_3 is the multinomial PMF

$$P_{N_2,N_3,X}(n_2, n_3, x) = \binom{4}{n_2, n_3, x} \left(\frac{4}{15}\right)^{n_2} \left(\frac{2}{15}\right)^{n_3} \left(\frac{9}{15}\right)^{x}. \tag{4}$$

However, Since $X + 4 - N_2 - N_3$, we observe that

$$\begin{aligned} P_{N_2,N_3}(n_2, n_3) &= P_{N_2,N_3,X}(n_2, n_3, 4 - n_2 - n_3) \\ &= \binom{4}{n_2, n_3, 4 - n_2 - n_3} \left(\frac{4}{15}\right)^{n_2} \left(\frac{2}{15}\right)^{n_3} \left(\frac{9}{15}\right)^{4-n_2-n_3} \\ &= \left(\frac{9}{15}\right)^{4} \binom{4}{n_2, n_3, 4 - n_2 - n_3} \left(\frac{4}{9}\right)^{n_2} \left(\frac{2}{9}\right)^{n_3} \end{aligned} \tag{5}$$

Similarly, since each repair is a motherboard repair with probability $p_2 = 4/15$, the number of motherboard repairs has binomial PMF

$$P_{N_2}(n_2)\, n_2 = \binom{4}{n_2} \left(\frac{4}{15}\right)^{n_2} \left(\frac{11}{15}\right)^{4-n_2} \tag{6}$$

Finally, the probability that more laptops require motherboard repairs than keyboard repairs is

$$\begin{aligned} \mathrm{P}\left[N_2 > N_3\right] = P_{N_2,N_3}(1,0) + P_{N_2,N_3}(2,0) + P_{N_2,N_3}(2,1) \\ + P_{N_2}(3) + P_{N_2}(4)\,. \end{aligned} \tag{7}$$

where we use the fact that if $N_2 = 3$ or $N_2 = 4$, then we must have $N_2 > N_3$. Inserting the various probabilities, we obtain

$$\begin{aligned} \mathrm{P}\left[N_2 > N_3\right] &= \left(\frac{9}{15}\right)^4 \left(4\left(\frac{4}{9}\right) + 6\left(\frac{16}{81}\right) + 12\left(\frac{32}{81}\right)\right) \\ &\quad + 4\left(\frac{4}{15}\right)^3 \left(\frac{11}{15}\right) + \left(\frac{4}{15}\right)^4 \\ &= \frac{8{,}656}{16{,}875} \approx 0.5129. \end{aligned} \tag{8}$$

Problem 5.10.2 Solution

Whether a computer has feature i is a Bernoulli trial with success probability $p_i = 2^{-i}$. Given that n computers were sold, the number of computers sold with feature i has the binomial PMF

$$P_{N_i}(n_i) = \binom{n}{n_i} p_i^{n_i} (1-p_i)^{n_i}. \tag{1}$$

Since a computer has feature i with probability p_i independent of whether any other feature is on the computer, the number N_i of computers with feature i is independent of the number of computers with any other features. That is, $N_1, \ldots, N_4$ are mutually independent and have joint PMF

$$P_{N_1,\ldots,N_4}(n_1,\ldots,n_4) = P_{N_1}(n_1)\, P_{N_2}(n_2)\, P_{N_3}(n_3)\, P_{N_4}(n_4)\,. \tag{2}$$

Problem 5.10.3 Solution

(a) In terms of the joint PDF, we can write the joint CDF as

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(y_1,\ldots,y_n)\,dy_1\cdots dy_n. \tag{1}$$

However, simplifying the above integral depends on the values of each x_i. In particular, $f_{X_1,\ldots,X_n}(y_1,\ldots,y_n) = 1$ if and only if $0 \le y_i \le 1$ for each i. Since $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 0$ if any $x_i < 0$, we limit, for the moment, our attention to the case where $x_i \ge 0$ for all i. In this case, some thought will show that we can write the limits in the following way:

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \int_0^{\max(1,x_1)}\cdots\int_0^{\min(1,x_n)} dy_1\cdots dy_n \\ &= \min(1,x_1)\min(1,x_2)\cdots\min(1,x_n). \end{aligned} \tag{2}$$

A complete expression for the CDF of $X_1,\ldots,X_n$ is

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \begin{cases} \prod_{i=1}^n \min(1,x_i) & 0 \le x_i, i = 1,2,\ldots,n, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(b) For $n = 3$,

$$\begin{aligned} 1 - P\left[\min_i X_i \le 3/4\right] &= P\left[\min_i X_i > 3/4\right] \\ &= P\left[X_1 > 3/4, X_2 > 3/4, X_3 > 3/4\right] \\ &= \int_{3/4}^1\int_{3/4}^1\int_{3/4}^1 dx_1\,dx_2\,dx_3 \\ &= (1-3/4)^3 = 1/64. \end{aligned} \tag{4}$$

Thus $P[\min_i X_i \le 3/4] = 63/64$.

Problem 5.10.4 Solution

The random variables N_1, N_2, N_3 and N_4 are *dependent*. To see this we observe that $P_{N_i}(4) = p_i^4$. However,

$$P_{N_1,N_2,N_3,N_4}(4,4,4,4) = 0, \tag{1}$$

which does not equal

$$P_{N_1}(4)\, P_{N_2}(4)\, P_{N_3}(4)\, P_{N_4}(4) = p_1^4 p_2^4 p_3^4 p_4^4. \tag{2}$$

Problem 5.10.5 Solution

The value of each byte is an independent experiment with 255 possible outcomes. Each byte takes on the value b_i with probability $p_i = p = 1/255$. The joint PMF of $N_0, \ldots, N_{255}$ is the multinomial PMF

$$\begin{aligned} P_{N_0,\ldots,N_{255}}(n_0,\ldots,n_{255}) &= \binom{10000}{n_0, n_1, \ldots, n_{255}} p^{n_0} p^{n_1} \cdots p^{n_{255}} \\ &= \binom{10000}{n_0, n_1, \ldots, n_{255}} (1/255)^{10000}. \end{aligned} \tag{1}$$

Keep in mind that the multinomial coefficient is defined for nonnegative integers n_i such that

$$\binom{10000}{n_0, n_1, \ldots, n_{255}} = \begin{cases} \frac{10000}{n_0! n_1! \cdots n_{255}!} & n_0 + \cdots + n_{255} = 10000, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

To evaluate the joint PMF of N_0 and N_1, we define a new experiment with three categories: b_0, b_1 and "other." Let $\hat{N}$ denote the number of bytes that are "other." In this case, a byte is in the "other" category with probability $\hat{p} = 253/255$. The joint PMF of N_0, N_1, and $\hat{N}$ is

$$P_{N_0,N_1,\hat{N}}(n_0,n_1,\hat{n}) = \binom{10000}{n_0, n_1, \hat{n}} \left(\frac{1}{255}\right)^{n_0} \left(\frac{1}{255}\right)^{n_1} \left(\frac{253}{255}\right)^{\hat{n}}. \tag{3}$$

Now we note that the following events are one in the same:

$$\{N_0 = n_0, N_1 = n_1\} = \left\{N_0 = n_0, N_1 = n_1, \hat{N} = 10000 - n_0 - n_1\right\}. \tag{4}$$

Hence, for non-negative integers n_0 and n_1 satisfying $n_0 + n_1 \leq 10000$,

$$\begin{aligned} P_{N_0,N_1}(n_0, n_1) & \\ &= P_{N_0,N_1,\hat{N}}(n_0, n_1, 10000 - n_0 - n_1) \\ &= \frac{10000!}{n_0!n_1!(10000 - n_0 - n_1)!} \left(\frac{1}{255}\right)^{n_0+n_1} \left(\frac{253}{255}\right)^{10000-n_0-n_1}. \end{aligned} \tag{5}$$

Problem 5.10.6 Solution

(a) Note that Z is the number of three-page faxes. In principle, we can sum the joint PMF $P_{X,Y,Z}(x, y, z)$ over all x, y to find $P_Z(z)$. However, it is better to realize that each fax has 3 pages with probability 1/6, independent of any other fax. Thus, Z has the binomial PMF

$$P_Z(z) = \binom{5}{z} (1/6)^z (5/6)^{5-z}. \tag{1}$$

(b) From the properties of the binomial distribution given in Appendix A, we know that $\mathrm{E}[Z] = 5(1/6)$.

(c) We want to find the conditional PMF of the number X of 1-page faxes and number Y of 2-page faxes given $Z = 2$ 3-page faxes. Note that given $Z = 2$, $X + Y = 3$. Hence for non-negative integers x, y satisfying $x + y = 3$,

$$P_{X,Y|Z}(x, y|2) = \frac{P_{X,Y,Z}(x, y, 2)}{P_Z(2)} = \frac{\frac{5!}{x!y!2!}(1/3)^x(1/2)^y(1/6)^2}{\binom{5}{2}(1/6)^2(5/6)^3}. \tag{2}$$

With some algebra, this simplifies to

$$P_{X,Y|Z}(x, y|2) = \frac{3!}{x!y!}(2/5)^x(3/5)^y. \tag{3}$$

Using the properties of the multinomial coefficient, the complete expression of the conditional PMF is

$$P_{X,Y|Z}(x, y|2) = \binom{3}{x, y}(2/5)^x(3/5)^y. \tag{4}$$

In the above expression, the condition $Z = 2$ implies $Y = 3 - X$. Thus, because $y = 3 - x$, the multinomial coefficient

$$\binom{3}{x, y} = \frac{3!}{x!y!} \tag{5}$$

equals the binomial coefficient

$$\binom{3}{x} = \frac{3!}{x!(3-x)!}. \tag{6}$$

Moreover, given $Z = 2$, the conditional PMF of X is

$$P_{X|Z}(x|2) = P_{X,Y|Z}(x, 3-x|2) = \binom{3}{x}(2/5)^x(3/5)^{3-x}. \tag{7}$$

That is, given $Z = 2$, there are 3 faxes left, each of which independently could be a 1-page fax. The conditonal PMF of the number of 1-page faxes is binomial where 2/5 is the conditional probability that a fax has 1 page given that it either has 1 page or 2 pages. Moreover given $X = x$ and $Z = 2$ we must have $Y = 3 - x$.

(d) Given $Z = 2$, the conditional PMF of X is binomial for 3 trials and success probability 2/5. The conditional expectation of X given $Z = 2$ is $\mathrm{E}[X|Z = 2] = 3(2/5) = 6/5$.

(e) There are several ways to solve this problem. The most straightforward approach is to realize that for integers $0 \le x \le 5$ and $0 \le y \le 5$, the event $\{X = x, Y = y\}$ occurs iff $\{X = x, Y = y, Z = 5 - (x + y)\}$. For the rest of this problem, we assume x and y are non-negative integers so that

$$\begin{aligned} P_{X,Y}(x, y) &= P_{X,Y,Z}(x, y, 5 - (x + y)) \\ &= \binom{5}{x, y, 5 - x - y}\left(\frac{1}{3}\right)^x\left(\frac{1}{2}\right)^y\left(\frac{1}{6}\right)^{5-x-y}. \end{aligned} \tag{8}$$

The above expression may seem unwieldy. To simplify the expression, we observe that

$$
\begin{aligned}
P_{X,Y}(x,y) &= P_{X,Y,Z}(x,y,5-x-y) \\
&= P_{X,Y|Z}(x,y|5-x+y)\, P_Z(5-x-y)\,.
\end{aligned} \tag{9}
$$

Using $P_Z(z)$ found in part (a), we can calculate $P_{X,Y|Z}(x,y|5-x-y)$ for $0 \leq x+y \leq 5$, integer valued.

$$
\begin{aligned}
P_{X,Y|Z}(x,y|5-x-y) &= \frac{P_{X,Y,Z}(x,y,5-x-y)}{P_Z(5-x-y)} \\
&= \binom{x+y}{x}\left(\frac{1/3}{1/2+1/3}\right)^x\left(\frac{1/2}{1/2+1/3}\right)^y \\
&= \binom{x+y}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^{(x+y)-x}.
\end{aligned} \tag{10}
$$

In the above expression, it is wise to think of $x+y$ as some fixed value. In that case, we see that given $x+y$ is a fixed value, X and Y have a joint PMF given by a binomial distribution in x. This should not be surprising since it is just a generalization of the case when $Z = 2$. That is, given that there were a fixed number of faxes that had either one or two pages, each of those faxes is a one page fax with probability $(1/3)/(1/2+1/3)$ and so the number of one page faxes should have a binomial distribution, Moreover, given the number X of one page faxes, the number Y of two page faxes is completely specified.

Finally, by rewriting $P_{X,Y}(x,y)$ given above, the complete expression for the joint PMF of X and Y is

$$
\begin{aligned}
&P_{X,Y}(x,y) \\
&\quad = \binom{5}{5-x-y}\left(\frac{1}{6}\right)^{5-x-y}\left(\frac{5}{6}\right)^{x+y}\binom{x+y}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^y.
\end{aligned} \tag{11}
$$

It is perhaps a matter of taste whether (8) or (11) is the best way to express the joint PMF $P_{X,Y}(x,y)$.

Problem 5.10.7 Solution

We could use Theorem 8.2 to skip several of the steps below. However, it is also nice to solve the problem from first principles.

(a) We first derive the CDF of V. Since each X_i is non-negative, V is non-negative, thus $F_V(v) = 0$ for $v < 0$. For $v \geq 0$, independence of the X_i yields

$$\begin{aligned} F_V(v) = \mathrm{P}[V \leq v] &= \mathrm{P}[\min(X_1, X_2, X_3) \leq v] \\ &= 1 - \mathrm{P}[\min(X_1, X_2, X_3) > v] \\ &= 1 - \mathrm{P}[X_1 > v, X_2 > v, X_3 > v] \\ &= 1 - \mathrm{P}[X_1 > v]\,\mathrm{P}[X_2 > v]\,\mathrm{P}[X_3 > v]. \end{aligned} \tag{1}$$

Note that independence of the X_i was used in the final step. Since each X_i is an exponential (λ) random variable, for $v \geq 0$,

$$\mathrm{P}[X_i > v] = \mathrm{P}[X > v] = 1 - F_X(v) = e^{-\lambda v}. \tag{2}$$

Thus,

$$F_V(v) = 1 - \left(e^{-\lambda v}\right)^3 = 1 - e^{-3\lambda v}. \tag{3}$$

The complete expression for the CDF of V is

$$F_V(v) = \begin{cases} 0 & v < 0, \\ 1 - e^{-3\lambda v} & v \geq 0. \end{cases} \tag{4}$$

By taking the derivative of the CDF, we obtain the PDF

$$f_V(v) = \begin{cases} 0 & v < 0, \\ 3\lambda e^{-3\lambda v} & v \geq 0. \end{cases} \tag{5}$$

We see that V is an exponential (3λ) random variable.

(b) The CDF of W is found in a similar way. Since each X_i is non-negative, W is non-negative, thus $F_W(w) = 0$ for $w < 0$. For $w \geq 0$, independence of the X_i yields

$$\begin{aligned} F_W(w) = \mathrm{P}[W \leq w] &= \mathrm{P}[\max(X_1, X_2, X_3) \leq w] \\ &= \mathrm{P}[X_1 \leq w, X_2 \leq w, X_3 \leq w] \\ &= \mathrm{P}[X_1 \leq w]\,\mathrm{P}[X_2 \leq w]\,\mathrm{P}[X_3 \leq w]. \end{aligned} \tag{6}$$

Since each X_i is an exponential (λ) random variable, for $w \geq 0$,

$$\mathrm{P}[X_i \leq w] = 1 - e^{-\lambda w}. \tag{7}$$

Thus, $F_W(w) = (1 - e^{-\lambda w})^3$ for $w \geq 0$. The complete expression for the CDF of Y is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ \left(1 - e^{-\lambda w}\right)^3 & w \geq 0. \end{cases} \tag{8}$$

By taking the derivative of the CDF, we obtain the PDF

$$f_W(w) = \begin{cases} 0 & w < 0, \\ 3(1 - e^{-\lambda w})^2 e^{-\lambda w} & w \geq 0. \end{cases} \tag{9}$$

Problem 5.10.8 Solution

Let X_i denote the finishing time of boat i. Since finishing times of all boats are iid Gaussian random variables with expected value 35 minutes and standard deviation 5 minutes, we know that each X_i has CDF

$$F_{X_i}(x) = \mathrm{P}[X_i \leq x] = \mathrm{P}\left[\frac{X_i - 35}{5} \leq \frac{x - 35}{5}\right] = \Phi\left(\frac{x - 35}{5}\right). \tag{1}$$

(a) The time of the winning boat is

$$W = \min(X_1, X_2, \ldots, X_{10}). \tag{2}$$

To find the probability that $W \leq 25$, we will find the CDF $F_W(w)$ since this will also be useful for part (c).

$$\begin{aligned} F_W(w) &= \mathrm{P}[\min(X_1, X_2, \ldots, X_{10}) \leq w] \\ &= 1 - \mathrm{P}[\min(X_1, X_2, \ldots, X_{10}) > w] \\ &= 1 - \mathrm{P}[X_1 > w, X_2 > w, \ldots, X_{10} > w]. \end{aligned} \tag{3}$$

Since the X_i are iid,

$$\begin{aligned} F_W(w) = 1 - \prod_{i=1}^{10} \mathrm{P}[X_i > w] &= 1 - (1 - F_{X_i}(w))^{10} \\ &= 1 - \left(1 - \Phi\left(\frac{w - 35}{5}\right)\right)^{10}. \end{aligned} \tag{4}$$

Thus,

$$\begin{aligned}
P[W \leq 25] = F_W(25) &= 1 - (1 - \Phi(-2))^{10} \\
&= 1 - [\Phi(2)]^{10} = 0.2056. && (5)
\end{aligned}$$

(b) The finishing time of the last boat is $L = \max(X_1, \ldots, X_{10})$. The probability that the last boat finishes in more than 50 minutes is

$$\begin{aligned}
P[L > 50] &= 1 - P[L \leq 50] \\
&= 1 - P[X_1 \leq 50, X_2 \leq 50, \ldots, X_{10} \leq 50]. && (6)
\end{aligned}$$

Once again, since the X_i are iid Gaussian $(35, 5)$ random variables,

$$\begin{aligned}
P[L > 50] = 1 - \prod_{i=1}^{10} P[X_i \leq 50] &= 1 - (F_{X_i}(50))^{10} \\
&= 1 - (\Phi([50 - 35]/5))^{10} \\
&= 1 - (\Phi(3))^{10} = 0.0134. && (7)
\end{aligned}$$

(c) A boat will finish in negative time if and only iff the winning boat finishes in negative time, which has probability

$$\begin{aligned}
F_W(0) &= 1 - (1 - \Phi(-35/5))^{10} \\
&= 1 - (1 - \Phi(-7))^{10} = 1 - (\Phi(7))^{10}. && (8)
\end{aligned}$$

Unfortunately, the tables in the text have neither $\Phi(7)$ nor $Q(7)$. However, those with access to MATLAB, or a programmable calculator, can find out that $Q(7) = 1 - \Phi(7) = 1.28 \times 10^{-12}$. This implies that a boat finishes in negative time with probability

$$F_W(0) = 1 - (1 - 1.28 \times 10^{-12})^{10} = 1.28 \times 10^{-11}. \qquad (9)$$

Problem 5.10.9 Solution

(a) This is straightforward:

$$
\begin{aligned}
F_{U_n}(u) &= \mathrm{P}\left[\max(X_1, \ldots, X_n) \le u\right] \\
&= \mathrm{P}\left[X_1 \le u, \ldots, X_n \le u\right] \\
&= \mathrm{P}\left[X_1 \le u\right] \mathrm{P}\left[X_2 \le u\right] \cdots \mathrm{P}\left[X_n \le u\right] = \left(F_X(u)\right)^n. \qquad (1)
\end{aligned}
$$

(b) This is also straightforward.

$$
\begin{aligned}
F_{L_n}(l) &= 1 - \mathrm{P}\left[\min(X_1, \ldots, X_n) > l\right] \\
&= 1 - \mathrm{P}\left[X_1 > l, \ldots, X_n > l\right] \\
&= 1 - \mathrm{P}\left[X_1 > l\right] \mathrm{P}\left[X_2 > l\right] \cdots \mathrm{P}\left[X_n > l\right] \\
&= 1 - \left(1 - F_X(l)\right)^n. \qquad (2)
\end{aligned}
$$

(c) This part is a little more difficult. The key is to identify the "easy" joint probability

$$
\begin{aligned}
&\mathrm{P}\left[L_n > l, U_n \le u\right] \\
&\quad = \mathrm{P}\left[\min(X_1, \ldots, X_n) \ge l, \max(X_1, \ldots, X_n) \le u\right] \\
&\quad = \mathrm{P}\left[l < X_i \le u, i = 1, 2, \ldots, n\right] \\
&\quad = \mathrm{P}\left[l < X_1 \le u\right] \cdots \mathrm{P}\left[l < X_n \le u\right] \\
&\quad = \left[F_X(u) - F_X(l)\right]^n. \qquad (3)
\end{aligned}
$$

Next we observe by the law of total probability that

$$
\mathrm{P}\left[U_n \le u\right] = \mathrm{P}\left[L_n > l, U_n \le u\right] + \mathrm{P}\left[L_n \le l, U_n \le u\right]. \qquad (4)
$$

The final term on the right side is the joint CDF we desire and using the expressions we derived for the first two terms, we obtain

$$
\begin{aligned}
F_{L_n,U_n}(l, u) &= \mathrm{P}\left[U_n \le u\right] - \mathrm{P}\left[L_n > l, U_n \le u\right] \\
&= \left[F_X(u)\right]^n - \left[F_X(u) - F_X(l)\right]^n. \qquad (5)
\end{aligned}
$$

Problem 5.10.10 Solution

(a) You accept suitcase i when $X_i > \tau_i$. The values of past suitcases are irrelevant given that you have opened suitcase i. Thus, given $X_i \geq \tau_i$, X_i has a conditional PDF of a continuous uniform (τ_i, m) random variable. Since a uniform (a, b) random variable has expected value $(a+b)/2$, we can conclude that

$$\mathrm{E}\left[X_i | X_i \geq \tau_i\right] = \frac{\tau_i + m}{2}. \tag{1}$$

(b) When there is exactly one remaining suitcase, we must accept whatever reward it offers. Thus $W_1 = X_1$ and $\mathrm{E}[W_1] = \mathrm{E}[X_1] = m/2$.

(c) In this case, we condition on whether $X_k \geq \tau_k$. Since $0 \leq \tau_k \leq m$, we can write

$$\begin{aligned}
\mathrm{E}\left[W_k\right] &= \mathrm{E}\left[X_k | X_k \geq \tau_k\right] \mathrm{P}\left[X_k \geq \tau_k\right] + \mathrm{E}\left[W_{k-1} | X_k < \tau_k\right] \mathrm{P}\left[X_k < \tau_k\right] \\
&= \mathrm{E}\left[X_k | X_k \geq \tau_k\right] \mathrm{P}\left[X_k \geq \tau_k\right] + \mathrm{E}\left[W_{k-1}\right] \mathrm{P}\left[X_k < \tau_k\right] \\
&= \frac{\tau_k + m}{2}\left(1 - \frac{\tau_k}{m}\right) + \mathrm{E}\left[W_{k-1}\right] \frac{\tau_k}{m} \\
&= \frac{m^2 - \tau_k^2}{2m} + \mathrm{E}\left[W_{k-1}\right] \frac{\tau_k}{m}.
\end{aligned} \tag{2}$$

(d) From the recursive relationship (2), we can find τ_k to maximize $\mathrm{E}[W_k]$. In particular, solving

$$\frac{d\mathrm{E}\left[W_k\right]}{d\tau_k} = -\frac{\tau_k}{m} + \frac{\mathrm{E}\left[W_{k-1}\right]}{m} = 0. \tag{3}$$

implies the optimal threshold is $\tau_k^* = \mathrm{E}[W_{k-1}]$. That is, the optimal policy is to accept suitcase k if the reward X_k is higher than the expected reward we would receive from the remaining $k-1$ suitcases if we were to reject suitcase k. With one suitcase left, the optimal policy is $\tau_1^* = 0$ since we don't reject the last suitcase. The optimal reward is $\mathrm{E}[W_1^*] = \mathrm{E}[X_1] = m/2$. For two suitcases left, the optimal threshold is $\tau_2^* = \mathrm{E}[W_1^*]$. Using (2) we can recursively optimize the rewards:

$$\begin{aligned}
\mathrm{E}\left[W_k^*\right] &= \frac{m^2 - (\tau_k^*)^2}{2m} + \mathrm{E}\left[W_{k-1}^*\right]\frac{\tau_k^*}{m} \\
&= \frac{m}{2} - \frac{(\mathrm{E}\left[W_{k-1}^*\right])^2}{2m} + \frac{(\mathrm{E}\left[W_{k-1}^*\right])^2}{m} \\
&= \frac{m}{2} + \frac{(\mathrm{E}\left[W_{k-1}^*\right])^2}{2m}. \qquad (4)
\end{aligned}$$

The recursion becomes more clear by defining $\mathrm{E}[W_k^*] = m\alpha_k$. Since the reward cannot exceed m dollars, we know that $0 \leq \alpha_k \leq 1$. In addition, it follows from (4) that

$$\alpha_k = \frac{1}{2} + \frac{\alpha_{k-1}^2}{2}. \qquad (5)$$

Since $\alpha_1 = 1/2$,

$$\alpha_2 = \frac{1}{2} + \frac{1}{8} = \frac{5}{8}, \qquad (6)$$

$$\alpha_3 = \frac{1}{2} + \frac{\alpha_2^2}{2} = \frac{89}{128}, \qquad (7)$$

$$\alpha_4 = \frac{1}{2} + \frac{\alpha_3^2}{2} = \frac{24,305}{32768} = 0.74. \qquad (8)$$

The optimal thresholds are $\tau_1^* = 0$ and for $k > 1$,

$$\tau_k^* = \mathrm{E}\left[W_{k-1}^*\right] = \alpha_{k-1}m. \qquad (9)$$

Thus,

$$\tau_1^* = 0, \qquad \tau_2^* = \frac{5}{8}m \qquad \tau_3^* = \frac{89}{128}m, \qquad \tau_4^* = \frac{24305}{32768}m. \qquad (10)$$

Note that $\lim_{k\to\infty} \alpha_k = 1$. That is, if the number of suitcases k goes to infinity, the optimal rewards and thresholds satisfy $\mathrm{E}[W_k^*] = \tau_{k-1}^* \to m$.

Problem 5.10.11 Solution

Since $U_1, \ldots, U_n$ are iid uniform $(0,1)$ random variables,

$$f_{U_1,\ldots,U_n}(u_1,\ldots,u_n) = \begin{cases} 1/T^n & 0 \leq u_i \leq 1; i = 1,2,\ldots,n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Since $U_1, \ldots, U_n$ are continuous, $\mathrm{P}[U_i = U_j] = 0$ for all $i \neq j$. For the same reason, $\mathrm{P}[X_i = X_j] = 0$ for $i \neq j$. Thus we need only to consider the case when $x_1 < x_2 < \cdots < x_n$.

To understand the claim, it is instructive to start with the $n = 2$ case. In this case, $(X_1, X_2) = (x_1, x_2)$ (with $x_1 < x_2$) if either $(U_1, U_2) = (x_1, x_2)$ or $(U_1, U_2) = (x_2, x_1)$. For infinitesimal Δ,

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2)\,\Delta^2 &= \mathrm{P}\left[x_1 < X_1 \leq x_1 + \Delta, x_2 < X_2 \leq x_2 + \Delta\right] \\ &= \mathrm{P}\left[x_1 < U_1 \leq x_1 + \Delta, x_2 < U_2 \leq x_2 + \Delta\right] \\ &\qquad + \mathrm{P}\left[x_2 < U_1 \leq x_2 + \Delta, x_1 < U_2 \leq x_1 + \Delta\right] \\ &= f_{U_1,U_2}(x_1,x_2)\,\Delta^2 + f_{U_1,U_2}(x_2,x_1)\,\Delta^2. \end{aligned} \tag{2}$$

We see that for $0 \leq x_1 < x_2 \leq 1$ that

$$f_{X_1,X_2}(x_1,x_2) = 2/T^n. \tag{3}$$

For the general case of n uniform random variables, we define

$$\boldsymbol{\pi} = \begin{bmatrix} \pi(1) & \ldots & \pi(n) \end{bmatrix}'. \tag{4}$$

as a permutation vector of the integers $1, 2, \ldots, n$ and Π as the set of $n!$ possible permutation vectors. In this case, the event

$$\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \tag{5}$$

occurs if

$$U_1 = x_{\pi(1)}, U_2 = x_{\pi(2)}, \ldots, U_n = x_{\pi(n)} \tag{6}$$

for any permutation $\boldsymbol{\pi} \in \Pi$. Thus, for $0 \leq x_1 < x_2 < \cdots < x_n \leq 1$,

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\,\Delta^n = \sum_{\boldsymbol{\pi} \in \Pi} f_{U_1,\ldots,U_n}\left(x_{\pi(1)},\ldots,x_{\pi(n)}\right)\Delta^n. \tag{7}$$

Since there are $n!$ permutations and $f_{U_1,\ldots,U_n}(x_{\pi(1)},\ldots,x_{\pi(n)}) = 1/T^n$ for each permutation $\boldsymbol{\pi}$, we can conclude that

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = n!/T^n. \tag{8}$$

Since the order statistics are necessarily ordered, $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 0$ unless $x_1 < \cdots < x_n$.

Problem 5.11.1 Solution

The script `imagepmf` in Example 5.26 generates the grid variables `SX`, `SY`, and `PXY`. Recall that for each entry in the grid, `SX`. `SY` and `PXY` are the corresponding values of x, y and $P_{X,Y}(x,y)$. Displaying them as adjacent column vectors forms the list of all possible pairs x, y and the probabilities $P_{X,Y}(x,y)$. Since any MATLAB vector or matrix `x` is reduced to a column vector with the command `x(:)`, the following simple commands will generate the list:

```
>> format rat;
>> imagepmf;
>> [SX(:) SY(:) PXY(:)]
ans =
    800          400          1/5
   1200          400          1/20
   1600          400           0
    800          800          1/20
   1200          800          1/5
   1600          800          1/10
    800         1200          1/10
   1200         1200          1/10
   1600         1200          1/5
>>
```

Note that the command `format rat` wasn't necessary; it just formats the output as rational numbers, i.e., ratios of integers, which you may or may not find esthetically pleasing.

Problem 5.11.2 Solution

In this problem, we need to calculate $\mathrm{E}[X]$, $\mathrm{E}[Y]$, the correlation $\mathrm{E}[XY]$, and the covariance $\mathrm{Cov}[X,Y]$ for random variables X and Y in Example 5.26. In this case, we can use the script `imagepmf.m` in Example 5.26 to generate the grid variables `SX`, `SY` and `PXY` that describe the joint PMF $P_{X,Y}(x,y)$.

However, for the rest of the problem, a general solution is better than a specific solution. The general problem is that given a pair of finite random variables described by the grid variables `SX`, `SY` and `PXY`, we want MATLAB to calculate an expected value $E[g(X,Y)]$. This problem is solved in a few simple steps. First we write a function that calculates the expected value of any finite random variable.

```
function ex=finiteexp(sx,px);
%Usage: ex=finiteexp(sx,px)
%returns the expected value E[X]
%of finite random variable X described
%by samples sx and probabilities px
ex=sum((sx(:)).*(px(:)));
```

Note that `finiteexp` performs its calculations on the sample values `sx` and probabilities `px` using the column vectors `sx(:)` and `px(:)`. As a result, we can use the same `finiteexp` function when the random variable is represented by grid variables. For example, we can calculate the correlation $r = E[XY]$ as

```
r=finiteexp(SX.*SY,PXY)
```

It is also convenient to define a function that returns the covariance:

```
function covxy=finitecov(SX,SY,PXY);
%Usage: cxy=finitecov(SX,SY,PXY)
%returns the covariance of
%finite random variables X and Y
%given by grids SX, SY, and PXY
ex=finiteexp(SX,PXY);
ey=finiteexp(SY,PXY);
R=finiteexp(SX.*SY,PXY);
covxy=R-ex*ey;
```

The following script calculates the desired quantities:

```
%imageavg.m
%Solution for Problem 4.12.2
imagepmf; %defines SX, SY, PXY
ex=finiteexp(SX,PXY)
ey=finiteexp(SY,PXY)
rxy=finiteexp(SX.*SY,PXY)
cxy=finitecov(SX,SY,PXY)
```

```
>> imageavg
ex =
        1180
ey =
   860
rxy =
     1064000
cxy =
       49200
>>
```

The careful reader will observe that `imagepmf` is inefficiently coded in that the correlation $\mathrm{E}[XY]$ is calculated twice, once directly and once inside of `finitecov`. For more complex problems, it would be worthwhile to avoid this duplication.

Problem 5.11.3 Solution

In this problem `randn(1,2)` generates a 1×2 array of independent Gaussian $(0,1)$ random variables. If this array is $\begin{bmatrix} X & Y \end{bmatrix}$, then $W = 4(X + Y)$ and

$$\mathrm{Var}[W] = \mathrm{Var}[4(X + Y)] = 16(\mathrm{Var}[X] + \mathrm{Var}[Y]) = 16(1 + 1) = 32.$$

Problem 5.11.4 Solution

The script is just a MATLAB calculation of $F_{X,Y}(x, y)$ in Equation (5.26).

```
%trianglecdfplot.m
[X,Y]=meshgrid(0:0.05:1.5);
R=(0<=Y).*(Y<=X).*(X<=1).*(2*(X.*Y)-(Y.^2));
R=R+((0<=X).*(X<Y).*(X<=1).*(X.^2));
R=R+((0<=Y).*(Y<=1).*(1<X).*((2*Y)-(Y.^2)));
R=R+((X>1).*(Y>1));
mesh(X,Y,R);
xlabel('\it x');
ylabel('\it y');
```

For functions like $F_{X,Y}(x, y)$ that have multiple cases, we calculate the function for each case and multiply by the corresponding boolean condition so as to have a zero contribution when that case doesn't apply. Using this technique, its important to define the boundary conditions carefully to make sure that no point is included in two different boundary conditions.

Problem 5.11.5 Solution

By following the formulation of Problem 5.2.6, the code to set up the sample grid is reasonably straightforward:

```
function [SX,SY,PXY]=circuits(n,p);
%Usage: [SX,SY,PXY]=circuits(n,p);
%   (See Problem 4.12.4)
[SX,SY]=ndgrid(0:n,0:n);
PXY=0*SX;
PXY(find((SX==n) & (SY==n)))=p^n;
for y=0:(n-1),
    I=find((SY==y) &(SX>=SY) &(SX<n));
    PXY(I)=(p^y)*(1-p)* ...
        binomialpmf(n-y-1,p,SX(I)-y);
end;
```

The only catch is that for a given value of y, we need to calculate the binomial probability of $x-y$ successes in $(n-y-1)$ trials. We can do this using the function call

```
binomialpmf(n-y-1,p,x-y)
```

However, this function expects the argument `n-y-1` to be a scalar. As a result, we must perform a separate call to `binomialpmf` for each value of `y`.

An alternate solution is direct calculation of the PMF $P_{X,Y}(x,y)$ in Problem 5.2.6. Here we calculate $m!$ using the MATLAB function `gamma(m+1)`. Because `gamma(x)` function will calculate the gamma function for each element in a vector `x`, we can calculate the PMF without any loops:

```
function [SX,SY,PXY]=circuits2(n,p);
%Usage: [SX,SY,PXY]=circuits2(n,p);
%   (See Problem 4.12.4)
[SX,SY]=ndgrid(0:n,0:n);
PXY=0*SX;
PXY(find((SX==n) & (SY==n)))=p^n;
I=find((SY<=SX) &(SX<n));
PXY(I)=(gamma(n-SY(I))./(gamma(SX(I)-SY(I)+1)...
      .*gamma(n-SX(I)))).*(p.^SX(I)).*((1-p).^(n-SX(I)));
```

Some experimentation with `cputime` will show that `circuits2(n,p)` runs much faster than `circuits(n,p)`. As is typical, the `for` loop in `circuit` results in time wasted running the MATLAB interpretor and in regenerating the binomial PMF in each cycle.

To finish the problem, we need to calculate the correlation coefficient

$$\rho_{X,Y} = \frac{\text{Cov}\left[X, Y\right]}{\sigma_X \sigma_Y}. \tag{1}$$

In fact, this is one of those problems where a general solution is better than a specific solution. The general problem is that given a pair of finite random variables described by the grid variables `SX`, `SY` and PMF `PXY`, we wish to calculate the correlation coefficient

This problem is solved in a few simple steps. First we write a function that calculates the expected value of a finite random variable.

```
function ex=finiteexp(sx,px);
%Usage: ex=finiteexp(sx,px)
%returns the expected value E[X]
%of finite random variable X described
%by samples sx and probabilities px
ex=sum((sx(:)).*(px(:)));
```

Note that `finiteexp` performs its calculations on the sample values `sx` and probabilities `px` using the column vectors `sx(:)` and `px(:)`. As a result, we can use the same `finiteexp` function when the random variable is represented by grid variables.We can build on `finiteexp` to calculate the variance using `finitevar`:

```
function v=finitevar(sx,px);
%Usage: ex=finitevar(sx,px)
%   returns the variance Var[X]
%   of finite random variables X described by
%   samples sx and probabilities px
ex2=finiteexp(sx.^2,px);
ex=finiteexp(sx,px);
v=ex2-(ex^2);
```

Putting these pieces together, we can calculate the correlation coefficient.

```
function rho=finitecoeff(SX,SY,PXY);
%Usage:  rho=finitecoeff(SX,SY,PXY)
%Calculate the correlation coefficient rho of
%finite random variables X and Y
ex=finiteexp(SX,PXY); vx=finitevar(SX,PXY);
ey=finiteexp(SY,PXY); vy=finitevar(SY,PXY);
R=finiteexp(SX.*SY,PXY);
rho=(R-ex*ey)/sqrt(vx*vy);
```

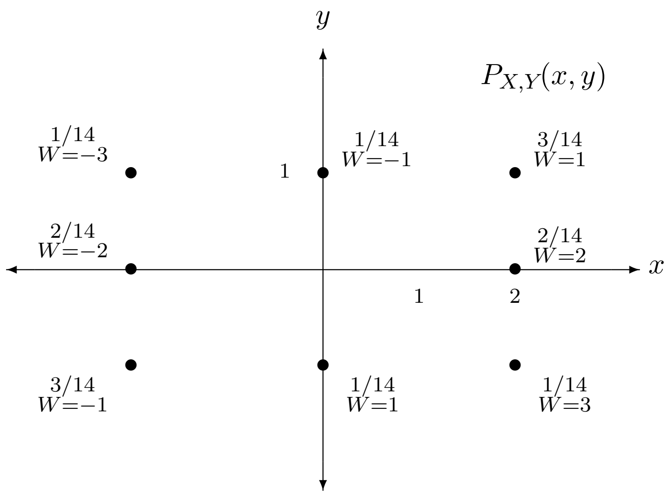
Calculating the correlation coefficient of X and Y, is now a two line exercise..

```
>> [SX,SY,PXY]=circuits2(50,0.9);
>> rho=finitecoeff(SX,SY,PXY)
rho =
    0.4451
>>
```

Problem Solutions – Chapter 6

Problem 6.1.1 Solution

In this problem, it is helpful to label possible points X, Y along with the corresponding values of $W = X - Y$. From the statement of Problem 6.1.1,



To find the PMF of W, we simply add the probabilities associated with each possible value of W:

$$P_W(-3) = P_{X,Y}(-2,1) = 1/14, \tag{1}$$

$$P_W(-2) = P_{X,Y}(-2,0) = 2/14, \tag{2}$$

$$P_W(-1) = P_{X,Y}(-2,-1) + P_{X,Y}(0,1) = 4/14, \tag{3}$$

$$P_W(1) = P_{X,Y}(0,-1) + P_{X,Y}(2,1) = 4/14, \tag{4}$$

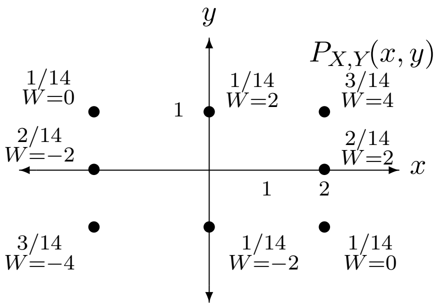
$$P_W(2) = P_{X,Y}(2,0) = 2/14, \tag{5}$$

$$P_W(3) = P_{X,Y}(2,1) = 1/14. \tag{6}$$

For all other values of w, $P_W(w) = 0$. A table for the PMF of W is

w	-3	-2	-1	1	2	3
$P_W(w)$	$1/14$	$2/14$	$4/14$	$4/14$	$2/14$	$1/14$

Problem 6.1.2 Solution



For this problem, we start by labeling each possible X, Y point with its probability $P_{X,Y}(x,y)$ and its corresponding value of $W = X + 2Y$.

From the above graph, we can calculate the probability of each possible value of w:

$$P_W(-4) = P_{X,Y}(-2,-1) = 3/14, \tag{1}$$
$$P_W(-2) = P_{X,Y}(-2,0) + P_{X,Y}(0,-1) = 3/14, \tag{2}$$
$$P_W(0) = P_{X,Y}(-2,1) + P_{X,Y}(2,-1) = 2/14, \tag{3}$$
$$P_W(2) = P_{X,Y}(0,1) + P_{X,Y}(2,0) = 3/14, \tag{4}$$
$$P_W(4) = P_{X,Y}(2,1) = 3/14. \tag{5}$$

We can summarize the PMF as

$$P_W(w) = \begin{cases} 3/14 & w = -4, -2, 2, 4, \\ 2/14 & w = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Problem 6.1.3 Solution

This is basically a trick problem. It looks like this problem should be in Section 6.5 since we have to find the PMF of the sum $L = N + M$. However, this problem is an special case since N and M are both binomial with the same success probability $p = 0.4$.

In this case, N is the number of successes in 100 independent trials with success probability $p = 0.4$. M is the number of successes in 50 independent trials with success probability $p = 0.4$. Thus $L = M + N$ is the number of successes in 150 independent trials with success probability $p = 0.4$. We conclude that L has the binomial $(n = 150, p = 0.4)$ PMF

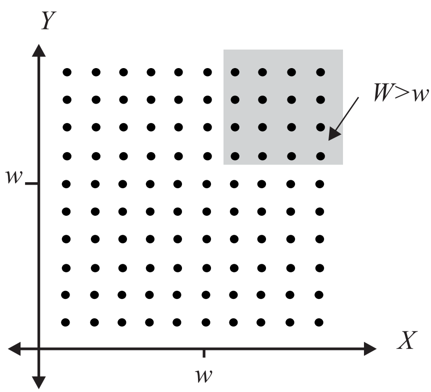
$$P_L(l) = \binom{150}{l}(0.4)^l(0.6)^{150-l}. \tag{1}$$

Problem 6.1.4 Solution

We observe that when $X = x$, we must have $Y = w - x$ in order for $W = w$. That is,

$$P_W(w) = \sum_{x=-\infty}^{\infty} \mathrm{P}[X = x, Y = w - x] = \sum_{x=-\infty}^{\infty} P_{X,Y}(x, w - x). \tag{1}$$

Problem 6.1.5 Solution



The x, y pairs with nonzero probability are shown in the figure. For $w = 0, 1, \ldots, 10$, we observe that

$$\begin{aligned} \mathrm{P}[W > w] &= \mathrm{P}[\min(X, Y) > w] \\ &= \mathrm{P}[X > w, Y > w] \\ &= 0.01(10 - w)^2. \end{aligned} \tag{1}$$

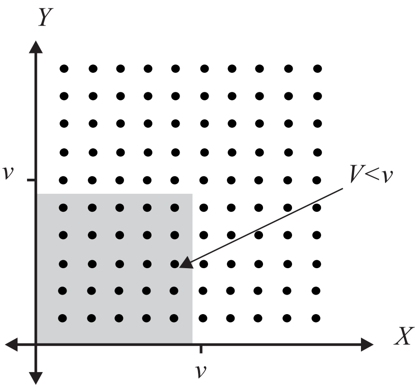
To find the PMF of W, we observe that for $w = 1, \ldots, 10$,

$$\begin{aligned} P_W(w) &= \mathrm{P}[W > w - 1] - \mathrm{P}[W > w] \\ &= 0.01[(10 - w - 1)^2 - (10 - w)^2] = 0.01(21 - 2w). \end{aligned} \tag{2}$$

The complete expression for the PMF of W is

$$P_W(w) = \begin{cases} 0.01(21 - 2w) & w = 1, 2, \ldots, 10, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 6.1.6 Solution



The x, y pairs with nonzero probability are shown in the figure. For $v = 1, \ldots, 11$, we observe that

$$
\begin{aligned}
\mathrm{P}[V < v] &= \mathrm{P}[\max(X, Y) < v] \\
&= \mathrm{P}[X < v, Y < v] \\
&= 0.01(v-1)^2. \qquad (1)
\end{aligned}
$$

To find the PMF of V, we observe that for $v = 1, \ldots, 10$,

$$
\begin{aligned}
P_V(v) &= \mathrm{P}[V < v+1] - \mathrm{P}[V < v] \\
&= 0.01[v^2 - (v-1)^2] \\
&= 0.01(2v-1). \qquad (2)
\end{aligned}
$$

The complete expression for the PMF of V is

$$
P_V(v) = \begin{cases} 0.01(2v-1) & v = 1, 2, \ldots, 10, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)
$$

Problem 6.2.1 Solution

Since $0 \leq X \leq 1$, $Y = X^2$ satisfies $0 \leq Y \leq 1$. We can conclude that $F_Y(y) = 0$ for $y < 0$ and that $F_Y(y) = 1$ for $y \geq 1$. For $0 \leq y < 1$,

$$
F_Y(y) = \mathrm{P}\left[X^2 \leq y\right] = \mathrm{P}[X \leq \sqrt{y}]. \qquad (1)
$$

Since $f_X(x) = 1$ for $0 \leq x \leq 1$, we see that for $0 \leq y < 1$,

$$
\mathrm{P}[X \leq \sqrt{y}] = \int_0^{\sqrt{y}} dx = \sqrt{y} \qquad (2)
$$

Hence, the CDF of Y is

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ \sqrt{y} & 0 \leq y < 1, \\ 1 & y \geq 1. \end{cases} \tag{3}$$

By taking the derivative of the CDF, we obtain the PDF

$$f_Y(y) = \begin{cases} 1/(2\sqrt{y}) & 0 \leq y < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 6.2.2 Solution

We start by finding the CDF $F_Y(y)$. Since $Y \geq 0$, $F_Y(y) = 0$ for $y < 0$. For $y \geq 0$,

$$\begin{aligned} F_Y(y) &= \mathrm{P}\left[|X| \leq y\right] \\ &= \mathrm{P}\left[-y \leq X \leq y\right] \\ &= \Phi(y) - \Phi(-y) = 2\Phi(y) - 1. \end{aligned} \tag{1}$$

It follows $f_Y(y) = 0$ for $y < 0$ and that for $y > 0$,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = 2f_X(y) = \frac{2}{\sqrt{2\pi}} e^{-y^2/2}. \tag{2}$$

The complete expression for the CDF of Y is

$$f_Y(y) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-y^2/2} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

From the definition of the expected value,

$$\begin{aligned} \mathrm{E}\left[Y\right] &= \int_{-\infty}^{\infty} y f_Y(y)\, dy \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y e^{-y^2/2}\, dy = \sqrt{\frac{2}{\pi}}\, e^{-y^2/2}\Big|_0^{\infty} = \sqrt{\frac{2}{\pi}}. \end{aligned} \tag{4}$$

Problem 6.2.3 Solution

Note that T has the continuous uniform PDF

$$f_T(t) = \begin{cases} 1/15 & 60 \le t \le 75, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The rider's maximum possible speed is $V = 3000/60 = 50$ km/hr while the rider's minimum speed is $V = 3000/75 = 40$ km/hr. For $40 \le v \le 50$,

$$\begin{aligned} F_V(v) &= \mathrm{P}\left[\frac{3000}{T} \le v\right] = \mathrm{P}\left[T \ge \frac{3000}{v}\right] \\ &= \int_{3000/v}^{75} \frac{1}{15}\,dt = \left.\frac{t}{15}\right|_{3000/v}^{75} = 5 - \frac{200}{v}. \end{aligned} \tag{2}$$

Thus the CDF, and via a derivative, the PDF are

$$F_V(v) = \begin{cases} 0 & v < 40, \\ 5 - 200/v & 40 \le v \le 50, \\ 1 & v > 50, \end{cases} \quad f_V(v) = \begin{cases} 0 & v < 40, \\ 200/v^2 & 40 \le v \le 50, \\ 0 & v > 50. \end{cases} \tag{3}$$

Problem 6.2.4 Solution

First, we observe that V achieves a maximum $V = 30$ when $W = -1$ and it achieves a minimum $V = 10$ when $W = 1$. Thus $10 \le V \le 30$. This implies that the CDF satisfies $F_V(v) = 0$ for $v \le 10$ and $F_V(v) = 1$ for $v \ge 30$. For $10 < v < 30$,

$$\begin{aligned} F_V(v) &= \mathrm{P}\left[20 - 10W^{1/3} \le v\right] \\ &= \mathrm{P}\left[W^{1/3} \ge 2 - \frac{v}{10}\right] \\ &= \mathrm{P}\left[W \ge 8\left(1 - \frac{v}{20}\right)^3\right]. \end{aligned} \tag{1}$$

Since W has the uniform $-1, 1)$ PDF

$$f_W(w) = \begin{cases} 1/2 & -1 \le w \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

it follows from (1) that for $10 \leq v \leq 30$,

$$F_V(v) = \int_{8(1-\frac{v}{20})^3}^{1} f_W(w)\, dw = \frac{1}{2} - 4(1 - v/20)^3. \tag{3}$$

The complete CDF is

$$F_V(v) = \begin{cases} 0 & v \leq 10, \\ 1/2 - 4(1 - v/20)^3 & 10 \leq v \leq 30, \\ 1 & v > 30. \end{cases} \tag{4}$$

Taking a derivative, we obtain

$$f_V(v) = \begin{cases} 3(1 - v/20)^2/5 & 10 \leq v \leq 30, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 6.2.5 Solution

Since X is non-negative, $W = X^2$ is also non-negative. Hence for $w < 0$, $f_W(w) = 0$. For $w \geq 0$,

$$\begin{aligned} F_W(w) = \mathrm{P}[W \leq w] &= \mathrm{P}\left[X^2 \leq w\right] \\ &= \mathrm{P}[X \leq w] \\ &= 1 - e^{-\lambda\sqrt{w}}. \end{aligned} \tag{1}$$

Taking the derivative with respect to w yields $f_W(w) = \lambda e^{-\lambda\sqrt{w}}/(2\sqrt{w})$. The complete expression for the PDF is

$$f_W(w) = \begin{cases} \frac{\lambda e^{-\lambda\sqrt{w}}}{2\sqrt{w}} & w \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 6.2.6 Solution

Let X denote the position of the pointer and Y denote the area within the arc defined by the stopping position of the pointer.

(a) If the disc has radius r, then the area of the disc is πr^2. Since the circumference of the disc is 1 and X is measured around the circumference, $Y = \pi r^2 X$. For example, when $X = 1$, the shaded area is the whole disc and $Y = \pi r^2$. Similarly, if $X = 1/2$, then $Y = \pi r^2/2$ is half the area of the disc. Since the disc has circumference 1, $r = 1/(2\pi)$ and

$$Y = \pi r^2 X = \frac{X}{4\pi}. \tag{1}$$

(b) The CDF of Y can be expressed as

$$F_Y(y) = \mathrm{P}[Y \le y] = \mathrm{P}\left[\frac{X}{4\pi} \le y\right] = \mathrm{P}[X \le 4\pi y] = F_X(4\pi y). \tag{2}$$

Therefore the CDF is

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ 4\pi y & 0 \le y \le \frac{1}{4\pi}, \\ 1 & y \ge \frac{1}{4\pi}. \end{cases} \tag{3}$$

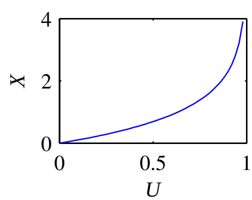
(c) By taking the derivative of the CDF, the PDF of Y is

$$f_Y(y) = \begin{cases} 4\pi & 0 \le y \le \frac{1}{4\pi}, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(d) The expected value of Y is $\mathrm{E}[Y] = \int_0^{1/(4\pi)} 4\pi y \, dy = 1/(8\pi)$.

Problem 6.2.7 Solution

Before solving for the PDF, it is helpful to have a sketch of the function $X = -\ln(1 - U)$.



(a) From the sketch, we observe that X will be nonnegative. Hence $F_X(x) = 0$ for $x < 0$. Since U has a uniform distribution on $[0, 1]$, for $0 \leq u \leq 1$, $\mathrm{P}[U \leq u] = u$. We use this fact to find the CDF of X. For $x \geq 0$,

$$\begin{aligned} F_X(x) &= \mathrm{P}\left[-\ln(1-U) \leq x\right] \\ &= \mathrm{P}\left[1 - U \geq e^{-x}\right] = \mathrm{P}\left[U \leq 1 - e^{-x}\right]. \end{aligned} \tag{1}$$

For $x \geq 0$, $0 \leq 1 - e^{-x} \leq 1$ and so

$$F_X(x) = F_U\left(1 - e^{-x}\right) = 1 - e^{-x}. \tag{2}$$

The complete CDF can be written as

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-x} & x \geq 0. \end{cases} \tag{3}$$

(b) By taking the derivative, the PDF is

$$f_X(x) = \begin{cases} e^{-x} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Thus, X has an exponential PDF. In fact, since most computer languages provide uniform $[0, 1]$ random numbers, the procedure outlined in this problem provides a way to generate exponential random variables from uniform random variables.

(c) Since X is an exponential random variable with parameter $a = 1$, $\mathrm{E}[X] = 1$.

Problem 6.2.8 Solution

We wish to find a transformation that takes a uniformly distributed random variable on [0,1] to the following PDF for Y.

$$f_Y(y) = \begin{cases} 3y^2 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We begin by realizing that in this case the CDF of Y must be

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ y^3 & 0 \leq y \leq 1, \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

Therefore, for $0 \leq y \leq 1$,

$$\mathrm{P}[Y \leq y] = \mathrm{P}[g(X) \leq y] = y^3. \tag{3}$$

Thus, using $g(X) = X^{1/3}$, we see that for $0 \leq y \leq 1$,

$$\mathrm{P}[g(X) \leq y] = \mathrm{P}\left[X^{1/3} \leq y\right] = \mathrm{P}\left[X \leq y^3\right] = y^3, \tag{4}$$

which is the desired answer.

Problem 6.2.9 Solution

Since X is constrained to the interval $[-1, 1]$, we see that $20 \leq Y \leq 35$. Thus , $F_Y(y) = 0$ for $y < 20$ and $F_Y(y) = 1$ fpr $y > 35$. For $20 \leq y \leq 35$,

$$\begin{aligned} F_Y(y) &= \mathrm{P}\left[20 + 15X^2 \leq y\right] \\ &= \mathrm{P}\left[X^2 \leq \frac{y-20}{15}\right] \\ &= \mathrm{P}\left[-\sqrt{\frac{y-20}{15}} \leq X \leq \sqrt{\frac{y-20}{15}}\right] \\ &= \int_{-\sqrt{\frac{y-20}{15}}}^{\sqrt{\frac{y-20}{15}}} \frac{1}{2}\,dx = \sqrt{\frac{y-20}{15}}. \end{aligned} \tag{1}$$

The complete expression for the CDF and, by taking the derivative, the PDF are

$$F_Y(y) = \begin{cases} 0 & y < 20, \\ \sqrt{\frac{y-20}{15}} & 20 \leq y \leq 35, \\ 1 & y > 35, \end{cases} \qquad f_Y(y) = \begin{cases} \frac{1}{\sqrt{60(y-20)}} & 20 \leq y \leq 35, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 6.2.10 Solution

Theorem 6.2 states that for a constant $a > 0$, $Y = aX$ has CDF and PDF

$$F_Y(y) = F_X(y/a), \qquad f_Y(y) = \frac{1}{a} f_X(y/a). \tag{1}$$

(a) If X is uniform (b, c), then $Y = aX$ has PDF

$$\begin{aligned} f_Y(y) = \frac{1}{a} f_X(y/a) &= \begin{cases} \frac{1}{a(c-b)} & b \leq y/a \leq c, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} \frac{1}{ac-ab} & ab \leq y \leq ac, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

Thus Y has the PDF of a uniform (ab, ac) random variable.

(b) Using Theorem 6.2, the PDF of $Y = aX$ is

$$\begin{aligned} f_Y(y) = \frac{1}{a} f_X(y/a) &= \begin{cases} \frac{\lambda}{a} e^{-\lambda(y/a)} & y/a \geq 0, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} (\lambda/a) e^{-(\lambda/a)y} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

Hence Y is an exponential (λ/a) exponential random variable.

(c) Using Theorem 6.2, the PDF of $Y = aX$ is

$$\begin{aligned} f_Y(y) = \frac{1}{a} f_X(y/a) &= \begin{cases} \frac{\lambda^n (y/a)^{n-1} e^{-\lambda(y/a)}}{a(n-1)!} & y/a \geq 0 \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \frac{(\lambda/a)^n y^{n-1} e^{-(\lambda/a)y}}{(n-1)!} & y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{4}$$

which is an Erlang (n, λ) PDF.

(d) If X is a Gaussian (μ,σ) random variable, then $Y = aX$ has PDF

$$
\begin{aligned}
f_Y(y) = f_X(y/a) &= \frac{1}{a\sqrt{2\pi\sigma^2}} e^{-((y/a)-\mu)^2/2\sigma^2} \\
&= \frac{1}{\sqrt{2\pi a^2\sigma^2}} e^{-(y-a\mu)^2/2(a^2\sigma^2)}. \qquad (5)
\end{aligned}
$$

Thus Y is a Gaussian random variable with expected value $\mathrm{E}[Y] = a\mu$ and $\mathrm{Var}[Y] = a^2\sigma^2$. That is, Y is a Gaussian $(a\mu, a\sigma)$ random variable.

Problem 6.2.11 Solution

If X has a uniform distribution from 0 to 1 then the PDF and corresponding CDF of X are

$$
f_X(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(x) = \begin{cases} 0 & x < 0, \\ x & 0 \le x \le 1, \\ 1 & x > 1. \end{cases} \qquad (1)
$$

For $b - a > 0$, we can find the CDF of the function $Y = a + (b-a)X$

$$
\begin{aligned}
F_Y(y) = \mathrm{P}[Y \le y] &= \mathrm{P}[a + (b-a)X \le y] \\
&= \mathrm{P}\left[X \le \frac{y-a}{b-a}\right] \\
&= F_X\left(\frac{y-a}{b-a}\right) = \frac{y-a}{b-a}. \qquad (2)
\end{aligned}
$$

Therefore the CDF of Y is

$$
F_Y(y) = \begin{cases} 0 & y < a, \\ \frac{y-a}{b-a} & a \le y \le b, \\ 1 & y \ge b. \end{cases} \qquad (3)
$$

By differentiating with respect to y we arrive at the PDF

$$
f_Y(y) = \begin{cases} 1/(b-a) & a \le x \le b, \\ 0 & \text{otherwise,} \end{cases} \qquad (4)
$$

which we recognize as the PDF of a uniform (a, b) random variable.

Problem 6.2.12 Solution

Since $X = F^{-1}(U)$, it is desirable that the function $F^{-1}(u)$ exist for all $0 \leq u \leq 1$. However, for the continuous uniform random variable U, $P[U = 0] = P[U = 1] = 0$. Thus, it is a zero probability event that $F^{-1}(U)$ will be evaluated at $U = 0$ or $U = 1$. Asa result, it doesn't matter whether $F^{-1}(u)$ exists at $u = 0$ or $u = 1$.

Problem 6.2.13 Solution

We can prove the assertion by considering the cases where $a > 0$ and $a < 0$, respectively. For the case where $a > 0$ we have

$$F_Y(y) = P[Y \leq y] = P\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right). \tag{1}$$

Therefore by taking the derivative we find that

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right), \qquad a > 0. \tag{2}$$

Similarly for the case when $a < 0$ we have

$$F_Y(y) = P[Y \leq y] = P\left[X \geq \frac{y-b}{a}\right] = 1 - F_X\left(\frac{y-b}{a}\right). \tag{3}$$

And by taking the derivative, we find that for negative a,

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right), \qquad a < 0. \tag{4}$$

A valid expression for both positive and negative a is

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \tag{5}$$

Therefore the assertion is proved.

Problem 6.2.14 Solution

Understanding this claim may be harder than completing the proof. Since $0 \leq F(x) \leq 1$, we know that $0 \leq U \leq 1$. This implies $F_U(u) = 0$ for $u < 0$ and

$F_U(u) = 1$ for $u \geq 1$. Moreover, since $F(x)$ is an increasing function, we can write for $0 \leq u \leq 1$,

$$F_U(u) = \mathrm{P}[F(X) \leq u] = \mathrm{P}\left[X \leq F^{-1}(u)\right] = F_X\left(F^{-1}(u)\right). \tag{1}$$

Since $F_X(x) = F(x)$, we have for $0 \leq u \leq 1$,

$$F_U(u) = F(F^{-1}(u)) = u. \tag{2}$$

Hence the complete CDF of U is

$$F_U(u) = \begin{cases} 0 & u < 0, \\ u & 0 \leq u < 1, \\ 1 & u \geq 1. \end{cases} \tag{3}$$

That is, U is a uniform $[0, 1]$ random variable.

Problem 6.3.1 Solution

From Problem 4.7.1, random variable X has CDF

$$F_X(x) = \begin{cases} 0 & x < -1, \\ x/3 + 1/3 & -1 \leq x < 0, \\ x/3 + 2/3 & 0 \leq x < 1, \\ 1 & 1 \leq x. \end{cases} \tag{1}$$

(a) We can find the CDF of Y, $F_Y(y)$ by noting that Y can only take on two possible values, 0 and 100. And the probability that Y takes on these two values depends on the probability that $X < 0$ and $X \geq 0$, respectively. Therefore

$$F_Y(y) = \mathrm{P}[Y \leq y] = \begin{cases} 0 & y < 0, \\ \mathrm{P}[X < 0] & 0 \leq y < 100, \\ 1 & y \geq 100. \end{cases} \tag{2}$$

The probabilities concerned with X can be found from the given CDF $F_X(x)$. This is the general strategy for solving problems of this type: to express the

CDF of Y in terms of the CDF of X. Since $\mathrm{P}[X < 0] = F_X(0^-) = 1/3$, the CDF of Y is

$$F_Y(y) = \mathrm{P}[Y \leq y] = \begin{cases} 0 & y < 0, \\ 1/3 & 0 \leq y < 100, \\ 1 & y \geq 100. \end{cases} \tag{3}$$

(b) The CDF $F_Y(y)$ has jumps of $1/3$ at $y = 0$ and $2/3$ at $y = 100$. The corresponding PDF of Y is

$$f_Y(y) = \delta(y)/3 + 2\delta(y - 100)/3. \tag{4}$$

(c) The expected value of Y is

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy = 0 \cdot \frac{1}{3} + 100 \cdot \frac{2}{3} = 66.66. \tag{5}$$

Problem 6.3.2 Solution

The PDF of T is

$$f_T(t) = \begin{cases} 1/10 & 50 \leq t < 60, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

If $T = 50$, then $R = 3000$. If $T = 60$, then $R = 3600$. Since R is an integer, it must be in the set $\{3000, 3001, \ldots, 3600\}$. However, we note that $R = 3000$ if and only if $T = 50$. This implies $\mathrm{P}[R = 3000] = \mathrm{P}[T = 50] = 0$ since T is a continuous random variable. Thus for integers $r \in S_R = \{3001, 3002, \ldots, 3600\}$, we observe that

$$\begin{aligned} \mathrm{P}[R = r] = \mathrm{P}[r - 1 < 60T \leq r] &= \mathrm{P}\left[\frac{r-1}{60} < T \leq \frac{r}{60}\right] \\ &= \int_{(r-1)/60}^{r/60} f_T(t)\, dt \\ &= \int_{(r-1)/60}^{r/60} \frac{1}{10}\, dt \\ &= \frac{r/60 - (r-1)/60}{10} = \frac{1}{600}. \end{aligned} \tag{2}$$

It follows that

$$P_R(r) = \begin{cases} 1/600 & r = 3001, 3002, \ldots, 3600, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 6.3.3 Solution

Since the microphone voltage V is uniformly distributed between -1 and 1 volts, V has PDF and CDF

$$f_V(v) = \begin{cases} 1/2 & -1 \leq v \leq 1, \\ 0 & \text{otherwise,} \end{cases} \qquad F_V(v) = \begin{cases} 0 & v < -1, \\ (v+1)/2 & -1 \leq v \leq 1, \\ 1 & v > 1. \end{cases} \tag{1}$$

The voltage is processed by a limiter whose output magnitude is given by below

$$L = \begin{cases} |V| & |V| \leq 0.5, \\ 0.5 & \text{otherwise.} \end{cases} \tag{2}$$

(a)

$$\begin{aligned} \mathrm{P}[L = 0.5] &= \mathrm{P}[|V| \geq 0.5] = \mathrm{P}[V \geq 0.5] + \mathrm{P}[V \leq -0.5] \\ &= 1 - F_V(0.5) + F_V(-0.5) \\ &= 1 - 1.5/2 + 0.5/2 = 1/2. \end{aligned} \tag{3}$$

(b) For $0 \leq l \leq 0.5$,

$$\begin{aligned} F_L(l) = \mathrm{P}[|V| \leq l] &= \mathrm{P}[-l \leq v \leq l] \\ &= F_V(l) - F_V(-l) \\ &= 1/2(l+1) - 1/2(-l+1) = l. \end{aligned} \tag{4}$$

So the CDF of L is

$$F_L(l) = \begin{cases} 0 & l < 0, \\ l & 0 \leq l < 0.5, \\ 1 & l \geq 0.5. \end{cases} \tag{5}$$

(c) By taking the derivative of $F_L(l)$, the PDF of L is

$$f_L(l) = \begin{cases} 1 + (0.5)\delta(l - 0.5) & 0 \le l \le 0.5, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The expected value of L is

$$\begin{aligned} \mathrm{E}[L] &= \int_{-\infty}^{\infty} l f_L(l)\, dl \\ &= \int_0^{0.5} l\, dl + 0.5 \int_0^{0.5} l(0.5)\delta(l - 0.5)\, dl = 0.375. \end{aligned} \tag{7}$$

Problem 6.3.4 Solution

The uniform $(0, 2)$ random variable U has PDF and CDF

$$f_U(u) = \begin{cases} 1/2 & 0 \le u \le 2, \\ 0 & \text{otherwise,} \end{cases} \qquad F_U(u) = \begin{cases} 0 & u < 0, \\ u/2 & 0 \le u < 2, \\ 1 & u > 2. \end{cases} \tag{1}$$

The uniform random variable U is subjected to the following clipper.

$$W = g(U) = \begin{cases} U & U \le 1, \\ 1 & U > 1. \end{cases} \tag{2}$$

To find the CDF of the output of the clipper, W, we remember that $W = U$ for $0 \le U \le 1$ while $W = 1$ for $1 \le U \le 2$. First, this implies W is nonnegative, i.e., $F_W(w) = 0$ for $w < 0$. Furthermore, for $0 \le w \le 1$,

$$F_W(w) = \mathrm{P}[W \le w] = \mathrm{P}[U \le w] = F_U(w) = w/2 \tag{3}$$

Lastly, we observe that it is always true that $W \le 1$. This implies $F_W(w) = 1$ for $w \ge 1$. Therefore the CDF of W is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ w/2 & 0 \le w < 1, \\ 1 & w \ge 1. \end{cases} \tag{4}$$

From the jump in the CDF at $w = 1$, we see that $P[W = 1] = 1/2$. The corresponding PDF can be found by taking the derivative and using the delta function to model the discontinuity.

$$f_W(w) = \begin{cases} 1/2 + (1/2)\delta(w-1) & 0 \leq w \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The expected value of W is

$$\begin{aligned} E[W] = \int_{-\infty}^{\infty} w f_W(w)\, dw &= \int_0^1 w[1/2 + (1/2)\delta(w-1)]\, dw \\ &= 1/4 + 1/2 = 3/4. \end{aligned} \tag{6}$$

Problem 6.3.5 Solution

Given the following function of random variable X,

$$Y = g(X) = \begin{cases} 10 & X < 0, \\ -10 & X \geq 0. \end{cases} \tag{1}$$

we follow the same procedure as in Problem 6.3.1. We attempt to express the CDF of Y in terms of the CDF of X. We know that Y is always less than -10. We also know that $-10 \leq Y < 10$ when $X \geq 0$, and finally, that $Y = 10$ when $X < 0$. Therefore

$$F_Y(y) = P[Y \leq y] = \begin{cases} 0 & y < -10, \\ P[X \geq 0] = 1 - F_X(0) & -10 \leq y < 10, \\ 1 & y \geq 10. \end{cases} \tag{2}$$

Problem 6.3.6 Solution

(a) The PDF and CDF of T are

$$f_T(t) = \begin{cases} (1/200)e^{-t/200} & t \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$F_T(t) = \begin{cases} 0 & t < 0, \\ 1 - e^{-t/200} & t \geq 0. \end{cases} \tag{2}$$

You are charged \$30 if $T \leq 300$, which has probability

$$\mathrm{P}[C = 30] = \mathrm{P}[T \leq 300] = F_T(300) = 1 - e^{-300/200} = 0.777. \tag{3}$$

(b) First we find the CDF $F_C(c)$. Since $C \geq 30$, $F_C(c) = 0$ for $c < 30$. When $T \geq 300$ and $C \geq 30$, we observe that $C = 30 + 0.50(T - 300)$. It follows that for $c \geq 30$,

$$\begin{aligned} F_C(c) = \mathrm{P}[C \leq c] &= \mathrm{P}[30 + 0.5(T - 300) \leq c] \\ &= \mathrm{P}[T \leq 2c + 240] = F_T(2c + 240). \end{aligned} \tag{4}$$

From the exponential CDF of T,

$$\begin{aligned} F_C(c) &= \begin{cases} 0 & c < 30, \\ 1 - e^{-(2c+240)/200} & c \geq 30, \end{cases} \\ &= \begin{cases} 0 & c < 30, \\ 1 - e^{-(c+120)/100} & c \geq 30. \end{cases} \end{aligned} \tag{5}$$

By taking the derivative, we obtain the PDF

$$f_C(c) = \begin{cases} (1 - e^{-3/2})\delta(c - 30) + (1/100)e^{-(c+120)/100} & c \geq 30, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

(c) We could use the PDF of C to write $\mathrm{E}[C] = \int_{-\infty}^{\infty} c f_C(c)\, dc$ but the resulting integral is prety messy. A slightly simpler way to do the integral is to view C as the function

$$C = C(T) = \begin{cases} 30 & T < 300, \\ 30 + (T - 300)/2 & T \geq 300. \end{cases} \tag{7}$$

The expected value of C is then

$$\begin{aligned} \mathrm{E}[C] &= \int_{-\infty}^{\infty} C(t) f_T(t)\ dt \\ &= \int_0^{300} 30 f_T(t)\ dt + \int_{300}^{\infty} [30 + (T - 300)/2] f_T(t)\ dt \\ &= \int_0^{\infty} 30 f_T(t)\ dt + \frac{1}{2} \int_{300}^{\infty} (t - 300) \frac{1}{200} e^{-t/200}\, dt. \end{aligned} \tag{8}$$

With the variable substitution $x = t - 300$, we obtain

$$\begin{aligned}
\mathrm{E}[C] &= 30 + \frac{1}{2}\int_0^\infty x \frac{1}{200} e^{-(x+300)/200}\,dx \\
&= 30 + \frac{e^{-3/2}}{2}\int_0^\infty x \frac{1}{200} e^{-x/200}\,dx \\
&= 30 + \frac{e^{-3/2}}{2} 200 \\
&= 30 + 100e^{-3/2} = 52.31. \qquad (9)
\end{aligned}$$

Note that the last integral equals 200 because its the expected value of an exponential ($\lambda = 1/200$) random variable. This problem shows that there is a lot of money in overages. Even though the customer thinks he is paying 10 cents per minute (300 minutes for \$30), in fact his average bill is \$52.31 while he uses 200 minutes on average per month. His average cost per minute is actually around 52.31/200, or about 26 cents per minute.

Problem 6.3.7 Solution

The PDF of U is

$$f_U(u) = \begin{cases} 1/2 & -1 \le u \le 1, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

Since $W \ge 0$, we see that $F_W(w) = 0$ for $w < 0$. Next, we observe that the rectifier output W is a mixed random variable since

$$\mathrm{P}[W = 0] = \mathrm{P}[U < 0] = \int_{-1}^0 f_U(u)\,du = 1/2. \qquad (2)$$

The above facts imply that

$$F_W(0) = \mathrm{P}[W \le 0] = \mathrm{P}[W = 0] = 1/2. \qquad (3)$$

Next, we note that for $0 < w < 1$,

$$F_W(w) = \mathrm{P}[U \le w] = \int_{-1}^w f_U(u)\,du = (w+1)/2. \qquad (4)$$

Finally, $U \leq 1$ implies $W \leq 1$, which implies $F_W(w) = 1$ for $w \geq 1$. Hence, the complete expression for the CDF is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ (w+1)/2 & 0 \leq w \leq 1, \\ 1 & w > 1. \end{cases} \tag{5}$$

By taking the derivative of the CDF, we find the PDF of W; however, we must keep in mind that the discontinuity in the CDF at $w = 0$ yields a corresponding impulse in the PDF.

$$f_W(w) = \begin{cases} (\delta(w)+1)/2 & 0 \leq w \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

From the PDF, we can calculate the expected value

$$\mathrm{E}[W] = \int_0^1 w(\delta(w)+1)/2\, dw = 0 + \int_0^1 (w/2)\, dw = 1/4. \tag{7}$$

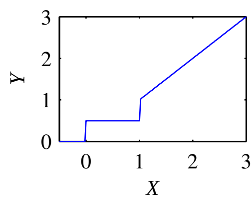
Perhaps an easier way to find the expected value is to use Theorem 3.10. In this case,

$$\mathrm{E}[W] = \int_{-\infty}^{\infty} g(u) f_W(w)\ du = \int_0^1 u(1/2)\, du = 1/4. \tag{8}$$

As we expect, both approaches give the same answer.

Problem 6.3.8 Solution

The relationship between X and Y is shown in the following figure:



(a) Note that $Y = 1/2$ if and only if $0 \le X \le 1$. Thus,

$$\mathrm{P}[Y = 1/2] = \mathrm{P}[0 \le X \le 1] = \int_0^1 f_X(x)\, dx = \int_0^1 (x/2)\, dx = 1/4. \tag{1}$$

(b) Since $Y \ge 1/2$, we can conclude that $F_Y(y) = 0$ for $y < 1/2$. Also, $F_Y(1/2) = \mathrm{P}[Y = 1/2] = 1/4$. Similarly, for $1/2 < y \le 1$,

$$F_Y(y) = \mathrm{P}[0 \le X \le 1] = \mathrm{P}[Y = 1/2] = 1/4. \tag{2}$$

Next, for $1 < y \le 2$,

$$F_Y(y) = \mathrm{P}[X \le y] = \int_0^y f_X(x)\, dx = y^2/4 \tag{3}$$

Lastly, since $Y \le 2$, $F_Y(y) = 1$ for $y \ge 2$. The complete expression of the CDF is

$$F_Y(y) = \begin{cases} 0 & y < 1/2, \\ 1/4 & 1/2 \le y \le 1, \\ y^2/4 & 1 < y < 2, \\ 1 & y \ge 2. \end{cases} \tag{4}$$

Problem 6.3.9 Solution

You may find it helpful to plot W as a function of V for the following calculations. We start by finding the CDF $F_W(w) = \mathrm{P}[W \le w]$. Since $0 \le W \le 10$, we know that

$$F_W(w) = 0 \qquad (w < 0) \tag{1}$$

and that

$$F_W(w) = 1 \qquad (w \ge 10). \tag{2}$$

Next we recall that continuous uniform V has the CDF

$$F_V(v) = \begin{cases} 0 & v < -15, \\ (v+15)/30 & -15 \le v \le 15, \\ 1 & v > 15. \end{cases} \tag{3}$$

Now we can write for $w = 0$ that

$$F_W(0) = P[V \leq 0] = F_V(0) = 15/30 = 1/2. \tag{4}$$

For $0 < w < 10$,

$$F_W(w) = P[W \leq w] = P[V \leq w] = F_V(w) = \frac{w + 15}{30}. \tag{5}$$

Thus the complete CDF of W is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ (w + 15)/30 & 0 \leq w < 10, \\ 1 & w \geq 10. \end{cases} \tag{6}$$

If you study (and perhaps plot) $F_W(w)$, you'll see that it has a jump discontinuity of height $1/2$ at $w = 0$ and also has a second jump of height $1/6$ at $w = 10$. Thus when we take a derivative of the CDF, we obtain the PDF

$$f_W(w) = \begin{cases} (1/2)\delta(w) & w = 0, \\ 1/30 & 0 < w < 10, \\ (1/6)\delta(w - 10) & w = 10, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Problem 6.3.10 Solution

Note that X has PDF

$$f_X(x) = \begin{cases} 1/4 & -2 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) We first find the CDF $F_Y(y)$. We start by observing that $0 \leq Y \leq 36$. This implies $F_Y(y) = 0$ for $y < 0$ and $F_Y(y) = 1$ for $y \geq 36$. For $0 \leq y < 36$,

$$\begin{aligned} F_Y(y) = P[Y \leq y] &= P\left[9X^2 \leq y\right] \\ &= P[-\sqrt{y}/3 \leq X \leq \sqrt{y}/3] \\ &= \int_{-\sqrt{y}/3}^{\sqrt{y}/3} \frac{1}{4}\, dx = \left.\frac{x}{4}\right|_{-\sqrt{y}/3}^{\sqrt{y}/3} = \frac{\sqrt{y}}{6}. \end{aligned} \tag{2}$$

The complete expression for the CDF is

$$F_Y(y) = \begin{cases} 0 & y \leq 0, \\ \sqrt{y}/6 & 0 < y \leq 36, \\ 1 & y > 36. \end{cases} \tag{3}$$

Taking the derivative of the CDF, we find the PDF is

$$f_Y(y) = \begin{cases} 1/[12\sqrt{y}] & 0 < y \leq 36, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(b) To solve this part, you may find it helpful to sketch W as a function of Y. As always, we first find the CDF of W. We observe that $0 \leq W \leq 16$ so that $F_W(w) = 0$ for $w < 0$ and $F_W(w) = 1$ for $w \geq 16$. For $0 \leq w < 16$,

$$F_W(w) = \mathrm{P}[W \leq w] = \mathrm{P}[Y \leq w] = F_Y(w) = \sqrt{w}/6. \tag{5}$$

The complete expression for the CDF of W is

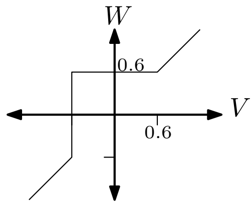
$$F_W(w) = \begin{cases} 0 & w \leq 0, \\ \sqrt{w}/6 & 0 < w < 16, \\ 1 & w \geq 16. \end{cases} \tag{6}$$

Before taking the derivative to find the PDF, we observe that there is a jump discontinuity in $F_W(w)$ at $w = 16$ since $F_W(16^-) = 4/6$ but $F_W(16) = 1$. This creates an impulse with weight $1 - 4/6 = 1/3$ at $w = 16$ in the PDF so that

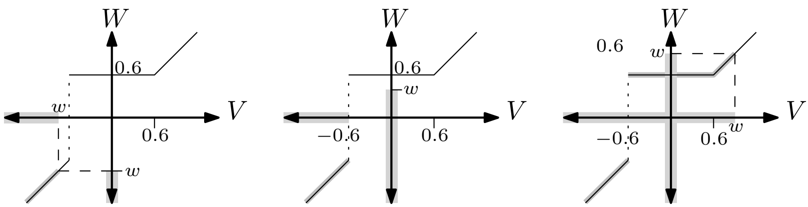
$$f_W(w) = \begin{cases} 1/[12\sqrt{w}] & 0 \leq w < 16, \\ (1/3)\delta(w-16) & w = 16, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Problem 6.3.11 Solution

A key to this problem is recognizing that W is a mixed random variable. Here is the mapping from V to W:



To find the CDF $F_W(w) = \mathrm{P}[W \leq w]$, careful study of the function shows there are three different cases for w that must be considered:



(a) $w < -0.6$ **(b)** $-0.6 \leq w < 0.6$ **(c)** $w \geq 0.6$

(a) If $w < -0.6$, then the event $\{W \leq w\}$, shown as the highlighted range on the vertical axis of graph (a) corresponds to the event that the the pair (V, W) is on the gray highlighted segment of the function $W = g(V)$, which corresponds to the event $\{V \leq w\}$. In this case, $F_W(w) = \mathrm{P}[V \leq w] = F_V(w)$.

(b) If $-0.6 < w < 0.6$, then the event $\{W \leq w\}$, shown as the highlighted range on the vertical axis of graph (b) corresponds to the event that the pair (V, W) is on the gray highlighted segment of the function $W = g(V)$, which corresponds to the event $\{V < 0.6\}$. In this case, $F_W(w) = \mathrm{P}[V < 0.6] = F_V(0.6^-)$.

(c) If $w \geq 0.6$, then the event $\{W \leq w\}$, shown as the highlighted range on the vertical axis of graph (c) corresponds to the event that the pair (V, W) is on the gray highlighted segment of the function $W = g(V)$, which now includes pairs v, w on the horizontal segment such that $w = 0.6$, and this corresponds to the event $\{V \leq w\}$. In this case, $F_W(w) = \mathrm{P}[V \leq w] = F_V(w)$.

We combine these three cases in the CDF

$$F_W(w) = \begin{cases} F_V(w) & w < -0.6, \\ F_V(0.6^-) & -0.6 \leq w < 0.6, \\ F_V(w) & w \geq 0.6. \end{cases} \tag{1}$$

Thus far, our answer is valid for any CDF $F_V(v)$. Now we specialize the result to the given CDF for V. Since V is a continuous uniform $(-5, 5)$ random variable, V has CDF

$$F_V(v) = \begin{cases} 0 & v < -5, \\ (v+5)/10 & -5 \le v \le 5, \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

The given V causes the case $w < 0.6$ to split into two cases: $w < -5$ and $-5 \le w < 0.6$. Similarly, the $w \ge 0.6$ case gets split into two cases. Applying this CDF to Equation (1), we obtain

$$F_W(w) = \begin{cases} 0 & w < -5, \\ (w+5)/10 & -5 \le w < -0.6, \\ 0.44 & -0.6 \le w < 0.6, \\ (w+5)/10 & 0.6 \le w < 5, \\ 1 & w > 1. \end{cases} \tag{3}$$

In this CDF, there is a jump from 0.44 to 0.56 at $w = 0.6$. This jump of height 0.12 corresponds precisely to $\mathrm{P}[W = 0.6] = 0.12$.

Since the CDF has a jump at $w = 0.6$, we have an impulse at $w = 0$ when we take the derivative:

$$f_W(w) = \begin{cases} 0.1 & -5 \le w < -0.6, \\ 0 & -0.6 \le w < 0.6, \\ 0.12\delta(w-0.6) & w = 0.6, \\ 0.1 & 0.6 < w \le 5, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 6.3.12 Solution

(a) Note that $\hat{X} = -c$ with probability $\mathrm{P}[X < 0] = 1/2$. The other possibility is $\hat{X} = c$ with probability $\mathrm{P}[X \ge 0] = 1/2$. No other values of $\hat{X}$ are possible and thus $\hat{X}$ has PMF

$$P_{\hat{X}}(x) = \begin{cases} 1/2 & x = -c, c, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) Note that X has the uniform $(-3, 3)$ PDF

$$f_X(x) = \begin{cases} 1/6 & -3 \leq x \leq 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

If this problem isn't clear, you should plot the function $g(x)$. The expected distortion is

$$\begin{aligned} \mathrm{E}[D] = \mathrm{E}[d(X)] = \mathrm{E}\left[(X - g(X))^2\right] &= \int_{-\infty}^{\infty} (x - g(x))^2 f_X(x)\, dx \\ &= \frac{1}{6}\int_{-3}^{0} (x + c)^2\, dx + \frac{1}{6}\int_{0}^{3} (x - c)^2\, dx \\ &= \frac{1}{18}(x + c)^3\Big|_{-3}^{0} + \frac{1}{18}(x - c)^3\Big|_{0}^{3} \\ &= \frac{1}{9}\left(c^3 + (3 - c)^3\right). \end{aligned} \tag{3}$$

To find the value of c that minimizes $\mathrm{E}[d(X)]$, we set the derivative $d\,\mathrm{E}[D]/dc$ to zero yielding $c^2 - (3 - c)^2 = 0$. Ths implies $c = 3 - c$ or $c = 3/2$. This should be intuitively pleasing, because for positive values of X, all values in the interval $[0, 3]$ are equally likely. The value of c that best approximates a positive value of X is the middle value $c = 3/2$.

(c) Here you were supposed to know that a continuous uniform (a, b) random variable has expected value $(a + b)/2$ and variance $(b - a)^2/12$. Thus X has expected value and variance

$$\mathrm{E}[X] = \frac{-3 + 3}{2} = 0, \qquad \sigma_X^2 = \frac{(3 - -3)^2}{12} = 3. \tag{4}$$

Thus Y is a Gaussian $(0, \sqrt{3})$ random variable with PDF

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-y^2/2\sigma_Y^2} = \frac{1}{\sqrt{6\pi}} e^{-y^2/6}. \tag{5}$$

(d) This part is actually hard. The expected distortion is

$$\begin{aligned}
E[D] &= E\left[(Y - g(Y))^2\right] \\
&= \int_{-\infty}^{\infty} (y - g(y))^2 f_Y(y)\, dy \\
&= \int_{-\infty}^{0} (y + c)^2 f_Y(y)\, dy + \int_{0}^{\infty} (y - c)^2 f_Y(y)\, dy.
\end{aligned} \tag{6}$$

Before plugging in the Gaussian PDF $f_Y(y)$, it's best to try and simplify as much as possible.

$$\begin{aligned}
E[D] &= \int_{-\infty}^{0} (y^2 + 2cy + c^2) f_Y(y)\, dy + \int_{0}^{\infty} (y^2 - 2cy + c^2) f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} (y^2 + c^2) f_Y(y)\, dy + 2c\left(\int_{-\infty}^{0} y f_Y(y)\, dy - \int_{0}^{\infty} y f_Y(y)\, dy\right) \\
&= E\left[Y^2\right] + c^2 + 2c\left(-\int_{0}^{\infty} u f_Y(u)\, du - \int_{0}^{\infty} y f_Y(y)\, dy\right) \\
&= E\left[Y^2\right] + c^2 - 4c\int_{0}^{\infty} y f_Y(y)\, dy.
\end{aligned} \tag{7}$$

where we used the substitution $u = -y$ and the fact that $f_Y(-u) = f_Y(u)$ in the integral from $-\infty$ to 0. Since $E[Y] = 0$, $E[Y^2] = \sigma_Y^2$. Also,

$$\begin{aligned}
E[D] &= \sigma_Y^2 + c^2 - 4c\frac{1}{\sqrt{2\pi\sigma_Y^2}} \int_{0}^{\infty} y e^{-y^2/2\sigma_Y^2}\, dy \\
&= \sigma_Y^2 + c^2 + 4c\frac{1}{\sqrt{2\pi\sigma_Y^2}} \sigma_Y^2 e^{-y^2/2\sigma_Y^2}\Big|_0^{\infty} \\
&= \sigma_Y^2 + c^2 - 4c\frac{\sigma_Y}{\sqrt{2\pi}}.
\end{aligned} \tag{8}$$

To find the minimizing c, we set the derivative $d\,\mathrm{E}[D]/dc = 0$, yielding

$$2c - 4\frac{\sigma_Y}{\sqrt{2\pi}} = 0. \tag{9}$$

Thus, the minimizing c is

$$c = \sigma_Y \sqrt{\frac{2}{\pi}} = \sqrt{\frac{6}{\pi}}. \tag{10}$$

Problem 6.3.13 Solution

(a) Given $F_X(x)$ is a continuous function, there exists x_0 such that $F_X(x_0) = u$. For each value of u, the corresponding x_0 is unique. To see this, suppose there were also x_1 such that $F_X(x_1) = u$. Without loss of generality, we can assume $x_1 > x_0$ since otherwise we could exchange the points x_0 and x_1. Since $F_X(x_0) = F_X(x_1) = u$, the fact that $F_X(x)$ is nondecreasing implies $F_X(x) = u$ for all $x \in [x_0, x_1]$, i.e., $F_X(x)$ is flat over the interval $[x_0, x_1]$, which contradicts the assumption that $F_X(x)$ has no flat intervals. Thus, for any $u \in (0,1)$, there is a unique x_0 such that $F_X(x) = u$. Moreiver, the same x_0 is the minimum of all x' such that $F_X(x') \geq u$. The uniqueness of x_0 such that $F_X(x)x_0 = u$ permits us to define $\tilde{F}(u) = x_0 = F_X^{-1}(u)$.

(b) In this part, we are given that $F_X(x)$ has a jump discontinuity at x_0. That is, there exists $u_0^- = F_X(x_0^-)$ and $u_0^+ = F_X(x_0^+)$ with $u_0^- < u_0^+$. Consider any u in the interval $[u_0^-, u_0^+]$. Since $F_X(x_0) = F_X(x_0^+)$ and $F_X(x)$ is nondecreasing,

$$F_X(x) \geq F_X(x_0) = u_0^+, \qquad x \geq x_0. \tag{1}$$

Moreover,

$$F_X(x) < F_X\left(x_0^-\right) = u_0^-, \qquad x < x_0. \tag{2}$$

Thus for any u satisfying $u_o^- \leq u \leq u_0^+$, $F_X(x) < u$ for $x < x_0$ and $F_X(x) \geq u$ for $x \geq x_0$. Thus, $\tilde{F}(u) = \min\{x | F_X(x) \geq u\} = x_0$.

(c) We note that the first two parts of this problem were just designed to show the properties of $\tilde{F}(u)$. First, we observe that

$$\mathrm{P}\left[\hat{X} \leq x\right] = \mathrm{P}\left[\tilde{F}(U) \leq x\right] = \mathrm{P}\left[\min\left\{x' | F_X\left(x'\right) \geq U\right\} \leq x\right]. \tag{3}$$

To prove the claim, we define, for any x, the events

$$A: \quad \min\left\{x' | F_X\left(x'\right) \geq U\right\} \leq x, \tag{4}$$
$$B: \quad U \leq F_X\left(x\right). \tag{5}$$

Note that $\mathrm{P}[A] = \mathrm{P}[\hat{X} \leq x]$. In addition, $\mathrm{P}[B] = \mathrm{P}[U \leq F_X(x)] = F_X(x)$ since $\mathrm{P}[U \leq u] = u$ for any $u \in [0,1]$.

We will show that the events A and B are the same. This fact implies

$$\mathrm{P}\left[\hat{X} \leq x\right] = \mathrm{P}\left[A\right] = \mathrm{P}\left[B\right] = \mathrm{P}\left[U \leq F_X\left(x\right)\right] = F_X\left(x\right). \tag{6}$$

All that remains is to show A and B are the same. As always, we need to show that $A \subset B$ and that $B \subset A$.

- To show $A \subset B$, suppose A is true and $\min\{x' | F_X(x') \geq U\} \leq x$. This implies there exists $x_0 \leq x$ such that $F_X(x_0) \geq U$. Since $x_0 \leq x$, it follows from $F_X(x)$ being nondecreasing that $F_X(x_0) \leq F_X(x)$. We can thus conclude that

$$U \leq F_X\left(x_0\right) \leq F_X\left(x\right). \tag{7}$$

  That is, event B is true.

- To show $B \subset A$, we suppose event B is true so that $U \leq F_X(x)$. We define the set

$$L = \left\{x' | F_X\left(x'\right) \geq U\right\}. \tag{8}$$

  We note $x \in L$. It follows that the minimum element satisfies $\min\{x' | x' \in L\}$ x. That is,

$$\min\left\{x' | F_X\left(x'\right) \geq U\right\} \leq x, \tag{9}$$

  which is simply event A.

Problem 6.4.1 Solution

Since $0 \leq X \leq 1$, and $0 \leq Y \leq 1$, we have $0 \leq V \leq 1$. This implies $F_V(v) = 0$ for $v < 0$ and $F_V(v) = 1$ for $v \geq 1$. For $0 \leq v \leq 1$,

$$\begin{aligned} F_V(v) = \mathrm{P}[\max(X,Y) \leq v] &= \mathrm{P}[X \leq v, Y \leq v] \\ &= \int_0^v \int_0^v 6xy^2\,dx\,dy \\ &= \left(\int_0^v 2x\,dx\right)\left(\int_0^v 3y^2\,dy\right) \\ &= (v^2)(v^3) = v^5. \end{aligned} \tag{1}$$

The CDF and (by taking the derivative) PDF of V are

$$F_V(v) = \begin{cases} 0 & v < 0, \\ v^5 & 0 \leq v \leq 1, \\ 1 & v > 1, \end{cases} \qquad f_V(v) = \begin{cases} 5v^4 & 0 \leq v \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 6.4.2 Solution

Since $0 \leq X \leq 1$, and $0 \leq Y \leq 1$, we have $0 \leq W \leq 1$. This implies $F_W(w) = 0$ for $w < 0$ and $F_W(w) = 1$ for $w \geq 1$. For $0 \leq w \leq 1$,

$$\begin{aligned} F_W(w) &= \mathrm{P}[\min(X,Y) \leq w] \\ &= 1 - \mathrm{P}[\min(X,Y) \geq w] = 1 - \mathrm{P}[X \geq w, Y \geq w]. \end{aligned} \tag{1}$$

Now we calculate

$$\begin{aligned} \mathrm{P}[X \geq w, Y \geq w] &= \int_w^1 \int_w^1 6xy^2\,dx\,dy \\ &= \left(\int_w^1 2x\,dx\right)\left(\int_w^1 3y^2\,dy\right) \\ &= (1-w^2)(1-w^3) = 1 - w^2 - w^3 + w^5. \end{aligned} \tag{2}$$

The complete expression for the CDF of W is

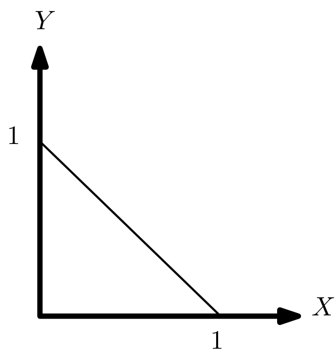
$$F_W(w) = 1 - P[X \geq w, Y \geq w]$$
$$= \begin{cases} 0 & w < 0, \\ w^2 + w^3 - w^5 & 0 \leq w \leq 1, \\ 1 & w > 1. \end{cases} \tag{3}$$

Taking the derivative of the CDF, we obtain the PDF

$$f_W(w) = \begin{cases} 2w + 3w^2 - 5w^4 & 0 \leq w \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 6.4.3 Solution

The key to the solution is to draw the triangular region where the PDF is nonzero:

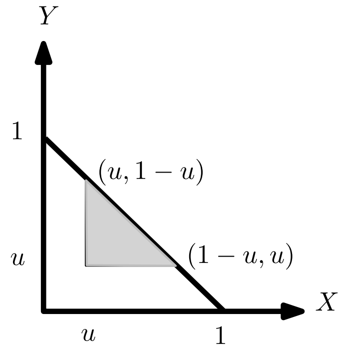


(a) X and Y are not independent. For example, note that $f_X(3/4) = f_Y(3/4) > 0$ and thus $f_X(3/4)f_Y(3/4) > 0$. However, $f_{X,Y}(3/4, 3/4) = 0$.

(b) First we find the CDF. Since $X \geq 0$ and $Y \geq 0$, we know that $F_U(u) = 0$ for $u < 0$. Next, for non-negative u, we see that

$$F_U(u) = P[\min(X, Y) \leq u] = 1 - P[\min(X, Y) > u]$$
$$= 1 - P[X > u, Y > u]. \tag{1}$$

At this point is is instructive to draw the region for small u:



We see that this area exists as long as $u \leq 1-u$, or $u \leq 1/2$. This is because if both $X > 1/2$ and $Y > 1/2$ then $X + Y > 1$ which violates the constraint $X + Y \leq 1$. For $0 \leq u \leq 1/2$,

$$\begin{aligned} F_U(u) &= 1 - \int_u^{1-u} \int_u^{1-x} 2\, dy\, dx \\ &= 1 - 2\frac{1}{2}[(1-u)-u]^2 = 1 - [1-2u]^2. \end{aligned} \tag{2}$$

Note that we wrote the integral expression but we calculated the integral as c times the area of integration. Thus the CDF of U is

$$F_U(u) = \begin{cases} 0 & u < 0, \\ 1 - [1-2u]^2 & 0 \leq u \leq 1/2, \\ 1 & u > 1/2. \end{cases} \tag{3}$$

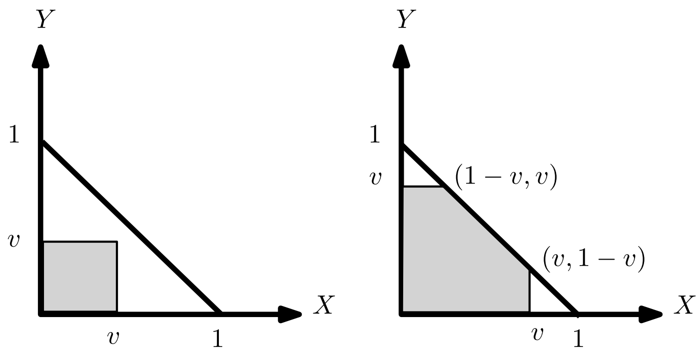
Taking the derivative, we find the PDF of U is

$$F_U(u) = \begin{cases} 4(1-2u) & 0 \leq u \leq 1/2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(c) For the CDF of V, we can write

$$\begin{aligned} F_V(v) = \mathrm{P}[V \leq v] &= \mathrm{P}[\max(X,Y) \leq v] \\ &= \mathrm{P}[X \leq v, Y \leq v] \\ &= \int_0^v \int_0^v f_{X,Y}(x,y)\; dx, dy. \end{aligned} \tag{5}$$

This is tricky because there are two distinct cases:



$0 \leq v \leq 1/2$ $\qquad$ $1/2 \leq v \leq 1$

For $0 \leq v \leq 1/2$,

$$F_V(v) = \int_0^v \int_0^v 2\,dx\,dy = 2v^2. \tag{6}$$

For $1/2 \leq v \leq 1$, you can write the integral as

$$\begin{aligned} F_V(v) &= \int_0^{1-v} \int_0^v 2\,dy\,dx + \int_{1-v}^v \int_0^{1-x} 2\,dy\,dx \\ &= 2\left[v^2 - \frac{1}{2}[v-(1-v)]^2\right] \\ &= 2v^2 - (2v-1)^2 = 4v - 2v^2 - 1, \end{aligned} \tag{7}$$

where we skipped the steps of the integral by observing that the shaded area of integration is a square of area v^2 minus the cutoff triangle on the upper right corner. The full expression for the CDF of V is

$$F_V(v) = \begin{cases} 0 & v < 0, \\ 2v^2 & 0 \leq v \leq 1/2, \\ 4v - 2v^2 - 1 & 1/2 \leq v \leq 1, \\ 1 & v > 1. \end{cases} \tag{8}$$

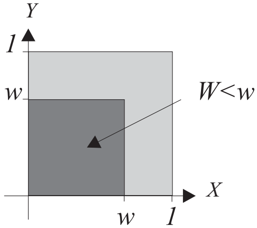
Taking a derivative, the PDF of V is

$$f_V(v) = \begin{cases} 4v & 0 \leq v \leq 1/2, \\ 4(1-v) & 1/2 \leq v \leq 1. \end{cases} \tag{9}$$

Problem 6.4.4 Solution

(a) The minimum value of W is $W = 0$, which occurs when $X = 0$ and $Y = 0$. The maximum value of W is $W = 1$, which occurs when $X = 1$ or $Y = 1$. The range of W is $S_W = \{w|0 \leq w \leq 1\}$.

(b) For $0 \leq w \leq 1$, the CDF of W is



$$\begin{aligned} F_W(w) &= P[\max(X, Y) \leq w] \\ &= P[X \leq w, Y \leq w] \\ &= \int_0^w \int_0^w f_{X,Y}(x, y)\, dy\, dx. \end{aligned} \tag{1}$$

Substituting $f_{X,Y}(x, y) = x + y$ yields

$$\begin{aligned} F_W(w) &= \int_0^w \int_0^w (x + y)\, dy\, dx \\ &= \int_0^w \left(xy + \frac{y^2}{2} \bigg|_{y=0}^{y=w} \right) dx = \int_0^w (wx + w^2/2)\, dx = w^3. \end{aligned} \tag{2}$$

The complete expression for the CDF is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ w^3 & 0 \leq w \leq 1, \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

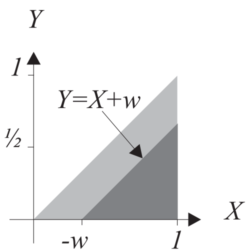
The PDF of W is found by differentiating the CDF.

$$f_W(w) = \frac{dF_W(w)}{dw} = \begin{cases} 3w^2 & 0 \leq w \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 6.4.5 Solution

(a) Since the joint PDF $f_{X,Y}(x,y)$ is nonzero only for $0 \le y \le x \le 1$, we observe that $W = Y - X \le 0$ since $Y \le X$. In addition, the most negative value of W occurs when $Y = 0$ and $X = 1$ and $W = -1$. Hence the range of W is $S_W = \{w | -1 \le w \le 0\}$.

(b) For $w < -1$, $F_W(w) = 0$. For $w > 0$, $F_W(w) = 1$. For $-1 \le w \le 0$, the CDF of W is



$$F_W(w) = \mathrm{P}[Y - X \le w] = \int_{-w}^{1} \int_{0}^{x+w} 6y\, dy\, dx = \int_{-w}^{1} 3(x+w)^2\, dx = (x+w)^3 \Big|_{-w}^{1} = (1+w)^3. \quad (1)$$

Therefore, the complete CDF of W is

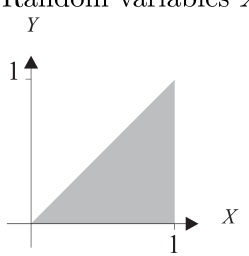
$$F_W(w) = \begin{cases} 0 & w < -1, \\ (1+w)^3 & -1 \le w \le 0, \\ 1 & w > 0. \end{cases} \quad (2)$$

By taking the derivative of $f_W(w)$ with respect to w, we obtain the PDF

$$f_W(w) = \begin{cases} 3(w+1)^2 & -1 \le w \le 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Problem 6.4.6 Solution

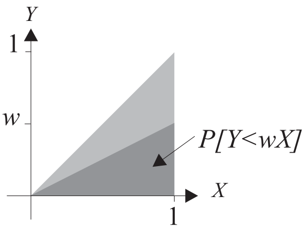
Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

(a) Since X and Y are both nonnegative, $W = Y/X \geq 0$. Since $Y \leq X$, $W = Y/X \leq 1$. Note that $W = 0$ can occur if $Y = 0$. Thus the range of W is $S_W = \{w|0 \leq w \leq 1\}$.

(b) For $0 \leq w \leq 1$, the CDF of W is



$$\begin{aligned} F_W(w) &= \mathrm{P}\left[Y/X \leq w\right] \\ &= \mathrm{P}\left[Y \leq wX\right] = w. \end{aligned} \tag{2}$$

The complete expression for the CDF is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ w & 0 \leq w < 1, \\ 1 & w \geq 1. \end{cases} \tag{3}$$

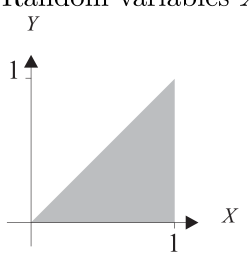
By taking the derivative of the CDF, we find that the PDF of W is

$$f_W(w) = \begin{cases} 1 & 0 \leq w < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We see that W has a uniform PDF over $[0,1]$. Thus $\mathrm{E}[W] = 1/2$.

Problem 6.4.7 Solution

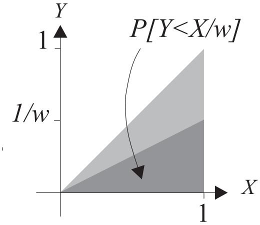
Random variables X and Y have joint PDF



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

(a) Since $f_{X,Y}(x,y) = 0$ for $y > x$, we can conclude that $Y \leq X$ and that $W = X/Y \geq 1$. Since Y can be arbitrarily small but positive, W can be arbitrarily large. Hence the range of W is $S_W = \{w|w \geq 1\}$.

(b) For $w \geq 1$, the CDF of W is



$$
\begin{aligned}
F_W(w) &= \mathrm{P}\left[X/Y \leq w\right] \\
&= 1 - \mathrm{P}\left[X/Y > w\right] \\
&= 1 - \mathrm{P}\left[Y < X/w\right] \\
&= 1 - 1/w. \qquad (2)
\end{aligned}
$$

Note that we have used the fact that $\mathrm{P}[Y < X/w]$ equals $1/2$ times the area of the corresponding triangle. The complete CDF is

$$
F_W(w) = \begin{cases} 0 & w < 1, \\ 1 - 1/w & w \geq 1. \end{cases} \qquad (3)
$$

The PDF of W is found by differentiating the CDF.

$$
f_W(w) = \frac{dF_W(w)}{dw} = \begin{cases} 1/w^2 & w \geq 1, \\ 0 & \text{otherwise.} \end{cases} \qquad (4)
$$

To find the expected value $\mathrm{E}[W]$, we write

$$
\mathrm{E}\left[W\right] = \int_{-\infty}^{\infty} w f_W(w)\, dw = \int_1^{\infty} \frac{dw}{w}. \qquad (5)
$$

However, the integral diverges and $\mathrm{E}[W]$ is undefined.

Problem 6.4.8 Solution

The position of the mobile phone is equally likely to be anywhere in the area of a circle with radius 16 km. Let X and Y denote the position of the mobile. Since we are given that the cell has a radius of 4 km, we will measure X and Y in kilometers. Assuming the base station is at the origin of the X, Y plane, the joint PDF of X and Y is

$$
f_{X,Y}(x,y) = \begin{cases} \frac{1}{16\pi} & x^2 + y^2 \leq 16, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)
$$

Since the mobile's radial distance from the base station is $R = \sqrt{X^2 + Y^2}$, the CDF of R is

$$F_R(r) = \mathrm{P}\left[R \le r\right] = \mathrm{P}\left[X^2 + Y^2 \le r^2\right]. \tag{2}$$

By changing to polar coordinates, we see that for $0 \le r \le 4$ km,

$$F_R(r) = \int_0^{2\pi} \int_0^r \frac{r'}{16\pi}\, dr'\, d\theta' = r^2/16. \tag{3}$$

So

$$F_R(r) = \begin{cases} 0 & r < 0, \\ r^2/16 & 0 \le r < 4, \\ 1 & r \ge 4. \end{cases} \tag{4}$$

Then by taking the derivative with respect to r we arrive at the PDF

$$f_R(r) = \begin{cases} r/8 & 0 \le r \le 4, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 6.4.9 Solution

Since X_1 and X_2 are iid Gaussian $(0, 1)$, each has PDF

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{1}$$

For $w < 0$, $F_W(w) = 0$. For $w \ge 0$, we define the disc

$$\mathcal{R}(w) = \left\{(x_1, x_2) | x_1^2 + x_2^2 \le w\right\}. \tag{2}$$

and we write

$$\begin{aligned} F_W(w) = \mathrm{P}\left[X_1^2 + X_2^2 \le w\right] &= \iint_{\mathcal{R}(w)} f_{X_1}(x_1)\, f_{X_2}(x_2)\, dx_1\, dx_2 \\ &= \iint_{\mathcal{R}(w)} \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}\, dx_1\, dx_2. \end{aligned} \tag{3}$$

Changing to polar coordinates with $r = \sqrt{x_1^2 + x_2^2}$ yields

$$\begin{aligned} F_W(w) &= \int_0^{2\pi} \int_0^{\sqrt{w}} \frac{1}{2\pi} e^{-r^2/2} r \, dr \, d\theta \\ &= \int_0^{\sqrt{w}} r e^{-r^2/2} \, dr = -e^{-r^2/2} \Big|_0^{\sqrt{w}} = 1 - e^{-w/2}. \end{aligned} \tag{4}$$

Taking the derivative of $F_W(w)$, the complete expression for the PDF of W is

$$f_W(w) = \begin{cases} 0 & w < 0, \\ \frac{1}{2} e^{-w/2} & w \geq 0. \end{cases} \tag{5}$$

Thus W is an exponential $(\lambda = 1/2)$ random variable.

Problem 6.4.10 Solution

First we observe that since X and Z are non-negative, $Y = ZX$ is non-negative. This implies $F_Y(y) = 0$ for $y < 0$. Second, for $y \geq 0$, we find the CDF $F_Y(y)$ by finding all (X, Z) pairs satisfying $ZX \leq y$. In particular, we notice that if $Z = 0$ then $Y = XZ = 0 \leq y$. That is, for $y \geq 0$,

$$\{Y \leq y\} = \{Z = 0\} \cup \{X \leq y, Z = 1\}. \tag{1}$$

Since this is a mutually exclusive union, we have for $y \geq 0$,

$$\begin{aligned} F_Y(y) = \mathrm{P}[Y \leq y] &= \mathrm{P}[Z = 0] + \mathrm{P}[X \leq y, Z = 1] \\ &= \frac{1}{2} + \mathrm{P}[X \leq y] \mathrm{P}[Z = 1] \\ &= \frac{1}{2} + \frac{1}{2} F_X(y), \end{aligned} \tag{2}$$

where independence of X and Z was used in (2). We note that this expression can be generalized to both positive and negative y. When $y < 0$, the first term of $1/2$ needs to disappear. We do this by writing

$$F_Y(y) = \frac{1}{2} u(y) + \frac{1}{2} F_X(y). \tag{3}$$

Taking the derivative, we obtain the PDF

$$f_Y(y) = \frac{1}{2} \delta(y) + \frac{1}{2} f_X(y) = \begin{cases} 0 & y < 0, \\ (\delta(y) + e^{-y})/2 & y \geq 0. \end{cases} \tag{4}$$

Problem 6.4.11 Solution

Although Y is a function of two random variables X and Z, it is not similar to other problems of the form $Y = g(X, Z)$ because Z is discrete. However, we can still use the same approach to find the CDF $F_Y(y)$ by identifying those pairs (X, Z) that belong to the event $\{Y \leq y\}$. In particular, since $Y = ZX$, we can write the event $\{Y \leq y\}$ as the disjoint union

$$\{Y \leq y\} = \{X \leq y, Z = 1\} \cup \{X \geq -y, Z = -1\}. \tag{1}$$

In particular, we note that if $X \geq -y$ and $Z = -1$, then $Y = ZX = -X \leq y$. It follows that

$$\begin{aligned} F_Y(y) &= \mathrm{P}[Y \leq y] \\ &= \mathrm{P}[X \leq y, Z = 1] + \mathrm{P}[X \geq -y, Z = -1] \\ &= \mathrm{P}[X \leq y] \mathrm{P}[Z = 1] + \mathrm{P}[X \geq -y] \mathrm{P}[Z = -1] \qquad (2) \\ &= p \mathrm{P}[X \leq y] + (1-p) \mathrm{P}[-X \leq y] \\ &= p \mathrm{P}[X \leq y] + (1-p) \mathrm{P}[-X \leq y] \\ &= p\Phi(y) + (1-p)\Phi(y) = \Phi(y). \qquad (3) \end{aligned}$$

Note that we use independence of X and Z to write (2). It follows that Y is Gaussian $(0, 1)$ and has PDF

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}. \tag{4}$$

Note that what has happened here is that as often as Z turns a negative X into a positive $Y = -X$, it also turns a positive X into a negative $Y = -X$. Because the PDF of X is an even function, these switches probabilistically cancel each other out.

Problem 6.4.12 Solution

(a) The exponential $(\lambda = 1/5)$ random variable Y has $\mathrm{E}[Y] = 1/\lambda$ and PDF

$$f_Y(y) = \begin{cases} (1/5)e^{-y/5} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Similarly, the exponential X has PDF

$$f_X(x) = \begin{cases} (1/10)e^{-x/10} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Since X and Y are independent,

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) = \begin{cases} (1/50)e^{-(x/10+y/5)} & x \geq 0, y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(b) If the following steps are not clear, sketch the region $X > Y \geq 0$:

$$\begin{aligned} \mathrm{P}[L] = \mathrm{P}[X > Y] &= \int_0^\infty \int_y^\infty f_{X,Y}(x,y)\; dx\, dy \\ &= \frac{1}{50} \int_0^\infty e^{-y/5} \int_y^\infty e^{-x/10}\, dx,\; dy \\ &= \frac{1}{50} \int_0^\infty e^{-y/5} \left(-10e^{-x/10} \Big|_y^\infty \right) dy \\ &= \frac{1}{5} \int_0^\infty e^{-3y/10}\, dy = -\frac{10}{15} e^{-3y/10} \Big|_0^\infty = \frac{2}{3}. \end{aligned} \tag{4}$$

(c) First we find the CDF. For $w \geq 0$,

$$\begin{aligned} F_W(w) = 1 - \mathrm{P}[W > w] &= 1 - \mathrm{P}[\min(X,Y) > w] \\ &= 1 - \mathrm{P}[X > w, Y > w] \\ &= 1 - \mathrm{P}[X > w]\, \mathrm{P}[Y > w]. \end{aligned} \tag{5}$$

where the last step follows from independence of X and Y. Since X and Y are exponential,

$$\begin{aligned} \mathrm{P}[X > w] &= \int_w^\infty (1/10)e^{-x/10}\, dx = e^{-w/10}, \\ \mathrm{P}[Y > w] &= \int_w^\infty (1/5)e^{-y/5}\, dy = e^{-w/5}. \end{aligned} \tag{6}$$

Combining steps, we obtain for $w \geq 0$,

$$F_W(w) = 1 - e^{-w/10}e^{-w/5} = 1 - e^{-3w/10}. \tag{7}$$

Taking a derivative, we obtain for $w \geq 0$ that $f_W(w) = (3/10)e^{-3w/10}$. The complete expression for the PDF is

$$f_W(w) = \begin{cases} (3/10)e^{-3w/10} & w \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We see that W is an exponential ($\lambda = 3/10$) random variable.

(d) First we find the CDF. For $t \geq 0$,

$$\begin{aligned} F_T(t) = 1 - \mathrm{P}[T > t] &= 1 - \mathrm{P}[\min(X+5, Y+15) > t] \\ &= 1 - \mathrm{P}[X > t-5, Y > t-15] \\ &= 1 - \mathrm{P}[X > t-5]\,\mathrm{P}[Y > t-15], \end{aligned} \tag{9}$$

where the last step follows from independence of X and Y. In the previous part, we found for $w \geq 0$ that $\mathrm{P}[X > w] = e^{-w/10}$ and $\mathrm{P}[Y > w] = e^{-w/5}$. For $w < 0$, $\mathrm{P}[X > w] = \mathrm{P}[Y > w] = 1$. This implies

$$\mathrm{P}[X > t-5] = \begin{cases} 1 & t < 5, \\ e^{-(t-5)/10} & t \geq 5, \end{cases} \tag{10}$$

$$\mathrm{P}[Y > t-15] = \begin{cases} 1 & t < 15, \\ e^{-(t-15)/5} & t \geq 15. \end{cases} \tag{11}$$

It follows that

$$\begin{aligned} F_T(t) &= 1 - \mathrm{P}[X > t-5]\,\mathrm{P}[Y > t-15] \\ &= \begin{cases} 0 & t < 5, \\ 1 - e^{-(t-5)/10} & 5 \leq t < 15, \\ 1 - e^{-(t-5)/10}e^{-(t-15)/5} & t \geq 15, \end{cases} \\ &= \begin{cases} 0 & t < 5, \\ 1 - e^{-(t-5)/10} & 5 \leq t < 15, \\ 1 - e^{-(3t-35)/10} & t \geq 15. \end{cases} \end{aligned} \tag{12}$$

Taking a derivative with respect to t, we obtain the PDF

$$f_T(t) = \begin{cases} 0 & t < 5, \\ (1/10)e^{-(t-5)/10} & 5 \leq t < 15, \\ (3/10)e^{-(3t-35)/10} & t \geq 15. \end{cases} \tag{13}$$

(e) Suppose the local train arrives at time w. If you ride the local train, you reach your destination at exactly 15 minutes later. If you skip the local and wait for the express, you will wait a time $\hat{X} + 5$, which is the residual (ie remaining) waiting time for the express. However, by the memoryless property of exponential random variables, $\hat{X}$ is identical to X. Thus is you wait for the express, your expected time to the finalstop is $\mathrm{E}[\hat{X} + 5] = \mathrm{E}[\hat{X}] + 5 = 15$. On average, the time until you reach your destination is the same whether you ride the express or the local. On the other hand, if you absolutely, positively must be at the destination in 20 minutes or less, ride the local since you are guaranteed to arrive in 15 minutes. Yet, if you earn some additional reward by getting home early and there is no penalty for being a few minutes late, then take the express since

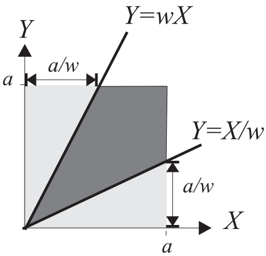
$$\mathrm{P}\left[\hat{X} + 5 \leq 15\right] = \mathrm{P}\left[X \leq 10\right] = 1 - e^{-10/10} = 1 - e^{-1} \approx 0.63. \tag{14}$$

That is, there is still a 63 percent probability the express will get you home sooner.

Problem 6.4.13 Solution

Following the hint, we observe that either $Y \geq X$ or $X \geq Y$, or, equivalently, $(Y/X) \geq 1$ or $(X/Y) \geq 1$. Hence, $W \geq 1$. To find the CDF $F_W(w)$, we know that $F_W(w) = 0$ for $w < 1$. For $w \geq 1$, we solve

$$\begin{aligned} F_W(w) &= \mathrm{P}\left[\max[(X/Y), (Y/X)] \leq w\right] \\ &= \mathrm{P}\left[(X/Y) \leq w, (Y/X) \leq w\right] \\ &= \mathrm{P}\left[Y \geq X/w, Y \leq wX\right] \\ &= \mathrm{P}\left[X/w \leq Y \leq wX\right]. \end{aligned} \tag{1}$$



We note that in the middle of the above steps, nonnegativity of X and Y was essential. We can depict the given set $\{X/w \leq Y \leq wX\}$ as the dark region on the X, Y plane. Because the PDF is uniform over the square, it is easier to use geometry to calculate the probability. In particular, each of the lighter triangles that are not part of the region of interest has area $a^2/2w$.

This implies

$$\mathrm{P}\left[X/w \leq Y \leq wX\right] = 1 - \frac{a^2/2w + a^2/2w}{a^2} = 1 - \frac{1}{w}. \tag{2}$$

The final expression for the CDF of W is

$$F_W(w) = \begin{cases} 0 & w < 1, \\ 1 - 1/w & w \geq 1. \end{cases} \tag{3}$$

By taking the derivative, we obtain the PDF

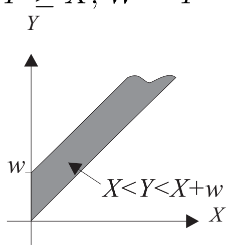
$$f_W(w) = \begin{cases} 0 & w < 1, \\ 1/w^2 & w \geq 1. \end{cases} \tag{4}$$

Problem 6.4.14 Solution

Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & 0 \leq x \leq y, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For $W = Y - X$ we can find $f_W(w)$ by integrating over the region indicated in the figure below to get $F_W(w)$ then taking the derivative with respect to w. Since $Y \geq X$, $W = Y - X$ is nonnegative. Hence $F_W(w) = 0$ for $w < 0$. For $w \geq 0$,



$$\begin{aligned} F_W(w) &= 1 - \mathrm{P}\left[W > w\right] = 1 - \mathrm{P}\left[Y > X + w\right] \\ &= 1 - \int_0^\infty \int_{x+w}^\infty \lambda^2 e^{-\lambda y}\, dy\, dx \\ &= 1 - e^{-\lambda w}. \end{aligned} \tag{2}$$

The complete expressions for the joint CDF and corresponding joint PDF are

$$F_W(w) = \begin{cases} 0 & w < 0, \\ 1 - e^{-\lambda w} & w \geq 0, \end{cases} \qquad f_W(w) = \begin{cases} 0 & w < 0, \\ \lambda e^{-\lambda w} & w \geq 0. \end{cases} \tag{3}$$

Problem 6.4.15 Solution

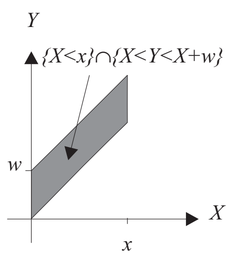
(a) To find if W and X are independent, we must be able to factor the joint density function $f_{X,W}(x,w)$ into the product $f_X(x) f_W(w)$ of marginal density functions. To verify this, we must find the joint PDF of X and W. First we find the joint CDF.

$$\begin{aligned} F_{X,W}(x,w) &= \mathrm{P}[X \leq x, W \leq w] \\ &= \mathrm{P}[X \leq x, Y - X \leq w] = \mathrm{P}[X \leq x, Y \leq X + w]. \end{aligned} \tag{1}$$

Since $Y \geq X$, the CDF of W satisfies

$$F_{X,W}(x,w) = \mathrm{P}[X \leq x, X \leq Y \leq X + w]. \tag{2}$$

Thus, for $x \geq 0$ and $w \geq 0$,



$$\begin{aligned} F_{X,W}(x,w) &= \int_0^x \int_{x'}^{x'+w} \lambda^2 e^{-\lambda y}\, dy\, dx' \\ &= \int_0^x \left(-\lambda e^{-\lambda y} \Big|_{x'}^{x'+w} \right) dx' \\ &= \int_0^x \left(-\lambda e^{-\lambda(x'+w)} + \lambda e^{-\lambda x'} \right) dx' \\ &= e^{-\lambda(x'+w)} - e^{-\lambda x'} \Big|_0^x \\ &= (1 - e^{-\lambda x})(1 - e^{-\lambda w}) \end{aligned} \tag{3}$$

We see that $F_{X,W}(x,w) = F_X(x) F_W(w)$. Moreover, by applying Theorem 5.5,

$$f_{X,W}(x,w) = \frac{\partial^2 F_{X,W}(x,w)}{\partial x\, \partial w} = \lambda e^{-\lambda x} \lambda e^{-\lambda w} = f_X(x)\, f_W(w). \tag{4}$$

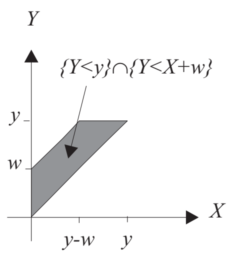
Since we have our desired factorization, W and X are independent.

(b) Following the same procedure, we find the joint CDF of Y and W.

$$F_{W,Y}(w,y) = \mathrm{P}[W \le w, Y \le y] = \mathrm{P}[Y - X \le w, Y \le y]$$
$$= \mathrm{P}[Y \le X + w, Y \le y]. \qquad (5)$$

The region of integration for the event $\{Y \le x+w, Y \le y\}$ depends on whether $y < w$ or $y \ge w$. Keep in mind that although $W = Y - X \le Y$, the dummy arguments y and w of $f_{W,Y}(w,y)$ need not obey the same constraints. In any case, we must consider each case separately.

For $y > w$, the integration is

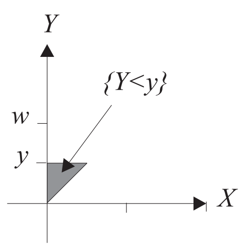


$$\begin{aligned} F_{W,Y}(w,y) &= \int_0^{y-w} \int_u^{u+w} \lambda^2 e^{-\lambda v}\, dv\, du \\ &\quad + \int_{y-w}^{y} \int_u^{y} \lambda^2 e^{-\lambda v}\, dv\, du \\ &= \lambda \int_0^{y-w} \left[e^{-\lambda u} - e^{-\lambda(u+w)}\right] du \\ &\quad + \lambda \int_{y-w}^{y} \left[e^{-\lambda u} - e^{-\lambda y}\right] du. \end{aligned} \qquad (6)$$

It follows that

$$\begin{aligned} F_{W,Y}(w,y) &= \left[-e^{-\lambda u} + e^{-\lambda(u+w)}\right]\Big|_0^{y-w} + \left[-e^{-\lambda u} - u\lambda e^{-\lambda y}\right]\Big|_{y-w}^{y} \\ &= 1 - e^{-\lambda w} - \lambda w e^{-\lambda y}. \end{aligned} \qquad (7)$$

For $y \le w$,



$$\begin{aligned} F_{W,Y}(w,y) &= \int_0^{y} \int_u^{y} \lambda^2 e^{-\lambda v}\, dv\, du \\ &= \int_0^{y} \left[-\lambda e^{-\lambda y} + \lambda e^{-\lambda u}\right] du \\ &= -\lambda u e^{-\lambda y} - e^{-\lambda u}\Big|_0^{y} \\ &= 1 - (1 + \lambda y) e^{-\lambda y}. \end{aligned} \qquad (8)$$

The complete expression for the joint CDF is

$$F_{W,Y}(w,y) = \begin{cases} 1 - e^{-\lambda w} - \lambda w e^{-\lambda y} & 0 \le w \le y, \\ 1 - (1+\lambda y)e^{-\lambda y} & 0 \le y \le w, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Applying Theorem 5.5 yields

$$f_{W,Y}(w,y) = \frac{\partial^2 F_{W,Y}(w,y)}{\partial w\,\partial y} = \begin{cases} 2\lambda^2 e^{-\lambda y} & 0 \le w \le y \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The joint PDF $f_{W,Y}(w,y)$ doesn't factor and thus W and Y are dependent.

Problem 6.4.16 Solution

We need to define the events $A = \{U \le u\}$ and $B = \{V \le v\}$. In this case,

$$F_{U,V}(u,v) = \mathrm{P}[AB] = \mathrm{P}[B] - \mathrm{P}[A^c B] = \mathrm{P}[V \le v] - \mathrm{P}[U > u, V \le v] \tag{1}$$

Note that $U = \min(X,Y) > u$ if and only if $X > u$ and $Y > u$. In the same way, since $V = \max(X,Y)$, $V \le v$ if and only if $X \le v$ and $Y \le v$. Thus

$$\begin{aligned} \mathrm{P}[U > u, V \le v] &= \mathrm{P}[X > u, Y > u, X \le v, Y \le v] \\ &= \mathrm{P}[u < X \le v, u < Y \le v]. \end{aligned} \tag{2}$$

Thus, the joint CDF of U and V satisfies

$$\begin{aligned} F_{U,V}(u,v) &= \mathrm{P}[V \le v] - \mathrm{P}[U > u, V \le v] \\ &= \mathrm{P}[X \le v, Y \le v] - \mathrm{P}[u < X \le v, u < X \le v]. \end{aligned} \tag{3}$$

Since X and Y are independent random variables,

$$\begin{aligned} F_{U,V}(u,v) &= \mathrm{P}[X \le v]\,\mathrm{P}[Y \le v] - \mathrm{P}[u < X \le v]\,\mathrm{P}[u < X \le v] \\ &= F_X(v)\,F_Y(v) - (F_X(v) - F_X(u))\,(F_Y(v) - F_Y(u)) \\ &= F_X(v)\,F_Y(u) + F_X(u)\,F_Y(v) - F_X(u)\,F_Y(u). \end{aligned} \tag{4}$$

The joint PDF is

$$
\begin{aligned}
f_{U,V}(u,v) &= \frac{\partial^2 F_{U,V}(u,v)}{\partial u \partial v} \\
&= \frac{\partial}{\partial u}\left[f_X(v) F_Y(u) + F_X(u) f_Y(v)\right] \\
&= f_X(u) f_Y(v) + f_X(v) f_Y(v). \qquad (5)
\end{aligned}
$$

Problem 6.5.1 Solution

Since X and Y are take on only integer values, $W = X + Y$ is integer valued as well. Thus for an integer w,

$$P_W(w) = \mathrm{P}[W = w] = \mathrm{P}[X + Y = w]. \qquad (1)$$

Suppose $X = k$, then $W = w$ if and only if $Y = w - k$. To find all ways that $X + Y = w$, we must consider each possible integer k such that $X = k$. Thus

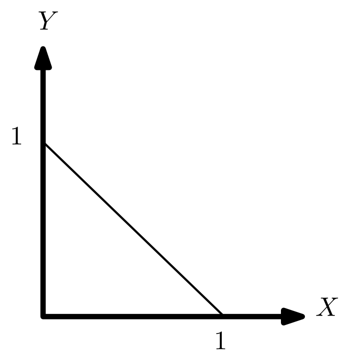
$$P_W(w) = \sum_{k=-\infty}^{\infty} \mathrm{P}[X = k, Y = w - k] = \sum_{k=-\infty}^{\infty} P_{X,Y}(k, w - k). \qquad (2)$$

Since X and Y are independent, $P_{X,Y}(k, w-k) = P_X(k)P_Y(w-k)$. It follows that for any integer w,

$$P_W(w) = \sum_{k=-\infty}^{\infty} P_X(k)\, P_Y(w - k). \qquad (3)$$

Problem 6.5.2 Solution

The key to the solution is to draw the triangular region where the PDF is nonzero:



For the PDF of $W = X + Y$, we could use the usual procedure to derive the CDF of W and take a derivative, but it is much easier to use Theorem 6.4 to write

$$f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, w - x)\, dx. \tag{1}$$

For $0 \leq w \leq 1$,

$$f_W(w) = \int_0^w 2\, dx = 2w. \tag{2}$$

For $w < 0$ or $w > 1$, $f_W(w) = 0$ since $0 \leq W \leq 1$. The complete expression is

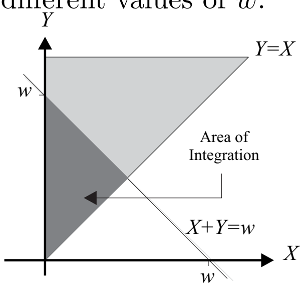
$$f_W(w) = \begin{cases} 2w & 0 \leq w \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 6.5.3 Solution

The joint PDF of X and Y is

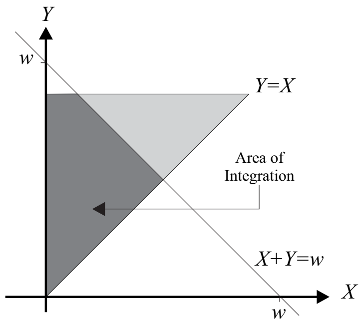
$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \leq x \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We wish to find the PDF of W where $W = X + Y$. First we find the CDF of W, $F_W(w)$, but we must realize that the CDF will require different integrations for different values of w.



For values of $0 \leq w \leq 1$ we look to integrate the shaded area in the figure to the right.

$$F_W(w) = \int_0^{\frac{w}{2}} \int_x^{w-x} 2\, dy\, dx = \frac{w^2}{2}. \tag{2}$$



For values of w in the region $1 \leq w \leq 2$ we look to integrate over the shaded region in the graph to the right. From the graph we see that we can integrate with respect to x first, ranging y from 0 to $w/2$, thereby covering the lower right triangle of the shaded region and leaving the upper trapezoid, which is accounted for in the second term of the following expression:

$$F_W(w) = \int_0^{\frac{w}{2}} \int_0^y 2\,dx\,dy + \int_{\frac{w}{2}}^1 \int_0^{w-y} 2\,dx\,dy$$
$$= 2w - 1 - \frac{w^2}{2}. \tag{3}$$

Putting all the parts together gives the CDF

$$F_W(w) = \begin{cases} 0 & w < 0, \\ \frac{w^2}{2} & 0 \leq w \leq 1, \\ 2w - 1 - \frac{w^2}{2} & 1 \leq w \leq 2, \\ 1 & w > 2, \end{cases} \tag{4}$$

and (by taking the derivative) the PDF

$$f_W(w) = \begin{cases} w & 0 \leq w \leq 1, \\ 2 - w & 1 \leq w \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 6.5.4 Solution

The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \leq x, y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Proceeding as in Problem 6.5.3, we must first find $F_W(w)$ by integrating over the square defined by $0 \leq x, y \leq 1$. Again we are forced to find $F_W(w)$ in parts as we did in Problem 6.5.3 resulting in the following integrals for their appropriate regions. For $0 \leq w \leq 1$,

$$F_W(w) = \int_0^w \int_0^{w-x} dx\,dy = w^2/2. \tag{2}$$

For $1 \leq w \leq 2$,

$$F_W(w) = \int_0^{w-1} \int_0^1 dx\,dy + \int_{w-1}^1 \int_0^{w-y} dx\,dy = 2w - 1 - w^2/2. \tag{3}$$

The complete CDF is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ w^2/2 & 0 \leq w \leq 1, \\ 2w - 1 - w^2/2 & 1 \leq w \leq 2, \\ 1 & \text{otherwise.} \end{cases} \tag{4}$$

The corresponding PDF, $f_W(w) = dF_W(w)/dw$, is

$$f_W(w) = \begin{cases} w & 0 \leq w \leq 1, \\ 2 - w & 1 \leq w \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 6.5.5 Solution

By using Theorem 6.9, we can find the PDF of $W = X + Y$ by convolving the two exponential distributions. For $\mu \neq \lambda$,

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} f_X(x)\, f_Y(w - x)\, dx \\ &= \int_0^w \lambda e^{-\lambda x} \mu e^{-\mu(w-x)}\, dx \\ &= \lambda \mu e^{-\mu w} \int_0^w e^{-(\lambda - \mu)x}\, dx \\ &= \begin{cases} \frac{\lambda\mu}{\lambda - \mu}\left(e^{-\mu w} - e^{-\lambda w}\right) & w \geq 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{1}$$

When $\mu = \lambda$, the previous derivation is invalid because of the denominator term $\lambda - \mu$. For $\mu = \lambda$, we have

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} f_X(x)\, f_Y(w-x)\ dx \\ &= \int_0^w \lambda e^{-\lambda x} \lambda e^{-\lambda(w-x)}\, dx \\ &= \lambda^2 e^{-\lambda w} \int_0^w dx \\ &= \begin{cases} \lambda^2 w e^{-\lambda w} & w \geq 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

Note that when $\mu = \lambda$, W is the sum of two iid exponential random variables and has a second order Erlang PDF.

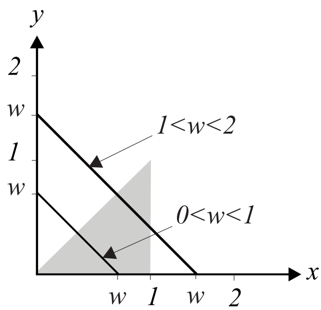
Problem 6.5.6 Solution

In this problem, X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 8xy & 0 \leq y \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We can find the PDF of W using Theorem 6.8: $f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)\, dx$. The only tricky part remaining is to determine the limits of the integration. First, for $w < 0$, $f_W(w) = 0$. The two remaining cases are shown in the accompanying figure. The shaded area shows where the joint PDF $f_{X,Y}(x,y)$ is nonzero. The diagonal lines depict $y = w - x$ as a function of x. The intersection of the diagonal line and the shaded area define our limits of integration.

For $0 \leq w \leq 1$,



$$f_W(w) = \int_{w/2}^{w} 8x(w-x)\,dx,$$
$$= 4wx^2 - 8x^3/3\Big|_{w/2}^{w} = 2w^3/3. \qquad (2)$$

For $1 \leq w \leq 2$,

$$f_W(w) = \int_{w/2}^{1} 8x(w-x)\,dx$$
$$= 4wx^2 - 8x^3/3\Big|_{w/2}^{1} \qquad (3)$$
$$= 4w - 8/3 - 2w^3/3. \qquad (4)$$

Since $X + Y \leq 2$, $f_W(w) = 0$ for $w > 2$. Hence the complete expression for the PDF of W is

$$f_W(w) = \begin{cases} 2w^3/3 & 0 \leq w \leq 1, \\ 4w - 8/3 - 2w^3/3 & 1 \leq w \leq 2, \\ 0 & \text{otherwise.} \end{cases} \qquad (5)$$

Problem 6.5.7 Solution

We first find the CDF of W following the same procedure as in the proof of Theorem 6.8.

$$F_W(w) = P[X \leq Y + w] = \int_{-\infty}^{\infty} \int_{-\infty}^{y+w} f_{X,Y}(x,y)\, dx\, dy. \qquad (1)$$

By taking the derivative with respect to w, we obtain

$$f_W(w) = \frac{dF_W(w)}{dw} = \int_{-\infty}^{\infty} \frac{d}{dw}\left(\int_{-\infty}^{y+w} f_{X,Y}(x,y)\, dx\right) dy$$
$$= \int_{-\infty}^{\infty} f_{X,Y}(w+y,y)\, dy. \qquad (2)$$

With the variable substitution $y = x - w$, we have $dy = dx$ and

$$f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, x-w)\, dx. \qquad (3)$$

Problem 6.5.8 Solution

The random variables K and J have PMFs

$$P_J(j) = \begin{cases} \frac{\alpha^j e^{-\alpha}}{j!} & j = 0, 1, 2, \ldots, \\ 0 & \text{otherwise}, \end{cases} \qquad P_K(k) = \begin{cases} \frac{\beta^k e^{-\beta}}{k!} & k = 0, 1, 2, \ldots, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

For $n \geq 0$, we can find the PMF of $N = J + K$ via

$$\mathrm{P}[N = n] = \sum_{k=-\infty}^{\infty} \mathrm{P}[J = n - k, K = k]. \tag{2}$$

Since J and K are independent, non-negative random variables,

$$\begin{aligned} \mathrm{P}[N = n] &= \sum_{k=0}^{n} P_J(n-k)\, P_K(k) \\ &= \sum_{k=0}^{n} \frac{\alpha^{n-k} e^{-\alpha}}{(n-k)!} \frac{\beta^k e^{-\beta}}{k!} \\ &= \frac{(\alpha+\beta)^n e^{-(\alpha+\beta)}}{n!} \underbrace{\sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \left(\frac{\alpha}{\alpha+\beta}\right)^{n-k} \left(\frac{\beta}{\alpha+\beta}\right)^{k}}_{1}. \end{aligned} \tag{3}$$

The marked sum above equals 1 because it is the sum of a binomial PMF over all possible values.The PMF of N is the Poisson PMF

$$P_N(n) = \begin{cases} \frac{(\alpha+\beta)^n e^{-(\alpha+\beta)}}{n!} & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise}. \end{cases} \tag{4}$$

Problem 6.6.1 Solution

Given $0 \leq u \leq 1$, we need to find the "inverse" function that finds the value of w satisfying $u = F_W(w)$. The problem is that for $u = 1/4$, any w in the interval $[-3, 3]$ satisfies $F_W(w) = 1/4$. However, in terms of generating samples of random variable W, this doesn't matter. For a uniform $(0, 1)$ random variable U, $\mathrm{P}[U = 1/4] = 0$. Thus we can choose any $w \in [-3, 3]$. In particular, we define the inverse CDF as

$$w = F_W^{-1}(u) = \begin{cases} 8u - 5 & 0 \leq u \leq 1/4, \\ (8u + 7)/3 & 1/4 < u \leq 1. \end{cases} \tag{1}$$

Note that because $0 \le F_W(w) \le 1$, the inverse $F_W^{-1}(u)$ is defined only for $0 \le u \le 1$. Careful inspection will show that $u = (w+5)/8$ for $-5 \le w < -3$ and that $u = 1/4 + 3(w-3)/8$ for $-3 \le w \le 5$. Thus, for a uniform $(0,1)$ random variable U, the function $W = F_W^{-1}(U)$ produces a random variable with CDF $F_W(w)$. To implement this solution in MATLAB, we define

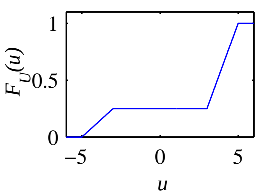
```
function w=iwcdf(u);
w=((u>=0).*(u <= 0.25).*(8*u-5))+...
    ((u > 0.25).*(u<=1).*((8*u+7)/3));
```

so that the MATLAB code `W=icdfrv(@iwcdf,m)` generates m samples of random variable W.

Problem 6.6.2 Solution

To solve this problem, we want to use Theorem 6.5. One complication is that in the theorem, U denotes the uniform random variable while X is the derived random variable. In this problem, we are using U for the random variable we want to derive. As a result, we will use Theorem 6.5 with the roles of X and U reversed. Given U with CDF $F_U(u) = F(u)$, we need to find the inverse functon $F^{-1}(x) = F_U^{-1}(x)$ so that for a uniform $(0,1)$ random variable X, $U = F^{-1}(X)$.

Recall that random variable U defined in Problem 4.4.7 has CDF



$$F_U(u) = \begin{cases} 0 & u < -5, \\ (u+5)/8 & -5 \le u < -3, \\ 1/4 & -3 \le u < 3, \\ 1/4 + 3(u-3)/8 & 3 \le u < 5, \\ 1 & u \ge 5. \end{cases} \tag{1}$$

At $x = 1/4$, there are multiple values of u such that $F_U(u) = 1/4$. However, except for $x = 1/4$, the inverse $F_U^{-1}(x)$ is well defined over $0 < x < 1$. At $x = 1/4$, we can arbitrarily define a value for $F_U^{-1}(1/4)$ because when we produce sample values of $F_U^{-1}(X)$, the event $X = 1/4$ has probability zero. To generate the inverse CDF, given a value of x, $0 < x < 1$, we ave to find the value of u such that $x = F_U(u)$.

From the CDF we see that

$$0 \le x \le \frac{1}{4} \qquad \Rightarrow x = \frac{u+5}{8}, \tag{2}$$

$$\frac{1}{4} < x \le 1 \qquad \Rightarrow x = \frac{1}{4} + \frac{3}{8}(u-3). \tag{3}$$

These conditions can be inverted to express u as a function of x.

$$u = F^{-1}(x) = \begin{cases} 8x - 5 & 0 \le x \le 1/4, \\ (8x+7)/3 & 1/4 < x \le 1. \end{cases} \tag{4}$$

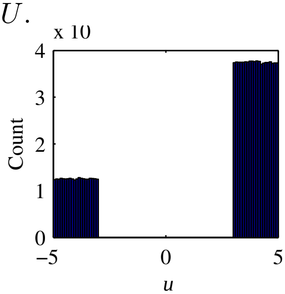
In particular, when X is a uniform $(0,1)$ random variable, $U = F^{-1}(X)$ will generate samples of the rndom variable U. A MATLAB program to implement this solution is now straightforward:

```
function u=urv(m)
%Usage: u=urv(m)
%Generates m samples of the random
%variable U defined in Problem 3.3.7
x=rand(m,1);
u=(x<=1/4).*(8*x-5);
u=u+(x>1/4).*(8*x+7)/3;
```

To see that this generates the correct output, we can generate a histogram of a million sample values of U using the commands

```
u=urv(1000000); hist(u,100);
```

The output is shown in the following graph, alongside the corresponding PDF of U.



$$f_U(u) = \begin{cases} 0 & u < -5, \\ 1/8 & -5 \le u < -3, \\ 0 & -3 \le u < 3, \\ 3/8 & 3 \le u < 5, \\ 0 & u \ge 5. \end{cases} \tag{5}$$

Note that the scaling constant 10^4 on the histogram plot comes from the fact that the histogram was generated using 10^6 sample points and 100 bins. The width of each bin is $\Delta = 10/100 = 0.1$. Consider a bin of idth Δ centered at u_0. A sample value of U would fall in that bin with probability $f_U(u_0)\Delta$. Given that we generate $m = 10^6$ samples, we would expect about $mf_U(u_0)\Delta = 10^5 f_U(u_0)$ samples in each bin. For $-5 < u_0 < -3$, we would expect to see about 1.25×10^4 samples in each bin. For $3 < u_0 < 5$, we would expect to see about 3.75×10^4 samples in each bin. As can be seen, these conclusions are consistent with the histogam data.

Finally, we comment that if you generate histograms for a range of values of m, the number of samples, you will see that the histograms will become more and more similar to a scaled version of the PDF. This gives the (false) impression that any bin centered on u_0 has a number of samples increasingly close to $mf_U(u_0)\Delta$. Because the histpgram is always the same height, what is actually happening is that the vertical axis is effectively scaled by $1/m$ and the height of a histogram bar is proportional to *the fraction* of m samples that land in that bin. We will see in Chapter 10 that the fraction of samples in a bin does converge to the probability of a sample being in that bin as the number of samples m goes to infinity.

Problem 6.6.3 Solution

In the first approach X is an exponential (λ) random variable, Y is an independent exponential (μ) random variable, and $W = Y/X$. we implement this approach in the function `wrv1.m` shown below.

In the second approach, we use Theorem 6.5 and generate samples of a uniform $(0,1)$ random variable U and calculate $W = F_W^{-1}(U)$. In this problem,

$$F_W(w) = 1 - \frac{\lambda/\mu}{\lambda/\mu + w}. \tag{1}$$

Setting $u = F_W(w)$ and solving for w yields

$$w = F_W^{-1}(u) = \frac{\lambda}{\mu}\left(\frac{u}{1-u}\right). \tag{2}$$

We implement this solution in the function `wrv2`. Here are the two solutions:

```
function w=wrv1(lambda,mu,m)
%Usage: w=wrv1(lambda,mu,m)
%Return m samples of W=Y/X
%X is exponential (lambda)
%Y is exponential (mu)

x=exponentialrv(lambda,m);
y=exponentialrv(mu,m);
w=y./x;
```

```
function w=wrv2(lambda,mu,m)
%Usage: w=wrv1(lambda,mu,m)
%Return m samples of W=Y/X
%X is exponential (lambda)
%Y is exponential (mu)
%Uses CDF of F_W(w)

u=rand(m,1);
w=(lambda/mu)*u./(1-u);
```

We would expect that `wrv2` would be faster simply because it does less work. In fact, its instructive to account for the work each program does.

- `wrv1` Each exponential random sample requires the generation of a uniform random variable, and the calculation of a logarithm. Thus, we generate $2m$ uniform random variables, calculate $2m$ logarithms, and perform m floating point divisions.

- `wrv2` Generate m uniform random variables and perform m floating points divisions.

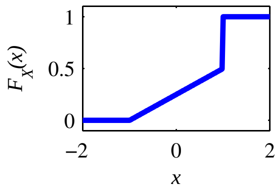
This quickie analysis indicates that `wrv1` executes roughly $5m$ operations while `wrv2` executes about $2m$ operations. We might guess that `wrv2` would be faster by a factor of 2.5. Experimentally, we calculated the execution time associated with generating a million samples:

```
>> t2=cputime;w2=wrv2(1,1,1000000);t2=cputime-t2
t2 =
    0.2500
>> t1=cputime;w1=wrv1(1,1,1000000);t1=cputime-t1
t1 =
    0.7610
>>
```

We see in our simple experiments that `wrv2` is faster by a rough factor of 3. (Note that repeating such trials yielded qualitatively similar results.)

Problem 6.6.4 Solution

From Quiz 4.7, random variable X has CDF The CDF of X is



$$F_X(x) = \begin{cases} 0 & x < -1, \\ (x+1)/4 & -1 \leq x < 1, \\ 1 & x \geq 1. \end{cases} \tag{1}$$

Following the procedure outlined in Problem 6.3.13, we define for $0 < u \leq 1$,

$$\tilde{F}(u) = \min\{x | F_X(x) \geq u\}. \tag{2}$$

We observe that if $0 < u < 1/4$, then we can choose x so that $F_X(x) = u$. In this case, $(x+1)/4 = u$, or equivalently, $x = 4u - 1$. For $1/4 \leq u \leq 1$, the minimum x that satisfies $F_X(x) \geq u$ is $x = 1$. These facts imply

$$\tilde{F}(u) = \begin{cases} 4u - 1 & 0 < u < 1/4, \\ 1 & 1/4 \leq u \leq 1. \end{cases} \tag{3}$$

It follows that if U is a uniform $(0,1)$ random variable, then $\tilde{F}(U)$ has the same CDF as X. This is trivial to implement in MATLAB.

```
function x=quiz36rv(m)
%Usage x=quiz36rv(m)
%Returns the vector x holding m samples
%of the random variable X of Quiz 3.6
u=rand(m,1);
x=((4*u-1).*(u< 0.25))+(1.0*(u>=0.25));
```

Problem Solutions – Chapter 7

Problem 7.1.1 Solution

Given the CDF $F_X(x)$, we can write

$$\begin{aligned} F_{X|X>0}(x) &= \mathrm{P}\left[X \le x | X > 0\right] \\ &= \frac{\mathrm{P}\left[X \le x, X > 0\right]}{\mathrm{P}\left[X > 0\right]} \\ &= \frac{\mathrm{P}\left[0 < X \le x\right]}{\mathrm{P}\left[X > 0\right]} = \begin{cases} 0 & x \le 0, \\ \frac{F_X(x) - F_X(0)}{\mathrm{P}[X>0]}. & x > 0. \end{cases} \end{aligned} \tag{1}$$

From $F_X(x)$, we know that $F_X(0) = 0.4$ and $\mathrm{P}[X > 0] = 1 - F_X(0) = 0.6$. Thus

$$\begin{aligned} F_{X|X>0}(x) &= \begin{cases} 0 & x \le 0, \\ \frac{F_X(x) - 0.4}{0.6} & x > 0, \end{cases} \\ &= \begin{cases} 0 & x < 5, \\ \frac{0.8-0.4}{0.6} = \frac{2}{3} & 5 \le x < 7, \\ 1 & x \ge 7. \end{cases} \end{aligned} \tag{2}$$

From the jumps in the conditional CDF $F_{X|X>0}(x)$, we can write down the conditional PMF

$$P_{X|X>0}(x) = \begin{cases} 2/3 & x = 5, \\ 1/3 & x = 7, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Alternatively, we can start with the jumps in $F_X(x)$ and read off the PMF of X as

$$P_X(x) = \begin{cases} 0.4 & x = -3, \\ 0.4 & x = 5, \\ 0.2 & x = 7, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The event $\{X > 0\}$ has probability $\mathrm{P}[X > 0] = P_X(5) + P_X(7) = 0.6$. From Theorem 7.1, the conditional PMF of X given $X > 0$ is

$$P_{X|X>0}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[X>0]} & x \in B, \\ 0 & \text{otherwise}, \end{cases} = \begin{cases} 2/3 & x = 5, \\ 1/3 & x = 7, \\ 0 & \text{otherwise}. \end{cases} \tag{5}$$

Problem 7.1.2 Solution

Since

$$P_X(x) = \begin{cases} 1/5 & x = 1, 2, \ldots, 5, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

and $\mathrm{E}[X] = 2.5$, note that

$$\mathrm{P}\left[X \geq \mathrm{E}\left[X\right]\right] = \mathrm{P}\left[X \geq 2.5\right] = P_X(3) + P_X(4) + P_X(5) = 3/5. \tag{2}$$

The formula for conditional PMF is

$$P_{X|X\geq 2.5}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[X\geq 2.5]} & x \geq 2.5, \\ 0 & \text{otherwise}, \end{cases} = \begin{cases} 1/3 & x = 3, 4, 5, \\ 0 & \text{otherwise}. \end{cases} \tag{3}$$

It follows that

$$\mathrm{E}\left[X|X \geq 2.5\right] = \sum_x x P_{X|X\geq 2.5}(x) = \frac{3 + 4 + 5}{3} = 4. \tag{4}$$

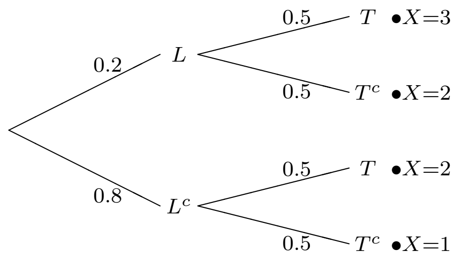
Problem 7.1.3 Solution

The event $B = \{X \neq 0\}$ has probability $\mathrm{P}[B] = 1 - \mathrm{P}[X = 0] = 15/16$. The conditional PMF of X given B is

$$P_{X|B}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[B]} & x \in B, \\ 0 & \text{otherwise}, \end{cases} = \binom{4}{x} \frac{1}{15}. \tag{1}$$

Problem 7.1.4 Solution

(a) Let L denote the event that the layup is good and T denote the event that the free throw is good. Let L^c and T^c denote the complementary events. The tree is



Reading from the tree, we see that

$$\begin{aligned} \mathrm{P}[X = 1] &= (0.8)(0.45) = 0.4, && (1)\\ \mathrm{P}[X = 2] &= (0.2)(0.5) + (0.8)(0.5) = 0.5, && (2)\\ \mathrm{P}[X = 3] &= (0.2)(0.5) = 0.1. && (3) \end{aligned}$$

The complete PMF is

$$P_X(x) = \begin{cases} 0.4 & x = 1, \\ 0.5 & x = 2, \\ 0.1 & x = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(b) There are only two outcomes LT (where $X = 3$) and L^cT (where $X = 2$) for which event T occurs. Since $\mathrm{P}[T] = (0.2)(0.5) + (0.8)(0.5) = 0.5$, the definition of conditional probability says

$$P_{X|T}(x) = \frac{\mathrm{P}[X = x, T]}{\mathrm{P}[T]} = \begin{cases} \frac{(0.8)(0.5)}{0.5} & x = 2, \\ \frac{(0.2)(0.5)}{0.5} & x = 3, \\ 0 & \text{otherwise,} \end{cases}$$

$$= \begin{cases} 0.8 & x = 2, \\ 0.2 & x = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 7.1.5 Solution

(a) You run $M = m$ miles if (with probability $(1-q)^m$) you choose to run the first m miles and then (with probability q) you choose to quite just prior to mile $m+1$. The PMF of M, the number of miles run on an arbitrary day is

$$P_M(m) = \begin{cases} q(1-q)^m & m = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The probability that we run a marathon on any particular day is the probability that $M \geq 26$.

$$r = \mathrm{P}[M \geq 26] = \sum_{m=26}^{\infty} q(1-q)^m = (1-q)^{26}. \tag{2}$$

(c) We run a marathon on each day with probability equal to r, and we do not run a marathon with probability $1-r$. Therefore in a year we have 365 tests of our jogging resolve, and thus 365 chances to run a marathon. So the PMF of the number of marathons run in a year, J, can be expressed as

$$P_J(j) = \binom{365}{j} r^j (1-r)^{365-j}. \tag{3}$$

(d) The random variable K is defined as the number of miles we run above that required for a marathon, $K = M - 26$. Given the event, A, that we have run a marathon, we wish to know how many miles in excess of 26 we in fact ran. So we want to know the conditional PMF $P_{K|A}(k)$.

$$P_{K|A}(k) = \frac{\mathrm{P}[K = k, A]}{\mathrm{P}[A]} = \frac{\mathrm{P}[M = 26 + k]}{\mathrm{P}[A]}. \tag{4}$$

Since $\mathrm{P}[A] = r$, for $k = 0, 1, \ldots$,

$$P_{K|A}(k) = \frac{(1-q)^{26+k} q}{(1-q)^{26}} = (1-q)^k q. \tag{5}$$

The complete expression of for the conditional PMF of K is

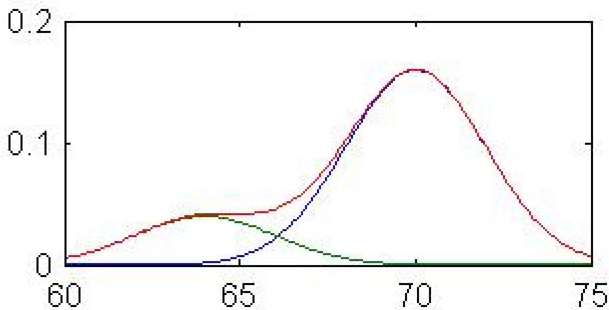
$$P_{K|A}(k) = \begin{cases} (1-q)^k q & k = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Problem 7.1.6 Solution

(a) The PDF of X is weighted combination of the Gaussian PDFs

$$\frac{1}{\sqrt{8\pi}} e^{-(x-70)^2/8} \qquad \text{and} \qquad \frac{1}{\sqrt{8\pi}} e^{-(x-65)^2/8}. \tag{1}$$

The key is to plot each of these Gaussian PDFs (with the weight factors 0.8 and 0.2) and then sketch the combination. Here is a plot:



(b) There are two ways to solve this problem. The first way is just by calculus:

$$
\begin{aligned}
P[X \leq 68] &= \int_{-\infty}^{68} \left[\frac{0.8}{\sqrt{8\pi}} e^{-(x-70)^2/8} + \frac{0.2}{\sqrt{8\pi}} e^{-(x-65)^2/8} \right] dx \\
&= \frac{0.8}{\sqrt{8\pi}} \int_{-\infty}^{68} e^{-(x-70)^2/8}\, dx + \frac{0.2}{\sqrt{8\pi}} \int_{-\infty}^{68} e^{-(x-65)^2/8}\, dx. \qquad (2)
\end{aligned}
$$

Treating the two integrals separately, we make the substitution $u = (x-70)/2$ in the first integral and the substitution $v = (x-65)/2$ in the second integral. This yields

$$
\begin{aligned}
P[X \leq 68] &= 0.8 \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du + 0.2 \int_{-\infty}^{3/2} \frac{1}{\sqrt{2\pi}} e^{-v^2/2}\, dv \\
&= 0.8\Phi(-1) + 0.2\Phi(3/2) \\
&= 0.8(1 - \Phi(1)) + 0.2\Phi(1.5) \\
&= 0.8(1 - 0.841) + 0.2(0.933) = 0.3138. \qquad (3)
\end{aligned}
$$

A second way to solve this problem is to view $f_X(x)$ as a combination of two conditional PDFs

$$
f_{X|M}(x) = \frac{1}{\sqrt{8\pi}} e^{-(x-70)^2/8} \quad \text{and} \quad f_{X|W}(x) = \frac{1}{\sqrt{8\pi}} e^{-(x-65)^2/8}. \qquad (4)
$$

With these conditional PDFs, we define the event partition (M, W) where $P[M] = 0.8$ and $P[W] = 0.2$ and we can write

$$
f_X(x) = P[M]\, f_{X|M}(x) + P[W]\, f_{X|W}(x). \qquad (5)
$$

That is given M, X is Gaussian with conditional expected value $\mu_{X|M} = 70$ and conditional standard deviation $\sigma_{X|M} = 2$ while given W, X is Gaussian with conditional expected value $\mu_{X|M} = 65$ and conditional standard deviation $\sigma_{X|W} = 2$. Similarly, we can write the CDF $F_X(x)$ in terms of the corresponding conditional CDFs as

$$
F_X(x) = P[M]\, F_{X|M}(x) + P[W]\, F_{X|W}(x). \qquad (6)
$$

Since these conditional CDFs are Gaussian, we can write

$$F_X(x) = P[M]\,\Phi\left(\frac{x-\mu_{X|M}}{\sigma_{X|M}}\right) + P[W]\,\Phi\left(\frac{x-\mu_{X|W}}{\sigma_{X|W}}\right)$$
$$= 0.8\Phi\left(\frac{x-70}{2}\right) + 0.2\Phi\left(\frac{x-65}{2}\right). \tag{7}$$

Finally,

$$P[X \le 68] = F_X(68) = 0.8\Phi\left(\frac{68-70}{2}\right) + 0.2\Phi\left(\frac{68-65}{2}\right)$$
$$= 0.8\Phi(-1) + 0.2\Phi(3/2), \tag{8}$$

which is the answer we obtained from calculus.

(c) This is a reasonable model. The idea is that an engineering class might be 80 percent men and 20 percent women. The men have an average height of 70 inches while the women have an average height of 65 inches. For both men and women, the conditional standard deviation of height is 2 inches. The PDF $f_X(x)$ reflects that a randomly chosen student will be a man with probability 0.8 or a woman with probability 0.2.

Problem 7.1.7 Solution

(a) Given that a person is healthy, X is a Gaussian ($\mu = 90, \sigma = 20$) random variable. Thus,

$$f_{X|H}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{20\sqrt{2\pi}}e^{-(x-90)^2/800}. \tag{1}$$

(b) Given the event H, we use the conditional PDF $f_{X|H}(x)$ to calculate the required probabilities

$$P\left[T^+|H\right] = P[X \ge 140|H] = P[X - 90 \ge 50|H]$$
$$= P\left[\frac{X-90}{20} \ge 2.5|H\right]$$
$$= 1 - \Phi(2.5) = 0.006. \tag{2}$$

Similarly,

$$\begin{aligned} P\left[T^-|H\right] = P\left[X \leq 110|H\right] &= P\left[X - 90 \leq 20|H\right] \\ &= P\left[\frac{X-90}{20} \leq 1|H\right] \\ &= \Phi(1) = 0.841. \end{aligned} \tag{3}$$

(c) Using Bayes Theorem, we have

$$P\left[H|T^-\right] = \frac{P\left[T^-|H\right]P\left[H\right]}{P\left[T^-\right]} = \frac{P\left[T^-|H\right]P\left[H\right]}{P\left[T^-|D\right]P\left[D\right] + P\left[T^-|H\right]P\left[H\right]}. \tag{4}$$

In the denominator, we need to calculate

$$\begin{aligned} P\left[T^-|D\right] = P\left[X \leq 110|D\right] &= P\left[X - 160 \leq -50|D\right] \\ &= P\left[\frac{X-160}{40} \leq -1.25|D\right] \\ &= \Phi(-1.25) = 1 - \Phi(1.25) = 0.106. \end{aligned} \tag{5}$$

Thus,

$$\begin{aligned} P\left[H|T^-\right] &= \frac{P\left[T^-|H\right]P\left[H\right]}{P\left[T^-|D\right]P\left[D\right] + P\left[T^-|H\right]P\left[H\right]} \\ &= \frac{0.841(0.9)}{0.106(0.1) + 0.841(0.9)} = 0.986. \end{aligned} \tag{6}$$

(d) Since T^-, T^0, and T^+ are mutually exclusive and collectively exhaustive,

$$\begin{aligned} P\left[T^0|H\right] &= 1 - P\left[T^-|H\right] - P\left[T^+|H\right] \\ &= 1 - 0.841 - 0.006 = 0.153. \end{aligned} \tag{7}$$

We say that a test is a failure if the result is T^0. Thus, given the event H, each test has conditional failure probability of $q = 0.153$, or success probability $p = 1 - q = 0.847$. Given H, the number of trials N until a success is a geometric (p) random variable with PMF

$$P_{N|H}(n) = \begin{cases} (1-p)^{n-1}p & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Problem 7.1.8 Solution

(a) The event B_i that $Y = \Delta/2 + i\Delta$ occurs if and only if $i\Delta \leq X < (i+1)\Delta$. In particular, since X has the uniform $(-r/2, r/2)$ PDF

$$f_X(x) = \begin{cases} 1/r & -r/2 \leq x < r/2, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

we observe that

$$\mathrm{P}[B_i] = \int_{i\Delta}^{(i+1)\Delta} \frac{1}{r}\, dx = \frac{\Delta}{r}. \tag{2}$$

In addition, the conditional PDF of X given B_i is

$$\begin{aligned} f_{X|B_i}(x) &= \begin{cases} f_X(x) / \mathrm{P}[B] & x \in B_i, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 1/\Delta & i\Delta \leq x < (i+1)\Delta, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

It follows that given B_i, $Z = X - Y = X - \Delta/2 - i\Delta$, which is a uniform $(-\Delta/2, \Delta/2)$ random variable. That is,

$$f_{Z|B_i}(z) = \begin{cases} 1/\Delta & -\Delta/2 \leq z < \Delta/2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(b) We observe that $f_{Z|B_i}(z)$ is the same for every i. This implies

$$f_Z(z) = \sum_i \mathrm{P}[B_i] f_{Z|B_i}(z) = f_{Z|B_0}(z) \sum_i \mathrm{P}[B_i] = f_{Z|B_0}(z). \tag{5}$$

Thus, Z is a uniform $(-\Delta/2, \Delta/2)$ random variable. From the definition of a uniform (a, b) random variable, Z has mean and variance

$$\mathrm{E}[Z] = 0, \qquad \mathrm{Var}[Z] = \frac{(\Delta/2 - (-\Delta/2))^2}{12} = \frac{\Delta^2}{12}. \tag{6}$$

Problem 7.1.9 Solution

For this problem, almost any non-uniform random variable X will yield a non-uniform random variable Z. For example, suppose X has the "triangular" PDF

$$f_X(x) = \begin{cases} 8x/r^2 & 0 \leq x \leq r/2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this case, the event B_i that $Y = i\Delta + \Delta/2$ occurs if and only if $i\Delta \leq X < (i+1)\Delta$. Thus

$$\mathrm{P}[B_i] = \int_{i\Delta}^{(i+1)\Delta} \frac{8x}{r^2}\, dx = \frac{8\Delta(i\Delta + \Delta/2)}{r^2}. \tag{2}$$

It follows that the conditional PDF of X given B_i is

$$f_{X|B_i}(x) = \begin{cases} \frac{f_X(x)}{\mathrm{P}[B_i]} & x \in B_i, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} \frac{x}{\Delta(i\Delta+\Delta/2)} & i\Delta \leq x < (i+1)\Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Given event B_i, $Y = i\Delta + \Delta/2$, so that $Z = X - Y = X - i\Delta - \Delta/2$. This implies

$$f_{Z|B_i}(z) = f_{X|B_i}(z + i\Delta + \Delta/2) = \begin{cases} \frac{z+i\Delta+\Delta/2}{\Delta(i\Delta+\Delta/2)} & -\Delta/2 \leq z < \Delta/2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We observe that the PDF of Z depends on which event B_i occurs. Moreover, $f_{Z|B_i}(z)$ is non-uniform for all B_i.

Problem 7.2.1 Solution

The probability of the event B is

$$\begin{aligned} \mathrm{P}[B] = \mathrm{P}[X \geq \mu_X] = \mathrm{P}[X \geq 3] &= P_X(3) + P_X(4) + P_X(5) \\ &= \frac{\binom{5}{3} + \binom{5}{4} + \binom{5}{5}}{32} = 21/32. \end{aligned} \tag{1}$$

The conditional PMF of X given B is

$$P_{X|B}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[B]} & x \in B, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} \binom{5}{x}\frac{1}{21} & x = 3, 4, 5, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The conditional first and second moments of X are

$$\begin{aligned} \mathrm{E}[X|B] = \sum_{x=3}^{5} x P_{X|B}(x) &= 3\binom{5}{3}\frac{1}{21} + 4\binom{5}{4}\frac{1}{21} + 5\binom{5}{5}\frac{1}{21} \\ &= [30 + 20 + 5]/21 = 55/21, \end{aligned} \tag{3}$$

$$\begin{aligned} \mathrm{E}\left[X^2|B\right] = \sum_{x=3}^{5} x^2 P_{X|B}(x) &= 3^2\binom{5}{3}\frac{1}{21} + 4^2\binom{5}{4}\frac{1}{21} + 5^2\binom{5}{5}\frac{1}{21} \\ &= [90 + 80 + 25]/21 = 195/21 = 65/7. \end{aligned} \tag{4}$$

The conditional variance of X is

$$\begin{aligned} \mathrm{Var}[X|B] &= \mathrm{E}\left[X^2|B\right] - (\mathrm{E}[X|B])^2 \\ &= 65/7 - (55/21)^2 = 1070/441 = 2.43. \end{aligned} \tag{5}$$

Problem 7.2.2 Solution

From the solution to Problem 3.4.2, the PMF of X is

$$P_X(x) = \begin{cases} 0.2 & x = -1, \\ 0.5 & x = 0, \\ 0.3 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The event $B = \{|X| > 0\}$ has probability $\mathrm{P}[B] = \mathrm{P}[X \neq 0] = 0.5$. From Theorem 7.1, the conditional PMF of X given B is

$$P_{X|B}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[B]} & x \in B, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} 0.4 & x = -1, \\ 0.6 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The conditional first and second moments of X are

$$\mathrm{E}[X|B] = \sum_x x P_{X|B}(x) = (-1)(0.4) + 1(0.6) = 0.2, \tag{3}$$

$$\mathrm{E}\left[X^2|B\right] = \sum_x x^2 P_{X|B}(x) = (-1)^2(0.4) + 1^2(0.6) = 1. \tag{4}$$

The conditional variance of X is

$$\operatorname{Var}[X|B] = \operatorname{E}\left[X^2|B\right] - (\operatorname{E}[X|B])^2 = 1 - (0.2)^2 = 0.96. \tag{5}$$

Problem 7.2.3 Solution

The PDF of X is

$$f_X(x) = \begin{cases} 1/10 & -5 \le x \le 5, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The event B has probability

$$\operatorname{P}[B] = \operatorname{P}[-3 \le X \le 3] = \int_{-3}^{3} \frac{1}{10}\, dx = \frac{3}{5}. \tag{2}$$

From Definition 7.3, the conditional PDF of X given B is

$$f_{X|B}(x) = \begin{cases} f_X(x) / \operatorname{P}[B] & x \in B, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} 1/6 & |x| \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(b) Given B, we see that X has a uniform PDF over $[a, b]$ with $a = -3$ and $b = 3$. From Theorem 4.6, the conditional expected value of X is $\operatorname{E}[X|B] = (a+b)/2 = 0$.

(c) From Theorem 4.6, the conditional variance of X is $\operatorname{Var}[X|B] = (b-a)^2/12 = 3$.

Problem 7.2.4 Solution

From Definition 4.6, the PDF of Y is

$$f_Y(y) = \begin{cases} (1/5)e^{-y/5} & y \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The event A has probability

$$\begin{aligned} P[A] = P[Y < 2] &= \int_0^2 (1/5)e^{-y/5}\,dy \\ &= -e^{-y/5}\Big|_0^2 = 1 - e^{-2/5}. \end{aligned} \tag{2}$$

From Definition 7.3, the conditional PDF of Y given A is

$$\begin{aligned} f_{Y|A}(y) &= \begin{cases} f_Y(y)\,/\,P[A] & x \in A, \\ 0 & \text{otherwise}, \end{cases} \\ &= \begin{cases} (1/5)e^{-y/5}/(1 - e^{-2/5}) & 0 \le y < 2, \\ 0 & \text{otherwise}. \end{cases} \end{aligned} \tag{3}$$

(b) The conditional expected value of Y given A is

$$E[Y|A] = \int_{-\infty}^{\infty} y f_{Y|A}(y)\,dy = \frac{1/5}{1 - e^{-2/5}} \int_0^2 y e^{-y/5}\,dy. \tag{4}$$

Using the integration by parts formula $\int u\,dv = uv - \int v\,du$ with $u = y$ and $dv = e^{-y/5}\,dy$ yields

$$\begin{aligned} E[Y|A] &= \frac{1/5}{1 - e^{-2/5}} \left(-5ye^{-y/5}\Big|_0^2 + \int_0^2 5e^{-y/5}\,dy \right) \\ &= \frac{1/5}{1 - e^{-2/5}} \left(-10e^{-2/5} - 25e^{-y/5}\Big|_0^2 \right) \\ &= \frac{5 - 7e^{-2/5}}{1 - e^{-2/5}}. \end{aligned} \tag{5}$$

Problem 7.2.5 Solution

The condition *right side of the circle* is $R = [0, 1/2]$. Using the PDF in Example 4.5, we have

$$P[R] = \int_0^{1/2} f_Y(y)\,dy = \int_0^{1/2} 3y^2\,dy = 1/8. \tag{1}$$

Therefore, the conditional PDF of Y given event R is

$$f_{Y|R}(y) = \begin{cases} 24y^2 & 0 \leq y \leq 1/2 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The conditional expected value and mean square value are

$$\mathrm{E}\left[Y|R\right] = \int_{-\infty}^{\infty} y f_{Y|R}(y)\ dy = \int_0^{1/2} 24y^3\,dy = 3/8 \text{ meter}, \tag{3}$$

$$\mathrm{E}\left[Y^2|R\right] = \int_{-\infty}^{\infty} y^2 f_{Y|R}(y)\ dy = \int_0^{1/2} 24y^4\,dy = 3/20 \text{ m}^2. \tag{4}$$

The conditional variance is

$$\mathrm{Var}\left[Y|R\right] = \mathrm{E}\left[Y^2|R\right] - \left(\mathrm{E}\left[Y|R\right]\right)^2 = \frac{3}{20} - \left(\frac{3}{8}\right)^2 = 3/320 \text{ m}^2. \tag{5}$$

The conditional standard deviation is $\sigma_{Y|R} = \sqrt{\mathrm{Var}[Y|R]} = 0.0968$ meters.

Problem 7.2.6 Solution

Recall that the PMF of the number of pages in a fax is

$$P_X(x) = \begin{cases} 0.15 & x = 1, 2, 3, 4, \\ 0.1 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The event that a fax was sent to machine A can be expressed mathematically as the event that the number of pages X is an even number. Similarly, the event that a fax was sent to B is the event that X is an odd number. Since $S_X = \{1, 2, \ldots, 8\}$, we define the set $A = \{2, 4, 6, 8\}$. Using this definition for A, we have that the event that a fax is sent to A is equivalent to the event $X \in A$. The event A has probability

$$\mathrm{P}\left[A\right] = P_X(2) + P_X(4) + P_X(6) + P_X(8) = 0.5. \tag{2}$$

Given the event A, the conditional PMF of X is

$$P_{X|A}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[A]} & x \in A, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} 0.3 & x = 2, 4, \\ 0.2 & x = 6, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The conditional first and second moments of X given A is

$$\begin{aligned}
\mathrm{E}[X|A] &= \sum_x x P_{X|A}(x) \\
&= 2(0.3) + 4(0.3) + 6(0.2) + 8(0.2) = 4.6, && (4)\\
\mathrm{E}\left[X^2|A\right] &= \sum_x x^2 P_{X|A}(x) \\
&= 4(0.3) + 16(0.3) + 36(0.2) + 64(0.2) = 26. && (5)
\end{aligned}$$

The conditional variance and standard deviation are

$$\begin{aligned}
\mathrm{Var}[X|A] &= \mathrm{E}\left[X^2|A\right] - (\mathrm{E}[X|A])^2 = 26 - (4.6)^2 = 4.84, && (6)\\
\sigma_{X|A} &= \sqrt{\mathrm{Var}[X|A]} = 2.2. && (7)
\end{aligned}$$

(b) Let the event B' denote the event that the fax was sent to B and that the fax had no more than 6 pages. Hence, the event $B' = \{1, 3, 5\}$ has probability

$$\mathrm{P}\left[B'\right] = P_X(1) + P_X(3) + P_X(5) = 0.4. \tag{8}$$

The conditional PMF of X given B' is

$$P_{X|B'}(x) = \begin{cases} \frac{P_X(x)}{\mathrm{P}[B']} & x \in B', \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} 3/8 & x = 1, 3, \\ 1/4 & x = 5, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Given the event B', the conditional first and second moments are

$$\begin{aligned}
\mathrm{E}\left[X|B'\right] &= \sum_x x P_{X|B'}(x) \\
&= 1(3/8) + 3(3/8) + 5(1/4) + = 11/4, && (10)\\
\mathrm{E}\left[X^2|B'\right] &= \sum_x x^2 P_{X|B'}(x) \\
&= 1(3/8) + 9(3/8) + 25(1/4) = 10. && (11)
\end{aligned}$$

The conditional variance and standard deviation are

$$\begin{aligned}
\mathrm{Var}[X|B'] &= \mathrm{E}\left[X^2|B'\right] - (\mathrm{E}\left[X|B'\right])^2 = 10 - (11/4)^2 = 39/16, && (12)\\
\sigma_{X|B'} &= \sqrt{\mathrm{Var}[X|B']} = \sqrt{39}/4 \approx 1.56. && (13)
\end{aligned}$$

Problem 7.2.7 Solution

(a) Consider each circuit test as a Bernoulli trial such that a failed circuit is called a success. The number of trials until the first success (i.e. a failed circuit) has the geometric PMF

$$P_N(n) = \begin{cases} (1-p)^{n-1}p & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) The probability there are at least 20 tests is

$$\mathrm{P}[B] = \mathrm{P}[N \geq 20] = \sum_{n=20}^{\infty} P_N(n) = (1-p)^{19}. \tag{2}$$

Note that $(1-p)^{19}$ is just the probability that the first 19 circuits pass the test, which is what we would expect since there must be at least 20 tests if the first 19 circuits pass. The conditional PMF of N given B is

$$\begin{aligned} P_{N|B}(n) &= \begin{cases} \frac{P_N(n)}{\mathrm{P}[B]} & n \in B, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} (1-p)^{n-20}p & n = 20, 21, \ldots, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

(c) Given the event B, the conditional expectation of N is

$$\mathrm{E}[N|B] = \sum_n n P_{N|B}(n) = \sum_{n=20}^{\infty} n(1-p)^{n-20}p. \tag{4}$$

Making the substitution $j = n - 19$ yields

$$\mathrm{E}[N|B] = \sum_{j=1}^{\infty} (j+19)(1-p)^{j-1}p = 1/p + 19. \tag{5}$$

We see that in the above sum, we effectively have the expected value of $J + 19$ where J is geometric random variable with parameter p. This is not surprising since the $N \geq 20$ iff we observed 19 successful tests. After 19 successful tests, the number of additional tests needed to find the first failure is still a geometric random variable with mean $1/p$.

Problem 7.2.8 Solution

From Definition 4.8, the PDF of W is

$$f_W(w) = \frac{1}{\sqrt{32\pi}} e^{-w^2/32}. \tag{1}$$

(a) Since W has expected value $\mu = 0$, $f_W(w)$ is symmetric about $w = 0$. Hence $\mathrm{P}[C] = \mathrm{P}[W > 0] = 1/2$. From Definition 7.3, the conditional PDF of W given C is

$$\begin{aligned} f_{W|C}(w) &= \begin{cases} f_W(w) / \mathrm{P}[C] & w \in C, \\ 0 & \text{otherwise}, \end{cases} \\ &= \begin{cases} 2e^{-w^2/32}/\sqrt{32\pi} & w > 0, \\ 0 & \text{otherwise}. \end{cases} \end{aligned} \tag{2}$$

(b) The conditional expected value of W given C is

$$\begin{aligned} \mathrm{E}[W|C] &= \int_{-\infty}^{\infty} w f_{W|C}(w)\, dw \\ &= \frac{2}{4\sqrt{2\pi}} \int_0^{\infty} w e^{-w^2/32}\, dw. \end{aligned} \tag{3}$$

Making the substitution $v = w^2/32$, we obtain

$$\mathrm{E}[W|C] = \frac{32}{\sqrt{32\pi}} \int_0^{\infty} e^{-v}\, dv = \frac{32}{\sqrt{32\pi}}. \tag{4}$$

(c) The conditional second moment of W is

$$\mathrm{E}\left[W^2|C\right] = \int_{-\infty}^{\infty} w^2 f_{W|C}(w)\, dw = 2\int_0^{\infty} w^2 f_W(w)\, dw. \tag{5}$$

We observe that $w^2 f_W(w)$ is an even function. Hence

$$\begin{aligned}\mathrm{E}\left[W^2|C\right] &= 2\int_0^{\infty} w^2 f_W(w)\, dw \\ &= \int_{-\infty}^{\infty} w^2 f_W(w)\, dw = \mathrm{E}\left[W^2\right] = \sigma^2 = 16.\end{aligned} \tag{6}$$

Lastly, the conditional variance of W given C is

$$\mathrm{Var}[W|C] = \mathrm{E}\left[W^2|C\right] - (\mathrm{E}\left[W|C\right])^2 = 16 - 32/\pi = 5.81. \tag{7}$$

Problem 7.2.9 Solution

(a) We first find the conditional PDF of T. The PDF of T is

$$f_T(t) = \begin{cases} 100e^{-100t} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The conditioning event has probability

$$\mathrm{P}\left[T > 0.02\right] = \int_{0.02}^{\infty} f_T(t)\, dt = -e^{-100t}\Big|_{0.02}^{\infty} = e^{-2}. \tag{2}$$

From Definition 7.3, the conditional PDF of T is

$$\begin{aligned} f_{T|T>0.02}(t) &= \begin{cases} \frac{f_T(t)}{\mathrm{P}[T>0.02]} & t \geq 0.02, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 100e^{-100(t-0.02)} & t \geq 0.02, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

The conditional expected value of T is

$$\mathrm{E}\left[T|T > 0.02\right] = \int_{0.02}^{\infty} t(100)e^{-100(t-0.02)}\, dt. \tag{4}$$

The substitution $\tau = t - 0.02$ yields

$$\begin{aligned}
\mathrm{E}\left[T|T > 0.02\right] &= \int_{0}^{\infty} (\tau + 0.02)(100)e^{-100\tau}\, d\tau \\
&= \int_{0}^{\infty} (\tau + 0.02) f_T(\tau)\; d\tau \\
&= \mathrm{E}\left[T + 0.02\right] = 0.03. \qquad (5)
\end{aligned}$$

(b) The conditional second moment of T is

$$\mathrm{E}\left[T^2|T > 0.02\right] = \int_{0.02}^{\infty} t^2(100)e^{-100(t-0.02)}\, dt. \tag{6}$$

The substitution $\tau = t - 0.02$ yields

$$\begin{aligned}
\mathrm{E}\left[T^2|T > 0.02\right] &= \int_{0}^{\infty} (\tau + 0.02)^2(100)e^{-100\tau}\, d\tau \\
&= \int_{0}^{\infty} (\tau + 0.02)^2 f_T(\tau)\; d\tau \\
&= \mathrm{E}\left[(T + 0.02)^2\right]. \qquad (7)
\end{aligned}$$

Now we can calculate the conditional variance.

$$\begin{aligned}
\mathrm{Var}[T|T > 0.02] &= \mathrm{E}\left[T^2|T > 0.02\right] - \left(\mathrm{E}\left[T|T > 0.02\right]\right)^2 \\
&= \mathrm{E}\left[(T + 0.02)^2\right] - \left(\mathrm{E}\left[T + 0.02\right]\right)^2 \\
&= \mathrm{Var}[T + 0.02] \\
&= \mathrm{Var}[T] = 0.01. \qquad (8)
\end{aligned}$$

Problem 7.2.10 Solution

(a) We have to calculate Roy's probability of winning for each bike. For both bikes

$$\mathrm{P}[W] = \mathrm{P}[60/V \leq 1] = \mathrm{P}[V \geq 60]. \tag{1}$$

Let's use B_i to denote the event that he choses bike i and then derive the conditional PDF of V given B_i for each bike. If he chooses the carbon fiber bike, V is a continuous uniform (a, b) random variable with

$$\mathrm{E}[V|B_1] = \frac{a+b}{2} = 58, \qquad \mathrm{Var}[V|B_1] = \frac{(b-a)^2}{12} = 12. \tag{2}$$

We can solve these equations and find $a = 52$ and $b = 64$ so that the PDF of V given B_1 is

$$f_{V|B_1}(v) = \begin{cases} 1/12 & 52 \leq v \leq 64, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

It follows that if Roy rides the carbon bike then

$$\mathrm{P}[W|B_1] = \mathrm{P}[V \geq 60 B_1] = \int_{60}^{64} \frac{1}{12}\, dv = 1/3. \tag{4}$$

If Roy rides the titanium bike (event B_2), then V is an exponential $(\lambda = 1/60)$ random variable with conditional PDF

$$f_{V|B_2}(v) = \begin{cases} \lambda e^{-\lambda v} & v \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

In this case,

$$\begin{aligned} \mathrm{P}[W|B_2] &= \mathrm{P}[V \geq 60|B_2] \\ &= \int_{60}^{\infty} f_{V|B_2}(v)\, dv \\ &= -e^{-\lambda v}\Big|_{60}^{\infty} = e^{-60\lambda} = e^{-1}. \end{aligned} \tag{6}$$

Since $e^{-1} = 0.368 > 1/3$, Roy chooses the titanium bike and $\mathrm{P}[W] = \mathrm{P}[W|B_2] = e^{-1}$.

(b) This is just the law of total probability with $\mathrm{P}[B_1] = \mathrm{P}[B_2] = 1/2$ so that

$$\begin{aligned}
\mathrm{P}[W] &= \mathrm{P}[B_1]\,\mathrm{P}[W|B_1] + \mathrm{P}[B_2]\,\mathrm{P}[W|B_2] \\
&= \frac{1}{2}\frac{1}{3} + \frac{1}{2}e^{-1} = \frac{1+3e^{-1}}{6} = 0.3506.
\end{aligned} \tag{7}$$

Problem 7.2.11 Solution

(a) In Problem 4.7.8, we found that the PDF of D is

$$f_D(y) = \begin{cases} 0.3\delta(y) & y < 60, \\ 0.07e^{-(y-60)/10} & y \geq 60. \end{cases} \tag{1}$$

First, we observe that $D > 0$ if the throw is good so that $\mathrm{P}[D > 0] = 0.7$. A second way to find this probability is

$$\mathrm{P}[D > 0] = \int_{0^+}^{\infty} f_D(y)\; dy = 0.7. \tag{2}$$

From Definition 7.3, we can write

$$\begin{aligned}
f_{D|D>0}(y) &= \begin{cases} \frac{f_D(y)}{\mathrm{P}[D>0]} & y > 0, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} (1/10)e^{-(y-60)/10} & y \geq 60, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{3}$$

(b) If instead we learn that $D \leq 70$, we can calculate the conditional PDF by first calculating

$$\begin{aligned}
\mathrm{P}[D \leq 70] &= \int_0^{70} f_D(y)\; dy \\
&= \int_0^{60} 0.3\delta(y)\, dy + \int_{60}^{70} 0.07e^{-(y-60)/10}\, dy \\
&= 0.3 + -0.7e^{-(y-60)/10}\Big|_{60}^{70} = 1 - 0.7e^{-1}.
\end{aligned} \tag{4}$$

The conditional PDF is

$$f_{D|D\leq 70}(y) = \begin{cases} \frac{f_D(y)}{\mathrm{P}[D\leq 70]} & y \leq 70, \\ 0 & \text{otherwise,} \end{cases}$$

$$= \begin{cases} \frac{0.3}{1-0.7e^{-1}}\delta(y) & 0 \leq y < 60, \\ \frac{0.07}{1-0.7e^{-1}}e^{-(y-60)/10} & 60 \leq y \leq 70, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 7.3.1 Solution

X and Y each have the discrete uniform PMF

$$P_X(x) = P_Y(x) = \begin{cases} 0.1 & x = 1, 2, \ldots, 10, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The joint PMF of X and Y is

$$P_{X,Y}(x, y) = P_X(x)\, P_Y(y)$$

$$= \begin{cases} 0.01 & x = 1, 2, \ldots, 10; y = 1, 2, \ldots, 10, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The event A occurs iff $X > 5$ and $Y > 5$ and has probability

$$\mathrm{P}[A] = \mathrm{P}[X > 5, Y > 5] = \sum_{x=6}^{10}\sum_{y=6}^{10} 0.01 = 0.25. \tag{3}$$

Alternatively, we could have used independence of X and Y to write $\mathrm{P}[A] = \mathrm{P}[X > 5]\,\mathrm{P}[Y > 5] = 1/4$. From Theorem 7.6,

$$P_{X,Y|A}(x, y) = \begin{cases} \frac{P_{X,Y}(x,y)}{\mathrm{P}[A]} & (x, y) \in A, \\ 0 & \text{otherwise,} \end{cases}$$

$$= \begin{cases} 0.04 & x = 6, \ldots, 10; y = 6, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 7.3.2 Solution

X and Y each have the discrete uniform PMF

$$P_X(x) = P_Y(x) = \begin{cases} 0.1 & x = 1, 2, \ldots, 10, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The joint PMF of X and Y is

$$\begin{aligned} P_{X,Y}(x, y) &= P_X(x) P_Y(y) \\ &= \begin{cases} 0.01 & x = 1, 2, \ldots, 10; y = 1, 2, \ldots, 10, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

The event B occurs iff $X \leq 5$ and $Y \leq 5$ and has probability

$$\mathrm{P}[B] = \mathrm{P}[X \leq 5, Y \leq 5] = \sum_{x=1}^{5} \sum_{y=1}^{5} 0.01 = 0.25. \tag{3}$$

From Theorem 7.6,

$$\begin{aligned} P_{X,Y|B}(x, y) &= \begin{cases} \frac{P_{X,Y}(x,y)}{\mathrm{P}[B]} & (x, y) \in A, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 0.04 & x = 1, \ldots, 5; y = 1, \ldots, 5, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

Problem 7.3.3 Solution

Given the event $A = \{X + Y \leq 1\}$, we wish to find $f_{X,Y|A}(x, y)$. First we find

$$\mathrm{P}[A] = \int_0^1 \int_0^{1-x} 6e^{-(2x+3y)}\, dy\, dx = 1 - 3e^{-2} + 2e^{-3}. \tag{1}$$

So then

$$f_{X,Y|A}(x, y) = \begin{cases} \frac{6e^{-(2x+3y)}}{1-3e^{-2}+2e^{-3}} & x + y \leq 1, x \geq 0, y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 7.3.4 Solution

(a) First we observe that for $n = 1, 2, \ldots$, the marginal PMF of N satisfies

$$P_N(n) = \sum_{k=1}^{n} P_{N,K}(n,k)$$
$$= (1-p)^{n-1} p \sum_{k=1}^{n} \frac{1}{n} = (1-p)^{n-1} p. \tag{1}$$

Thus, the event B has probability

$$\mathrm{P}[B] = \sum_{n=10}^{\infty} P_N(n)$$
$$= (1-p)^9 p[1 + (1-p) + (1-p)^2 + \cdots]$$
$$= (1-p)^9. \tag{2}$$

From Theorem 7.6,

$$P_{N,K|B}(n,k) = \begin{cases} \frac{P_{N,K}(n,k)}{\mathrm{P}[B]} & n, k \in B, \\ 0 & \text{otherwise}, \end{cases}$$
$$= \begin{cases} (1-p)^{n-10} p/n & n = 10, 11, \ldots; k = 1, \ldots, n, \\ 0 & \text{otherwise}. \end{cases} \tag{3}$$

The conditional PMF $P_{N|B}(n|b)$ could be found directly from $P_N(n)$ using Theorem 7.1. However, we can also find it just by summing the conditional joint PMF.

$$P_{N|B}(n) = \sum_{k=1}^{n} P_{N,K|B}(n,k)$$
$$= \begin{cases} (1-p)^{n-10} p & n = 10, 11, \ldots, \\ 0 & \text{otherwise}. \end{cases} \tag{4}$$

(b) From the conditional PMF $P_{N|B}(n)$, we can calculate directly the conditional moments of N given B. Instead, however, we observe that given B, $N' = N - 9$ has a geometric PMF with mean $1/p$. That is, for $n = 1, 2, \ldots$,

$$P_{N'|B}(n) = \mathrm{P}\left[N = n + 9|B\right] = P_{N|B}(n + 9) = (1 - p)^{n-1}p. \tag{5}$$

Hence, given B, $N = N' + 9$ and we can calculate the conditional expectations

$$\mathrm{E}\left[N|B\right] = \mathrm{E}\left[N' + 9|B\right] = \mathrm{E}\left[N'|B\right] + 9 = 1/p + 9, \tag{6}$$

$$\mathrm{Var}[N|B] = \mathrm{Var}[N' + 9|B] = \mathrm{Var}[N'|B] = (1 - p)/p^2. \tag{7}$$

Note that further along in the problem we will need $\mathrm{E}[N^2|B]$ which we now calculate.

$$\mathrm{E}\left[N^2|B\right] = \mathrm{Var}[N|B] + (\mathrm{E}\left[N|B\right])^2, \tag{8}$$

$$= \frac{2}{p^2} + \frac{17}{p} + 81. \tag{9}$$

For the conditional moments of K, we work directly with the conditional PMF $P_{N,K|B}(n, k)$.

$$\mathrm{E}\left[K|B\right] = \sum_{n=10}^{\infty} \sum_{k=1}^{n} k \frac{(1 - p)^{n-10}p}{n} = \sum_{n=10}^{\infty} \frac{(1 - p)^{n-10}p}{n} \sum_{k=1}^{n} k. \tag{10}$$

Since $\sum_{k=1}^{n} k = n(n + 1)/2$,

$$\mathrm{E}\left[K|B\right] = \sum_{n=1}^{\infty} \frac{n + 1}{2} (1 - p)^{n-1}p = \frac{1}{2} \mathrm{E}\left[N + 1|B\right] = \frac{1}{2p} + 5. \tag{11}$$

We now can calculate the conditional expectation of the sum.

$$\mathrm{E}\left[N + K|B\right] = \mathrm{E}\left[N|B\right] + \mathrm{E}\left[K|B\right]$$

$$= 1/p + 9 + 1/(2p) + 5 = \frac{3}{2p} + 14. \tag{12}$$

The conditional second moment of K is

$$\mathrm{E}\left[K^2|B\right] = \sum_{n=10}^{\infty} \sum_{k=1}^{n} k^2 \frac{(1 - p)^{n-10}p}{n}$$

$$= \sum_{n=10}^{\infty} \frac{(1 - p)^{n-10}p}{n} \sum_{k=1}^{n} k^2. \tag{13}$$

Using the identity $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$, we obtain

$$\begin{aligned} \mathrm{E}\left[K^2|B\right] &= \sum_{n=10}^{\infty} \frac{(n+1)(2n+1)}{6}(1-p)^{n-10}p \\ &= \frac{1}{6}\,\mathrm{E}\left[(N+1)(2N+1)|B\right]. \end{aligned} \tag{14}$$

Applying the values of $\mathrm{E}[N|B]$ and $\mathrm{E}[N^2|B]$ found above, we find that

$$\mathrm{E}\left[K^2|B\right] = \frac{\mathrm{E}\left[N^2|B\right]}{3} + \frac{\mathrm{E}\left[N|B\right]}{2} + \frac{1}{6} = \frac{2}{3p^2} + \frac{37}{6p} + 31\,\frac{2}{3}. \tag{15}$$

Thus, we can calculate the conditional variance of K.

$$\mathrm{Var}[K|B] = \mathrm{E}\left[K^2|B\right] - \left(\mathrm{E}\left[K|B\right]\right)^2 = \frac{5}{12p^2} - \frac{7}{6p} + 6\frac{2}{3}. \tag{16}$$

To find the conditional correlation of N and K,

$$\begin{aligned} \mathrm{E}\left[NK|B\right] &= \sum_{n=10}^{\infty}\sum_{k=1}^{n} nk\frac{(1-p)^{n-10}p}{n} \\ &= \sum_{n=10}^{\infty}(1-p)^{n-1}p\sum_{k=1}^{n} k. \end{aligned} \tag{17}$$

Since $\sum_{k=1}^{n} k = n(n+1)/2$,

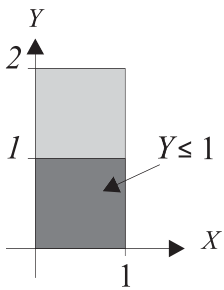
$$\begin{aligned} \mathrm{E}\left[NK|B\right] &= \sum_{n=10}^{\infty} \frac{n(n+1)}{2}(1-p)^{n-10}p \\ &= \frac{1}{2}\,\mathrm{E}\left[N(N+1)|B\right] = \frac{1}{p^2} + \frac{9}{p} + 45. \end{aligned} \tag{18}$$

Problem 7.3.5 Solution

The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} (x+y)/3 & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The probability that $Y \leq 1$ is



$$
\begin{aligned}
\mathrm{P}[A] = \mathrm{P}[Y \leq 1] &= \iint_{y \leq 1} f_{X,Y}(x,y)\, dx\, dy \\
&= \int_0^1 \int_0^1 \frac{x+y}{3}\, dy\, dx \\
&= \int_0^1 \left(\frac{xy}{3} + \frac{y^2}{6} \Big|_{y=0}^{y=1} \right) dx \\
&= \int_0^1 \frac{2x+1}{6}\, dx \\
&= \frac{x^2}{6} + \frac{x}{6} \Big|_0^1 = \frac{1}{3}.
\end{aligned} \tag{2}
$$

(b) By Definition 7.7, the conditional joint PDF of X and Y given A is

$$
\begin{aligned}
f_{X,Y|A}(x,y) &= \begin{cases} \frac{f_{X,Y}(x,y)}{\mathrm{P}[A]} & (x,y) \in A, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} x+y & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{3}
$$

From $f_{X,Y|A}(x,y)$, we find the conditional marginal PDF $f_{X|A}(x)$. For $0 \leq x \leq 1$,

$$
\begin{aligned}
f_{X|A}(x) &= \int_{-\infty}^{\infty} f_{X,Y|A}(x,y)\, dy \\
&= \int_0^1 (x+y)\, dy = xy + \frac{y^2}{2} \Big|_{y=0}^{y=1} = x + \frac{1}{2}.
\end{aligned} \tag{4}
$$

The complete expression is

$$
f_{X|A}(x) = \begin{cases} x + 1/2 & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}
$$

For $0 \leq y \leq 1$, the conditional marginal PDF of Y is

$$
\begin{aligned}
f_{Y|A}(y) &= \int_{-\infty}^{\infty} f_{X,Y|A}(x,y)\ dx \\
&= \int_0^1 (x+y)\,dx = \frac{x^2}{2} + xy\Big|_{x=0}^{x=1} = y + \frac{1}{2}.
\end{aligned} \tag{6}
$$

The complete expression is

$$
f_{Y|A}(y) = \begin{cases} y + 1/2 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}
$$

Problem 7.3.6 Solution

Random variables X and Y have joint PDF

$$
f_{X,Y}(x,y) = \begin{cases} (4x+2y)/3 & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}
$$

(a) The probability of event $A = \{Y \leq 1/2\}$ is

$$
\begin{aligned}
\mathrm{P}[A] &= \iint_{y \leq 1/2} f_{X,Y}(x,y)\ dy\,dx \\
&= \int_0^1 \int_0^{1/2} \frac{4x+2y}{3}\,dy\,dx.
\end{aligned} \tag{2}
$$

With some calculus,

$$
\begin{aligned}
\mathrm{P}[A] &= \int_0^1 \frac{4xy+y^2}{3}\Big|_{y=0}^{y=1/2} dx \\
&= \int_0^1 \frac{2x+1/4}{3}\,dx = \frac{x^2}{3} + \frac{x}{12}\Big|_0^1 = \frac{5}{12}.
\end{aligned} \tag{3}
$$

(b) The conditional joint PDF of X and Y given A is

$$
\begin{aligned}
f_{X,Y|A}(x,y) &= \begin{cases} \frac{f_{X,Y}(x,y)}{\mathrm{P}[A]} & (x,y) \in A, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} 8(2x+y)/5 & 0 \leq x \leq 1, 0 \leq y \leq 1/2, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{4}
$$

For $0 \leq x \leq 1$, the PDF of X given A is

$$f_{X|A}(x) = \int_{-\infty}^{\infty} f_{X,Y|A}(x,y)\ dy = \frac{8}{5}\int_0^{1/2} (2x+y)\,dy$$
$$= \frac{8}{5}\left(2xy + \frac{y^2}{2}\right)\Big|_{y=0}^{y=1/2} = \frac{8x+1}{5}. \tag{5}$$

The complete expression is

$$f_{X|A}(x) = \begin{cases} (8x+1)/5 & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

For $0 \leq y \leq 1/2$, the conditional marginal PDF of Y given A is

$$f_{Y|A}(y) = \int_{-\infty}^{\infty} f_{X,Y|A}(x,y)\ dx = \frac{8}{5}\int_0^{1} (2x+y)\,dx$$
$$= \frac{8x^2+8xy}{5}\Big|_{x=0}^{x=1} = \frac{8y+8}{5}. \tag{7}$$

The complete expression is

$$f_{Y|A}(y) = \begin{cases} (8y+8)/5 & 0 \leq y \leq 1/2, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Problem 7.3.7 Solution

(a) For a woman, B and T are independent and thus

$$P_{B,T|W}(b,t) = P_{B|W}(b)\,P_{T|W}(t)\,. \tag{1}$$

Expressed in the form of a table, we have

$P_{B,T\|W}(b,t)$	$t=0$	$t=1$	$t=2$
$b=0$	0.36	0.12	0.12
$b=1$	0.18	0.06	0.06
$b=2$	0.06	0.02	0.02

(2)

(b) For a man, B and T are independent and thus

$$P_{B,T|M}(b,t) = P_{B|M}(b)\, P_{T|M}(t)\,. \tag{3}$$

Expressed in the form of a table, we have

$P_{B,T\|M}(b,t)$	$t=0$	$t=1$	$t=2$
$b=0$	0.04	0.04	0.12
$b=1$	0.06	0.06	0.18
$b=2$	0.10	0.10	0.30

(4)

(c) To find the joint PMF, we use the law of total probability to write

$$\begin{aligned} P_{B,T}(b,t) &= \mathrm{P}[W]\, P_{B,T|W}(b,t) + \mathrm{P}[M]\, P_{B,T|M}(b,t) \\ &= \frac{1}{2} P_{B,T|W}(b,t) + \frac{1}{2} P_{B,T|M}(b,t)\,. \end{aligned} \tag{5}$$

Equation (5) says to add the tables for $P_{B,T|W}(b,t)$ and $P_{B,T|M}(b,t)$ and divide by two. This yields

$P_{B,T}(b,t)$	$t=0$	$t=1$	$t=2$
$b=0$	0.20	0.08	0.12
$b=1$	0.12	0.06	0.12
$b=2$	0.08	0.06	0.16

(6)

(d) To check independence, we compute the marginal PMFs by writing the row and column sums:

$P_{B,T}(b,t)$	$t=0$	$t=1$	$t=2$	$P_B(b)$
$b=0$	0.20	0.08	0.12	0.40
$b=1$	0.12	0.06	0.12	0.30
$b=2$	0.08	0.06	0.16	0.30
$P_T(t)$	0.40	0.20	0.40	

(7)

We see that B and T are dependent since $P_{B,T}(b,t) \neq P_B(b)P_T(t)$. For example, $P_{B,T}(0,0) = 0.20$ but $P_B(0)P_T(0) = (0.40)(0.40) = 0.16$.

Now we calculate the covariance $\text{Cov}[B,T] = \text{E}[BT] - \text{E}[B]\,\text{E}[T]$ via

$$\text{E}\,[B] = \sum_{b=0}^{2} bP_B(b) = (1)(0.30) + (2)(0.30) = 0.90, \tag{8}$$

$$\text{E}\,[T] = \sum_{t=0}^{2} tP_T(t) = (1)(0.20) + (2)(0.40) = 1.0, \tag{9}$$

$$\begin{aligned}\text{E}\,[BT] &= \sum_{b=0}^{2}\sum_{t=0}^{2} btP_{B,T}(b,t) \\ &= (1\cdot 1)(0.06) + (1\cdot 2)(0.12) \\ &\quad + (2\cdot 1)(0.06) + (2\cdot 2)(0.16) \\ &= 1.06. \end{aligned} \tag{10}$$

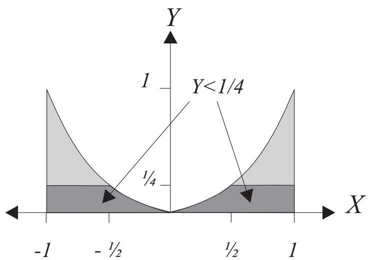
Thus the covariance is

$$\text{Cov}\,[B,T] = \text{E}\,[BT] - \text{E}\,[B]\,\text{E}\,[T] = 1.06 - (0.90)(1.0) = 0.16. \tag{11}$$

Thus baldness B and time watching football T are positively correlated. This should seem odd since B and T are uncorrelated for men and also uncorrelated for women. However, men tend to be balder and watch more football while women tend to be not bald and watch less football. Averaged over mean and women, the result is that baldness and football watching are positively correlated because given that a person watches more football, the person is more likely to be a man and thus is more likely to be balder than average (since the average includes women who tend not to be bald.)

Problem 7.3.8 Solution

(a) The event $A = \{Y \leq 1/4\}$ has probability



$$
\begin{aligned}
\mathrm{P}[A] &= 2\int_0^{1/2}\int_0^{x^2} \frac{5x^2}{2}\,dy\,dx \\
&\quad + 2\int_{1/2}^{1}\int_0^{1/4} \frac{5x^2}{2}\,dy\,dx \\
&= \int_0^{1/2} 5x^4\,dx + \int_{1/2}^{1} \frac{5x^2}{4}\,dx \\
&= x^5\Big|_0^{1/2} + 5x^3/12\Big|_{1/2}^{1} \\
&= 19/48. \qquad (1)
\end{aligned}
$$

This implies

$$
\begin{aligned}
f_{X,Y|A}(x,y) &= \begin{cases} f_{X,Y}(x,y)\,/\,\mathrm{P}[A] & (x,y)\in A, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} 120x^2/19 & -1\le x\le 1, 0\le y\le x^2, y\le 1/4, \\ 0 & \text{otherwise.} \end{cases} \qquad (2)
\end{aligned}
$$

(b)

$$
\begin{aligned}
f_{Y|A}(y) &= \int_{-\infty}^{\infty} f_{X,Y|A}(x,y)\;dx \\
&= 2\int_{\sqrt{y}}^{1} \frac{120x^2}{19}\,dx \\
&= \begin{cases} \frac{80}{19}(1-y^{3/2}) & 0\le y\le 1/4, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)
\end{aligned}
$$

(c) The conditional expectation of Y given A is

$$
\begin{aligned}
\mathrm{E}[Y|A] &= \int_0^{1/4} y\frac{80}{19}(1-y^{3/2})\,dy \\
&= \frac{80}{19}\left(\frac{y^2}{2} - \frac{2y^{7/2}}{7}\right)\Bigg|_0^{1/4} = \frac{65}{532}. \qquad (4)
\end{aligned}
$$

(d) To find $f_{X|A}(x)$, we can write $f_{X|A}(x) = \int_{-\infty}^{\infty} f_{X,Y|A}(x,y)\,dy$. However, when we substitute $f_{X,Y|A}(x,y)$, the limits will depend on the value of x. When $|x| \leq 1/2$,

$$f_{X|A}(x) = \int_0^{x^2} \frac{120x^2}{19}\,dy = \frac{120x^4}{19}. \tag{5}$$

When $-1 \leq x \leq -1/2$ or $1/2 \leq x \leq 1$,

$$f_{X|A}(x) = \int_0^{1/4} \frac{120x^2}{19}\,dy = \frac{30x^2}{19}. \tag{6}$$

The complete expression for the conditional PDF of X given A is

$$f_{X|A}(x) = \begin{cases} 30x^2/19 & -1 \leq x \leq -1/2, \\ 120x^4/19 & -1/2 \leq x \leq 1/2, \\ 30x^2/19 & 1/2 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

(e) The conditional mean of X given A is

$$\mathrm{E}\left[X|A\right] = \int_{-1}^{-1/2} \frac{30x^3}{19}\,dx + \int_{-1/2}^{1/2} \frac{120x^5}{19}\,dx + \int_{1/2}^{1} \frac{30x^3}{19}\,dx = 0. \tag{8}$$

Problem 7.3.9 Solution

X and Y are independent random variables with PDFs

$$f_X(x) = \begin{cases} 2x & 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 3y^2 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For the event $A = \{X > Y\}$, this problem asks us to calculate the conditional expectations $\mathrm{E}[X|A]$ and $\mathrm{E}[Y|A]$. We will do this using the conditional joint PDF $f_{X,Y|A}(x,y)$. Since X and Y are independent, it is tempting to argue that the event $X > Y$ does not alter the probability model for X and Y. Unfortunately, this is not the case. When we learn that $X > Y$, it increases the probability that X is large and Y is small. We will see this when we compare the conditional expectations $\mathrm{E}[X|A]$ and $\mathrm{E}[Y|A]$ to $\mathrm{E}[X]$ and $\mathrm{E}[Y]$.

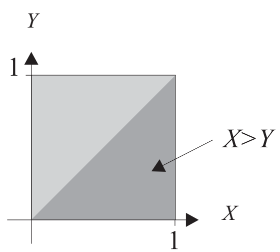
(a) We can calculate the unconditional expectations, E$[X]$ and E$[Y]$, using the marginal PDFs $f_X(x)$ and $f_Y(y)$.

$$\mathrm{E}\left[X\right] = \int_{-\infty}^{\infty} f_X\left(x\right)\, dx = \int_0^1 2x^2\, dx = 2/3, \tag{2}$$

$$\mathrm{E}\left[Y\right] = \int_{-\infty}^{\infty} f_Y\left(y\right)\, dy = \int_0^1 3y^3\, dy = 3/4. \tag{3}$$

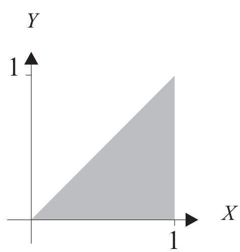
(b) First, we need to calculate the conditional joint PDF $ipdf X, Y|Ax, y$. The first step is to write down the joint PDF of X and Y:

$$f_{X,Y}\left(x,y\right) = f_X\left(x\right) f_Y\left(y\right) = \begin{cases} 6xy^2 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$



The event A has probability

$$\begin{aligned} \mathrm{P}\left[A\right] &= \iint_{x>y} f_{X,Y}\left(x,y\right)\, dy\, dx \\ &= \int_0^1 \int_0^x 6xy^2\, dy\, dx \\ &= \int_0^1 2x^4\, dx = 2/5. \end{aligned} \tag{5}$$



The conditional joint PDF of X and Y given A is

$$\begin{aligned} f_{X,Y|A}(x,y) &= \begin{cases} \frac{f_{X,Y}(x,y)}{\mathrm{P}[A]} & (x,y) \in A, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 15xy^2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

The triangular region of nonzero probability is a signal that given A, X and Y are no longer independent. The conditional expected value of X given A

is

$$
\begin{aligned}
\mathrm{E}\left[X|A\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y|A}(x,y|a)\, x, y\, dy\, dx \\
&= 15\int_0^1 x^2 \int_0^x y^2\, dy\, dx \\
&= 5\int_0^1 x^5\, dx = 5/6. \qquad (7)
\end{aligned}
$$

The conditional expected value of Y given A is

$$
\begin{aligned}
\mathrm{E}\left[Y|A\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y|A}(x,y)\ dy\, dx \\
&= 15\int_0^1 x \int_0^x y^3\, dy\, dx \\
&= \frac{15}{4}\int_0^1 x^5\, dx = 5/8. \qquad (8)
\end{aligned}
$$

We see that $\mathrm{E}[X|A] > \mathrm{E}[X]$ while $\mathrm{E}[Y|A] < \mathrm{E}[Y]$. That is, learning $X > Y$ gives us a clue that X may be larger than usual while Y may be smaller than usual.

Problem 7.4.1 Solution

These conditional PDFs require no calculation. Straight from the problem statement,

$$
\begin{aligned}
f_{Y_1|X}(y_1|1) &= \frac{1}{\sqrt{2\pi}} e^{-(y_1-x)^2/2}, \qquad (1) \\
f_{Y_2|X}(y_2|x) &= \frac{1}{\sqrt{2\pi x^2}} e^{-(y_2-x)^2/2x^2}. \qquad (2)
\end{aligned}
$$

Conditional PDFs like $f_{Y_1|X}(y_1|x)$ occur often. Conditional PDFs resembling $f_{Y_2|X}(y_2|x$ are fairly uncommon.

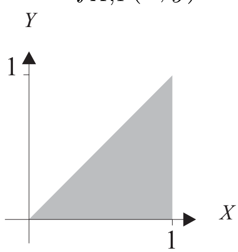
Problem 7.4.2 Solution

From the problem statement, we know that

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$f_{Y|X}(y|x) = \begin{cases} 1/x & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Since $f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x)$, the joint PDF is simply



$$f_{X,Y}(x,y) = \begin{cases} 1/x & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The triangular region of the X,Y plane where $f_{X,Y}(x,y) > 0$ is shown on the left.

Problem 7.4.3 Solution

This problem is mostly about translating words to math. From the words, we learned that

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$f_{Y|X}(y|x) = \begin{cases} 1/(1+x) & 0 \le y \le 1+x, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It follows that the joint PDF is

$$\begin{aligned} f_{X,Y}(x,y) &= f_{Y|X}(y|x) f_X(x) \\ &= \begin{cases} 1/(1+x) & 0 \le x \le 1, 0 \le y \le 1+x, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

Problem 7.4.4 Solution

Given $X = x$, $Y = x + Z$. Since Z is independent of X, we see that conditioned on $X = x$, Y is the sum of a constant x and additive Gaussian noise Z with standard deviation 1. That is, given $X = x$, Y is a Gaussian $(x, 1)$ random variable. The conditional PDF of Y is

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}. \tag{1}$$

If this argument is unconvincing, we observe that the conditional CDF of Y given $X = x$ is

$$\begin{aligned} F_{Y|X}(y|x) &= \mathrm{P}\left[Y \leq y | X = x\right] \\ &= \mathrm{P}\left[x + Z \leq y | X = x\right] \\ &= \mathrm{P}\left[x + Z \leq y\right] \\ &= \mathrm{P}\left[Z \leq y - x\right] \\ &= F_Z(y - x), \end{aligned} \tag{2}$$

where the key step is that $X = x$ conditioning is removed in (2) because Z is independent of X. Taking a derivative with respect to y, we obtain

$$f_{Y|X}(y|x) = \frac{dF_{Y|X}(y|x)}{dy} = \frac{dF_Z(y-x)}{dy} = f_Z(y - x). \tag{3}$$

Since Z is Gaussian $(0, 1)$, the conditional PDF of Y is

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}. \tag{4}$$

Problem 7.4.5 Solution

The main part of this problem is just interpreting the problem statement. No calculations are necessary. Since a trip is equally likely to last 2, 3 or 4 days,

$$P_D(d) = \begin{cases} 1/3 & d = 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Given a trip lasts d days, the weight change is equally likely to be any value between $-d$ and d pounds. Thus,

$$P_{W|D}(w|d) = \begin{cases} 1/(2d+1) & w = -d, -d+1, \ldots, d, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The joint PMF is simply

$$\begin{aligned} P_{D,W}(d,w) &= P_{W|D}(w|d)\, P_D(d) \\ &= \begin{cases} 1/(6d+3) & d = 2,3,4; w = -d, \ldots, d, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

Problem 7.4.6 Solution

$$f_{X,Y}(x,y) = \begin{cases} (x+y) & 0 \le x, y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The conditional PDF $f_{X|Y}(x|y)$ is defined for all y such that $0 \le y \le 1$. For $0 \le y \le 1$,

$$f_{X|Y}(x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{(x+y)}{\int_0^1 (x+y)\, dy} = \begin{cases} \frac{(x+y)}{x+1/2} & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) The conditional PDF $f_{Y|X}(y|x)$ is defined for all values of x in the interval $[0,1]$. For $0 \le x \le 1$,

$$f_{Y|X}(y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{(x+y)}{\int_0^1 (x+y)\, dx} = \begin{cases} \frac{(x+y)}{y+1/2} & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 7.4.7 Solution

We are told in the problem statement that if we know r, the number of feet a student sits from the blackboard, then we also know that that student's grade is a Gaussian random variable with mean $80 - r$ and standard deviation r. This is exactly

$$f_{X|R}(x|r) = \frac{1}{\sqrt{2\pi r^2}} e^{-(x-[80-r])^2/2r^2}. \tag{1}$$

Problem 7.4.8 Solution

(a) TRUE:

$$\begin{aligned}
F_{Y|Z=1}(y|Z=1) &= \mathrm{P}[Y \le y|Z=1] \\
&= \mathrm{P}[X \le y|Z=1] = \mathrm{P}[X \le y] = \Phi(y). && (1) \\
F_{Y|Z=-1}(y|Z=-1) &= \mathrm{P}[Y \le y|Z=-1] \\
&= \mathrm{P}[-X \le y] = \mathrm{P}[X \ge -y] = \Phi(y). && (2)
\end{aligned}$$

Taking derivatives, we have

$$f_{Y|Z}(y|1) = f_{Y|Z}(y|-1) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}. \tag{3}$$

(b) FALSE: Given $X = x$, $Y = Zx$ is equally likely to be x or $-x$. That is,

$$P_{Y|X}(y|x) = \mathrm{P}[Y = y|X = x] = \begin{cases} 1-p & y = -x, \\ p & y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Since the conditional PMF of Y given X depends on X, Y and X are dependent.

Problem 7.4.9 Solution

This problem is fairly easy when we use conditional PMF's. In particular, given that $N = n$ pizzas were sold before noon, each of those pizzas has mushrooms with probability 1/3. The conditional PMF of M given N is the binomial distribution

$$P_{M|N}(m|n) = \binom{n}{m} (1/3)^m (2/3)^{n-m}. \tag{1}$$

Since $P_{M|N}(m|n)$ depends on the event $N = n$, we see that M and N are dependent.

The other fact we know is that for each of the 100 pizzas sold, the pizza is sold before noon with probability 1/2. Hence, N has the binomial PMF

$$P_N(n) = \binom{100}{n} (1/2)^n (1/2)^{100-n}. \tag{2}$$

The joint PMF of N and M is for integers m, n,

$$P_{M,N}(m,n) = P_{M|N}(m|n)\, P_N(n) \\ = \binom{n}{m}\binom{100}{n}(1/3)^m(2/3)^{n-m}(1/2)^{100}. \qquad (3)$$

Problem 7.4.10 Solution

(a) Normally, checking independence requires the marginal PMFs. However, in this problem, the zeroes in the table of the joint PMF $P_{X,Y}(x,y)$ allows us to verify very quickly that X and Y are dependent. In particular, $P_X(-1) = 1/4$ and $P_Y(1) = 14/48$ but

$$P_{X,Y}(-1,1) = 0 \neq P_X(-1)\, P_Y(1)\,. \qquad (1)$$

(b) To fill in the tree diagram, we need the marginal PMF $P_X(x)$ and the conditional PMFs $P_{Y|X}(y|x)$. By summing the rows on the table for the joint PMF, we obtain

$P_{X,Y}(x,y)$	$y=-1$	$y=0$	$y=1$	$P_X(x)$
$x=-1$	3/16	1/16	0	1/4
$x=0$	1/6	1/6	1/6	1/2
$x=1$	0	1/8	1/8	1/4

(2)

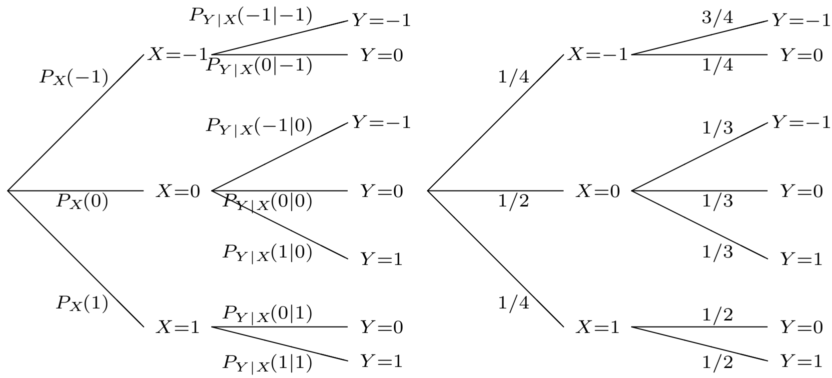
Now we use the conditional PMF $P_{Y|X}(y|x) = P_{X,Y}(x,y)/P_X(x)$ to write

$$P_{Y|X}(y|-1) = \begin{cases} 3/4 & y=-1, \\ 1/4 & y=0, \\ 0 & \text{otherwise}, \end{cases} \qquad (3)$$

$$P_{Y|X}(y|0) = \begin{cases} 1/3 & y=-1,0,1, \\ 0 & \text{otherwise}, \end{cases} \qquad (4)$$

$$P_{Y|X}(y|1) = \begin{cases} 1/2 & y=0,1, \\ 0 & \text{otherwise}. \end{cases} \qquad (5)$$

Now we can us these probabilities to label the tree. The generic solution and the specific solution with the exact values are



Problem 7.4.11 Solution

We can make a table of the possible outcomes and the corresponding values of W and Y

outcome	$P[\cdot]$	W	Y
hh	p^2	0	2
ht	$p(1-p)$	1	1
th	$p(1-p)$	-1	1
tt	$(1-p)^2$	0	0

(1)

In the following table, we write the joint PMF $P_{W,Y}(w,y)$ along with the marginal PMFs $P_Y(y)$ and $P_W(w)$.

$P_{W,Y}(w,y)$	$w=-1$	$w=0$	$w=1$	$P_Y(y)$
$y=0$	0	$(1-p)^2$	0	$(1-p)^2$
$y=1$	$p(1-p)$	0	$p(1-p)$	$2p(1-p)$
$y=2$	0	p^2	0	p^2
$P_W(w)$	$p(1-p)$	$1-2p+2p^2$	$p(1-p)$	

(2)

Using the definition $P_{W|Y}(w|y) = P_{W,Y}(w,y)/P_Y(y)$, we can find the conditional PMFs of W given Y:

$$P_{W|Y}(w|0) = \begin{cases} 1 & w = 0, \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

$$P_{W|Y}(w|1) = \begin{cases} 1/2 & w = -1, 1, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

$$P_{W|Y}(w|2) = \begin{cases} 1 & w = 0, \\ 0 & \text{otherwise}. \end{cases} \tag{5}$$

Similarly, the conditional PMFs of Y given W are

$$P_{Y|W}(y|-1) = \begin{cases} 1 & y = 1, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

$$P_{Y|W}(y|0) = \begin{cases} \frac{(1-p)^2}{1-2p+2p^2} & y = 0, \\ \frac{p^2}{1-2p+2p^2} & y = 2, \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

$$P_{Y|W}(y|1) = \begin{cases} 1 & y = 1, \\ 0 & \text{otherwise}. \end{cases} \tag{8}$$

Problem 7.4.12 Solution

To use the hint, we start by using Bayes' rule to write

$$\begin{aligned} \mathrm{P}\left[x_1 < X \le x_2 | y < Y \le y + \Delta\right] &= \frac{\mathrm{P}\left[x_1 < X \le x_2, y < Y \le y + \Delta\right]}{\mathrm{P}\left[y < Y \le y + \Delta\right]} \\ &= \frac{\displaystyle\int_{x_1}^{x_2} \left(\int_y^{y+\Delta} f_{X,Y}(x,u)\, du\right) dx}{\displaystyle\int_y^{y+\Delta} f_Y(v)\, dv}. \end{aligned} \tag{1}$$

We observe in (1) that in the inner integral in the numerator that $f_{X,Y}(x,u)$ is just a function of u for each x and that

$$\int_y^{y+\Delta} f_{X,Y}(x,u)\, du \to f_{X,Y}(x,y)\,\Delta \tag{2}$$

as $\Delta \to 0$. Similarly, in the denominator of (1),

$$\int_{y}^{y+\Delta} f_Y(v)\,dv \to f_Y(y)\,\Delta \tag{3}$$

as $\Delta \to 0$. Thus,

$$\begin{aligned}
\lim_{\Delta\to 0} \mathrm{P}\left[x_1 < X \le x_2 | y < Y \le y+\Delta\right] &= \lim_{\Delta\to 0} \frac{\int_{x_1}^{x_2} \left(f_{X,Y}(x,y)\,\Delta\right)\,dx}{f_Y(y)\,\Delta} \\
&= \frac{\int_{x_1}^{x_2} f_{X,Y}(x,y)\,dx}{f_Y(y)} \\
&= \int_{x_1}^{x_2} f_{X|Y}(x|y)\,dx.
\end{aligned} \tag{4}$$

Problem 7.4.13 Solution

The key to solving this problem is to find the joint PMF of M and N. Note that $N \ge M$. For $n > m$, the joint event $\{M = m, N = n\}$ has probability

$$\begin{aligned}
\mathrm{P}\left[M = m, N = n\right] &= \mathrm{P}[\overbrace{dd\cdots d}^{\substack{m-1\\ \text{calls}}} v\ \overbrace{dd\cdots d}^{\substack{n-m-1\\ \text{calls}}}\ v] \\
&= (1-p)^{m-1}p(1-p)^{n-m-1}p \\
&= (1-p)^{n-2}p^2.
\end{aligned} \tag{1}$$

A complete expression for the joint PMF of M and N is

$$P_{M,N}(m,n) = \begin{cases} (1-p)^{n-2}p^2 & m = 1,2,\ldots,n-1; \\ & n = m+1, m+2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The marginal PMF of N satisfies

$$P_N(n) = \sum_{m=1}^{n-1} (1-p)^{n-2}p^2 = (n-1)(1-p)^{n-2}p^2, \qquad n = 2,3,\ldots. \tag{3}$$

Similarly, for $m = 1, 2, \ldots$, the marginal PMF of M satisfies

$$P_M(m) = \sum_{n=m+1}^{\infty} (1-p)^{n-2} p^2$$
$$= p^2[(1-p)^{m-1} + (1-p)^m + \cdots]$$
$$= (1-p)^{m-1} p. \tag{4}$$

The complete expressions for the marginal PMF's are

$$P_M(m) = \begin{cases} (1-p)^{m-1} p & m = 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$P_N(n) = \begin{cases} (n-1)(1-p)^{n-2} p^2 & n = 2, 3, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Not surprisingly, if we view each voice call as a successful Bernoulli trial, M has a geometric PMF since it is the number of trials up to and including the first success. Also, N has a Pascal PMF since it is the number of trials required to see 2 successes. The conditional PMF's are now easy to find.

$$P_{N|M}(n|m) = \frac{P_{M,N}(m,n)}{P_M(m)} = \begin{cases} (1-p)^{n-m-1} p & n = m+1, m+2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The interpretation of the conditional PMF of N given M is that given $M = m$, $N = m + N'$ where N' has a geometric PMF with mean $1/p$. The conditional PMF of M given N is

$$P_{M|N}(m|n) = \frac{P_{M,N}(m,n)}{P_N(n)} = \begin{cases} 1/(n-1) & m = 1, \ldots, n-1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Given that call $N = n$ was the second voice call, the first voice call is equally likely to occur in any of the previous $n-1$ calls.

Problem 7.4.14 Solution

(a) The number of buses, N, must be greater than zero. Also, the number of minutes that pass cannot be less than the number of buses. Thus, $\mathrm{P}[N = n, T = t] > 0$ for integers n, t satisfying $1 \le n \le t$.

(b) First, we find the joint PMF of N and T by carefully considering the possible sample paths. In particular, $P_{N,T}(n,t) = \mathrm{P}[ABC] = \mathrm{P}[A]\,\mathrm{P}[B]\,\mathrm{P}[C]$ where the events A, B and C are

$$A = \{n-1 \text{ buses arrive in the first } t-1 \text{ minutes}\}, \tag{1}$$
$$B = \{\text{none of the first } n-1 \text{ buses are boarded}\}, \tag{2}$$
$$C = \{\text{at time } t \text{ a bus arrives and is boarded}\}. \tag{3}$$

These events are independent since each trial to board a bus is independent of when the buses arrive. These events have probabilities

$$\mathrm{P}[A] = \binom{t-1}{n-1} p^{n-1}(1-p)^{t-1-(n-1)}, \tag{4}$$
$$\mathrm{P}[B] = (1-q)^{n-1}, \tag{5}$$
$$\mathrm{P}[C] = pq. \tag{6}$$

Consequently, the joint PMF of N and T is

$$P_{N,T}(n,t) = \begin{cases} \binom{t-1}{n-1} p^{n-1}(1-p)^{t-n}(1-q)^{n-1}pq & n \ge 1, t \ge n \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

(c) It is possible to find the marginal PMF's by summing the joint PMF. However, it is much easier to obtain the marginal PMFs by consideration of the experiment. Specifically, when a bus arrives, it is boarded with probability q. Moreover, the experiment ends when a bus is boarded. By viewing whether each arriving bus is boarded as an independent trial, N is the number of trials until the first success. Thus, N has the geometric PMF

$$P_N(n) = \begin{cases} (1-q)^{n-1}q & n = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

To find the PMF of T, suppose we regard each minute as an independent trial in which a success occurs if a bus arrives and that bus is boarded. In this case, the success probability is pq and T is the number of minutes up to and including the first success. The PMF of T is also geometric.

$$P_T(t) = \begin{cases} (1-pq)^{t-1}pq & t = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

(d) Once we have the marginal PMFs, the conditional PMFs are easy to find.

$$\begin{aligned} P_{N|T}(n|t) &= \frac{P_{N,T}(n,t)}{P_T(t)} \\ &= \binom{t-1}{n-1} \left(\frac{p(1-q)}{1-pq}\right)^{n-1} \left(\frac{1-p}{1-pq}\right)^{t-1-(n-1)}. \end{aligned} \tag{10}$$

That is, given you depart at time $T = t$, the number of buses that arrive during minutes $1, \ldots, t-1$ has a binomial PMF since in each minute a bus arrives with probability p. Similarly, the conditional PMF of T given N is

$$\begin{aligned} P_{T|N}(t|n) &= \frac{P_{N,T}(n,t)}{P_N(n)} \\ &= \binom{t-1}{n-1} p^n (1-p)^{t-n}. \end{aligned} \tag{11}$$

This result can be explained. Given that you board bus $N = n$, the time T when you leave is the time for n buses to arrive. If we view each bus arrival as a success of an independent trial, the time for n buses to arrive has the above Pascal PMF.

Problem 7.4.15 Solution

If you construct a tree describing what type of packet (if any) that arrived in any 1 millisecond period, it will be apparent that an email packet arrives with probability $\alpha = pqr$ or no email packet arrives with probability $1 - \alpha$. That is, whether an email packet arrives each millisecond is a Bernoulli trial with success probability α. Thus, the time required for the first success has the geometric PMF

$$P_T(t) = \begin{cases} (1-\alpha)^{t-1}\alpha & t = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Note that N is the number of trials required to observe 100 successes. Moreover, the number of trials needed to observe 100 successes is $N = T + N'$ where N' is the number of trials needed to observe successes 2 through 100. Since N' is just the number of trials needed to observe 99 successes, it has the Pascal $(k = 99, p)$ PMF

$$P_{N'}(n) = \binom{n-1}{98}\alpha^{99}(1-\alpha)^{n-99}. \tag{2}$$

Since the trials needed to generate successes 2 though 100 are independent of the trials that yield the first success, N' and T are independent. Hence

$$P_{N|T}(n|t) = P_{N'|T}(n-t|t) = P_{N'}(n-t)\,. \tag{3}$$

Applying the PMF of N' found above, we have

$$P_{N|T}(n|t) = \binom{n-t-1}{98}\alpha^{99}(1-\alpha)^{n-t-99}. \tag{4}$$

Finally the joint PMF of N and T is

$$\begin{aligned} P_{N,T}(n,t) &= P_{N|T}(n|t)\,P_T(t) \\ &= \binom{n-t-1}{98}\alpha^{100}(1-\alpha)^{n-100}. \end{aligned} \tag{5}$$

This solution can also be found a consideration of the sample sequence of Bernoulli trials in which we either observe or do not observe an email packet.

To find the conditional PMF $P_{T|N}(t|n)$, we first must recognize that N is simply the number of trials needed to observe 100 successes and thus has the Pascal PMF

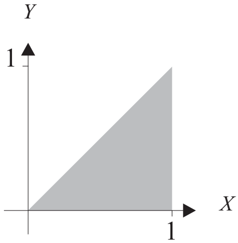
$$P_N(n) = \binom{n-1}{99}\alpha^{100}(1-\alpha)^{n-100}. \tag{6}$$

Hence for any integer $n \geq 100$, the conditional PMF is

$$P_{T|N}(t|n) = \frac{P_{N,T}(n,t)}{P_N(n)}. = \frac{\binom{n-t-1}{98}}{\binom{n-1}{99}}. \tag{7}$$

Problem 7.5.1 Solution

Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_y^1 2\, dx = 2(1-y). \tag{2}$$

Also, for $y < 0$ or $y > 1$, $f_Y(y) = 0$. The complete expression for the marginal PDF is

$$f_Y(y) = \begin{cases} 2(1-y) & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

By Theorem 7.10, the conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{1-y} & y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

That is, since $Y \le X \le 1$, X is uniform over $[y, 1]$ when $Y = y$. The conditional expectation of X given $Y = y$ can be calculated as

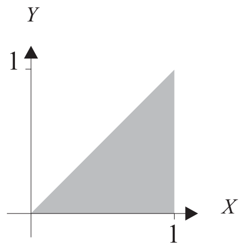
$$\mathrm{E}\left[X|Y=y\right] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\ dx \tag{5}$$

$$= \int_y^1 \frac{x}{1-y}\, dx = \left.\frac{x^2}{2(1-y)}\right|_y^1 = \frac{1+y}{2}. \tag{6}$$

In fact, since we know that the conditional PDF of X is uniform over $[y, 1]$ when $Y = y$, it wasn't really necessary to perform the calculation.

Problem 7.5.2 Solution

Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For $0 \le x \le 1$, the marginal PDF for X satisfies

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_0^x 2\, dy = 2x. \tag{2}$$

Note that $f_X(x) = 0$ for $x < 0$ or $x > 1$. Hence the complete expression for the marginal PDF of X is

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The conditional PDF of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} 1/x & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Given $X = x$, Y has a uniform PDF over $[0,x]$ and thus has conditional expected value $E[Y|X=x] = x/2$. Another way to obtain this result is to calculate $\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy$.

Problem 7.5.3 Solution

(a) First we observe that A takes on the values $S_A = \{-1, 1\}$ while B takes on values from $S_B = \{0, 1\}$. To construct a table describing $P_{A,B}(a,b)$ we build a table for all possible values of pairs (A,B). The general form of the entries is

$$\begin{array}{c|cc} P_{A,B}(a,b) & b=0 & b=1 \\ \hline a=-1 & P_{B|A}(0|-1)\,P_A(-1) & P_{B|A}(1|-1)\,P_A(-1) \\ a=1 & P_{B|A}(0|1)\,P_A(1) & P_{B|A}(1|1)\,P_A(1) \end{array} \tag{1}$$

Now we fill in the entries using the conditional PMFs $P_{B|A}(b|a)$ and the marginal PMF $P_A(a)$. This yields

$$\begin{array}{c|cc} P_{A,B}(a,b) & b=0 & b=1 \\ \hline a=-1 & (1/3)(1/3) & (2/3)(1/3) \\ a=1 & (1/2)(2/3) & (1/2)(2/3) \end{array}, \tag{2}$$

which simplifies to

$$\begin{array}{c|cc} P_{A,B}(a,b) & b=0 & b=1 \\ \hline a=-1 & 1/9 & 2/9 \\ a=1 & 1/3 & 1/3 \end{array}. \tag{3}$$

(b) Since $P_A(1) = P_{A,B}(1,0) + P_{A,B}(1,1) = 2/3$,

$$P_{B|A}(b|1) = \frac{P_{A,B}(1,b)}{P_A(1)} = \begin{cases} 1/2 & b=0,1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

If $A=1$, the conditional expectation of B is

$$\mathrm{E}[B|A=1] = \sum_{b=0}^{1} b P_{B|A}(b|1) = P_{B|A}(1|1) = 1/2. \tag{5}$$

(c) Before finding the conditional PMF $P_{A|B}(a|1)$, we first sum the columns of the joint PMF table to find

$$P_B(b) = \begin{cases} 4/9 & b=0, \\ 5/9 & b=1, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The conditional PMF of A given $B = 1$ is

$$P_{A|B}(a|1) = \frac{P_{A,B}(a,1)}{P_B(1)} = \begin{cases} 2/5 & a = -1, \\ 3/5 & a = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

(d) Now that we have the conditional PMF $P_{A|B}(a|1)$, calculating conditional expectations is easy.

$$\mathrm{E}\left[A|B=1\right] = \sum_{a=-1,1} aP_{A|B}(a|1) = -1(2/5) + (3/5) = 1/5, \tag{8}$$

$$\mathrm{E}\left[A^2|B=1\right] = \sum_{a=-1,1} a^2 P_{A|B}(a|1) = 2/5 + 3/5 = 1. \tag{9}$$

The conditional variance is then

$$\begin{aligned} \mathrm{Var}[A|B=1] &= \mathrm{E}\left[A^2|B=1\right] - (\mathrm{E}\left[A|B=1\right])^2 \\ &= 1 - (1/5)^2 = 24/25. \end{aligned} \tag{10}$$

(e) To calculate the covariance, we need

$$\mathrm{E}\left[A\right] = \sum_{a=-1,1} aP_A(a) = -1(1/3) + 1(2/3) = 1/3, \tag{11}$$

$$\mathrm{E}\left[B\right] = \sum_{b=0}^{1} bP_B(b) = 0(4/9) + 1(5/9) = 5/9, \tag{12}$$

$$\begin{aligned} \mathrm{E}\left[AB\right] &= \sum_{a=-1,1} \sum_{b=0}^{1} abP_{A,B}(a,b) \\ &= -1(0)(1/9) + -1(1)(2/9) + 1(0)(1/3) + 1(1)(1/3) \\ &= 1/9. \end{aligned} \tag{13}$$

The covariance is just

$$\begin{aligned} \mathrm{Cov}\left[A,B\right] &= \mathrm{E}\left[AB\right] - \mathrm{E}\left[A\right]\mathrm{E}\left[B\right] \\ &= 1/9 - (1/3)(5/9) = -2/27. \end{aligned} \tag{14}$$

Problem 7.5.4 Solution

First we need to find the conditional expectations

$$E[B|A=-1] = \sum_{b=0}^{1} b P_{B|A}(b|-1) = 0(1/3) + 1(2/3) = 2/3, \tag{1}$$

$$E[B|A=1] = \sum_{b=0}^{1} b P_{B|A}(b|1) = 0(1/2) + 1(1/2) = 1/2. \tag{2}$$

Keep in mind that $E[B|A]$ is a random variable that is a function of A. that is we can write

$$E[B|A] = g(A) = \begin{cases} 2/3 & A=-1, \\ 1/2 & A=1. \end{cases} \tag{3}$$

We see that the range of U is $S_U = \{1/2, 2/3\}$. In particular,

$$P_U(1/2) = P_A(1) = 2/3, \qquad P_U(2/3) = P_A(-1) = 1/3. \tag{4}$$

The complete PMF of U is

$$P_U(u) = \begin{cases} 2/3 & u = 1/2, \\ 1/3 & u = 2/3, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Note that

$$\begin{aligned} E[E[B|A]] = E[U] &= \sum_u u P_U(u) \\ &= (1/2)(2/3) + (2/3)(1/3) = 5/9. \end{aligned} \tag{6}$$

You can check that $E[U] = E[B]$.

Problem 7.5.5 Solution

Random variables N and K have the joint PMF

$$P_{N,K}(n,k) = \begin{cases} \frac{100^n e^{-100}}{(n+1)!} & k = 0,1,\ldots,n; \\ & n = 0,1,\ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) We can find the marginal PMF for N by summing over all possible K. For $n \geq 0$,

$$P_N(n) = \sum_{k=0}^{n} \frac{100^n e^{-100}}{(n+1)!} = \frac{100^n e^{-100}}{n!}. \tag{2}$$

We see that N has a Poisson PMF with expected value 100. For $n \geq 0$, the conditional PMF of K given $N = n$ is

$$P_{K|N}(k|n) = \frac{P_{N,K}(n,k)}{P_N(n)} = \begin{cases} 1/(n+1) & k = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

That is, given $N = n$, K has a discrete uniform PMF over $\{0, 1, \ldots, n\}$. Thus,

$$\mathrm{E}[K|N = n] = \sum_{k=0}^{n} k/(n+1) = n/2. \tag{4}$$

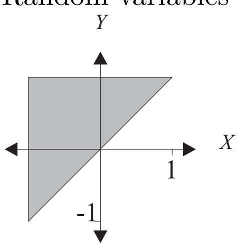
(b) Since $\mathrm{E}[K|N = n] = n/2$, we can conclude that $\mathrm{E}[K|N] = N/2$. Thus, by Theorem 7.13,

$$\mathrm{E}[K] = \mathrm{E}[\mathrm{E}[K|N]] = \mathrm{E}[N/2] = 50, \tag{5}$$

since N is Poisson with $\mathrm{E}[N] = 100$.

Problem 7.5.6 Solution

Random variables X and Y have joint PDF



$$f_{X,Y}(x,y) = \begin{cases} 1/2 & -1 \leq x \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) For $-1 \leq y \leq 1$, the marginal PDF of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \frac{1}{2}\int_{-1}^{y} dx = (y+1)/2. \tag{2}$$

The complete expression for the marginal PDF of Y is

$$f_Y(y) = \begin{cases} (y+1)/2 & -1 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(b) The conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{1+y} & -1 \leq x \leq y, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(c) Given $Y = y$, the conditional PDF of X is uniform over $[-1, y]$. Hence the conditional expected value is $\mathrm{E}[X|Y = y] = (y-1)/2$.

Problem 7.5.7 Solution

We are given that the joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 1/(\pi r^2) & 0 \leq x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The marginal PDF of X is

$$f_X(x) = 2\int_0^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2}\, dy = \begin{cases} \frac{2\sqrt{r^2-x^2}}{\pi r^2} & -r \leq x \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} 1/(2\sqrt{r^2 - x^2}) & y^2 \leq r^2 - x^2, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(b) Given $X = x$, we observe that over the interval $[-\sqrt{r^2 - x^2}, \sqrt{r^2 - x^2}]$, Y has a uniform PDF. Since the conditional PDF $f_{Y|X}(y|x)$ is symmetric about $y = 0$,

$$\mathrm{E}\left[Y|X = x\right] = 0. \tag{4}$$

Problem 7.5.8 Solution

We buy the option at the beginning of the month if $D \leq d^*$ and we sell at the end of the month for price V. To proceed, we require a few additional assumptions. We will suppose that D and V are independent and that D has PDF $f_D(d)$. This assumption is not unreasonable because it assumes that the expected return $\mathrm{E}[V]$ is embedded in the expected option price $\mathrm{E}[D]$. Using τ to denote the threshold d^*, at the end of the thirty days, the return is

$$R = \begin{cases} V - D & D \leq \tau, \\ 0 & D > \tau. \end{cases}$$

Thus the conditional expected return is

$$\mathrm{E}[R|D = x] = \begin{cases} \mathrm{E}[V] - x & x \leq \tau, \\ 0 & x > \tau. \end{cases}$$

It follows that the expected return is

$$\begin{aligned} \mathrm{E}[R] &= \int_{-\infty}^{\infty} \mathrm{E}[R|D = x] \, f_D(x) \, dx \\ &= \int_{-\infty}^{\tau} (\mathrm{E}[V] - x) \, f_D(x) \, dx. \end{aligned} \tag{1}$$

To find the value of the threshold τ that maximizes $\mathrm{E}[R]$, we calculate

$$\frac{d\mathrm{E}[R]}{d\tau} = (\mathrm{E}[V] - \tau) f_D(\tau). \tag{2}$$

We see that $d\,\mathrm{E}[R]/d\tau \geq 0$ for all $\tau \leq \mathrm{E}[V]$ and that the derivative is zero at $\tau = \mathrm{E}[V]$. Hence $\mathrm{E}[R]$ is maximized at $\tau = d^* = \mathrm{E}[V]$. In fact, this answer can be found by intuition. When the option has price $D < \mathrm{E}[V]$, your choice is either to reject the option and earn 0 reward or to earn a reward R with $\mathrm{E}[R] > 0$. On an expected value basis, it's always better to buy the call whenever $D < \mathrm{E}[V]$. Hence we should set the threshold at $d^* = \mathrm{E}[V]$.

We note, however, that implementation of this strategy requires us to calculate $\mathrm{E}[V]$. Let $Y = X - k$. At time t, Y is a Gaussian $(0, \sqrt{t})$ random variable and since $V = Y^+$,

$$\mathrm{E}[V] = \mathrm{E}\left[Y^+\right] = \int_0^{\infty} y f_Y(y) \, dy = \frac{1}{\sqrt{2\pi t}} \int_0^{\infty} y e^{-y^2/2t} \, dy. \tag{3}$$

With the variable substitution $w = y^2/2t$, we have $dw = (y/t)\, dy$ and

$$\mathrm{E}\left[V\right] = \frac{t}{\sqrt{2\pi t}} \int_0^\infty e^{-w}\, dw = \sqrt{\frac{t}{2\pi}}. \tag{4}$$

Problem 7.5.9 Solution

Since 50 cents of each dollar ticket is added to the jackpot,

$$J_{i-1} = J_i + \frac{N_i}{2}. \tag{1}$$

Given $J_i = j$, N_i has a Poisson distribution with mean j. It follows that $\mathrm{E}[N_i|J_i = j] = j$ and that $\mathrm{Var}[N_i|J_i = j] = j$. This implies

$$\begin{aligned}\mathrm{E}\left[N_i^2|J_i = j\right] &= \mathrm{Var}[N_i|J_i = j] + \left(\mathrm{E}\left[N_i|J_i = j\right]\right)^2 \\ &= j + j^2. \end{aligned} \tag{2}$$

In terms of the conditional expectations given J_i, these facts can be written as

$$\mathrm{E}\left[N_i|J_i\right] = J_i \qquad \mathrm{E}\left[N_i^2|J_i\right] = J_i + J_i^2. \tag{3}$$

This permits us to evaluate the moments of J_{i-1} in terms of the moments of J_i. Specifically,

$$\mathrm{E}\left[J_{i-1}|J_i\right] = \mathrm{E}\left[J_i|J_i\right] + \frac{1}{2}\mathrm{E}\left[N_i|J_i\right] = J_i + \frac{J_i}{2} = \frac{3J_i}{2}. \tag{4}$$

Using the iterated expectation, this implies

$$\mathrm{E}\left[J_{i-1}\right] = \mathrm{E}\left[\mathrm{E}\left[J_{i-1}|J_i\right]\right] = \frac{3}{2}\mathrm{E}\left[J_i\right]. \tag{5}$$

We can use this the calculate $\mathrm{E}[J_i]$ for all i. Since the jackpot starts at 1 million dollars, $J_6 = 10^6$ and $\mathrm{E}[J_6] = 10^6$. This implies

$$\mathrm{E}\left[J_i\right] = (3/2)^{6-i}10^6 \tag{6}$$

Now we will find the second moment $\mathrm{E}[J_i^2]$. Since

$$J_{i-1}^2 = J_i^2 + N_iJ_i + N_i^2/4, \tag{7}$$

we have

$$\begin{aligned}
\mathrm{E}\left[J_{i-1}^2|J_i\right] &= \mathrm{E}\left[J_i^2|J_i\right] + \mathrm{E}\left[N_i J_i|J_i\right] + \mathrm{E}\left[N_i^2|J_i\right]/4 \\
&= J_i^2 + J_i\,\mathrm{E}\left[N_i|J_i\right] + (J_i + J_i^2)/4 \\
&= (3/2)^2 J_i^2 + J_i/4. \qquad (8)
\end{aligned}$$

By taking the expectation over J_i we have

$$\mathrm{E}\left[J_{i-1}^2\right] = (3/2)^2\,\mathrm{E}\left[J_i^2\right] + \mathrm{E}\left[J_i\right]/4 \qquad (9)$$

This recursion allows us to calculate $\mathrm{E}[J_i^2]$ for $i = 6, 5, \ldots, 0$. Since $J_6 = 10^6$, $\mathrm{E}[J_6^2] = 10^{12}$. From the recursion, we obtain

$$\begin{aligned}
\mathrm{E}\left[J_5^2\right] &= (3/2)^2\,\mathrm{E}\left[J_6^2\right] + \mathrm{E}\left[J_6\right]/4 \\
&= (3/2)^2 10^{12} + \frac{1}{4}10^6, \qquad (10) \\
\mathrm{E}\left[J_4^2\right] &= (3/2)^2\,\mathrm{E}\left[J_5^2\right] + \mathrm{E}\left[J_5\right]/4 \\
&= (3/2)^4 10^{12} + \frac{1}{4}\left[(3/2)^2 + (3/2)\right]10^6, \qquad (11) \\
\mathrm{E}\left[J_3^2\right] &= (3/2)^2\,\mathrm{E}\left[J_4^2\right] + \mathrm{E}\left[J_4\right]/4 \\
&= (3/2)^6 10^{12} + \frac{1}{4}\left[(3/2)^4 + (3/2)^3 + (3/2)^2\right]10^6. \qquad (12)
\end{aligned}$$

The same recursion will also allow us to show that

$$\begin{aligned}
\mathrm{E}\left[J_2^2\right] &= (3/2)^8 10^{12} + \frac{1}{4}\left[(3/2)^6 + (3/2)^5 + (3/2)^4 + (3/2)^3\right]10^6, \qquad (13) \\
\mathrm{E}\left[J_1^2\right] &= (3/2)^{10} 10^{12} \\
&\quad + \frac{1}{4}\left[(3/2)^8 + (3/2)^7 + (3/2)^6 + (3/2)^5 + (3/2)^4\right]10^6, \qquad (14) \\
\mathrm{E}\left[J_0^2\right] &= (3/2)^{12} 10^{12} + \frac{1}{4}\left[(3/2)^{10} + (3/2)^9 + \cdots + (3/2)^5\right]10^6. \qquad (15)
\end{aligned}$$

Finally, day 0 is the same as any other day in that $J = J_0 + N_0/2$ where N_0 is a Poisson random variable with mean J_0. By the same argument that we used to develop recursions for $\mathrm{E}[J_i]$ and $\mathrm{E}[J_i^2]$, we can show

$$\mathrm{E}\left[J\right] = (3/2)\,\mathrm{E}\left[J_0\right] = (3/2)^7 10^6 \approx 17 \times 10^6. \qquad (16)$$

and

$$
\begin{aligned}
\mathrm{E}\left[J^2\right] &= (3/2)^2\,\mathrm{E}\left[J_0^2\right] + \mathrm{E}\left[J_0\right]/4 \\
&= (3/2)^{14}10^{12} + \frac{1}{4}\left[(3/2)^{12} + (3/2)^{11} + \cdots + (3/2)^6\right]10^6 \\
&= (3/2)^{14}10^{12} + \frac{10^6}{2}(3/2)^6[(3/2)^7 - 1]. \qquad (17)
\end{aligned}
$$

Finally, the variance of J is

$$
\mathrm{Var}[J] = \mathrm{E}\left[J^2\right] - (\mathrm{E}\left[J\right])^2 = \frac{10^6}{2}(3/2)^6[(3/2)^7 - 1]. \qquad (18)
$$

Since the variance is hard to interpret, we note that the standard deviation of J is $\sigma_J \approx 9572$. Although the expected jackpot grows rapidly, the standard deviation of the jackpot is fairly small.

Problem 7.6.1 Solution

This problem is actually easy and short if you think carefully.

(a) Since Z is Gaussian $(0, 2)$ and Z and X are independent,

$$
f_{Z|X}(z|x) = f_Z(z) = \frac{1}{\sqrt{8\pi}}e^{-z^2/8}. \qquad (1)
$$

(b) Using the hint, we observe that if $X = 2$, then $Y = 2 + Z$. Furthermore, independence of X and Z implies that given $X = 2$, Z still has the Gaussian PDF $f_Z(z)$. Thus, given $X = x = 2$, $Y = 2 + Z$ is conditionally Gaussian with

$$
\begin{aligned}
\mathrm{E}\left[Y|X = 2\right] &= 2 + \mathrm{E}\left[Z|X = 2\right] = 2, \qquad (2) \\
\mathrm{Var}[Y|X = 2] &= \mathrm{Var}[2 + Z|X = 2] = \mathrm{Var}[Z|X = 2] = 2. \qquad (3)
\end{aligned}
$$

The conditional PDF of Y is

$$
f_{Y|X}(y|2) = \frac{1}{\sqrt{8\pi}}e^{-(y-2)^2/8}. \qquad (4)
$$

Problem 7.6.2 Solution

From the problem statement, we learn that

$$\mu_X = \mu_Y = 0, \qquad \sigma_X^2 = \sigma_Y^2 = 1. \tag{1}$$

From Theorem 7.16, the conditional expectation of Y given X is

$$\mathrm{E}\left[Y|X\right] = \tilde{\mu}_Y(X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) = \rho X. \tag{2}$$

In the problem statement, we learn that $\mathrm{E}[Y|X] = X/2$. Hence $\rho = 1/2$. From Definition 5.10, the joint PDF is

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{3\pi^2}} e^{-2(x^2 - xy + y^2)/3}. \tag{3}$$

Problem 7.6.3 Solution

We need to calculate

$$\mathrm{Cov}\left[\hat{X}, \hat{Y}\right] = \mathrm{E}\left[\hat{X}\hat{Y}\right] - \mathrm{E}\left[\hat{X}\right]\mathrm{E}\left[\hat{Y}\right]. \tag{1}$$

To do so, we need to condition on whether a cyclist is male (event M) or female (event F):

$$\begin{aligned}
\mathrm{E}\left[\hat{X}\right] &= p\,\mathrm{E}\left[\hat{X}|M\right] + (1-p)\,\mathrm{E}\left[\hat{X}|F\right] \\
&= p\,\mathrm{E}\left[X\right] + (1-p)\,\mathrm{E}\left[X'\right] = 0.8(20) + (0.2)(15) = 16, && (2)\\
\mathrm{E}\left[\hat{Y}\right] &= p\,\mathrm{E}\left[Y|M\right] + (1-p)\,\mathrm{E}\left[Y|F\right] \\
&= p\,\mathrm{E}\left[Y\right] + (1-p)\,\mathrm{E}\left[Y'\right] = 0.8(75) + (0.2)(50) = 70. && (3)
\end{aligned}$$

Similarly, for the correlation term,

$$\begin{aligned}
\mathrm{E}\left[\hat{X}\hat{Y}\right] &= p\,\mathrm{E}\left[\hat{X}\hat{Y}|M\right] + (1-p)\,\mathrm{E}\left[\hat{X}\hat{Y}|F\right] \\
&= 0.8\,\mathrm{E}\left[XY\right] + 0.2\,\mathrm{E}\left[X'Y'\right]. && (4)
\end{aligned}$$

However, each of these terms needs some additional calculation:

$$\begin{aligned}
\mathrm{E}[XY] &= \mathrm{Cov}[X,Y] + \mathrm{E}[X]\,\mathrm{E}[Y] \\
&= \rho_{XY}\sigma_X\sigma_Y + \mathrm{E}[X]\,\mathrm{E}[Y] \\
&= -0.6(10) + (20)(75) = 1494. \qquad (5)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[X'Y'\right] &= \mathrm{Cov}\left[X',Y'\right] + \mathrm{E}\left[X'\right]\mathrm{E}\left[Y'\right] \\
&= \rho_{X'Y'}\sigma_{X'}\sigma_{Y'} + \mathrm{E}\left[X'\right]\mathrm{E}\left[Y'\right] \\
&= -0.6(10) + (15)(50) = 744. \qquad (6)
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{E}\left[\hat{X}\hat{Y}\right] &= 0.8\,\mathrm{E}[XY] + 0.2\,\mathrm{E}\left[X'Y'\right] \\
&= 0.8(1494) + 0.2(744) = 1344. \qquad (7)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}[X,Y] &= \mathrm{E}\left[\hat{X}\hat{Y}\right] - \mathrm{E}\left[\hat{X}\right]\mathrm{E}\left[\hat{Y}\right] \\
&= 1344 - (70)(19) = 14. \qquad (8)
\end{aligned}$$

Thus we see that the covariance of $\hat{X}$ and $\hat{Y}$ is positive. It follows that $\rho_{\hat{X}\hat{Y}} > 0$. Hence speed $\hat{X}$ and weight $\hat{Y}$ are positively correlated when we choose a cyclist randomly among men and women even though they are negatively correlated for women and negatively correlated for men. The reason for this is that men are heavier but they also ride faster than women. When we mix the populations, a fast rider is likely to be a male rider who is likely to be a relatively heavy rider (compared to a woman).

Problem 7.6.4 Solution

In this problem, X_1 and X_2 are jointly Gaussian random variables with $\mathrm{E}[X_i] = \mu_i$, $\mathrm{Var}[X_i] = \sigma_i^2$, and correlation coefficient $\rho_{12} = \rho$. The goal is to show that $Y = X_1X_2$ has variance

$$\mathrm{Var}[Y] = (1+\rho^2)\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + 2\rho\mu_1\mu_2\sigma_1\sigma_2. \qquad (1)$$

Since $\text{Var}[Y] = \text{E}[Y^2] - (\text{E}[Y])^2$, we will find the moments of Y. The first moment is

$$\text{E}[Y] = \text{E}[X_1 X_2] = \text{Cov}[X_1, X_2] + \text{E}[X_1]\,\text{E}[X_2] = \rho\sigma_1\sigma_2 + \mu_1\mu_2. \tag{2}$$

For the second moment of Y, we follow the problem hint and use the iterated expectation

$$\text{E}\left[Y^2\right] = \text{E}\left[X_1^2 X_2^2\right] = \text{E}\left[\text{E}\left[X_1^2 X_2^2 | X_2\right]\right] = \text{E}\left[X_2^2\,\text{E}\left[X_1^2 | X_2\right]\right]. \tag{3}$$

Given $X_2 = x_2$, we observe from Theorem 7.16 that X_1 is is Gaussian with

$$\text{E}[X_1 | X_2 = x_2] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \qquad \text{Var}[X_1 | X_2 = x_2] = \sigma_1^2(1 - \rho^2). \tag{4}$$

Thus, the conditional second moment of X_1 is

$$\begin{aligned}
\text{E}\left[X_1^2 | X_2\right] &= (\text{E}[X_1 | X_2])^2 + \text{Var}[X_1 | X_2] \\
&= \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)\right)^2 + \sigma_1^2(1 - \rho^2), \qquad (5)\\
&= [\mu_1^2 + \sigma_1^2(1 - \rho^2)] + 2\rho\mu_1\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2) + \rho^2\frac{\sigma_1^2}{\sigma_2^2}(X_2 - \mu_2)^2. \qquad (6)
\end{aligned}$$

It follows that

$$\begin{aligned}
&\text{E}\left[X_1^2 X_2^2\right] \\
&= \text{E}\left[X_2^2\,\text{E}\left[X_1^2 | X_2^2\right]\right] \\
&= \text{E}\left[[\mu_1^2 + \sigma_1^2(1 - \rho^2)]X_2^2 + 2\rho\mu_1\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)X_2^2 + \rho^2\frac{\sigma_1^2}{\sigma_2^2}(X_2 - \mu_2)^2 X_2^2\right]. \qquad (7)
\end{aligned}$$

Since $\text{E}[X_2^2] = \sigma_2^2 + \mu_2^2$,

$$\begin{aligned}
\text{E}\left[X_1^2 X_2^2\right] &= \left(\mu_1^2 + \sigma_1^2(1 - \rho^2)\right)(\sigma_2^2 + \mu_2^2) \\
&\quad + 2\rho\mu_1\frac{\sigma_1}{\sigma_2}\,\text{E}\left[(X_2 - \mu_2)X_2^2\right] + \rho^2\frac{\sigma_1^2}{\sigma_2^2}\,\text{E}\left[(X_2 - \mu_2)^2 X_2^2\right]. \qquad (8)
\end{aligned}$$

We observe that

$$
\begin{aligned}
&\mathrm{E}\left[(X_2-\mu_2)X_2^2\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)(X_2-\mu_2+\mu_2)^2\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)\left((X_2-\mu_2)^2+2\mu_2(X_2-\mu_2)+\mu_2^2\right)\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)^3\right]+2\mu_2\,\mathrm{E}\left[(X_2-\mu_2)^2\right]+\mu_2\,\mathrm{E}\left[(X_2-\mu_2)\right].
\end{aligned} \tag{9}
$$

We recall that $\mathrm{E}[X_2-\mu_2]=0$ and that $\mathrm{E}[(X_2-\mu_2)^2]=\sigma_2^2$. We now look ahead to Problem 9.2.4 to learn that

$$
\mathrm{E}\left[(X_2-\mu_2)^3\right]=0, \qquad \mathrm{E}\left[(X_2-\mu_2)^4\right]=3\sigma_2^4. \tag{10}
$$

This implies

$$
\mathrm{E}\left[(X_2-\mu_2)X_2^2\right]=2\mu_2\sigma_2^2. \tag{11}
$$

Following this same approach, we write

$$
\begin{aligned}
&\mathrm{E}\left[(X_2-\mu_2)^2X_2^2\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)^2(X_2-\mu_2+\mu_2)^2\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)^2\left((X_2-\mu_2)^2+2\mu_2(X_2-\mu_2)+\mu_2^2\right)\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)^2\left((X_2-\mu_2)^2+2\mu_2(X_2-\mu_2)+\mu_2^2\right)\right] \\
&\qquad = \mathrm{E}\left[(X_2-\mu_2)^4\right]+2\mu_2\,\mathrm{E}\left[X_2-\mu_2)^3\right]+\mu_2^2\,\mathrm{E}\left[(X_2-\mu_2)^2\right].
\end{aligned} \tag{12}
$$

It follows that

$$
\mathrm{E}\left[(X_2-\mu_2)^2X_2^2\right]=3\sigma_2^4+\mu_2^2\sigma_2^2. \tag{13}
$$

Combining the above results, we can conclude that

$$
\begin{aligned}
\mathrm{E}\left[X_1^2X_2^2\right] &= \left(\mu_1^2+\sigma_1^2(1-\rho^2)\right)(\sigma_2^2+\mu_2^2) \\
&\qquad + 2\rho\mu_1\frac{\sigma_1}{\sigma_2}(2\mu_2\sigma_2^2)+\rho^2\frac{\sigma_1^2}{\sigma_2^2}(3\sigma_2^4+\mu_2^2\sigma_2^2) \\
&= (1+2\rho^2)\sigma_1^2\sigma_2^2+\mu_2^2\sigma_1^2+\mu_1^2\sigma_2^2+\mu_1^2\mu_2^2+4\rho\mu_1\mu_2\sigma_1\sigma_2.
\end{aligned} \tag{14}
$$

Finally, combining Equations (2) and (14) yields

$$
\begin{aligned}
\mathrm{Var}[Y] &= \mathrm{E}\left[X_1^2X_2^2\right]-\left(\mathrm{E}\left[X_1X_2\right]\right)^2 \\
&= (1+\rho^2)\sigma_1^2\sigma_2^2+\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2+2\rho\mu_1\mu_2\sigma_1\sigma_2.
\end{aligned} \tag{15}
$$

Problem 7.6.5 Solution

The key to this problem is to see that the integrals in the given proof of Theorem 5.19 are actually iterated expectation. We start with the definition

$$\rho_{X,Y} = \frac{\mathrm{E}\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X\sigma_Y}. \tag{1}$$

To evaluate this expected value, we use the method of iterated expectation in Theorem 7.14 to write

$$\begin{aligned}\rho_{X,Y} &= \frac{\mathrm{E}\left[\mathrm{E}\left[(X-\mu_X)(Y-\mu_Y)|Y\right]\right]}{\sigma_X\sigma_Y}\\ &= \frac{\mathrm{E}\left[(Y-\mu_Y)\,\mathrm{E}\left[(X-\mu_X)|Y\right]\right]}{\sigma_X\sigma_Y}.\end{aligned} \tag{2}$$

In Equation (2), the "given Y" conditioning allows us to treat $Y-\mu_Y$ as a given that comes outside of the inner expectation. Next, Theorem 7.16 implies

$$\mathrm{E}\left[(X-\mu_X)|Y\right] = \mathrm{E}\left[X|Y\right]-\mu_X = \rho\frac{\sigma_X}{\sigma_Y}(Y-\mu_Y). \tag{3}$$

Therefore, (2) and (2) imply

$$\begin{aligned}\rho_{X,Y} &= \frac{\mathrm{E}\left[(Y-\mu_Y)\rho\frac{\sigma_X}{\sigma_Y}(Y-\mu_Y)\right]}{\sigma_X\sigma_Y}\\ &= \frac{\rho\,\mathrm{E}\left[(Y-\mu_Y)^2\right]}{\sigma_Y^2}. = \rho.\end{aligned} \tag{4}$$

Problem 7.7.1 Solution

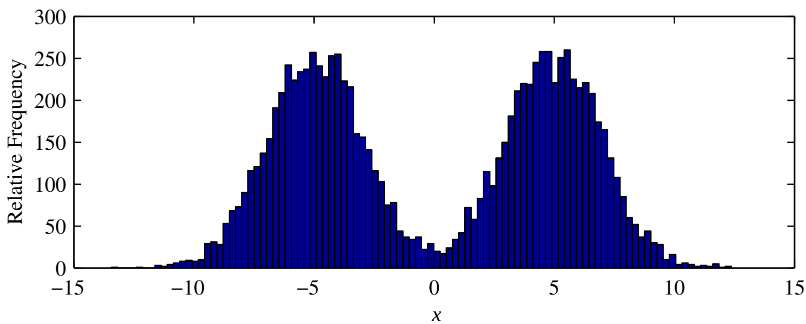
The modem receiver voltage is genrated by taking a ± 5 voltage representing data, and adding to it a Gaussian $(0,2)$ noise variable. Although siuations in which two random variables are added together are not analyzed until Chapter 5, generating samples of the receiver voltage is easy in MATLAB. Here is the code:

```
function x=modemrv(m);
%Usage: x=modemrv(m)
%generates m samples of X, the modem
%receiver voltage in Exampe 3.32.
%X=+-5 + N where N is Gaussian (0,2)
sb=[-5; 5]; pb=[0.5; 0.5];
b=finiterv(sb,pb,m);
noise=gaussrv(0,2,m);
x=b+noise;
```

The commands

```
x=modemrv(10000); hist(x,100);
```

generate 10,000 sample of the modem receiver voltage and plots the relative frequencies using 100 bins. Here is an example plot:



As expected, the result is qualitatively similar ("hills" around $X = -5$ and $X = 5$) to the sketch in Figure 4.3.

Problem 7.7.2 Solution

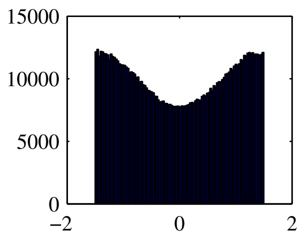
First we need to build a uniform $(-r/2, r/2)$ b-bit quantizer. The function `uquantize` does this.

```
function y=uquantize(r,b,x)
%uniform (-r/2,r/2) b bit quantizer
n=2^b;
delta=r/n;
x=min(x,(r-delta/2)/2);
x=max(x,-(r-delta/2)/2);
y=(delta/2)+delta*floor(x/delta);
```

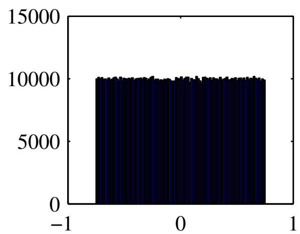
Note that if $|x| > r/2$, then x is truncated so that the quantizer output has maximum amplitude. Next, we generate Gaussian samples, quantize them and record the errors:

```
function stdev=quantizegauss(r,b,m)
x=gaussrv(0,1,m);
x=x((x<=r/2)&(x>=-r/2));
y=uquantize(r,b,x);
z=x-y;
hist(z,100);
stdev=sqrt(sum(z.^2)/length(z));
```

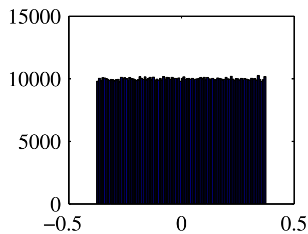
For a Gaussian random variable X, $\mathrm{P}[|X| > r/2] > 0$ for any value of r. When we generate enough Gaussian samples, we will always see some quantization errors due to the finite $(-r/2, r/2)$ range. To focus our attention on the effect of b bit quantization, `quantizegauss.m` eliminates Gaussian samples outside the range $(-r/2, r/2)$. Here are outputs of `quantizegauss` for $b = 1, 2, 3$ bits.



$b = 1$



$b = 2$



$b = 3$

It is obvious that for $b = 1$ bit quantization, the error is decidely not uniform. However, it appears that the error is uniform for $b = 2$ and $b = 3$. You can verify that uniform errors is a reasonable model for larger values of b.

Problem Solutions – Chapter 8

Problem 8.1.1 Solution

This problem is very simple. In terms of the vector $\mathbf{X}$, the PDF is

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1 & \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

However, just keep in mind that the inequalities $\mathbf{0} \leq \mathbf{x}$ and $\mathbf{x} \leq \mathbf{1}$ are vector inequalities that must hold for every component x_i.

Problem 8.1.2 Solution

In this problem, we find the constant c from the requirement that that the integral of the vector PDF over all possible values is 1. That is,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) \, dx_1 \cdots dx_n = 1. \tag{1}$$

Since

$$f_{\mathbf{X}}(\mathbf{x}) = c\mathbf{a}'\mathbf{x} = c \sum_{i=1}^{n} a_i x_i, \tag{2}$$

we have that

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) \, dx_1 \cdots dx_n \\ &\quad = c \int_0^1 \cdots \int_0^1 \left(\sum_{i=1}^{n} a_i x_i \right) dx_1 \cdots dx_n \\ &\quad = c \sum_{i=1}^{n} \left(\int_0^1 \cdots \int_0^1 a_i x_i \, dx_1 \cdots dx_n \right) \\ &\quad = c \sum_{i=1}^{n} a_i \left[\left(\int_0^1 dx_1 \right) \cdots \left(\int_0^1 x_i \, dx_i \right) \cdots \left(\int_0^1 dx_n \right) \right] \\ &\quad = c \sum_{i=1}^{n} a_i \left(\left. \frac{x_i^2}{2} \right|_0^1 \right) = c \sum_{i=1}^{n} \frac{a_i}{2}. \end{aligned} \tag{3}$$

The requirement that the PDF integrate to unity thus implies

$$c = \frac{2}{\sum_{i=1}^{n} a_i}. \tag{4}$$

Problem 8.1.3 Solution

Filling in the parameters in Problem 8.1.2, we obtain the vector PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{2}{3}(x_1 + x_2 + x_3) & 0 \le x_1, x_2, x_3 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this case, for $0 \le x_3 \le 1$, the marginal PDF of X_3 is

$$\begin{aligned} f_{X_3}(x_3) &= \frac{2}{3} \int_0^1 \int_0^1 (x_1 + x_2 + x_3)\, dx_1\, dx_2 \\ &= \frac{2}{3} \int_0^1 \left(\frac{x_1^2}{2} + x_2 x_1 + x_3 x_1 \right) \Bigg|_{x_1=0}^{x_1=1} dx_2 \\ &= \frac{2}{3} \int_0^1 \left(\frac{1}{2} + x_2 + x_3 \right) dx_2 \\ &= \frac{2}{3} \left(\frac{x_2}{2} + \frac{x_2^2}{2} + x_3 x_2 \right) \Bigg|_{x_2=0}^{x_2=1} = \frac{2}{3} \left(\frac{1}{2} + \frac{1}{2} + x_3 \right) \end{aligned} \tag{2}$$

The complete expresion for the marginal PDF of X_3 is

$$f_{X_3}(x_3) = \begin{cases} 2(1 + x_3)/3 & 0 \le x_3 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 8.1.4 Solution

First we note that each marginal PDF is nonzero only if any subset of the x_i obeys the ordering contraints $0 \le x_1 \le x_2 \le x_3 \le 1$. Within these constraints, we have

$$f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\, dx_3 = \int_{x_2}^{1} 6\, dx_3 = 6(1 - x_2), \tag{1}$$

and

$$f_{X_2,X_3}(x_2,x_3) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\ dx_1 = \int_0^{x_2} 6\,dx_1 = 6x_2, \tag{2}$$

and

$$f_{X_1,X_3}(x_1,x_3) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\ dx_2 = \int_{x_1}^{x_3} 6\,dx_2 = 6(x_3 - x_1). \tag{3}$$

In particular, we must keep in mind that $f_{X_1,X_2}(x_1,x_2) = 0$ unless $0 \le x_1 \le x_2 \le 1$, $f_{X_2,X_3}(x_2,x_3) = 0$ unless $0 \le x_2 \le x_3 \le 1$, and that $f_{X_1,X_3}(x_1,x_3) = 0$ unless $0 \le x_1 \le x_3 \le 1$. The complete expressions are

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 6(1-x_2) & 0 \le x_1 \le x_2 \le 1, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_{X_2,X_3}(x_2,x_3) = \begin{cases} 6x_2 & 0 \le x_2 \le x_3 \le 1, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

$$f_{X_1,X_3}(x_1,x_3) = \begin{cases} 6(x_3-x_1) & 0 \le x_1 \le x_3 \le 1, \\ 0 & \text{otherwise}. \end{cases} \tag{5}$$

Now we can find the marginal PDFs. When $0 \le x_i \le 1$ for each x_i,

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\ dx_2 \\ &= \int_{x_1}^{1} 6(1-x_2)\,dx_2 = 3(1-x_1)^2. \end{aligned} \tag{6}$$

$$\begin{aligned} f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_2,X_3}(x_2,x_3)\ dx_3 \\ &= \int_{x_2}^{1} 6x_2\,dx_3 = 6x_2(1-x_2). \end{aligned} \tag{7}$$

$$\begin{aligned} f_{X_3}(x_3) &= \int_{-\infty}^{\infty} f_{X_2,X_3}(x_2,x_3)\ dx_2 \\ &= \int_{0}^{x_3} 6x_2\,dx_2 = 3x_3^2. \end{aligned} \tag{8}$$

The complete expressions are

$$f_{X_1}(x_1) = \begin{cases} 3(1-x_1)^2 & 0 \le x_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

$$f_{X_2}(x_2) = \begin{cases} 6x_2(1-x_2) & 0 \le x_2 \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

$$f_{X_3}(x_3) = \begin{cases} 3x_3^2 & 0 \le x_3 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Problem 8.1.5 Solution

Since J_1, J_2 and J_3 are independent, we can write

$$P_{\mathbf{K}}(\mathbf{k}) = P_{J_1}(k_1)\, P_{J_2}(k_2 - k_1)\, P_{J_3}(k_3 - k_2)\,. \tag{1}$$

Since $P_{J_i}(j) > 0$ only for integers $j > 0$, we have that $P_{\mathbf{K}}(\mathbf{k}) > 0$ only for $0 < k_1 < k_2 < k_3$; otherwise $P_{\mathbf{K}}(\mathbf{k}) = 0$. Finally, for $0 < k_1 < k_2 < k_3$,

$$\begin{aligned} P_{\mathbf{K}}(\mathbf{k}) &= (1-p)^{k_1-1} p (1-p)^{k_2-k_1-1} p (1-p)^{k_3-k_2-1} p \\ &= (1-p)^{k_3-3} p^3. \end{aligned} \tag{2}$$

Problem 8.1.6 Solution

The joint PMF is

$$P_{\mathbf{K}}(\mathbf{k}) = P_{K_1,K_2,K_3}(k_1,k_2,k_3) = \begin{cases} p^3(1-p)^{k_3-3} & 1 \le k_1 < k_2 < k_3, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) We start by finding $P_{K_1,K_2}(k_1,k_2)$. For $1 \le k_1 < k_2$,

$$\begin{aligned} P_{K_1,K_2}(k_1,k_2) &= \sum_{k_3=-\infty}^{\infty} P_{K_1,K_2,K_3}(k_1,k_2,k_3) \\ &= \sum_{k_3=k_2+1}^{\infty} p^3(1-p)^{k_3-3} \\ &= p^3(1-p)^{k_2-2}\left(1 + (1-p) + (1-p)^2 + \cdots\right) \\ &= p^2(1-p)^{k_2-2}. \end{aligned} \tag{2}$$

The complete expression is

$$P_{K_1,K_2}(k_1,k_2) = \begin{cases} p^2(1-p)^{k_2-2} & 1 \le k_1 < k_2, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Next we find $P_{K_1,K_3}(k_1,k_3)$. For $k_1 \ge 1$ and $k_3 \ge k_1 + 2$, we have

$$\begin{aligned} P_{K_1,K_3}(k_1,k_3) &= \sum_{k_2=-\infty}^{\infty} P_{K_1,K_2,K_3}(k_1,k_2,k_3) \\ &= \sum_{k_2=k_1+1}^{k_3-1} p^3(1-p)^{k_3-3} \\ &= (k_3-k_1-1)p^3(1-p)^{k_3-3}. \end{aligned} \tag{4}$$

The complete expression of the PMF of K_1 and K_3 is

$$P_{K_1,K_3}(k_1,k_3) = \begin{cases} (k_3-k_1-1)p^3(1-p)^{k_3-3} & 1 \le k_1 \le k_3-2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The next marginal PMF is

$$\begin{aligned} P_{K_2,K_3}(k_2,k_3) &= \sum_{k_1=-\infty}^{\infty} P_{K_1,K_2,K_3}(k_1,k_2,k_3) \\ &= \sum_{k_1=1}^{k_2-1} p^3(1-p)^{k_3-3} = (k_2-1)p^3(1-p)^{k_3-3}. \end{aligned} \tag{6}$$

The complete expression of the PMF of K_2 and K_3 is

$$P_{K_2,K_3}(k_2,k_3) = \begin{cases} (k_2-1)p^3(1-p)^{k_3-3} & 1 \le k_2 < k_3, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

(b) Going back to first principles, we note that K_n is the number of trials up to and including the nth success. Thus K_1 is a geometric (p) random variable, K_2 is an Pascal $(2,p)$ random variable, and K_3 is an Pascal $(3,p)$ random variable. We could write down the respective marginal PMFs of K_1, K_2 and K_3 just by looking up the Pascal (n, p) PMF. Nevertheless, it is instructive to derive these PMFs from the joint PMF $P_{K_1,K_2,K_3}(k_1, k_2, k_3)$.

For $k_1 \geq 1$, we can find $P_{K_1}(k_1)$ via

$$\begin{aligned} P_{K_1}(k_1) &= \sum_{k_2=-\infty}^{\infty} P_{K_1,K_2}(k_1, k_2) \\ &= \sum_{k_2=k_1+1}^{\infty} p^2(1-p)^{k_2-2} \\ &= p^2(1-p)^{k_1-1}[1 + (1-p) + (1-p)^2 + \cdots] \\ &= p(1-p)^{k_1-1}. \end{aligned} \tag{8}$$

The complete expression for the PMF of K_1 is the usual geometric PMF

$$P_{K_1}(k_1) = \begin{cases} p(1-p)^{k_1-1} & k_1 = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Following the same procedure, the marginal PMF of K_2 is

$$\begin{aligned} P_{K_2}(k_2) = \sum_{k_1=-\infty}^{\infty} P_{K_1,K_2}(k_1, k_2) &= \sum_{k_1=1}^{k_2-1} p^2(1-p)^{k_2-2} \\ &= (k_2 - 1)p^2(1-p)^{k_2-2}. \end{aligned} \tag{10}$$

Since $P_{K_2}(k_2) = 0$ for $k_2 < 2$, the complete PMF is the Pascal $(2, p)$ PMF

$$P_{K_2}(k_2) = \binom{k_2 - 1}{1} p^2(1-p)^{k_2-2}. \tag{11}$$

Finally, for $k_3 \geq 3$, the PMF of K_3 is

$$\begin{aligned}
P_{K_3}(k_3) &= \sum_{k_2=-\infty}^{\infty} P_{K_2,K_3}(k_2, k_3) \\
&= \sum_{k_2=2}^{k_3-1} (k_2 - 1) p^3 (1-p)^{k_3-3} \\
&= [1 + 2 + \cdots + (k_3 - 2)] p^3 (1-p)^{k_3-3} \\
&= \frac{(k_3-2)(k_3-1)}{2} p^3 (1-p)^{k_3-3}. \qquad (12)
\end{aligned}$$

Since $P_{K_3}(k_3) = 0$ for $k_3 < 3$, the complete expression for $P_{K_3}(k_3)$ is the Pascal $(3, p)$ PMF

$$P_{K_3}(k_3) = \binom{k_3 - 1}{2} p^3 (1-p)^{k_3-3}. \qquad (13)$$

Problem 8.1.7 Solution

In Example 5.21, random variables $N_1, \ldots, N_r$ have the multinomial distribution

$$P_{N_1,\ldots,N_r}(n_1, \ldots, n_r) = \binom{n}{n_1, \ldots, n_r} p_1^{n_1} \cdots p_r^{n_r} \qquad (1)$$

where $n > r > 2$.

(a) To evaluate the joint PMF of N_1 and N_2, we define a new experiment with mutually exclusive events: s_1, s_2 and "other" Let $\hat{N}$ denote the number of trial outcomes that are "other". In this case, a trial is in the "other" category with probability $\hat{p} = 1 - p_1 - p_2$. The joint PMF of N_1, N_2, and $\hat{N}$ is

$$P_{N_1,N_2,\hat{N}}(n_1, n_2, \hat{n}) = \binom{n}{n_1, n_2, \hat{n}} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{\hat{n}}. \qquad (2)$$

Now we note that the following events are one in the same:

$$\{N_1 = n_1, N_2 = n_2\} = \left\{N_1 = n_1, N_2 = n_2, \hat{N} = n - n_1 - n_2\right\}. \qquad (3)$$

Hence,

$$
\begin{aligned}
P_{N_1,N_2}(n_1, n_2) &= P_{N_1,N_2,\hat{N}}(n_1, n_2, n - n_1 - n_2) \\
&= \binom{n}{n_1, n_2, n - n_1 - n_2} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n-n_1-n_2}. \qquad (4)
\end{aligned}
$$

From the definition of the multinomial coefficient, $P_{N_1,N_2}(n_1, n_2) \neq 0$ only for non-negative integers n_1 and n_2 satisfying $n_1 + n_2 \leq n$.

(b) We could find the PMF of T_i by summing $P_{N_1,\ldots,N_r}(n_1, \ldots, n_r)$. However, it is easier to start from first principles. Suppose we say a success occurs if the outcome of the trial is in the set $\{s_1, s_2, \ldots, s_i\}$ and otherwise a failure occurs. In this case, the success probability is $q_i = p_1 + \cdots + p_i$ and T_i is the number of successes in n trials. Thus, T_i has the binomial PMF

$$
P_{T_i}(t) = \begin{cases} \binom{n}{t} q_i^t (1 - q_i)^{n-t} & t = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \qquad (5)
$$

(c) The joint PMF of T_1 and T_2 satisfies

$$
\begin{aligned}
P_{T_1,T_2}(t_1, t_2) &= \mathrm{P}\left[N_1 = t_1, N_1 + N_2 = t_2\right] \\
&= \mathrm{P}\left[N_1 = t_1, N_2 = t_2 - t_1\right] \\
&= P_{N_1,N_2}(t_1, t_2 - t_1). \qquad (6)
\end{aligned}
$$

By the result of part (a),

$$
P_{T_1,T_2}(t_1, t_2) = \binom{n}{t_1, t_2 - t_1, n - t_2} p_1^{t_1} p_2^{t_2 - t_1} (1 - p_1 - p_2)^{n - t_2}. \qquad (7)
$$

Similar to the previous parts, keep in mind that $P_{T_1,T_2}(t_1, t_2)$ is nonzero only if $0 \leq t_1 \leq t_2 \leq n$.

Problem 8.1.8 Solution

For $0 \le y_1 \le y_4 \le 1$, the marginal PDF of Y_1 and Y_4 satisfies

$$\begin{aligned} f_{Y_1,Y_4}(y_1,y_4) &= \iint f_{\mathbf{Y}}(\mathbf{y})\, dy_2\, dy_3 \\ &= \int_{y_1}^{y_4} \left(\int_{y_2}^{y_4} 24\, dy_3 \right) dy_2 \\ &= \int_{y_1}^{y_4} 24(y_4 - y_2)\, dy_2 \\ &= -12(y_4 - y_2)^2 \Big|_{y_2=y_1}^{y_2=y_4} = 12(y_4 - y_1)^2. \end{aligned} \tag{1}$$

The complete expression for the joint PDF of Y_1 and Y_4 is

$$f_{Y_1,Y_4}(y_1,y_4) = \begin{cases} 12(y_4 - y_1)^2 & 0 \le y_1 \le y_4 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

For $0 \le y_1 \le y_2 \le 1$, the marginal PDF of Y_1 and Y_2 is

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \iint f_{\mathbf{Y}}(\mathbf{y})\, dy_3\, dy_4 \\ &= \int_{y_2}^{1} \left(\int_{y_3}^{1} 24\, dy_4 \right) dy_3 \\ &= \int_{y_2}^{1} 24(1 - y_3)\, dy_3 = 12(1 - y_2)^2. \end{aligned} \tag{3}$$

The complete expression for the joint PDF of Y_1 and Y_2 is

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 12(1 - y_2)^2 & 0 \le y_1 \le y_2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

For $0 \le y_1 \le 1$, the marginal PDF of Y_1 can be found from

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\, dy_2 \\ &= \int_{y_1}^{1} 12(1 - y_2)^2\, dy_2 = 4(1 - y_1)^3. \end{aligned} \tag{5}$$

The complete expression of the PDF of Y_1 is

$$f_{Y_1}(y_1) = \begin{cases} 4(1-y_1)^3 & 0 \le y_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Note that the integral $f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_4}(y_1, y_4)\, dy_4$ would have yielded the same result. This is a good way to check our derivations of $f_{Y_1,Y_4}(y_1, y_4)$ and $f_{Y_1,Y_2}(y_1, y_2)$.

Problem 8.1.9 Solution

In Problem 8.1.5, we found that the joint PMF of $\mathbf{K} = \begin{bmatrix} K_1 & K_2 & K_3 \end{bmatrix}'$ is

$$P_{\mathbf{K}}(\mathbf{k}) = \begin{cases} p^3(1-p)^{k_3-3} & k_1 < k_2 < k_3, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this problem, we generalize the result to n messages.

(a) For $k_1 < k_2 < \cdots < k_n$, the joint event

$$\{K_1 = k_1, K_2 = k_2, \cdots, K_n = k_n\} \tag{2}$$

occurs if and only if all of the following events occur

A_1 $k_1 - 1$ failures, followed by a successful transmission,
A_2 $(k_2 - 1) - k_1$ failures followed by a successful transmission,
A_3 $(k_3 - 1) - k_2$ failures followed by a successful transmission,
$\vdots$
A_n $(k_n - 1) - k_{n-1}$ failures followed by a successful transmission.

Note that the events $A_1, A_2, \ldots, A_n$ are independent and

$$\mathrm{P}[A_j] = (1-p)^{k_j - k_{j-1} - 1} p. \tag{3}$$

Thus

$$\begin{aligned} &P_{K_1,\ldots,K_n}(k_1, \ldots, k_n) \\ &\quad = \mathrm{P}[A_1]\,\mathrm{P}[A_2] \cdots \mathrm{P}[A_n] \\ &\quad = p^n (1-p)^{(k_1-1)+(k_2-k_1-1)+(k_3-k_2-1)+\cdots+(k_n-k_{n-1}-1)} \\ &\quad = p^n (1-p)^{k_n - n}. \end{aligned} \tag{4}$$

To clarify subsequent results, it is better to rename $\mathbf{K}$ as

$$\mathbf{K}_n = \begin{bmatrix} K_1 & K_2 & \cdots & K_n \end{bmatrix}'. \tag{5}$$

We see that

$$P_{\mathbf{K}_n}(\mathbf{k}_n) = \begin{cases} p^n(1-p)^{k_n-n} & 1 \le k_1 < k_2 < \cdots < k_n, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

(b) For $j < n$,

$$P_{K_1,K_2,\ldots,K_j}(k_1, k_2, \ldots, k_j) = P_{\mathbf{K}_j}(\mathbf{k}_j). \tag{7}$$

Since $\mathbf{K}_j$ is just $\mathbf{K}_n$ with $n = j$, we have

$$P_{\mathbf{K}_j}(\mathbf{k}_j) = \begin{cases} p^j(1-p)^{k_j-j} & 1 \le k_1 < k_2 < \cdots < k_j, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

(c) Rather than try to deduce $P_{K_i}(k_i)$ from the joint PMF $P_{\mathbf{K}_n}(\mathbf{k}_n)$, it is simpler to return to first principles. In particular, K_i is the number of trials up to and including the ith success and has the Pascal (i, p) PMF

$$P_{K_i}(k_i) = \binom{k_i - 1}{i - 1} p^i (1-p)^{k_i - i}. \tag{9}$$

Problem 8.2.1 Solution

For $i \neq j$, X_i and X_j are independent and $\mathrm{E}[X_i X_j] = \mathrm{E}[X_i]\,\mathrm{E}[X_j] = 0$ since $\mathrm{E}[X_i] = 0$. Thus the i,jth entry in the covariance matrix $\mathbf{C_X}$ is

$$C_{\mathbf{X}}(i,j) = \mathrm{E}[X_i X_j] = \begin{cases} \sigma_i^2 & i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Thus for random vector $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}'$, all the off-diagonal entries in the covariance matrix are zero and the covariance matrix is

$$\mathbf{C_X} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}. \tag{2}$$

Problem 8.2.2 Solution

We will use the PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1 & 0 \leq x_i \leq 1, i = 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

to find the marginal PDFs $f_{X_i}(x_i)$. In particular, for $0 \leq x_1 \leq 1$,

$$\begin{aligned} f_{X_1}(x_1) &= \int_0^1 \int_0^1 \int_0^1 f_{\mathbf{X}}(\mathbf{x})\, dx_2\, dx_3\, dx_4 \\ &= \left(\int_0^1 dx_2\right)\left(\int_0^1 dx_3\right)\left(\int_0^1 dx_4\right) = 1. \end{aligned} \tag{2}$$

Thus,

$$f_{X_1}(x_1) = \begin{cases} 1 & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Following similar steps, one can show that

$$f_{X_1}(x) = f_{X_2}(x) = f_{X_3}(x) = f_{X_4}(x) = \begin{cases} 1 & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Thus

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x)\, f_{X_2}(x)\, f_{X_3}(x)\, f_{X_4}(x)\,. \tag{5}$$

We conclude that X_1, X_2, X_3 and X_4 are independent.

Problem 8.2.3 Solution

We will use the PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6e^{-(x_1+2x_2+3x_3)} & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

to find the marginal PDFs $f_{X_i}(x_i)$. In particular, for $x_1 \geq 0$,

$$\begin{aligned} f_{X_1}(x_1) &= \int_0^\infty \int_0^\infty f_{\mathbf{X}}(\mathbf{x})\, dx_2\, dx_3 \\ &= 6e^{-x_1} \left(\int_0^\infty e^{-2x_2} dx_2 \right) \left(\int_0^\infty \right) e^{-3x_3}\, dx_3 \\ &= 6e^{-x_1} \left(-\frac{1}{2} e^{-2x_2} \Big|_0^\infty \right) \left(-\frac{1}{3} e^{-3x_3} \Big|_0^\infty \right) = e^{-x_1}. \end{aligned} \tag{2}$$

Thus,

$$f_{X_1}(x_1) = \begin{cases} e^{-x_1} & x_1 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Following similar steps, one can show that

$$f_{X_2}(x_2) = \int_0^\infty \int_0^\infty f_{\mathbf{X}}(\mathbf{x})\, dx_1\, dx_3 = \begin{cases} 2^{-2x_2} & x_2 \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

$$f_{X_3}(x_3) = \int_0^\infty \int_0^\infty f_{\mathbf{X}}(\mathbf{x})\, dx_1\, dx_2 = \begin{cases} 3^{-3x_3} & x_3 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Thus

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3)\,. \tag{6}$$

We conclude that X_1, X_2, and X_3 are independent.

Problem 8.2.4 Solution

This problem can be solved without any real math. Some thought should convince you that for any $x_i > 0$, $f_{X_i}(x_i) > 0$. Thus, $f_{X_1}(10) > 0$, $f_{X_2}(9) > 0$, and $f_{X_3}(8) > 0$. Thus $f_{X_1}(10) f_{X_2}(9) f_{X_3}(8) > 0$. However, from the definition of the joint PDF

$$f_{X_1,X_2,X_3}(10, 9, 8) = 0 \neq f_{X_1}(10)\, f_{X_2}(9)\, f_{X_3}(8)\,. \tag{1}$$

It follows that X_1, X_2 and X_3 are dependent. Readers who find this quick answer dissatisfying are invited to confirm this conclusions by solving Problem 8.2.5 for the exact expressions for the marginal PDFs $f_{X_1}(x_1)$, $f_{X_2}(x_2)$, and $f_{X_3}(x_3)$.

Problem 8.2.5 Solution

We find the marginal PDFs using Theorem 5.26. First we note that for $x < 0$, $f_{X_i}(x) = 0$. For $x_1 \geq 0$,

$$f_{X_1}(x_1) = \int_{x_1}^{\infty} \left(\int_{x_2}^{\infty} e^{-x_3}\, dx_3 \right) dx_2 = \int_{x_1}^{\infty} e^{-x_2}\, dx_2 = e^{-x_1}. \tag{1}$$

Similarly, for $x_2 \geq 0$, X_2 has marginal PDF

$$f_{X_2}(x_2) = \int_{0}^{x_2} \left(\int_{x_2}^{\infty} e^{-x_3}\, dx_3 \right) dx_1 = \int_{0}^{x_2} e^{-x_2}\, dx_1 = x_2 e^{-x_2}. \tag{2}$$

Lastly,

$$\begin{aligned} f_{X_3}(x_3) &= \int_{0}^{x_3} \left(\int_{x_1}^{x_3} e^{-x_3}\, dx_2 \right) dx_1 \\ &= \int_{0}^{x_3} (x_3 - x_1) e^{-x_3}\, dx_1 \\ &= -\frac{1}{2}(x_3 - x_1)^2 e^{-x_3} \Big|_{x_1=0}^{x_1=x_3} = \frac{1}{2} x_3^2 e^{-x_3}. \end{aligned} \tag{3}$$

The complete expressions for the three marginal PDFs are

$$f_{X_1}(x_1) = \begin{cases} e^{-x_1} & x_1 \geq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

$$f_{X_2}(x_2) = \begin{cases} x_2 e^{-x_2} & x_2 \geq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

$$f_{X_3}(x_3) = \begin{cases} (1/2) x_3^2 e^{-x_3} & x_3 \geq 0, \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

In fact, each X_i is an Erlang $(n, \lambda) = (i, 1)$ random variable.

Problem 8.3.1 Solution

For discrete random vectors, it is true in general that

$$P_{\mathbf{Y}}(\mathbf{y}) = \mathrm{P}[\mathbf{Y} = \mathbf{y}] = \mathrm{P}[\mathbf{AX} + \mathbf{b} = \mathbf{y}] = \mathrm{P}[\mathbf{AX} = \mathbf{y} - \mathbf{b}]. \tag{1}$$

For an arbitrary matrix $\mathbf{A}$, the system of equations $\mathbf{Ax} = \mathbf{y} - \mathbf{b}$ may have no solutions (if the columns of $\mathbf{A}$ do not span the vector space), multiple solutions (if the columns of $\mathbf{A}$ are linearly dependent), or, when $\mathbf{A}$ is invertible, exactly one solution. In the invertible case,

$$P_{\mathbf{Y}}(\mathbf{y}) = \mathrm{P}\left[\mathbf{AX} = \mathbf{y} - \mathbf{b}\right] = \mathrm{P}\left[\mathbf{X} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})\right] = P_{\mathbf{X}}\left(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})\right). \tag{2}$$

As an aside, we note that when $\mathbf{Ax} = \mathbf{y} - \mathbf{b}$ has multiple solutions, we would need to do some bookkeeping to add up the probabilities $P_{\mathbf{X}}(\mathbf{x})$ for all vectors $\mathbf{x}$ satisfying $\mathbf{Ax} = \mathbf{y} - \mathbf{b}$. This can get disagreeably complicated.

Problem 8.3.2 Solution

The random variable J_n is the number of times that message n is transmitted. Since each transmission is a success with probability p, independent of any other transmission, the number of transmissions of message n is independent of the number of transmissions of message m. That is, for $m \neq n$, J_m and J_n are independent random variables. Moreover, because each message is transmitted over and over until it is transmitted succesfully, each J_m is a geometric (p) random variable with PMF

$$P_{J_m}(j) = \begin{cases} (1-p)^{j-1}p & j = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Thus the PMF of $\mathbf{J} = \begin{bmatrix} J_1 & J_2 & J_3 \end{bmatrix}'$ is

$$P_{\mathbf{J}}(\mathbf{j}) = P_{J_1}(j_1)\, P_{J_2}(j_2)\, P_{J_3}(j_3) = \begin{cases} p^3(1-p)^{j_1+j_2+j_3-3} & j_i = 1, 2, \ldots; \\ & i = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 8.3.3 Solution

The response time X_i of the ith truck has PDF $f_{X_i}(x_i)$ and CDF $F_{X_i}(x_i)$ given by

$$f_{X_i}(x) = \begin{cases} \frac{1}{2}e^{-x/2} & x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$F_{X_i}(x) = \begin{cases} 1 - e^{-x/2} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Let $R = \max(X_1, X_2, \ldots, X_6)$ denote the maximum response time. From Theorem 8.2, R has PDF

$$F_R(r) = (F_X(r))^6. \tag{2}$$

(a) The probability that all six responses arrive within five seconds is

$$P[R \leq 5] = F_R(5) = (F_X(5))^6 = (1 - e^{-5/2})^6 = 0.5982. \tag{3}$$

(b) This question is worded in a somewhat confusing way. The "expected response time" refers to $E[X_i]$, the response time of an individual truck, rather than $E[R]$. If the expected response time of a truck is τ, then each X_i has CDF

$$F_{X_i}(x) = F_X(x) = \begin{cases} 1 - e^{-x/\tau} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The goal of this problem is to find the maximum permissible value of τ. When each truck has expected response time τ, the CDF of R is

$$F_R(r) = (F_X(x)\, r)^6 = \begin{cases} (1 - e^{-r/\tau})^6 & r \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

We need to find τ such that

$$P[R \leq 3] = (1 - e^{-3/\tau})^6 = 0.9. \tag{6}$$

This implies

$$\tau = \frac{-3}{\ln\left(1 - (0.9)^{1/6}\right)} = 0.7406 \text{ s.} \tag{7}$$

Problem 8.3.4 Solution

Let A denote the event $X_n = \max(X_1, \ldots, X_n)$. We can find $\mathrm{P}[A]$ by conditioning on the value of X_n.

$$\begin{aligned}
\mathrm{P}[A] &= \mathrm{P}[X_1 \leq X_n, X_2 \leq X_n, \cdots, X_{n_1} \leq X_n] \\
&= \int_{-\infty}^{\infty} \mathrm{P}[X_1 < X_n, X_2 < X_n, \cdots, X_{n-1} < X_n | X_n = x] f_{X_n}(x)\, dx \\
&= \int_{-\infty}^{\infty} \mathrm{P}[X_1 < x, X_2 < x, \cdots, X_{n-1} < x | X_n = x] f_X(x)\, dx. \qquad (1)
\end{aligned}$$

Since $X_1, \ldots, X_{n-1}$ are independent of X_n,

$$\mathrm{P}[A] = \int_{-\infty}^{\infty} \mathrm{P}[X_1 < x, X_2 < x, \cdots, X_{n-1} < x] f_X(x)\, dx. \qquad (2)$$

Since $X_1, \ldots, X_{n-1}$ are iid,

$$\begin{aligned}
\mathrm{P}[A] &= \int_{-\infty}^{\infty} \mathrm{P}[X_1 \leq x] \mathrm{P}[X_2 \leq x] \cdots \mathrm{P}[X_{n-1} \leq x] f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} [F_X(x)]^{n-1} f_X(x)\, dx = \frac{1}{n} [F_X(x)]^n \Big|_{-\infty}^{\infty} = \frac{1}{n}(1-0) \qquad (3)
\end{aligned}$$

Not surprisingly, since the X_i are identical, symmetry would suggest that X_n is as likely as any of the other X_i to be the largest. Hence $\mathrm{P}[A] = 1/n$ should not be surprising.

Problem 8.4.1 Solution

(a) The coavariance matrix of $\mathbf{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}'$ is

$$\mathbf{C_X} = \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] \\ \mathrm{Cov}[X_1, X_2] & \mathrm{Var}[X_2] \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}. \qquad (1)$$

(b) From the problem statement,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} \mathbf{X} = \mathbf{A}\mathbf{X}. \qquad (2)$$

By Theorem 8.8, $\mathbf{Y}$ has covariance matrix

$$\mathbf{C_Y} = \mathbf{A}\mathbf{C_X}\mathbf{A}' = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 28 & -66 \\ -66 & 252 \end{bmatrix}. \tag{3}$$

Problem 8.4.2 Solution

The mean value of a sum of random variables is always the sum of their individual means.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} \mathrm{E}[X_i] = 0. \tag{1}$$

The variance of any sum of random variables can be expressed in terms of the individual variances and co-variances. Since the $\mathrm{E}[Y]$ is zero, $\mathrm{Var}[Y] = \mathrm{E}[Y^2]$. Thus,

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] \\ &= \sum_{i=1}^{n} \mathrm{E}\left[X_i^2\right] + \sum_{i=1}^{n}\sum_{j\neq i} \mathrm{E}[X_i X_j]. \end{aligned} \tag{2}$$

Since $\mathrm{E}[X_i] = 0$, $\mathrm{E}[X_i^2] = \mathrm{Var}[X_i] = 1$ and for $i \neq j$,

$$\mathrm{E}[X_i X_j] = \mathrm{Cov}[X_i, X_j] = \rho \tag{3}$$

Thus, $\mathrm{Var}[Y] = n + n(n-1)\rho$.

Problem 8.4.3 Solution

Since $\mathbf{X}$ and $\mathbf{Y}$ are independent and $\mathrm{E}[Y_j] = 0$ for all components Y_j, we observe that $\mathrm{E}[X_i Y_j] = \mathrm{E}[X_i]\,\mathrm{E}[Y_j] = 0$. This implies that the cross-covariance matrix is

$$\mathrm{E}\left[\mathbf{X}\mathbf{Y}'\right] = \mathrm{E}[\mathbf{X}]\,\mathrm{E}\left[\mathbf{Y}'\right] = \mathbf{0}. \tag{1}$$

Problem 8.4.4 Solution

Inspection of the vector PDF $f_{\mathbf{X}}(\mathbf{x})$ will show that X_1, X_2, X_3, and X_4 are iid uniform $(0,1)$ random variables. That is,

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3)\, f_{X_4}(x_4)\,, \tag{1}$$

where each X_i has the uniform $(0,1)$ PDF

$$f_{X_i}(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It follows that for each i, $\mathrm{E}[X_i] = 1/2$, $\mathrm{E}[X_i^2] = 1/3$ and $\mathrm{Var}[X_i] = 1/12$. In addition, X_i and X_j have correlation

$$\mathrm{E}[X_i X_j] = \mathrm{E}[X_i]\,\mathrm{E}[X_j] = 1/4. \tag{3}$$

and covariance $\mathrm{Cov}[X_i, X_j] = 0$ for $i \neq j$ since independent random variables always have zero covariance.

(a) The expected value vector is

$$\begin{aligned} \mathrm{E}[\mathbf{X}] &= \begin{bmatrix} \mathrm{E}[X_1] & \mathrm{E}[X_2] & \mathrm{E}[X_3] & \mathrm{E}[X_4] \end{bmatrix}' \\ &= \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix}'. \end{aligned} \tag{4}$$

(b) The correlation matrix is

$$\begin{aligned} \mathbf{R}_X = \mathrm{E}[\mathbf{X}\mathbf{X}'] &= \begin{bmatrix} \mathrm{E}[X_1^2] & \mathrm{E}[X_1X_2] & \mathrm{E}[X_1X_3] & \mathrm{E}[X_1X_4] \\ \mathrm{E}[X_2X_1] & \mathrm{E}[X_2^2] & \mathrm{E}[X_2X_3] & \mathrm{E}[X_2X_4] \\ \mathrm{E}[X_3X_1] & \mathrm{E}[X_3X_2] & \mathrm{E}[X_3^2] & \mathrm{E}[X_3X_4] \\ \mathrm{E}[X_4X_1] & \mathrm{E}[X_4X_2] & \mathrm{E}[X_4X_3] & \mathrm{E}[X_4^2] \end{bmatrix} \\ &= \begin{bmatrix} 1/3 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/3 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/3 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/3 \end{bmatrix}. \end{aligned} \tag{5}$$

(c) The covariance matrix for $\mathbf{X}$ is the diagonal matrix

$$\mathbf{C}_X = \begin{bmatrix} \text{Var}[X_1] & \text{Cov}[X_1, X_2] & \text{Cov}[X_1, X_3] & \text{Cov}[X_1, X_4] \\ \text{Cov}[X_2, X_1] & \text{Var}[X_2] & \text{Cov}[X_2, X_3] & \text{Cov}[X_2, X_4] \\ \text{Cov}[X_3, X_1] & \text{Cov}[X_3, X_2] & \text{Var}[X_3] & \text{Cov}[X_3, X_4] \\ \text{Cov}[X_4, X_1] & \text{Cov}[X_4, X_2] & \text{Cov}[X_4, X_3] & \text{Var}[X_4] \end{bmatrix}$$

$$= \begin{bmatrix} 1/12 & 0 & 0 & 0 \\ 0 & 1/12 & 0 & 0 \\ 0 & 0 & 1/12 & 0 \\ 0 & 0 & 0 & 1/12 \end{bmatrix}. \tag{6}$$

Note that its easy to verify that $\mathbf{C}_X = \mathbf{R}_X - \boldsymbol{\mu}_X \boldsymbol{\mu}'_X$.

Problem 8.4.5 Solution

From $\mathbf{C}_\mathbf{Y}$ we see that

$$\rho_{Y_1 Y_2} = \frac{\text{Cov}[Y_1, Y_2]}{\sqrt{\text{Var}[Y_1]\,\text{Var}[Y_2]}} = \frac{\gamma}{\sqrt{(25)(4)}} = \gamma/10. \tag{1}$$

The requirement $|\rho_{Y_1 Y_2}| \leq 1$ implies $|\gamma| \leq 10$. Note that you can instead require that the eigenvalues of $\mathbf{C}_\mathbf{Y}$ are non-negative. This will lead to the same condition.

Problem 8.4.6 Solution

The random variable J_m is the number of times that message m is transmitted. Since each transmission is a success with probability p, independent of any other transmission, J_1, J_2 and J_3 are iid geometric (p) random variables with

$$\text{E}[J_m] = \frac{1}{p}, \qquad \text{Var}[J_m] = \frac{1-p}{p^2}. \tag{1}$$

Thus the vector $\mathbf{J} = \begin{bmatrix} J_1 & J_2 & J_3 \end{bmatrix}'$ has expected value

$$\text{E}[\mathbf{J}] = \begin{bmatrix} \text{E}[J_1] & \text{E}[J_2] & EJ_3 \end{bmatrix}' = \begin{bmatrix} 1/p & 1/p & 1/p \end{bmatrix}'. \tag{2}$$

For $m \neq n$, the correlation matrix $\mathbf{R}_\mathbf{J}$ has m, nth entry

$$R_\mathbf{J}(m, n) = \text{E}[J_m J_n] = \text{E}[J_m J_n] = 1/p^2. \tag{3}$$

For $m = n$,

$$R_{\mathbf{J}}(m,m) = \mathrm{E}\left[J_m^2\right] = \mathrm{Var}[J_m] + (\mathrm{E}\left[J_m^2\right])^2 = \frac{1-p}{p^2} + \frac{1}{p^2} = \frac{2-p}{p^2}. \tag{4}$$

Thus

$$\mathbf{R_J} = \frac{1}{p^2}\begin{bmatrix} 2-p & 1 & 1 \\ 1 & 2-p & 1 \\ 1 & 1 & 2-p \end{bmatrix}. \tag{5}$$

Because J_m and J_n are independent, off-diagonal terms in the covariance matrix are

$$C_{\mathbf{J}}(m,n) = \mathrm{Cov}\left[J_m, J_n\right] = 0. \tag{6}$$

Since $C_{\mathbf{J}}(m,m) = \mathrm{Var}[J_m]$, we have that

$$\mathbf{C_J} = \frac{1-p}{p^2}\mathbf{I} = \frac{1-p}{p^2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

Problem 8.4.7 Solution

This problem is quite difficult unless one uses the observation that the vector $\mathbf{K}$ can be expressed in terms of the vector $\mathbf{J} = \begin{bmatrix} J_1 & J_2 & J_3 \end{bmatrix}'$ where J_i is the number of transmissions of message i. Note that we can write

$$\mathbf{K} = \mathbf{AJ} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{J}. \tag{1}$$

We also observe that since each transmission is an independent Bernoulli trial with success probability p, the components of $\mathbf{J}$ are iid geometric (p) random variables. Thus $\mathrm{E}[J_i] = 1/p$ and $\mathrm{Var}[J_i] = (1-p)/p^2$. Thus $\mathbf{J}$ has expected value

$$\mathrm{E}\left[\mathbf{J}\right] = \boldsymbol{\mu}_J = \begin{bmatrix} \mathrm{E}\left[J_1\right] & \mathrm{E}\left[J_2\right] & \mathrm{E}\left[J_3\right] \end{bmatrix}' = \begin{bmatrix} 1/p & 1/p & 1/p \end{bmatrix}'. \tag{2}$$

Since the components of $\mathbf{J}$ are independent, it has the diagonal covariance matrix

$$\mathbf{C}_J = \begin{bmatrix} \mathrm{Var}[J_1] & 0 & 0 \\ 0 & \mathrm{Var}[J_2] & 0 \\ 0 & 0 & \mathrm{Var}[J_3] \end{bmatrix} = \frac{1-p}{p^2}\mathbf{I}. \tag{3}$$

Given these properties of $\mathbf{J}$, finding the same properties of $\mathbf{K} = \mathbf{AJ}$ is simple.

(a) The expected value of $\mathbf{K}$ is

$$\mathrm{E}[\mathbf{K}] = \mathbf{A}\boldsymbol{\mu}_J = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/p \\ 1/p \\ 1/p \end{bmatrix} = \begin{bmatrix} 1/p \\ 2/p \\ 3/p \end{bmatrix}. \tag{4}$$

(b) From Theorem 8.8, the covariance matrix of $\mathbf{K}$ is

$$\begin{aligned} \mathbf{C}_K &= \mathbf{A}\mathbf{C}_J\mathbf{A}' \\ &= \frac{1-p}{p^2}\mathbf{A}\mathbf{I}\mathbf{A}' \\ &= \frac{1-p}{p^2} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1-p}{p^2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}. \end{aligned} \tag{5}$$

(c) Given the expected value vector $\boldsymbol{\mu}_K$ and the covariance matrix $\mathbf{C}_K$, we can use Theorem 8.7 to find the correlation matrix

$$\begin{aligned} \mathbf{R}_K &= \mathbf{C}_K + \boldsymbol{\mu}_K\boldsymbol{\mu}_K' \\ &= \frac{1-p}{p^2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} 1/p \\ 2/p \\ 3/p \end{bmatrix} \begin{bmatrix} 1/p & 2/p & 3/p \end{bmatrix} && (6) \\ &= \frac{1-p}{p^2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} + \frac{1}{p^2} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} && (7) \\ &= \frac{1}{p^2} \begin{bmatrix} 2-p & 3-p & 4-p \\ 3-p & 6-2p & 8-2p \\ 4-p & 8-2p & 12-3p \end{bmatrix}. && (8) \end{aligned}$$

Problem 8.4.8 Solution

(a) Note that for $x_1, x_2 \geq 0$, $f_{X_1,X_2}(x_1,x_2) = 10e^{-5x_1}e^{-2x_2}$. It follows that for $x_1 \geq 0$,

$$
\begin{aligned}
f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\, dx_2 \\
&= \int_0^{\infty} 10e^{-5x_1}e^{-2x_2}\, dx_2 \\
&= 5e^{-5x_1}\left(-e^{-2x_2}\Big|_0^{\infty}\right) = 5e^{-5x_1}. \qquad (1)
\end{aligned}
$$

Since $f_{X_1}(x_1) = 0$ for $x_1 < 0$, we see that X_1 has the exponential PDF

$$
f_{X_1}(x_1) = \begin{cases} 5e^{-5x_1} & x_1 \geq 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (2)
$$

Following the same procedure, one can show

$$
f_{X_2}(x_2) = \begin{cases} 2e^{-2x_2} & x_2 \geq 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (3)
$$

(b) Since X_i is exponential (λ_i), $\mathrm{E}[X_i] = 1/\lambda_i$ and $\mathrm{Var}[X_i] = 1/\lambda_i^2$. In addition, we found in the previous part that

$$
f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2). \qquad (4)
$$

and thus X_1 and X_2 are independent. This implies $\mathrm{Cov}[X_1,X_2] = 0$. The expected value vector is

$$
\boldsymbol{\mu}_X = \begin{bmatrix} \mathrm{E}[X_1] \\ \mathrm{E}[X_2] \end{bmatrix} = \begin{bmatrix} 1/\lambda_1 \\ 1/\lambda_2 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 1/2 \end{bmatrix}. \qquad (5)
$$

The covariance matrix is

$$
\begin{aligned}
\mathbf{C}_\mathbf{X} &= \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1,X_2] \\ \mathrm{Cov}[X_1,X_2] & \mathrm{Var}[X_2] \end{bmatrix} \\
&= \begin{bmatrix} 1/\lambda_1^2 & 0 \\ 0 & 1/\lambda_2^2 \end{bmatrix} = \begin{bmatrix} 1/25 & 0 \\ 0 & 1/4 \end{bmatrix}. \qquad (6)
\end{aligned}
$$

(c) Since

$$\mathbf{Z} = \mathbf{A}\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{X}, \tag{7}$$

the vector $\mathbf{Z}$ has covariance matrix

$$\begin{aligned} \mathbf{C_Z} &= \mathbf{A}\mathbf{C_X}\mathbf{A}' \\ &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/25 & 0 \\ 0 & 1/4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{100} \begin{bmatrix} 29 & -21 \\ -21 & 29 \end{bmatrix}. \end{aligned} \tag{8}$$

Problem 8.4.9 Solution

In Example 5.23, we found the marginal PDF of Y_3 is

$$f_{Y_3}(y_3) = \begin{cases} 2(1-y_3) & 0 \leq y_3 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We also need to find the marginal PDFs of Y_1, Y_2, and Y_4. In Equation (5.78) of Example 5.23, we found the marginal PDF

$$f_{Y_1,Y_4}(y_1, y_4) = \begin{cases} 4(1-y_1)y_4 & 0 \leq y_1 \leq 1, 0 \leq y_4 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We can use this result to show that

$$f_{Y_1}(y_1) = \int_0^1 f_{Y_1,Y_4}(y_1, y_4)\, dy_4 = 2(1-y_1), \qquad 0 \leq y_1 \leq 1, \tag{3}$$

$$f_{Y_4}(y_4) = \int_0^1 f_{Y_1,Y_4}(y_1, y_4)\, dy_1 = 2y_4, \qquad 0 \leq y_4 \leq 1. \tag{4}$$

The full expressions for the marginal PDFs are

$$f_{Y_1}(y_1) = \begin{cases} 2(1-y_1) & 0 \leq y_1 \leq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$f_{Y_4}(y_4) = \begin{cases} 2y_4 & 0 \leq y_4 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Similarly, we found in Equation (5.80) of Example 5.23 the marginal PDF

$$f_{Y_2,Y_3}(y_2,y_3) = \begin{cases} 4y_2(1-y_3) & 0 \le y_2 \le 1, 0 \le y_3 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

This implies that for $0 \le y_2 \le 1$,

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_2,Y_3}(y_2,y_3)\ dy_3 = \int_0^1 4y_2(1-y_3)\, dy_3 = 2y_2 \tag{8}$$

It follows that the marginal PDF of Y_2 is

$$f_{Y_2}(y_2) = \begin{cases} 2y_2 & 0 \le y_2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Equations (1), (5), (6), and (9) imply

$$\mathrm{E}[Y_1] = \mathrm{E}[Y_3] = \int_0^1 2y(1-y)\, dy = 1/3, \tag{10}$$

$$\mathrm{E}[Y_2] = \mathrm{E}[Y_4] = \int_0^1 2y^2\, dy = 2/3. \tag{11}$$

Thus $\mathbf{Y}$ has expected value $\mathrm{E}[\mathbf{Y}] = \begin{bmatrix} 1/3 & 2/3 & 1/3 & 2/3 \end{bmatrix}'$. The second part of the problem is to find the correlation matrix $\mathbf{R_Y}$. In fact, we need to find $R_{\mathbf{Y}}(i,j) = \mathrm{E}[Y_iY_j]$ for each i,j pair. We will see that these are seriously tedious calculations. For $i = j$, the second moments are

$$\mathrm{E}\left[Y_1^2\right] = \mathrm{E}\left[Y_3^2\right] = \int_0^1 2y^2(1-y)\, dy = 1/6, \tag{12}$$

$$\mathrm{E}\left[Y_2^2\right] = \mathrm{E}\left[Y_4^2\right] = \int_0^1 2y^3\, dy = 1/2. \tag{13}$$

In terms of the correlation matrix,

$$R_{\mathbf{Y}}(1,1) = R_{\mathbf{Y}}(3,3) = 1/6, \qquad R_{\mathbf{Y}}(2,2) = R_{\mathbf{Y}}(4,4) = 1/2. \tag{14}$$

To find the off-diagonal terms $R_{\mathbf{Y}}(i,j) = \mathrm{E}[Y_iY_j]$, we need to find the marginal PDFs $f_{Y_i,Y_j}(y_i,y_j)$. Example 5.23 showed that

$$f_{Y_1,Y_4}(y_1,y_4) = \begin{cases} 4(1-y_1)y_4 & 0 \le y_1 \le 1, 0 \le y_4 \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

$$f_{Y_2,Y_3}(y_2,y_3) = \begin{cases} 4y_2(1-y_3) & 0 \le y_2 \le 1, 0 \le y_3 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Inspection will show that Y_1 and Y_4 are independent since $f_{Y_1,Y_4}(y_1,y_4) = f_{Y_1}(y_1)f_{Y_4}(y_4)$ Similarly, Y_2 and Y_3 are independent since $f_{Y_2,Y_3}(y_2,y_3) = f_{Y_2}(y_2)f_{Y_3}(y_3)$. This implies

$$R_{\mathbf{Y}}(1,4) = \mathrm{E}[Y_1Y_4] = \mathrm{E}[Y_1]\,\mathrm{E}[Y_4] = 2/9, \tag{17}$$

$$R_{\mathbf{Y}}(2,3) = \mathrm{E}[Y_2Y_3] = \mathrm{E}[Y_2]\,\mathrm{E}[Y_3] = 2/9. \tag{18}$$

We also need to calculate the marginal PDFs

$$f_{Y_1,Y_2}(y_1,y_2),\ f_{Y_3,Y_4}(y_3,y_4),\ f_{Y_1,Y_3}(y_1,y_3),\ \text{and } f_{Y_2,Y_4}(y_2,y_4).$$

To start, for $0 \le y_1 \le y_2 \le 1$,

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_1,Y_2,Y_3,Y_4}(y_1,y_2,y_3,y_4)\; dy_3\,dy_4 \\ &= \int_0^1\int_0^{y_4} 4\,dy_3\,dy_4 = \int_0^1 4y_4\,dy_4 = 2. \end{aligned} \tag{19}$$

Similarly, for $0 \le y_3 \le y_4 \le 1$,

$$\begin{aligned} f_{Y_3,Y_4}(y_3,y_4) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_1,Y_2,Y_3,Y_4}(y_1,y_2,y_3,y_4)\; dy_1\,dy_2 \\ &= \int_0^1\int_0^{y_2} 4\,dy_1\,dy_2 = \int_0^1 4y_2\,dy_2 = 2. \end{aligned} \tag{20}$$

In fact, these PDFs are the same in that

$$f_{Y_1,Y_2}(x,y) = f_{Y_3,Y_4}(x,y) = \begin{cases} 2 & 0 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

This implies $R_{\mathbf{Y}}(1,2) = R_{\mathbf{Y}}(3,4) = \mathrm{E}[Y_3 Y_4]$ and that

$$\mathrm{E}\left[Y_3 Y_4\right] = \int_0^1 \int_0^y 2xy \, dx \, dy = \int_0^1 \left(yx^2 \big|_0^y \right) \, dy = \int_0^1 y^3 \, dy = \frac{1}{4}. \tag{22}$$

Continuing in the same way, we see for $0 \le y_1 \le 1$ and $0 \le y_3 \le 1$ that

$$\begin{aligned} f_{Y_1,Y_3}(y_1, y_3) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Y_1,Y_2,Y_3,Y_4}(y_1, y_2, y_3, y_4) \, dy_2 \, dy_4 \\ &= 4 \left(\int_{y_1}^1 dy_2 \right) \left(\int_{y_3}^1 dy_4 \right) \\ &= 4(1-y_1)(1-y_3). \end{aligned} \tag{23}$$

We observe that Y_1 and Y_3 are independent since $f_{Y_1,Y_3}(y_1, y_3) = f_{Y_1}(y_1) f_{Y_3}(y_3)$. It follows that

$$R_{\mathbf{Y}}(1,3) = \mathrm{E}\left[Y_1 Y_3\right] = \mathrm{E}\left[Y_1\right] \mathrm{E}\left[Y_3\right] = 1/9. \tag{24}$$

Finally, we need to calculate

$$\begin{aligned} f_{Y_2,Y_4}(y_2, y_4) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Y_1,Y_2,Y_3,Y_4}(y_1, y_2, y_3, y_4) \, dy_1 \, dy_3 \\ &= 4 \left(\int_0^{y_2} dy_1 \right) \left(\int_0^{y_4} dy_3 \right) \\ &= 4 y_2 y_4. \end{aligned} \tag{25}$$

We observe that Y_2 and Y_4 are independent since $f_{Y_2,Y_4}(y_2, y_4) = f_{Y_2}(y_2) f_{Y_4}(y_4)$. It follows that $R_{\mathbf{Y}}(2,4) = \mathrm{E}[Y_2 Y_4] = \mathrm{E}[Y_2]\,\mathrm{E}[Y_4] = 4/9$. The above results give $R_{\mathbf{Y}}(i,j)$ for $i \le j$. Since $\mathbf{R_Y}$ is a symmetric matrix,

$$\mathbf{R_Y} = \begin{bmatrix} 1/6 & 1/4 & 1/9 & 2/9 \\ 1/4 & 1/2 & 2/9 & 4/9 \\ 1/9 & 2/9 & 1/6 & 1/4 \\ 2/9 & 4/9 & 1/4 & 1/2 \end{bmatrix}. \tag{26}$$

Since $\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix}1/3 & 2/3 & 1/3 & 2/3\end{bmatrix}'$, the covariance matrix is

$$\begin{aligned}
\mathbf{C}_{\mathbf{Y}} &= \mathbf{R}_{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{X}}' \\
&= \begin{bmatrix}1/6 & 1/4 & 1/9 & 2/9\\ 1/4 & 1/2 & 2/9 & 4/9\\ 2/9 & 4/9 & 1/4 & 1/2\end{bmatrix} - \begin{bmatrix}1/3\\ 2/3\\ 1/3\\ 2/3\end{bmatrix}\begin{bmatrix}1/3 & 2/3 & 1/3 & 2/3\end{bmatrix} \\
&= \begin{bmatrix}1/18 & 1/36 & 0 & 0\\ 1/36 & 1/18 & 0 & 0\\ 0 & 0 & 1/18 & 1/36\\ 0 & 0 & 1/36 & 1/18\end{bmatrix}.
\end{aligned} \tag{27}$$

The off-diagonal zero blocks are a consequence of $\begin{bmatrix}Y_1 & Y_2\end{bmatrix}'$ being independent of $\begin{bmatrix}Y_3 & Y_4\end{bmatrix}'$. Along the diagonal, the two identical sub-blocks occur because $f_{Y_1,Y_2}(x,y) = f_{Y_3,Y_4}(x,y)$. In short, the matrix structure is the result of $\begin{bmatrix}Y_1 & Y_2\end{bmatrix}'$ and $\begin{bmatrix}Y_3 & Y_4\end{bmatrix}'$ being iid random vectors.

Problem 8.4.10 Solution

Since $\mathrm{Var}[X_1] = \mathrm{Var}[X_2] = 1$, and since $\mathrm{E}[X_1] = \mathrm{E}[X_2] = 1$,

$$\rho = \mathrm{Cov}\left[X_1, X_2\right] = \mathrm{E}\left[X_1 X_2\right]. \tag{1}$$

This implies

$$\begin{aligned}
\mathrm{E}\left[Y^2\right] &= \omega\,\mathrm{E}\left[X_1^2\right] + 2\sqrt{\omega(1-\omega)}\,\mathrm{E}\left[X_1X_2\right] + (1-\omega)\,\mathrm{E}\left[X_2^2\right] \\
&= \omega + 2\rho\sqrt{\omega(1-\omega)} + (1-\omega) \\
&= 1 + 2\rho\sqrt{\omega(1-\omega)}.
\end{aligned} \tag{2}$$

If $\rho > 0$, then $\mathrm{E}[Y^2]$ is maximized at $\omega = 1/2$. If $\rho < 0$, then $\mathrm{E}[Y^2]$ is maximized at $\omega = 0$ and $\omega = 1$. If $\rho = 0$, then $\mathrm{E}[Y^2]$ is the same for all ω.

Problem 8.4.11 Solution

The 2-dimensional random vector $\mathbf{Y}$ has PDF

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} 2 & \mathbf{y} \geq \mathbf{0}, \begin{bmatrix}1 & 1\end{bmatrix}\mathbf{y} \leq 1, \\ 0 & \text{otherwise.}\end{cases} \tag{1}$$

Rewritten in terms of the variables y_1 and y_2,

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} 2 & y_1 \geq 0, y_2 \geq 0, y_1 + y_2 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In this problem, the PDF is simple enough that we can compute $\mathrm{E}[Y_i^n]$ for arbitrary integers $n \geq 0$.

$$\mathrm{E}\left[Y_1^n\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1^n f_{Y_1,Y_2}(y_1, y_2)\; dy_1\, dy_2 = \int_0^1 \int_0^{1-y_2} 2y_1^n\, dy_1\, dy_2. \tag{3}$$

A little calculus yields

$$\begin{aligned} \mathrm{E}\left[Y_1^n\right] &= \int_0^1 \left(\frac{2}{n+1} y_1^{n+1} \Big|_0^{1-y_2} \right) dy_2 \\ &= \frac{2}{n+1} \int_0^1 (1 - y_2)^{n+1}\, dy_2 = \frac{2}{(n+1)(n+2)}. \end{aligned} \tag{4}$$

Symmetry of the joint PDF $f_{Y_{1,2}}(y_{1,2})$ implies that $\mathrm{E}[Y_2^n] = \mathrm{E}[Y_1^n]$. Thus, $\mathrm{E}[Y_1] = \mathrm{E}[Y_2] = 1/3$ and

$$\mathrm{E}\left[\mathbf{Y}\right] = \boldsymbol{\mu}_{\mathbf{Y}} = \begin{bmatrix} 1/3 & 1/3 \end{bmatrix}'. \tag{5}$$

In addition,

$$R_{\mathbf{Y}}(1, 1) = \mathrm{E}\left[Y_1^2\right] = 1/6, \qquad R_{\mathbf{Y}}(2, 2) = \mathrm{E}\left[Y_2^2\right] = 1/6. \tag{6}$$

To complete the correlation matrix, we find

$$\begin{aligned} R_{\mathbf{Y}}(1, 2) &= \mathrm{E}\left[Y_1 Y_2\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1,Y_2}(y_1, y_2)\; dy_1\, dy_2 \\ &= \int_0^1 \int_0^{1-y_2} 2y_1 y_2\, dy_1\, dy_2. \end{aligned} \tag{7}$$

Following through on the calculus, we obtain

$$\begin{aligned} R_{\mathbf{Y}}(1,2) &= \int_0^1 \left(y_1^2 \Big|_0^{1-y-2} \right) y_2 \, dy_2 \\ &= \int_0^1 y_2(1-y_2)^2 \, dy_2 \\ &= \frac{1}{2}y_2^2 - \frac{2}{3}y_2^3 + \frac{1}{4}y_2^4 \Big|_0^1 = \frac{1}{12}. \end{aligned} \tag{8}$$

Thus we have found that

$$\mathbf{R_Y} = \begin{bmatrix} \mathrm{E}\left[Y_1^2\right] & \mathrm{E}\left[Y_1 Y_2\right] \\ \mathrm{E}\left[Y_2 Y_1\right] & \mathrm{E}\left[Y_2^2\right] \end{bmatrix} = \begin{bmatrix} 1/6 & 1/12 \\ 1/12 & 1/6 \end{bmatrix}. \tag{9}$$

Lastly, $\mathbf{Y}$ has covariance matrix

$$\begin{aligned} \mathbf{C_Y} = \mathbf{R_Y} - \boldsymbol{\mu}_\mathbf{Y}\boldsymbol{\mu}_\mathbf{Y}' &= \begin{bmatrix} 1/6 & 1/12 \\ 1/12 & 1/6 \end{bmatrix} - \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 \end{bmatrix} \\ &= \begin{bmatrix} 1/9 & -1/36 \\ -1/36 & 1/9 \end{bmatrix}. \end{aligned} \tag{10}$$

Problem 8.4.12 Solution

Given an arbitrary random vector $\mathbf{X}$, we can define $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}_\mathbf{X}$ so that

$$\mathbf{C_X} = \mathrm{E}\left[(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})'\right] = \mathrm{E}\left[\mathbf{Y}\mathbf{Y}'\right] = \mathbf{R_Y}. \tag{1}$$

It follows that the covariance matrix $\mathbf{C_X}$ is positive semi-definite if and only if the correlation matrix $\mathbf{R_Y}$ is positive semi-definite. Thus, it is sufficient to show that every correlation matrix, whether it is denoted $\mathbf{R_Y}$ or $\mathbf{R_X}$, is positive semi-definite.

To show a correlation matrix $\mathbf{R_X}$ is positive semi-definite, we write

$$\mathbf{a}'\mathbf{R_X}\mathbf{a} = \mathbf{a}'\,\mathrm{E}\left[\mathbf{X}\mathbf{X}'\right]\mathbf{a} = \mathrm{E}\left[\mathbf{a}'\mathbf{X}\mathbf{X}'\mathbf{a}\right] = \mathrm{E}\left[(\mathbf{a}'\mathbf{X})(\mathbf{X}'\mathbf{a})\right] = \mathrm{E}\left[(\mathbf{a}'\mathbf{X})^2\right]. \tag{2}$$

We note that $W = \mathbf{a}'\mathbf{X}$ is a random variable. Since $\mathrm{E}[W^2] \geq 0$ for any random variable W,

$$\mathbf{a}'\mathbf{R_X}\mathbf{a} = \mathrm{E}\left[W^2\right] \geq 0. \tag{3}$$

Problem 8.5.1 Solution

(a) From Theorem 8.7, the correlation matrix of $\mathbf{X}$ is

$$\begin{aligned}\mathbf{R}_X &= \mathbf{C}_X + \boldsymbol{\mu}_X\boldsymbol{\mu}_X' \\ &= \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} + \begin{bmatrix} 4 \\ 8 \\ 6 \end{bmatrix} \begin{bmatrix} 4 & 8 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} + \begin{bmatrix} 16 & 32 & 24 \\ 32 & 64 & 48 \\ 24 & 48 & 36 \end{bmatrix} = \begin{bmatrix} 20 & 30 & 25 \\ 30 & 68 & 46 \\ 25 & 46 & 40 \end{bmatrix}. \end{aligned} \tag{1}$$

(b) Let $\mathbf{Y} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}'$. Since $\mathbf{Y}$ is a subset of the components of $\mathbf{X}$, it is a Gaussian random vector with expected velue vector

$$\boldsymbol{\mu}_Y = \begin{bmatrix} \mathrm{E}[X_1] & \mathrm{E}[X_2] \end{bmatrix}' = \begin{bmatrix} 4 & 8 \end{bmatrix}'. \tag{2}$$

and covariance matrix

$$\mathbf{C}_Y = \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] \\ \mathbf{C}_{X_1 X_2} & \mathrm{Var}[X_2] \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}. \tag{3}$$

We note that $\det(\mathbf{C}_Y) = 12$ and that

$$\mathbf{C}_Y^{-1} = \frac{1}{12} \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/6 \\ 1/6 & 1/3 \end{bmatrix}. \tag{4}$$

This implies that

$$\begin{aligned}(\mathbf{y} - \boldsymbol{\mu}_Y)'\mathbf{C}_Y^{-1}(\mathbf{y} - \boldsymbol{\mu}_Y) &= \begin{bmatrix} y_1 - 4 & y_2 - 8 \end{bmatrix} \begin{bmatrix} 1/3 & 1/6 \\ 1/6 & 1/3 \end{bmatrix} \begin{bmatrix} y_1 - 4 \\ y_2 - 8 \end{bmatrix} \\ &= \begin{bmatrix} y_1 - 4 & y_2 - 8 \end{bmatrix} \begin{bmatrix} y_1/3 + y_2/6 - 8/3 \\ y_1/6 + y_2/3 - 10/3 \end{bmatrix} \\ &= \frac{y_1^2}{3} + \frac{y_1 y_2}{3} - \frac{16 y_1}{3} - \frac{20 y_2}{3} + \frac{y_2^2}{3} + \frac{112}{3}. \end{aligned} \tag{5}$$

The PDF of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi\sqrt{12}} e^{-(\mathbf{y}-\boldsymbol{\mu}_Y)'\mathbf{C}_Y^{-1}(\mathbf{y}-\boldsymbol{\mu}_Y)/2}$$
$$= \frac{1}{\sqrt{48\pi^2}} e^{-(y_1^2+y_1y_2-16y_1-20y_2+y_2^2+112)/6} \tag{6}$$

Since $\mathbf{Y} = \left[X_1, X_2\right]'$, the PDF of X_1 and X_2 is simply

$$f_{X_1,X_2}(x_1,x_2) = f_{Y_1,Y_2}(x_1,x_2)$$
$$= \frac{1}{\sqrt{48\pi^2}} e^{-(x_1^2+x_1x_2-16x_1-20x_2+x_2^2+112)/6}. \tag{7}$$

(c) We can observe directly from $\boldsymbol{\mu}_X$ and $\mathbf{C}_X$ that X_1 is a Gaussian $(4,2)$ random variable. Thus,

$$\mathrm{P}\left[X_1 > 8\right] = \mathrm{P}\left[\frac{X_1-4}{2} > \frac{8-4}{2}\right] = Q(2) = 0.0228. \tag{8}$$

Problem 8.5.2 Solution

Since $Y = 2X_1 + X_2$ is Gaussian, all we need to do s calculate its mean and variance and then write down the Gaussian PDF. Since $\mathrm{E}[Y] = 2\,\mathrm{E}[X_1] + \mathrm{E}[X_2] = 0$, Y has variance

$$\begin{aligned}
\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] &= \mathrm{E}\left[(2X_1+X_2)^2\right] \\
&= \mathrm{E}\left[4X_1^2 + 4X_1X_2 + X_2^2\right] \\
&= \mathrm{Var}\left[X_1\right] + 4\,\mathrm{Cov}\left[X_1,X_2\right] + \mathrm{Var}\left[X_2\right] \\
&= 4 + 4(1) + 2 = 10.
\end{aligned} \tag{1}$$

Thus Y has PDF

$$f_Y(y) = \frac{1}{\sqrt{20\pi} e^{-y^2/20}}. \tag{2}$$

Note that if we write $Y = \mathbf{a}'\mathbf{X}$, we can use the fact that Y is a scalar and so $Y = Y'$ to calculate its variance as

$$\begin{aligned}
\mathrm{Var}[Y] = \mathrm{E}\left[YY'\right] &= \mathrm{E}\left[\mathbf{a}'\mathbf{X}\mathbf{X}'\mathbf{a}\right] \\
&= \mathbf{a}'\,\mathrm{E}\left[\mathbf{X}\mathbf{X}'\right]\mathbf{a} = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 10.
\end{aligned} \tag{3}$$

Problem 8.5.3 Solution

We are given that $\mathbf{X}$ is a Gaussian random vector with

$$\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} 4 \\ 8 \\ 6 \end{bmatrix}, \qquad \mathbf{C}_{\mathbf{X}} = \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix}. \tag{1}$$

We are also given that $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} 1 & 1/2 & 2/3 \\ 1 & -1/2 & 2/3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}. \tag{2}$$

Since the two rows of $\mathbf{A}$ are linearly independent row vectors, $\mathbf{A}$ has rank 2. By Theorem 8.11, $\mathbf{Y}$ is a Gaussian random vector. Given these facts, the various parts of this problem are just straightforward calculations using Theorem 8.11.

(a) The expected value of $\mathbf{Y}$ is

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{Y}} &= \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{b} \\ &= \begin{bmatrix} 1 & 1/2 & 2/3 \\ 1 & -1/2 & 2/3 \end{bmatrix} \begin{bmatrix} 4 \\ 8 \\ 6 \end{bmatrix} + \begin{bmatrix} -4 \\ -4 \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \end{bmatrix}. \end{aligned} \tag{3}$$

(b) The covariance matrix of $\mathbf{Y}$ is

$$\begin{aligned} \mathbf{C}_{\mathbf{Y}} &= \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A} \\ &= \begin{bmatrix} 1 & 1/2 & 2/3 \\ 1 & -1/2 & 2/3 \end{bmatrix} \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1/2 & -1/2 \\ 2/3 & 2/3 \end{bmatrix} \\ &= \frac{1}{9} \begin{bmatrix} 43 & 55 \\ 55 & 103 \end{bmatrix}. \end{aligned} \tag{4}$$

(c) $\mathbf{Y}$ has correlation matrix

$$\begin{aligned} \mathbf{R}_{\mathbf{Y}} &= \mathbf{C}_{\mathbf{Y}} + \boldsymbol{\mu}_{\mathbf{Y}}\boldsymbol{\mu}_{\mathbf{Y}}' \\ &= \frac{1}{9} \begin{bmatrix} 43 & 55 \\ 55 & 103 \end{bmatrix} + \begin{bmatrix} 8 \\ 0 \end{bmatrix} \begin{bmatrix} 8 & 0 \end{bmatrix} \\ &= \frac{1}{9} \begin{bmatrix} 619 & 55 \\ 55 & 103 \end{bmatrix}. \end{aligned} \tag{5}$$

(d) From $\boldsymbol{\mu}_{\mathbf{Y}}$, we see that $\mathrm{E}[Y_2] = 0$. From the covariance matrix $\mathbf{C}_{\mathbf{Y}}$, we learn that Y_2 has variance $\sigma_2^2 = C_{\mathbf{Y}}(2,2) = 103/9$. Since Y_2 is a Gaussian random variable,

$$\begin{aligned} \mathrm{P}\left[-1 \leq Y_2 \leq 1\right] &= \mathrm{P}\left[-\frac{1}{\sigma_2} \leq \frac{Y_2}{\sigma_2} \leq \frac{1}{\sigma_2}\right] \\ &= \Phi\left(\frac{1}{\sigma_2}\right) - \Phi\left(\frac{-1}{\sigma_2}\right) \\ &= 2\Phi\left(\frac{1}{\sigma_2}\right) - 1 \\ &= 2\Phi\left(\frac{3}{\sqrt{103}}\right) - 1 = 0.2325. \end{aligned} \tag{6}$$

Problem 8.5.4 Solution

This problem is just a special case of Theorem 8.11 with the matrix $\mathbf{A}$ replaced by the row vector $\mathbf{a}'$ and a 1 element vector $\mathbf{b} = b = 0$. In this case, the vector $\mathbf{Y}$ becomes the scalar Y. The expected value vector $\boldsymbol{\mu}_Y = [\mu_Y]$ and the covariance "matrix" of Y is just the 1×1 matrix $[\sigma_Y^2]$. Directly from Theorem 8.11, we can conclude that Y is a length 1 Gaussian random vector, which is just a Gaussian random variable. In addition, $\mu_Y = \mathbf{a}'\boldsymbol{\mu}_{\mathbf{X}}$ and

$$\mathrm{Var}[Y] = \mathbf{C}_Y = \mathbf{a}'\mathbf{C}_{\mathbf{X}}\mathbf{a}. \tag{1}$$

Problem 8.5.5 Solution

(a) $\mathbf{C}$ must be symmetric since

$$\alpha = \beta = \mathrm{E}\left[X_1 X_2\right]. \tag{1}$$

In addition, α must be chosen so that $\mathbf{C}$ is positive semi-definite. Since the characteristic equation is

$$\begin{aligned} \det\left(\mathbf{C} - \lambda\mathbf{I}\right) &= (1-\lambda)(4-\lambda) - \alpha^2 \\ &= \lambda^2 - 5\lambda + 4 - \alpha^2 = 0, \end{aligned} \tag{2}$$

the eigenvalues of $\mathbf{C}$ are

$$\lambda_{1,2} = \frac{5 \pm \sqrt{25 - 4(4 - \alpha^2)}}{2}. \tag{3}$$

The eigenvalues are non-negative as long as $\alpha^2 \leq 4$, or $|\alpha| \leq 2$. Another way to reach this conclusion is through the requirement that $|\rho_{X_1 X_2}| \leq 1$.

(b) It remains true that $\alpha = \beta$ and $\mathbf{C}$ must be positive semi-definite. For $\mathbf{X}$ to be a Gaussian vector, $\mathbf{C}$ also must be positive definite. For the eigenvalues of $\mathbf{C}$ to be strictly positive, we must have $|\alpha| < 2$.

(c) Since $\mathbf{X}$ is a Gaussian vector, W is a Gaussian random variable. Thus, we need only calculate

$$\mathrm{E}[W] = 2\,\mathrm{E}[X_1] - \mathrm{E}[X_2] = 0, \tag{4}$$

and

$$\begin{aligned} \mathrm{Var}[W] = \mathrm{E}\left[W^2\right] &= \mathrm{E}\left[4X_1^2 - 4X_1X_2 + X_2^2\right] \\ &= 4\,\mathrm{Var}[X_1] - 4\,\mathrm{Cov}[X_1, X_2] + \mathrm{Var}[X_2] \\ &= 4 - 4\alpha + 4 = 4(2 - \alpha). \end{aligned} \tag{5}$$

The PDF of W is

$$f_W(w) = \frac{1}{\sqrt{8(2-\alpha)\pi}} e^{-w^2/8(2-\alpha)}. \tag{6}$$

Problem 8.5.6 Solution

(a) Since the matrix is always symmetric, we need to make sure that $\mathbf{C}_X$ is positive definite. From first principles, we can find the eigenvalues of $\mathbf{C}$ by solving

$$\det(\mathbf{C} - \lambda \mathbf{I}) = (\sigma_1^2 - \lambda)(\sigma_2^2 - \lambda) - 1 = 0. \tag{1}$$

Equivalently

$$\lambda^2 - (\sigma_1^2 + \sigma_2^2)\lambda + (\sigma_1^2\sigma_2^2 - 1) = 0. \tag{2}$$

By the quadratic formula,

$$\lambda = \frac{\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4(\sigma_1^2\sigma_2^2 - 1)}}{2}. \tag{3}$$

Some algebra yields

$$\lambda = \frac{\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4}}{2}. \tag{4}$$

For both roots to be strictly positive, we must have

$$(\sigma_1^2 + \sigma_2^2)^2 > (\sigma_1^2 - \sigma_2^2)^2 + 4, \tag{5}$$

which occurs if and only if $\sigma_1^2\sigma_2^2 > 1$.

Note you get this same answer if you check the requirement that the correlation coefficient of X_1 and X_2 satisfies $\rho^2 < 1$.

(b) Y_1 and Y_2 are jointly Gaussian, so they are independent if they have zero covariance. To check the covariance, we calculate the covariance matrix

$$\begin{aligned}\mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}' &= \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\sigma_1^2 & 1\\ 1 & \sigma_2^2\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\\ &= \begin{bmatrix}\sigma_1^2 + \sigma_2^2 + 2 & \sigma_1^2 - \sigma_2^2\\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 - 2\end{bmatrix}.\end{aligned} \tag{6}$$

Thus Y_1 and Y_2 are independent if $[\mathbf{C}_Y]_{12} = \sigma_1^2 - \sigma_2^2 = 0$.

Problem 8.5.7 Solution

(a) Since $\mathbf{X}$ is Gaussian, W is also Gaussian. Thus we need only compute the expected value

$$\mathrm{E}[W] = \mathrm{E}[X_1] + 2\,\mathrm{E}[X_2] = 0$$

and variance

$$\begin{aligned}\operatorname{Var}[W] = \mathrm{E}\left[W^2\right] &= \mathrm{E}\left[(X_1 + 2X_2)^2\right] \\ &= \mathrm{E}\left[X_1^2 + 4X_1X_2 + 4X_2^2\right] \\ &= C_{11} + 4C_{12} + 4C_{22} = 10. \end{aligned} \tag{1}$$

Thus W has the Gaussian $(0, \sqrt{10})$ PDF

$$f_W(w) = \frac{1}{\sqrt{20\pi}} e^{-w^2/20}.$$

(b) We first calculate

$$\mathrm{E}[V] = 0, \qquad \operatorname{Var}[V] = 4\operatorname{Var}[X_1] = 8, \tag{2}$$

$$\mathrm{E}[W] = 0, \qquad \operatorname{Var}[W] = 10, \tag{3}$$

and that V and W have correlation coefficient

$$\begin{aligned}\rho_{VW} &= \frac{\mathrm{E}[VW]}{\sqrt{\operatorname{Var}[V]\operatorname{Var}[W]}} \\ &= \frac{\mathrm{E}[2X_1(X_1 + 2X_2)]}{\sqrt{80}} \\ &= \frac{2C_{11} + 4C_{12}}{\sqrt{80}} = \frac{8}{\sqrt{80}} = \frac{2}{\sqrt{5}}. \end{aligned} \tag{4}$$

Now we recall that the conditional PDF $f_{V|W}(v|w)$ is Gaussian with conditional expected value

$$\begin{aligned}\mathrm{E}[V|W = w] &= \mathrm{E}[V] + \rho_{VW}\frac{\sigma_V}{\sigma_W}(w - \mathrm{E}[W]) \\ &= \frac{2}{\sqrt{5}}\frac{\sqrt{8}}{\sqrt{10}} w = 4w/5 \end{aligned} \tag{5}$$

and conditional variance

$$\operatorname{Var}[V|W] = \operatorname{Var}[V](1 - \rho_{VW}^2) = \frac{8}{5}. \tag{6}$$

It follows that

$$
\begin{aligned}
f_{V|W}(v|w) &= \frac{1}{\sqrt{2\pi \operatorname{Var}[V|W]}} e^{-(v-\mathrm{E}[V|W])^2/2\operatorname{Var}[V|W]} \\
&= \sqrt{\frac{5}{16\pi}} e^{-5(v-4w/5)^2/16}. \qquad (7)
\end{aligned}
$$

Problem 8.5.8 Solution

From Definition 8.12, the $n = 2$ dimensional Gaussian vector $\mathbf{X}$ has PDF

$$
f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi[\det(\mathbf{C_X})]^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})'\mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})\right) \qquad (1)
$$

where $\mathbf{C_X}$ has determinant

$$
\det(\mathbf{C_X}) = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2). \qquad (2)
$$

Thus,

$$
\frac{1}{2\pi[\det(\mathbf{C_X})]^{1/2}} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}. \qquad (3)
$$

Using the 2×2 matrix inverse formula

$$
\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \qquad (4)
$$

we obtain

$$
\begin{aligned}
\mathbf{C}_{\mathbf{X}}^{-1} &= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \\
&= \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}. \qquad (5)
\end{aligned}
$$

Thus

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})'\mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}}) = -\frac{\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2}\\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\end{bmatrix}\begin{bmatrix} x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}}{2(1-\rho^2)}$$

$$= -\frac{\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2\end{bmatrix}\begin{bmatrix}\frac{x_1-\mu_1}{\sigma_1^2} - \frac{\rho(x_2-\mu_2)}{\sigma_1\sigma_2}\\ -\frac{\rho(x_1-\mu_1)}{\sigma_1\sigma_2} + \frac{x_2-\mu_2}{\sigma_2^2}\end{bmatrix}}{2(1-\rho^2)}$$

$$= -\frac{\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)}. \tag{6}$$

Combining Equations (1), (3), and (6), we see that

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)}\right], \tag{7}$$

which is the bivariate Gaussian PDF in Definition 5.10.

Problem 8.5.9 Solution

(a) First $b = c$ since a covariance matrix is always symmetric. Second, $a = \text{Var}[X_1]$ and $b = \text{Var}[X_2]$. Hence we must have $a > 0$ and $d > 0$. Third, $\mathbf{C}$ must be positive definite, i.e. the eigenvalues of $\mathbf{C}$ must be positive. This can be tackled directly from first principles by solving for the eigenvalues using $\det(()\mathbf{C}-\lambda\mathbf{I}) = 0$. If you do this, you will find, after some algebra that the eigenvalues are

$$\lambda = \frac{(a+d) \pm \sqrt{(a-d)^2 + 4b^2}}{2}. \tag{1}$$

The requirement $\lambda > 0$ holds iff $b^2 < ad$. As it happens, this is precisely the same conditon as requiring the correlation coefficient to have magnitude less than 1:

$$|\rho_{X_1X_2}| = \left|\frac{b}{\sqrt{ad}}\right| < 1. \tag{2}$$

To summarize, there are four requirements:

$$a > 0, \qquad d > 0, \qquad b = c, \qquad b^2 < ad. \tag{3}$$

(b) This is easy: for Gaussian random variables, zero covariance implies X_1 and X_2 are independent. Hence the answer is $b = 0$.

(c) X_1 and X_2 are identical if they have the same variance: $a = d$.

Problem 8.5.10 Solution

Suppose that $\mathbf{X}$ Gaussian $(\mathbf{0}, \mathbf{I})$ random vector. Since

$$\mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A} \end{bmatrix} \mathbf{X} = \mathbf{D}\mathbf{X}. \tag{1}$$

Theorem 8.8 implies $\boldsymbol{\mu}_{\mathbf{W}} = \mathbf{D}\,\mathrm{E}[\mathbf{X}] = \mathbf{0}$ and $\mathbf{C}_{\mathbf{W}} = \mathbf{D}\mathbf{D}'$. The matrix $\mathbf{D}$ is $(m+n) \times n$ and has rank n. That is, the rows of $\mathbf{D}$ are dependent and there exists a vector $\mathbf{y}$ such that $\mathbf{y}'\mathbf{D} = \mathbf{0}$. This implies $\mathbf{y}'\mathbf{D}\mathbf{D}'\mathbf{y} = 0$. Hence $\det(\mathbf{C}_{\mathbf{W}}) = 0$ and $\mathbf{C}_{\mathbf{W}}^{-1}$ does not exist. Hence $\mathbf{W}$ is **not** a Gaussian random vector.

The point to keep in mind is that the definition of a Gaussian random vector does not permit a component random variable to be a deterministic linear combination of other components.

Problem 8.5.11 Solution

(a) From Theorem 8.8, $\mathbf{Y}$ has covariance matrix

$$\begin{aligned} \mathbf{C}_{\mathbf{Y}} &= \mathbf{Q}\mathbf{C}_{\mathbf{X}}\mathbf{Q}' \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2\cos^2\theta + \sigma_2^2\sin^2\theta & (\sigma_1^2 - \sigma_2^2)\sin\theta\cos\theta \\ (\sigma_1^2 - \sigma_2^2)\sin\theta\cos\theta & \sigma_1^2\sin^2\theta + \sigma_2^2\cos^2\theta \end{bmatrix}. \end{aligned} \tag{1}$$

We conclude that Y_1 and Y_2 have covariance

$$\mathrm{Cov}\left[Y_1, Y_2\right] = C_{\mathbf{Y}}(1,2) = (\sigma_1^2 - \sigma_2^2)\sin\theta\cos\theta. \tag{2}$$

Since Y_1 and Y_2 are jointly Gaussian, they are independent if and only if $\text{Cov}[Y_1, Y_2] = 0$. Thus, Y_1 and Y_2 are independent for all θ if and only if $\sigma_1^2 = \sigma_2^2$. In this case, when the joint PDF $f_{\mathbf{X}}(\mathbf{x})$ is symmetric in x_1 and x_2. In terms of polar coordinates, the PDF $f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,X_2}(x_1, x_2)$ depends on $r = \sqrt{x_1^2 + x_2^2}$ but for a given r, is constant for all $\phi = \tan^{-1}(x_2/x_1)$. The transformation of $\mathbf{X}$ to $\mathbf{Y}$ is just a rotation of the coordinate system by θ preserves this circular symmetry.

(b) If $\sigma_2^2 > \sigma_1^2$, then Y_1 and Y_2 are independent if and only if $\sin\theta\cos\theta = 0$. This occurs in the following cases:

- $\theta = 0$: $Y_1 = X_1$ and $Y_2 = X_2$
- $\theta = \pi/2$: $Y_1 = -X_2$ and $Y_2 = -X_1$
- $\theta = \pi$: $Y_1 = -X_1$ and $Y_2 = -X_2$
- $\theta = -\pi/2$: $Y_1 = X_2$ and $Y_2 = X_1$

In all four cases, Y_1 and Y_2 are just relabeled versions, possibly with sign changes, of X_1 and X_2. In these cases, Y_1 and Y_2 are independent because X_1 and X_2 are independent. For other values of θ, each Y_i is a linear combination of both X_1 and X_2. This mixing results in correlation between Y_1 and Y_2.

Problem 8.5.12 Solution

The difficulty of this problem is overrated since its a pretty simple application of Problem 8.5.11. In particular,

$$\mathbf{Q} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\Bigg|_{\theta=45^\circ} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \tag{1}$$

Since $\mathbf{X} = \mathbf{QY}$, we know from Theorem 8.11 that $\mathbf{X}$ is Gaussian with covariance matrix

$$
\begin{aligned}
\mathbf{C_X} &= \mathbf{Q}\mathbf{C_Y}\mathbf{Q}' \\
&= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+\rho & -(1-\rho) \\ 1+\rho & 1-\rho \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \qquad (2)
\end{aligned}
$$

Problem 8.5.13 Solution

As given in the problem statement, we define the m-dimensional vector $\mathbf{X}$, the n-dimensional vector $\mathbf{Y}$ and $\mathbf{W} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Y}' \end{bmatrix}'$. Note that $\mathbf{W}$ has expected value

$$
\boldsymbol{\mu}_{\mathbf{W}} = \mathrm{E}[\mathbf{W}] = \mathrm{E}\left[\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}\right] = \begin{bmatrix} \mathrm{E}[\mathbf{X}] \\ \mathrm{E}[\mathbf{Y}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix}. \qquad (1)
$$

The covariance matrix of $\mathbf{W}$ is

$$
\begin{aligned}
\mathbf{C_W} &= \mathrm{E}\left[(\mathbf{W}-\boldsymbol{\mu}_{\mathbf{W}})(\mathbf{W}-\boldsymbol{\mu}_{\mathbf{W}})'\right] \\
&= \mathrm{E}\left[\begin{bmatrix} \mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}} \\ \mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix}\begin{bmatrix} (\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})' & (\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})' \end{bmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})'] & \mathrm{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})'] \\ \mathrm{E}[(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})'] & \mathrm{E}[(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})'] \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{C_X} & \mathbf{C_{XY}} \\ \mathbf{C_{YX}} & \mathbf{C_Y} \end{bmatrix}. \qquad (2)
\end{aligned}
$$

The assumption that $\mathbf{X}$ and $\mathbf{Y}$ are independent implies that

$$
\mathbf{C_{XY}} = \mathrm{E}\left[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y}'-\boldsymbol{\mu}_{\mathbf{Y}}')\right] = (\mathrm{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})]\,\mathrm{E}\left[(\mathbf{Y}'-\boldsymbol{\mu}_{\mathbf{Y}}')\right] = \mathbf{0}. \qquad (3)
$$

This also implies $\mathbf{C_{YX}} = \mathbf{C}_{\mathbf{XY}}' = \mathbf{0}'$. Thus

$$
\mathbf{C_W} = \begin{bmatrix} \mathbf{C_X} & \mathbf{0} \\ \mathbf{0}' & \mathbf{C_Y} \end{bmatrix}. \qquad (4)
$$

Problem 8.5.14 Solution

(a) If you are familiar with the Gram-Schmidt procedure, the argument is that applying Gram-Schmidt to the rows of $\mathbf{A}$ yields m orthogonal row vectors. It is then possible to augment those vectors with an additional $n-m$ orothogonal vectors. Those orthogonal vectors would be the rows of $\tilde{\mathbf{A}}$.

An alternate argument is that since $\mathbf{A}$ has rank m the nullspace of $\mathbf{A}$, i.e., the set of all vectors $\mathbf{y}$ such that $\mathbf{A}\mathbf{y} = \mathbf{0}$ has dimension $n-m$. We can choose any $n-m$ linearly independent vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-m}$ in the nullspace $\mathbf{A}$. We then define $\tilde{\mathbf{A}}'$ to have columns $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-m}$. It follows that $\mathbf{A}\tilde{\mathbf{A}}' = \mathbf{0}$.

(b) To use Theorem 8.11 for the case $m = n$ to show

$$\bar{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \hat{\mathbf{Y}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \hat{\mathbf{A}} \end{bmatrix} \mathbf{X}. \tag{1}$$

is a Gaussian random vector requires us to show that

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \hat{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1} \end{bmatrix} \tag{2}$$

is a rank n matrix. To prove this fact, we will suppose there exists $\mathbf{w}$ such that $\bar{\mathbf{A}}\mathbf{w} = \mathbf{0}$, and then show that $\mathbf{w}$ is a zero vector. Since $\mathbf{A}$ and $\tilde{\mathbf{A}}$ together have n linearly independent rows, we can write the row vector $\mathbf{w}'$ as a linear combination of the rows of $\mathbf{A}$ and $\tilde{\mathbf{A}}$. That is, for some $\mathbf{v}$ and $\tilde{\mathbf{v}}$,

$$\mathbf{w}' = \mathbf{v}\mathbf{t}'\mathbf{A} + \tilde{\mathbf{v}}'\tilde{\mathbf{A}}. \tag{3}$$

The condition $\bar{\mathbf{A}}\mathbf{w} = \mathbf{0}$ implies

$$\begin{bmatrix} \mathbf{A} \\ \tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1} \end{bmatrix} \left(\mathbf{A}'\mathbf{v} + \tilde{\mathbf{A}}'\tilde{\mathbf{v}}' \right) = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{4}$$

This implies

$$\mathbf{A}\mathbf{A}'\mathbf{v} + \mathbf{A}\tilde{\mathbf{A}}'\tilde{\mathbf{v}} = \mathbf{0}, \tag{5}$$

$$\tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\mathbf{A}\mathbf{v} + \tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\tilde{\mathbf{v}} = \mathbf{0}. \tag{6}$$

Since $\mathbf{A}\tilde{\mathbf{A}}' = \mathbf{0}$, Equation (5) implies that $\mathbf{A}\mathbf{A}'\mathbf{v} = \mathbf{0}$. Since $\mathbf{A}$ is rank m, $\mathbf{A}\mathbf{A}'$ is an $m \times m$ rank m matrix. It follows that $\mathbf{v} = \mathbf{0}$. We can then conclude from Equation (6) that

$$\tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\tilde{\mathbf{v}} = \mathbf{0}. \tag{7}$$

This would imply that $\tilde{\mathbf{v}}'\tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\tilde{\mathbf{v}} = 0$. Since $\mathbf{C}_{\mathbf{X}}^{-1}$ is invertible, this would imply that $\tilde{\mathbf{A}}'\tilde{\mathbf{v}} = \mathbf{0}$. Since the rows of $\tilde{\mathbf{A}}$ are linearly independent, it must be that $\tilde{\mathbf{v}} = \mathbf{0}$. Thus $\bar{\mathbf{A}}$ is full rank and $\bar{\mathbf{Y}}$ is a Gaussian random vector.

(c) We note that By Theorem 8.11, the Gaussian vector $\bar{\mathbf{Y}} = \bar{\mathbf{A}}\mathbf{X}$ has covariance matrix

$$\bar{\mathbf{C}} = \bar{\mathbf{A}}\mathbf{C}_{\mathbf{X}}\bar{\mathbf{A}}'. \tag{8}$$

Since $(\mathbf{C}_{\mathbf{X}}^{-1})' = \mathbf{C}_{\mathbf{X}}^{-1}$,

$$\bar{\mathbf{A}}' = \begin{bmatrix}\mathbf{A}' & (\tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1})'\end{bmatrix} = \begin{bmatrix}\mathbf{A}' & \mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\end{bmatrix}. \tag{9}$$

Applying this result to Equation (8) yields

$$\begin{aligned}
\bar{\mathbf{C}} &= \begin{bmatrix}\mathbf{A} \\ \tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\end{bmatrix}\mathbf{C}_{\mathbf{X}}\begin{bmatrix}\mathbf{A}' & \mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{A}\mathbf{C}_{\mathbf{X}} \\ \tilde{\mathbf{A}}\end{bmatrix}\begin{bmatrix}\mathbf{A}' & \mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}' & \mathbf{A}\tilde{\mathbf{A}}' \\ \tilde{\mathbf{A}}\mathbf{A}' & \tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\end{bmatrix}.
\end{aligned} \tag{10}$$

Since $\tilde{\mathbf{A}}\mathbf{A}' = \mathbf{0}$,

$$\bar{\mathbf{C}} = \begin{bmatrix}\mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}' & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}\tilde{\mathbf{A}}'\end{bmatrix} = \begin{bmatrix}\mathbf{C}_{\mathbf{Y}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\hat{\mathbf{Y}}}\end{bmatrix}. \tag{11}$$

We see that $\bar{\mathbf{C}}$ is block diagonal covariance matrix. From the claim of Problem 8.5.13, we can conclude that $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are independent Gaussian random vectors.

Problem 8.6.1 Solution

We can use Theorem 8.11 since the scalar Y is also a 1-dimensional vector. To do so, we write

$$Y = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \mathbf{X} = \mathbf{A}\mathbf{X}. \tag{1}$$

By Theorem 8.11, Y is a Gaussian vector with expected value

$$\begin{aligned} \mathrm{E}[Y] = \mathbf{A}\boldsymbol{\mu}_X &= (\mathrm{E}[X_1] + \mathrm{E}[X_2] + \mathrm{E}[X_3])/3 \\ &= (4 + 8 + 6)/3 = 6. \end{aligned} \tag{2}$$

and covariance matrix

$$\begin{aligned} \mathbf{C}_Y = \mathrm{Var}[Y] &= \mathbf{A}\mathbf{C}_X\mathbf{A}' \\ &= \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \frac{2}{3}. \end{aligned} \tag{3}$$

Thus Y is a Gaussian $(6, \sqrt{2/3})$ random variable, implying

$$\begin{aligned} \mathrm{P}[Y > 4] &= \mathrm{P}\left[\frac{Y - 6}{\sqrt{2/3}} > \frac{4 - 6}{\sqrt{2/3}}\right] \\ &= 1 - \Phi(-\sqrt{6}) = \Phi(\sqrt{6}) = 0.9928. \end{aligned} \tag{4}$$

Problem 8.6.2 Solution

(a) The covariance matrix $\mathbf{C}_X$ has $\mathrm{Var}[X_i] = 25$ for each diagonal entry. For $i \neq j$, the i,jth entry of $\mathbf{C}_X$ is

$$\begin{aligned} [\mathbf{C}_X]_{ij} &= \rho_{X_iX_j}\sqrt{\mathrm{Var}[X_i]\,\mathrm{Var}[X_j]} \\ &= (0.8)(25) = 20 \end{aligned} \tag{1}$$

The covariance matrix of X is a 10×10 matrix of the form

$$\mathbf{C}_X = \begin{bmatrix} 25 & 20 & \cdots & 20 \\ 20 & 25 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 20 \\ 20 & \cdots & 20 & 25 \end{bmatrix}. \tag{2}$$

(b) We observe that

$$Y = \begin{bmatrix}1/10 & 1/10 & \cdots & 1/10\end{bmatrix} \mathbf{X} = \mathbf{A}\mathbf{X}. \tag{3}$$

Since Y is the average of 10 iid random variables, $\mathrm{E}[Y] = \mathrm{E}[X_i] = 5$. Since Y is a scalar, the 1×1 covariance matrix $\mathbf{C}_Y = \mathrm{Var}[Y]$. By Theorem 8.8, the variance of Y is

$$\mathrm{Var}[Y] = \mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}' = 20.5. \tag{4}$$

Since Y is Gaussian,

$$\mathrm{P}[Y \leq 25] = \mathrm{P}\left[\frac{Y-5}{\sqrt{20.5}} \leq \frac{25-20.5}{\sqrt{20.5}}\right] = \Phi(0.9939) = 0.8399. \tag{5}$$

Problem 8.6.3 Solution

Under the model of Quiz 8.6, the temperature on day i and on day j have covariance

$$\mathrm{Cov}[T_i, T_j] = C_T[i-j] = \frac{36}{1+|i-j|}. \tag{1}$$

From this model, the vector $\mathbf{T} = \begin{bmatrix}T_1 & \cdots & T_{31}\end{bmatrix}'$ has covariance matrix

$$\mathbf{C_T} = \begin{bmatrix} C_T[0] & C_T[1] & \cdots & C_T[30] \\ C_T[1] & C_T[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_T[1] \\ C_T[30] & \cdots & C_T[1] & C_T[0] \end{bmatrix}. \tag{2}$$

If you have read the solution to Quiz 8.6, you know that $\mathbf{C_T}$ is a symmetric Toeplitz matrix and that MATLAB has a `toeplitz` function to generate Toeplitz matrices. Using the `toeplitz` function to generate the covariance matrix, it is easy to use `gaussvector` to generate samples of the random vector $\mathbf{T}$.Here is the code for estimating $\mathrm{P}[A]$ using m samples.

```
function p=julytemp583(m);
c=36./(1+(0:30));
CT=toeplitz(c);
mu=80*ones(31,1);
T=gaussvector(mu,CT,m);
Y=sum(T)/31;
Tmin=min(T);
p=sum((Tmin>=72) & (Y <= 82))/m;
```

```
julytemp583(100000)
ans =
    0.0684
>> julytemp583(100000)
ans =
    0.0706
>> julytemp583(100000)
ans =
    0.0714
>> julytemp583(100000)
ans =
    0.0701
```

We see from repeated experiments with $m = 100{,}000$ trials that $P[A] \approx 0.07$.

Problem 8.6.4 Solution

The covariance matrix $\mathbf{C}_X$ has $\text{Var}[X_i] = 25$ for each diagonal entry. For $i \neq j$, the i, jth entry of $\mathbf{C}_X$ is

$$[\mathbf{C}_X]_{ij} = \rho_{X_i X_j} \sqrt{\text{Var}[X_i] \text{Var}[X_j]} = (0.8)(25) = 20 \tag{1}$$

The covariance matrix of X is a 10×10 matrix of the form

$$\mathbf{C}_X = \begin{bmatrix} 25 & 20 & \cdots & 20 \\ 20 & 25 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 20 \\ 20 & \cdots & 20 & 25 \end{bmatrix}. \tag{2}$$

A program to estimate $P[W \leq 25]$ uses `gaussvector` to generate m sample vector of race times $\mathbf{X}$. In the program `sailboats.m`, `X` is an $10 \times m$ matrix such that each column of `X` is a vector of race times. In addition `min(X)` is a row vector indicating the fastest time in each race.

```
function p=sailboats(w,m)
%Usage: p=sailboats(f,m)
%In Problem 5.8.4, W is the
%winning time in a 10 boat race.
%We use m trials to estimate
%P[W<=w]
CX=(5*eye(10))+(20*ones(10,10));
mu=35*ones(10,1);
X=gaussvector(mu,CX,m);
W=min(X);
p=sum(W<=w)/m;
```

```
>> sailboats(25,10000)
ans =
    0.0827
>> sailboats(25,100000)
ans =
    0.0801
>> sailboats(25,100000)
ans =
    0.0803
>> sailboats(25,100000)
ans =
    0.0798
```

We see $P[W \leq 25] \approx 0.08$ from repeated experiments with $m = 100{,}000$ trials.

Problem 8.6.5 Solution

When we built `poissonrv.m`, we went to some trouble to be able to generate m iid samples at once. In this problem, each Poisson random variable that we generate has an expected value that is different from that of any other Poisson random variables. Thus, we must generate the daily jackpots sequentially. Here is a simple program for this purpose.

```
function jackpot=lottery1(jstart,M,D)
%Usage: function j=lottery1(jstart,M,D)
%Perform M trials of the D day lottery
%of Problem 5.5.5 and initial jackpot jstart
jackpot=zeros(M,1);
for m=1:M,
    disp('trm)
jackpot(m)=jstart;
for d=1:D,
    jackpot(m)=jackpot(m)+(0.5*poissonrv(jackpot(m),1));
end
end
```

The main problem with `lottery1` is that it will run *very* slowly. Each call to `poissonrv` generates an entire Poisson PMF $P_X(x)$ for $x = 0, 1, \ldots, x_{\max}$ where $x_{\max} \geq 2 \cdot 10^6$. This is slow in several ways. First, we repeating the calculation of $\sum_{j=1}^{x_{\max}} \log j$ with each call to `poissonrv`. Second, each call to `poissonrv` asks for a Poisson sample value with expected value $\alpha > 1 \cdot 10^6$. In these cases, for small values of x, $P_X(x) = \alpha^x e^{-\alpha x}/x!$ is so small that it is less than the smallest nonzero number that MATLAB can store!

To speed up the simulation, we have written a program `bigpoissonrv` which generates Poisson (α) samples for large α. The program makes an approximation that for a Poisson (α) random variable X, $P_X(x) \approx 0$ for $|x-\alpha| > 6\sqrt{\alpha}$. Since X has standard deviation $\sqrt{\alpha}$, we are assuming that X cannot be more than six standard deviations away from its mean value. The error in this approximation is very small. In fact, for a Poisson (a) random variable, the program `poissonsigma(a,k)` calculates the error $\text{P}[|X - a| > k\sqrt{a}]$. Here is `poissonsigma.m` and some simple calculations:

```
function err=poissonsigma(a,k);
xmin=max(0,floor(a-k*sqrt(a)));
xmax=a+ceil(k*sqrt(a));
sx=xmin:xmax;
logfacts =cumsum([0,log(1:xmax)]);
  %logfacts includes 0 in case xmin=0
  %Now we extract needed values:
logfacts=logfacts(sx+1);
  %pmf(i,:) is a Poisson a(i) PMF
  %    from xmin to xmax
pmf=exp(-a+ (log(a)*sx)-(logfacts));
err=1-sum(pmf);
```

```
>> poissonsigma(1,6)
ans =
  1.0249e-005
>> poissonsigma(10,6)
ans =
  2.5100e-007
>> poissonsigma(100,6)
ans =
  1.2620e-008
>> poissonsigma(1000,6)
ans =
  2.6777e-009
>> poissonsigma(10000,6)
ans =
  1.8081e-009
>> poissonsigma(100000,6)
ans =
 -1.6383e-010
```

The error reported by `poissonsigma(a,k)` should always be positive. In fact, we observe negative errors for very large `a`.For large α and x, numerical calculation of $P_X(x) = \alpha^x e^{-\alpha}/x!$ is tricky because we are taking ratios of very large numbers. In fact, for $\alpha = x = 1{,}000{,}000$, MATLAB calculation of α^x and $x!$ will report infinity while $e^{-\alpha}$ will evaluate as zero. Our method of calculating the Poisson (α) PMF is to use the fact that $\ln x! = \sum_{j=1}^{x} \ln j$ to calculate

$$\exp\left(\ln P_X(x)\right) = \exp\left(x \ln \alpha - \alpha - \sum_{j=1}^{x} \ln j\right). \tag{1}$$

This method works reasonably well except that the calculation of the logarithm has finite precision. The consequence is that the calculated sum over the PMF can vary from 1 by a very small amount, on the order of 10^{-7} in our experiments. In our problem, the error is inconsequential, however, one should keep in mind that this may not be the case in other other experiments using large Poisson random variables. In any case, we can conclude that within the accuracy of MATLAB's simulated experiments, the approximations to be used by `bigpoissonrv` are not significant.

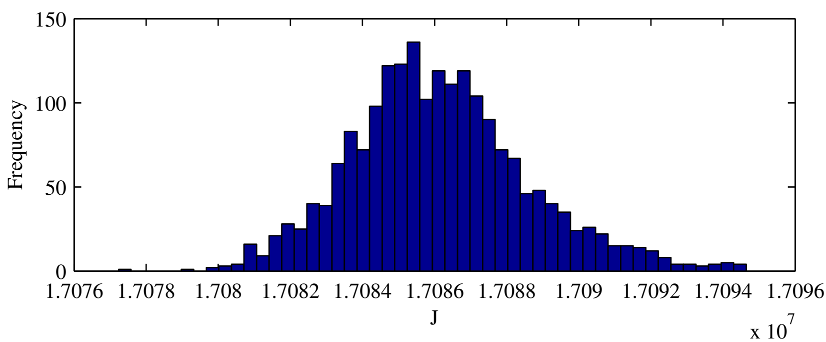
The other feature of `bigpoissonrv` is that for a vector `alpha` corresponding to expected values $\begin{bmatrix}\alpha_1 & \cdots & \alpha_m\end{bmatrix}'$, `bigpoissonrv` returns a vector `X` such that `X(i)` is a Poisson `alpha(i)` sample. The work of calculating the sum of logarithms is done only once for all calculated samples. The result is a significant savings in cpu time as long as the values of `alpha` are reasonably close to each other.

```
function x=bigpoissonrv(alpha)
%for vector alpha, returns a vector x such that
% x(i) is a Poisson (alpha(i)) rv
%set up Poisson CDF from xmin to xmax for each alpha(i)
alpha=alpha(:);
amin=min(alpha(:));
amax=max(alpha(:));
%Assume Poisson PMF is negligible +-6 sigma from the average
xmin=max(0,floor(amin-6*sqrt(amax)));
xmax=amax+ceil(6*sqrt(amax));%set max range
sx=xmin:xmax;
%Now we include the basic code of poissonpmf (but starting at xmin)
logfacts =cumsum([0,log(1:xmax)]); %include 0 in case xmin=0
logfacts=logfacts(sx+1); %extract needed values
%pmf(i,:) is a Poisson alpha(i) PMF from xmin to xmax
pmf=exp(-alpha*ones(size(sx))+ ...
        (log(alpha)*sx)-(ones(size(alpha))*logfacts));
cdf=cumsum(pmf,2); %each row is a cdf
x=(xmin-1)+sum((rand(size(alpha))*ones(size(sx)))<=cdf,2);
```

Finally, given `bigpoissonrv`, we can write a short program `lottery` that simulates trials of the jackpot experiment. Ideally, we would like to use `lottery` to perform $m = 1{,}000$ trials in a single pass. In general, MATLAB is more efficient when calculations are executed in parallel using vectors. However, in `bigpoissonrv`, the matrix `pmf` will have m rows and at least $12\sqrt{\alpha} = 12{,}000$ columns. For m more than several hundred, MATLAB running on my laptop reported an "Out of Memory" error. Thus, we wrote the program lottery to perform M trials at once and to repeat that N times. The output is an $M \times N$ matrix where each i, j entry is a sample jackpot after seven days.

```
function jackpot=lottery(jstart,M,N,D)
%Usage: function j=lottery(jstart,M,N,D)
%Perform M trials of the D day lottery
%of Problem 5.5.5 and initial jackpot jstart
jackpot=zeros(M,N);
for n=1:N,
jackpot(:,n)=jstart*ones(M,1);
for d=1:D,
    disp(d);
    jackpot(:,n)=jackpot(:,n)+(0.5*bigpoissonrv(jackpot(:,n)));
end
end
```

Executing `J=lottery(1e6,200,10,7)` generates a matrix `J` of 2,000 sample jackpots. The command `hist(J(:),50)` generates a histogram of the values with 50 bins. An example is shown here:



Problem Solutions – Chapter 9

Problem 9.1.1 Solution

Let $Y = X_1 - X_2$.

(a) Since $Y = X_1 + (-X_2)$, Theorem 9.1 says that the expected value of the difference is

$$\mathrm{E}[Y] = \mathrm{E}[X_1] + \mathrm{E}[-X_2] = \mathrm{E}[X] - \mathrm{E}[X] = 0. \tag{1}$$

(b) By Theorem 9.2, the variance of the difference is

$$\mathrm{Var}[Y] = \mathrm{Var}[X_1] + \mathrm{Var}[-X_2] = 2\,\mathrm{Var}[X]. \tag{2}$$

Problem 9.1.2 Solution

The random variable X_{33} is a Bernoulli random variable that indicates the result of flip 33. The PMF of X_{33} is

$$P_{X_{33}}(x) = \begin{cases} 1-p & x = 0, \\ p & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Note that each X_i has expected value $\mathrm{E}[X] = p$ and variance $\mathrm{Var}[X] = p(1-p)$. The random variable $Y = X_1 + \cdots + X_{100}$ is the number of heads in 100 coin flips. Hence, Y has the binomial PMF

$$P_Y(y) = \begin{cases} \binom{100}{y} p^y (1-p)^{100-y} & y = 0, 1, \ldots, 100, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Since the X_i are independent, by Theorems 9.1 and 9.3, the mean and variance of Y are

$$\mathrm{E}[Y] = 100\,\mathrm{E}[X] = 100p, \qquad \mathrm{Var}[Y] = 100\,\mathrm{Var}[X] = 100p(1-p). \tag{3}$$

Problem 9.1.3 Solution

(a) The PMF of N_1, the number of phone calls needed to obtain the correct answer, can be determined by observing that if the correct answer is given on the nth call, then the previous $n-1$ calls must have given wrong answers so that

$$P_{N_1}(n) = \begin{cases} (3/4)^{n-1}(1/4) & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) N_1 is a geometric random variable with parameter $p = 1/4$. In Theorem 3.5, the mean of a geometric random variable is found to be $1/p$. For our case, $\mathrm{E}[N_1] = 4$.

(c) Using the same logic as in part (a) we recognize that in order for n to be the fourth correct answer, that the previous $n-1$ calls must have contained exactly 3 correct answers and that the fourth correct answer arrived on the n-th call. This is described by a Pascal random variable.

$$P_{N_4}(n_4) = \begin{cases} \binom{n-1}{3}(3/4)^{n-4}(1/4)^4 & n = 4, 5, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(d) Using the hint given in the problem statement we can find the mean of N_4 by summing up the means of the 4 identically distributed geometric random variables each with mean 4. This gives $\mathrm{E}[N_4] = 4\,\mathrm{E}[N_1] = 16$.

Problem 9.1.4 Solution

Each X_i has PDF identical to a random variable X. First we observe that $\mathrm{E}[Y] = 3\,\mathrm{E}[X] = 0$, implying $\mathrm{E}[X] = 0$. Second, we observe that since the X_i are independent, Y has variance

$$\mathrm{Var}[Y] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \mathrm{Var}[X_3] = 3\,\mathrm{Var}[X]. \tag{1}$$

Thus $\text{Var}[X] = \sigma_Y^2/3 = 4/3$. Finally, since X is a continuous uniform random variable, we need to find the parameters (a, b) which satisfy

$$\text{E}[X] = \frac{b+a}{2} = 0, \qquad \text{Var}[X] = \frac{(b-a)^2}{12} = \frac{4}{3}. \tag{2}$$

Thus $b + a = 0$ and $b - a = 4$, implying $b = 2$ and $a = -2$. The PDF of X_1 is thus

$$f_{X_1}(x) = f_X(x) = \begin{cases} 1/4 & -2 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 9.1.5 Solution

We can solve this problem using Theorem 9.2 which says that

$$\text{Var}[W] = \text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}[X, Y]. \tag{1}$$

The first two moments of X are

$$\text{E}[X] = \int_0^1 \int_0^{1-x} 2x\,dy\,dx = \int_0^1 2x(1-x)\,dx = 1/3, \tag{2}$$

$$\text{E}[X^2] = \int_0^1 \int_0^{1-x} 2x^2\,dy\,dx = \int_0^1 2x^2(1-x)\,dx = 1/6. \tag{3}$$

Thus the variance of X is $\text{Var}[X] = \text{E}[X^2] - (\text{E}[X])^2 = 1/18$. By symmetry, it should be apparent that $\text{E}[Y] = \text{E}[X] = 1/3$ and $\text{Var}[Y] = \text{Var}[X] = 1/18$. To find the covariance, we first find the correlation

$$\text{E}[XY] = \int_0^1 \int_0^{1-x} 2xy\,dy\,dx = \int_0^1 x(1-x)^2\,dx = 1/12. \tag{4}$$

The covariance is

$$\text{Cov}[X, Y] = \text{E}[XY] - \text{E}[X]\,\text{E}[Y] = 1/12 - (1/3)^2 = -1/36. \tag{5}$$

Finally, the variance of the sum $W = X + Y$ is

$$\begin{aligned} \text{Var}[W] &= \text{Var}[X] + \text{Var}[Y] - 2\,\text{Cov}[X, Y] \\ &= 2/18 - 2/36 = 1/18. \end{aligned} \tag{6}$$

For this specific problem, it's arguable whether it would easier to find Var[W] by first deriving the CDF and PDF of W. In particular, for $0 \le w \le 1$,

$$\begin{aligned} F_W(w) &= \mathrm{P}\left[X + Y \le w\right] \\ &= \int_0^w \int_0^{w-x} 2\,dy\,dx \\ &= \int_0^w 2(w-x)\,dx = w^2. \end{aligned} \tag{7}$$

Hence, by taking the derivative of the CDF, the PDF of W is

$$f_W(w) = \begin{cases} 2w & 0 \le w \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

From the PDF, the first and second moments of W are

$$\mathrm{E}\left[W\right] = \int_0^1 2w^2\,dw = 2/3, \qquad \mathrm{E}\left[W^2\right] = \int_0^1 2w^3\,dw = 1/2. \tag{9}$$

The variance of W is $\mathrm{Var}[W] = \mathrm{E}[W^2] - (\mathrm{E}[W])^2 = 1/18$. Not surprisingly, we get the same answer both ways.

Problem 9.2.1 Solution

For a constant $a > 0$, a zero mean Laplace random variable X has PDF

$$f_X(x) = \frac{a}{2} e^{-a|x|} \qquad -\infty < x < \infty \tag{1}$$

The moment generating function of X is

$$\begin{aligned} \phi_X(s) = \mathrm{E}\left[e^{sX}\right] &= \frac{a}{2}\int_{-\infty}^0 e^{sx} e^{ax}\,dx + \frac{a}{2}\int_0^\infty e^{sx}e^{-ax}\,dx \\ &= \frac{a}{2}\frac{e^{(s+a)x}}{s+a}\bigg|_{-\infty}^0 + \frac{a}{2}\frac{e^{(s-a)x}}{s-a}\bigg|_0^\infty \\ &= \frac{a}{2}\left(\frac{1}{s+a} - \frac{1}{s-a}\right) \\ &= \frac{a^2}{a^2 - s^2}. \end{aligned} \tag{2}$$

Problem 9.2.2 Solution

(a) By summing across the rows of the table, we see that J has PMF

$$P_J(j) = \begin{cases} 0.6 & j = -2, \\ 0.4 & j = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The MGF of J is $\phi_J(s) = \mathrm{E}[e^{sJ}] = 0.6e^{-2s} + 0.4e^{-s}$.

(b) Summing down the columns of the table, we see that K has PMF

$$P_K(k) = \begin{cases} 0.7 & k = -1, \\ 0.2 & k = 0, \\ 0.1 & k = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The MGF of K is $\phi_K(s) = 0.7e^{-s} + 0.2 + 0.1e^{s}$.

(c) To find the PMF of $M = J + K$, it is easist to annotate each entry in the table with the coresponding value of M:

$P_{J,K}(j,k)$	$k = -1$	$k = 0$	$k = 1$
$j = -2$	$0.42(M = -3)$	$0.12(M = -2)$	$0.06(M = -1)$
$j = -1$	$0.28(M = -2)$	$0.08(M = -1)$	$0.04(M = 0)$

(3)

We obtain $P_M(m)$ by summing over all j, k such that $j + k = m$, yielding

$$P_M(m) = \begin{cases} 0.42 & m = -3, \\ 0.40 & m = -2, \\ 0.14 & m = -1, \\ 0.04 & m = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(d) One way to solve this problem, is to find the MGF $\phi_M(s)$ and then take four derivatives. Sometimes its better to just work with definition of $E[M^4]$:

$$\begin{aligned} E\left[M^4\right] &= \sum_m P_M(m)\, m^4 \\ &= 0.42(-3)^4 + 0.40(-2)^4 + 0.14(-1)^4 + 0.04(0)^4 \\ &= 40.434. \end{aligned} \tag{5}$$

As best I can tell, the prupose of this problem is to check that you know when not to use the methods in this chapter.

Problem 9.2.3 Solution

We find the MGF by calculating $E[e^{sX}]$ from the PDF $f_X(x)$.

$$\phi_X(s) = E\left[e^{sX}\right] = \int_a^b e^{sX} \frac{1}{b-a}\, dx = \frac{e^{bs} - e^{as}}{s(b-a)}. \tag{1}$$

Now to find the first moment, we evaluate the derivative of $\phi_X(s)$ at $s = 0$.

$$E[X] = \left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \left.\frac{s\left[be^{bs} - ae^{as}\right] - \left[e^{bs} - e^{as}\right]}{(b-a)s^2}\right|_{s=0}. \tag{2}$$

Direct evaluation of the above expression at $s = 0$ yields $0/0$ so we must apply l'Hôpital's rule and differentiate the numerator and denominator.

$$\begin{aligned} E[X] &= \lim_{s\to 0} \frac{be^{bs} - ae^{as} + s\left[b^2e^{bs} - a^2e^{as}\right] - \left[be^{bs} - ae^{as}\right]}{2(b-a)s} \\ &= \lim_{s\to 0} \frac{b^2e^{bs} - a^2e^{as}}{2(b-a)} = \frac{b+a}{2}. \end{aligned} \tag{3}$$

To find the second moment of X, we first find that the second derivative of $\phi_X(s)$ is

$$\frac{d^2\phi_X(s)}{ds^2} = \frac{s^2\left[b^2e^{bs} - a^2e^{as}\right] - 2s\left[be^{bs} - ae^{as}\right] + 2\left[be^{bs} - ae^{as}\right]}{(b-a)s^3}. \tag{4}$$

Substituting $s = 0$ will yield 0/0 so once again we apply l'Hôpital's rule and differentiate the numerator and denominator.

$$\begin{aligned} \mathrm{E}\left[X^2\right] = \lim_{s \to 0} \frac{d^2\phi_X(s)}{ds^2} &= \lim_{s \to 0} \frac{s^2\left[b^3 e^{bs} - a^3 e^{as}\right]}{3(b-a)s^2} \\ &= \frac{b^3 - a^3}{3(b-a)} = (b^2 + ab + a^2)/3. \end{aligned} \tag{5}$$

In this case, it is probably simpler to find these moments without using the MGF.

Problem 9.2.4 Solution

Using the moment generating function of X, $\phi_X(s) = e^{\sigma^2 s^2/2}$. We can find the nth moment of X, $\mathrm{E}[X^n]$ by taking the nth derivative of $\phi_X(s)$ and setting $s = 0$.

$$\mathrm{E}\left[X\right] = \sigma^2 s e^{\sigma^2 s^2/2}\Big|_{s=0} = 0, \tag{1}$$

$$\mathrm{E}\left[X^2\right] = \sigma^2 e^{\sigma^2 s^2/2} + \sigma^4 s^2 e^{\sigma^2 s^2/2}\Big|_{s=0} = \sigma^2. \tag{2}$$

Continuing in this manner we find that

$$\mathrm{E}\left[X^3\right] = \left(3\sigma^4 s + \sigma^6 s^3\right) e^{\sigma^2 s^2/2}\Big|_{s=0} = 0, \tag{3}$$

$$\mathrm{E}\left[X^4\right] = \left(3\sigma^4 + 6\sigma^6 s^2 + \sigma^8 s^4\right) e^{\sigma^2 s^2/2}\Big|_{s=0} = 3\sigma^4. \tag{4}$$

To calculate the moments of Y, we define $Y = X + \mu$ so that Y is Gaussian (μ, σ). In this case the second moment of Y is

$$\mathrm{E}\left[Y^2\right] = \mathrm{E}\left[(X+\mu)^2\right] = \mathrm{E}\left[X^2 + 2\mu X + \mu^2\right] = \sigma^2 + \mu^2. \tag{5}$$

Similarly, the third moment of Y is

$$\begin{aligned} \mathrm{E}\left[Y^3\right] &= \mathrm{E}\left[(X+\mu)^3\right] \\ &= \mathrm{E}\left[X^3 + 3\mu X^2 + 3\mu^2 X + \mu^3\right] = 3\mu\sigma^2 + \mu^3. \end{aligned} \tag{6}$$

Finally, the fourth moment of Y is

$$\begin{aligned} \mathrm{E}\left[Y^4\right] &= \mathrm{E}\left[(X+\mu)^4\right] \\ &= \mathrm{E}\left[X^4 + 4\mu X^3 + 6\mu^2 X^2 + 4\mu^3 X + \mu^4\right] \\ &= 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4. \end{aligned} \tag{7}$$

Problem 9.2.5 Solution

The PMF of K is

$$P_K(k) = \begin{cases} 1/n & k = 1, 2, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The corresponding MGF of K is

$$\begin{aligned} \phi_K(s) = \mathrm{E}\left[e^{sK}\right] &= \frac{1}{n}\left(e^s + e^2 s + \cdots + e^{ns}\right) \\ &= \frac{e^s}{n}\left(1 + e^s + e^{2s} + \cdots + e^{(n-1)s}\right) \\ &= \frac{e^s(e^{ns}-1)}{n(e^s-1)}. \end{aligned} \tag{2}$$

We can evaluate the moments of K by taking derivatives of the MGF. Some algebra will show that

$$\frac{d\phi_K(s)}{ds} = \frac{ne^{(n+2)s} - (n+1)e^{(n+1)s} + e^s}{n(e^s-1)^2}. \tag{3}$$

Evaluating $d\phi_K(s)/ds$ at $s = 0$ yields $0/0$. Hence, we apply l'Hôpital's rule twice (by twice differentiating the numerator and twice differentiating the denominator) when we write

$$\begin{aligned} \left.\frac{d\phi_K(s)}{ds}\right|_{s=0} &= \lim_{s\to 0} \frac{n(n+2)e^{(n+2)s} - (n+1)^2 e^{(n+1)s} + e^s}{2n(e^s-1)} \\ &= \lim_{s\to 0} \frac{n(n+2)^2 e^{(n+2)s} - (n+1)^3 e^{(n+1)s} + e^s}{2ne^s} \\ &= (n+1)/2. \end{aligned} \tag{4}$$

A significant amount of algebra will show that the second derivative of the MGF is

$$\frac{d^2\phi_K(s)}{ds^2} = \frac{n^2 e^{(n+3)s} - (2n^2+2n-1)e^{(n+2)s} + (n+1)^2 e^{(n+1)s} - e^{2s} - e^s}{n(e^s-1)^3}. \tag{5}$$

Evaluating $d^2\phi_K(s)/ds^2$ at $s = 0$ yields $0/0$.Because $(e^s - 1)^3$ appears in the denominator, we need to use l'Hôpital's rule three times to obtain our answer.

$$
\begin{aligned}
&\left.\frac{d^2\phi_K(s)}{ds^2}\right|_{s=0} \\
&= \lim_{s\to 0} \frac{n^2(n+3)^3 e^{(n+3)s} - (2n^2+2n-1)(n+2)^3 e^{(n+2)s} + (n+1)^5 - 8e^{2s} - e^s}{6ne^s} \\
&= \frac{n^2(n+3)^3 - (2n^2+2n-1)(n+2)^3 + (n+1)^5 - 9}{6n} \\
&= (2n+1)(n+1)/6. \qquad (6)
\end{aligned}
$$

We can use these results to derive two well known results. We observe that we can directly use the PMF $P_K(k)$ to calculate the moments

$$\mathrm{E}[K] = \frac{1}{n}\sum_{k=1}^{n} k, \qquad \mathrm{E}\left[K^2\right] = \frac{1}{n}\sum_{k=1}^{n} k^2. \qquad (7)$$

Using the answers we found for $\mathrm{E}[K]$ and $\mathrm{E}[K^2]$, we have the formulas

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \qquad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}. \qquad (8)$$

Problem 9.3.1 Solution

N is a binomial $(n = 100, p = 0.4)$ random variable. M is a binomial $(n = 50, p = 0.4)$ random variable. Thus N is the sum of 100 independent Bernoulli $(p = 0.4)$ and M is the sum of 50 independent Bernoulli $(p = 0.4)$ random variables. Since M and N are independent, $L = M + N$ is the sum of 150 independent Bernoulli $(p = 0.4)$ random variables. Hence L is a binomial $n = 150, p = 0.4)$ and has PMF

$$P_L(l) = \binom{150}{l}(0.4)^l(0.6)^{150-l}. \qquad (1)$$

Problem 9.3.2 Solution

Random variable Y has the moment generating function $\phi_Y(s) = 1/(1-s)$. Random variable V has the moment generating function $\phi_V(s) = 1/(1-s)^4$. Y and V are independent. $W = Y + V$.

(a) From Table 9.1, Y is an exponential ($\lambda = 1$) random variable. For an exponential (λ) random variable, Example 9.4 derives the moments of the exponential random variable. For $\lambda = 1$, the moments of Y are

$$\mathrm{E}[Y] = 1, \qquad \mathrm{E}\left[Y^2\right] = 2, \qquad \mathrm{E}\left[Y^3\right] = 3! = 6. \tag{1}$$

(b) Since Y and V are independent, $W = Y + V$ has MGF

$$\begin{aligned} \phi_W(s) &= \phi_Y(s)\phi_V(s) \\ &= \left(\frac{1}{1-s}\right)\left(\frac{1}{1-s}\right)^4 = \left(\frac{1}{1-s}\right)^5. \end{aligned} \tag{2}$$

W is the sum of five independent exponential ($\lambda = 1$) random variables $X_1, \ldots, X_5$. (That is, W is an Erlang ($n = 5, \lambda = 1$) random variable.) Each X_i has expected value $\mathrm{E}[X] = 1$ and variance $\mathrm{Var}[X] = 1$. From Theorem 9.1 and Theorem 9.3,

$$\mathrm{E}[W] = 5\,\mathrm{E}[X] = 5, \qquad \mathrm{Var}[W] = 5\,\mathrm{Var}[X] = 5. \tag{3}$$

It follows that

$$\mathrm{E}\left[W^2\right] = \mathrm{Var}[W] + (\mathrm{E}[W])^2 = 5 + 25 = 30. \tag{4}$$

Problem 9.3.3 Solution

In the iid random sequence $K_1, K_2, \ldots$, each K_i has PMF

$$P_K(k) = \begin{cases} 1-p & k = 0, \\ p & k = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The MGF of K is $\phi_K(s) = \mathrm{E}[e^{sK}] = 1 - p + pe^s$.

(b) By Theorem 9.6, $M = K_1 + K_2 + \ldots + K_n$ has MGF

$$\phi_M(s) = [\phi_K(s)]^n = [1 - p + pe^s]^n. \tag{2}$$

(c) Although we could just use the fact that the expectation of the sum equals the sum of the expectations, the problem asks us to find the moments using $\phi_M(s)$. In this case,

$$\begin{aligned} \mathrm{E}[M] &= \left.\frac{d\phi_M(s)}{ds}\right|_{s=0} \\ &= n(1-p+pe^s)^{n-1}pe^s\big|_{s=0} = np. \end{aligned} \tag{3}$$

The second moment of M can be found via

$$\begin{aligned} \mathrm{E}\left[M^2\right] &= \left.\frac{d\phi_M(s)}{ds}\right|_{s=0} \\ &= np\left((n-1)(1-p+pe^s)pe^{2s} + (1-p+pe^s)^{n-1}e^s\right)\big|_{s=0} \\ &= np[(n-1)p+1]. \end{aligned} \tag{4}$$

The variance of M is

$$\mathrm{Var}[M] = \mathrm{E}\left[M^2\right] - (\mathrm{E}[M])^2 = np(1-p) = n\,\mathrm{Var}[K]. \tag{5}$$

Problem 9.3.4 Solution

Based on the problem statement, the number of points X_i that you earn for game i has PMF

$$P_{X_i}(x) = \begin{cases} 1/3 & x = 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The MGF of X_i is

$$\phi_{X_i}(s) = \mathrm{E}\left[e^{sX_i}\right] = 1/3 + e^s/3 + e^{2s}/3. \tag{2}$$

Since $Y = X_1 + \cdots + X_n$, Theorem 9.6 implies

$$\phi_Y(s) = [\phi_{X_i}(s)]^n = [1 + e^s + e^{2s}]^n/3^n. \tag{3}$$

(b) First we observe that first and second moments of X_i are

$$\mathrm{E}[X_i] = \sum_x x P_{X_i}(x) = 1/3 + 2/3 = 1, \tag{4}$$

$$\mathrm{E}\left[X_i^2\right] = \sum_x x^2 P_{X_i}(x) = 1^2/3 + 2^2/3 = 5/3. \tag{5}$$

Hence,

$$\mathrm{Var}[X_i] = \mathrm{E}\left[X_i^2\right] - \left(\mathrm{E}\left[X_i\right]\right)^2 = 2/3. \tag{6}$$

By Theorems 9.1 and 9.3, the mean and variance of Y are

$$\mathrm{E}\left[Y\right] = n\,\mathrm{E}\left[X\right] = n, \tag{7}$$
$$\mathrm{Var}[Y] = n\,\mathrm{Var}[X] = 2n/3. \tag{8}$$

Another more complicated way to find the mean and variance is to evaluate derivatives of $\phi_Y(s)$ as $s = 0$.

Problem 9.3.5 Solution

K_i has PMF

$$P_{K_i}(k) = \begin{cases} 2^k e^{-2}/k! & k = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Let $R_i = K_1 + K_2 + \ldots + K_i$

(a) From Table 9.1, we find that the Poisson ($\alpha = 2$) random variable K has MGF $\phi_K(s) = e^{2(e^s-1)}$.

(b) The MGF of R_i is the product of the MGFs of the K_i's.

$$\phi_{R_i}(s) = \prod_{n=1}^{i} \phi_K(s) = e^{2i(e^s-1)}. \tag{2}$$

(c) Since the MGF of R_i is of the same form as that of the Poisson with parameter, $\alpha = 2i$. Therefore we can conclude that R_i is in fact a Poisson random variable with parameter $\alpha = 2i$. That is,

$$P_{R_i}(r) = \begin{cases} (2i)^r e^{-2i}/r! & r = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(d) Because R_i is a Poisson random variable with parameter $\alpha = 2i$, the mean and variance of R_i are then both $2i$.

Problem 9.3.6 Solution

The total energy stored over the 31 days is

$$Y = X_1 + X_2 + \cdots + X_{31}. \tag{1}$$

The random variables $X_1, \ldots, X_{31}$ are Gaussian and independent but not identically distributed. However, since the sum of independent Gaussian random variables is Gaussian, we know that Y is Gaussian. Hence, all we need to do is find the mean and variance of Y in order to specify the PDF of Y. The mean of Y is

$$\begin{aligned} \mathrm{E}[Y] &= \sum_{i=1}^{31} \mathrm{E}[X_i] = \sum_{i=1}^{31} (32 - i/4) \\ &= 32(31) - \frac{31(32)}{8} = 868 \text{ kW-hr.} \end{aligned} \tag{2}$$

Since each X_i has variance of $100(\text{kW-hr})^2$, the variance of Y is

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_{31}] \\ &= 31\,\mathrm{Var}[X_i] = 3100. \end{aligned} \tag{3}$$

Since $\mathrm{E}[Y] = 868$ and $\mathrm{Var}[Y] = 3100$, the Gaussian PDF of Y is

$$f_Y(y) = \frac{1}{\sqrt{6200\pi}} e^{-(y-868)^2/6200}. \tag{4}$$

Problem 9.3.7 Solution

By Theorem 9.6, we know that $\phi_M(s) = [\phi_K(s)]^n$.

(a) The first derivative of $\phi_M(s)$ is

$$\frac{d\phi_M(s)}{ds} = n\,[\phi_K(s)]^{n-1}\,\frac{d\phi_K(s)}{ds}. \tag{1}$$

We can evaluate $d\phi_M(s)/ds$ at $s = 0$ to find $\mathrm{E}[M]$.

$$\begin{aligned} \mathrm{E}[M] &= \left.\frac{d\phi_M(s)}{ds}\right|_{s=0} \\ &= n\,[\phi_K(s)]^{n-1} \left.\frac{d\phi_K(s)}{ds}\right|_{s=0} = n\,\mathrm{E}[K]. \end{aligned} \tag{2}$$

(b) The second derivative of $\phi_M(s)$ is

$$\frac{d^2\phi_M(s)}{ds^2} = n(n-1)\left[\phi_K(s)\right]^{n-2}\left(\frac{d\phi_K(s)}{ds}\right)^2 + n\left[\phi_K(s)\right]^{n-1}\frac{d^2\phi_K(s)}{ds^2}. \tag{3}$$

Evaluating the second derivative at $s=0$ yields

$$\begin{aligned} \mathrm{E}\left[M^2\right] &= \left.\frac{d^2\phi_M(s)}{ds^2}\right|_{s=0} \\ &= n(n-1)\left(\mathrm{E}\left[K\right]\right)^2 + n\,\mathrm{E}\left[K^2\right]. \end{aligned} \tag{4}$$

Problem 9.4.1 Solution

(a) From Table 9.1, we see that the exponential random variable X has MGF

$$\phi_X(s) = \frac{\lambda}{\lambda - s}. \tag{1}$$

(b) Note that K is a geometric random variable identical to the geometric random variable X in Table 9.1 with parameter $p = 1-q$. From Table 9.1, we know that random variable K has MGF

$$\phi_K(s) = \frac{(1-q)e^s}{1-qe^s}. \tag{2}$$

Since K is independent of each X_i, $V = X_1 + \cdots + X_K$ is a random sum of random variables. From Theorem 9.10,

$$\begin{aligned} \phi_V(s) &= \phi_K(\ln\phi_X(s)) \\ &= \frac{(1-q)\frac{\lambda}{\lambda-s}}{1-q\frac{\lambda}{\lambda-s}} = \frac{(1-q)\lambda}{(1-q)\lambda - s}. \end{aligned} \tag{3}$$

We see that the MGF of V is that of an exponential random variable with parameter $(1-q)\lambda$. The PDF of V is

$$f_V(v) = \begin{cases} (1-q)\lambda e^{-(1-q)\lambda v} & v \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 9.4.2 Solution

The number N of passes thrown has the Poisson PMF and MGF

$$P_N(n) = \begin{cases} (30)^n e^{-30}/n! & n = 0, 1, \ldots, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$\phi_N(s) = e^{30(e^s-1)}. \tag{2}$$

Let $X_i = 1$ if pass i is thrown and completed and otherwise $X_i = 0$. The PMF and MGF of each X_i is

$$P_{X_i}(x) = \begin{cases} 1/3 & x = 0, \\ 2/3 & x = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$\phi_{X_i}(s) = 1/3 + (2/3)e^s. \tag{4}$$

The number of completed passes can be written as the random sum of random variables

$$K = X_1 + \cdots + X_N. \tag{5}$$

Since each X_i is independent of N, we can use Theorem 9.10 to write

$$\phi_K(s) = \phi_N(\ln \phi_X(s)) = e^{30(\phi_X(s)-1)} = e^{30(2/3)(e^s-1)}. \tag{6}$$

We see that K has the MGF of a Poisson random variable with mean $\mathrm{E}[K] = 30(2/3) = 20$, variance $\mathrm{Var}[K] = 20$, and PMF

$$P_K(k) = \begin{cases} (20)^k e^{-20}/k! & k = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Problem 9.4.3 Solution

In this problem, $Y = X_1 + \cdots + X_N$ is not a straightforward random sum of random variables because N and the X_i's are dependent. In particular, given $N = n$, then we know that there were exactly 100 heads in N flips. Hence, given N, $X_1 + \cdots + X_N = 100$ *no matter what is the actual value of* N. Hence $Y = 100$ every time and the PMF of Y is

$$P_Y(y) = \begin{cases} 1 & y = 100, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 9.4.4 Solution

Donovan McNabb's passing yardage is the random sum of random variables

$$V + Y_1 + \cdots + Y_K, \tag{1}$$

where Y_i has the exponential PDF

$$f_{Y_i}(y) = \begin{cases} \frac{1}{15} e^{-y/15} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

From Table 9.1, the MGFs of Y and K are

$$\phi_Y(s) = \frac{1/15}{1/15 - s} = \frac{1}{1 - 15s}, \tag{3}$$

$$\phi_K(s) = e^{20(e^s - 1)}. \tag{4}$$

From Theorem 9.10, V has MGF

$$\phi_V(s) = \phi_K(\ln \phi_Y(s)) = e^{20(\phi_Y(s) - s)} = e^{300s/(1-15s)}. \tag{5}$$

The PDF of V cannot be found in a simple form. However, we can use the MGF to calculate the mean and variance. In particular,

$$\begin{aligned} \mathrm{E}[V] &= \left. \frac{d\phi_V(s)}{ds} \right|_{s=0} \\ &= \left. e^{300s/(1-15s)} \frac{300}{(1-15s)^2} \right|_{s=0} = 300, \end{aligned} \tag{6}$$

$$\begin{aligned} \mathrm{E}\left[V^2\right] &= \left. \frac{d^2\phi_V(s)}{ds^2} \right|_{s=0} \\ &= \left. e^{300s/(1-15s)} \left(\frac{300}{(1-15s)^2} \right)^2 + e^{300s/(1-15s)} \frac{9000}{(1-15s)^3} \right|_{s=0} \\ &= 99,000. \end{aligned} \tag{7}$$

Thus, V has variance $\mathrm{Var}[V] = \mathrm{E}[V^2] - (\mathrm{E}[V])^2 = 9,000$ and standard deviation $\sigma_V \approx 94.9$.

A second way to calculate the mean and variance of V is to use Theorem 9.11 which says

$$\mathrm{E}[V] = \mathrm{E}[K]\,\mathrm{E}[Y] = 20(15) = 200, \tag{8}$$

$$\begin{aligned} \mathrm{Var}[V] &= \mathrm{E}[K]\,\mathrm{Var}[Y] + \mathrm{Var}[K](\mathrm{E}[Y])^2 \\ &= (20)15^2 + (20)15^2 = 9000. \end{aligned} \tag{9}$$

Problem 9.4.5 Solution

Since each ticket is equally likely to have one of $\binom{46}{6}$ combinations, the probability a ticket is a winner is

$$q = \frac{1}{\binom{46}{6}}. \tag{1}$$

Let $X_i = 1$ if the ith ticket sold is a winner; otherwise $X_i = 0$. Since the number K of tickets sold has a Poisson PMF with $\mathrm{E}[K] = r$, the number of winning tickets is the random sum

$$V = X_1 + \cdots + X_K. \tag{2}$$

From Appendix A,

$$\phi_X(s) = (1-q) + qe^s, \qquad \phi_K(s) = e^{r[e^s-1]}. \tag{3}$$

By Theorem 9.10,

$$\phi_V(s) = \phi_K(\ln \phi_X(s)) = e^{r[\phi_X(s)-1]}. = e^{rq(e^s-1)} \tag{4}$$

Hence, we see that V has the MGF of a Poisson random variable with mean $\mathrm{E}[V] = rq$. The PMF of V is

$$P_V(v) = \begin{cases} (rq)^v e^{-rq}/v! & v = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Problem 9.4.6 Solution

(a) We can view K as a shifted geometric random variable. To find the MGF, we start from first principles with Definition 9.1:

$$\begin{aligned}\phi_K(s) &= \sum_{k=0}^{\infty} e^{sk} p(1-p)^k \\ &= p\sum_{n=0}^{\infty}[(1-p)e^s]^k = \frac{p}{1-(1-p)e^s}. \qquad (1)\end{aligned}$$

(b) First, we need to recall that each X_i has MGF $\phi_X(s) = e^{s+s^2/2}$. From Theorem 9.10, the MGF of R is

$$\begin{aligned}\phi_R(s) &= \phi_K(\ln \phi_X(s)) \\ &= \phi_K(s+s^2/2) = \frac{p}{1-(1-p)e^{s+s^2/2}}. \qquad (2)\end{aligned}$$

(c) To use Theorem 9.11, we first need to calculate the mean and variance of K:

$$\begin{aligned}\mathrm{E}[K] &= \left.\frac{d\phi_K(s)}{ds}\right|_{s=0} \\ &= \left.\frac{p(1-p)e^s}{1-(1-p)e^s}^{2}\right|_{s=0} = \frac{1-p}{p}, \qquad (3) \\ \mathrm{E}\left[K^2\right] &= \left.\frac{d^2\phi_K(s)}{ds^2}\right|_{s=0} \\ &= p(1-p)\left.\frac{[1-(1-p)e^s]e^s + 2(1-p)e^{2s}}{[1-(1-p)e^s]^3}\right|_{s=0} \\ &= \frac{(1-p)(2-p)}{p^2}. \qquad (4)\end{aligned}$$

Hence, $\mathrm{Var}[K] = \mathrm{E}[K^2] - (\mathrm{E}[K])^2 = (1-p)/p^2$. Finally. we can use Theorem 9.11 to write

$$\begin{aligned}\mathrm{Var}[R] &= \mathrm{E}[K]\,\mathrm{Var}[X] + (\mathrm{E}[X])^2\,\mathrm{Var}[K] \\ &= \frac{1-p}{p} + \frac{1-p}{p^2} = \frac{1-p^2}{p^2}.\end{aligned} \tag{5}$$

Problem 9.4.7 Solution

The way to solve for the mean and variance of U is to use conditional expectations. Given $K = k$, $U = X_1 + \cdots + X_k$ and

$$\begin{aligned}\mathrm{E}[U|K=k] &= \mathrm{E}[X_1 + \cdots + X_k | X_1 + \cdots + X_n = k] \\ &= \sum_{i=1}^{k} \mathrm{E}[X_i | X_1 + \cdots + X_n = k].\end{aligned} \tag{1}$$

Since X_i is a Bernoulli random variable,

$$\begin{aligned}\mathrm{E}[X_i | X_1 + \cdots + X_n = k] &= \mathrm{P}\left[X_i = 1 | \sum_{j=1}^{n} X_j = k\right] \\ &= \frac{\mathrm{P}\left[X_i = 1, \sum_{j \neq i} X_j = k-1\right]}{\mathrm{P}\left[\sum_{j=1}^{n} X_j = k\right]}.\end{aligned} \tag{2}$$

Note that $\sum_{j=1}^{n} X_j$ is just a binomial random variable for n trials while $\sum_{j \neq i} X_j$ is a binomial random variable for $n-1$ trials. In addition, X_i and $\sum_{j \neq i} X_j$ are independent random variables. This implies

$$\begin{aligned}\mathrm{E}[X_i | X_1 + \cdots + X_n = k] &= \frac{\mathrm{P}[X_i = 1]\,\mathrm{P}\left[\sum_{j \neq i} X_j = k-1\right]}{\mathrm{P}\left[\sum_{j=1}^{n} X_j = k\right]} \\ &= \frac{p\binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{k}{n}.\end{aligned} \tag{3}$$

A second way is to argue that symmetry implies

$$\mathrm{E}[X_i | X_1 + \cdots + X_n = k] = \gamma, \tag{4}$$

the same for each i. In this case,

$$
\begin{aligned}
n\gamma &= \sum_{i=1}^{n} \mathrm{E}\left[X_i | X_1 + \cdots + X_n = k\right] \\
&= \mathrm{E}\left[X_1 + \cdots + X_n | X_1 + \cdots + X_n = k\right] = k.
\end{aligned} \tag{5}
$$

Thus $\gamma = k/n$. At any rate, the conditional mean of U is

$$
\begin{aligned}
\mathrm{E}\left[U | K = k\right] &= \sum_{i=1}^{k} \mathrm{E}\left[X_i | X_1 + \cdots + X_n = k\right] \\
&= \sum_{i=1}^{k} \frac{k}{n} = \frac{k^2}{n}.
\end{aligned} \tag{6}
$$

This says that the random variable $\mathrm{E}[U|K] = K^2/n$. Using iterated expectations, we have

$$
\mathrm{E}\left[U\right] = \mathrm{E}\left[\mathrm{E}\left[U|K\right]\right] = \mathrm{E}\left[K^2/n\right]. \tag{7}
$$

Since K is a binomial random variable, we know that $\mathrm{E}[K] = np$ and $\mathrm{Var}[K] = np(1-p)$. Thus,

$$
\mathrm{E}\left[U\right] = \frac{\mathrm{E}\left[K^2\right]}{n} = \frac{\mathrm{Var}[K] + (\mathrm{E}\left[K\right])^2}{n} = p(1-p) + np^2. \tag{8}
$$

On the other hand, V is just and ordinary random sum of independent random variables and the mean of $\mathrm{E}[V] = \mathrm{E}[X]\,\mathrm{E}[M] = np^2$.

Problem 9.4.8 Solution

Using N to denote the number of games played, we can write the total number of points earned as the random sum

$$
Y = X_1 + X_2 + \cdots + X_N. \tag{1}
$$

(a) It is tempting to use Theorem 9.10 to find $\phi_Y(s)$; however, this would be wrong since each X_i is not independent of N. In this problem, we must start from first principles using iterated expectations.

$$\begin{aligned}\phi_Y(s) &= \mathrm{E}\left[\mathrm{E}\left[e^{s(X_1+\cdots+X_N)}|N\right]\right] \\ &= \sum_{n=1}^{\infty} P_N(n)\,\mathrm{E}\left[e^{s(X_1+\cdots+X_n)}|N=n\right]. \end{aligned} \tag{2}$$

Given $N = n$, $X_1, \ldots, X_n$ are independent so that

$$\begin{aligned}&\mathrm{E}\left[e^{s(X_1+\cdots+X_n)}|N=n\right] \\ &\quad = \mathrm{E}\left[e^{sX_1}|N=n\right]\mathrm{E}\left[e^{sX_2}|N=n\right]\cdots\mathrm{E}\left[e^{sX_n}|N=n\right]. \end{aligned} \tag{3}$$

Given $N = n$, we know that games 1 through $n-1$ were either wins or ties and that game n was a loss. That is, given $N = n$, $X_n = 0$ and for $i < n$, $X_i \neq 0$. Moreover, for $i < n$, X_i has the conditional PMF

$$P_{X_i|N=n}(x) = P_{X_i|X_i\neq 0}(x) = \begin{cases} 1/2 & x = 1, 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

These facts imply

$$\mathrm{E}\left[e^{sX_n}|N=n\right] = e^0 = 1, \tag{5}$$

and that for $i < n$,

$$\begin{aligned}\mathrm{E}\left[e^{sX_i}|N=n\right] &= (1/2)e^s + (1/2)e^{2s} \\ &= e^s/2 + e^{2s}/2. \end{aligned} \tag{6}$$

Now we can find the MGF of Y.

$$\begin{aligned}\phi_Y(s) &= \sum_{n=1}^{\infty} P_N(n)\,\mathrm{E}\left[e^{sX_1}|N=n\right]\mathrm{E}\left[e^{sX_2}|N=n\right]\cdots\mathrm{E}\left[e^{sX_n}|N=n\right] \\ &= \sum_{n=1}^{\infty} P_N(n)\left[e^s/2 + e^{2s}/2\right]^{n-1} \\ &= \frac{1}{e^s/2 + e^{2s}/2}\sum_{n=1}^{\infty} P_N(n)\left[e^s/2 + e^{2s}/2\right]^{n}. \end{aligned} \tag{7}$$

It follows that

$$\begin{aligned}\phi_Y(s) &= \frac{1}{e^s/2 + e^{2s}/2} \sum_{n=1}^{\infty} P_N(n)\, e^{n \ln[(e^s+e^{2s})/2]} \\ &= \frac{\phi_N(\ln[e^s/2 + e^{2s}/2])}{e^s/2 + e^{2s}/2}. \end{aligned} \tag{8}$$

The tournament ends as soon as you lose a game. Since each game is a loss with probability 1/3 independent of any previous game, the number of games played has the geometric PMF and corresponding MGF

$$P_N(n) = \begin{cases} (2/3)^{n-1}(1/3) & n = 1, 2, \ldots, \\ 0 & \text{otherwise}, \end{cases} \tag{9}$$

$$\phi_N(s) = \frac{(1/3)e^s}{1 - (2/3)e^s}. \tag{10}$$

Thus, the MGF of Y is

$$\phi_Y(s) = \frac{1/3}{1 - (e^s + e^{2s})/3}. \tag{11}$$

(b) To find the moments of Y, we evaluate the derivatives of the MGF $\phi_Y(s)$. Since

$$\frac{d\phi_Y(s)}{ds} = \frac{e^s + 2e^{2s}}{9\left[1 - e^s/3 - e^{2s}/3\right]^2}, \tag{12}$$

we see that

$$\mathrm{E}\left[Y\right] = \left.\frac{d\phi_Y(s)}{ds}\right|_{s=0} = \frac{3}{9(1/3)^2} = 3. \tag{13}$$

You may have noticed that $\mathrm{E}[Y] = 3$ precisely equals $\mathrm{E}[N]\,\mathrm{E}[X_i]$, the answer you would get if you mistakenly assumed that N and each X_i were independent. Although this may seem like a coincidence, its actually the result of theorem known as Wald's equality.

The second derivative of the MGF is

$$\frac{d^2\phi_Y(s)}{ds^2} = \frac{(1 - e^s/3 - e^{2s}/3)(e^s + 4e^{2s}) + 2(e^s + 2e^{2s})^2/3}{9(1 - e^s/3 - e^{2s}/3)^3}. \tag{14}$$

The second moment of Y is

$$\mathrm{E}\left[Y^2\right] = \left.\frac{d^2\phi_Y(s)}{ds^2}\right|_{s=0} = \frac{5/3+6}{1/3} = 23. \tag{15}$$

The variance of Y is $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = 23 - 9 = 14$.

Problem 9.5.1 Solution

We know that the waiting time, W is uniformly distributed on [0,10] and therefore has the following PDF.

$$f_W(w) = \begin{cases} 1/10 & 0 \le w \le 10, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We also know that the total time is 3 milliseconds plus the waiting time, that is $X = W + 3$.

(a) The expected value of X is $\mathrm{E}[X] = \mathrm{E}[W+3] = \mathrm{E}[W] + 3 = 5 + 3 = 8$.

(b) The variance of X is $\mathrm{Var}[X] = \mathrm{Var}[W+3] = \mathrm{Var}[W] = 25/3$.

(c) The expected value of A is $\mathrm{E}[A] = 12\,\mathrm{E}[X] = 96$.

(d) The standard deviation of A is $\sigma_A = \sqrt{\mathrm{Var}[A]} = \sqrt{12(25/3)} = 10$.

(e) $P[A > 116] = 1 - \Phi(\frac{116-96}{10}) = 1 - \Phi(2) = 0.02275$.

(f) $P[A < 86] = \Phi(\frac{86-96}{10}) = \Phi(-1) = 1 - \Phi(1) = 0.1587$.

Problem 9.5.2 Solution

Knowing that the probability that voice call occurs is 0.8 and the probability that a data call occurs is 0.2 we can define the random variable D_i as the number of data calls in a single telephone call. It is obvious that for any i there are only two possible values for D_i, namely 0 and 1. Furthermore for all i the D_i's are independent and identically distributed withe the following PMF.

$$P_D(d) = \begin{cases} 0.8 & d = 0, \\ 0.2 & d = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

From the above we can determine that

$$E[D] = 0.2, \qquad \mathrm{Var}[D] = 0.2 - 0.04 = 0.16. \tag{2}$$

With these facts, we can answer the questions posed by the problem.

(a) $E[K_{100}] = 100\, E[D] = 20$.

(b) $\mathrm{Var}[K_{100}] = \sqrt{100\, \mathrm{Var}[D]} = \sqrt{16} = 4$.

(c)

$$\begin{aligned} P[K_{100} \geq 18] &= 1 - \Phi\left(\frac{18-20}{4}\right) \\ &= 1 - \Phi(-1/2) = \Phi(1/2) = 0.6915. \end{aligned} \tag{3}$$

(d)

$$\begin{aligned} P[16 \leq K_{100} \leq 24] &= \Phi\left(\frac{24-20}{4}\right) - \Phi\left(\frac{16-20}{4}\right) \\ &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 = 0.6826. \end{aligned} \tag{4}$$

Problem 9.5.3 Solution

(a) Let $X_1, \ldots, X_{120}$ denote the set of call durations (measured in minutes) during the month. From the problem statement, each $X - I$ is an exponential (λ) random variable with $E[X_i] = 1/\lambda = 2.5$ min and $\mathrm{Var}[X_i] = 1/\lambda^2 = 6.25$ min^2. The total number of minutes used during the month is $Y = X_1 + \cdots + X_{120}$. By Theorem 9.1 and Theorem 9.3,

$$\begin{aligned} E[Y] &= 120\, E[X_i] = 300 \\ \mathrm{Var}[Y] &= 120\, \mathrm{Var}[X_i] = 750. \end{aligned} \tag{1}$$

The subscriber's bill is $30 + 0.4(y - 300)^+$ where $x^+ = x$ if $x \geq 0$ or $x^+ = 0$ if $x < 0$. the subscribers bill is exactly \$36 if $Y = 315$. The probability the subscribers bill exceeds \$36 equals

$$\begin{aligned} P[Y > 315] &= P\left[\frac{Y - 300}{\sigma_Y} > \frac{315 - 300}{\sigma_Y}\right] \\ &= Q\left(\frac{15}{\sqrt{750}}\right) = 0.2919. \end{aligned} \tag{2}$$

(b) If the actual call duration is X_i, the subscriber is billed for $M_i = \lceil X_i \rceil$ minutes. Because each X_i is an exponential (λ) random variable, Theorem 4.9 says that M_i is a geometric (p) random variable with $p = 1 - e^{-\lambda} = 0.3297$. Since M_i is geometric,

$$E[M_i] = \frac{1}{p} = 3.033, \qquad \text{Var}[M_i] = \frac{1-p}{p^2} = 6.167. \tag{3}$$

The number of billed minutes in the month is $B = M_1 + \cdots + M_{120}$. Since $M_1, \ldots, M_{120}$ are iid random variables,

$$E[B] = 120\,E[M_i] = 364.0, \qquad \text{Var}[B] = 120\,\text{Var}[M_i] = 740.08. \tag{4}$$

Similar to part (a), the subscriber is billed \$36 if $B = 315$ minutes. The probability the subscriber is billed more than \$36 is

$$\begin{aligned} P[B > 315] &= P\left[\frac{B - 364}{\sqrt{740.08}} > \frac{315 - 365}{\sqrt{740.08}}\right] \\ &= Q(-1.8) = \Phi(1.8) = 0.964. \end{aligned} \tag{5}$$

Problem 9.5.4 Solution

In Theorem 9.7, we learned that a sum of iid Poisson random variables is a Poisson random variable. Hence W_n is a Poisson random variable with mean $E[W_n] = n\,E[K] = n$. Thus W_n has variance $\text{Var}[W_n] = n$ and PMF

$$P_{W_n}(w) = \begin{cases} n^w e^{-n}/w! & w = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

All of this implies that we can exactly calculate

$$\mathrm{P}\left[W_n = n\right] = P_{W_n}(n) = n^n e^{-n}/n! \tag{2}$$

Since we can perform the exact calculation, using a central limit theorem may seem silly; however for large n, calculating n^n or $n!$ is difficult for large n. Moreover, it's interesting to see how good the approximation is. In this case, the approximation is

$$\begin{aligned}\mathrm{P}\left[W_n = n\right] &= \mathrm{P}\left[n \leq W_n \leq n\right] \\ &\approx \Phi\left(\frac{n+0.5-n}{\sqrt{n}}\right) - \Phi\left(\frac{n-0.5-n}{\sqrt{n}}\right) \\ &= 2\Phi\left(\frac{1}{2\sqrt{n}}\right) - 1. \end{aligned} \tag{3}$$

The comparison of the exact calculation and the approximation are given in the following table.

$\mathrm{P}\left[W_n = n\right]$	$n = 1$	$n = 4$	$n = 16$	$n = 64$
exact	0.3679	0.1954	0.0992	0.0498
approximate	0.3829	0.1974	0.0995	0.0498

(4)

Problem 9.5.5 Solution

(a) Since the number of requests N has expected value $\mathrm{E}[N] = 300$ and variance $\mathrm{Var}[N] = 300$, we need C to satisfy

$$\begin{aligned}\mathrm{P}\left[N > C\right] &= \mathrm{P}\left[\frac{N-300}{\sqrt{300}} > \frac{C-300}{\sqrt{300}}\right] \\ &= 1 - \Phi\left(\frac{C-300}{\sqrt{300}}\right) = 0.05. \end{aligned} \tag{1}$$

From Table 4.2, we note that $\Phi(1.65) = 0.9505$. Thus,

$$C = 300 + 1.65\sqrt{300} = 328.6. \tag{2}$$

(b) For $C = 328.6$, the exact probability of overload is

$$\begin{aligned} P[N > C] &= 1 - P[N \leq 328] \\ &= 1 - \texttt{poissoncdf(300,328)} = 0.0516, \end{aligned} \tag{3}$$

which shows the central limit theorem approximation is reasonable.

(c) This part of the problem could be stated more carefully. Re-examining Definition 3.9 for the Poisson random variable and the accompanying discussion in Chapter 3, we observe that the webserver has an arrival rate of $\lambda = 300$ hits/min, or equivalently $\lambda = 5$ hits/sec. Thus in a one second interval, the number of requests N' is a Poisson ($\alpha = 5$) random variable.

However, since the server "capacity" in a one second interval is not precisely defined, we will make the somewhat arbitrary definition that the server capacity is $C' = 328.6/60 = 5.477$ packets/sec. With this somewhat arbitrary definition, the probability of overload in a one second interval is

$$P\left[N' > C'\right] = 1 - P\left[N' \leq 5.477\right] = 1 - P\left[N' \leq 5\right]. \tag{4}$$

Because the number of arrivals in the interval is small, it would be a mistake to use the Central Limit Theorem to estimate this overload probability. However, the direct calculation of the overload probability is not hard. For $E[N'] = \alpha = 5$,

$$\begin{aligned} 1 - P\left[N' \leq 5\right] &= 1 - \sum_{n=0}^{5} P_N(n) \\ &= 1 - e^{-\alpha} \sum_{n=0}^{5} \frac{\alpha^n}{n!} = 0.3840. \end{aligned} \tag{5}$$

(d) Here we find the smallest C such that $P[N' \leq C] \geq 0.95$. From the previous step, we know that $C > 5$. Since N' is a Poisson ($\alpha = 5$) random variable, we need to find the smallest C such that

$$P[N \leq C] = \sum_{n=0}^{C} \alpha^n e^{-\alpha}/n! \geq 0.95. \tag{6}$$

Some experiments with `poissoncdf(alpha,c)` will show that

$$\mathrm{P}[N \leq 8] = 0.9319, \qquad \mathrm{P}[N \leq 9] = 0.9682. \tag{7}$$

Hence $C = 9$.

(e) If we use the Central Limit theorem to estimate the overload probability in a one second interval, we would use the facts that $\mathrm{E}[N'] = 5$ and $\mathrm{Var}[N'] = 5$ to estimate the the overload probability as

$$1 - \mathrm{P}\left[N' \leq 5\right] = 1 - \Phi\left(\frac{5-5}{\sqrt{5}}\right) = 0.5, \tag{8}$$

which overestimates the overload probability by roughly 30 percent. We recall from Chapter 3 that a Poisson random is the limiting case of the (n, p) binomial random variable when n is large and $np = \alpha$.In general, for fixed p, the Poisson and binomial PMFs become closer as n increases. Since large n is also the case for which the central limit theorem applies, it is not surprising that the the CLT approximation for the Poisson (α) CDF is better when $\alpha = np$ is large.

Comment: Perhaps a more interesting question is why the overload probability in a one-second interval is so much higher than that in a one-minute interval? To answer this, consider a T-second interval in which the number of requests N_T is a Poisson (λT) random variable while the server capacity is cT hits. In the earlier problem parts, $c = 5.477$ hits/sec. We make the assumption that the server system is reasonably well-engineered in that $c > \lambda$. (In fact, to assume otherwise means that the backlog of requests will grow without bound.) Further, assuming T is fairly large, we use the CLT to estimate the probability of overload in a T-second interval as

$$\mathrm{P}[N_T \geq cT] = \mathrm{P}\left[\frac{N_T - \lambda T}{\sqrt{\lambda T}} \geq \frac{cT - \lambda T}{\sqrt{\lambda T}}\right] = Q\left(k\sqrt{T}\right), \tag{9}$$

where $k = (c - \lambda)/\sqrt{\lambda}$. As long as $c > \lambda$, the overload probability decreases with increasing T. In fact, the overload probability goes rapidly to zero as T becomes large. The reason is that the gap $cT - \lambda T$ between server capacity cT and the expected number of requests λT grows linearly in T while the standard deviation of the number of requests grows proportional to $\sqrt{T}$.

However, one should add that the definition of a T-second overload is somewhat arbitrary. In fact, one can argue that as T becomes large, the requirement for no overloads simply becomes less stringent. Using more advanced techniques found in the Markov Chains Supplement, a system such as this webserver can be evaluated in terms of the average backlog of requests and the average delay in serving in serving a request. These statistics won't depend on a particular time period T and perhaps better describe the system performance.

Problem 9.5.6 Solution

(a) The number of tests L needed to identify 500 acceptable circuits is a Pascal $(k = 500, p = 0.8)$ random variable, which has expected value $\mathrm{E}[L] = k/p = 625$ tests.

(b) Let K denote the number of acceptable circuits in $n = 600$ tests. Since K is binomial $(n = 600, p = 0.8)$, $\mathrm{E}[K] = np = 480$ and $\mathrm{Var}[K] = np(1-p) = 96$. Using the CLT, we estimate the probability of finding at least 500 acceptable circuits as

$$\mathrm{P}\left[K \geq 500\right] = \mathrm{P}\left[\frac{K-480}{\sqrt{96}} \geq \frac{20}{\sqrt{96}}\right] \approx Q\left(\frac{20}{\sqrt{96}}\right) = 0.0206. \tag{1}$$

(c) Using MATLAB, we observe that

```
1.0-binomialcdf(600,0.8,499)
ans =
    0.0215
```

(d) We need to find the smallest value of n such that the binomial (n, p) random variable K satisfies $\mathrm{P}[K \geq 500] \geq 0.9$. Since $\mathrm{E}[K] = np$ and $\mathrm{Var}[K] = np(1-p)$, the CLT approximation yields

$$\begin{aligned}\mathrm{P}\left[K \geq 500\right] &= \mathrm{P}\left[\frac{K-np}{\sqrt{np(1-p)}} \geq \frac{500-np}{\sqrt{np(1-p)}}\right] \\ &\approx 1 - \Phi(z) = 0.90. \end{aligned} \tag{2}$$

where $z = (500 - np)/\sqrt{np(1-p)}$. It follows that $1 - \Phi(z) = \Phi(-z) \geq 0.9$, implying $z = -1.29$. Since $p = 0.8$, we have that

$$np - 500 = 1.29\sqrt{np(1-p)}. \tag{3}$$

Equivalently, for $p = 0.8$, solving the quadratic equation

$$\left(n - \frac{500}{p}\right)^2 = (1.29)^2 \frac{1-p}{p} n \tag{4}$$

yields $n = 641.3$. Thus we should test $n = 642$ circuits.

Problem 9.5.7 Solution

Random variable K_n has a binomial distribution for n trials and success probability $P[V] = 3/4$.

(a) The expected number of video packets out of 48 packets is

$$E[K_{48}] = 48\,P[V] = 36. \tag{1}$$

(b) The variance of K_{48} is

$$\text{Var}[K_{48}] = 48\,P[V](1 - P[V]) = 48(3/4)(1/4) = 9 \tag{2}$$

Thus K_{48} has standard deviation $\sigma_{K_{48}} = 3$.

(c) Using the ordinary central limit theorem and Table 4.2 yields

$$\begin{aligned} P[30 \leq K_{48} \leq 42] &\approx \Phi\left(\frac{42-36}{3}\right) - \Phi\left(\frac{30-36}{3}\right) \\ &= \Phi(2) - \Phi(-2) \end{aligned} \tag{3}$$

Recalling that $\Phi(-x) = 1 - \Phi(x)$, we have

$$P[30 \leq K_{48} \leq 42] \approx 2\Phi(2) - 1 = 0.9545. \tag{4}$$

(d) Since K_{48} is a discrete random variable, we can use the De Moivre-Laplace approximation to estimate

$$\begin{aligned} P[30 \leq K_{48} \leq 42] &\approx \Phi\left(\frac{42+0.5-36}{3}\right) - \Phi\left(\frac{30-0.5-36}{3}\right) \\ &= 2\Phi(2.16666) - 1 = 0.9687. \end{aligned} \tag{5}$$

Problem 9.5.8 Solution

We start by finding E[V] and Var[V] as follows.

$$
\begin{aligned}
\mathrm{E}[V] = \mathrm{E}\left[20 - 10W^3\right] &= 20 - 10\,\mathrm{E}\left[W^3\right] \\
&= 20 - 10\int_{-1}^{1} w^3 \frac{1}{2}\,dw \\
&= 20 - 10\left(\left.\frac{w^4}{8}\right|_{-1}^{1}\right) = 20. \qquad (1)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}[V] &= \mathrm{E}\left[(V - \mathrm{E}[V])^2\right] \\
&= \mathrm{E}\left[\left(20 - 10W^3 - 20\right)^2\right] \\
&= \mathrm{E}\left[100W^6\right] = 50\int_{-1}^{1} w^6\,dw = \frac{100}{7}. \qquad (2)
\end{aligned}
$$

To use a central limit theorem approximation, we find the expected value and variance of X.

$$
\mathrm{E}[X] = \frac{\mathrm{E}[V_1] + \mathrm{E}[V_2] + \cdots + \mathrm{E}[V_{30}]}{6} = 5\,\mathrm{E}[V] = 100. \qquad (3)
$$

Since the V_i are iid,

$$
\begin{aligned}
\mathrm{Var}[X] &= \frac{\mathrm{Var}[V_1 + \cdots + V_{30}]}{36} \\
&= \frac{\mathrm{Var}[V_1] + \cdots + \mathrm{Var}[V_{30}]}{36} = \frac{30}{36}\,\mathrm{Var}[V] = \frac{250}{21}. \qquad (4)
\end{aligned}
$$

Since X is a sum of 30 iid random variables, it is reasonable to make the central limit theorem approximation

$$
\begin{aligned}
\mathrm{P}[X \geq 95] &= \mathrm{P}\left[\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}} \geq \frac{95 - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}}\right] \\
&\approx \mathrm{P}\left[Z \geq \frac{95 - 100}{\sqrt{250/21}}\right] \\
&= Q\left(-\sqrt{\frac{21}{10}}\right) = \Phi(\sqrt{2.1}) = 0.9264. \qquad (5)
\end{aligned}
$$

Problem 9.5.9 Solution

By symmetry, $\mathrm{E}[X] = 0$. Since X is a continuous $(a = -1, b = 1)$ uniform random variable, its variance is $\mathrm{Var}[X] = (b-a)^2/12 = 1/3$. Working with the moments of X, we can write

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}\left[20 + 15X^2\right] \\ &= 20 + 15\,\mathrm{E}\left[X^2\right] \\ &= 20 + 15\,\mathrm{Var}[X^2] = 25, \end{aligned} \tag{1}$$

where we recall that $\mathrm{E}[X] = 0$ implies $\mathrm{E}[X^2] = \mathrm{Var}[X]$. Next we observe that

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{Var}[20 + 15X^2] \\ &= \mathrm{Var}[15X^2] \\ &= 225\,\mathrm{Var}[X^2] = 225\left(\mathrm{E}\left[(X^2)^2\right] - \left(\mathrm{E}\left[X^2\right]\right)^2\right) \end{aligned} \tag{2}$$

Since $\mathrm{E}[X^2] = \mathrm{Var}[X] = 1/3$, and since $\mathrm{E}[(X^2)^2] = \mathrm{E}[X^4]$, it follows that

$$\begin{aligned} \mathrm{Var}[Y] &= 225\left(\mathrm{E}\left[X^4\right] - (\mathrm{Var}[X])^2\right) \\ &= 225\left(\int_{-1}^{1} \frac{1}{2} x^4\,dx - \left(\frac{1}{3}\right)^2\right) = 225\left(\frac{1}{5} - \frac{1}{9}\right) = 20. \end{aligned} \tag{3}$$

To use the central limit theorem, we approximate the CDF of W by a Gaussian CDF with expected value $\mathrm{E}[W]$ and variance $\mathrm{Var}[W]$. Since the expectation of the sum equals the sum of the expectations,

$$\begin{aligned} \mathrm{E}[W] &= \mathrm{E}\left[\frac{1}{100}\sum_{i=1}^{100} Y_i\right] \\ &= \frac{1}{100}\sum_{i=1}^{100} \mathrm{E}[Y_i] = \mathrm{E}[Y] = 25. \end{aligned} \tag{4}$$

Since the independence of the Y_i follows from the independence of the X_i, we can write

$$\mathrm{Var}[W] = \frac{1}{100^2}\,\mathrm{Var}\left[\sum_{i=1}^{100} Y_i\right] = \frac{1}{100^2}\sum_{i=1}^{100} \mathrm{Var}[Y_i] = \frac{\mathrm{Var}[Y]}{100} = 0.2. \tag{5}$$

By a CLT approximation,

$$\begin{aligned}
\mathrm{P}[W \leq 25.4] &= \mathrm{P}\left[\frac{W-25}{\sqrt{0.2}} \leq \frac{25.4-25}{\sqrt{0.2}}\right] \\
&\approx \Phi\left(\frac{0.4}{\sqrt{0.2}}\right) = \Phi(2/\sqrt{5}) = 0.8145. \qquad (6)
\end{aligned}$$

Problem 9.5.10 Solution

(a) If $v = 15$, then

$$Y = 50 + (15 + W - 15)^3 = 50 + W^3 \quad \text{Watts.} \qquad (1)$$

First we note that the PDF and CDF of W are

$$f_W(w) = \begin{cases} 1/10 & 0 \leq w \leq 10, \\ 0 & \text{otherwise,} \end{cases}$$

$$F_W(w) = \begin{cases} 0 & w < 0, \\ w/10 & 0 \leq w \leq 10, \\ 1 & w > 10. \end{cases} \qquad (2)$$

Since $Y = 50 + W^3$, we can calculate

$$\begin{aligned}
\mathrm{E}[Y] &= 50 + \mathrm{E}\left[W^3\right] \\
&= 50 + \int_0^{10} w^3(1/10)\,dw \\
&= 50 + \left(w^4/40\Big|_0^{10}\right) = 300. \qquad (3)
\end{aligned}$$

(b) Since Lance rides at a constant 15 mi/hr, he finishes in $3000/15 = 200$ hours. Using W_i for the wind speed over mile i, Ashwin's speed over mile i is $\hat{V}_i = a - W_i$ where $a = 15 + (\hat{y} - 50)^{1/3}$.

Ashwin's time for mile i is

$$T_i = \frac{1}{\hat{V}_i} = \frac{1}{a - W_i}. \tag{4}$$

Ashwin's total race time is

$$T = T_1 + T_2 + \cdots + T_{3000}. \tag{5}$$

Since the W_i are iid, the T_i are also iid and T is an iid sum. We can use the central limit theorem to estimate $P[A] = P[T < 200]$. However, because T can be approximated as Gaussian, $P[T < 200] = 1/2$ if $E[T] = 200$.

The expected time for mile i is

$$\begin{aligned} E[T_i] = E\left[\frac{1}{a - W_i}\right] &= \frac{1}{10}\int_0^{10} \frac{1}{a-w}\, dw \\ &= -\frac{\ln(a-w)}{10)}\bigg|_0^{10} = \frac{1}{10}\ln\left(\frac{a}{a-10}\right). \end{aligned} \tag{6}$$

Thus T has expected value

$$E[T] = 3000\, E[T_i] = 300 \ln \frac{a}{a-10} \tag{7}$$

and $P[A] = 1/2$ if

$$300 \ln \frac{a}{a-10} = 200 \qquad \Longrightarrow \qquad a = 10\frac{e^{2/3}}{e^{2/3}-1} = 20.5515. \tag{8}$$

This implies

$$\hat{y} = 50 + (a - 15)^3 = 221.09 \text{ Watts}. \tag{9}$$

Ashwin has an equal chance of winning the race even though his power output is almost 79 Watts less than Lance. In fact, Ashwin rides at an average speed of $E[\hat{V}] = a - E[W_i] = 15.5515$ mi/hr, or 0.55 mi/hr faster than Lance. However, because his speed varies, his higher average speed only results in an equal chance of winning.

Problem 9.5.11 Solution

(a) On quiz i, your score X_i is the sum of $n = 10$ independent Bernoulli trials and so X_i is a binomial $(n = 10, p = 0.8)$ random variable, which has PMF

$$P_{X_i}(x) = \binom{10}{x}(0.8)^x(0.2)^{10-x}. \tag{1}$$

(b) First we note that $\mathrm{E}[X_i] = np = 8$ and that

$$\mathrm{Var}[X_i] = np(1-p) = 10(0.8)(0.2) = 1.6. \tag{2}$$

Since X is a scaled sum of 100 Bernoulli trials, it is appropriate to use a central limit theorem approximation. All we need to do is calculate the expected value and variance of X:

$$\mu_X = \mathrm{E}[X] = 0.01\sum_{i=1}^{10}\mathrm{E}[X_i] = 0.8, \tag{3}$$

$$\begin{aligned}\sigma_X^2 = \mathrm{Var}[X] &= (0.01)^2\,\mathrm{Var}\left[\sum_{i=1}^{10} X_i\right]\\ &= 10^{-4}\sum_{i=1}^{10}\mathrm{Var}[X_i] = 16\times 10^{-4}.\end{aligned} \tag{4}$$

To use the central limit theorem, we write

$$\begin{aligned}\mathrm{P}[A] = \mathrm{P}[X \geq 0.9] &= \mathrm{P}\left[\frac{X-\mu_X}{\sigma_X} \geq \frac{0.9-\mu_X}{\sigma_X}\right]\\ &\approx \mathrm{P}\left[Z \geq \frac{0.9-0.8}{0.04}\right]\\ &= \mathrm{P}[Z \geq 2.5] = Q(2.5).\end{aligned} \tag{5}$$

A nicer way to do this same calculation is to observe that

$$\mathrm{P}[A] = \mathrm{P}[X \geq 0.9] = \mathrm{P}\left[\sum_{i=1}^{10} X_i \geq 90\right]. \tag{6}$$

Now we define $W = \sum_{i=1}^{10} X_i$ and use the central limit theorem on W. In this case,

$$\mathrm{E}[W] = 10\,\mathrm{E}[X_i] = 80, \qquad \mathrm{Var}[W] = 10\,\mathrm{Var}[X_i] = 16. \tag{7}$$

Our central limit theorem approximation can now be written as

$$\begin{aligned} \mathrm{P}[A] = \mathrm{P}[W \geq 90] &= \mathrm{P}\left[\frac{W-80}{\sqrt{16}} \geq \frac{90-80}{\sqrt{16}}\right] \\ &\approx \mathrm{P}[Z \geq 2.5] = Q(2.5). \end{aligned} \tag{8}$$

We will see that this second approach is more useful in the next problem.

(c) With n attendance quizzes,

$$\begin{aligned} \mathrm{P}[A] &= \mathrm{P}\left[X' \geq 0.9\right] \\ &= \mathrm{P}\left[10n + \sum_{i=1}^{10} X_i \geq 9n + 90\right] = \mathrm{P}[W \geq 90 - n], \end{aligned} \tag{9}$$

where $W = \sum_{i=1}^{10} X_i$ is the same as in the previous part. Thus

$$\begin{aligned} \mathrm{P}[A] &= \mathrm{P}\left[\frac{W - \mathrm{E}[W]}{\sqrt{\mathrm{Var}[W]}} \geq \frac{90 - n - \mathrm{E}[W]}{\sqrt{\mathrm{Var}[W]}}\right] \\ &= Q\left(\frac{10-n}{4}\right) = Q(2.5 - 0.25n). \end{aligned} \tag{10}$$

(d) Without the scoring change on quiz 1, your grade will be based on

$$X = \frac{8 + \sum_{i=2}^{10} X_i}{100} = \frac{8+Y}{100}. \tag{11}$$

With the corrected scoring, your grade will be based on

$$X' = \frac{9 + \sum_{i=2}^{10} X_i}{100} = \frac{9+Y}{100} = 0.01 + X. \tag{12}$$

The only time this change will matter is when X is on the borderline between two grades. Specifically, your grade will change if $X \in \{0.59, 0.69, 0.79, 0.89\}$. Equivalently,

$$\begin{aligned} \mathrm{P}[U^c] &= \mathrm{P}[Y = 51] + \mathrm{P}[Y = 61] + \mathrm{P}[Y = 71] + \mathrm{P}[Y = 81] \\ &= P_Y(51) + P_Y(61) + P_Y(71) + P_Y(81). \end{aligned} \tag{13}$$

If you're curious, we note that since Y is binomial with $\mathrm{E}[Y] = 72$, the dominant term in the above sum is $P_Y(71)$ and that

$$\mathrm{P}[U^c] \approx \binom{90}{71}(0.8)^{71}(0.2)^{19} \approx 0.099. \tag{14}$$

This near 10 percent probability is fairly high because the student is a borderline B/C student. That is, the point matters if you are a borderline student. Of course, in real life, you don't know if you're a borderline student.

Problem 9.6.1 Solution

Note that W_n is a binomial $(10^n, 0.5)$ random variable. We need to calculate

$$\begin{aligned} \mathrm{P}[B_n] &= \mathrm{P}[0.499 \times 10^n \leq W_n \leq 0.501 \times 10^n] \\ &= \mathrm{P}[W_n \leq 0.501 \times 10^n] - \mathrm{P}[W_n < 0.499 \times 10^n]. \end{aligned} \tag{1}$$

A complication is that the event $W_n < w$ is not the same as $W_n \leq w$ when w is an integer. In this case, we observe that

$$\mathrm{P}[W_n < w] = \mathrm{P}[W_n \leq \lceil w \rceil - 1] = F_{W_n}(\lceil w \rceil - 1). \tag{2}$$

Thus

$$\mathrm{P}[B_n] = F_{W_n}(0.501 \times 10^n) - F_{W_n}\left(\lceil 0.499 \times 10^9 \rceil - 1\right). \tag{3}$$

For $n = 1, \ldots, N$, we can calculate $\mathrm{P}[B_n]$ in this MATLAB program:

```
function pb=binomialcdftest(N);
pb=zeros(1,N);
for n=1:N,
    w=[0.499 0.501]*10^n;
    w(1)=ceil(w(1))-1;
    pb(n)=diff(binomialcdf(10^n,0.5,w));
end
```

Unfortunately, on this user's machine (a Windows XP laptop), the program fails for $N = 4$. The problem, as noted earlier is that `binomialcdf.m` uses `binomialpmf.m`, which fails for a binomial $(10000, p)$ random variable. Of course, your mileage may vary. A slightly better solution is to use the `bignomialcdf.m` function, which is identical to `binomialcdf.m` except it calls `bignomialpmf.m` rather than `binomialpmf.m`. This enables calculations for larger values of n, although at some cost in numerical accuracy. Here is the code:

```
function pb=bignomialcdftest(N);
pb=zeros(1,N);
for n=1:N,
    w=[0.499 0.501]*10^n;
    w(1)=ceil(w(1))-1;
    pb(n)=diff(bignomialcdf(10^n,0.5,w));
end
```

For comparison, here are the outputs of the two programs:

```
>> binomialcdftest(4)
ans =
  0.2461   0.0796   0.0756      NaN
>> bignomialcdftest(6)
ans =
  0.2461   0.0796   0.0756   0.1663   0.4750   0.9546
```

The result 0.9546 for $n = 6$ corresponds to the exact probability in Example 9.14 which used the CLT to estimate the probability as 0.9544. Unfortunately for this user, `bignomialcdftest(7)` failed.

Problem 9.6.2 Solution

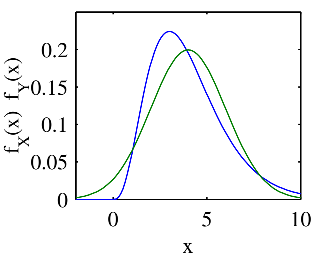
The Erlang $(n, \lambda = 1)$ random variable X has expected value $E[X] = n/\lambda = n$ and variance $\mathrm{Var}[X] = n/\lambda^2 = n$. The PDF of X as well as the PDF of a Gaussian random variable Y with the same expected value and variance are

$$f_X(x) = \begin{cases} \dfrac{x^{n-1}e^{-x}}{(n-1)!} & x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(x) = \frac{1}{\sqrt{2\pi n}} e^{-x^2/2n}. \tag{1}$$

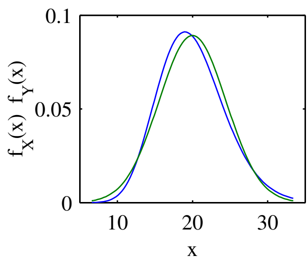
```
function df=erlangclt(n);
r=3*sqrt(n);
x=(n-r):(2*r)/100:n+r;
fx=erlangpdf(n,1,x);
fy=gausspdf(n,sqrt(n),x);
plot(x,fx,x,fy);
df=fx-fy;
```

From the forms of the functions, it not likely to be apparent that $f_X(x)$ and $f_Y(x)$ are similar. The following program plots $f_X(x)$ and $f_Y(x)$ for values of x within three standard deviations of the expected value n. Below are sample outputs of `erlangclt(n)` for $n = 4, 20, 100$.

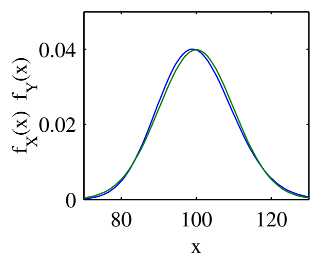
In the graphs we will see that as n increases, the Erlang PDF becomes increasingly similar to the Gaussian PDF of the same expected value and variance. This is not surprising since the Erlang (n, λ) random variable is the sum of n of exponential random variables and the CLT says that the Erlang CDF should converge to a Gaussian CDF as n gets large.



`erlangclt(4)`



`erlangclt(20)`



`erlangclt(100)`

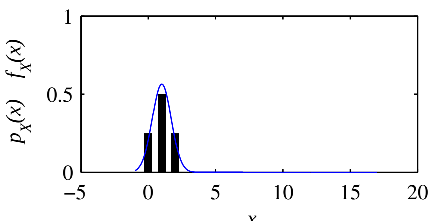
On the other hand, the convergence should be viewed with some caution. For example, the mode (the peak value) of the Erlang PDF occurs at $x = n - 1$ while the mode of the Gaussian PDF is at $x = n$. This difference only appears to go away for $n = 100$ because the graph x-axis range is expanding. More important, the two PDFs are quite different far away from the center of the distribution. The Erlang PDF is always zero for $x < 0$ while the Gaussian PDF is always positive. For large postive x, the two distributions do not have the same exponential decay. Thus it's not a good idea to use the CLT to estimate probabilities of rare events such as $\{X > x\}$ for extremely large values of x.

Problem 9.6.3 Solution

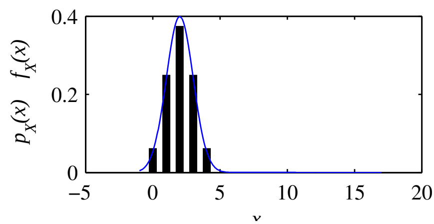
In this problem, we re-create the plots of Figure 9.3 except we use the binomial PMF and corresponding Gaussian PDF. Here is a MATLAB program that compares the binomial (n, p) PMF and the Gaussian PDF with the same expected value and variance.

```
function y=binomcltpmf(n,p)
x=-1:17;
xx=-1:0.05:17;
y=binomialpmf(n,p,x);
std=sqrt(n*p*(1-p));
clt=gausspdf(n*p,std,xx);
hold off;
pmfplot(x,y,'\it x','\it p_X(x)    f_X(x)');
hold on; plot(xx,clt); hold off;
```

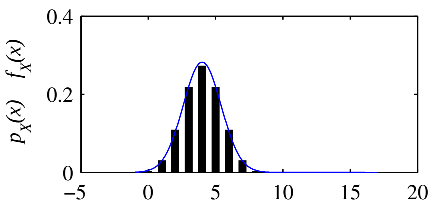
Here are the output plots for $p = 1/2$ and $n = 2, 4, 8, 16$.



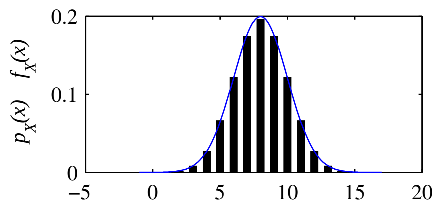
binomcltpmf(2,0.5)



binomcltpmf(4,0.5)



binomcltpmf(8,0.5)



binomcltpmf(16,0.5)

To see why the values of the PDF and PMF are roughly the same, consider the Gaussian random variable Y. For small Δ,

$$f_Y(x)\,\Delta \approx \frac{F_Y(x + \Delta/2) - F_Y(x - \Delta/2)}{\Delta}. \tag{1}$$

For $\Delta = 1$, we obtain

$$f_Y(x) \approx F_Y(x + 1/2) - F_Y(x - 1/2). \tag{2}$$

Since the Gaussian CDF is approximately the same as the CDF of the binomial (n, p) random variable X, we observe for an integer x that

$$f_Y(x) \approx F_X(x + 1/2) - F_X(x - 1/2) = P_X(x) . \tag{3}$$

Although the equivalence in heights of the PMF and PDF is only an approximation, it can be useful for checking the correctness of a result.

Problem 9.6.4 Solution

Since the `conv` function is for convolving signals in time, we treat $P_{X_1}(x)$ and $P_{X_2}(x_2)x$, or as though they were signals in time starting at time $x = 0$. That is,

$$\texttt{px1} = \begin{bmatrix} P_{X_1}(0) & P_{X_1}(1) & \cdots & P_{X_1}(25) \end{bmatrix}, \tag{1}$$
$$\texttt{px2} = \begin{bmatrix} P_{X_2}(0) & P_{X_2}(1) & \cdots & P_{X_2}(100) \end{bmatrix}. \tag{2}$$

```
%convx1x2.m
sw=(0:125);
px1=[0,0.04*ones(1,25)];
px2=zeros(1,101);
px2(10*(1:10))=10*(1:10)/550;
pw=conv(px1,px2);
h=pmfplot(sw,pw,...
    '\itw','\itP_W(w)');
set(h,'LineWidth',0.25);
```

In particular, between its minimum and maximum values, the vector `px2` must enumerate all integer values, including those which have zero probability. In addition, we write down `sw=0:125` directly based on knowledge that the range enumerated by `px1` and `px2` corresponds to $X_1 + X_2$ having a minimum value of 0 and a maximum value of 125.

The resulting plot will be essentially identical to Figure 9.4. One final note, the command `set(h,'LineWidth',0.25)` is used to make the bars of the PMF thin enough to be resolved individually.

Problem 9.6.5 Solution

In Example 10.4, the height of a storm surge X is a Gaussian $(5.5, 1)$ random variable. Here were are asked to estimate

$$\mathrm{P}[X > 7] = \mathrm{P}[X - 5.5 > 1.5] = 1 - \Phi(1.5). \tag{1}$$

using the `uniform12.m` function defined in Example 9.18.

The exact correct value is $1 - \Phi(1.5) = 0.0668$. You may wonder why this problem asks you to estimate $1-\Phi(1.5)$ when we can calculate it exactly. The goal of this exercise is really to understand the limitations of using a sum of 12 uniform random variables as an approximation to a Gaussian.

Unfortunately, in the function `uniform12.m`, the vector `T=(-3:3)` is hard-coded, making it hard to directly reuse the function to solve our problem. So instead, let's redefine a new `unif12sum.m` function that accepts the number of trials `m` and the threshold value `T` as arguments:

```
function FT = unif12sum(m,T)
%Using m samples of a sum of 12 uniform random variables,
%FT is an estimate of P(X<T) for a Gaussian (0,1) rv X
x=sum(rand(12,m))-6;
FT=(count(x,T)/m)';
end
```

Before looking at some experimental runs, we note that `unif12sum` is making two different approximations. First, samples consisting of the sum of 12 uniform random variables are being used as an approximation for a Gaussian $(0, 1)$ random variable X. Second, we are using the relative frequency of samples below the threshold T as an approximation or estimate of $\mathrm{P}[X < T]$.

(a) Here are some sample runs for $m = 1000$ sample values:

```
>> m=1000;t=1.5;
>> 1-[unif12sum(m,t) unif12sum(m,t) unif12sum(m,t)]
ans =
    0.0640    0.0620    0.0610
>> 1-[unif12sum(m,t) unif12sum(m,t) unif12sum(m,t)]
ans =
    0.0810    0.0670    0.0690
```

We see that six trials yields six close but different estimates.

(b) Here are some sample runs for $m = 10{,}000$ sample values:

```
>> m=10000;t=1.5;
>> 1-[unif12sum(m,t) unif12sum(m,t) unif12sum(m,t)]
ans =
    0.0667    0.0709    0.0697
>> 1-[unif12sum(m,t) unif12sum(m,t) unif12sum(m,t)]
ans =
    0.0686    0.0672    0.0708
```

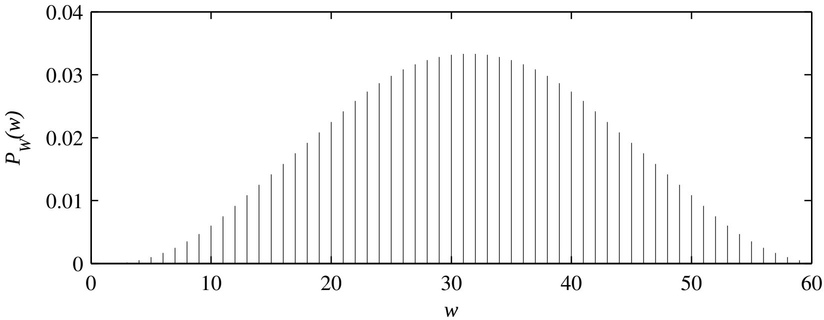
Casual inspection gives the impression that 10,000 samples provide better estimates than 1000 samples. Although the small number of tests here is definitely not sufficient to make such an assertion, we will see in Chapter 10 that this is indeed true.

Problem 9.6.6 Solution

```
sx1=(1:10);px1=0.1*ones(1,10);
sx2=(1:20);px2=0.05*ones(1,20);
sx3=(1:30);px3=ones(1,30)/30;
[SX1,SX2,SX3]=ndgrid(sx1,sx2,sx3);
[PX1,PX2,PX3]=ndgrid(px1,px2,px3);
SW=SX1+SX2+SX3;
PW=PX1.*PX2.*PX3;
sw=unique(SW);
pw=finitepmf(SW,PW,sw);
h=pmfplot(sw,pw,'\itw','\itP_W(w)');
set(h,'LineWidth',0.25);
```

Since the `mdgrid` function extends naturally to higher dimensions, this solution follows the logic of `sumx1x2` in Example 9.17.

The output of `sumx1x2x3` is the plot of the PMF of W shown below. We use the command `set(h,'LineWidth',0.25)` to ensure that the bars of the PMF are thin enough to be resolved individually.



Problem 9.6.7 Solution

The function `sumfinitepmf` generalizes the method of Example 9.17.

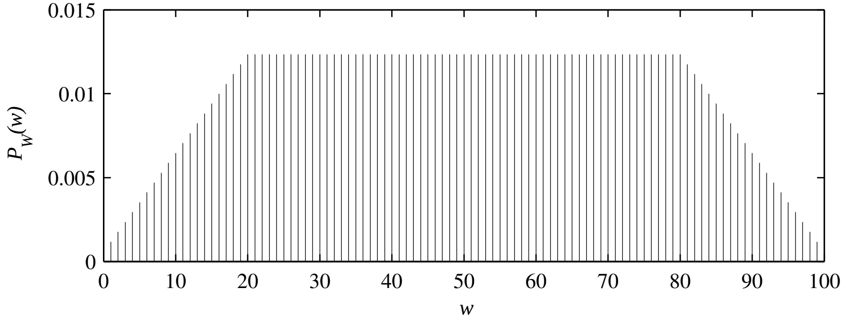
```
function [pw,sw]=sumfinitepmf(px,sx,py,sy);
[SX,SY]=ndgrid(sx,sy);
[PX,PY]=ndgrid(px,py);
SW=SX+SY;PW=PX.*PY;
sw=unique(SW);
pw=finitepmf(SW,PW,sw);
```

The only difference is that the PMFs `px` and `py` and ranges `sx` and `sy` are not hard coded, but instead are function inputs.

As an example, suppose X is a discrete uniform $(0, 20)$ random variable and Y is an independent discrete uniform $(0, 80)$ random variable. The following program `sum2unif` will generate and plot the PMF of $W = X + Y$.

```
%sum2unif.m
sx=0:20;px=ones(1,21)/21;
sy=0:80;py=ones(1,81)/81;
[pw,sw]=sumfinitepmf(px,sx,py,sy);
h=pmfplot(sw,pw,'\it w','\it P_W(w)');
set(h,'LineWidth',0.25);
```

Here is the graph generated by `sum2unif`.


0.015
0.01
0.005
0
$P_W(w)$
0
10
20
30
40
50
60
70
80
90
100
w

Problem Solutions – Chapter 10

Problem 10.1.1 Solution

Recall that $X_1, X_2 \ldots X_n$ are independent exponential random variables with mean value $\mu_X = 5$ so that for $x \geq 0$, $F_X(x) = 1 - e^{-x/5}$.

(a) Using Theorem 10.1, $\sigma^2_{M_n(x)} = \sigma^2_X/n$. Realizing that $\sigma^2_X = 25$, we obtain

$$\text{Var}[M_9(X)] = \frac{\sigma^2_X}{9} = \frac{25}{9}. \tag{1}$$

(b)

$$\begin{aligned} P[X_1 \geq 7] &= 1 - P[X_1 \leq 7] \\ &= 1 - F_X(7) \\ &= 1 - (1 - e^{-7/5}) = e^{-7/5} \approx 0.247. \end{aligned} \tag{2}$$

(c) First we express $P[M_9(X) > 7]$ in terms of $X_1, \ldots, X_9$.

$$\begin{aligned} P[M_9(X) > 7] &= 1 - P[M_9(X) \leq 7] \\ &= 1 - P[(X_1 + \ldots + X_9) \leq 63]. \end{aligned} \tag{3}$$

Now the probability that $M_9(X) > 7$ can be approximated using the Central Limit Theorem (CLT).

$$\begin{aligned} P[M_9(X) > 7] &= 1 - P[(X_1 + \ldots + X_9) \leq 63] \\ &\approx 1 - \Phi\left(\frac{63 - 9\mu_X}{\sqrt{9}\sigma_X}\right) \\ &= 1 - \Phi(6/5). \end{aligned} \tag{4}$$

Consulting with Table 4.2 yields $P[M_9(X) > 7] \approx 0.1151$.

Problem 10.1.2 Solution

$X_1, X_2 \ldots X_n$ are independent uniform random variables with mean value $\mu_X = 7$ and $\sigma_X^2 = 3$

(a) Since X_1 is a uniform random variable, it must have a uniform PDF over an interval $[a, b]$. From Appendix A, we can look up that $\mu_X = (a+b)/2$ and that $\text{Var}[X] = (b-a)^2/12$. Hence, given the mean and variance, we obtain the following equations for a and b.

$$(b-a)^2/12 = 3, \qquad (a+b)/2 = 7. \tag{1}$$

Solving these equations yields $a = 4$ and $b = 10$ from which we can state the distribution of X.

$$f_X(x) = \begin{cases} 1/6 & 4 \le x \le 10, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) From Theorem 10.1, we know that

$$\text{Var}[M_{16}(X)] = \frac{\text{Var}[X]}{16} = \frac{3}{16}. \tag{3}$$

(c)

$$\text{P}[X_1 \ge 9] = \int_9^\infty f_{X_1}(x)\, dx = \int_9^{10} (1/6)\, dx = 1/6. \tag{4}$$

(d) The variance of $M_{16}(X)$ is much less than $\text{Var}[X_1]$. Hence, the PDF of $M_{16}(X)$ should be much more concentrated about $\text{E}[X]$ than the PDF of X_1. Thus we should expect $\text{P}[M_{16}(X) > 9]$ to be much less than $\text{P}[X_1 > 9]$.

$$\begin{aligned} \text{P}[M_{16}(X) > 9] &= 1 - \text{P}[M_{16}(X) \le 9] \\ &= 1 - \text{P}[(X_1 + \cdots + X_{16}) \le 144]. \end{aligned} \tag{5}$$

By a Central Limit Theorem approximation,

$$\begin{aligned} \text{P}[M_{16}(X) > 9] &\approx 1 - \Phi\left(\frac{144 - 16\mu_X}{\sqrt{16}\sigma_X}\right) \\ &= 1 - \Phi(2.66) = 0.0039. \end{aligned} \tag{6}$$

As we predicted, $\text{P}[M_{16}(X) > 9] \ll \text{P}[X_1 > 9]$.

Problem 10.1.3 Solution

This problem is in the wrong section since the *standard error* isn't defined until Section 10.4. However is we peek ahead to this section, the problem isn't very hard. Given the sample mean estimate $M_n(X)$, the standard error is defined as the standard deviation $e_n = \sqrt{\text{Var}[M_n(X)]}$. In our problem, we use samples X_i to generate $Y_i = X_i^2$. For the sample mean $M_n(Y)$, we need to find the standard error

$$e_n = \sqrt{\text{Var}[M_n(Y)]} = \sqrt{\frac{\text{Var}[Y]}{n}}. \tag{1}$$

Since X is a uniform $(0, 1)$ random variable,

$$\text{E}[Y] = \text{E}\left[X^2\right] = \int_0^1 x^2\,dx = 1/3, \tag{2}$$

$$\text{E}\left[Y^2\right] = \text{E}\left[X^4\right] = \int_0^1 x^4\,dx = 1/5. \tag{3}$$

Thus $\text{Var}[Y] = 1/5 - (1/3)^2 = 4/45$ and the sample mean $M_n(Y)$ has standard error

$$e_n = \sqrt{\frac{4}{45n}}. \tag{4}$$

Problem 10.1.4 Solution

(a) Since $Y_n = X_{2n-1} + (-X_{2n})$, Theorem 9.1 says that the expected value of the difference is

$$\text{E}[Y] = \text{E}[X_{2n-1}] + \text{E}[-X_{2n}] = \text{E}[X] - \text{E}[X] = 0. \tag{1}$$

By Theorem 9.2, the variance of the difference between X_{2n-1} and X_{2n} is

$$\text{Var}[Y_n] = \text{Var}[X_{2n-1}] + \text{Var}[-X_{2n}] = 2\,\text{Var}[X]. \tag{2}$$

(b) Each Y_n is the difference of two samples of X that are independent of the samples used by any other Y_m. Thus $Y_1, Y_2, \ldots$ is an iid random sequence. By Theorem 10.1, the mean and variance of $M_n(Y)$ are

$$\mathrm{E}[M_n(Y)] = \mathrm{E}[Y_n] = 0, \tag{3}$$

$$\mathrm{Var}[M_n(Y)] = \frac{\mathrm{Var}[Y_n]}{n} = \frac{2\,\mathrm{Var}[X]}{n}. \tag{4}$$

Problem 10.2.1 Solution

If the average weight of a Maine black bear is 500 pounds with standard deviation equal to 100 pounds, we can use the Chebyshev inequality to upper bound the probability that a randomly chosen bear will be more then 200 pounds away from the average.

$$\mathrm{P}[|W - \mathrm{E}[W]| \geq 200] \leq \frac{\mathrm{Var}[W]}{200^2} \leq \frac{100^2}{200^2} = 0.25. \tag{1}$$

Problem 10.2.2 Solution

We know from the Chebyshev inequality that

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\sigma_X^2}{c^2}. \tag{1}$$

Choosing $c = k\sigma_X$, we obtain

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq k\sigma] \leq \frac{1}{k^2}. \tag{2}$$

The actual probability the Gaussian random variable Y is more than k standard deviations from its expected value is

$$\begin{aligned}\mathrm{P}[|Y - \mathrm{E}[Y]| \geq k\sigma_Y] &= \mathrm{P}[Y - \mathrm{E}[Y] \leq -k\sigma_Y] + \mathrm{P}[Y - \mathrm{E}[Y] \geq k\sigma_Y] \\ &= 2\,\mathrm{P}\left[\frac{Y - \mathrm{E}[Y]}{\sigma_Y} \geq k\right] \\ &= 2Q(k). \end{aligned} \tag{3}$$

The following table compares the upper bound and the true probability:

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Chebyshev bound	1	0.250	0.111	0.0625	0.040
$2Q(k)$	0.317	0.046	0.0027	6.33×10^{-5}	5.73×10^{-7}

The Chebyshev bound gets increasingly weak as k goes up. As an example, for $k = 4$, the bound exceeds the true probability by a factor of 1,000 while for $k = 5$ the bound exceeds the actual probability by a factor of nearly 100,000.

Problem 10.2.3 Solution

The arrival time of the third elevator is $W = X_1+X_2+X_3$. Since each X_i is uniform $(0, 30)$, $\mathrm{E}[X_i] = 15$ and $\mathrm{Var}[X_i] = (30 - 0)^2/12 = 75$. Thus $\mathrm{E}[W] = 3\,\mathrm{E}[X_i] = 45$, and $\mathrm{Var}[W] = 3\,\mathrm{Var}[X_i] = 225$.

(a) By the Markov inequality,

$$\mathrm{P}[W > 75] \leq \frac{\mathrm{E}[W]}{75} = \frac{45}{75} = \frac{3}{5}. \tag{1}$$

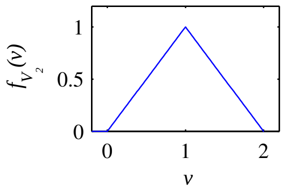
(b) By the Chebyshev inequality,

$$\begin{aligned} \mathrm{P}[W > 75] &= \mathrm{P}[W - \mathrm{E}[W] > 30] \\ &\leq \mathrm{P}[|W - \mathrm{E}[W]| > 30] \\ &\leq \frac{\mathrm{Var}[W]}{30^2} = \frac{1}{4}. \end{aligned} \tag{2}$$

Problem 10.2.4 Solution

The hard part of this problem is to derive the PDF of the sum $W = X_1 + X_2 + X_3$ of iid uniform $(0, 30)$ random variables. In this case, we need to use the techniques of Chapter 9 to convolve the three PDFs. To simplify our calculations, we will instead find the PDF of $V = Y_1 + Y_2 + Y_3$ where the Y_i are iid uniform $(0, 1)$ random variables. By Theorem 6.3 to conclude that $W = 30V$ is the sum of three iid uniform $(0, 30)$ random variables.

To start, let $V_2 = Y_1 + Y_2$. Since each Y_1 has a PDF shaped like a unit area pulse, the PDF of V_2 is the triangular function



$$f_{V_2}(v) = \begin{cases} v & 0 \leq v \leq 1, \\ 2 - v & 1 < v \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

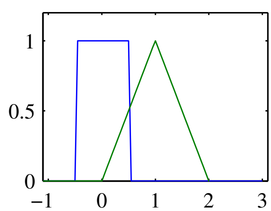
The PDF of V is the convolution integral

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_{V_2}(y)\, f_{Y_3}(v-y)\, dy \\ &= \int_0^1 y f_{Y_3}(v-y)\, dy + \int_1^2 (2-y) f_{Y_3}(v-y)\, dy. \end{aligned} \tag{2}$$

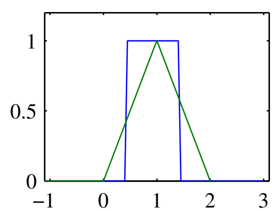
Evaluation of these integrals depends on v through the function

$$f_{Y_3}(v-y) = \begin{cases} 1 & v-1 < v < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

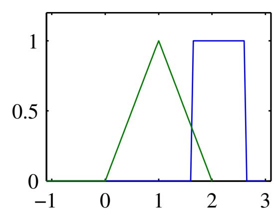
To compute the convolution, it is helpful to depict the three distinct cases. In each case, the square "pulse" is $f_{Y_3}(v-y)$ and the triangular pulse is $f_{V_2}(y)$.



$0 \leq v < 1$



$1 \leq v < 2$



$2 \leq v < 3$

From the graphs, we can compute the convolution for each case:

$$0 \leq v < 1: \qquad f_{V_3}(v) = \int_0^v y\, dy = \frac{1}{2} v^2, \tag{4}$$

$$1 \leq v < 2: \qquad f_{V_3}(v) = \int_{v-1}^1 y\, dy + \int_1^v (2-y)\, dy = -v^2 + 3v - \frac{3}{2}, \tag{5}$$

$$2 \leq v < 3: \qquad f_{V_3}(v) = \int_{v-1}^2 (2-y)\, dy = \frac{(3-v)^2}{2}. \tag{6}$$

To complete the problem, we use Theorem 6.3 to observe that $W = 30V_3$ is the sum of three iid uniform $(0,30)$ random variables. From Theorem 6.2,

$$\begin{aligned} f_W(w) &= \frac{1}{30} f_{V_3}(v_3)\, v/30 \\ &= \begin{cases} (w/30)^2/60 & 0 \le w < 30, \\ [-(w/30)^2 + 3(w/30) - 3/2]/30 & 30 \le w < 60, \\ [3 - (w/30)]^2/60 & 60 \le w < 90, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{7}$$

Finally, we can compute the exact probability

$$\begin{aligned} P[W \ge 75] &= \frac{1}{60} \int_{75}^{9} 0[3 - (w/30)]^2 \, dw \\ &= -\frac{(3 - w/30)^3}{6} \bigg|_{75}^{90} = \frac{1}{48}. \end{aligned} \tag{8}$$

For comparison, the Markov inequality indicated that $P[W < 75] \le 3/5$ and the Chebyshev inequality showed that $P[W < 75] \le 1/4$. As we see, both inequalities are quite weak in this case.

Problem 10.2.5 Solution

On each roll of the dice, a success, namely snake eyes, occurs with probability $p = 1/36$. The number of trials, R, needed for three successes is a Pascal $(k = 3, p)$ random variable with

$$E[R] = \frac{3}{p} = 108, \qquad \text{Var}[R] = \frac{3(1-p)}{p^2} = 3780. \tag{1}$$

(a) By the Markov inequality,

$$P[R \ge 250] \le \frac{E[R]}{250} = \frac{54}{125} = 0.432. \tag{2}$$

(b) By the Chebyshev inequality,

$$\begin{aligned} P[R \ge 250] = P[R - 108 \ge 142] &= P[|R - 108| \ge 142] \\ &\le \frac{\text{Var}[R]}{(142)^2} = 0.1875. \end{aligned} \tag{3}$$

(c) The exact value is $P[R \geq 250] = 1 - \sum_{r=3}^{249} P_R(r)$. Since there is no way around summing the Pascal PMF to find the CDF, this is what `pascalcdf` does.

```
>> 1-pascalcdf(3,1/36,249)
ans =
    0.0299
```

Thus the Markov and Chebyshev inequalities are valid bounds but not good estimates of $P[R \geq 250]$.

Problem 10.2.6 Solution

The $N[0,1]$ random variable Z has MGF $\phi_Z(s) = e^{s^2/2}$. Hence the Chernoff bound for Z is

$$P[Z \geq c] \leq \min_{s \geq 0} e^{-sc} e^{s^2/2} = \min_{s \geq 0} e^{s^2/2 - sc}. \tag{1}$$

We can minimize $e^{s^2/2-sc}$ by minimizing the exponent $s^2/2 - sc$. By setting

$$\frac{d}{ds}\left(s^2/2 - sc\right) = 2s - c = 0 \tag{2}$$

we obtain $s = c$. At $s = c$, the upper bound is $P[Z \geq c] \leq e^{-c^2/2}$. The table below compares this upper bound to the true probability. Note that for $c = 1, 2$ we use Table 4.2 and the fact that $Q(c) = 1 - \Phi(c)$.

	$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$
Chernoff bound	0.606	0.135	0.011	3.35×10^{-4}	3.73×10^{-6}
$Q(c)$	0.1587	0.0228	0.0013	3.17×10^{-5}	2.87×10^{-7}

We see that in this case, the Chernoff bound typically overestimates the true probability by roughly a factor of 10.

Problem 10.2.7 Solution

For an $N[\mu, \sigma^2]$ random variable X, we can write

$$P[X \geq c] = P[(X - \mu)/\sigma \geq (c - \mu)/\sigma] = P[Z \geq (c - \mu)/\sigma]. \tag{1}$$

Since Z is $N[0,1]$, we can apply the result of Problem 10.2.6 with c replaced by $(c - \mu)/\sigma$. This yields

$$P[X \geq c] = P[Z \geq (c - \mu)/\sigma] \leq e^{-(c-\mu)^2/2\sigma^2} \tag{2}$$

Problem 10.2.8 Solution

From Appendix A, we know that the MGF of K is

$$\phi_K(s) = e^{\alpha(e^s-1)}. \tag{1}$$

The Chernoff bound becomes

$$P[K \geq c] \leq \min_{s\geq 0} e^{-sc} e^{\alpha(e^s-1)} = \min_{s\geq 0} e^{\alpha(e^s-1)-sc}. \tag{2}$$

Since e^y is an increasing function, it is sufficient to choose s to minimize $h(s) = \alpha(e^s - 1) - sc$. Setting $dh(s)/ds = \alpha e^s - c = 0$ yields $e^s = c/\alpha$ or $s = \ln(c/\alpha)$. Note that for $c < \alpha$, the minimizing s is negative. In this case, we choose $s = 0$ and the Chernoff bound is $P[K \geq c] \leq 1$. For $c \geq \alpha$, applying $s = \ln(c/\alpha)$ yields $P[K \geq c] \leq e^{-\alpha}(\alpha e/c)^c$. A complete expression for the Chernoff bound is

$$P[K \geq c] \leq \begin{cases} 1 & c < \alpha, \\ \alpha^c e^c e^{-\alpha}/c^c & c \geq \alpha. \end{cases} \tag{3}$$

Problem 10.2.9 Solution

This problem is solved completely in the solution to Quiz 10.2! We repeat that solution here. Since $W = X_1 + X_2 + X_3$ is an Erlang $(n = 3, \lambda = 1/2)$ random variable, Theorem 4.11 says that for any $w > 0$, the CDF of W satisfies

$$F_W(w) = 1 - \sum_{k=0}^{2} \frac{(\lambda w)^k e^{-\lambda w}}{k!} \tag{1}$$

Equivalently, for $\lambda = 1/2$ and $w = 20$,

$$\begin{aligned} P[W > 20] &= 1 - F_W(20) \\ &= e^{-10}\left(1 + \frac{10}{1!} + \frac{10^2}{2!}\right) = 61e^{-10} = 0.0028. \end{aligned} \tag{2}$$

Problem 10.2.10 Solution

Let $W_n = X_1 + \cdots + X_n$. Since $M_n(X) = W_n/n$, we can write

$$P[M_n(X) \geq c] = P[W_n \geq nc]. \tag{1}$$

Since $\phi_{W_n}(s) = (\phi_X(s))^n$, applying the Chernoff bound to W_n yields

$$P[W_n \geq nc] \leq \min_{s \geq 0} e^{-snc}\phi_{W_n}(s) = \min_{s \geq 0} \left(e^{-sc}\phi_X(s)\right)^n. \tag{2}$$

For $y \geq 0$, y^n is a nondecreasing function of y. This implies that the value of s that minimizes $e^{-sc}\phi_X(s)$ also minimizes $(e^{-sc}\phi_X(s))^n$. Hence

$$P[M_n(X) \geq c] = P[W_n \geq nc] \leq \left(\min_{s \geq 0} e^{-sc}\phi_X(s)\right)^n. \tag{3}$$

Problem 10.3.1 Solution

$X_1, X_2, \ldots$ are iid random variables each with mean 75 and standard deviation 15.

(a) We would like to find the value of n such that

$$P[74 \leq M_n(X) \leq 76] = 0.99. \tag{1}$$

When we know only the mean and variance of X_i, our only real tool is the Chebyshev inequality which says that

$$\begin{aligned} P[74 \leq M_n(X) \leq 76] &= 1 - P[|M_n(X) - E[X]| \geq 1] \\ &\geq 1 - \frac{\text{Var}[X]}{n} = 1 - \frac{225}{n} \geq 0.99. \end{aligned} \tag{2}$$

This yields $n \geq 22{,}500$.

(b) If each X_i is a Gaussian, the sample mean, $M_n(X)$ will also be Gaussian with mean and variance

$$E[M_{n'}(X)] = E[X] = 75, \tag{3}$$

$$\text{Var}[M_{n'}(X)] = \text{Var}[X]/n' = 225/n' \tag{4}$$

In this case,

$$\begin{aligned}
\mathrm{P}\left[74 \leq M_{n'}(X) \leq 76\right] &= \Phi\left(\frac{76-\mu}{\sigma}\right) - \Phi\left(\frac{74-\mu}{\sigma}\right) \\
&= \Phi(\sqrt{n'}/15) - \Phi(-\sqrt{n'}/15) \\
&= 2\Phi(\sqrt{n'}/15) - 1 = 0.99. \qquad (5)
\end{aligned}$$

Thus, $n' = 1{,}521$.

Since even under the Gaussian assumption, the number of samples n' is so large that even if the X_i are not Gaussian, the sample mean may be approximated by a Gaussian. Hence, about 1500 samples probably is about right. However, in the absence of any information about the PDF of X_i beyond the mean and variance, we cannot make any guarantees stronger than that given by the Chebyshev inequality.

Problem 10.3.2 Solution

(a) Since X_A is a Bernoulli ($p = \mathrm{P}[A]$) random variable,

$$\mathrm{E}\left[X_A\right] = \mathrm{P}\left[A\right] = 0.8, \qquad \mathrm{Var}[X_A] = \mathrm{P}\left[A\right](1 - \mathrm{P}\left[A\right]) = 0.16. \qquad (1)$$

(b) Let $X_{A,i}$ to denote X_A on the ith trial. Since

$$\hat{P}_n(A) = M_n(X_A) = \frac{1}{n}\sum_{i=1}^{n} X_{A,i}, \qquad (2)$$

is a sum of n independent random variables,

$$\mathrm{Var}[\hat{P}_n(A)] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_{A,i}] = \frac{\mathrm{P}\left[A\right](1 - \mathrm{P}\left[A\right])}{n}. \qquad (3)$$

(c) Since $\hat{P}_{100}(A) = M_{100}(X_A)$, we can use Theorem 10.5(b) to write

$$\begin{aligned}
\mathrm{P}\left[\left|\hat{P}_{100}(A) - \mathrm{P}\left[A\right]\right| < c\right] &\geq 1 - \frac{\mathrm{Var}[X_A]}{100c^2} \\
&= 1 - \frac{0.16}{100c^2} = 1 - \alpha. \qquad (4)
\end{aligned}$$

For $c = 0.1$, $\alpha = 0.16/[100(0.1)^2] = 0.16$. Thus, with 100 samples, our confidence coefficient is $1 - \alpha = 0.84$.

(d) In this case, the number of samples n is unknown. Once again, we use Theorem 10.5(b) to write

$$\mathrm{P}\left[\left|\hat{P}_n(A) - \mathrm{P}\left[A\right]\right| < c\right] \geq 1 - \frac{\mathrm{Var}[X_A]}{nc^2}$$
$$= 1 - \frac{0.16}{nc^2} = 1 - \alpha. \tag{5}$$

For $c = 0.1$, we have confidence coefficient $1-\alpha = 0.95$ if $\alpha = 0.16/[n(0.1)^2] = 0.05$, or $n = 320$.

Problem 10.3.3 Solution

(a) As $n \to \infty$, Y_{2n} is a sum of a large number of iid random variables, so we can use the central limit theorem. Since $\mathrm{E}[Y_{2n}] = n$ and $\mathrm{Var}[Y_{2n}] = 2np(1-p) = n/2$,

$$\begin{aligned}
\mathrm{P}\left[|Y_{2n} - n| \leq \sqrt{n/2}\right] &= \mathrm{P}\left[Y_{2n} - \mathrm{E}\left[Y_{2n}\right] \leq \sqrt{n/2}\right] \\
&= \mathrm{P}\left[\frac{|Y_{2n} - \mathrm{E}\left[Y_{2n}\right]|}{\sqrt{\mathrm{Var}[Y_{2n}]}} \leq 1\right] \\
&= \mathrm{P}\left[-1 \leq Z_n \leq 1\right]. \qquad (1)
\end{aligned}$$

By the central limit theorem, $Z_n = (Y_{2n} - \mathrm{E}[Y_{2n}])/\sqrt{\mathrm{Var}[Y_{2n}]}$ is converging to a Gaussian $(0,1)$ random variable Z. Thus

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{P}\left[|Y_{2n} - n| \leq \sqrt{n/2}\right] &= \mathrm{P}\left[-1 \leq Z \leq 1\right] \\
&= \Phi(1) - \Phi(-1) \\
&= 2\Phi(1) - 1 = 0.68. \qquad (2)
\end{aligned}$$

(b) Note that $Y_{2n}/(2n)$ is a sample mean for $2n$ samples of X_n. Since $\mathrm{E}[X_n] = 1/2$, the weak law says that given any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left[\left|\frac{Y_{2n}}{2n} - \frac{1}{2}\right| > \epsilon\right] = 0. \tag{3}$$

An equivalent statement is

$$\lim_{n\to\infty} \mathrm{P}\left[|Y_{2n} - n| > 2n\epsilon\right] = 0. \tag{4}$$

Problem 10.3.4 Solution

Since the relative frequency of the error event E is $\hat{P}_n(E) = M_n(X_E)$ and $\mathrm{E}[M_n(X_E)] = \mathrm{P}[E]$, we can use Theorem 10.5(a) to write

$$\mathrm{P}\left[\left|\hat{P}_n(A) - \mathrm{P}\left[E\right]\right| \geq c\right] \leq \frac{\mathrm{Var}[X_E]}{nc^2}. \tag{1}$$

Note that $\mathrm{Var}[X_E] = \mathrm{P}[E](1 - \mathrm{P}[E])$ since X_E is a Bernoulli ($p = \mathrm{P}[E]$) random variable. Using the additional fact that $\mathrm{P}[E] \leq \epsilon$ and the fairly trivial fact that $1 - \mathrm{P}[E] \leq 1$, we can conclude that

$$\mathrm{Var}[X_E] = \mathrm{P}\left[E\right]\left(1 - \mathrm{P}\left[E\right]\right) \leq \mathrm{P}\left[E\right] \leq \epsilon. \tag{2}$$

Thus

$$\mathrm{P}\left[\left|\hat{P}_n(A) - \mathrm{P}\left[E\right]\right| \geq c\right] \leq \frac{\mathrm{Var}[X_E]}{nc^2} \leq \frac{\epsilon}{nc^2}. \tag{3}$$

Problem 10.3.5 Solution

Given $N_0, N_1, \ldots$, there are jN_j chips used in cookies with j chips. The total number of chips used is $\sum_{k=0}^{\infty} kN_k$. You are equally likely to be any one of the chips in a cookie, so the probability you landed in a cookie with j chips is

$$\mathrm{P}\left[J = j\right] = \frac{jN_j}{\sum_{k=0}^{\infty} kN_k} = \frac{j\frac{N_j}{n}}{\sum_{k=0}^{\infty} k\frac{N_k}{n}}. \tag{1}$$

First, we note that $\mathrm{P}[J = 0] = 0$ since a chip cannot land in a cookie that has zero chips. Second, we note that N_j/n is the relative frequency of cookies with j chips out of all cookies. By comparison, $P_K(j)$ is the probability a cookie has j chips. As $n \to \infty$, the law of large numbers implies $N_j/n \to P_K(j)$. It follows for $j \geq 1$ that as $n \to \infty$,

$$P_J(j) \to \frac{jP_K(j)}{\sum_{k=0}^{\infty} kP_K(k)} = \frac{j(10)^j e^{-10}/j!}{\mathrm{E}\left[K\right]} = \frac{(10)^{j-1} e^{-10}}{(j-1)!}. \tag{2}$$

Problem 10.3.6 Solution

(a) From Theorem 9.2, we have

$$\mathrm{Var}[X_1 + \cdots + X_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{Cov}\left[X_i, X_j\right] \tag{1}$$

Note that $\mathrm{Var}[X_i] = \sigma^2$ and for $j > i$, $\mathrm{Cov}[X_i, X_j] = \sigma^2 a^{j-i}$. This implies

$$\begin{aligned}
\mathrm{Var}[X_1 + \cdots + X_n] &= n\sigma^2 + 2\sigma^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a^{j-i} \\
&= n\sigma^2 + 2\sigma^2 \sum_{i=1}^{n-1} \left(a + a^2 + \cdots + a^{n-i}\right) \\
&= n\sigma^2 + \frac{2a\sigma^2}{1-a} \sum_{i=1}^{n-1} (1 - a^{n-i}). \qquad (2)
\end{aligned}$$

With some more algebra, we obtain

$$\begin{aligned}
&\mathrm{Var}[X_1 + \cdots + X_n] \\
&\quad = n\sigma^2 + \frac{2a\sigma^2}{1-a}(n-1) - \frac{2a\sigma^2}{1-a}\left(a + a^2 + \cdots + a^{n-1}\right) \\
&\quad = \left(\frac{n(1+a)\sigma^2}{1-a}\right) - \frac{2a\sigma^2}{1-a} - 2\sigma^2 \left(\frac{a}{1-a}\right)^2 (1 - a^{n-1}). \qquad (3)
\end{aligned}$$

Since $a/(1-a)$ and $1 - a^{n-1}$ are both nonnegative,

$$\mathrm{Var}[X_1 + \cdots + X_n] \le n\sigma^2 \left(\frac{1+a}{1-a}\right). \tag{4}$$

(b) Since the expected value of a sum equals the sum of the expected values,

$$\mathrm{E}\left[M(X_1, \ldots, X_n)\right] = \frac{\mathrm{E}\left[X_1\right] + \cdots + \mathrm{E}\left[X_n\right]}{n} = \mu. \tag{5}$$

The variance of $M(X_1, \dots, X_n)$ is

$$\begin{aligned}
\operatorname{Var}[M(X_1, \dots, X_n)] &= \frac{\operatorname{Var}[X_1 + \cdots + X_n]}{n^2} \\
&\leq \frac{\sigma^2(1+a)}{n(1-a)}. \qquad (6)
\end{aligned}$$

Applying the Chebyshev inequality to $M(X_1, \dots, X_n)$ yields

$$\begin{aligned}
P\left[|M(X_1, \dots, X_n) - \mu| \geq c\right] &\leq \frac{\operatorname{Var}[M(X_1, \dots, X_n)]}{c^2} \\
&\leq \frac{\sigma^2(1+a)}{n(1-a)c^2}. \qquad (7)
\end{aligned}$$

(c) Taking the limit as n approaches infinity of the bound derived in part (b) yields

$$\lim_{n\to\infty} P\left[|M(X_1, \dots, X_n) - \mu| \geq c\right] \leq \lim_{n\to\infty} \frac{\sigma^2(1+a)}{n(1-a)c^2} = 0. \qquad (8)$$

Thus

$$\lim_{n\to\infty} P\left[|M(X_1, \dots, X_n) - \mu| \geq c\right] = 0. \qquad (9)$$

Problem 10.3.7 Solution

(a) We start by observing that $E[R_1] = E[X_1] = q$. Next, we write

$$R_n = \frac{(n-1)R_{n-1}}{n} + \frac{X_n}{n}. \qquad (1)$$

If follows that

$$\begin{aligned}
E\left[R_n | R_{n-1} = r\right] &= \frac{(n-1)\, E\left[R_{n-1} | R_{n-1} = r\right]}{n} + \frac{E\left[X_n | R_{n-1} = r\right]}{n} \\
&= \frac{(n-1)r}{n} + \frac{r}{n} = r. \qquad (2)
\end{aligned}$$

Thus $E[R_n | R_{n-1}] = R_{n-1}$ and by iterated expectation, $E[R_n] = E[R_{n-1}]$. By induction, it follows that $E[R_n] = E[R_1] = q$.

(b) From the start, $R_1 = X_1$ is Gaussian. Given $R_1 = r$, $R_2 = R_1/2 + X_2/2$ where X_2 is conditionally Gaussian given R_1. Since R_1 is Gaussian, it follows that R_1 and X_2 are jointly Gaussian. It follows that R_2 is also Gaussian since it is a linear combination of jointly Gaussian random variables. Similarly, X_n is conditionally Gaussian given R_{n-1} and thus X_n and R_{n-1} are jointly Gaussian. Thus R_n which is a linear combination of X_n and R_{n-1} is Gaussian. Since $\mathrm{E}[R_n] = q$, we can define $\sigma_n^2 = \mathrm{Var}[R_n]$ and write the PDF of R_n as

$$f_{R_n}(r) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-(r-q)^2/2\sigma_n^2}. \tag{3}$$

The parameter σ_n^2 still needs to be determined.

(c) Following the hint, given $R_{n-1} = r$, $R_n = (n-1)r/n + X_n/n$. It follows that

$$\begin{aligned}
\mathrm{E}\left[R_n^2 | R_{n-1} = r\right] &= \mathrm{E}\left[\left(\frac{(n-1)r}{n} + \frac{X_n}{n}\right)^2 | R_{n-1} = r\right] \\
&= \mathrm{E}\left[\frac{(n-1)^2 r^2}{n^2} + 2r\frac{(n-1)X_n}{n^2} + \frac{X_n^2}{n^2} | R_{n-1} = r\right] \\
&= \frac{(n-1)^2 r^2}{n^2} + \frac{2(n-1)r^2}{n^2} + \frac{\mathrm{E}\left[X_n^2 | R_{n-1} = r\right]}{n^2}.
\end{aligned} \tag{4}$$

Given $R_{n-1} = r$, X_n is Gaussian $(r, 1)$. Since $\mathrm{Var}[X_n | R_{n-1} = r] = 1$,

$$\begin{aligned}
\mathrm{E}\left[X_n^2 | R_{n-1} = r\right] &= \mathrm{Var}[X_n | R_{n-1} = r] + \left(\mathrm{E}\left[X_n | R_{n-1} = r\right]\right)^2 \\
&= 1 + r^2.
\end{aligned} \tag{5}$$

This implies

$$\begin{aligned}
\mathrm{E}\left[R_n^2 | R_{n-1} = r\right] &= \frac{(n-1)^2 r^2}{n^2} + \frac{2(n-1)r^2}{n^2} + \frac{1 + r^2}{n^2} \\
&= r^2 + \frac{1}{n^2},
\end{aligned} \tag{6}$$

and thus

$$\mathrm{E}\left[R_n^2 | R_{n-1}\right] = R_{n-1}^2 + \frac{1}{n^2}. \tag{7}$$

By the iterated expectation,

$$\mathrm{E}\left[R_n^2\right] = \mathrm{E}\left[R_{n-1}^2\right] + \frac{1}{n^2}. \tag{8}$$

Since $\mathrm{E}[R_1^2] = \mathrm{E}[X_1^2] = 1 + q^2$, it follows that

$$\mathrm{E}\left[R_n^2\right] = q^2 + \sum_{j=1}^{n} \frac{1}{j^2}. \tag{9}$$

Hence

$$\mathrm{Var}[R_n] = \mathrm{E}\left[R_n^2\right] - (\mathrm{E}\left[R_n\right])^2 = \sum_{j=1}^{n} \frac{1}{j^2}. \tag{10}$$

Note that $\mathrm{Var}[R_n]$ is an increasing sequence and that $\lim_{n\to\infty} \mathrm{Var}[R_n] \approx 1.645$.

(d) When the prior ratings have no influence, the review scores X_n are iid and by the law of large numbers, R_n will converge to q, the "true quality" of the movie. In our system, the possibility of early misjudgments will lead to randomness in the final rating. The eventual rating $R_n = \lim_{n\to\infty} R_n$ is a random variable with $\mathrm{E}[R] = q$ and $\mathrm{Var}[R] \approx 1.645$. This may or may not be representative of how bad movies can occasionally get high ratings.

Problem 10.4.1 Solution

For an an arbitrary Gaussian (μ, σ) random variable Y,

$$\begin{aligned} \mathrm{P}\left[\mu - \sigma \leq Y \leq \mu + \sigma\right] &= \mathrm{P}\left[-\sigma \leq Y - \mu \leq \sigma\right] \\ &= \mathrm{P}\left[-1 \leq \frac{Y-\mu}{\sigma} \leq 1\right] \\ &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6827. \end{aligned} \tag{1}$$

Note that Y can be any Gaussian random variable, including, for example, $M_n(X)$ when X is Gaussian. When X is not Gaussian, the same claim holds to the extent that the central limit theorem promises that $M_n(X)$ is nearly Gaussian for large n.

Problem 10.4.2 Solution

It should seem obvious that the result is true since $\text{Var}[\hat{R}_n]$ going to zero implies the probablity that $\hat{R}_n$ differs from $\text{E}[\hat{R}_n]$ is going to zero. Similarly, the difference between $\text{E}[\hat{R}_n]$ and r is also going to zero deterministically. Hence it ought to follow that $\hat{R}_n$ is converging to r in probability. Here are the details:

We must show that $\lim_{N\to\infty} \text{P}[|\hat{R}_n - r| \geq \epsilon] = 0$. First we note that $\hat{R}_n$ being asymptotically unbiased implies that $\lim_{n\to\infty} \text{E}[\hat{R}_n] = r$. Equivalently, given $\epsilon > 0$, there exists n_0 such that $|\text{E}[\hat{R}_n] - r| \leq \epsilon^2/2$ for all $n \geq n_0$.

Second, we observe that

$$\begin{aligned} \left|\hat{R}_n - r\right|^2 &= \left|(\hat{R}_n - \text{E}\left[\hat{R}_n\right]) + (\text{E}\left[\hat{R}_n\right] - r)\right|^2 \\ &\leq \left|\hat{R}_n - \text{E}\left[\hat{R}_n\right]\right|^2 + \left|\text{E}\left[\hat{R}_n\right] - r\right|^2. \end{aligned} \tag{1}$$

Thus for all $n \geq n_0$,

$$\left|\hat{R}_n - r\right|^2 \leq\leq \left|\hat{R}_n - \text{E}\left[\hat{R}_n\right]\right|^2 + \epsilon^2/2. \tag{2}$$

It follows for $n \geq n_0$ that

$$\begin{aligned} \text{P}\left[\left|\hat{R}_n - r\right|^2 \geq \epsilon^2\right] &\leq \text{P}\left[\left|\hat{R}_n - \text{E}\left[\hat{R}_n\right]\right|^2 + \epsilon^2/2 \geq \epsilon^2\right] \\ &= \text{P}\left[\left|\hat{R}_n - \text{E}\left[\hat{R}_n\right]\right|^2 \geq \epsilon^2/2\right]. \end{aligned} \tag{3}$$

By the Chebyshev inequality, we have that

$$\text{P}\left[\left|\hat{R}_n - \text{E}\left[\hat{R}_n\right]\right|^2 \geq \epsilon^2/2\right] \leq \frac{\text{Var}[\hat{R}_n]}{(\epsilon/\sqrt{2})^2}. \tag{4}$$

Combining these facts, we see for $n \geq n_0$ that

$$\text{P}\left[\left|\hat{R}_n - r\right|^2 \geq \epsilon^2\right] \leq \frac{\text{Var}[\hat{R}_n]}{(\epsilon/\sqrt{2})^2}. \tag{5}$$

It follows that

$$\lim_{n\to\infty} \text{P}\left[\left|\hat{R}_n - r\right|^2 \geq \epsilon^2\right] \leq \lim_{n\to\infty} \frac{\text{Var}[\hat{R}_n]}{(\epsilon/\sqrt{2})^2} = 0. \tag{6}$$

This proves that $\hat{R}_n$ is a consistent estimator.

Problem 10.4.3 Solution

This problem is really very simple. If we let $Y = X_1 X_2$ and for the ith trial, let $Y_i = X_1(i)X_2(i)$, then $\hat{R}_n = M_n(Y)$, the sample mean of random variable Y. By Theorem 10.9, $M_n(Y)$ is unbiased. Since $\text{Var}[Y] = \text{Var}[X_1X_2] < \infty$, Theorem 10.11 tells us that $M_n(Y)$ is a consistent sequence.

Problem 10.4.4 Solution

(a) Since the expectation of a sum equals the sum of the expectations also holds for vectors,

$$\mathrm{E}\left[\mathbf{M}(n)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\mathbf{X}(i)\right] = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mu}_{\mathbf{X}} = \boldsymbol{\mu}_{\mathbf{X}}. \tag{1}$$

(b) The jth component of $\mathbf{M}(n)$ is $M_j(n) = \frac{1}{n}\sum_{i=1}^{n} X_j(i)$, which is just the sample mean of X_j. Defining $A_j = \{|M_j(n) - \mu_j| \geq c\}$, we observe that

$$\mathrm{P}\left[\max_{j=1,\ldots,k} |M_j(n) - \mu_j| \geq c\right] = \mathrm{P}\left[A_1 \cup A_2 \cup \cdots \cup A_k\right]. \tag{2}$$

Applying the Chebyshev inequality to $M_j(n)$, we find that

$$\mathrm{P}\left[A_j\right] \leq \frac{\text{Var}[M_j(n)]}{c^2} = \frac{\sigma_j^2}{nc^2}. \tag{3}$$

By the union bound,

$$\mathrm{P}\left[\max_{j=1,\ldots,k} |M_j(n) - \mu_j| \geq c\right] \leq \sum_{j=1}^{k}\mathrm{P}\left[A_j\right] \leq \frac{1}{nc^2}\sum_{j=1}^{k}\sigma_j^2. \tag{4}$$

Since $\sum_{j=1}^{k}\sigma_j^2 < \infty$, $\lim_{n\to\infty}\mathrm{P}[\max_{j=1,\ldots,k}|M_j(n) - \mu_j| \geq c] = 0$.

Problem 10.4.5 Solution

Note that we can write Y_k as

$$Y_k = \left(\frac{X_{2k-1} - X_{2k}}{2}\right)^2 + \left(\frac{X_{2k} - X_{2k-1}}{2}\right)^2 \\ = \frac{(X_{2k} - X_{2k-1})^2}{2}. \tag{1}$$

Hence,

$$\mathrm{E}\left[Y_k\right] = \frac{1}{2}\,\mathrm{E}\left[X_{2k}^2 - 2X_{2k}X_{2k-1} + X_{2k-1}^2\right] \\ = \mathrm{E}\left[X^2\right] - (\mathrm{E}\left[X\right])^2 = \mathrm{Var}[X]. \tag{2}$$

Next we observe that $Y_1, Y_2, \ldots$ is an iid random sequence. If this independence is not obvious, consider that Y_1 is a function of X_1 and X_2, Y_2 is a function of X_3 and X_4, and so on. Since $X_1, X_2, \ldots$ is an idd sequence, $Y_1, Y_2, \ldots$ is an iid sequence. Hence, $\mathrm{E}[M_n(Y)] = \mathrm{E}[Y] = \mathrm{Var}[X]$, implying $M_n(Y)$ is an unbiased estimator of $\mathrm{Var}[X]$. We can use Theorem 10.9 to prove that $M_n(Y)$ is consistent if we show that $\mathrm{Var}[Y]$ is finite.Since $\mathrm{Var}[Y] \leq \mathrm{E}[Y^2]$, it is sufficient to prove that $\mathrm{E}[Y^2] < \infty$. Note that

$$Y_k^2 = \frac{X_{2k}^4 - 4X_{2k}^3 X_{2k-1} + 6X_{2k}^2 X_{2k-1}^2 - 4X_{2k}X_{2k-1}^3 + X_{2k-1}^4}{4}. \tag{3}$$

Taking expectations yields

$$\mathrm{E}\left[Y_k^2\right] = \frac{1}{2}\,\mathrm{E}\left[X^4\right] - 2\,\mathrm{E}\left[X^3\right]\mathrm{E}\left[X\right] + \frac{3}{2}\left(\mathrm{E}\left[X^2\right]\right)^2. \tag{4}$$

Hence, if the first four moments of X are finite, then $\mathrm{Var}[Y] \leq \mathrm{E}[Y^2] < \infty$. By Theorem 10.9, the sequence $M_n(Y)$ is consistent.

Problem 10.4.6 Solution

(a) Since the expectation of the sum equals the sum of the expectations,

$$\mathrm{E}\left[\hat{\mathbf{R}}(n)\right] = \frac{1}{n}\sum_{m=1}^{n}\mathrm{E}\left[\mathbf{X}(m)\mathbf{X}'(m)\right] = \frac{1}{n}\sum_{m=1}^{n}\mathbf{R} = \mathbf{R}. \tag{1}$$

(b) This proof follows the method used to solve Problem 10.4.4. The i,jth element of $\hat{\mathbf{R}}(n)$ is $\hat{R}_{i,j}(n) = \frac{1}{n}\sum_{m=1}^{n} X_i(m)X_j(m)$, which is just the sample mean of X_iX_j. Defining the event

$$A_{i,j} = \left\{ \left| \hat{R}_{i,j}(n) - \mathrm{E}\left[X_iX_j\right] \right| \geq c \right\}, \tag{2}$$

we observe that

$$\mathrm{P}\left[\max_{i,j} \left| \hat{R}_{i,j}(n) - \mathrm{E}\left[X_iX_j\right] \right| \geq c \right] = \mathrm{P}\left[\cup_{i,j} A_{i,j}\right]. \tag{3}$$

Applying the Chebyshev inequality to $\hat{R}_{i,j}(n)$, we find that

$$\mathrm{P}\left[A_{i,j}\right] \leq \frac{\mathrm{Var}[\hat{R}_{i,j}(n)]}{c^2} = \frac{\mathrm{Var}[X_iX_j]}{nc^2}. \tag{4}$$

By the union bound,

$$\begin{aligned} \mathrm{P}\left[\max_{i,j} \left| \hat{R}_{i,j}(n) - \mathrm{E}\left[X_iX_j\right] \right| \geq c \right] &\leq \sum_{i,j} \mathrm{P}\left[A_{i,j}\right] \\ &\leq \frac{1}{nc^2} \sum_{i,j} \mathrm{Var}[X_iX_j]. \end{aligned} \tag{5}$$

By the result of Problem 7.6.4, X_iX_j, the product of jointly Gaussian random variables, has finite variance. Thus

$$\begin{aligned} \sum_{i,j} \mathrm{Var}[X_iX_j] &= \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{Var}[X_iX_j] \\ &\leq k^2 \max_{i,j} \mathrm{Var}[X_iX_j] < \infty. \end{aligned} \tag{6}$$

It follows that

$$\begin{aligned} \lim_{n\to\infty} \mathrm{P}\left[\max_{i,j} \left| \hat{R}_{i,j}(n) - \mathrm{E}\left[X_iX - j\right] \right| \geq c \right] &\leq \lim_{n\to\infty} \frac{k^2 \max_{i,j} \mathrm{Var}[X_iX_j]}{nc^2} \\ &= 0. \end{aligned} \tag{7}$$

Problem 10.5.1 Solution

X has the Bernoulli (0.9) PMF

$$P_X(x) = \begin{cases} 0.1 & x = 0, \\ 0.9 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) $\mathrm{E}[X]$ is in fact the same as $P_X(1)$ because X is Bernoulli.

(b) We can use the Chebyshev inequality to find

$$\begin{aligned} \mathrm{P}\left[|M_{90}(X) - P_X(1)| \geq .05\right] &= \mathrm{P}\left[|M_{90}(X) - \mathrm{E}\left[X\right]| \geq .05\right] \\ &\leq \alpha. \end{aligned} \tag{2}$$

In particular, the Chebyshev inequality states that

$$\alpha = \frac{\sigma_X^2}{90(.05)^2} = \frac{.09}{90(.05)^2} = 0.4. \tag{3}$$

(c) Now we wish to find the value of n such that

$$\mathrm{P}\left[|M_n(X) - P_X(1)| \geq .03\right] \leq 0.1. \tag{4}$$

From the Chebyshev inequality, we write

$$0.1 = \frac{\sigma_X^2}{n(.03)^2}. \tag{5}$$

Since $\sigma_X^2 = 0.09$, solving for n yields $n = 100$.

Problem 10.5.2 Solution

Since $\mathrm{E}[X] = \mu_X = p$ and $\mathrm{Var}[X] = p(1-p)$, we use Theorem 10.5(b) to write

$$\mathrm{P}\left[|M_{100}(X) - p| < c\right] \geq 1 - \frac{p(1-p)}{100c^2} = 1 - \alpha. \tag{1}$$

For confidence coefficient 0.99, we require

$$\frac{p(1-p)}{100c^2} \leq 0.01 \quad \text{or} \quad c \geq \sqrt{p(1-p)}. \tag{2}$$

Since p is unknown, we must ensure that the constraint is met for every value of p. The worst case occurs at $p = 1/2$ which maximizes $p(1-p)$. In this case, $c = \sqrt{1/4} = 1/2$ is the smallest value of c for which we have confidence coefficient of at least 0.99.

If $M_{100}(X) = 0.06$, our interval estimate for p is

$$M_{100}(X) - c < p < M_{100}(X) + c. \tag{3}$$

Since $p \geq 0$, $M_{100}(X) = 0.06$ and $c = 0.5$ imply that our interval estimate is

$$0 \leq p < 0.56. \tag{4}$$

Our interval estimate is not very tight because because 100 samples is not very large for a confidence coefficient of 0.99.

Problem 10.5.3 Solution

First we observe that the interval estimate can be expressed as

$$\left|\hat{P}_n(A) - \mathrm{P}[A]\right| < 0.05. \tag{1}$$

Since $\hat{P}_n(A) = M_n(X_A)$ and $\mathrm{E}[M_n(X_A)] = \mathrm{P}[A]$, we can use Theorem 10.5(b) to write

$$\mathrm{P}\left[\left|\hat{P}_n(A) - \mathrm{P}[A]\right| < 0.05\right] \geq 1 - \frac{\mathrm{Var}[X_A]}{n(0.05)^2}. \tag{2}$$

Note that $\mathrm{Var}[X_A] = \mathrm{P}[A](1 - \mathrm{P}[A]) \leq 0.25$. Thus for confidence coefficient 0.9, we require that

$$1 - \frac{\mathrm{Var}[X_A]}{n(0.05)^2} \geq 1 - \frac{0.25}{n(0.05)^2} \geq 0.9. \tag{3}$$

This implies $n \geq 1{,}000$ samples are needed.

Problem 10.5.4 Solution

Both questions can be answered using the following equation from Example 10.7:

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] \leq \frac{P[A](1 - P[A])}{nc^2}. \tag{1}$$

The unusual part of this problem is that we are given the true value of $P[A]$. Since $P[A] = 0.01$, we can write

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] \leq \frac{0.0099}{nc^2}. \tag{2}$$

(a) In this part, we meet the requirement by choosing $c = 0.001$ yielding

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq 0.001\right] \leq \frac{9900}{n}. \tag{3}$$

Thus to have confidence level 0.01, we require that $9900/n \leq 0.01$. This requires $n \geq 990{,}000$.

(b) In this case, we meet the requirement by choosing $c = 10^{-3}\,P[A] = 10^{-5}$. This implies

$$\begin{aligned} P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] &\leq \frac{P[A](1 - P[A])}{nc^2} \\ &= \frac{0.0099}{n10^{-10}} = \frac{9.9 \times 10^7}{n}. \end{aligned} \tag{4}$$

The confidence level 0.01 is met if $9.9 \times 10^7/n = 0.01$ or $n = 9.9 \times 10^9$.

Problem 10.6.1 Solution

In this problem, we have to keep straight that the Poisson expected value $\alpha = 1$ is a different α than the confidence coefficient $1 - \alpha$. That said, we will try avoid using α for the confidence coefficient. Using X to denote the Poisson ($\alpha = 1$) random variable, the trace of the sample mean is the sequence $M_1(X), M_2(X), \ldots$ The confidence interval estimate of α has the form

$$M_n(X) - c \leq \alpha \leq M_n(X) + c. \tag{1}$$

The confidence coefficient of the estimate based on n samples is

$$\begin{aligned} \mathrm{P}\left[M_n(X) - c \le \alpha \le M_n(X) + c\right] &= \mathrm{P}\left[\alpha - c \le M_n(X) \le \alpha + c\right] \\ &= \mathrm{P}\left[-c \le M_n(X) - \alpha \le c\right]. \end{aligned} \tag{2}$$

Since $\mathrm{Var}[M_n(X)] = \mathrm{Var}[X]/n = 1/n$, the 0.9 confidence interval shrinks with increasing n. In particular, $c = c_n$ will be a decreasing sequence. Using a Central Limit Theorem approximation, a 0.9 confidence implies

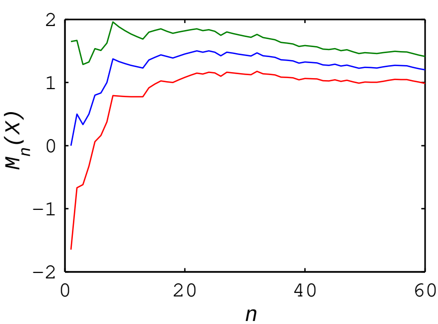
$$\begin{aligned} 0.9 &= \mathrm{P}\left[\frac{-c_n}{\sqrt{1/n}} \le \frac{M_n(X) - \alpha}{\sqrt{1/n}} \le \frac{c_n}{\sqrt{1/n}}\right] \\ &= \Phi(c_n\sqrt{n}) - \Phi(-c_n\sqrt{n}) = 2\Phi(c_n\sqrt{n}) - 1. \end{aligned} \tag{3}$$

Equivalently, $\Phi(c_n\sqrt{n}) = 0.95$ or $c_n = 1.65/\sqrt{n}$.

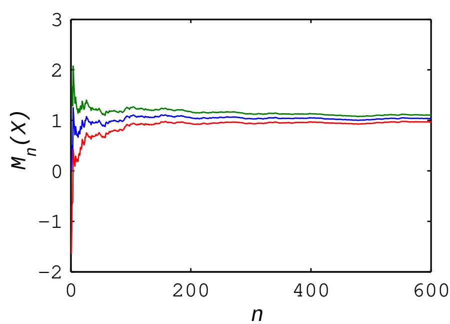
Thus, as a function of the number of samples n, we plot three functions: the sample mean $M_n(X)$, and the upper limit $M_n(X) + 1.65/\sqrt{n}$ and lower limit $M_n(X) - 1.65/\sqrt{n}$ of the 0.9 confidence interval. We use the MATLAB function `poissonmeanseq(n)` to generate these sequences for n sample values.

```
function M=poissonmeanseq(n);
x=poissonrv(1,n);
nn=(1:n)';
M=cumsum(x)./nn;
r=(1.65)./sqrt(nn);
plot(nn,M,nn,M+r,nn,M-r);
```

Here are two output graphs:



poissonmeanseq(60)



poissonmeanseq(600)

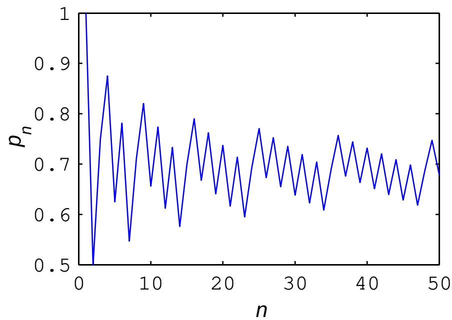
Problem 10.6.2 Solution

For a Bernoulli ($p = 1/2$) random variable X, the sample mean $M_n(X)$ is the fraction of successes in n Bernoulli trials. That is, $M_n(X) = K_n/n$ where K_n is a binomial ($n, p = 1/2$) random variable. Thus the probability the sample mean is within one standard error of ($p = 1/2$) is

$$\begin{aligned} p_n &= \mathrm{P}\left[\frac{n}{2} - \frac{\sqrt{n}}{2} \le K_n \le \frac{n}{2} + \frac{\sqrt{n}}{2}\right] \\ &= \mathrm{P}\left[K_n \le \frac{n}{2} + \frac{\sqrt{n}}{2}\right] - \mathrm{P}\left[K_n < \frac{n}{2} - \frac{\sqrt{n}}{2}\right] \\ &= F_{K_n}\left(\frac{n}{2} + \frac{\sqrt{n}}{2}\right) - F_{K_n}\left(\left\lceil \frac{n}{2} - \frac{\sqrt{n}}{2}\right\rceil - 1\right). \end{aligned} \tag{1}$$

Here is a MATLAB function that graphs p_n as a function of n for N steps alongside the output graph for `bernoullistderr(50)`.

```
function p=bernoullistderr(N);
p=zeros(1,N);
for n=1:N,
  r=[ceil((n-sqrt(n))/2)-1; ...
          (n+sqrt(n))/2];
  p(n)=diff(binomialcdf(n,0.5,r));
end
plot(1:N,p);
ylabel('\it p_n');
xlabel('\it n');
```



The alternating up-down sawtooth pattern is a consequence of the CDF of K_n jumping up at integer values. In particular, depending on the value of $\sqrt{n}/2$, every other value of $(n \pm \sqrt{n})/2$ exceeds the next integer and increases p_n. The less frequent but larger jumps in p_n occur when $\sqrt{n}$ is an integer. For very large n, the central lmit theorem takes over since the CDF of $M_n(X)$ converges to a Gaussian CDF. In this case, the sawtooth pattern dies out and p_n will converge to 0.68, the probability that a Gaussian random variable is within one standard deviation of its expected value.

Finally, we note that the sawtooth pattern is essentially the same as the sawtooth pattern observed in Quiz 10.6 where we graphed the fraction of 1,000 Bernoulli sample mean traces that were within one standard error of the true expected value.

For 1,000 traces in Quiz 10.6 was so large, the fraction of the number of traces within one standard error was always very close to the actual probability p_n.

Problem 10.6.3 Solution

First, we need to determine whether the relative performance of the two estimators depends on the actual value of λ. To address this, we observe that if Y is an exponential (1) random variable, then Theorem 6.3 tells us that $X = Y/\lambda$ is an exponential (λ) random variable. Thus if $Y_1, Y_2, \ldots$ are iid samples of Y, then $Y_1/\lambda, Y_2/\lambda, \ldots$ are iid samples of X. Moreover, the sample mean of X is

$$M_n(X) = \frac{1}{n\lambda}\sum_{i=1}^{n} Y_i = \frac{1}{\lambda} M_n(Y). \tag{1}$$

Similarly, the sample variance of X satisfies

$$\begin{aligned} V_n'(X) &= \frac{1}{n-1}\sum_{i=1}^{n} (X_i - M_n(X))^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n} \left(\frac{Y_i}{\lambda} - \frac{1}{\lambda} M_n(Y)\right)^2 = \frac{V_n'(Y)}{\lambda^2}. \end{aligned} \tag{2}$$

We can conclude that

$$\hat{\lambda} = \frac{\lambda}{M_n(Y)}, \qquad \tilde{\lambda} = \frac{\lambda}{\sqrt{V_n'(Y)}}. \tag{3}$$

For $\lambda \neq 1$, the estimators $\hat{\lambda}$ and $\tilde{\lambda}$ are just scaled versions of the estimators for the case $\lambda = 1$. Hence it is sufficient to consider only the $\lambda = 1$ case. The function `z=lamest(n,m)` returns the estimation errors for m trials of each estimator where each trial uses n iid exponential (1) samples.

```
function z=lamest(n,m);
x=exponentialrv(1,n*m);
x=reshape(x,n,m);
mx=sum(x)/n;
MX=ones(n,1)*mx;
vx=sum((x-MX).^2)/(n-1);
z=[(1./mx); (1./sqrt(vx))]-1;
```

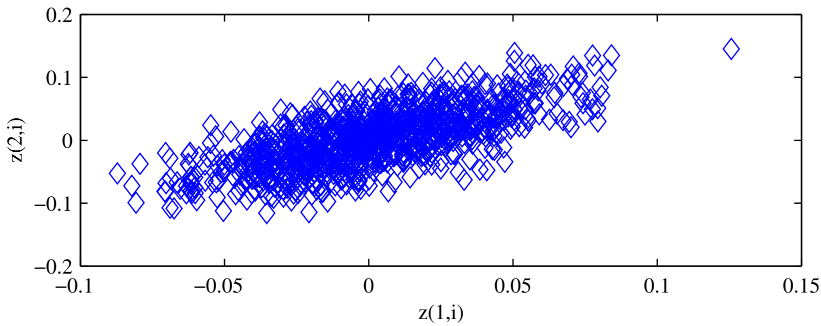
In `lamest.m`, each column of matrix `x` represents one trial. Note that `mx` is a row vector such that `mx(i)` is the sample mean for trial i. The matrix `MX` has `mx(i)` for every element in column i. Thus `vx` is a row vector such that `vx(i)` is the sample variance for trial i.

Finally, `z` is a $2 \times m$ matrix such that column i of `z` records the estimation errors for trial i. If $\hat{\lambda}_i$ and $\tilde{\lambda}_i$ are the estimates for for trial i, then `z(1,i)` is the error $\hat{Z}_i = \hat{\lambda}_i - 1$ while `z(2,i)` is the error $\tilde{Z}_i = \tilde{\lambda}_i - 1$.

Now that we can simulate the errors generated by each estimator, we need to determine which estimator is better. We start by using the commands

```
z=lamest(1000,1000);
plot(z(1,:),z(2,:),'bd')
```

to perform 1,000 trials, each using 1,000 samples. The `plot` command generates a scatter plot of the error pairs $(\hat{Z}_i, \tilde{Z}_i)$ for each trial. Here is an example of the resulting scatter plot:



In the scatter plot, each diamond marks an independent pair $(\hat{Z}, \tilde{Z})$ where $\hat{Z}$ is plotted on the x-axis and $\tilde{Z}$ is plotted on the y-axis. (Although it is outside the scope of this solution, it is interesting to note that the errors $\hat{Z}$ and $\tilde{Z}$ appear to be positively correlated.) From the plot, it may not be obvious that one estimator is better than the other. However, by reading the axis ticks carefully, one can observe that it appears that typical values for $\hat{Z}$ are in the interval $(-0.05, 0.05)$ while typical values for $\tilde{Z}$ are in the interval $(-0.1, 0.1)$. This suggests that $\hat{Z}$ may be superior.To verify this observation, we calculate the sample mean for each squared errors

$$M_m(\hat{Z}^2) = \frac{1}{m}\sum_{i=1}^{m} \hat{Z}_i^2, \qquad M_m(\tilde{Z}^2) = \frac{1}{m}\sum_{i=1}^{m} \tilde{Z}_i^2. \tag{4}$$

From our MATLAB experiment with $m = 1{,}000$ trials, we calculate

```
>> sum(z.^2,2)/1000
ans =
    0.0010
    0.0021
```

That is, $M_{1,000}(\hat{Z}^2) = 0.0010$ and $M_{1,000}(\tilde{Z}^2) = 0.0021$. In fact, one can show (with a lot of work) for large m that

$$M_m(\hat{Z}^2) \approx 1/m, \qquad M_m(\tilde{Z}^2) = 2/m, \tag{5}$$

and that

$$\lim_{m\to\infty} \frac{M_m(\tilde{Z}^2)}{M_m(\hat{Z}^2)} = 2. \tag{6}$$

In short, the mean squared error of the $\tilde{\lambda}$ estimator is twice that of the $\hat{\lambda}$ estimator.

Problem 10.6.4 Solution

For the sample covariance matrix

$$\hat{\mathbf{R}}(n) = \frac{1}{n}\sum_{m=1}^{n} \mathbf{X}(m)\mathbf{X}'(m), \tag{1}$$

we wish to use a MATLAB simulation to estimate the probability

$$p(n) = \mathrm{P}\left[\max_{i,j}\left|\hat{R}_{ij} - I_{ij}\right| \geq 0.05\right]. \tag{2}$$

(In the original printing, 0.05 was 0.01 but that requirement demanded that n be so large that most installations of MATLAB would grind to a halt on the calculations.)

The MATLAB program uses a matrix algebra identity that may (or may not) be familiar. For a matrix

$$\mathbf{X} = \begin{bmatrix}\mathbf{x}(1) & \mathbf{x}(2) & \cdots & \mathbf{x}(n)\end{bmatrix}, \tag{3}$$

with columns $\mathbf{x}(i)$, we can write

$$\mathbf{X}\mathbf{X}' = \sum_{i=1}^{n} \mathbf{x}(i)\mathbf{x}'(i). \tag{4}$$

The MATLAB function `diagtest(n,t)` performs the calculation by generating t sample covariance matrices, each using n vectors $\mathbf{x}(i)$ in each sample covariance matrix.

```
function p=diagtest(n,t);
y=zeros(t,1); p=[ ];
for ntest=n,
  disp(ntest)
  for i=1:t,
    X=gaussvector(0,eye(10),ntest);
    R=X*(X')/ntest;
    y(i)=(max(max(abs(R-eye(10))))>= 0.05);
  end
p=[p, sum(y)/t];
end
semilogy(n,p);
xlabel('\it n'); ylabel('\it p(n)');
```

These n vectors are generated by the `gaussvector` function. The vector $\mathbf{x}(i)$ is the ith column of the matrix `X`. The function `diagtest` estimates $p(n)$ as the fraction of t trials for which the threshold 0.05 is exceeded. In addition, if the input `n` to `diagtest` is a vector, then the experiment is repeated for each value in the vector `n`.

The following commands, estimate $p(n)$ for $n \in \{10, 100, 1000, 10000\}$ using $t = 2000$ trials:

```
n=[10 100 1000 10000];
p=diagtest(n,2000);
```

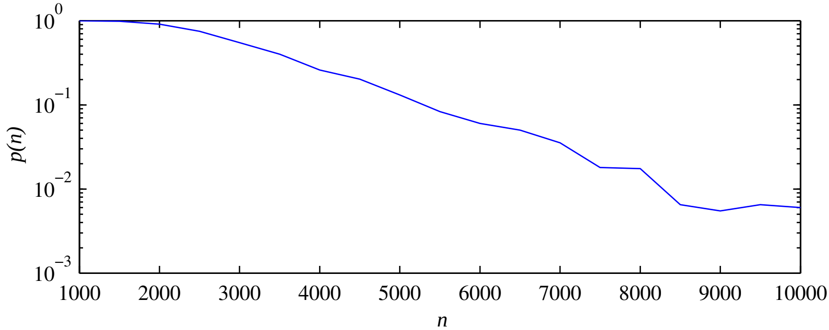
The output is

```
p=
    1.0000    1.0000    1.0000    0.0035
```

We see that $p(n)$ goes from roughly 1 to almost 0 in going from $n = 1{,}000$ to $n = 10{,}000$. To investigate this transition more carefully, we execute the commands

```
nn=1000:500:10000;
p=diagtest(nn,2000);
```

The output is shown in the following graph. We use a semilog plot to emphasize differences when $p(n)$ is close to zero.



Beyond $n = 1{,}000$, the probability $p(n)$ declines rapidly. The "bumpiness" of the graph for large n occurs because the probability $p(n)$ is small enough that out of 2,000 trials, the 0.05 threshold is exceeded only a few times.

Note that if $\mathbf{x}$ has dimension greater than 10, then the value of n needed to ensure that $p(n)$ is small would increase.

Problem 10.6.5 Solution

The difficulty in this problem is that although $\mathrm{E}[X]$ exists, EX^2 and higher order moments are infinite. Thus $\mathrm{Var}[X]$ is also infinite. It also follows for any finite n that the sample mean $M_n(X)$ has infinite variance. In this case, we *cannot* apply the Chebyshev inequality to the sample mean to show the convergence in probability of $M_n(X)$ to $\mathrm{E}[X]$.

If $\lim_{n\to\infty} \mathrm{P}[|M_n(X) - \mathrm{E}[X]| \geq \epsilon] = p$, then there are two distinct possibilities:

- $p > 0$, or
- $p = 0$ but the Chebyshev inequality isn't a sufficient powerful technique to verify this fact.

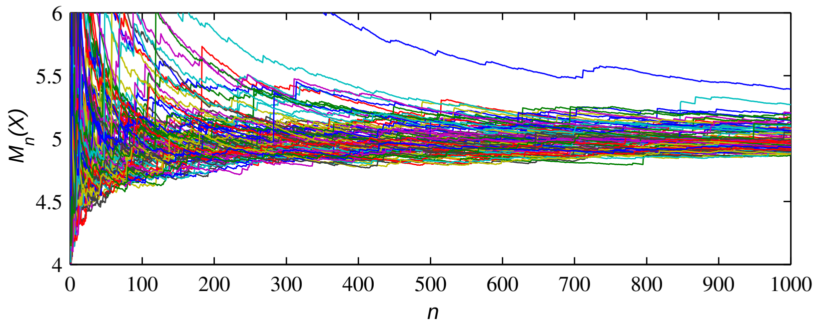
To resolve whether $p = 0$ (and the sample mean converges to the expected value) one can spend time trying to prove either $p = 0$ or $p > 0$. At this point, we try some simulation experiments to see if the experimental evidence points one way or the other.

As requested by the problem, the MATLAB function `samplemeantest(n,a)` simulates one hundred traces of the sample mean when $\mathrm{E}[X] = a$. Each trace is a length n sequence $M_1(X), M_2(X), \ldots, M_n(X)$.

```
function mx=samplemeantest(n,a);
u=rand(n,100);
x=a-2+(1./sqrt(1-u));
d=((1:n)')*ones(1,100);
mx=cumsum(x)./d;
plot(mx);
xlabel('\it n'); ylabel('\it M_n(X)');
axis([0 n a-1 a+1]);
```

The $n \times 100$ matrix `x` consists of iid samples of X. Taking cumulative sums along each column of x, and dividing row i by i, each column of `mx` is a length n sample mean trace. we then plot the traces.

The following graph was generated by `samplemeantest(1000,5)`:



Frankly, it is difficult to draw strong conclusions from the graph. If the sample sequences $M_n(X)$ are converging to $\mathrm{E}[X]$, the convergence is fairly slow. Even after averaging 1,000 samples, typical values for the sample mean appear to range from $a-0.5$ to $a+0.5$. There may also be outlier sequences which are still off the charts since we truncated the y-axis range. On the other hand, the sample mean sequences do not appear to be diverging (which is also possible since $\mathrm{Var}[X] = \infty$.) Note the above graph was generated using 10^5 sample values. Repeating the experiment with more samples, say `samplemeantest(10000,5)`, will yield a similarly inconclusive result. Even if your version of MATLAB can support the generation of 100 times as many samples, you won't know for sure whether the sample mean sequence *always* converges. On the other hand, the experiment is probably enough that if you pursue the analysis, you should start by trying to prove that $p = 0$.

Problem Solutions – Chapter 11

Problem 11.1.1 Solution

Assuming the coin is fair, we must choose a rejection region R such that $\alpha = P[R] = 0.05$. We can choose a rejection region $R = \{L > r\}$. What remains is to choose r so that $P[R] = 0.05$. Note that $L > l$ if we first observe l tails in a row. Under the hypothesis that the coin is fair, l tails in a row occurs with probability

$$P[L > l] = (1/2)^l. \tag{1}$$

Thus, we need

$$P[R] = P[L > r] = 2^{-r} = 0.05. \tag{2}$$

Thus, $r = -\log_2(0.05) = \log_2(20) = 4.32$. In this case, we reject the hypothesis that the coin is fair if $L \geq 5$. The significance level of the test is $\alpha = P[L > 4] = 2^{-4} = 0.0625$ which close to but not exactly 0.05.

The shortcoming of this test is that we always accept the hypothesis that the coin is fair whenever heads occurs on the first, second, third or fourth flip. If the coin was biased such that the probability of heads was much higher than 1/2, say 0.8 or 0.9, we would hardly ever reject the hypothesis that the coin is fair. In that sense, our test cannot identify that kind of biased coin.

Problem 11.1.2 Solution

There can be a variety of valid answers to this problems. Nevertheless, reasonable explanations for why one TA was no better than the other tend to fall into one of the following three categories:

- The students in section 1 were actually better than students in section 2. For example, because honors student all took a certain class, all the honor students were forced by the schedule to be in section 1.
- The students in the two sections were equal but it was simply a random outcome that the students in section 1 did better.
- There was a structural reason that students in section 1 were able to learn more. For example, maybe section 2 was very early in the morning when students had a hard time staying awake.

Problem 11.1.3 Solution

We reject the null hypothesis when the call rate M is too high. Note that

$$\mathrm{E}\left[M\right] = \mathrm{E}\left[N_i\right] = 2.5, \qquad \mathrm{Var}[M] = \frac{\mathrm{Var}[N_i]}{T} = \frac{2.5}{T}. \tag{1}$$

For large T, we use a central limit theorem approximation to calculate the rejection probability

$$\begin{aligned} \mathrm{P}\left[R\right] &= \mathrm{P}\left[M \geq m_0\right] \\ &= \mathrm{P}\left[\frac{M - 2.5}{\sigma_M} \geq \frac{m_0 - 2.5}{\sigma_M}\right] \\ &= 1 - \Phi\left(\frac{m_0 - 2.5}{\sqrt{\frac{2.5}{T}}}\right) = 0.05. \end{aligned} \tag{2}$$

It follows that

$$\frac{m_0 - 2.5}{\sqrt{2.5/T}} = 1.65 \implies m_0 = 2.5 + \frac{1.65\sqrt{2.5}}{\sqrt{T}} = 2.5 + \frac{2.6}{\sqrt{T}}. \tag{3}$$

That is, we reject the null hypothesis whenever

$$M \geq 2.5 + \frac{2.6}{\sqrt{T}}. \tag{4}$$

As T gets larger, smaller deviations of M above the expected value $\mathrm{E}[M] = 2.5$ are sufficient to reject the null hypothesis.

Problem 11.1.4 Solution

(a) The rejection region is $R = \{T > t_0\}$. The duration of a voice call has exponential PDF

$$f_T(t) = \begin{cases} (1/3)e^{-t/3} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The significance level of the test is

$$\alpha = \mathrm{P}\left[T > t_0\right] = \int_{t_0}^{\infty} f_T(t)\ dt = e^{-t_0/3}. \tag{2}$$

(b) The significance level is $\alpha = 0.05$ if $t_0 = -3 \ln \alpha = 8.99$ minutes.

Problem 11.1.5 Solution

In order to test just a small number of pacemakers, we test n pacemakers and we reject the null hypothesis if *any* pacemaker fails the test. Moreover, we choose the smallest n such that we meet the required significance level of the test.

The number of pacemakers that fail the test is X, a binomial $(n, q_0 = 10^{-4})$ random variable. The significance level of the test is

$$\alpha = \mathrm{P}[X > 0] = 1 - \mathrm{P}[X = 0] = 1 - (1 - q_0)^n. \tag{1}$$

For a significance level $\alpha = 0.01$, we have that

$$n = \frac{\ln(1-\alpha)}{\ln(1-q_0)} = 100.5. \tag{2}$$

Comment: For $\alpha = 0.01$, keep in mind that there is a one percent probability that a normal factory will fail the test. That is, a test failure is quite unlikely if the factory is operating normally.

Problem 11.1.6 Solution

(a) We wish to develop a hypothesis test of the form

$$\mathrm{P}[|K - \mathrm{E}[K]| > c] = 0.05. \tag{1}$$

to determine if the coin we've been flipping is indeed a fair one. We would like to find the value of c, which will determine the upper and lower limits on how many heads we can get away from the expected number out of 100 flips and still accept our hypothesis. Under our fair coin hypothesis, the expected number of heads, and the standard deviation of the process are

$$\mathrm{E}[K] = 50, \qquad \sigma_K = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5. \tag{2}$$

Now in order to find c we make use of the central limit theorem and divide the above inequality through by σ_K to arrive at

$$\mathrm{P}\left[\frac{|K - \mathrm{E}[K]|}{\sigma_K} > \frac{c}{\sigma_K}\right] = 0.05. \tag{3}$$

Taking the complement, we get

$$\mathrm{P}\left[-\frac{c}{\sigma_K} \leq \frac{K - \mathrm{E}[K]}{\sigma_K} \leq \frac{c}{\sigma_K}\right] = 0.95. \tag{4}$$

Using the Central Limit Theorem we can write

$$\Phi\left(\frac{c}{\sigma_K}\right) - \Phi\left(\frac{-c}{\sigma_K}\right) = 2\Phi\left(\frac{c}{\sigma_K}\right) - 1 = 0.95. \tag{5}$$

This implies $\Phi(c/\sigma_K) = 0.975$ or $c/5 = 1.96$. That is, $c = 9.8$ flips. So we see that if we observe more then $50 + 10 = 60$ or less then $50 - 10 = 40$ heads, then with significance level $\alpha \approx 0.05$ we should reject the hypothesis that the coin is fair.

(b) Now we wish to develop a test of the form

$$\mathrm{P}[K > c] = 0.01. \tag{6}$$

Thus we need to find the value of c that makes the above probability true. This value will tell us that if we observe more than c heads, then with significance level $\alpha = 0.01$, we should reject the hypothesis that the coin is fair. To find this value of c we look to evaluate the CDF

$$F_K(k) = \sum_{i=0}^{k} \binom{100}{i} (1/2)^{100}. \tag{7}$$

Computation reveals that $c \approx 62$ flips. So if we observe 62 or greater heads, then with a significance level of 0.01 we should reject the fair coin hypothesis. Another way to obtain this result is to use a Central Limit Theorem approximation. First, we express our rejection region in terms of a zero mean, unit variance random variable.

$$\begin{aligned}\mathrm{P}[K > c] &= 1 - \mathrm{P}[K \leq c] \\ &= 1 - \mathrm{P}\left[\frac{K - \mathrm{E}[K]}{\sigma_K} \leq \frac{c - \mathrm{E}[K]}{\sigma_K}\right] = 0.01. \end{aligned}\tag{8}$$

Since $\mathrm{E}[K] = 50$ and $\sigma_K = 5$, the CLT approximation is

$$\mathrm{P}\left[K > c\right] \approx 1 - \Phi\left(\frac{c-50}{5}\right) = 0.01. \tag{9}$$

From Table 4.2, we have $(c-50)/5 = 2.35$ or $c = 61.75$. Once again, we see that we reject the hypothesis if we observe 62 or more heads.

Problem 11.1.7 Solution

A reasonable test would reject the null hypothesis that the plant is operating normally if one or more of the chips fail the one-day test. Exactly how many shold be tested and how many failures N are needed to reject the null hypothesis would depend on the significance level of the test.

(a) The lifetime of a chip is X, an exponential (λ) random variable with $\lambda = (T/200)^2$. The probability p that a chip passes the one-day test is

$$p = \mathrm{P}\left[X \geq 1/365\right] = e^{-\lambda/365}. \tag{1}$$

For an m chip test, the significance level of the test is

$$\begin{aligned}\alpha = \mathrm{P}\left[N > 0\right] &= 1 - \mathrm{P}\left[N = 0\right] \\ &= 1 - p^m = 1 - e^{-m\lambda/365}.\end{aligned} \tag{2}$$

(b) At $T = 100^\circ$, $\lambda = 1/4$ and we obtain a significance level of $\alpha = 0.01$ if

$$m = -\frac{365\ln(0.99)}{\lambda} = \frac{3.67}{\lambda} = 14.74. \tag{3}$$

In fact, at $m = 15$ chips, the significance level is $\alpha = 0.0102$.

(c) Raising T raises the failure rate $\lambda = (T/200)^2$ and thus lowers $m = 3.67/\lambda$. In essence, raising the temperature makes a "tougher" test and thus requires fewer chips to be tested for the same significance level.

Problem 11.1.8 Solution

This problem has a lot of words, but is not all that hard.

(a) The pool leader has picked $W = \max(W_1, \ldots, W_n)$ games correctly. Under hypothesis H_0, the W_i are iid with PDF

$$P_{W_i|H_0}(w) = \binom{16m}{w} \left(\frac{1}{2}\right)^w \left(\frac{1}{2}\right)^{16m-w}. \tag{1}$$

Since $\mathrm{E}[W_i|H_0] = 8m$ and $\mathrm{Var}[W_i|H_0] = 16m(1/2)(1/2) = 4m$, we can use a Central Limit theorem approximation to write

$$\mathrm{P}[W_i \leq w|H_0] = \mathrm{P}\left[\frac{W_i - 8m}{2\sqrt{m}} \leq \frac{w - 8m}{2\sqrt{m}}\right] = \Phi\left(\frac{w - 8m}{s\sqrt{m}}\right). \tag{2}$$

Given H_0, the conditional CDF for W is

$$\begin{aligned}
\mathrm{P}[W \leq w|H_0] &= \mathrm{P}[\max(W_1, \ldots, W_n) \leq w|H_0] \\
&= \mathrm{P}[W_1 \leq w, \ldots, W_n \leq w|H_0] \\
&= \mathrm{P}[W_1 \leq w|H_0] \cdots \mathrm{P}[W_n \leq w|H_0] \\
&= (\mathrm{P}[W_i \leq w|H_0])^n = \Phi^n\left(\frac{w - 8m}{2\sqrt{m}}\right).
\end{aligned} \tag{3}$$

We choose the rejection region such that $W > w*$ because we want to reject the hypothesis H_0 that everyone is merely guessing if the leader does exceptionally well. Thus,

$$\alpha = \mathrm{P}[R] = \mathrm{P}[W > w^*|H_0] = 1 - \Phi^n\left(\frac{w^* - 8m}{2\sqrt{m}}\right). \tag{4}$$

For $\alpha = 0.05$, we find that

$$\Phi\left(\frac{w^* - 8m}{2\sqrt{m}}\right) = (0.95)^{1/n}. \tag{5}$$

For $n = 38$,

$$Q\left(\frac{w^* - 8m}{2\sqrt{m}}\right) = 1 - (0.95)^{1/38} = 1.35 \times 10^{-3}. \tag{6}$$

It follows that

$$\frac{w^* - 8m}{2\sqrt{m}} = 3, \tag{7}$$

or

$$w^* = 8m + 6\sqrt{m}. \tag{8}$$

(b) After $m = 14$, we require $w^* = 134.5$. Thus, if the leader Narayan has $w \geq 135$ winning picks after 14 weeks, then we reject the null hypothesis that Narayan has no ability to pick winners better than random selection. With only 199 correct picks, we conclude that Narayan has no special skills in picking games.

(c) The purpose of the point spread is to get a random bettor to be equally likely to pick either team. The point spread incorporates any side information about which team is better and thus more likely to win. With the point spread, it is a reasonable model to assume that all players' picks are just random guesses. In the absence of a point spread, all contestants in the pool would have some ability to pick winners beyond just guessing.

Problem 11.1.9 Solution

Since the null hypothesis H_0 asserts that the two exams have the same mean and variance, we reject H_0 if the difference in sample means is large. That is, $R = \{|D| \geq d_0\}$.

Under H_0, the two sample means satisfy

$$\begin{aligned} \mathrm{E}\left[M_A\right] &= \mathrm{E}\left[M_B\right] = \mu, \\ \mathrm{Var}[M_A] &= \mathrm{Var}[M_B] = \frac{\sigma^2}{n} = \frac{100}{n}. \end{aligned} \tag{1}$$

Since n is large, it is reasonable to use the Central Limit Theorem to approximate M_A and M_B as Gaussian random variables. Since M_A and M_B are independent, D is also Gaussian with

$$\begin{aligned} \mathrm{E}\left[D\right] &= \mathrm{E}\left[M_A\right] - \mathrm{E}\left[M_B\right] = 0, \\ \mathrm{Var}[D] &= \mathrm{Var}[M_A] + \mathrm{Var}[M_B] = \frac{200}{n}. \end{aligned} \tag{2}$$

Under the Gaussian assumption, we can calculate the significance level of the test as

$$\alpha = \mathrm{P}\left[|D| \geq d_0\right] = 2\left(1 - \Phi(d_0/\sigma_D)\right). \tag{3}$$

For $\alpha = 0.05$, $\Phi(d_0/\sigma_D) = 0.975$, or $d_0 = 1.96\sigma_D = 1.96\sqrt{200/n}$. If $n = 100$ students take each exam, then $d_0 = 2.77$ and we reject the null hypothesis that the exams are the same if the sample means differ by more than 2.77 points.

Problem 11.2.1 Solution

For the MAP test, we must choose acceptance regions A_0 and A_1 for the two hypotheses H_0 and H_1. From Theorem 11.2, the MAP rule is

$$n \in A_0 \text{ if } \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \frac{\mathrm{P}\left[H_1\right]}{\mathrm{P}\left[H_0\right]}; \qquad n \in A_1 \text{ otherwise.} \tag{1}$$

Since $P_{N|H_i}(n) = \lambda_i^n e^{-\lambda_i}/n!$, the MAP rule becomes

$$n \in A_0 \text{ if } \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{-(\lambda_0 - \lambda_1)} \geq \frac{\mathrm{P}\left[H_1\right]}{\mathrm{P}\left[H_0\right]}; \qquad n \in A_1 \text{ otherwise.} \tag{2}$$

By taking logarithms and assuming $\lambda_1 > \lambda_0$ yields the final form of the MAP rule

$$n \in A_0 \text{ if } n \leq n^* = \frac{\lambda_1 - \lambda_0 + \ln(\mathrm{P}\left[H_0\right] / \mathrm{P}\left[H_1\right])}{\ln(\lambda_1/\lambda_0)}; \qquad n \in A_1 \text{ otherwise.} \tag{3}$$

From the MAP rule, we can get the ML rule by setting the a priori probabilities to be equal. This yields the ML rule

$$n \in A_0 \text{ if } n \leq n^* = \frac{\lambda_1 - \lambda_0}{\ln(\lambda_1/\lambda_0)}; \qquad n \in A_1 \text{ otherwise.} \tag{4}$$

Problem 11.2.2 Solution

Hypotheses H_0 and H_1 have a priori probabilities $\mathrm{P}[H_0] = 0.8$ and $\mathrm{P}[H_1] = 0.2$ and likelihood functions

$$f_{T|H_0}(t) = \begin{cases} (1/3)e^{-t/3} & t \geq 0, \\ & \text{otherwise,} \end{cases}$$

$$f_{T|H_1}(t) = \begin{cases} (1/\mu_D)e^{-t/\mu_D} & t \geq 0, \\ & \text{otherwise.} \end{cases} \tag{1}$$

The acceptance regions are $A_0 = \{t|T \leq t_0\}$ and $A_1 = \{t|t > t_0\}$.

(a) The false alarm probability is

$$P_{\mathrm{FA}} = \mathrm{P}\left[A_1|H_0\right] = \int_{t_0}^{\infty} f_{T|H_0}(t)\ dt = e^{-t_0/3}. \tag{2}$$

(b) The miss probability is

$$P_{\mathrm{MISS}} = \mathrm{P}\left[A_0|H_1\right] = \int_{0}^{t_0} f_{T|H_1}(t)\ dt = 1 - e^{-t_0/\mu_D}. \tag{3}$$

(c) From Theorem 11.6, the maximum likelihood decision rule is

$$t \in A_0 \text{ if } \frac{f_{T|H_0}(t)}{f_{T|H_1}(t)} \geq 1; \qquad t \in A_1 \text{ otherwise.} \tag{4}$$

After some algebra, this rule simplifies to

$$t \in A_0 \text{ if } t \leq t_{ML} = \frac{\ln(\mu_D/3)}{1/3 - 1/\mu_D}; \qquad t \in A_1 \text{ otherwise.} \tag{5}$$

When $\mu_D = 6$ minutes, $t_{ML} = 6 \ln 2 = 4.16$ minutes. When $\mu_D = 10$ minutes, $t_{ML} = (30/7) \ln(10/3) = 5.16$ minutes.

(d) The ML rule is the same as the MAP rule when $\mathrm{P}[H_0] = \mathrm{P}[H_1]$. When $\mathrm{P}[H_0] > \mathrm{P}[H_1]$, the MAP rule (which minimizes the probability of an error) should enlarge the A_0 acceptance region. Thus we would expect $t_{\mathrm{MAP}} > t_{ML}$.

(e) From Theorem 11.2, the MAP rule is

$$t \in A_0 \text{ if } \frac{f_{T|H_0}(t)}{f_{T|H_1}(t)} \geq \frac{\mathrm{P}\left[H_1\right]}{\mathrm{P}\left[H_0\right]} = \frac{1}{4}; \qquad t \in A_1 \text{ otherwise.} \tag{6}$$

This rule simplifies to

$$t \in A_0 \text{ if } t \leq t_{\mathrm{MAP}} = \frac{\ln(4\mu_D/3)}{1/3 - 1/\mu_D}; \qquad t \in A_1 \text{ otherwise.} \tag{7}$$

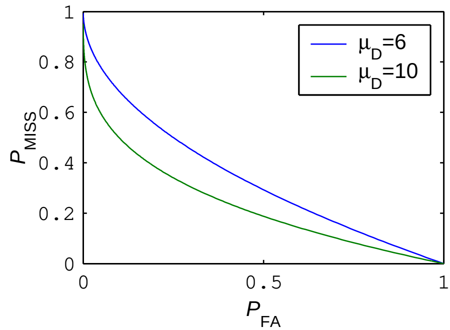
When $\mu_D = 6$ minutes, $t_{\mathrm{MAP}} = 6 \ln 8 = 12.48$ minutes. When $\mu_D = 10$ minutes, $t_{ML} = (30/7) \ln(40/3) = 11.1$ minutes.

(f) For a given threshold t_0, we learned in parts (a) and (b) that

$$P_{\rm FA} = e^{-t_0/3}, \qquad\qquad P_{\rm MISS} = 1 - e^{-t_0/\mu_D}. \tag{8}$$

The MATLAB program `rocvoicedataout` graphs both receiver operating curves. The program and the resulting ROC are shown here.

```
t=0:0.05:30;
PFA= exp(-t/3);
PMISS6= 1-exp(-t/6);
PMISS10=1-exp(-t/10);
plot(PFA,PMISS6,PFA,PMISS10);
legend('\mu_D=6','\mu_D=10');
xlabel('\itP_{\rmFA}');
ylabel('\itP_{\rmMISS}');
```



As one might expect, larger μ_D resulted in reduced $P_{\rm MISS}$ for the same $P_{\rm FA}$.

Problem 11.2.3 Solution

By Theorem 11.5, the decision rule is

$$n \in A_0 \text{ if } L(n) = \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \gamma; \qquad n \in A_1 \text{ otherwise}, \tag{1}$$

where γ is the largest possible value such that $\sum_{L(n)<\gamma} P_{N|H_0}(n) \leq \alpha$.

Given H_0, N is Poisson ($a_0 = 1{,}000$) while given H_1, N is Poisson ($a_1 = 1{,}300$). We can solve for the acceptance set A_0 by observing that $n \in A_0$ if

$$\frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} = \frac{a_0^n e^{-a_0}/n!}{a_1^n e^{-a_1}/n!} \geq \gamma. \tag{2}$$

Cancelling common factors and taking the logarithm, we find that $n \in A_0$ if

$$n \ln \frac{a_0}{a_1} \geq (a_0 - a_1) + \ln \gamma. \tag{3}$$

Since $\ln(a_0/a_1) < 0$, dividing through reverses the inequality and shows that

$$n \in A_0 \text{ if } n \leq n^* = \frac{(a_0 - a_1) + \ln\gamma}{\ln(a_0/a_1)} = \frac{(a_1 - a_0) - \ln\gamma}{\ln(a_1/a_0)}; \qquad n \in A_1 \text{ otherwise.}$$

However, we still need to determine the constant γ. In fact, it is easier to work with the threshold n^* directly. Note that $L(n) < \gamma$ if and only if $n > n^*$. Thus we choose the smallest n^* such that

$$\mathrm{P}\left[N > n^* | H_0\right] = \sum_{n > n^*} P_{N|H_0}(n)\, \alpha \leq 10^{-6}. \tag{4}$$

To find n^* a reasonable approach would be to use Central Limit Theorem approximation since given H_0, N is a Poisson (1,000) random variable, which has the same PDF as the sum on 1,000 independent Poisson (1) random variables. Given H_0, N has expected value a_0 and variance a_0. From the CLT,

$$\begin{aligned} \mathrm{P}\left[N > n^* | H_0\right] &= \mathrm{P}\left[\frac{N - a_0}{\sqrt{a_0}} > \frac{n^* - a_0}{\sqrt{a_0}} | H_0\right] \\ &\approx Q\left(\frac{n^* - a_0}{\sqrt{a_0}}\right) \leq 10^{-6}. \end{aligned} \tag{5}$$

From Table 4.3, $Q(4.75) = 1.02 \times 10^{-6}$ and $Q(4.76) < 10^{-6}$, implying

$$n^* = a_0 + 4.76\sqrt{a_0} = 1150.5. \tag{6}$$

On the other hand, perhaps the CLT should be used with some caution since $\alpha = 10^{-6}$ implies we are using the CLT approximation far from the center of the distribution. In fact, we can check out answer using the `poissoncdf` functions:

```
>> nstar=[1150 1151 1152 1153 1154 1155];
>> (1.0-poissoncdf(1000,nstar))'
ans =
 1.0e-005 *
   0.1644   0.1420   0.1225   0.1056   0.0910   0.0783
>>
```

Thus we see that n^*1154. Using this threshold, the miss probability is

$$\begin{aligned} \mathrm{P}\left[N \leq n^* | H_1\right] &= \mathrm{P}\left[N \leq 1154 | H_1\right] \\ &= \texttt{poissoncdf(1300,1154)} = 1.98 \times 10^{-5}. \end{aligned} \tag{7}$$

Keep in mind that this is the smallest possible P_{MISS} subject to the constraint that $P_{\text{FA}} \leq 10^{-6}$.

Problem 11.2.4 Solution

(a) Given H_0, X is Gaussian $(0,1)$. Given H_1, X is Gaussian $(4,1)$. From Theorem 11.2, the MAP hypothesis test is

$$x \in A_0 \text{ if } \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} = \frac{e^{-x^2/2}}{e^{-(x-4)^2/2}} \geq \frac{\mathrm{P}[H_1]}{\mathrm{P}[H_0]}; \qquad x \in A_1 \text{ otherwise.} \tag{1}$$

Since a target is present with probability $\mathrm{P}[H_1] = 0.01$, the MAP rule simplifies to

$$x \in A_0 \text{ if } x \leq x_{\mathrm{MAP}}; \qquad x \in A_1 \text{ otherwise} \tag{2}$$

where

$$x_{\mathrm{MAP}} = 2 - \frac{1}{4}\ln\left(\frac{\mathrm{P}[H_1]}{\mathrm{P}[H_0]}\right) = 3.15. \tag{3}$$

The false alarm and miss probabilities are

$$P_{\mathrm{FA}} = \mathrm{P}[X \geq x_{\mathrm{MAP}}|H_0] = Q(x_{\mathrm{MAP}}) = 8.16 \times 10^{-4}, \tag{4}$$

$$\begin{aligned} P_{\mathrm{MISS}} &= \mathrm{P}[X < x_{\mathrm{MAP}}|H_1] \\ &= \Phi(x_{\mathrm{MAP}} - 4) = 1 - \Phi(0.85) = 0.1977. \end{aligned} \tag{5}$$

The average cost of the MAP policy is

$$\begin{aligned} \mathrm{E}[C_{\mathrm{MAP}}] &= C_{10} P_{\mathrm{FA}} \mathrm{P}[H_0] + C_{01} P_{\mathrm{MISS}} \mathrm{P}[H_1] \\ &= (1)(8.16 \times 10^{-4})(0.99) + (10^4)(0.1977)(0.01) \\ &= 19.77. \end{aligned} \tag{6}$$

(b) The cost of a false alarm is $C_{10} = 1$ unit while the cost of a miss is $C_{01} = 10^4$ units. From Theorem 11.3, we see that the Minimum Cost test is the same as the MAP test except the $\mathrm{P}[H_0]$ is replaced by $C_{10}\,\mathrm{P}[H_0]$ and $\mathrm{P}[H_1]$ is replaced by $C_{01}\,\mathrm{P}[H_1]$. Thus, we see from the MAP test that the minimum cost test is

$$x \in A_0 \text{ if } x \le x_{\mathrm{MC}}; \qquad x \in A_1 \text{ otherwise.} \tag{7}$$

where

$$x_{\mathrm{MC}} = 2 - \frac{1}{4} \ln\left(\frac{C_{01}\,\mathrm{P}[H_1]}{C_{10}\,\mathrm{P}[H_0]}\right) = 0.846. \tag{8}$$

The false alarm and miss probabilities are

$$P_{\mathrm{FA}} = \mathrm{P}[X \ge x_{\mathrm{MC}}|H_0] = Q(x_{\mathrm{MC}}) = 0.1987, \tag{9}$$

$$\begin{aligned} P_{\mathrm{MISS}} &= \mathrm{P}[X < x_{\mathrm{MC}}|H_1] \\ &= \Phi(x_{\mathrm{MC}} - 4) = 1 - \Phi(3.154) = 8.06 \times 10^{-4}. \end{aligned} \tag{10}$$

The average cost of the minimum cost policy is

$$\begin{aligned} \mathrm{E}[C_{\mathrm{MC}}] &= C_{10} P_{\mathrm{FA}}\,\mathrm{P}[H_0] + C_{01} P_{\mathrm{MISS}}\,\mathrm{P}[H_1] \\ &= (1)(0.1987)(0.99) + (10^4)(8.06 \times 10^{-4})(0.01) \\ &= 0.2773. \end{aligned} \tag{11}$$

Because the cost of a miss is so high, the minimum cost test greatly reduces the miss probability, resulting in a much lower average cost than the MAP test.

Problem 11.2.5 Solution

Given H_0, X is Gaussian $(0,1)$. Given H_1, X is Gaussian $(v,1)$. By Theorem 11.4, the Neyman-Pearson test is

$$x \in A_0 \text{ if } L(x) = \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} = \frac{e^{-x^2/2}}{e^{-(x-v)^2/2}} \ge \gamma; \qquad x \in A_1 \text{ otherwise.} \tag{1}$$

This rule simplifies to

$$x \in A_0 \text{ if } L(x) = e^{-[x^2-(x-v)^2]/2} = e^{-vx+v^2/2} \ge \gamma; \qquad x \in A_1 \text{ otherwise.} \tag{2}$$

Taking logarithms, the Neyman-Pearson rule becomes

$$x \in A_0 \text{ if } x \le x_0 = \frac{v}{2} - \frac{1}{v}\ln\gamma; \qquad x \in A_1 \text{ otherwise.} \tag{3}$$

The choice of γ has a one-to-one correspondence with the choice of the threshold x_0. Moreoever $L(x) \ge \gamma$ if and only if $x \le x_0$. In terms of x_0, the false alarm probability is

$$P_{\text{FA}} = \mathrm{P}\left[L(X) < \gamma | H_0\right] = \mathrm{P}\left[X \ge x_0 | H_0\right] = Q(x_0). \tag{4}$$

Thus we choose x_0 such that $Q(x_0) = \alpha$.

Problem 11.2.6 Solution

(a) The sysadmin is doing something fairly tricky to insert packets to generate a Poisson process of rate $2\lambda_0$. Fortunately, we don't have to worry about how he does that here. Given $W = w$, the Poisson process rate is specified as either λ_0 if $w = 0$ or $\lambda_1 = 2\lambda_0$ if $w = 1$. Given $W = w$, we have a Poisson process of rate λ_w and $X_1, \ldots, X_n$ are iid exponential (λ_w) random variables, each with conditional PDF

$$f_{X_i|W=w}(x) = \begin{cases} \lambda_w e^{-\lambda_w x} & x \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

By conditional independence of the X_i, the conditional PDF of the vector $\mathbf{X}$ is

$$f_{\mathbf{X}|W=w}(\mathbf{x}) = f_{X_1,\ldots,X_n|W=w}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i|W=w}(x_i). \tag{2}$$

For non-negative vectors $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$, we can write

$$f_{\mathbf{X}|W=w}(\mathbf{x}) = \prod_{i=1}^{n} \left(\lambda_w e^{-\lambda_w x_i}\right) = \lambda_w^n e^{-\lambda_w \sum_{i=1}^{n} x_i}. \tag{3}$$

(b) Since the hypotheses are equiprobable, the MAP and ML tests are the same. Also keep in mind that the test has the same structure no matter whether the observation is a vector or a scalar. Thus the ML rule is simply that

$$\mathbf{x} \in A_0 \quad \text{if} \quad f_{\mathbf{X}|W=0}(\mathbf{x}) \geq f_{\mathbf{X}|W=1}(\mathbf{x}). \tag{4}$$

This simplifies to

$$\mathbf{x} \in A_0 \quad \text{if} \quad \lambda_0^n e^{-\lambda_0 \sum_{i=1}^n X_i} \geq (2\lambda_0)^n e^{-2\lambda_0 \sum_{i=1}^n X_i}, \tag{5}$$

or equivalently,

$$\mathbf{x} \in A_0 \quad \text{if} \quad \sum_{i=1}^n X_i \geq n \frac{\ln 2}{\lambda_0}. \tag{6}$$

In short, if the time needed for the router to produce n packets is sufficiently long, then the accomplice concludes that the router is in "normal mode' and that $W = 0$.

(c) For a random variable Y and constant c, recall that the Chernoff bound says $\mathrm{P}[Y \geq c] \leq \min_{s \geq 0} e^{-sc}\phi_Y(s)$. It will also be useful to recall that an Erlang (n, λ) random variable has MGF $[\lambda/(\lambda - s)]^n$. The error probability is

$$P_e = \mathrm{P}\left[\hat{W} = 0|W = 1\right] = \mathrm{P}\left[\sum_{i=1}^n X_i > n\frac{\ln 2}{\lambda_0}|W = 1\right]. \tag{7}$$

Using the hint, we define $Y = \sum_{i=1}^n X_i$. Given $W = 1$, Y is the sum of n exponential $(\lambda = 2\lambda_0)$ random variables. The hint should remind you to recall that Y is thus an Erlang $(n, 2\lambda_0)$ random variable with MGF

$$\phi_Y(s) = \frac{2\lambda_0}{2\lambda_0 - s}. \tag{8}$$

Defining $c = n\ln(2)/\lambda_0$, the Chernoff bound says

$$\begin{aligned} P_e = \mathrm{P}[Y > c] &\le \min_{s\ge 0} e^{-sc}\phi_Y(s) \\ &= \min_{s\ge 0} \left(e^{-(\ln 2)s/\lambda_0}\right)^n \left(\frac{2\lambda_0}{2\lambda_0 - s}\right)^n \\ &= \min_{s\ge 0} \left(\frac{2\lambda_0 e^{-(\ln 2)s/\lambda_0}}{2\lambda_0 - s}\right)^n \\ &= \min_{\tau\ge 0} \left(\frac{2e^{-\tau\ln 2}}{2-\tau}\right)^n \\ &= \left(2\min_{\tau\ge 0} \frac{e^{-\tau\ln 2}}{2-\tau}\right)^n. \end{aligned} \tag{9}$$

where we have defined $\tau = s/\lambda_0$. Setting a derivative to zero yields the minimizing τ is

$$\tau^* = 2 - \frac{1}{\ln 2}. \tag{10}$$

With some additional algebra, you can show that

$$P_e \le \left(\frac{2e^{-\tau^*\ln 2}}{2-\tau^*}\right)^n = [(2\ln 2)e^{1-2\ln 2}]^n \approx (0.94)^n. \tag{11}$$

This may seem not so good. However, if $n = 100$, then $P_e \le 0.0026$ and bits can be communicated with reasonable reliability to the accomplice.

Problem 11.2.7 Solution

Given H_0, $M_n(T)$ has expected value $\mathrm{E}[V]/n = 3/n$ and variance $\mathrm{Var}[V]/n = 9/n$. Given H_1, $M_n(T)$ has expected value $\mathrm{E}[D]/n = 6/n$ and variance $\mathrm{Var}[D]/n = 36/n$.

(a) Using a Central Limit Theorem approximation, the false alarm probability is

$$\begin{aligned} P_{\mathrm{FA}} &= \mathrm{P}[M_n(T) > t_0|H_0] \\ &= \mathrm{P}\left[\frac{M_n(T)-3}{\sqrt{9/n}} > \frac{t_0-3}{\sqrt{9/n}}\right] = Q(\sqrt{n}[t_0/3-1]). \end{aligned} \tag{1}$$

(b) Again, using a CLT Approximation, the miss probability is

$$\begin{aligned} P_{\text{MISS}} &= \text{P}\left[M_n(T) \leq t_0 | H_1\right] \\ &= \text{P}\left[\frac{M_n(T) - 6}{\sqrt{36/n}} \leq \frac{t_0 - 6}{\sqrt{36/n}}\right] = \Phi(\sqrt{n}[t_0/6 - 1]). \end{aligned} \tag{2}$$

(c) From Theorem 11.6, the maximum likelihood decision rule is

$$t \in A_0 \text{ if } \frac{f_{M_n(T)|H_0}(t)}{f_{M_n(T)|H_1}(t)} \geq 1; \qquad t \in A_1 \text{ otherwise.} \tag{3}$$

We will see shortly that using a CLT approximation for the likelihood functions is something of a detour. Nevertheless, with a CLT approximation, the likelihood functions are

$$\begin{aligned} f_{M_n(T)|H_0}(t) &= \sqrt{\frac{n}{18\pi}} e^{-n(t-3)^2/18}, \\ f_{M_n(T)|H_1}(t) &= \sqrt{\frac{n}{72\pi}} e^{-n(t-6)^2/72}. \end{aligned} \tag{4}$$

From the CLT approximation, the ML decision rule is

$$t \in A_0 \text{ if } \sqrt{\frac{72}{18}} \frac{e^{-n(t-3)^2/18}}{e^{-n(t-6)^2/72}} \geq 1; \qquad t \in A_1 \text{ otherwise.} \tag{5}$$

This simplifies to

$$t \in A_0 \text{ if } 2e^{-n[4(t-3)^2-(t-6)^2]/72} \geq 1; \qquad t \in A_1 \text{ otherwise.} \tag{6}$$

After more algebra, this rule further simplifies to

$$t \in A_0 \text{ if } t^2 - 4t - \frac{24 \ln 2}{n} \leq 0; \qquad t \in A_1 \text{ otherwise.} \tag{7}$$

Since the quadratic $t^2 - 4t - 24\ln(2)/n$ has two zeros, we use the quadratic formula to find the roots. One root corresponds to a negative value of t and can be discarded since $M_n(T) \geq 0$. Thus the ML rule (for $n = 9$) becomes

$$t \in A_0 \text{ if } t \leq t_{ML} = 2 + 2\sqrt{1 + 6\ln(2)/n} = 4.42; \qquad t \in A_1 \text{ otherwise.}$$

The negative root of the quadratic is the result of the Gaussian assumption which allows for a nonzero probability that $M_n(T)$ will be negative. In this case, hypothesis H_1 which has higher variance becomes more likely. However, since $M_n(T) \geq 0$, we can ignore this root since it is just an artifact of the CLT approximation.

In fact, the CLT approximation gives an incorrect answer. Note that $M_n(T) = Y_n/n$ where Y_n is a sum of iid exponential random variables. Under hypothesis H_0, Y_n is an Erlang $(n, \lambda_0 = 1/3)$ random variable. Under hypothesis H_1, Y_n is an Erlang $(n, \lambda_1 = 1/6)$ random variable. Since $M_n(T) = Y_n/n$ is a scaled version of Y_n, Theorem 6.3 tells us that given hypothesis H_i, $M_n(T)$ is an Erlang $(n, n\lambda_i)$ random variable. Thus $M_n(T)$ has likelihood functions

$$f_{M_n(T)|H_i}(t) = \begin{cases} \frac{(n\lambda_i)^n t^{n-1} e^{-n\lambda_i t}}{(n-1)!} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Using the Erlang likelihood functions, the ML rule becomes

$$t \in \begin{cases} A_0 & \text{if } \dfrac{f_{M_n(T)|H_0}(t)}{f_{M_n(T)|H_1}(t)} = \left(\dfrac{\lambda_0}{\lambda_1}\right)^n e^{-n(\lambda_0-\lambda_1)t} \geq 1; \\ A_1 & \text{otherwise.} \end{cases} \tag{9}$$

This rule simplifies to

$$t \in A_0 \text{ if } t \leq t_{\mathrm{ML}}; \qquad t \in A_1 \text{ otherwise.} \tag{10}$$

where

$$t_{\mathrm{ML}} = \frac{\ln(\lambda_0/\lambda_1)}{\lambda_0 - \lambda_1} = 6\ln 2 = 4.159. \tag{11}$$

Since $6\ln 2 = 4.159$, this rule is not the same as the rule derived using a CLT approximation. Using the exact Erlang PDF, the ML rule does not depend on n. Moreoever, even if $n \to \infty$, the exact Erlang-derived rule and the CLT approximation rule remain different. In fact, the CLT-based rule is simply an approximation to the correct rule. This highlights that we should first check whether a CLT approximation is necessary before we use it.

(d) In this part, we will use the exact Erlang PDFs to find the MAP decision rule. From 11.2, the MAP rule is

$$t \in \begin{cases} A_0 & \text{if } \dfrac{f_{M_n(T)|H_0}(t)}{f_{M_n(T)|H_1}(t)} = \left(\dfrac{\lambda_0}{\lambda_1}\right)^n e^{-n(\lambda_0-\lambda_1)t} \geq \dfrac{\mathrm{P}[H_1]}{\mathrm{P}[H_0]}; \\ A_1 & \text{otherwise.} \end{cases} \tag{12}$$

Since $\mathrm{P}[H_0] = 0.8$ and $\mathrm{P}[H_1] = 0.2$, the MAP rule simplifies to

$$t \in \begin{cases} A_0 & \text{if } t \leq t_{\mathrm{MAP}} = \dfrac{\ln \frac{\lambda_0}{\lambda_1} - \frac{1}{n} \ln \frac{\mathrm{P}[H_1]}{\mathrm{P}[H_0]}}{\lambda_0 - \lambda_1} = 6\left[\ln 2 + \dfrac{\ln 4}{n}\right]; \\ A_1 & \text{otherwise.} \end{cases} \tag{13}$$

For $n = 9$, $t_{\mathrm{MAP}} = 5.083$.

(e) Although we have seen it is incorrect to use a CLT approximation to derive the decision rule, the CLT approximation used in parts (a) and (b) remains a good way to estimate the false alarm and miss probabilities. However, given H_i, $M_n(T)$ is an Erlang $(n, n\lambda_i)$ random variable. In particular, given H_0, $M_n(T)$ is an Erlang $(n, n/3)$ random variable while given H_1, $M_n(T)$ is an Erlang $(n, n/6)$. Thus we can also use `erlangcdf` for an exact calculation of the false alarm and miss probabilities. To summarize the results of parts (a) and (b), a threshold t_0 implies that

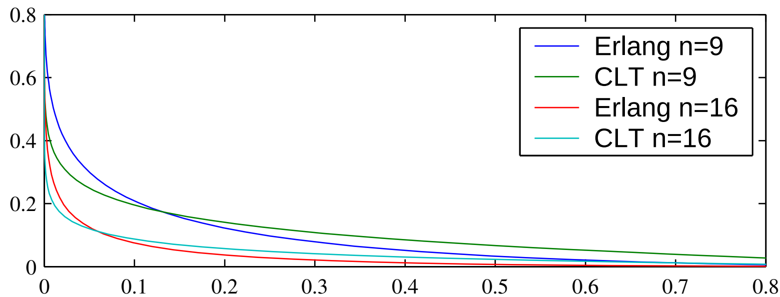
$$\begin{aligned} P_{\mathrm{FA}} &= \mathrm{P}[M_n(T) > t_0 | H_0] \\ &= \texttt{1-erlangcdf(n,n/3,t0)} \approx Q(\sqrt{n}[t_0/3 - 1]), \end{aligned} \tag{14}$$

$$\begin{aligned} P_{\mathrm{MISS}} &= \mathrm{P}[M_n(T) \leq t_0 | H_1] \\ &= \texttt{erlangcdf(n,n/6,t0)} \approx \Phi(\sqrt{n}[t_0/6 - 1]). \end{aligned} \tag{15}$$

Here is a program that generates the receiver operating curve.

```
%voicedatroc.m
t0=1:0.1:8';
n=9;
PFA9=1.0-erlangcdf(n,n/3,t0);
PFA9clt=1-phi(sqrt(n)*((t0/3)-1));
PM9=erlangcdf(n,n/6,t0);
PM9clt=phi(sqrt(n)*((t0/6)-1));
n=16;
PFA16=1.0-erlangcdf(n,n/3,t0);
PFA16clt=1.0-phi(sqrt(n)*((t0/3)-1));
PM16=erlangcdf(n,n/6,t0);
PM16clt=phi(sqrt(n)*((t0/6)-1));
plot(PFA9,PM9,PFA9clt,PM9clt,PFA16,PM16,PFA16clt,PM16clt);
axis([0 0.8 0 0.8]);
legend('Erlang n=9','CLT n=9','Erlang n=16','CLT n=16');
```

Here are the resulting ROCs.



Both the true curve and CLT-based approximations are shown. The graph makes it clear that the CLT approximations are somewhat innaccurate. It is also apparent that the ROC for $n = 16$ is clearly better than for $n = 9$.

Problem 11.2.8 Solution

This problem is a continuation of Problem 11.2.7. In this case, we say a call is a "success" if $T > t_0$. The success probability depends on which hypothesis is true. In particular,

$$p_0 = \mathrm{P}\left[T > t_0 | H_0\right] = e^{-t_0/3}, \qquad p_1 = \mathrm{P}\left[T > t_0 | H_1\right] = e^{-t_0/6}. \tag{1}$$

Under hypothesis H_i, K has the binomial (n, p_i) PMF

$$P_{K|H_i}(k) = \binom{n}{k} p_i^k (1-p_i)^{n-k}. \tag{2}$$

(a) A false alarm occurs if $K > k_0$ under hypothesis H_0. The probability of this event is

$$P_{\text{FA}} = \text{P}\left[K > k_0 | H_0\right] = \sum_{k=k_0+1}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k}, \tag{3}$$

(b) From Theorem 11.6, the maximum likelihood decision rule is

$$k \in \begin{cases} A_0 & \text{if } \dfrac{P_{K|H_0}(k)}{P_{K|H_1}(k)} = \dfrac{p_0^k(1-p_0)^{n-k}}{p_1^k(1-p_1)^{n-k}} \geq 1; \\ A_1 & \text{otherwise.} \end{cases} \tag{4}$$

This rule simplifies to

$$k \in \begin{cases} A_0 & \text{if } k \ln\left(\dfrac{p_0/(1-p_0)}{p_1/(1-p_1)}\right) \geq \ln\left(\dfrac{1-p_1}{1-p_0}\right) n; \\ A_1 & \text{otherwise.} \end{cases} \tag{5}$$

To proceed further, we need to know if $p_0 < p_1$ or if $p_0 \geq p_1$. For $t_0 = 4.5$,

$$p_0 = e^{-1.5} = 0.2231 < e^{-0.75} = 0.4724 = p_1. \tag{6}$$

In this case, the ML rule becomes

$$k \in \begin{cases} A_0 & \text{if } k \leq k_{ML} = \dfrac{\ln\left(\frac{1-p_0}{1-p_1}\right)}{\ln\left(\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right)} n = (0.340)n; \\ A_1 & \text{otherwise.} \end{cases} \tag{7}$$

For $n = 16$, $k_{ML} = 5.44$.

(c) From Theorem 11.2, the MAP test is

$$k \in \begin{cases} A_0 & \text{if } \dfrac{P_{K|H_0}(k)}{P_{K|H_1}(k)} = \dfrac{p_0^k(1-p_0)^{n-k}}{p_1^k(1-p_1)^{n-k}} \geq \dfrac{\mathrm{P}[H_1]}{\mathrm{P}[H_0]}; \\ A_1 & \text{otherwise.} \end{cases} \tag{8}$$

Since $\mathrm{P}[H_0] = 0.8$ and $\mathrm{P}[H_1] = 0.2$, this rule simplifies to

$$k \in \begin{cases} A_0 & \text{if } k \ln\left(\dfrac{p_0/(1-p_0)}{p_1/(1-p_1)}\right) \geq \ln\left(\dfrac{1-p_1}{1-p_0}\right) n; \\ A_1 & \text{otherwise.} \end{cases}. \tag{9}$$

For $t_0 = 4.5$, $p_0 = 0.2231 < p_1 = 0.4724$, the MAP rule becomes

$$k \in A_0 \text{ if } k \leq k_{\mathrm{MAP}}; \qquad k \in A_1 \text{ otherwise.} \tag{10}$$

where

$$k_{\mathrm{MAP}} = \frac{\ln\left(\frac{1-p_0}{1-p_1}\right) + \ln 4}{\ln\left(\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right)} n = (0.340)n + 1.22. \tag{11}$$

For $n = 16$, $k_{\mathrm{MAP}} = 6.66$.

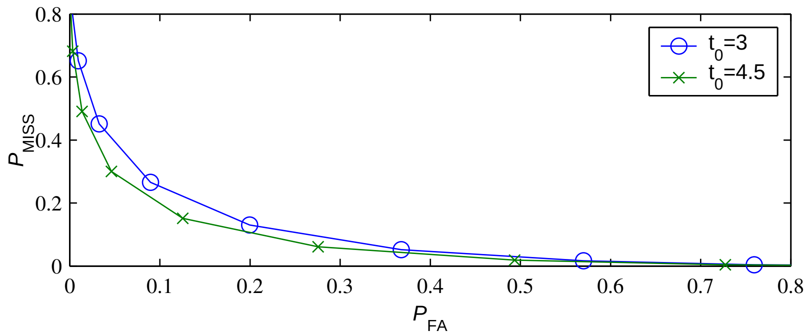
(d) For threshold k_0, the false alarm and miss probabilities are

$$P_{\mathrm{FA}} = \mathrm{P}[K > k_0 | H_0] = \texttt{1-binomialcdf(n,p0,k0)}, \tag{12}$$
$$P_{\mathrm{MISS}} = \mathrm{P}[K \leq k_0 | H_1] = \texttt{binomialcdf(n,p1,k0)}. \tag{13}$$

The ROC is generated by evaluating P_{FA} and P_{MISS} for each value of k_0. Here is a MATLAB program that does this task and plots the ROC.

```
function [PFA,PMISS]=binvoicedataroc(n);
t0=[3; 4.5];
p0=exp(-t0/3); p1=exp(-t0/6);
k0=(0:n)';
PFA=zeros(n+1,2);
for j=1:2,
    PFA(:,j) = 1.0-binomialcdf(n,p0(j),k0);
    PM(:,j)=binomialcdf(n,p1(j),k0);
end
plot(PFA(:,1),PM(:,1),'-o',PFA(:,2),PM(:,2),'-x');
legend('t_0=3','t_0=4.5');
axis([0 0.8 0 0.8]);
xlabel('\itP_{\rmFA}');
ylabel('\itP_{\rmMISS}');
```

and here is the resulting ROC:



As we see, the test works better with threshold $t_0 = 4.5$ than with $t_0 = 3$.

Problem 11.2.9 Solution

Given hypothesis H_0 that $X = 0$, $Y = W$ is an exponential ($\lambda = 1$) random variable. Given hypothesis H_1 that $X = 1$, $Y = V + W$ is an Erlang ($n = 2, \lambda = 1$) random variable. That is,

$$f_{Y|H_0}(y) = \begin{cases} e^{-y} & y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad f_{Y|H_1}(y) = \begin{cases} ye^{-y} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The probability of a decoding error is minimized by the MAP rule. Since $\mathrm{P}[H_0] = \mathrm{P}[H_1] = 1/2$, the MAP rule is

$$y \in A_0 \text{ if } \frac{f_{Y|H_0}(y)}{f_{Y|H_1}(y)} = \frac{e^{-y}}{ye^{-y}} \geq \frac{\mathrm{P}[H_1]}{\mathrm{P}[H_0]} = 1; \qquad y \in A_1 \text{ otherwise.} \tag{2}$$

Thus the MAP rule simplifies to

$$y \in A_0 \text{ if } y \leq 1; \qquad y \in A_1 \text{ otherwise.} \tag{3}$$

The probability of error is

$$\begin{aligned} P_{\mathrm{ERR}} &= \mathrm{P}[Y > 1|H_0]\,\mathrm{P}[H_0] + \mathrm{P}[Y \leq 1|H_1]\,\mathrm{P}[H_1] \\ &= \frac{1}{2}\int_1^\infty e^{-y}\,dy + \frac{1}{2}\int_0^1 ye^{-y}\,dy \\ &= \frac{e^{-1}}{2} + \frac{1-2e^{-1}}{2} = \frac{1-e^{-1}}{2}. \end{aligned} \tag{4}$$

Problem 11.2.10 Solution

(a) Composing the source and relay signals, we obtain

$$\begin{aligned} Y_i = \beta_i X_i + Z_i &= \beta_i(\alpha_i V + W_i) + Z_i \\ &= \alpha_i\beta_i V + (\beta_i W_i + Z_i) \\ &= \gamma_i V + \hat{Z}_i, \end{aligned} \tag{1}$$

where $\gamma_i = \alpha_i\beta_i$ describes the effective channel gain and $\hat{Z}_i = \beta_i W_i + Z_i$ is the effective receiver noise. We note that $\hat{Z}_i$ is the sum of independent Gaussian random variables so that $\mathrm{E}[\hat{Z}_i] = 0$ and $\mathrm{Var}[\hat{Z}_i] = 1 + \beta_i^2$. The receiver decisions statistic is

$$Y = \sum_{i=1}^n Y_i = \left(\sum_{i=1}^n \gamma_i\right) V + \sum_{i=1}^n \hat{Z}_i = \Gamma V + \hat{Z}, \tag{2}$$

where $\Gamma = \sum_{i=1}^n \gamma_i$ and $\hat{Z} = \sum_{i=1}^n \hat{Z}_i$. In this case, the components of $\hat{Z}$ are independent and thus $\hat{Z}$ is Gaussian with $\mathrm{E}[\hat{Z}] = 0$ and

$$\mathrm{Var}[\hat{Z}] = \sum_{i=1}^n \mathrm{Var}[\hat{Z}_i] = n + \sum_{i=1}^n \beta_i^2. \tag{3}$$

By symmetry the error probability is

$$\begin{aligned} P_e &= \mathrm{P}\left[Y > 0 | V = -1\right] \\ &= \mathrm{P}\left[-\Gamma + \hat{Z} > 0 | V = -1\right] \\ &= \mathrm{P}\left[-\Gamma + \hat{Z} > 0\right] = \mathrm{P}\left[\hat{Z} > \gamma\right]. \end{aligned} \tag{4}$$

Since $\hat{Z}$ is Gaussian,

$$P_e = Q\left(\frac{\Gamma}{\sqrt{\mathrm{Var}[\hat{Z}]}}\right) = Q\left(\frac{\sum_{i=1}^n \alpha_i \beta_i}{\sqrt{n + \sum_{i=1}^n \beta_i^2}}\right). \tag{5}$$

(b) The key in this problem is that the receiver observes and uses the components Y_i of the vector $\mathbf{Y}$. Note that given hypothesis H_0 that $V = -1$,

$$Y_i = -\gamma_i + \hat{Z}_i, \qquad i = 1, 2, \ldots, n. \tag{6}$$

Since the $\hat{Z}_i$ are independent, the Y_i are conditionally independent given H_0. In addition, given $V = -1$, Y_i is Gaussian $(-\gamma_i, \hat{\sigma}_i)$ where $\hat{\sigma}_i = \sqrt{1 + \beta_i^2}$. This implies

$$f_{\mathbf{Y}|H_0}(\mathbf{y}) = \prod_{i=1}^n f_{Y_i|V=-1}(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} e^{-(y_i+\gamma_i)^2/2\hat{\sigma}_i^2}. \tag{7}$$

Similarly, when $V = 1$, $Y_i = \gamma_i + \hat{Z}_i$ and

$$f_{\mathbf{Y}|H_1}(\mathbf{y}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} e^{-(y_i-\gamma_i)^2/2\hat{\sigma}_i^2} \tag{8}$$

To find the MAP test based on the vector $\mathbf{Y}$, we need to find the optimal acceptance regions A_i for hypothesis H_i. In general, these are given by

$$\mathbf{y} \in A_0 \quad \text{if} \quad \mathrm{P}[H_0] f_{\mathbf{Y}|H_0}(\mathbf{y}) \geq \mathrm{P}[H_1] f_{\mathbf{Y}|H_1}(\mathbf{y}). \tag{9}$$

Since H_0 and H_1 are equally likely, this simplifies to

$$\mathbf{y} \in A_0 \quad \text{if} \quad \frac{f_{\mathbf{Y}|H_0}(\mathbf{y})}{f_{\mathbf{Y}|H_0}(\mathbf{y})} = \frac{\prod_{i=1}^n e^{-(y_i+\gamma_i)^2/2\hat{\sigma}_i^2}}{\prod_{i=1}^n e^{-(y_i-\gamma_i)^2/2\hat{\sigma}_i^2}} \geq 1. \tag{10}$$

Taking the logarithm of both sides, we obtain

$$\mathbf{y} \in A_0 \quad \text{if} \quad -\sum_{i=1}^n \frac{(y_i+\gamma_i)^2}{2\hat{\sigma}_i^2} + \sum_{i=1}^n \frac{(y_i-\gamma_i)^2}{2\hat{\sigma}_i^2} \geq 0, \tag{11}$$

$$\text{if} \quad \sum_{i=1}^n \left(\frac{-(y_i^2 + 2\gamma_i y_i + \gamma_i^2)}{2\hat{\sigma}_i^2} + \frac{y_i^2 - 2\gamma_i y_i + \gamma_i^2}{2\hat{\sigma}_i^2} \right) \geq 0. \tag{12}$$

Defining $\omega_i = \gamma_i/\hat{\sigma}_i^2$, this simplifies to

$$\mathbf{y} \in A_0 \quad \text{if} \quad \sum_{i=1}^n \omega_i y_i \leq 0. \tag{13}$$

(c) In terms of the observation $\mathbf{Y}$, the decision statistic is

$$Y^* = \sum_{i=1} \omega_i Y_i = \underbrace{\left(\sum_{i=1}^n \omega_i \gamma_i \right)}_{\Gamma^*} V + \underbrace{\sum_{i=1}^n \omega_i \hat{Z}_i}_{Z^*}. \tag{14}$$

This is the same detection problem and decision rule as in part (a) with Γ and $\hat{Z}$ replaced by Γ^* and Z^*. In particular, note that

$$\Gamma^* = \sum_{i=1}^n \omega_i \gamma_i = \sum_{i=1}^n \frac{\gamma_i^2}{\hat{\sigma}_i^2} = \sum_{i=1}^n \frac{\alpha_i \beta_i}{1+\beta_i^2} \tag{15}$$

and that Z^* is Gaussian with variance

$$\mathrm{Var}[Z^*] = \sum_{i=1}^n \mathrm{Var}[\omega_i \hat{Z}_i] = \sum_{i=1}^n \omega_i^2 \hat{\sigma}_i^2 = \sum_{i=1}^n \frac{\gamma_i^2}{\hat{\sigma}_i^2} = \Gamma^*. \tag{16}$$

It follows that the MAP detector has error probability

$$
\begin{aligned}
P_e^* &= \mathrm{P}\left[Z^* > \Gamma^*\right] \\
&= Q\left(\frac{\Gamma^*}{\sqrt{\mathrm{Var}[Z^*]}}\right) = Q(\sqrt{\Gamma^*}) = Q\left(\sum_{i=1}^{n} \frac{\alpha_i \beta_i}{1+\beta_i^2}\right).
\end{aligned} \tag{17}
$$

(d) Plugging in the values, we obtain

$$
P_e = Q\left(\frac{1+10+10+100}{\sqrt{4+1+1+100+100}}\right) = Q(8.43). \tag{18}
$$

By comparison

$$
\Gamma^* = \frac{1}{2} + \frac{100}{2} + \frac{100}{101} + \frac{10^4}{101} = 150.5, \tag{19}
$$

and

$$
P_e^* = Q(\sqrt{150.5}) = Q(12.27). \tag{20}
$$

Although the numbers are semi-ridiculous, $Q(12.7) \ll Q(8.4)$. The reason is that relay 3 is amplifying a very noisy signal and the destination receives Y_3 as a high-power but low SNR signal. In the equal gain combiner, this hurts the simple detector.

Problem 11.2.11 Solution

(a) Since $\mathbf{Y}$ is continuous, the conditioning event is $\mathbf{y} < \mathbf{Y} \leq \mathbf{y} + d\mathbf{y}$. We then write

$$
\begin{aligned}
&\mathrm{P}\left[X = 1 | \mathbf{y} < \mathbf{Y} \leq \mathbf{y} + d\mathbf{y}\right] \\
&\qquad = \frac{\mathrm{P}\left[\mathbf{y} < \mathbf{Y} \leq \mathbf{y} + d\mathbf{y} | X = 1\right] \mathrm{P}\left[X = 1\right]}{\mathrm{P}\left[\mathbf{y} < \mathbf{Y} \leq \mathbf{y} + d\mathbf{y}\right]} \\
&\qquad = \frac{\frac{1}{2} f_{\mathbf{Y}|X}(\mathbf{y}|1)\, d\mathbf{y}}{\frac{1}{2} f_{\mathbf{Y}|X}(\mathbf{y}|1)\ d\mathbf{y} + \frac{1}{2} f_{\mathbf{Y}|X}(\mathbf{y}|-1)\ d\mathbf{y}}.
\end{aligned} \tag{1}
$$

We conclude that

$$\mathrm{P}\left[X=1|\mathbf{Y}=\mathbf{y}\right] = \frac{f_{\mathbf{Y}|X}(\mathbf{y}|1)}{f_{\mathbf{Y}|X}(\mathbf{y}|1)+f_{\mathbf{Y}|X}(\mathbf{y}|-1)} = \frac{1}{1+\frac{f_{\mathbf{Y}|X}(\mathbf{y}|-1)}{f_{\mathbf{Y}|X}(\mathbf{y}|1)}}. \tag{2}$$

Given $X=x$, $Y_i = x+W_i$ and

$$f_{Y_i|X}(y_i|x) = \frac{1}{\sqrt{2\pi}} e^{(y_i-x)^2/2}. \tag{3}$$

Since the W_i are iid and independent of X, given $X=x$, the Y_i are conditionally iid. That is,

$$f_{\mathbf{Y}|X}(\mathbf{y}|x) = \prod_{i=1}^{n} f_{Y_i|X}(y_i|x) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^{n}(y_i-x)^2/2}. \tag{4}$$

This implies

$$\begin{aligned} L(\mathbf{y}) = \frac{f_{\mathbf{Y}|X}(\mathbf{y}|-1)}{f_{\mathbf{Y}|X}(\mathbf{y}|1)} &= \frac{e^{-\sum_{i=1}^{n}(y_i+1)^2/2}}{e^{-\sum_{i=1}^{n}(y_i-1)^2/2}} \\ &= e^{-\sum_{i=1}^{n}[(y_i+1)^2-(y_i-1)^2]/2} = e^{-2\sum_{i=1}^{n} y_i} \end{aligned} \tag{5}$$

and that

$$\mathrm{P}\left[X=1|\mathbf{Y}=\mathbf{y}\right] = \frac{1}{1+L(\mathbf{y})} = \frac{1}{1+e^{-2\sum_{i=1}^{n} y_i}}. \tag{6}$$

(b)

$$\begin{aligned} P_e = \mathrm{P}\left[X^* \neq X|X=1\right] &= \mathrm{P}\left[X^*=-1|X=1\right] \\ &= \mathrm{P}\left[\frac{1}{1+L(\mathbf{Y})} < \frac{1}{2}|X=1\right] \\ &= \mathrm{P}\left[L(\mathbf{Y})>1|X=1\right] \\ &= \mathrm{P}\left[e^{-2\sum_{i=1}^{n} Y_i} > 1|X=1\right] \\ &= \mathrm{P}\left[\sum_{i=1}^{n} Y_i < 0|X=1\right]. \end{aligned} \tag{7}$$

Given $X = 1$, $Y_i = 1 + W_i$ and $\sum_{i=1}^{n} Y_i = n + W$ where $W = \sum_{i=1}^{n} W_n$ is a Gaussian $(0, \sqrt{n})$ random variable. Since, W is independent of X,

$$P_e = \mathrm{P}\left[n + W < 0 | X = 1\right] = \mathrm{P}\left[W < -n\right] = \mathrm{P}\left[W > n\right] = Q(\sqrt{n}). \tag{8}$$

(c) First we observe that we decide $X^* = 1$ on stage n iff

$$\begin{aligned} \hat{X}_n(\mathbf{Y}) > 1 - \epsilon &\Rightarrow 2\,\mathrm{P}\left[X = 1|\mathbf{Y} = \mathbf{y}\right] - 1 < 1 - \epsilon \\ &\Rightarrow \mathrm{P}\left[X = 1|\mathbf{Y} = \mathbf{y}\right] > 1 - \epsilon/2 = \epsilon_2. \end{aligned} \tag{9}$$

However, when we decide $X^* = 1$ given $\mathbf{Y} = \mathbf{y}$, the probability of a correct decision is $\mathrm{P}[X = 1|\mathbf{Y} = \mathbf{y}]$. The probability of an error thus satisfies

$$P_e = 1 - \mathrm{P}\left[X = 1|\mathbf{Y} = \mathbf{y}\right] < \epsilon/2. \tag{10}$$

This answer is simple if the logic occurs to you. In some ways, the following lower bound derivation is more straightforward. If $X = 1$, an error occurs after transmission $n = 1$ if $\hat{X}_1(\mathbf{y}) < -1 + \epsilon$. Thus

$$\begin{aligned} P_e &\geq \mathrm{P}\left[\hat{X}_1(\mathbf{y}) < -1 + \epsilon | X = 1\right] \\ &= \mathrm{P}\left[\frac{1 - L(\mathbf{y})}{1 + L(\mathbf{y})} < -1 + \epsilon | X = 1\right] \\ &= \mathrm{P}\left[L(\mathbf{y}) > (2/\epsilon) - 1 | X = 1\right]. \end{aligned} \tag{11}$$

For $n = 1$, $L(\mathbf{y}) = e^{-2Y_1} = e^{-2(X+W_1)}$. This implies

$$\begin{aligned} P_e &\geq \mathrm{P}\left[e^{-2(X+W_1)} > 2/\epsilon - 1 | X = 1\right] \\ &= \mathrm{P}\left[1 + W_1 < -\frac{1}{2}\ln\left(\frac{2}{\epsilon} - 1\right)\right] \\ &= \mathrm{P}\left[W_1 < -1 - \ln\sqrt{\frac{2}{\epsilon} - 1}\right] = Q\left(1 + \sqrt{\frac{2}{\epsilon} - 1}\right) = \epsilon_1. \end{aligned} \tag{12}$$

Problem 11.2.12 Solution

Given hypothesis H_i, K has the binomial PMF

$$P_{K|H_i}(k) = \binom{n}{k} q_i^k (1-q_i)^{n-k}. \tag{1}$$

(a) The ML rule is

$$k \in A_0 \text{ if } P_{K|H_0}(k) > P_{K|H_1}(k)\,; \qquad k \in A_1 \text{ otherwise.} \tag{2}$$

When we observe $K = k \in \{0, 1, \ldots, n\}$, plugging in the conditional PMF's yields the rule

$$k \in \begin{cases} A_0 & \text{if } \binom{n}{k} q_0^k (1-q_0)^{n-k} > \binom{n}{k} q_1^k (1-q_1)^{n-k}; \\ A_1 & \text{otherwise.} \end{cases} \tag{3}$$

Canceling common factors, taking the logarithm of both sides, and rearranging yields

$$k \in \begin{cases} A_0 & \text{if } k \ln\left(\dfrac{q_1/(1-q_1)}{q_0/(1-q_0)}\right) < n \ln\left(\dfrac{1-q_0}{1-q_1}\right); \\ A_1 & \text{otherwise.} \end{cases} \tag{4}$$

Note that $q_1 > q_0$ implies $q_1/(1-q_1) > q_0/(1-q_0)$. Thus, we can rewrite our ML rule as

$$k \in \begin{cases} A_0 & \text{if } k < k^* = n \dfrac{\ln[(1-q_0)/(1-q_1)]}{\ln[q_1/q_0] + \ln[(1-q_0)/(1-q_1)]}; \\ A_1 & \text{otherwise.} \end{cases} \tag{5}$$

(b) Let k^* denote the threshold given in part (a). Using $n = 500$, $q_0 = 10^{-4}$, and $q_1 = 10^{-2}$, we have

$$k^* = 500 \frac{\ln[(1-10^{-4})/(1-10^{-2})]}{\ln[10^{-2}/10^{-4}] + \ln[(1-10^{-4})/(1-10^{-2})]} \approx 1.078. \tag{6}$$

Thus the ML rule is that if we observe $K \leq 1$, then we choose hypothesis H_0; otherwise, we choose H_1. The false alarm probability is

$$\begin{aligned} P_{\mathrm{FA}} = \mathrm{P}\left[A_1|H_0\right] &= \mathrm{P}\left[K > 1|H_0\right] \\ &= 1 - P_{K|H_0}(0) - P_{K|H_1}(1) \\ &= 1 - (1-q_0)^{500} - 500q_0(1-q_0)^{499} = 0.0012. \end{aligned} \tag{7}$$

and the miss probability is

$$\begin{aligned} P_{\mathrm{MISS}} = \mathrm{P}\left[A_0|H_1\right] &= \mathrm{P}\left[K \leq 1|H_1\right] \\ &= P_{K|H_1}(0) + P_{K|H_1}(1) \\ &= (1-q_1)^{500} + 500q_1(1-q_1)^{499} = 0.0398. \end{aligned} \tag{8}$$

(c) In the test of Example 11.8, the geometric random variable N, the number of tests needed to find the first failure, was used. In this problem, the binomial random variable K, the number of failures in 500 tests, was used. We will call these two procedures the geometric and the binomial tests. Also, we will use $P_{\mathrm{FA}}^{(N)}$ and $P_{\mathrm{MISS}}^{(N)}$ to denote the false alarm and miss probabilities using the geometric test. We also use $P_{\mathrm{FA}}^{(K)}$ and $P_{\mathrm{MISS}}^{(K)}$ for the error probabilities of the binomial test. From Example 11.8, we have the following comparison:

$$\begin{array}{cc} \text{geometric test} & \text{binomial test} \\ P_{\mathrm{FA}}^{(N)} = 0.0045, & P_{\mathrm{FA}}^{(K)} = 0.0012, \\ P_{\mathrm{MISS}}^{(N)} = 0.0087, & P_{\mathrm{MISS}}^{(K)} = 0.0398. \end{array} \tag{9}$$

When making comparisons between tests, we want to judge both the reliability of the test as well as the cost of the testing procedure. With respect to the reliability, we see that the conditional error probabilities appear to be comparable in that

$$\frac{P_{\mathrm{FA}}^{(N)}}{P_{\mathrm{FA}}^{(K)}} = 3.75 \quad \text{but} \quad \frac{P_{\mathrm{MISS}}^{(K)}}{P_{\mathrm{MISS}}^{(N)}} = 4.57. \tag{10}$$

Roughly, the false alarm probability of the geometric test is about four times higher than that of the binomial test. However, the miss probability of the binomial test is about four times that of the geometric test. As for the cost of the test, it is reasonable to assume the cost is proportional to the number of disk drives that are tested. For the geometric test of Example 11.8, we test either until the first failure or until 46 drives pass the test. For the binomial test, we test until either 2 drives fail or until 500 drives the pass the test! You can, if you wish, calculate the expected number of drives tested under each test method for each hypothesis. However, it isn't necessary in order to see that a lot more drives will be tested using the binomial test. If we knew the a priori probabilities $\mathrm{P}[H_i]$ and also the relative costs of the two type of errors, then we could determine which test procedure was better. However, without that information, it would not be unreasonable to conclude that the geometric test offers performance roughly comparable to that of the binomial test but with a significant reduction in the expected number of drives tested.

Problem 11.2.13 Solution

The key to this problem is to observe that

$$\mathrm{P}\left[A_0|H_0\right] = 1 - \mathrm{P}\left[A_1|H_0\right], \qquad \mathrm{P}\left[A_1|H_1\right] = 1 - \mathrm{P}\left[A_0|H_1\right]. \tag{1}$$

The total expected cost can be written as

$$\begin{aligned} \mathrm{E}\left[C'\right] &= \mathrm{P}\left[A_1|H_0\right]\mathrm{P}\left[H_0\right]C'_{10} + (1 - \mathrm{P}\left[A_1|H_0\right])\,\mathrm{P}\left[H_0\right]C'_{00} \\ &\quad + \mathrm{P}\left[A_0|H_1\right]\mathrm{P}\left[H_1\right]C'_{01} + (1 - \mathrm{P}\left[A_0|H_1\right])\,\mathrm{P}\left[H_1\right]C'_{11}. \end{aligned} \tag{2}$$

Rearranging terms, we have

$$\begin{aligned} \mathrm{E}\left[C'\right] &= \mathrm{P}\left[A_1|H_0\right]\mathrm{P}\left[H_0\right](C'_{10} - C'_{00}) + \mathrm{P}\left[A_0|H_1\right]\mathrm{P}\left[H_1\right](C'_{01} - C'_{11}) \\ &\quad + \mathrm{P}\left[H_0\right]C'_{00} + \mathrm{P}\left[H_1\right]C'_{11}. \end{aligned} \tag{3}$$

Since $\mathrm{P}[H_0]C'_{00} + \mathrm{P}[H_1]C'_{11}$ does not depend on the acceptance sets A_0 and A_1, the decision rule that minimizes $\mathrm{E}[C']$ is the same decision rule that minimizes

$$\mathrm{E}\left[C''\right] = \mathrm{P}\left[A_1|H_0\right]\mathrm{P}\left[H_0\right](C'_{10} - C'_{00}) + \mathrm{P}\left[A_0|H_1\right]\mathrm{P}\left[H_1\right](C'_{01} - C'_{11}). \tag{4}$$

The decision rule that minimizes $\mathrm{E}[C'']$ is the same as the minimum cost test in Theorem 11.3 with the costs C_{01} and C_{10} replaced by the differential costs $C'_{01} - C'_{11}$ and $C'_{10} - C'_{00}$.

Problem 11.3.1 Solution

Since the three hypotheses H_0, H_1, and H_2 are equally likely, the MAP and ML hypothesis tests are the same. From Theorem 11.8, the MAP rule is

$$x \in A_m \text{ if } f_{X|H_m}(x) \geq f_{X|H_j}(x) \text{ for all } j. \tag{1}$$

Since N is Gaussian with zero mean and variance σ_N^2, the conditional PDF of X given H_i is

$$f_{X|H_i}(x) = \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-(x-a(i-1))^2/2\sigma_N^2}. \tag{2}$$

Thus, the MAP rule is

$$x \in A_m \text{ if } (x - a(m-1))^2 \leq (x - a(j-1))^2 \text{ for all } j. \tag{3}$$

This implies that the rule for membership in A_0 is

$$x \in A_0 \text{ if } (x+a)^2 \leq x^2 \text{ and } (x+a)^2 \leq (x-a)^2. \tag{4}$$

This rule simplifies to

$$x \in A_0 \text{ if } x \leq -a/2. \tag{5}$$

Similar rules can be developed for A_1 and A_2. These are:

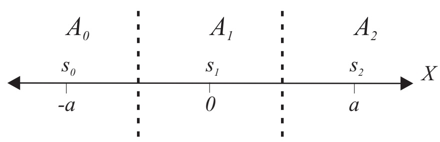
$$x \in A_1 \text{ if } -a/2 \leq x \leq a/2, \tag{6}$$

$$x \in A_2 \text{ if } x \geq a/2. \tag{7}$$

To summarize, the three acceptance regions are

$$A_0 = \{x | x \leq -a/2\}, A_1 = \{x | -a/2 < x \leq a/2\}, A_2 = \{x | x > a/2\}. \tag{8}$$

Graphically, the signal space is one dimensional and the acceptance regions are



Just as in the QPSK system of Example 11.13, the additive Gaussian noise dictates that the acceptance region A_i is the set of observations x that are closer to $s_i = (i-1)a$ than any other s_j.

Problem 11.3.2 Solution

Let the components of $\mathbf{s}_{ijk}$ be denoted by $s_{ijk}^{(1)}$ and $s_{ijk}^{(2)}$ so that given hypothesis H_{ijk},

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} s_{ijk}^{(1)} \\ s_{ijk}^{(2)} \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix}. \tag{1}$$

As in Example 11.13, we will assume N_1 and N_2 are iid zero mean Gaussian random variables with variance σ^2. Thus, given hypothesis H_{ijk}, X_1 and X_2 are independent and the conditional joint PDF of X_1 and X_2 is

$$\begin{aligned} f_{X_1,X_2|H_{ijk}}(x_1,x_2) &= f_{X_1|H_{ijk}}(x_1)\, f_{X_2|H_{ijk}}(x_2) \\ &= \frac{1}{2\pi\sigma^2} e^{-(x_1-s_{ijk}^{(1)})^2/2\sigma^2} e^{-(x_2-s_{ijk}^{(2)})^2/2\sigma^2} \\ &= \frac{1}{2\pi\sigma^2} e^{-[(x_1-s_{ijk}^{(1)})^2+(x_2-s_{ijk}^{(2)})^2]/2\sigma^2}. \end{aligned} \tag{2}$$

In terms of the distance $\|\mathbf{x}-\mathbf{s}_{ijk}\|$ between vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \mathbf{s}_{ijk} = \begin{bmatrix} s_{ijk}^{(1)} \\ s_{ijk}^{(2)}. \end{bmatrix} \tag{3}$$

we can write

$$f_{X_1,X_2|H_i}(x_1,x_2) = \frac{1}{2\pi\sigma^2} e^{-\left\|\mathbf{x}-\mathbf{s}_{ijk}\right\|^2/2\sigma^2}. \tag{4}$$

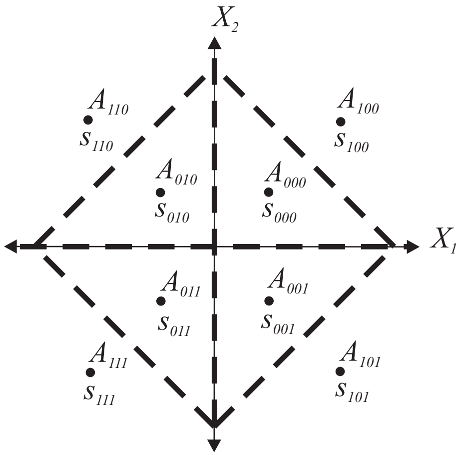
Since all eight symbols $s_{000},\ldots,s_{111}$ are equally likely, the MAP and ML rules are

$$\mathbf{x} \in A_{ijk} \text{ if } f_{X_1,X_2|H_{ijk}}(x_1,x_2) \geq f_{X_1,X_2|H_{i'j'k'}}(x_1,x_2) \text{ for all other } H_{i'j'k'}. \tag{5}$$

This rule simplifies to

$$\mathbf{x} \in A_{ijk} \text{ if } \|\mathbf{x}-\mathbf{s}_{ijk}\| \leq \|\mathbf{x}-\mathbf{s}_{ijk}\| \text{ for all other } i'j'k'. \tag{6}$$

This means that A_{ijk} is the set of all vectors $\mathbf{x}$ that are closer to $\mathbf{s}_{ijk}$ than any other signal. Graphically, to find the boundary between points closer to $\mathbf{s}_{ijk}$ than $\mathbf{s}_{i'j'k'}$, we draw the line segment connecting $\mathbf{s}_{ijk}$ and $\mathbf{s}_{i'j'k'}$. The boundary is then the perpendicular bisector. The resulting boundaries are shown in this figure:



Problem 11.3.3 Solution

Let H_i denote the hypothesis that symbol a_i was transmitted. Since the four hypotheses are equally likely, the ML tests will maximize the probability of a correct decision. Given H_i, N_1 and N_2 are independent and thus X_1 and X_2 are independent. This implies

$$\begin{aligned} f_{X_1,X_2|H_i}(x_1,x_2) &= f_{X_1|H_i}(x_1)\, f_{X_2|H_i}(x_2) \\ &= \frac{1}{2\pi\sigma^2} e^{-(x_1-s_{i1})^2/2\sigma^2} e^{-(x_2-s_{i2})^2/2\sigma^2} \\ &= \frac{1}{2\pi\sigma^2} e^{-[(x_1-s_{i1})^2+(x_2-s_{i2})^2]/2\sigma^2}. \end{aligned} \tag{1}$$

From Definition 11.2 the acceptance regions A_i for the ML multiple hypothesis test must satisfy

$$(x_1,x_2) \in A_i \text{ if } f_{X_1,X_2|H_i}(x_1,x_2) \geq f_{X_1,X_2|H_j}(x_1,x_2) \text{ for all } j. \tag{2}$$

Equivalently, the ML acceptance regions are

$$(x_1,x_2) \in A_i \text{ if } (x_1-s_{i1})^2 + (x_2-s_{i2})^2 \leq \min_j (x_1-s_{j1})^2 + (x_2-s_{j2})^2. \tag{3}$$

In terms of the vectors $\mathbf{x}$ and $\mathbf{s}_i$, the acceptance regions are defined by the rule

$$\mathbf{x} \in A_i \text{ if } \|\mathbf{x}-\mathbf{s}_i\|^2 \leq \|\mathbf{x}-\mathbf{s}_j\|^2. \tag{4}$$

Just as in the case of QPSK, the acceptance region A_i is the set of vectors $\mathbf{x}$ that are closest to $\mathbf{s}_i$.

Problem 11.3.4 Solution

From the signal constellation depicted in Problem 11.3.4, each signal s_{ij1} is below the x-axis while each signal s_{ij0} is above the x-axis. The event B_3 of an error in the third bit occurs if we transmit a signal s_{ij1} but the receiver output $\mathbf{x}$ is above the x-axis or if we transmit a signal s_{ij0} and the receiver output is below the x-axis. By symmetry, we need only consider the case when we transmit one of the four signals s_{ij1}. In particular,

- Given H_{011} or H_{001}, $X_2 = -1 + N_2$.
- Given H_{101} or H_{111}, $X_2 = -2 + N_2$.

This implies

$$P[B_3|H_{011}] = P[B_3|H_{001}] = P[-1 + N_2 > 0] = Q(1/\sigma_N), \tag{1}$$
$$P[B_3|H_{101}] = P[B_3|H_{111}] = P[-2 + N_2 > 0] = Q(2/\sigma_N). \tag{2}$$

Assuming all four hypotheses are equally likely, the probability of an error decoding the third bit is

$$\begin{aligned} P[B_3] &= \frac{P[B_3|H_{011}] + P[B_3|H_{001}] + P[B_3|H_{101}] + P[B_3|H_{111}]}{4} \\ &= \frac{Q(1/\sigma_N) + Q(2/\sigma_N)}{2}. \end{aligned} \tag{3}$$

Problem 11.3.5 Solution

(a) Hypothesis H_i is that $\mathbf{X} = \mathbf{s}_i + \mathbf{N}$, where $\mathbf{N}$ is a Gaussian random vector independent of which signal was transmitted. Thus, given H_i, $\mathbf{X}$ is a Gaussian $(\mathbf{s}_i, \sigma^2\mathbf{I})$ random vector. Since $\mathbf{X}$ is two-dimensional,

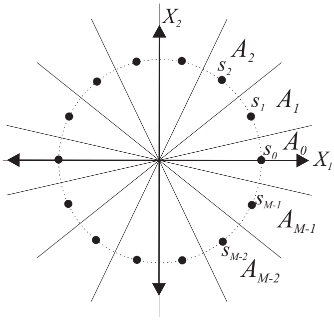
$$f_{\mathbf{X}|H_i}(\mathbf{x}) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{s}_i)'\sigma^2\mathbf{I}^{-1}(\mathbf{x}-\mathbf{s}_i)} = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\|\mathbf{x}-\mathbf{s}_i\|^2}. \tag{1}$$

Since the hypotheses H_i are equally likely, the MAP and ML rules are the same and achieve the minimum probability of error. In this case, from the vector version of Theorem 11.8, the MAP rule is

$$\mathbf{x} \in A_m \text{ if } f_{\mathbf{X}|H_m}(\mathbf{x}) \geq f_{\mathbf{X}|H_j}(\mathbf{x}) \text{ for all } j. \tag{2}$$

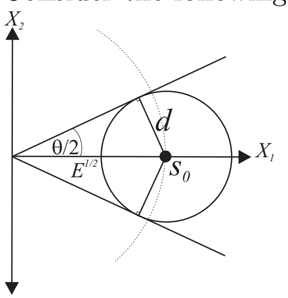
Using the conditional PDFs $f_{\mathbf{X}|H_i}(\mathbf{x})$, the MAP rule becomes

$$\mathbf{x} \in A_m \text{ if } \|\mathbf{x} - \mathbf{s}_m\|^2 \leq \|\mathbf{x} - \mathbf{s}_j\|^2 \text{ for all } j. \tag{3}$$



In terms of geometry, the interpretation is that all vectors $\mathbf{x}$ closer to $\mathbf{s}_m$ than to any other signal $\mathbf{s}_j$ are assigned to A_m. In this problem, the signal constellation (i.e., the set of vectors $\mathbf{s}_i$) is the set of vectors on the circle of radius E. The acceptance regions are the "pie slices" around each signal vector.

(b) Consider the following sketch to determine d.



Geometrically, the largest d such that $\|\mathbf{x} - \mathbf{s}_i\| \leq d$ defines the largest circle around $\mathbf{s}_i$ that can be inscribed into the pie slice A_i. By symmetry, this is the same for every A_i, hence we examine A_0. Each pie slice has angle $\theta = 2\pi/M$. Since the length of each signal vector is $\sqrt{E}$, the sketch shows that $\sin(\theta/2) = d/\sqrt{E}$. Thus $d = \sqrt{E}\sin(\pi/M)$.

(c) By symmetry, P_{ERR} is the same as the conditional probability of error $1 - \mathrm{P}[A_i|H_i]$, no matter which $\mathbf{s}_i$ is transmitted. Let B denote a circle of radius d at the origin and let B_i denote the circle of radius d around $\mathbf{s}_i$. Since $B_0 \subset A_0$,

$$\mathrm{P}[A_0|H_0] = \mathrm{P}[\mathbf{X} \in A_0|H_0] \geq \mathrm{P}[\mathbf{X} \in B_0|H_0] = \mathrm{P}[\mathbf{N} \in B]. \tag{4}$$

Since the components of $\mathbf{N}$ are iid Gaussian $(0, \sigma^2)$ random variables,

$$\begin{aligned} \mathrm{P}[\mathbf{N} \in B] &= \iint_B f_{N_1,N_2}(n_1, n_2)\, dn_1\, dn_2 \\ &= \frac{1}{2\pi\sigma^2} \iint_B e^{-(n_1^2+n_2^2)/2\sigma^2}\, dn_1\, dn_2. \end{aligned} \tag{5}$$

By changing to polar coordinates,

$$\begin{aligned} \mathrm{P}[\mathbf{N} \in B] &= \frac{1}{2\pi\sigma^2} \int_0^d \int_0^{2\pi} e^{-r^2/2\sigma^2} r\, d\theta\, dr \\ &= \frac{1}{\sigma^2} \int_0^d re^{-r^2/2\sigma^2} r\, dr \\ &= -e^{-r^2/2\sigma^2}\Big|_0^d = 1 - e^{-d^2/2\sigma^2} \\ &= 1 - e^{-E\sin^2(\pi/M)/2\sigma^2}. \end{aligned} \tag{6}$$

Thus

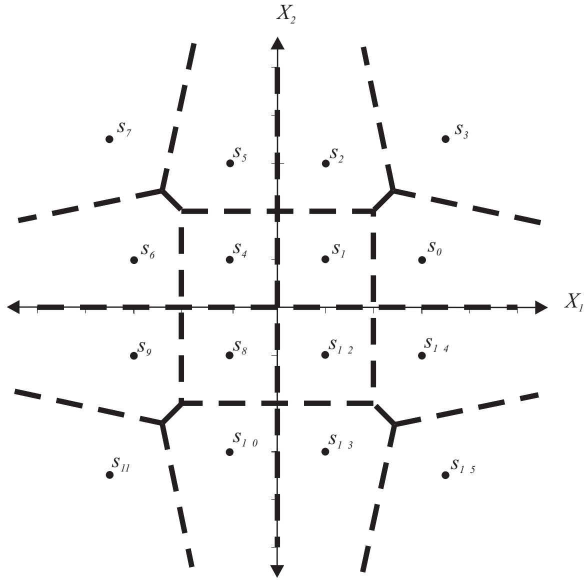
$$P_{\mathrm{ERR}} = 1 - \mathrm{P}[A_0|H_0] \leq 1 - \mathrm{P}[\mathbf{N} \in B] = e^{-E\sin^2(\pi/M)/2\sigma^2}. \tag{7}$$

Problem 11.3.6 Solution

(a) In Problem 11.3.3, we found that in terms of the vectors $\mathbf{x}$ and $\mathbf{s}_i$, the acceptance regions are defined by the rule

$$\mathbf{x} \in A_i \text{ if } \|\mathbf{x} - \mathbf{s}_i\|^2 \leq \|\mathbf{x} - \mathbf{s}_j\|^2 \text{ for all } j. \tag{1}$$

Just as in the case of QPSK, the acceptance region A_i is the set of vectors $\mathbf{x}$ that are closest to $\mathbf{s}_i$. Graphically, these regions are easily found from the sketch of the signal constellation:



(b) For hypothesis A_1, we see that the acceptance region is

$$A_1 = \{(X_1, X_2)|0 < X_1 \leq 2, 0 < X_2 \leq 2\} \tag{2}$$

Given H_1, a correct decision is made if $(X_1, X_2) \in A_1$. Given H_1, $X_1 = 1+N_1$ and $X_2 = 1 + N_2$. Thus,

$$\begin{aligned} \mathrm{P}[C|H_1] &= \mathrm{P}[(X_1, X_2) \in A_1|H_1] \\ &= \mathrm{P}[0 < 1 + N_1 \leq 2, 0 < 1 + N_2 \leq 2] \\ &= (\mathrm{P}[-1 < N_1 \leq 1])^2 \\ &= (\Phi(1/\sigma_N) - \Phi(-1/\sigma_N))^2 \\ &= (2\Phi(1/\sigma_N) - 1)^2 . \end{aligned} \tag{3}$$

(c) Surrounding each signal s_i is an acceptance region A_i that is no smaller than the acceptance region A_1. That is,

$$\begin{aligned}
\mathrm{P}\left[C|H_i\right] &= \mathrm{P}\left[(X_1, X_2) \in A_i|H_1\right] \\
&\geq \mathrm{P}\left[-1 < N_1 \leq 1, -1 < N_2 \leq 1\right] \\
&= \left(\mathrm{P}\left[-1 < N_1 \leq 1\right]\right)^2 = \mathrm{P}\left[C|H_1\right].
\end{aligned} \tag{4}$$

This implies

$$\begin{aligned}
\mathrm{P}\left[C\right] &= \sum_{i=0}^{15} \mathrm{P}\left[C|H_i\right]\mathrm{P}\left[H_1\right] \\
&\geq \sum_{i=0}^{15} \mathrm{P}\left[C|H_1\right]\mathrm{P}\left[H_i\right] \\
&= \mathrm{P}\left[C|H_1\right]\sum_{i=0}^{15} \mathrm{P}\left[H_i\right] = \mathrm{P}\left[C|H_1\right].
\end{aligned} \tag{5}$$

Problem 11.3.7 Solution

Let $p_i = \mathrm{P}[H_i]$. From Theorem 11.8, the MAP multiple hypothesis test is

$$(x_1, x_2) \in A_i \text{ if } p_i f_{X_1,X_2|H_i}(x_1, x_2) \geq p_j f_{X_1,X_2|H_j}(x_1, x_2) \text{ for all } j. \tag{1}$$

From Example 11.13, the conditional PDF of X_1, X_2 given H_i is

$$f_{X_1,X_2|H_i}(x_1, x_2) = \frac{1}{2\pi\sigma^2} e^{-[(x_1-\sqrt{E}\cos\theta_i)^2+(x_2-\sqrt{E}\sin\theta_i)^2]/2\sigma^2}. \tag{2}$$

Using this conditional joint PDF, the MAP rule becomes

- $(x_1, x_2) \in A_i$ if for all j,

$$\begin{aligned}
&-\frac{(x_1-\sqrt{E}\cos\theta_i)^2+(x_2-\sqrt{E}\sin\theta_i)^2}{2\sigma^2} \\
&\qquad+\frac{(x_1-\sqrt{E}\cos\theta_j)^2+(x_2-\sqrt{E}\sin\theta_j)^2}{2\sigma^2} \geq \ln\frac{p_j}{p_i}.
\end{aligned} \tag{3}$$

Expanding the squares and using the identity $\cos^2\theta+\sin^2\theta = 1$ yields the simplified rule

- $(x_1, x_2) \in A_i$ if for all j,

$$x_1[\cos\theta_i - \cos\theta_j] + x_2[\sin\theta_i - \sin\theta_j] \geq \frac{\sigma^2}{\sqrt{E}} \ln \frac{p_j}{p_i}. \tag{4}$$

Note that the MAP rules define linear constraints in x_1 and x_2. Since $\theta_i = \pi/4 + i\pi/2$, we use the following table to enumerate the constraints:

	$\cos\theta_i$	$\sin\theta_i$
$i = 0$	$1/\sqrt{2}$	$1/\sqrt{2}$
$i = 1$	$-1/\sqrt{2}$	$1/\sqrt{2}$
$i = 2$	$-1/\sqrt{2}$	$-1/\sqrt{2}$
$i = 3$	$1/\sqrt{2}$	$-1/\sqrt{2}$

(5)

To be explicit, to determine whether $(x_1, x_2) \in A_i$, we need to check the MAP rule for each $j \neq i$. Thus, each A_i is defined by three constraints. Using the above table, the acceptance regions are

- $(x_1, x_2) \in A_0$ if

$$x_1 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_1}{p_0}, \quad x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_3}{p_0}, \quad x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_0}. \tag{6}$$

- $(x_1, x_2) \in A_1$ if

$$x_1 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_1}{p_0}, \quad x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_1}, \quad -x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_3}{p_1}. \tag{7}$$

- $(x_1, x_2) \in A_2$ if

$$x_1 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_3}, \quad x_2 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_1}, \quad x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_0}. \tag{8}$$

- $(x_1, x_2) \in A_3$ if

$$x_1 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_3}, \quad x_2 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_3}{p_0}, \quad -x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_3}. \tag{9}$$

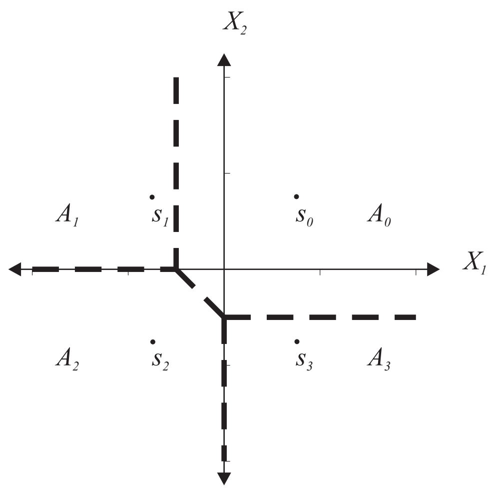
Using the parameters

$$\sigma = 0.8, \quad E = 1, \quad p_0 = 1/2, \quad p_1 = p_2 = p_3 = 1/6, \tag{10}$$

the acceptance regions for the MAP rule are

$$A_0 = \{(x_1, x_2) | x_1 \geq -0.497, x_2 \geq -0.497, x_1 + x_2 \geq -0.497\}, \tag{11}$$
$$A_1 = \{(x_1, x_2) | x_1 \leq -0.497, x_2 \geq 0, -x_1 + x_2 \geq 0\}, \tag{12}$$
$$A_2 = \{(x_1, x_2) | x_1 \leq 0, x_2 \leq 0, x_1 + x_2 \geq -0.497\}, \tag{13}$$
$$A_3 = \{(x_1, x_2) | x_1 \geq 0, x_2 \leq -0.497, -x_1 + x_2 \geq 0\}. \tag{14}$$

Here is a sketch of these acceptance regions:



Note that the boundary between A_1 and A_3 defined by $-x_1 + x_2 \geq 0$ plays no role because of the high value of p_0.

Problem 11.3.8 Solution

(a) First we note that

$$\mathbf{P}^{1/2}\mathbf{X} = \begin{bmatrix} \sqrt{p_1} & & \\ & \ddots & \\ & & \sqrt{p_k} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} = \begin{bmatrix} \sqrt{p_1}X_1 \\ \vdots \\ \sqrt{p_k}X_k \end{bmatrix}. \tag{1}$$

Since each $\mathbf{S}_i$ is a column vector,

$$\mathbf{SP}^{1/2}\mathbf{X} = \begin{bmatrix}\mathbf{S}_1 & \cdots & \mathbf{S}_k\end{bmatrix}\begin{bmatrix}\sqrt{p_1}X_1 \\ \vdots \\ \sqrt{p_k}X_k\end{bmatrix}$$
$$= \sqrt{p_1}X_1\mathbf{S}_1 + \cdots + \sqrt{p_k}X_k\mathbf{S}_k. \tag{2}$$

Thus $\mathbf{Y} = \mathbf{SP}^{1/2}\mathbf{X} + \mathbf{N} = \sum_{i=1}^{k}\sqrt{p_i}X_i\mathbf{S}_i + \mathbf{N}$.

(b) Given the observation $\mathbf{Y} = \mathbf{y}$, a detector must decide which vector $\mathbf{X} = \begin{bmatrix}X_1 & \cdots & X_k\end{bmatrix}'$ was (collectively) sent by the k transmitters. A hypothesis H_j must specify whether $X_i = 1$ or $X_i = -1$ for each i. That is, a hypothesis H_j corresponds to a vector $\mathbf{x}_j \in B_k$ which has ± 1 components. Since there are 2^k such vectors, there are 2^k hypotheses which we can enumerate as $H_1, \ldots, H_{2^k}$. Since each X_i is independently and equally likely to be ± 1, each hypothesis has probability 2^{-k}. In this case, the MAP and and ML rules are the same and achieve minimum probability of error. The MAP/ML rule is

$$\mathbf{y} \in A_m \text{ if } f_{\mathbf{Y}|H_m}(\mathbf{y}) \geq f_{\mathbf{Y}|H_j}(\mathbf{y}) \text{ for all } j. \tag{3}$$

Under hypothesis H_j, $\mathbf{Y} = \mathbf{SP}^{1/2}\mathbf{x}_j + \mathbf{N}$ is a Gaussian $(\mathbf{SP}^{1/2}\mathbf{x}_j, \sigma^2\mathbf{I})$ random vector. The conditional PDF of $\mathbf{Y}$ is

$$f_{\mathbf{Y}|H_j}(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{SP}^{1/2}\mathbf{x}_j)'(\sigma^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{SP}^{1/2}\mathbf{x}_j)}$$
$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\left\|\mathbf{y}-\mathbf{SP}^{1/2}\mathbf{x}_j\right\|^2/2\sigma^2}. \tag{4}$$

The MAP rule is

$$\mathbf{y} \in A_m \text{ if } e^{-\left\|\mathbf{y}-\mathbf{SP}^{1/2}\mathbf{x}_m\right\|^2/2\sigma^2} \geq e^{-\left\|\mathbf{y}-\mathbf{SP}^{1/2}\mathbf{x}_j\right\|^2/2\sigma^2} \text{ for all } j, \tag{5}$$

or equivalently,

$$\mathbf{y} \in A_m \text{ if } \left\|\mathbf{y} - \mathbf{SP}^{1/2}\mathbf{x}_m\right\| \leq \left\|\mathbf{y} - \mathbf{SP}^{1/2}\mathbf{x}_j\right\| \text{ for all } j. \tag{6}$$

That is, we choose the vector $\mathbf{x}^* = \mathbf{x}_m$ that minimizes the distance $\left\|\mathbf{y} - \mathbf{S}\mathbf{P}^{1/2}\mathbf{x}_j\right\|$ among all vectors $\mathbf{x}_j \in B_k$. Since this vector $\mathbf{x}^*$ is a function of the observation $\mathbf{y}$, this is described by the math notation

$$\mathbf{x}^*(\mathbf{y}) = \arg\min_{\mathbf{x}\in B_k} \left\|\mathbf{y} - \mathbf{S}\mathbf{P}^{1/2}\mathbf{x}\right\|, \tag{7}$$

where $\arg\min_{\mathbf{x}} g(\mathbf{x})$ returns the argument $\mathbf{x}$ that minimizes $g(\mathbf{x})$.

(c) To implement this detector, we must evaluate $\left\|\mathbf{y} - \mathbf{S}\mathbf{P}^{1/2}\mathbf{x}\right\|$ for each $\mathbf{x} \in B_k$. Since there 2^k vectors in B_k, we have to evaluate 2^k hypotheses. Because the number of hypotheses grows exponentially with the number of users k, the maximum likelihood detector is considered to be computationally intractable for a large number of users k.

Problem 11.3.9 Solution

A short answer is that the decorrelator cannot be the same as the optimal maximum likelihood (ML) detector. If they were the same, that means we have reduced the 2^k comparisons of the optimal detector to a linear transformation followed by k single bit comparisons.

However, as this is not a satisfactory answer, we will build a simple example with $k = 2$ users and precessing gain $n = 2$ to show the difference between the ML detector and the decorrelator. In particular, suppose user 1 transmits with code vector $\mathbf{S}_1 = \begin{bmatrix}1 & 0\end{bmatrix}'$ and user transmits with code vector $\mathbf{S}_2 = \begin{bmatrix}\cos\theta & \sin\theta\end{bmatrix}'$ In addition, we assume that the users powers are $p_1 = p_2 = 1$. In this case, $\mathbf{P} = \mathbf{I}$ and

$$\mathbf{S} = \begin{bmatrix}1 & \cos\theta \\ 0 & \sin\theta\end{bmatrix}. \tag{1}$$

For the ML detector, there are four hypotheses corresponding to each possible transmitted bit of each user. Using H_i to denote the hypothesis that $\mathbf{X} = \mathbf{x}_i$, we have

$$\mathbf{X} = \mathbf{x}_1 = \begin{bmatrix}1 \\ 1\end{bmatrix}, \qquad \mathbf{X} = \mathbf{x}_3 = \begin{bmatrix}-1 \\ 1\end{bmatrix}, \tag{2}$$

$$\mathbf{X} = \mathbf{x}_2 = \begin{bmatrix}1 \\ -1\end{bmatrix}, \qquad \mathbf{X} = \mathbf{x}_4 = \begin{bmatrix}-1 \\ -1\end{bmatrix}. \tag{3}$$

When $\mathbf{X} = \mathbf{x}_i$, $\mathbf{Y} = \mathbf{y}_i + \mathbf{N}$ where $\mathbf{y}_i = \mathbf{S}\mathbf{x}_i$. Thus under hypothesis H_i, $\mathbf{Y} = \mathbf{y}_i + \mathbf{N}$ is a Gaussian $(\mathbf{y}_i, \sigma^2\mathbf{I})$ random vector with PDF

$$f_{\mathbf{Y}|H_i}(\mathbf{y}) = \frac{1}{2\pi\sigma^2} e^{-(\mathbf{y}-\mathbf{y}_i)'(\sigma^2\mathbf{I})^{-1}(\mathbf{y}-\mathbf{y}_i)/2} = \frac{1}{2\pi\sigma^2} e^{-\|\mathbf{y}-\mathbf{y}_i\|^2/2\sigma^2}. \tag{4}$$

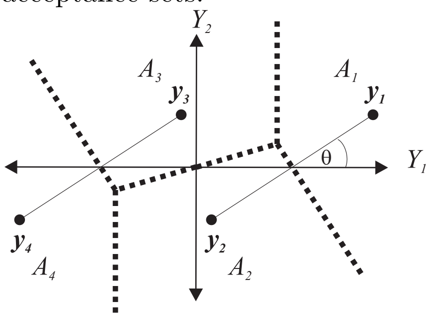
With the four hypotheses equally likely, the MAP and ML detectors are the same and minimize the probability of error. From Theorem 11.8, this decision rule is

$$\mathbf{y} \in A_m \text{ if } f_{\mathbf{Y}|H_m}(\mathbf{y}) \geq f_{\mathbf{Y}|H_j}(\mathbf{y}) \text{ for all } j. \tag{5}$$

This rule simplifies to

$$\mathbf{y} \in A_m \text{ if } \|\mathbf{y} - \mathbf{y}_m\| \leq \|\mathbf{y} - \mathbf{y}_j\| \text{ for all } j. \tag{6}$$

It is useful to show these acceptance sets graphically. In this plot, the area around $\mathbf{y}_i$ is the acceptance set A_i and the dashed lines are the boundaries between the acceptance sets.



$$\mathbf{y}_1 = \begin{bmatrix} 1 + \cos\theta \\ \sin\theta \end{bmatrix}, \quad \mathbf{y}_3 = \begin{bmatrix} -1 + \cos\theta \\ \sin\theta \end{bmatrix},$$
$$\mathbf{y}_2 = \begin{bmatrix} 1 - \cos\theta \\ -\sin\theta \end{bmatrix}, \quad \mathbf{y}_4 = \begin{bmatrix} -1 - \cos\theta \\ -\sin\theta \end{bmatrix}.$$

The probability of a correct decision is

$$\mathrm{P}[C] = \frac{1}{4} \sum_{i=1}^{4} \int_{A_i} f_{\mathbf{Y}|H_i}(\mathbf{y})\, d\mathbf{y}. \tag{7}$$

Even though the components of $\mathbf{Y}$ are conditionally independent given H_i, the four integrals $\int_{A_i} f_{\mathbf{Y}|H_i}(\mathbf{y})\, d\mathbf{y}$ cannot be represented in a simple form. Moreoever, they cannot even be represented by the $\Phi(\cdot)$ function. Note that the probability of a correct decision is the probability that the bits X_1 and X_2 transmitted by both users are detected correctly.

The probability of a bit error is still somewhat more complex. For example if $X_1 = 1$, then hypotheses H_1 and H_3 are equally likely. The detector guesses $\hat{X}_1 = 1$ if $\mathbf{Y} \in A_1 \cup A_3$. Given $X_1 = 1$, the conditional probability of a correct decision on this bit is

$$\begin{aligned} \mathrm{P}\left[\hat{X}_1 = 1 | X_1 = 1\right] &= \frac{1}{2}\mathrm{P}\left[\mathbf{Y} \in A_1 \cup A_3 | H_1\right] + \frac{1}{2}\mathrm{P}\left[\mathbf{Y} \in A_1 \cup A_3 | H_3\right] \\ &= \frac{1}{2}\int_{A_1 \cup A_3} f_{\mathbf{Y}|H_1}(\mathbf{y})\, d\mathbf{y} + \frac{1}{2}\int_{A_1 \cup A_3} f_{\mathbf{Y}|H_3}(\mathbf{y})\, d\mathbf{y}. \end{aligned} \tag{8}$$

By comparison, the decorrelator does something simpler. Since $\mathbf{S}$ is a square invertible matrix,

$$(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' = \mathbf{S}^{-1}(\mathbf{S}')^{-1}\mathbf{S}' = \mathbf{S}^{-1} = \frac{1}{\sin\theta}\begin{bmatrix} 1 & -\cos\theta \\ 0 & 1 \end{bmatrix}. \tag{9}$$

We see that the components of $\tilde{\mathbf{Y}} = \mathbf{S}^{-1}\mathbf{Y}$ are

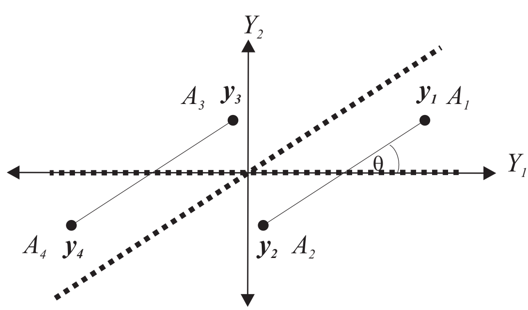
$$\tilde{Y}_1 = Y_1 - \frac{\cos\theta}{\sin\theta}Y_2, \qquad \tilde{Y}_2 = \frac{Y_2}{\sin\theta}. \tag{10}$$

Assuming (as in earlier sketch) that $0 < \theta < \pi/2$, the decorrelator bit decisions are

$$\hat{X}_1 = \operatorname{sgn}(\tilde{Y}_1) = \operatorname{sgn}\left(Y_1 - \frac{\cos\theta}{\sin\theta}Y_2\right), \tag{11}$$

$$\hat{X}_2 = \operatorname{sgn}(\tilde{Y}_2) = \operatorname{sgn}\left(\frac{Y_2}{\sin\theta}\right) = \operatorname{sgn}(Y_2). \tag{12}$$

Graphically, these regions are:



Because we chose a coordinate system such that $\mathbf{S}_1$ lies along the x-axis, the effect of the decorrelator on the rule for bit X_2 is particularly easy to understand. For bit X_2, we just check whether the vector $\mathbf{Y}$ is in the upper half plane. Generally, the boundaries of the decorrelator decision regions are determined by straight lines, they are easy to implement and probability of error is easy to calculate. However, these regions are suboptimal in terms of probability of error.

Problem 11.4.1 Solution

Under hypothesis H_i, the conditional PMF of X is

$$P_{X|H_i}(x) = \begin{cases} (1-p_i)p_i^{x-1}/(1-p_i^{20}) & x = 1, 2, \ldots, 20, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where $p_0 = 0.99$ and $p_1 = 0.9$. It follows that for $x_0 = 0, 1, \ldots, 19$ that

$$\begin{aligned} \mathrm{P}\left[X > x_0 | H_i\right] &= \frac{1-p_i}{1-p_i^{20}} \sum_{x=x_0+1}^{20} p_i^{x-1} \\ &= \frac{1-p_i}{1-p_i^{20}} \left[p_i^{x_0} + \cdots + p_i^{19}\right] \\ &= \frac{p_i^{x_0}(1-p_i)}{1-p_i^{20}} \left[1 + p_i + \cdots + p_i^{19-x_0}\right] \\ &= \frac{p_i^{x_0}(1-p_i^{20-x_0})}{1-p_i^{20}} = \frac{p_i^{x_0} - p_i^{20}}{1-p_i^{20}}. \end{aligned} \tag{2}$$

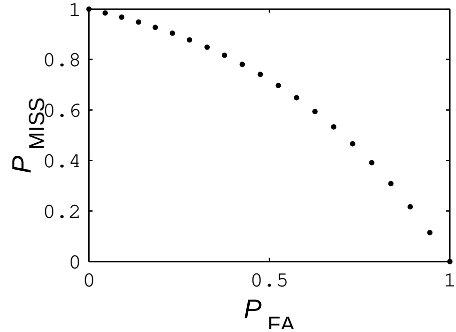
We note that the above formula is also correct for $x_0 = 20$. Using this formula, the false alarm and miss probabilities are

$$P_{\text{FA}} = \mathrm{P}\left[X > x_0 | H_0\right] = \frac{p_0^{x_0} - p_0^{20}}{1-p_0^{20}}, \tag{3}$$

$$P_{\text{MISS}} = 1 - \mathrm{P}\left[X > x_0 | H_1\right] = \frac{1-p_1^{x_0}}{1-p_1^{20}}. \tag{4}$$

The MATLAB program `rocdisc(p0,p1)` returns the false alarm and miss probabilities and also plots the ROC. Here is the program and the output for `rocdisc(0.9,0.99`

```
function [PFA,PMISS]=rocdisc(p0,p1);
x=0:20;
PFA= (p0.^x-p0^(20))/(1-p0^(20));
PMISS= (1.0-(p1.^x))/(1-p1^(20));
plot(PFA,PMISS,'k.');
xlabel('\itP_{\rm FA}');
ylabel('\itP_{\rm MISS}');
```



From the receiver operating curve, we learn that we have a fairly lousy sensor. No matter how we set the threshold x_0, either the false alarm probability or the miss probability (or both!) exceed 0.5.

Problem 11.4.2 Solution

From Example 11.7, the probability of error is

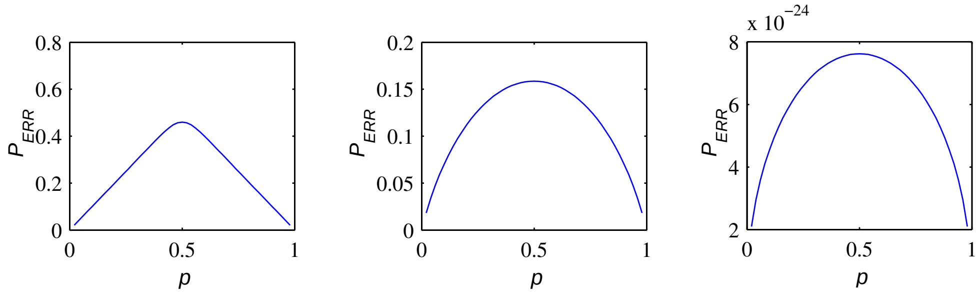
$$P_{\mathrm{ERR}} = pQ\left(\frac{\sigma}{2v}\ln\frac{p}{1-p}+\frac{v}{\sigma}\right)+(1-p)\Phi\left(\frac{\sigma}{2v}\ln\frac{p}{1-p}-\frac{v}{\sigma}\right). \tag{1}$$

It is straightforward to use MATLAB to plot P_{ERR} as a function of p. The function `bperr` calculates P_{ERR} for a vector `p` and a scalar signal to noise ratio `snr` corresponding to v/σ. A second program `bperrplot(snr)` plots P_{ERR} as a function of p. Here are the programs

```
function perr=bperr(p,snr);
%Problem 8.4.2 Solution
r=log(p./(1-p))/(2*snr);
perr=(p.*(qfunction(r+snr))) ...
    +((1-p).*phi(r-snr));
```

```
function pe=bperrplot(snr);
p=0.02:0.02:0.98;
pe=bperr(p,snr);
plot(p,pe);
xlabel('\it p');
ylabel('\it P_{ERR}');
```

Here are three outputs of `bperrplot` for the requested SNR values.



`bperrplot(0.1)` `bperrplot(0.1)` `bperrplot(0.1)`

In all three cases, we see that P_{ERR} is *maximum* at $p = 1/2$. When $p \neq 1/2$, the optimal (minimum probability of error) decision rule is able to exploit the one hypothesis having higher a priori probability than the other.

This gives the wrong impression that one should consider building a communication system with $p \neq 1/2$. To see this, consider the most extreme case in which the error probability goes to zero as $p \to 0$ or $p \to 1$. However, in these extreme cases, no information is being communicated. When $p = 0$ or $p = 1$, the detector can simply guess the transmitted bit. In fact, there is no need to tansmit a bit; however, it becomes impossible to transmit any information.

Finally, we note that v/σ is an SNR voltage ratio. For communication systems, it is common to measure SNR as a power ratio. In particular, $v/\sigma = 10$ corresponds to a SNR of $10\log_1 0(v^2/\sigma^2) = 20$ dB.

Problem 11.4.3 Solution

With $v = 1.5$ and $d = 0.5$, it appeared in Example 11.14 that $T = 0.5$ was best among the values tested. However, it also seemed likely the error probability P_e would decrease for larger values of T. To test this possibility we use sqdistor with 100,000 transmitted bits by trying the following:

```
>> T=[0.4:0.1:1.0];Pe=sqdistor(1.5,0.5,100000,T);
>> [Pmin,Imin]=min(Pe);T(Imin)
ans =
   0.80000000000000
```

Thus among $\{0.4, 0.5, \cdots, 1.0\}$, it appears that $T = 0.8$ is best. Now we test values of T in the neighborhood of 0.8:

```
>> T=[0.70:0.02:0.9];Pe=sqdistor(1.5,0.5,100000,T);
>>[Pmin,Imin]=min(Pe);T(Imin)
ans =
   0.78000000000000
```

This suggests that $T = 0.78$ is best among these values. However, inspection of the vector `Pe` shows that all values are quite close. If we repeat this experiment a few times, we obtain:

```
>> T=[0.70:0.02:0.9];Pe=sqdistor(1.5,0.5,100000,T);
>> [Pmin,Imin]=min(Pe);T(Imin)
ans =
   0.78000000000000
>> T=[0.70:0.02:0.9];Pe=sqdistor(1.5,0.5,100000,T);
>> [Pmin,Imin]=min(Pe);T(Imin)
ans =
   0.80000000000000
>> T=[0.70:0.02:0.9];Pe=sqdistor(1.5,0.5,100000,T);
>> [Pmin,Imin]=min(Pe);T(Imin)
ans =
   0.76000000000000
>> T=[0.70:0.02:0.9];Pe=sqdistor(1.5,0.5,100000,T);
>> [Pmin,Imin]=min(Pe);T(Imin)
ans =
   0.78000000000000
```

This suggests that the best value of T is in the neighborhood of 0.78. If someone were paying you to find the best T, you would probably want to do more testing. The only useful lesson here is that when you try to optimize parameters using simulation results, you should repeat your experiments to get a sense of the variance of your results.

Problem 11.4.4 Solution

Since the a priori probabilities $P[H_0]$ and $P[H_1]$ are unknown, we use a Neyamn-Pearson formulation to find the ROC. For a threshold γ, the decision rule is

$$x \in A_0 \text{ if } \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \geq \gamma; \qquad x \in A_1 \text{ otherwise.} \tag{1}$$

Using the given conditional PDFs, we obtain

$$x \in A_0 \text{ if } e^{-(8x-x^2)/16} \geq \gamma x/4; \qquad x \in A_1 \text{ otherwise.} \tag{2}$$

Taking logarithms yields

$$x \in A_0 \text{ if } x^2 - 8x \geq 16\ln(\gamma/4) + 16\ln x; \qquad x \in A_1 \text{ otherwise.} \tag{3}$$

With some more rearranging,

$$x \in A_0 \text{ if } (x-4)^2 \geq \underbrace{16\ln(\gamma/4) + 16}_{\gamma_0} + 16\ln x; \qquad x \in A_1 \text{ otherwise.} \tag{4}$$

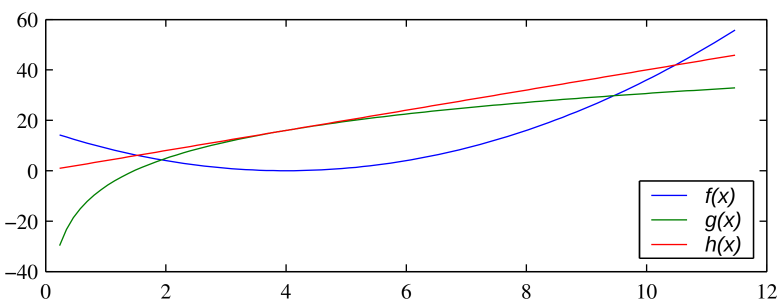
When we plot the functions $f(x) = (x-4)^2$ and $g(x) = \gamma_0 + 16\ln x$, we see that there exist x_1 and x_2 such that $f(x_1) = g(x_1)$ and $f(x_2) = g(x_2)$. In terms of x_1 and x_2,

$$A_0 = [0, x_1] \cup [x_2, \infty), \qquad\qquad A_1 = (x_1, x_2). \tag{5}$$

Using a Taylor series expansion of $\ln x$ around $x = x_0 = 4$, we can show that

$$g(x) = \gamma_0 + 16\ln x \leq h(x) = \gamma_0 + 16(\ln 4 - 1) + 4x. \tag{6}$$

Since $h(x)$ is linear, we can use the quadratic formula to solve $f(x) = h(x)$, yielding a solution $\bar{x}_2 = 6 + \sqrt{4 + 16\ln 4 + \gamma_0}$. One can show that $x_2 \leq \bar{x}_2$. In the example shown below corresponding to $\gamma = 1$ shown here, $x_1 = 1.95$, $x_2 = 9.5$ and $\bar{x}_2 = 6 + \sqrt{20} = 10.47$.



Given x_1 and x_2, the false alarm and miss probabilities are

$$P_{\text{FA}} = \text{P}\left[A_1|H_0\right] = \int_{x_1}^{2} \frac{1}{2} e^{-x/2}\,dx = e^{-x_1/2} - e^{-x_2/2}, \tag{7}$$

$$\begin{aligned} P_{\text{MISS}} &= 1 - \text{P}\left[A_1|H_1\right] = 1 - \int_{x_1}^{x_2} \frac{x}{8} e^{-x^2/16}\,dx \\ &= 1 - e^{-x_1^2/16} + e^{-x_2^2/16}. \end{aligned} \tag{8}$$

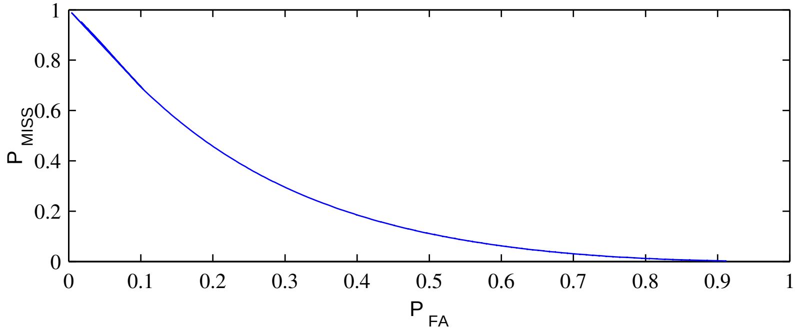
To calculate the ROC, we need to find x_1 and x_2. Rather than find them exactly, we calculate $f(x)$ and $g(x)$ for discrete steps over the interval $[0, 1 + \bar{x}_2]$ and find the discrete values closest to x_1 and x_2. However, for these approximations to x_1 and x_2, we calculate the exact false alarm and miss probabilities. As a result, the optimal detector using the exact x_1 and x_2 cannot be worse than the ROC that we calculate.

In terms of Matlab, we divide the work into the functions `gasroc(n)` which generates the ROC by calling `[x1,x2]=gasrange(gamma)` to calculate x_1 and x_2 for a given value of γ.

```
function [pfa,pmiss]=gasroc(n);
a=(400)^(1/(n-1));
k=1:n;
g=0.05*(a.^(k-1));
pfa=zeros(n,1);
pmiss=zeros(n,1);
for k=1:n,
  [x1,x2]=gasrange(g(k));
  pmiss(k)=1-(exp(-x1^2/16)...
    -exp(-x2^2/16));
  pfa(k)=exp(-x1/2)-exp(-x2/2);
end
plot(pfa,pmiss);
ylabel('P_{\rm MISS}');
xlabel('P_{\rm FA}');
```

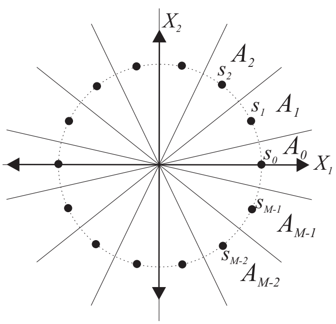
```
function [x1,x2]=gasrange(gamma);
g=16+16*log(gamma/4);
xmax=7+ ...
  sqrt(max(0,4+(16*log(4))+g));
dx=xmax/500;
x=dx:dx:4;
y=(x-4).^2-g-16*log(x);
[ym,i]=min(abs(y));
x1=x(i);
x=4:dx:xmax;
y=(x-4).^2-g-16*log(x);
[ym,i]=min(abs(y));
x2=x(i);
```

The argment `n` of `gasroc(n)` generates the ROC for `n` values of γ, ranging from from 1/20 to 20 in multiplicative steps. Here is the resulting ROC:



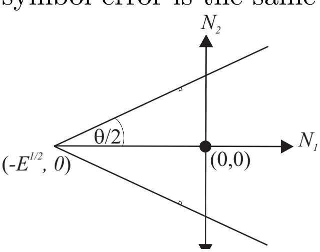
After all of this work, we see that the sensor is not particularly good since no matter how we choose the thresholds, we cannot reduce both the miss and false alarm probabilities under 30 percent.

Problem 11.4.5 Solution



In the solution to Problem 11.3.5, we found that the signal constellation and acceptance regions shown in the adjacent figure. We could solve this problem by a general simulation of an M-PSK system. This would include a random sequence of data sysmbols, mapping symbol i to vector $\mathbf{s}_i$, adding the noise vector $\mathbf{N}$ to produce the receiver output $\mathbf{X} = \mathbf{s}_i + \mathbf{N}$.

However, we are only asked to find the probability of symbol error, but not the probability that symbol i is decoded as symbol j at the receiver. Because of the symmetry of the signal constellation and the acceptance regions, the probability of symbol error is the same no matter what symbol is transmitted.



Thus it is simpler to assume that $\mathbf{s}_0$ is transmitted every time and check that the noise vector $\mathbf{N}$ is in the pie slice around $\mathbf{s}_0$. In fact by translating $\mathbf{s}_0$ to the origin, we obtain the "pie slice" geometry shown in the figure. Because the lines marking the boundaries of the pie slice have slopes $\pm \tan \theta/2$.

The pie slice region is given by the constraints

$$N_2 \leq \tan(\theta/2)\left[N_1 + \sqrt{E}\right], \qquad N_2 \geq -\tan(\theta/2)\left[N_1 + \sqrt{E}\right]. \tag{1}$$

We can rearrange these inequalities to express them in vector form as

$$\begin{bmatrix} -\tan\theta/2 & 1 \\ -\tan\theta/2 & -1 \end{bmatrix} \begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix} \sqrt{E}\tan\theta/2. \tag{2}$$

Finally, since each N_i has variance σ^2, we define the Gaussian $(\mathbf{0}, \mathbf{I})$ random vector $\mathbf{Z} = \mathbf{N}/\sigma$ and write our constraints as

$$\begin{bmatrix} -\tan\theta/2 & 1 \\ -\tan\theta/2 & -1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix} \sqrt{\gamma}\tan\theta/2, \tag{3}$$

where $\gamma = E/\sigma^2$ is the signal to noise ratio of the system.

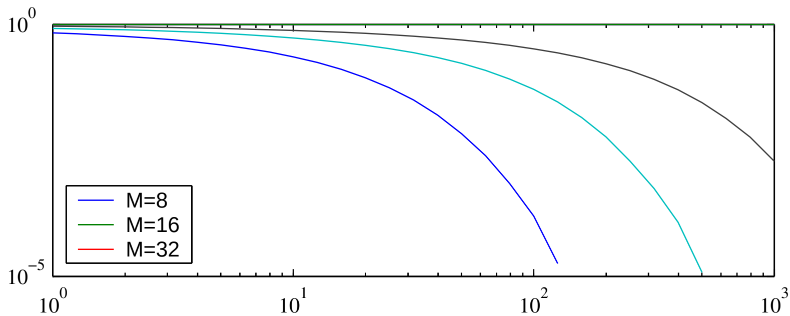
The MATLAB "simulation" simply generates many pairs $\begin{bmatrix} Z_1 & Z_2 \end{bmatrix}'$ and checks what fraction meets these constraints. the function `mpsksim(M,snr,n)` simulates the M-PSK system with SNR `snr` for n bit transmissions. The script `mpsktest` graphs the symbol error probability for $M = 8, 16, 32$.

```
function Pe=mpsksim(M,snr,n);
%Problem 8.4.5 Solution:
%Pe=mpsksim(M,snr,n)
%n bit M-PSK simulation
t=tan(pi/M);
A =[-t 1; -t -1];
Z=randn(2,n);
PC=zeros(length(snr));
for k=1:length(snr),
    B=(A*Z)<=t*sqrt(snr(k));
    PC(k)=sum(min(B))/n;
end
Pe=1-PC;
```

```
%mpsktest.m;
snr=10.^((0:30)/10);
n=500000;
Pe8=mpsksim(8,snr,n);
Pe16=mpsksim(16,snr,n);
Pe32=mpsksim(32,snr,n);
loglog(snr,Pe8,snr,Pe16,snr,Pe32);
legend('M=8','M=16','M=32',3);
```

In `mpsksim`, each column of the matrix `Z` corresponds to a pair of noise variables $\begin{bmatrix} Z_1 & Z_2 \end{bmatrix}'$. The code `B=(A*Z)<=t*sqrt(snr(k))` checks whether each pair of noise variables is in the pie slice region. That is, `B(1,j)` and `B(2,j)` indicate if the ith pair meets the first and second constraints. Since `min(B)` operates on each column of `B`, `min(B)` is a row vector indicating which pairs of noise variables passed the test.

Here is the output of `mpsktest`:



The curves for $M = 8$ and $M = 16$ end prematurely because for high SNR, the error rate is so low that no errors are generated in 500,000 symbols. In this case, the measured P_e is zero and since $\log 0 = -\infty$, the `loglog` function simply ignores the zero values.

Problem 11.4.6 Solution

When the transmitted bit vector is $\mathbf{X} = \mathbf{x}$, the received signal vector $\mathbf{Y} = \mathbf{SP}^{1/2}\mathbf{x} + \mathbf{N}$ is a Gaussian $(\mathbf{SP}^{1/2}\mathbf{x}, \sigma^2\mathbf{I})$ random vector with conditional PDF

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-\left\|\mathbf{y}-\mathbf{SP}^{1/2}\mathbf{x}\right\|^2/2\sigma^2}. \tag{1}$$

The transmitted data vector $\mathbf{x}$ belongs to the set B_k of all binary ± 1 vectors of length k. In principle, we can enumerate the vectors in B_k as $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{2^k-1}$. Moreover, each possible data vector $\mathbf{x}_m$ represents a hypothesis. Since there are 2^k possible data vectors, there are 2^k acceptance sets A_m. The set A_m is the set of all vectors $\mathbf{y}$ such that the decision rule is to guess $\hat{\mathbf{X}} = \mathbf{x}_m$, Our normal procedure is to write a decision rule as "$\mathbf{y} \in A_m$ if ..." however this problem has so many has so many hypotheses that it is more staightforward to refer to a hypothesis $\mathbf{X} = \mathbf{x}_m$ by the function $\hat{\mathbf{x}}(\mathbf{y})$ which returns the vector $\mathbf{x}_m$ when $\mathbf{y} \in A_m$. In short, $\hat{\mathbf{x}}(\mathbf{y})$ is our best guess as to which vector $\mathbf{x}$ was transmitted when $\mathbf{y}$ is received.

Because each hypotheses has a priori probability 2^{-k}, the probability of error is minimized by the maximum likelihood (ML) rule

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg\max_{\mathbf{x}\in B_k} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}). \tag{2}$$

Keep in mind that $\arg\max_{\mathbf{x}} g(\mathbf{x})$ returns the argument $\mathbf{x}$ that maximizes $g(\mathbf{x})$. In any case, the form of $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ implies that the ML rule should minimize the negative exponent of $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. That is, the ML rule is

$$\begin{aligned}
\hat{\mathbf{x}}(\mathbf{y}) &= \arg\min_{\mathbf{x}\in B_k} \left\|\mathbf{y} - \mathbf{S}\mathbf{P}^{1/2}\mathbf{x}\right\| \\
&= \arg\min_{\mathbf{x}\in B_k} (\mathbf{y} - \mathbf{S}\mathbf{P}^{1/2}\mathbf{x})'(\mathbf{y} - \mathbf{S}\mathbf{P}^{1/2}\mathbf{x}) \\
&= \arg\min_{\mathbf{x}\in B_k} \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{S}\mathbf{P}^{1/2}\mathbf{x} + \mathbf{x}'\mathbf{P}^{1/2}\mathbf{S}'\mathbf{S}\mathbf{P}^{1/2}\mathbf{x}. \qquad (3)
\end{aligned}$$

Since the term $\mathbf{y}'\mathbf{y}$ is the same for every $\mathbf{x}$, we can define the function

$$h(\mathbf{x}) = -2\mathbf{y}'\mathbf{S}\mathbf{P}^{1/2}\mathbf{x} + \mathbf{x}'\mathbf{P}^{1/2}\mathbf{S}'\mathbf{S}\mathbf{P}^{1/2}\mathbf{x}, \tag{4}$$

In this case, the ML rule can be expressed as $\hat{\mathbf{x}}(\mathbf{y}) = \arg\min_{\mathbf{x}\in B_k} h(\mathbf{x})$. We use MATLAB to evaluate $h(\mathbf{x})$ for each $\mathbf{x} \in B_k$. Since for $k = 10$, B_k has $2^{10} = 1024$ vectors, it is desirable to make the calculation as easy as possible. To this end, we define $\mathbf{w} = \mathbf{S}\mathbf{P}^{1/2}\mathbf{x}$ and and we write, with some abuse of notation, $h(\cdot)$ as a function of $\mathbf{w}$:

$$h(\mathbf{w}) = -2\mathbf{y}'\mathbf{w} + \mathbf{w}'\mathbf{w}. \tag{5}$$

Still, given $\mathbf{y}$, we need to evaluate $h(\mathbf{w})$ for each vector $\mathbf{w}$. In MATLAB, this will be convenient because we can form the matrices $\mathbf{X}$ and $\mathbf{W}$ with columns consisting of all possible vectors $\mathbf{x}$ and $\mathbf{w}$. In MATLAB, it is easy to calculate $\mathbf{w}'\mathbf{w}$ by operating on the matrix $\mathbf{W}$ without looping through all columns $\mathbf{w}$.

```
function X=allbinaryseqs(n)
%See Problem 8.4.6
%X: n by 2^n matrix of all
%length n binary vectors
%Thanks to Jasvinder Singh
A=repmat([0:2^n-1],[n,1]);
P=repmat([1:n]',[1,2^n]);
X = bitget(A,P);
X=(2*X)-1;
```

In terms of MATLAB, we start by defining `X=allbinaryseqs(n)` which returns an $n \times 2^n$ matrix `X`, corresponding to $\mathbf{X}$, such that the columns of `X` enumerate all possible binary ± 1 sequences of length n. How `allbinaryseqs` works will be clear by generating the matrices `A` and `P` and reading the help for `bitget`.

```
function S=randomsignals(n,k);
%S is an n by k matrix, columns are
%random unit length signal vectors
S=(rand(n,k)>0.5);
S=((2*S)-1.0)/sqrt(n);
```

Next is a short program that generates k random signals, each of length n. Each random signal is just a binary ± 1 sequence normalized to have length 1.

Next, for a set of signal vectors (spreading sequences in CDMA parlance) given by the $n \times k$ matrix `S`, `err=cdmasim(S,P,m)` simulates the transmission of a frame of `m` symbols through a k user CDMA system with additive Gaussian noise. A "symbol," is just a vector $\mathbf{x}$ corresponding to the k transmitted bits of the k users.

In addition, the function `Pe=rcdma(n,k,snr,s,m)` runs `cdmasim` for the pairs of values of users `k` and SNR `snr`. Here is the pair of functions:

```
function err=cdmasim(S,P,m);
%err=cdmasim(P,S,m);
%S= n x k matrix of signals
%P= diag matrix of SNRs (power
% normalized by noise variance)
%See Problem 8.4.6
k=size(S,2); %number of users
n=size(S,1); %processing gain
X=allbinaryseqs(k);%all data
Phalf=sqrt(P);
W=S*Phalf*X;
WW=sum(W.*W);
err=0;
for j=1:m,
  s=duniformrv(1,2^k,1);
  y=S*Phalf*X(:,s)+randn(n,1);
  [hmin,imin]=min(-2*y'*W+WW);
  err=err+sum(X(:,s)~=X(:,imin));
end
```

```
function Pe=rcdma(n,k,snr,s,m);
%Pe=rcdma(n,k,snr,s,m);
%R-CDMA simulation:
%  proc gain=n, users=k
%  rand signal set/frame
%  s frames, m symbols/frame
%See Problem 8.4.6 Solution
[K,SNR]=ndgrid(k,snr);
Pe=zeros(size(SNR));
for j=1:prod(size(SNR)),
   p=SNR(j);k=K(j);
   e=0;
   for i=1:s,
      S=randomsignals(n,k);
      e=e+cdmasim(S,p*eye(k),m);
   end
   Pe(j)=e/(s*m*k);
%   disp([p k e Pe(j)]);
end
```

In `cdmasim`, the kth diagonal element of `P` is the "power" p_k of user k. Technically, we assume that the additive Gaussian noise variable have variance 1, and thus p_k is actually the signal to noise ratio of user k. In addition, `WW` is a length 2^k row vector, with elements $\mathbf{w}'\mathbf{w}$ for each possible $\mathbf{w}$. For each of the `m` random data symbols, represented by $\mathbf{x}$ (or `X(:,s)` in MATLAB), `cdmasim` calculates a received signal $\mathbf{y}$ (y). Finally, hmin is the minimum $h(\mathbf{w})$ and imin is the index of the column of W that minimizes $h(\mathbf{w})$. Thus imin is also the index of the minimizing column of X. Finally, cdmasim compares $\hat{\mathbf{x}}(\mathbf{y})$ and the transmitted vector $\mathbf{x}$ bit by bit and counts the total number of bit errors.

The function rcdma repeats cdmasim for s frames, with a random signal set for each frame. Dividing the total number of bit errors over s frames by the total number of transmitted bits, we find the bit error rate P_e. For an SNR of 4 dB and processing gain 16, the requested tests are generated with the commands

```
>> n=16;
>> k=[2 4 8 16];
>> Pe=rcdma(n,k,snr,100,1000);
>>Pe
Pe =
   0.0252   0.0272    0.0385    0.0788
>>
```

To answer part (b), the code for the matched filter (MF) detector is much simpler because there is no need to test 2^k hypotheses for every transmitted symbol. Just as for the case of the ML detector, we define a function err=mfcdmasim(S,P,m) that simulates the MF detector for m symbols for a given set of signal vectors S. In mfcdmasim, there is no need for looping. The mth transmitted symbol is represented by the mth column of X and the corresponding received signal is given by the mth column of Y. The matched filter processing can be applied to all m columns at once. A second function Pe=mfrcdma(n,k,snr,s,m) cycles through all combinations of users k and SNR snr and calculates the bit error rate for each pair of values. Here are the functions:

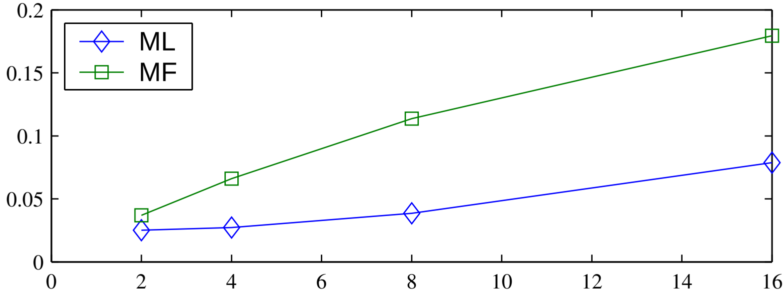
```
function err=mfcdmasim(S,P,m);
%err=mfcdmasim(P,S,m);
%S= n x k matrix of signals
%P= diag matrix of SNRs
% SNR=power/var(noise)
%See Problem 8.4.6b
k=size(S,2); %no. of users
n=size(S,1); %proc. gain
Phalf=sqrt(P);
X=randombinaryseqs(k,m);
Y=S*Phalf*X+randn(n,m);
XR=sign(S'*Y);
err=sum(sum(XR ~= X));
```

```
function Pe=mfrcdma(n,k,snr,s,m);
%Pe=rcdma(n,k,snr,s,m);
%R-CDMA, MF detection
%  proc gain=n, users=k
%  rand signal set/frame
%  s frames, m symbols/frame
%See Problem 8.4.6 Solution
[K,SNR]=ndgrid(k,snr);
Pe=zeros(size(SNR));
for j=1:prod(size(SNR)),
  p=SNR(j);kt=K(j);
  e=0;
  for i=1:s,
    S=randomsignals(n,kt);
    e=e+mfcdmasim(S,p*eye(kt),m);
  end
  Pe(j)=e/(s*m*kt);
  disp([snr kt e]);
end
```

Here is a run of `mfrcdma`.

```
>> pemf=mfrcdma(16,k,4,1000,1000);
     4         2        73936
     4         4        264234
     4         8        908558
     4         16      2871356
>> pemf'
ans =
    0.0370    0.0661    0.1136    0.1795
>>
```

The following plot compares the maximum likelihood (ML) and matched filter (MF) detectors.



As the ML detector offers the minimum probability of error, it should not surprising that it has a lower bit error rate. Although the MF detector is worse, the reduction in detctor complexity makes it attractive. In fact, in practical CDMA-based cellular phones, the processing gain ranges from roughly 128 to 512. In such case, the complexity of the ML detector is prohibitive and thus only matched filter detectors are used.

Problem 11.4.7 Solution

For the CDMA system of Problem 11.3.9, the received signal resulting from the transmissions of k users was given by

$$\mathbf{Y} = \mathbf{S}\mathbf{P}^{1/2}\mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{S}$ is an $n \times k$ matrix with ith column $\mathbf{S}_i$ and $\mathbf{P}^{1/2} = \mathrm{diag}[\sqrt{p_1}, \ldots, \sqrt{p_k}]$ is a $k \times k$ diagonal matrix of received powers, and $\mathbf{N}$ is a Gaussian $(\mathbf{0}, \sigma^2\mathbf{I})$ Gaussian noise vector.

(a) When $\mathbf{S}$ has linearly independent columns, $\mathbf{S}'\mathbf{S}$ is invertible. In this case, the decorrelating detector applies a transformation to $\mathbf{Y}$ to generate

$$\tilde{\mathbf{Y}} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{Y} = \mathbf{P}^{1/2}\mathbf{X} + \tilde{\mathbf{N}}, \tag{2}$$

where $\tilde{\mathbf{N}} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{N}$ is still a Gaussian noise vector with expected value $\mathrm{E}[\tilde{\mathbf{N}}] = \mathbf{0}$. Decorrelation separates the signals in that the ith component of $\tilde{\mathbf{Y}}$ is

$$\tilde{Y}_i = \sqrt{p_i}X_i + \tilde{N}_i. \tag{3}$$

This is the same as a single user-receiver output of the binary communication system of Example 11.6. The single-user decision rule $\hat{X}_i = \operatorname{sgn}(\tilde{Y}_i)$ for the transmitted bit X_i has probability of error

$$P_{e,i} = \mathrm{P}\left[\tilde{Y}_i > 0 | X_i = -1\right]$$
$$= \mathrm{P}\left[-\sqrt{p_i} + \tilde{N}_i > 0\right] = Q\left(\sqrt{\frac{p_i}{\mathrm{Var}[\tilde{N}_i]}}\right). \tag{4}$$

However, since $\tilde{\mathbf{N}} = \mathbf{AN}$ where $\mathbf{A} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$, Theorem 8.11 tells us that $\tilde{\mathbf{N}}$ has covariance matrix $\mathbf{C}_{\tilde{\mathbf{N}}} = \mathbf{A}\mathbf{C}_{\mathbf{N}}\mathbf{A}'$. We note that the general property that $(\mathbf{B}^{-1})' = (\mathbf{B}')^{-1}$ implies that $\mathbf{A}' = \mathbf{S}((\mathbf{S}'\mathbf{S})')^{-1} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}$. These facts imply

$$\mathbf{C}_{\tilde{\mathbf{N}}} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\sigma^2\mathbf{I})\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1} = \sigma^2(\mathbf{S}'\mathbf{S})^{-1}. \tag{5}$$

Note that $\mathbf{S}'\mathbf{S}$ is called the correlation matrix since its i,jth entry is $\mathbf{S}_i'\mathbf{S}_j$ is the correlation between the signal of user i and that of user j. Thus $\mathrm{Var}[\tilde{N}_i] = \sigma^2(\mathbf{S}'\mathbf{S})^{-1}_{ii}$ and the probability of bit error for user i is for user i is

$$P_{e,i} = Q\left(\sqrt{\frac{p_i}{\mathrm{Var}[\tilde{N}_i]}}\right) = Q\left(\sqrt{\frac{p_i}{(\mathbf{S}'\mathbf{S})^{-1}_{ii}}}\right). \tag{6}$$

To find the probability of error for a randomly chosen but, we average over the bits of all users and find that

$$P_e = \frac{1}{k}\sum_{i=1}^{k} P_{e,i} = \frac{1}{k}\sum_{i=1}^{k} Q\left(\sqrt{\frac{p_i}{(\mathbf{S}'\mathbf{S})^{-1}_{ii}}}\right). \tag{7}$$

(b) When $\mathbf{S}'\mathbf{S}$ is not invertible, the detector flips a coin to decide each bit. In this case, $P_{e,i} = 1/2$ and thus $P_e = 1/2$.

(c) When $\mathbf{S}$ is chosen randomly, we need to average over all possible matrices $\mathbf{S}$ to find the average probability of bit error. However, there are 2^{kn} possible matrices $\mathbf{S}$ and averaging over all of them is too much work. Instead, we randomly generate m matrices $\mathbf{S}$ and estimate the average P_e by averaging over these m matrices.

A function berdecorr uses this method to evaluate the decorrelator BER. The code has a lot of lines because it evaluates the BER using m signal sets for each combination of users k and SNRs snr. However, because the program generates signal sets and calculates the BER asssociated with each, there is no need for the simulated transmission of bits. Thus the program runs quickly. Since there are only 2^n distinct columns for matrix $\mathbf{S}$, it is quite possible to generate signal sets that are not linearly independent. In this case, berdecorr assumes the "flip a coin" rule is used. Just to see whether this rule dominates the error probability, we also display counts of how often $\mathbf{S}$ is rank deficient.

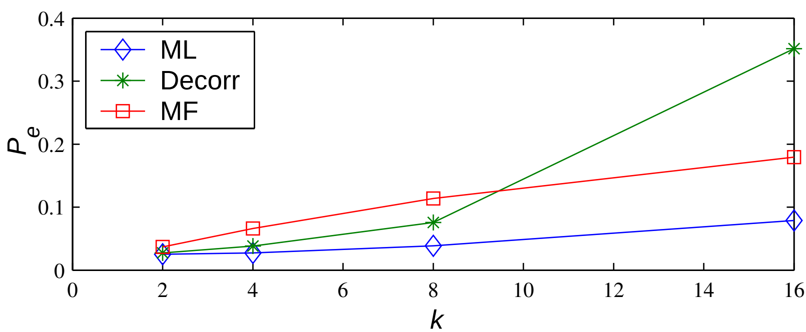
Here is the (somewhat tedious) code:

```
function Pe=berdecorr(n,k,snr,m);
%Problem 8.4.7 Solution: R-CDMA with decorrelation
%proc gain=n, users=k, average Pe for m signal sets
count=zeros(1,length(k)); %counts rank<k signal sets
Pe=zeros(length(k),length(snr)); snr=snr(:)';
for mm=1:m,
   for i=1:length(k),
   S=randomsignals(n,k(i)); R=S'*S;
   if (rank(R)<k(i))
      count(i)=count(i)+1;
      Pe(i,:)=Pe(i,:)+0.5*ones(1,length(snr));
   else
      G=diag(inv(R));
      Pe(i,:)=Pe(i,:)+sum(qfunction(sqrt((1./G)*snr)))/k(i);
   end
end
end
disp('Rank deficiency count:');disp(k);disp(count);
Pe=Pe/m;
```

Running berdecorr with processing gains $n = 16$ and $n = 32$ yields the following output:

```
>> k=[1 2 4 8 16 32];
>> pe16=berdecorr(16,k,4,10000);
Rank deficiency count:
  1       2       4       8       16      32
  0       2       2       12      454     10000
>> pe16'
ans =
  0.0228 0.0273 0.0383 0.0755 0.3515 0.5000
>> pe32=berdecorr(32,k,4,10000);
Rank deficiency count:
  1       2       4       8       16      32
  0       0       0       0       0       0
>> pe32'
ans =
  0.0228 0.0246 0.0290 0.0400 0.0771 0.3904
>>
```

As you might expect, the BER increases as the number of users increases. This occurs because the decorrelator must suppress a large set of interferers. Also, in generating 10,000 signal matrices $\mathbf{S}$ for each value of k we see that rank deficiency is fairly uncommon, however it occasionally occurs for processing gain $n = 16$, even if $k = 4$ or $k = 8$. Finally, here is a plot of these same BER statistics for $n = 16$ and $k \in \{2, 4, 8, 16\}$. Just for comparison, on the same graph is the BER for the matched filter detector and the maximum likelihood detector found in Problem 11.4.6.



We see from the graph that the decorrelator is better than the matched filter for a small number of users. However, when the number of users k is large (relative to the processing gain n), the decorrelator suffers because it must suppress all interfering users. Finally, we note that these conclusions are specific to this scenario when all users have equal SNR. When some users have very high SNR, the decorrelator is good for the low-SNR user because it zeros out the interference from the high-SNR user.

Problem 11.4.8 Solution

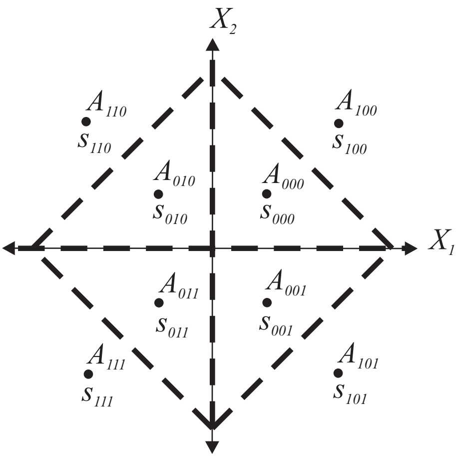
Each transmitted symbol $\mathbf{s}_{b_1 b_2 b_3}$ corresponds to the transmission of three bits given by the vector $\mathbf{b} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}'$. Note that $\mathbf{s}_{b_1 b_2 b_3}$ is a two dimensional vector with components $\begin{bmatrix} \mathbf{s}_{b_1 b_2 b_3}^{(1)} & \mathbf{s}_{b_1 b_2 b_3}^{(2)} \end{bmatrix}'$. The key to this problem is the mapping from bits to symbol components and then back to bits. From the signal constellation shown with Problem 11.3.2, we make the following observations:

- $\mathbf{s}_{b_1 b_2 b_3}$ is in the right half plane if $b_2 = 0$; otherwise it is in the left half plane.
- $\mathbf{s}_{b_1 b_2 b_3}$ is in the upper half plane if $b_3 = 0$; otherwise it is in the lower half plane.
- There is an inner ring and an outer ring of signals. $\mathbf{s}_{b_1 b_2 b_3}$ is in the inner ring if $b_1 = 0$; otherwise it is in the outer ring.

Given a bit vector $\mathbf{b}$, we use these facts by first using b_2 and b_3 to map $\mathbf{b} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}'$ to an inner ring signal vector

$$\mathbf{s} \in \left\{ \begin{bmatrix} 1 & 1 \end{bmatrix}', \begin{bmatrix} -1 & 1 \end{bmatrix}', \begin{bmatrix} -1 & -1 \end{bmatrix}', \begin{bmatrix} 1 & -1 \end{bmatrix}' \right\}. \tag{1}$$

In the next step we scale $\mathbf{s}$ by $(1 + b_1)$. If $b_1 = 1$, then $\mathbf{s}$ is stretched to the outer ring. Finally, we add a Gaussian noise vector $\mathbf{N}$ to generate the received signal $\mathbf{X} = \mathbf{s}_{b_1 b_2 b_3} + \mathbf{N}$.



In the solution to Problem 11.3.2, we found that the acceptance set for the hypothesis $H_{b_1b_2b_3}$ that $\mathbf{s}_{b_1b_2b_3}$ is transmitted is the set of signal space points closest to $\mathbf{s}_{b_1b_2b_3}$. Graphically, these acceptance sets are given in the adjacent figure. These acceptance sets correspond an inverse mapping of the received signal vector $\mathbf{X}$ to a bit vector guess $\hat{\mathbf{b}} = \begin{bmatrix}\hat{b}_1 & \hat{b}_2 & \hat{b}_3\end{bmatrix}'$ using the following rules:

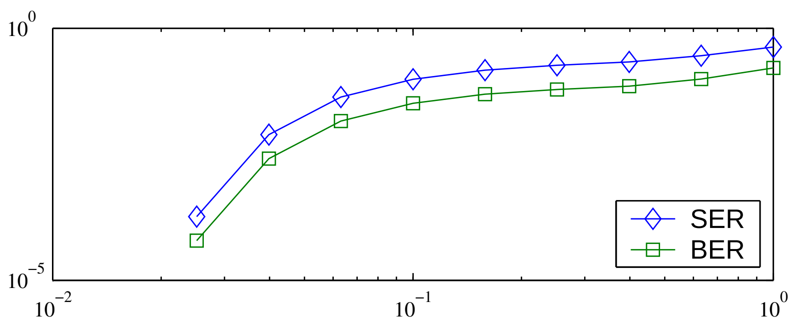
- $\hat{b}_2 = 1$ if $X_1 < 0$; otherwise $\hat{b}_2 = 0$.
- $\hat{b}_3 = 1$ if $X_2 < 0$; otherwise $\hat{b}_3 = 0$.
- If $|X_1| + |X_2| > 3\sqrt{2}/2$, then $\hat{b}_1 = 1$; otherwise $\hat{b}_1 = 0$.

We implement these steps with the function `[Pe,ber]=myqam(sigma,m)` which simulates the transmission of m symbols for each value of the vector `sigma`. Each column of `B` corresponds to a bit vector $\mathbf{b}$. Similarly, each column of `S` and `X` corresponds to a transmitted signal $\mathbf{s}$ and received signal $\mathbf{X}$.We calculate both the symbol decision errors that are made as well as the bit decision errors. Finally, a script `myqamplot.m` plots the symbol error rate `Pe` and bit error rate `ber` as a function of `sigma`. Here are the programs:

```
function [Pe,ber]=myqam(sigma,m);
Pe=zeros(size(sigma)); ber=Pe;
B=reshape(bernoullirv(0.5,3*m),3,m);
%S(1,:)=1-2*B(2,:);
%S(2,:)=1-2*B(3,:);
S=1-2*B([2; 3],:);
S=([1;1]*(1+B(1,:))).*S;
N=randn(2,m);
for i=1:length(sigma),
  X=S+sigma(i)*N;
  BR=zeros(size(B));
  BR([2;3],:)=(X<0);
  BR(1,:)=sum(abs(X))>(3/sqrt(2));
  E=(BR~=B);
  Pe(i)=sum(max(E))/m;
  ber(i)=sum(sum(E))/(3*m);
end
```

```
%myqamplot.m
sig=10.^(0.2*(-8:0));
[Pe,ber]=myqam(sig,1e6);
loglog(sig,Pe,'-d', ...
    sig,ber,'-s');
legend('SER','BER',4);
```

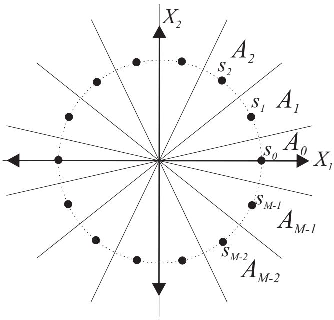
Note that we generate the bits and transmitted signals, and normalized noise only once. However for each value of **sigma**, we rescale the additive noise, recalculate the received signal and receiver bit decisions. The output of **myqamplot** is shown in this figure:



Careful reading of the figure will show that the ratio of the symbol error rate to the bit error rate is always very close to 3. This occurs because in the acceptance set for $b1b_2b_3$, the adjacent acceptance sets correspond to a one bit difference. Since the usual type of symbol error occurs when the vector $\mathbf{X}$ is in the adjacent set, a symbol error typically results in one bit being in error but two bits being received correctly. Thus the bit error rate is roughly one third the symbol error rate.

Problem 11.4.9 Solution

(a) For the M-PSK communication system with additive Gaussian noise, A_j denoted the hypothesis that signal $\mathbf{s}_j$ was transmitted. The solution to Problem 11.3.5 derived the MAP decision rule



$$\mathbf{x} \in A_m \text{ if } \|\mathbf{x} - \mathbf{s}_m\|^2 \leq \|\mathbf{x} - \mathbf{s}_j\|^2 \text{ for all } j.$$

In terms of geometry, the interpretation is that all vectors $\mathbf{x}$ closer to $\mathbf{s}_m$ than to any other signal $\mathbf{s}_j$ are assigned to A_m. In this problem, the signal constellation (i.e., the set of vectors $\mathbf{s}_i$) is the set of vectors on the circle of radius E. The acceptance regions are the "pie slices" around each signal vector.

We observe that

$$\|\mathbf{x} - \mathbf{s}_j\|^2 = (\mathbf{x} - \mathbf{s}_j)'(\mathbf{x} - \mathbf{s}_j) = \mathbf{x}'\mathbf{x} - 2\mathbf{x}'\mathbf{s}_j + \mathbf{s}_j'\mathbf{s}'. \tag{1}$$

Since all the signals are on the same circle, $\mathbf{s}_j'\mathbf{s}_j$ is the same for all j. Also, $\mathbf{x}'\mathbf{x}$ is the same for all j. Thus

$$\min_j \|\mathbf{x} - \mathbf{s}_j\|^2 = \min_j -\mathbf{x}'\mathbf{s}_j = \max_j \mathbf{x}'\mathbf{s}_j. \tag{2}$$

Since $\mathbf{x}'\mathbf{s}_j = \|\mathbf{x}\| \, \|\mathbf{s}_j\| \cos\phi$ where ϕ is the angle between $\mathbf{x}$ and $\mathbf{s}_j$. Thus maximizing $\mathbf{x}'\mathbf{s}_j$ is equivalent to minimizing the angle between $\mathbf{x}$ and $\mathbf{s}_j$.

(b) In Problem 11.4.5, we estimated the probability of symbol error without building a complete simulation of the M-PSK system. In this problem, we need to build a more complete simulation to determine the probabilities P_{ij}. By symmetry, it is sufficient to transmit $\mathbf{s}_0$ repeatedly and count how often the receiver guesses $\mathbf{s}_j$. This is done by the functionp=mpskerr(M,snr,n).

```
function p=mpskerr(M,snr,n);
%Problem 8.4.5 Solution:
%Pe=mpsksim(M,snr,n)
%n bit M-PSK simulation
t=(2*pi/M)*(0:(M-1));
S=sqrt(snr)*[cos(t);sin(t)];
X=repmat(S(:,1),1,n)+randn(2,n);
[y,e]=max(S'*X);
p=countequal(e-1,(0:(M-1)))'/n;
```

Note that column i of `S` is the signal $\mathbf{s}_{i-1}$. The kth column of `X` corresponds to $\mathbf{X}_k = \mathbf{s}_0 + \mathbf{N}_k$, the received signal for the kth transmission. Thus `y(k)` corresponds to $\max_j \mathbf{X}_k' \mathbf{s}_j$ and `e(k)` reports the receiver decision for the kth transmission. The vector `p` calculates the relative frequency of each receiver decision.

The next step is to translate the vector $\begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0,M-1} \end{bmatrix}'$ (corresponding to `p` in MATLAB) into an entire matrix $\mathbf{P}$ with elements P_{ij}. The symmetry of the phase rotiation dictates that each row of $\mathbf{P}$ should be a one element cyclic rotation of the previous row. Moreover, by symmetry we observe that $P_{01} = P_{0,M-1}$, $P_{02} = P_{0,M-2}$ and so on. However, because `p` is derived from a simulation experiment, it will exhibit this symmetry only approximately.

```
function P=mpskmatrix(p);
M=length(p);
r=[0.5 zeros(1,M-2)];
A=toeplitz(r)+...
    hankel(fliplr(r));
A=[zeros(1,M-1);A];
A=[[1; zeros(M-1,1)] A];
P=toeplitz(A*(p(:)));
```

Our ad hoc (and largely unjustified) solution is to take the average of estimates of probabilities that symmetry says should be identical. (Why this is might be a good thing to do would make an interesting exam problem.) In `mpskmatrix(p)`, the matrix `A` implements the averaging. The code will become clear by examining the matrices `A` and the output `P`.

(c) The next step is to determine the effect of the mapping of bits to transmission vectors $\mathbf{s}_j$. The matrix $\mathbf{D}$ with i,jth element d_{ij} that indicates the number of bit positions in which the bit string assigned to $\mathbf{s}_i$ differs from the bit string assigned to $\mathbf{s}_j$. In this case, the integers provide a compact representation of this mapping. For example the binary mapping is

$\mathbf{s}_0$	$\mathbf{s}_1$	$\mathbf{s}_2$	$\mathbf{s}_3$	$\mathbf{s}_4$	$\mathbf{s}_5$	$\mathbf{s}_6$	$\mathbf{s}_7$
000	001	010	011	100	101	110	111
0	1	2	3	4	5	6	7

The Gray mapping is

$\mathbf{s}_0$	$\mathbf{s}_1$	$\mathbf{s}_2$	$\mathbf{s}_3$	$\mathbf{s}_4$	$\mathbf{s}_5$	$\mathbf{s}_6$	$\mathbf{s}_7$
000	001	011	010	110	111	101	100
0	1	3	2	6	7	5	4

Thus the binary mapping can be represented by a vector

$$\mathbf{c}_1 = \begin{bmatrix} 0 & 1 & \cdots & 7 \end{bmatrix}', \tag{3}$$

while the Gray mapping is described by

$$\mathbf{c}_2 = \begin{bmatrix} 0 & 1 & 3 & 2 & 6 & 7 & 5 & 4 \end{bmatrix}'. \tag{4}$$

```
function D=mpskdist(c);
L=length(c);m=log2(L);
[C1,C2]=ndgrid(c,c);
B1=dec2bin(C1,m);
B2=dec2bin(C2,m);
D=reshape(sum((B1~=B2),2),L,L);
```

The function `D=mpskdist(c)` translates the mapping vector `c` into the matrix `D` with entries d_{ij}. The method is to generate grids `C1` and `C2` for the pairs of integers, convert each integer into a length $\log_2 M$ binary string, and then to count the number of bit positions in which each pair differs.

Given matrices $\mathbf{P}$ and $\mathbf{D}$, the rest is easy. We treat BER as as a finite random variable that takes on value d_{ij} with probability P_{ij}. the expected value of this finite random variable is the expected number of bit errors. Note that the BER is a "rate" in that

$$\text{BER} = \frac{1}{M} \sum_i \sum_j P_{ij} d_{ij}. \tag{5}$$

is the expected number of bit errors per transmitted symbol.

```
function Pb=mpskmap(c,snr,n);
M=length(c);
D=mpskdist(c);
Pb=zeros(size(snr));
for i=1:length(snr),
    p=mpskerr(M,snr(i),n);
    P=mpskmatrix(p);
    Pb(i)=finiteexp(D,P)/M;
end
```

Given the integer mapping vector `c`, we estimate the BER of the a mapping using just one more function `Pb=mpskmap(c,snr,n)`. First we calculate the matrix `D` with elements dij. Next, for each value of the vector `snr`, we use n transmissions to estimate the probabilities P_{ij}. Last, we calculate the expected number of bit errors per transmission.

(d) We evaluate the binary mapping with the following commands:

```
>> c1=0:7;
>>snr=[4     8     16     32     64];
>>Pb=mpskmap(c1,snr,1000000);
>> Pb
Pb =
    0.7640    0.4878    0.2198    0.0529    0.0038
```

(e) Here is the performance of the Gray mapping:

```
>> c2=[0 1 3 2 6 7 5 4];
>>snr=[4     8     16     32     64];
>>Pg=mpskmap(c2,snr,1000000);
>> Pg
Pg =
    0.4943    0.2855    0.1262    0.0306    0.0023
```

Experimentally, we observe that the BER of the binary mapping is higher than the BER of the Gray mapping by a factor in the neighborhood of 1.5 to 1.7

In fact, this approximate ratio can be derived by a quick and dirty analysis. For high SNR, suppose that that $\mathbf{s}_i$ is decoded as $\mathbf{s}_{i+1}$ or $\mathbf{s}_{i-1}$ with probability $q = P_{i,i+1} = P_{i,i-1}$ and all other types of errors are negligible. In this case, the BER formula based on this approximation corresponds to summing the matrix D for the first off-diagonals and the corner elements. Here are the calculations:

```
>> D=mpskdist(c1);
>> sum(diag(D,1))+sum(diag(D,-1))+D(1,8)+D(8,1)
ans =
    28
>> DG=mpskdist(c2);
>> sum(diag(DG,1))+sum(diag(DG,-1))+DG(1,8)+DG(8,1)
ans =
    16
```

Thus in high SNR, we would expect

$$\mathrm{BER(binary)} \approx 28q/M, \qquad \mathrm{BER(Gray)} \approx 16q/M. \tag{6}$$

The ratio of BERs is $28/16 = 1.75$. Experimentally, we found at high SNR that the ratio of BERs was $0.0038/0.0023 = 1.65$, which seems to be in the right ballpark.

Problem 11.4.10 Solution

As this problem is a continuation of Problem 11.4.9, this solution is also a continuation. In this problem, we want to determine the error probability for each bit k in a mapping of bits to the M-PSK signal constellation. The bit error rate associated with bit k is

$$\mathrm{BER}(k) = \frac{1}{M} \sum_i \sum_j P_{ij} d_{ij}(k) \tag{1}$$

where $d_{ij}(k)$ indicates whether the bit strings mapped to $\mathbf{s}_i$ and $\mathbf{s}_j$ differ in bit position k.

As in Problem 11.4.9, we describe the mapping by the vector of integers d. For example the binary mapping is

$\mathbf{s}_0$	$\mathbf{s}_1$	$\mathbf{s}_2$	$\mathbf{s}_3$	$\mathbf{s}_4$	$\mathbf{s}_5$	$\mathbf{s}_6$	$\mathbf{s}_7$
000	001	010	011	100	101	110	111
0	1	2	3	4	5	6	7

The Gray mapping is

$\mathbf{s}_0$	$\mathbf{s}_1$	$\mathbf{s}_2$	$\mathbf{s}_3$	$\mathbf{s}_4$	$\mathbf{s}_5$	$\mathbf{s}_6$	$\mathbf{s}_7$
000	001	011	010	110	111	101	100
0	1	3	2	6	7	5	4

Thus the binary mapping can be represented by a vector $\mathbf{c}_1 = \begin{bmatrix}0 & 1 & \cdots & 7\end{bmatrix}'$ while the Gray mapping is described by $\mathbf{c}_2 = \begin{bmatrix}0 & 1 & 3 & 2 & 6 & 7 & 5 & 4\end{bmatrix}'$.

```
function D=mpskdbit(c,k);
%See Problem 8.4.10: For mapping
%c, calculate BER of bit k
L=length(c);m=log2(L);
[C1,C2]=ndgrid(c,c);
B1=bitget(C1,k);
B2=bitget(C2,k);
D=(B1~=B2);
```

The function `D=mpskdbit(c,k)` translates the mapping vector `c` into the matrix `D` with entries d_{ij} that indicates whether bit k is in error when transmitted symbol $\mathbf{s}_i$ is decoded by the receiver as $\mathbf{s}_j$. The method is to generate grids `C1` and `C2` for the pairs of integers, identify bit k in each integer, and then check if the integers differ in bit k.

Thus, there is a matrix `D` associated with each bit position and we calculate the expected number of bit errors associated with each bit position. For each bit, the rest of the solution is the same as in Problem 11.4.9. We use the commands `p=mpskerr(M,snr,n)` and `P=mpskmatrix(p)` to calculate the matrix $\mathbf{P}$ which holds an estimate of each probability P_{ij}. Finally, using matrices $\mathbf{P}$ and $\mathbf{D}$, we treat $\text{BER}(k)$ as a finite random variable that takes on value d_{ij} with probability P_{ij}. the expected value of this finite random variable is the expected number of bit errors.

```
function Pb=mpskbitmap(c,snr,n);
%Problem 8.4.10: Calculate prob. of
%bit error for each bit position for
%an MPSK bit to symbol mapping c
M=length(c);m=log2(M);
p=mpskerr(M,snr,n);
P=mpskmatrix(p);
Pb=zeros(1,m);
for k=1:m,
    D=mpskdbit(c,k);
    Pb(k)=finiteexp(D,P)/M;
end
```

Given the integer mapping vector `c`, we estimate the BER of the a mapping using just one more function `Pb=mpskmap(c,snr,n)`. First we calculate the matrix `D` with elements dij. Next, for a given value of `snr`, we use n transmissions to estimate the probabilities P_{ij}. Last, we calculate the expected number of bit k errors per transmission.

For an SNR of 10dB, we evaluate the two mappings with the following commands:

```
>> c1=0:7;
>> mpskbitmap(c1,10,100000)
ans =
    0.2247    0.1149    0.0577

>> c2=[0 1 3 2 6 7 5 4];
>> mpskbitmap(c2,10,100000)
ans =
    0.1140    0.0572    0.0572
```

We see that in the binary mapping, the 0.22 error rate of bit 1 is roughly double that of bit 2, which is roughly double that of bit 3. For the Gray mapping, the error rate of bit 1 is cut in half relative to the binary mapping. However, the bit error rates at each position a re still not identical since the error rate of bit 1 is still double that for bit 2 or bit 3. One might surmise that careful study of the matrix **D** might lead one to prove for the Gray map that the error rate for bit 1 is exactly double that for bits 2 and 3 ... but that would be some other homework problem.

Problem Solutions – Chapter 12

Problem 12.1.1 Solution

First we note that the event $T > t_0$ has probability

$$\mathrm{P}\left[T > t_0\right] = \int_{t_0}^{\infty} \lambda e^{-\lambda t}\, dt = e^{-\lambda t_0}. \tag{1}$$

Given $T > t_0$, the conditional PDF of T is

$$f_{T|T>t_0}(t) = \begin{cases} \frac{f_T(t)}{\mathrm{P}[T>t_0]} & t > t_0, \\ 0 & \text{otherwise}, \end{cases} = \begin{cases} \lambda e^{-\lambda(t-t_0)} & t > t_0, \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

Given $T > t_0$, the minimum mean square error estimate of T is

$$\hat{T} = \mathrm{E}\left[T|T > t_0\right] = \int_{-\infty}^{\infty} t f_{T|T>t_0}(t)\; dt = \int_{t_0}^{\infty} \lambda t e^{-\lambda(t-t_0)}\, dt. \tag{3}$$

With the substitution $t' = t - t_0$, we obtain

$$\begin{aligned} \hat{T} &= \int_0^{\infty} \lambda(t_0 + t')e^{-\lambda t'}\, dt' \\ &= t_0 \underbrace{\int_0^{\infty} \lambda e^{-\lambda t'}\, dt'}_{1} + \underbrace{\int_0^{\infty} t' \lambda e^{-\lambda t'}\, dt'}_{\mathrm{E}[T]} = t_0 + \mathrm{E}\left[T\right]. \end{aligned} \tag{4}$$

Problem 12.1.2 Solution

(a) For $0 \le x \le 1$,

$$\begin{aligned} f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy &= \int_x^1 6(y-x)\, dy \\ &= 3y^2 - 6xy\Big|_{y=x}^{y=1} \\ &= 3(1 - 2x + x^2) = 3(1-x)^2. \end{aligned} \tag{1}$$

The complete expression for the marginal PDF of X is

$$f_X(x) = \begin{cases} 3(1-x)^2 & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) The blind estimate of X is

$$\hat{x}_B = \mathrm{E}[X] = \int_0^1 3x(1-x)^2\,dx = \frac{3}{2}x^2 - 2x^3 + \frac{3}{4}x^4\Big|_0^1 = \frac{1}{4}. \tag{3}$$

(c) First we calculate

$$\begin{aligned} \mathrm{P}[X < 0.5] &= \int_0^{0.5} f_X(x)\,dx \\ &= \int_0^{0.5} 3(1-x)^2\,dx = -(1-x)^3\big|_0^{0.5} = \frac{7}{8}. \end{aligned} \tag{4}$$

The conditional PDF of X given $X < 0.5$ is

$$\begin{aligned} f_{X|X<0.5}(x) &= \begin{cases} \frac{f_X(x)}{\mathrm{P}[X<0.5]} & 0 \le x < 0.5, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} \frac{24}{7}(1-x)^2 & 0 \le x < 0.5, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

The minimum mean square estimate of X given $X < 0.5$ is

$$\begin{aligned} \mathrm{E}[X|X < 0.5] &= \int_{-\infty}^{\infty} x f_{X|X<0.5}(x)\,dx \\ &= \int_0^{0.5} \frac{24x}{7}(1-x)^2\,dx \\ &= \frac{12x^2}{7} - \frac{16x^3}{7} + \frac{6x^4}{7}\Big|_0^{0.5} = \frac{11}{56}. \end{aligned} \tag{6}$$

(d) For $y < 0$ or $y > 1$, $f_Y(y) = 0$. For $0 \le y \le 1$,

$$f_Y(y) = \int_0^y 6(y-x)\,dx = 6xy - 3x^2\big|_0^y = 3y^2. \tag{7}$$

The complete expression for the marginal PDF of Y is

$$f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

(e) The blind estimate of Y is

$$\hat{y}_B = \mathrm{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^1 3y^3\, dy = \frac{3}{4}. \tag{9}$$

(f) First we calculate

$$\mathrm{P}[Y > 0.5] = \int_{0.5}^{\infty} f_Y(y)\, dy = \int_{0.5}^1 3y^2\, dy = \frac{7}{8}. \tag{10}$$

The conditional PDF of Y given $Y > 0.5$ is

$$\begin{aligned} f_{Y|Y>0.5}(y) &= \begin{cases} \frac{f_Y(y)}{\mathrm{P}[Y>0.5]} & y > 0.5, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 24y^2/7 & y > 0.5, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{11}$$

The minimum mean square estimate of Y given $Y > 0.5$ is

$$\begin{aligned} \mathrm{E}[Y|Y > 0.5] &= \int_{-\infty}^{\infty} y f_{Y|Y>0.5}(y)\, dy \\ &= \int_{0.5}^1 \frac{24y^3}{7}\, dy = \left.\frac{6y^4}{7}\right|_{0.5}^1 = \frac{45}{56}. \end{aligned} \tag{12}$$

Problem 12.1.3 Solution

(a) For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_x^1 2\, dy = 2(1 - x). \tag{1}$$

The complete expression of the PDF of X is

$$f_X(x) = \begin{cases} 2(1-x) & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) The blind estimate of X is

$$\hat{X}_B = \mathrm{E}[X] = \int_0^1 2x(1-x)\,dx = \left(x^2 - \frac{2x^3}{3}\right)\Big|_0^1 = \frac{1}{3}. \tag{3}$$

(c) First we calculate

$$\begin{aligned} \mathrm{P}[X > 1/2] &= \int_{1/2}^1 f_X(x)\ dx \\ &= \int_{1/2}^1 2(1-x)\,dx = (2x - x^2)\big|_{1/2}^1 = \frac{1}{4}. \end{aligned} \tag{4}$$

Now we calculate the conditional PDF of X given $X > 1/2$.

$$\begin{aligned} f_{X|X>1/2}(x) &= \begin{cases} \frac{f_X(x)}{\mathrm{P}[X>1/2]} & x > 1/2, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 8(1-x) & 1/2 < x \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

The minimum mean square error estimate of X given $X > 1/2$ is

$$\begin{aligned} \mathrm{E}[X|X > 1/2] &= \int_{-\infty}^{\infty} x f_{X|X>1/2}(x)\ dx \\ &= \int_{1/2}^1 8x(1-x)\,dx = \left(4x^2 - \frac{8x^3}{3}\right)\Big|_{1/2}^1 = \frac{2}{3}. \end{aligned} \tag{6}$$

(d) For $0 \le y \le 1$, the marginal PDF of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_0^y 2\,dx = 2y. \tag{7}$$

The complete expression for the marginal PDF of Y is

$$f_Y(y) = \begin{cases} 2y & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

(e) The blind estimate of Y is

$$\hat{y}_B = \mathrm{E}[Y] = \int_0^1 2y^2\, dy = \frac{2}{3}. \tag{9}$$

(f) We already know that $\mathrm{P}[X > 1/2] = 1/4$. However, this problem differs from the other problems in this section because we will estimate Y based on the observation of X. In this case, we need to calculate the conditional joint PDF

$$\begin{aligned} f_{X,Y|X>1/2}(x,y) &= \begin{cases} \frac{f_{X,Y}(x,y)}{\mathrm{P}[X>1/2]} & x > 1/2, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} 8 & 1/2 < x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{10}$$

The MMSE estimate of Y given $X > 1/2$ is

$$\begin{aligned} \mathrm{E}[Y|X > 1/2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y|X>1/2}(x,y)\; dx\, dy \\ &= \int_{1/2}^{1} y \left(\int_{1/2}^{y} 8\, dx \right) dy \\ &= \int_{1/2}^{1} y(8y-4)\, dy = \frac{5}{6}. \end{aligned} \tag{11}$$

Problem 12.1.4 Solution

The joint PDF of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 6(y-x) & 0 \le x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The conditional PDF of X given Y is found by dividing the joint PDF by the marginal with respect to Y. For $y < 0$ or $y > 1$, $f_Y(y) = 0$. For $0 \le y \le 1$,

$$f_Y(y) = \int_0^y 6(y-x)\,dx = 6xy - 3x^2\big|_0^y = 3y^2 \tag{2}$$

The complete expression for the marginal PDF of Y is

$$f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Thus for $0 < y \le 1$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{6(y-x)}{3y^2} & 0 \le x \le y, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

(b) The minimum mean square estimator of X given $Y = y$ is

$$\begin{aligned} \hat{X}_M(y) = \mathrm{E}\left[X|Y=y\right] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\; dx \\ &= \int_0^y \frac{6x(y-x)}{3y^2}\,dx \\ &= \left.\frac{3x^2y - 2x^3}{3y^2}\right|_{x=0}^{x=y} = y/3. \end{aligned} \tag{5}$$

(c) First we must find the marginal PDF for X. For $0 \le x \le 1$,

$$\begin{aligned} f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy &= \int_x^1 6(y-x)\,dy = 3y^2 - 6xy\big|_{y=x}^{y=1} \\ &= 3(1-2x+x^2) = 3(1-x)^2. \end{aligned} \tag{6}$$

The conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{2(y-x)}{1-2x+x^2} & x \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

(d) The minimum mean square estimator of Y given X is

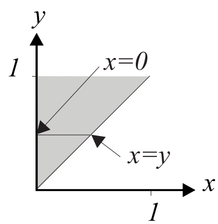
$$\hat{Y}_M(x) = \mathrm{E}\left[Y|X=x\right] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\ dy$$
$$= \int_x^1 \frac{2y(y-x)}{1-2x+x^2}\,dy$$
$$= \left.\frac{(2/3)y^3 - y^2 x}{1-2x+x^2}\right|_{y=x}^{y=1} = \frac{2-3x+x^3}{3(1-x)^2}. \tag{8}$$

Perhaps surprisingly, this result can be simplified to

$$\hat{Y}_M(x) = \frac{x}{3} + \frac{2}{3}. \tag{9}$$

Problem 12.1.5 Solution

(a) First we find the marginal PDF $f_Y(y)$. For $0 \le y \le 2$,



$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_0^y 2\,dx = 2y. \tag{1}$$

Hence, for $0 \le y \le 2$, the conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} 1/y & 0 \le x \le y, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(b) The optimum mean squared error estimate of X given $Y=y$ is

$$\hat{x}_M(y) = \mathrm{E}\left[X|Y=y\right] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\ dx = \int_0^y \frac{x}{y}\,dx = y/2. \tag{3}$$

(c) The MMSE estimator of X given Y is $\hat{X}_M(Y) = \mathrm{E}[X|Y] = Y/2$. The mean squared error is

$$e^*_{X,Y} = \mathrm{E}\left[(X - \hat{X}_M(Y))^2\right]$$
$$= \mathrm{E}\left[(X-Y/2)^2\right] = \mathrm{E}\left[X^2 - XY + Y^2/4\right]. \tag{4}$$

Of course, the integral must be evaluated.

$$\begin{aligned} e^*_{X,Y} &= \int_0^1 \int_0^y 2(x^2 - xy + y^2/4)\, dx\, dy \\ &= \int_0^1 \left(2x^3/3 - x^2 y + xy^2/2\right)\Big|_{x=0}^{x=y} dy \\ &= \int_0^1 \frac{y^3}{6}\, dy = 1/24. \end{aligned} \tag{5}$$

Another approach to finding the mean square error is to recognize that the MMSE estimator is a linear estimator and thus must be the optimal linear estimator. Hence, the mean square error of the optimal linear estimator given by Theorem 12.3 must equal $e^*_{X,Y}$. That is, $e^*_{X,Y} = \text{Var}[X](1 - \rho^2_{X,Y})$. However, calculation of the correlation coefficient $\rho_{X,Y}$ is at least as much work as direct calculation of $e^*_{X,Y}$.

Problem 12.1.6 Solution

(a) The MMSE estimator of Z given Y is always the conditional expectation. Random variables Y and Z are jointly Gaussian. For jointly (bivariate) Gaussian random variables Y and Z, the conditional expectation is

$$\text{E}\left[Z|Y\right] = \rho_{ZY} \frac{\sigma_Z}{\sigma_Y}\left(Y - \mu_Y\right) + \mu_Z. \tag{1}$$

In this case, it is given that $\mu_Z = 0$, $\sigma_Z = 1$ and

$$\mu_Y = \text{E}\left[Y\right] = \text{E}\left[X\right] + \text{E}\left[Z\right] = 0. \tag{2}$$

Since X and Z are independent,

$$\sigma^2_Y = \sigma^2_X + \sigma^2_Z = 2, \tag{3}$$

$$\begin{aligned} \rho_{Z,Y} &= \frac{\text{E}\left[ZY\right]}{\sigma_Y \sigma_Z} \\ &= \frac{\text{E}\left[Z(X+Z)\right]}{\sigma_Y \sigma_Z} \\ &= \frac{\text{E}\left[Z\right]\text{E}\left[X\right] + \text{E}\left[Z^2\right]}{\sigma_Y \sigma_Z} = \frac{\sigma_Z}{\sigma_Y} = \frac{1}{\sqrt{2}}. \end{aligned} \tag{4}$$

It follows that

$$\hat{Z}(Y) = \mathrm{E}\left[Z|Y\right] = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}(Y-0)+0 = \frac{Y}{2}. \tag{5}$$

(b) If you solved the previous part correctly, you can find the mean squared error from first principles:

$$e = \mathrm{E}\left[(Z-Y/2)^2\right] = \mathrm{E}\left[Z^2\right] - \mathrm{E}\left[ZY\right] + \mathrm{E}\left[Y^2\right]/4. \tag{6}$$

Since $\mathrm{E}[Z^2] = \sigma_Z^2 = 1$, $\mathrm{E}[Y^2] = \sigma_Y^2 = 2$ and, from the previous part, $\mathrm{E}[ZY] = \mathrm{E}[Z^2] = 1$, we see that $e = 1 - 1 + 2/4 = 1/2$.

On the other hand, if you didn't solve the previous part, you might still remember that

$$e = \mathrm{Var}[Z|Y] = (1-\rho_{Z,Y}^2)\sigma_Z^2 = 1/2.$$

Problem 12.1.7 Solution

We need to find the conditional estimate

$$\mathrm{E}\left[X|Y=y\right] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx. \tag{1}$$

Replacing y by Y in $\mathrm{E}[X|Y=y]$ will yield the requested $\mathrm{E}[X|Y]$. We start by finding $f_{Y|X}(y|x)$. Given $X=x$, $Y = x - Z$ so that

$$\begin{aligned} \mathrm{P}\left[Y \le y|X=x\right] &= \mathrm{P}\left[x - Z \le y|X=x\right] \\ &= \mathrm{P}\left[Z \ge x-y|X=x\right] = 1 - F_Z(x-y). \end{aligned} \tag{2}$$

Note the last inequality follows because Z and X are independent random variables. Taking derivatives, we have

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{d\mathrm{P}\left[Z \le x-y|X=x\right]}{dy} \\ &= \frac{d}{dy}\left(1 - F_Z(x-y)\right) = f_Z(x-y). \end{aligned} \tag{3}$$

It follows that X and Y have joint PDF

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x) = f_Z(x-y)\, f_X(x). \tag{4}$$

By the definition of conditional PDF,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_Z(x-y)\, f_X(x)}{f_Y(y)}, \tag{5}$$

and thus

$$\begin{aligned} \mathrm{E}[X|Y=y] &= \int_{-\infty}^{\infty} x \frac{f_Z(x-y)\, f_X(x)}{f_Y(y)}\, dx \\ &= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} x f_Z(x-y)\, f_X(x)\, dx. \end{aligned} \tag{6}$$

Without more information, this is the simplest possible answer. Also note that the denominator $f_Y(y)$ is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \int_{-\infty}^{\infty} f_Z(x-y)\, f_X(x)\, dx. \tag{7}$$

For a given PDF $f_Z(z)$, it is sometimes possible to compute these integrals in closed form; Gaussian Z is one such example.

Problem 12.1.8 Solution

(a) The MMSE estimate is always the conditional expected value. Since X is binary, given $\mathbf{Y} = \mathbf{y}$,

$$\begin{aligned} \hat{X}_n(\mathbf{y}) &= \mathrm{E}[X|\mathbf{Y}=\mathbf{y}] \\ &= (1)\,\mathrm{P}[X=1|\mathbf{Y}=\mathbf{y}] + (-1)\,\mathrm{P}[X=-1|\mathbf{Y}=-\mathbf{y}] \\ &= (1)\,\mathrm{P}[X=1|\mathbf{Y}=\mathbf{y}] + (-1)(1-\mathrm{P}[X=1|\mathbf{Y}=-\mathbf{y}]) \\ &= 2\,\mathrm{P}[X=1|\mathbf{Y}=\mathbf{y}] - 1. \end{aligned} \tag{1}$$

Since $\mathbf{Y}$ is continuous, we write the conditioning event as $\mathbf{y} < \mathbf{Y} \leq \mathbf{y} + d\mathbf{y}$ so that

$$\begin{aligned} &\mathrm{P}[X=1|\mathbf{y} < \mathbf{Y} \leq \mathbf{y}+d\mathbf{y}] \\ &\quad = \frac{\mathrm{P}[\mathbf{y} < \mathbf{Y} \leq \mathbf{y}+d\mathbf{y}|X=1]\,\mathrm{P}[X=1]}{\mathrm{P}[\mathbf{y} < \mathbf{Y} \leq \mathbf{y}+d\mathbf{y}]} \\ &\quad = \frac{(1/2) f_{\mathbf{Y}|X}(\mathbf{y}|1)\, d\mathbf{y}}{(1/2) f_{\mathbf{Y}|X}(\mathbf{y}|1)\, d\mathbf{y} + (1/2) f_{\mathbf{Y}|X}(\mathbf{y}|-1)\, d\mathbf{y}}. \end{aligned} \tag{2}$$

We conclude that

$$\begin{aligned}
\mathrm{P}\left[X=1|\mathbf{Y}=\mathbf{y}\right] &= \frac{f_{\mathbf{Y}|X}(\mathbf{y}|1)}{f_{\mathbf{Y}|X}(\mathbf{y}|1)+f_{\mathbf{Y}|X}(\mathbf{y}|-1)} \\
&= \frac{1}{1+\frac{f_{\mathbf{Y}|X}(\mathbf{y}|-1)}{f_{\mathbf{Y}|X}(\mathbf{y}|1)}} = \frac{1}{1+L(\mathbf{y})}. \qquad (3)
\end{aligned}$$

Combining (1) and (3), we see that

$$\begin{aligned}
\hat{X}_n(\mathbf{y}) &= 2\,\mathrm{P}\left[X=1|\mathbf{Y}=\mathbf{y}\right]-1 \\
&= 2\left(\frac{1}{1+L(\mathbf{y})}\right)-1 = \frac{1-L(\mathbf{y})}{1+L(\mathbf{y})}. \qquad (4)
\end{aligned}$$

(b) To simplify the likelihood ratio, we observe that given $X=x$, $Y_i = x+W_i$ and

$$f_{Y_i|X}(y_i|x) = \frac{1}{\sqrt{2\pi}} e^{(y_i-x)^2/2}. \qquad (5)$$

Since the W_i are iid and independent of X, given $X=x$, the Y_i are conditionally iid. That is,

$$f_{\mathbf{Y}|X}(\mathbf{y}|x) = \prod_{i=1}^{n} f_{Y_i|X}(y_i|x) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^{n}(y_i-x)^2/2}. \qquad (6)$$

Thus

$$\begin{aligned}
L(\mathbf{y}) = \frac{f_{\mathbf{Y}|X}(\mathbf{y}|-1)}{f_{\mathbf{Y}|X}(\mathbf{y}|1)} &= \frac{e^{-\sum_{i=1}^{n}(y_i+1)^2/2}}{e^{-\sum_{i=1}^{n}(y_i-1)^2/2}} \\
&= e^{-\sum_{i=1}^{n}[(y_i+1)^2-(y_i-1)^2]/2} = e^{-2\sum_{i=1}^{n} y_i}. \qquad (7)
\end{aligned}$$

Problem 12.2.1 Solution

(a) The marginal PMFs of X and Y are listed below

$$P_X(x) = \begin{cases} 1/3 & x = -1, 0, 1, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

$$P_Y(y) = \begin{cases} 1/4 & y = -3, -1, 0, 1, 3, \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

(b) No, the random variables X and Y are not independent since

$$P_{X,Y}(1, -3) = 0 \neq P_X(1)\, P_Y(-3) \tag{3}$$

(c) Direct evaluation leads to

$$\mathrm{E}[X] = 0, \qquad \mathrm{Var}[X] = 2/3, \tag{4}$$

$$\mathrm{E}[Y] = 0, \qquad \mathrm{Var}[Y] = 5. \tag{5}$$

This implies

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[XY] = 7/6. \tag{6}$$

(d) From Theorem 12.3, the optimal linear estimate of X given Y is

$$\hat{X}_L(Y) = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y) + \mu_X = \frac{7}{30}Y + 0. \tag{7}$$

Therefore, $a^* = 7/30$ and $b^* = 0$.

(e) From the previous part, X and Y have correlation coefficient

$$\rho_{X,Y} = \mathrm{Cov}[X, Y] / \sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = \sqrt{49/120}. \tag{8}$$

From Theorem 12.3, the minimum mean square error of the optimum linear estimate is

$$e_L^* = \sigma_X^2(1 - \rho_{X,Y}^2) = \frac{2}{3}\frac{71}{120} = \frac{71}{180}. \tag{9}$$

(f) The conditional probability mass function is

$$P_{X|Y}(x|-3) = \frac{P_{X,Y}(x,-3)}{P_Y(-3)} = \begin{cases} \frac{1/6}{1/4} = 2/3 & x = -1, \\ \frac{1/12}{1/4} = 1/3 & x = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

(g) The minimum mean square estimator of X given that $Y = 3$ is

$$\hat{x}_M(-3) = \mathrm{E}\left[X|Y=-3\right] = \sum_x x P_{X|Y}(x|-3) = -2/3. \tag{11}$$

(h) The mean squared error of this estimator is

$$\begin{aligned} \hat{e}_M(-3) &= \mathrm{E}\left[(X - \hat{x}_M(-3))^2 | Y = -3\right] \\ &= \sum_x (x + 2/3)^2 P_{X|Y}(x|-3) \\ &= (-1/3)^2(2/3) + (2/3)^2(1/3) = 2/9. \end{aligned} \tag{12}$$

Problem 12.2.2 Solution

The problem statement tells us that

$$f_V(v) = \begin{cases} 1/12 & -6 \le v \le 6, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Furthermore, we are also told that $R = V + X$ where X is a Gaussian $(0, \sqrt{3})$ random variable.

(a) The expected value of R is the expected value V plus the expected value of X. We already know that X has zero expected value, and that V is uniformly distributed between -6 and 6 volts and therefore also has zero expected value. So

$$\mathrm{E}\left[R\right] = \mathrm{E}\left[V + X\right] = \mathrm{E}\left[V\right] + \mathrm{E}\left[X\right] = 0. \tag{2}$$

(b) Because X and V are independent random variables, the variance of R is the sum of the variance of V and the variance of X.

$$\text{Var}[R] = \text{Var}[V] + \text{Var}[X] = 12 + 3 = 15. \tag{3}$$

(c) Since $\text{E}[R] = \text{E}[V] = 0$,

$$\text{Cov}\left[V, R\right] = \text{E}\left[VR\right] = \text{E}\left[V(V+X)\right] = \text{E}\left[V^2\right] = \text{Var}[V]. \tag{4}$$

(d) The correlation coefficient of V and R is

$$\rho_{V,R} = \frac{\text{Cov}\left[V, R\right]}{\sqrt{\text{Var}[V]\,\text{Var}[R]}} = \frac{\text{Var}[V]}{\sqrt{\text{Var}[V]\,\text{Var}[R]}} = \frac{\sigma_V}{\sigma_R}. \tag{5}$$

The LMSE estimate of V given R is

$$\hat{V}(R) = \rho_{V,R}\frac{\sigma_V}{\sigma_R}\left(R - \text{E}\left[R\right]\right) + \text{E}\left[V\right] = \frac{\sigma_V^2}{\sigma_R^2}R = \frac{12}{15}R. \tag{6}$$

Therefore $a^* = 12/15 = 4/5$ and $b^* = 0$.

(e) The minimum mean square error in the estimate is

$$e^* = \text{Var}[V](1 - \rho_{V,R}^2) = 12(1 - 12/15) = 12/5. \tag{7}$$

Problem 12.2.3 Solution

The solution to this problem is to simply calculate the various quantities required for the optimal linear estimator given by Theorem 12.3. First we calculate the necessary moments of X and Y.

$$\text{E}\left[X\right] = -1(1/4) + 0(1/2) + 1(1/4) = 0, \tag{1}$$
$$\text{E}\left[X^2\right] = (-1)^2(1/4) + 0^2(1/2) + 1^2(1/4) = 1/2, \tag{2}$$
$$\text{E}\left[Y\right] = -1(17/48) + 0(17/48) + 1(14/48) = -1/16, \tag{3}$$
$$\text{E}\left[Y^2\right] = (-1)^2(17/48) + 0^2(17/48) + 1^2(14/48) = 31/48, \tag{4}$$
$$\text{E}\left[XY\right] = 3/16 - 0 - 0 + 1/8 = 5/16. \tag{5}$$

The variances and covariance are

$$\operatorname{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 1/2, \tag{6}$$

$$\operatorname{Var}[Y] = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 493/768, \tag{7}$$

$$\operatorname{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 5/16, \tag{8}$$

$$\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} = \frac{5\sqrt{6}}{\sqrt{493}}. \tag{9}$$

By reversing the labels of X and Y in Theorem 12.3, we find that the optimal linear estimator of Y given X is

$$\hat{Y}_L(X) = \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}(X - \mathrm{E}[X]) + \mathrm{E}[Y] = \frac{5}{8}X - \frac{1}{16}. \tag{10}$$

The mean square estimation error is

$$e_L^* = \operatorname{Var}[Y](1 - \rho_{X,Y}^2) = 343/768. \tag{11}$$

Problem 12.2.4 Solution

To solve this problem, we use Theorem 12.3. The only difficulty is in computing $\mathrm{E}[X]$, $\mathrm{E}[Y]$, $\operatorname{Var}[X]$, $\operatorname{Var}[Y]$, and $\rho_{X,Y}$. First we calculate the marginal PDFs

$$f_X(x) = \int_x^1 2(y+x)\,dy = y^2 + 2xy\Big|_{y=x}^{y=1} = 1 + 2x - 3x^2, \tag{1}$$

$$f_Y(y) = \int_0^y 2(y+x)\,dx = 2xy + x^2\Big|_{x=0}^{x=y} = 3y^2. \tag{2}$$

The first and second moments of X are

$$\mathrm{E}[X] = \int_0^1 (x + 2x^2 - 3x^3)\,dx = x^2/2 + 2x^3/3 - 3x^4/4\Big|_0^1 = 5/12, \tag{3}$$

$$\mathrm{E}\left[X^2\right] = \int_0^1 (x^2 + 2x^3 - 3x^4)\,dx = x^3/3 + x^4/2 - 3x^5/5\Big|_0^1 = 7/30. \tag{4}$$

The first and second moments of Y are

$$\mathrm{E}[Y] = \int_0^1 3y^3\,dy = 3/4, \qquad \mathrm{E}\left[Y^2\right] = \int_0^1 3y^4\,dy = 3/5. \tag{5}$$

Thus, X and Y each have variance

$$\operatorname{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}\left[X\right])^2 = \frac{129}{2160}, \tag{6}$$

$$\operatorname{Var}[Y] = \mathrm{E}\left[Y^2\right] - (\mathrm{E}\left[Y\right])^2 = \frac{3}{80}. \tag{7}$$

To calculate the correlation coefficient, we first must calculate the correlation

$$\begin{aligned} \mathrm{E}\left[XY\right] &= \int_0^1 \int_0^y 2xy(x+y)\,dx\,dy \\ &= \int_0^1 \left[2x^3y/3 + x^2y^2\right]\Big|_{x=0}^{x=y} dy = \int_0^1 \frac{5y^4}{3}\,dy = 1/3. \end{aligned} \tag{8}$$

Hence, the correlation coefficient is

$$\rho_{X,Y} = \frac{\operatorname{Cov}\left[X,Y\right]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} \frac{\mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} = \frac{5}{\sqrt{129}}. \tag{9}$$

Finally, we use Theorem 12.3 to combine these quantities in the optimal linear estimator.

$$\begin{aligned} \hat{X}_L(Y) &= \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}\left(Y - \mathrm{E}\left[Y\right]\right) + \mathrm{E}\left[X\right] \\ &= \frac{5}{\sqrt{129}} \frac{\sqrt{129}}{9}\left(Y - \frac{3}{4}\right) + \frac{5}{12} = \frac{5}{9}Y. \end{aligned} \tag{10}$$

Problem 12.2.5 Solution

The linear mean square estimator of X given Y is

$$\hat{X}_L(Y) = \left(\frac{\mathrm{E}\left[XY\right] - \mu_X\mu_Y}{\operatorname{Var}[Y]}\right)(Y - \mu_Y) + \mu_X. \tag{1}$$

To find the parameters of this estimator, we calculate

$$f_Y(y) = \int_0^y 6(y-x)\,dx = 6xy - 3x^2\Big|_0^y = 3y^2 \qquad (0 \le y \le 1), \tag{2}$$

$$f_X(x) = \int_x^1 6(y-x)\,dy = \begin{cases} 3(1 + -2x + x^2) & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The moments of X and Y are

$$\mathrm{E}[Y] = \int_0^1 3y^3\,dy = 3/4, \tag{4}$$

$$\mathrm{E}[X] = \int_0^1 3x(1-2x+x^2)\,dx = 1/4, \tag{5}$$

$$\mathrm{E}\left[Y^2\right] = \int_0^1 3y^4\,dy = 3/5, \tag{6}$$

$$\mathrm{E}\left[X^2\right] = \int_0^1 3x^2(1+-2x+x^2)\,dx = 1/10. \tag{7}$$

The correlation between X and Y is

$$\mathrm{E}[XY] = 6\int_0^1\int_x^1 xy(y-x)\,dy\,dx = 1/5. \tag{8}$$

Putting these pieces together, the optimal linear estimate of X given Y is

$$\hat{X}_L(Y) = \left(\frac{1/5-3/16}{3/5-(3/4)^2}\right)\left(Y-\frac{3}{4}\right)+\frac{1}{4} = \frac{Y}{3}. \tag{9}$$

Problem 12.2.6 Solution

We are told that random variable X has a second order Erlang distribution

$$f_X(x) = \begin{cases} \lambda x e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We also know that given $X = x$, random variable Y is uniform on $[0, x]$ so that

$$f_{Y|X}(y|x) = \begin{cases} 1/x & 0 \leq y \leq x, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(a) Given $X = x$, Y is uniform on $[0, x]$. Hence $\mathrm{E}[Y|X = x] = x/2$. Thus the minimum mean square estimate of Y given X is

$$\hat{Y}_M(X) = \mathrm{E}[Y|X] = X/2. \tag{3}$$

(b) The minimum mean square estimate of X given Y can be found by finding the conditional probability density function of X given Y. First we find the joint density function.

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x) = \begin{cases} \lambda e^{-\lambda x} & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Now we can find the marginal of Y

$$f_Y(y) = \int_y^\infty \lambda e^{-\lambda x}\, dx = \begin{cases} e^{-\lambda y} & y \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

By dividing the joint density by the marginal density of Y we arrive at the conditional density of X given Y.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \lambda e^{-\lambda(x-y)} & x \ge y, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Now we are in a position to find the minimum mean square estimate of X given Y. Given $Y = y$, the conditional expected value of X is

$$\mathrm{E}\left[X|Y=y\right] = \int_y^\infty \lambda x e^{-\lambda(x-y)}\, dx. \tag{7}$$

Making the substitution $u = x - y$ yields

$$\mathrm{E}\left[X|Y=y\right] = \int_0^\infty \lambda(u+y) e^{-\lambda u}\, du. \tag{8}$$

We observe that if U is an exponential random variable with parameter λ, then

$$\mathrm{E}\left[X|Y=y\right] = \mathrm{E}\left[U+y\right] = \frac{1}{\lambda} + y. \tag{9}$$

The minimum mean square error estimate of X given Y is

$$\hat{X}_M(Y) = \mathrm{E}\left[X|Y\right] = \frac{1}{\lambda} + Y. \tag{10}$$

(c) Since the MMSE estimate of Y given X is the linear estimate $\hat{Y}_M(X) = X/2$, the optimal linear estimate of Y given X must also be the MMSE estimate. That is, $\hat{Y}_L(X) = X/2$.

(d) Since the MMSE estimate of X given Y is the linear estimate $\hat{X}_M(Y) = Y + 1/\lambda$, the optimal linear estimate of X given Y must also be the MMSE estimate. That is, $\hat{X}_L(Y) = Y + 1/\lambda$.

Problem 12.2.7 Solution

From the problem statement, we learn the following facts:

$$f_R(r) = \begin{cases} e^{-r} & r \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad f_{X|R}(x|r) = \begin{cases} re^{-rx} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Note that $f_{X,R}(x,r) > 0$ for all non-negative X and R. Hence, for the remainder of the problem, we assume both X and R are non-negative and we omit the usual "zero otherwise" considerations.

(a) To find $\hat{r}_M(X)$, we need the conditional PDF

$$f_{R|X}(r|x) = \frac{f_{X|R}(x|r)\, f_R(r)}{f_X(x)}. \tag{2}$$

The marginal PDF of X is

$$f_X(x) = \int_0^\infty f_{X|R}(x|r)\, f_R(r)\; dr = \int_0^\infty re^{-(x+1)r}\, dr. \tag{3}$$

We use the integration by parts formula $\int u\, dv = uv - \int v\, du$ by choosing $u = r$ and $dv = e^{-(x+1)r}\, dr$. Thus $v = -e^{-(x+1)r}/(x+1)$ and

$$\begin{aligned} f_X(x) &= \frac{-r}{x+1} e^{-(x+1)r}\Big|_0^\infty + \frac{1}{x+1} \int_0^\infty e^{-(x+1)r}\, dr \\ &= \frac{-1}{(x+1)^2} e^{-(x+1)r}\Big|_0^\infty = \frac{1}{(x+1)^2}. \end{aligned} \tag{4}$$

Now we can find the conditional PDF of R given X.

$$f_{R|X}(r|x) = \frac{f_{X|R}(x|r)\, f_R(r)}{f_X(x)} = (x+1)^2 r e^{-(x+1)r}. \tag{5}$$

By comparing, $f_{R|X}(r|x)$ to the Erlang PDF shown in Appendix A, we see that given $X = x$, the conditional PDF of R is an Erlang PDF with parameters $n = 1$ and $\lambda = x + 1$. This implies

$$\mathrm{E}[R|X = x] = \frac{1}{x+1}, \qquad \mathrm{Var}[R|X = x] = \frac{1}{(x+1)^2}. \tag{6}$$

Hence, the MMSE estimator of R given X is

$$\hat{r}_M(X) = \mathrm{E}[R|X] = \frac{1}{X+1}. \tag{7}$$

(b) The MMSE estimate of X given $R = r$ is $\mathrm{E}[X|R = r]$. From the initial problem statement, we know that given $R = r$, X is exponential with expectred value $1/r$. That is, $\mathrm{E}[X|R = r] = 1/r$. Another way of writing this statement is

$$\hat{x}_M(R) = \mathrm{E}[X|R] = 1/R. \tag{8}$$

(c) Note that the expected value of X is

$$\mathrm{E}[X] = \int_0^\infty x f_X(x)\, dx = \int_0^\infty \frac{x}{(x+1)^2}\, dx = \infty. \tag{9}$$

Because $\mathrm{E}[X]$ doesn't exist, the LMSE estimate of X given R doesn't exist.

(d) Just as in part (c), because $\mathrm{E}[X]$ doesn't exist, the LMSE estimate of R given X doesn't exist.

Problem 12.2.8 Solution

(a) As a function of a, the mean squared error is

$$e = \mathrm{E}\left[(aY - X)^2\right] = a^2\,\mathrm{E}\left[Y^2\right] - 2a\,\mathrm{E}\left[XY\right] + \mathrm{E}\left[X^2\right]. \tag{1}$$

Setting $de/da|_{a=a^*} = 0$ yields

$$a^* = \frac{\mathrm{E}\left[XY\right]}{\mathrm{E}\left[Y^2\right]}. \tag{2}$$

(b) Using $a = a^*$, the mean squared error is

$$e^* = \mathrm{E}\left[X^2\right] - \frac{\left(\mathrm{E}\left[XY\right]\right)^2}{\mathrm{E}\left[Y^2\right]}. \tag{3}$$

(c) We can write the LMSE estimator given in Theorem 12.3 in the form

$$\hat{x}_L(()Y) = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}Y - b, \tag{4}$$

where

$$b = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}\,\mathrm{E}\left[Y\right] - \mathrm{E}\left[X\right]. \tag{5}$$

When $b = 0$, $\hat{X}(Y)$ is the LMSE estimate. Note that the typical way that $b = 0$ occurs when $\mathrm{E}[X] = \mathrm{E}[Y] = 0$. However, it is possible that the right combination of expected values, variances, and correlation coefficent can also yield $b = 0$.

Problem 12.2.9 Solution

These four joint PMFs are actually related to each other. In particular, completing the row sums and column sums shows that each random variable has the same marginal PMF. That is,

$$\begin{aligned} P_X(x) = P_Y(x) = P_U(x) = P_V(x) = P_S(x) = P_T(x) = P_Q(x) = P_R(x) \\ = \begin{cases} 1/3 & x = -1, 0, 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{1}$$

This implies

$$\mathrm{E}[X] = \mathrm{E}[Y] = \mathrm{E}[U] = \mathrm{E}[V] = \mathrm{E}[S] = \mathrm{E}[T] = \mathrm{E}[Q] = \mathrm{E}[R] = 0, \tag{2}$$

and that

$$\begin{aligned} \mathrm{E}\left[X^2\right] &= \mathrm{E}\left[Y^2\right] = \mathrm{E}\left[U^2\right] = \mathrm{E}\left[V^2\right] \\ &= \mathrm{E}\left[S^2\right] = \mathrm{E}\left[T^2\right] = \mathrm{E}\left[Q^2\right] = \mathrm{E}\left[R^2\right] = 2/3. \end{aligned} \tag{3}$$

Since each random variable has zero expected value, the second moment equals the variance. Also, the standard deviation of each random variable is $\sqrt{2/3}$. These common properties will make it much easier to answer the questions.

(a) Random variables X and Y are independent since for all x and y,

$$P_{X,Y}(x, y) = P_X(x)\, P_Y(y). \tag{4}$$

Since each other pair of random variables has the same marginal PMFs as X and Y but a different joint PMF, all of the other pairs of random variables must be dependent. Since X and Y are independent, $\rho_{X,Y} = 0$. For the other pairs, we must compute the covariances.

$$\mathrm{Cov}[U, V] = \mathrm{E}[UV] = (1/3)(-1) + (1/3)(-1) = -2/3, \tag{5}$$

$$\mathrm{Cov}[S, T] = \mathrm{E}[ST] = 1/6 - 1/6 + 0 + -1/6 + 1/6 = 0, \tag{6}$$

$$\mathrm{Cov}[Q, R] = \mathrm{E}[QR] = 1/12 - 1/6 - 1/6 + 1/12 = -1/6 \tag{7}$$

The correlation coefficient of U and V is

$$\rho_{U,V} = \frac{\mathrm{Cov}[U, V]}{\sqrt{\mathrm{Var}[U]}\sqrt{\mathrm{Var}[V]}} = \frac{-2/3}{\sqrt{2/3}\sqrt{2/3}} = -1 \tag{8}$$

In fact, since the marginal PMF's are the same, the denominator of the correlation coefficient will be $2/3$ in each case. The other correlation coefficients are

$$\rho_{S,T} = \frac{\mathrm{Cov}[S, T]}{2/3} = 0, \qquad \rho_{Q,R} = \frac{\mathrm{Cov}[Q, R]}{2/3} = -1/4. \tag{9}$$

(b) From Theorem 12.3, the least mean square linear estimator of U given V is

$$\hat{U}_L(V) = \rho_{U,V}\frac{\sigma_U}{\sigma_V}(V - \mathrm{E}[V]) + \mathrm{E}[U] = \rho_{U,V}V = -V. \tag{10}$$

Similarly for the other pairs, all expected values are zero and the ratio of the standard deviations is always 1. Hence,

$$\hat{X}_L(Y) = \rho_{X,Y}Y = 0, \tag{11}$$
$$\hat{S}_L(T) = \rho_{S,T}T = 0, \tag{12}$$
$$\hat{Q}_L(R) = \rho_{Q,R}R = -R/4. \tag{13}$$

From Theorem 12.3, the mean square errors are

$$e^*_L(X,Y) = \mathrm{Var}[X](1 - \rho^2_{X,Y}) = 2/3, \tag{14}$$
$$e^*_L(U,V) = \mathrm{Var}[U](1 - \rho^2_{U,V}) = 0, \tag{15}$$
$$e^*_L(S,T) = \mathrm{Var}[S](1 - \rho^2_{S,T}) = 2/3, \tag{16}$$
$$e^*_L(Q,R) = \mathrm{Var}[Q](1 - \rho^2_{Q,R}) = 5/8. \tag{17}$$

Problem 12.2.10 Solution

Following the hint, we first find the LMSE estimate $\hat{X} = aY$. We can find the optimal $\hat{a}$ using Theorem 12.3 or we can solve for $\hat{a}$ using first principles. Since $\hat{X} = a(X - Z)$, the linear MSE is

$$\begin{aligned} e(a) &= \mathrm{E}\left[(X - a(X - Z))^2\right] \\ &= \mathrm{E}\left[((1-a)X + aZ)^2\right] \\ &= \mathrm{E}\left[(1-a)^2X^2 + a(1-a)XZ + a^2Z^2\right]. \end{aligned} \tag{1}$$

Since, $\mathrm{E}[Z] = 0$, $\mathrm{E}[Z^2] = \mathrm{Var}[Z] = 1$. Also, since X and Z are independent, $\mathrm{E}[XZ] = \mathrm{E}[X]\,\mathrm{E}[Z] = 0$. These facts imply

$$e(a) = (1-a)^2\,\mathrm{E}\left[X^2\right] + a^2\,\mathrm{E}\left[Z^2\right] = (1-a)^2\,\mathrm{E}\left[X^2\right] + a^2. \tag{2}$$

We find $a = \hat{a}$ by solving

$$\left.\frac{de(a)}{da}\right|_{a=\hat{a}} = -2(1-\hat{a})\,\mathrm{E}\left[X^2\right] + 2\hat{a} = 0. \tag{3}$$

This implies

$$\hat{a} = \frac{\mathrm{E}\left[X^2\right]}{1 + \mathrm{E}\left[X^2\right]}. \tag{4}$$

From Equation (4), we see that $\hat{a} < 1$.

To prove the statement is wrong, we observe that if the best nonlinear estimator $\mathrm{E}[X|Y]$ is a linear function $\tilde{X} = \tilde{a}Y$, then the best linear estimator $\hat{a}Y$ must have $\hat{a} = \tilde{a}$. Here we have a contradiction since the claim is that $\hat{X}(Y) = Y$ is optimal while we have shown that the best linear estimator $Xhat = aY$ has an optimal value of $a = \hat{a} < 1$. In particular, the estimator $\tilde{Y} = \tilde{a}Y = Y$ has mean square error $e(1) = 1$ while the optimal linear estimator $\hat{X} = \hat{a}Y$ can be shown to have mean square error $e(\hat{a}) = \hat{a} < 1$. Thus there is definitely an error in the logic of the given argument.

The goal of this question was to identify this logical error. It is true that given $Y = y$, then $X = y + Z$ and that

$$\begin{aligned}
\mathrm{P}\left[X \leq x | Y = y\right] &= \mathrm{P}\left[Y + Z \leq x | Y = y\right] \\
&= \mathrm{P}\left[y + Z \leq x | Y = y\right] \\
&= \mathrm{P}\left[Z \leq x - y | Y = y\right]. \qquad (5)
\end{aligned}$$

The key mistake in the argument was to assume that Z and Y are independent, which wrongly implied $\mathrm{P}[Z \leq x - y | Y = y] = \mathrm{P}[Z \leq x - y] = F_Z(x - y)$. Taking a derivative with respect to x leads to the false conclusion that $f_{X|Y}(x|y) = f_Z(x - y)$.

In fact, Z and Y are *dependent*. If this is not obvious consider the special case when $\mathrm{E}[X] = 0$. In this case, Y and Z are correlated (and thus dependent) since

$$\begin{aligned}
\mathrm{Cov}\left[Y, Z\right] = \mathrm{E}\left[YZ\right] &= \mathrm{E}\left[(X - Z)Z\right] \\
&= \mathrm{E}\left[XZ\right] - \mathrm{E}\left[Z^2\right] = -\mathrm{E}\left[Z^2\right] < 0. \qquad (6)
\end{aligned}$$

As a result, we cannot use part (c) in that $f_{X|Y}(x|y) \neq f_Z(x - y)$.

Problem 12.3.1 Solution

In this case, the joint PDF of X and R is

$$\begin{aligned}
f_{X,R}(x, r) &= f_{X|R}(x|r)\, f_R(r) \\
&= \begin{cases} \frac{1}{r_0\sqrt{128\pi}} e^{-(x+40+40\log_{10} r)^2/128} & 0 \leq r \leq r_0, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)
\end{aligned}$$

From Theorem 12.5, the MAP estimate of R given $X = x$ is the value of r that maximizes $f_{X|R}(x|r) f_R(r)$. Since R has a uniform PDF over $[0, 1000]$,

$$\hat{r}_{\mathrm{MAP}}(x) = \arg\max_{0 \le r} f_{X|R}(x|r)\, f_R(r) = \arg\max_{0 \le r \le 1000} f_{X|R}(x|r) \tag{2}$$

Hence, the maximizing value of r is the same as for the ML estimate in Quiz 12.3 unless the maximizing r exceeds 1000 m. In this case, the maximizing value is $r = 1000$ m. From the solution to Quiz 12.3, the resulting ML estimator is

$$\hat{r}_{\mathrm{ML}}(x) = \begin{cases} 1000 & x < -160, \\ (0.1)10^{-x/40} & x \ge -160. \end{cases} \tag{3}$$

Problem 12.3.2 Solution

From the problem statement we know that R is an exponential random variable with expected value $1/\mu$. Therefore it has the following probability density function.

$$f_R(r) = \begin{cases} \mu e^{-\mu r} & r \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

It is also known that, given $R = r$, the number of phone calls arriving at a telephone switch, N, is a Poisson ($\alpha = rT$) random variable. So we can write the following conditional probability mass function of N given R.

$$P_{N|R}(n|r) = \begin{cases} \frac{(rT)^n e^{-rT}}{n!} & n = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(a) The minimum mean square error estimate of N given R is the conditional expected value of N given $R = r$. This is given directly in the problem statement as r.

$$\hat{N}_M(r) = \mathrm{E}\left[N|R = r\right] = rT. \tag{3}$$

(b) The maximum a posteriori estimate of N given R is simply the value of n that will maximize $P_{N|R}(n|r)$. That is,

$$\hat{n}_{MAP(r)} = \arg\max_{n \ge 0} P_{N|R}(n|r) = \arg\max_{n \ge 0} (rT)^n e^{-rT}/n! \tag{4}$$

Usually, we set a derivative to zero to solve for the maximizing value. In this case, that technique doesn't work because n is discrete.Since e^{-rT} is a common factor in the maximization, we can define $g(n) = (rT)^n/n!$ so that $\hat{n}_{MAP} = \arg\max_n g(n)$. We observe that

$$g(n) = \frac{rT}{n} g(n-1). \tag{5}$$

this implies that for $n \leq rT$, $g(n) \geq g(n-1)$. Hence the maximizing value of n is the largest n such that $n \leq rT$. That is, $\hat{n}_{MAP} = \lfloor rT \rfloor$.

(c) The maximum likelihood estimate of N given R selects the value of n that maximizes $f_{R|N=n}(r)$, the conditional PDF of R given N. When dealing with situations in which we mix continuous and discrete random variables, its often helpful to start from first principles. In this case,

$$\begin{aligned} f_{R|N}(r|n)\; dr &= \mathrm{P}\left[r < R \leq r + dr | N = n\right] \\ &= \frac{\mathrm{P}\left[r < R \leq r + dr, N = n\right]}{\mathrm{P}\left[N = n\right]} \\ &= \frac{\mathrm{P}\left[N = n | R = r\right] \mathrm{P}\left[r < R \leq r + dr\right]}{\mathrm{P}\left[N = n\right]}. \end{aligned} \tag{6}$$

In terms of PDFs and PMFs, we have

$$f_{R|N}(r|n) = \frac{P_{N|R}(n|r)\, f_R(r)}{P_N(n)}. \tag{7}$$

To find the value of n that maximizes $f_{R|N}(r|n)$, we need to find the denominator $P_N(n)$.

$$\begin{aligned} P_N(n) &= \int_{-\infty}^{\infty} P_{N|R}(n|r)\, f_R(r)\; dr \\ &= \int_0^{\infty} \frac{(rT)^n e^{-rT}}{n!} \mu e^{-\mu r}\; dr \\ &= \frac{\mu T^n}{n!(\mu + T)} \int_0^{\infty} r^n (\mu + T) e^{-(\mu+T)r}\; dr \\ &= \frac{\mu T^n}{n!(\mu + T)} \mathrm{E}\left[X^n\right]. \end{aligned} \tag{8}$$

where X is an exponential random variable with expected value $1/(\mu+T)$. There are several ways to derive the nth moment of an exponential random variable including integration by parts. In Example 9.4, the MGF is used to show that $\mathrm{E}[X^n] = n!/(\mu+T)^n$. Hence, for $n \geq 0$,

$$P_N(n) = \frac{\mu T^n}{(\mu+T)^{n+1}}. \tag{9}$$

Finally, the conditional PDF of R given N is

$$\begin{aligned} f_{R|N}(r|n) = \frac{P_{N|R}(n|r)\, f_R(r)}{P_N(n)} &= \frac{\frac{(rT)^n e^{-rT}}{n!}\mu e^{-\mu r}}{\frac{\mu T^n}{(\mu+T)^{n+1}}} \\ &= (\mu+T)\frac{[(\mu+T)r]^n e^{-(\mu+T)r}}{n!}. \end{aligned} \tag{10}$$

The ML estimate of N given R is

$$\begin{aligned} \hat{n}_{ML}(r) &= \arg\max_{n\geq 0} f_{R|N}(r|n) \\ &= \arg\max_{n\geq 0} (\mu+T)\frac{[(\mu+T)r]^n e^{-(\mu+T)r}}{n!}. \end{aligned} \tag{11}$$

This maximization is exactly the same as in the previous part except rT is replaced by $(\mu+T)r$. The maximizing value of n is $\hat{n}_{ML} = \lfloor(\mu+T)r\rfloor$.

Problem 12.3.3 Solution

Both parts (a) and (b) rely on the conditional PDF of R given $N = n$. When dealing with situations in which we mix continuous and discrete random variables, its often helpful to start from first principles.

$$\begin{aligned} f_{R|N}(r|n)\; dr &= \mathrm{P}\left[r < R \leq r + dr | N = n\right] \\ &= \frac{\mathrm{P}\left[r < R \leq r + dr, N = n\right]}{\mathrm{P}\left[N = n\right]} \\ &= \frac{\mathrm{P}\left[N = n | R = r\right] \mathrm{P}\left[r < R \leq r + dr\right]}{\mathrm{P}\left[N = n\right]}. \end{aligned} \tag{1}$$

In terms of PDFs and PMFs, we have

$$f_{R|N}(r|n) = \frac{P_{N|R}(n|r)\, f_R(r)}{P_N(n)}. \tag{2}$$

To find the value of n that maximizes $f_{R|N}(r|n)$, we need to find the denominator $P_N(n)$.

$$\begin{aligned} P_N(n) &= \int_{-\infty}^{\infty} P_{N|R}(n|r)\, f_R(r)\; dr \\ &= \int_0^{\infty} \frac{(rT)^n e^{-rT}}{n!} \mu e^{-\mu r}\, dr \\ &= \frac{\mu T^n}{n!(\mu+T)} \int_0^{\infty} r^n (\mu+T) e^{-(\mu+T)r}\, dr \\ &= \frac{\mu T^n}{n!(\mu+T)} \mathrm{E}\left[X^n\right]. \end{aligned} \tag{3}$$

where X is an exponential random variable with expected value $1/(\mu+T)$.There are several ways to derive the nth moment of an exponential random variable including integration by parts. In Example 9.4, the MGF is used to show that $\mathrm{E}[X^n] = n!/(\mu+T)^n$. Hence, for $n \geq 0$,

$$P_N(n) = \frac{\mu T^n}{(\mu+T)^{n+1}}. \tag{4}$$

Finally, the conditional PDF of R given N is

$$\begin{aligned} f_{R|N}(r|n) &= \frac{P_{N|R}(n|r)\, f_R(r)}{P_N(n)} \\ &= \frac{\frac{(rT)^n e^{-rT}}{n!} \mu e^{-\mu r}}{\frac{\mu T^n}{(\mu+T)^{n+1}}} = \frac{(\mu+T)^{n+1} r^n e^{-(\mu+T)r}}{n!}. \end{aligned} \tag{5}$$

(a) The MMSE estimate of R given $N = n$ is the conditional expected value $\mathrm{E}[R|N=n]$. Given $N = n$, the conditional PDF oF R is that of an Erlang random variable or order $n+1$. From Appendix A, we find that $\mathrm{E}[R|N=n] = (n+1)/(\mu+T)$. The MMSE estimate of R given N is

$$\hat{R}_M(N) = \mathrm{E}\left[R|N\right] = \frac{N+1}{\mu+T}. \tag{6}$$

(b) The MAP estimate of R given $N = n$ is the value of r that maximizes $f_{R|N}(r|n)$.

$$\begin{aligned}\hat{R}_{\text{MAP}}(n) &= \arg\max_{r \geq 0} f_{R|N}(r|n) \\ &= \arg\max_{r \geq 0} \frac{(\mu+T)^{n+1}}{n!} r^n e^{-(\mu+T)r}.\end{aligned} \tag{7}$$

By setting the derivative with respect to r to zero, we obtain the MAP estimate

$$\hat{R}_{\text{MAP}}(n) = \frac{n}{\mu+T}. \tag{8}$$

(c) The ML estimate of R given $N = n$ is the value of R that maximizes $P_{N|R}(n|r)$. That is,

$$\hat{R}_{\text{ML}}(n) = \arg\max_{r \geq 0} \frac{(rT)^n e^{-rT}}{n!}. \tag{9}$$

Seting the derivative with respect to r to zero yields

$$\hat{R}_{\text{ML}}(n) = n/T. \tag{10}$$

Problem 12.3.4 Solution

This problem is closely related to Example 12.7.

(a) Given $Q = q$, the conditional PMF of K is

$$P_{K|Q}(k|q) = \begin{cases} \binom{n}{k} q^k (1-q)^{n-k} & k = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The ML estimate of Q given $K = k$ is

$$\hat{q}_{\text{ML}}(k) = \arg\max_{0 \leq q \leq 1} P_{Q|K}(q|k). \tag{2}$$

Differentiating $P_{Q|K}(q|k)$ with respect to q and setting equal to zero yields

$$\begin{aligned} \frac{dP_{Q|K}(q|k)}{dq} &= \binom{n}{k}\left(kq^{k-1}(1-q)^{n-k} - (n-k)q^k(1-q)^{n-k-1}\right) \\ &= 0. \end{aligned} \tag{3}$$

The maximizing value is $q = k/n$ so that

$$\hat{Q}_{\mathrm{ML}}(K) = \frac{K}{n}. \tag{4}$$

(b) To find the PMF of K, we average over all q.

$$P_K(k) = \int_{-\infty}^{\infty} P_{K|Q}(k|q)\, f_Q(q)\; dq = \int_0^1 \binom{n}{k} q^k(1-q)^{n-k}\, dq. \tag{5}$$

We can evaluate this integral by expressing it in terms of the integral of a beta PDF. Since $\beta(k+1, n-k+1) = \frac{(n+1)!}{k!(n-k)!}$, we can write

$$\begin{aligned} P_K(k) &= \frac{1}{n+1}\int_0^1 \beta(k+1, n-k+1) q^k(1-q)^{n-k}\, dq \\ &= \frac{1}{n+1}. \end{aligned} \tag{6}$$

That is, K has the uniform PMF

$$P_K(k) = \begin{cases} 1/(n+1) & k = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Hence, $\mathrm{E}[K] = n/2$.

(c) The conditional PDF of Q given K is

$$\begin{aligned} f_{Q|K}(q|k) &= \frac{P_{K|Q}(k|q)\, f_Q(q)}{P_K(k)} \\ &= \begin{cases} \frac{(n+1)!}{k!(n-k)!} q^k(1-q)^{n-k} & 0 \le q \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{8}$$

That is, given $K = k$, Q has a beta $(k+1, n-k+1)$ PDF.

(d) The MMSE estimate of Q given $K = k$ is the conditional expectation $\mathrm{E}[Q|K = k]$ From the beta PDF described in Appendix A, $\mathrm{E}[Q|K = k] = (k+1)/(n+2)$. The MMSE estimator is

$$\hat{Q}_M(K) = \mathrm{E}\left[Q|K\right] = \frac{K+1}{n+2}. \tag{9}$$

Problem 12.4.1 Solution

(a) Since $Y_1 = X + N_1$, we see that

$$D_1 = Y_1 - X = (X + N_1) - X = N_1. \tag{1}$$

Thus $\mathrm{E}[D_1] = \mathrm{E}[N_1] = 0$ and $\mathrm{E}[D_1^2] = \mathrm{E}[N_1^2]$. Since $\mathrm{E}[N_1] = 0$, we know that $\mathrm{E}[N_1^2] = \mathrm{Var}[N_1] = 1$. That is, $\mathrm{E}[D_1^2] = 1$.

(b) Note that

$$\begin{aligned} Y_3 = \frac{Y_1}{2} + \frac{Y_2}{2} &= \frac{X + N_1}{2} + \frac{X + N_2}{2} \\ &= X + \frac{N_1}{2} + \frac{N_2}{2}. \end{aligned} \tag{2}$$

It follows that

$$D_3 = Y_3 - X = \frac{N_1}{2} + \frac{N_2}{2}. \tag{3}$$

Since N_1 and N_2 are independent Gaussian random variables, D_3 is Gaussian with expected value and variance

$$\mathrm{E}\left[D_3\right] = \frac{\mathrm{E}\left[N_1\right]}{2} + \frac{\mathrm{E}\left[N_2\right]}{2} = 0, \tag{4}$$

$$\mathrm{Var}[D_3] = \frac{\mathrm{Var}[N_1]}{4} + \frac{\mathrm{Var}[N_2]}{4} = \frac{1}{4} + \frac{4}{4} = \frac{5}{4}. \tag{5}$$

Since $\mathrm{E}[D_3] = 0$, D_3 has second moment $\mathrm{E}[D_3^2] = \mathrm{Var}[D_3^2] = 5/4$. In terms of expected squared error, the estimator Y_3 is worse than the estimator Y_1. Even though Y_3 gets to average two noisy observations Y_1 and Y_2, the large variance of N_2 makes Y_2 a lousy estimate. As a result, including Y_2 as part of the estimate Y_3 is worse than just using the estimate of Y_1 by itself.

(c) In this problem,

$$\begin{aligned} Y_4 &= aY_1 + (1-a)Y_2 \\ &= a(X+N_1) + (1-a)(X+N_2) \\ &= X + aN_1 + (1-a)N_2. \end{aligned} \tag{6}$$

This implies

$$D_4 = Y_4 - X = aN_1 + (1-a)N_2. \tag{7}$$

Thus the error D_4 is a linear combination of the errors N_1 and N_2. Since N_1 and N_2 are independent, $\mathrm{E}[D_4] = 0$ and

$$\begin{aligned} \mathrm{Var}[D_4] &= a^2 \mathrm{Var}[N_1] + (1-a)^2 \mathrm{Var}[N_2] \\ &= a^2 + 4(1-a)^2. \end{aligned} \tag{8}$$

Since $\mathrm{E}[D_4] = 0$, the second moment of the error is simply

$$\mathrm{E}\left[D_4^2\right] = \mathrm{Var}[D_4] = a^2 + 4(1-a)^2. \tag{9}$$

Since $\mathrm{E}[D_4^2]$ is a quadratic function in a, we can choose a to minimize the error. In this case, taking the derivative with respect to a and setting it equal to zero yields $2a - 8(1-a) = 0$, implying $a = 0.8$. Although the problem does not request this, it is interesting to note that for $a = 0.8$, the expected squared error is $\mathrm{E}[D_4^2] = 0.80$, which is significantly less than the error obtained by using either just Y_1 or an average of Y_1 and Y_2.

Problem 12.4.2 Solution

In this problem, we view $\mathbf{Y} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}'$ as the observation and $X = X_3$ as the variable we wish to estimate. Since $\mathrm{E}[\mathbf{X}] = \mathbf{0}$, we can use Theorem 12.6 to find the minimum mean square error estimate $\hat{X}_L(\mathbf{Y}) = \mathbf{R}_{X\mathbf{Y}} \mathbf{R}_{\mathbf{Y}}^{-1} \mathbf{Y}$.

(a) In this case,

$$\begin{aligned} \mathbf{R}_{\mathbf{Y}} = \mathrm{E}\left[\mathbf{Y}\mathbf{Y}'\right] &= \begin{bmatrix} \mathrm{E}\left[X_1^2\right] & \mathrm{E}\left[X_1X_2\right] \\ \mathrm{E}\left[X_2X_1\right] & \mathrm{E}\left[X_2^2\right] \end{bmatrix} \\ &= \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}. \end{aligned} \tag{1}$$

Similarly,

$$\begin{aligned}
\mathbf{R}_{X\mathbf{Y}} = \mathrm{E}\left[X\mathbf{Y}'\right] &= \mathrm{E}\left[X_3 \begin{bmatrix} X_1 & X_2 \end{bmatrix}\right] \\
&= \begin{bmatrix} \mathrm{E}\left[X_1 X_3\right] & \mathrm{E}\left[X_2 X_3\right] \end{bmatrix} \\
&= \begin{bmatrix} r_{13} & r_{23} \end{bmatrix} = \begin{bmatrix} 0.64 & -0.8 \end{bmatrix}. \qquad (2)
\end{aligned}$$

Thus

$$\mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1} = \begin{bmatrix} 0.64 & -0.8 \end{bmatrix} \begin{bmatrix} 25/9 & 20/9 \\ 20/9 & 25/9 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & -0.8 \end{bmatrix}. \qquad (3)$$

The optimum linear estimator of X_3 given X_1 and X_2 is

$$\hat{X}_3 = -0.8X_2. \qquad (4)$$

The fact that this estimator depends only on X_2 while ignoring X_1 is an example of a result to be proven in Problem 12.4.8.

By Theorem 12.6(b), the mean squared error of the optimal estimator is

$$\begin{aligned}
e_L^* &= \mathrm{Var}[X_3] - \mathbf{R}_{X\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X} \\
&= 1 - \begin{bmatrix} 0 & -0.8 \end{bmatrix} \begin{bmatrix} 0.64 \\ -0.8 \end{bmatrix} = 0.36. \qquad (5)
\end{aligned}$$

(b) In the previous part, we found that the optimal linear estimate of X_3 based on the observation of random variables X_1 and X_2 employed only X_2. Hence this same estimate, $\hat{X}_3 = -0.8X_2$, is the optimal linear estimate of X_3 just using X_2. (This can be derived using Theorem 12.3, if you wish to do more algebra.)

Since the estimator is the same, the mean square error is still $e_L^* = 0.36$.

Problem 12.4.3 Solution

From the problem statement, we learn for vectors $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ and $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix}'$ that

$$\mathrm{E}\left[\mathbf{X}\right] = \mathbf{0}, \qquad \mathbf{R}_{\mathbf{X}} = \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 3/4 \\ 1/2 & 3/4 & 1 \end{bmatrix}, \qquad (1)$$

and

$$\mathbf{Y} = \mathbf{A}\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{X}. \tag{2}$$

(a) Since $\mathrm{E}[\mathbf{Y}] = \mathbf{A}\,\mathrm{E}[\mathbf{X}] = \mathbf{0}$, we can apply Theorem 12.6 which states that the minimum mean square error estimate of X_1 is $\hat{X}_1(\mathbf{Y}) = \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y}$ where $\hat{\mathbf{a}} =$. The rest of the solution is just calculation. (We note that even in the case of a 3×3 matrix, its convenient to use MATLAB with `format rat` mode to perform the calculations and display the results as nice fractions.) From Theorem 8.8,

$$\begin{aligned} \mathbf{R}_{\mathbf{Y}} = \mathbf{A}\mathbf{R}_{\mathbf{X}}\mathbf{A}' &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 3/4 \\ 1/2 & 3/4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 7/2 & 3 \\ 3 & 7/2 \end{bmatrix}. \end{aligned} \tag{3}$$

In addition, since $\mathbf{R}_{X_1\mathbf{Y}} = \mathrm{E}[X_1\mathbf{Y}'] = \mathrm{E}[X_1\mathbf{X}'\mathbf{A}'] = \mathrm{E}[X_1\mathbf{X}']\mathbf{A}'$,

$$\begin{aligned} \mathbf{R}_{X_1\mathbf{Y}} &= \begin{bmatrix} \mathrm{E}\left[X_1^2\right] & \mathrm{E}\left[X_1X_2\right] & \mathrm{E}\left[X_1X_3\right] \end{bmatrix} \mathbf{A}' \\ &= \begin{bmatrix} R_{\mathbf{X}}(1,1) & R_{\mathbf{X}}(2,1) & R_{\mathbf{X}}(3,1) \end{bmatrix} \mathbf{A}' \\ &= \begin{bmatrix} 1 & 3/4 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7/4 & 5/4 \end{bmatrix}. \end{aligned} \tag{4}$$

Finally,

$$\begin{aligned} \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1} &= \begin{bmatrix} 7/4 & 5/4 \end{bmatrix} \begin{bmatrix} 14/13 & -12/13 \\ -12/13 & 14/13 \end{bmatrix} \\ &= \begin{bmatrix} 19/26 & -7/26 \end{bmatrix}. \end{aligned} \tag{5}$$

Thus the linear MMSE estimator of X_1 given $\mathbf{Y}$ is

$$\begin{aligned} \hat{X}_1(\mathbf{Y}) = \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y} &= \frac{19}{26}Y_1 - \frac{7}{26}Y_2 \\ &= 0.7308Y_1 - 0.2692Y_2. \end{aligned} \tag{6}$$

(b) By Theorem 12.6(b), the mean squared error of the optimal estimator is

$$
\begin{aligned}
e_L^* &= \operatorname{Var}[X_1] - \hat{\mathbf{a}}'\mathbf{R}_{\mathbf{Y}X_1} \\
&= R_{\mathbf{X}}(1,1) - \mathbf{R}'_{\mathbf{Y}X_1}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X_1} \\
&= 1 - \begin{bmatrix} 7/4 & 5/4 \end{bmatrix} \begin{bmatrix} 14/13 & -12/13 \\ -12/13 & 14/13 \end{bmatrix} \begin{bmatrix} 7/4 \\ 5/4 \end{bmatrix} = \frac{3}{52}. \qquad (7)
\end{aligned}
$$

(c) We can estimate random variable X_1 based on the observation of random variable Y_1 using Theorem 12.3. Note that Theorem 12.3 is just a special case of Theorem 12.7 in which the observation is a random vector. In any case, from Theorem 12.3, the optimum linear estimate is $\hat{X}_1(Y_1) = a^*Y_1 + b^*$ where

$$
a^* = \frac{\operatorname{Cov}[X_1, Y_1]}{\operatorname{Var}[Y_1]}, \qquad b^* = \mu_{X_1} - a^*\mu_{Y_1}. \qquad (8)
$$

Since $Y_1 = X_1 + X_2$, we see that

$$
\begin{aligned}
\mu_{X_1} &= \mathrm{E}[X_1] = 0, & (9)\\
\mu_{Y_1} &= \mathrm{E}[Y_1] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 0. & (10)
\end{aligned}
$$

These facts, along with $\mathbf{R}_{\mathbf{X}}$ and $\mathbf{R}_{\mathbf{Y}}$ from part (a), imply

$$
\begin{aligned}
\operatorname{Cov}[X_1, Y_1] &= \mathrm{E}[X_1 Y_1] \\
&= \mathrm{E}[X_1(X_1 + X_2)] \\
&= R_{\mathbf{X}}(1,1) + R_{\mathbf{X}}(1,2) = 7/4, & (11)\\
\operatorname{Var}[Y_1] &= \mathrm{E}\left[Y_1^2\right] = R_{\mathbf{Y}}(1,1) = 7/2 & (12)
\end{aligned}
$$

Thus

$$
\begin{aligned}
a^* &= \frac{\operatorname{Cov}[X_1, Y_1]}{\operatorname{Var}[Y_1]} = \frac{7/4}{7/2} = \frac{1}{2}, & (13)\\
b^* &= \mu_{X_1} - a^*\mu_{Y_1} = 0. & (14)
\end{aligned}
$$

Thus the optimum linear estimate of X_1 given Y_1 is

$$
\hat{X}_1(Y_1) = \frac{1}{2}Y_1. \qquad (15)
$$

From Theorem 12.3(a), the mean square error of this estimator is

$$e_L^* = \sigma_{X_1}^2(1 - \rho_{X_1,Y_1}^2). \tag{16}$$

Since X_1 and Y_1 have zero expected value, $\sigma_{X_1}^2 = R_{\mathbf{X}}(1,1) = 1$ and $\sigma_{Y_1}^2 = R_{\mathbf{Y}}(1,1) = 7/2$. Also, since $\text{Cov}[X_1, Y_1] = 7/4$, we see that

$$\rho_{X_1,Y_1} = \frac{\text{Cov}\left[X_1, Y_1\right]}{\sigma_{X_1}\sigma_{Y_1}} = \frac{7/4}{\sqrt{7/2}} = \sqrt{\frac{7}{8}}. \tag{17}$$

Thus $e_L^* = 1 - (\sqrt{7/8})^2 = 1/8$. Note that $1/8 > 3/52$. As we would expect, the estimate of X_1 based on just Y_1 has larger mean square error than the estimate based on both Y_1 and Y_2.

Problem 12.4.4 Solution

From the problem statement, we learn for vectors $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ and $\mathbf{W} = \begin{bmatrix} W_1 & W_2 \end{bmatrix}'$ that

$$\text{E}\left[\mathbf{X}\right] = \mathbf{0}, \qquad \mathbf{R}_{\mathbf{X}} = \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 3/4 \\ 1/2 & 3/4 & 1 \end{bmatrix}, \tag{1}$$

$$\text{E}\left[\mathbf{W}\right] = \mathbf{0}, \qquad \mathbf{R}_{\mathbf{W}} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}. \tag{2}$$

In addition,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{AX} + \mathbf{W} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{X} + \mathbf{W}. \tag{3}$$

(a) Since $\text{E}[\mathbf{Y}] = \mathbf{A}\,\text{E}[\mathbf{X}] = \mathbf{0}$, we can apply Theorem 12.6 which states that the minimum mean square error estimate of X_1 is

$$\hat{X}_1(\mathbf{Y}) = \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y}. \tag{4}$$

First we find $\mathbf{R}_{\mathbf{Y}}$.

$$\begin{aligned} \mathbf{R}_{\mathbf{Y}} &= \text{E}\left[\mathbf{Y}\mathbf{Y}'\right] \\ &= \text{E}\left[(\mathbf{AX} + \mathbf{W})(\mathbf{AX} + \mathbf{W})'\right] \\ &= \text{E}\left[(\mathbf{AX} + \mathbf{W})(\mathbf{X}'\mathbf{A}' + \mathbf{W}')\right] \\ &= \text{E}\left[\mathbf{AXX}'\mathbf{A}'\right] + \text{E}\left[\mathbf{WX}'\mathbf{A}\right] + \text{E}\left[\mathbf{AXW}'\right] + \text{E}\left[\mathbf{WW}'\right]. \end{aligned} \tag{5}$$

Since $\mathbf{X}$ and $\mathbf{W}$ are independent, $\mathrm{E}[\mathbf{W}\mathbf{X}'] = \mathbf{0}$ and $\mathrm{E}[\mathbf{X}\mathbf{W}'] = \mathbf{0}$. This implies

$$\begin{aligned}
\mathbf{R}_{\mathbf{Y}} &= \mathbf{A}\,\mathrm{E}\left[\mathbf{X}\mathbf{X}'\right]\mathbf{A}' + \mathrm{E}\left[\mathbf{W}\mathbf{W}'\right] \\
&= \mathbf{A}\mathbf{R}_{\mathbf{X}}\mathbf{A}' + \mathbf{R}_{\mathbf{W}} \\
&= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 3/4 \\ 1/2 & 3/4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \\
&= \begin{bmatrix} 3.6 & 3 \\ 3 & 3.6 \end{bmatrix}.
\end{aligned} \tag{6}$$

Once again, independence of $\mathbf{W}$ and X_1 yields

$$\mathbf{R}_{X_1\mathbf{Y}} = \mathrm{E}\left[X_1\mathbf{Y}'\right] = \mathrm{E}\left[X_1(\mathbf{X}' + \mathbf{W}')\mathbf{A}'\right] = \mathrm{E}\left[X_1\mathbf{X}'\right]\mathbf{A}'. \tag{7}$$

This implies

$$\begin{aligned}
\mathbf{R}_{X_1\mathbf{Y}} &= \begin{bmatrix} \mathrm{E}\left[X_1^2\right] & \mathrm{E}\left[X_1X_2\right] & \mathrm{E}\left[X_1X_3\right] \end{bmatrix}\mathbf{A}' \\
&= \begin{bmatrix} R_{\mathbf{X}}(1,1) & R_{\mathbf{X}}(2,1) & R_{\mathbf{X}}(3,1) \end{bmatrix}\mathbf{A}' \\
&= \begin{bmatrix} 1 & 3/4 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7/4 & 5/4 \end{bmatrix}.
\end{aligned} \tag{8}$$

Putting these facts together, we find that

$$\mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1} = \begin{bmatrix} 7/4 & 5/4 \end{bmatrix} \begin{bmatrix} 10/11 & -25/33 \\ -25/33 & 10/11 \end{bmatrix} = \begin{bmatrix} \frac{85}{132} & -\frac{25}{132} \end{bmatrix}. \tag{9}$$

Thus the linear MMSE estimator of X_1 given $\mathbf{Y}$ is

$$\begin{aligned}
\hat{X}_1(\mathbf{Y}) = \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y} &= \frac{85}{132}Y_1 - \frac{25}{132}Y_2 \\
&= 0.6439Y_1 - 0.1894Y_2.
\end{aligned} \tag{10}$$

(b) By Theorem 12.6(b), the mean squared error of the optimal estimator is

$$\begin{aligned}
e_L^* &= \mathrm{Var}[X_1] - \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}'_{X_1\mathbf{Y}} \\
&= R_{\mathbf{X}}(1,1) - \mathbf{R}_{X_1\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}'_{X_1\mathbf{Y}} \\
&= 1 - \begin{bmatrix} 7/4 & 5/4 \end{bmatrix} \begin{bmatrix} 10/11 & -25/33 \\ -25/33 & 10/11 \end{bmatrix} \begin{bmatrix} 7/4 \\ 5/4 \end{bmatrix} \\
&= \frac{29}{264} = 0.1098. \qquad (11)
\end{aligned}$$

In Problem 12.4.3, we solved essentialy the same problem but the observations $\mathbf{Y}$ were not subjected to the additive noise $\mathbf{W}$. In comparing the estimators, we see that the additive noise perturbs the estimator somewhat but not dramatically because the correaltion structure of $\mathbf{X}$ and the mapping $\mathbf{A}$ from $\mathbf{X}$ to $\mathbf{Y}$ remains unchanged. On the other hand, in the noiseless case, the resulting mean square error was about half as much, $3/52 = 0.0577$ versus 0.1098.

(c) We can estimate random variable X_1 based on the observation of random variable Y_1 using Theorem 12.3. Note that Theorem 12.3 is a special case of Theorem 12.7 in which the observation is a random vector. In any case, from Theorem 12.3, the optimum linear estimate is $\hat{X}_1(Y_1) = a^* Y_1 + b^*$ where

$$a^* = \frac{\mathrm{Cov}\left[X_1, Y_1\right]}{\mathrm{Var}[Y_1]}, \qquad b^* = \mu_{X_1} - a^* \mu_{Y_1}. \qquad (12)$$

Since $\mathrm{E}[X_i] = \mu_{X_i} = 0$ and $Y_1 = X_1 + X_2 + W_1$, we see that

$$\mu_{Y_1} = \mathrm{E}\left[Y_1\right] = \mathrm{E}\left[X_1\right] + \mathrm{E}\left[X_2\right] + \mathrm{E}\left[W_1\right] = 0. \qquad (13)$$

These facts, along with independence of X_1 and W_1, imply

$$\begin{aligned}
\mathrm{Cov}\left[X_1, Y_1\right] = \mathrm{E}\left[X_1 Y_1\right] &= \mathrm{E}\left[X_1(X_1 + X_2 + W_1)\right] \\
&= R_{\mathbf{X}}(1,1) + R_{\mathbf{X}}(1,2) = 7/4. \qquad (14)
\end{aligned}$$

In addition, using $\mathbf{R}_{\mathbf{Y}}$ from part (a), we see that

$$\mathrm{Var}[Y_1] = \mathrm{E}\left[Y_1^2\right] = R_{\mathbf{Y}}(1,1) = 3.6. \qquad (15)$$

Thus

$$a^* = \frac{\text{Cov}[X_1, Y_1]}{\text{Var}[Y_1]} = \frac{7/4}{3.6} = \frac{35}{72}, \qquad b^* = \mu_{X_1} - a^*\mu_{Y_1} = 0. \tag{16}$$

Thus the optimum linear estimate of X_1 given Y_1 is

$$\hat{X}_1(Y_1) = \frac{35}{72} Y_1. \tag{17}$$

From Theorem 12.3(a), the mean square error of this estimator is

$$e_L^* = \sigma_{X_1}^2 (1 - \rho_{X_1,Y_1}^2). \tag{18}$$

Since X_1 and Y_1 have zero expected value, $\sigma_{X_1}^2 = R_{\mathbf{X}}(1,1) = 1$ and $\sigma_{Y_1}^2 = R_{\mathbf{Y}}(1,1) = 3.6$. Also, since $\text{Cov}[X_1, Y_1] = 7/4$, we see that

$$\rho_{X_1,Y_1} = \frac{\text{Cov}[X_1, Y_1]}{\sigma_{X_1}\sigma_{Y_1}} = \frac{7/4}{\sqrt{3.6}} = \frac{\sqrt{490}}{24}. \tag{19}$$

Thus $e_L^* = 1 - (490/24^2) = 0.1493$. As we would expect, the estimate of X_1 based on just Y_1 has larger mean square error than the estimate based on both Y_1 and Y_2.

Problem 12.4.5 Solution

The key to this problem is to write $\mathbf{Y}$ in terms of $\mathbf{Q}$. First we observe that

$$Y_1 = q_0 + 1q_1 + 1^2 q_2 + Z_1, \tag{1}$$
$$Y_2 = q_0 + 2q_1 + 2^2 q_2 + Z_2, \tag{2}$$
$$\vdots \qquad \vdots$$
$$Y_n = q_0 + nq_1 + n^2 q_2 + Z_n. \tag{3}$$

In terms of the vector $\mathbf{Q}$, we can write

$$\mathbf{Y} = \underbrace{\begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix}}_{\mathbf{K}_n} \mathbf{Q} + \mathbf{Z} = \mathbf{K}_n \mathbf{Q} + \mathbf{Z}. \tag{4}$$

From the problem statement we know that $\mathrm{E}[\mathbf{Q}] = \mathbf{0}$, $\mathrm{E}[\mathbf{Z}] = \mathbf{0}$, $\mathbf{R_Q} = \mathbf{I}$, and $\mathbf{R_Z} = \mathbf{I}$. Applying Theorem 12.8 as expressed in Equation (12.77), we obtain

$$\hat{\mathbf{Q}}_L(\mathbf{Y}) = \mathbf{R_{QY}}\mathbf{R_Y}^{-1}\mathbf{Y}. \tag{5}$$

Since $\mathbf{Q}$ and the noise $\mathbf{Z}$ are independent,

$$\mathrm{E}\left[\mathbf{Q}\mathbf{Z}'\right] = \mathrm{E}\left[\mathbf{Q}\right]\mathrm{E}\left[\mathbf{Z}'\right] = \mathbf{0}. \tag{6}$$

This implies

$$\begin{aligned}
\mathbf{R_{QY}} &= \mathrm{E}\left[\mathbf{Q}\mathbf{Y}'\right] \\
&= \mathrm{E}\left[\mathbf{Q}(\mathbf{K}_n\mathbf{Q} + \mathbf{Z})'\right] \\
&= \mathrm{E}\left[\mathbf{Q}\mathbf{Q}'\mathbf{K}_n' + \mathbf{Q}\mathbf{Z}'\right] = \mathbf{R_Q}\mathbf{K}_n'.
\end{aligned} \tag{7}$$

Again using (6), we have that

$$\begin{aligned}
\mathbf{R_Y} &= \mathrm{E}\left[\mathbf{Y}\mathbf{Y}'\right] \\
&= \mathrm{E}\left[(\mathbf{K}_n\mathbf{Q} + \mathbf{Z})(\mathbf{K}_n\mathbf{Q} + \mathbf{Z})'\right] \\
&= \mathrm{E}\left[(\mathbf{K}_n\mathbf{Q} + \mathbf{Z})(\mathbf{Q}'\mathbf{K}_n' + \mathbf{Z}')\right] \\
&= \mathbf{K}_n\,\mathrm{E}\left[\mathbf{Q}\mathbf{Q}'\right]\mathbf{K}_n' + \mathbf{K}_n\,\mathrm{E}\left[\mathbf{Q}\mathbf{Z}'\right] + \mathrm{E}\left[\mathbf{Z}\mathbf{Q}'\right]\mathbf{K}_n' + \mathrm{E}\left[\mathbf{Z}\mathbf{Z}'\right] \\
&= \mathbf{K}_n\mathbf{K}_n' + \mathbf{I}.
\end{aligned} \tag{8}$$

It follows that

$$\hat{\mathbf{Q}} = \mathbf{R_{QY}}\mathbf{R_Y}^{-1}\mathbf{Y} = \mathbf{K}_n'(\mathbf{K}_n\mathbf{K}_n' + \mathbf{I})^{-1}\mathbf{Y}. \tag{9}$$

Problem 12.4.6 Solution

From the problem statement, we learn for vectors $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ and $\mathbf{W} = \begin{bmatrix} W_1 & W_2 \end{bmatrix}'$ that

$$\mathrm{E}\left[\mathbf{X}\right] = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{R_X} = \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 1/2 \\ 1/2 & 3/4 & 1 \end{bmatrix}, \tag{1}$$

$$\mathrm{E}\left[\mathbf{W}\right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mathbf{R_W} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}. \tag{2}$$

In addition, from Theorem 8.7,

$$
\begin{aligned}
\mathbf{C_X} &= \mathbf{R_X} - \boldsymbol{\mu}_\mathbf{X}\boldsymbol{\mu}'_\mathbf{X} \\
&= \begin{bmatrix} 1 & 3/4 & 1/2 \\ 3/4 & 1 & 1/2 \\ 1/2 & 3/4 & 1 \end{bmatrix} - \begin{bmatrix} -0.1 \\ 0 \\ 0.1 \end{bmatrix} \begin{bmatrix} -0.1 & 0 & 0.1 \end{bmatrix} \\
&= \begin{bmatrix} 0.99 & 0.75 & 0.51 \\ 0.75 & 1.0 & 0.75 \\ 0.51 & 0.75 & 0.99 \end{bmatrix}, \qquad (3)
\end{aligned}
$$

and

$$
\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{AX} + \mathbf{W} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{X} + \mathbf{W}. \qquad (4)
$$

(a) Since $\mathrm{E}[\mathbf{X}]$ is nonzero, we use Theorem 12.7, in the form of Equation (12.75), which states that the minimum mean square error estimate of X_1 is

$$
\hat{X}_1(\mathbf{Y}) = \mathbf{C}_{X_1\mathbf{Y}}\mathbf{C}_\mathbf{Y}^{-1}\mathbf{Y} + \hat{b} \qquad (5)
$$

with

$$
\hat{b} = \mathrm{E}[X_1] - \mathbf{C}_{X_1\mathbf{Y}}\mathbf{C}_\mathbf{Y}^{-1}\mathrm{E}[\mathbf{Y}]. \qquad (6)
$$

First we find $\mathbf{C_Y}$. Since $\mathrm{E}[\mathbf{Y}] = \mathbf{A}\,\mathrm{E}[\mathbf{X}] + \mathrm{E}[\mathbf{W}] = \mathbf{A}\,\mathrm{E}[\mathbf{X}]$,

$$
\begin{aligned}
\mathbf{C_Y} &= \mathrm{E}\left[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])'\right] \\
&= \mathrm{E}\left[(\mathbf{A}(\mathbf{X} - \mathrm{E}[\mathbf{X}]) + \mathbf{W})(\mathbf{A}(\mathbf{X} - \mathrm{E}[\mathbf{X}]) + \mathbf{W})'\right] \\
&= \mathrm{E}\left[(\mathbf{A}(\mathbf{X} - \mathrm{E}[\mathbf{X}]) + \mathbf{W})((\mathbf{X} - \mathrm{E}[\mathbf{X}])'\mathbf{A}' + \mathbf{W}')\right] \\
&= \mathbf{A}\,\mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])'\right]\mathbf{A}' \\
&\qquad + \mathrm{E}\left[\mathbf{W}(\mathbf{X} - \mathrm{E}[\mathbf{X}])'\right]\mathbf{A}' + \mathbf{A}\,\mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])\mathbf{W}'\right] \\
&\qquad\qquad + \mathrm{E}\left[\mathbf{W}\mathbf{W}'\right]. \qquad (7)
\end{aligned}
$$

Since $\mathbf{X}$ and $\mathbf{W}$ are independent,

$$
\mathrm{E}\left[\mathbf{W}(\mathbf{X} - \mathrm{E}[\mathbf{X}])'\right] = \mathbf{0}, \qquad \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])\mathbf{W}'\right] = \mathbf{0}. \qquad (8)
$$

This implies

$$\begin{aligned}
\mathbf{C_Y} &= \mathbf{A}\mathbf{C_X}\mathbf{A}' + \mathbf{R_W} \\
&= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.99 & 0.75 & 0.51 \\ 0.75 & 1.0 & 0.75 \\ 0.51 & 0.75 & 0.99 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \\
&= \begin{bmatrix} 3.59 & 3.01 \\ 3.01 & 3.59 \end{bmatrix}.
\end{aligned} \tag{9}$$

Once again, independence of $\mathbf{W}$ and X_1 yields $\mathrm{E}[\mathbf{W}(X_1 - \mathrm{E}[X_1])] = \mathbf{0}$, implying

$$\begin{aligned}
\mathbf{C}_{X_1\mathbf{Y}} &= \mathrm{E}\left[(X_1 - \mathrm{E}\left[X_1\right])(\mathbf{Y}' - \mathrm{E}\left[\mathbf{Y}'\right])\right] \\
&= \mathrm{E}\left[(X_1 - \mathrm{E}\left[X_1\right])((\mathbf{X}' - \mathrm{E}\left[\mathbf{X}'\right])\mathbf{A}' + \mathbf{W}')\right] \\
&= \mathrm{E}\left[(X_1 - \mathrm{E}\left[X_1\right])(\mathbf{X}' - \mathrm{E}\left[\mathbf{X}'\right])\right]\mathbf{A}' \\
&= \begin{bmatrix} C_{\mathbf{X}}(1,1) & C_{\mathbf{X}}(2,1) & C_{\mathbf{X}}(3,1) \end{bmatrix} \mathbf{A}' \\
&= \begin{bmatrix} 0.99 & 0.75 & 0.51 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.74 & 1.26 \end{bmatrix}.
\end{aligned} \tag{10}$$

Finally,

$$\begin{aligned}
\mathbf{C}_{X_1\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1} &= \begin{bmatrix} 1.74 & 1.26 \end{bmatrix} \begin{bmatrix} 0.9378 & -0.7863 \\ -0.7863 & 0.9378 \end{bmatrix} \\
&= \begin{bmatrix} 0.6411 & -0.1865 \end{bmatrix}.
\end{aligned} \tag{11}$$

Next, we find that

$$\mathrm{E}\left[\mathbf{Y}\right] = \mathbf{A}\,\mathrm{E}\left[\mathbf{X}\right] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -0.1 \\ 0 \\ 0.1 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}. \tag{12}$$

This implies

$$\begin{aligned}
\hat{b} &= \mathrm{E}\left[X_1\right] - \mathbf{C}_{X_1\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1}\,\mathrm{E}\left[\mathbf{Y}\right] \\
&= -0.1 - \begin{bmatrix} 0.6411 & -0.1865 \end{bmatrix} \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix} = -0.0172.
\end{aligned} \tag{13}$$

Thus the linear MMSE estimator of X_1 given $\mathbf{Y}$ is

$$\begin{aligned}\hat{X}_1(\mathbf{Y}) &= \mathbf{C}_{X_1\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{Y} + \hat{b} \\ &= 0.6411Y_1 - 0.1865Y_2 + 0.0172. \end{aligned} \tag{14}$$

We note that although $\mathrm{E}[X] = 0.1$, the estimator's offset is only 0.0172. This is because the change in $\mathrm{E}[X]$ is also included in the change in $\mathrm{E}[Y]$.

(b) By Theorem 12.7(b), the mean square error of the optimal estimator is

$$\begin{aligned} e_L^* &= \mathrm{Var}[X_1] - \hat{\mathbf{a}}'\mathbf{C}_{\mathbf{Y}X_1} \\ &= C_{\mathbf{X}}(1,1) - \mathbf{C}'_{\mathbf{Y}X_1}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{Y}X_1} \\ &= 0.99 - \begin{bmatrix} 1.7400 & 1.2600 \end{bmatrix} \begin{bmatrix} 0.9378 & -0.7863 \\ -0.7863 & 0.9378 \end{bmatrix} \begin{bmatrix} 1.7400 \\ 1.2600 \end{bmatrix} \\ &= 0.1096. \end{aligned} \tag{15}$$

(c) We can estimate random variable X_1 based on the observation of random variable Y_1 using Theorem 12.3. Note that Theorem 12.3 is a special case of Theorem 12.7 in which the observation is a random vector. In any case, from Theorem 12.3, the optimum linear estimate is $\hat{X}_1(Y_1) = a^*Y_1 + b^*$ where

$$a^* = \frac{\mathrm{Cov}\left[X_1, Y_1\right]}{\mathrm{Var}[Y_1]}, \qquad b^* = \mu_{X_1} - a^*\mu_{Y_1}. \tag{16}$$

First we note that $\mathrm{E}[X_1] = \mu_{X_1} = -0.1$. Second, since

$$Y_1 = X_1 + X_2 + W_1, \qquad\qquad \mathrm{E}\left[W_1\right] = 0, \tag{17}$$

we see that

$$\begin{aligned} \mu_{Y_1} = \mathrm{E}\left[Y_1\right] &= \mathrm{E}\left[X_1\right] + \mathrm{E}\left[X_2\right] + \mathrm{E}\left[W_1\right] \\ &= \mathrm{E}\left[X_1\right] + \mathrm{E}\left[X_2\right] \\ &= -0.1. \end{aligned} \tag{18}$$

These facts, along with independence of X_1 and W_1, imply

$$\begin{aligned}
\operatorname{Cov}[X_1, Y_1] &= \mathrm{E}\left[(X_1 - \mathrm{E}[X_1])(Y_1 - \mathrm{E}[Y_1])\right] \\
&= \mathrm{E}\left[(X_1 - \mathrm{E}[X_1])((X_1 - \mathrm{E}[X_1]) + (X_2 - EX_2) + W_1)\right] \\
&= C_{\mathbf{X}}(1,1) + C_{\mathbf{X}}(1,2) = 1.74. \qquad (19)
\end{aligned}$$

In addition, using $\mathbf{C_Y}$ from part (a), we see that

$$\operatorname{Var}[Y_1] = \mathrm{E}\left[Y_1^2\right] = C_{\mathbf{Y}}(1,1) = 3.59. \qquad (20)$$

Thus

$$a^* = \frac{\operatorname{Cov}[X_1, Y_1]}{\operatorname{Var}[Y_1]} = \frac{1.74}{3.59} = 0.4847, \qquad (21)$$

$$b^* = \mu_{X_1} - a^* \mu_{Y_1} = -0.0515. \qquad (22)$$

Thus the optimum linear estimate of X_1 given Y_1 is

$$\hat{X}_1(Y_1) = 0.4847 Y_1 - 0.0515. \qquad (23)$$

From Theorem 12.3(a), the mean square error of this estimator is

$$e_L^* = \sigma_{X_1}^2 (1 - \rho_{X_1,Y_1}^2). \qquad (24)$$

Note that $\sigma_{X_1}^2 = R_{\mathbf{X}}(1,1) = 0.99$ and $\sigma_{Y_1}^2 = R_{\mathbf{Y}}(1,1) = 3.59$. Also, since $\operatorname{Cov}[X_1, Y_1] = 1.74$, we see that

$$\rho_{X_1,Y_1} = \frac{\operatorname{Cov}[X_1, Y_1]}{\sigma_{X_1}\sigma_{Y_1}} = \frac{1.74}{\sqrt{(0.99)(3.59)}} = 0.923. \qquad (25)$$

Thus $e_L^* = 0.99(1 - (0.923)^2) = 0.1466$. As we would expect, the estimate of X_1 based on just Y_1 has larger mean square error than the estimate based on both Y_1 and Y_2.

Problem 12.4.7 Solution

From Theorem 12.6, we know that the minimum mean square error estimate of X given $\mathbf{Y}$ is $\hat{X}_L(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y}$, where $\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X}$. In this problem, $\mathbf{Y}$ is simply a scalar Y and $\hat{\mathbf{a}}$ is a scalar $\hat{a}$. Since $\mathrm{E}[Y] = 0$,

$$\mathbf{R_Y} = \mathrm{E}\left[\mathbf{Y}\mathbf{Y}'\right] = \mathrm{E}\left[Y^2\right] = \sigma_Y^2. \qquad (1)$$

Similarly,

$$\mathbf{R}_{\mathbf{Y}X} = \mathrm{E}\left[\mathbf{Y}X\right] = \mathrm{E}\left[YX\right] = \mathrm{Cov}\left[X,Y\right]. \tag{2}$$

It follows that

$$\hat{a} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X} = \left(\sigma_Y^2\right)^{-1}\mathrm{Cov}\left[X,Y\right] = \frac{\sigma_X}{\sigma_Y}\frac{\mathrm{Cov}\left[X,Y\right]}{\sigma_X\sigma_Y} = \frac{\sigma_X}{\sigma_Y}\rho_{X,Y}. \tag{3}$$

Problem 12.4.8 Solution

For this problem, let $\mathbf{Y} = \begin{bmatrix} X_1 & X_2 & \cdots & _{n-1}\end{bmatrix}'$ and let $X = X_n$. Since $\mathrm{E}[\mathbf{Y}] = 0$ and $\mathrm{E}[X] = 0$, Theorem 12.6 tells us that the minimum mean square linear estimate of X given $\mathbf{Y}$ is $\hat{X}_n(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y}$, where $\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{R}_{\mathbf{Y}X}$. This implies that $\hat{\mathbf{a}}$ is the solution to

$$\mathbf{R}_{\mathbf{Y}}\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}X}. \tag{1}$$

Note that

$$\mathbf{R}_{\mathbf{Y}} = \mathrm{E}\left[\mathbf{Y}\mathbf{Y}'\right] = \begin{bmatrix} 1 & c & \cdots & c^{n-2} \\ c & c^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c \\ c^{n-2} & \cdots & c & 1 \end{bmatrix},$$

$$\mathbf{R}_{\mathbf{Y}X} = \mathrm{E}\left[\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \end{bmatrix} X_n\right] = \begin{bmatrix} c^{n-1} \\ c^{n-2} \\ \vdots \\ c \end{bmatrix}. \tag{2}$$

We see that the last column of $c\mathbf{R}_{\mathbf{Y}}$ equals $\mathbf{R}_{\mathbf{Y}X}$. Equivalently, if $\hat{\mathbf{a}} = \begin{bmatrix} 0 & \cdots & 0 & c \end{bmatrix}'$, then $\mathbf{R}_{\mathbf{Y}}\hat{\mathbf{a}} = \mathbf{R}_{\mathbf{Y}X}$. It follows that the optimal linear estimator of X_n given $\mathbf{Y}$ is

$$\hat{X}_n(\mathbf{Y}) = \hat{\mathbf{a}}'\mathbf{Y} = cX_{n-1}, \tag{3}$$

which completes the proof of the claim.

The mean square error of this estimate is

$$\begin{aligned} e_L^* &= \mathrm{E}\left[(X_n - cX_{n-1})^2\right] \\ &= R_{\mathbf{X}}(n,n) - cR_{\mathbf{X}}(n,n-1) - cR_{\mathbf{X}}(n-1,n) + c^2 R_{\mathbf{X}}(n-1,n-1) \\ &= 1 - 2c^2 + c^2 = 1 - c^2. \end{aligned} \tag{4}$$

When c is close to 1, X_{n-1} and X_n are highly correlated and the estimation error will be small.

Comment: Correlation matrices with this structure arise naturally in the study of wide sense stationary random sequences.

Problem 12.4.9 Solution

(a) In this case, we use the observation $\mathbf{Y}$ to estimate each X_i. Since $\mathrm{E}[X_i] = 0$,

$$\mathrm{E}\left[\mathbf{Y}\right] = \sum_{j=1}^{k} \mathrm{E}\left[X_j\right] \sqrt{p_j}\mathbf{S}_j + \mathrm{E}\left[\mathbf{N}\right] = \mathbf{0}. \tag{1}$$

Thus, Theorem 12.6 tells us that the MMSE linear estimate of X_i is $\hat{X}_i(\mathbf{Y}) = \mathbf{R}_{X_i\mathbf{Y}}\mathbf{R}_{\mathbf{Y}}^{-1}\mathbf{Y}$. First we note that

$$\mathbf{R}_{X_i\mathbf{Y}} = \mathrm{E}\left[X_i\mathbf{Y}'\right] = \mathrm{E}\left[X_i\left(\sum_{j=1}^{k} X_j\sqrt{p_j}\mathbf{S}'_j + \mathbf{N}'\right)\right] \tag{2}$$

Since $\mathbf{N}$ and X_i are independent, $\mathrm{E}[X_i\mathbf{N}'] = \mathrm{E}[X_i]\,\mathrm{E}[\mathbf{N}'] = \mathbf{0}$. Because X_i and X_j are independent for $i \neq j$, $\mathrm{E}[X_iX_j] = \mathrm{E}[X_i]\,\mathrm{E}[X_j] = 0$ for $i \neq j$. In addition, $\mathrm{E}[X_i^2] = 1$, and it follows that

$$\mathbf{R}_{X_i\mathbf{Y}} = \sum_{j=1}^{k} \mathrm{E}\left[X_iX_j\right] \sqrt{p_j}\mathbf{S}'_j + \mathrm{E}\left[X_i\mathbf{N}'\right] = \sqrt{p_i}\mathbf{S}'_i. \tag{3}$$

For the same reasons,

$$
\begin{aligned}
\mathbf{R}_{\mathbf{Y}} &= \mathrm{E}\left[\mathbf{Y}\mathbf{Y}'\right] \\
&= \mathrm{E}\left[\left(\sum_{j=1}^{k} \sqrt{p_j} X_j \mathbf{S}_j + \mathbf{N}\right)\left(\sum_{l=1}^{k} \sqrt{p_l} X_l \mathbf{S}'_l + \mathbf{N}'\right)\right] \\
&= \sum_{j=1}^{k}\sum_{l=1}^{k} \sqrt{p_j p_l}\, \mathrm{E}\left[X_j X_l\right] \mathbf{S}_j \mathbf{S}'_l \\
&\quad + \sum_{j=1}^{k} \sqrt{p_j}\, \underbrace{\mathrm{E}\left[X_j \mathbf{N}\right]}_{=0} \mathbf{S}_j + \sum_{l=1}^{k} \sqrt{p_l}\, \underbrace{\mathrm{E}\left[X_l \mathbf{N}'\right]}_{=0} \mathbf{S}'_l + \mathrm{E}\left[\mathbf{N}\mathbf{N}'\right] \\
&= \sum_{j=1}^{k} p_j \mathbf{S}_j \mathbf{S}'_j + \sigma^2 \mathbf{I}.
\end{aligned} \tag{4}
$$

Now we use a linear algebra identity. For a matrix $\mathbf{S}$ with columns $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_k$, and a diagonal matrix $\mathbf{P} = \mathrm{diag}[p_1, p_2, \ldots, p_k]$,

$$
\sum_{j=1}^{k} p_j \mathbf{S}_j \mathbf{S}'_j = \mathbf{S}\mathbf{P}\mathbf{S}'. \tag{5}
$$

Although this identity may be unfamiliar, it is handy in manipulating correlation matrices. (Also, if this is unfamiliar, you may wish to work out an example with $k = 2$ vectors of length 2 or 3.) Thus,

$$
\mathbf{R}_{\mathbf{Y}} = \mathbf{S}\mathbf{P}\mathbf{S}' + \sigma^2 \mathbf{I}, \tag{6}
$$

and

$$
\mathbf{R}_{X_i \mathbf{Y}} \mathbf{R}_{\mathbf{Y}}^{-1} = \sqrt{p_i}\, \mathbf{S}'_i \left(\mathbf{S}\mathbf{P}\mathbf{S}' + \sigma^2 \mathbf{I}\right)^{-1}. \tag{7}
$$

Recall that if $\mathbf{C}$ is symmetric, then $\mathbf{C}^{-1}$ is also symmetric. This implies the MMSE estimate of X_i given $\mathbf{Y}$ is

$$
\hat{X}_i(\mathbf{Y}) = \mathbf{R}_{X_i \mathbf{Y}} \mathbf{R}_{\mathbf{Y}}^{-1} \mathbf{Y} = \sqrt{p_i}\, \mathbf{S}'_i \left(\mathbf{S}\mathbf{P}\mathbf{S}' + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{Y}. \tag{8}
$$

(b) We observe that $\mathbf{V} = (\mathbf{SPS}' + \sigma^2\mathbf{I})^{-1}\mathbf{Y}$ is a vector that does not depend on which bit X_i that we want to estimate. Since $\hat{X}_i = \sqrt{p_i}\mathbf{S}'_i\mathbf{V}$, we can form the vector of estimates

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_k \end{bmatrix} = \begin{bmatrix} \sqrt{p_1}\mathbf{S}'_1\mathbf{V} \\ \vdots \\ \sqrt{p_k}\mathbf{S}'_k\mathbf{V} \end{bmatrix} = \begin{bmatrix} \sqrt{p_1} & & \\ & \ddots & \\ & & \sqrt{p_k} \end{bmatrix} \begin{bmatrix} \mathbf{S}'_1 \\ \vdots \\ \mathbf{S}'_k \end{bmatrix} \mathbf{V}$$

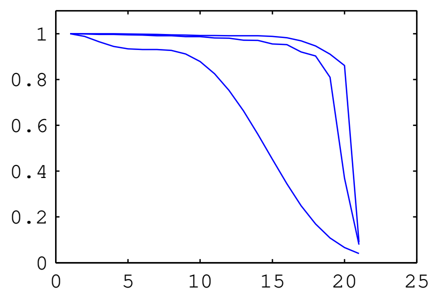
$$= \mathbf{P}^{1/2}\mathbf{S}'\mathbf{V}$$

$$= \mathbf{P}^{1/2}\mathbf{S}'\left(\mathbf{SPS}' + \sigma^2\mathbf{I}\right)^{-1}\mathbf{Y}. \tag{9}$$

Problem 12.5.1 Solution

This problem can be solved using the function `mse` defined in Example 12.10. All we need to do is define the correlation structure of the vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_{21} \end{bmatrix}'$. Just as in Example 12.10, we do this by defining just the first row of the correlation matrix. Here are the commands we need, and the resulting plot.

```
r1=sinc(0.1*(0:20)); mse(r1);
hold on;
r5=sinc(0.5*(0:20)); mse(r5);
r9=sinc(0.9*(0:20)); mse(r9);
```

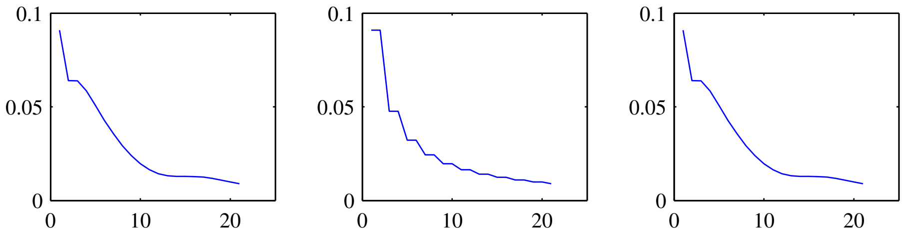


Although the plot lacks labels, there are three curves for the mean square error MSE(n) corresponding to $\phi_0 \in \{0.1, 0.5, 0.9\}$. Keep in mind that MSE($n$) is the MSE of the linear estimate of X_{21} using random variables $X_1, \ldots, X_n$.

If you run the commands, you'll find that the $\phi_0 = 0.1$ yields the lowest mean square error while $\phi_0 = 0.9$ results in the highest mean square error. When $\phi_0 = 0.1$, random variables X_n for $n = 10, 11, \ldots, 20$ are increasingly correlated with X_{21}. The result is that the MSE starts to decline rapidly for $n > 10$. As ϕ_0 increases, fewer observations X_n are correlated with X_{21}. The result is the MSE is simply worse as ϕ_0 increases. For example, when $\phi_0 = 0.9$, even X_{20} has only a small correlation with X_{21}. We only get a good estimate of X_{21} at time $n = 21$ when we observe $X_{21} + W_{21}$.

Problem 12.5.2 Solution

This problem can be solved using the function `mse` defined in Example 12.10. All we need to do is define the correlation structure of the vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_{21} \end{bmatrix}'$. Just as in Example 12.10, we do this by defining just the first row `r` of the correlation matrix. Here are the commands we need, and the resulting plots.




```
n=0:20;
r1=cos(0.1*pi*n);
mse(r1);
```

```
n=0:20;
r5=cos(0.5*pi*n);
mse(r5);
```

```
n=0:20;
r9=cos(0.9*pi*n);
mse(r9);
```

All three cases report similar results for the mean square error (MSE). the reason is that in all three cases, X_1 and X_{21} are completely correlated; that is, $\rho_{X_1,X_{21}} = 1$. As a result, $X_1 = X_{21}$ so that at time $n = 1$, the observation is

$$Y_1 = X_1 + W_1 = X_{21} + W_1. \tag{1}$$

The MSE at time $n = 1$ is 0.1, corresponding to the variance of the additive noise. Subsequent improvements in the estimates are the result of making other measurements of the form $Y_n = X_n + W_n$ where X_n is highly correlated with X_{21}. The result is a form of averaging of the additive noise, which effectively reduces its variance.

The choice of ϕ_0 changes the values of n for which X_n and X_{21} are highly correlated. However, in all three cases, there are several such values of n and the result in all cases is an improvement in the MSE due to averaging the additive noise.

Problem 12.5.3 Solution

The solution to this problem is almost the same as the solution to Example 12.10, except perhaps the MATLAB code is somewhat simpler. As in the example, let

$\mathbf{W}^{(n)}$, $\mathbf{X}^{(n)}$, and $\mathbf{Y}^{(n)}$ denote the vectors, consisting of the first n components of $\mathbf{W}$, $\mathbf{X}$, and $\mathbf{Y}$. Just as in Examples 12.8 and 12.10, independence of $\mathbf{X}^{(n)}$ and $\mathbf{W}^{(n)}$ implies that the correlation matrix of $\mathbf{Y}^{(n)}$ is

$$\mathbf{R}_{\mathbf{Y}^{(n)}} = \mathrm{E}\left[(\mathbf{X}^{(n)} + \mathbf{W}^{(n)})(\mathbf{X}^{(n)} + \mathbf{W}^{(n)})'\right] = \mathbf{R}_{\mathbf{X}^{(n)}} + \mathbf{R}_{\mathbf{W}^{(n)}} \tag{1}$$

Note that $\mathbf{R}_{\mathbf{X}^{(n)}}$ and $\mathbf{R}_{\mathbf{W}^{(n)}}$ are the $n \times n$ upper-left submatrices of $\mathbf{R}_{\mathbf{X}}$ and $\mathbf{R}_{\mathbf{W}}$. In addition,

$$\mathbf{R}_{\mathbf{Y}^{(n)}X} = \mathrm{E}\left[\begin{bmatrix} X_1 + W_1 \\ \vdots \\ X_n + W_n \end{bmatrix} X_1\right] = \begin{bmatrix} r_0 \\ \vdots \\ r_{n-1} \end{bmatrix}. \tag{2}$$

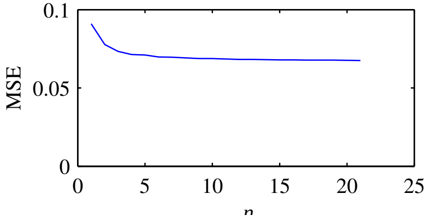
Compared to the solution of Example 12.10, the only difference in the solution is in the reversal of the vector $\mathbf{R}_{\mathbf{Y}^{(n)}X}$. The optimal filter based on the first n observations is $\hat{\mathbf{a}}^{(n)} = \mathbf{R}_{\mathbf{Y}^{(n)}}^{-1}\mathbf{R}_{\mathbf{Y}^{(n)}X}$, and the mean square error is

$$e_L^* = \mathrm{Var}[X_1] - (\hat{\mathbf{a}}^{(n)})'\mathbf{R}_{\mathbf{Y}^{(n)}X}. \tag{3}$$

```
function e=mse953(r)
N=length(r);
e=[];
for n=1:N,
    RYX=r(1:n)';
    RY=toeplitz(r(1:n))+0.1*eye(n);
    a=RY\RYX;
    en=r(1)-(a')*RYX;
    e=[e;en];
end
plot(1:N,e);
```

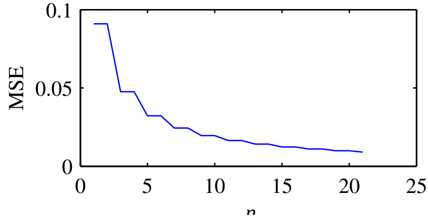
The program `mse953.m` simply calculates the mean square error e_L^*. The input is the vector `r` corresponding to the vector $\begin{bmatrix} r_0 & \cdots & r_{20} \end{bmatrix}$, which holds the first row of the Toeplitz correlation matrix $\mathbf{R}_{\mathbf{X}}$. Note that $\mathbf{R}_{\mathbf{X}^{(n)}}$ is the Toeplitz matrix whose first row is the first n elements of `r`.

To plot the mean square error as a function of the number of observations, n, we generate the vector `r` and then run `mse953(r)`. For the requested cases (a) and (b), the necessary MATLAB commands and corresponding mean square estimation error output as a function of n are shown here:




```
ra=sinc(0.1*pi*(0:20));
mse953(ra)
```

(a)




```
rb=cos(0.5*pi*(0:20));
mse953(rb)
```

(b)

In comparing the results of cases (a) and (b), we see that the mean square estimation error depends strongly on the correlation structure given by $r_{|i-j|}$. For case (a), Y_1 is a noisy observation of X_1 and is highly correlated with X_1. The MSE at $n = 1$ is just the variance of W_1. Additional samples of Y_n mostly help to average the additive noise. Also, samples X_n for $n \geq 10$ have very little correlation with X_1. Thus for $n \geq 10$, the samples of Y_n result in almost no improvement in the estimate of X_1.

In case (b), $Y_1 = X_1 + W_1$, just as in case (a), is simply a noisy copy of X_1 and the estimation error is due to the variance of W_1. On the other hand, for case (b), X_5, X_9, X_{13} and X_{17} and X_{21} are completely correlated with X_1. Other samples also have significant correlation with X_1. As a result, the MSE continues to go down with increasing n.

Problem 12.5.4 Solution

When the transmitted bit vector is $\mathbf{X}$, the received signal vector $\mathbf{Y} = \mathbf{S}\mathbf{P}^{1/2}\mathbf{X} + \mathbf{N}$. In Problem 12.4.9, we found that the linear MMSE estimate of the vector $\mathbf{X}$ is

$$\hat{\mathbf{X}} = \mathbf{P}^{1/2}\mathbf{S}' \left(\mathbf{S}\mathbf{P}\mathbf{S}' + \sigma^2\mathbf{I}\right)^{-1} \mathbf{Y}. \tag{1}$$

As indicated in the problem statement, the evaluation of the LMSE detector follows the approach outlined in Problem 11.4.6. Thus the solutions are essentially the same.

```
function S=randomsignals(n,k);
%S is an n by k matrix, columns are
%random unit length signal vectors
S=(rand(n,k)>0.5);
S=((2*S)-1.0)/sqrt(n);
```

The transmitted data vector $\mathbf{x}$ belongs to the set B_k of all binary ± 1 vectors of length k. This short program generates k random signals, each of length n. Each random signal is a binary ± 1 sequence normalized to length 1.

The evaluation of the LMSE detector is most similar to evaluation of the matched filter (MF) detector in Problem 11.4.6. We define a function `err=lmsecdmasim(S,P,m)` that simulates the LMSE detector for m symbols for a given set of signal vectors `S`. In `lmsecdmasim`, there is no need for looping. The mth transmitted symbol is represented by the mth column of `X` and the corresponding received signal is given by the mth column of `Y`. The matched filter processing can be applied to all m columns at once.A second function `Pe=lmsecdma(n,k,snr,s,m)` cycles through all combination of users `k` and SNR `snr` and calculates the bit error rate for each pair of values. Here are the functions:

```
function e=lmsecdmasim(S,P,m);
%err=lmsecdmasim(P,S,m);
%S= n x k matrix of signals
%P= diag matrix of SNRs
% SNR=power/var(noise)
k=size(S,2); %no. of users
n=size(S,1); %proc. gain
P2=sqrt(P);
X=randombinaryseqs(k,m);
Y=S*P2*X+randn(n,m);
L=P2*S'*inv((S*P*S')+eye(n));
XR=sign(L*Y);
e=sum(sum(XR ~= X));
```

```
function Pe=lmsecdma(n,k,snr,s,m);
%Pe=lmsecdma(n,k,snr,s,m);
%RCDMA, LMSE detector, users=k
%proc gain=n, rand signals/frame
%  s frames, m symbols/frame
%See Problem 9.5.4 Solution
[K,SNR]=ndgrid(k,snr);
Pe=zeros(size(SNR));
for j=1:prod(size(SNR)),
  p=SNR(j);kt=K(j); e=0;
  for i=1:s,
    S=randomsignals(n,kt);
    e=e+lmsecdmasim(S,p*eye(kt),m);
  end
  Pe(j)=e/(s*m*kt);
  disp([snr kt e]);
end
```

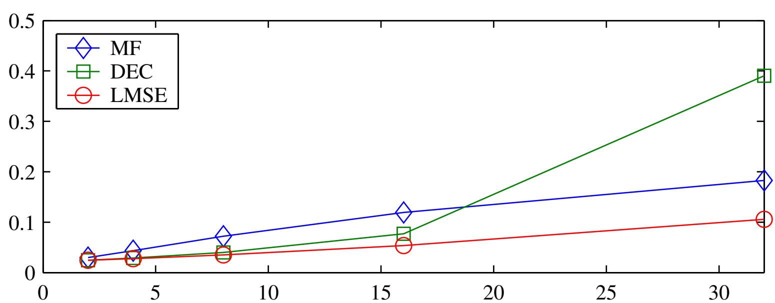
Here is a run of `lmsecdma`.

```
>> pelmse = lmsecdma(32,k,4,1000,1000);
           4            2        48542
           4            4       109203
           4            8       278266
           4           16       865358
           4           32      3391488
>> pelmse'
ans =
    0.0243    0.0273    0.0348    0.0541    0.1060
>>
```

For processing gain $n = 32$, the maximum likelihood detector is too complex for my version of MATLAB to run quickly. Instead we can compare the LMSE detector to the matched filter (MF) detector of Problem 11.4.6 and the decorrelator of Problem 11.4.7 with the following script:

```
k=[2 4 8 16 32];
pemf = mfrcdma(32,k,4,1000,1000);
pedec=berdecorr(32,k,4,10000);
pelmse = lmsecdma(32,k,4,1000,1000);
plot(k,pemf,'-d',k,pedec,'-s',k,pelmse);
legend('MF','DEC','LMSE',2);
axis([0 32 0 0.5]);
```

The resulting plot resembles



Compared to the matched filter and the decorrelator, the linear pre-processing of the LMSE detector offers an improvement in the bit error rate. Recall that for each bit X_i, the decorrelator zeroes out the interference from all users $j \neq i$ at the expense of enhancing the receiver noise. When the number of users is small, the decorrelator works well because the cost of suppressing other users is small. When the number of users equals the processing gain, the decorrelator owrks poorly because the noise is gratly enhanced. By comparison, the LMSE detector applies linear processing that results in an output that minimizes the mean square error between the output and the original transmitted bit. It works about as well as the decorrelator when the number of users is small. For a large number of users, it still works better than the matched filter detector.

Problem Solutions – Chapter 13

Problem 13.1.1 Solution

There are many correct answers to this question. A correct answer specifies enough random variables to specify the sample path exactly. One choice for an alternate set of random variables that would specify $m(t, s)$ is

- $m(0, s)$, the number of ongoing calls at the start of the experiment
- N, the number of new calls that arrive during the experiment
- $X_1, \ldots, X_N$, the interarrival times of the N new arrivals
- H, the number of calls that hang up during the experiment
- $D_1, \ldots, D_H$, the call completion times of the H calls that hang up

Problem 13.1.2 Solution

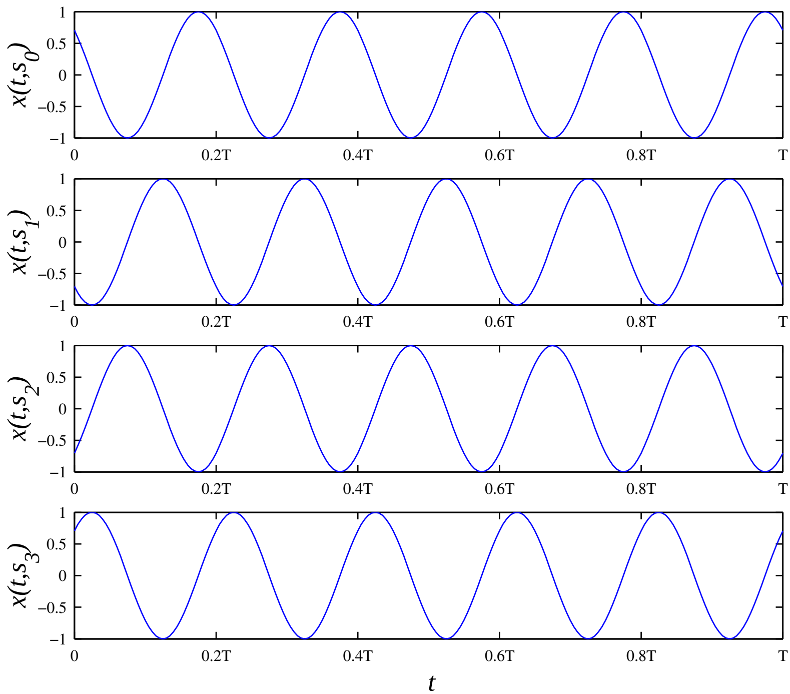
- In Example 13.3, the daily noontime temperature at Newark Airport is a discrete time, continuous value random process. However, if the temperature is recorded only in units of one degree, then the process was would be discrete value.
- In Example 13.4, the number of active telephone calls is discrete time and discrete value.
- The dice rolling experiment of Example 13.5 yields a discrete time, discrete value random process.
- The QPSK system of Example 13.6 is a continuous time and continuous value random process.

Problem 13.1.3 Solution

The sample space of the underlying experiment is $S = \{s_0, s_1, s_2, s_3\}$. The four elements in the sample space are equally likely. The ensemble of sample functions is $\{x(t, s_i) | i = 0, 1, 2, 3\}$ where

$$x(t, s_i) = \cos(2\pi f_0 t + \pi/4 + i\pi/2), \qquad 0 \le t \le T. \tag{1}$$

For $f_0 = 5/T$, this ensemble is shown below.

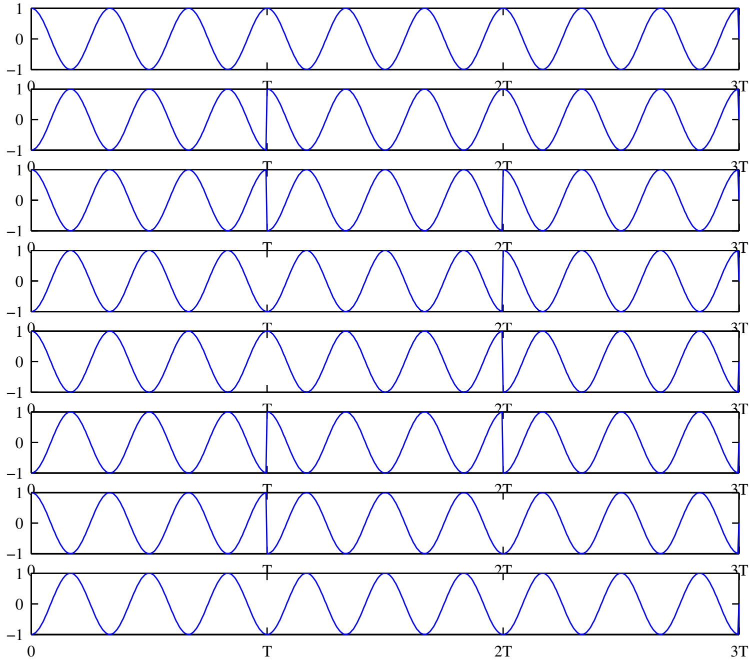


Problem 13.1.4 Solution

The eight possible waveforms correspond to the bit sequences

$$\{(0,0,0), (1,0,0), (1,1,0), \ldots, (1,1,1)\}. \tag{1}$$

The corresponding eight waveforms are:



Problem 13.1.5 Solution

The statement is *false*. As a counterexample, consider the rectified cosine waveform $X(t) = R|\cos 2\pi ft|$ of Example 13.9. When $t = \pi/2$, then $\cos 2\pi ft = 0$ so that $X(\pi/2) = 0$. Hence $X(\pi/2)$ has PDF

$$f_{X(\pi/2)}(x) = \delta(x). \tag{1}$$

That is, $X(\pi/2)$ is a discrete random variable.

Problem 13.2.1 Solution

In this problem, we start from first principles. What makes this problem fairly straightforward is that the ramp is defined for all time. That is, the ramp doesn't start at time $t = W$. Thus,

$$P[X(t) \leq x] = P[t - W \leq x] = P[W \geq t - x]. \tag{1}$$

Since $W \geq 0$, if $x \geq t$ then $P[W \geq t - x] = 1$. When $x < t$,

$$P[W \geq t - x] = \int_{t-x}^{\infty} f_W(w)\, dw = e^{-(t-x)}. \tag{2}$$

Combining these facts, we have

$$F_{X(t)}(x) = P[W \geq t - x] = \begin{cases} e^{-(t-x)} & x < t, \\ 1 & t \leq x. \end{cases} \tag{3}$$

We note that the CDF contain no discontinuities. Taking the derivative of the CDF $F_{X(t)}(x)$ with respect to x, we obtain the PDF

$$f_{X(t)}(x) = \begin{cases} e^{x-t} & x < t, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 13.2.2 Solution

(a) Each resistor has frequency W in Hertz with uniform PDF

$$f_R(r) = \begin{cases} 0.025 & 9980 \leq r \leq 1020, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The probability that a test yields a one part in 10^4 oscillator is

$$p = P[9999 \leq W \leq 10001] = \int_{9999}^{10001} (0.025)\, dr = 0.05. \tag{2}$$

(b) To find the PMF of T_1, we view each oscillator test as an independent trial. A success occurs on a trial with probability p if we find a one part in 10^4 oscillator. The first one part in 10^4 oscillator is found at time $T_1 = t$ if we observe failures on trials $1, \ldots, t-1$ followed by a success on trial t. Hence, just as in Example 3.9, T_1 has the geometric PMF

$$P_{T_1}(t) = \begin{cases} (1-p)^{t-1} p & t = 1, 2, \ldots, \\ 9 & \text{otherwise.} \end{cases} \tag{3}$$

A geometric random variable with success probability p has mean $1/p$. This is derived in Theorem 3.5. The expected time to find the first good oscillator is $\mathrm{E}[T_1] = 1/p = 20$ minutes.

(c) Since $p = 0.05$, the probability the first one part in 10^4 oscillator is found in exactly 20 minutes is $P_{T_1}(20) = (0.95)^{19}(0.05) = 0.0189$.

(d) The time T_5 required to find the 5th one part in 10^4 oscillator is the number of trials needed for 5 successes. T_5 is a Pascal random variable. If this is not clear, see Example 3.13 where the Pascal PMF is derived. When we are looking for 5 successes, the Pascal PMF is

$$P_{T_5}(t) = \begin{cases} \binom{t-1}{4} p^5 (1-p)^{t-5} & t = 5, 6, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Looking up the Pascal PMF in Appendix A, we find that $\mathrm{E}[T_5] = 5/p = 100$ minutes. The following argument is a second derivation of the mean of T_5. Once we find the first one part in 10^4 oscillator, the number of additional trials needed to find the next one part in 10^4 oscillator once again has a geometric PMF with mean $1/p$ since each independent trial is a success with probability p. Similarly, the time required to find 5 one part in 10^4 oscillators is the sum of five independent geometric random variables. That is,

$$T_5 = K_1 + K_2 + K_3 + K_4 + K_5, \tag{5}$$

where each K_i is identically distributed to T_1. Since the expectation of the sum equals the sum of the expectations,

$$\begin{aligned} \mathrm{E}[T_5] &= \mathrm{E}[K_1 + K_2 + K_3 + K_4 + K_5] \\ &= 5\,\mathrm{E}[K_i] = 5/p = 100 \text{ minutes.} \end{aligned} \tag{6}$$

Problem 13.2.3 Solution

Once we find the first one part in 10^4 oscillator, the number of additional tests needed to find the next one part in 10^4 oscillator once again has a geometric PMF with mean $1/p$ since each independent trial is a success with probability p. That is $T_2 = T_1 + T'$ where T' is independent and identically distributed to T_1. Thus,

$$\begin{aligned} \mathrm{E}[T_2|T_1 = 3] &= \mathrm{E}[T_1|T_1 = 3] + \mathrm{E}\left[T'|T_1 = 3\right] \\ &= 3 + \mathrm{E}\left[T'\right] = 23 \text{ minutes.} \end{aligned} \tag{1}$$

Problem 13.2.4 Solution

Since the problem states that the pulse is delayed, we will assume $T \geq 0$. This problem is difficult because the answer will depend on t. In particular, for $t < 0$, $X(t) = 0$ and $f_{X(t)}(x) = \delta(x)$. Things are more complicated when $t > 0$. For $x < 0$, $\mathrm{P}[X(t) > x] = 1$. For $x \geq 1$, $\mathrm{P}[X(t) > x] = 0$. Lastly, for $0 \leq x < 1$,

$$\begin{aligned}\mathrm{P}\left[X(t) > x\right] &= \mathrm{P}\left[e^{-(t-T)}u(t-T) > x\right] \\ &= \mathrm{P}\left[t + \ln x < T \leq t\right] \\ &= F_T(t) - F_T(t + \ln x)\,. \end{aligned} \tag{1}$$

Note that condition $T \leq t$ is needed to make sure that the pulse doesn't arrive after time t. The other condition $T > t + \ln x$ ensures that the pulse didn't arrrive too early and already decay too much. We can express these facts in terms of the CDF of $X(t)$.

$$\begin{aligned}F_{X(t)}(x) &= 1 - \mathrm{P}\left[X(t) > x\right] \\ &= \begin{cases} 0 & x < 0, \\ 1 + F_T(t + \ln x) - F_T(t) & 0 \leq x < 1, \\ 1 & x \geq 1. \end{cases}\end{aligned} \tag{2}$$

We can take the derivative of the CDF to find the PDF. However, we need to keep in mind that the CDF has a jump discontinuity at $x = 0$. In particular, since $\ln 0 = -\infty$,

$$F_{X(t)}(0) = 1 + F_T(-\infty) - F_T(t) = 1 - F_T(t)\,. \tag{3}$$

Hence, when we take a derivative, we will see an impulse at $x = 0$. The PDF of $X(t)$ is

$$f_{X(t)}(x) = \begin{cases} (1 - F_T(t))\delta(x) + f_T(t + \ln x)/x & 0 \leq x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Problem 13.3.1 Solution

Each Y_k is the sum of two identical independent Gaussian random variables. Hence, each Y_k must have the same PDF. That is, the Y_k are identically distributed. Next, we observe that the sequence of Y_k is independent. To see this, we observe that each Y_k is composed of two samples of X_k that are unused by any other Y_j for $j \neq k$.

Problem 13.3.2 Solution

Each W_n is the sum of two identical independent Gaussian random variables. Hence, each W_n must have the same PDF. That is, the W_n are identically distributed. However, since W_{n-1} and W_n both use X_{n-1} in their averaging, W_{n-1} and W_n are dependent. We can verify this observation by calculating the covariance of W_{n-1} and W_n. First, we observe that for all n,

$$\mathrm{E}[W_n] = (\mathrm{E}[X_n] + \mathrm{E}[X_{n-1}])/2 = 30. \tag{1}$$

Next, we observe that W_{n-1} and W_n have covariance

$$\begin{aligned}\mathrm{Cov}[W_{n-1}, W_n] &= \mathrm{E}[W_{n-1}W_n] - \mathrm{E}[W_n]\,\mathrm{E}[W_{n-1}] \\ &= \frac{1}{4}\mathrm{E}[(X_{n-1} + X_{n-2})(X_n + X_{n-1})] - 900. \end{aligned} \tag{2}$$

We observe that for $n \neq m$, $\mathrm{E}[X_n X_m] = \mathrm{E}[X_n]\,\mathrm{E}[X_m] = 900$ while

$$\mathrm{E}\left[X_n^2\right] = \mathrm{Var}[X_n] + (\mathrm{E}[X_n])^2 = 916. \tag{3}$$

Thus,

$$\mathrm{Cov}[W_{n-1}, W_n] = \frac{900 + 916 + 900 + 900}{4} - 900 = 4. \tag{4}$$

Since $\mathrm{Cov}[W_{n-1}, W_n] \neq 0$, W_n and W_{n-1} must be dependent.

Problem 13.3.3 Solution

The number Y_k of failures between successes $k-1$ and k is exactly $y \geq 0$ iff after success $k-1$, there are y failures followed by a success. Since the Bernoulli trials are independent, the probability of this event is $(1-p)^y p$. The complete PMF of Y_k is

$$P_{Y_k}(y) = \begin{cases} (1-p)^y p & y = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Since this argument is valid for all k including $k = 1$, we can conclude that $Y_1, Y_2, \ldots$ are identically distributed. Moreover, since the trials are independent, the failures between successes $k-1$ and k and the number of failures between successes $k'-1$ and k' are independent. Hence, $Y_1, Y_2, \ldots$ is an iid sequence.

Problem 13.4.1 Solution

This is a very straightforward problem. The Poisson process has rate $\lambda = 4$ calls per second. When t is measured in seconds, each $N(t)$ is a Poisson random variable with mean $4t$ and thus has PMF

$$P_{N(t)}(n) = \begin{cases} \frac{(4t)^n}{n!} e^{-4t} & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Using the general expression for the PMF, we can write down the answer for each part.

(a) $P_{N(1)}(0) = 4^0 e^{-4}/0! = e^{-4} \approx 0.0183$.

(b) $P_{N(1)}(4) = 4^4 e^{-4}/4! = 32e^{-4}/3 \approx 0.1954$.

(c) $P_{N(2)}(2) = 8^2 e^{-8}/2! = 32e^{-8} \approx 0.0107$.

Problem 13.4.2 Solution

Following the instructions given, we express each answer in terms of $N(m)$ which has PMF

$$P_{N(m)}(n) = \begin{cases} (6m)^n e^{-6m}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(a) The probability of no queries in a one minute interval is $P_{N(1)}(0) = 6^0 e^{-6}/0! = 0.00248$.

(b) The probability of exactly 6 queries arriving in a one minute interval is $P_{N(1)}(6) = 6^6 e^{-6}/6! = 0.161$.

(c) The probability of exactly three queries arriving in a one-half minute interval is $P_{N(0.5)}(3) = 3^3 e^{-3}/3! = 0.224$.

Problem 13.4.3 Solution

Since there is always a backlog an the service times are iid exponential random variables, The time between service completions are a sequence of iid exponential random variables. that is, the service completions are a Poisson process. Since the expected service time is 30 minutes, the rate of the Poisson process is $\lambda = 1/30$ per minute. Since t hours equals $60t$ minutes, the expected number serviced is $\lambda(60t)$ or $2t$. Moreover, the number serviced in the first t hours has the Poisson PMF

$$P_{N(t)}(n) = \begin{cases} \frac{(2t)^n e^{-2t}}{n!} & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 13.4.4 Solution

Since $D(t)$ is a Poisson process with rate 0.1 drops/day, the random variable $D(t)$ is a Poisson random variable with parameter $\alpha = 0.1t$. The PMF of $D(t)$. the number of drops after t days, is

$$P_{D(t)}(d) = \begin{cases} (0.1t)^d e^{-0.1t}/d! & d = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 13.4.5 Solution

Note that it matters whether $t \geq 2$ minutes. If $t \leq 2$, then any customers that have arrived must still be in service. Since a Poisson number of arrivals occur during $(0, t]$,

$$P_{N(t)}(n) = \begin{cases} (\lambda t)^n e^{-\lambda t}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \qquad (0 \leq t \leq 2.) \tag{1}$$

For $t \geq 2$, the customers in service are precisely those customers that arrived in the interval $(t-2, t]$. The number of such customers has a Poisson PMF with mean $\lambda[t - (t-2)] = 2\lambda$. The resulting PMF of $N(t)$ is

$$P_{N(t)}(n) = \begin{cases} (2\lambda)^n e^{-2\lambda}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \qquad (t \geq 2.) \tag{2}$$

Problem 13.4.6 Solution

(a) and (b) are Poisson, the others are not. For (a) and (b), $N'(t) = N(\alpha t)$ is Poisson because for any interval (t_1, t_2),

$$N'(t_2) - N'(t_1) = N(\alpha t_2) - N(\alpha t_1), \tag{1}$$

which is Poisson with expected value $\alpha\lambda(t_2 - t_1)$. Also, non-overlapping intervals with respect to $N'(t)$ correspond to non-overlapping intervals for $N(t)$ and thus the number of arrivals in non-overlapping intervals are independent. For case (c), $2N(t)$ is not Poisson because $2N(t)$ is always even. For case (d), $N(t)/2$ takes on fractional sample values, which a Poisson process cannot. For case (e), $N(t+2)$ may be nonzero at time $t = 0$ but a Poisson process is always zero at $t = 0$. Finally, for (f), $N(t) - N(t-1)$ is not necessarily increasing.

Problem 13.4.7 Solution

(a) N_τ is a Poisson ($\alpha = 10\tau$) random variable. You should know that $\mathrm{E}[N_\tau] = 10\tau$. Thus $\mathrm{E}[N_{60}] = 10 \cdot 60 = 600$.

(b) In a $\tau = 10$ minute interval N_{10} hamburgers are sold. Thus,

$$\mathrm{P}\left[N_{10} = 0\right] = P_{N_{10}}(0) = (100)^0 e^{-100}/0! = e^{-100}. \tag{1}$$

(c) Let t denote the time 12 noon. In this case, for $w > 0$, $W > w$ if and only if no burgers are sold in the time interval $[t, t+w]$. That is,

$$\begin{aligned} \mathrm{P}\left[W > w\right] &= \mathrm{P}\left[\text{No burgers are sold in } [t, t+w]\right] \\ &= \mathrm{P}\left[N_w = 0\right] \\ &= P_{N_w}(0) = (10w)^0 e^{-10w}/0! = e^{-10w}. \end{aligned} \tag{2}$$

For $w > 0$, $F_W(w) = 1 - \mathrm{P}[W > w] = 1 - e^{-10w}$. That is, the CDF of W is

$$F_W(w) = \begin{cases} 0 & w < 0, \\ 1 - e^{-10w} & w \geq 0. \end{cases} \tag{3}$$

Taking a derivative, we have

$$f_W(w) = \begin{cases} 0 & w < 0, \\ 10e^{-10w} & w \geq 0. \end{cases} \tag{4}$$

We see that W is an exponential $\lambda = 10$ random variable.

Problem 13.4.8 Solution

The time T between queries are independent exponential random variables with PDF

$$f_T(t) = \begin{cases} (1/8)e^{-t/8} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

From the PDF, we can calculate for $t > 0$,

$$\mathrm{P}[T \geq t] = \int_0^t f_T(t')\, dt' = e^{-t/8}. \tag{2}$$

Using this formula, each question can be easily answered.

(a) $\mathrm{P}[T \geq 4] = e^{-4/8} \approx 0.951$.

(b)

$$\begin{aligned} \mathrm{P}[T \geq 13 | T \geq 5] &= \frac{\mathrm{P}[T \geq 13, T \geq 5]}{\mathrm{P}[T \geq 5]} \\ &= \frac{\mathrm{P}[T \geq 13]}{\mathrm{P}[T \geq 5]} = \frac{e^{-13/8}}{e^{-5/8}} = e^{-1} \approx 0.368. \end{aligned} \tag{3}$$

(c) Although the time between queries are independent exponential random variables, $N(t)$ is not exactly a Poisson random process because the first query occurs at time $t = 0$. Recall that in a Poisson process, the first arrival occurs some time after $t = 0$. However $N(t) - 1$ is a Poisson process of rate 8. Hence, for $n = 0, 1, 2, \ldots$,

$$\mathrm{P}[N(t) - 1 = n] = (t/8)^n e^{-t/8}/n! \tag{4}$$

Thus, for $n = 1, 2, \ldots$, the PMF of $N(t)$ is

$$P_{N(t)}(n) = \mathrm{P}[N(t) - 1 = n - 1] = (t/8)^{n-1} e^{-t/8}/(n-1)! \tag{5}$$

The complete expression of the PMF of $N(t)$ is

$$P_{N(t)}(n) = \begin{cases} (t/8)^{n-1} e^{-t/8}/(n-1)! & n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Problem 13.4.9 Solution

This proof is just a simplified version of the proof given for Theorem 13.3. The first arrival occurs at time $X_1 > x \geq 0$ iff there are no arrivals in the interval $(0, x]$. Hence, for $x \geq 0$,

$$\mathrm{P}[X_1 > x] = \mathrm{P}[N(x) = 0] = (\lambda x)^0 e^{-\lambda x}/0! = e^{-\lambda x}. \tag{1}$$

Since $\mathrm{P}[X_1 \leq x] = 0$ for $x < 0$, the CDF of X_1 is the exponential CDF

$$F_{X_1}(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases} \tag{2}$$

Problem 13.4.10 Solution

(a) For $X_i = -\ln U_i$, we can write

$$\mathrm{P}[X_i > x] = \mathrm{P}[-\ln U_i > x] = \mathrm{P}[\ln U_i \leq -x] = \mathrm{P}\left[U_i \leq e^{-x}\right]. \tag{1}$$

When $x < 0$, $e^{-x} > 1$ so that $\mathrm{P}[U_i \leq e^{-x}] = 1$. When $x \geq 0$, we have $0 < e^{-x} \leq 1$, implying $\mathrm{P}[U_i \leq e^{-x}] = e^{-x}$. Combining these facts, we have

$$\mathrm{P}[X_i > x] = \begin{cases} 1 & x < 0, \\ e^{-x} & x \geq 0. \end{cases} \tag{2}$$

This permits us to show that the CDF of X_i is

$$F_{X_i}(x) = 1 - \mathrm{P}[X_i > x] = \begin{cases} 0 & x < 0, \\ 1 - e^{-x} & x > 0. \end{cases} \tag{3}$$

We see that X_i has an exponential CDF with mean 1.

(b) Note that $N = n$ iff

$$\prod_{i=1}^{n} U_i \geq e^{-t} > \prod_{i=1}^{n+1} U_i. \tag{4}$$

By taking the logarithm of both inequalities, we see that $N = n$ iff

$$\sum_{i=1}^{n} \ln U_i \geq -t > \sum_{i=1}^{n+1} \ln U_i. \tag{5}$$

Next, we multiply through by -1 and recall that $X_i = -\ln U_i$ is an exponential random variable. This yields $N = n$ iff

$$\sum_{i=1}^{n} X_i \leq t < \sum_{i=1}^{n+1} X_i. \tag{6}$$

Now we recall that a Poisson process $N(t)$ of rate 1 has independent exponential interarrival times $X_1, X_2, \ldots$. That is, the ith arrival occurs at time $\sum_{j=1}^{i} X_j$. Moreover, $N(t) = n$ iff the first n arrivals occur by time t but arrival $n+1$ occurs after time t. Since the random variable $N(t)$ has a Poisson distribution with mean t, we can write

$$\mathrm{P}\left[\sum_{i=1}^{n} X_i \leq t < \sum_{i=1}^{n+1} X_i\right] = \mathrm{P}\left[N(t) = n\right] = \frac{t^n e^{-t}}{n!}. \tag{7}$$

Problem 13.5.1 Solution

Customers entering (or not entering) the casino is a Bernoulli decomposition of the Poisson process of arrivals at the casino doors. By Theorem 13.6, customers entering the casino are a Poisson process of rate $100/2 = 50$ customers/hour. Thus in the two hours from 5 to 7 PM, the number, N, of customers entering the casino is a Poisson random variable with expected value $\alpha = 2 \cdot 50 = 100$. The PMF of N is

$$P_N(n) = \begin{cases} 100^n e^{-100}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 13.5.2 Solution

This problem is straightforward. When a train arrives, it is blue with probability $p = \lambda_B/(\lambda_R + \lambda_B)$ and otherwise it is red with probability $1 - p$. Since the arrival processes are Poisson, they are memoryless, and each train arrival is an independent trial. If we view the arrival of a blue train as a success, we see that n is the number of failures before the first success. We will see $N = n$ red trains if we have n failures followed by a success on trial $n + 1$; this occurs with probability $(1 - p)^n p$. Thus,

$$P_N(n) = \begin{cases} (1-p)^n p & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Problem 13.5.3 Solution

(a) The trains (red and blue together) arrive as a Poisson process of rate $\lambda_R + \lambda_B = 0.45$ trains per minute. In one hour, the number of trains that arrive N is a Poisson ($\alpha = 27$) random variable. The PMF is

$$P_N(=) \begin{cases} 27^n e^{-27}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) Each train that arrives is a red train with probability $p = \lambda_R/(\lambda_R+\lambda_B) = 1/3$. Given that $N = 30$ trains arrive, R is conditionally a binomial $(30, p)$ random variable. The conditional PMF is

$$P_{R|N}(r|30) = \binom{30}{r} p^r (1-p)^{30-r}. \tag{2}$$

Problem 13.5.4 Solution

(a) Since buses arrives as a Poisson process, the interarrival time X has the exponential PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

(b) Since the Poisson process is memoryless, the time until the next arrival W has the exponential PDF

$$f_W(w) = \begin{cases} \lambda e^{-\lambda w} & w \geq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

no matter how long it may have been since the prior arrival. One can deduce this directly from the observation that $W > w$ iff there are no arrival in the interval $[t, t + w]$.

(c) For $v \geq 0$,

$$\text{P}[V > v] = \text{P}[\text{no arrivals in } [t - v, t]] = e^{-\lambda v}. \tag{3}$$

Since $\text{P}[V > v] = 1$ for $v < 0$, it follows that the CDF of V satisfies

$$F_V(v) = 1 - \text{P}[V > v] = \begin{cases} 0 & v < 0, \\ 1 - e^{-\lambda v} & v \geq 0, \end{cases} \tag{4}$$

and that

$$f_V(v) = \frac{dF_V(v)}{dv} = \begin{cases} 0 & v < 0, \\ \lambda e^{-\lambda v} & v \geq 0. \end{cases} \tag{5}$$

(d) Note that $U = V + W$ is the sum of independent exponential (λ) random variables. It is fact (which you could derive if you wish) that U has the exponential ($n = 2, \lambda$) PDF

$$f_U(u) = \begin{cases} \lambda^2 u e^{-\lambda u} & u \geq 0, \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

Problem 13.5.5 Solution

In an interval $(t, t + \Delta]$ with an infinitesimal Δ, let A_i denote the event of an arrival of the process $N_i(t)$. Also, let $A = A_1 \cup A_2$ denote the event of an arrival of either process. Since $N_i(t)$ is a Poisson process, the alternative model says that

$\mathrm{P}[A_i] = \lambda_i \Delta$. Also, since $N_1(t) + N_2(t)$ is a Poisson process, the proposed Poisson process model says

$$\mathrm{P}[A] = (\lambda_1 + \lambda_2)\Delta. \tag{1}$$

Lastly, the conditional probability of a type 1 arrival given an arrival of either type is

$$\mathrm{P}[A_1|A] = \frac{\mathrm{P}[A_1 A]}{\mathrm{P}[A]} = \frac{\mathrm{P}[A_1]}{\mathrm{P}[A]} = \frac{\lambda_1 \Delta}{(\lambda_1 + \lambda_2)\Delta} = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \tag{2}$$

This solution is something of a cheat in that we have used the fact that the sum of Poisson processes is a Poisson process without using the proposed model to derive this fact.

Problem 13.5.6 Solution

We start with the case when $t \geq 2$. When each service time is equally likely to be either 1 minute or 2 minutes, we have the following situation. Let M_1 denote those customers that arrived in the interval $(t-1, 1]$. All M_1 of these customers will be in the bank at time t and M_1 is a Poisson random variable with mean λ.

Let M_2 denote the number of customers that arrived during $(t-2, t-1]$. Of course, M_2 is Poisson with expected value λ. We can view each of the M_2 customers as flipping a coin to determine whether to choose a 1 minute or a 2 minute service time. Only those customers that chooses a 2 minute service time will be in service at time t. Let M_2' denote those customers choosing a 2 minute service time. It should be clear that M_2' is a Poisson number of Bernoulli random variables. Theorem 13.6 verifies that using Bernoulli trials to decide whether the arrivals of a rate λ Poisson process should be counted yields a Poisson process of rate $p\lambda$. A consequence of this result is that a Poisson number of Bernoulli (success probability p) random variables has Poisson PMF with mean $p\lambda$. In this case, M_2' is Poisson with mean $\lambda/2$. Moreover, the number of customers in service at time t is $N(t) = M_1 + M_2'$. Since M_1 and M_2' are independent Poisson random variables, their sum $N(t)$ also has a Poisson PMF. This was verified in Theorem 9.7. Hence $N(t)$ is Poisson with mean $\mathrm{E}[N(t)] = \mathrm{E}[M_1] + \mathrm{E}[M_2'] = 3\lambda/2$. The PMF of $N(t)$ is

$$P_{N(t)}(n) = \begin{cases} (3\lambda/2)^n e^{-3\lambda/2}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise}, \end{cases} \qquad t \geq 2. \tag{1}$$

Now we can consider the special cases arising when $t < 2$. When $0 \le t < 1$, every arrival is still in service. Thus the number in service $N(t)$ equals the number of arrivals and has the PMF

$$P_{N(t)}(n) = \begin{cases} (\lambda t)^n e^{-\lambda t}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases} \qquad 0 \le t \le 1. \tag{2}$$

When $1 \le t < 2$, let M_1 denote the number of customers in the interval $(t-1, t]$. All M_1 customers arriving in that interval will be in service at time t. The M_2 customers arriving in the interval $(0, t-1]$ must each flip a coin to decide one a 1 minute or two minute service time. Only those customers choosing the two minute service time will be in service at time t. Since M_2 has a Poisson PMF with mean $\lambda(t-1)$, the number M_2' of those customers in the system at time t has a Poisson PMF with mean $\lambda(t-1)/2$. Finally, the number of customers in service at time t has a Poisson PMF with expected value $\mathrm{E}[N(t)] = \mathrm{E}[M_1] + \mathrm{E}[M_2'] = \lambda + \lambda(t-1)/2$. Hence, for $1 \le t \le 2$, the PMF of $N(t)$ becomes

$$P_{N(t)}(n) = \begin{cases} (\lambda(t+1)/2)^n e^{-\lambda(t+1)/2}/n! & n = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Problem 13.5.7 Solution

(a) The last runner's finishing time is $L = \max(R_1, \ldots, R_{10})$ and

$$\begin{aligned} \mathrm{P}[L \le 20] &= \mathrm{P}[\max(R_1, \ldots, R_{10}) \le 20] \\ &= \mathrm{P}[R_1 \le 20, R_2 \le 20, \ldots, R_{10} \le 20] \\ &= \mathrm{P}[R_1 \le 20]\,\mathrm{P}[R_2 \le 20] \cdots \mathrm{P}[R_{10} \le 20] \\ &= (\mathrm{P}[R_1 \le 20])^{10} \\ &= \left(1 - e^{-20\mu}\right)^{10} = \left(1 - e^{-2}\right)^{10} \approx 0.234. \end{aligned} \tag{1}$$

(b) At the start at time zero, we can view each runner as the first arrival of an independent Poisson process of rate μ. Thus, at time zero, the arrival of the first runner can be viewed as the first arrival of a process of rate 10μ. Hence, X_1 is exponential with expected value $1/(10\mu) = 1$ and has PDF

$$f_{X_1}(x_1) = \begin{cases} e^{-x_1} & x_1 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(c) We can view Y as the 10th arrival of a Poisson process of rate μ. Thus Y has the Erlang $(n = 10, \mu)$ PDF

$$f_Y(y) = \begin{cases} \frac{\mu^{10} y^9 e^{-\mu y}}{9!} & y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

(d) We already found the PDF of X_1. We observe that after the first runner finishes, there are still 9 runners on the course. Because each runner's time is memoryless, each runner has a residual running time that is an exponential (μ) random variable. Because these residual running times are independent X_2 is exponential with expected value $1/(9\mu) = 1/0.9$ and has PDF

$$f_{X_2}(x_2) = \begin{cases} 9\mu e^{-9\mu x_2} & x_2 \geq 0, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} 0.9 e^{-0.9 x_2} & x_2 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Similarly, for the ith arrival, there are $10 - i + 1 = 11 - i$ runners left on the course. The interarrival time for the ith arriving runner is the same as waiting for the first arrival of a Poisson process of rate $(11 - i)\mu$. Thus X_i has PDF

$$f_{X_i}(x_i) = \begin{cases} (11 - i)\mu e^{-(11-i)\mu x_i} & x_i \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Finally, we observe that the memoryless property of the runners' exponential running times ensures that the X_i are independent random variables. Hence,

$$\begin{aligned} &f_{X_1,\ldots,X_{10}}(x_1, \ldots, x_{10}) \\ &\quad = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_{10}}(x_{10}) \\ &\quad = \begin{cases} 10!\mu^{10} e^{-\mu(10x_1 + 9x_2 + \cdots + 2x_9 + x_{10})} & x_i \geq 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

Problem 13.5.8 Solution

Since the arrival times $S_1, \ldots, S_n$ are ordered in time and since a Poisson process cannot have two simultaneous arrivals, the conditional PDF $f_{S_1,\ldots,S_n|N}(S_1, \ldots, S_n|n)$ is nonzero only if $s_1 < s_2 < \cdots < s_n < T$. In this case, consider an arbitrarily small Δ; in particular, $\Delta < \min_i(s_{i+1} - s_i)/2$ implies that the intervals $(s_i, s_i + \Delta]$ are non-overlapping. We now find the joint probability

$$\mathrm{P}\left[s_1 < S_1 \leq s_1 + \Delta, \ldots, s_n < S_n \leq s_n + \Delta, N = n\right]$$

that each S_i is in the interval $(s_i, s_i + \Delta]$ and that $N = n$. This joint event implies that there were zero arrivals in each interval $(s_i + \Delta, s_{i+1}]$. That is, over the interval $[0, T]$, the Poisson process has exactly one arrival in each interval $(s_i, s_i + \Delta]$ and zero arrivals in the time period $T - \bigcup_{i=1}^{n}(s_i, s_i + \Delta]$. The collection of intervals in which there was no arrival had a total duration of $T - n\Delta$. Note that the probability of exactly one arrival in the interval $(s_i, s_i + \Delta]$ is $\lambda\Delta e^{-\lambda\delta}$ and the probability of zero arrivals in a period of duration $T - n\Delta$ is $e^{-\lambda(T_n - \Delta)}$. In addition, the event of one arrival in each interval $(s_i, s_i + \Delta)$ and zero events in the period of length $T - n\Delta$ are independent events because they consider non-overlapping periods of the Poisson process. Thus,

$$\begin{aligned}
\mathrm{P}&\left[s_1 < S_1 \leq s_1 + \Delta, \ldots, s_n < S_n \leq s_n + \Delta, N = n\right] \\
&= \left(\lambda\Delta e^{-\lambda\Delta}\right)^n e^{-\lambda(T - n\Delta)} \\
&= (\lambda\Delta)^n e^{-\lambda T}. \qquad (1)
\end{aligned}$$

Since $\mathrm{P}[N = n] = (\lambda T)^n e^{-\lambda T}/n!$, we see that

$$\begin{aligned}
\mathrm{P}&\left[s_1 < S_1 \leq s_1 + \Delta, \ldots, s_n < S_n \leq s_n + \Delta | N = n\right] \\
&= \frac{\mathrm{P}\left[s_1 < S_1 \leq s_1 + \Delta, \ldots, s_n < S_n \leq s_n + \Delta, N = n\right]}{\mathrm{P}\left[N = n\right]} \\
&= \frac{(\lambda\Delta)^n e^{-\lambda T}}{(\lambda T)^n e^{-\lambda T}/n!} \\
&= \frac{n!}{T^n}\Delta^n. \qquad (2)
\end{aligned}$$

Finally, for infinitesimal Δ, the conditional PDF of $S_1, \ldots, S_n$ given $N = n$ satisfies

$$\begin{aligned}
&f_{S_1,\ldots,S_n|N}(s_1, \ldots, s_n|n)\, \Delta^n \\
&\quad = \mathrm{P}\left[s_1 < S_1 \le s_1 + \Delta, \ldots, s_n < S_n \le s_n + \Delta | N = n\right] \\
&\quad = \frac{n!}{T^n} \Delta^n. \qquad (3)
\end{aligned}$$

Since the conditional PDF is zero unless $s_1 < s_2 < \cdots < s_n \le T$, it follows that

$$f_{S_1,\ldots,S_n|N}(s_1, \ldots, s_n|n) = \begin{cases} n!/T^n & 0 \le s_1 < \cdots < s_n \le T, \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

If it seems that the above argument had some "hand-waving," we now do the derivation of $\mathrm{P}[s_1 < S_1 \le s_1 + \Delta, \ldots, s_n < S_n \le s_n + \Delta | N = n]$ in somewhat excruciating detail. (Feel free to skip the following if you were satisfied with the earlier explanation.)

For the interval $(s, t]$, we use the shorthand notation $0_{(s,t)}$ and $1_{(s,t)}$ to denote the events of 0 arrivals and 1 arrival respectively. This notation permits us to write

$$\begin{aligned}
&\mathrm{P}\left[s_1 < S_1 \le s_1 + \Delta, \ldots, s_n < S_n \le s_n + \Delta, N = n\right] \\
&= \mathrm{P}\left[0_{(0,s_1)} 1_{(s_1,s_1+\Delta)} 0_{(s_1+\Delta,s_2)} 1_{(s_2,s_2+\Delta)} 0_{(s_2+\Delta,s_3)} \cdots 1_{(s_n,s_n+\Delta)} 0_{(s_n+\Delta,T)}\right]. \qquad (5)
\end{aligned}$$

The set of events $0_{(0,s_1)}$, $0_{(s_n+\Delta,T)}$, and for $i = 1, \ldots, n-1$, $0_{(s_i+\Delta,s_{i+1})}$ and $1_{(s_i,s_i+\Delta)}$ are independent because each event depends on the Poisson process in a time interval that overlaps none of the other time intervals. In addition, since the Poisson process has rate λ,

$$\mathrm{P}\left[0_{(s,t)}\right] = e^{-\lambda(t-s)}, \qquad \mathrm{P}\left[1_{(s_i,s_i+\Delta)}\right] = (\lambda\Delta)e^{-\lambda\Delta}. \qquad (6)$$

Thus,

$$\begin{aligned}
&\mathrm{P}\left[s_1 < S_1 \le s_1 + \Delta, \ldots, s_n < S_n \le s_n + \Delta, N = n\right] \\
&\quad = \mathrm{P}\left[0_{(0,s_1)}\right] \mathrm{P}\left[1_{(s_1,s_1+\Delta)}\right] \mathrm{P}\left[0_{(s_1+\Delta,s_2)}\right] \cdots \mathrm{P}\left[1_{(s_n,s_n+\Delta)}\right] \mathrm{P}\left[0_{(s_n+\Delta,T)}\right] \\
&\quad = e^{-\lambda s_1} \left(\lambda\Delta e^{-\lambda\Delta}\right) e^{-\lambda(s_2-s_1-\Delta)} \cdots \left(\lambda\Delta e^{-\lambda\Delta}\right) e^{-\lambda(T-s_n-\Delta)} \\
&\quad = (\lambda\Delta)^n e^{-\lambda T}. \qquad (7)
\end{aligned}$$

Problem 13.6.1 Solution

From the problem statement, the change in the stock price is $X(8)-X(0)$ and the standard deviation of $X(8)-X(0)$ is 1/2 point. In other words, the variance of $X(8)-X(0)$ is $\text{Var}[X(8)-X(0)] = 1/4$. By the definition of Brownian motion. $\text{Var}[X(8)-X(0)] = 8\alpha$. Hence $\alpha = 1/32$.

Problem 13.6.2 Solution

The key here is to realize that $Y_n = \sum_{i=1}^{n} X_i$. It follows that

$$\mathrm{E}[Y_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = 0, \tag{1}$$

$$\text{Var}[Y_n] = \sum_{i=1}^{n} \text{Var}[X_i] = n. \tag{2}$$

For $k \geq 0$, we can write

$$Y_{n+k} = Y_n + \sum_{i=n+1}^{n+k} X_i. \tag{3}$$

We note that for $n+1 \leq i \leq n+k$, that X_i is independent of $Y_n = \sum_{i=1}^{n} X_i$. It follows that

$$\begin{aligned} R_Y[n,k] = \mathrm{E}[Y_n Y_{n+k}] &= \mathrm{E}\left[Y_n(Y_n + \sum_{i=n+1}^{n+k} X_i)\right] \\ &= \mathrm{E}\left[Y_n^2\right] + \mathrm{E}[Y_n]\,\mathrm{E}\left[\sum_{i=n+1}^{n+k} X_i\right] \\ &= n. \end{aligned} \tag{4}$$

If $k < 0$, then $n+k < n$. In this case, let $n' = n+k$ so that $n = n'-k = n'+|k|$. This enables us to write

$$\begin{aligned} R_Y[n,k] = \mathrm{E}[Y_n Y_{n+k}] &= \mathrm{E}\left[Y_{n'+|k|} Y_{n'}\right] \\ &= R_Y\left[n', |k|\right] = n' = n-k. \end{aligned} \tag{5}$$

To summarize the $k \geq 0$ and $k < 0$ cases, we have

$$R_Y[n,k] = \min(n, n+k). \tag{6}$$

This should be familiar because the Y_n process is just Brownian motion sampled at each time unit. The independent increments property is preserved in discrete time and the result is that we get the usual correlation function associated with an independent increments process.

Problem 13.6.3 Solution

We need to verify that $Y(t) = X(ct)$ satisfies the conditions given in Definition 13.10. First we observe that $Y(0) = X(c\cdot 0) = X(0) = 0$. Second, we note that since $X(t)$ is Brownian motion process implies that $Y(t) - Y(s) = X(ct) - X(cs)$ is a Gaussian random variable. Further, $X(ct) - X(cs)$ is independent of $X(t')$ for all $t' \leq cs$. Equivalently, we can say that $X(ct) - X(cs)$ is independent of $X(c\tau)$ for all $\tau \leq s$. In other words, $Y(t) - Y(s)$ is independent of $Y(\tau)$ for all $\tau \leq s$. Thus $Y(t)$ is a Brownian motion process.

Problem 13.6.4 Solution

First we observe that $Y_n = X_n - X_{n-1} = X(n) - X(n-1)$ is a Gaussian random variable with mean zero and variance α. Since this fact is true for all n, we can conclude that $Y_1, Y_2, \ldots$ are identically distributed. By Definition 13.10 for Brownian motion, $Y_n = X(n) - X(n-1)$ is independent of $X(m)$ for any $m \leq n-1$. Hence Y_n is independent of $Y_m = X(m) - X(m-1)$ for any $m \leq n-1$. Equivalently, $Y_1, Y_2, \ldots$ is a sequence of independent random variables.

Problem 13.6.5 Solution

Recall that the vector $\mathbf{X}$ of increments has independent components $X_n = W_n - W_{n-1}$. Alternatively, each W_n can be written as the sum

$$W_1 = X_1, \tag{1}$$

$$W_2 = X_1 + X_2, \tag{2}$$

$$\vdots$$

$$W_k = X_1 + X_2 + \cdots + X_k. \tag{3}$$

In terms of matrices, $\mathbf{W} = \mathbf{A}\mathbf{X}$ where $\mathbf{A}$ is the lower triangular matrix

$$\mathbf{A} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & & \ddots & \\ 1 & \cdots & \cdots & 1 \end{bmatrix}. \tag{4}$$

Since $\mathrm{E}[\mathbf{W}] = \mathbf{A}\,\mathrm{E}[\mathbf{X}] = \mathbf{0}$, it folows from Theorem 8.11 that

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{1}{|\det(\mathbf{A})|} f_{\mathbf{X}}\left(\mathbf{A}^{-1}\mathbf{w}\right). \tag{5}$$

Since $\mathbf{A}$ is a lower triangular matrix, $\det(\mathbf{A})$ is the product of its diagonal entries. In this case, $\det(\mathbf{A}) = 1$. In addition, reflecting the fact that each $X_n = W_n - W_{n-1}$,

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ 0 & -1 & 1 & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{-1}\mathbf{W} = \begin{bmatrix} W_1 \\ W_2 - W_1 \\ W_3 - W_2 \\ \vdots \\ W_k - W_{k-1} \end{bmatrix}. \tag{6}$$

Combining these facts with the observation that $f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=1}^{k} f_{X_n}(x_n)$, we can write

$$f_{\mathbf{W}}(\mathbf{w}) = f_{\mathbf{X}}\left(\mathbf{A}^{-1}\mathbf{w}\right) = \prod_{n=1}^{k} f_{X_n}(w_n - w_{n-1}), \tag{7}$$

which completes the missing steps in the proof of Theorem 13.8.

Problem 13.7.1 Solution

The discrete time autocovariance function is

$$C_X[m,k] = \mathrm{E}\left[(X_m - \mu_X)(X_{m+k} - \mu_X)\right]. \tag{1}$$

For $k = 0$, $C_X[m,0] = \mathrm{Var}[X_m] = \sigma_X^2$. For $k \neq 0$, X_m and X_{m+k} are independent so that

$$C_X[m,k] = \mathrm{E}\left[(X_m - \mu_X)\right] \mathrm{E}\left[(X_{m+k} - \mu_X)\right] = 0. \tag{2}$$

Thus the autocovariance of X_n is

$$C_X[m,k] = \begin{cases} \sigma_X^2 & k = 0, \\ 0 & k \neq 0. \end{cases} \tag{3}$$

Problem 13.7.2 Solution

Recall that $X(t) = t - W$ where $\mathrm{E}[W] = 1$ and $\mathrm{E}[W^2] = 2$.

(a) The mean is $\mu_X(t) = \mathrm{E}[t - W] = t - \mathrm{E}[W] = t - 1$.

(b) The autocovariance is

$$\begin{aligned} C_X(t,\tau) &= \mathrm{E}[X(t)X(t+\tau)] - \mu_X(t)\mu_X(t+\tau) \\ &= \mathrm{E}[(t-W)(t+\tau-W)] - (t-1)(t+\tau-1) \\ &= t(t+\tau) - \mathrm{E}[(2t+\tau)W] + \mathrm{E}\left[W^2\right] - t(t+\tau) + 2t + \tau - 1 \\ &= -(2t+\tau)\,\mathrm{E}[W] + 2 + 2t + \tau - 1 \\ &= 1. \end{aligned} \tag{1}$$

Problem 13.7.3 Solution

In this problem, the daily temperature process results from

$$C_n = 16\left[1 - \cos\frac{2\pi n}{365}\right] + 4X_n, \tag{1}$$

where X_n is an iid random sequence of $N[0,1]$ random variables. The hardest part of this problem is distinguishing between the process C_n and the covariance function $C_C[k]$.

(a) The expected value of the process is

$$\begin{aligned} \mathrm{E}[C_n] &= 16\,\mathrm{E}\left[1 - \cos\frac{2\pi n}{365}\right] + 4\,\mathrm{E}[X_n] \\ &= 16\left[1 - \cos\frac{2\pi n}{365}\right]. \end{aligned} \tag{2}$$

(b) Note that (1) and (2) imply

$$C_n - \mathrm{E}[C_n] = 4X_n. \tag{3}$$

This implies that the autocovariance of C_n is

$$\begin{aligned} C_C[m,k] &= \mathrm{E}[(C_m - \mathrm{E}[C_m])(C_{m+k} - \mathrm{E}[C_{m+k}])] \\ &= 16\,\mathrm{E}[X_m X_{m+k}] = \begin{cases} 16 & k=0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

(c) A model of this type may be able to capture the mean and variance of the daily temperature. However, one reason this model is overly simple is because day to day temperatures are uncorrelated. A more realistic model might incorporate the effects of "heat waves" or "cold spells" through correlated daily temperatures.

Problem 13.7.4 Solution

By repeated application of the recursion $C_n = C_{n-1}/2 + 4X_n$, we obtain

$$\begin{aligned} C_n &= \frac{C_{n-2}}{4} + 4\left[\frac{X_{n-1}}{2} + X_n\right] \\ &= \frac{C_{n-3}}{8} + 4\left[\frac{X_{n-2}}{4} + \frac{X_{n-1}}{2} + X_n\right] \\ &\vdots \\ &= \frac{C_0}{2^n} + 4\left[\frac{X_1}{2^{n-1}} + \frac{X_2}{2^{n-2}} + \cdots + X_n\right] = \frac{C_0}{2^n} + 4\sum_{i=1}^{n} \frac{X_i}{2^{n-i}}. \end{aligned} \tag{1}$$

(a) Since $C_0, X_1, X_2, \ldots$ all have zero mean,

$$\mathrm{E}[C_n] = \frac{\mathrm{E}[C_0]}{2^n} + 4\sum_{i=1}^{n} \frac{\mathrm{E}[X_i]}{2^{n-i}} = 0. \tag{2}$$

(b) The autocovariance is

$$C_C[m,k] = \mathrm{E}\left[\left(\frac{C_0}{2^n} + 4\sum_{i=1}^{n} \frac{X_i}{2^{n-i}}\right)\left(\frac{C_0}{2^m + k} + 4\sum_{j=1}^{m+k} \frac{X_j}{2^{m+k-j}}\right)\right]. \tag{3}$$

Since $C_0, X_1, X_2, \ldots$ are independent (and zero mean), $\mathrm{E}[C_0 X_i] = 0$. This implies

$$C_C[m,k] = \frac{\mathrm{E}\left[C_0^2\right]}{2^{2m+k}} + 16 \sum_{i=1}^{m} \sum_{j=1}^{m+k} \frac{\mathrm{E}\left[X_i X_j\right]}{2^{m-i} 2^{m+k-j}}. \tag{4}$$

For $i \neq j$, $\mathrm{E}[X_i X_j] = 0$ so that only the $i = j$ terms make any contribution to the double sum. However, at this point, we must consider the cases $k \geq 0$ and $k < 0$ separately. Since each X_i has variance 1, the autocovariance for $k \geq 0$ is

$$\begin{aligned} C_C[m,k] &= \frac{1}{2^{2m+k}} + 16 \sum_{i=1}^{m} \frac{1}{2^{2m+k-2i}} \\ &= \frac{1}{2^{2m+k}} + \frac{16}{2^k} \sum_{i=1}^{m} (1/4)^{m-i} \\ &= \frac{1}{2^{2m+k}} + \frac{16}{2^k} \frac{1-(1/4)^m}{3/4}. \end{aligned} \tag{5}$$

For $k < 0$, we can write

$$\begin{aligned} C_C[m,k] &= \frac{\mathrm{E}\left[C_0^2\right]}{2^{2m+k}} + 16 \sum_{i=1}^{m} \sum_{j=1}^{m+k} \frac{\mathrm{E}\left[X_i X_j\right]}{2^{m-i} 2^{m+k-j}} \\ &= \frac{1}{2^{2m+k}} + 16 \sum_{i=1}^{m+k} \frac{1}{2^{2m+k-2i}} \\ &= \frac{1}{2^{2m+k}} + \frac{16}{2^{-k}} \sum_{i=1}^{m+k} (1/4)^{m+k-i} \\ &= \frac{1}{2^{2m+k}} + \frac{16}{2^k} \frac{1-(1/4)^{m+k}}{3/4}. \end{aligned} \tag{6}$$

A general expression that's valid for all m and k is

$$C_C[m,k] = \frac{1}{2^{2m+k}} + \frac{16}{2^{|k|}} \frac{1-(1/4)^{\min(m,m+k)}}{3/4}. \tag{7}$$

(c) Since $\mathrm{E}[C_i] = 0$ for all i, our model has a mean daily temperature of zero degrees Celsius for the entire year. This is not a reasonable model for a year.

(d) For the month of January, a mean temperature of zero degrees Celsius seems quite reasonable. we can calculate the variance of C_n by evaluating the covariance at $n = m$. This yields

$$\mathrm{Var}[C_n] = \frac{1}{4^n} + \frac{16}{4^n}\frac{4(4^n - 1)}{3}. \tag{8}$$

Note that the variance is upper bounded by

$$\mathrm{Var}[C_n] \leq 64/3. \tag{9}$$

Hence the daily temperature has a standard deviation of $8/\sqrt{3} \approx 4.6$ degrees. Without actual evidence of daily temperatures in January, this model is more difficult to discredit.

Problem 13.7.5 Solution

This derivation of the Poisson process covariance is almost identical to the derivation of the Brownian motion autocovariance since both rely on the use of independent increments. From the definition of the Poisson process, we know that $\mu_N(t) = \lambda t$. When $\tau \geq 0$, we can write

$$\begin{aligned} C_N(t, \tau) &= \mathrm{E}\left[N(t)N(t+\tau)\right] - (\lambda t)[\lambda(t+\tau)] \\ &= \mathrm{E}\left[N(t)[(N(t+\tau) - N(t)) + N(t)]\right] - \lambda^2 t(t+\tau) \\ &= \mathrm{E}\left[N(t)[N(t+\tau) - N(t)]\right] + \mathrm{E}\left[N^2(t)\right] - \lambda^2 t(t+\tau). \end{aligned} \tag{1}$$

By the definition of the Poisson process, $N(t+\tau) - N(t)$ is the number of arrivals in the interval $[t, t+\tau)$ and is independent of $N(t)$ for $\tau > 0$. This implies

$$\begin{aligned} \mathrm{E}\left[N(t)[N(t+\tau) - N(t)]\right] &= \mathrm{E}\left[N(t)\right]\mathrm{E}\left[N(t+\tau) - N(t)\right] \\ &= \lambda t[\lambda(t+\tau) - \lambda t]. \end{aligned} \tag{2}$$

Note that since $N(t)$ is a Poisson random variable, $\mathrm{Var}[N(t)] = \lambda t$. Hence

$$\mathrm{E}\left[N^2(t)\right] = \mathrm{Var}[N(t)] + (\mathrm{E}\left[N(t)\right]^2 = \lambda t + (\lambda t)^2. \tag{3}$$

Therefore, for $\tau \geq 0$,

$$C_N(t,\tau) = \lambda t[\lambda(t+\tau) - \lambda t) + \lambda t + (\lambda t)^2 - \lambda^2 t(t+\tau) = \lambda t. \tag{4}$$

If $\tau < 0$, then we can interchange the labels t and $t+\tau$ in the above steps to show $C_N(t,\tau) = \lambda(t+\tau)$. For arbitrary t and τ, we can combine these facts to write

$$C_N(t,\tau) = \lambda \min(t, t+\tau). \tag{5}$$

Problem 13.7.6 Solution

$$\begin{aligned} \mathrm{E}\left[Y(t)\right] &= \sum_{n=0}^{\infty} \mathrm{E}\left[Y(t)|N(t)=n\right] P_{N(t)}(n) \\ &= \sum_{n=0}^{\infty} \mathrm{E}\left[X_n|N(t)=n\right] P_{N(t)}(n)\,. \end{aligned} \tag{1}$$

Since $N(t)$ and X_n are independent, $\mathrm{E}[X_n|N(t)=n] = \mathrm{E}[X_n] = 0$. This implies

$$\mathrm{E}\left[Y(t)\right] = \sum_{n=0}^{\infty} 0 \cdot P_{N(t)}(n) = 0. \tag{2}$$

For the autocovariance function, we have

$$C_Y(t,\tau) = \mathrm{E}\left[Y(t)Y(t+\tau)\right] = \mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}\right]. \tag{3}$$

In this case, what matters is whether $N(t)$ and $N(t+\tau)$ are the same since if they are the same then $X_{N(t)}$ and $X_{N(t+\tau)}$ will be one in the same random variable. If $\tau > 0$ then $\mathrm{P}[N(t+\tau) = N(t)]$ is the probability of zero arrivals of the $N(t)$ process in the interval $[t, t+\tau]$. Thus,

$$\mathrm{P}\left[N(t+\tau) = N(t)\right] = \mathrm{P}\left[N(t+\tau) - N(t) = 0\right] = e^{-\tau}. \tag{4}$$

If $\tau < 0$, then

$$\mathrm{P}\left[N(t+\tau) = N(t)\right] = \mathrm{P}\left[N(t) - N(t+\tau) = 0\right] = e^{-|\tau|}. \tag{5}$$

That is, for all τ,

$$\mathrm{P}\left[N(t) = N(t+\tau)\right] = e^{-|\tau|}. \tag{6}$$

Further when $N(t) = N(t+\tau)$,

$$\begin{aligned}
\mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}|N(t) = N(t+\tau)\right] &= \mathrm{E}\left[X_{N(t)}^2|N(t)\right] \\
&= \mathrm{E}\left[X_{N(t)}^2\right] = \sigma^2,
\end{aligned} \tag{7}$$

and

$$\mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}|N(t) \neq N(t+\tau)\right] = \mathrm{E}\left[X_{N(t)}\right]\mathrm{E}\left[X_{N(t+\tau)}\right] = 0. \tag{8}$$

We now write

$$\begin{aligned}
C_Y(t,\tau) &= \mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}|N(t) = N(t+\tau)\right] e^{-|\tau|} \\
&\qquad + \mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}|N(t) \neq N(t+\tau)\right](1 - e^{-|\tau|}) \\
&= \sigma^2 e^{-|\tau|}.
\end{aligned} \tag{9}$$

A more mechanical procedure that leads to the same answer is, for $\tau > 0$, to define $N = N(t)$ and $K = N(t+\tau) - N(t)$. This implies N and K are independent Poisson random variables with expected values $\mathrm{E}[N] = t$ and $\mathrm{E}[K] = \tau$. Using iterated expectation, we have

$$\begin{aligned}
C_Y(t,\tau) &= \mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}\right] \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \mathrm{E}\left[X_{N(t)}X_{N(t+\tau)}|N = n, K = k\right] P_N(n)\, P_K(k) \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \mathrm{E}\left[X_n X_{n+k}|N = n, K = k\right] P_N(n)\, P_K(k) \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \mathrm{E}\left[X_n X_{n+k}\right] P_N(n)\, P_K(k) \\
&= \sum_{n=0}^{\infty} \mathrm{E}\left[X_n^2\right] P_N(n)\, P_K(0) \\
&= \sum_{n=0}^{\infty} \sigma^2 e^{-\tau} P_N(n) = \sigma^2 e^{-\tau}.
\end{aligned} \tag{10}$$

The key step in the above calculation was that for $k > 0$,

$$\mathrm{E}[X_n X_{n+k}] = \mathrm{E}[X_n]\,\mathrm{E}[X_{n+k}] = 0. \tag{11}$$

For $\tau < 0$, we define $N = N(t+\tau)$ and $K = N(t) - N(t+\tau)$. In this case, K is Poisson with expected value $|\tau|$ and $P_K(0) = e^{-|\tau|}$. The end result gives the same answer as the previous method: $C_Y(t,\tau) = \sigma^2 e^{-|\tau|}$.

Problem 13.7.7 Solution

Since the X_n are independent,

$$\mathrm{E}[Y_n] = \mathrm{E}[X_{n-1}X_n] = \mathrm{E}[X_{n-1}]\,\mathrm{E}[X_n] = 0. \tag{1}$$

Thus the autocovariance function is

$$C_Y[n,k] = \mathrm{E}[Y_n Y_{n+k}] = \mathrm{E}[X_{n-1}X_n X_{n+k-1}X_{n+k}]. \tag{2}$$

To calculate this expectation, what matters is whether any of the four terms in the product are the same. This reduces to five cases:

1. $n+k-1 > n$, or equivalently $k > 1$:

   In this case, we have

   $$n-1 < n < n+k-1 < n+k, \tag{3}$$

   implying that X_{n-1}, X_n, X_{n+k-1} and X_{n+k} are independent. It follows that

   $$\begin{aligned}\mathrm{E}[X_{n-1}X_n X_{n+k-1}X_{n+k}] &= \mathrm{E}[X_{n-1}]\,\mathrm{E}[X_n]\,\mathrm{E}[X_{n+k-1}]\,\mathrm{E}[X_{n+k}]\\ &= 0. \end{aligned}\tag{4}$$

2. $n+k < n-1$, or equivalently $k < -1$:

   In this case, we have

   $$n+k-1 < n+k < n-1 < n, \tag{5}$$

   implying that X_{n+k-1} X_{n+k}, X_{n-1}, and X_n are independent. It follows that

   $$\begin{aligned}\mathrm{E}[X_{n-1}X_n X_{n+k-1}X_{n+k}] &= \mathrm{E}[X_{n-1}]\,\mathrm{E}[X_n]\,\mathrm{E}[X_{n+k-1}]\,\mathrm{E}[X_{n+k}]\\ &= 0. \end{aligned}\tag{6}$$

3. $k = -1$:

   In this case, we have

$$\begin{aligned} \mathrm{E}\left[X_{n-1}X_nX_{n+k-1}X_{n+k}\right] &= \mathrm{E}\left[X_{n-1}^2X_nX_{n-2}\right] \\ &= \mathrm{E}\left[X_{n-1}^2\right]\mathrm{E}\left[X_n\right]\mathrm{E}\left[X_{n-2}\right] = 0. \end{aligned} \tag{7}$$

4. $k = 0$:

   In this case, we have

$$\begin{aligned} \mathrm{E}\left[X_{n-1}X_nX_{n+k-1}X_{n+k}\right] &= \mathrm{E}\left[X_{n-1}^2X_n^2\right] \\ &= \mathrm{E}\left[X_{n-1}^2\right]\mathrm{E}\left[X_n^2\right] = 9. \end{aligned} \tag{8}$$

5. $k = 1$:

   In this case, we have

$$\begin{aligned} \mathrm{E}\left[X_{n-1}X_nX_{n+k-1}X_{n+k}\right] &= \mathrm{E}\left[X_{n-1}X_n^2X_{n+1}\right] \\ &= \mathrm{E}\left[X_{n-1}\right]\mathrm{E}\left[X_n^2\right]\mathrm{E}\left[X_{n+1}\right] = 0. \end{aligned} \tag{9}$$

Combining these case, we find that

$$C_Y[n,k] = \begin{cases} 9 & k = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Problem 13.8.1 Solution

For a set of samples $Y(t_1), \ldots, Y(t_k)$, we observe that $Y(t_j) = X(t_j + a)$. This implies

$$f_{Y(t_1),\ldots,Y(t_k)}(y_1, \ldots, y_k) = f_{X(t_1+a),\ldots,X(t_k+a)}(y_1, \ldots, y_k). \tag{1}$$

Thus,

$$f_{Y(t_1+\tau),\ldots,Y(t_k+\tau)}(y_1, \ldots, y_k) = f_{X(t_1+\tau+a),\ldots,X(t_k+\tau+a)}(y_1, \ldots, y_k). \tag{2}$$

Since $X(t)$ is a stationary process,

$$f_{X(t_1+\tau+a),\ldots,X(t_k+\tau+a)}(y_1,\ldots,y_k) = f_{X(t_1+a),\ldots,X(t_k+a)}(y_1,\ldots,y_k)\,. \tag{3}$$

This implies

$$\begin{aligned} f_{Y(t_1+\tau),\ldots,Y(t_k+\tau)}(y_1,\ldots,y_k) &= f_{X(t_1+a),\ldots,X(t_k+a)}(y_1,\ldots,y_k) \\ &= f_{Y(t_1),\ldots,Y(t_k)}(y_1,\ldots,y_k)\,. \end{aligned} \tag{4}$$

We can conclude that $Y(t)$ is a stationary process.

Problem 13.8.2 Solution

The short answer is No. For the given process $X(t)$,

$$\mathrm{Var}[X(t_1)] = C_{11} = 2, \qquad \mathrm{Var}[X(t_2)] = C_{22} = 1. \tag{1}$$

However, stationarity of $X(t)$ requires $\mathrm{Var}[X(t_1)] = \mathrm{Var}[X(t_2)]$, which is a contradiction.

Problem 13.8.3 Solution

For an arbitrary set of samples $Y(t_1),\ldots,Y(t_k)$, we observe that $Y(t_j) = X(at_j)$. This implies

$$f_{Y(t_1),\ldots,Y(t_k)}(y_1,\ldots,y_k) = f_{X(at_1),\ldots,X(at_k)}(y_1,\ldots,y_k)\,. \tag{1}$$

Thus,

$$f_{Y(t_1+\tau),\ldots,Y(t_k+\tau)}(y_1,\ldots,y_k) = f_{X(at_1+a\tau),\ldots,X(at_k+a\tau)}(y_1,\ldots,y_k)\,. \tag{2}$$

We see that a time offset of τ for the $Y(t)$ process corresponds to an offset of time $\tau' = a\tau$ for the $X(t)$ process. Since $X(t)$ is a stationary process,

$$\begin{aligned} f_{Y(t_1+\tau),\ldots,Y(t_k+\tau)}(y_1,\ldots,y_k) &= f_{X(at_1+\tau'),\ldots,X(at_k+\tau')}(y_1,\ldots,y_k) \\ &= f_{X(at_1),\ldots,X(at_k)}(y_1,\ldots,y_k) \\ &= f_{Y(t_1),\ldots,Y(t_k)}(y_1,\ldots,y_k)\,. \end{aligned} \tag{3}$$

We can conclude that $Y(t)$ is a stationary process.

Problem 13.8.4 Solution

Since $Y_{n_i+k} = X((n_i+k)\Delta)$ for a set of time samples $n_1, \ldots, n_m$ and an offset k,

$$f_{Y_{n_1+k},\ldots,Y_{n_m+k}}(y_1,\ldots,y_m) = f_{X((n_1+k)\Delta),\ldots,X((n_m+k)\Delta)}(y_1,\ldots,y_m). \tag{1}$$

Since $X(t)$ is a stationary process,

$$f_{X((n_1+k)\Delta),\ldots,X((n_m+k)\Delta)}(y_1,\ldots,y_m) = f_{X(n_1\Delta),\ldots,X(n_m\Delta)}(y_1,\ldots,y_m). \tag{2}$$

Since $X(n_i\Delta) = Y_{n_i}$, we see that

$$f_{Y_{n_1+k},\ldots,Y_{n_m+k}}(y_1,\ldots,y_m) = f_{Y_{n_1},\ldots,Y_{n_m}}(y_1,\ldots,y_m). \tag{3}$$

Hence Y_n is a stationary random sequence.

Problem 13.8.5 Solution

Since $Y_n = X_{kn}$,

$$f_{Y_{n_1+l},\ldots,Y_{n_m+l}}(y_1,\ldots,y_m) = f_{X_{kn_1+kl},\ldots,X_{kn_m+kl}}(y_1,\ldots,y_m) \tag{1}$$

Stationarity of the X_n process implies

$$\begin{aligned} f_{X_{kn_1+kl},\ldots,X_{kn_m+kl}}(y_1,\ldots,y_m) &= f_{X_{kn_1},\ldots,X_{kn_m}}(y_1,\ldots,y_m) \\ &= f_{Y_{n_1},\ldots,Y_{n_m}}(y_1,\ldots,y_m). \end{aligned} \tag{2}$$

We combine these steps to write

$$f_{Y_{n_1+l},\ldots,Y_{n_m+l}}(y_1,\ldots,y_m) = f_{Y_{n_1},\ldots,Y_{n_m}}(y_1,\ldots,y_m). \tag{3}$$

Thus Y_n is a stationary process.

Problem 13.8.6 Solution

Given $A = a$, $Y(t) = aX(t)$ which is a special case of $Y(t) = aX(t) + b$ given in Theorem 13.10. Applying the result of Theorem 13.10 with $b = 0$ yields

$$f_{Y(t_1),\ldots,Y(t_n)|A}(y_1,\ldots,y_n|a) = \frac{1}{a^n} f_{X(t_1),\ldots,X(t_n)}\left(\frac{y_1}{a},\ldots,\frac{y_n}{a}\right). \tag{1}$$

Integrating over the PDF $f_A(a)$ yields

$$\begin{aligned} f_{Y(t_1),\ldots,Y(t_n)}(y_1,\ldots,y_n) &= \int_0^\infty f_{Y(t_1),\ldots,Y(t_n)|A}(y_1,\ldots,y_n|a)\, f_A(a)\; da \\ &= \int_0^\infty \frac{1}{a^n} f_{X(t_1),\ldots,X(t_n)}\left(\frac{y_1}{a},\ldots,\frac{y_n}{a}\right) f_A(a)\; da. \qquad (2) \end{aligned}$$

This complicated expression can be used to find the joint PDF

$$\begin{aligned} &f_{Y(t_1+\tau),\ldots,Y(t_n+\tau)}(y_1,\ldots,y_n) \\ &\quad = \int_0^\infty \frac{1}{a^n} f_{X(t_1+\tau),\ldots,X(t_n+\tau)}\left(\frac{y_1}{a},\ldots,\frac{y_n}{a}\right) f_A(a)\; da. \qquad (3) \end{aligned}$$

Since $X(t)$ is a stationary process, the joint PDF of $X(t_1+\tau),\ldots,X(t_n+\tau)$ is the same as the joint PDF of $X(t_1),\ldots,X(t_n)$. Thus

$$\begin{aligned} &f_{Y(t_1+\tau),\ldots,Y(t_n+\tau)}(y_1,\ldots,y_n) \\ &\quad = \int_0^\infty \frac{1}{a^n} f_{X(t_1+\tau),\ldots,X(t_n+\tau)}\left(\frac{y_1}{a},\ldots,\frac{y_n}{a}\right) f_A(a)\; da \\ &\quad = \int_0^\infty \frac{1}{a^n} f_{X(t_1),\ldots,X(t_n)}\left(\frac{y_1}{a},\ldots,\frac{y_n}{a}\right) f_A(a)\; da \\ &\quad = f_{Y(t_1),\ldots,Y(t_n)}(y_1,\ldots,y_n). \qquad (4) \end{aligned}$$

We can conclude that $Y(t)$ is a stationary process.

Problem 13.8.7 Solution

Since $g(\cdot)$ is an unspecified function, we will work with the joint CDF of $Y(t_1+\tau),\ldots,Y(t_n+\tau)$. To show $Y(t)$ is a stationary process, we will show that for all τ,

$$F_{Y(t_1+\tau),\ldots,Y(t_n+\tau)}(y_1,\ldots,y_n) = F_{Y(t_1),\ldots,Y(t_n)}(y_1,\ldots,y_n). \qquad (1)$$

By taking partial derivatives with respect to $y_1,\ldots,y_n$, it should be apparent that this implies that the joint PDF $f_{Y(t_1+\tau),\ldots,Y(t_n+\tau)}(y_1,\ldots,y_n)$ will not depend on τ. To proceed, we write

$$\begin{aligned} &F_{Y(t_1+\tau),\ldots,Y(t_n+\tau)}(y_1,\ldots,y_n) \\ &\quad = \mathrm{P}\left[Y(t_1+\tau)\le y_1,\ldots,Y(t_n+\tau)\le y_n\right] \\ &\quad = \mathrm{P}\Bigg[\underbrace{g(X(t_1+\tau))\le y_1,\ldots,g(X(t_n+\tau))\le y_n}_{A_\tau}\Bigg]. \qquad (2) \end{aligned}$$

In principle, we can calculate $P[A_\tau]$ by integrating $f_{X(t_1+\tau),\ldots,X(t_n+\tau)}(x_1,\ldots,x_n)$ over the region corresponding to event A_τ. Since $X(t)$ is a stationary process,

$$f_{X(t_1+\tau),\ldots,X(t_n+\tau)}(x_1,\ldots,x_n) = f_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n). \tag{3}$$

This implies $P[A_\tau]$ does not depend on τ. In particular,

$$\begin{aligned} F_{Y(t_1+\tau),\ldots,Y(t_n+\tau)}(y_1,\ldots,y_n) &= P[A_\tau] \\ &= P[g(X(t_1)) \le y_1,\ldots,g(X(t_n)) \le y_n] \\ &= F_{Y(t_1),\ldots,Y(t_n)}(y_1,\ldots,y_n). \end{aligned} \tag{4}$$

Problem 13.9.1 Solution

The autocorrelation function $R_X(\tau) = \delta(\tau)$ is mathematically valid in the sense that it meets the conditions required in Theorem 13.12. That is,

$$R_X(\tau) = \delta(\tau) \ge 0, \tag{1}$$
$$R_X(\tau) = \delta(\tau) = \delta(-\tau) = R_X(-\tau), \tag{2}$$
$$R_X(\tau) \le R_X(0) = \delta(0). \tag{3}$$

However, for a process $X(t)$ with the autocorrelation $R_X(\tau) = \delta(\tau)$, Definition 13.16 says that the average power of the process is

$$E\left[X^2(t)\right] = R_X(0) = \delta(0) = \infty. \tag{4}$$

Processes with infinite average power cannot exist in practice.

Problem 13.9.2 Solution

Since $Y(t) = A + X(t)$, the mean of $Y(t)$ is

$$E[Y(t)] = E[A] + E[X(t)] = E[A] + \mu_X. \tag{1}$$

The autocorrelation of $Y(t)$ is

$$\begin{aligned} R_Y(t,\tau) &= E[(A + X(t))(A + X(t+\tau))] \\ &= E\left[A^2\right] + E[A]\,E[X(t)] + A\,E[X(t+\tau)] + E[X(t)X(t+\tau)] \\ &= E\left[A^2\right] + 2\,E[A]\,\mu_X + R_X(\tau). \end{aligned} \tag{2}$$

We see that neither $E[Y(t)]$ nor $R_Y(t,\tau)$ depend on t. Thus $Y(t)$ is a wide sense stationary process.

Problem 13.9.3 Solution

TRUE: First we observe that $\mathrm{E}[Y_n] = \mathrm{E}[X_n] - \mathrm{E}[X_{n-1}] = 0$, which doesn't depend on n. Second, we verify that

$$\begin{aligned} C_Y[n,k] &= \mathrm{E}[Y_n Y_{n+k}] \\ &= \mathrm{E}[(X_n - X_{n-1})(X_{n+k} - X_{n+k-1})] \\ &= \mathrm{E}[X_n X_{n+k}] - \mathrm{E}[X_n X_{n+k-1}] \\ &\quad - \mathrm{E}[X_{n-1} X_{n+k}] - \mathrm{E}[X_{n-1} X_{n+k-1}] \\ &= C_X[k] - C_X[k-1] - C_X[k+1] + C_X[k], \end{aligned} \tag{1}$$

which doesn't depend on n. Hence Y_n is WSS.

Problem 13.9.4 Solution

Since X_n and X_{n+k} are independent for all $k \neq 0$,

$$C_X[n,k] = \mathrm{Cov}[X_n, X_{n+k}, =] 0, \qquad k \neq 0. \tag{1}$$

For $k = 0$,

$$C_X[n,0] = \mathrm{Var}[X_n] = \mathrm{E}\left[X_n^2\right] - (\mathrm{E}[X_n])^2 = p(1-p). \tag{2}$$

Thus

$$C_X[n,k] = C_X[k] = p(1-p)\delta[k] \tag{3}$$

where $\delta[k]$ is the Kronecker delta function. The autocorrelation function is simply

$$R_X[n,k] = C_X[n,k] + \mu_X^2 = p(1-p)\delta[k] + p^2. \tag{4}$$

Problem 13.9.5 Solution

For $k \neq 0$, X_n and X_{n+k} are independent so that

$$R_X[n,k] = \mathrm{E}[X_n X_{n+k}] = \mathrm{E}[X_n]\,\mathrm{E}[X_{n+k}] = \mu^2. \tag{1}$$

For $k = 0$,

$$R_X[n,0] = \mathrm{E}[X_n X_n] = \mathrm{E}\left[X_n^2\right] = \sigma^2 + \mu^2. \tag{2}$$

Combining these expressions, we obtain

$$R_X[n,k] = R_X[k] = \mu^2 + \sigma^2\delta[k], \tag{3}$$

where $\delta[k] = 1$ if $k = 0$ and is otherwise zero.

Problem 13.9.6 Solution

(a) TRUE: First we observe that $\mathrm{E}[V(t)] = \mathrm{E}[X(t)] + \mathrm{E}[Y(t)] = \mu_X + \mu_Y$. Next, we observe that

$$\begin{aligned} R_V(t,\tau) &= \mathrm{E}\left[V(t)V(t+\tau)\right] \\ &= \mathrm{E}\left[(X(t)+Y(t))(X(t+\tau)+Y(t+\tau))\right] \\ &= \mathrm{E}\left[X(t)X(t+\tau)\right] + \mathrm{E}\left[Y(t)X(t+\tau)\right] \\ &\qquad + \mathrm{E}\left[X(t)Y(t+\tau)\right] + \mathrm{E}\left[Y(t)Y(t+\tau)\right]. \end{aligned} \tag{1}$$

Since $X(t)$ and $Y(t)$ are independent processes,

$$\mathrm{E}\left[X(t_1)Y(t_2)\right] = \mathrm{E}\left[X(t_1)\right]\mathrm{E}\left[Y(t_2)\right] = \mu_X\mu_Y \tag{2}$$

for all t_1 and t_2. If follows that

$$R_V(t,\tau) = R_X(\tau) + R_Y(\tau) + 2\mu_X\mu_Y. \tag{3}$$

Thus the process $V(t)$ is wide sense stationary since $\mathrm{E}[V(t)]$ is constant and the autocorrelation $R_V(t,\tau)$ depends only on τ.

(b) TRUE: Independence of $X(t)$ and $Y(t)$ implies

$$\mathrm{E}\left[W(t)\right] = \mathrm{E}\left[X(t)Y(t)\right] = \mathrm{E}\left[X(t)\right]\mathrm{E}\left[Y(t)\right] = \mu_X\mu_Y \tag{4}$$

and

$$\begin{aligned} R_W(t,\tau) &= \mathrm{E}\left[W(t)W(t+\tau)\right] \\ &= \mathrm{E}\left[X(t)Y(t)X(t+\tau)Y(t+\tau)\right] \\ &= \mathrm{E}\left[X(t)X(t+\tau)\right]\mathrm{E}\left[Y(t)Y(t+\tau)\right] \qquad \text{(by independence)} \\ &= C_X(\tau)C_Y(\tau). \end{aligned} \tag{5}$$

Since $W(t)$ has constant expected value and the autocorrelation depends only on the time difference τ, $W(t)$ is wide-sense stationary.

Problem 13.9.7 Solution

FALSE: The autocorrelation of Y_n is

$$\begin{aligned}
R_Y[n,k] &= \mathrm{E}[Y_n Y_{n+k}] \\
&= \mathrm{E}\left[(X_n + (-1)^{n-1}X_{n-1})(X_{n+k} + (-1)^{n+k-1}X_{n+k-1})\right] \\
&= \mathrm{E}[X_n X_{n+k}] + \mathrm{E}\left[(-1)^{n-1}X_{n-1}X_{n+k}\right] \\
&\quad + \mathrm{E}\left[X_n(-1)^{n+k-1}X_{n+k-1}\right] + \mathrm{E}\left[(-1)^{2n+k-2}X_{n-1}X_{n+k-1}\right] \\
&= R_X[k] + (-1)^{n-1}R_X[k+1] \\
&\qquad + (-1)^{n+k-1}R_X[k+1] + (-1)^k R_X[k] \\
&= [1+(-1)^k](R_X[k] + (-1)^{n-1}R_X[k+1]), \qquad (1)
\end{aligned}$$

which depends on n.

Problem 13.9.8 Solution

In this problem, we find the autocorrelation $R_W(t,\tau)$ when

$$W(t) = X\cos 2\pi f_0 t + Y\sin 2\pi f_0 t, \qquad (1)$$

and X and Y are uncorrelated random variables with $\mathrm{E}[X] = \mathrm{E}[Y] = 0$.

We start by writing

$$\begin{aligned}
&R_W(t,\tau) \\
&= \mathrm{E}[W(t)W(t+\tau)] \\
&= \mathrm{E}[(X\cos 2\pi f_0 t + Y\sin 2\pi f_0 t)(X\cos 2\pi f_0(t+\tau) + Y\sin 2\pi f_0(t+\tau))]. \qquad (2)
\end{aligned}$$

Since X and Y are uncorrelated, $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y] = 0$. Thus, when we expand $\mathrm{E}[W(t)W(t+\tau)]$ and take the expectation, all of the XY cross terms will be zero. This implies

$$\begin{aligned}
R_W(t,\tau) = \mathrm{E}\left[X^2\right]&\cos 2\pi f_0 t\cos 2\pi f_0(t+\tau) \\
&+ \mathrm{E}\left[Y^2\right]\sin 2\pi f_0 t\sin 2\pi f_0(t+\tau). \qquad (3)
\end{aligned}$$

Since $\mathrm{E}[X] = \mathrm{E}[Y] = 0$,

$$\mathrm{E}\left[X^2\right] = \mathrm{Var}[X] - (\mathrm{E}[X])^2 = \sigma^2, \qquad (4)$$

$$\mathrm{E}\left[Y^2\right] = \mathrm{Var}[Y] - (\mathrm{E}[Y])^2 = \sigma^2. \qquad (5)$$

In addition, from Math Fact B.2, we use the formulas

$$\cos A \cos B = \frac{1}{2}\left[\cos(A-B) + \cos(A+B)\right], \tag{6}$$

$$\sin A \sin B = \frac{1}{2}\left[\cos(A-B) - \cos(A+B)\right], \tag{7}$$

to write

$$\begin{aligned} R_W(t,\tau) &= \frac{\sigma^2}{2}\left(\cos 2\pi f_0\tau + \cos 2\pi f_0(2t+\tau)\right) \\ &\quad + \frac{\sigma^2}{2}\left(\cos 2\pi f_0\tau - \cos 2\pi f_0(2t+\tau)\right) \\ &= \sigma^2 \cos 2\pi f_0\tau \end{aligned} \tag{8}$$

Thus $R_W(t,\tau) = R_W(\tau)$. Since

$$\mathrm{E}\left[W(t)\right] = \mathrm{E}\left[X\right]\cos 2\pi f_0 t + \mathrm{E}\left[Y\right]\sin 2\pi f_0 t = 0, \tag{9}$$

we can conclude that $W(t)$ is a wide sense stationary process. However, we note that if $\mathrm{E}[X^2] \neq \mathrm{E}[Y^2]$, then the $\cos 2\pi f_0(2t+\tau)$ terms in $R_W(t,\tau)$ would not cancel and $W(t)$ would not be wide sense stationary.

Problem 13.9.9 Solution

(a) In the problem statement, we are told that $X(t)$ has average power equal to 1. By Definition 13.16, the average power of $X(t)$ is $\mathrm{E}[X^2(t)] = 1$.

(b) Since Θ has a uniform PDF over $[0, 2\pi]$,

$$f_\Theta(\theta) = \begin{cases} 1/(2\pi) & 0 \leq \theta \leq 2\pi, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expected value of the random phase cosine is

$$
\begin{aligned}
\mathrm{E}\left[\cos(2\pi f_c t + \Theta)\right] &= \int_{-\infty}^{\infty} \cos(2\pi f_c t + \theta) f_\Theta(\theta)\ d\theta \\
&= \int_0^{2\pi} \cos(2\pi f_c t + \theta) \frac{1}{2\pi} d\theta \\
&= \frac{1}{2\pi} \sin(2\pi f_c t + \theta)\Big|_0^{2\pi} \\
&= \frac{1}{2\pi} \left(\sin(2\pi f_c t + 2\pi) - \sin(2\pi f_c t)\right) = 0. \qquad (2)
\end{aligned}
$$

(c) Since $X(t)$ and Θ are independent,

$$
\begin{aligned}
\mathrm{E}\left[Y(t)\right] &= \mathrm{E}\left[X(t)\cos(2\pi f_c t + \Theta)\right] \\
&= \mathrm{E}\left[X(t)\right] \mathrm{E}\left[\cos(2\pi f_c t + \Theta)\right] = 0. \qquad (3)
\end{aligned}
$$

Note that the mean of $Y(t)$ is zero no matter what the mean of $X(t)$ since the random phase cosine has zero mean.

(d) Independence of $X(t)$ and Θ results in the average power of $Y(t)$ being

$$
\begin{aligned}
\mathrm{E}\left[Y^2(t)\right] &= \mathrm{E}\left[X^2(t)\cos^2(2\pi f_c t + \Theta)\right] \\
&= \mathrm{E}\left[X^2(t)\right] \mathrm{E}\left[\cos^2(2\pi f_c t + \Theta)\right] \\
&= \mathrm{E}\left[\cos^2(2\pi f_c t + \Theta)\right]. \qquad (4)
\end{aligned}
$$

Note that we have used the fact from part (a) that $X(t)$ has unity average power. To finish the problem, we use the trigonometric identity $\cos^2\phi = (1 + \cos 2\phi)/2$. This yields

$$
\mathrm{E}\left[Y^2(t)\right] = \mathrm{E}\left[\frac{1}{2}\left(1 + \cos(2\pi(2f_c)t + \Theta)\right)\right] = 1/2. \qquad (5)
$$

Note that $\mathrm{E}[\cos(2\pi(2f_c)t + \Theta)] = 0$ by the argument given in part (b) with $2f_c$ replacing f_c.

Problem 13.9.10 Solution

This proof simply parallels the proof of Theorem 13.12. For the first item, $R_X[0] = R_X[m,0] = \mathrm{E}[X_m^2]$. Since $X_m^2 \geq 0$, we must have $\mathrm{E}[X_m^2] \geq 0$. For the second item, Definition 13.13 implies that

$$\begin{aligned} R_X[k] = R_X[m,k] &= \mathrm{E}[X_m X_{m+k}] \\ &= \mathrm{E}[X_{m+k} X_m] = R_X[m+k,-k]. \end{aligned} \tag{1}$$

Since X_m is wide sense stationary, $R_X[m+k,-k] = R_X[-k]$. The final item requires more effort. First, we note that when X_m is wide sense stationary, $\mathrm{Var}[X_m] = C_X[0]$, a constant for all t. Second, Theorem 5.14 says that

$$|C_X[m,k]| \leq \sigma_{X_m}\sigma_{X_{m+k}} = C_X[0]. \tag{2}$$

Note that $C_X[m,k] \leq |C_X[m,k]|$, and thus it follows that

$$C_X[m,k] \leq \sigma_{X_m}\sigma_{X_{m+k}} = C_X[0]. \tag{3}$$

(This little step was unfortunately omitted from the proof of Theorem 13.12.)

Now for any numbers a, b, and c, if $a \leq b$ and $c \geq 0$, then $(a+c)^2 \leq (b+c)^2$. Choosing $a = C_X[m,k]$, $b = C_X[0]$, and $c = \mu_X^2$ yields

$$\left(C_X[m,m+k] + \mu_X^2\right)^2 \leq \left(C_X[0] + \mu_X^2\right)^2. \tag{4}$$

In the above expression, the left side equals $(R_X[k])^2$ while the right side is $(R_X[0])^2$, which proves the third part of the theorem.

Problem 13.9.11 Solution

The solution to this problem is essentially the same as the proof of Theorem 13.13 except integrals are replaced by sums. First we verify that $\overline{X}_m$ is unbiased:

$$\begin{aligned} \mathrm{E}\left[\overline{X}_m\right] &= \frac{1}{2m+1}\mathrm{E}\left[\sum_{n=-m}^{m} X_n\right] \\ &= \frac{1}{2m+1}\sum_{n=-m}^{m}\mathrm{E}[X_n] = \frac{1}{2m+1}\sum_{n=-m}^{m}\mu_X = \mu_X. \end{aligned} \tag{1}$$

To show consistency, it is sufficient to show that $\lim_{m\to\infty} \text{Var}[\overline{X}_m] = 0$. First, we observe that $\overline{X}_m - \mu_X = \frac{1}{2m+1}\sum_{n=-m}^{m}(X_n - \mu_X)$. This implies

$$\begin{aligned}
\text{Var}[\overline{X}(T)] &= \text{E}\left[\left(\frac{1}{2m+1}\sum_{n=-m}^{m}(X_n-\mu_X)\right)^2\right] \\
&= \text{E}\left[\frac{1}{(2m+1)^2}\left(\sum_{n=-m}^{m}(X_n-\mu_X)\right)\left(\sum_{n'=-m}^{m}(X_{n'}-\mu_X)\right)\right] \\
&= \frac{1}{(2m+1)^2}\sum_{n=-m}^{m}\sum_{n'=-m}^{m}\text{E}\left[(X_n-\mu_X)(X_{n'}-\mu_X)\right] \\
&= \frac{1}{(2m+1)^2}\sum_{n=-m}^{m}\sum_{n'=-m}^{m}C_X\left[n'-n\right]. \qquad (2)
\end{aligned}$$

We note that

$$\begin{aligned}
\sum_{n'=-m}^{m} C_X\left[n'-n\right] &\le \sum_{n'=-m}^{m}\left|C_X\left[n'-n\right]\right| \\
&\le \sum_{n'=-\infty}^{\infty}\left|C_X\left[n'-n\right]\right| = \sum_{k=-\infty}^{\infty}|C_X(k)| < \infty. \qquad (3)
\end{aligned}$$

Hence there exists a constant K such that

$$\text{Var}[\overline{X}_m] \le \frac{1}{(2m+1)^2}\sum_{n=-m}^{m}K = \frac{K}{2m+1}. \qquad (4)$$

Thus $\lim_{m\to\infty}\text{Var}[\overline{X}_m] \le \lim_{m\to\infty}\frac{K}{2m+1} = 0$.

Problem 13.9.12 Solution

(a) TRUE: To verify this, we need to show that $\text{E}[V_n]$ and $R_V[n,k]$ do not depend on t. First for the expected value, independence of X_n and Y_n imply

$$\text{E}\left[V_n\right] = \text{E}\left[\frac{X_n}{Y_n}\right] = \text{E}\left[X_n\right]\text{E}\left[\frac{1}{Y_n}\right]. \qquad (1)$$

It then follows from Y_n being stationary that

$$\mathrm{E}\left[V_n\right] = \mu_X \int_{-\infty}^{\infty} \frac{1}{y} f_{Y_n}(y)\; dy = \mu_X \int_{-\infty}^{\infty} \frac{1}{y} f_{Y(0)}(y)\; dy. \tag{2}$$

Note that the above expression cannot be simplified, however, all that matters is that it does not depend on the index n. Now for the autocorrelation, independence of the X_n and Y_n processes implies

$$\begin{aligned} R_V\left[n,k\right] &= \mathrm{E}\left[\frac{X_n X_{n+k}}{Y_n Y_{n+k}}\right] \\ &= \mathrm{E}\left[X_n X_{n+k}\right] \mathrm{E}\left[\frac{1}{Y_n Y_{n+k}}\right] \\ &= R_X\left[k\right] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{y_1 y_2} f_{Y_n,Y_{n+k}}(y_1,y_2)\; dy_1\, dy_2. \end{aligned} \tag{3}$$

Stationarity of Y_n implies $f_{Y_n,Y_{n+k}}(y_1,y_2) = f_{Y_0,Y_k}(y_1,y_2)$ and thus

$$R_V\left[n,k\right] = R_X\left[k\right] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{y_1 y_2} f_{Y(0),Y_k}(y_1,y_2)\; dy_1\, dy_2, \tag{4}$$

which depends only on the time shift k. Thus V_n is wide sense stationary.

(b) FALSE: To verify this, we will construct a counterexample in which X_n and Y_n are independent and wide sense stationary, but $\mathrm{E}[W_n]$ depends on n. Let X_n denote any stationary process with $\mathrm{E}[X_n] = \mu_X \neq 0$. Note that independence of X_n and Y_n implies

$$\mathrm{E}\left[W_n\right] = \mathrm{E}\left[\frac{X_n}{Y_n}\right] = \mathrm{E}\left[X_n\right] \mathrm{E}\left[\frac{1}{Y_n}\right] = \mu_X\, \mathrm{E}\left[\frac{1}{Y_n}\right]. \tag{5}$$

Consider the random sequence Y_n such that

$$f_{Y_{2n}}(y) = f_{Y_0}(y) = \begin{cases} 2y/9 & 0 \leq y \leq 3, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

$$f_{Y_{2n+1}}(y) = f_{Y_1}(y) = \begin{cases} \frac{1}{2\gamma} & 2-\gamma \leq y \leq 2+\gamma, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $\gamma = \sqrt{3/2}$. We observe that

$$\mathrm{E}\left[Y_0\right] = \int_{-\infty}^{\infty} y f_{Y_0}(y)\ dy = \int_0^3 2y^2/9\, dy = 2, \tag{8}$$

$$\mathrm{E}\left[Y_0^2\right] = \int_{-\infty}^{\infty} y^2 f_{Y_0}(y)\ dy = \int_0^3 2y^3/9\, dy = 9/2. \tag{9}$$

Thus $\mathrm{Var}[Y_0] = 1/2$. In addition, since Y_1 is a uniform $(2-\gamma, 2+\gamma)$ random variable,

$$\mathrm{E}\left[Y_1\right] = 2, \qquad \mathrm{Var}[Y_1] = \frac{4\gamma^2}{12} = \frac{1}{2}. \tag{10}$$

We now verify that Y_n is WSS. Although the PDF of Y_n depends on whether n is even or odd, we see that $\mathrm{E}[Y_n] = \mu_Y = 2$ and $\mathrm{Var}[Y_n] = \sigma_Y^2 = 1/2$. Thus

$$R_Y\left[0\right] = \mathrm{Var}[Y_n] + (\mathrm{E}\left[Y_n\right])^2 = 9/2. \tag{11}$$

For $k \neq 0$, Y_n and Y_{n+k} are independent so that

$$R_Y\left[n,k\right] = \mathrm{E}\left[Y_n Y_{n+k}\right] = \mathrm{E}\left[Y_n\right]\mathrm{E}\left[Y_{n+k}\right] = \mu_Y^2 = 4. \tag{12}$$

Combining these expressions, we obtain

$$R_Y\left[n,k\right] = R_X\left[k\right] = 4 + \frac{1}{2}\delta[k], \tag{13}$$

where $\delta[k] = 1$ if $k = 0$ and is otherwise zero. Now that we have verified that Y_n is WSS, we recall that Equation (5) showed

$$\mathrm{E}\left[W_n\right] = \mu_X\, \mathrm{E}\left[\frac{1}{Y_n}\right]. \tag{14}$$

When n is odd

$$\begin{aligned}\mathrm{E}\left[\frac{1}{Y_n}\right] = \mathrm{E}\left[\frac{1}{Y_1}\right] &= \int_{-\infty}^{\infty} \frac{1}{y} f_{Y_1}(y)\ dy \\ &= \int_{2-\gamma}^{2+\gamma} \frac{1}{2\gamma y}\, dy = \frac{1}{2\gamma}\ln\left(\frac{2+\gamma}{2-\gamma}\right) = 0.581.\end{aligned} \tag{15}$$

When n is even,

$$\mathrm{E}\left[\frac{1}{Y_n}\right] = \mathrm{E}\left[\frac{1}{Y_0}\right] = \int_{-\infty}^{\infty} \frac{1}{y} f_{Y_0}(y)\, dy = \int_0^3 \frac{2}{9}\, dy = \frac{2}{3}. \tag{16}$$

Thus $\mathrm{E}[1/Y_n]$ depends on n and so $\mathrm{E}[W_n]$ depends on n, implying W_n cannot be WSS.

Problem 13.10.1 Solution

(a) Since $X(t)$ and $Y(t)$ are independent processes,

$$\mathrm{E}[W(t)] = \mathrm{E}[X(t)Y(t)] = \mathrm{E}[X(t)]\,\mathrm{E}[Y(t)] = \mu_X \mu_Y. \tag{1}$$

In addition,

$$\begin{aligned} R_W(t,\tau) &= \mathrm{E}[W(t)W(t+\tau)] \\ &= \mathrm{E}[X(t)Y(t)X(t+\tau)Y(t+\tau)] \\ &= \mathrm{E}[X(t)X(t+\tau)]\,\mathrm{E}[Y(t)Y(t+\tau)] \\ &= R_X(\tau)R_Y(\tau). \end{aligned} \tag{2}$$

We can conclude that $W(t)$ is wide sense stationary.

(b) To examine whether $X(t)$ and $W(t)$ are jointly wide sense stationary, we calculate

$$R_{WX}(t,\tau) = \mathrm{E}[W(t)X(t+\tau)] = \mathrm{E}[X(t)Y(t)X(t+\tau)]. \tag{3}$$

By independence of $X(t)$ and $Y(t)$,

$$R_{WX}(t,\tau) = \mathrm{E}[X(t)X(t+\tau)]\,\mathrm{E}[Y(t)] = \mu_Y R_X(\tau). \tag{4}$$

Since $W(t)$ and $X(t)$ are both wide sense stationary and since $R_{WX}(t,\tau)$ depends only on the time difference τ, we can conclude from Definition 13.18 that $W(t)$ and $X(t)$ are jointly wide sense stationary.

Problem 13.10.2 Solution

To show that $X(t)$ and $X_i(t)$ are jointly wide sense stationary, we must first show that $X_i(t)$ is wide sense stationary and then we must show that the cross correlation $R_{XX_i}(t,\tau)$ is only a function of the time difference τ. For each $X_i(t)$, we have to check whether these facts are implied by the fact that $X(t)$ is wide sense stationary.

(a) Since $\mathrm{E}[X_1(t)] = \mathrm{E}[X(t+a)] = \mu_X$ and

$$\begin{aligned} R_{X_1}(t,\tau) &= \mathrm{E}\left[X_1(t)X_1(t+\tau)\right] \\ &= \mathrm{E}\left[X(t+a)X(t+\tau+a)\right] \\ &= R_X(\tau), \end{aligned} \tag{1}$$

we have verified that $X_1(t)$ is wide sense stationary. Now we calculate the cross correlation

$$\begin{aligned} R_{XX_1}(t,\tau) &= \mathrm{E}\left[X(t)X_1(t+\tau)\right] \\ &= \mathrm{E}\left[X(t)X(t+\tau+a)\right] \\ &= R_X(\tau+a). \end{aligned} \tag{2}$$

Since $R_{XX_1}(t,\tau)$ depends on the time difference τ but not on the absolute time t, we conclude that $X(t)$ and $X_1(t)$ are jointly wide sense stationary.

(b) Since $\mathrm{E}[X_2(t)] = \mathrm{E}[X(at)] = \mu_X$ and

$$\begin{aligned} R_{X_2}(t,\tau) &= \mathrm{E}\left[X_2(t)X_2(t+\tau)\right] \\ &= \mathrm{E}\left[X(at)X(a(t+\tau))\right] \\ &= \mathrm{E}\left[X(at)X(at+a\tau)\right] = R_X(a\tau), \end{aligned} \tag{3}$$

we have verified that $X_2(t)$ is wide sense stationary. Now we calculate the cross correlation

$$\begin{aligned} R_{XX_2}(t,\tau) &= \mathrm{E}\left[X(t)X_2(t+\tau)\right] \\ &= \mathrm{E}\left[X(t)X(a(t+\tau))\right] \\ &= R_X((a-1)t+\tau). \end{aligned} \tag{4}$$

Except for the trivial case when $a=1$ and $X_2(t)=X(t)$, $R_{XX_2}(t,\tau)$ depends on both the absolute time t and the time difference τ, we conclude that $X(t)$ and $X_2(t)$ are not jointly wide sense stationary.

Problem 13.10.3 Solution

(a) $Y(t)$ has autocorrelation function

$$\begin{aligned} R_Y(t,\tau) &= \mathrm{E}\left[Y(t)Y(t+\tau)\right] \\ &= \mathrm{E}\left[X(t-t_0)X(t+\tau-t_0)\right] \\ &= R_X(\tau). \end{aligned} \tag{1}$$

(b) The cross correlation of $X(t)$ and $Y(t)$ is

$$\begin{aligned} R_{XY}(t,\tau) &= \mathrm{E}\left[X(t)Y(t+\tau)\right] \\ &= \mathrm{E}\left[X(t)X(t+\tau-t_0)\right] \\ &= R_X(\tau-t_0). \end{aligned} \tag{2}$$

(c) We have already verified that $R_Y(t,\tau)$ depends only on the time difference τ. Since $\mathrm{E}[Y(t)] = \mathrm{E}[X(t-t_0)] = \mu_X$, we have verified that $Y(t)$ is wide sense stationary.

(d) Since $X(t)$ and $Y(t)$ are wide sense stationary and since we have shown that $R_{XY}(t,\tau)$ depends only on τ, we know that $X(t)$ and $Y(t)$ are jointly wide sense stationary.

Comment: This problem is badly designed since the conclusions don't depend on the specific $R_X(\tau)$ given in the problem text. (Sorry about that!)

Problem 13.11.1 Solution

For the $X(t)$ process to be stationary, we must have $f_{X(t_1)}(x) = f_{X(t_2)}(x)$. Since $X(t_1)$ and $X(t_2)$ are both Gaussian and zero mean, this requires that

$$\sigma_1^2 = \mathrm{Var}[X(t_1)] = \mathrm{Var}[X(t_2)] = \sigma_2^2. \tag{1}$$

In addition the correlation coefficient of $X(t_1)$ and $X(t_2)$ must satisfy

$$\left|\rho_{X(t_1),X(t_2)}\right| \leq 1. \tag{2}$$

This implies

$$\rho_{X(t_1),X(t_2)} = \frac{\mathrm{Cov}\left[X(t_1),X(t_2)\right]}{\sigma_1\sigma_2} = \frac{1}{\sigma_2^2} \leq 1. \tag{3}$$

Thus $\sigma_1^2 = \sigma_2^2 \geq 1$.

Problem 13.11.2 Solution

All are Gaussian. In particular, (a), (d), and (e) are Gaussian because linear processing of a Gaussian process yields a Gaussian process. Cases (b) and (e) are really the same case because they involve rescaling time. Let $Y(t) = X(\alpha t)$ where $\alpha = 1/2$ for (b) or $\alpha = 2$ for (e). In this case, consider a vector of samples

$$\mathbf{Y} = [Y(t_1), \ldots, Y(t_k)] = [X(\alpha t_1), \ldots X(\alpha t_k)]. \tag{1}$$

Since $X(t)$ is a Gaussian process, the vector $[X(\alpha t_1), \ldots X(\alpha t_k)]$ is a Gaussian random vector. Hence, $\mathbf{Y}$ is a Gaussian random vector and thus $Y(t)$ is a Gaussian process.

Problem 13.11.3 Solution

Writing $Y(t+\tau) = \int_0^{t+\tau} N(v)\,dv$ permits us to write the autocorrelation of $Y(t)$ as

$$\begin{aligned} R_Y(t,\tau) = \mathrm{E}\left[Y(t)Y(t+\tau)\right] &= \mathrm{E}\left[\int_0^t \int_0^{t+\tau} N(u)N(v)\,dv\,du\right] \\ &= \int_0^t \int_0^{t+\tau} \mathrm{E}\left[N(u)N(v)\right]\,dv\,du \\ &= \int_0^t \int_0^{t+\tau} \alpha\delta(u-v)\,dv\,du. \end{aligned} \tag{1}$$

At this point, it matters whether $\tau \geq 0$ or if $\tau < 0$. When $\tau \geq 0$, then v ranges from 0 to $t+\tau$ and at some point in the integral over v we will have $v = u$. That is, when $\tau \geq 0$,

$$R_Y(t,\tau) = \int_0^t \alpha\,du = \alpha t. \tag{2}$$

When $\tau < 0$, then we must reverse the order of integration. In this case, when the inner integral is over u, we will have $u = v$ at some point so that

$$R_Y(t,\tau) = \int_0^{t+\tau} \int_0^t \alpha\delta(u-v)\,du\,dv = \int_0^{t+\tau} \alpha\,dv = \alpha(t+\tau). \tag{3}$$

Thus we see the autocorrelation of the output is

$$R_Y(t,\tau) = \alpha \min\{t, t+\tau\}. \tag{4}$$

Perhaps surprisingly, $R_Y(t,\tau)$ is what we found in Example 13.19 to be the autocorrelation of a Brownian motion process. In fact, Brownian motion is the integral of the white noise process.

Problem 13.11.4 Solution

Let $\mu_i = \mathrm{E}[X(t_i)]$.

(a) Since $C_X(t_1, t_2 - t_1) = \rho\sigma_1\sigma_2$, the covariance matrix is

$$\mathbf{C} = \begin{bmatrix} C_X(t_1, 0) & C_X(t_1, t_2 - t_1) \\ C_X(t_2, t_1 - t_2) & C_X(t_2, 0) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{1}$$

Since $\mathbf{C}$ is a 2×2 matrix, it has determinant $|\mathbf{C}| = \sigma_1^2\sigma_2^2(1 - \rho^2)$.

(b) Is is easy to verify that

$$\mathbf{C}^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} \dfrac{1}{\sigma_1^2} & \dfrac{-\rho}{\sigma_1\sigma_2} \\ \dfrac{-\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_1^2}. \end{bmatrix} \tag{2}$$

(c) The general form of the multivariate density for $X(t_1), X(t_2)$ is

$$f_{X(t_1),X(t_2)}(x_1, x_2) = \frac{1}{(2\pi)^{k/2} |\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{X})' \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{X})}. \tag{3}$$

where $k = 2$ and $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}'$ and $\boldsymbol{\mu}_\mathbf{X} = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}'$. Hence,

$$\frac{1}{(2\pi)^{k/2} |\mathbf{C}|^{1/2}} = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1 - \rho^2)}}. \tag{4}$$

Furthermore, the exponent is

$$\begin{aligned}
&-\frac{1}{2}(\bar{x} - \bar{\mu}_X)^\top \mathbf{C}^{-1} (\bar{x} - \bar{\mu}_X) \\
&\quad = -\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix} \frac{1}{1 - \rho^2} \begin{bmatrix} \dfrac{1}{\sigma_1^2} & \dfrac{-\rho}{\sigma_1\sigma_2} \\ \dfrac{-\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_1^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\
&\quad = -\frac{\left(\dfrac{x_1 - \mu_1}{\sigma_1}\right)^2 - \dfrac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \left(\dfrac{x_2 - \mu_2}{\sigma_2}\right)^2}{2(1 - \rho^2)}.
\end{aligned} \tag{5}$$

Plugging in each piece into the joint PDF $f_{X(t_1),X(t_2)}(x_1, x_2)$ given above, we obtain the bivariate Gaussian PDF.

Problem 13.11.5 Solution

Let $\mathbf{W} = \begin{bmatrix} W(t_1) & W(t_2) & \cdots & W(t_n) \end{bmatrix}'$ denote a vector of samples of a Brownian motion process. To prove that $W(t)$ is a Gaussian random process, we must show that $\mathbf{W}$ is a Gaussian random vector. To do so, let

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}' \\ &= \begin{bmatrix} W(t_1) & W(t_2) - W(t_1) & \cdots & W(t_n) - W(t_{n-1}) \end{bmatrix}' \end{aligned} \tag{1}$$

denote the vector of increments. By the definition of Brownian motion, $X_1, \ldots, X_n$ is a sequence of independent Gaussian random variables. Thus $\mathbf{X}$ is a Gaussian random vector. Finally,

$$\mathbf{W} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_1 + X_2 \\ \vdots \\ X_1 + \cdots + X_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & & \ddots & \\ 1 & \cdots & \cdots & 1 \end{bmatrix}}_{\mathbf{A}} \mathbf{X}. \tag{2}$$

Since $\mathbf{X}$ is a Gaussian random vector and $\mathbf{W} = \mathbf{A}\mathbf{X}$ with $\mathbf{A}$ a rank n matrix, Theorem 8.11 implies that $\mathbf{W}$ is a Gaussian random vector.

Problem 13.11.6 Solution

(a) Given $N(t) = n$, $Y(t) = X_0 + \cdots + X_n$ is a sum of $n+1$ Gaussian $(0,1)$ random variables. Since

$$\mathrm{E}[Y(t)] = (n+1)\,\mathrm{E}[X] = 0, \tag{1}$$
$$\mathrm{Var}[Y(t)] = (n+1)\,\mathrm{Var}[X] = n+1, \tag{2}$$

$Y(t)$ has conditional CDF

$$\begin{aligned} F_{Y(t)|N(t)}(y|n) &= \mathrm{P}\left[Y(t) \leq y | N(t) = n\right] \\ &= \mathrm{P}\left[\frac{Y(t)}{\sqrt{n+1}} \leq \frac{y}{\sqrt{n+1}} | N(t) = n\right] \\ &= \Phi\left(\frac{y}{\sqrt{n+1}}\right). \end{aligned} \tag{3}$$

(b) A necessary condition for $Y(t)$ to be a Gaussian process is that the random variable $Y(t)$ be Gaussian for every time instant t. In particular $Y(t)$ is Gaussian if it has a CDF of the form $F_{Y(t)}(y) = \Phi((y-\mu)/\sigma)$. For the given process, we can write the CDF of $Y(t)$ as

$$\begin{aligned} F_{Y(t)}(y) = \mathrm{P}[Y(t) \leq y] &= \sum_{n=0}^{\infty} \mathrm{P}[Y(t) \leq y | N(t) = n] \, \mathrm{P}[N(t) = n] \\ &= \sum_{n=0}^{\infty} \Phi\left(\frac{y}{\sqrt{n+1}}\right) \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \end{aligned} \tag{4}$$

Unfortunately, this sum cannot be reduced to a single $\Phi(\cdot)$ function. If this is not clear, you should take a derivative of the CDF and see that you do not obtain a Gaussian PDF. Thus $Y(t)$ is not a Gaussian random variable and thus the process is not Gaussian.

(c) The process is **not** stationary because $F_{Y(t)}(y)$ depends on the time t.

(d) First we find the expected value and autocorrelation, and then we will know whether the process is wide-sense stationary. For the expected value, we recall from part (a) that conditioned on $N(t) = n$, $Y(t)$ was a zero mean Gaussian. That is, $\mathrm{E}[Y(t)N(t) = n] = 0$. This implies

$$\mathrm{E}[Y(t)] = \sum_{n=0}^{\infty} \mathrm{E}[Y(t) | N(t) = n] \, \mathrm{P}[N(t) = n] = 0. \tag{5}$$

To find the autocovariance, we use the same trick as for finding the autocorrelation of the Poisson process and the Brownian motion process. To start, we assume $\tau > 0$. Since $\mathrm{E}[Y(t)] = 0$,

$$\begin{aligned} C_Y(t,\tau) &= \mathrm{E}[Y(t)Y(t+\tau)] \\ &= \mathrm{E}[Y(t)((Y(t+\tau) - Y(t)) + Y(t))] \\ &= \mathrm{E}[Y(t)(Y(t+\tau) - Y(t))] + \mathrm{E}\left[Y^2(t)\right]. \end{aligned} \tag{6}$$

Note that $Y(t+\tau) - Y(t)$ depends on X_n corresponding to arrivals of the Poisson process in the interval $(t, t+\tau]$, which is independent of the arrivals prior to time t. In addition the X_n corresponding to arrivals in the interval

$(t, t+\tau]$ are independent of the X_n corresponding to arrivals prior to time t. Thus $Y(t+\tau) - Y(t)$ is independent of $Y(t)$. It follows from (6) that

$$\begin{aligned} C_Y(t,\tau) &= \mathrm{E}\left[Y(t)\right]\mathrm{E}\left[Y(t+\tau) - Y(t)\right] + \mathrm{E}\left[Y^2(t)\right] \\ &= \mathrm{E}\left[Y^2(t)\right] = \mathrm{Var}[Y(t)]. \end{aligned} \tag{7}$$

To calculate $\mathrm{Var}[Y(t)]$, we observe that it will be convenient to redefine $Y(t)$ as

$$Y(t) = X_1 + X_2 + \cdots + X_N, \tag{8}$$

where $N = N(t) + 1$. This makes no difference since $Y(t)$ is still the sum of $N = N(t) + 1$ iid Gaussian $(0, 1)$ random variables. Written this way, we see that $Y(t)$ is a random sum of random variables such that N is independent of $X_1, X_2, \ldots$ The variance of a random sum of random variables is given by

$$\mathrm{Var}[Y(t)] = \mathrm{E}\left[N\right]\mathrm{Var}[X] + \mathrm{Var}[N](\mathrm{E}\left[X\right])^2. \tag{9}$$

Since $\mathrm{E}[X] = 0$ and $\mathrm{Var}[X] = 1$, we have

$$\mathrm{Var}[Y(t)] = \mathrm{E}\left[N\right] = \mathrm{E}\left[N(t) + 1\right] = \lambda t + 1. \tag{10}$$

This same result can also be obtained by careful use of the iterative expectation with conditioning on $N(t)$ and $N(t+\tau)$. Thus for $\tau \geq 0$, $C_Y(t,\tau) = 1 + \lambda t$. For $\tau < 0$, $t + \tau < t$. In the above argument, we reverse all labels of t and $t + \tau$ and we can conclude that $C_Y(t,\tau) = 1 + \lambda(t + \tau)$. If you don't trust this argument, here are the details:

$$\begin{aligned} C_Y(t,\tau) &= \mathrm{E}\left[Y(t)Y(t+\tau)\right] \\ &= \mathrm{E}\left[(Y(t) - Y(t+\tau) + Y(t+\tau))Y(t+\tau)\right] \\ &= \mathrm{E}\left[(Y(t) - Y(t+\tau))Y(t+\tau)\right] + \mathrm{E}\left[Y^2(t+\tau)\right] \\ &= \mathrm{E}\left[Y^2(t+\tau)\right] = \mathrm{Var}[Y(t+\tau)]. \end{aligned} \tag{11}$$

Note that $\mathrm{Var}[Y(t+\tau)]$ is the same as $\mathrm{Var}[Y(t)]$ but with t replace by $t + \tau$. Thus, for $\tau < 0$, $C_Y(t,\tau) = 1 + \lambda(t + \tau)$. A general expression for the autocovariance is

$$C_Y(t,\tau) = 1 + \lambda \min(t, t+\tau). \tag{12}$$

Since $C_Y(t,\tau)$ depends on t, $Y(t)$ is not wide-sense stationary.

Problem 13.12.1 Solution

From the instructions given in the problem, the program `noisycosine.m` will generate the four plots.

```
n=1000; t=0.001*(-n:n);
w=gaussrv(0,0.01,(2*n)+1);
%Continuous Time, Continuous Value
xcc=2*cos(2*pi*t) + w';
plot(t,xcc);
xlabel('\it t');ylabel('\it X_{cc}(t)');
axis([-1 1 -3 3]);
figure; %Continuous Time, Discrete Value
xcd=round(xcc); plot(t,xcd);
xlabel('\it t');ylabel('\it X_{cd}(t)');
axis([-1 1 -3 3]);
figure; %Discrete time, Continuous Value
ts=subsample(t,100); xdc=subsample(xcc,100);
plot(ts,xdc,'b.');
xlabel('\it t');ylabel('\it X_{dc}(t)');
axis([-1 1 -3 3]);
figure; %Discrete Time, Discrete Value
xdd=subsample(xcd,100); plot(ts,xdd,'b.');
xlabel('\it t');ylabel('\it X_{dd}(t)');
axis([-1 1 -3 3]);
```

In `noisycosine.m`, we use a function `subsample.m` to obtain the discrete time sample functions. In fact, `subsample` is hardly necessary since it's such a simple one-line MATLAB function:

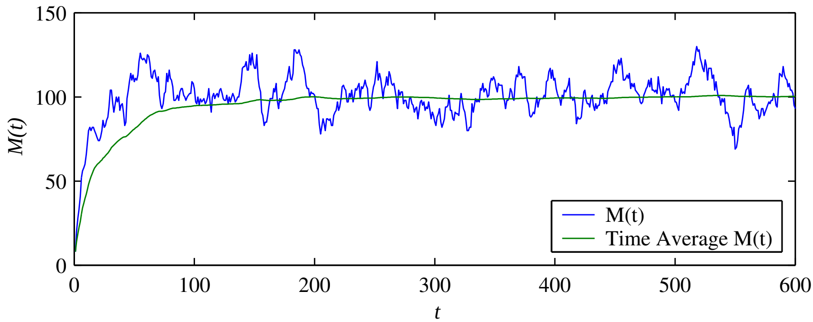
```
function y=subsample(x,n)
%input x(1), x(2) ...
%output y(1)=x(1), y(2)=x(1+n), y(3)=x(2n+1)
y=x(1:n:length(x));
```

However, we use it just to make `noisycosine.m` a little more clear.

Problem 13.12.2 Solution

```
>> t=(1:600)';
>> M=simswitch(10,0.1,t);
>> Mavg=cumsum(M)./t;
>> plot(t,M,t,Mavg);
```

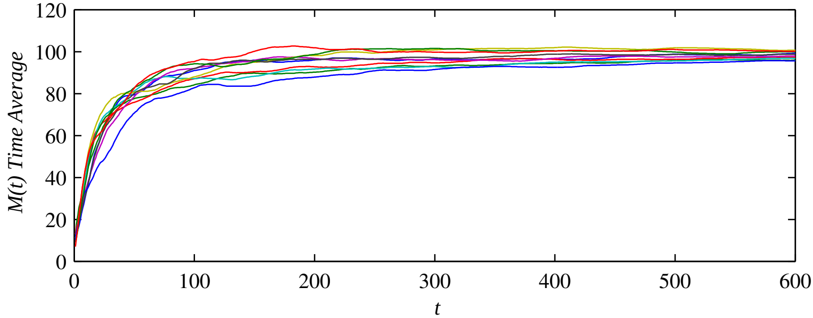
These commands will simulate the switch for 600 minutes, producing the vector `M` of samples of $M(t)$ each minute, the vector `Mavg` which is the sequence of time average estimates, and a plot resembling this one:



From the figure, it appears that the time average is converging to a value in th neighborhood of 100. In particular, because the switch is initially empty with $M(0) = 0$, it takes a few hundred minutes for the time average to climb to something close to 100.Following the problem instructions, we can write the following short program to examine ten simulation runs:

```
function Mavg=simswitchavg(T,k)
%Usage: Mavg=simswitchavg(T,k)
%simulate k runs of duration T of the
%telephone switch in Chapter 10
%and plot the time average of each run
t=(1:k)';
%each column of Mavg is a time average sample run
Mavg=zeros(T,k);
for n=1:k,
    M=simswitch(10,0.1,t);
    Mavg(:,n)=cumsum(M)./t;
end
plot(t,Mavg);
```

The command `simswitchavg(600,10)` produced this graph:



From the graph, one can see that even after $T = 600$ minutes, each sample run produces a time average $\overline{M}_{600}$ around 100. Note that Markov chain analysis can be used to prove that the expected number of calls in the switch is in fact 100. However, note that even if T is large, $\overline{M}_T$ is still a random variable. From the above plot, one might guess that $\overline{M}_{600}$ has a standard deviation of perhaps $\sigma = 2$ or $\sigma = 3$. An exact calculation of the variance of $\overline{M}_{600}$ is fairly difficult because it is a sum of dependent random variables, each of which has a PDF that is in itself reasonably difficult to calculate.

Problem 13.12.3 Solution

In this problem, our goal is to find out the average number of ongoing calls in the switch. Before we use the approach of Problem 13.12.2, its worth a moment to consider the physical situation. In particular, calls arrive as a Poisson process of rate $\lambda = 100$ call/minute and each call has duration of *exactly* one minute. As a result, if we inspect the system at an arbitrary time t at least one minute past initialization, the number of calls at the switch will be exactly the number of calls N_1 that arrived in the previous minute. Since calls arrive as a Poisson proces of rate $\lambda = 100$ calls/minute. N_1 is a Poisson random variable with $\mathrm{E}[N_1] = 100$.

In fact, this should be true for every inspection time t. Hence it should surprising if we compute the time average and find the time average number in the queue to be something other than 100. To check out this quickie analysis, we use the method of Problem 13.12.2. However, unlike Problem 13.12.2, we cannot directly use the function `simswitch.m` because the call duration are no longer exponential random variables. Instead, we must modify `simswitch.m` for the deterministic one minute call durations, yielding the function `simswitchd.m`:

```
function M=simswitchd(lambda,T,t)
%Poisson arrivals, rate lambda
%Deterministic (T) call duration
%For vector t of times
%M(i) = no. of calls at time t(i)
s=poissonarrivals(lambda,max(t));
y=s+T;
A=countup(s,t);
D=countup(y,t);
M=A-D;
```

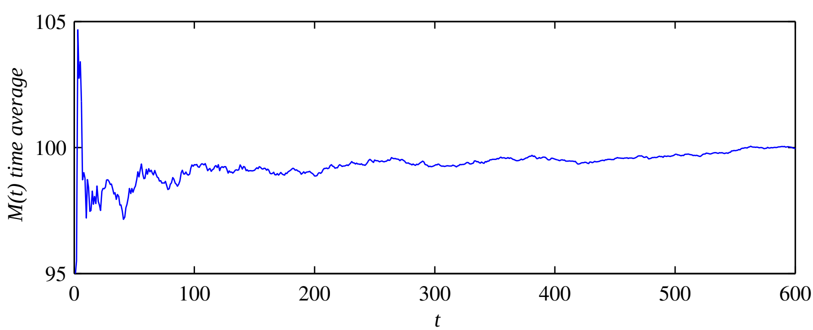
Note that if you compare `simswitch.m` in the text with `simswitchd.m` here, two changes occurred. The first is that the exponential call durations are replaced by the deterministic time T. The other change is that `count(s,t)` is replaced by `countup(s,t)`. In fact, `n=countup(x,y)` does exactly the same thing as `n=count(x,y)`; in both cases, `n(i)` is the number of elements less than or equal to $y(i)$. The difference is that `countup` requires that the vectors `x` and `y` be nondecreasing.

Now we use the same procedure as in Problem 13.12.2 and form the time average

$$\overline{M}(T) = \frac{1}{T}\sum_{t=1}^{T} M(t). \qquad (1)$$

```
>> t=(1:600)';
>> M=simswitchd(100,1,t);
>> Mavg=cumsum(M)./t;
>> plot(t,Mavg);
```

We form and plot the time average using these commands will yield a plot vaguely similar to that shown below.



We used the word "vaguely" because at $t = 1$, the time average is simply the number of arrivals in the first minute, which is a Poisson ($\alpha = 100$) random variable which has not been averaged. Thus, the left side of the graph will be random for each run. As expected, the time average appears to be converging to 100.

Problem 13.12.4 Solution

The random variable S_n is the sum of n exponential (λ) random variables. That is, S_n is an Erlang (n, λ) random variable. Since $K = 1$ if and only if $S_n > T$, $\mathrm{P}[K = 1] = \mathrm{P}[S_n > T]$. Typically, $\mathrm{P}[K = 1]$ is fairly high because

$$\mathrm{E}\left[S_n\right] = \frac{n}{\lambda} = \frac{\lceil 1.1\lambda T \rceil}{\lambda} \approx 1.1T. \tag{1}$$

Increasing n increases $\mathrm{P}[K = 1]$; however, `poissonarrivals` then does more work generating exponential random variables. Although we don't want to generate more exponential random variables than necessary, if we need to generate a lot of arrivals (ie a lot of exponential interarrival times), then MATLAB is typically faster generating a vector of them all at once rather than generating them one at a time. Choosing $n = \lceil 1.1\lambda T \rceil$ generates about 10 percent more exponential random variables than we typically need. However, as long as $\mathrm{P}[K = 1]$ is high, a ten percent penalty won't be too costly.

When n is small, it doesn't much matter if we are efficient because the amount of calculation is small. The question that must be addressed is to estimate $\mathrm{P}[K = 1]$ when n is large. In this case, we can use the central limit theorem because S_n is the sum of n exponential random variables. Since $\mathrm{E}[S_n] = n/\lambda$ and $\mathrm{Var}[S_n] = n/\lambda^2$,

$$\begin{aligned} \mathrm{P}\left[S_n > T\right] &= \mathrm{P}\left[\frac{S_n - n/\lambda}{\sqrt{n/\lambda^2}} > \frac{T - n/\lambda}{\sqrt{n/\lambda^2}}\right] \\ &\approx Q\left(\frac{\lambda T - n}{\sqrt{n}}\right). \end{aligned} \tag{2}$$

To simplify our algebra, we assume for large n that $0.1\lambda T$ is an integer. In this case, $n = 1.1\lambda T$ and

$$\mathrm{P}\left[S_n > T\right] \approx Q\left(-\frac{0.1\lambda T}{\sqrt{1.1\lambda T}}\right) = \Phi\left(\sqrt{\frac{\lambda T}{110}}\right). \tag{3}$$

Thus for large λT, $\mathrm{P}[K = 1]$ is very small. For example, if $\lambda T = 1{,}000$, $\mathrm{P}[S_n > T] \approx \Phi(3.01) = 0.9987$. If $\lambda T = 10{,}000$, $\mathrm{P}[S_n > T] \approx \Phi(9.5)$.

Problem 13.12.5 Solution

Following the problem instructions, we can write the function `newarrivals.m`. For convenience, here are `newarrivals` and `poissonarrivals` side by side.

```
function s=newarrivals(lam,T)
%Usage s=newarrivals(lam,T)
%Returns Poisson arrival times
%s=[s(1) ... s(n)] over [0,T]
n=poissonrv(lam*T,1);
s=sort(T*rand(n,1));
```

```
function s=poissonarrivals(lam,T)
%arrival times s=[s(1) ... s(n)]
%  s(n)<= T < s(n+1)
n=ceil(1.1*lam*T);
s=cumsum(exponentialrv(lam,n));
while (s(length(s))< T),
  s_new=s(length(s))+ ...
    cumsum(exponentialrv(lam,n));
  s=[s; s_new];
end
s=s(s<=T);
```

Clearly the code for `newarrivals` is shorter, more readable, and perhaps, with the help of Problem 13.5.8, more logical than `poissonarrivals`. Unfortunately this doesn't mean the code runs better. Here are some `cputime` comparisons:

```
>> t=cputime;s=poissonarrivals(1,100000);t=cputime-t
t =
    0.1110
>> t=cputime;s=newarrivals(1,100000);t=cputime-t
t =
    0.5310
>> t=cputime;poissonrv(100000,1);t=cputime-t
t =
    0.5200
>>
```

Unfortunately, these results were highly repeatable; `poissonarrivals` generated 100,000 arrivals of a rate 1 Poisson process required roughly 0.1 seconds of cpu time. The same task took `newarrivals` about 0.5 seconds, or roughly 5 times as long! In the `newarrivals` code, the culprit is the way `poissonrv` generates a single Poisson random variable with expected value 100,000. In this case, `poissonrv` generates the first 200,000 terms of the Poisson PMF! This required calculation is so large that it dominates the work need to generate 100,000 uniform random numbers. In fact, this suggests that a more efficient way to generate a Poisson (α) random variable N is to generate arrivals of a rate α Poisson process until the Nth arrival is after time 1.

Problem 13.12.6 Solution

We start with `brownian.m` to simulate the Brownian motion process with barriers, Since the goal is to estimate the barrier probability $\mathrm{P}[|X(t)| = b]$, we don't keep track of the value of the process over all time. Also, we simply assume that a unit time step $\tau = 1$ for the process. Thus, the process starts at $n = 0$ at position $W_0 = 0$ at each step n, the position, if we haven't reached a barrier, is $W_n = W_{n-1} + X_n$, where $X_1, \ldots, X_T$ are iid Gaussian $(0, \sqrt{\alpha})$ random variables. Accounting for the effect of barriers,

$$W_n = \max(\min(W_{n-1} + X_n, b), -b). \tag{1}$$

To implement the simulation, we can generate the vector `x` of increments all at once. However to check at each time step whether we are crossing a barrier, we need to proceed sequentially. (This is analogous to the problem in Quiz 13.12.)

In `brownbarrier` shown below, `pb(1)` tracks how often the process touches the left barrier at $-b$ while `pb(2)` tracks how often the right side barrier at b is reached. By symmetry, $\mathrm{P}[X(t) = b] = \mathrm{P}[X(t) = -b]$. Thus if T is chosen very large, we should expect `pb(1)=pb(2)`. The extent to which this is not the case gives an indication of the extent to which we are merely estimating the barrier probability. Here is the code and for each $T \in \{10{,}000, 100{,}000, 1{,}000{,}000\}$, here two sample runs:

```
function pb=brownwall(alpha,b,T)
%pb=brownwall(alpha,b,T)
%Brownian motion, param. alpha
%walls at [-b, b], sampled
%unit of time until time T
%each Returns vector pb:
%pb(1)=fraction of time at -b
%pb(2)=fraction of time at b
T=ceil(T);
x=sqrt(alpha).*gaussrv(0,1,T);
w=0;pb=zeros(1,2);
for k=1:T,
    w=w+x(k);
    if (w <= -b)
        w=-b;
        pb(1)=pb(1)+1;
    elseif (w >= b)
        w=b;
        pb(2)=pb(2)+1;
    end
end
pb=pb/T;
```

```
>> pb=brownwall(0.01,1,1e4)
pb =
    0.0301    0.0353
>> pb=brownwall(0.01,1,1e4)
pb =
    0.0417    0.0299
>> pb=brownwall(0.01,1,1e5)
pb =
    0.0333    0.0360
>> pb=brownwall(0.01,1,1e5)
pb =
    0.0341    0.0305
>> pb=brownwall(0.01,1,1e6)
pb =
    0.0323    0.0342
>> pb=brownwall(0.01,1,1e6)
pb =
    0.0333    0.0324
>>
```

The sample runs show that for $\alpha = 0.1$ and $b = 1$ that the

$$\mathrm{P}\left[X(t) = -b\right] \approx \mathrm{P}\left[X(t) = b\right] \approx 0.03. \tag{2}$$

Otherwise, the numerical simulations are not particularly instructive. Perhaps the most important thing to understand is that the Brownian motion process with barriers is *very* different from the ordinary Brownian motion process. Remember that for ordinary Brownian motion, the variance of $X(t)$ always increases linearly with t. For the process with barriers, $X^2(t) \leq b^2$ and thus $\mathrm{Var}[X(t)] \leq b^2$. In fact, for the process with barriers, the PDF of $X(t)$ converges to a limit as t becomes large. If you're curious, you shouldn't have much trouble digging in the library to find out more.

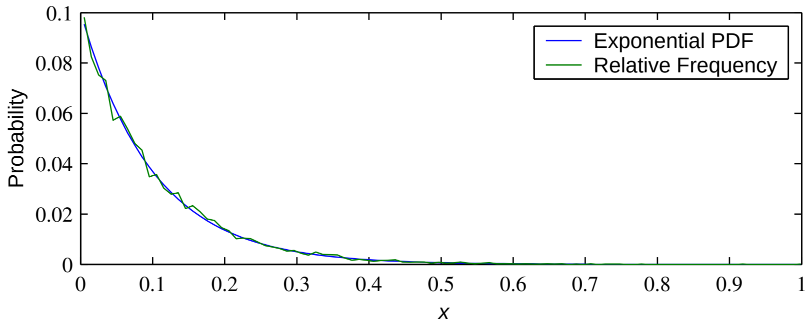
Problem 13.12.7 Solution

In this problem, we start with the `simswitch.m` code to generate the vector of departure times `y`. We then construct the vector `I` of inter-departure times. The command `hist,20` will generate a 20 bin histogram of the departure times. The fact that this histogram resembles an exponential PDF suggests that perhaps it is reasonable to try to match the PDF of an exponential (μ) random variable against the histogram.

In most problems in which one wants to fit a PDF to measured data, a key issue is how to choose the parameters of the PDF. In this problem, choosing μ is simple. Recall that the switch has a Poisson arrival process of rate λ so interarrival times are exponential (λ) random variables. If $1/\mu < 1/\lambda$, then the average time between departures from the switch is less than the average time between arrivals to the switch. In this case, calls depart the switch faster than they arrive which is impossible because each departing call was an arriving call at an earlier time. Similarly, if $1/\mu > 1/\lambda$, then calls would be departing from the switch more slowly than they arrived. This can happen to an overloaded switch; however, it's impossible in this system because each arrival departs after an exponential time. Thus the only possibility is that $1/\mu = 1/\lambda$. In the program `simswitchdepart.m`, we plot a histogram of departure times for a switch with arrival rate λ against the scaled exponential (λ) PDF $\lambda e^{-\lambda x} b$ where b is the histogram bin size. Here is the code:

```
function I=simswitchdepart(lambda,mu,T)
%Usage: I=simswitchdepart(lambda,mu,T)
%Poisson arrivals, rate lambda
%Exponential (mu) call duration
%Over time [0,T], returns I,
%the vector of inter-departure times
%M(i) = no. of calls at time t(i)
s=poissonarrivals(lambda,T);
y=s+exponentialrv(mu,length(s));
y=sort(y);
n=length(y);
I=y-[0; y(1:n-1)]; %interdeparture times
imax=max(I);b=ceil(n/100);
id=imax/b; x=id/2:id:imax;
pd=hist(I,x); pd=pd/sum(pd);
px=exponentialpdf(lambda,x)*id;
plot(x,px,x,pd);
xlabel('\it x');ylabel('Probability');
legend('Exponential PDF','Relative Frequency');
```

Here is an example of the `simswitchdepart(10,1,1000)` output:



As seen in the figure, the match is quite good. Although this is not a carefully designed statistical test of whether the inter-departure times are exponential random variables, it is enough evidence that one may want to pursue whether such a result can be proven.

In fact, the switch in this problem is an example of an $M/M/\infty$ queuing system for which it has been shown that not only do the inter-departure have an exponential distribution, but the steady-state departure process is a Poisson process. For the curious reader, details can be found, for example, in the text *Stochastic Processes: Theory for Applications* by Gallager.