Appendix A

Solutions to selected exercises

Exercise 1.2: This exercise derives the probability of an arbitrary (non-disjoint) union of events, derives the union bound, and derives some useful limit expressions.

a) For 2 arbitrary events \( A_1 \) and \( A_2 \), show that

\[
A_1 \cup A_2 = A_1 \cup (A_2 - A_1),
\]

where \( A_2 - A_1 = A_2A_1^c \). Show that \( A_1 \) and \( A_2 - A_1 \) are disjoint. Hint: This is what Venn diagrams were invented for.

Solution: Note that each sample point \( \omega \) is in \( A_1 \) or \( A_1^c \), but not both. Thus each \( \omega \) is in exactly one of \( A_1 \), \( A_1^cA_2 \) or \( A_1^cA_2^c \). In the first two cases, \( \omega \) is in both sides of (A.1) and in the last case it is in neither. Thus the two sides of (A.1) are identical. Also, as pointed out above, \( A_1 \) and \( A_2 - A_1 \) are disjoint. These results are intuitively obvious from the Venn diagram,

\[
A_1A_2^c \ A_1A_2 \ A_2A_1^c = A_2 - A_1
\]

b) For any \( n \geq 2 \) and arbitrary events \( A_1, \ldots, A_n \), define \( B_n = A_n - \bigcup_{i=1}^{n-1} A_i \). Show that \( B_1, B_2, \ldots \) are disjoint events and show that for each \( n \geq 2 \), \( \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i \). Hint: Use induction.

Solution: Let \( B_1 = A_1 \). Part a) then showed that \( B_1 \) and \( B_2 \) are disjoint and (from (A.1)) that \( A_1 \cup A_2 = B_1 \cup B_2 \). Let \( C_n = \bigcup_{i=1}^{n} A_i \). We use induction to prove that \( C_n = \bigcup_{i=1}^{n} B_i \) and that the \( B_n \) are disjoint. We have seen that \( C_2 = B_1 \cup B_2 \), which forms the basis for the induction. We assume that \( C_{n-1} = \bigcup_{i=1}^{n-1} B_i \) and prove that \( C_n = \bigcup_{i=1}^{n} B_i \).

\[
C_n = C_{n-1} \cup A_n = C_{n-1} \cup A_n C_{n-1}^c = C_{n-1} \cup B_n = \bigcup_{i=1}^{n} B_i.
\]

In the second equality, we used (A.1), letting \( C_{n-1} \) play the role of \( A_1 \) and \( A_n \) play the role of \( A_2 \). From this same application of (A.1), we also see that \( C_{n-1} \) and \( B_n = A_n - C_{n-1} \) are disjoint. Since \( C_{n-1} = \bigcup_{i=1}^{n-1} B_i \), this also shows that \( B_n \) is disjoint from \( B_1, \ldots, B_{n-1} \).
c) Show that
\[
\Pr\left\{ \bigcup_{n=1}^{\infty} A_n \right\} = \Pr\left\{ \bigcup_{n=1}^{\infty} B_n \right\} = \sum_{n=1}^{\infty} \Pr\{B_n\}.
\]

**Solution:** If \( \omega \in \bigcup_{n=1}^{\infty} A_n \), then it is in \( A_n \) for some \( n \geq 1 \). Thus \( \omega \in \bigcup_{i=1}^{n} B_i \), and thus \( \omega \in \bigcup_{n=1}^{\infty} B_n \). The same argument works the other way, so \( \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \). This establishes the first equality above, and the second is valid by the axioms of probability.

d) Show that for each \( n \), \( \Pr\{B_n\} \leq \Pr\{A_n\} \). Use this to show that
\[
\Pr\left\{ \bigcup_{n=1}^{\infty} A_n \right\} \leq \sum_{n=1}^{\infty} \Pr\{A_n\}.
\]

**Solution:** Since \( B_n = A_n - \bigcup_{i=1}^{n-1} A_i \), we see that \( \omega \in B_n \) implies that \( \omega \in A_n \). From (1.5), this implies that \( \Pr\{B_n\} \leq \Pr\{A_n\} \) for each \( n \). Thus
\[
\Pr\left\{ \bigcup_{n=1}^{\infty} A_n \right\} = \sum_{n=1}^{\infty} \Pr\{B_n\} \leq \sum_{n=1}^{\infty} \Pr\{A_n\}.
\]

e) Show that \( \Pr\left\{ \bigcup_{n=1}^{\infty} A_n \right\} = \lim_{n \to \infty} \Pr\{\bigcup_{i=1}^{n} A_i\} \). Hint: Use part c). Note that this says that the probability of a limit is equal to the limit of the probabilities. This might well appear to be obvious without a proof, but you will see situations later where similar appearing interchanges cannot be made.

**Solution:** From part c),
\[
\Pr\left\{ \bigcup_{n=1}^{\infty} A_n \right\} = \sum_{n=1}^{\infty} \Pr\{B_n\} = \lim_{k \to \infty} \sum_{n=1}^{k} \Pr\{B_n\}.
\]

From part b), however,
\[
\sum_{n=1}^{k} \Pr\{B_n\} = \Pr\left\{ \bigcup_{n=1}^{k} B_n \right\} = \Pr\left\{ \bigcup_{n=1}^{k} A_n \right\}.
\]

f) Show that \( \Pr\{\bigcap_{n=1}^{\infty} A_n\} = \lim_{n \to \infty} \Pr\{\bigcap_{i=1}^{n} A_i\} \). Hint: Remember De Morgan’s equalities.

**Solution:** Using De Morgan’s equalities,
\[
\Pr\left\{ \bigcap_{n=1}^{\infty} A_n \right\} = 1 - \Pr\left\{ \bigcup_{n=1}^{\infty} A_n^c \right\} = 1 - \lim_{k \to \infty} \Pr\left\{ \bigcup_{n=1}^{k} A_n^c \right\} = \lim_{k \to \infty} \Pr\left\{ \bigcap_{n=1}^{k} A_n \right\}.
\]

**Exercise 1.4:** Consider a sample space of 8 equiprobable sample points and let \( A_1, A_2, A_3 \) be three events each of probability 1/2 such that \( \Pr\{A_1 \cap A_2 \cap A_3\} = \Pr\{A_1\} \Pr\{A_2\} \Pr\{A_3\} \).

a) Create an example where \( \Pr\{A_1 \cap A_2\} = \Pr\{A_1 \cap A_3\} = \frac{1}{4} \) but \( \Pr\{A_2 \cap A_3\} = \frac{1}{8} \). Hint: Make a table with a row for each sample point and a column for each of the above 3 events and try different ways of assigning sample points to events (the answer is not unique).
**Solution:** Note that exactly one sample point must be in $A_1, A_2,$ and $A_3$ in order to make $\Pr\{A_1A_2A_3\} = 1/8$. In order to make $\Pr\{A_1A_2\} = 1/4$, there must be an additional sample point that contains $A_1$ and $A_2$ but not $A_3$. Similarly, there must be a sample point that contains $A_1$ and $A_3$ but not $A_2$. These give rise to the first three rows in the table below. The other sample points can each be in at most 1 of the events $A_1, A_2,$ and $A_3$, and in order to make each event have probability 1/2, each of these sample points must be in exactly one of the three events. This uniquely specifies the table below except for which sample point lies in each event.

<table>
<thead>
<tr>
<th>Sample points</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

b) Show that, for your example, $A_2$ and $A_3$ are not independent. Note that the definition of statistical independence would be very strange if it allowed $A_1, A_2, A_3$ to be independent while $A_2$ and $A_3$ are dependent. This illustrates why the definition of independence requires (1.14) rather than just (1.15).

**Solution:** Note that $\Pr\{A_2A_3\} = 1/8 \neq \Pr\{A_2\} \Pr\{A_3\}$, so $A_2$ and $A_3$ are dependent. We also note that $\Pr\{A_1A_2A_3\} = 0$, which also violates our intuition about statistical independence. Although the definition in (1.14) of statistical independence of more than 2 events looks strange, it is clear from this example that (1.15) is insufficient in the sense that it only specifies part of the above table.

**Exercise 1.5:** This exercise shows that for all rv’s $X$, $F_X(x)$ is continuous from the right.

a) For any given rv $X$, any real number $x$, and each integer $n \geq 1$, let $A_n = \{\omega : X > x + 1/n\}$, and show that $A_1 \subseteq A_2 \subseteq \cdots$. Use this and the corollaries to the axioms of probability to show that $\Pr\{\bigcup_{n \geq 1} A_n\} = \lim_{n \to \infty} \Pr\{A_n\}$.

**Solution:** If $X(\omega) > x + 1/n$, then (since $1/n > 1/(n+1)$), we also have $X(\omega) > x + 3/(n+1)$. Thus $A_n \subseteq A_{n+1}$ for all $n \geq 1$. Thus from (1.9), $\Pr\{\bigcup_{n \geq 1} A_n\} = \lim_{n \to \infty} \Pr\{A_n\}$.

b) Show that $\Pr\{\bigcup_{n \geq 1} A_n\} = \Pr\{X > x\}$ and that $\Pr\{X > x\} = \lim_{n \to \infty} \Pr\{X > x + 1/n\}$.

d) Show that for $\epsilon > 0$, $\lim_{n \to \infty} \Pr\{X \leq x + \epsilon\} = \Pr\{X \leq x\}$. 

**Solution:** If $X(\omega) > x$, then there must be an $n$ sufficiently large that $X(\omega) > x + 1/n$. Thus $\{\omega : X > x\} \subseteq \bigcup_{n \geq 1} A_n$. The subset inequality goes the other way also since $X(\omega) > x + 1/n$ for any $n \geq 1$ implies that $X(\omega) > x$. Since these represent the same events, they have the same probability and $\Pr\{\bigcup_{n \geq 1} A_n\} = \Pr\{X > x\}$. Then from part a) we also have

$$\Pr\{X > x\} = \lim_{n \to \infty} \Pr\{A_n\} = \lim_{n \to \infty} \Pr\{X > x + 1/n\}.$$ 

e) Show that for $\epsilon > 0$, $\lim_{n \to \infty} \Pr\{X \leq x + \epsilon\} = \Pr\{X \leq x\}$. 


Solution: Taking the complement of both sides of the above equation, \( \Pr\{X \leq x\} = \lim_{n \to \infty} \Pr\{X \leq x + 1/n\} \). Since \( \Pr\{X \leq x + \epsilon\} \) is non-decreasing in \( \epsilon \), it also follows that for \( \epsilon > 0 \), \( \Pr\{X \leq x\} = \lim_{\epsilon \to 0} \Pr\{X \leq x + \epsilon\} \).

Exercise 1.6: Show that for a continuous nonnegative rv \( X \),

\[
\int_0^\infty \Pr\{X > x\} \, dx = \int_0^\infty x f_X(x) \, dx. \tag{A.2}
\]

Hint 1: First rewrite \( \Pr\{X > x\} \) on the left side of (1.98) as \( \int_x^\infty f_X(y) \, dy \). Then think through, to your level of comfort, how and why the order of integration can be interchanged in the resulting expression.

Solution: We have \( \Pr\{X > x\} = \int_x^\infty f_X(y) \, dy \) from the definition of a continuous rv. We look at \( \mathbb{E}[X] = \int_0^\infty \Pr\{X > x\} \, dx \) as \( \lim_{A \to \infty} \int_0^A F_X^c(x) \, dx \) since the limiting operation \( A \to \infty \) is where the interesting issue is.

\[
\int_0^A F_X^c(x) \, dx = \int_0^A \int_x^\infty f_X(y) \, dy \, dx \\
= \int_0^A \int_x^A f_X(y) \, dy \, dx + \int_0^A \int_A^\infty f_X(y) \, dy \, dx \\
= \int_0^A \int_0^y f_X(y) \, dx \, dy + A F_X^c(A).
\]

We first broke the second integral into two parts and then interchanged the order of integration in the first part. The inner integral of the first portion is \( y f_X(y) \), so we have

\[
\lim_{A \to \infty} \int_0^A F_X^c(x) \, dx = \lim_{A \to \infty} \int_0^A y f_X(y) \, dy + \lim_{A \to \infty} A F_X^c(A).
\]

Assuming that \( \mathbb{E}[X] \) exists, the first integral above is nondecreasing in \( A \) and has a limit. The second integral is also nondecreasing and upper bounded by the first integral, so it also has a limit. This means that \( \lim_{A \to \infty} A F_X^c(A) \) must also have a limit. This limit must be 0, since otherwise \( F_X(A) \) would be asymptotically equal to \( \alpha/A \) for some \( \alpha \neq 0 \), contradicting the assumption that \( \mathbb{E}[X] \) is finite.

This establishes (A.2) for the case where \( \mathbb{E}[X] \) is finite. The case where \( \mathbb{E}[X] \) is infinite is a minor perturbation.

The result that \( \lim_{A \to \infty} A F_X^c(A) = 0 \) is also important and can be seen intuitively from Figure 1.3.

Hint 2: As an alternate approach, derive (1.98) using integration by parts.

Solution: Using integration by parts and being less careful,

\[
\int_0^\infty d\left(x F_X^c(x)\right) = -\int_0^\infty x f_X(x) \, dx + \int_0^\infty F_X^c(x) \, dx.
\]

The left side is \( \lim_{A \to \infty} A F_X(A) - 0 F_X(0) \) so this shows the same thing, again requiring the fact that \( \lim_{A \to \infty} A F_X(A) = 0 \) when \( \mathbb{E}[X] \) exists.
**Exercise 1.8:** A variation of Example 1.5.1 is to let $M$ be a random variable that takes on both positive and negative values with the PMF
\[
p_M(m) = \frac{1}{2|m|(|m| + 1)}.
\]
In other words, $M$ is symmetric around 0 and $|M|$ has the same PMF as the nonnegative rv $N$ of Example 1.5.1.

a) Show that $\sum_{m \geq 0} m p_M(m) = \infty$ and $\sum_{m < 0} m p_M(m) = -\infty$. (Thus show that the expectation of $M$ not only does not exist but is undefined even in the extended real number system.)

**Solution:**
\[
\sum_{m \geq 0} m p_M(m) = \sum_{m \geq 0} \frac{1}{2(|m| + 1)} = \infty
\]
\[
\sum_{m < 0} m p_M(m) = \sum_{m < 0} \frac{-1}{2(|m| + 1)} = -\infty.
\]

b) Suppose that the terms in $\sum_{m = -\infty}^{\infty} m p_M(m)$ are summed in the order of 2 positive terms for each negative term (i.e., in the order $1, 2, -1, 3, 4, -2, 5, \cdots$). Find the limiting value of the partial sums in this series. Hint: You may find it helpful to know that
\[
\lim_{n \to \infty} \left[ \sum_{i=1}^{n} \frac{1}{i} - \int_{1}^{n} \frac{1}{x} \, dx \right] = \gamma,
\]
where $\gamma$ is the Euler-Mascheroni constant, $\gamma = 0.57721 \cdots$.

**Solution:** The sum after $3n$ terms is
\[
\sum_{m=1}^{2n} \frac{1}{2(m+1)} - \sum_{m=1}^{n} \frac{1}{2(m+1)} = \frac{1}{2} \sum_{i=n+2}^{2n+1} \frac{1}{i}.
\]
Taking the limit as $n \to \infty$, the Euler-Mascheroni constant cancels out and
\[
\lim_{n \to \infty} \frac{1}{2} \sum_{i=n+2}^{2n+1} \frac{1}{i} = \lim_{n \to \infty} \frac{1}{2} \int_{n+2}^{2n+1} \frac{1}{x} \, dx = \lim_{n \to \infty} \ln \frac{2n+1}{n+2} = \ln 2.
\]

c) Repeat part b) where, for any given integer $k > 0$, the order of summation is $k$ positive terms for each negative term.

**Solution:** This is done the same way, and the answer is $\ln k$. What the exercise essentially shows is that in a sum for which both the positive terms sum to infinity and the negative terms sum to $-\infty$, one can get any desired limit by summing terms in an appropriate order. In fact, to reach any desired limit, one alternates between positive terms until exceeding the desired limit and then negative terms until falling below the desired limit.

**Exercise 1.9:** (Proof of Theorem 1.4.1) The bounds on the binomial in this theorem are based on the Stirling bounds. These say that for all $n \geq 1$, $n!$ is upper and lower bounded by
\[
\sqrt{2\pi n} \left( \frac{n}{e} \right)^n < n! < \sqrt{2\pi n} \left( \frac{n}{e} \right)^n e^{1/12n}.
\]
(A.3)
The ratio, $\sqrt{2\pi n(n/e)^n}/n!$, of the first two terms is monotonically increasing with $n$ toward the limit 1, and the ratio $\sqrt{2\pi n(n/e)^n} \exp(1/12n)/n!$ is monotonically decreasing toward 1. The upper bound is more
accurate, but the lower bound is simpler and known as the Stirling approximation. See [8] for proofs and further discussion of the above facts.

a) Show from (A.3) and from the above monotone property that

\[
\binom{n}{k} < \frac{n}{2\pi k(n-k)} \frac{n^n}{k^k(n-k)^{n-k}}.
\]

Hint: First show that \(n!/k! < \sqrt{n/k}n^{n-k}e^{-n+k}\) for \(k < n\).

**Solution:** Since the ratio of the first two terms of (A.3) is increasing in \(n\), we have

\[
\sqrt{2\pi k(e/k)^k} < \sqrt{2\pi n(n/e)^n/n!}.
\]

Rearranging terms, we have the result in the hint. Applying the first inequality of (A.3) to \(n - k\) and combining this with the result on \(n!/k!\) yields the desired result.

b) Use the result of part a) to upper bound \(p_{SN}(k)\) by

\[
p_{SN}(k) < \frac{n}{2\pi k(n-k)} \frac{p^k(1-p)^{n-k}n^n}{k^k(n-k)^{n-k}}.
\]

Show that this is equivalent to the upper bound in Theorem 1.4.1.

**Solution:** Using the binomial equation and then part a),

\[
p_{SN}(k) = \binom{n}{k} p^k(1-p)^{n-k} < \frac{n}{2\pi k(n-k)} \frac{n^n}{k^k(n-k)^{n-k}} p^k(1-p)^{n-k}.
\]

This is the desired bound on \(p_{SN}(k)\). Letting \(\tilde{p} = k/n\), this becomes

\[
p_{SN}(\tilde{p}n) \leq \frac{1}{2\pi n \tilde{p}(1-\tilde{p})} \frac{\tilde{p}^n(1-\tilde{p})^{n(1-\tilde{p})}}{\tilde{p}^n(1-\tilde{p})^{n(1-\tilde{p})}} = \sqrt{\frac{1}{2\pi n \tilde{p}(1-\tilde{p})}} \exp \left( n \left[ \tilde{p} \ln \frac{\tilde{p}}{1-\tilde{p}} + \tilde{p} \ln \frac{1-\tilde{p}}{1-\tilde{p}} \right] \right),
\]

which is the same as the upper bound in Theorem 1.4.1.

c) Show that

\[
\binom{n}{k} > \frac{n}{2\pi k(n-k)} \frac{n^n}{k^k(n-k)^{n-k}} \left[ 1 - n \frac{n}{12k(n-k)} \right].
\]

**Solution:** Use the factorial lower bound on \(n!\) and the upper bound on \(k\) and \((n-k)!\). This yields

\[
\binom{n}{k} > \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k(n-k)^{n-k}} \exp \left( -\frac{1}{12k} - \frac{1}{12(n-k)} \right) > \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k(n-k)^{n-k}} \left[ 1 - n \frac{n}{12k(n-k)} \right],
\]

where the latter equation comes from combining the two terms in the exponent and then using the bound \(e^{-x} > 1 - x\).
d) Derive the lower bound in Theorem 1.4.1.

**Solution:** This follows by substituting \( \hat{p}n \) for \( k \) in the solution to c) and substituting this in the binomial formula.

e) Show that \( \phi(p, \hat{p}) = \hat{p} \ln(\hat{p}) + (1 - \hat{p}) \ln(\frac{1 - \hat{p}}{1 - \hat{p}}) \) is 0 at \( \hat{p} = p \) and nonnegative elsewhere.

**Solution:** Taking the first two derivatives of \( \phi(p, \hat{p}) \) with respect to \( \hat{p} \),

\[
\frac{\partial \phi(p, \hat{p})}{\partial \hat{p}} = -\ln \left( \frac{p(1 - \hat{p})}{\hat{p}(1 - \hat{p})} \right) \quad \frac{\partial^2 \phi(p, \hat{p})}{\partial \hat{p}^2} = \frac{1}{\hat{p}(1 - \hat{p})}.
\]

Since the second derivative is positive for \( 0 < \hat{p} < 1 \), the minimum of \( \phi(p, \hat{p}) \) with respect to \( \hat{p} \) is 0, is achieved where the first derivative is 0, i.e., at \( \hat{p} = p \). Thus \( \phi(p, \hat{p}) > 0 \) for \( \hat{p} \neq p \). Furthermore, \( \phi(p, \hat{p}) \) increases as \( \hat{p} \) moves in either direction away from \( p \).

**Exercise 1.10:** Let \( X \) be a ternary rv taking on the 3 values 0, 1, 2 with probabilities \( p_0, p_1, p_2 \) respectively. Find the median of \( X \) for each of the cases below.

- a) \( p_0 = 0.2, \ p_1 = 0.4, \ p_2 = 0.4 \).
- b) \( p_0 = 0.2, \ p_1 = 0.2, \ p_2 = 0.6 \).
- c) \( p_0 = 0.2, \ p_1 = 0.3, \ p_2 = 0.5 \).

Note 1: The median is not unique in part c). Find the interval of values that are medians. Note 2: Some people force the median to be distinct by defining it as the midpoint of the interval satisfying the definition given here.

**Solution:** The median of \( X \) is 1 for part a), 2 for part b), and the interval \([1, 2]\) for part c).

d) Now suppose that \( X \) is nonnegative and continuous with the density \( f_X(x) = 1 \) for \( 0 \leq x \leq 0.5 \) and \( f_X(x) = 0 \) for \( 0.5 < x \leq 1 \). We know that \( f_X(x) \) is positive for all \( x > 1 \), but it is otherwise unknown. Find the median or interval of medians.

The median is sometimes (incorrectly) defined as that \( \alpha \) for which \( \Pr\{X > \alpha\} = \Pr\{X < \alpha\} \). Show that it is possible for no such \( \alpha \) to exist. Hint: Look at the examples above.

**Solution:** The interval of medians is \([0.5, 1]\). In particular, \( \Pr\{X \leq x\} = 1/2 \) for all \( x \) in this interval and \( \Pr\{X \geq x\} = 1/2 \) in this interval.

There is no \( \alpha \) for which \( \Pr\{X < \alpha\} = \Pr\{X > \alpha\} \) in all of the first 3 examples. One should then ask why there must always be an \( x \) such that \( \Pr\{X \geq x\} \geq 1/2 \) and \( \Pr\{X \leq x\} \geq 1/2 \).

To see this, let \( x_0 = \inf\{x : F_X(x) \geq 1/2\} \). We must have \( F_X(x_0) \geq 1/2 \) since \( F_X \) is continuous on the right. Because of the infimum, we must have \( F_X(x_0 - \epsilon) < 1/2 \) for all \( \epsilon > 0 \), and therefore \( \Pr\{X \geq x_0 - \epsilon\} \geq 1/2 \). But \( \Pr\{X \geq x\} \) is continuous on the left for the same reason that \( F_X(x) \) is continuous on the right, and thus \( x_0 \) is a median of \( X \). This is the kind of argument that makes people hate analysis.

**Exercise 1.11:** a) For any given rv \( Y \), express \( \mathbb{E}[|Y|] \) in terms of \( \int_{y<0} F_Y(y) \, dy \) and \( \int_{y>0} F_Y^c(y) \, dy \). Hint: Review the argument in Figure 1.4.

**Solution:** We have seen in (1.34) that

\[
\mathbb{E}[Y] = -\int_{y<0} F_Y(y) \, dy + \int_{y\geq0} F_Y^c(y) \, dy
\]
Since all negative values of \( Y \) become positive in \( |Y| \),

\[
E[|Y|] = + \int_{y<0} F_Y(y) \, dy + \int_{y\geq0} F_Y^c(y) \, dy
\]

To spell this out in greater detail, let \( Y = Y^+ + Y^- \) where \( Y^+ = \max(0,Y) \) and \( Y^- = \min(Y,0) \). Then \( Y = Y^+ + Y^- \) and \( |Y| = Y^+ - Y^- \). Since \( E[Y^+] = \int_{y\geq0} F_Y(y) \, dy \) and \( E[Y^-] = -\int_{y\geq0} F_Y^c(y) \, dy \), the above results follow.

b) For some given rv \( X \) with \( E[|X|] < \infty \), let \( Y = X - \alpha \). Using part a), show that

\[
E[|X - \alpha|] = \int_{-\infty}^{\alpha} F_X(x) \, dx + \int_{\alpha}^{\infty} F_X(x) \, dx.
\]

**Solution:** This follows by changing the variable of integration in part a). That is,

\[
E[|X - \alpha|] = E[|Y|] = + \int_{y<0} F_Y(y) \, dy + \int_{y\geq0} F_Y^c(y) \, dy
\]

\[
= \int_{-\infty}^{\alpha} F_X(x) \, dx + \int_{\alpha}^{\infty} F_X(x) \, dx.
\]

where in the last step, we have changed the variable of integration from \( y \) to \( x - \alpha \).

c) Show that \( E[|X - \alpha|] \) is minimized over \( \alpha \) by choosing \( \alpha \) to be a median of \( X \). Hint: Both the easy way and the most instructive way to do this is to use a graphical argument illustrating the above two integrals. Be careful to show that when the median is an interval, all points in this interval achieve the minimum.

**Solution:** As illustrated in the picture, we are minimizing an integral for which the integrand changes from \( F_X(x) \) to \( F_X^c(x) \) at \( x = \alpha \). If \( F_X(x) \) is strictly increasing in \( x \), then \( F_X^c = 1 - F_X \) is strictly decreasing. We then minimize the integrand over all \( x \) by choosing \( \alpha \) to be the point where the curves cross, i.e., where \( F_X(x) = .5 \). Since the integrand has been minimized at each point, the integral must also be minimized.

If \( F_X \) is continuous but not strictly increasing, then there might be an interval over which \( F_X(x) = .5 \); all points on this interval are medians and also minimize the integral; Exercise 1.10 part c) gives an example where \( F_X \) is not continuous. Finally, if \( F_X(\alpha) \geq .5 \) and \( F_X(\alpha - \epsilon) < .5 \) for some \( \alpha \) and all \( \epsilon > 0 \) (as in parts a) and b) of Exercise 1.10), then the integral is minimized at that \( \alpha \) and that \( \alpha \) is also the median.

**Exercise 1.12:** Let \( X \) be a rv with CDF \( F_X(x) \). Find the CDF of the following rv’s.

a) The maximum of \( n \) IID rv’s, each with CDF \( F_X(x) \).

**Solution:** Let \( M_n \) be the maximum of the \( n \) rv’s \( X_1, \ldots, X_n \). Note that for any real \( x \),
\(M_+\) is less than or equal to \(x\) if and only if \(X_j \leq x\) for each \(j, 1 \leq j \leq n\). Thus

\[
\Pr\{M_+ \leq x\} = \Pr\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\} = \prod_{j=1}^{n} \Pr\{X_j \leq x\},
\]

where we have used the independence of the \(X_j\)'s. Finally, since \(\Pr\{X_j \leq x\} = F_X(x)\) for each \(j\), we have \(F_{M_+}(x) = \Pr\{M_+ \leq x\} = (F_X(x))^n\).

b) The minimum of \(n\) IID rv's, each with CDF \(F_X(x)\).

**Solution:** Let \(M_-\) be the minimum of \(X_1, \ldots, X_n\). Then, in the same way as in part (a), \(M_- > y\) if and only if \(X_j > y\) for \(1 \leq j \leq n\) and for all choice of \(y\). We could make the same statement using greater than or equal in place of strictly greater than, but the strict inequality is what is needed for the CDF. Thus,

\[
\Pr\{M_- > y\} = \Pr\{X_1 > y, X_2 > y, \ldots, X_n > y\} = \prod_{j=1}^{n} \Pr\{X_j > y\},
\]

It follows that \(1 - F_{M_-}(y) = \left(1 - F_X(y)\right)^n\).

c) The difference of the rv's defined in a) and b); assume \(X\) has a density \(f_X(x)\).

**Solution:** There are many difficult ways to do this, but also a simple way, based on first conditioning on the event that \(X_1 = x\). Then \(X_1 = M_+\) if and only if \(X_j \leq x\) for \(2 \leq j \leq n\). Also, given \(X_1 = M_+ = x\), we have \(R = M_+ - M_- \leq r\) if and only if \(X_j > x - r\) for \(2 \leq j \leq n\). Thus, since the \(X_j\) are IID,

\[
\Pr\{M_+ = X_1, R \leq r \mid X_1 = x\} = \prod_{j=2}^{n} \Pr\{x - r < X_j \leq x\}
= [\Pr\{x - r < X \leq x\}]^{n-1} = [F_X(x) - F_X(x - r)]^{n-1}
\]

We can now remove the conditioning by averaging over \(X_1 = x\). Assuming that \(X\) has the density \(f_X(x)\),

\[
\Pr\{X_1 = M_+, R \leq r\} = \int_{-\infty}^{\infty} f_X(x) [F_X(x) - F_X(x - r)]^{n-1} dx
\]

Finally, we note that the probability that two of the \(X_j\) are the same is 0 so the events \(X_j = M_+\) are disjoint except with zero probability. Also we could condition on \(X_j = x\) instead of \(X_1\) with the same argument (i.e., by using symmetry), so \(\Pr\{X_j = M_+, R \leq r\} = \Pr\{X_1 = M_+ R \leq r\}\) It follows that

\[
\Pr\{R \leq r\} = \int_{-\infty}^{\infty} n f_X(x) [F_X(x) - F_X(x - r)]^{n-1} dx
\]

The only place we really needed the assumption that \(X\) has a PDF was in asserting that the probability that two or more of the \(X_j\)'s are jointly equal to the maximum is 0. The formula can be extended to arbitrary CDF's by being careful about this possibility.
These expressions can be evaluated easily if $X$ is exponential with the PDF $\lambda e^{-\lambda x}$ for $x \geq 0$. Then $M_\cdot$ has the PDF $2\lambda e^{-2\lambda x}$ and $R$ is exponential with the PDF $\lambda e^{-\lambda x}$. We will look at this in a number of ways when we study Poisson processes.

**Exercise 1.13:** Let $X$ and $Y$ be rv’s in some sample space $\Omega$ and let $Z = X + Y$, i.e., for each $\omega \in \Omega$, $Z(\omega) = X(\omega) + Y(\omega)$. The purpose of this exercise is to show that $Z$ is a rv. This is a mathematical fine point that many readers may prefer to simply accept without proof.

a) Show that the set of $\omega$ for which $Z(\omega)$ is infinite or undefined has probability 0.

**Solution:** Note that $Z$ can be infinite (either $\pm \infty$) or undefined only when either $X$ or $Y$ are infinite or undefined. Since these are events of zero probability, $Z$ is infinite or undefined with probability 0.

b) We must show that $\{\omega \in \Omega : Z(\omega) \leq \alpha\}$ is an event for each real $\alpha$, and we start by approximating that event. To show that $Z = X + Y$ is a rv, we must show that for each real number $\alpha$, the set $\{\omega \in \Omega : X(\omega) + Y(\omega) \leq \alpha\}$ is an event. Let $B(n, k) = \{\omega : X(\omega) \leq k/n\} \cap \{Y(\omega) \leq \alpha + (1 - k)/n\}$ for integer $k > 0$. Let $D(n) = \bigcup_k B(n, k)$, and show that $D(n)$ is an event.

**Solution:** For each $n$, $D(n)$ is a countable union (over $k$) of the sets $B(n, k)$. Each such set is an intersection of two events, namely the event $\{\omega : X(\omega) \leq k/n\}$ and the event $\{Y(\omega) \leq \alpha + (1 - k)/n\}$; these must be events since $X$ and $Y$ are rv’s.

c) On a 2 dimensional sketch for a given $\alpha$, show the values of $X(\omega)$ and $Y(\omega)$ for which $\omega \in D(n)$. Hint: This set of values should be bounded by a staircase function.

**Solution:**

The region $D(n)$ is sketched for $\alpha n = 5$; it is the region to the left of the staircase function above.

d) Show that

$$\{\omega : X(\omega) + Y(\omega) \leq \alpha\} = \bigcap_{n \geq 1} D(n).$$

(A.4)

**Solution:** The region $\{\omega : X(\omega) + Y(\omega) \leq \alpha\}$ is the region to the left of a diagonal line of slope -1 that passes through the point $(0, \alpha)$. This region is thus contained in $D(n)$ for each $n \geq 1$ and is thus contained in $\bigcap_{n \geq 1} D(n)$. On the other hand, each point $\omega$ for which $X(\omega) + Y(\omega) > \alpha$ is not contained in $D(n)$ for sufficiently large $n$. This verifies (A.4). Since $D(n)$ is an event, the countable intersection is also an event, so $\{\omega : X(\omega) + Y(\omega) \leq \alpha\}$ is an event. This applies for all $\alpha$. This in conjunction with part a) shows that $Z$ is a rv.
e) Explain why this implies that if \( Y = X_1 + X_2 + \cdots + X_n \) and if \( X_1, X_2, \ldots, X_n \) are rv's, then \( Y \) is a rv.

**Solution:** We have shown that \( X_1 + X_2 \) is a rv, so \( (X_1 + X_2) + X_3 \) is a rv, etc.

**Exercise 1.14:** a) Let \( X_1, X_2, \ldots, X_n \) be rv's with expected values \( \overline{X}_1, \ldots, \overline{X}_n \). Show that \( E[X_1 + \cdots + X_n] = \overline{X}_1 + \cdots + \overline{X}_n \). You may assume that the rv's have a joint density function, but do not assume that the rv's are independent.

**Solution:** We assume that the rv's have a joint density, and we ignore all mathematical fine points here. Then

\[
E[X_1 + \cdots + X_n] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1 + \cdots + x_n) f_{X_1 \cdots X_n}(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]

\[
= \sum_{j=1}^{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_j f_{X_1 \cdots X_n}(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]

\[
= \sum_{j=1}^{n} \int_{-\infty}^{\infty} x_j f_{X_j}(x_j) \, dx_j = \sum_{j=1}^{n} E[X_j].
\]

Note that the separation into a sum of integrals simply used the properties of integration and that no assumption of statistical independence was made.

b) Now assume that \( X_1, \ldots, X_n \) are statistically independent and show that the expected value of the product is equal to the product of the expected values.

**Solution:** From the independence, \( f_{X_1 \cdots X_n}(x_1, \ldots, x_n) = \prod_{j=1}^{n} f_{X_j}(x_j) \). Thus

\[
E[X_1X_2 \cdots X_n] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{n} x_j \prod_{j=1}^{n} f_{X_j}(x_j) \, dx_1 \cdots dx_n
\]

\[
= \prod_{j=1}^{n} \int_{-\infty}^{\infty} x_j f_{X_j}(x_j) \, dx_j = \prod_{j=1}^{n} E[X_j].
\]

c) Again assuming that \( X_1, \ldots, X_n \) are statistically independent, show that the variance of the sum is equal to the sum of the variances.

**Solution:** Since part a) shows that \( E \left[ \sum_{j=1}^{n} X_j \right] = \sum_{j=1}^{n} \overline{X}_j \), we have

\[
\text{VAR} \left[ \sum_{j=1}^{n} X_j \right] = E \left[ \left( \sum_{j=1}^{n} X_j - \sum_{j=1}^{n} \overline{X}_j \right)^2 \right]
\]

\[
= E \left[ \sum_{j=1}^{n} \sum_{i=1}^{n} (X_j - \overline{X}_j)(X_i - \overline{X}_i) \right]
\]

\[
= \sum_{j=1}^{n} \sum_{i=1}^{n} E \left[ (X_j - \overline{X}_j)(X_i - \overline{X}_i) \right] \quad (A.5)
\]
where we have again used part a). Now from part b), and using the independence of the $X_j$, $E[(X_j - \bar{X}_j)(X_i - \bar{X}_i)] = 0$ for $i \neq j$. Thuds(A.5) simplifies to

$$\text{VAR} \left[ \sum_{j=1}^{n} X_j \right] = \sum_{j=1}^{n} E \left[ (X_j - \bar{X}_j)^2 \right] = \sum_{j=1}^{n} \text{VAR}[X_j]$$

**Exercise 1.15:** (Stieltjes integration) a) Let $h(x) = u(x)$ and $F_X(x) = u(x)$ where $u(x)$ is the unit step, i.e., $u(x) = 0$ for $-\infty < x < 0$ and $u(x) = 1$ for $x \geq 0$. Using the definition of the Stieltjes integral in Footnote 19, show that $\int_{-1}^{1} h(x) dF_X(x)$ does not exist. Hint: Look at the term in the Riemann sum including $x = 0$ and look at the range of choices for $h(x)$ in that interval. Intuitively, it might help initially to view $dF_X(x)$ as a unit impulse at $x = 0$.

**Solution:** The Riemann sum for this Stieltjes integral is $\sum_{n} h(x_n)[F(y_n) - F(y_{n-1})]$ where $y_{n-1} < x_n \leq y_n$. For any partition $\{y_n; n \geq 1\}$, consider the $k$ such that $y_{k-1} < 0 \leq y_k$ and consider choosing either $x_n < 0$ or $x_n \geq 0$. In the first case $h(x_n)[F(y_n) - F(y_{n-1})] = 0$ and in the second $h(x_n)[F(y_n) - F(y_{n-1})] = 1$. All other terms are 0 and this can be done for all partitions as $\delta \to 0$, so the integral is undefined.

b) Let $h(x) = u(x-a)$ and $F_X(x) = u(x-b)$ where $a$ and $b$ are in $(-1, +1)$. Show that $\int_{-1}^{1} h(x) dF_X(x)$ exists if and only if $a \neq b$. Show that the integral has the value 1 for $a < b$ and the value 0 for $a > b$. Argue that this result is still valid in the limit of integration over $(-\infty, \infty)$.

**Solution:** Using the same argument as in part a) for any given partition $\{y_n; n \geq 1\}$, consider the $k$ such that $y_{k-1} < b \leq y_k$. If $a = b$, $x_n$ can be chosen to make $h(x_n)$ either 0 or 1, causing the integral to be undefined as in part a). If $a < b$, then for a sufficiently fine partition, $h(x_k) = 1$ for all $x_k$ such that $y_{k-1} < x_k \leq y_k$. Thus that term in the Riemann sum is 1. For all other $n$, $F_Y(y_n) - F_Y(n-1) = 0$, so the Riemann sum is 1. For $a > b$ and $k$ as before, $h(x_k) = 0$ for a sufficiently fine partition, and the integral is 0. The argument does not involve the finite limits of integration, so the integral remains the same as the limits become infinite.

c) Let $X$ and $Y$ be independent discrete rv’s, each with a finite set of possible values. Show that $\int_{-\infty}^{\infty} F_X(z-y) dF_Y(y)$, defined as a Stieltjes integral, is equal to the distribution of $Z = X + Y$ at each $z$ other than the possible sample values of $Z$, and is undefined at each sample value of $Z$. Hint: Express $F_X$ and $F_Y$ as sums of unit steps. Note: This failure of Stieltjes integration is not a serious problem; $F_Z(z)$ is a step function, and the integral is undefined at its points of discontinuity. We automatically define $F_Z(z)$ at those step values so that $F_Z$ is a CDF (i.e., is continuous from the right). This problem does not arise if either $X$ or $Y$ is continuous.

**Solution:** Let $X$ have the PMF $\{p(x_1), \ldots, p(x_K)\}$ and $Y$ have the PMF $\{p_Y(y_1), \ldots, p_Y(y_J)\}$. Then $F_X(x) = \sum_{k=1}^{K} p(x_k)u(x-x_k)$ and $F_Y(y) = \sum_{j=1}^{J} q(y_j)u(y-y_j)$. Then

$$\int_{-\infty}^{\infty} F_X(z-y) dF_Y(y) = \sum_{k=1}^{K} \sum_{j=1}^{J} \int_{-\infty}^{\infty} p(x_k)q(y_j)u(z-y_j-x_k)du(y-y_j).$$

From part b), the integral above for a given $k, j$ exists unless $z = x_k + y_j$. In other words, the Stieltjes integral gives the CDF of $X + Y$ except at those $z$ equal to $x_k + y_j$ for some
Exercise 1.16: Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of IID continuous rv’s with the common probability density function $f_X(x)$; note that $\Pr\{X=\alpha\} = 0$ for all $\alpha$ and that $\Pr\{X_i=X_j\} = 0$ for all $i \neq j$. For $n \geq 2$, define $X_n$ as a record-to-date of the sequence if $X_n > X_i$ for all $i < n$.

a) Find the probability that $X_2$ is a record-to-date. Use symmetry to obtain a numerical answer without computation. A one or two line explanation should be adequate).

Solution: $X_2$ is a record-to-date with probability 1/2. The reason is that $X_1$ and $X_2$ are IID, so either one is larger with probability 1/2; this uses the fact that they are equal with probability 0 since they have a density.

b) Find the probability that $X_n$ is a record-to-date, as a function of $n \geq 1$. Again use symmetry.

Solution: By the same symmetry argument, each $X_i, 1 \leq i \leq n$ is equally likely to be the largest, so that each is largest with probability 1/n. Since $X_n$ is a record-to-date if an only if it is the largest of $X_1, \ldots, X_n$, it is a record-to-date with probability 1/n.

c) Find a simple expression for the expected number of records-to-date that occur over the first $m$ trials for any given integer $m$. Hint: Use indicator functions. Show that this expected number is infinite in the limit $m \to \infty$.

Solution: Let $I_n$ be 1 if $X_n$ is a record-to-date and be 0 otherwise. Thus $\mathbb{E}[I_i]$ is the expected value of the ‘number’ of records-to-date (either 1 or 0) on trial $i$. That is

$$
\mathbb{E}[I_n] = \Pr\{I_n = 1\} = \Pr\{X_n\text{is a record-to-date}\} = 1/n.
$$

Thus

$$
\mathbb{E}[\text{records-to-date up to } m] = \sum_{n=1}^{m} \mathbb{E}[I_n] = \sum_{n=1}^{m} \frac{1}{n}.
$$

This is the harmonic series, which goes to $\infty$ in the limit $m \to \infty$. If you are unfamiliar with this, note that $\sum_{n=1}^{\infty} 1/n \geq \int_{1}^{\infty} \frac{1}{x} \, dx = \infty$.

Exercise 1.17: (Continuation of Exercise 1.16) a) Let $N_1$ be the index of the first record-to-date in the sequence. Find $\Pr\{N_1 > n\}$ for each $n \geq 2$. Hint: There is a far simpler way to do this than working from part b in Exercise 1.16.

Solution: The event $\{N_1 > n\}$ is the event that no record-to-date occurs in the first $n$ trials, which means that $X_1$ is the largest of $\{X_1, X_2, \ldots, X_n\}$, which by symmetry has probability 1/n. Thus $\Pr\{N_1 > n\} = 1/n$. 

k, j, i.e., equal to the values of $Z$ at which $F_Z(z)$ (as found by discrete convolution) has step discontinuities.

To give a more intuitive explanation, $F_X(x) = \Pr\{X \leq x\}$ for any discrete rv $X$ has jumps at the sample values of $X$ and the value of $F_X(x_k)$ at any such $x_k$ includes $p(x_k)$, i.e., $F_X$ is continuous to the right. The Riemann sum used to define the Stieltjes integral is not sensitive enough to ‘see’ this step discontinuity at the step itself. Thus, the knowledge that $Z$ is a CDF, and thus continuous on the right, must be used in addition to the Stieltjes integral to define $F_Z$ at its jumps.
b) Show that $N_1$ is a rv (i.e., that $N_1$ is not defective).

Solution: Every sample sequence for $X_1, X_2, \ldots$, maps into either a positive integer or infinity for $N_1$. The probability that $N_1$ is infinite is $\lim_{n \to \infty} \Pr\{N > n\}$, which is 0. Thus $N_1$ is finite with probability 1 and is thus a rv.

c) Show that $E[N_1] = \infty$.

Solution: Since $N_1$ is a nonnegative rv,

$$E[N_1] = \int_0^\infty \Pr\{N_1 > x\} \, dx = 1 + \sum_{n=1}^\infty \frac{1}{n} = \infty.$$ 

d) Let $N_2$ be the index of the second record-to-date in the sequence. Show that $N_2$ is a rv. You need not find the CDF of $N_2$ here.

Solution: We will show that $N_2$ is finite conditioned on the sample value $n_1$ of $N_1$ and the sample value $x_{n_1}$ of $X_{N_1}$. Each $X_n$ for $n > n_1$ will independently exceed $x_{n_1}$ with probability $1 - F_X(x_{n_1})$. Thus, assuming $1 - F_X(x_{n_1}) > 0$, $N_2$ is finite with probability 1. But $1 - F_X(x_{n_1} > 0)$ must be positive with probability 1 since $X$ is a continuous rv.

Another approach is to calculate $\Pr\{N_2 > n\}$ by first calculating $\Pr\{N_2 > n, N_1 = j\}$ for $2 \leq j \leq n$. The event $N_2 > n, N_1 = j$ means that $X_j > X_i$ both for $1 \leq i < j$ and for $j < i \leq n$. It also means that $X_1$ is the largest $X_i$ for $1 \leq i < j$. Using symmetry as before, $\Pr\{\bigcup_{i=1}^{n} \{X_j \geq X_i\}\} = 1/n$. Independent of this, $\Pr\{\bigcup_{i=2}^{j-1} \{X_1 \geq X_i\}\} = 1/j$. Thus,

$$\Pr\{N_2 > n, N_1 = j\} = \frac{1}{jn}.$$ 

Now the event $N_2 > n$ can happen with $N_1 = j$ for each $j$, $2 \leq j \leq n$ and also for $N_1 > n$, which has probability $1/n$. Thus

$$\Pr\{N_2 > n\} = \sum_{j=2}^{n} \frac{1}{nj} + \frac{1}{n} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{j}.$$ 

Since $\sum_{j=1}^{n} 1/j$ grows as $\ln n$, we see that $\lim_{n \to \infty} \Pr\{N_2 > n\} = 0$, showing that $N_2$ is a rv.

Note that $\Pr\{N_2 > n\}$ is considerably larger than $\Pr\{N_1 > n\}$, and their ratio grows as $\ln n$. The reason for this is that after one record occurs, the ‘number to beat,’ i.e., $X_{N_1}$ is larger than $X_1$. One should think through why this effect does not depend on the CDF of $X$.

e) Contrast your result in part c to the result from part c of Exercise 1.16 saying that the expected number of records-to-date is infinite over an an infinite number of trials. Note: this should be a shock to your intuition — there is an infinite expected wait for the first of an infinite sequence of occurrences.

Solution: Even though the expected wait for the first record-to-date is infinite, it is still a random variable, and thus the first record-to-date must eventually occur. We have also shown that the second record-to-date eventually occurs, and it can be shown that the $n$th
eventually occurs for all \( n \). This makes the result in Exercise 1.16 unsurprising once it is understood.

Exercise 1.18: (Another direction from Exercise 1.16) a) For any given \( n \geq 2 \), find the probability that \( X_n \) and \( X_{n+1} \) are both records-to-date. Hint: The idea in part b of 1.16 is helpful here, but the result is not.

Solution: For both \( X_{n+1} \) and \( X_n \) to be records-to-date it is necessary and sufficient for \( X_{n+1} \) to be larger than all the earlier \( X_i \) (including \( X_n \)) and for \( X_n \) to be larger than all of the \( X_i \) earlier than it. Since any one of the first \( n + 1 \) \( X_i \) is equally likely to be the largest, the probability that \( X_{n+1} \) is the largest is \( 1/(n+1) \). For \( X_n \) to also be a record-to-date, it must be the second largest of the \( n + 1 \). Since all the first \( n \) terms are equally likely to be the second largest (given that \( X_{n+1} \) is the largest), the conditional probability that \( X_n \) is the second largest is \( 1/n \). Thus,

\[
\Pr \{X_{n+1} \text{ and } X_n \text{ are records-to-date}\} = \frac{1}{n(n+1)}.
\]

Note that there might be earlier records-to-date before \( n \); we have simply calculated the probability that \( X_n \) and \( X_{n+1} \) are records-to-date.

b) Is the event that \( X_n \) is a record-to-date statistically independent of the event that \( X_{n+1} \) is a record-to-date?

Solution: Yes, we have found the joint probability that \( X_n \) and \( X_{n+1} \) are records, and it is the product of the events that each are records individually.

c) Find the expected number of adjacent pairs of records-to-date over the sequence \( X_1, X_2, \ldots \). Hint: A helpful fact here is that \( \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1} \).

Solution: Let \( \mathbb{I}_n \) be the indicator function of the event that \( X_n \) and \( X_{n+1} \) are records-to-date. Then \( \mathbb{E}[\mathbb{I}_n] = \Pr \{\mathbb{I}_n = 1\} = \frac{1}{n(n+1)} \). The expected number of pairs of records (for example, counting records at 2, 3, and 4 as two pairs of records), the expected number of pairs over the sequence is

\[
\mathbb{E}[\text{Number of pairs}] = \sum_{n=2}^{\infty} \frac{1}{n(n+1)} = \sum_{n=2}^{\infty} \left( \frac{1}{n} - \frac{1}{n+1} \right)
\]

\[
= \sum_{n=2}^{\infty} \frac{1}{n} - \sum_{n=3}^{\infty} \frac{1}{n} = \frac{1}{2},
\]

where we have used the hint and then summed the terms separately. This hint is often useful in analyzing stochastic processes.

The intuition here is that records-to-date tend to become more rare with increasing \( n \) (\( X_n \) is a record with probability \( 1/n \)). As we have seen, the expected number of records from 2 to \( m \) is on the order of \( \ln m \), which grows very slowly with \( m \). The probability of an adjacent pair of records, as we have seen, decreases as \( 1/n(n+1) \) with \( n \), which means that if one does not occur for small \( n \), it will probably not occur at all. It can be seen from this that the time until the first pair of records is a defective random variable.
Exercise 1.19: a) Assume that $X$ is a nonnegative discrete rv taking on values $a_1, a_2, \ldots$, and let $Y = h(X)$ for some nonnegative function $h$. Let $b_i = h(a_i)$, $i \geq 1$ be the $i^{th}$ value taken on by $Y$. Show that $E[Y] = \sum_i b_i p_Y(b_i) = \sum_i h(a_i) p_X(a_i)$. Find an example where $E[X]$ exists but $E[Y] = \infty$.

Solution: If we make the added assumption that $b_i \neq b_j$ for all $i \neq j$, then $Y$ has the sample value $b_i$ if and only if $X$ has the sample value $a_i$; thus $p_Y(b_i) = p_X(a_i)$ for each $i$. It then follows that $\sum_i b_i p_Y(b_i) = \sum_i h(a_i) p_X(a_i)$. This must be $E[Y]$ (which might be finite or infinite).

A simple example where $E[X]$ is finite and $E[y] = \infty$ is to choose $a_1, a_2, \ldots$, to be 1, 2, \ldots and choose $p_X(i) = 2^{-i}$. Then $E[X] = 2$. Choosing $h(i) = 2^i$, we have $b_i = 2^i$ and $E[Y] = \sum_i 2^i \cdot 2^{-i} = \infty$. Without the assumption that $b_i \neq b_j$, the set of sample points of $Y$ is the set of distinct values of $b_i$. We still have $E[Y] = \sum_i h(a_i) p_X(a_i)$ but it is no longer correct that $E[Y] = \sum_i b_i p_Y(b_i)$.

b) Let $X$ be a nonnegative continuous rv with density $f_X(x)$ and let $h(x)$ be differentiable, nonnegative, and nondecreasing in $x$. Let $A(\delta) = \sum_{n \geq 1} h(n \delta) [F(n \delta) - F(n \delta - \delta)]$, i.e., $A(\delta)$ is a $\delta$th order approximation to the Stieltjes integral $\int h(x) dF(x)$. Show that if $A(1) < \infty$, then $A(2^{-k}) \leq A(2^{-(k-1)}) < \infty$ for $k \geq 1$.

Solution: Let $\delta = 2^{-k}$ for $k \geq 1$. We take the expression for $A(2\delta)$ and break each interval of size $2\delta$ into two intervals each of size $\delta$; we use this to relate $A(2\delta)$ to $A(\delta)$.

$$A(2\delta) = \sum_{n \geq 1} h(2n\delta) \left[ F(2n\delta) - F(2n\delta - 2\delta) \right]$$

$$= \sum_{n \geq 1} h(2n\delta) \left[ F(2n\delta) - F(2n\delta - \delta) \right] + \sum_{n \geq 1} h(2n\delta) \left[ F(2n\delta - \delta) - F(2n\delta - 2\delta) \right]$$

$$\geq \sum_{n \geq 1} h(2n\delta) \left[ F(2n\delta) - F(2n\delta - \delta) \right] + \sum_{n \geq 1} h(2n\delta - \delta) \left[ F(2n\delta - \delta) - F(2n\delta - 2\delta) \right]$$

$$= \sum_{k \geq 1} h(k\delta) \left[ F(k\delta) - F(k\delta - \delta) \right] = A(\delta)$$

For $k = 1$, $\delta = 2^{-k} = 1/2$, so $A(2\delta) = A(1) < \infty$. Thus $A(1/2) < \infty$. Using induction, $A(2^{-k}) < \infty$ for all $k > 1$. Also $A(2^{-k})$ is nonnegative and nonincreasing in $k$, so $\lim_{k \to \infty} A(2^{-k})$ has a limit, which must be the value of the Stieltjes integral.

To be a little more precise about the Stieltjes integral, we see that $A(\delta)$ as defined above uses the largest value, $h(n \delta)$, of $h(x)$ over the interval $x \in [n\delta - \delta, n\delta]$. By replacing $h(n \delta)$ by $h(n \delta - \delta)$, we get the smallest value in each interval, and then the sequence is nonincreasing with the same limit. For an arbitrary partition of the real line, rather than the equi-spaced partition here, the argument here would have to be further extended.

Exercise 1.20: a) Consider a positive, integer-valued rv whose CDF is given at integer values by

$$F_Y(y) = 1 - \frac{2}{(y+1)(y+2)} \quad \text{for integer } y \geq 0.$$

Use (1.31) to show that $E[Y] = 2$. Hint: Note that $1/[(y+1)(y+2)] = 1/(y+1) - 1/(y+2)$. 

$$E[Y] = \int_0^\infty 1 - \frac{2}{(y+1)(y+2)} \, dy$$

$$= \left. \left[ y \right] \right|_1^\infty - \left. \left[ \frac{2}{y+1} \right] \right|_1^\infty$$

$$= \infty - 2 \left[ \frac{2}{y+1} \right]_1^\infty$$

$$= \infty - 2 \left( \frac{2}{\infty} - \frac{2}{2} \right)$$

$$= \infty - 2(1)$$

$$= \infty$$

$$= 2$$
Solution: Combining (1.31) with the hint, we have
\[
E[Y] = \sum_{y \geq 0} F_Y(y) = \sum_{y \geq 0} \frac{2}{y + 1} - \sum_{y \geq 0} \frac{2}{y + 2}
\]
\[
= \sum_{y \geq 0} \frac{2}{y + 1} - \sum_{y \geq 1} \frac{2}{y + 1} = 2,
\]
where the second sum in the second line eliminates all but the first term of the first sum.

b) Find the PMF of \( Y \) and use it to check the value of \( E[Y] \).

Solution: For \( y = 0 \), \( p_Y(0) = F_Y(0) = 0 \). For integer \( y \geq 1 \), \( p_Y(y) = F_Y(y) - F_Y(y - 1) \). Thus for \( y \geq 1 \),
\[
p_Y(y) = \frac{2}{y(y + 1)} - \frac{2}{(y + 1)(y + 2)} = \frac{4}{y(y + 1)(y + 2)}.
\]
Finding \( E[Y] \) from the PMF, we have
\[
E[Y] = \sum_{y=1}^{\infty} y p_Y(y) = \sum_{y=1}^{\infty} \frac{4}{(y + 1)(y + 2)}
\]
\[
= \sum_{y=1}^{\infty} \frac{4}{y + 1} - \sum_{y=2}^{\infty} \frac{4}{y + 1} = 2.
\]

c) Let \( X \) be another positive, integer-valued rv. Assume its conditional PMF is given by
\[
p_{X|Y}(x|y) = \frac{1}{y} \quad \text{for } 1 \leq x \leq y.
\]
Find \( E[X \mid Y = y] \) and use it to show that \( E[X] = 3/2 \). Explore finding \( p_X(x) \) until you are convinced that using the conditional expectation to calculate \( E[X] \) is considerably easier than using \( p_X(x) \).

Solution: Conditioned on \( Y = y \), \( X \) is uniform over \( \{1, 2, \ldots, y\} \) and thus has the conditional mean \((y + 1)/2 \). If you are unfamiliar with this fact, think of adding \( \{1 + 2 + \cdots + y\} \) to \( \{y + y - 1 + \cdots + 1\} \), getting \( y(y + 1) \). Thus \( \{1 + 2 + \cdots + y\} = y(y + 1)/2 \). It follows that
\[
E[X] = E[E[X \mid Y]] = E\left[\frac{Y + 1}{2}\right] = 3/2.
\]
Simply writing out a formula for \( p_X(x) \) is messy enough to realize that this is a poor way to calculate \( E[X] \).

d) Let \( Z \) be another integer-valued rv with the conditional PMF
\[
p_{Z|Y}(z|y) = \frac{1}{y^2} \quad \text{for } 1 \leq z \leq y^2.
\]
Find \( E[Z \mid Y = y] \) for each integer \( y \geq 1 \) and find \( E[Z] \).

Solution: As in part c), \( E[Z | Y] = \frac{Y^2 + 1}{2} \). Since \( p_Y(y) \) approaches 0 as \( y^{-3} \), we see that \( E[Y^2] \) is infinite and thus \( E[Z] = \infty \).
Exercise 1.20: Consider a positive, integer-valued rv whose CDF is given at integer values by
\[ F_Y(y) = 1 - \frac{2}{(y+1)(y+2)} \quad \text{for integer } y \geq 0. \]

Use (1.34) to show that \( \mathbb{E}[Y] = 2 \). Hint: Note the PMF given in (1.29).

Solution: The complementary CDF of \( Y \) is \( 2/[(y + 1)(y + 2)] \) and is piecewise constant with changes only at integer values. Thus
\[
\mathbb{E}[Y] = \int_0^\infty F_Y^c(y) \, dy = \sum_{y=0}^{\infty} \frac{2}{(y+1)(y+2)}
\]
where we have substituted \( n \) for \( y + 1 \) and then used the fact from (1.29) that \( 1/[n(n+1)] \) for \( 1 \leq n \) is a PMF.

b) Find the PMF of \( Y \) and use it to check the value of \( \mathbb{E}[Y] \).

Solution: \( p_Y(y) = F_Y(y) - F_Y(y-1) \) for each integer \( y \geq 1 \). Thus
\[
p_Y(y) = \frac{2}{y(y+1)} - \frac{2}{(y+1)(y+2)} = \frac{4}{y(y+1)(y+2)}.
\]

\[
\mathbb{E}[Y] = \sum_{y=1}^{\infty} \frac{4y}{y(y+1)(y+2)} = \sum_{y=1}^{\infty} \frac{4}{(y+1)(y+2)}
\]
\[
= \frac{4}{n(n+1)} = \sum_{n=1}^{\infty} \frac{4}{n(n+1)} - 2 = 2.
\]

c) Let \( X \) be another positive integer-valued rv. Assume its conditional PMF is given by
\[ p_{X|Y}(x|y) = \frac{1}{y} \quad \text{for } 1 \leq x \leq y. \]

Find \( \mathbb{E}[X|Y = y] \) and show that \( \mathbb{E}[X] = 3/2 \). Explore finding \( p_X(x) \) until you are convinced that using the conditional expectation to calculate \( \mathbb{E}[X] \) is considerably easier than using \( p_X(x) \).

Solution: Recall that \( 1 + 2 + \cdots + y = y(y+1)/2 \). Thus
\[
\mathbb{E}[X|Y = y] = \frac{1 + 2 + \cdots + y}{y} = \frac{y + 1}{2}.
\]

\[
\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \sum_{y=1}^{\infty} p_Y(y) \frac{y + 1}{2}
\]
\[
= \frac{1}{2} \sum_{y=1}^{\infty} y p_Y(y) + \frac{1}{2} \sum_{y=1}^{\infty} p_Y(y) = \frac{\mathbb{E}[Y]}{2} + \frac{1}{2} = \frac{3}{2}.
\]
Calculating this expectation in the conventional way would require first calculating $p_X(x)$ and then calculating the expectation. Calculating $p_X(x)$,

$$p_X(x) = \sum_{y=x}^{\infty} p_Y(y)p_{X|Y}(x|y) = \sum_{y=x}^{\infty} \frac{4}{y(y+1)(y+2)} \frac{1}{y}$$

It might be possible to calculate this in closed form, but it certainly does not look attractive. Only a dedicated algebraic masochist would pursue this further given the other approach.

d) Let $Z$ be another integer-valued rv with the conditional PMF

$$p_{Z|Y}(z|y) = \frac{1}{y^2} \quad \text{for } 1 \leq z \leq y^2.$$ 

Find $E[Z|Y=y]$ for each integer $y \geq 1$ and find $E[Z]$.

**Solution:** Using the same argument as in part c),

$$E[Z|Y] = \frac{y^2(y^2+1)}{2y^2} = \frac{y^2+1}{2}.$$ 

$$E[Z] = \sum_{y \geq 1} p_Y(y)E[Z|Y=y] = \sum_{y \geq 1} \frac{2(y^2+1)}{y(y+1)(y+2)}.$$ 

These terms decrease with increasing $y$ as $2/y$ and thus the sum is infinite. To spell this out, the denominator term in the sum is $y^3 + 3y^2 + 2y \leq 5y^3$ for $y \geq 1$. Also the numerator is $2y^2 + 2 > 2y^2$. Thus each term exceeds $\frac{2}{5y^3}$, so the sum diverges as an harmonic series.

**Exercise 1.22:** Suppose $X$ has the Poisson PMF, $p_X(n) = \lambda^n \exp(-\lambda)/n!$ for $n \geq 0$ and $Y$ has the Poisson PMF, $p_Y(n) = \mu^n \exp(-\mu)/n!$ for $n \geq 0$. Assume that $X$ and $Y$ are independent. Find the distribution of $Z = X + Y$ and find the conditional distribution of $Y$ conditional on $Z = n$.

**Solution:** The seemingly straightforward approach is to take the discrete convolution of $X$ and $Y$ (i.e., the sum of the joint PMF’s of $X$ and $Y$ for which $X + Y$ has a given value $Z = n$). Thus

$$p_Z(n) = \sum_{k=0}^{n} p_X(k)p_Y(n-k) = \sum_{k=0}^{n} \frac{\lambda^k e^{-\lambda}}{k!} \times \frac{\mu^{n-k} e^{-\mu}}{(n-k)!}$$

$$= e^{-(\lambda+\mu)} \sum_{k=0}^{n} \frac{\lambda^k \mu^{n-k}}{k!(n-k)!}$$

At this point, one needs some added knowledge or luck. One might hypothesize (correctly) that $Z$ is also a Poisson rv with parameter $\lambda + \mu$; one might recognize the sum above, or one might look at an old solution. We multiply and divide the right hand expression above by $(\lambda + \mu)^n/n!$.

$$p_Z(n) = \frac{(\lambda + \mu)^n e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \left( \frac{\lambda}{\lambda + \mu} \right)^k \left( \frac{\mu}{\lambda + \mu} \right)^{n-k}$$

$$= \frac{(\lambda + \mu)^n e^{-(\lambda+\mu)}}{n!},$$
where we have recognized the sum on the right as a binomial sum.

Another approach that is actually more straightforward uses the fact (see (1.52)) that the MGF of the sum of independent rv’s is the product of the MGF’s of those rv’s. From Table 1.2 (or a simple derivation), \( g_X(r) = \exp \{ \lambda(e^r - 1) \} \). Similarly, \( g_Y(r) = \exp \mu(e^r - 1) \), so \( g_Z(r) = \exp \{ \lambda + \mu)(e^r - 1) \). Since the MGF specifies the PMF, \( Z \) is a Poisson rv with parameter \( \lambda + \mu \).

Finally, we must find \( p_{Y|Z}(i | n) \). As a prelude to using Bayes’ law, note that

\[
p_{Z|Y}(n | i) = \Pr \{ X + Y = n | Y = i \} = \Pr \{ X = n - i \}.\]

Thus

\[
p_{Y|Z}(i | n) = \frac{p_Y(i)p_X(n-i)}{p_{Z}(n)} = \frac{\lambda^{n-i}e^{-\lambda}}{(n-i)!} \times \frac{\mu^i e^{-\mu}}{i!} \times \frac{n!}{(\mu + \lambda)^n e^{-(\lambda + \mu)}}
= \binom{n}{i} \left( \frac{\lambda}{\lambda + \mu} \right)^{n-i} \left( \frac{\mu}{\lambda + \mu} \right)^i.
\]

Why this turns out to be a binomial PMF will be clarified when we study Poisson processes.

**Exercise 1.23:** a) Suppose \( X, Y \) and \( Z \) are binary rv’s, each taking on the value 0 with probability 1/2 and the value 1 with probability 1/2. Find a simple example in which \( X, Y, Z \) are statistically dependent but are pairwise statistically independent (i.e., \( X, Y \) are statistically independent, \( X, Z \) are statistically independent, and \( Y, Z \) are statistically independent). Give \( p_{XYZ}(x,y,z) \) for your example. Hint: In the simplest example, there are four joint values for \( x, y, z \) that have probability 1/4 each.

**Solution:** The simplest solution is also a very common relationship between 3 binary rv’s. The relationship is that \( X \) and \( Y \) are IID and \( Z = X \oplus Y \) where \( \oplus \) is modulo two addition, i.e., addition with the table \( 0 \oplus 0 = 1 \oplus 1 = 0 \) and \( 0 \oplus 1 = 1 \oplus 0 = 1 \). Since \( Z \) is a function of \( X \) and \( Y \), there are only 4 sample values, each of probability 1/4. The 4 possible sample values for \((XYZ)\) are then \((000)\), \((011)\), \((101)\) and \((110)\). It is seen from this that all pairs of \( X,Y,Z \) are statistically independent.

b) Is pairwise statistical independence enough to ensure that

\[
\mathbb{E} \left[ \prod_{i=1}^{n} X_i \right] = \prod_{i=1}^{n} \mathbb{E} [X_i].
\]

for a set of rv’s \( X_1, \ldots, X_n \)?

**Solution:** No, part a) gives an example, i.e., \( \mathbb{E} [XYZ] = 0 \) and \( \mathbb{E} [X] \mathbb{E} [Y] \mathbb{E} [Z] = 1/8 \).

**Exercise 1.24:** Show that \( \mathbb{E} [X] \) is the value of \( \alpha \) that minimizes \( \mathbb{E} [ (X - \alpha)^2 ] \).

**Solution:** \( \mathbb{E} [ (X - \alpha)^2 ] = \mathbb{E} [X^2] - 2\alpha \mathbb{E} [X] + \alpha^2 \). This is clearly minimized over \( \alpha \) by \( \alpha = \mathbb{E} [X] \).

**Exercise 1.25:** For each of the following random variables, find the endpoints \( r_- \) and \( r_+ \) of the interval for which the moment generating function \( g(r) \) exists. Determine in each case whether \( g_X(r) \) exists at \( r_- \) and \( r_+ \). For parts a) and b) you should also find and sketch \( g(r) \). For parts c) and d), \( g(r) \) has no closed form.
a) Let \( \lambda, \theta \), be positive numbers and let \( X \) have the density.

\[
f_X(x) = \begin{cases} 
\frac{1}{2} \exp(-\lambda x); & x \geq 0; \\
\frac{1}{2} \theta \exp(\theta x); & x < 0.
\end{cases}
\]

**Solution:** Integrating to find \( g_X(r) \) as a function of \( \lambda \) and \( \theta \), we get

\[
g_X(r) = \int_{-\infty}^{0} \frac{1}{2} \theta e^{\theta x + rx} \, dx = \frac{\theta}{2(\theta + r)} + \frac{\lambda}{2(\lambda - r)}.
\]

The first integral above converges for \( r > -\theta \) and the second for \( r < \lambda \). Thus \( r_- = -\theta \) and \( r_+ = \lambda \). The MGF does not exist at either end point.

b) Let \( Y \) be a Gaussian random variable with mean \( m \) and variance \( \sigma^2 \).

**Solution:** Calculating the MGF by completing the square in the exponent,

\[
g_Y(r) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( \frac{-(y - m)^2}{2\sigma^2} + ry \right) \, dy
\]

\[
= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( \frac{-(y - m - r\sigma)^2}{2\sigma^2} + 2rm + \frac{r^2\sigma^2}{2} \right) \, dy
\]

\[
= \exp \left( rm + \frac{r^2\sigma^2}{2} \right),
\]

where the final equality arises from realizing that the other terms in the equation above represent a Gaussian density and thus have unit integral. Note that this is the same as the result in Table 1.1. This MGF is finite for all \( r \) so \( r_- = -\infty \) and \( r_+ = \infty \). Also \( g_Y(r) \) is infinite at each endpoint.

c) Let \( Z \) be a nonnegative random variable with density

\[
f_Z(z) = k(1 + z)^{-2} \exp(-\lambda z); \quad z \geq 0.
\]

where \( \lambda > 0 \) and \( k = \left[ \int_{z \geq 0} (1 + z)^{-2} \exp(-\lambda z) \, dz \right]^{-1} \). Hint: Do not try to evaluate \( g_Z(r) \). Instead, investigate values of \( r \) for which the integral is finite and infinite.

**Solution:** Writing out the formula for \( g_Z(r) \), we have

\[
g_Z(r) = \int_{0}^{\infty} k(1 + z)^{-2} \exp\left( (r - \lambda)z \right) \, dz.
\]

This integral is clearly infinite for \( r > \lambda \) and clearly finite for \( r < \lambda \). For \( r = \lambda \), the exponential term disappears, and we note that \( (1 + z)^{-2} \) is bounded for \( z \leq 1 \) and goes to 0 as \( z^{-2} \) as \( z \to \infty \), so the integral is finite. Thus \( r_+ \) belongs to the region where \( g_Z(r) \) is finite.

The whole point of this is that the random variables for which \( r_+ = \lambda \) are those for which the density or PMF go to 0 with increasing \( z \) as \( e^{-\lambda z} \). Whether or not \( g_Z(\lambda) \) is finite depends on the coefficient of \( e^{-\lambda z} \).

d) For the \( Z \) of part c), find the limit of \( \gamma'(r) \) as \( r \) approaches \( \lambda \) from below. Then replace \( (1 + z)^2 \) with \(|1 + z|^3\) in the definition of \( f_Z(z) \) and \( K \) and show whether the above limit is then finite or not. Hint: no integration is required.
Solution: Differentiating $g_Z(r)$ with respect to $r$,

$$g_Z'(r) = \int_0^\infty k z (1 + z)^{-2} \exp\left((r - \lambda)z\right) dz.$$ 

As $z \to \infty$, this approaches 0 as $1/z$ and thus the integral does not converge. In other words, although $g_Z(\lambda)$ is finite, the slope of $g_Z(r)$ is unbounded as $r \to \lambda$ from below. If $(1 + z)^{-2}$ is replaced with $(1 + z)^{-3}$ (with $k$ modified to maintain a probability density), we see that as $z \to \infty$, $z(1 + z)^{-3}$ goes to 0 as $1/z^2$, so the integral converges. Thus in this case the slope of $g_Z(r)$ remains bounded for $r < \lambda$.

Exercise 1.26: a) Assume that the random variable $X$ has a moment generating function $g_X(r)$ that is finite in the interval $(r_-, r_+)$, $r_- < 0 < r_+$, and assume $r_- < r < r_+$ throughout. For any finite constant $c$, express the moment generating function of $X - c$, i.e., $g_{X-c}(r)$ in terms of the moment generating function of $X$. Show that $g'_{X-c}(r) \geq 0$.

Solution: Note that $g_{X-c}(r) = \mathbb{E}[\exp(r(X - c))]$. Thus the derivatives of $g_{X-c}(r)$ with respect to $r$ are

$$g'_{X-c}(r) = \mathbb{E}[(X - c) \exp(r(X - c))]$$

$$g''_{X-c}(r) = \mathbb{E}[(X - c)^2 \exp(r(X - c))].$$

Since the argument of the expectation above is nonnegative for all sample values of $x$, $g''_{X-c}(r) \geq 0$.

b) Show that $g''_{X-c}(r) = [g''_X(r) - 2c g'_X(r) + c^2 g_X(r)] e^{-rc}$.

Solution: Writing $(X - c)^2$ as $X^2 - 2cX + c^2$, we get

$$g''_{X-c}(r) = \mathbb{E}[X^2 \exp(r(X - c))] - 2c \mathbb{E}[X \exp(r(X - c))] + c^2 \mathbb{E}[\exp(r(X - c))]$$

$$= \left[\mathbb{E}[X^2 \exp(rX)] - 2c \mathbb{E}[X \exp(rX)] + c^2 \mathbb{E}[\exp(rX)]\right] \exp(-rc)$$

$$= \left[g''_X(r) - 2c g'_X(r) + g(r)\right] \exp(-rc).$$

c) Use a) and b) to show that $g''_X(r)g_X(r) - [g'_X(r)]^2 \geq 0$, and that $\gamma''_X(r) \geq 0$. Hint: Let $c = g'_X(r)/g_X(r)$.

Solution: With the suggested choice for $c$,

$$g''_{X-c}(r) = \left[g''_X(r) - 2\left(\frac{g'_X(r)}{g_X(r)}\right)^2 + \left(\frac{g'_X(r)}{g_X(r)}\right)^2\right] \exp(-rc)$$

$$= \left[\frac{g_X(r)g''_X(r) - [g'_X(r)]^2}{g_X(r)}\right] \exp(-rc).$$

Since this is nonnegative from part a), we see that

$$\gamma''_X(r) = g_X(r)g''_X(r) - [g'_X(r)]^2 \geq 0.$$ 

d) Assume that $X$ is non-atomic, i.e., that there is no value of $c$ such that $\Pr\{X = c\} = 1$. Show that the inequality sign “$\geq$” may be replaced by “$>$” everywhere in a), b) and c).
Solution: Since $X$ is non-atomic, $(X - c)$ must be non-zero with positive probability, and thus from part a), $g''_{X-c}(r) > 0$. Thus the inequalities in parts b) and c) are positive also.

Exercise 1.27: A computer system has $n$ users, each with a unique name and password. Due to a software error, the $n$ passwords are randomly permuted internally (i.e. each of the $n!$ possible permutations are equally likely. Only those users lucky enough to have had their passwords unchanged in the permutation are able to continue using the system.

a) What is the probability that a particular user, say user 1, is able to continue using the system?

Solution: The obvious (and perfectly correct) solution is that user 1 is equally likely to be assigned any one of the $n$ passwords in the permutation, and thus has probability $1/n$ of being able to continue using the system.

This might be somewhat unsatisfying since it does not reason directly from the property that all $n!$ permutations are equally likely. Thus a more direct argument is that the number of permutations starting with 1 is the number of permutations of the remaining $(n-1)$ users, i.e., $(n-1)!$. These are equally likely, each with probability $1/n!$, so again the answer is $1/n$.

b) What is the expected number of users able to continue using the system? Hint: Let $X_i$ be a rv with the value 1 if user $i$ can use the system and 0 otherwise.

Solution: We have just seen that $Pr\{X_i = 1\} = 1/n$, so $Pr\{X_i = 0\} = 1 - 1/n$. Thus $E[X_i] = 1/n$. The number of users who can continue to use the system is $\sum_{i=1}^n X_i$, so $E[S_n] = \sum_i E[X_i] = n/n = 1$. It is important to understand here that the $X_i$ are not independent, and that the expected value of a finite sum of rv’s is equal to the sum of the expected values whether or not the rv’s are independent or not.

Exercise 1.29: Let $Z$ be an integer-valued rv with the PMF $p_Z(n) = 1/k$ for $0 \leq n \leq k-1$. Find the mean, variance, and moment generating function of $Z$. Hint: An elegant way to do this is to let $U$ be a uniformly distributed continuous rv over $(0, 1]$ that is independent of $Z$. Then $U + Z$ is uniform over $(0, k]$. Use the known results about $U$ and $U + Z$ to find the mean, variance, and MGF for $Z$.

Solution: It is not particularly hard to find the mean, variance, and MGF of $Z$ directly, but the technique here is often useful in relating discrete rv’s to continuous rv’s. We have

$$E[Z] = E[U + Z] - E[U] = k/2 - 1/2 = (k - 1)/2$$
$$\sigma_Z^2 = \sigma_{U+Z}^2 - \sigma_U^2 = k^2/12 - 1/12 = (k^2 - 1)/12$$
$$g_Z(r) = g_{U+Z}(r)/g_U(r) = \frac{\exp(kr) - 1}{rk} \times \frac{r}{\exp(r) - 1}$$
$$= \frac{\exp(kr) - 1}{k[\exp(r) - 1]}$$

Exercise 1.31: (Alternate approach 2 to the Markov inequality) a) Minimize $E[Y]$ over all nonnegative rv’s such that $Pr\{Y \geq b\} = \beta$ for some given $b > 0$ and $0 < \beta < 1$. Hint: Use a graphical argument similar to that in Figure 1.7. What is the rv that achieves the minimum. Hint: It is binary.

Solution: Since $E[Y] = \int_0^\infty F^c(y) dy$, we can try to minimize $E[Y]$ by minimizing $F^c(y)$ at each $y$ subject to the constraints on $F^c(y)$. One constraint is that $F^c(y)$ is nonincreasing in
Solution: We have seen that \( E[Y] \geq \beta b = b \Pr\{Y \geq b\} \), which, on dividing by \( b \), is the Markov inequality. Thus, for any \( b > 0 \) and any \( \Pr\{Y \geq b\} \) there is a binary rv that meets the Markov bound with equality.

Exercise 1.34: We stressed the importance of the mean of a rv \( X \) in terms of its association with the sample average via the WLLN. Here we show that there is a form of WLLN for the median and for the entire CDF, say \( F_X(x) \) of \( X \) via sufficiently many independent sample values of \( X \).

a) For any given \( x \), let \( \mathbb{I}_j(x) \) be the indicator function of the event \( \{X_j \leq x\} \) where \( X_1, X_2, \ldots, X_j, \ldots \) are IID rv’s with the CDF \( F_X(x) \). State the WLLN for the IID rv’s \( \{\mathbb{I}_1(x), \mathbb{I}_2(x), \ldots\} \).

Solution: The mean value of \( \mathbb{I}_j(x) \) is \( F_X(x) \) and the variance (after a short calculation) is \( F_X(x)F_c(x) \). This is finite (and in fact at most 1/4), so Theorem 1.7.1 applies and

\[
\lim_{n \to \infty} \Pr\left\{ \left| \frac{1}{n} \sum_{j=1}^n \mathbb{I}_j(x) - F_X(x) \right| > \epsilon \right\} = 0 \quad \text{for all } x \text{ and } \epsilon > 0. \tag{A.6}
\]

This says that if we take \( n \) samples of \( X \) and use \( (1/n) \sum_{j=1}^n \mathbb{I}_j(x) \) to approximate the CDF \( F_X(x) \) at each \( x \), then the probability that the approximation error exceeds \( \epsilon \) at any given \( x \) approaches 0 with increasing \( n \).

b) Does the answer to part a) require \( X \) to have a mean or variance?

Solution: No. As pointed out in a), \( \mathbb{I}_j(x) \) has a mean and variance whether or not \( X \) does, so Theorem 1.7.1 applies.

c) Suggest a procedure for evaluating the median of \( X \) from the sample values of \( X_1, X_2, \ldots \). Assume that \( X \) is a continuous rv and that its PDF is positive in an open interval around the median. You need not be precise, but try to think the issue through carefully.

What you have seen here, without stating it precisely or proving it is that the median has a law of large numbers associated with it, saying that the sample median of \( n \) IID samples of a rv is close to the true median with high probability.

Solution: Let \( \alpha \) be the median and suppose for any given \( n \) we estimate the median by the median, say \( \hat{\alpha}_n \), of the CDF \( (1/n) \sum_{j=1}^n \mathbb{I}_j(y) \) as a function of \( y \). Let \( \delta > 0 \) be arbitrary. Note that if \( (1/n) \sum_{j=1}^n \mathbb{I}_j(\alpha - \delta) < .5 \), then \( \hat{\alpha}_n > \alpha - \delta \). Similarly, if \( (1/n) \sum_{j=1}^n \mathbb{I}_j(\alpha + \delta) > .5 \),
then $\hat{\alpha}_n < \alpha + \delta$. Thus,

$$\Pr\{|\hat{\alpha}_n - \alpha| \geq \delta\} \leq \Pr \left\{ \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_j (\alpha - \delta) \geq .5 \right\} + \Pr \left\{ \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_j (\alpha + \delta) \leq .5 \right\}.$$  

Because of the assumption of a nonzero density, There is some $\epsilon_1 > 0$ such that $F_X (\alpha - \delta) < .5 - \epsilon_1$ and some $\epsilon_2 > 0$ such that $F_X (\alpha - \delta) > .5 + \epsilon_1$. Thus,

$$\Pr\{|\hat{\alpha}_n - \alpha| \geq \delta\} \leq \Pr \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_j (\alpha - \delta) - F_X (\alpha - \delta) \right| > \epsilon_1 \right\} + \Pr \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_j (\alpha + \delta) - F_X (\alpha + \delta) \right| > \epsilon_2 \right\}.$$

From (A.6), the limit of this as $n \to \infty$ is 0, which is a WLLN for the median. With a great deal more fussing, the same result holds true without the assumption of a positive density if we allow $\alpha$ above to be any median in cases where the median is nonunique.

**Exercise 1.37:** Represent the MGF of a rv $X$ by

$$g_X (r) = \int_{-\infty}^{0} e^{rx} dF(x) + \int_{0}^{\infty} e^{rx} dF(x).$$

In each of the following parts, you are welcome to restrict $X$ to be either discrete or continuous.

a) Show that the first integral always exists (i.e., is finite) for $r \geq 0$ and that the second integral always exists for $r \leq 0$.

**Solution:** For $r \geq 0$ and $x \leq 0$, we have $0 \leq e^{rx} \leq 1$. Thus the first integral lies between 0 and 1 and is thus finite. Similarly, for $r \leq 0$ and $x \geq 0$, we have $0 \leq e^{rx} \leq 1$, so the second integral is finite in this range.

b) Show that if the second integral exists for a given $r_1 > 0$, then it also exists for all $r$ in the range $0 \leq r \leq r_1$.

**Solution:** For $x \geq 0$ (i.e., in the domain of the first integral), we have $e^{rx} \leq e^{r_1 x}$, so we also have $0 \leq \int_{0}^{\infty} e^{rx} dF(x) \leq \int_{0}^{\infty} e^{r_1 x} dF(x)$. Thus the second integral exists for all $r \in [0, r_1]$.

c) Show that if the first integral exists for a given $r_2 < 0$, then it also exists for all $r$ in the range $r_2 \leq r \leq 0$.

**Solution:** This follows in the same way as part b).

d) Show that the domain of $r$ over which $g_X (r)$ exists is an interval from some $r_+ \geq 0$ to some $r_- \leq 0$ (the interval might or might not include each endpoint, and either or both end point might be 0, $\infty$, or any point between).

**Solution:** Combining parts a) and b), we see that if the second integral exists for some $r_1 > 0$, then it exists for all $r \leq r_1$. Defining $r_+$ as the supremum of values of $r$ for which the second integral exists, we see that the second integral exists for all $r < r_+$. Since the second integral exists for $r = 0$, we see that $r_+ = 0$. 
Similarly, the first integral exists for all \( r > r_- \) where \( r_- \leq 0 \) is the infimum of \( r \) such that the first integral exists. From part a), both integrals then exist for \( r \in (r_-, r_+) \)

e) Find an example where \( r_+ = 1 \) and the MGF does not exist for \( r = 1 \). Find another example where \( r_+ = 1 \) and the MGF does exist for \( r = 1 \). Hint: Consider \( f_X(x) = e^{-x} \) for \( x \geq 0 \) and figure out how to modify it to \( f_Y(y) \) so that \( \int_0^\infty e^{ry}f_Y(y) \, dy < \infty \) but \( \int_0^\infty e^{x+y}f_Y(y) = \infty \) for all \( \epsilon > 0 \).

**Solution:** Note: The hint should not have been necessary here. Anyone trying to imagine why \( \int_0^\infty e^{rx}f(x) \, dx \) should converge for some \( r \) and not for others who doesn’t consider exponentials was up too late the previous night. We observe that the second integral, \( \int_0^\infty \exp(rx-x) \, dx \) converges for \( r < 1 \) and diverges for \( r \geq 1 \), so \( r_+ \) is not in the domain of convergence.

How do we modify \( e^{-x} \) for \( x \geq 0 \) to make \( \int_0^\infty \exp(r-1) \, dx \) converge for \( r = 1 \) but to diverge for \( r > 1 \)? We need a function that goes to 0 with increasing \( x \) just a little faster than \( e^{-x} \), but it has to go to zero more slowly than \( e^{-\left(1+\epsilon\right)x} \) for any \( \epsilon > 0 \) There are many choices, such as \( \exp(-x - \sqrt{x}) \) or \( \left(1 + x^2\right)^{-1} \exp-x \). Each of these requires a normalizing constant to make the function a probability density, but there is no need to calculate the constant.

**Exercise 1.38:** Let \( \{X_n; n \geq 1\} \) be a sequence of independent but not identically distributed rv’s. We say that the weak law of large numbers (WLLN) holds for this sequence if for all \( \epsilon > 0 \)

\[
\lim_{n \to \infty} \Pr \left\{ \left| \frac{S_n}{n} - \frac{E[S_n]}{n} \right| \geq \epsilon \right\} = 0 \quad \text{where} \quad S_n = X_1 + X_2 + \cdots + X_n.
\]

a) Show that the WLLN holds if there is some constant \( A \) such that \( \sigma^2_{X_n} \leq A \) for all \( n \).

**Solution:** From (1.41), the variance of \( S_n \) is given by \( \sigma^2_{S_n} = \sum_{i=1}^{n} \sigma^2_{X_i} \). Since \( \sigma^2_{X_i} \leq A \), we have \( \sigma^2_{S_n} \leq nA \) and thus \( \sigma^2_{S_n/n} \leq A/n \). Thus, from the Chebyshev inequality,

\[
\Pr \left\{ \left| \frac{S_n}{n} - \frac{E[S_n]}{n} \right| \geq \epsilon \right\} \leq \frac{A}{n \epsilon^2}
\]

Then, exactly as in the proof of the WLLN for IID rv’s, the quantity on the right approaches 0 as \( n \to \infty \) for all \( \epsilon > 0 \), completing the demonstration.

b) Suppose that \( \sigma^2_{X_n} \leq A n^{1-\alpha} \) for some \( \alpha, 0 < \alpha < 1 \) and for all \( n \geq 1 \). Show that the WLLN holds in this case.

**Solution:** This is similar to part (a), but it is considerably stronger, since it says that the variance of the \( X_i \) can be slowly increasing in \( i \), and if this increase is bounded as above, the WLLN still holds. Note that \( An^{1-\alpha} \) is increasing in \( n \), so \( \sigma^2_{X_i} \leq An^{1-\alpha} \). Thus

\[
\sigma^2_{S_n} = \sum_{i=1}^{n} \sigma^2_{X_i} \leq nAn^{1-\alpha} \leq An^{2-\alpha}
\]

\[
\sigma^2_{S_n/n} \leq An^{-\alpha}
\]

From the Chebyshev inequality,

\[
\Pr \left\{ \left| \frac{S_n}{n} - \frac{E[S_n]}{n} \right| \geq \epsilon \right\} \leq \frac{An^{-\alpha}}{\epsilon^2},
\]
which for any $\epsilon > 0$ approaches 0 as $n \to \infty$.

**Exercise 1.39:** Let $\{X_i; i \geq 1\}$ be IID binary rv’s. Let $\Pr\{X_i = 1\} = \delta$, $\Pr\{X_i = 0\} = 1 - \delta$. Let $S_n = X_1 + \cdots + X_n$. Let $m$ be an arbitrary but fixed positive integer. Think! then evaluate the following and explain your answers:

a) $\lim_{n \to \infty} \sum_{i: n\delta - m \leq i \leq n\delta + m} \Pr\{S_n = i\}$.

**Solution:** It is easier to reason about the problem if we restate the sum in the following way:

$$
\sum_{i: n\delta - m \leq i \leq n\delta + m} \Pr\{S_n = i\} = \Pr\{n\delta - m \leq S_n \leq n\delta + m\} = \Pr\{-m \leq S_n - n\bar{X} \leq m\} = \Pr\left\{\frac{-m}{\sigma/\sqrt{n}} \leq \frac{S_n - n\bar{X}}{\sigma/\sqrt{n}} \leq \frac{m}{\sigma/\sqrt{n}}\right\},
$$

where $\sigma$ is the standard deviation of $X$. Now in the limit $n \to \infty$, $(S_n - n\bar{X})/\sigma\sqrt{n}$ approaches a normalized Gaussian rv in distribution, i.e.,

$$
\lim_{n \to \infty} \Pr\left\{\frac{-m}{\sigma/\sqrt{n}} \leq \frac{S_n - n\bar{X}}{\sigma/\sqrt{n}} \leq \frac{m}{\sigma/\sqrt{n}}\right\} = \lim_{n \to \infty} \left[\Phi\left(\frac{m}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-m}{\sigma/\sqrt{n}}\right)\right] = 0.
$$

This can also be seen immediately from the binomial distribution as it approaches a discrete Gaussian distribution. We are looking only at essentially the central $2m$ terms of the binomial, and each of those terms goes to 0 as $1/\sqrt{n}$ with increasing $n$.

b) $\lim_{n \to \infty} \sum_{i: 0 \leq i \leq n\delta + m} \Pr\{S_n = i\}$.

**Solution:** Here all terms on lower side of the distribution are included and the upper side is bounded as in part a). Arguing in the same way as in part a), we see that

$$
\sum_{i: 0 \leq i \leq n\delta + m} \Pr\{S_n = i\} = \Pr\left\{\frac{S_n - n\bar{X}}{\sigma/\sqrt{n}} \leq \frac{m}{\sigma/\sqrt{n}}\right\}.
$$

In the limit, this is $\Phi(0) = 1/2$.

c) $\lim_{n \to \infty} \sum_{i: n(\delta - 1/m) \leq i \leq n(\delta + 1/m)} \Pr\{S_n = i\}$.

**Solution:** Here the number of terms included in the sum is increasing linearly with $n$, and the appropriate mechanism is the WLLN.

$$
\sum_{i: n(\delta - 1/m) \leq i \leq n(\delta + 1/m)} \Pr\{S_n = i\} = \Pr\left\{\frac{1}{m} \leq \frac{S_n - n\bar{X}}{n} \leq \frac{1}{m}\right\}.
$$

In the limit $n \to \infty$, this is 1 by the WLLN. The essence of this exercise has been to scale the random variables properly to go the limit. We have used the CLT and the WLLN, but one could guess the answers immediately by recognizing what part of the distribution is being looked at.
Exercise 1.40: The WLLN is said to hold for a zero-mean sequence \( \{ S_n; n \geq 1 \} \) if
\[
\lim_{n \to \infty} \Pr \left\{ \left| \frac{S_n}{n} \right| > \epsilon \right\} = 0 \quad \text{for every } \epsilon > 0.
\]
The CLT is said to hold for \( \{ S_n; n \geq 1 \} \) if for some \( \sigma > 0 \) and all \( z \in \mathbb{R} \),
\[
\lim_{n \to \infty} \Pr \left\{ \frac{S_n}{\sigma \sqrt{n}} \leq z \right\} = \Phi(z),
\]
where \( \Phi(z) \) is the normal CDF. Show that if the CLT holds, then the WLLN holds also. Note 1: If you hate \( \epsilon, \delta \) arguments, you will hate this. Note 2: It will probably ease the pain if you convert the WLLN statement to: For every \( \epsilon > 0, \delta > 0 \) there exists an \( n(\epsilon, \delta) \) such that for every \( n \geq n(\epsilon, \delta) \),
\[
\Pr \left\{ \frac{S_n}{n} < -\epsilon \right\} \leq \delta \quad \text{and} \quad \Pr \left\{ \frac{S_n}{n} > 1 - \epsilon \right\} \leq \delta. \tag{i}
\]
Solution: The alternate form above simply splits the two tails of the distribution and uses the definition of a limit. Note that if this statement is satisfied for one value of \( \delta > 0 \), it is satisfied for all larger \( \delta \). Thus we restrict attention to \( 0 < \delta < 1 \). For any such \( \delta \) and any \( \epsilon > 0 \), we now derive this alternate form from the CLT. To do this, choose \( z < 0 \) to satisfy \( \delta = 2\Phi(z) \). Note that this also satisfies \( \delta = 2(1 - \Phi(|z|)) \) (it might help to look at Figure 1.12 while following this).

From the CLT, using this particular \( z \), we see that since the limit is \( \Phi(z) \), there must be an \( n_1 \) large enough that
\[
\Pr \left\{ \frac{S_n}{\sigma \sqrt{n}} < z \right\} < 2\Phi(z) = \delta \quad \text{for } n \geq n_1.
\]
Rewriting this,
\[
\Pr \left\{ \frac{S_n}{n} < \frac{z\sigma}{\sqrt{n}} \right\} < \delta \quad \text{for } n \geq n_1.
\]
Let \( n(\epsilon, \delta) \) be any integer both greater than \( n_1 \) and such that \( z\sigma/\sqrt{n} < \epsilon \). Then the first part of (i) is satisfied. The second part is a minor variation.

Exercise 1.41: (Details in the proof of Theorem 1.7.4)
a) Show that if \( X_1, X_2, \ldots \) are IID, then the truncated versions \( \bar{X}_1, \bar{X}_2, \ldots \) are also IID.

Solution: Readers with an engineering orientation can be forgiven for saying this is obvious. For an actual proof, note that
\[
F_{\bar{X}_1, \ldots, \bar{X}_n}(y_1, \ldots, y_n) = \Pr \left\{ \bigcap_{j=1}^n \{ \bar{X}_j \leq y_j \} \right\}
\]
The rv \( \bar{X}_j \) is a function of the rv \( X_j \) and for each real number \( y_j \), we have
\[
\{ \bar{X}_j \leq y_j \} = \begin{cases} 
\emptyset & \text{for } y < \bar{X} - b \\
\{ X_j \leq y_j \} & \text{for } \bar{X} - b \leq y < \bar{X} + b \\
\Omega & \text{for } y \geq \bar{X} + b.
\end{cases}
\]
If all the \( y_j \) are in the range \([X-b, X+b]\), then each event \( \{\bar{X}_j \leq y_j\} \) is the same as \( \{X_j \leq y_j\} \), so it is clear that these \( n \) events are statistically independent. If some of the \( y_j \) are outside of this range, then the events are still independent since \( \emptyset \) and \( \Omega \) are independent of all other events, including themselves. Thus

\[
F_{\bar{X}_1 \ldots \bar{X}_n}(y_1, \ldots, y_n) = \prod_{j=1}^{n} \Pr\{\bar{X}_j \leq y_j\} = \prod_{j=1}^{n} F_{\bar{X}_j}(y_j)
\]

Finally, since the \( X_j \) are identically distributed, the \( \bar{X}_j \) are also.

b) Show that each \( \bar{X}_j \) has a finite mean \( E[\bar{X}] \) and finite variance \( \sigma^2_{\bar{X}} \). Show that the variance is upper bounded by the second moment around the original mean \( X \), i.e., show that \( \sigma^2_{\bar{X}} \leq E[(\bar{X} - E[X])^2] \).

**Solution:** Since \( X - b \leq \bar{X} \leq X + b \), we have

\[
\overline{X} - b \leq E[\bar{X}] \leq X + b
\]

so \( E[\bar{X}] \) is finite (and in the same way, \( E[|\bar{X}|] \) is finite).

Also, since \( (\bar{X})^2 \leq (|X| + b)^2 \), we have \( E[\overline{X}^2] \leq (|X| + b)^2 \) so the variance \( \sigma^2_{\bar{X}} \) is also finite.

Finally, for any rv \( Y \) with a variance, and any real \( \alpha \), \( E[(Y - \alpha)^2] = E[Y^2] - 2\alpha Y + \alpha^2 \), which is minimized by \( \alpha = \overline{Y} \), i.e., \( E[(Y - \alpha)^2] \geq \sigma^2_Y \). Letting \( Y \) be \( \bar{X} \) and \( \alpha \) be \( X \) shows that \( \sigma^2_{\bar{X}} \leq E[(\bar{X} - E[X])^2] \).

c) Assume that \( \bar{X}_j \) is \( X_j \) truncated to \( X \pm b \). Show that \( |\bar{X} - X| \leq b \) and that \( |\bar{X} - X| \leq |X - X| \). Use this to show that \( \sigma^2_{\bar{X}} \leq bE[|\bar{X} - X|] \leq 2bE[|X|] \).

**Solution:** Since \( X - b \leq \bar{X} \leq X + b \), we have \( -b \leq \bar{X} - X \leq b \), i.e., \( |\bar{X} - X| \leq b \). To show that \( |\bar{X} - X| \leq |X - X| \), we look at the two cases, \( |X - X| \leq b \), \( |X - X| > b \). In the first case, \( X = X \), so \( \bar{X} - X = X - X \). In the second case, \( |\bar{X} - X| = b \) so \( |\bar{X} - X| < |X - X| \).

\[
|\bar{X} - X|^2 \leq b|X - X|
\]

where we have used the upper bound \( b \) on the first \( |\bar{X} - X| \) above and the bound \( X - X \) on the other. Combining this with part b),

\[
\sigma^2_{\bar{X}} \leq E[|\bar{X} - X|^2] \leq bE[|X - X|] \leq bE[|X| + |X|] = 2bE[|X|]
\]

(A.7)

d) Let \( \bar{S}_n = \bar{X}_1 + \cdots + \bar{X}_n \) and show that for any \( \epsilon > 0 \),

\[
Pr\left\{ \left| \frac{\bar{S}_n}{n} - E[\bar{X}] \right| \geq \frac{\epsilon}{2} \right\} \leq \frac{8bE[|X|]}{n\epsilon^2}.
\]

(A.8)

**Solution:** Note that \( \bar{S}_n/n \) has variance \( \sigma^2_{\bar{X}}/n \), so the Chebyshev inequality says that

\[
Pr\left\{ \left| \frac{\bar{S}_n}{n} - E[\bar{X}] \right| \geq \frac{\epsilon}{2} \right\} \leq \frac{4\sigma^2_{\bar{X}}}{n\epsilon^2}.
\]
Substituting (A.7) in this gives us the desired bound.

e) Use Figure 1.13, along with (1.34), to show that for all sufficiently large \( b \), \( \left| E \left[ X - \bar{X} \right] \right| \leq \epsilon / 2 \). Use this to show that

\[
\Pr \left\{ \frac{S_n}{n} - E[X] \geq \epsilon \right\} \leq \frac{8\delta ||Z||}{n \epsilon^2} \quad \text{for all large enough } b.
\]

(A.9)

**Solution:** Note that \( \bar{X} \) and \( X \) have the same CDF from \( \bar{X} - b \) to \( \bar{X} + b \). From (1.34), then,

\[
E \left[ X - \bar{X} \right] = - \int_{-\infty}^{\bar{X} - b} F_X(x) \, dx + \int_{\bar{X} + b}^{\infty} F_X(x) \, dx
\]

Since \( X \) has a finite mean, \( \int_{0}^{\infty} F_X(x) \) is finite, so \( \lim_{b \to \infty} \int_{\bar{X} + b}^{\infty} F_X(x) \, dx = 0 \). This shows that \( E \left[ X - \bar{X} \right] \leq \epsilon / 2 \) for all sufficiently large \( b \). In the same way \( E \left[ X - \bar{X} \right] \geq -\epsilon / 2 \) for all sufficiently large \( b \) and thus \( \left| E \left[ X - \bar{X} \right] \right| \leq \epsilon / 2 \). Next, recall that for all numbers \( a, b \), we have \( |a - b| \leq |a| + |b| \). Thus,

\[
\left| \frac{S_n}{n} - \bar{X} \right| = \left| \frac{S_n}{n} - E[\bar{X}] + E[\bar{X}] - \bar{X} \right| \leq \left| \frac{S_n}{n} - E[\bar{X}] \right| + \left| E[\bar{X}] - \bar{X} \right|
\]

For all sample points in which the first term exceeds \( \epsilon \), one of the two final terms must exceed \( \epsilon / 2 \). Thus

\[
\Pr \left\{ \left| \frac{S_n}{n} - \bar{X} \right| \geq \epsilon \right\} \leq \Pr \left\{ \left| \frac{S_n}{n} - E[\bar{X}] \right| \geq \epsilon / 2 \right\} + \Pr \left\{ \left| E[\bar{X}] - \bar{X} \right| \geq \epsilon / 2 \right\}
\]

The second term on the right is 0 for all sufficiently large \( b \) and the first term is upper bounded in (A.8).

f) Use the following equation to justify (1.96).

\[
\Pr \left\{ \left| \frac{S_n}{n} - E[X] \right| > \epsilon \right\} = \Pr \left\{ \left| \frac{S_n}{n} - E[X] \right| > \epsilon \cap S_n = \bar{S}_n \right\} + \Pr \left\{ \left| \frac{S_n}{n} - E[X] \right| > \epsilon \cap S_n \neq \bar{S}_n \right\}
\]

**Solution:** The first term on the right above is unchanged if the first \( S_n \) is replaced by \( \bar{S}_n \), and therefore it is upper-bounded by (A.9). The second term is upper-bounded by \( \Pr \left\{ S_n \neq \bar{S}_n \right\} \), which is the probability that one or more of the \( X_j \) are outside of \( [\bar{X} - b, \bar{X} + b] \). Using the union bound results in (1.96).

**Exercise 1.43:** (MS convergence \( \implies \) convergence in probability) Assume that \( \{Z_n; n \geq 1\} \) is a sequence of r.v.’s and \( \alpha \) is a number with the property that \( \lim_{n \to \infty} E \left[ (Z_n - \alpha)^2 \right] = 0 \).

a) Let \( \epsilon > 0 \) be arbitrary and show that for each \( n \geq 0 \),

\[
\Pr \{ |Z_n - \alpha| \geq \epsilon \} \leq \frac{E \left[ (Z_n - \alpha)^2 \right]}{\epsilon^2}.
\]

(A.10)
Solution: Nitpickers will point out that it is possible for \( E[(Z_n - \alpha)^2] \) to be infinite for some \( n \), although because of the limit, it is finite for all large enough \( n \). We consider only such large enough \( n \) in what follows. The statement to be shown is almost the Chebyshev inequality, except that \( \alpha \) need not be \( E[Z_n] \). However, the desired statement follows immediately from applying the Markov inequality to \((Z_n - \alpha)^2\). This gives us a further reminder that it is the Markov inequality that is the more fundamental and useful between Markov and Chebyshev.

b) For the \( \epsilon \) above, let \( \delta > 0 \) be arbitrary. Show that there is an integer \( m \) such that \( E[(Z_n - \alpha)^2] \leq \epsilon^2 \delta \) for all \( n \geq m \).

Solution: By the definition of a limit, \( \lim_{n \to \infty} E[(Z_n - \alpha)^2] = 0 \) means that for all \( \epsilon_1 > 0 \), there is an \( m \) large enough that \( E[(Z_n - \alpha)^2] < \epsilon_1 \) for all \( n \geq m \). Choosing \( \epsilon_1 = \epsilon^2 \delta \) establishes the desired result.

c) Show that this implies convergence in probability.

Solution: Substituting \( E[(Z_n - \alpha)^2] \leq \epsilon^2 \delta \) into (A.10), we see that for all \( \epsilon, \delta > 0 \), there is an \( m \) such that \( \Pr\{|S_n/n - \alpha|^2 \geq \epsilon\} \leq \delta \) for all \( n \geq m \). This is convergence in probability

Exercise 1.45: Verify (1.57), i.e., verify that \( \lim_{y \to \infty} y \Pr\{Y \geq y\} = 0 \) if \( Y \geq 0 \) and \( Y < \infty \). Hint: Show that \( y \Pr\{Y \geq y\} \leq \int_{z \geq y} zdF_Y(z) \) and show that \( \lim_{y \to \infty} \int_{z \geq y} zdF_Y(z) = 0 \).

Solution: We can write \( \Pr\{Y \geq y\} \) as the Stieltjes integral \( \int_y^\infty dF_Y(z) \). Since \( z \geq y \) over the range of integration,

\[
y \Pr\{Y \geq y\} = \int_y^\infty y dF_Y(z) \leq \int_y^\infty zdF_Y(z)
\]

We can express \( Y \) as the Stieltjes integral

\[
\int_0^\infty zdF_Y(z) = \lim_{y \to \infty} \int_0^y zdF_Y(z)
\]

Since \( Y < \infty \), the limit above exists, so that \( \lim_{y \to \infty} \int_y^\infty zdF_Y(y) = 0 \). Combining these results, \( \lim_{y \to \infty} \int_{z \geq y} y \Pr\{Y \geq y\} = 0 \).

The result can also be proven without using Stieltjes integration. Since \( Y = \lim_{y \to \infty} \int_0^y F_Y^z(dy) \) and \( Y < \infty \), the above limit exists, so

\[
\lim_{y \to \infty} \int_y^\infty F_Y^z(z) dz = 0.
\]

For any \( \epsilon > 0 \), we can thereby choose \( y_0 \) such that

\[
\int_{y_0}^\infty F_Y^z(z) dz \leq \epsilon/2.
\]
For \( y > y_0 \), we can express \( y \Pr \{ Y \geq y \} \) as

\[
y \Pr \{ Y \geq y \} = (y - y_0) \Pr \{ Y \geq y \} + y_0 \Pr \{ Y \geq y \}
\]

\[
= \int_{y_0}^{y} \Pr \{ Y \geq y \} \, dz + y_0 \Pr \{ Y \geq y \} \\
\leq \int_{y_0}^{y} F_Y^c(z) \, dz + y_0 \Pr \{ Y \geq y \} \tag{A.12}
\]

The first term in (A.12) was bounded by using \( F_Y^c(z) \leq \Pr \{ Y \geq y \} \) for \( z < y \). From (A.11), the first term in (A.13) is upper bounded by \( \epsilon/2 \). Since \( \lim_{y \to \infty} \Pr \{ Y \geq y \} = 0 \), \( y_0 \Pr \{ Y \geq y \} \leq \epsilon/2 \) for all sufficiently large \( y \). Thus \( y \Pr \{ Y \geq y \} \leq \epsilon \) for all sufficiently large \( y \). Since \( \epsilon \) is arbitrary, \( \lim_{y \to \infty} y \Pr \{ Y \geq y \} = 0 \).

**Exercise 1.47:** Consider a discrete rv \( X \) with the PMF

\[
p_X(-1) = (1 - 10^{-10})/2, \\
p_X(1) = (1 - 10^{-10})/2, \\
p_X(10^{12}) = 10^{-10}.
\]

a) Find the mean and variance of \( X \). Assuming that \( \{X_m; m \geq 1\} \) is an IID sequence with the distribution of \( X \) and that \( S_n = X_1 + \cdots + X_n \) for each \( n \), find the mean and variance of \( S_n \). (no explanations needed.)

**Solution:** \( \bar{X} = 100 \) and \( \sigma_X^2 = 10^{14} + (1 - 10^{-10}) - 10^4 \approx 10^{14} \). Thus \( \overline{S}_n = 100n \) and \( \sigma_{S_n}^2 \approx n \times 10^{14} \).

b) Let \( n = 10^6 \) and describe the event \( \{S_n \leq 10^6\} \) in words. Find an exact expression for \( \Pr \{ S_n \leq 10^6 \} = F_{S_n}(10^6) \).

**Solution:** This is the event that all \( 10^6 \) trials result in \pm 1. That is, there are no occurrences of \( 10^{12} \). Thus \( \Pr \{ S_n \leq 10^6 \} = (1 - 10^{-10})^{10^6} \)

c) Find a way to use the union bound to get a simple upper bound and approximation of \( 1 - F_{S_n}(10^6) \).

**Solution:** From the union bound, the probability of one or more occurrences of the sample value \( 10^{12} \) out of \( 10^6 \) trials is bounded by a sum over \( 10^6 \) terms, each of which is \( 10^{-10} \), i.e., \( 1 - F_{S_n}(10^6) \leq 10^{-4} \). This is also a good approximation, since we can write

\[
(1 - 10^{-10})^{10^6} = \exp \left( 10^6 \ln(1 - 10^{-10}) \right) \approx \exp(10^6 \cdot 10^{-10}) \approx 1 - 10^{-4}.
\]

d) Sketch the CDF of \( S_n \) for \( n = 10^6 \). You can choose the horizontal axis for your sketch to go from \(-1\) to \(+1\) or from \(-3 \times 10^{3}\) to \(3 \times 10^{3}\) or from \(-10^{6}\) to \(10^{6}\) or from 0 to \(10^{12}\), whichever you think will best describe this CDF.

**Solution:** Conditional on no occurrences of \( 10^{12} \), \( S_n \) simply has a binomial distribution. We know from the central limit theorem for the binomial case that \( S_n \) will be approximately Gaussian with mean 0 and standard deviation \( 10^3 \). Since one or more occurrences of \( 10^{12} \) occur only with probability \( 10^{-4} \), this can be neglected in the sketch, so the CDF is approximately Gaussian with 3 sigma points at \( \pm 3 \times 10^3 \). There is a little blip out at \( 10^{12} \) which doesn’t show up well in the sketch, but of course could be important for some purposes.
e) Now let $n = 10^{10}$. Give an exact expression for $\Pr\{S_n \leq 10^{10}\}$ and show that this can be approximated by $e^{-1}$. Sketch the CDF of $S_n$ for $n = 10^{10}$, using a horizontal axis going from slightly below 0 to slightly more than $2 \times 10^{12}$. Hint: First view $S_n$ as conditioned on an appropriate rv.

**Solution:** First consider the PMF $p_B(j)$ of the number $B = j$ of occurrences of the value $10^{12}$. We have

$$p_B(j) = \binom{10^{10}}{j} p^j (1-p)^{10^{10}-j} \quad \text{where } p = 10^{-10}$$

- $p_B(0) = (1-p)^{10^{10}} = \exp\{10^{10} \ln(1-p)\} \approx \exp(-10^{10}p) = \exp(-1)$
- $p_B(1) = 10^{10}p(1-p)^{10^{10}-1} = (1-p)^{10^{10}-1} \approx \exp(-1)$
- $p_B(2) = \binom{10^{10}}{2} p^2 (1-p)^{10^{10}-2} \approx \frac{1}{2} \exp(-1)$.

Conditional on $B = j$, $S_n$ will be approximately Gaussian with mean $10^{12}j$ and standard deviation $10^5$. Thus $F_{S_n}(s)$ rises from 0 to $e^{-1}$ over a range from about $-3 \times 10^5$ to $+3 \times 10^5$. It then stays virtually constant up to about $10^{12} - 3 \times 10^5$. It rises to $2/e$ by $10^{12} + 3 \times 10^5$. It stays virtually constant up to $2 \times 10^{12} - 3 \times 10^5$ and rises to $2.5/e$ by $2 \times 10^{12} + 3 \times 10^5$. When we sketch this, it looks like a staircase function, rising from 0 to $1/e$ at 0, from $1/e$ to $2/e$ at $10^{12}$ and from $2/e$ to $2.5/e$ at $2 \times 10^{12}$. There are smaller steps at larger values, but they would not show up on the sketch.

d) Can you make a qualitative statement about how the distribution function of a rv $X$ affects the required size of $n$ before the WLLN and the CLT provide much of an indication about $S_n$.

**Solution:** It can be seen that for this peculiar rv, $S_n/n$ is not concentrated around its mean even for $n = 10^{10}$ and $S_n/\sqrt{n}$ does not look Gaussian even for $n = 10^{10}$. For this particular distribution, $n$ has to be so large that $B$, the number of occurrences of $10^{12}$, is large, and this requires $n >> 10^{10}$. This illustrates a common weakness of limit theorems. They say what happens as a parameter ($n$ in this case) becomes sufficiently large, but it takes extra work to see how large that is.

**Exercise 1.48:** Let $\{Y_n; n \geq 1\}$ be a sequence of rv’s and assume that $\lim_{n \to \infty} \mathbb{E}[\|Y_n\|] = 0$. Show that $\{Y_n; n \geq 1\}$ converges to 0 in probability. Hint 1: Look for the easy way. Hint 2: The easy way uses the Markov inequality.

**Solution:** Applying the Markov inequality to $|Y_n|$ for arbitrary $n$ and arbitrary $\epsilon > 0$, we
have
\[ \Pr\{|Y_n| \geq \epsilon\} \leq \frac{\mathbb{E}[|Y_n|]}{\epsilon}. \]

Thus going to the limit \( n \to \infty \) for the given \( \epsilon \),
\[ \lim_{n \to \infty} \Pr\{|Y_n| \geq \epsilon\} = 0. \]

Since this is true for every \( \epsilon > 0 \), this satisfies the definition for convergence to 0 in probability.
Exercise 2.1: a) Find the Erlang density \( f_{S_n}(t) \) by convolving \( f_X(x) = \lambda \exp(-\lambda x) \) with itself \( n \) times.

**Solution:** For \( n = 2 \), we convolve \( f_X(x) \) with itself.

\[
 f_{S_2}(t) = \int_0^t f_{X_1}(x) f_{X_2}(t-x) \, dx = \int_0^t \lambda e^{-\lambda x} \lambda e^{-(t-x)} \, dx = \lambda^2 t e^{-\lambda t}
\]

For larger \( n \), convolving \( f_X(x) \) with itself \( n \) times is found by taking the convolution \( n-1 \) times, \textit{i.e.}, \( f_{S_{n-1}}(t) \), and convolving this with \( f_X(x) \). Starting with \( n = 3 \),

\[
 f_{S_3}(t) = \int_0^t f_{S_2}(x) f_{X_3}(t-x) \, dx = \int_0^t \lambda^2 x e^{-\lambda x} \lambda e^{-(t-x)} \, dx = \frac{\lambda^3 t^2}{2} e^{-\lambda t}
\]

\[
 f_{S_4}(t) = \int_0^t \frac{\lambda^3 t^2}{2} e^{-\lambda x} \cdot \lambda e^{t-x} \, dx = \frac{\lambda^4 t^3}{3!} e^{-\lambda t}
\]

We now see the pattern; each additional integration increases the power of \( \lambda \) and \( t \) by 1 and multiplies the denominator by \( n - 1 \). Thus we hypothesize that \( f_{S_n}(t) = \frac{\lambda^n t^{n-1}}{n!} e^{-\lambda t} \).

If one merely wants to verify the well-known Erlang density, one can simply use induction from the beginning, but it is more satisfying, and not that much more difficult, to actually derive the Erlang density, as done above.

b) Find the moment generating function of \( X \) (or find the Laplace transform of \( f_X(x) \)), and use this to find the moment generating function (or Laplace transform) of \( S_n = X_1 + X_2 + \cdots + X_n \).

**Solution:** The formula for the MGF is almost trivial here,

\[
 g_X(r) = \int_0^\infty \lambda e^{-\lambda x} x^r \, dx = \frac{\lambda}{\lambda - r} \quad \text{for } r < \lambda.
\]

Since \( S_n \) is the sum of \( n \) IID rv’s,

\[
 g_{S_n}(r) = [g_X(r)]^n = \left( \frac{\lambda}{\lambda - r} \right)^n.
\]

c) Find the Erlang density by starting with (2.15) and then calculating the marginal density for \( S_n \).

**Solution:** To find the marginal density, \( f_{S_n}(s_n) \), we start with the joint density in (2.15) and integrate over the region of space where \( s_1 \leq s_2 \leq \cdots \leq s_n \). It is a peculiar integral, since the integrand is constant and we are just finding the volume of the \( n - 1 \) dimensional space in \( s_1, \ldots, s_{n-1} \) with the inequality constraints above. For \( n = 2 \) and \( n = 3 \), we have

\[
 f_{S_2}(s_2) = \lambda^2 e^{-\lambda s_2} \int_0^{s_2} ds_1 = \left( \lambda^2 e^{-\lambda s_2} \right) s_2
\]

\[
 f_{S_3}(s_3) = \lambda^3 e^{-\lambda s_3} \int_0^{s_3} \left[ \int_0^{s_2} ds_1 \right] ds_2 = \lambda^3 e^{-\lambda s_3} \int_0^{s_3} s_2 ds_2 = \left( \lambda^3 e^{-\lambda s_3} \right) \frac{s_3^2}{2}.
\]

The critical part of these calculations is the calculation of the volume, and we can do this inductively by guessing from the previous equation that the volume, given \( s_n \), of the \( n - 1 \) dimensional space where \( 0 < s_1 < \cdots < s_{n-1} < s_n \) is \( s_n^{n-1}/(n-1)! \). We can check that by

\[
 \int_0^{s_n} \left[ \int_0^{s_{n-1}} \cdots \int_0^{s_2} ds_1 \cdots ds_{n-2} \right] ds_{n-1} = \int_0^{s_n} \frac{s_n^{n-2}}{(n-2)!} ds_{n-1} = \frac{s_n^{n-1}}{(n-1)!}.
\]
This volume integral, multiplied by $\lambda^n e^{-\lambda s_n}$, is then the desired marginal density.

A more elegant and instructive way to calculate this volume is by first observing that the volume of the $n - 1$ dimensional cube, $s_n$ on a side, is $s_n^{n-1}$. Each point in this cube can be visualized as a vector $(s_1, s_2, \ldots, s_{n-1})$. Each component lies in $(0, s_n)$, but the cube doesn’t have the ordering constraint $s_1 < s_2, \ldots s_{n-1}$. By symmetry, the volume of points in the cube satisfying this ordering constraint is the same as the volume in which the components $s_1, \ldots s_{n-1}$ are ordered in any other particular way. There are $(n - 1)!$ different ways to order these $n - 1$ components (i.e., there are $(n - 1)!$ permutations of the components), and thus the volume with the ordering constraint, is $s_n^{n-1}/(n - 1)!$.

**Exercise 2.3:** The purpose of this exercise is to give an alternate derivation of the Poisson distribution for $N(t)$, the number of arrivals in a Poisson process up to time $t$. Let $\lambda$ be the rate of the process.

a) Find the conditional probability $\Pr\{N(t) = n \mid S_n = \tau\}$ for all $\tau \leq t$.

**Solution:** The condition $S_n = \tau$ means that the epoch of the $n$th arrival is $\tau$. Conditional on this, the event $N(t) = n$ for some $t > \tau$ means there have been no subsequent arrivals from $\tau$ to $t$. In other words, it means that the $(n+1)$th interarrival time, $X_{n+1}$ exceeds $t - \tau$. This interarrival time is independent of $S_n$ and thus

$$\Pr\{N(t) = n \mid S_n = \tau\} = \Pr\{X_{n+1} > t - \tau\} = e^{t-\tau} \quad \text{for } t > \tau. \quad (A.14)$$

b) Using the Erlang density for $S_n$, use (a) to find $\Pr\{N(t) = n\}$.

**Solution:** We find $\Pr\{N(t) = n\}$ simply by averaging (A.14) over $S_n$.

$$\Pr\{N(t)=n\} = \int_0^\infty \Pr\{N(t)=n \mid S_n=\tau\} f_{S_n}(\tau) d\tau = \int_0^t e^{-\lambda(t-\tau)} \frac{\lambda^n \tau^{n-1} e^{-\lambda \tau}}{(n - 1)!} d\tau = \frac{\lambda^n e^{-\lambda t}}{(n - 1)!} \int_0^t \tau^{n-1} d\tau = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

**Exercise 2.4:** Assume that a counting process $\{N(t); t > 0\}$ has the independent and stationary increment properties and satisfies (2.17) for all $t > 0$. Let $X_1$ be the epoch of the first arrival and $X_n$ be the interarrival time between the $n - 1$st and the $n$th arrival. Use only these assumptions in doing the following parts of this exercise.

a) Show that $\Pr\{X_1 > x\} = e^{-\lambda x}$.

**Solution:** The event $\{X_1 > x\}$ is the same as the event $\{N(x) = 0\}$. Thus, from (2.17), $\Pr\{X_1 > x\} = \Pr\{N(x) = 0\} = e^{-\lambda x}$.

b) Let $S_{n-1}$ be the epoch of the $n - 1$st arrival. Show that $\Pr\{X_n > x \mid S_{n-1} = \tau\} = e^{-\lambda x}$.

**Solution:** The conditioning event $\{S_{n-1} = \tau\}$ is somewhat messy to deal with in terms of the parameters of the counting process, so we start by solving an approximation of the desired result and then go to a limit as the approximation becomes increasingly close. Let $\delta > 0$ be a small positive number which we later allow to approach 0. We replace the event $\{S_{n-1} = \tau\}$ with $\{\tau - \delta < S_{n-1} \leq \tau\}$. Since the occurrence of two arrivals in a very small
interval is very unlikely, we also include the condition that \( S_{n-2} \leq \tau - \delta \) and \( S_n \geq \tau \). With this, the approximate conditioning event becomes

\[
\{S_{n-2} \leq \tau - \delta < S_{n-1} \leq \tau < S_n\} = \{N(\tau - \delta) = n-2, \tilde{N}(\tau - \delta, \tau) = 1\}
\]

Since we are irretrievably deep in approximations, we also replace the event \( \{X_n > x\} \) with \( \tilde{N}(\tau, \tau+x) = 0 \). Note that this approximation is exact for \( \delta = 0 \), since in that case \( S_{n-1} = \tau \), so \( X_n > x \) means that no arrivals occur in \( (\tau, \tau+x) \).

We can now solve this approximate problem precisely,

\[
\Pr\left\{ \tilde{N}(\tau, \tau+x) = 0 \mid N(\tau - \delta) = n-2, \tilde{N}(\tau - \delta, \tau) = 1 \right\}
= \Pr\{\tilde{N}(\tau, \tau+x) = 0\}
= e^{-\lambda x}
\]

In the first step, we used the independent increment property and in the second, the stationary increment property along with part (a).

In the limit \( \delta \to 0 \), the conditioning event becomes \( S_{n-1} = \tau \) and the conditioned event becomes \( X_n > x \). The argument is very convincing, and becomes more convincing the more one thinks about it. At the same time, it is somewhat unsatisfactory since both the conditioned and conditioning event are being approximated. One can easily upper and lower bound the probability that \( X_n > x \) for each \( \delta \) but the ‘proof’ does not quite follow from the axioms of probability without a great deal of extra unsignificant details. We leave these out since it is so much cleaner to simply start with a definition of Poisson processes that uses independent interarrival times.

c) For each \( n > 1 \), show that \( \Pr\{X_n > x\} = e^{-\lambda x} \) and that \( X_n \) is independent of \( S_{n-1} \).

Solution: We have seen that \( \Pr\{X_n > x \mid S_{n-1} = \tau\} = e^{-\lambda x} \). Since the value of this probability conditioned on \( S_{n-1} = \tau \) does not depend on \( \tau \), \( X_n \) must be independent of \( S_{n-1} \).

d) Argue that \( X_n \) is independent of \( X_1, X_2, \ldots X_{n-1} \).

Solution: Equivalently, we show that \( X_n \) is independent of \( \{S_1=s_1, S_2=s_2, \ldots, S_{n-1}=s_{n-1}\} \) for all choices of \( 0 < s_1 < s_2 < \cdots < s_{n-1} \). Using the same artifice as in part (b), this latter event is the same as the limit as \( \delta \to 0 \) of the event

\[
N(s_1-\delta)=0, \quad \tilde{N}(s_1-\delta, s_1)=1, \quad \tilde{N}(s_1, s_2-\delta)=0, \quad \tilde{N}(s_2-\delta, s_2)=1, \quad \ldots, \quad \tilde{N}(s_{n-1}-\delta, s_{n-1})=1
\]

From the independent increment property, the above event is independent of the event \( \tilde{N}(s_{n-1}, s_{n-1} + x) \) for each \( x > 0 \). As in part (b), this shows that \( X_n \) is independent of \( S_1, \ldots, S_{n-1} \) and thus of \( X_1, \ldots, X_{n-1} \).

The most interesting part of this entire exercise is that the Poisson CDF was used only to derive the fact that \( X_1 \) has an exponential CDF. In other words, we have shown quite a bit more than Definition 2 of a Poisson process. We have shown that if \( X_1 \) is exponential and the stationary and independent increment properties hold, then the process is Poisson.

Exercise 2.5: The point of this exercise is to show that the sequence of PMF’s for the counting process of a Bernoulli process does not specify the process. In other words, knowing that \( N(t) \) satisfies the binomial
distribution for all \( t \) does not mean that the process is Bernoulli. This helps us understand why the second definition of a Poisson process requires stationary and independent increments along with the Poisson distribution for \( N(t) \).

a) For a sequence of binary rv's \( Y_1, Y_2, Y_3, \ldots \), in which each rv is 0 or 1 with equal probability, find a joint distribution for \( Y_1, Y_2, Y_3 \) that satisfies the binomial distribution, \( p_{N(t)}(k) = \binom{t}{k} 2^{-k} \) for \( t = 1, 2, 3 \) and \( 0 \leq k \leq \), but for which \( Y_1, Y_2, Y_3 \) are not independent.

Your solution should contain four 3-tuples with probability 1/8 each, two 3-tuples with probability 1/4 each, and two 3-tuples with probability 0. Note that by making the subsequent arrivals IID and equiprobable, you have an example where \( N(t) \) is binomial for all \( t \) but the process is not Bernoulli. Hint: Use the binomial for \( t = 3 \) to find two 3-tuples that must have probability 1/8. Combine this with the binomial for \( t = 2 \) to find two other 3-tuples with probability 1/8. Finally look at the constraints imposed by the binomial distribution on the remaining four 3-tuples.

Solution: The 3-tuples 000 and 111 have probability 1/8 as the unique tuples for which \( N(3) = 0 \) and \( N(3) = 3 \) respectively. In the same way, \( N(2) = 0 \) only for \((Y_1, Y_2) = (0, 0)\), so \((0,0)\) has probability 1/4. Since \((0,0,0)\) has probability 1/8, it follows that \((0,0,1)\) has probability 1/8. In the same way, looking at \( N(2) = 2 \), we see that \((1,1,0)\) has probability 1/8.

The four remaining 3-tuples are illustrated below, with the constraints imposed by \( N(1) \) and \( N(2) \) on the left and those imposed by \( N(3) \) on the right.

\[
\begin{array}{cccc}
1/4 & 0 & 1 & 0 \\
0 & 1 & 1 & 1/4 \\
1/4 & 1 & 0 & 0 \\
1 & 0 & 1 & 1/4 \\
\end{array}
\]

It can be seen by inspection from the figure that if \((0,1,0)\) and \((1,0,1)\) each have probability 1/4, then the constraints are satisfied. There is one other solution, which is to choose \((0,1,1)\) and \((1,0,0)\) to each have probability 1/4.

b) Generalize part a) to the case where \( Y_1, Y_2, Y_3 \) satisfy \( \Pr\{Y_i = 1\} = q \) and \( \Pr\{Y_i = 0\} = 1 - q \). Assume \( q < 1/2 \) and find a joint distribution on \( Y_1, Y_2, Y_3 \) that satisfies the binomial distribution, but for which the 3-tuple \((0,1,1)\) has zero probability.

Solution: Arguing as in part a), we see that \( \Pr\{(0,0,0)\} = (1 - q)^3 \), \( \Pr\{(0,0,1)\} = (1 - q)^2 p \), \( \Pr\{(1,1,1)\} = q^3 \), and \( \Pr\{(1,1,0)\} = q^2 (1 - q) \). The remaining four 3-tuples are constrained as shown below.

\[
\begin{array}{cccc}
q(1 - q) & 0 & 1 & 0 \\
0 & 1 & 1 & 2q(1 - q)^2 \\
q(1 - q) & 1 & 0 & 0 \\
1 & 0 & 1 & 2q^2 (1 - q) \\
\end{array}
\]

If we set \( \Pr\{(0,1,1)\} = 0 \), then \( \Pr\{0,1,0\} = q(1 - q) \), \( \Pr\{1,0,1\} = 2q^2 (1 - q) \), and \( \Pr\{(1,0,0)\} = (1 - q) - 2q^2 (1 - q) = (1 - q)(1 - 2q) \). This satisfies all the binomial
c) More generally yet, view a joint PMF on binary $t$-tuples as a nonnegative vector in a $2^t$ dimensional vector space. Each binomial probability $p_{N(\tau)}(k) = \binom{\tau}{k} q^k (1 - q)^{\tau-k}$ constitutes a linear constraint on this vector. For each $\tau$, show that one of these constraints may be replaced by the constraint that the components of the vector sum to 1.

**Solution:** There are $2^t$ binary $n$-tuples and each has a probability, so the joint PMF can be viewed as a vector of $2^t$ numbers. The binomial probability $p_{N(\tau)}(k) = \binom{\tau}{k} q^k (1 - q)^{\tau-k}$ specifies the sum of the probabilities of the $n$-tuples in the event $N(\tau) = k$, and thus is a linear constraint on the joint PMF. Since $\sum_{k=0}^{t} \binom{\tau}{k} p^k q^{\tau-k} = 1$, one of these $\tau + 1$ constraints can be replaced by the constraint that the sum of all $2^t$ components of the PMF is 1.

d) Using part c), show that at most $(t + 1)t/2 + 1$ of the binomial constraints are linearly independent. Note that this means that the linear space of vectors satisfying these binomial constraints has dimension at least $2^t - (t + 1)t/2 - 1$. This linear space has dimension 1 for $t = 3$ explaining the results in parts a) and b). It has a rapidly increasing dimension for $t > 3$, suggesting that the binomial constraints are relatively ineffectual for constraining the joint PMF of a joint distribution. More work is required for the case of $t > 3$ because of all the inequality constraints, but it turns out that this large dimensionality remains.

**Solution:** We know that the sum of all the $2^t$ components of the PMF is 1, and we saw in part c) that for each integer $\tau$, $1 \leq t$, there are $\tau$ additional linear constraints on the PMF established by the binomial terms $N(\tau = k)$ for $0 \leq k \leq \tau$. Since $\sum_{\tau=1}^{t} \tau = (t + 1)t/2$, we see that there are $t(t + 1)/2$ independent linear constraints on the joint PMF imposed by the binomial terms, in addition to the overall constraint that the components sum to 1. Thus the dimensionality of the $2^t$ vectors satisfying these linear constraints is at least $2^t - 1 - (t + 1)t/2$.

**Exercise 2.6:** Let $h(x)$ be a positive function of a real variable that satisfies $h(x + t) = h(x) + h(t)$ and let $h(1) = c$.

a) Show that for integer $k > 0$, $h(k) = kc$.

**Solution:** We use induction. We know $h(1) = c$ and the inductive hypothesis is that $h(n) = nc$, which is satisfied for $n = 1$. We then have $h(n + 1) = h(n) + h(1) = nc + c = (n + 1)c$. Thus if the hypothesis is satisfied for $n$ it is also satisfied for $n + 1$, which verifies that it is satisfied for all positive integer $n$.

b) Show that for integer $j > 0$, $h(1/j) = c/j$.

**Solution:** Repeatedly adding $h(1/j)$ to itself, we get $h(2/j) = h(1/j) + h(1/j) = 2h(1/j)$, $h(3/j) = h(2/j) + h(1/j) = 3h(1/j)$ and so forth to $h(1) = h(1/j) = jh(1/j)$. Thus $h(1/j) = c/j$.

c) Show that for all integer $k, j$, $h(k/j) = ck/j$.

**Solution:** Since $h(1/j) = c/j$, for each positive integer $j$, we can use induction on positive integers $k$ for any given $j > 0$ to get $h(k/j) = ck/j$.

d) The above parts show that $h(x)$ is linear in positive rational numbers. For very picky mathematicians, this does not guarantee that $h(x)$ is linear in positive real numbers. Show that if $h(x)$ is also monotonic in $x$, then $h(x)$ is linear in $x > 0$. 

**Exercise 2.7**: Assume that a counting process \( \{N(t); t>0\} \) has the independent and stationary increment properties and, for all \( t > 0 \), satisfies
\[
\begin{align*}
\Pr\{\bar{N}(t, t + \delta) = 0\} &= 1 - \lambda\delta + o(\delta) \\
\Pr\{\bar{N}(t, t + \delta) = 1\} &= \lambda\delta + o(\delta) \\
\Pr\{\bar{N}(t, t + \delta) > 1\} &= o(\delta). 
\end{align*}
\] (A.15)

a) Let \( F_1^c(\tau) = \Pr\{N(\tau) = 0\} \) and show that \( \frac{dF_1^c(\tau)}{d\tau} = -\lambda F_1^c(\tau) \).

**Solution**: Note that \( F_1^c \) is the complementary CDF of \( X_1 \). Using the fundamental definition of a derivative,
\[
\frac{dF_1^c(\tau)}{d\tau} = \lim_{\delta \to 0} \frac{F_1^c(\tau + \delta) - F_1^c(\tau)}{\delta} = \lim_{\delta \to 0} \frac{\Pr\{N(\tau + \delta) = 0\} - \Pr\{N(\tau) = 0\}}{\delta} \\
= \lim_{\delta \to 0} \frac{\Pr\{N(\tau) = 0\} \left(\Pr\{\bar{N}(\tau + \delta) = 0\} - 1\right)}{\delta} \\
= \lim_{\delta \to 0} \frac{\Pr\{N(\tau) = 0\} \left(1 - \lambda\delta + o(\delta) - 1\right)}{\delta} \\
= \Pr\{N(\tau) = 0\} (-\lambda) = -\lambda F_1^c(\tau),
\] (A.16) (A.17)

where (A.16) resulted from the independent increment property and (A.17) resulted from (A.15).

b) Show that \( X_1 \), the time of the first arrival, is exponential with parameter \( \lambda \).

**Solution**: The complementary CDF of \( X_1 \) is \( F_1^c(\tau) \), which satisfies the first order linear differential equation in (a). The solution to that equation, with the boundary point \( F_1^c(0) = 1 \) is \( e^{-\lambda\tau} \), showing that \( X_1 \) is exponential.

c) Let \( F_n^c(\tau) = \Pr\{\bar{N}(t, t + \tau) = 0 \mid S_{n-1} = t\} \) and show that \( \frac{dF_n^c(\tau)}{d\tau} = -\lambda F_n^c(\tau) \).

**Solution**: We are to show that \( F_n^c(\tau) = e^{\lambda\tau} \), which is equivalent to the desired differential equation. This was shown in the solution to Exercise 2.4 part (b), which was based on \( X_1 \) being exponential, which we established here in part (a).

d) Argue that \( X_n \) is exponential with parameter \( \lambda \) and independent of earlier arrival times.

**Solution**: This was shown in the solution to Exercise 2.4 parts (c) and (d). In other words, definitions 2 and 3 of a Poisson process both follow from the assumptions that \( X_1 \) is exponential and that the stationary and independent increment properties hold.

**Exercise 2.8**: For a Poisson process, let \( t > 0 \) be arbitrary and let \( Z_1 \) be the duration of the interval from \( t \) until the next arrival after \( t \). Let \( Z_m \), for each \( m > 1 \), be the interarrival time from the epoch of the \((m-1)\)st arrival after \( t \) until the \( m\)th arrival after \( t \).

a) Given that \( N(t) = n \), explain why \( Z_m = X_{m+n} \) for \( m > 1 \) and \( Z_1 = X_{n+1} - t + S_n \).

**Solution**: Given \( N(t) = n \), the \( m\)th arrival after \( t \) for \( m \geq 1 \) must be the \((n+m)\)th arrival overall. Its arrival epoch is denoted as
\[
S_{n+m} = S_{n+m-1} + X_{n+m}. \tag{A.18}
\]
By definition for \( m > 1 \), \( Z_m \) is the interval from \( S_{n+m+1} \) (the time of the \((m-1)\)st arrival after \( t \)) to \( S_{n+m} \). Thus, from (A.18), \( Z_m = X_{n+m} \) for \( m > 1 \). For \( m = 1 \), \( Z_1 \) is the interval from \( t \) until the next arrival, i.e., \( Z_1 = S_{n+1} - t \). From (A.18) for \( m = 1 \), \( Z_1 = S_n + X_{n+1} - t \).

b) Conditional on \( N(t) = n \) and \( S_n = \tau \), show that \( Z_1, Z_2, \ldots \) are IID.

**Solution:** Conditional on \( N(t) = n \) and \( S_n = \tau \), \( Z_1 \) is simply a translation of \( X_{n+1} \), for \( m > 1 \), \( Z_m = X_{n+m} \). Thus \( Z_1, \ldots, Z_m \) are independent of each other. The fact that \( Z_1 \) is exponential with the same rate as each interarrival time is proved in Theorem 2.2.1, and thus, conditionally, \( Z_1, \ldots \), are also identically distributed.

c) Show that \( Z_1, Z_2, \ldots \) are IID.

**Solution:** We have shown that \( \{Z_m; m \geq 1\} \) are IID and exponential conditional on \( N(t) = n \) and \( S_n = \tau \). The joint distribution of \( Z_1, \ldots, Z_m \) is then given as a function of \( N(t) = n \) and \( S_n = \tau \). Since this function is constant in \( n \) and \( \tau \), the joint conditional distribution must be the same as the joint unconditional distribution, and therefore \( Z_1, \ldots, Z_m \) are IID for all \( m > 0 \).

**Exercise 2.9:** Consider a “shrinking Bernoulli” approximation \( N_t(m\delta) = Y_1 + \cdots + Y_m \) to a Poisson process as described in Subsection 2.2.5.

a) Show that
\[
\Pr\{N_t(m\delta) = n\} = \frac{m^n}{n^n} (\lambda\delta)^n (1 - \lambda\delta)^{m-n}.
\]

**Solution:** This is just the binomial PMF in (1.23)

b) Let \( t = m\delta \), and let \( t \) be fixed for the remainder of the exercise. Explain why
\[
\lim_{\delta \to 0} \Pr\{N_t(t) = n\} = \lim_{m \to \infty} \left( \frac{m^n}{n^n} \right) \left( \frac{\lambda t}{m} \right)^n \left( 1 - \frac{\lambda t}{m} \right)^{m-n},
\]
where the limit on the left is taken over values of \( \delta \) that divide \( t \).

**Solution:** This is the binomial PMF in (a) with \( \delta = t/m \)

c) Derive the following two equalities:
\[
\lim_{m \to \infty} \left( \frac{m^n}{n^n} \right) \frac{1}{m^n} = \frac{1}{n!}; \quad \text{and} \quad \lim_{m \to \infty} \left( 1 - \frac{\lambda t}{m} \right)^{m-n} = e^{-\lambda t}.
\]

**Solution:** Note that
\[
\binom{m}{n} = \frac{m!}{n!(m-n)!} = \frac{1}{n!} \prod_{i=0}^{n-1} (m - i)
\]
When this is divided by \( m^n \), each term in the product above is divided by \( m \), so
\[
\binom{m}{n} \frac{1}{m^n} = \frac{1}{n!} \prod_{i=0}^{n-1} \frac{(m - i)}{m} = \frac{1}{n!} \prod_{i=0}^{n-1} \left( 1 - \frac{i}{m} \right)
\]
(A.19)
Taking the limit as \( m \to \infty \), each of the \( n \) terms in the product approaches 1, so the limit is \( 1/n! \), verifying the first equality in (c). For the second,

\[
\left( 1 - \frac{\lambda t}{m} \right)^{m-n} = \exp \left[ (m-n) \ln \left( 1 - \frac{\lambda t}{m} \right) \right] = \exp \left[ (m-n) \left( -\frac{\lambda t}{m} + o(1/m) \right) \right]
\]

\[
= \exp \left[ -\lambda t + \frac{n\lambda t}{m} + (m-n)o(1/m) \right]
\]

In the second equality, we expanded \( \ln(1 - x) = -x + x^2/2 \cdots \). In the limit \( m \to \infty \), the final expression is \( \exp(-\lambda t) \), as was to be shown.

If one wishes to see how the limit in (A.19) is approached, we have

\[
\frac{1}{n!} \prod_{i=0}^{n-1} \left( 1 - \frac{i}{m} \right) = \frac{1}{n!} \exp \left( \sum_{i=1}^{n-1} \ln \left( 1 - \frac{i}{m} \right) \right) = \frac{1}{n!} \exp \left( -\frac{n(n-1)}{2m} + o(1/m) \right)
\]

d) Conclude from this that for every \( t \) and every \( n \), \( \lim_{\delta \to 0} \Pr \{ N_\delta(t) = n \} = \Pr \{ N(t) = n \} \) where \( \{ N(t); t > 0 \} \) is a Poisson process of rate \( \lambda \).

**Solution:** We simply substitute the results of part (c) into the expression in part (b), getting

\[
\lim_{\delta \to 0} \Pr \{ N_\delta(t) = n \} = \frac{(\lambda t)^n e^{-\lambda t}}{n!}
\]

**Exercise 2.10:** Let \( \{ N(t); t > 0 \} \) be a Poisson process of rate \( \lambda \).

a) Find the joint probability mass function (PMF) of \( N(t) \), \( N(t+s) \) for \( s > 0 \).

**Solution:** Note that \( N(t+s) \) is the number of arrivals in \( (0, t] \) plus the number in \( (t, t+s) \). In order to find the joint distribution of \( N(t) \) and \( N(t+s) \), it makes sense to express \( N(t+s) \) as \( N(t) + \tilde{N}(t, t+s) \) and to use the independent increment property to see that \( \tilde{N}(t, t+s) \) is independent of \( N(t) \). Thus for \( m > n \),

\[
p_{N(t)N(t+s)}(n, m) = \Pr \{ N(t) = n \} \Pr \left\{ \tilde{N}(t, t+s) = m-n \right\}
\]

\[
= \frac{(\lambda t)^n e^{-\lambda t}}{n!} \times \frac{(\lambda s)^{m-n} e^{-\lambda s}}{(m-n)!},
\]

where we have used the stationary increment property to see that \( \tilde{N}(t, t+s) \) has the same distribution as \( N(s) \). This solution can be simplified in various minor ways, of which the most interesting is

\[
p_{N(t)N(t+s)}(n, m) = \frac{(\lambda(t+s))^m e^{-\lambda s}}{m!} \times \left( \frac{t}{m} \right) \left( \frac{s}{t+s} \right)^{m-n},
\]

where the first term is \( p_{N(t+s)}(m) \) (the probability of \( m \) arrivals in \( (0, t+s) \)) and the second, conditional on the first, is the binomial probability that \( n \) of those \( m \) arrivals occur in \( (0, t) \).
b) Find $\mathbb{E} [N(t) \cdot N(t+s)]$ for $s > 0$.

**Solution:** Again expressing $N(t+s) = N(t) + \tilde{N}(t,t+s)$,

\[
\mathbb{E} [N(t) \cdot N(t+s)] = \mathbb{E} [N^2(t)] + \mathbb{E} [N(t)\tilde{N}(t,t+s)]
\]

\[
= \mathbb{E} [N^2(t)] + \mathbb{E} [N(t)]\mathbb{E} [N(s)]
\]

\[
= \lambda t + \lambda^2 t^2 + \lambda t s.
\]

In the final step, we have used the fact (from Table 1.2) that the mean of a Poisson rv with PMF $(\lambda t)^n \exp(-\lambda t)/n!$ is $\lambda t$ and the variance is also $\lambda t$ (thus making the second moment $\lambda t + (\lambda t)^2$). This mean and variance can also be derived by a simple calculation from the PMF. Finally this mean and variance can be seen most insightfully by looking at the limit of shrinking Bernoulli processes.

e) Find $\mathbb{E} \left[ \tilde{N}(t_1,t_3) \cdot \tilde{N}(t_2,t_4) \right]$ where $\tilde{N}(t, \tau)$ is the number of arrivals in $(t, \tau)$ and $t_1 < t_2 < t_3 < t_4$.

**Solution:** This is a straightforward generalization of what was done in part (b). We break up $\tilde{N}(t_1,t_3)$ as $\tilde{N}(t_1,t_2) + \tilde{N}(t_2,t_3)$ and break up $\tilde{N}(t_2,t_4)$ as $\tilde{N}(t_2,t_3) + \tilde{N}(t_3,t_4)$. The interval $(t_2, t_3)$ is shared. Thus

\[
\mathbb{E} \left[ \tilde{N}(t_1,t_3)\tilde{N}(t_2,t_4) \right] = \mathbb{E} \left[ \tilde{N}(t_1,t_2)\tilde{N}(t_2,t_4) \right] + \mathbb{E} \left[ \tilde{N}(t_2,t_3)\tilde{N}(t_2,t_4) \right] + \mathbb{E} \left( \tilde{N}(t_2,t_3)\tilde{N}(t_3,t_4) \right]
\]

\[
= \lambda^2 (t_2-t_1)(t_4-t_2) + \lambda^2 (t_3-t_2)^2 + \lambda (t_3-t_2) + \lambda^2 (t_3-t_2)(t_4-t_3)
\]

\[
= \lambda^2 (t_3-t_1)(t_4-t_2) + \lambda (t_3-t_2)
\]

**Exercise 2.12:** Starting from time 0, northbound buses arrive at 77 Mass. Ave according to a Poisson process of rate $\lambda$. Customers arrive according to an independent Poisson process of rate $\mu$. When a bus arrives, all waiting customers instantly enter the bus and subsequent customers wait for the next bus.

a) Find the PMF for the number of customers entering a bus (more specifically, for any given $m$, find the PMF for the number of customers entering the $m$th bus).

**Solution:** Since the customer arrival process and the bus arrival process are independent Poisson processes, the sum of the two counting processes is a Poisson counting process of rate $\lambda + \mu$. Each arrival for the combined process is a bus with probability $\lambda/(\lambda + \mu)$ and a customer with probability $\mu/(\lambda + \mu)$. The sequence of bus/customer choices is an IID sequence. Thus, starting immediately after bus $m-1$ (or at time 0 for $m=1$), the probability of $n$ customers in a row followed by a bus, for any $n \geq 0$, is $[\mu/(\lambda + \mu)]^n \lambda (\lambda + \mu)$. This is the probability that $n$ customers enter the $m$th bus, $i.e.$, defining $N_m$ as the number of customers entering the $m$th bus, the PMF of $N_m$ is

\[
p_{N_m}(n) = \left( \frac{\mu}{\lambda + \mu} \right)^n \frac{\lambda}{\lambda + \mu}
\]

(A.20)

b) Find the PMF for the number of customers entering the $m$th bus given that the interarrival interval between bus $m-1$ and bus $m$ is $x$.  

Solution: For any given interval of size $x$ (i.e., for the interval $(s, s+x]$ for any given $s$), the number of customer arrivals in that interval has a Poisson distribution of rate $\mu$. Since the customer arrival process is independent of the bus arrivals, this is also the distribution of customer arrivals between the arrival of bus $m - 1$ and the arrival of bus $m$ given that the interval $X_m$ between these bus arrivals is $x$. Thus letting $X_m$ be the interval between the arrivals of bus $m - 1$ and $m$,

$$p_{X_m\mid X_m}(n\mid x) = (\mu x)^n e^{-\mu x} / n!$$

c) Given that a bus arrives at time 10:30 PM, find the PMF for the number of customers entering the next bus.

Solution: First assume that for some given $m$, bus $m - 1$ arrives at 10:30 PM. The number of customers entering bus $m$ is still determined by the argument in part (a) and has the PMF in (A.20). In other words, $N_m$ is independent of the arrival time of bus $m - 1$. From the formula in (A.20), the PMF of the number entering a bus is also independent of $m$. Thus the desired PMF is that on the right side of (A.20).

d) Given that a bus arrives at 10:30 PM and no bus arrives between 10:30 and 11, find the PMF for the number of customers on the next bus.

Solution: Using the same reasoning as in part (b), the number of customer arrivals from 10:30 to 11 is a Poisson rv, say $N'$ with PMF $p_{N'}(n) = (\mu/2)^n e^{-\mu/2} / n!$ (we are measuring time in hours so that $\mu$ is the customer arrival rate per hour.) Since this is independent of bus arrivals, it is also the PMF of customer arrivals in (10:30 to 11) given no bus arrival in that interval.

The number of customers to enter the next bus is $N'$ plus the number of customers arriving between 11 and the next bus arrival. By the argument in part (a), $N''$ has the PMF in (A.20). Since $N'$ and $N''$ are independent, the PMF of $N' + N''$ (the number entering the next bus given this conditioning) is the convolution of the PMF’s of $N'$ and $N''$, i.e.,

$$p_{N'+N''}(n) = \sum_{k=0}^{n} \left( \frac{\mu}{\lambda + \mu} \right)^k \frac{\lambda}{\lambda + \mu} \frac{(\mu/2)^{n-k} e^{-\mu/2}}{(n-k)!}$$

This does not simplify in any nice way.

e) Find the PMF for the number of customers waiting at some given time, say 2:30 PM (assume that the processes started infinitely far in the past). Hint: think of what happens moving backward in time from 2:30 PM.

Solution: Let $\{Z_i; -\infty < i < \infty\}$ be the (doubly infinite) IID sequence of bus/customer choices where $Z_i = 0$ if the $i$th combined arrival is a bus and $Z_i = 1$ if it is a customer. Indexing this sequence so that $-1$ is the index of the most recent combined arrival before 2:30, we see that if $Z_{-1} = 0$, then no customers are waiting at 2:30. If $Z_{-1} = 1$ and $Z_{-2} = 0$, then one customer is waiting. In general, if $Z_{-n} = 0$ and $Z_{-m} = 1$ for $1 \leq m < n$, then $n$ customers are waiting. Since the $Z_i$ are IID, the PMF of the number $N_{\text{past}}$ waiting at 2:30 is

$$p_{N_{\text{past}}}(n) = \left( \frac{\mu}{\lambda + \mu} \right)^n \frac{\lambda}{\lambda + \mu}$$
This is intuitive in one way, i.e., the number of customers looking back toward the previous bus should be the same as the number of customers looking forward to the next bus since the bus/customer choices are IID. It is paradoxical in another way since if we visualize a sample path of the process, we see waiting customers gradually increasing until a bus arrival, then going to 0 and gradually increasing again, etc. It is then surprising that the number of customers at an arbitrary time is statistically the same as the number immediately before a bus arrival. This paradox is partly explained at the end of part (f) and fully explained in Chapter 5.

Mathematically inclined readers may also be concerned about the notion of ‘starting infinitely far in the past.’ A more precise way of looking at this is to start the Poisson process at time 0 (in accordance with the definition of a Poisson process). We can then find the PMF of the number waiting at time t and take the limit of this PMF as t → ∞. For very large t, the number M of combined arrivals before t is large with high probability. Given M = m, the geometric distribution above is truncated at m, which is a negligible correction for t large. This type of issue is handled more cleanly in Chapter 5.

f) Find the PMF for the number of customers getting on the next bus to arrive after 2:30. Hint: this is different from part a); look carefully at part e).

Solution: The number getting on the next bus after 2:30 is the sum of the number $N_p$ waiting at 2:30 and the number of future customer arrivals $N_t$ (found in part c)) until the next bus after 2:30. Note that $N_p$ and $N_t$ are IID. Convolving these PMF’s, we get

$$p_{N_p+N_t}(n) = \sum_{m=0}^{n} \left( \frac{\mu}{\lambda+\mu} \right)^m \frac{\lambda}{\lambda+\mu} \left( \frac{\mu}{\lambda+\mu} \right)^{n-m} \frac{\lambda}{\lambda+\mu}$$

$$= (n+1) \left( \frac{\mu}{\lambda+\mu} \right)^n \left( \frac{\lambda}{\lambda+\mu} \right)^2$$

This is very surprising. It says that the number of people getting on the first bus after 2:30 is the sum of two IID rv’s, each with the same distribution as the number to get on the mth bus. This is an example of the ‘paradox of residual life,’ which we discuss very informally here and then discuss carefully in Chapter 5.

Consider a very large interval of time $(0, t_o]$ over which a large number of bus arrivals occur. Then choose a random time instant $T$, uniformly distributed in $(0, t_o]$. Note that $T$ is more likely to occur within one of the larger bus interarrival intervals than within one of the smaller intervals, and thus, given the randomly chosen time instant $T$, the bus interarrival interval around that instant will tend to be larger than that from a given bus arrival, $m-1$ say, to the next bus arrival $m$. Since 2:30 is arbitrary, it is plausible that the interval around 2:30 behaves like that around $T$, making the result here also plausible.

g) Given that I arrive to wait for a bus at 2:30 PM, find the PMF for the number of customers getting on the next bus.

Solution: My arrival at 2:30 is in additions to the Poisson process of customers, and thus the number entering the next bus is $1 + N_p + N_t$. This has the sample value $n$ if $N_p + N_t$
has the sample value \( n - 1 \), so from (f),

\[
p_{1+N_p+N_1}(n) = n \left( \frac{\mu}{\lambda+\mu} \right)^{n-1} \left( \frac{\lambda}{\lambda+\mu} \right)^2
\]

Do not be discouraged if you made a number of errors in this exercise and if it still looks very strange. This is a first exposure to a difficult set of issues which will become clear in Chapter 5.

**Exercise 2.15:** Consider generalizing the bulk arrival process in Figure 2.5. Assume that the epochs at which arrivals occur form a Poisson process \( \{N(t); t > 0\} \) of rate \( \lambda \). At each arrival epoch, \( S_n \), the number of arrivals, \( Z_n \), satisfies \( \Pr\{Z_n=1\} = p \), \( \Pr\{Z_n=2\} = 1 - p \). The variables \( Z_n \) are IID.

a) Let \( \{N_1(t); t > 0\} \) be the counting process of the epochs at which single arrivals occur. Find the PMF of \( N_1(t) \) as a function of \( t \). Similarly, let \( \{N_2(t); t \geq 0\} \) be the counting process of the epochs at which double arrivals occur. Find the PMF of \( N_2(t) \) as a function of \( t \).

**Solution:** Since the process of arrival epochs is Poisson, and these epochs are split into single and double arrival epochs by an IID splitting, the process of single arrival epochs is Poisson with rate \( \lambda p \) and epochs with two arrivals per epoch is Poisson with rate \( \lambda(1-p) \). Thus, letting \( q = 1 - p \),

\[
p_{N_1(t)}(n) = \frac{(\lambda p)^n e^{-\lambda p}}{n!}; \quad p_{N_2(t)}(m) = \frac{(\lambda q)^m e^{-\lambda q}}{m!}
\]

b) Let \( \{N_B(t); t \geq 0\} \) be the counting process of the total number of arrivals. Give an expression for the PMF of \( N_B(t) \) as a function of \( t \).

**Solution:** Since there are two arrivals at each arrival epoch for the double arrivals, we have \( N_B(t) = N_1(t) + 2N_2(t) \), and as seen above \( N_1(t) \) and \( N_2(t) \) are independent. This can be done as a digital convolution of the PMF’s for \( N_1(t) \) and \( 2N_2(t) \), but this can be slightly confusing since \( 2N_2(t) \) is nonzero only for even integers. Thus we revert to the general approach, which also reminds you where convolution comes from and thus how it can be done in general (PDF’s, PMF’s, etc.)

\[
\Pr\{N_B(t) = n, N_2(t) = m\} = p_{N_1(t)}(n-2m)p_{N_2(t)}(m)
\]

The marginal PMF for \( N_B(t) \) is then given by

\[
p_{N_B(t)}(n) = \sum_{m=0}^{[m/2]} p_{N_1(t)}(n-2m)p_{N_2(t)}(m)
\]

\[
= \sum_{m=0}^{[m/2]} \frac{(\lambda pt)^{n-2m} e^{-\lambda pt}}{(n-2m)!} \frac{(\lambda qt)^m e^{-\lambda qt}}{m!} m!
\]

\[
= e^{-\lambda t} \sum_{m=0}^{[m/2]} \frac{(\lambda pt)^{n-2m}(\lambda qt)^m}{(n-2m)!m!}
\]
**Exercise 2.18:** Consider a counting process in which the rate is a rv $\Lambda$ with probability density $f_\Lambda(\lambda) = \alpha e^{-\alpha \lambda}$ for $\lambda > 0$. Conditional on a given sample value $\lambda$ for the rate, the counting process is a Poisson process of rate $\lambda$ (i.e., nature first chooses a sample value $\lambda$ and then generates a sample path of a Poisson process of that rate $\lambda$).

a) What is $\Pr\{N(t)=n \mid \Lambda=\lambda\}$, where $N(t)$ is the number of arrivals in the interval $(0,t]$ for some given $t > 0$?

**Solution:** Conditional on $\Lambda = \lambda$, $\{N(t); t > 0\}$ is a Poisson process, so

$$\Pr\{N(t)=n \mid \Lambda=\lambda\} = (\lambda t)^n e^{-\lambda t} / n!$$

b) Show that $\Pr\{N(t)=n\}$, the unconditional PMF for $N(t)$, is given by

$$\Pr\{N(t)=n\} = \frac{\alpha t^n}{(t+\alpha)^{n+1}}$$

**Solution:** The straightforward approach is to average the conditional distribution over $\lambda$,

$$\Pr\{N(t) = n\} = \int_0^\infty \frac{(\lambda t)^n e^{-\lambda t}}{n!} \alpha e^{-\lambda \alpha} d\lambda$$

$$= \frac{\alpha t^n}{(t+\alpha)^{n+1}} \int_0^\infty \frac{[\lambda(t+\alpha)]^n e^{-\lambda(t+\alpha)}}{n!} d\lambda$$

$$= \frac{\alpha t^n}{(t+\alpha)^{n+1}} \int_0^\infty \frac{x^n e^{-x}}{n!} dx = \frac{\alpha t^n}{(t+\alpha)^{n+1}}$$

where we changed the variable of integration form $\lambda$ to $x = \lambda(t+\alpha)$ and then recognized the integral as the integral of an Erlang density of order $n+1$ with unit rate.

The solution can also be written as $pq^n$ where $p = \alpha/(t+\alpha)$ and $q = t/(t+\alpha)$. This suggests a different interpretation for this result. $\Pr\{N(t)=n\}$ for a Poisson process (PP) of rate $\lambda$ is a function only of $\lambda t$ and $n$. The rv $N(t)$ for a PP of rate $\lambda$ thus has the same distribution as $N(\lambda t)$ for a PP of unit rate and thus $N(t)$ for a PP of variable rate $\Lambda$ has the same distribution as $N(\Lambda t)$ for a PP of rate 1.

Since $\Lambda t$ is an exponential rv of parameter $\alpha/t$, we see that $N(\Lambda t)$ is the number of arrivals of a PP of unit rate before the first arrival from an independent PP of rate $\alpha/t$. This unit rate PP and rate $\alpha/t$ PP are independent and the combined process has rate $1 + \alpha/t$. The event $\{N(\Lambda t) = n\}$ then has the probability of $n$ arrivals from the unit rate PP followed by one arrival from the $\alpha/t$ rate process, thus yielding the probability $q^n p$.

c) Find $f_\lambda(\lambda \mid N(t)=n)$, the density of $\lambda$ conditional on $N(t)=n$.

**Solution:** Using Bayes’ law with the answers in (a) and (b), we get

$$f_{\lambda|N(t)}(\lambda \mid n) = \frac{\lambda^n e^{-\lambda(\alpha+t)} (\alpha + t)^{n+1}}{n!}$$

This is an Erlang PDF of order $n+1$ and can be interpreted (after a little work) in the same way as part (b).

d) Find $E[\Lambda \mid N(t)=n]$ and interpret your result for very small $t$ with $n = 0$ and for very large $t$ with $n$ large.
Solution: Since $\Lambda$ conditional on $N(t) = n$ is Erlang, it is the sum of $n + 1$ IID rv’s, each of mean $1/(t + \alpha)$. Thus
\[
E[\Lambda | N(t)=n] = \frac{n + 1}{\alpha + t}
\]
For $N(t) = 0$ and $t < < \alpha$, this is close to $1/\alpha$, which is $E[\Lambda]$. This is not surprising since it has little effect on the distribution of $\Lambda$. For $n$ large and $t >> \alpha$, $E[\Lambda | N(t)=n] \approx n/t$.

e) Find $E[\Lambda | N(t)=n, S_1, S_2, \ldots, S_n]$. (Hint: consider the distribution of $S_1, \ldots, S_n$ conditional on $N(t)$ and $\Lambda$). Find $E[\Lambda | N(t)=n, N(\tau)=m]$ for some $\tau < t$.

Solution: From Theorem 2.5.1, $S_1, \ldots, S_n$ are uniformly distributed, subject to $0 < S_1 < \cdots < S_n < t$, given $N(t) = n$ and $\Lambda = \lambda$. Thus, conditional on $N(t) = n$, $\Lambda$ is statistically independent of $S_1, \ldots, S_n$.
\[
E[\Lambda | N(t)=n, S_1, S_2, \ldots, S_n] = E[\Lambda | N(t)=n] = \frac{n + 1}{\alpha + t}
\]
Conditional on $N(t) = n$, $N(\tau)$ is determined by $S_1, \ldots, S_n$ for $\tau < t$, and thus
\[
E[\Lambda | N(t)=n, N(\tau) = m] = E[\Lambda | N(t)=n] = \frac{n + 1}{\alpha + t}
\]
This corresponds to one’s intuition; given the number of arrivals in $(0, t]$, it makes no difference where the individual arrivals occur.

Exercise 2.23: Let $\{N_1(t); t > 0\}$ be a Poisson counting process of rate $\lambda$. Assume that the arrivals from this process are switched on and off by arrivals from a second independent Poisson process $\{N_2(t); t > 0\}$ of rate $\gamma$.

rate $\lambda$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $N_1(t)$

rate $\gamma$ \quad $\leftarrow$On \quad $\times$ \quad $\leftarrow$On \quad $\times$ \quad $\leftarrow$On \quad $\times$ \quad $N_2(t)$

$\leftarrow$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $\times$ \quad $N_A(t)$

Let $\{N_A(t); t \geq 0\}$ be the switched process; that is $N_A(t)$ includes the arrivals from $\{N_1(t); t > 0\}$ during periods when $N_2(t)$ is even and excludes the arrivals from $\{N_1(t); t > 0\}$ while $N_2(t)$ is odd.

a) Find the PMF for the number of arrivals of the first process, $\{N_1(t); t > 0\}$, during the $n$th period when the switch is on.

Solution: We have seen that the combined process $\{N_1(t) + N_2(t)\}$ is a Poisson process of rate $\lambda + \gamma$. For any even numbered arrival to process 2, subsequent arrivals to the combined process independently come from process 1 or 2, and come from process 1 with probability $\lambda/(\lambda + \gamma)$. The number $N_s$ of such arrivals before the next arrival to process 2 is geometric with PMF $p_{N_s}(n) = [\lambda/(\lambda + \gamma)]^n[\gamma/(\lambda + \gamma)]$ for integer $k \geq 0$.

b) Given that the first arrival for the second process occurs at epoch $\tau$, find the conditional PMF for the number of arrivals $N_a$ of the first process up to $\tau$. 
Solution: Since processes 1 and 2 are independent, this is equal to the PMF for the number of arrivals of the first process up to \( \tau \). This number has a Poisson PMF, \((\lambda \tau)^n e^{-\lambda \tau}/n!\).

c) Given that the number of arrivals of the first process, up to the first arrival for the second process, is \( n \), find the density for the epoch of the first arrival from the second process.

Solution: Let \( N_a \) be the number of process 1 arrivals before the first process 2 arrival and let \( X_2 \) be the time of the first process 2 arrival. In part (a), we showed that \( p_{N_a}(n) = \left[ \lambda/(\lambda+\gamma) \right]^n \left[ \gamma/(\lambda+\gamma) \right] \) and in part (b) we showed that \( p_{N_a|X_2}(n|\tau) = (\lambda \tau)^n e^{-\lambda \tau}/n! \). We can then use Bayes’ law to find \( f_{X_2|N_a}(\tau | n) \), which is the desired solution. We have

\[
f_{X_2|N_a}(\tau | n) = f_X(\tau) \frac{p_{N_a|X_2}(n|\tau)}{p_{N_a}(n)} = \frac{(\lambda+\gamma)^n+1}{n!} \tau^n e^{-(\lambda+\gamma)\tau}
\]

where we have used the fact that \( X_2 \) is exponential with PDF \( \gamma \exp(-\gamma \tau) \) for \( \tau \geq 0 \). It can be seen that the solution is an Erlang rv of order \( n+1 \). To interpret this (and to solve the exercise in a perhaps more elegant way), note that this is the same as the Erlang density for the epoch of the \( (n+1) \)th arrival in the combined process. This arrival epoch is independent of the process 1/process 2 choices for these \( n+1 \) arrivals, and thus is the arrival epoch for the particular choice of \( n \) successive arrivals to process 1 followed by 1 arrival to process 2.

d) Find the density of the interarrival time for \( \{N_a(t); t \geq 0\} \). Note: This part is quite messy and is done most easily via Laplace transforms.

Solution: The process \( \{N_a(t); t > 0\} \) is not a Poisson process, but, perhaps surprisingly, it is a renewal process; that is, the interarrival times are independent and identically distributed. One might prefer to postpone trying to understand this until starting to study renewal processes, but we have the necessary machinery already.

Starting at a given arrival to \( \{N_a(t); t > 0\} \), let \( X_A \) be the interval until the next arrival to \( \{N_a(t); t > 0\} \) and let \( X \) be the interval until the next arrival to the combined process. Given that the next arrival in the combined process is from process 1, it will be an arrival to \( \{N_a(t); t > 0\} \), so that under this condition, \( X_A = X \). Alternatively, given that this next arrival is from process 2, \( X_A \) will be the sum of three independent rv’s, first \( X \), next, the interval \( X_2 \) to the following arrival for process 2, and next the interval from that point to the following arrival to \( \{N_a(t); t > 0\} \). This final interarrival time will have the same distribution as \( X_A \). Thus the unconditional PDF for \( X_A \) is given by

\[
f_{X_A}(x) = \frac{\lambda}{\lambda+\gamma} f_X(x) + \frac{\gamma}{\lambda+\gamma} f_X(x) \odot f_{X_2}(x) \odot f_{X_A}(x) \\
= \lambda \exp(-\lambda+\gamma)x + \gamma \exp(-\lambda+\gamma)x \odot \gamma \exp(-\gamma x) \odot f_{X_A}(x) f_{X_2}(x) \odot f_{X_A}(x)
\]

where \( \odot \) is the convolution operator and all functions are 0 for \( x < 0 \).

Solving this by Laplace transforms is a mechanical operation of no real interest here. The solution is

\[
f_{X_A}(x) = B \exp \left[ -x \left( 2\gamma + \lambda + \sqrt{4\gamma^2 + \lambda^2} \right) \right] + C \exp \left[ -x \left( 2\gamma + \lambda - \sqrt{4\gamma^2 + \lambda^2} \right) \right]
\]

where

\[
B = \frac{\lambda}{2} \left( 1 + \frac{\lambda}{\sqrt{4\gamma^2 + \lambda^2}} \right) \quad C = \frac{\lambda}{2} \left( 1 - \frac{\lambda}{\sqrt{4\gamma^2 + \lambda^2}} \right)
\]
Exercise 2.24: Let us model the chess tournament between Fisher and Spassky as a stochastic process. Let $X_i$, for $i \geq 1$, be the duration of the $i$th game and assume that $\{X_i; i \geq 1\}$ is a set of IID exponentially distributed rv’s each with density $f_X(x) = \lambda e^{-\lambda x}$. Suppose that each game (independently of all other games, and independently of the length of the games) is won by Fisher with probability $p$, by Spassky with probability $q$, and is a draw with probability $1 - p - q$. The first player to win $n$ games is defined to be the winner, but we consider the match up to the point of winning as being embedded in an unending sequence of games.

a) Find the distribution of time, from the beginning of the match, until the completion of the first game that is won (i.e., that is not a draw). Characterize the process of the number $\{N(t); t > 0\}$ of games won up to and including time $t$. Characterize the process of the number $\{N_F(t); t \geq 0\}$ of games won by Fisher and the number $\{N_S(t); t \geq 0\}$ won by Spassky.

Solution: The Poisson game process is split into 3 independent Poisson processes, namely the draw process of rate $(1 - p - q)\lambda$, the Fisher win process of rate $p\lambda$ and the Spassky win process of rate $q\lambda$. The process of wins is the sum of the Fisher and Spassky win processes, and is independent of the draw process. Thus the time to the first win is an exponential rv $X_w$ of rate $(p + q)\lambda$. Thus the density and CDF are

$$f_{X_w}(x) = (p + q)\lambda \exp(-(p + q)\lambda x), \quad F_{X_w}(x) = 1 - \exp(-(p + q)\lambda x)$$

b) For the remainder of the problem, assume that the probability of a draw is zero; i.e., that $p + q = 1$. How many of the first $2n - 1$ games must be won by Fisher in order to win the match?

Solution: Note that (a) shows that the process of wins is Poisson within the process of games including draws, and thus the assumption that there are no draws (i.e., $p + q = 1$) only simplifies the notation slightly. Note also that in assuming that $2n - 1$ games are played, we are also assuming that the play continues beyond where one or the other wins the match. Now, Fisher must win at least $n$ of the first $2n - 1$ games to win the match. To see this, note that if Fisher wins at least $n$ of the first $2n - 1$, then Spassky wins fewer than $n$ so Fisher wins his $n$th game before Spassky. Conversely if he wins fewer than $n$ of the first $2n - 1$, then Spassky wins more and wins his $n$th before Fisher.

c) What is the probability that Fisher wins the match? Your answer should not involve any integrals. Hint: consider the unending sequence of games and use part b).

Solution: The sequence of games is Bernoulli with probability $p$ of a Fisher game win. Thus, using the binomial PMF, the probability that Fisher wins the match is

$$\Pr\{\text{Fisher wins}\} = \sum_{k=n}^{2n-1} \binom{2n-1}{k} p^k q^{(2n-1) - k}$$

Without the hint, the problem is more tricky, but no harder computationally. The probability that Fisher wins the match at the end of game $k$, for $n \leq k \leq 2n - 1$, is the probability that he wins $n - 1$ games out the first $k - 1$ and then wins the $k$th. This is $p$ times $\binom{n-1}{k-1} p^{n-1} q^{k-1-1}$. Thus

$$\Pr\{\text{Fisher wins}\} = \sum_{k=n}^{2n-1} \binom{k-1}{n-1} p^n q^{k-n}$$
It is surprising that these very different appearing expressions are the same.

d) Let $T$ be the epoch at which the match is completed (i.e., either Fisher or Spassky wins). Find the CDF of $T$.

**Solution:** Let $T_f$ be the time at which Fisher wins his $n$th game and $T_s$ be the time at which Spassky wins his $n$th game (again assuming that playing continues beyond the winning of the match). The Poisson process of Fisher wins is independent of that of Spassky wins. Also, the time $T$ at which the match ends is the minimum of $T_f$ and $T_s$, so, for any $t > 0$, $\Pr\{T > t\} = \Pr\{T_f > t, T_s > t\}$. Thus

$$\Pr\{T > t\} = \Pr\{T_f > t\} \Pr\{T_s > t\} \quad (A.21)$$

Now $T_f$ has an Erlang distribution so its complementary CDF is equal to the probability that fewer than $n$ Fisher wins have occurred by time $t$. The number of Fisher wins is a Poisson rv, and Spassky wins are handled the same way. Thus,

$$\Pr\{T > t\} = \sum_{k=0}^{n-1} \frac{e^{-\lambda t} \lambda^k t^k}{k!} \sum_{j=0}^{n-1} \frac{e^{-\lambda q t} (\lambda q t)^j}{j!} = e^{-\lambda t} \sum_{k=0}^{n-1} \frac{e^{-\lambda q t} \lambda^k t^k}{k!} \sum_{j=0}^{n-1} \frac{e^{-\lambda q t} (\lambda q t)^j}{j!}$$

Finally $F_T(t) = 1 - \Pr\{T > t\}$.

e) Find the probability that Fisher wins and that $T$ lies in the interval $(t, t+\delta)$ for arbitrarily small $\delta$.

**Solution:** From (A.21), the PDF $f_T(t)$ is given by

$$f_T(t) = f_{T_f}(t)\Pr\{T_s > t\} + f_{T_s}(t)\Pr\{T_f > t\}.$$

The first term on the right is associated with a Fisher win, and the second with a Spassky win. Thus

$$\lim_{\delta \to 0} \frac{\Pr\{\text{Fisher wins, } T \in [t,t+\delta]\}}{\delta} = \frac{e^{-\lambda t} \lambda^k t^k (n-1)!}{(n-1)!} \sum_{j=0}^{n-1} \frac{(\lambda q t)^j e^{-\lambda q t}}{j!} \quad (A.22)$$

The main reason for calculating this is to point out that $T$ and the event that Fisher wins the match are not independent events. The number of games that Fisher wins up to time $t$ is independent of the number won by Spassky up to $t$. Also, given that $k$ games have been played by time $t$, the distribution of the number of those games won by Fisher does not vary with $t$. The problem here, given a Fisher win with $T = t$, the most likely number of Spassky wins is 0 when $t$ is very small and $n-1$ when $t$ is very large. This can be seen by comparing the terms in the sum of (A.22). This is not something you could reasonably be expected to hypothesize intuitively; it is simply something that indicates that one must sometimes be very careful.

**Exercise 2.28:** The purpose of this problem is to illustrate that for an arrival process with independent but not identically distributed interarrival intervals, $X_1, X_2, \ldots$, the number of arrivals $N(t)$ in the interval $(0, t]$ can be a defective rv. In other words, the ‘counting process’ is not a stochastic process according to our definitions. This illustrates that it is necessary to prove that the counting rv’s for a renewal process are actually rv’s.
a) Let the CDF of the $i$th interarrival interval for an arrival process be $F_{X_i}(x_i) = 1 - \exp(-\alpha^{-i}x_i)$ for some fixed $\alpha \in (0, 1)$. Let $S_n = X_1 + \cdots + X_n$ and show that

$$E[S_n] = \frac{\alpha(1 - \alpha^n)}{1 - \alpha}.$$

**Solution:** Each $X_i$ is an exponential rv, of rate $\alpha^i$, so $E[X_i] = \alpha^i$. Thus

$$E[S_n] = \alpha + \alpha^2 + \cdots \alpha^n.$$

Recalling that $1 + \alpha + \alpha^2 + \cdots = 1/(1 - \alpha)$,

$$E[S_n] = \frac{\alpha(1 + \alpha + \cdots + \alpha^{n-1})}{1 - \alpha} = \frac{\alpha(1 - \alpha^n)}{1 - \alpha} < \frac{\alpha}{1 - \alpha}.$$

In other words, not only is $E[X_i]$ decaying to 0 geometrically with increasing $i$, but $E[S_n]$ is upper bounded, for all $n$, by $\alpha/(1 - \alpha)$.

**b)** Sketch a ‘reasonable’ sample function for $N(t)$.

**Solution:** Since the expected interarrival times are decaying geometrically and the expected arrival epochs are bounded for all $n$, it is reasonable for a sample path to have the following shape:

```
0     S1     S2     S3  t
```

Note that the question here is not precise (there are obviously many sample paths, and which are ‘reasonable’ is a matter of interpretation). The reason for drawing such sketches is to acquire understanding to guide the solution to the following parts of the problem.

c) Find $\sigma_{S_n}^2$.

**Solution:** Since $X_i$ is exponential, $\sigma_{X_i}^2 = \alpha^{2i}$. Since the $X_i$ are independent,

$$\sigma_{S_n}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \cdots + \sigma_{X_n}^2 = \alpha^2 + \alpha^4 + \cdots + \alpha^{2n} = \alpha^2(1 + \alpha^2 + \cdots + \alpha^{2(n-1)}) = \frac{\alpha^2(1 - \alpha^{2n})}{1 - \alpha^2} < \frac{\alpha^2}{1 - \alpha^2}.$$

d) Use the Markov inequality on $\Pr\{S_n \geq t\}$ to find an upper bound on $\Pr\{N(t) \leq n\}$ that is smaller than 1 for all $n$ and for large enough $t$. Use this to show that $N(t)$ is defective for large enough $t$. 

Solution: The figure suggests (but does not prove) that for typical sample functions (and in particular for a set of sample functions of non-zero probability), \( N(t) \) goes to infinity for finite values of \( t \). If the probability that \( N(t) \leq n \) (for a given \( t \)) is bounded, independent of \( n \), by a number strictly less than 1, then that \( N(t) \) is a defective rv rather than a true rv.

By the Markov inequality,
\[
\Pr\{S_n \geq t\} \leq \frac{\mathbb{E}[S_n]}{t} \leq \frac{\alpha}{t(1-\alpha)}
\]
\[
\Pr\{N(t) < n\} = \Pr\{S_n > t\} \leq \Pr\{S_n \geq t\} \leq \frac{\alpha}{t(1-\alpha)}.
\]
where we have used (2.3). Since this is valid for all \( n \), it is also valid for \( \Pr\{N(t) \leq n\} \). For any \( t > \alpha/(1-\alpha) \), we see that \( \frac{\alpha}{t(1-\alpha)} < 1 \). Thus \( N(t) \) is defective for any such \( t \), i.e., for any \( t \) greater than \( \lim_{n \to \infty} \mathbb{E}[S_n] \).

Actually, by working a good deal harder, it can be shown that \( N(t) \) is defective for all \( t > 0 \). The outline of the argument is as follows: for any given \( t \), we choose an \( m \) such that \( \Pr\{S_m \leq t/2\} > 0 \) and such that \( \Pr\{S_{\infty} - S_m \leq t/2\} > 0 \) where \( S_{\infty} - S_m = \sum_{i=m+1}^{\infty} X_i \).

The second inequality can be satisfied for \( m \) large enough by the Markov inequality. The first inequality is then satisfied since \( S_m \) has a density that is positive for \( t > 0 \).

Exercise 3.6: a) Let \( Z \sim N(0, [K]) \) be \( n \)-dimensional. By expanding in a power series in \((1/2)r^T[K]r\), show that
\[
g_z(r) = \exp\left[ \frac{r^T[K]r}{2} \right] = 1 + \frac{\sum_{j,k} r_j r_k K_{j,k}}{2} + \cdots + \frac{\left(\sum_{j,k} r_j r_k K_{j,k}\right)^m}{2^m m!} + \cdots. \tag{A.23}
\]

Solution: This looks complicated, but is really quite simple. For each \( r = (r_1, \ldots, r_n)^T \), we know that \( r^T[K]r = \sum_{j,k} r_j r_k K_{j,k} \). For any given \( r \), this is simply a number, and we can use the series expansion of \( \exp x = 1 + x + x^2/2 + \cdots \) (which always converges) on that number divided by 2, getting the desired expression.

b) By expanding \( e^{r_j Z_j} \) in a power series in \( r_j Z_j \) for each \( j \), show that
\[
g_z(r) = \mathbb{E} \left[ \exp\left( \sum_j r_j Z_j \right) \right] = \sum_{\ell_1=0}^{\infty} \cdots \sum_{\ell_n=0}^{\infty} \frac{r_1^{\ell_1}}{(\ell_1)!} \cdots \frac{r_n^{\ell_n}}{(\ell_n)!} \mathbb{E}\left[ Z_1^{\ell_1} \cdots Z_n^{\ell_n} \right]. \tag{A.24}
\]

Solution: From the first part of (3.11),
\[
g_z(r) = \mathbb{E} \left[ \exp \sum_{j=0}^{n} r_j Z_j \right] = \mathbb{E} \left[ \prod_{j=0}^{n} \exp(r_j Z_j) \right]
\]
\[
= \mathbb{E} \left[ \prod_{j=0}^{n} \left( 1 + r_j Z_j + \frac{r_j^2 Z_j^2}{2} + \cdots + \frac{r_j^{\ell_j} Z_j^{\ell_j}}{\ell_j!} + \cdots \right) \right]
\]

The product multiplies into a sum of multinomials, and when we interchange the expectation with this sum, we get the desired expression. The interchange above is valid but we won’t justify it here.
c) Let $D = \{i_1, i_2, \ldots, i_{2m}\}$ be a set of $2m$ distinct integers each between 1 and $n$ and let $(j_1, \ldots, j_{2m})$ be a permutation of $D$. Consider the term $r_{j_1}r_{j_2} \cdots r_{j_{2m}} E[Z_{j_1}Z_{j_2} \cdots Z_{j_{2m}}]$ in part (b). By comparing with the set of terms in part (a) containing the same product $r_{j_1}r_{j_2} \cdots r_{j_{2m}}$, show that

$$E[Z_{i_1}Z_{i_2} \cdots Z_{i_{2m}}] = \frac{\sum_{j_1,j_2, \ldots, j_{2m}} K_{j_1j_2}K_{j_3j_4} \cdots K_{j_{2m-1}j_{2m}}}{2^m m!},$$  

(A.25)

where the sum is over all permutations $(j_1, j_2, \ldots, j_{2m})$ of $D$.

**Solution:** If each term of the form $\left( \sum_{j,k} r_j r_k K_{j,k} \right)^m$ in (A.23) is multiplied out and simplified by combining terms, the entire right hand side of (A.23) becomes a multinomial in $r_1, \ldots, r_n$. We know (although we have not proven it) that if the MGF of a single rv converges in a region containing 0 in the interior, then the MGF specifies the moments, and the moments collectively specify the MGF. The same situation occurs for random vectors, and this means that each multinomial term in (A.23) is equal to the corresponding term in (A.24).

We now must understand what is meant by simplifying (A.23) after multiplying it, and how this leads to (A.25). First, look at the term in (A.23) for $m = 1$.

$$\frac{1}{2} \sum_{j,k} r_j r_k K_{j,k} = \frac{1}{2} \sum_j r_j^2 K_{j,j} + \frac{1}{2} \sum_{j,k>j} 2r_j r_k K_{j,k},$$

Thus, for example, the multinomial term for $r_1 r_2$ in (A.23) is $\frac{1}{2} 2r_1 r_2 K_{1,2} = r_1 r_2 K_{1,2}$. This agrees with (A.24), where the $r_1r_2$ term is the term with $\ell_1 = \ell_2 = 0$ and $\ell_3 = \cdots = \ell_n = 0$.

Next consider the sum for the $m = 2$ term in (A.23).

$$\left( \sum_{j_1,k_1} r_{j_1} r_{k_1} K_{j_1,k_1} \right) \left( \sum_{j_2,k_2} r_{j_2} r_{k_2} K_{j_2,k_2} \right) = \sum_{j_1j_2,j_1k_2} r_{j_1} r_{k_1} r_{j_2} r_{k_2} K_{j_1,k_1}K_{j_2,k_2}$$  

(A.26)

Here the multinomial in this sum for, say, $r_1, r_2, r_3, r_4$ is the sum of all the terms in (A.26) for which $(j_1, j_2, j_2, k_2)$ is a permutation of $(1, 2, 3, 4)$. The corresponding term in (A.24) is that for which $\ell_1 = \cdots = \ell_4 = 1$ and $\ell_5 = \cdots = \ell_n = 0$.

The analysis is the same for all $m$. All the product terms for which $(j_1, j_1, j_2, \ldots, j_m)$ is a permutation of $D$ are added together into the multinomial term for $r_1, \ldots, r_{2m}$. Finally it is easy to see that the only product in (A.23) that gives rise to multinomial terms with $2m$ distinct components is the product of $m$ terms.

d) Find the number of permutations of $D$ that contain the same set of unordered pairs $(\{j_1, j_2\}, \ldots, \{j_{2m-1}, j_{2m}\})$. For example, $(\{1, 2\}, \{3, 4\})$ is the same set of unordered pairs as $(\{3, 4\}, \{2, 1\})$. Show that

$$E[Z_{j_1}Z_{j_2} \cdots Z_{j_{2m}}] = \sum_{j_1,j_2, \ldots, j_{2m}} K_{j_1j_2}K_{j_3j_4} \cdots K_{j_{2m-1}j_{2m}},$$  

(A.27)

where the sum is over distinct sets of unordered pairs of the set $D$. Note: another way to say the same thing is that the sum is over the set of all permutations of $D$ for which $j_{2k-1} < j_{2k}$ for $1 \leq k \leq m$ and $j_{2k-1} < j_{2k+1}$ for $1 \leq k \leq m - 1$.

**Solution:** There are 2 ways to order the elements within each pair, so $2^m$ ways to order the elements within all the $m$ pairs. There are then $m!$ ways to order the $m$ pairs. Thus
there are $2^m m!$ permutations in (A.26) that all correspond to the same set of unordered pairs. This cancels the denominator in (A.25), giving rise to (A.27).

ey) To find $E [Z_1^n \cdots Z_m^n]$, where $j_1 + j_2 + \cdots + j_n = 2m$, construct the random variables $U_1, \ldots, U_{2m}$, where $U_1, \ldots, U_{j_1}$ are all identically equal to $Z_1$, where $U_{j_1+1}, \ldots, U_{j_1+j_2}$ are identically equal to $Z_2$, etc., and use (A.27) to find $E [U_1 U_2 \cdots U_{2m}]$. Use this formula to find $E [Z_1^2 Z_2 Z_3]$, $E [Z_1^2 Z_2^2]$, and $E [Z_1^4]$.

**Solution:** All of the expected values to be found involve $E [U_1 U_2 \cdots U_{2m}]$ for $m = 2$. From (A.27),


Applying this to the 3 desired expectations,


$$E [Z_1^2 Z_2^2] = E [Z_1^2] E [Z_2^2] + 2(E [Z_1 Z_2])^2$$

$$E [Z_1^4] = 3(E [Z_1^2])^2$$

**Exercise 3.16:** a) Express $[B]$, $[C]$, and $[D]$, as defined in (3.39), in terms of $[K_X]$, $[K_Y]$ and $[K_{X,Y}]$ by multiplying the block expression for $[K]$ by that for $[K]^{-1}$. You can check your solutions against those in (3.46) to (3.48). Hint: You can solve for $[B]$ and $[C]$ by looking at only two of the four block equations in $[KK^{-1}]$. You can use the symmetry between $X$ and $Y$ to solve for $[D]$

**Solution:** One reason for going through this exercise, for those not very familiar with matrix manipulations, is to realize that algebraic manipulations on matrices are very similar to those on equations of numbers and real variables. One major difference is that matrices are not in general commutative (i.e., $AB \neq BA$ in many cases), and thus premultiplication and postmultiplication are different. Another is that invertibility involves much more than being non-zero. Recognizing this, we proceed with slightly guided plug and chug.

Multiplying out two of the block terms in $[KK^{-1}]$, we get $[K_X B] + [K_X Y C^T] = [I]$ and $[K_{X,Y} B] + [K_Y C^T] = 0$. These involve only two of the unknown terms, and we now solve for those terms. Recognizing that $[K_X]$ and $[K_Y]$ are invertible, we can rearrange the second equation as

$$[C^T] = -[K_Y^{-1} K_{X,Y} B]$$

Substituting this into the first equation, we get

$$[K_X] - [K_{X,Y} K_Y^{-1} K_{X,Y}^T] [B] = [I].$$

Now $[B]$ must be invertible since $[K^{-1}]$ is. Thus the matrix preceeding $[B]$ above is also invertible, so

$$[B] = \left([K_X] - [K_{X,Y} K_Y^{-1} K_{X,Y}^T]\right)^{-1}.$$ 

This agrees with the solution (derived very differently) in (3.46). Next, to solve for $[C]$, we take the transpose of $[C^T]$ above, leading to $[C] = -[BK_{X,Y} K_Y^{-1}]$. This agrees with (3.47).
We could solve for $[D]$ in the same way, but it is easier to use the symmetry and simply interchange the roles of $X$ and $Y$ to get (3.48).

b) Use your result in part a) for $[C]$ plus the symmetry between $X$ and $Y$ to show that

$$[BK_X Y K_Y^{-1}] = [K_X^{-1} K_X Y D].$$

**Solution:** The quantity on the left above is $-[C]$ as derived in part a). By using the symmetry between $X$ and $Y$, we see that $[DK_X^{-1} Y K_Y^{-1}]$ is $-C^T$, and taking the transpose completes the argument.

c) Show that $[K_Y^{-1} G] = [H^T K_Z^{-1}]$ for the formulations $X = [G] Y + V$ and $Y = [H] X + Z$ where $X$ and $Y$ are zero-mean, jointly Gaussian and have a non-singular combined covariance matrix. Hint: This is almost trivial from part b), (3.43), (3.44), and the symmetry.

**Solution:** From (3.43), $[K_Y] = [B^{-1}]$ and from (3.44), $[G] = [K_X Y K_Y]$. Substituting this into the left side of part b) and the symmetric relations for $X$ and $Y$ interchanged into the right side completes the demonstration.

**Exercise 4.1:** Let $[P]$ be the transition matrix for a finite state Markov chain and let state $i$ be recurrent. Prove that $i$ is aperiodic if $P_{ii} > 0$.

**Solution:** Since $P_{11} > 0$, there is a walk of length 2 from state 1 to 1 again, so $P_{11}^2 > 0$. In the same way, there is a walk of length $n$ for each $n > 1$, passing through state 1 $n$ times repeatedly. Thus $P_{11}^n > 0$ and the greatest common denominator of $\{n : P_{11}^n > 0\}$ is 1.

**Exercise 4.2:** Show that every Markov chain with $M < \infty$ states contains at least one recurrent set of states. Explaining each of the following statements is sufficient.

**Solution:**

a) If state $i_1$ is transient, then there is some other state $i_2$ such that $i_1 \rightarrow i_2$ and $i_2 \not\rightarrow i_1$.

**Solution:** If there is no such state $i_2$, then $i_1$ is recurrent by definition. We must have $i_1 \neq i_2$ since otherwise $i_1 \rightarrow i_2$ would imply $i_2 \rightarrow i_1$.

b) If the $i_2$ of part a) is also transient, there is a third state $i_3$ such that $i_2 \rightarrow i_3$, $i_3 \not\rightarrow i_2$; that state must satisfy $i_3 \neq i_2$, $i_3 \neq i_1$.

**Solution:** The argument why $i_3$ exists with $i_2 \rightarrow i_3$, $i_3 \not\rightarrow i_2$ and with $i_3 \neq i_2$ is the same as part a). Since $i_1 \rightarrow i_2$ and $i_2 \rightarrow i_3$, we have $i_1 \rightarrow i_3$. We must also have $i_3 \not\rightarrow i_1$, since otherwise $i_3 \rightarrow i_1$ and $i_1 \rightarrow i_2$ would imply the contradiction $i_3 \rightarrow i_2$. It follows that $i_3 \neq i_1$.

c) Continue iteratively to repeat part b) for successive states, $i_1, i_2, \ldots$. That is, if $i_1, \ldots, i_k$ are generated as above and are all transient, generate $i_{k+1}$ such that $i_k \rightarrow i_{k+1}$ and $i_{k+1} \not\rightarrow i_k$. Then $i_{k+1} \neq i_j$ for $1 \leq j \leq k$.

**Solution:** The argument why $i_{k+1}$ exists with $i_k \rightarrow i_{k+1}$, $i_{k+1} \not\rightarrow i_k$ and with $i_{k+1} \neq i_k$ is the same as before. To show that $i_{k+1} \neq i_j$ for each $j < k$, we use contradiction, noting that if $i_{k+1} = i_j$, then $i_{k+1} \rightarrow i_{j+1} \rightarrow i_k$.

d) Show that for some $k \leq M, k$ is not transient, i.e., it is recurrent, so a recurrent class exists.
Solution: For transient states $i_1, \ldots, i_k$ generated in part c), state $i_{k+1}$ found in part c) must be distinct from the distinct states $i_j, j \leq k$. Since there are only $M$ states, there cannot be $M$ transient states, since then, with $k = M$, a new distinct state $i_{M+1}$ would be generated, which is impossible. Thus there must be some $k < M$ for which the extension to $i_{k+1}$ leads to a recurrent state.

Exercise 4.3: Consider a finite-state Markov chain in which some given state, say state 1, is accessible from every other state. Show that the chain has at most one recurrent class $R$ of states and state $1 \in R$. (Note that, combined with Exercise 4.2, there is exactly one recurrent class and the chain is then a unichain.)

Solution: Since $j \rightarrow 1$ for each $j$, there can be no state $j$ for which $1 \rightarrow j$ and $j \neq 1$. Thus state 1 is recurrent. Next, for any given $j$, if $1 \not\sim j$, then $j$ must be transient since $j \rightarrow 1$. On the other hand, if $1 \sim j$, then 1 and $j$ communicate and $j$ must be in the same recurrent class as 1. Thus each state is either transient or in the same recurrent class as 1.

Exercise 4.4: Show how to generalize the graph in Figure 4.4 to an arbitrary number of states $M \geq 3$ with one cycle of $M$ nodes and one of $M - 1$ nodes. For $M = 4$, let node 1 be the node not in the cycle of $M - 1$ nodes. List the set of states accessible from node 1 in $n$ steps for each $n \leq 12$ and show that the bound in Theorem 4.2.4 is met with equality. Explain why the same result holds for all larger $M$.

Solution: As illustrated below, the generalization has the set of nonzero transition probabilities $P_{i,j}$ for $0 < i < M$ and has $P_{M,0} > 0$ and $P_{M,2} > 0$.

\[ \begin{array}{c}
M-1 \\
\vdots \\
4, \ldots, M-2 \\
3 \\
2 \\
1 \\
M
\end{array} \]

Let $T(n)$ be the set of states accessible from state 1 in $n$ steps.

\[
\begin{align*}
T(1) &= \{2\} & T(5) &= \{2, 3\} & T(9) &= \{2, 3, 4\} \\
T(2) &= \{3\} & T(6) &= \{3, 4\} & T(10) &= \{1, 2, 3, 4\} \\
T(3) &= \{4\} & T(7) &= \{1, 2, 4\} & T(11) &= \{1, 2, 3, 4\} \\
T(4) &= \{1, 2\} & T(8) &= \{1, 2, 3\} & T(12) &= \{1, 2, 3, 4\}
\end{align*}
\]

What we observe above is that for the first three steps, $T(n)$ is a singleton set, and for arbitrary $M$, $T(n)$ for the first $M - 1$ steps are singleton states. There are then $M - 1$ doubleton states, and so forth up to $M - 1$ steps with $M - 1$ states included in $T(n)$. Thus it is on the $(M - 1)^2 + 1$th step that all states are first included.

Exercise 4.5: (Proof of Theorem 4.2.4)

a) Show that an ergodic Markov chain with $M > 1$ states must contain a cycle with $\tau < M$ states. Hint: Use ergodicity to show that the smallest cycle cannot contain $M$ states.

Solution: The states in any cycle are distinct and thus a cycle contains at most $M$ states. An ergodic chain must contain cycles, since for each pair of states $\ell \neq j$, there is a walk
from $\ell$ to $j$ and then back to $\ell$; if any state $i$ other than $\ell$ is repeated in this walk, the first $i$ and all subsequent states before the second $i$ can be eliminated. This can be done repeatedly until a cycle remains.

Finally, suppose a cycle contains $M$ states. If there is any transition $P_{im} > 0$ for which $(i, m)$ is not a transition on that cycle, then that transition can be added to the cycle and all the transitions between $i$ and $m$ on the existing cycle can be omitted, thus creating a cycle with fewer than $M$ states. If there are no nonzero transitions other than those in the original cycle with $M$ states, then the Markov is periodic with period $M$ and thus not ergodic.

b) Let $\ell$ be a fixed state on a cycle of length $\tau < M$. Let $T(m)$ be the set of states accessible from $\ell$ in $m$ steps. Show that for each $m \geq 1$, $T(m) \subseteq T(m + \tau)$. Hint: For any given state $j \in T(m)$, show how to construct a walk of $m + \tau$ steps from $\ell$ to $j$ from the assumed walk of $m$ steps.

**Solution:** Let $j$ be any state in $T(m)$. Then there is an $m$-step walk from $\ell$ to $j$. There is also a cycle of length $\tau$ from state $\ell$ to $\ell$. Concatenate this cycle (as a walk) with the above $m$ step walk from $\ell$ to $j$, yielding a walk of length $\tau + m$ from $\ell$ to $j$. Thus $j \in T(m + \tau)$ and it follows that $T(m) \subseteq T(m + \tau)$.

c) Define $T(0)$ to be the singleton set $\{\ell\}$ and show that

$$T(0) \subseteq T(\tau) \subseteq T(2\tau) \subseteq \cdots \subseteq T(n\tau) \subseteq \cdots.$$  \hfill (i)

**Solution:** Since $T(0) = \{\ell\}$ and $\ell \in T(\tau)$, we see that $T(0) \subseteq T(\tau)$. Next, for each $n \geq 1$, use part b), with $m = n\tau$, to see that $T(n\tau) \subseteq T(n\tau + \tau)$. Thus each subset inequality above is satisfied.

d) Show that if one of the inclusions above is satisfied with equality, then all subsequent inclusions are satisfied with equality. Show from this that at most the first $M - 1$ inclusions can be satisfied with strict inequality and that $T(n\tau) = T((M - 1)\tau)$ for all $n \geq M - 1$.

**Solution:** We first show that if $T((n+1)\tau) = T(n\tau)$ for some $n$, then $T((m+1)\tau) = T(m\tau)$ for all $m > n$. Note that $T((n+1)\tau)$ is the set of states reached in $\tau$ steps from $T(n\tau)$ and $T((n+2)\tau)$ is the set of states reached in $\tau$ steps from $T((n+1)\tau)$. Thus if $T((n+1)\tau) = T(n\tau)$ then also $T((n+2)\tau) = T((n+1)\tau)$. Using induction,

$$T((n+m+1)\tau) = T((n+m)\tau) = T(n\tau) \quad \text{for all } m \geq 1.$$  

Now if $n$ is the smallest integer for which $T(n\tau) = T((n+1)\tau)$, then the size of $T(m\tau)$ must increase for each $m < n$. Since $|T(0)| = 1$, we see that $|T(m\tau)| \geq m + 1$ for $m \leq n$. Since $M$ is the total number of states, we see that $n \leq M - 1$. Thus $T(n\tau) = T((M - 1)\tau)$ for all $n \geq M - 1$.

e) Show that all states are included in $T((M - 1)\tau)$.

**Solution:** For any $t$ such that $P_{\ell\ell}(t) > 0$, we can substitute $t$ for $\tau$ in part (b) to see that for any $m \geq 1$, $T(m) \subseteq T(m + t)$. Thus we have

$$T((M-1)\tau) \subseteq T((M-1)\tau + t) \subseteq \cdots \subseteq T((M-1)\tau + tr) = T((M-1)\tau),$$

where part (d) was used in the final equality. This shows that all the inclusions above are satisfied with equality and thus that $T((M-1)\tau) = T((M-1)\tau + kt)$ for all $k \leq \tau$. Using
t in place of τ in the argument in part (d), this can be extended to
\[ T((M-1)\tau) = T((M-1)\tau + kt) \quad \text{for all } k \geq 1. \]
Since the chain is ergodic, we can choose t so that both \( P_{ij}^k > 0 \) and \( \gcd(t, \tau) = 1 \). From elementary number theory, integers \( k \geq 1 \) and \( j \geq 1 \) can then be chosen so that \( kt = j\tau + 1 \). Thus
\[ T((M-1)\tau) = T((M-1)\tau + kt) = T((M-1 + j)\tau + 1) = T((M-1)\tau + 1). \]
From the argument in part (d), it follows that
\[ T((M-1)\tau) = T((M-1)\tau + m) \quad \text{for all } m \geq 1. \]
This means that \( T((M-1)\tau) \) contains all states that can ever occur from time \((M-1)\tau\) on, and thus must contain all states since the chain is recurrent.

f) Show that \( P_{ij}^{(M-1)^2 + 1} > 0 \) for all \( i, j \).

**Solution:** We have shown that all states are accessible from state \( \ell \) at all times \( \tau(M-1) \) or later, and since \( \tau \leq M-1 \), all are accessible at times \((M-1)^2 \) or later. The same applies to any state on a cycle of length at most \( M-1 \). It is possible (as in Figure 4.4), for some states to be only on a cycle of length \( M \). Any such state can reach the cycle in the proof in at most \( M - \tau \) steps. Using this path to reach a state on the cycle and following this by paths of length \( \tau(M-1) \), all states can reach all other states at all times greater than or equal to
\[ \tau(M-1) + M - \tau \leq (M-1)^2 + 1. \]

**Exercise 4.8:** A transition probability matrix \([P]\) is said to be doubly stochastic if
\[ \sum_j P_{ij} = 1 \quad \text{for all } i; \sum_i P_{ij} = 1 \quad \text{for all } j. \]
That is, all row sums and all column sums each equal 1. If a doubly stochastic chain has \( M \) states and is ergodic (i.e., has a single class of states and is aperiodic), calculate its steady-state probabilities.

**Solution:** It is easy to see that if the row sums are all equal to 1, then \([P]e = e\). If the column sums are also equal to 1, then \( e^T[P] = e^T \). Thus \( e^T \) is a left eigenvector of \([P]\) with eigenvalue 1, and it is unique within a scale factor since the chain is ergodic. Scaling \( e^T \) to be the steady-state probabilities, \( \pi = (1/M, 1/M, \ldots, 1/M) \).

**Exercise 4.9:** a) Find the steady-state probabilities \( \pi_0, \ldots, \pi_{k-1} \) for the Markov chain below. Express your answer in terms of the ratio \( \rho = p/q \) where \( q = 1 - p \). Pay particular attention to the special case \( \rho = 1 \).

**Solution:** The steady-state equations, using the abbreviation \( q = 1 - p \) are
\[
\begin{align*}
\pi_0 &= q\pi_0 + q\pi_1 \\
\pi_j &= p\pi_{j-1} + q\pi_{j+1}; \quad \text{for } 1 \leq j \leq k-2 \\
\pi_{k-1} &= p\pi_{k-2} + p\pi_{k-1}
\end{align*}
\]
Simplifying the first equation, we get \( p\pi_0 = q\pi_1 \).

Substituting \( q\pi_1 \) for \( p\pi_0 \) in the second equation, we get \( \pi_1 = q\pi_1 + q\pi_2 \). Simplifying the second equation, then, we get \( p\pi_1 = q\pi_2 \).

We can then use induction. Using the inductive hypothesis \( p\pi_{j-1} = q\pi_j \) (which has been verified for \( j = 1, 2 \)) on \( \pi_j = p\pi_{j-1} + q\pi_{j+1} \), we get

\[
p\pi_j = q\pi_{j+1} \quad \text{for } 1 \leq j \leq k - 2
\]

Combining these equations, \( \pi_j = \rho\pi_{j-1} \) for \( 1 \leq j \leq k-1 \), so \( \pi_j = \rho^j\pi_0 \) for \( 1 \leq j \leq k-1 \). Normalizing by the fact that the steady-state probabilities sum to 1, and taking \( \rho \neq 1 \),

\[
\pi_0 \left( \sum_{j=0}^{k-1} \rho^j \right) = 1 \quad \text{so } \pi_0 = \frac{1 - \rho}{1 - \rho^k}; \quad \pi_j = \rho^j \frac{1 - \rho}{1 - \rho^k} \quad (A.28)
\]

For \( \rho = 1 \), \( \rho^j = 1 \) and \( \pi_j = 1/k \) for \( 0 \leq j \leq k - 1 \). Note that we did not use the final steady-state equation. The reason is that \( \pi = \pi[P] \) specifies \( \pi \) only within a scale factor, so the \( k \) equations can not be linearly independent and the \( k \)th equation can not provide anything new.

Note also that the general result here is mostly simply stated as \( p\pi_{j-1} = q\pi_j \). This says that the steady-state probability of a transition from \( j-1 \) to \( j \) is the same as that from \( j \) to \( j-1 \). This is intuitively obvious since in any sample path, the total number of transitions from \( j-1 \) to \( j \) is within 1 of the transitions in the opposite direction. This important intuitive notion will become precise after studying the strong law of numbers.

b) Sketch \( \pi_0, \ldots, \pi_{k-1} \). Give one sketch for \( \rho = 1/2 \), one for \( \rho = 1 \), and one for \( \rho = 2 \).

Solution: We see from (A.28) that for \( \rho \neq 1 \), \( \pi_j \) is geometrically distributed for \( 1 \leq j \leq k-1 \). For \( \rho > 1 \), it is geometrically increasing in \( j \) and for \( \rho < 1 \), it is geometrically decreasing. For \( \rho = 1 \), it is constant. If you can’t visualize this without a sketch, you should draw the sketch yourself.

c) Find the limit of \( \pi_0 \) as \( k \) approaches \( \infty \); give separate answers for \( \rho < 1 \), \( \rho = 1 \), and \( \rho > 1 \). Find limiting values of \( \pi_{k-1} \) for the same cases.

\[
\begin{array}{c}
& & 0 & p & 1 & p & 2 & p & \cdots & 0 & p & 1 & p & k & -1 & p \\
\end{array}
\]

Solution:

\[
\lim_{k \to \infty} \pi_0 = \begin{cases} 
\lim_{k \to \infty} \frac{1 - \rho}{1 - \rho^k} = 1 - \rho & \text{for } \rho < 1 \\
\lim_{k \to \infty} \frac{1}{k} = 0 & \text{for } \rho = 1 \\
\lim_{k \to \infty} \frac{\rho - 1}{\rho^k - 1} = 0 & \text{for } \rho > 1 
\end{cases}
\]

Note that the approach to 0 for \( \rho = 1 \) is harmonic in \( k \) and that for \( \rho > 1 \) is geometric in \( k \).

Exercise 4.14: Answer the following questions for the following stochastic matrix \([P]\)

\[
[P] = \begin{bmatrix} 
1/2 & 1/2 & 0 \\
0 & 1/2 & 1/2 \\
0 & 0 & 1 
\end{bmatrix}.
\]
a) Find \([P^n]\) in closed form for arbitrary \(n > 1\).

**Solution:** There are several approaches here. We first give the brute-force solution of simply multiplying \([P]\) by itself multiple times (which is reasonable for a first look), and then give the elegant solution.

\[
[P^2] = \begin{bmatrix}
0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
0.25 & 0.25 & 0.25 \\
0.25 & 0.25 & 0.25 \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
0.25 & 0.25 & 0.25 \\
0.25 & 0.25 & 0.25 \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & 1
\end{bmatrix}.
\]

\[
[P^3] = \begin{bmatrix}
0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
0.25 & 0.25 & 0.25 \\
0.25 & 0.25 & 0.25 \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
0.25 & 0.25 & 0.25 \\
0.25 & 0.25 & 0.25 \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{8} & \frac{3}{8} & \frac{5}{8} \\
\frac{1}{8} & \frac{3}{8} & \frac{5}{8} \\
0 & 0 & 1
\end{bmatrix}.
\]

We could proceed to \([P^4]\), but it is natural to stop and think whether this is telling us something. The bottom row of \([P^n]\) is clearly \((0, 0, 1)\) for all \(n\), and we can easily either reason or guess that the first two main diagonal elements are \(2^{-n}\). The final column is whatever is required to make the rows sum to 1. The only questionable element is \(P_{12}^n\). We guess that is \(n2^{-n}\) and verify it by induction,

\[
[P^{n+1}] = [P][P^n] = \begin{bmatrix}
0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
2^{-n} & n2^{-n} & 1 - (n+1)2^{-n} \\
0 & 2^{-n} & 1 - 2^{-n} \\
0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
2^{-n-1} & (n+1)2^{-n-1} & \frac{1}{2}(n+2)2^{-n-1} \\
0 & 2^{-n} & 1 - 2^{-n} \\
0 & 0 & 1
\end{bmatrix}.
\]

This solution is not very satisfying, first because it is tedious, second because it required a guess that was not very well motivated, and third because no clear rhyme or reason emerged.

The elegant solution, which can be solved with no equations, requires looking at the graph of the Markov chain,

\[
\begin{array}{c}
1 \\
\frac{1}{2}
\end{array}
\begin{array}{c}
2 \\
\frac{1}{2}
\end{array}
\begin{array}{c}
3 \\
\frac{1}{2}
\end{array}
\begin{array}{c}
4 \\
1
\end{array}
\]

It is now clear that \(P_{11}^n = 2^{-n}\) is the probability of taking the lower loop for \(n\) successive steps starting in state 1. Similarly \(P_{22}^n = 2^{-n}\) is the probability of taking the lower loop at state 2 for \(n\) successive steps.

Finally, \(P_{12}^n\) is the probability of taking the transition from state 1 to 2 exactly once out the \(n\) transitions starting in state 1 and of staying in the same state (first 1 and then 2) for the other \(n - 1\) transitions. There are \(n\) such paths, corresponding to the \(n\) possible steps at which the 1 \(\rightarrow\) 2 transition can occur, and each path has probability \(2^{-n}\). Thus \(P_{12}^n = n2^{-n}\), and we ‘see’ why this factor of \(n\) appears. The transitions \(P_{13}^n\) are then chosen to make the rows sum to 1, yielding the same solution as above..
b) Find all distinct eigenvalues and the multiplicity of each distinct eigenvalue for \([P]\).

**Solution:** Note that \([P]\) is an upper triangular matrix, and thus \([P - \lambda I]\) is also upper triangular. Thus its determinant is the product of the terms on the diagonal, \(\det[P - \lambda I] = (\frac{1}{2} - \lambda)^2(1 - \lambda)\). It follows that \(\lambda = 1\) is an eigenvalue of multiplicity 1 and \(\lambda = 1/2\) is an eigenvalue of multiplicity 2.

c) Find a right eigenvector for each distinct eigenvalue, and show that the eigenvalue of multiplicity 2 does not have 2 linearly independent eigenvectors.

**Solution:** For any Markov chain, \(e = (1, 1, 1)^T\) is a right eigenvector. This is unique here within a scale factor, since \(\lambda = 1\) has multiplicity 1. For \(\nu\) to be a right eigenvector of eigenvalue 1/2, it must satisfy

\[
\frac{1}{2}\nu_1 + \frac{1}{2}\nu_2 + 0\nu_3 = \frac{1}{2}\nu_1 \\
0\nu_1 + \frac{1}{2}\nu_2 + \frac{1}{2}\nu_3 = \frac{1}{2}\nu_2 \\
\nu_3 = \frac{1}{2}\nu_3.
\]

From the first equation, \(\nu_2 = 0\) and from the third \(\nu_3 = 0\), so \(\nu_2 = 1\) is the right eigenvector, unique within a scale factor.

d) Use (c) to show that there is no diagonal matrix \([\Lambda]\) and no invertible matrix \([U]\) for which \([P]\)[\(U\)] = \([U]\)[\(\Lambda\)].

**Solution:** Letting \(\nu_1, \nu_2, \nu_3\) be the columns of an hypothesized matrix \([U]\), we see that \([P]\)[\(U\)] = \([U]\)[\(\Lambda\)] can be written out as \([P]\)[\(\nu_i\)] = \(\lambda_i\)[\(\nu_i\)] for \(i = 1, 2, 3\). For \([U]\) to be invertible, \(\nu_1, \nu_2, \nu_3\) must be linearly independent eigenvectors of \([P]\). Part (c) however showed that 3 such eigenvectors do not exist.

e) Rederive the result of part d) using the result of a) rather than c).

**Solution:** If the \([U]\) and \([\Lambda]\) of part (d) exist, then \([P^n]\) = \([U]\)[\(\Lambda^n\)][\(U^{-1}\)]. Then, as in (4.30), \([P^n]\) = \(\sum_{i=1}^{3} \lambda_i^n [\nu^{(i)}]\pi^{(i)}\) where \(\nu^{(i)}\) is the ith column of \([U]\) and \(\pi^{(i)}\) is the ith row of \([U^{-1}]\). Since \(P_{12}^n = n(1/2)^n\), the factor of \(n\) means that it cannot have the form \(a\lambda_1^n + b\lambda_2^n + c\lambda_3^n\) for any choice of \(\lambda_1, \lambda_2, \lambda_3, a, b, c\).

Note that the argument here is quite general. If \([P^n]\) has any terms containing a polynomial in \(n\) times \(\lambda_i^n\), then the eigenvectors can’t span the space and a Jordan form decomposition is required.

**Exercise 4.15:** a) Let \([J_i]\) be a 3 by 3 block of a Jordan form, i.e.,

\[
[J_i] = \begin{bmatrix}
\lambda_i & 1 & 0 \\
0 & \lambda_i & 1 \\
0 & 0 & \lambda_i \\
\end{bmatrix}.
\]

Show that the nth power of \([J_i]\) is given by

\[
[J_i^n] = \begin{bmatrix}
\lambda_i^n & n\lambda_i^{n-1} & \binom{n}{2}\lambda_i^{n-2} \\
0 & \lambda_i^n & n\lambda_i^{n-1} \\
0 & 0 & \lambda_i^n \\
\end{bmatrix}.
\]

Hint: Perhaps the easiest way is to calculate \([J_i^2]\) and \([J_i^3]\) and then use iteration.
Solution: It is probably worthwhile to multiply \([J_i]\) by itself one or two times to gain familiarity with the problem, but note that the proposed formula for \([J^n_i]\) is equal to \([J_i]\) for \(n = 1\), so we can use induction to show that it is correct for all larger \(n\). That is, for each \(n \geq 1\), we assume the given formula for \([J^n_i]\) is correct and demonstrate that \([J^n_i][J^n_i]\) is the given formula evaluated at \(n + 1\). We suppress the subscript \(i\) in the evaluation.

We start with the elements on the first row, \(i.e.,\)
\[
\sum_j J_{1j}J_{1j}^n = \lambda \cdot \lambda^n = \lambda^{n+1}
\]
\[
\sum_j J_{1j}J_{1j}^n = \lambda \cdot n\lambda^{n-1} + 1 \cdot \lambda^n = (n+1)\lambda^n
\]
\[
\sum_j J_{1j}J_{1j}^n = \left(\frac{n}{2}\right) \lambda^{n-2} + 1 \cdot n\lambda^{n-1} = \left(\frac{n+1}{2}\right)\lambda^{n-1}.
\]

In the last equation, we used
\[
\left(\frac{n}{2}\right) + n = \frac{n(n-1)}{2} + n = \frac{n(n+1)}{2} = \left(\frac{n+1}{2}\right).
\]

The elements in the second and third row can be handled the same way, although with slightly less work. The solution in part b) provides a more elegant and more general solution.

b) Generalize a) to a \(k \times k\) block of a Jordan form \(J\). Note that the \(n\)th power of an entire Jordan form is composed of these blocks along the diagonal of the matrix.

Solution: This is rather difficult unless looked at in just the right way, but the right way is quite instructive and often useful. Thus we hope you didn’t spin your wheels too much on this, but hope you will learn from the solution. The idea comes from the elegant way to solve part a) of Exercise 4.14, namely taking a graphical approach rather than a matrix multiplication approach. Forgetting for the moment that \(J\) is not a stochastic matrix, we can draw its graph, \(i.e.,\) the representation of the non-zero elements of the matrix, as

```
1 \(\lambda\) 2 \(\lambda\) 1 \(\lambda\) 3 \(\lambda\) \ldots \(\lambda\) k
```

Each edge, say from \(i \rightarrow j\) in the graph, represents a non-zero element of \(J\) and is labelled as \(J_{ij}\). If we look at \([J^2]\), then \(J_{ij}^2 = \sum_k J_{ik}J_{kj}\) is the sum over the walks of length 2 in the graph of the ‘measure’ of that walk, where the measure of a walk is the product of the labels on the edges of that walk. Similarly, we can see that \(J_{ij}^n\) is the sum over walks of length \(n\) of the product of the labels for that walk. In short, \(J_{ij}^n\) can be found from the graph in the same way as for stochastic matrices; we simply ignore the fact that the outgoing edges from each node do not sum to 1.

We can now see immediately how to evaluate these elements. For \(J_{ij}^n\), we must look at all the walks of length \(n\) that go from \(i\) to \(j\). For \(i > j\), the number is 0, so we assume \(j \geq i\). Each such walk must include \(j - i\) rightward transitions and thus must include \(n - j + i\)
self loops. Thus the measure of each such walk is $\lambda^{n-j+i}$. The number of such walks is the number of ways that $j - i$ rightward transitions can be selected out the $n$ transitions altogether, i.e., $\binom{n}{j-i}$. Thus, the general formula is

$$J^n_{ij} = \left( \begin{array}{c} n \\ j-i \end{array} \right) \lambda^{j-i} \quad \text{for } j > i.$$  

c) Let $[P]$ be a stochastic matrix represented by a Jordan form $[J]$ as $[P] = U[J][U^{-1}]$ and consider $[U^{-1}][P][U] = [J]$. Show that any repeated eigenvalue of $[P]$ that is represented by a Jordan block of 2 by 2 or more must be strictly less than 1. Hint: Upper bound the elements of $[U^{-1}][P^n][U]$ by taking the magnitude of the elements of $[U]$ and $[U^{-1}]$ and upper bounding each element of a stochastic matrix by 1.

**Solution:** Each element of the matrix $[U^{-1}][P^n][U]$ is a sum of products of terms and the magnitude of that sum is upper bounded by the sum of products of the magnitudes of the terms. Representing the matrix whose elements are the magnitudes of a matrix $[U]$ as $|[U]|$, and recognizing that $[P^n] = ||P^n||$ since its elements are nonnegative, we have

$$|[J^n]| \leq ||U^{-1}||[P^n][|U|] \leq ||U^{-1}||[M][|U|].$$  

where $[M]$ is a matrix whose elements are all equal to 1.

The matrix on the right is independent of $n$, so each of its elements is simply a finite number. If $[P]$ has an eigenvalue $\lambda$ of magnitude 1 whose multiplicity exceeds its number of linearly independent eigenvectors, then $[J^n]$ contains an element $n\lambda^{n-1}$ of magnitude $n$, and for large enough $n$, this is larger than the finite number above, yielding a contradiction. Thus, for any stochastic matrix, any repeated eigenvalue with a defective number of linearly independent eigenvectors has a magnitude strictly less than 1.

d) Let $\lambda_s$ be the eigenvalue of largest magnitude less than 1. Assume that the Jordan blocks for $\lambda_s$ are at most of size $k$. Show that each ergodic class of $[P]$ converges at least as fast as $n^k \lambda_s^n$.

**Solution:** What we want to show is that $|[P^n] - e\pi| \leq bn^k |\lambda_s^n|$ for some sufficiently large constant $b$ depending on $[P]$ but not $n$.

The states of an ergodic class have no transitions out, and for questions of convergence within the class we can ignore transitions in. Thus we will get rid of some excess notation by simply assuming an ergodic Markov chain. Thus there is a single eigenvalue equal to 1 and all other eigenvalues are strictly less than 1. The largest such in magnitude is denoted as $\lambda_s$. Assume that the eigenvalues are arranged in $[J]$ in decreasing magnitude, so that $J_{11} = 1$.

For all $n$, we have $[P^n] = [U][J^n][U^{-1}]$. Note that $\lim_{n \to \infty} [J^n]$ is a matrix with $J_{ij} = 0$ except for $i = j = 1$. Thus it can be seen that the first column of $[U]$ is $e$ and the first row of $[U^{-1}]$ is $\pi$. It follows from this that $[P^n] - e\pi = [U][J^n][U^{-1}]$ where $[J^n]$ is the same as $[J^n]$ except that the eigenvalue 1 has been replaced by 0. This means, however, that the magnitude of each element of $[J^n]$ is upper bounded by $n^k \lambda_s^{n-k}$. It follows that when the magnitude of the elements of $[U][J^n][U^{-1}]$ the resulting elements are at most $bn^k |\lambda_s^n|$ for large enough $b$ (the value of $b$ takes account of $[U], [U^{-1}]$, and $\lambda_s^k$).

**Exercise 4.19:** Suppose a recurrent Markov chain has period $d$ and let $S_m$, $0 \leq m \leq d-1$, be the $m$th subset in the sense of Theorem 4.2.3. Assume the states are numbered so that the first $s_0$ states are the
states of $S_0$, the next $s_1$ are those of $S_1$, and so forth. Thus the matrix $[P]$ for the chain has the block form given by

$$
[P] = 
\begin{bmatrix}
0 & [P_0] & \cdots & 0 \\
0 & 0 & [P_1] & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & [P_{d-2}] \\
[P_{d-1}] & 0 & \cdots & 0 \\
\end{bmatrix},
$$

where $[P_m]$ has dimension $s_m \times s_{m+1}$ for $0 \leq m \leq d - 1$, where $(d - 1) + 1$ is interpreted as 0 throughout. In what follows it is often more convenient to express $[P_m]$ as an $M \times M$ matrix $[P'_m]$ whose entries are 0 except for the rows of $S_m$ and the columns of $S_{m+1}$, where the entries are equal to those of $[P_m]$. In this view, $[P] = \sum_{m=0}^{d-1} [P'_m]$.

a) Show that $[P^d]$ has the form

$$
[P^d] = 
\begin{bmatrix}
[Q_0] & 0 & \cdots & 0 \\
0 & [Q_1] & \cdots & \vdots \\
0 & 0 & \cdots & [Q_{d-1}] \\
\end{bmatrix},
$$

where $[Q_m] = [P_m][P_{m+1}][P_{d-1}][P_0] \cdots [P_{m-1}]$. Expressing $[Q_m]$ as an $M \times M$ matrix $[Q'_m]$ whose entries are 0 except for the rows and columns of $S_m$ where the entries are equal to those of $[Q_m]$, this becomes $[P^d] = \sum_{m=0}^{d-1} [Q'_m]$.

Solution: As explained in Section 4.3.5, $P^d_{ij}$ must be 0 for $(i, j)$ not in the same subset and thus $S_j$. Since the subsets are arranged as contiguous sets of state numbers, $[Q_j]$ is the $d$th order transition matrix of the states in class $S_j$. To see this more clearly and to understand why $[Q_m] = [P_m][P_{m+1}][P_{d-1}][P_0] \cdots [P_{m-1}]$, note that $[P^2]$ is given by

$$
[P^2] = 
\begin{bmatrix}
0 & 0 & \cdots & [P_0P_1] \\
0 & 0 & \cdots & [P_1P_2] \\
\vdots & \vdots & \ddots & \vdots \\
[P_{d-2}P_{d-1}] & 0 & \cdots & 0 \\
0 & [P_{d-1}P_0] & \cdots & \vdots \\
\end{bmatrix}.
$$

b) Show that $[Q_m]$ is the matrix of an ergodic Markov chain and let its eigenvector of eigenvalue 1 (scaled to be a probability vector) be $\hat{\pi}_m$ and let $\hat{\nu}_m$ be the corresponding right eigenvector scaled so that $\hat{\pi}_m \hat{\nu}_m = 1$. Let $\hat{\pi}'_m$ and $\hat{\nu}'_m$ be the corresponding $M$-vectors. Show that $\lim_{n\to\infty}[P'^d] = \sum_{m=0}^{d-1} \hat{\nu}'_m \hat{\pi}'_m$.

Solution: Note that $[Q_m]$ for $0 \leq m \leq d$ must be recurrent since $[P]$ is recurrent and thus each pair of states in $S_m$ communicate, and communicate only within periods of $d$. Also $[Q_m]$ is aperiodic since otherwise $P^k_{ii}$ for $i \in S_m$ would be nonzero only for $k = \ell d$ for some integer $\ell > 1$, which is contrary to the given fact that the chain, and thus $i$, is periodic with period $d$. Thus $[Q_m]$ is the matrix of an ergodic Markov chain. It follows that $[Q_m]$ has a unique left eigenvector $\hat{\pi}_m$ whose components sum to 1, i.e., $\hat{\pi}_m \hat{\nu}_m = 1$ where $\hat{\nu}_m$ is the unit vector of $s_m$ dimensions (i.e., $\hat{\nu}_m = \hat{\epsilon}$). Since $\lim_{n\to\infty}[Q^n_m] = \hat{\nu}_m \hat{\pi}_m$, we have $\lim_{n\to\infty}[P'^d] = \sum_{m=0}^{d-1} \hat{\nu}'_m \hat{\pi}'_m$. 


c) Show that $\hat{\pi}'_m[P'_m] = \hat{\pi}'_{m+1}$. Note that $\hat{\pi}'_m$ is an M-tuple that is non-zero only on the components of $S_m$.

**Solution:** Note that $\hat{\pi}'_m[Q'_m] = \hat{\pi}'_m$, i.e.,

$$\hat{\pi}'_m[P'_m P'_{m+1} \cdots P'_d P'_0 \cdots P'_{m-1}] = \hat{\pi}'_m$$

For $m < d - 1$, we can postmultiply both sides of this equation by $[P'_m]$ to get

$$\hat{\pi}'_m[P'_m][Q_{m+1}] = \hat{\pi}'_m[P'_m].$$

This means that $\hat{\pi}'_m[P'_m]$ is a left eigenvector of $[Q'_{m+1}]$. For $m = d - 1$, the same argument shows that $\hat{\pi}'_{d-1}[P'_{d-1}]$ is a left eigenvector of $[Q'_0]$.

d) Let $\phi = \frac{2\pi \sqrt{-1}}{d}$ and let $\pi^{(k)} = \sum_{m=0}^{d-1} \hat{\pi}'_m e^{mk\phi}$. Show that $\pi^{(k)}$ is a left eigenvector of $[P]$ of eigenvalue $e^{-k\phi}$.

**Solution:** Note that $\{k\phi; k = 0, 1, \ldots, d - 1\}$ is the set of $d$ equi-spaced points around the unit circle. We want to show that these points are (possibly complex) eigenvalues of $[P]$ and that $\pi^{(k)}$ is a corresponding left eigenvector. We have

$$\pi^{(k)}[P] = (\hat{\pi}'_0, \hat{\pi}'_1 e^{\phi}, \ldots, \hat{\pi}'_{d-1} e^{d(d-1)\phi})[P]$$

$$= (\hat{\pi}'_{d-1} e^{d(d-1)\phi}[P'_d], \hat{\pi}'_0[P'_0], \ldots, \hat{\pi}'_{d-1} e^{d(d-2)\phi}[P'_1])$$

$$= (e^{(d-1)\phi}\hat{\pi}'_0, e^{\phi}\hat{\pi}'_1, \ldots, e^{d(d-2)\phi}\hat{\pi}'_{d-2})$$

$$= e^{-k\phi}\pi^{(k)}.$$

In the second step, we used the block form structure of $[P]$ and in the third step above, we used part (c).

**Exercise 4.21:** Suppose $A$ and $B$ are each ergodic Markov chains with transition probabilities $\{P_{A_i, A_j}\}$ and $\{P_{B_i, B_j}\}$ respectively. Denote the steady-state probabilities of $A$ and $B$ by $\{\pi_{A_i}\}$ and $\{\pi_{B_i}\}$ respectively.

The chains are now connected and modified as shown below. In particular, states $A_1$ and $B_1$ are now connected and the new transition probabilities $P'$ for the combined chain are given by

$$P_{A_1, B_1} = \epsilon, \quad P'_{A_1, A_j} = (1 - \epsilon)P_{A_1, A_j} \quad \text{for all } A_j$$

$$P'_{B_1, A_1} = \delta, \quad P'_{B_1, B_j} = (1 - \delta)P_{B_1, B_j} \quad \text{for all } B_j.$$

All other transition probabilities remain the same. Think intuitively of $\epsilon$ and $\delta$ as being small, but do not make any approximations in what follows. Give your answers to the following questions as functions of $\epsilon$, $\delta$, $\{\pi_{A_i}\}$ and $\{\pi_{B_i}\}$.

**a)** Assume that $\epsilon > 0$, $\delta = 0$ (i.e., that $A$ is a set of transient states in the combined chain). Starting in state $A_1$, find the conditional expected time to return to $A_1$ given that the first transition is to some state in chain $A$. 

---

<Diagram>
Solution: Conditional on the first transition from state $A_1$ being to a state in $A$, these conditional transition probabilities are the same as the original transition probabilities for $A$. If we look at a long sequence of transitions in chain $A$ alone, the relative frequency of state $A_1$ tends to $\pi_{A_1}$, so we might hypothesize that the expected time to return to $A_1$ starting from $A_1$ is $1/\pi_{A_1}$. This hypothesis is correct and we will verify it by a somewhat simpler argument when we study renewal theory. Here, however, we verify it by looking at first passage times within the chain $A$. For now, label the states in $A$ as $(1, 2, \ldots, M)$ where $1$ stands for $A_1$. For $2 \leq i \leq M$, let $v_i$ be the expected time to first reach state $1$ starting in state $i$. As in (4.31),

$$v_i = 1 + \sum_{j \neq 1} P_{ij}v_j; \quad 2 \leq i \leq M$$

We can use these equations to write an equation for the expected time $T$ to return to state 1 given that we start in state 1. The first transition goes to each state $j$ with probability $P_{1j}$ and the remaining time to reach state 1 from state $j$ is $v_j$. If we define $v_1 = 0$ (i.e., for a transition from 1 to 1, there is no remaining time required to return to state 1), we then have

$$T = 1 + \sum_{j=1}^{M} P_{ij}v_j$$

Note that this is very different from (4.32) where $[P]$ is a Markov chain in which 1 is a trapping state. Note also that $v_1 = 0$ is interpreted here as the time to reach state 1 when in state 1 and $T$ is the expected time for the first return to 1 starting in 1. We can now write these equations as a vector equation:

$$Te_1 + v = e + [P]v$$

where $e_1 = (1, 0, 0, \ldots, 0)^T$ and $e = (1, 1, \ldots, 1)^T$. Motivated by the hypothesis that $T = 1/\pi_1$, we premultiply this vector equation by the steady state row vector $\pi$, getting

$$T\pi_1 + \pi v = 1 + \pi [P]v = 1 + \pi v$$

Cancelling terms, we get $T = 1/\pi_1$ as hypothesized.

b) Assume that $\epsilon > 0$, $\delta = 0$. Find $T_{A,B}$, the expected time to first reach state $B_1$ starting from state $A_1$. Your answer should be a function of $\epsilon$ and the original steady state probabilities $\{\pi_{A_i}\}$ in chain $A$.

Solution: Starting in state $A_1$, we reach $B_1$ in a single step with probability $\epsilon$. With probability $1 - \epsilon$, we wait for a return to $A_1$ and then have expected time $T_{A,B}$ remaining. Thus $T_{A,B} = \epsilon + (1 - \epsilon)(1/\pi_{A_1} + T_{A,B})$. Solving this equation,

$$T_{A,B} = 1 + \frac{1 - \epsilon}{\epsilon \pi_{A_1}}$$

c) Assume $\epsilon > 0$, $\delta > 0$. Find $T_{B,A}$, the expected time to first reach state $A_1$, starting in state $B_1$. Your answer should depend only on $\delta$ and $\{\pi_{B_i}\}$.
Solution: The fact that $\epsilon > 0$ here is irrelevant since that transition can never be used in the first passage from $B_1$ to $A_1$. Thus the answer is the reversed version of the answer to (b), where now $\pi_{B_1}$ is the steady state probability of $B_1$ for chain $B$ alone.

$$T_{B,A} = 1 + \frac{1 - \delta}{\delta \pi_{B_1}}$$

d) Assume $\epsilon > 0$ and $\delta > 0$. Find $P'(A)$, the steady-state probability that the combined chain is in one of the states $\{A_j\}$ of the original chain $A$.

Solution: With $1 > \epsilon > 0$ and $1 > \delta > 0$, it is clear that the combined chain is ergodic, so we can in principal calculate the steady state probabilities. In parts (d), (e), and (f), we are interested in those probabilities in chain $A$. We denote those states, as before, as $(1, \ldots, M)$ where 1 is state $A_1$. Steady state probabilities for $A$ in the combined chain for given $\epsilon$, $\delta$ are denoted $\pi_j'$, whereas they are denoted as $\pi_j$ in the original chain. We first find $\pi_1'$ and then $\pi_j'$ for $2 \leq j \leq M$.

As we saw in part (a), the expected first return time from a state to itself is the reciprocal of the steady state probability, so we first find $T_{AA}$, the expected time of first return from $A$ to $A$. Given that the first transition from state 1 goes to a state in $A$, the expected first-return time is $1/\pi_1$ from part (a). If the transition goes to $B_2$, the expected first-return time is $1 + T_{BA}$, where $T_{BA}$ is found in part (c). Combining these, with the a priori probabilities of going to $A$ or $B$, $T_{AA} = (1 - \epsilon)/\pi_1 + \epsilon + \epsilon(1 - \delta)/(\delta \pi_{B_1})$. Thus

$$\pi_1' = \left[ \frac{1 - \epsilon}{\pi_1} + 2\epsilon + \frac{\epsilon(1 - \delta)}{\delta \pi_{B_1}} \right]^{-1}$$

Next we find $\pi_j'$ for the other states in $A$ in terms of the $\pi_j$ for the uncombined chains and $\pi_1'$. The original and the combined steady state equations for $2 \leq j \leq M$ are

$$\pi_j = \sum_{i \neq 1} \pi_i P_{ij} + \pi_1 P_{1j}; \quad \pi_j' = \sum_{i \neq 1} \pi_i' P_{ij} + \pi_1'(1 - \epsilon) P_{1j}$$

These equations, as $M - 1$ equations in the unknowns $\pi_j$; $j \geq 2$ for known $\pi_1$, uniquely specify $\pi_2, \ldots, \pi_M$ and they differ in $\pi_1$ being replaced by $(1 - \epsilon)\pi_1'$. From this, we see that the second set of equations is satisfied if we choose

$$\pi_j' = \frac{\pi_j (1 - \epsilon) \pi_1'}{\pi_1}$$

We can now sum the steady-state probabilities in $A$ to get

$$\Pr\{A\} = \sum_{j=1}^{M} \pi_j' = \pi_1' \left[ \epsilon + \frac{1 - \epsilon}{\pi_1} \right]$$

e) Assume $\epsilon > 0$, $\delta = 0$. For each state $A_j \neq A_1$ in $A$, find $v_{A_j}$, the expected number of visits to state $A_j$, starting in state $A_1$, before reaching state $B_1$. Your answer should depend only on $\epsilon$ and $\{\pi_{A_i}\}$.

Solution: Here we use a variation on the first passage time problem in part (a) to find the expected number of visits to state $j$, $E[N_j]$, in the original chain starting in state 1 before
the first return to 1. Here we let \( v_i \) be the expected number of visits to \( j \), starting in state \( i \), before the first visit to 1. The equations are

\[
v_i = \sum_{k \neq 1} P_{ik} v_k + P_{ij} \quad \text{for} \ i \neq 1; \quad E[N_j] = \sum_{k \neq 1} P_{ik} v_k + P_{ij}
\]

Writing this as a vector equation,

\[
E[N_j] \ e_1 + \ v = [P] e_j + [P] v
\]

Premultiplying by \( \pi \), we see that \( E[N_j] = \pi_j/\pi_1 \). Finally, to find \( v_{A_j} \), the expected number of visits to state \( j \) before the first to \( B_1 \), we have

\[
v_{A_j} = (1 - \epsilon)[E[N_j] + v_{A_j}] = \frac{(1 - \epsilon)E[N_j]}{\epsilon} = \frac{(1 - \epsilon)\pi_j}{\epsilon \pi_i}
\]

f) Assume \( \epsilon > 0, \delta > 0 \). For each state \( A_j \) in \( A \), find \( \pi'_{A_j} \), the steady-state probability of being in state \( A_j \) in the combined chain. Hint: Be careful in your treatment of state \( A_1 \).

**Solution:** This was solved in part (d). Readers might want to come back to this exercise later and re-solve it using renewal theory.

**Exercise 4.28:** Consider finding the expected time until a given string appears in a IID binary sequence with \( \Pr\{X_i = 1\} = p_1 \), \( \Pr\{X_i = 0\} = p_0 = 1 - p_1 \).

a) Following the procedure in Example 4.5.1, draw the 3 state Markov chain for the string \((0, 1)\). Find the expected number of trials until the first occurrence of the string.

**Solution:**

Let \( v_i \) be the expected first passage time from node \( i \) to node 2. Then

\[
\begin{align*}
\ v_0 &= 1 + p_1 v_0 + p_0 v_1; \quad v_1 = 1 + p_0 v_1 + p_1 v_2 \\
\ v_0 &= 1/p_0 + v_1; \quad v_1 = 1/p_1 + v_2.
\end{align*}
\]

Combining these equations to eliminate \( v_1 \),

\[
v_0 = 1/p_0 + 1/p_1 + v_2 = 1/p_0 p_1 + v_2.
\]

Finally, state 2 is the trapping state, so \( v_2 = 0 \) and \( v_0 = 1/p_0 p_1 \).

b) For parts b) to d), let \( (a_1, a_2, a_3, \ldots, a_k) = (0, 1, 1, \ldots, 1) \), i.e., zero followed by \( k - 1 \) ones. Draw the corresponding Markov chain for \( k = 4 \).

**Solution:**
c) Let \( v_i, 1 \leq i \leq k \) be the expected first-passage time from state \( i \) to state \( k \). Note that \( v_k = 0 \). For each \( i, 1 \leq i < k \), show that \( v_i = \alpha_i + v_{i+1} \) and \( v_0 = \beta_i + v_{i+1} \) where \( \alpha_i \) and \( \beta_i \) are each expressed as a product of powers of \( p_0 \) and \( p_1 \). Hint: use induction on \( i \) using \( i = 1 \) as the base. For the inductive step, first find \( \beta_{i+1} \) as a function of \( \beta_i \) starting with \( i = 1 \) and using the equation \( v_0 = 1/p_0 + v_1 \).

**Solution:** Part a) solved the problem for \( i = 1 \). The fact that the string length was 2 there was of significance only at the end where we set \( v_2 = 0 \). We found that \( \alpha_1 = 1/p_1 \) and \( \beta_1 = 1/p_0p_1 \).

For the inductive step, assume \( v_i = \alpha_i + v_{i+1} \) and \( v_0 = \beta_i + v_{i+1} \) for a given \( i \). Using the basic first-passage-time equation,

\[
v_{i+1} = 1 + p_0v_1 + p_1v_{i+2}
\]

\[
= p_0v_0 + p_1v_{i+2}
\]

\[
= p_0\beta_i + p_0v_{i+1} + p_1v_{i+2}.
\]

Combining the terms in \( v_{i+1} \)

\[
v_{i+1} = \frac{p_0\beta_i}{p_1} + v_{i+2}.
\]

This completes half the inductive step, showing that \( a_{i+1} = p_0\beta_i/p_1 \). Now we use the inductive assumption on \( v_{i+1} \),

\[
v_0 = \beta_i + v_{i+1} = \beta_i + \frac{p_0\beta_i}{p_1} + v_{i+2} = \frac{\beta_i}{p_1} + v_{i+2}.
\]

This completes the second half of the induction, showing that \( \beta_{i+1} = \beta_i/p_1 \). Iterating on these equations for \( \beta_i \) and \( \alpha_i \), we find the explicit expression

\[
\alpha_i = \frac{1}{p_1}; \quad \beta_i = \frac{1}{p_0p_1}.
\]

Note that the algebra here was quite simple, but if one did not follow the hints precisely, one could get into a terrible mess. In addition, the whole thing was quite unmotivated. We look at this again in Exercise 5.35 and find quite an intuitive way to get the same answer.

d) Let \( a = (0, 1, 0) \). Draw the corresponding Markov chain for this string. Evaluate \( v_0 \), the expected time for \( (0, 1, 0) \) to occur.

**Solution:**
The solution for \( v_0 \) and \( v_1 \) in terms of \( v_2 \) is the same as part a). The basic equation for \( v_2 \) in terms of its outward transitions is

\[
v_2 = 1 + p_0 v_0 + p_1 v_3 = 1 + p_0 \left[ \frac{1}{p_0 p_1} + v_2 \right].
\]

Combining the terms in \( v_2 \), we get

\[ p_1 v_2 = 1 + 1/p_1. \]

Using \( v_0 = 1/p_0 p_1 + v_2 \),

\[
v_0 = \frac{1}{p_1^2 p_0} + \frac{1}{p_1} + \frac{1}{p_1} = 1 + \frac{1}{p_1 p_0^2}.
\]

This solution will become more transparent after doing Exercise 5.35.

**Exercise 5.2:** The purpose of this exercise is to show that, for an arbitrary renewal process, \( N(t) \), the number of renewals in \((0, t]\), has finite expectation.

a) Let the inter-renewal intervals have the CDF \( F_X(x) \), with, as usual, \( F_X(0) = 0 \). Using whatever combination of mathematics and common sense is comfortable for you, show that for any \( \epsilon \), \( 0 < \epsilon < 1 \), there is a \( \delta > 0 \) such that \( F_X(\delta) \leq 1 - \epsilon \). In other words, you are to show that a positive rv must lie in some range of positive values bounded away from 0 with positive probability.

**Solution:** Consider the sequence of events \( \{X > 1/k; k \geq 1\} \). The union of these events is the event \( \{X > 0\} \). Since \( \Pr\{X \leq 0\} = 0 \), \( \Pr\{X > 0\} = 1 \). The events \( \{X > 1/k\} \) are nested in \( k \), so that, from (1.9),

\[
1 = \Pr\left\{ \bigcup_k \{X > 1/k\} \right\} = \lim_{k \to \infty} \Pr\{X > 1/k\}.
\]

Thus, for any \( 0 < \epsilon < 1 \), and any \( k \) large enough, \( \Pr\{X > 1/k\} > \epsilon \). Taking \( \delta \) to be \( 1/k \) for that value of \( k \) shows that \( \Pr\{X \leq \delta\} \leq 1 - \epsilon \). Another equally good approach is to use the continuity from the right of \( F_X \).

b) Show that \( \Pr\{S_n \leq \delta\} \leq (1 - \epsilon)^n \).

**Solution:** \( S_n \) is the sum of \( n \) interarrival times, and, bounding very loosely, \( S_n \leq \delta \) implies that for each \( i, 1 \leq i \leq n, X_i \leq \delta \). The \( X_i \) are independent, so, since \( \Pr\{X_i \leq \delta\} \leq (1 - \epsilon) \), we have \( \Pr\{S_n \leq \delta\} \leq (1 - \epsilon)^n \).

c) Show that \( \mathbb{E}[N(\delta)] \leq 1/\epsilon \).
Solution: Since $N(t)$ is nonnegative,
\[
E[N(\delta)] = \sum_{n=1}^{\infty} \Pr\{N(\delta) \geq n\} = \sum_{n=1}^{\infty} \Pr\{S_n \leq \delta\} \leq \sum_{n=1}^{\infty} (1 - \epsilon)^n = \frac{1 - \epsilon}{1 - (1 - \epsilon)} = \frac{1 - \epsilon}{\epsilon} \leq \frac{1}{\epsilon}.
\]

\[\text{d) For the } \epsilon, \delta \text{ of part a), show that for every integer } k, E[N(k\delta)] \leq k/\epsilon \text{ and thus that } E[N(t)] \leq \frac{t+\delta}{\epsilon} \text{ for any } t > 0.\]

Solution: The solution of part c) suggests breaking the interval $(0, k\delta]$ into $k$ intervals each of size $\delta$. Letting $\tilde{N}_i = N(i\delta) - N((i-1)\delta)$ be the number of arrivals in the $i$th of these intervals, we have $E[N(k\delta)] = \sum_{i=1}^{k} E[\tilde{N}_i]$. For the first of these intervals, we have shown that $E[N_1] \leq 1/\epsilon$, but that argument does not quite work for the subsequent intervals, since the first arrival in that interval does not necessarily correspond to an interarrival interval less than $\delta$. In order to have $n$ arrivals in the interval $((i-1)\delta, i\delta]$, however, the final $n-1$ arrivals must be at the end of interarrival intervals at most $\delta$. Thus if we let $S_n^{(i)}$ be the number of arrivals in the $i$th interval, we have
\[
\Pr\{S_n^{(i)} \leq \delta\} \leq (1 - \epsilon)^{n-1}.
\]
Repeating the argument in part c), then,
\[
E[\tilde{N}_i] = \sum_{n=1}^{\infty} \Pr\{\tilde{N}_i \geq n\} = \sum_{n=1}^{\infty} \Pr\{S_n^{(i)} \leq \delta\} \leq \sum_{n=1}^{\infty} (1 - \epsilon)^{n-1} = \frac{1}{1 - (1 - \epsilon)} = \frac{1}{\epsilon}.
\]
Since $E[N(k\delta)] = \sum_{i=1}^{k} E[N_i]$, we then have
\[
E[N(k\delta)] \leq k/\epsilon.
\]
Since $N(t)$ is non-decreasing in $t$, it can be upper bounded by the integer multiple of $1/\delta$ that is just larger than $t$, i.e.,
\[
E[N(t)] \leq E[N(\delta\lceil t/\delta \rceil)]] \leq \frac{[t/\delta]}{\epsilon} \leq \frac{t/\delta + 1}{\epsilon}.
\]

\[\text{e) Use the result here to show that } N(t) \text{ is non-defective.}\]
Solution: Since $N(t)$, for each $t$, is nonnegative and has finite expectation, it certainly must be non-defective. One way to see this is that $E[N(t)]$ is the integral of the complementary CDF, $F_{N(t)}^c(n)$ of $N(t)$. Since this integral is finite, $F_{N(t)}^c(n)$ must approach 0 with increasing $n$.

Exercise 5.4: Is it true for a renewal process that:

a) $N(t) < n$ if and only if $S_n > t$?

b) $N(t) \leq n$ if and only if $S_n \geq t$?

c) $N(t) > n$ if and only if $S_n < t$?

Solution: Part a) is true, as pointed out in (5.1). It is the obverse of the statement that $N(t) \geq n$ if and only if $S_n \leq t$.

Parts b) and c) are false, as seen by any situation where $S_n < t$ and $S_{n+1} > t$. In these cases, $N(t) = n$.

Exercise 5.5: (This shows that convergence WP1 implies convergence in probability.) a) Let $\{Y_n; n \geq 1\}$ be a sequence of rv's that converges to 0 WP1. For any positive integers $m$ and $k$, let

$$A(m,k) = \{ \omega : |Y_n(\omega)| \leq 1/k \text{ for all } n \geq m \}.$$

a) Show that if $\lim_{n \to \infty} Y_n(\omega) = 0$ for some given $\omega$, then (for any given $k$) $\omega \in A(m,k)$ for some positive integer $m$.

Solution: Note that for a given $\omega$, $\{Y_n(\omega; n \geq 1\}$ is simply a sequence of real numbers. The definition of convergence of such a sequence to 0 says that for any $\epsilon$ (or any $1/k$ where $k > 0$ is an integer), there must be an $m$ large enough that $Y_n(\omega) \leq 1/k$ for all $n \geq m$. In other words, the given $\omega$ is contained in $A(m,k)$ for that $m$.

b) Show that for all $k \geq 1$

$$\Pr\left\{ \bigcup_{m=1}^{\infty} A(m,k) \right\} = 1.$$

Solution: The set of $\omega$ for which $\lim_{n \to \infty} Y_n(\omega) = 0$ has probability 1, and each such $\omega$ is in $A(m,k)$ for some integer $m > 0$. Thus each such $\omega$ lies in the above union, implying that the union has probability 1.

c) Show that, for all $m \geq 1$, $A(m,k) \subseteq A(m+1,k)$. Use this (plus (1.9)) to show that

$$\lim_{m \to \infty} \Pr\{A(m,k)\} = 1.$$

Solution: Note that if $|Y_n(\omega)| \leq 1/k$ for all $n \geq m$, then also $|Y_n(\omega)| \leq 1/k$ for all $n \geq m+1$. This means that $A(m,k) \subseteq A(m+1,k)$. From (1.9) then

$$1 = \Pr\left\{ \bigcup_m A(m,k) \right\} = \lim_{m \to \infty} \Pr\{A(m,k)\}.$$

d) Show that if $\omega \in A(m,k)$, then $|Y_m(\omega)| \leq 1/k$. Use this (plus part c) to show that

$$\lim_{m \to \infty} \Pr\{|Y_m| > 1/k\} = 0.$$
**Solution:** Note that if $|Y_n(\omega)| \leq 1/k$ for all $n \geq m$, then certainly $|Y_n(\omega)| \leq 1/k$ for $n = m$. It then follows from part c) that $\lim_{m \to \infty} \Pr\{|Y_m| \leq 1/k\} = 1$, which is equivalent to the desired statement. This shows that $\{Y_n; n \geq 1\}$ converges in probability.

**Exercise 5.8:** The Borel-Cantelli lemma Consider the event $\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$ where $A_1, A_2, \ldots$ are arbitrary events.

a) Show that

$$\lim_{m \to \infty} \Pr\{\bigcup_{n=m}^{\infty} A_n\} = 0 \iff \Pr\{\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\} = 0.$$  

Hint: Apply (1.10).

**Solution:** Let $B_m = \bigcup_{n=m}^{\infty} A_n$. Then we are to show that

$$\lim_{m \to \infty} \Pr\{B_m\} = 0 \iff \Pr\{\bigcap_{m=1}^{\infty} B_m\} = 0$$

Since $B_m \subseteq B_{m+1}$ for each $m \geq 1$, this is implied by (1.10).

b) Show that $\sum_{m=1}^{\infty} \Pr\{A_m\} < \infty$ implies that $\lim_{m \to \infty} \Pr\{\bigcup_{n=m}^{\infty} A_n\} = 0$. Hint: First explain why $\sum_{m=1}^{\infty} \Pr\{A_m\} < \infty$ implies that $\lim_{m \to \infty} \sum_{n=m}^{\infty} \Pr\{A_n\} = 0$.

**Solution:** Let $S_m = \sum_{n=1}^{m} A_n$. Then the meaning of $\sum_{m=1}^{\infty} A_m < \infty$ is that $\lim_{m \to \infty} S_m$ has a finite limit, say $S_\infty$. Thus $\lim_{m \to \infty} (S_\infty - S_m) = 0$, i.e., $\lim_{m \to \infty} \sum_{n=m}^{\infty} \Pr\{A_n\} = 0$. Now, using the union bound,

$$\Pr\{\bigcup_{n=m}^{\infty} A_n\} \leq \sum_{m=1}^{\infty} \Pr\{A_n\}$$

Taking the limit of both sides, we get the desired result.

c) Show that if $\sum_{m=1}^{\infty} \Pr\{A_m\} < \infty$, then $\Pr\{\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\} = 0$. This well-known result is called the Borel-Cantelli lemma.

**Solution:** If $\sum_{m=1}^{\infty} \Pr\{A_m\} < \infty$, then from part b), $\lim_{m \to \infty} \Pr\{\bigcup_{n=m}^{\infty} A_n\} = 0$. From part a), this implies that $\Pr\{\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\} = 0$.

d) The set $\bigcap_{n=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$ is often referred to as the set of sample points $\omega$ that are contained in infinitely many of the $A_n$. Consider an $\omega$ that is contained in a finite number $k$ of the sets $A_n$ and argue that there must be an integer $m$ such that $\omega \notin A_n$ for all $n > m$. Conversely, if $\omega$ is contained in infinitely many of the $A_n$, show that there cannot be such an $m$.

**Solution:** If a given $\omega$ is in any of the sets $A_n$, let $b_1$ be the smallest $n$ for which $\omega \in A_n$. If it is in other sets, let $b_2$ be the next smallest, and so forth. Thus $\omega \in A_{b_i}$ for $1 \leq i \leq k$. If $k$ is finite, then $\omega \notin A_n$ for all $n > b_k$. If $k$ is infinite, then for all $m$, $\omega \in A_n$ for some $n > m$.

**Exercise 5.9:** (Strong law for renewals where $\overline{X} = \infty$) Let $\{X_i; i \geq 1\}$ be the inter-renewal intervals of a renewal process and assume that $E[X_i] = \infty$. Let $b > 0$ be an arbitrary number and $\tilde{X}_i$ be a truncated random variable defined by $\tilde{X}_i = X_i$ if $X_i \leq b$ and $\tilde{X}_i = b$ otherwise.

a) Show that for any constant $M > 0$, there is a $b$ sufficiently large so that $E[\tilde{X}_i] \geq M$. 

Solution: Since \(E[X] = \int_0^\infty F_X^c(x) \, dx = \infty\), we know from the definition of an integral over an infinite limit that
\[
E[X] = \lim_{b \to \infty} \int_0^b F_X^c(x) \, dx = \infty.
\]
For \(\bar{X} = \min(X, b)\), we see that \(F_X(x) = F_X(x)\) for \(x \leq b\) and \(F_X(x) = 1\) for \(x > b\). Thus \(E[\bar{X}] = \int_0^b F_X^c(x) \, dx\). Since \(E[\bar{X}]\) is increasing with \(b\) toward \(\infty\), we see that for any \(M > 0\), there is a \(b\) sufficiently large that \(E[\bar{X}] \geq M\).

b) Let \(\{\bar{N}(t); t \geq 0\}\) be the renewal counting process with inter-renewal intervals \(\{\bar{X}_i; i \geq 1\}\) and show that for all \(t > 0\), \(\bar{N}(t) \geq N(t)\).

Solution: Note that \(X - \bar{X}\) is a non-negative rv, i.e., it is 0 for \(X \leq b\) and greater than \(b\) otherwise. Thus \(\bar{X} \leq X\). It follows then that for all \(n \geq 1\),
\[
\bar{S}_n = \bar{X}_1 + \bar{X}_2 + \cdots \bar{X}_n \leq X_1 + X_2 + \cdots X_n = S_n.
\]
Since \(\bar{S}_n \leq S_n\), it follows for all \(t > 0\) that if \(S_n \leq t\) then also \(\bar{S}_n \leq t\). This then means that if \(N(t) \geq n\), then also \(\bar{N}(t) \geq n\). Since this is true for all \(n\), \(\bar{N}(t) \geq N(t)\), i.e., the reduction of inter-renewal lengths causes an increase in the number of renewals.

c) Show that for all sample functions \(N(t, \omega)\), except a set of probability 0, \(N(t, \omega)/t < 2/M\) for all sufficiently large \(t\). Note: Since \(M\) is arbitrary, this means that \(\lim_{t \to \infty} N(t)/t = 0\) with probability 1.

Solution: Let \(M\) and \(b < \infty\) such that \(E[\bar{X}] \geq M\) be fixed in what follows. Since \(\bar{X} \leq b\), we see that \(E[\bar{X}] < \infty\), so we can apply Theorem 4.3.1, which asserts that
\[
\Pr\left\{ \omega : \lim_{t \to \infty} \frac{\bar{N}(t, \omega)}{t} = \frac{1}{E[\bar{X}]} \right\} = 1.
\]
Let \(A\) denote the set of sample points above for which the above limit exists, i.e., for which \(\lim_{t \to \infty} \bar{N}(t, \omega)/t = 1/E[\bar{X}]\). We will show that, for each \(\omega \in A\), \(\lim_{t} N(t, \omega)/t \leq 1/2M\).

We know that any for \(\omega \in A\), \(\lim_{t} \bar{N}(t, \omega)/t = 1/E[\bar{X}]\). The definition of the limit of a real valued function states that for any \(\epsilon > 0\), there is a \(\tau(\epsilon)\) such that
\[
\left| \frac{\bar{N}(t, \omega)}{t} - \frac{1}{E[\bar{X}]} \right| \leq \epsilon \quad \text{for all } t \geq \tau(\epsilon).
\]
Note that \(\tau(\epsilon)\) depends on \(b\) and \(\omega\) as well as \(\epsilon\), so we denote it as \(\tau(\epsilon, b, \omega)\). Using only one side of this inequality, \(N(t, \omega)/t \leq \epsilon + 1/E[\bar{X}]\) for all \(t \geq \tau(\epsilon, b, \omega)\). Since we have seen that \(N(t, \omega) \leq \bar{N}(t, \omega)\) and \(\bar{X} \leq M\), we have
\[
\frac{N(t, \omega)}{t} \leq \epsilon + \frac{1}{M} \quad \text{for all } t \geq \tau(\epsilon, b, \omega).
\]
Since $\epsilon$ is arbitrary, we can choose it as $1/M$, giving the desired inequality for all $\omega \in A$. Now for each choice of integer $M$, let $A(M)$ be the set of probability 1 above. The intersection of these sets also has probability 1, and each $\omega$ in all of these sets has $\lim_{t} N(t, \omega)/t = 0$. If you did this correctly, you should surely be proud of yourself!!!

**Exercise 5.12:** Consider a variation of an $M/G/1$ queueing system in which there is no facility to save waiting customers. Assume customers arrive according to a Poisson process of rate $\lambda$. If the server is busy, the customer departs and is lost forever; if the server is not busy, the customer enters service with a service time CDF denoted by $F_{Y}(y)$.

Successive service times (for those customers that are served) are IID and independent of arrival times. Assume that customer number 0 arrives and enters service at time $t = 0$.

a) Show that the sequence of times $S_{1}, S_{2}, \ldots$ at which successive customers enter service are the renewal times of a renewal process. Show that each inter-renewal interval $X_{i} = S_{i} - S_{i-1}$ (where $S_{0} = 0$) is the sum of two independent random variables, $Y_{i} + U_{i}$ where $Y_{i}$ is the $i$th service time; find the probability density of $U_{i}$.

**Solution:** Let $Y_{1}$ be the first service time, i.e., the time spent serving customer 0. Customers who arrive during $(0, Y_{1}]$ are lost, and, given that $Y_{1} = y$, the residual time until the next customer arrives is memoryless and exponential with rate $\lambda$. Thus the time $X_{1} = S_{1}$ at which the next customer enters service is $Y_{1} + U_{1}$ where $U_{1}$ is exponential with rate $\lambda$, i.e., $f_{U_{1}}(u) = \lambda \exp(-\lambda u)$.

At time $X_{1}$, the arriving customer enters service, customers are dropped until $X_{1} + Y_{2}$, and after an exponential interval $U_{2}$ of rate $\lambda$ a new customer enters service at time $X_{1} + X_{2}$ where $X_{2} = Y_{2} + U_{2}$. Both $Y_{2}$ and $U_{2}$ are independent of $X_{1}$, so $X_{2}$ and $X_{1}$ are independent. Since the $Y_{i}$ are IID and the $U_{i}$ are IID, $X_{1}$ and $X_{2}$ are IID. In the same way, the sequence $X_{1}, X_{2}, \ldots$ are IID intervals between successive services. Thus $\{X_{i}; i \geq 1\}$ is a sequence of inter-renewal intervals for a renewal process and $S_{1}, S_{2}, \ldots$ are the renewal epochs.

b) Assume that a reward (actually a cost in this case) of one unit is incurred for each customer turned away. Sketch the expected reward function as a function of time for the sample function of inter-renewal intervals and service intervals shown below; the expectation is to be taken over those (unshown) arrivals of customers that must be turned away.

<table>
<thead>
<tr>
<th>S₀ = 0</th>
<th>S₁</th>
<th>S₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y₁ →</td>
<td>Y₂ →</td>
<td>Y₃ →</td>
</tr>
</tbody>
</table>

**Solution:** Customers are turned away at rate $\lambda$ during the service times, so that $R(t)$, the reward (the reward at time $t$ averaged over dropped customer arrivals but for a given sample path of services and residual times) is given by $R(t) = \lambda$ for $t$ in a service interval and $R(t) = 0$ otherwise.

<table>
<thead>
<tr>
<th>S₀ = 0</th>
<th>S₁</th>
<th>S₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$ → Y₁ →</td>
<td>Y₂ →</td>
<td>Y₃ →</td>
</tr>
</tbody>
</table>

Note that the number of arrivals within a service interval are dependent on the length of the service interval but independent of arrivals outside of that interval and independent of other service intervals.
c) Let $\int_0^t R(\tau)\,d\tau$ denote the accumulated reward (i.e., cost) from 0 to $t$ and find the limit as $t \to \infty$ of $(1/t) \int_0^t R(\tau)\,d\tau$. Explain (without any attempt to be rigorous or formal) why this limit exists with probability 1.

**Solution:** The reward within the $n$th inter-renewal interval (as a random variable over that interval) is $R_n = \lambda Y_n$. Using Theorem 4.4.1, then, the sample average reward, WP1, is $\frac{\lambda E[Y]}{E[Y] + 1/\lambda}$.

d) In the limit of large $t$, find the expected reward from time $t$ until the next renewal. Hint: Sketch this expected reward as a function of $t$ for a given sample of inter-renewal intervals and service intervals; then find the time average.

**Solution:** For the sample function above, the reward to the next inter-renewal (again averaged over dropped arrivals) is given by

![Diagram showing reward over inter-renewal intervals]

The reward over the $n$th inter-renewal interval is then $\lambda S_n^2/2$ so the sample path average of the expected reward per unit time is

$$\frac{E[R(t)]}{X} = \frac{\lambda E[Y^2]}{2(Y + 1/\lambda)}.$$

e) Now assume that the arrivals are deterministic, with the first arrival at time 0 and the $n$th arrival at time $n - 1$. Does the sequence of times $S_1, S_2, \ldots$ at which subsequent customers start service still constitute the renewal times of a renewal process? Draw a sketch of arrivals, departures, and service time intervals. Again find $\lim_{t \to \infty} \left( \int_0^t R(\tau)\,d\tau \right) / t$.

**Solution:** Since the arrivals are deterministic at unit intervals, the packet to be served at the completion of $Y_1$ is the packet arriving at $[Y_1]$ (the problem statement was not sufficiently precise to specify what happens if a service completion and a customer arrival are simultaneous, so we assume here that such a customer is served). The customers arriving from time 1 to $[Y_1] - 1$ are then dropped as illustrated below.

![Diagram showing deterministic arrivals and departures]

Finding the sample path average as before,

$$\frac{E[R(t)]}{X} = \frac{E[Y] - 1}{E[[Y]]}.$$

**Exercise 5.13:** Let $Z(t) = t - S_{N(t)}$ be the age of a renewal process and $Y(t) = S_{N(t)+1} - t$ be the residual life. Let $F_X(x)$ be the CDF of the inter-renewal interval and find the following as a function of $F_X(x)$:
a) \( \Pr\{Y(t) > x \mid Z(t) = s\} \)

b) \( \Pr\{Y(t) > x \mid Z(t+x/2) = s\} \)

Solution: If \( s \geq x/2 \) and if, to be specific, we take \( N(t) = n \), then the condition \( \{Z(t + \frac{x}{2}) = s, N(t) = n\} \) means that both \( S_n = t - s + \frac{x}{2} \) and \( S_{n+1} > t + \frac{x}{2} \), i.e., \( \{Z(t + \frac{x}{2}) = s, N(t) = n\} = \{S_n = t - s + \frac{x}{2}, X_{n+1} > s\} \).

The condition \( Y(t) > x \) can be translated under these conditions to \( X_{n+1} > x + s \).

c) \( \Pr\{Y(t) > x \mid Z(t+x) > s\} \) for a Poisson process.

Exercise 5.14: Let \( F_Z(z) \) be the fraction of time (over the limiting interval \( (0, \infty) \)) that the age of a renewal process is at most \( z \). Show that \( F_Z(z) \) satisfies

\[
F_Z(z) = \frac{1}{\bar{X}} \int_{x=0}^{y} \Pr\{X > x\} \, dx \quad \text{WP1.}
\]

Hint: Follow the argument in Example 5.4.5.

Exercise 5.17: A gambler with an initial finite capital of \( d > 0 \) dollars starts to play a dollar slot machine. At each play, either his dollar is lost or is returned with some additional number of dollars. Let \( X_i \) be his change of capital on the \( i \)th play. Assume that \( \{X_i; i=1,2,\ldots\} \) is a set of IID random variables taking on integer values \( \{-1,0,1,\ldots\} \). Assume that \( \mathbb{E}[X_i] < 0 \). The gambler plays until losing all his money (i.e., the initial \( d \) dollars plus subsequent winnings).

a) Let \( J \) be the number of plays until the gambler loses all his money. Is the weak law of large numbers sufficient to argue that \( \lim_{n \to \infty} \Pr\{J > n\} = 0 \) (i.e., that \( J \) is a random variable) or is the strong law necessary?

Solution: We show below that the weak law is sufficient. The event \( \{J > n\} \) is the event that \( S_i > -d \) for \( 1 \leq i \leq n \). Thus \( \Pr\{J > n\} \leq \Pr\{S_n > -d\} \). Since \( \mathbb{E}[S_n] = n\bar{X} \) and \( \bar{X} < 0 \), we see that the event \( \{S_n > -d\} \) for large \( n \) is an event in which \( S_n \) is very far above its mean. Putting this event in terms of distance from the sample average to the mean,

\[
\Pr\{S_n > -d\} = \Pr\left\{ \frac{S_n}{n} - \bar{X} > \frac{-d}{n} - \bar{X} \right\}.
\]

The WLLN says that \( \lim_{n \to \infty} \Pr\left\{ \left| \frac{S_n}{n} - \bar{X} \right| > \epsilon \right\} = 0 \) for all \( \epsilon > 0 \), and this also implies that \( \Pr\left\{ \frac{S_n}{n} - \bar{X} > \epsilon \right\} = 0 \) for all \( \epsilon > 0 \). If we choose \( \epsilon = -\bar{X}/2 \) for the equation above, then \( \epsilon \geq -d/n - \bar{X} \) for \( n \geq -d/2\epsilon \), and thus

\[
\lim_{n \to \infty} \Pr\{S_n > -d\} \leq \lim_{n \to \infty} \Pr\left\{ \frac{S_n}{n} - \bar{X} > \epsilon \right\} = 0.
\]

b) Find \( \mathbb{E}[J] \). Hint: The fact that there is only one possible negative outcome is important here.

Solution: b) One stops playing on trial \( J = n \) if one’s capital reaches 0 for the first time on the \( n \)th trial, i.e., if \( S_n = -d \) for the first time at trial \( n \). This is clearly a function of \( X_1, X_2, \ldots, X_n \), so \( J \) is a stopping rule. Note that stopping occurs exactly on reaching \( -d \) since \( S_n \) can decrease with \( n \) only in increments of -1 and \( S_n \) is always integer. Thus \( S_J = -d \).
Using Wald’s equality, which is valid since $E[J] < \infty$, we have
\[ E[J] = -\frac{d}{X}, \]
which is positive since $X$ is negative. You should note from the exercises we have done with Wald’s equality that it is often used to solve for $E[J]$ after determining $E[S_J]$.

**Exercise 5.19:** Let $J = \min\{n \mid S_n \leq b \text{ or } S_n \geq a\}$, where $a$ is a positive integer, $b$ is a negative integer, and $S_n = X_1 + X_2 + \cdots + X_n$. Assume that $\{X_i; i \geq 1\}$ is a set of zero-mean IID rv’s that can take on only the set of values $\{-1, 0, +1\}$, each with positive probability.

**a)** Is $J$ a stopping rule? Why or why not? Hint: The more difficult part of this is to argue that $J$ is a random variable (i.e., non-defective); you do not need to construct a proof of this, but try to argue why it must be true.

**Solution:** For $J$ to be a stopping trial, it must be a random variable and also $\mathbb{I}_{J=n}$ must be a function of $X_1, \ldots, X_n$. Now $S_n$ is a function of $X_1, \ldots, X_n$, so the event that $S_n \geq a$ or $S_n \leq b$ is a function of $S_n$ and the first $n$ at which this occurs is a function of $S_1, \ldots, S_n$. Thus $\mathbb{I}_{J=n}$ must be a function of $X_1, \ldots, X_n$. For $J$ to be a rv, we must show that $\lim_{n \to \infty} \Pr\{J \leq n\} = 1$ The central limit theorem states that $(S_n - n\overline{X})/\sqrt{n\sigma}$ approaches a normalized Gaussian rv in distribution as $n \to \infty$. Since $\overline{X} = 0$, $S_n/\sqrt{n\sigma}$ must approach normal. Now both $a/\sqrt{n\sigma}$ and $b/\sqrt{n\sigma}$ approach 0, so the probability that $S_n$ (i.e., the process without stopping) is between these limits goes to 0 as $n \to \infty$. Thus the probability that the process has not stopped by time $n$ goes to 0 as $n \to \infty$.

An alternate approach here is to model $\{S_i; n \geq 1\}$ for the stopped process as a Markov chain where $a$ and $b$ are recurrent states and the other states are transient. Then we know that one of the recurrent states are reached eventually with probability 1.

**b)** What are the possible values of $S_J$?

**Solution:** Since $S_n$ can change only in integer steps, it cannot exceed $a$ without first equaling $a$ and it cannot be less than $b$ without first equaling $b$. Thus $S_J$ is only $a$ or $b$.

**c)** Find an expression for $E[S_J]$ in terms of $p$, $a$, and $b$, where $p = \Pr\{S_J \geq a\}$.

**Solution:** $E[S_J] = a \Pr\{S_J = a\} + b \Pr\{S_J = b\} = pa + (1 - p)b$.

**d)** Find an expression for $E[S_J]$ from Wald’s equality. Use this to solve for $p$.

**Solution:** Since $J$ is a stopping trial for $X_1, X_2, \ldots$ and the $X_i$ are IID, we have $E[S_J] = \overline{X}E[J]$. Since $\overline{X} = 0$, we conclude that $E[S_J] = 0$. Combining this with part c), we have $0 = pa + (1 - p)b$, so $p = -b/(a - b)$. This is easier to interpret as $p = |b|/(a + |b|)$.

This approach works only for $\overline{X} = 0$, but we will see in Chapter 9 how to solve the problem for an arbitrary distribution on $X$. We also note that the solution is independent of the probability of the self loop. Finally we note that this helps explain the peculiar behavior of the ‘stop when you’re ahead’ example. The negative threshold $b$ represents the capital of the gambler in that example and shows that as $b \to -\infty$, the probability of reaching the threshold $a$ increases, but at the expense of a larger catastrophe if the gamblers capital is wiped out.
Exercise 5.21: Consider a \( G/G/1 \) queue with inter-arrival sequence \( \{X_i; i \geq 1\} \) and service time sequence \( \{V_i; i \geq 0\} \). Assume that \( \mathbb{E}[X] < \infty \) and that \( \mathbb{E}[V] < \mathbb{E}[X] \).

a) Let \( Z_n = \sum_{i=1}^{n} (X_i - V_{i-1}) \) and let \( I_n \) be the sum of the intervals between 0 and \( S_n = \sum_{i=1}^{n} X_i \) during which the server is empty. Show that \( I_n \geq Z_n \) and explain the difference between them. Show that if arrival \( n \) enters an empty system, then \( I_n = Z_n \).

**Solution:** \( \sum_{i=0}^{n-1} V_i \) is the aggregate time required by the server to serve customers 0 to \( n-1 \). The amount of time the server has been busy is equal to this sum less the unfinished work still in the queue and server. The aggregate time the server has been idle over \( (0, S_n] \) is thus \( Z_n \) plus the unfinished work. If customer \( n \) enters an empty system, then there is no unfinished work and \( I_n = Z_n \).

b) Assume that the inter-arrival distribution is bounded in the sense that for some \( b < \infty \), \( \Pr\{X > b\} = 0 \). Show that the duration of each idle period for the \( G/G/1 \) queue has a duration less than or equal to \( b \).

**Solution:** Note that \( S_n - S_{n-1} \leq b \). The server was busy just after the arrival at \( S_{n-1} \) and thus the idle period was at most \( b \).

c) Show that the number of idle intervals for the server up to time \( S_n \) is at least \( Z_n/b \). Show that the number of renewals up to time \( S_n \) (\( N^r(S_n) \)) at which an arrival sees an empty system is at least \( (Z_n/b) \).

**Solution:** Letting \( M_n \) be the number of idle intervals over \( (0, S_n] \), we see that \( M_n b \geq I_n \geq Z_n \), so \( M_n \geq Z_n/b \). Each idle interval is ended by a customer arrival, so the final idle interval either terminates a renewal before \( S_n \) or terminates a renewal at \( S_n \).

d) Show that the renewal counting process satisfies \( \lim_{t \to \infty} \frac{N^r(t)}{t} = a \) WP1 for some \( a > 0 \). Show that the mean number of arrivals per arrival to an empty system is \( J = 1/a < \infty \).

**Solution:** It was shown in Section 5.5.3 that \( \{N^r(t); t > 0\} \) is a renewal counting process, so \( \lim_{t \to \infty} \frac{N^r(t)}{t} \) exists WP1 and its value \( a \) is the reciprocal of the interval between renewals. We must show that \( a > 0 \). We have

\[
a = \lim_{t \to \infty} \frac{N^r(t)}{t} = \lim_{n \to \infty} \frac{N^r(S_n)}{S_n} = \lim_{n \to \infty} \frac{M_n}{S_n} \\
= \lim_{n \to \infty} \frac{M_n}{n} \lim_{n \to \infty} \frac{n}{S_n} \\
\geq \lim_{n \to \infty} \frac{Z_n}{bn} \lim_{n \to \infty} \frac{n}{S_n} = \frac{\mathbb{E}[X] - \mathbb{E}[V]}{b} \cdot \frac{1}{\mathbb{E}[X]} > 0,
\]

where we have used the SLLN.

e) Assume everything above except that \( \Pr\{X > b\} = 0 \) for some \( b \). Use the truncation method on \( X \) to establish part d) for this new case.

**Solution:** Let \( \mathbb{E}[X] - \mathbb{E}[V] = \epsilon \) and choose \( b \) large enough that when we choose \( \bar{X} = \min\{X, b\} \), we have \( \mathbb{E}[\bar{X}] - \mathbb{E}[V] \geq \epsilon/2 \). Now compare an arbitrary sample function of \( \{X_i; i \geq 1\}, \{V_i; i \geq 0\} \) with the truncated version \( \{\bar{X}_i; i \geq 1\}, \{\bar{V}_i; i \geq 0\} \). For each \( n \) for which the sample values \( \bar{X}_n = \bar{x}_n < X_n - x_n \), we see that the unfinished work in the truncated system at \( \bar{S}_n = \bar{s}_n \) increases relative to \( S_n = s_n \). When the original system has a period where the server is idle, the unfinished work in the truncated system decreases
toward 0 while that in the original system remains equal to 0. Thus, if the truncated system has an idle period in the interval before $\tilde{S}_n$, the untruncated system has an idle period there also. In other words, the number of idle intervals in the truncated system up to $\tilde{S}_n$ must be less than or equal to the number of idle periods in the original system up to $S_n$. Thus $N'(\tilde{S}_n) \leq N'(S_n)$ for all $n$. Since the rate of arrivals seeing an empty system in the truncated version exceeds 0, that in the original system does also.

**Exercise 5.23:** a) Consider a renewal process for which the inter-renewal intervals have the PMF $p_X(1) = p_X(2) = 1/2$. Let $A_n = 1$ if an arrival occurs at time $n$ and $A_n = 0$ otherwise. Use elementary combinatorics to find $E[A_n]$ for $1 \leq n \leq 3$. Use this to show that $m(1) = 1/2$, $m(2) = 5/4$, and $m(3) = 15/8$.

**Solution:** Note that $m(n) = \sum_{i=1}^{n} E[A_i]$. Now $A_1 = 1$ if and only if (iff) $X_1 = 1$, so $E[A_1] = 1/2$ and $m(1) = 1/2$. Next $A_2 = 1$ if either $X_1 = 2$ or both $X_1 = 1$ and $X_2 = 1$. Thus $E[A_2] = 3/4$ and $m(2) = 5/4$. To calculate $m(3)$, we get a little fancier and note that if $A_2 = 0$, then $A(3)$ must be 1 and if $A(2) = 1$, then $A(3)$ is 1 with probability 1/2. Thus

$$\Pr\{A_3 = 1\} = \Pr\{A_2 = 0\} + \frac{1}{2} \Pr\{A_2 = 1\} = \frac{5}{8}.$$  


b) Use elementary means to show that $E[S_{N(1)}] = 1/2$ and $E[S_{N(1)+1}] = 9/4$. Verify (5.37) in this case (i.e., for $t = 1$). Also explain why $N(1)$ is not a stopping trial and show that Wald’s equality is not satisfied if one stops at $N(1)$.

**Solution:** Note that $S_{N(1)}$ is the epoch of the last arrival up to and including time 1 (and is 0 if there are no such arrivals). With probability 1/2, there is an arrival at epoch 1, so $S_{N(1)} = 1$ and with probability 1/2, there are no prior arrivals and by convention $S_{N(1)} = 0$. Thus $E[S_{N(1)}] = 1/2$.

The first arrival epoch after time 1, i.e., $S_{N(1)+1}$, is at time 2 if $X_1 = 2$ and is at time 2 or 3 with equal probability if $X_1 = 1$. Thus it is 2 with probability 3/4 and 3 with probability 1/4, so $E[S_{N(1)+1}] = 2.25$.

Eq. (5.37) says that $E[S_{N(1)+1}] = E[X](E[N(1)]+1)$. Since $\bar{X} = 1.5$ and $E[N(1)] = 3/2$, (5.37) is satisfied.

The event $\{N(1) = 0\}$ has non-zero probability but is not determined before the observation of any $X$’s. Thus $N(1)$ cannot be a stopping trial for $\{X_n; n \geq 1\}$. Also, Wald’s equality (if it applied) would equate $E[S_{N(1)}] = 1/2$ with $\bar{X}E[N(1)] = 3/4$, so it is not satisfied.

c) Consider a more general form of part a) where $\Pr\{X = 1\} = 1-p$ and $\Pr\{X = 2\} = p$. Let $\Pr\{W_n = 1\} = x_n$ and show that $x_n$ satisfies the difference equation $x_n = 1 - px_{n-1}$ for $n \geq 1$ where by convention $x_0 = 1$.

Use this to show that

$$x_n = \frac{1 - (-p)^{n+1}}{1 + p}.$$  

From this, solve for $m(n)$ for $n \geq 1$.

**Solution:** Note that if $W_{n-1} = 0$, then, since each interarrival interval is either 1 or 2, we must have $W_n = 1$. Alternatively, if $W_{n-1} = 1$, then $W_n$ is 1 with probability $1-p$. Thus $x_n$ is given by $(1 - x_{n-1}) + (1-p)x_{n-1}$

$$\Pr\{W_n = 1\} = \Pr\{W_{n-1} = 0\} + (1-p)\Pr\{W_{n-1} = 1\}.$$  


In terms of \( x_n = \Pr\{W_n = 1\} \), this is
\[
x_n = (1 - x_{n-1}) + (1 - p)x_{n-1} = 1 - px_{n-1}.
\]
Expanding this expression,
\[
x_n = 1 - p + p^2 - p^3 + \cdots + (-p)^n.
\]
Now for any number \( y \), note that
\[
(1 - y)(1 + y^2 + \cdots + y^n) = 1 - y^{n+1}.
\](A.29)
Taking \(-p\) for \( y \), we get (5.115). We can find \( m(n) = \sum_{i=1}^n x_i \) by using (A.29) to sum the powers of \(-p\) in (A.29). The result is
\[
m(n) = \frac{m}{1 + p} - \frac{p^{2}(1 - (-p)^{n})}{(1 + p)^2}.
\]
With some patience, this agrees with the solution in part a). This same approach, combined with more tools about difference equations, can be used to solve for \( m(n) \) for other integer valued renewal processes. Finally note that we can solve for \( \mathbb{E}[S_{N(n)+1}] \) for all \( n \) by using Wald’s equality.

**Exercise 5.35:** Consider a sequence \( X_1, X_2, \ldots \) of IID binary rv’s with \( \Pr\{X_n = 1\} = p_1 \) and \( \Pr\{X_n = 0\} = p_0 = 1 - p_1 \). A renewal is said to occur at time \( n \geq 2 \) if \( X_{n-1} = 0 \) and \( X_n = 1 \).

a) Show that \( \{N(n); n > 0\} \) is a renewal counting process where \( N(n) \) is the number of renewals up to and including time \( n \).

**Solution:** Let \( Y_1, Y_2, \ldots \) be the intervals between successive occurrences of the pair \((0,1)\). Note that \( Y_1 \geq 2 \) since the first binary rv occurs at time 1. Also \( Y_n \geq 2 \) for each \( n > 1 \), since if there is an occurrence at time \( n \) (i.e., if \( X_{n-1} = 0, X_n = 1 \), then the pair \( X_n = 0, X_{n+1} = 1 \) is impossible. The length of each inter-arrival interval is determined solely by the bits within that interval, which are IID both among themselves and between intervals. Thus the inter-arrival intervals are IID and are thus renewals, so the counting process is also a renewal process.

b) What is the probability that a renewal occurs at time \( n, n \geq 2 \)?

**Solution:** The probability of a renewal at time \( n \geq 2 \) is \( \Pr\{X_{n-1} = 0, X_n = 1\} = p_0p_1 \).

c) Find the expected inter-renewal interval; use Blackwell’s theorem.

**Solution:** The inter-renewal distribution is arithmetic with span 1 and the expected number of renewals in the interval \([n, n + 1]\) is simply the probability of a renewal at time \( n + 1 \), i.e., \( p_0p_1 \). Thus, from Blackwell’s theorem,
\[
\mathbb{E}[N(n+1)] - \mathbb{E}[N(n)] = p_0p_1 = \frac{1}{\mathbb{E}[Y]}.
\]
Thus the expected inter-renewal interval is \( 1/p_0p_1 \); this is also the expected time until the first occurrence of \((0,1)\). This method of using Blackwell’s theorem gets increasingly useful as the string of interest becomes longer.
d) Now change the definition of renewal; a renewal now occurs at time \( n \) if \( X_{n-1} = 1 \) and \( X_n = 1 \). Show that \( \{N^*(n); n \geq 0\} \) is a delayed renewal counting process where \( N^*_n \) is the number of renewals up to and including \( n \) for this new definition of renewal.

**Solution:** Let \( Y_1, Y_2, \ldots \) now be the inter-arrival intervals between successive occurrences of the pair \( (1,1) \). Note that \( Y_1 \geq 2 \) as before since the first binary rv occurs at time 1, so the string \( (1,1) \) cannot appear until time 2. In this case, however, it is possible to have \( Y_i = 1 \) for \( i \geq 2 \) since if \( (X_{m-1}, X_m, X_{m+1}) = (1,1,1) \), then a renewal occurs at time \( m \) and also at \( m + 1 \).

Since successive occurrences of \( (1,1) \) can overlap, we must be a little careful to show that \( Y_1, Y_2, \ldots \) are independent. Given \( Y_1, \ldots, Y_i \), we see that \( Y_{i+1} = 1 \) with probability \( p_1 \), and this is independent of \( Y_1, \ldots, Y_i \), so the event \( \{Y_{i+1} = 1\} \) is independent of \( Y_1, \ldots, Y_i \). In the same way \( \{Y_{i+1} = k\} \) is independent of \( Y_1, \ldots, Y_i \), so \( Y_{i+1} = 1 \) is independent of \( Y_1, \ldots, Y_i \) for all \( i \geq 2 \), establishing the independence. The fact that \( Y_2, \ldots \) are equally distributed is obvious as before. Thus \( N^* \) is a delayed renewal process.

e) Find \( E[Y_i] \) for \( i \geq 2 \) for the case in part d).

**Solution:** Using the same argument as in part c), \( E[Y_i] = \frac{1}{p_1} \) for \( i \geq 2 \).

f) Find \( E[Y_1] \) for the case in part d), Hint: Show that \( E[Y_1|X_1 = 1] = 1 + E[Y_2] \) and \( E[Y_1|X_1 = 0] = 1 + E[Y_1] \).

**Solution:** If \( X_1 = 1 \), then the wait for a renewal, starting at time 1, has the same probabilities as the wait for a renewal starting at the previous renewal. If \( X_1 = 0 \), then the wait, starting at time 1 is the same as it was before. Thus

\[
E[Y_1] = 1 + p_1 E[Y_2] + p_0 E[Y_1].
\]

Combining the terms in \( E[Y_1] \), we have \( p_1 E[Y_1] = 1 + p_1 E[Y_0] \), so

\[
E[Y_1] = \frac{1}{p_1} + E[Y_2] = \frac{1}{p_1} + \frac{1}{p_1^2}.
\]

g) Looking at your results above for the strings \((0,1)\) and \((1,1)\), show that for an arbitrary string \( a = (a_1, \ldots, a_k) \), the arrival process of successive occurrences of the string is a renewal process if no proper suffix of \( a \) is a prefix of \( a \). Otherwise it is a delayed renewal process.

**Solution:** For the 2 strings looked at, \((0,1)\) has the property that no proper suffix is a prefix, and \((1,1)\) has the property that the suffix \((1)\) is equal to the prefix \((1)\). In general, if there is a proper suffix of length \( i \) equal to a prefix of length \( i \), then, when the string occurs, the final \( i \) bits of the string provide a prefix for the next string, which can then have a length as small as \( k - i \). Since the first occurrence of the string can not appear until time \( k \), The arrival process is a delayed renewal process, where the independence and the identical distribution after the first occurrence follow the arguments for the strings of length 1 above. If no proper suffix of \( a \) is equal to a prefix of \( a \), then the end of one string provides no beginning for a new string and the first renewal is statistically identical to the future renewals; i.e., the process is a renewal process.
h) Suppose a string \( a = (a_1, \ldots, a_k) \) of length \( k \) has no proper suffixes equal to a prefix. Show that the time to the first renewal satisfies

\[
E[Y_1] = \frac{1}{\prod_{t=1}^{k} p_{a_t}}.
\]

**Solution:** The successive occurrences of \( a \) form a renewal process from part g). The result then arises from Blackwell’s theorem as in the earlier parts.

i) Suppose the string \( a = (a_1, \ldots, a_k) \) has at least one proper suffix equal to a prefix, and suppose \( i \) is the length of the longest such suffix. Show that the expected time until the first occurrence of \( a \) is given by

\[
E[Y_i] = \frac{1}{\prod_{t=1}^{i} p_{a_t}} + E[U_i],
\]

where \( E[U_i] \) is the expected time until the first occurrence of the string \( (a_1, \ldots, a_i) \).

**Solution:** The first term above is the expected inter-renewal interval, which follows from Blackwell. If a renewal occurs at time \( n \), then, since \((a_1, \ldots, a_i)\) is both a prefix and suffix of \( a \), we must have \((X_{n-i+1}, \ldots, X_n) = (a_1, \ldots, a_i)\). This means that an inter-renewal interval is the time from an occurrence of \((a_1, \ldots, a_i)\) until the first occurrence of \((a_1, \ldots, a_k)\).

Starting from time 0, it is necessary for \((a_1, \ldots, a_i)\) to occur before \((a_1, \ldots, a_k)\), so we can characterize the time until the first occurrence of \( a \) as the sum of the time to the first occurrence of \((a_1, \ldots, a_i)\) plus the time from that first occurrence to the first occurrence of \((a_1, \ldots, a_k)\); as explained in greater detail in Example 4.5.1, this is statistically the same as an inter-renewal time.

j) Show that the expected time until the first occurrence of \( a = (a_1, \ldots, a_k) \) is given by

\[
E[Y_1] = \sum_{i=1}^{k} \frac{l_i}{\prod_{t=1}^{i} p_{a_t}},
\]

where, for \( 1 \leq i \leq k \), \( l_i \) is 1 if the prefix of \( a \) of length \( i \) is equal to the suffix of length \( i \). Hint: Use part h) recursively. Also show that if \( a \) has a suffix of length \( i \) equal to the prefix of length \( i \) and also a suffix of length \( j \) equal to a prefix of length \( j \) where \( j < i \), then the suffix of \((a_1, \ldots, a_i)\) of length \( j \) is also equal to the prefix of both \( a \) and \((a_1, \ldots, a_i)\) of length \( j \).

**Solution:** Note that the suffix of \( a \) of length \( k \) is not a proper suffix and is simply equal to \( a \) itself and thus also equal to the prefix of \( a \) of length \( k \). This is the final term of the expression above and also the first term in part h).

The next smaller value of \( i \) for which \( l_i = 1 \) is the largest \( i \) for which a proper suffix of \( a \) of length \( i \) is equal to a prefix. If we apply part h) to finding the expected first occurrence of \((a_1, \ldots, a_i)\) for that \( i \), we see that it is the next to final non-zero term in the above sum plus the expected time for the first occurrence of the largest \( j \) for which the proper suffix of \((a_1, \ldots, a_i)\) is equal to the prefix of \((a_1, \ldots, a_i)\) of length \( j \). Now the first \( j \) bits of \((a_1, \ldots, a_i)\) are also the first \( j \) bit of \( a \). Also, since \((a_1, \ldots, a_i) = (a_{k-i+1}, \ldots, a_k)\), we see that the last \( j \) bits of \((a_1, \ldots, a_i)\) are also the last \( j \) bits of \( a \). Thus \( j \) is the third from last value for which \( l_i = 1 \).

This argument continues (recursively) through all the terms in the above sum. We can also interpret the partial sums above as the expected times of occurrence for each prefix that is equal to a suffix.
APPENDIX A. SOLUTIONS TO SELECTED EXERCISES

k) Use part i) to find, first, the expected time until the first occurrence of (1,1,1,1,1,1,0) and, second, that of (1,1,1,1,1). Use (4.31) to check the relationship between these answers.

Solution: The first string has no proper suffixes equal to a prefix, so the expected time is \( p_1^{-6} p_0^{-1} \). For the second string, all suffixes are also prefixes, so the expected time is

\[
\sum_{j=1}^{6} p_1^{-j} = p_1^{-6} (1 + p_1 + \cdots + p_1^5) = p_1^{-6} (1 - p_1^6) / (1 - p_1)
\]

\[= p_1^{-6} p_0^{-1} - p_0^{-1}.\]

Adapting (4.31) to the notation here,

\[
E[\text{time to 111111}] = 1 + p_1 E[\text{time to 111111}] + p_0 E[\text{time to 1111110}]
\]

\[p_0 E[\text{time to 111111}] = 1 + p_0 E[\text{time to 1111110}],\]

which checks with the answer above.

**Exercise 5.46:** Consider a ferry that carries cars across a river. The ferry holds an integer number \( k \) of cars and departs the dock when full. At that time, a new ferry immediately appears and begins loading newly arriving cars ad infinitum. The ferry business has been good, but customers complain about the long wait for the ferry to fill up.

a) Assume that cars arrive according to a renewal process. The IID interarrival times have mean \( \bar{X} \), variance \( \sigma^2 \) and moment generating function \( g_X(r) \). Does the sequence of departure times of the ferries form a renewal process? Explain carefully.

Solution: Yes, the ferry departure times form a renewal process. The reason is that the \( \ell \)th ferry departure is immediately after the \( k\ell \)th customer arrival. The time from the \( \ell \)th to \( \ell + 1 \) ferry departure is the time from the \((k\ell + 1)\)th to \(((k + 1)\ell)\)th customer arrival, which is clearly independent of all previous ferry departure times.

b) Find the expected time that a customer waits, starting from its arrival at the ferry terminal and ending at the departure of its ferry. Note 1: Part of the problem here is to give a reasonable definition of the expected customer waiting time. Note 2: It might be useful to consider \( k = 1 \) and \( k = 2 \) first.

Solution: For \( k = 1 \), the ferry leaves immediately when a customer arrives, so the expected waiting time for each customer is 0. For \( k = 2 \), odd numbered customers wait for the following even numbered customer, and even number customers don’t wait at all, so the average waiting time over customers (which we take as the only sensible definition of expected waiting time) is \( \bar{X}/2 \).

We next find the expected waiting time, averaged over customers, for the \( \ell \)th ferry. To simplify notation, we look at the first ferry. Later ferries could be shown to be the same by renewal reward theory, but more simply they are the same simply by adding extra notation. The average expected wait over the \( k \) customers is the sum of their expected waits divided by \( k \) (recall that this is true even if (as here) the wait of different customers are statistically dependent. The expected wait of customer 1 is \((k - 1)\bar{X}\), that of customer 2 \((k - 2)\bar{X}\), etc. Recall (or derive) that \( 1 + 2 + \cdots + (k - 1) = (k - 1)k/2 \). Thus the expected wait per customer is \((k - 1)\bar{X}/2\), which checks with the result for \( k = 1 \) and \( k = 2 \).
c) Is there a ‘slow truck’ phenomenon (a dependence on $E[X^2]$) here? Give an intuitive explanation. Hint: Look at $k = 1$ and $k = 2$ again.

**Solution:** Clearly, there is no ‘slow truck’ phenomenon for the ferry wait here since the answer depends only on $k$ and $X$. The reason is most evident for $k = 1$, where the wait is 0. The arrival of a car at the ferry terminal could have delayed by an arbitrary amount by a slow truck in *getting to* the ferry terminal, but is not delayed at all in *getting on* the ferry since the slow truck took an earlier ferry. For larger $k$, a vehicle could be delayed by a later arriving truck, but at most $k - 1$ vehicles could be delayed that way, while the $E[X^2]$ effect arises from the potentially unbounded number of customers delayed by a slow truck.

d) In an effort to decrease waiting, the ferry managers institute a policy where no customer ever has to wait more than one hour. Thus, the first customer to arrive after a ferry departure waits for either one hour or the time at which the ferry is full, whichever comes first, and then the ferry leaves and a new ferry starts to accumulate new customers. Does the sequence of ferry departures form a renewal process under this new system? Does the sequence of times at which each successive empty ferry is entered by its first customer form a renewal process? You can assume here that $t = 0$ is the time of the first arrival to the first ferry. Explain carefully.

**Solution:** The sequence of ferry departures does not form a renewal process (nor a delayed renewal process). As an example, suppose that $k = 2$ and $X$ is 1 minute with probability $9/10$ and 2.1 hours with probability $1/10$. Then a ferry interdeparture interval of 2 minutes followed by one of an hour and one minute implies that the next departure is more than an hour away. The times at which each successive ferry is entered by its first customer is a renewal process since customer arrivals form a renewal process.

e) Give an expression for the expected waiting time of the first new customer to enter an empty ferry under this new strategy.

**Solution:** The first new customer waits for a time $T$ in hours which is the minimum of 1 hour and a rv with the CDF of $S_{k-1}$. Since the expected value of a rv is the integral of its complementary CDF,

$$E[T] = \int_0^1 [1 - F_{S_{k-1}}(s)] \, ds.$$ 

**Exercise 5.47:** Consider a (G/G/∞) ’queueing’ system. That is, the arriving customers form a renewal process, i.e., the interarrival intervals $\{X_n; n \geq 1\}$ are IID. You may assume throughout that $E[X] < \infty$. Each arrival immediately enters service; there are infinitely many servers, so one is immediately available for each arrival. The service time $Y_i$ of the $i$th customer is a rv of expected value $\bar{Y} < \infty$ and is IID with all other service times and independent of all inter-arrival times. There is no queueing in such a system, and one can easily intuit that the number in service never becomes infinite since there are always available servers.
a) Give a simple example of distributions for $X$ and $Y$ in which this system never becomes empty. Hint: Deterministic rv's are fair game.

Solution: Suppose, for example that the arrivals are deterministic, one arrival each second. Suppose that each service takes 2 second. Then when the second arrival occurs, the first service is only half over. When the third arrival arrives, the first service is just finishing, but the second arrival is only half finished, etc. This also illustrates the unusual behavior of systems with an unlimited number of servers. If the arrival rate is increased, there are simply more servers at work simultaneously; there is no possibility of the system becoming overloaded. The example here can be made less trivial by making the interarrival distribution arbitrary but upper bounded by 1 second. The service distribution could similarly be arbitrary but lower bounded by 2 seconds.

b) We want to prove Little’s theorem for this type of system, but there are no renewal instants for the entire system. As illustrated above, let $N(t)$ be the renewal counting process for the arriving customers and let $L(t)$ be the number of customers in the system (i.e., receiving service) at time $t$. In distinction to our usual view of queueing systems, assume that there is no arrival at time 0 and the first arrival occurs at time $S_1 = X_1$. The $n$th arrival occurs at time $S_n = X_1 + \cdots + X_n$.

Carefully explain why, for each sample point $\omega$ and each time $t > 0$,

$$
\int_0^t L(t, \omega) \, dt \leq \sum_{i=1}^{N(t, \omega)} Y_i(\omega).
$$

Solution: For the case in the above figure, $\int_0^t L(t, \omega) \, dt$ is $Y_1$ plus $Y_2$ plus the part of $Y_3$ on which service has been completed by $t$. Thus $\int_0^t L(t, \omega) \, dt \leq Y_1 + Y_2 + Y_3$. Since $N(t) = 3$ for this example, $\int_0^t L(t, \omega) \, dt \leq \sum_{i=1}^{N(t)} Y_i$.

In general, $\sum_{i=1}^{N(t)} Y_i$ is the total amount of service required by all customers arriving before or at $t$. $\int_0^t L(t, \omega) \, dt$ is the amount of service already provided by $t$. Since not all the required service has necessarily been provided by $t$, the inequality is as given.

c) Find the limit as $t \to \infty$ of $\frac{1}{t} \sum_{i=1}^{N(t, \omega)} Y_i(\omega)$ and show that this limit exists WP1.

Solution: This is virtually the same as many limits we have taken. The trick is to multiply and divide by $N(t, \omega)$ and then use the SLLN and the strong law for renewals, i.e.,

$$
\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{N(t, \omega)} Y_i(\omega) = \lim_{t \to \infty} \frac{N(t, \omega)}{t} \sum_{i=1}^{N(t, \omega)} \frac{Y_i(\omega)}{N(t, \omega)}.
$$

For a given $\omega$, $N(t, \omega)$ is a real valued function of $t$ which approaches $1/X$ for all $\omega$ except perhaps a set of probability 0. For each of those sample points, $\lim_{t \to \infty} N(t, \omega) = \infty$ and
thus
\[
\sum_{i=1}^{N(t,\omega)} Y_i(\omega) \over N(t,\omega) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i(\omega).
\]

Since the \(Y_i\) are IID, the SLLN says that this limit is \(\overline{Y}\) WP1. The set of \(\omega\) for which both the first limit and next exist then has probability 1, so
\[
\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{N(t,\omega)} Y_i(\omega) = \frac{\overline{Y}}{\overline{X}} \quad \text{for a set of } \omega \text{ of probability 1.}
\]

d) Assume that the service time distribution is bounded between 0 and some \(b > 0\), i.e., that \(F_Y(b) = 1\). Carefully explain why
\[
\int_{0}^{t+b} L(\tau, \omega) \, d\tau \geq \sum_{i=1}^{N(t,\omega)} Y_i(\omega).
\]

**Solution:** This is almost the same as part b). All customers that have entered service by \(t\) have completed service by \(t + b\), and thus the difference between the two terms above is the service provided to customers that have arrived between \(t\) and \(t + b\).

e) Find the limit as \(t \to \infty\) of \(\frac{1}{t} \int_{0}^{t} L(\tau, \omega) \, d\tau\) and indicate why this limit exists WP1.

**Solution:** Combining the results of b) and d),
\[
\sum_{i=1}^{N(t,\omega)} Y_i(\omega) \leq \int_{0}^{t+b} L(\tau, \omega) \, d\tau \leq \sum_{i=1}^{N(t+b,\omega)} Y_i(\omega)
\]

Dividing by \(t\) and going to the limit as \(t \to \infty\), the limit on the left is \(\frac{\overline{Y}}{\overline{X}}\) WP1 and the limit on the right, after multiplying and dividing by \(t + b\), is the same. Thus,
\[
\lim_{t \to \infty} \frac{1}{t} \int_{0}^{t+b} L(\tau, \omega) \, d\tau = \lim_{t \to \infty} \frac{t + b}{t + b} \int_{0}^{t+b} L(\tau, \omega) \, d\tau = \frac{\overline{Y}}{\overline{X}}.
\]

This is Little’s theorem.

**Exercise 6.1:** Let \(\{P_{ij} \colon i, j \geq 0\}\) be the set of transition probabilities for a countable-state Markov chain. For each \(i, j\), let \(F_{ij}(n)\) be the probability that state \(j\) occurs sometime between time 1 and \(n\) inclusive, given \(X_0 = i\). For some given \(j\), assume that \(\{x_i \colon i \geq 0\}\) is a set of nonnegative numbers satisfying \(x_i = P_{ij} + \sum_{k \neq j} P_{ik} x_k\) for all \(i \geq 0\). Show that \(x_i \geq F_{ij}(n)\) for all \(n\) and \(i\), and hence that \(x_i \geq F_{ij}(\infty)\) for all \(i\). Hint: use induction.

**Solution:** We use induction on \(n\). As the basis for the induction, we know that \(F_{ij}(1) = P_{ij}\). Since the \(x_i\) are by assumption nonnegative, it follows for all \(i\) that
\[
F_{ij}(1) = P_{ij} \leq P_{ij} + \sum_{k \neq j} P_{kj} x_j = x_i.
\]
Now assume that for a given \( n \geq 1 \), \( F_{ij}(n) \leq x_i \) for all \( i \). Using (6.9),
\[
F_{ij}(n+1) = P_{ij} + \sum_{k \neq j} P_{kj} F_{kj}(n)
\]
\[
\leq P_{ij} + \sum_{k \neq j} P_{kj} x_j = x_i \quad \text{for all } i.
\]
From (6.7), \( F_{ij}(n) \) is non-decreasing in \( n \) and thus has a limit, \( F_{ij}(\infty) \leq x_i \) for all \( i \).

**Exercise 6.2:**

**a)** For the Markov chain in Figure 6.2, show that, for \( p \geq 1/2 \), \( F_{00}(\infty) = 2(1-p) \) and show that \( F_{ij}(\infty) = [(1 - p)/p]^i \) for \( i \geq 1 \). Hint: first show that this solution satisfies (6.9) and then show that (6.9) has no smaller solution (Exercise 6.1 shows that \( F_{ij}(\infty) \) is the smallest solution to (6.9)). Note that you have shown that the chain is transient for \( p > 1/2 \) and that it is recurrent for \( p = 1/2 \).

**Solution:** It may be helpful before verifying these equations to explain where they come from. This derivation essentially solves the problem also, but the verification to follow, using the hint, is valuable to verify the solution, especially for those just becoming familiar with this topic.

First let \( F_{10}(\infty) \), the probability of ever reaching state 0 starting from state 1, be denoted as \( \alpha \). Since \( \alpha \) is determined solely by transitions from states \( i \geq 1 \), and since each state \( i \) “sees” the same Markov chain for states \( j \geq i \), we know that \( F_{i,i-1}(\infty) = \alpha \) for all \( i \geq 1 \).

Thus, using a trick we have used before, \( \alpha = q + pa \).

That is, the probability of ever reaching 0 from 1 is \( q \) (the probability of going there immediately) plus \( p \) (the probability of going to state 2) times \( \alpha \) (the probability of ever getting back to 1 from 2) times \( \alpha \) (the probability of then getting from 1 to 0).

This quadratic has two solutions, \( \alpha = 1 \) and \( \alpha = q/p \). With the help of Exercise 5.1, we know that \( F_{10}(\infty) = q/p \), i.e., the smaller of the two solutions. From this solution, we immediately determine that \( F_{00}(\infty) = q + pa \). Also, for each \( i > 1 \), \( F_{i0}(\infty) \) is the probability of ever moving from \( i \) to \( i - 1 \) times that of ever moving from \( i - 1 \) to \( i - 2 \), etc. down to 1. Thus \( F_{i0}(\infty) = \alpha^i \). We next verify this solution algebraically.

**b)** Under the same conditions as part a), show that \( F_{ij}(\infty) \) equals \( 2(1-p) \) for \( j = i \), equals \( [(1 - p)/p]^i \) for \( i > j \), and equals 1 for \( i < j \).

**Solution:**

**Exercise 6.3 a):** Show that the \( n \)th order transition probabilities, starting in state 0, for the Markov chain in Figure 6.2 satisfy
\[
P^n_{0j} = pP^{n-1}_{0,j-1} + qP^{n-1}_{0,j+1} \quad j \neq 0; \quad P^n_{00} = qP^{n-1}_{00} + qP^{n-1}_{01}.
\]

Hint: Use the Chapman-Kolmogorov equality, (4.7).

**Solution:** This is the Chapman-Kolmogorov equality in the form \( P^n_{ij} = \sum_k P^{n-1}_{ik} P_kj \) where \( P_{kj} = p \) for \( k = j - 1 \), \( P_{kj} = q \) for \( k = j + 1 \) and \( P_{kj} = 0 \) elsewhere.
b) For $p = 1/2$, use this equation to calculate $P_{0j}^n$ iteratively for $n = 1, 2, 3, 4$. Verify (6.3) for $n = 4$ and then use induction to verify (6.3) in general. Note: this becomes an absolute mess for $p \neq 1/2$, so don’t attempt this in general.

Solution: This is less tedious if organized in an array of terms. Each term (except $P_{00}^n$) for each $n$ is then half the term to the upper left plus half the term to the upper right. $P_{00}^n$ is half the term above plus half the term to the upper right.

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>3</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{0j}^1$</td>
<td>1/2</td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{0j}^2$</td>
<td>1/2</td>
<td>1/4</td>
<td>1/4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{0j}^3$</td>
<td>3/8</td>
<td>3/8</td>
<td>1/8</td>
<td>1/8</td>
<td></td>
</tr>
<tr>
<td>$P_{0j}^4$</td>
<td>3/8</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/16</td>
</tr>
</tbody>
</table>

From (6.3), for $n = 4$, $j = 0$, $j + n$ is even, so $P_{00}^4 = \binom{4}{2}2^{-4} = 6/16$, which agrees with the table above. For $j = 1$, $j + n$ is odd, so according to (6.3), $P_{01}^4 = \binom{4}{3}2^{-4} = 4/16$, which again agrees with the table. The other terms are similar. Using induction to validate (6.3) in general, we assume that (6.3) is valid for a given $n$ and all $j$. First, for $j = 0$ and $n$ even, we have

$$P_{00}^{n+1} = \frac{1}{2} [P_{00}^n + P_{00}^n] = \frac{1}{2} \left( \binom{n}{n/2} 2^{-n} + \binom{n}{(n/2)+1} 2^{-n} \right).$$

For arbitrary positive integers $n > k$, a useful combinatorial identity is

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

This can be seen by writing out all the factorials and manipulating the equations, but can be seen more simply by viewing $\binom{n+1}{k+1}$ as the number of ways to arrange $k+1$ ones in a binary $n+1$ tuple. These arrangements can be separated into those that start with a one followed by $k$ ones out of $n$ and those that start with a zero followed by $k+1$ ones out of $n$.

Applying this identity, we have $P_{00}^{n+1} = \binom{n+1}{(n/2)+1} 2^{-(n+1)}$. Since $n + 1 + j$ is odd with $j = 0$, this agrees with (6.3). Next, for $n$ even and $j \geq 2$ even, we have

$$P_{0j}^{n+1} = \frac{1}{2} [P_{0,j-1}^n + P_{0,j+1}^n] = \frac{1}{2} \left[ \binom{n}{(j+n)/2} 2^{-n} + \binom{n}{((j+n)/2)+1} 2^{-n} \right] = \binom{n+1}{((j+n)/2)+1} 2^{-(n+1)}.$$

c) As a more interesting approach, which brings out the relationship of Figures 6.2 and 6.1, note that (6.3), with $j + n$ even, is the probability that $S_n = j$ for the chain in 6.1 and (6.3) with $j + n$ odd is the probability that $S_n = -j - 1$ for the chain in 6.1. By viewing each transition over the self loop at state 0 as a sign reversal for the chain in 6.1, explain why this surprising result is true. (Again, this doesn’t work for $p \neq 1/2$, since the sign reversals also reverse the $+1$, $-1$ transitions.)
Solution: The symmetry to be used here will be more clear if the states are labelled as \( \ldots, -3/2, -1/2, 1/2, 3/2, \ldots \) for the Bernoulli type chain and \( 1/2, 3/2, 5/2, \ldots \) for the M/M/1 type chain.

Now let \( \{X_n; n \geq 0\} \) be the sequence of states for the Bernoulli type chain above and let \( Y_n = |X_n| \) for each \( n \). Since all the transition probabilities are 1/2 above, it can be seen that \( \{Y_n; n \geq 0\} \) is a Markov chain, and is in fact the M/M/1 type chain. That is, except for state 1/2, there is an equiprobable move up by 1 or down by 1. If \( X_n = 1/2 \), then \( X_{n+1} \) is 3/2 or -1/2 with equal probability, so \( |X_n| \) is 1/2 or 3/2 with equal probability.

This means that to find \( P^n_{1/2,j+1/2} \) for the M/M/1 type chain with \( p = 1/2 \), we find \( P^n_{1/2,j+1/2} \) and \( P^n_{1/2,j-1/2} \) for the Bernoulli chain and add them together.

There is one final peculiarity here: the Bernoulli chain is periodic, but each positive state \( j + 1/2 \) is in the opposite subclass from \( -j - 1/2 \), so that only one of these two terms is non-zero for each \( n \).

Exercise 6.8: Let \( \{X_n; n \geq 0\} \) be a branching process with \( X_0 = 1 \). Let \( \bar{Y}, \sigma^2 \) be the mean and variance of the number of offspring of an individual.

a) Argue that \( \lim_{n \to \infty} X_n \) exists with probability 1 and either has the value 0 (with probability \( F_{10}(\infty) \)) or the value \( \infty \) (with probability \( 1 - F_{10}(\infty) \)).

Solution We consider 2 special, rather trivial, cases before considering the important case (the case covered in the text). Let \( p_i \) be the PMF of the number of offspring of each individual. Then if \( p_1 = 1 \), we see that \( X_n = 1 \) for all \( n \), so the statement to be argued is simply false. It is curious that this exercise has been given many times over the years with no one pointing this out.

The next special case is where \( p_0 = 0 \) and \( p_1 < 1 \). Then \( X_{n+1} \geq X_n \) (i.e., the population never shrinks but can grow). Since \( X_n(\omega) \) is non-decreasing for each sample path, either \( \lim_{n \to \infty} X_n(\omega) = \infty \) or \( \lim_{n \to \infty} X_n(\omega) = j \) for some \( j < \infty \). The latter case is impossible, since \( P_{jj} = p_j^1 \) and thus \( P_{jj}^m = p_j^1 \to 0 \).

Ruling out these two trivial cases, we have \( p_0 > 0 \) and \( p_1 < 1 - p_0 \). In this case, state 0 is recurrent (i.e., it is a trapping state) and states 1, 2, \ldots are in a transient class. To see this, note that \( P_{10} = p_0 > 0 \), so \( F_{11}(\infty) \leq 1 - p_0 < 1 \), which means by definition that state 1 is transient. All states \( i > 1 \) communicate with state 1, so by Theorem 6.2.1, all states \( j \geq 1 \) are transient. Thus one can argue that the process has ‘no place to go’ other than 0 or \( \infty \).

The following ugly analysis makes this precise. Note from Theorem 6.3.1 part 4 that

\[
\lim_{t \to \infty} \sum_{n \leq t} P_{jj}^t \neq \infty.
\]
Since this sum is non-decreasing in \(t\), the limit must exist and the limit must be finite. This means that

\[
\lim_{t \to \infty} \sum_{n \geq t} P_{1j}^n = 0.
\]

Now we can write \(P_{1j}^n = \sum_{\ell \leq n} f_{1j}^\ell P_{1j}^{n-\ell}\), from which it can be seen that \(\lim_{t \to \infty} \sum_{n \geq t} P_{1j}^n = 0\).

From this, we see that for every finite integer \(\ell\),

\[
\lim_{t \to \infty} \sum_{n \geq t} \sum_{j=1}^\ell P_{1j}^n = 0.
\]

This says that for every \(\epsilon > 0\), there is a \(t\) sufficiently large that the probability of ever entering states 1 to \(\ell\) or after step \(t\) is less than \(\epsilon\). Since \(\epsilon > 0\) is arbitrary, all sample paths (other than a set of probability 0) never enter states 1 to \(\ell\) after some finite time. Since \(\ell\) is arbitrary, \(\lim_{n \to \infty} X_n\) exists WP1 and is either 0 or \(\infty\). By definition, it is 0 with probability \(F_{10}(\infty)\).

b) Show that \(\text{VAR}[X_n] = \sigma^2 \frac{\Phi^{n-1}}{\Phi^n - 1} \frac{1}{\Phi - 1}\) for \(\Phi \neq 1\) and \(\text{VAR}[X_n] = n \sigma^2\) for \(\Phi = 1\).

**Exercise 7.5:** Consider the Markov process illustrated below. The transitions are labelled by the rate \(q_{ij}\) at which those transitions occur. The process can be viewed as a single server queue where arrivals become increasingly discouraged as the queue lengthens. The word *time-average* below refers to the limiting time-average over each sample-path of the process, except for a set of sample paths of probability 0.

![Diagram](image)

a) Find the time-average fraction of time \(p_i\) spent in each state \(i > 0\) in terms of \(p_0\) and then solve for \(p_0\). Hint: First find an equation relating \(p_i\) to \(p_{i+1}\) for each \(i\). It also may help to recall the power series expansion of \(e^x\).

**Solution:** The \(p_i, i \geq 0\) for a birth-death chain are related by \(p_{i+1}q_{i+1,i} = p_iq_{i,i+1}\), which in this case is \(p_{i+1}\mu = p_i/\lambda(i+1)\). Iterating this equation,

\[
p_i = p_{i-1} \left( \frac{\lambda}{\mu} \right) = p_{i-2} \left( \frac{\lambda^2}{\mu^2 (i-1)} \right) = \cdots = p_0 \left( \frac{\lambda^i}{i!} \right).
\]

Denoting \(\lambda/\mu\) by \(\rho\),

\[
1 = \sum_{i=0}^{\infty} p_i = p_0 \left[ \sum_{i=0}^{\infty} \frac{\rho^i}{i!} \right] = p_0 e^\rho.
\]

Thus,

\[
p_0 = e^{-\rho}; \quad p_i = \frac{\rho^i e^{-\rho}}{i!}.
\]
b) Find a closed form solution to \( \sum_i p_i \nu_i \) where \( \nu_i \) is the rate at which transitions out of state \( i \) occur. Show that the process is positive recurrent for all choices of \( \lambda > 0 \) and \( \mu > 0 \) and explain intuitively why this must be so.

Solution: For the embedded chain, \( P_{01} = 1 \), and for all \( i > 0 \),

\[
P_{i,i+1} = \frac{\lambda}{\lambda + \mu(i+1)}; \quad P_{i,i-1} = \frac{\mu(i+1)}{\lambda + \mu(i+1)}.
\]

All other transition probabilities are 0. The departure rate from state \( i \) is

\[
\nu_i = \lambda; \quad \nu_i = \mu + \frac{\lambda}{i+1} \quad \text{for all } i > 0.
\]

We now calculate \( \sum_j p_j \nu_j \) by separating out the \( j = 0 \) term and then summing separately over the two terms, \( \mu \) and \( \lambda/(j+1) \), of \( \nu_j \).

\[
\sum_{j=0}^{\infty} p_j \nu_j = e^{-\rho} \lambda + \sum_{j=1}^{\infty} e^{-\rho} \frac{\rho^j \mu}{j!} + \sum_{j=1}^{\infty} e^{-\rho} \frac{\rho^j \lambda}{j!(j+1)}.
\]

Substituting \( \mu \rho \) for \( \lambda \) and combining the first and third term,

\[
\sum_{j=0}^{\infty} p_j \nu_j = \sum_{j=1}^{\infty} e^{-\rho} \frac{\rho^j \mu}{j!} + \sum_{j=0}^{\infty} e^{-\rho} \frac{\rho^{j+1} \mu}{(j+1)!} = 2 \sum_{j=1}^{\infty} e^{-\rho} \frac{\rho^j \mu}{j!} = 2 \mu(1 - e^{-\rho}).
\]

Intuitively, we saw that \( P_{i+1} = \frac{\lambda P_i}{\mu(i+1)} \), so \( p_i \) must decrease rapidly in \( i \) for sufficiently large \( i \). Thus the fraction of time spent in very high numbered states must be negligible. This suggests that the steady-state equations for the \( p_i \) must have a solution. Since \( \nu_i \) is bounded between \( \mu \) and \( \mu + \lambda \) for all \( i \), it is intuitively clear that \( \sum_i \nu_i p_i \) is finite, so the embedded chain must be positive recurrent.

c) For the embedded Markov chain corresponding to this process, find the steady-state probabilities \( \pi_i \) for each \( i \geq 0 \) and the transition probabilities \( P_{ij} \) for each \( i, j \).

Solution The steady state probabilities for the embedded chain are then given by

\[
\pi_0 = \frac{\rho}{2(\rho^0 - 1)}; \quad \pi_i = \frac{\rho^i}{2i!(\rho^0 - 1)} \left( \frac{\rho}{i+1} + 1 \right); \quad \text{for } i > 1.
\]

There are many forms for this answer. One sanity check is to observe that the embedded chain probabilities do not change if \( \lambda \) and \( \mu \) are both multiplied by the same constant, and thus the \( \pi_i \) must be a function of \( \rho \) alone. Another sanity check is to observe that in the limit \( \rho \to 0 \), the embedded chain is dominated by an alternation between states 0 and 1, so that in this limit \( \pi_0 = \pi_1 = 1/2 \).

d) For each \( i \), find both the time-average interval and the time-average number of overall state transitions between successive visits to \( i \).
Solution: The time-average interval between visits to state $j$ is $W_j = 1/(\rho_j \nu_j)$. This is explained in detail in section 6.2.3 of the class notes, but the essence of the result is that for renewals at successive entries to state $j$, $\rho_j \nu_j$ must be the ratio of the expected time $1/\nu_j$ spent in state $j$ to the overall expected renewal interval $W_j$. Thus $W_j = 1/(\nu_j \rho_j)$.

$$W_0 = \frac{e^\rho}{\lambda}; \quad W_j = \frac{(j+1)!e^\rho}{\rho^j(\lambda + (j+1)\mu)}.$$  

The time-average number of state transitions per visit to state $j$ is $T_{jj} = 1/\pi_j$. This is proven in Theorem 6.3.5, but the more interesting way to see it here is to use the same argument as used for $W_j$ above. That is, consider the embedded Markov chain as a discrete-time process with one unit of time per transition. Then $\pi_j$ is the ratio of the unit time spent on a visit to $j$ to the expected number of transitions per visit to $j$. Thus

$$T_{00} = \frac{2(e^\rho - 1)}{\rho}; \quad T_{jj} = \frac{2(j+1)!e^\rho}{\rho^j(\rho + j + 1)} \quad \text{for } j \geq 1.$$  

Exercise 7.9: Let $q_{i,i+1} = 2^{i-1}$ for all $i \geq 0$ and let $q_{i,i-1} = 2^{i-1}$ for all $i \geq 1$. All other transition rates are 0.

a) Solve the steady-state equations and show that $p_i = 2^{-i-1}$ for all $i \geq 0$.

b) Find the transition probabilities for the embedded Markov chain and show that the chain is null recurrent.

c) For any state $i$, consider the renewal process for which the Markov process starts in state $i$ and renewals occur on each transition to state $i$. Show that, for each $i \geq 1$, the expected inter-renewal interval is equal to 2. Hint: Use renewal-reward theory.

d) Show that the expected number of transitions between each entry into state $i$ is infinite. Explain why this does not mean that an infinite number of transitions can occur in a finite time.

Solution: We have seen in part b) that the embedded chain is null recurrent. This means that, given $X_0 = i$, for any given $i$, that a return to $i$ must happen in a finite number of transitions (i.e., $\lim_{n \to \infty} F_{ii}(n) = 1$). We have seen many rv’s that have an infinite expectation, but, being rv’s, have a finite sample value WP1.

Exercise 6.15: This exercise explores a continuous time version of a simple branching process.

Consider a population of primitive organisms which do nothing but procreate and die. In particular, the population starts at time 0 with one organism. This organism has an exponentially distributed lifespan $T_0$ with rate $\mu$ (i.e., $\Pr\{T_0 \geq \tau\} = e^{-\mu\tau}$). While this organism is alive, it gives birth to new organisms according to a Poisson process of rate $\lambda$. Each of these new organisms, while alive, gives birth to yet other organisms. The lifespan and birthrate for each of these new organisms are independently and identically distributed to those of the first organism. All these and subsequent organisms give birth and die in the same way, again independently of all other organisms.
a) Let \( X(t) \) be the number of (live) organisms in the population at time \( t \). Show that \( \{X(t); \ t \geq 0\} \) is a Markov process and specify the transition rates between the states.

b) Find the embedded Markov chain \( \{X_n; n \geq 0\} \) corresponding to the Markov process in part a). Find the transition probabilities for this Markov chain.

c) Explain why the Markov process and Markov chain above are not irreducible. Note: The majority of results you have seen for Markov processes assume the process is irreducible, so be careful not to use those results in this exercise.

d) For purposes of analysis, add an additional transition of rate \( \lambda \) from state 0 to state 1. Show that the Markov process and the embedded chain are irreducible. Find the values of \( \lambda \) and \( \mu \) for which the modified chain is positive recurrent, null recurrent, and transient.

e) Assume that \( \lambda < \mu \). Find the steady state process probabilities for the modified Markov process.

f) Find the mean recurrence time between visits to state 0 for the modified Markov process.

g) Find the mean time \( \overline{T} \) for the population in the original Markov process to die out. Note: We have seen before that removing transitions from a Markov chain or process to create a trapping state can make it easier to find mean recurrence times. This is an example of the opposite, where adding an exit from a trapping state makes it easy to find the recurrence time.

Solution: If \( X(t) = j > 0 \), then the state will change at the instant when the first birth or death among the \( j \) living organisms occurs. The first of the \( j \) deaths is a random variable that is exponential with rate \( j\lambda \), and the first birth is independently exponential with rate \( j\mu \). If \( X(t) = 0 \), then no deaths or births can occur at any \( \tau > t \). Thus in general, state transitions are exponential (with no transitions in state 0), and the rate of each transition at time \( t \) depends only on the state at time \( t \). Thus the process is a Markov process.

b) Let \( X_n \) be the state at the end of the \( n \)th transition. Given that \( X_n = j \) for any given \( j > 0 \), \( X_{n+1} \) can be only \( j+1 \) or \( j-1 \). Also

\[
\Pr\{X_{n+1} = j+1 \mid X_n = j\} = \frac{j\lambda}{j\lambda + j\mu} = \frac{\lambda}{\lambda + \mu}.
\]

There are no transitions from state 0 in the Markov process, and thus there are no transitions from state 0 in the embedded chain.

c) State 0 is a trapping state in the Markov chain, \( i.e. \), it communicates with no other state. Thus the Markov chain is not irreducible. By definition, the Markov process is then also not irreducible.

d) With the additional transition from state 0 to 1, there is a path from each state to each other state (both in the process and the embedded chain), and thus both are irreducible. Curiously enough, the embedded chain is now the same as the embedded chain of an M/M/1 queue.

e) The transitions in the Markov process are illustrated below.
Let $q_j$ be the steady-state process probability of state $j > 0$. Then

$$q_1 = \frac{\lambda q_0}{\mu}; \quad q_j = \frac{(j-1)\lambda q_{j-1}}{j\mu} \quad \text{for } j > 1. \tag{A.30}$$

Let $\rho = \lambda/\mu$ (so $0 < \rho < 1$). Iterating (A.30), $q_j$ can be expressed in terms of $q_0$ as

$$q_j = \frac{(j-1)\rho}{j} \frac{(j-2)\rho}{j-1} \ldots \frac{\rho q_0}{1} = \frac{\rho^j q_0}{j}. \tag{A.31}$$

The simplicity of (A.31) is quite surprising, but there is another way to derive the same result that helps explain it. This process has the same embedded chain as the M/M/1 process for which the steady-state process probabilities are $p_j = \rho^j p_0$. For the M/M/1 case, the rate of transitions out of each state $j$ is $\nu_j = \lambda + \mu$, whereas here the corresponding rate is $\nu_j' = j(\lambda + \mu)$. From the relationship between the steady-state probabilities for the embedded chain and the process, however, $p_j = \alpha \pi_j/\nu_j$ and $q_j = \beta \pi_j/\nu_j'$ where $\alpha$ and $\beta$ are normalizing constants. From this, we see that $q_j = (\alpha/\beta)p_j/j = (\alpha/\beta)p_0\rho^j/j$. The normalizing constants can be ignored, since they merely insure that $\sum_j q_j = 1$, and thus this provides another derivation of (A.31).

We still need to solve for $q_0$ to find the steady-state process probabilities. We have

$$1 = \sum_j q_j = q_0 \left[ 1 + \sum_{j=1}^{\infty} \frac{\rho^j}{j} \right] = q_0 \left[ 1 - \ln(1 - \rho) \right].$$

To see the equality of $\sum_{j=1}^{\infty} \frac{\rho^j}{j}$ and $-\ln(1 - \rho)$, note that they agree at $\rho = 0$ and the derivative of each with respect to $\rho$ is $1/(1 - \rho)$. We then have

$$q_0 = \frac{1}{1 - \ln(1 - \rho)}; \quad q_j = \frac{\rho^j}{j[1 - \ln(1 - \rho)]} \quad \text{for } j \geq 1.$$

f) The mean recurrence time between visits to state in the modified Markov process is

$$W_0 = \frac{1}{\nu_0 q_0} = \frac{1 - \ln[1 - \rho]}{\lambda}.$$

g) The mean recurrence time for state 0 consists of two parts — the mean time $(1/\lambda$ to first reach state 1 from state 0 and then the mean time to first reach state 0 again. This latter time is the same whether or not the extra transition from 0 to 1 is present.

**Exercise 7.30:** a) Assume that the Markov process in Exercise 7.5 is changed in the following way: whenever the process enters state 0, the time spent before leaving state 0 is now a uniformly distributed rv, taking values from 0 to $2/\lambda$. All other transitions remain
the same. For this new process, determine whether the successive epochs of entry to state 0 form renewal epochs, whether the successive epochs of exit from state 0 form renewal epochs, and whether the successive entries to any other given state \( i \) form renewal epochs.

**Solution** On each successive entry to state 0, the embedded chain is in state 0 and the duration in that state is a uniform rv independent of everything except the fact of its existence due to entering state 0. This duration in state 0 is finite and IID between successive visits to 0. All subsequent states and durations are independent of those after each subsequent visit to state 0 and the time to return to state 0 is finite with probability 1 from the same argument used for the Markov process. Thus successive returns to state 0 form a renewal process. The same argument works for exits from state 0.

For successive entries to some arbitrary given state \( j \), the argument is slightly more complex, but it also forces us to understand the issue better. The argument that the epochs of successive entries to a given state in a Markov process form a renewal process does not depend at all on the exponential holding times. It depends only on the independence of the holding times and their expected values. Thus changing one holding time from an exponential to a uniform rv of the same expected value changes neither the expected number of transitions per unit time nor the expected time between entries to a given state. Thus successive visits to any given state form renewal epochs.

**b)** For each \( i \), find both the time-average interval and the time-average number of overall state transitions between successive visits to \( i \).

**Solution:** As explained in part a, the expected time between returns to any given state \( j \) is the same as in part c of Exercise 7.5. Since the embedded Markov chain has not changed at all, the expected number of transitions between visits to any given state \( j \) is also given in part c of Exercise 7.5.

**c)** Is this modified process a Markov process in the sense that \( \Pr \{X(t) = i \mid X(\tau) = j, X(s) = k\} = \Pr \{X(t) = i \mid X(\tau) = j\} \) for all \( 0 < s < \tau < t \) and all \( i, j, k \)? Explain.

**Solution:** The modified process is not a Markov process. Essentially, the fact that the holding time in state 0 is no longer exponential, and thus not memoryless, means that if \( X(t) = 0 \), the time at which state 0 is entered provides information beyond knowing \( X(\tau) \) at some given \( \tau < t \)

As an example, suppose that \( \mu << \lambda \) so we can ignore multiple visits to state 0 within an interval of length \( 2/\lambda \). We then see that \( \Pr \{X(t) = 0 \mid X(t - 1/\lambda) = 0, X(t - 2/\lambda) = 0\} = 0 \) since there is 0 probability of a visit to state 0 lasting for more than time \( 2/\lambda \). On the other hand,

\[
\Pr \{X(t) = 0 \mid X(t - 1/\lambda) = 0, X(t - 2/\lambda) = 1\} > 0.
\]

In fact, this latter probability is 1/8 in the limit of large \( t \) and small \( \mu \).

**Exercise 8.15:** Consider a binary hypothesis testing problem where \( X \) is 0 or 1 and a one dimensional observation \( Y \) is given by \( Y = X + U \) where \( U \) is uniformly distributed over \([-1, 1]\) and is independent of \( X \).

**a)** Find \( f_{Y \mid X}(y \mid 0) \), \( f_{Y \mid X}(y \mid 1) \) and the likelihood ratio \( \Lambda(y) \).
Solution: Note that \( f_{Y \mid X} \) is simply the density of \( U \) shifted by \( X \), i.e.,

\[
f_{Y \mid X}(y \mid 0) = \begin{cases} 
1/2; & -1 \leq y \leq 1 \\
0; & \text{elsewhere}
\end{cases} \quad f_{Y \mid X}(y \mid 1) = \begin{cases} 
1/2; & 0 \leq y \leq 2 \\
0; & \text{elsewhere}
\end{cases}
\]

The likelihood ratio \( \Lambda(y) \) is defined only for \(-1 \leq y \leq 2\) since neither conditional density is non-zero outside this range.

\[
\Lambda(y) = \frac{f_{Y \mid X}(y \mid 0)}{f_{Y \mid X}(y \mid 1)} = \begin{cases} 
\infty; & -1 \leq y < 0 \\
1; & 0 \leq y \leq 1 \\
0; & 1 < y \leq 2
\end{cases}
\]

b) Find the threshold test at \( \eta \) for each \( \eta, 0 < \eta < \infty \) and evaluate the conditional error probabilities, \( q_0(\eta) \) and \( q_1(\eta) \).

Solution: Since \( \Lambda(y) \) has finitely many (3) possible values, all values of \( \eta \) between any adjacent pair lead to the same threshold test. Thus, for \( \eta > 1 \), \( \Lambda(y) > \eta \), leads to the decision \( \hat{x} = 0 \) if and only if (iff) \( \Lambda(y) = \infty \), i.e., iff \(-1 \leq y < 0 \). For \( \eta = 1 \), the rule is the same, \( \Lambda(y) > \eta \) iff \( \Lambda(y) = \infty \), but here there is a ‘don’t care’ case \( \Lambda(y) = 1 \) where \( 0 \leq y \leq 1 \) leads to \( \hat{x} = 1 \) simply because of the convention for the equal case taken in (8.11). Finally for all \( \eta < 1 \), \( \Lambda(Y) > \eta \) iff \(-1 \leq y \leq 1 \).

Consider \( q_0(\eta) \) (the error probability conditional on \( X = 0 \) when a threshold \( \eta \) is used) for \( \eta > 1 \). Then \( \hat{x} = 0 \) iff \(-1 \leq y < 0 \), and thus an error occurs (for \( X = 0 \)) iff \( y \geq 0 \). Thus \( q_0(\eta) = \Pr\{Y \geq 0 \mid X = 0\} = 1/2 \). An error occurs given \( X = 1 \) (still assuming \( \eta > 1 \)) iff \(-1 \leq y < 0 \). These values of \( y \) are impossible under \( X = 1 \) so \( q_1(\eta) = 0 \). These error probabilities are the same if \( \eta = 1 \) because of the handling of the don’t care cases.

For \( \eta < 1 \), \( \hat{x} = 0 \) if and only if \( y \leq 1 \). Thus \( q_0(\eta) = \Pr\{Y > 1 \mid X = 0\} = 0 \). Also \( q_1(\eta) = \Pr\{Y \leq 1 \mid X = 1\} = 1/2 \).

c) Find the error curve \( u(\alpha) \) and explain carefully how \( u(0) \) and \( u(1/2) \) are found (hint: \( u(0) = 1/2 \)).

Solution: We have seen that each \( \eta \geq 1 \) maps into the pair of error probabilities \( (q_0(\eta), q_1(\eta)) = (1/2, 0) \). Similarly, each \( \eta < 1 \) maps into the pair of error probabilities \( (q_0(\eta), q_1(\eta)) = (0, 1/2) \). The error curve contains these points and also contains the straight lines joining these points as shown below (see Figure 8.7). The point \( u(\alpha) \) is the value of \( q_0(\eta) \) for which \( q_1(\eta) = \alpha \). Since \( q_1(\eta) = 0 \) for \( \eta \geq 1 \), \( q_0(\eta) = 1/2 \) for those values of \( \eta \) and thus \( u(0) = 1/2 \). Similarly, \( u(1/2) = 0 \).

\[
\begin{array}{c}
1 \\
1/2 \\
q_0(\eta) \\
q_1(\eta) \\
(0, 1) \\
(1, 0) \\
\end{array}
\]

\[
(0, 1) \\
1/2 \quad q_0(\eta) \\
q_1(\eta) \quad 1/2 \\
(1, 0) \\
\]

d) Describe a decision rule for which the error probability under each hypothesis is 1/4. You need not use a randomized rule, but you need to handle the don’t-care cases under the threshold test carefully.

Solution: The don’t care cases arise for \( 0 \leq y \leq 1 \) when \( \eta = 1 \). With the decision rule of (8.11), these don’t care cases result in \( \hat{x} = 1 \). If half of those don’t care cases are decided
as \( \hat{x} = 0 \), then the error probability given \( X = 1 \) is increased to 1/4 and that for \( X = 0 \) is decreased to 1/4. This could be done by random choice, or just as easily, by mapping \( y > 1/2 \) into \( \hat{x} = 1 \) and \( y \leq 1/2 \) into \( \hat{x} = 0 \).

Exercise 9.10: In this exercise, we show that the optimized Chernoff bound is tight for the 3rd case in (9.12) as well as the first case. That is, we show that if \( r_+ < \infty \) and \( a \geq \sup_{r < r_+} \gamma'(r) \), then for any \( \epsilon > 0 \),
\[
\Pr\{S_n \geq na\} \geq \exp\{n[\gamma(r_+) - r_+a - \epsilon]\}
\]
for all large enough \( n \).

a) Let \( Y_i \) be the truncated version of \( X_i \), truncated for some given \( b \) to \( Y_i = X_i \) for \( X_i \leq b \) and \( Y_i = b \) otherwise. Let \( W_n = Y_1 + \cdots + Y_n \). Show that \( \Pr\{S_n \geq na\} \geq \Pr\{W_n \geq na\} \).

Solution: Since \( Y_i - X_i \) is a nonnegative rv for each \( i \), \( S_n - W_n \) is also nonnegative. Thus for all sample points such that \( W_n \geq a \), \( S_n \geq a \) also; thus \( \Pr\{W_n \geq a\} \leq \Pr\{S_n \geq 0\} \).

b) Let \( g_b(r) \) be the MGF of \( Y \). Show that \( g_b(r) < \infty \) and that \( g_b(r) \) is non-decreasing in \( b \) for all \( r < \infty \).

Solution: Assuming a PMF for notational simplicity, we have
\[
g_b(r) = \sum_{x \leq b} p_X(x)e^{rx} + \sum_{x > b} p_X(x)e^{rb}r \leq \sum_{x} p_X(x)e^{rb} = e^{rb} < \infty.
\]
If \( b \) is replaced by \( b_1 > b \), the terms above for \( x \leq b \) are unchanged and the terms with \( x > b \) are increased from \( p_X(x)e^{rb} \) to either \( p_X(x)e^{rb} \) or \( e^{rb_1} \), increasing the sum. For those who prefer calculus to logic, \( \frac{d}{dr}g_b(r) = re^{rb}\Pr\{X > b\} \), which is nonnegative.

c) Show that \( \lim_{b \to \infty} g_b(r) = \infty \) for all \( r > r_+ \) and that \( \lim_{b \to \infty} g_b(r) = g(r) \) for all \( r \leq r_+ \).

Solution: For \( r > r_+ \), we have \( g(r) = \infty \) and \( g_b(r) \geq \sum_{x \leq b} p_X(x)e^{rx} \), which goes to \( g(r) = \infty \) in the limit \( b \to \infty \). Since \( g_b(r) \) is non-decreasing in \( b \), the same argument works for \( r \leq r_+ \).

d) Let \( \gamma_b(r) = \ln g_b(r) \). Show that \( \gamma_b(r) < \infty \) for all \( r < \infty \). Also show that \( \lim_{b \to \infty} \gamma_b(r) = \infty \) for \( r > r_+ \) and \( \lim_{b \to \infty} \gamma_b(r) = \gamma(r) \) for \( r \leq r_+ \). Hint: Use b) and c).

Solution: Since \( g_b(r) < \infty \), its log is also less than infinity. Also, since the log is monotonic, the limiting results of part c) extend to \( \gamma_b(r) \).

e) Let \( \gamma'_b(r) = \frac{d}{dr} \gamma_b(r) \) and let \( \delta > 0 \) be arbitrary. Show that for all large enough \( b \), \( \gamma'_b(r_+ + \delta) > a \). Hint: First show that \( \gamma'_b(r_+ + \delta) \geq [\gamma_b(r_+ + \delta) - \gamma(r_+)]/\delta \).

Solution: The result in the hint follows since \( \gamma'_b(r) \) is the slope of \( \gamma_b(r) \) and is non-decreasing in \( r \). The main result follows from using part d), which shows that \( \gamma_b(r_+ + \delta) \) is unbounded with increasing \( b \).

f) Show that the optimized Chernoff bound for \( \Pr\{W_n \geq na\} \) is exponentially tight for the values of \( b \) in part e). Show that the optimizing \( r \) is less than \( r_+ + \delta \).

Solution: Applying Theorem 9.3.2 to \( \Pr\{W_n \geq na\} \), we see that the result is exponentially tight if \( a = \gamma'_b(r) \) for some \( r \). For any \( b \) large enough to satisfy part e), i.e., such that \( \gamma'_b(r_+ + \delta) > a \), we can use the continuity of \( \gamma'_b(r) \) in \( r \) and the fact that \( \gamma'_b(0) < \frac{1}{2}X \) to see that such an \( r \) must exist. Since \( \gamma'_b(r_+ + \delta) > a \), the optimizing \( r \) is less than \( r_+ + \delta \).

g) Show that for any \( \epsilon > 0 \) and all sufficiently large \( b \),
\[
\gamma_b(r) - ra \geq \gamma(r) - ra - \epsilon \geq \gamma(r_+) - r_+a - \epsilon \quad \text{for} \ 0 < r \leq r_+.
\]
Solution: It is clear that \(g_b(r)\) must converge to \(g(r)\) uniformly in \(r\) for \(0 \leq r \leq r_+\) and the uniform convergence of \(\gamma_b(r)\) follows from this. For any \(\epsilon > 0\), and for large enough \(b\), this means that \(\gamma(r) - \gamma_b(r) \leq \epsilon\) for all \(r \in (0, r_+)\). This is equivalent to the first part of the desired statement. The second part follows from the fact that \(a > \sup_{r < r_+} \gamma'(r)\).

h) Show that for arbitrary \(\epsilon > 0\), there exists \(\delta > 0\) and \(b_0\) such that
\[
\gamma_b(r) - ra \geq \gamma(r_+) - r_+a - \epsilon \quad \text{for } r_+ < r < r_+ + \delta \text{ and } b \geq b_0.
\]

Solution: For the given \(\epsilon > 0\), choose \(b_0\) so that \(\gamma_b(r_+) > \gamma(r_+) - \epsilon/3\) for \(b \geq b_0\). Define \(\gamma'(r_+)\) as \(\sup_{r < r_+} \gamma'(r)\). Using the same uniform convergence argument on \(\gamma_b'(r)\) as used on \(\gamma_b(r)\), we can also choose \(b_0\) large enough that \(\gamma_b'(r_+) \geq \gamma'(r_+) - \epsilon/3\delta\) for \(b \geq b_0\). We also choose \(\delta = \epsilon/[3(a - \gamma'(r_+))]\). For \(r_+ < r \leq r_+ + \delta\), we then have
\[
\gamma_b(r) - ra \geq \gamma_b(r_+) + (r - r_+) \gamma_b'(r_+) - ra \\
\geq \gamma(r_+) - \epsilon/3 + (r - r_+) \gamma'(r_+) - \epsilon/3\delta - ra \\
\geq \gamma(r_+) - r_+a - \epsilon/3 + (r - r_+) \gamma'(r_+) - a - \epsilon/3\delta \\
\geq \gamma(r_+) - r_+a - \epsilon/3 + \delta \gamma'(r_+) - a - \epsilon/3\delta \\
= \gamma(r_+) - r_+a - \epsilon.
\]

i) Note that if we put together parts g) and h) and d), and use the \(\delta\) of part h) in d), then we have shown that the optimized exponent in the Chernoff bound for \(\Pr\{W_n \geq na\}\) satisfies \(\mu_b(a) \geq \gamma(r_+) - r_+a - \epsilon\) for sufficiently large \(b\). Show that this means that \(\Pr\{S_n \geq na\} \geq \gamma(r_+) - r_+a - 2\epsilon\) for sufficiently large \(n\).

Solution: Theorem 9.3.2 applies to \(\{\Pr\{W_n \geq na\} ; n \geq 1\}\), so for sufficiently large \(n\) and large enough \(b\),
\[
\Pr\{W_n \geq na\} \geq \exp[n(\mu_b(a) - \epsilon)] \geq \exp[n(\gamma(r_+) - r_+a - 2\epsilon)].
\]

Using the result of part a) completes the argument.

Exercise 9.11: Assume that \(X\) is discrete, with possible values \(\{a_i ; i \geq 1\}\) and probabilities \(\Pr\{X = a_i\} = p_i\). Let \(X_r\) be the corresponding tilted random variable as defined in Exercise 9.6. Let \(S_n = X_1 + \ldots + X_n\) be the sum of \(n\) IID rv’s with the distribution of \(X\), and let \(S_{n,r} = X_1 + \ldots + X_n\) be the sum of \(n\) IID tilted rv’s with the distribution of \(X_r\). Assume that \(X < 0\) and that \(r > 0\) is such that \(\gamma(r)\) exists.

a) Show that \(\Pr\{S_{n,r} = s\} = \Pr\{S_n = s\} \exp[\gamma(r)]\). Hint: first show that
\[
\Pr\{X_1r = v_1, \ldots, X_nr = v_n\} = \Pr\{X_1 = v_1, \ldots, X_n = v_n\} \exp[\gamma(r)]
\]
where \(s = v_1 + \ldots + v_n\).

Solution: Since \(X_r\) is the tilted random variable of \(X\), it satisfies
\[
\Pr\{X_r = a_i\} = \Pr\{X = a_i\} \exp[a_i r - \gamma(r)].
\]

Let \(v_1, \ldots, v_n\) be any set of values drawn from the set \(\{a_i\}\) of possible values for \(X\). Let \(s = v_1 + v_2 + \ldots + v_n\). For any such set, since \(\{X_i\}\) is a set of IID variables, each with PMF \(p_{X_i}(v_i) = p_X(v_i) \exp[v_i r - \gamma(r)]\),
\[
\Pr\{X_1r = v_1, \ldots, X_nr = v_n\} = \Pr\{X_1 = v_1, \ldots, X_n = v_n\} \exp[\gamma(r)].
\]
Let $V(s)$ be the set of all vectors $\mathbf{v} = (v_1, \ldots, v_n)$ such that $v_1 + v_2 + \ldots + v_n = s$. Then

$$
\Pr\{S_{n,r} = s\} = \sum_{\mathbf{v} \in V(s)} \Pr\{X_1 = v_1, \ldots, X_n = v_n\} = \sum_{\mathbf{v} \in V(s)} \Pr\{X_1 = v_1, \ldots, X_n = v_n\} \exp[sr - n\gamma(r)] = \Pr\{S_n = s\} \exp[sr - n\gamma(r)].$

b) Find the mean and variance of $S_{n,r}$ in terms of $\gamma(r)$.

**Solution:** $E[S_{n,r}] = nE[X_r] = n\gamma'(r)$ and $\text{VAR}[S_{n,r}] = n\text{VAR}[X_r] = n\gamma''(r)$, as shown in Exercise 9.6.

c) Define $a = \gamma'(r)$ and $\sigma^2 = \gamma''(r)$. Show that $\Pr\{|S_{n,r} - na| \leq \sqrt{2n} \sigma_r\} > 1/2$. Use this to show that

$$
\Pr\{|S_n - na| \leq \sqrt{2n} \sigma_r\} > (1/2) \exp[-r(an + \sqrt{2n} \sigma_r) + n\gamma(r)].
$$

**Solution:** Using the Chebyshev inequality, $\Pr\{|Y - \overline{Y}| \geq \epsilon\} \leq \epsilon^{-2} \text{VAR}[Y]$,

$$
\Pr\{|S_{n,r} - na| \geq \sqrt{2n} \sigma_r\} \leq \frac{n\sigma^2_r}{2n\sigma^2_r} = \frac{1}{2}.
$$

Thus the complementary event satisfies

$$
\Pr\{|S_{n,r} - na| < \sqrt{2n} \sigma_r\} \geq \frac{n\sigma_r}{2n\sigma^2_r} = \frac{1}{2}.
$$

From part a), we see that $\Pr\{S_n = s\} = \Pr\{S_{n,r} = s\} \exp[-sr + n\gamma(r)]$. In the range where $|S_{n,r} - na| < \sqrt{2n} \sigma_r$, we see that $S_{n,r} < na + \sqrt{2n} \sigma_r$, and thus for all $s$ such that $\Pr\{S_{n,r} = s\} > 0$ in this range,

$$
\Pr\{S_n = s\} > \Pr\{S_{n,r} = s\} \exp[-r(na + \sqrt{2n} \sigma_r) + n\gamma(r)].
$$

Summing over all $s$ in this range,

$$
\Pr\{|S_n - na| < \sqrt{2n} \sigma_r\} \geq \frac{1}{2} \exp[-r(na + \sqrt{2n} \sigma_r) + n\gamma(r)].
$$

d) Use this to show that for any $\epsilon$ and for all sufficiently large $n$,

$$
\Pr\{S_n \geq n(\gamma'(r) - \epsilon)\} > \frac{1}{2} \exp[-rn(\gamma'(r) + \epsilon) + n\gamma(r)].
$$

**Solution:** Note that the event $\{|S_n - na| < \sqrt{2n} \sigma_r\}$ is contained in the event $\{S_n > na - \sqrt{2n} \sigma_r\}$. Also, for any $\epsilon > 0$, we can choose $n_0$ such that $n_0\epsilon > \sqrt{2n} \sigma_r$. For all $n \geq n_0$, the event $\{S_n > na - \sqrt{2n} \sigma_r\}$ is contained in the event $\{S_n > na - n\epsilon\} = \{S_n > n(\gamma'(r) - \epsilon)\}$. Thus

$$
\Pr\{|S_n > n(\gamma'(r) - \epsilon)\} > \frac{1}{2} \exp[-r(na + \sqrt{2n} \sigma_r) + n\gamma(r)].$$
Also, for \( n \geq n_0 \), the right hand side of this equation can be further bounded by
\[
\Pr\{ |S_n > n(\gamma'(r) - \epsilon) | \} \geq \frac{1}{2} \exp[-rn(\gamma'(r) + \epsilon)] + n\gamma(r)].
\]

**Exercise 9.13:** Consider a random walk \( \{S_n; n \geq 1\} \) where \( S_n = X_1 + \cdots + X_n \) and \( \{X_i; i \geq 1\} \) is a sequence of IID exponential rv’s with the PDF \( f(x) = \lambda e^{-\lambda x} \) for \( x \geq 0 \). In other words, the random walk is the sequence of arrival epochs in a Poisson process.

a) Show that for \( \lambda > 1 \), the optimized Chernoff bound for \( \Pr\{S_n \geq na\} \) is given by
\[
\Pr\{S_n \geq na\} \leq (a\lambda)^n e^{-n(a\lambda-1)}.
\]

**Solution:** The moment generating function is \( g(r) = \mathbb{E}[e^{Xr}] = \lambda/(\lambda - r) \) for \( r < \lambda \). Thus
\[
\gamma(r) = \ln g(r) = \ln(\lambda/(\lambda - r)) \quad \text{and} \quad \gamma'(r) = 1/(\lambda - r).
\]
The optimizing \( r \) for the Chernoff bound is then the solution to \( a = 1/(\lambda - r) \), which is \( r = \lambda - 1/a \). Using this \( r \) in the Chernoff bound,
\[
\Pr\{S_n \geq na\} = \exp\left[n\ln\left(\frac{\lambda}{\lambda - r}\right) - nra\right] = \exp[n\ln(a\lambda) - n(a\lambda - 1)],
\]
which is equivalent to the desired expression.

b) Show that the exact value of \( \Pr\{S_n \geq na\} \) is given by
\[
\Pr\{S_n \geq na\} = \sum_{i=0}^{n-1} \frac{(na\lambda)^i e^{-na\lambda}}{i!}.
\]

**Solution:** For a Poisson counting process \( \{N(t); t > 0\} \), the event \( \{S_n > na\} \) is the same as \( \{N(na) < n\} = \bigcup_{i=1}^{n-1} \{N(na) = i\} \). Thus the probabilities of these events are the same, which is the statement to be established. Note that \( S_n \) is continuous, so the probability that it equals \( na \) is 0.

c) By upper and lower bounding the quantity on the right above, show that
\[
\frac{(na\lambda)^n e^{-na\lambda}}{n! a\lambda} \leq \Pr\{S_n \geq na\} \leq \frac{(na\lambda)^n e^{-na\lambda}}{n!(a\lambda - 1)}.
\]

Hint: Use the term at \( i = n - 1 \) for the lower bound and note that the term on the right can be bounded by a geometric series starting at \( i = n - 1 \).

**Solution:** The lower bound on the left is the single term with \( i = n - 1 \) of the sum in part b). For the upper bound, rewrite the sum in b) as
\[
\sum_{i=0}^{n-1} \frac{(na\lambda)^i e^{-na\lambda}}{i!} = \frac{(na\lambda)^n e^{-na\lambda}}{n!} \left[ \frac{n}{na\lambda} + \frac{n(n-1)}{(na\lambda)^2} + \cdots \right]
\leq \frac{(na\lambda)^n e^{-na\lambda}}{n!} \left[ \frac{1}{a\lambda} + \frac{1}{(a\lambda)^2} + \cdots \right] = \frac{(na\lambda)^n e^{-na\lambda}}{n!(a\lambda - 1)}.
\]

d) Use the Stirling bounds on \( n! \) to show that
\[
\frac{(a\lambda)^n e^{-n(a\lambda-1)}}{\sqrt{2\pi n a\lambda} \exp(1/12n)} \leq \Pr\{S_n \geq na\} \leq \frac{(a\lambda)^n e^{-n(a\lambda-1)}}{\sqrt{2\pi n} (a\lambda - 1)}.
\]
Solution: The Stirling bounds are
\[
\sqrt{2\pi n} \left( \frac{n}{e} \right)^n < n! < \sqrt{2\pi n} \left( \frac{n}{e} \right)^n e^{1/12n}.
\]
Substituting these for \( n! \) and cancelling terms gives the desired expression. Note that the Chernoff bound contains all the factors that vary exponentially with \( n \). Note also that the Erlang expression for \( S_n \) and the Poisson expression for \( N(t) \) are quite simple, but the corresponding CDF’s are quite messy, and this makes the Chernoff bound more attractive in this case.

Exercise 9.14: Consider a random walk with thresholds \( \alpha > 0, \beta < 0 \). We wish to find \( \Pr \{ S_J \geq \alpha \} \) in the absence of a lower threshold. Use the upper bound in (9.46) for the probability that the random walk crosses \( \alpha \) before \( \beta \).

a) Given that the random walk crosses \( \beta \) first, find an upper bound to the probability that \( \alpha \) is now crossed before a yet lower threshold at \( 2\beta \) is crossed.

Solution: Let \( J_1 \) be the stopping trial at which the walk first crosses either \( \alpha \) or \( \beta \). Let \( J_2 \) be the stopping trial at which the random walk first crosses either \( \alpha \) or \( 2\beta \) (assuming the random walk continues forever rather than actually stopping at any stopping trial. Note that if \( S_{J_1} \geq \alpha \), then \( S_{J_2} = S_{J_1} \), but if \( S_{J_1} \leq \beta \), then it is still possible to have \( S_{J_2} \geq \alpha \). In order for this to happen, a random walk starting at trial \( J_1 \) must reach a threshold of \( \alpha - S_{J_1} \) before reaching \( 2\beta - S_{J_1} \). Putting this into equations,

\[
\Pr \{ S_{J_2} \geq \alpha \} = \Pr \{ S_{J_1} \geq \alpha \} + \Pr \{ S_{J_2} \geq \alpha \mid S_{J_1} \leq \beta \} \Pr \{ S_{J_1} \leq \beta \}.
\]

\[
\Pr \{ S_{J_2} \geq \alpha \mid S_{J_1} \leq \beta \} \leq \exp[\alpha - \beta],
\]

where the latter equation upper bounds the probability that the RW starting at trial \( J_1 \) reaches \( \alpha - S_{J_1} \) before \( 2\beta - S_{J_1} \), given that \( S_{J_1} \leq \beta \).

b) Given that \( 2\beta \) is crossed before \( \alpha \), upper bound the probability that \( \alpha \) is crossed before a threshold at \( 3\beta \). Extending this argument to successively lower thresholds, find an upper bound to each successive term, and find an upper bound on the overall probability that \( \alpha \) is crossed. By observing that \( \beta \) is arbitrary, show that (9.46) is valid with no lower threshold.

Solution: Let \( J_k \) for each \( k \geq 1 \) be the stopping trial for crossing \( \alpha \) before \( k\beta \). By the same argument as above,

\[
\Pr \{ S_{J_{k+1}} \geq \alpha \} = \Pr \{ S_{J_k} \geq \alpha \} + \Pr \{ S_{J_{k+1}} \geq \alpha \mid S_{J_k} \leq k\beta \} \Pr \{ S_{J_k} \leq k\beta \}
\]

\[
\leq \Pr \{ S_{J_k} \geq \alpha \} + \exp[\alpha - k\beta],
\]

Finally, let \( J_\infty \) be the defective stopping time at which \( \alpha \) is first crossed. We see from above that the event \( S_{J_\infty} > \alpha \) is the union of the the events \( S_{J_k} \geq \alpha \) over all \( k \geq 1 \). We can upper bound this by

\[
\Pr \{ S_{J_\infty} \geq \alpha \} \leq \Pr \{ S_{J_1} \geq \alpha \} + \sum_{k=1}^\infty \Pr \{ S_{J_{k+1}} \geq \alpha \mid S_{J_k} \leq k\beta \}
\]

\[
\leq \exp[\alpha] \frac{1}{1 - \exp[\alpha/\beta]}.
\]
Since this is true for all $\beta < 0$, it is valid in the limit $\beta \to -\infty$, yielding $e^{-r^*\alpha}$.

The reason why we did not simply take the limit $\beta \to -\infty$ in the first place is that such a limit would not define a defective stopping rule as any specific type of limit. The approach here was to define it as a union of non-defective stopping rules.

**Exercise 9.16:** a) Use Wald’s equality to show that if $X = 0$, then $E[S_J] = 0$ where $J$ is the time of threshold crossing with one threshold at $\alpha > 0$ and another at $\beta < 0$.

**Solution:** Wald’s equality holds since $E[|J|] < \infty$, which follows from Lemma 9.4.1. Thus $E[S_J] = XE[J]$. Since $X = 0$, it follows that $E[S_J] = 0$.

b) Obtain an expression for $Pr\{S_J \geq \alpha\}$. Your expression should involve the expected value of $S_J$ conditional on crossing the individual thresholds (you need not try to calculate these expected values).

**Solution:** Writing out $E[S_J] = 0$ in terms of conditional expectations,

$$E[S_J] = Pr\{S_J \geq \alpha\} \cdot E[S_J \mid S_J \geq \alpha] + Pr\{S_J \leq \beta\} \cdot E[S_J \mid S_J \leq \beta]$$

Using $E[S_J] = 0$, we can solve this for $Pr\{S_J \geq \alpha\}$,

$$Pr\{S_J \geq \alpha\} = \frac{E[-S_J \mid S_J \leq \beta]}{E[-S_J \mid S_J \leq \beta] + E[S_J \mid S_J \geq \alpha]}.$$

c) Evaluate your expression for the case of a simple random walk.

**Solution:** A simple random walk moves up or down only by unit steps, Thus if $\alpha$ and $\beta$ are integers, there can be no overshoot when a threshold is crossed. Thus $E[S_J \mid S_J \geq \alpha] = \alpha$ and $E[S_J \mid S_J \leq \beta] = \beta$. Thus $Pr\{S_J \geq \alpha\} = \frac{\beta^{\lfloor \beta \rfloor}}{\beta^{\lfloor \beta \rfloor} + \alpha}$. If $\alpha$ is non-integer, then a positive threshold crossing occurs at $\lfloor \alpha \rfloor$ and a lower threshold crossing at $\lfloor \beta \rfloor$. Thus, in this general case $Pr\{S_J \geq \alpha\} = \frac{\beta^{\lfloor \beta \rfloor}}{\beta^{\lfloor \beta \rfloor} + \alpha}$.

d) Evaluate your expression when $X$ has an exponential density, $f_X(x) = a_1e^{-\lambda x}$ for $x \geq 0$ and $f_X(x) = a_2e^{\mu x}$ for $x < 0$ and where $a_1$ and $a_2$ are chosen so that $X = 0$.

**Solution:** Let us condition on $J = n, S_n \geq \alpha$, and $S_{n-1} = s$, for $s < \alpha$. The overshoot, $V = S_J - \alpha$ is then $V = X_n + s - \alpha$. Because of the memoryless property of the exponential, the density of $V$, conditioned as above, is exponential and $f_V(v) = \lambda e^{-\lambda v}$ for $v \geq 0$. This does not depend on $n$ or $s$, and is thus the overshoot density conditioned only on $S_J \geq \alpha$. Thus $E[S_J \mid J \geq \alpha] = \alpha + 1/\lambda$. In the same way, $E[S_J \mid S_J \leq \beta] = \beta - 1/\mu$. Thus

$$Pr\{S_J \geq \alpha\} = \frac{\beta^{\lfloor \beta \rfloor} + \mu^{-1}}{\alpha + \lambda^{-1} + \beta^{\lfloor \beta \rfloor} + \mu^{-1}}.$$

Note that it is not necessary to calculate $a_1$ or $a_2$.

**Exercise 9.21:** (The secretary problem or marriage problem) This illustrates a very different type of sequential decision problem from those of Section 9.5. Alice is looking for a spouse and dates a set of $n$ suitors sequentially, one per week. For simplicity, assume that Alice must make her decision to accept
a suitor for marriage immediately after dating that suitor; she can not come back later to accept a formerly rejected suitor. Her decision must be based only on whether the current suitor is more suitable than all previous suitors. Mathematically, we can view the dating as continuing for all \( n \) weeks, but the choice at week \( m \) is a stopping rule. Assume that each suitor’s suitability is represented by a real number and that all \( n \) numbers are different. Alice does not observe the suitability numbers, but only observes at each time \( m \) whether suitor \( m \) has the highest suitability so far. The suitors are randomly permuted before Alice dates them.

**Solution:** Since the suitability numbers are randomly permuted (i.e., permuted so that each permutation is equally likely), the largest appears at position \( m \) with probability \( 1/n \) for \( 1 \leq m \leq n \).

Conditional on the largest number appearing at some given position \( m \), note that for \( m \leq k \),

\[
\Pr\{\text{choose } m \mid m \text{ best}\} = 0
\]

since the first \( k \) suitors are rejected out of hand. Next suppose that the position \( m \) of the largest number is \( m > k \). Then Alice will choose \( m \) (and thus choose the best suitor) unless Alice has chosen someone earlier, i.e., unless some suitor \( j \), \( k < j < m \), is more suitable than suitors 1 to \( k \). To state this slightly more simply, Alice chooses someone earlier if the best suitor from 1 to \( m - 1 \) lies in a position from \( k + 1 \) to \( m - 1 \). Since the best suitor from 1 to \( m - 1 \) is in each position with equal probability,

\[
\Pr\{\text{choose } m \mid m \text{ best}\} = \frac{k}{m-1} \quad k < j < m.
\]

Averaging over \( m \), we get

\[
q_k = \sum_{m=k+1}^{n} \frac{k}{n(m-1)}.
\]

**b)** Approximating \( \sum_{i=1}^{k} 1/i \) by \( \ln j \), show that for \( n \) and \( k \) large,

\[
q_k \approx \frac{k}{n} \ln(n/k).
\]

Ignoring the constraint that \( n \) and \( k \) are integers, show that the right hand side above is maximized at \( k/n = e \) and that \( \max_k q_k \approx 1/e \).

**Solution:** Using the solution for \( q_k \) in part a), and letting \( i = m - 1 \),

\[
q_k = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i} = \frac{k}{n} \left[ \sum_{i=1}^{n-1} \frac{1}{i} - \sum_{i=1}^{k-1} \frac{1}{i} \right] \approx \frac{k}{n} \left[ \ln(n-1) - \ln(k-1) \right] \approx \frac{k}{n} \left[ \ln \left( \frac{n}{k} \right) \right].
\]

The approximations here are pretty crude. The first one can be improved by using the Euler-Mascheroni constant, but \( q_k \) is still approximately optimized at \( k = n/e \), and at that
point \( q_k \approx 1/e \). The exact value is too easy to calculate to scrutinize the approximations closely.

c) (Optional) Show that the algorithm of part a), optimized over \( k \), is optimal over all algorithms (given
the constraints of the problem). Hint: Let \( p_m \) be the maximum probability of choosing the optimal suitor
given that no choice has been made before time \( m \). Show that \( p_m = \max \left( \frac{1}{n}, \frac{m-1}{m} p_{m+1}, p_{m+1} \right) \); part of
the problem here is understanding exactly what \( p_m \) means.

**Solution outline:** We have looked at the class of algorithms for which the first \( k \) suitors are
rejected and the first subsequent ‘best-yet’ suitor is chosen. Alice has many other choices,
including choosing the suitor at time \( m \) using the past history of best-yet suitors. It turns
out, however, that if we start at time \( m = n \) and work back inductively, we can find the
optimal probability of success conditional on no suitor being chosen before time \( m \). Denote
this optimal conditional probability as \( p_m \). As we shall see, the remarkable fact which
makes this possible is that this conditional probability depends neither on the decisions
made before time \( m \) nor history of best-yet suitors before \( m \).

Given that if no decision is made before time \( n \), then a correct decision can be made only
if suitor \( n \) is the best suitor, and this is an event of probability \( 1/n \), independent of past
history and past decisions. Given no prior decision, the optimal choice is for Alice to chose
suitor \( n \) if \( n \) is best-yet and thus best overall. Thus \( p_n = 1/n \).

Next suppose no decision has been made before time \( n - 1 \). The suitor at \( n - 1 \) is optimal
with probability \( 1/n \), and in this case must be a best-yet. Consider the following possible
decisions at time \( n - 1 \): first, choose \( n - 1 \) if best-yet, or, second, wait until time \( n \) whether
or not \( n - 1 \) is best-yet. We evaluate \( p_{n-1} \) under the first of these decisions and then show
it is optimal. The decisions at \( n - 1 \) and \( n \) are successful overall if \( n - 1 \) is optimal (an
event of probability \( 1/n \) independent of past history and choices) and also if \( n - 1 \) is not
best-yet and \( n \) is best. The event that \( n - 1 \) is not best-yet has probability \( (n - 2)/(n - 1) \)
and is independent of the event that \( n \) is best, so

\[
p_{n-1} = \frac{1}{n} + \frac{n - 2}{n(n - 1)}. \tag{A.32}\]

If Alice waits until time \( n \) even if \( n - 1 \) is best-yet, the probability of success drops to \( 1/n \).
Finally, if Alice bases her decision on the past history of best-yet suitors, the probability
of success is a linear combination of \( p_{n-1} \) above and \( 1/n \), thus verifying (A.32) in general.

We next find \( p_m \) in terms of \( p_{m+1} \) for arbitrary \( m \). We first assume that Alice’s optimal
strategy at time \( m \), given that no priori decision has been made, is to choose \( m \) if it is best-
yet, and if not best-yet to wait until time \( m + 1 \). As before, \( m \) is optimal with probability
\( 1/n \) and if so must be a best-yet choice. If \( m \) is not best-yet (an event of probability
\( (m - 1)/m \)), then the event that success will occur at time \( m + 1 \) or later has probability
\( p_{m+1} \). This event is independent of the event that \( m \) is not best-yet, so if this is the best
decision,

\[
p_m = \frac{1}{n} + \frac{m - 1}{m} p_{m+1} \quad \text{tentatively.} \tag{A.33}\]

Now recall that Alice can also reject suitor \( m \) even if best-yet, and in this case the probability
of success becomes \( p_{m+1} \). If the choice is based on past history, then the probability of success
is a linear combination of $p_{m+1}$ and the value in (A.33). We then have

$$p_m = \max \left[ \frac{1}{n} + \frac{m-1}{m} p_{m+1}, \quad p_{m+1} \right].$$

(A.34)

If we subtract $p_{m+1}$ from both sides, above, we see that $p_m - p_{m+1} = \max \left[ 1/n - p_{m+1}/m, \quad 0 \right]$, from which we conclude that $p_m$ increases as $m$ decreases from $n$, and eventually becomes constant, with the strategy rejecting all suitors for small $m$. If we expand the solution in (A.34), we see that it is simply the strategy in part a), which is thus optimal.

This proves the desired result, but it is of interest to see what happens if Alice receives more information. If at each $m$, she receives the ordering of the first $m$ numbers, then it is easy to see by an extension of the argument above, that the additional information is irrelevant and that the algorithm of part a), optimized over $k$, is still optimal.

On the other hand if Alice sees the actual suitability numbers, she can do better. For example, if the sequence of suitability numbers is a sequence of $n$ IID rv’s, then if Alice knows the distribution of those numbers, she can solve the problem in a very different way and increase the probability of best choice significantly. If she doesn’t know the distribution, and $n$ is very large, she can estimate the distribution and improve the algorithm. The case of an ‘unknown’ distribution is flaky, however, and the notion of optimality becomes undefined.

**d)** Explain why this is a poor model for choosing a spouse (or for making a best choice in a wide variety of similar problems). Caution: It is not enough to explain why this is not closely related to these real problems, You should also explain why this gives very little insight into such real problems.

**Solution:** The assumptions that there are a given number $n$ of candidates, that they are randomly permuted, that they can be evaluated on a one-dimensional scale, that Alice is unaware of that evaluation, but only of a best-yet ranking, and finally that decisions must be made immediately are often individually appropriate. but, as discussed below, are very strange when taken together.

We often model problems where some given parameter ($n$ in this case) is taken to be either large or infinite according to analytical convenience, and a stopping rule is needed to terminate an essentially infinite search. Here, however, the optimal stopping rule depends critically on $n$, so that $n$ should be a meaningful parameter of the underlying problem. In such cases, the reason for stopping and making an early decision is usually related to a cost of making measurements, which does not exist here. Here, the need to make decisions sequentially is quite artificial.

We next ask whether this model can illustrate limited aspects about choosing a spouse (or similar problems of choice). The many artificial and arbitrary aspects of the model make this difficult. The worst part is that most real world problems of sequential choice involve a combination of learning, decision making, and avoidance of delay. This model, in rejecting the first $k$ suitors, appears to be like this, but in fact no learning is involved and no criterion of time-saving is involved.

On the other hand, this model is clearly interesting from a mathematical standpoint — there is a large literature that illustrates its inherent interest. Problems that are inherently interesting are certainly worthy of attention. In addition, limiting the decision maker to
best-yet data can serve as a lower bound to situations of immediate choice where the decision maker has more available data. What is objectionable is claiming applicability to a broad set of real-world decision problems where no such applicability exists.

**Exercise 9.23:** Suppose \( \{Z_n; n \geq 1\} \) is a martingale. Show that

\[
E \left[ Z_m \mid Z_{n_1}, Z_{n_1-1}, \ldots, Z_{n_1} \right] = Z_{n_1} \quad \text{for all } 0 < n_1 < n_2 < \ldots < n_i < m.
\]

**Solution:** First observe from Lemma 9.6.1 that

\[
E \left[ Z_m \mid Z_{n_i}, Z_{n_i-1}, Z_{n_i-2}, Z_1 \right] = Z_{n_i}.
\]

This is valid for every sample value of every conditioning variable. Thus consider \( Z_{n_i-1} \) for example. Since this equation has the same value for each sample value of \( Z_{n_i-1} \), we could take the expected value of this conditional expectation over \( Z_{n_i-1} \), getting

\[
E \left[ Z_m \mid Z_{n_i}, Z_{n_i-1}, Z_1 \right] = E \left[ Z_{n_i-1} \mid Z_{n_i}, Z_{n_i-2}, Z_1 \right] = Z_{n_i}.
\]

In the same way, any subset of these conditioning rv’s could be removed, leaving us with the desired form.

**Exercise 9.26:** This exercise uses a martingale to find the expected number of successive trials \( E[J] \) until some fixed string, \( a_1, a_2, \ldots, a_k \), of successive binary digits occurs within a sequence of IID binary random variables \( X_1, X_2, \ldots \) (see Example 4.5.1 and Exercises 4.28 and 5.35 for alternate approaches). We take the stopping time \( J \) to be the time at which the string occurs, which is defined as the smallest \( n \) for which \( (X_{n-k+1}, \ldots, X_n) = (a_1, \ldots, a_k) \). A mythical casino and sequence of gamblers who follow a prescribed strategy will be used to determine \( E[J] \). The outcomes of the plays (trials), \( \{X_n; n \geq 1\} \) at the casino is a binary IID sequence for which \( \Pr\{X_n = i\} = p_i \) for \( i \in \{0, 1\} \).

If a gambler places a bet \( s \) on 1 at play \( n \), the return is \( s/p_1 \) if \( X_n = 1 \) and 0 otherwise. With a bet \( s \) on 0, the return is \( s/p_0 \) if \( X_n = 0 \) and 0 otherwise; i.e., the game is fair.

a) Assume an arbitrary choice of bets on 0 and 1 by the various gamblers on the various trials. Let \( Y_n \) be the net gain of the casino on trial \( n \). Show that \( E[Y_n] = 0 \). Let \( Z_n = Y_1 + Y_2 + \cdots + Y_n \) be the aggregate gain of the casino over \( n \) trials. Show that for any given pattern of bets, \( \{Z_n; n \geq 1\} \) is a martingale.

**Solution:** The net gain of the casino on trial \( n \) is the sum of the gains on each gambler. If a gambler bets \( s \) on outcome 1, the expected gain for the casino is \( s - p_1 s/p_1 = 0 \). Similarly, it is 0 for a bet on outcome 0. Since the expected gain from each gambler is 0, independent of earlier gains, we have \( E[Y_n|Y_{n-1}, \ldots, Y_1] = 0 \). As seen in Example 9.6.2, this implies that \( \{Z_n; n \geq 1\} \) is a martingale.

b) In order to determine \( E[J] \) for a given string \( (a_1, a_2, \ldots, a_k) \), we program our gamblers to bet as follows:

i) Gambler 1 has an initial capital of 1 which is bet on \( a_1 \) at trial 1. If \( X_1 = a_1 \), the capital grows to \( 1/p_{a_1} \), all of which is bet on \( a_2 \) at trial 2. If \( X_2 = a_2 \), the capital grows to \( 1/(p_{a_1} p_{a_2}) \), all of which is bet on \( a_3 \) at trial 3. Gambler 1 continues in this way until either losing at some trial (in which case he leaves with no money) or winning on \( k \) successive trials (in which case he leaves with \( 1/p_{a_1} \ldots p_{a_k} \)).

ii) Gambler \( m \), for each \( m > 1 \), follows the same strategy, but starts at trial \( m \). Note that if the string \( (a_1, \ldots, a_k) \) appears for the first time at trials \( n-k+1, n-k+2, \ldots, n \), i.e., if \( J = n \), then gambler \( n-k+1 \) leaves at time \( n \) with capital \( 1/[p_{a_1} \ldots p_{a_k}] \) and gamblers \( j < n-k+1 \) have all lost their capital. We will come back later to investigate the capital at time \( n \) for gamblers \( n-k+2 \) to \( n \).

First consider the string \( a_1=0, a_2=1 \) with \( k = 2 \). Find the sample values of \( Z_1, Z_2, Z_3 \) for the sample sequence \( X_1 = 1, X_2 = 0, X_3 = 1, \ldots \). Note that gamblers 1 and 3 have lost their capital, but gambler 2
now has capital $1/p_0p_1$. Show that the sample value of the stopping time for this case is $J = 3$. Given an arbitrary sample value $n \geq 2$ for $J$, show that $Z_n = n - 1/p_0p_1$.

**Solution:** Since gambler 1 bets on 0 at the first trial and $X_1 = 1$, gambler 1 loses and $Z_1 = 1$. At trial 2, gambler 2 bets on 0 and $X_2 = 0$. Gambler 2’s capital increases from 1 to $1/p_0$ so $Y_2 = 1 - 1/p_0$. Thus $Z_2 = 2 - 1/p_0$. On trial 3, gambler 1 is broke and doesn’t bet, gambler 2’s capital increases from $1/p_0$ to $1/p_0p_1$ and gambler 3 loses. Thus $Y_3 = 1 + 1/p_0 - 1/p_0p_1$ and $Z_3 = 3 - 1/p_0p_1$. It is preferable here to look only at the casino’s aggregate gain $Z_3$ and not the immediate gain $Y_3$. In aggregate, the casino keeps all 3 initial bets, and pays out $1/p_0p_1$.

$J = 3$ since $(X_2, X_3) = (a_1, a_2) = (0, 1)$ and this is the first time that the string $(0, 1)$ has appeared. For an arbitrary sample value $n$ for $J$, note that each gambler before $n - 1$ loses unit capital, gambler $n - 1$ retires to Maui with capital increased from 1 to $1/p_0p_1$, and gambler $n$ loses. Thus the casino has $n - 1/p_0p_1$ as its gain.

c) Find $E[Z_J]$ from part a). Use this plus part b) to find $E[J]$. Hint: This uses the special form of the solution in part b, not the Wald equality.

**Solution:** The casino’s expected gain at each time $n$ is $E[Z_n] = 0$, so it follows that $E[Z_J] = 0$ (It is easy to verify that the condition in 9.104 is satisfied in this case). We saw in part b) that $E[Z_n | J = n] = n - 1/p_0p_1$, so $E[Z_J] = E[J] - 1/p_0p_1$. Thus $E[J] = 1/p_0p_1$.

Note that this is the mean first passage time for the same problem in Exercise 4.28. The approach there was simpler than this for this short string. For long strings, the approach here will be simpler.

d) Repeat parts b) and c) using the string $(a_1, \ldots, a_k) = (1, 1)$ and initially assuming $(X_1X_2X_3) = (011)$.

Be careful about gambler 3 for $J = 3$. Show that $E[J] = 1/p_1^2 + 1/p_1$.

**Solution:** This is almost the same as part b) except that here gambler 3 wins at time 3. In other words, since $a_1 = a_2$, gamblers 2 and 3 both bet on 1 at time 3. As before, $J = 3$ for this sample outcome. We also see that for $J$ equal to an arbitrary $n$, gamblers $n - 1$ and $n$ both bet on 1 and since $X_n = 1$, both win. Thus $E[J] = 1/p_1^2 + 1/p_1$.

e) Repeat parts b) and c) for $(a_1, \ldots, a_k) = (1, 1, 1, 0, 1, 1)$.

**Solution:** Given that $J = n$, we know that $(X_{n-5}, \ldots, X_n) = (110111)$ so gambler $n - 5$ leaves with $1/p_1^5p_0$ and all previous gamblers lose their capital. For the gamblers after $n - 5$, note that gambler $n$ makes a winning bet on 1 at time $n$ and gambler $n - 1$ makes winning bets on $(1, 1)$ at times $(n-1, n)$. Thus the casino wins $n - 1/p_1^5p_0 - 1/p_1 - 1/p_1^2$. Averaging over $J$, we see that $E[J] = 1/(p_1^5p_0) + 1/p_1 + 1/p_1^2$. In general, we see that, given $J = n$, gambler $n$ wins if $a_1 = a_k$, gambler 2 wins if $(a_1, a_2) = (a_{k-1}, a_k)$ and so forth.

f) Consider an arbitrary binary string $a_1, \ldots, a_k$ and condition on $J = n$ for some $n \geq k$. Show that the sample capital of gambler $m$ is then equal to

- $0$ for $m < n - k$.
- $1/p_1^mp_0^m$ for $m = n - k + 1$.
- $1/p_1^mp_0^m$ for $m = n - i + 1$, $1 \leq i < k$ if $(a_1, \ldots, a_i) = (a_{k-i+1}, \ldots, a_k)$.
- $0$ for $m = n - i + 1$, $1 \leq i < k$ if $(a_1, \ldots, a_i) \neq (a_{k-i+1}, \ldots, a_k)$.

Verify that this general formula agrees with parts c), d), and e).
Solution: Gamer $m$ for $m \leq n - k$ bets (until losing a bet) on $a_1, a_2, \ldots, a_k$. Since the first occurrence of $(a_1, \ldots, a_k)$ occurs at $n$, we see that each of these gamers loses at some point and thus is reduced to 0 capital at that point and stays there. Gamer $n - k + 1$ bets on $a_1, \ldots, a_k$ at times $n-k+1, \ldots, n$ and thus wins each bet for $J = n$. Finally, gamer $m = n - i + 1$ bets (until losing) on $a_1, a_2, \ldots, a_i$ at times $n - i + 1$ to $n$. Since $J = n$ implies that $X_{n-k+1}, \ldots, X_n = a_1, \ldots, a_k$, gamer $n - i + 1$ is successful on all $i$ bets if and only if $(a_1, \ldots, a_i) = (a_{k-i+1}, \ldots, a_k)$.

For part b), gamer $n$ is unsuccessful and in part c), gamer $n$ is successful. In part d), gamers $n - 1$ and $n$ are each successful. It might be slightly mystifying at first that conditioning on $J$ is enough to specify what happens to each gamer after time $n - k + 1$, but the sample values of $X_{n-k+1}$ to $X_n$ are specified by $J = n$, and the bets of the gamers are also specified.

g) For a given binary string $(a_1, \ldots, a_k)$, and each $i, 1 \leq i \leq k$ let $I_i = 1$ if $(a_1, \ldots, a_i) = (a_{k-i+1}, \ldots, a_k)$ and let $I_j = 0$ otherwise. Show that

$$E[J] = \sum_{i=1}^{k} \frac{I_i}{\prod_{l=1}^{k} p_{u_l}}.$$ 

Note that this is the same as the final result in Exercise 5.35. The argument is shorter here, but more motivated and insightful there. Both approaches are useful and lead to simple generalizations.

Solution: The $i$th term in the above expansion is the capital of gamer $n - i + 1$ at time $n$. The final term at $i = k$ corresponds to the gamer who retires to Maui and $I_k = 1$ in all cases. How many other terms are non-zero depends on the choice of string. These other terms can all be set to zero by choosing a string for which no prefix is equal to the suffix of the same length.

Exercise 9.27: a) This exercise shows why the condition $E[|Z_J|] < \infty$ is required in Lemma 9.8.1. Let $Z_1 = -2$ and, for $n \geq 1$, let $Z_{n+1} = Z_n [1 + X_n (3n + 1)/(n+1)]$ where $X_1, X_2, \ldots$ are IID and take on the values $+1$ and $-1$ with probability $1/2$ each. Show that $\{Z_n; n \geq 1\}$ is a martingale.

Solution: From the definition of $Z_n$ above,

$$E[Z_n | Z_{n-1}, Z_{n-2}, \ldots, Z_1] = E[Z_{n-1} [1 + X_{n-1} (3n - 2)/n] | Z_{n-1}, \ldots, Z_1].$$

Since the $X_n$ are zero mean and IID, this is just $E[Z_{n-1} | Z_{n-1}, \ldots, Z_1]$, which is $Z_{n-1}$. Thus $\{Z_n; n \geq 1\}$ is a martingale.

b) Consider the stopping trial $J$ such that $J$ is the smallest value of $n > 1$ for which $Z_n$ and $Z_{n-1}$ have the same sign. Show that, conditional on $n < J$, $Z_n = (-2)^n/n$ and, conditional on $n = J$, $Z_J = -(-2)^n (n - 2)/(n^2 - n)$.

Solution: It can be seen from the iterative definition of $Z_n$ that $Z_n$ and $Z_{n-1}$ have different signs if $X_{n-1} = -1$ and have the same sign if $X_{n-1} = 1$. Thus the stopping trial is the smallest trial $n \geq 2$ for which $X_{n-1} = 1$. Thus for $n < J$, we must have $X_1 = -1$ for $1 \leq i < n$. For $n = 2 < J$, $X_1$ must be $-1$, so from the formula above, $Z_2 = Z_1 [1 - 4/2] = 2$. Thus $Z_n = (-2)^{n}/n$ for $n = 2 < J$. We can use induction now for arbitrary $n < J$. Thus for $X_n = -1$,

$$Z_{n+1} = Z_n \left[ 1 - \frac{3n + 1}{n + 1} \right] = \frac{(-2)^n}{n} \frac{-2n}{n+1} = \frac{(-2)^{n+1}}{n+1}.$$
The remaining task is to compute $Z_n$ for $n = J$. Using the result just derived for $n = J - 1$ and using $X_{J-1} = 1$,

$$Z_J = Z_{J-1} \left[ 1 + \frac{3(J-1)+1}{J} \right] = \frac{(-2)^{J-1} 4J - 2}{J - 1} = \frac{-(2)^J (2J - 1)}{J(J - 1)}.$$  

c) Show that $E[\|Z_J\|]$ is infinite, so that $E[Z_J]$ does not exist according to the definition of expectation, and show that $\lim_{n \to \infty} E[Z_n|J > n] \Pr\{J > n\} = 0$.

**Solution:** We have seen that $J = n$ if and only if $X_i = -1$ for $1 \leq i \leq n - 2$ and $X_{J-1} = 1$. Thus $\Pr\{J = n\} = 2^{-n+1}$. So

$$E[\|Z_J\|] = \sum_{n=2}^{\infty} 2^{n-1} \cdot \frac{2^n(2n - 1)}{n(n - 1)} = \sum_{n=2}^{\infty} \frac{2(2n - 1)}{n(n - 1)} \geq \sum_{n=2}^{\infty} \frac{4}{n} = \infty,$$

since the harmonic series diverges.

Finally, we see that $\Pr\{J > n\} = 2^{n-1}$ since this event occurs if and only if $X_i = -1$ for $1 \leq i < n$. Thus

$$E[Z_n \mid J > n] \Pr\{J > n\} = \frac{2^{-n+1} 2^n}{n} = \frac{2}{n} \to 0.$$  

Section 9.8 explains the significance of this exercise.

**Exercise 9.35:** (The double-or-quarter game) Consider an investment example similar to that of Example 9.10.1 in which there is only one investment other than cash. The ratio $X$ of that investment value at the end of an epoch to that at the beginning is either 2 or 1/4, each with equal probability. Thus $\Pr\{X = 2\} = 1/2$ and $\Pr\{X = 1/4\} = 1/2$. Successive ratios, $X_1, X_2, \ldots$, are IID.

a) In parts a) to c), assume a constant allocation strategy where a fraction $\lambda$ is constantly placed in the double-or-quarter investment and the remaining $1 - \lambda$ is kept in cash. Find the expected wealth, $E[W_n(\lambda)]$ and the expected log-wealth $E[L_n(\lambda)]$ as a function of $\lambda \in [0, 1]$ and $n \geq 1$. Assume unit initial wealth.

**Solution:** With probability 1/2, the fraction $\lambda$ that is invested during epoch 1 becomes $2\lambda$ and the $1 - \lambda$ in cash remains $1 - \lambda$, leading to wealth $W_1 = 1 + \lambda$. Similarly, with probability 1/2, the fraction $\lambda$ that is invested becomes $\lambda/4$, yielding the final wealth $W_1 = 1 - 3\lambda/4$.

The expected wealth at the end of the first epoch is then

$$E[W_1(n)] = (1/2)(1 + \lambda) + (1/2)(1 - 3\lambda/4) = 1 + \lambda/8.$$  

As in Example 9.10.1, this grows geometrically with $n$, so

$$E[W_n] = (1 + \lambda/8)^n.$$  

The log-wealth $L_n(\lambda)$ is the log of $W_n(\lambda)$, which is $\ln(1 + \lambda)$ with probability 1/2 and $\ln(1 - 3\lambda/4)$ with probability 1/2. Thus $E[L_1(\lambda)] = (1/2) \log ((1 + \lambda)(1 - 3\lambda/n))$. Thus

$$L_n(\lambda) = \frac{n}{2} \log ((1 + \lambda)(1 - 3\lambda/4)).$$  

b) For $\lambda = 1$, find the PMF for $W_4(1)$ and give a brief explanation of why $E[W_n(1)]$ is growing exponentially with $n$ but $E[L_n(1)]$ is decreasing linearly toward $-\infty$. 
**Solution:** The probability of $k$ successes out of $n$ is given by the binomial $\binom{n}{k}2^{-n}$, which for $n = 4$ is, respectively, $1/16$, $4/16$, $6/16$, $4/16$, $1/16$ for $k = 0, 1, 2, 3, 4$. Thus,

\[
\Pr\{L_n = 4 \log(1 - 3\lambda/4)\} = 1/16 \\
\Pr\{L_n = 3 \log(1 - 3\lambda/4) + \log(1 + \lambda)\} = 4/16 \\
\Pr\{L_n = 2 \log(1 - 3\lambda/4) + 2 \log(1 + \lambda)\} = 6/16 \\
\Pr\{L_n = 1 \log(1 - 3\lambda/4) + 3 \log(1 + \lambda)\} = 4/16 \\
\Pr\{L_n = 4 \log(1 + \lambda)\} = 1/16.
\]

The point of this is to be able to see how, as $n$ increases, the probability distribution of $L_n(\lambda)$ starts to peak around the mean value.

c) Using the same approach as in Example 9.10.1, find the value of $\lambda^*$ that maximizes $\mathbb{E}[L_n(\lambda)]$. Show that your solution satisfies the optimality conditions in (9.136).

**Solution:** The easiest way to do this is to use the fact that the log function is monotonic. Therefore we can maximize $L_n(\lambda)$ over $\lambda$ by maximizing $(1 + \lambda)(1 - 3\lambda/4)$. The maximum occurs at $\lambda^* = 1/6$. We then see that $\mathbb{E}[L_n(\lambda^*)] = (n/2)\ln(49/48)$.

To see that this actually satisfies the optimality conditions, let $X$ be the price relative for the double or quarter investment and 1 be used for the price relative of cash. Using $\lambda^* = 1/6$, we see that with probability $1/2$, $X = 2$ and

\[
\sum \lambda^*(k)X(k) = 2\lambda^* + (1 - \lambda^*) = 7/6.
\]

Similarly, with probability $1/2$, $X = 1/4$ and

\[
\sum \lambda^*(k)X(k) = (1/4)\lambda^* + 1 - \lambda^* = 7/8.
\]

For the double or quarter investment, then, the optimality condition is

\[
\mathbb{E}\left[\frac{X}{\sum \lambda(k)X(k)}\right] = \frac{1}{2} \cdot \frac{2}{7/6} + \frac{1}{2} \cdot \frac{1/4}{7/8} = 1.
\]

The same type of calculation works for the cash investment. The lesson to be learned from this is that the optimality conditions, while simple appearing and highly useful theoretically, is not trivial to interpret or work with.

d) Find the range of $\lambda$ over which $\mathbb{E}[L_n(\lambda)]$ is positive.

**Solution:** $\mathbb{E}[L_n(\lambda)]$ is positive where $(1 + \lambda)(1 - 3\lambda/4)$ is greater than 1. Solving for $(1 + \lambda)(1 - 3\lambda/4) = 1$, we get $\lambda = 1/3$ and $\lambda = 0$. We then see that $\mathbb{E}[L_n(\lambda)]$ is positive between those limits, i.e., for $0 < \lambda < 1/3$.

e) Find the PMF of the rv $Z_n/Z_{n-1}$ as given in (9.138) for any given $\lambda_n$.

**Solution:** We use (9.138) and then substitute in the result of part c), first for $X = 2$ and
then for $X = 1/4$.

\[ \frac{Z_n}{Z_{n-1}} = \sum_j \lambda_n(j) \frac{X_n(j)}{\sum_k \lambda^*(k)X_n(k)} \]

\[ = \lambda_n \frac{2}{7} + (1 - \lambda_n) \frac{1}{7/6} = \frac{6}{7} + \frac{6\lambda}{7} \quad \text{for } X = 2 \]

\[ \frac{Z_n}{Z_{n-1}} = \lambda_n \frac{1/4}{7/8} + (1 - \lambda_n) \frac{1}{7/8} = \frac{8}{7} - \frac{6\lambda}{7} \quad \text{for } X = 1/2. \]

Thus, $Z_n/Z_{n-1}$ is $6/7 + 6\lambda/7$ with probability $1/2$ and $8/7 - 6\lambda/7$ with probability $1/2$. As a check, $\mathbb{E}[Z_n/Z_{n-1}] = 1$ for all $\lambda$, since as we have seen $\{Z_n; n \geq 1\}$ is a product martingale in this case.

**Exercise 9.36:** (Kelly’s horse-racing game) An interesting special case of this simple theory of investment is the horse-racing game due to J. Kelly and described in [5]. There are $\ell - 1$ horses in a race and each horse $j$ wins with some probability $p(j) > 0$. One and only one horse wins, and if $j$ wins, the gambler receives $r(j) > 0$ for each dollar bet on $j$ and nothing for the bets on other horses. In other words, the price-relative $X(j)$ for each $j$, $1 \leq j \leq \ell - 1$, is $r(j)$ with probability $p(j)$ and 0 otherwise. For cash, $X(\ell) = 1$.

The gambler’s allocation of wealth on the race is denoted by $\lambda(j)$ on each horse $j$ with $\lambda(\ell)$ kept in cash. As usual, $\sum_j \lambda(j) = 1$ and $\lambda(j) \geq 0$ for $1 \leq j \leq \ell$. Note that $X(1), \ldots, X(\ell - 1)$ are highly dependent, since only one is nonzero in any sample race.

a) For any given allocation $\lambda$ find the expected wealth and the expected log-wealth at the end of a race for unit initial wealth.

**Solution:** With probability $p(j)$, horse $j$ wins and the resulting value of $W_1(\lambda)$ is $\lambda(j)r(j) + \lambda(\ell)$. Thus

\[ \mathbb{E}[W_1(\lambda)] = \sum_{j=1}^{\ell-1} p_j [\lambda(j)r(j) + \lambda(\ell)], \]

\[ \mathbb{E}[L_1(\lambda)] = \sum_{j=1}^{\ell-1} p_j \ln [\lambda(j)r(j) + \lambda(\ell)]. \]

b) Assume that a statistically identical sequence of races are run, i.e., $X_1, X_2, \ldots$, are IID where each $X_n = (X_n(1), \ldots, X_n(\ell))^T$. Assuming unit initial wealth and a constant allocation $\lambda$ on each race, find the expected log-wealth $\mathbb{E}[L_n(\lambda)]$ at the end of the $n$th race and express it as $n\mathbb{E}[Y(\lambda)]$.

**Solution:** Using (9.128) to express $\mathbb{E}[L_n(\lambda)]$ as $n\mathbb{E}[Y(\lambda)]$, we have

\[ \mathbb{E}[Y(\lambda)] = \sum_{j=1}^{\ell-1} p(j) \ln [\lambda(j)r(j) + \lambda(\ell)]. \quad (A.35) \]

c) Let $\lambda^*$ maximize $\mathbb{E}[Y(\lambda)]$. Use the necessary and sufficient condition (9.136) on $\lambda^*$ for horse $j$, $1 \leq j < \ell$ to show that $\lambda^*(j)$ can be expressed in the following two equivalent ways; each uniquely specifies each $\lambda^*(j)$ in terms of $\lambda^*(\ell)$.

\[ \lambda^*(j) \geq p(j) - \frac{\lambda^*(\ell)}{r(j)}, \quad \text{with equality if } \lambda^*(j) > 0 \quad (A.36) \]

\[ \lambda^*(j) = \max \left\{ p(j) - \frac{\lambda^*(\ell)}{r(j)}, 0 \right\}. \quad (A.37) \]
Solution: The necessary and sufficient condition for $\lambda^*$ in (9.136) for horse $j$ is

$$E \left[ \frac{X(j)}{\sum_k \lambda^*(k)X(k)} \right] \leq 1; \quad \text{with equality if } \lambda^*(j) > 0.$$  

In the event that horse $j$ wins, $X(j) = r(j)$ while $X(k) = 0$ for horses $k \neq j$. Also $X(\ell) = 1$. Thus in the event that horse $j$ wins, $X(j) = \frac{r(j)}{\sum_k \lambda^*(k)X(k)}$. If any other horse wins, $X(j) = 0$. Thus, since $j$ wins with probability $p(j)$,

$$E \left[ \frac{X(j)}{\sum_k \lambda^*(k)X(k)} \right] = \frac{p(j)}{\lambda^*(j)r(j) + \lambda^*(\ell)} \leq 1; \quad \text{with equality if } \lambda^*(j) > 0. \quad (A.38)$$

Rearranging this inequality yields (A.36); (A.37) is then easily verified by looking separately at the cases $\lambda^*(j) > 0$ and $\lambda^*(j) = 0$.

Solving for $\lambda^*(\ell)$ (which in turn specifies the other components of $\lambda^*$) breaks into 3 special cases which are treated below in parts d), e), and f) respectively. The first case, in part d), shows that if $\sum_{j<\ell} 1/r(j) < 1$, then $\lambda^*(\ell) = 0$. The second case, in part e), shows that if $\sum_{j<\ell} 1/r(j) > 1$, then $\lambda^*(\ell) = \min_j (p(j)r(j))$, with the specific value specified by the unique solution to (A.40). The third case, in part f), shows that if $\sum_{j<\ell} 1/r(j) = 1$, then $\lambda^*(\ell)$ is nonunique, and its set of possible values occupy the range $[0, \min_j (p(j)r(j))].$

d) Sum (A.36) over $j$ to show that if $\lambda^*(\ell) > 0$, then $\sum_{j<\ell} 1/r(j) \geq 1$. Note that the logical obverse of this is that $\sum_{j<\ell} 1/r(j) < 1$ implies that $\lambda^*(\ell) = 0$ and thus that $\lambda^*(j) = p(j)$ for each horse $j$.

Solution: Summing (A.36) over $j < \ell$ and using the fact that $\sum_{j<\ell} \lambda^*(j) = 1 - \lambda^*(\ell)$, we get

$$1 - \lambda^*(\ell) \geq 1 - \lambda^*(\ell) \sum_{j<\ell} 1/r(j).$$

If $\lambda^*(\ell) > 0$, this shows that $\sum_j 1/r(j) \geq 1$. The logical obverse is that if $\sum_j 1/r(j) < 1$, then $\lambda^*(\ell) = 0$. This is the precise way of saying that if the returns on the horses are sufficiently large, then no money should be retained in cash.

When $\lambda^*(\ell) = 0$ is substituted into (A.36), we see that each $\lambda^*(j)$ must be positive and thus equal to $p(j)$. This is very surprising, since it says that the allocation of bets does not depend on the rewards $r(j)$ (so long as the rewards are large enough to satisfy $\sum_j 1/r(j) < 1$). This will be further explained by example in part g).

e) In part c), $\lambda^*(\ell)$ for each $j < \ell$ was specified in terms of $\lambda^*(\ell)$; here you are to use the necessary and sufficient condition (9.136) on cash to specify $\lambda^*(\ell)$. More specifically, you are to show that $\lambda^*(\ell)$ satisfies each of the following two equivalent inequalities:

$$\sum_{j<\ell} \frac{p(j)}{\lambda^*(j)r(j) + \lambda^*(\ell)} \leq 1; \quad \text{with equality if } \lambda^*(\ell) > 0 \quad (A.39)$$

$$\sum_{j<\ell} \frac{p(j)}{\max [p(j)r(j), \lambda^*(\ell)]} \leq 1; \quad \text{with equality if } \lambda^*(\ell) > 0. \quad (A.40)$$

Show from (A.40) that if $\lambda^*(\ell) \leq p(j)r(j)$ for each $j$, then $\sum_j 1/r(j) \leq 1$. Point out that the logical obverse of this is that if $\sum_j 1/r(j) > 1$, then $\lambda^*(\ell) > \min_j (p(j)r(j))$. Explain why (A.40) has a unique solution for $\lambda^*(\ell)$ in this case. Note that $\lambda^*(j) = 0$ for each $j$ such that $p(j)r(j) < \lambda^*(\ell)$. 
**Solution:** The necessary and sufficient condition for cash (investment \( \ell \)) is

\[
E \left[ \frac{X(\ell)}{\sum_k \lambda^*(k)X(k)} \right] \leq 1; \quad \text{with equality if } \lambda^*(\ell) > 0. \quad (A.41)
\]

In the event that horse \( j \) wins, \( X(\ell) \) has the sample value 1 and \( \sum_k \lambda^*(k)X(k) \) has the sample value \( \lambda^*(j)r(j) + \lambda^*(\ell) \). Taking the expectation by multiplying by \( p(j) \) and summing over \( j < \ell \), (A.41) reduces to (A.39). Now if we multiply both sides of (A.37) by \( r(j) \) and then add \( \lambda^*(\ell) \) to both sides, we get

\[
\lambda^*(j)r(j) + \lambda^*(\ell) = \max \left[ p(j)r(j), \lambda^*(\ell) \right],
\]

which converts (A.39) into (A.40). Now assume that \( \lambda^*(\ell) \leq p(j)r(j) \) for each \( j \). Then the max in the denominator of the left side of (A.40) is simply \( p(j)r(j) \) for each \( j \) and (A.40) becomes \( \sum_{j < \ell} 1/r(j) \leq 1 \). The logical obverse is that \( \sum_{j < \ell} 1/r(j) > 1 \) implies that \( \lambda^*(\ell) > \min_j (p(j)r(j)) \), as was to be shown.

Finally, we must show that if \( \sum_{j < \ell} 1/r(j) > 1 \), then (A.40) has a unique solution for \( \lambda^*(\ell) \). The left side of (A.40), viewed as a function of \( \lambda^*(\ell) \), is \( \sum_{j < \ell} 1/r(j) > 1 \) for \( \lambda^*(\ell) = \min_j (p(j)r(j)) \). This quantity is strictly decreasing in \( \lambda^*(\ell) \) as \( \lambda^*(\ell) \) increases and is less than or equal to 1 at \( \lambda^*(\ell) = 1 \). Thus there must be a unique value of \( \lambda^*(\ell) \) at which (A.40) is satisfied.

It is interesting to observe from (A.37) that \( \lambda^*(j) = 0 \) for each \( j \) such that \( p(j)r(j) \leq \lambda^*(\ell) \). In other words, no bets are placed on any horse \( j \) for which \( E[X(j)] < \lambda^*(\ell) \). This is in marked contrast to the case in part d) where the allocation does not depend on \( r(j) \) (within the assumed range).

f) Now consider the case in which \( \sum_{j < \ell} 1/r(j) = 1 \). Show that (A.40) is satisfied with equality for each choice of \( \lambda^*(\ell) \), \( 0 \leq \lambda^*(\ell) \leq \min_{j < \ell} p(j)r(j) \).

**Solution:** Note that \( \max \left[ p(j)r(j), \lambda^*(\ell) \right] = p(j)r(j) \) over the given range of \( \lambda^*(\ell) \), so the left side of (A.40) is \( \sum_{j < \ell} 1/r(j) = 1 \) over this range. Thus the inequality in (A.40) is satisfied for all \( \lambda^*(\ell) \) in this range. Using \( \lambda^*(j) = p(j) - \lambda^*(\ell)/r(j) \), for each \( j < \ell \), all the necessary and sufficient conditions are satisfied for maximizing \( E[Y(\lambda)] \).

g) Consider the special case of a race with only two horses. Let \( p(1) = p(2) = 1/2 \). Assume that \( r(1) \) and \( r(2) \) are large enough to satisfy \( 1/r(1) + 1/r(2) < 1 \); thus no cash allotment is used in maximizing \( E[Y(\lambda)] \). With \( \lambda(3) = 0 \), we have

\[
E[Y(\lambda)] = \frac{1}{2} \ln[\lambda(1)r(1)] + \frac{1}{2} \ln[\lambda(2)r(2)] = \frac{1}{2} \ln[\lambda(1)r(1)(1 - \lambda(1))r(2)]. \quad (A.42)
\]

Use this equation to give an intuitive explanation for why \( \lambda^*(1) = 1/2 \), independent of \( r(1) \) and \( r(2) \).

**Solution:** Suppose that \( r(1) >> r(2) \). Choosing \( \lambda(1) \) to be large so as to enhance the profit when horse 1 wins is counter-productive, since (A.42) shows that there is a corresponding loss when horse 2 wins. This gain and loss cancel each other in the expected log wealth. To view this slightly differently, if each horse wins \( n/2 \) times, \( W_n \) is given by

\[
W_n = (\lambda(1))^{n/2}(1 - \lambda(1))^{n/2}(r(1))^{n/2}(r(2))^{n/2},
\]

which again makes it clear that \( \lambda^*(1) \) does not depend on \( r(1) \) and \( r(2) \).
h) Again consider the special case of two horses with \( p(1) = p(2) = 1/2 \), but let \( r(1) = 3 \) and \( r(2) = 3/2 \). Show that \( \lambda^* \) is nonunique with \( (1/2, 1/2, 0) \) and \( (1/4, 0, 3/4) \) as possible values. Show that if \( r(2) > 3/2 \), then the first solution above is unique, and if \( r(2) < 3/2 \), then the second solution is unique, assuming \( p(1) = 1/2 \) and \( r(1) = 3 \) throughout. Note: When \( 3/2 < r(2) < 2 \), this is an example where an investment in used to maximize log-wealth even though \( E[X(2)] = p(2)r(2) < 1 \), i.e., horse 2 is a lousy investment, but is preferable to cash in this case as a hedge against horse 1 losing.

Solution: Approach 1: Substitute \( \lambda^* = (1/2, 1/2, 0)^T \) and then \( (1/4, 0, 3/4)^T \) into the necessary and sufficient conditions; each satisfies those conditions. Approach 2: Note that \( 1/r(1) + 1/r(2) = 1 \). Thus, from part f), each of these values is satisfied.

Both choices of \( \lambda^* \) here lead to the same rv, i.e., \( Y(\lambda^*) = \ln[3/2] \) for the event that horse 1 wins and \( Y(\lambda^*) = \ln[3/4] \) for the event that horse 2 wins. In other words, the maximizing rv \( Y(\lambda^*) \) is uniquely specified, even though \( \lambda^* \) is not unique. All points on the straight line between these two choices of \( \lambda^* \), i.e., \( (1/2 - \alpha, 1/2 - 2\alpha, 3\alpha)^T \) for \( 0 \leq \alpha \leq 1/4 \) also lead to the same optimizing \( Y(\lambda^*) \).

For \( r(2) > 3/2 \), we have \( 1/r(1) + 1/r(2) < 1 \), so from part d), the solution \( (1/2, 1/2, 0) \) is valid and in this case unique. This can also be seen by substituting this choice of \( \lambda^* \) into the necessary and sufficient conditions, first with \( r(2) > 3/2 \) and then with \( r(2) < 3/2 \).

From part e), the choice \( \lambda^* = (1/4, 0, 3/4) \) is the unique solution for \( 1/r(1) + 1/r(2) > 0 \), i.e., for \( r(2) < 3/2 \). This can be recognized as the allocation that maximizes \( E[Y(\lambda)] \) for the triple-or-nothing investment.

i) For the case where \( \sum_{j<\ell} 1/r(j) = 1 \), define \( q(j) = 1/r(j) \) as a PMF on \( \{1, \ldots, \ell - 1\} \). Show that \( E[Y(\lambda^*)] = D(p \| q) \) for the conditions of part f). Note: To interpret this, we could view a horse race where each horse \( j \) has probability \( q(j) \) of winning the reward \( r(j) = 1/q(j) \) as a ‘fair game’. Our gambler, who knows that the true probabilities are \( \{p(j); 1 \leq j < \ell\} \), then has ‘inside information’ if \( p(j) \neq q(j) \), and can establish a positive rate of return equal to \( D(p \| q) \).

Solution: From part f), \( (p(1), \ldots, p(\ell - 1), 0)^T \) is one choice for the optimizing \( \lambda^* \). Using this choice,

\[
E[Y(\lambda^*)] = \sum_{j<\ell} p(j) \ln[p(j)r(j)] = D(p \| q).
\]

To further clarify the notion of a fair game, put on rose-colored glasses to envision a race track that simply accumulates the bets from all gamblers and distributes all income to the bets on the winning horse. In this sense, \( q(j) = 1/r(j) \) is the ‘odds’ on horse \( j \) as established by the aggregate of the gamblers. Fairness is not a word that is always used the same way, and here, rather than meaning anything about probabilities and expectations, it simply refers to the unrealistic assumption that neither the track nor the horse owners receive any return.

Exercise 9.37: (Proof of Theorem 9.10.1) Let \( Z_n = \frac{1}{n} L_n(\lambda) - E[Y(\lambda)] \) and note that \( \lim_{n \to \infty} Z_n = 0 \) WP1. Recall that this means that the set of sample points \( \omega \) for which \( \lim Z_n(\omega) = 0 \) has probability 1. Let \( \epsilon > 0 \) be arbitrary but fixed throughout the exercise and let \( n_0 \geq 1 \) be an integer. Let \( A(n_0, \epsilon) = \{ \omega : |Z_n(\omega)| \leq \epsilon \text{ for all } n \geq n_0 \} \).

a) Consider an \( \omega \) for which \( \lim_{n \to \infty} Z_n(\omega) = 0 \) and explain why \( \omega \in A(n_0, \epsilon) \) for some \( n_0 \).
**Solution:** By the definition of the limit (to 0) of a sequence of real numbers, \( i.e., \) of \( \lim Z_n(\omega) \), there must be an \( n_0 \) large enough that \( |Z_n(\omega)| \leq \epsilon \) for all \( n \geq n_0 \). Clearly, \( \omega \in A(n_0, \epsilon) \) for that \( n_0 \).

b) Show that \( \Pr \left\{ U_{n_0}^\infty A(n_0, \epsilon) \right\} = 1. \)

**Solution:** The set of \( \omega \) for which \( \lim_{n \to \infty} Z_n(\omega) = 0 \) has probability 1, and each such \( \omega \) is in \( A(n_0, \epsilon) \) for some \( n_0 \). Thus each such \( \omega \) is in the above union, so that the union has probability 1.

c) Show that for any \( \delta > 0 \), there is an \( n_0 \) large enough that \( \Pr \{ A(n_0, \epsilon) \} > 1 - \delta \). Hint: Use (1.9).

d) Show that for any \( \epsilon > 0 \) and \( \delta > 0 \), there is an \( n_0 \) such that

\[
\Pr \left\{ \exp \left[ n\left( E[Y(\lambda)] - \epsilon \right) \right] \leq W_n(\lambda) \leq \exp \left[ n\left( E[Y(\lambda)] + \epsilon \right) \right] \text{ for all } n \geq n_0 \right\} > 1 - \delta.
\]

**Solution:** Part c) applies to any given \( \epsilon \), and (from the definition of \( A(n_0, \epsilon) \)) says that, for any \( \epsilon, \delta > 0 \), there is an \( n_0 \) such that \( \Pr \{ |Z_n| \leq \epsilon \text{ for all } n \geq n_0 \} > 1 - \delta. \) Substituting the definition of \( Z_n \) into this,

\[
\Pr \{ nE[Y(\lambda)] - \epsilon \leq L_n(\lambda) \leq nE[Y(\lambda)] \text{ for all } n \geq n_0 \}.
\]

Finally, by exponentiating the terms above, we get the desired expression.

**Exercise 9.38:** In this exercise, you are to express the buy-and-hold strategy and the two-pile strategy of (9.137) as special cases of a time-varying allocation.

a) Suppose the buy-and-hold strategy starts with an allocation \( \lambda_1 \). Let \( W_n(k) \) be the investors wealth in investment \( k \) at time \( n \). Show that

\[
W_n(k) = \lambda_1(k) \prod_{m=1}^{n} X_m(k).
\]

**Solution:** There is hardly anything to be shown. \( \lambda_1(k) \) is the original investment, and each \( X_m(k) \) is the ratio of wealth in \( k \) at the beginning and end of epoch \( m \), so the product gives the final wealth at epoch \( n \).

b) Show that \( \lambda_n(k) \) can be expressed in each of the following ways.

\[
\lambda_n(k) = \frac{W_{n-1}(k)}{\sum_j W_{n-1}(j)} = \frac{\lambda_{n-1}(k)X_{n-1}(k)}{\sum_j \lambda_{n-1}(j)X_{n-1}(j)}, \quad (A.43)
\]

**Solution:** Note that \( \lambda_n(k) \) is the fraction of total wealth in investment \( k \) at the beginning of epoch \( n \). Thus, as shown in the first equality above, it is the wealth in \( k \) at the end of epoch \( n - 1 \) divided by the total wealth at the end of epoch \( n - 1 \). For the second equality,
we start by rewriting the terms in the first equality as

\[ \lambda_n(k) = \frac{W_{n-2}(k)X_{n-1}(k)}{\sum_j W_{n-2}(j)X_{n-1}(j)} \]

\[ = \frac{W_{n-2}(k)X_{n-1}(k)}{\sum_j W_{n-2}(j)X_{n-1}(j)} \]

\[ = \frac{\lambda_{n-1}(k)X_{n-1}(k)}{\sum_j \lambda_{n-1}(j)X_{n-1}(j)}. \]

where in the last equality we have used the first equality in (A.43), applied to \( n - 2 \).

Note that there is no real reason for wanting to know \( \lambda_n(k) \) for any given investment strategy, other than simply convincing oneself that \( \lambda_n(k) \) (which depends on \( X_m(i) \) for all \( m < n \) and investments \( i \)) exists for any strategy and in fact specifies any given strategy.

c) For the two-pile strategy of (9.137), verify (9.137) and find an expression for \( \lambda_n(k) \).

**Solution:** We omit the parenthesis for \( k \) since there is only one investment (other than cash). For a constant allocation strategy, we see from (9.126) that \( \mathbb{E}[W_n(\lambda)] = (1 + \lambda/2)^n \).

The two-pile strategy invests \( 1/2 \) the initial capital in a constant allocation with \( \lambda = 1 \) and \( 1/2 \) with \( \lambda = 1/4 \). Thus the resulting expected capital is \( \mathbb{E}[W_n] = \frac{1}{2}(1 + \frac{1}{2})^n + \frac{1}{2}(1 + \frac{1}{8})^n \) as desired.

To find \( \mathbb{E}[L_n] \), note that for the \( n \)-tuple \( X_m = 3 \) for \( 1 \leq m \leq n \), we have \( L_n = \ln \left[ \frac{3}{2}3^n + \frac{1}{2}(3/2)^n \right] \). This \( n \)-tuple occurs with probability \( 2^{-n} \) and thus this value of \( L_n \) can be neglected in approximating \( \mathbb{E}[L_n] \). For all other \( n \)-tuples, \( W_n \) is \( 1/2 \) its value with the constant allocation \( \lambda = 1/4 \), so \( L_n \) is reduced by \( \ln(2) \) from the pure strategy value, giving rise to (9.137).

To find \( \lambda_n \), we look first at \( \lambda_1 \) and \( \lambda_2 \). Since each pile initially has \( 1/2 \) (of the initial investment), and \( 1/4 \) of the first pile is bet, and all of the second pile is bet, \( \lambda_1 = \frac{1}{2}(1/4) + \frac{1}{2} = 5/8 \).

Now note that \( \lambda_2 \) is a rv depending on \( X_1 \). If \( X_1 = 3 \), then the piles contain \( 3/4 \) and \( 3/2 \) respectively. Since \( 1/4 \) of the first pile is bet and all the second pile is bet, we have \( 3/16 + 3/2 \) bet overall out of \( 3/4 \) and \( 3/2 \). Thus, for \( X_1 = 3 \), we have \( \lambda_2 = 3/4 \). For \( X_1 = 0 \), only the first pile is non-empty and \( \lambda_2 = 1/4 \).

In general, \( \lambda_n = 1/4 \) except where \( X_m = 3 \) for \( 1 \leq m < n \) and then

\[ \lambda_n = \frac{1 + 2^{n+1}}{4 + 2^{n+1}}. \]

**Exercise 9.39:** a) Consider the martingale \( \{Z_n; n \geq 1\} \) where \( Z_n = W_n/W_n^* \), \( W_n \) is the wealth at time \( n \) for the pure triple-or-nothing strategy of Example 9.10.1 and \( W_n^* \) is the wealth using the constant allocation \( \lambda^* \). Find the PMF of \( Z \) where \( Z = \lim_{n \to \infty} Z_n \) WP1.

**Solution:** The martingale convergence theorem applies, so \( Z \) actually exists as a rv. Actually, the power of the martingale convergence theorem is wasted here, since it is obvious
that \( \lim_{n \to \infty} Z_n(\omega) = 0 \) for all sample sequences \( \omega \) except that for which \( X_n = 3 \) for all \( n \). Thus \( Z \) is the constant rv \( Z = 0 \) WP1.

b) Now consider the two-pile strategy of (9.137) and find the PMF of the limiting rv \( Z \) for this strategy.

**Solution:** The martingale convergence theorem again applies, and again is unnecessary since \( Z_n(\omega) = 1/2 \) for all \( \omega \) other than those for which \( X_m = 3 \) for all \( m \leq n \). Thus \( Z = \lim_{n \to \infty} Z_n = 1/2 \) WP1.

c) Now consider a ‘tempt-the-fates’ strategy where one uses the pure triple-or-nothing strategy for the first 3 epochs and uses the constant allocation \( \lambda^* \) thereafter. Again find the PMF of the limiting rv \( Z \).

**Solution:** This is slightly less trivial since \( Z \) is now a non-trivial (i.e., non-constant) rv. The martingale convergence theorem again applies, although the answer can be seen without using that theorem. If \( X_1 = 3 \), then \( W_1 = 3 \) and \( W_1^* = 3/2 \) so \( Z_1 = 2 \). If \( X_1 = 0 \), then \( Z_n = 0 \) for all \( n \geq 1 \).

If \( X_1 = X_2 = 3 \), then \( W_2 = 9 \) and \( W_2^* = (3/2)^2 \) so \( Z_2 = 4 \). If \( X_2 = 0 \), then \( Z_n = 0 \) for all \( n \geq 2 \). Similarly, if \( X_m = 3 \) for \( 1 \leq m \leq 3 \), then \( Z_n = 8 \) and otherwise \( Z_n = 0 \) for all \( n \geq m \).

Since the allocation \( \lambda^* \) is used for \( m > 3 \), we see that \( Z_m = Z_3^* \) for \( m \geq 3 \). Thus \( Z = \lim_{n \to \infty} Z_n = 8 \) with probability 1/8 (i.e., if \( X_1 = X_2 = X_3 = 3 \)) and \( Z = 0 \) with probability 7/8.

**Exercise 9.42:** Let \( \{ Z_n; n \geq 1 \} \) be a martingale, and for some integer \( m \), let \( Y_n = Z_{n+m} - Z_m \).

a) Show that \( \mathbb{E}[Y_n \mid Z_{n+m-1} = z_{n+m-1}, Z_{n+m-2} = z_{n+m-2}, \ldots, Z_m = z_m, \ldots, Z_1 = z_1] = z_{n+m-1} - z_m \).

**Solution:** This is more straightforward if the desired result is written in the more abbreviated form

\[
\mathbb{E}[Y_n \mid Z_{n+m-1}, Z_{n+m-2}, \ldots, Z_m, \ldots, Z_1] = Z_{n+m-1} - Z_m.
\]

Since \( Y_n = Z_{n+m} - Z_m \), the left side above is

\[
\mathbb{E}[Z_{n+m} - Z_m | Z_{n+m-1}, \ldots, Z_1] = Z_{n+m-1} - \mathbb{E}[Z_m \mid Z_{n+m-1}, \ldots, Z_m, \ldots, Z_1].
\]

Given sample values for each conditioning rv on the right of the above expression, and particularly given that \( Z_m = z_m \), the expected value of \( Z_m \) is simply the conditioning value \( z_m \) for \( Z_m \). This is one of those strange things that are completely obvious, and yet somehow obscure. We then have \( \mathbb{E}[Y_n \mid Z_{n+m-1}, \ldots, Z_1] = Z_{n+m-1} - Z_m \).

b) Show that \( \mathbb{E}[Y_n \mid Y_{n-1} = y_{n-1}, \ldots, Y_1 = y_1] = y_{n-1} \).

**Solution:** In abbreviated form, we want to show that \( \mathbb{E}[Y_n \mid Y_{n-1}, \ldots, Y_1] = Y_{n-1} \). We showed in part a) that \( \mathbb{E}[Y_n \mid Z_{n+m-1}, \ldots, Z_1] = Y_{n-1} \). For each sample point \( \omega \), \( Y_{n-1}(\omega), \ldots, Y_1(\omega) \) is a function of \( Z_{n+m-1}(\omega), \ldots, Z_1(\omega) \). Thus, the rv \( \mathbb{E}[Y_n \mid Z_{n+m-1}, \ldots, Z_1] \) specifies the rv \( \mathbb{E}[Y_n \mid Y_{n-1}, \ldots, Y_1] \), which then must be \( Y_{n-1} \).

c) Show that \( \mathbb{E} | Y_n | < \infty \). Note that b) and c) show that \( \{ Y_n; n \geq 1 \} \) is a martingale.
Solution: Since \( Y_n = Z_{n+m} - Z_m \), we have |\( Y_n | \leq |Z_{n+m}| + |Z_m| \). Since \( \{Z_n; n \geq 1 \} \) is a martingale, \( \mathbb{E} [|Z_n|] < \infty \) for each \( n \) so

\[
\mathbb{E} [|Y_n|] \leq \mathbb{E} [|Z_{n+m}|] + \mathbb{E} [|Z_m|] < \infty.
\]

Exercise 9.43: This exercise views the continuous-time branching process of Exercise 7.15 as a stopped random walk. Recall that the process was specified there as a Markov process such that for each state \( j \), \( j \geq 0 \), the transition rate to \( j+1 \) is \( j\lambda \) and to \( j-1 \) is \( j\mu \). There are no other transitions, and in particular, there are no transitions out of state 0, so that the Markov process is reducible. Recall that the embedded Markov chain is the same as the embedded chain of an M/M/1 queue except that there is no transition from state 0 to state 1.

a) To model the possible extinction of the population, convert the embedded Markov chain above to a stopped random walk, \( \{S_n; n \geq 0 \} \). The stopped random walk starts at \( S_0 = 0 \) and stops on reaching a threshold at \(-1\). Before stopping, it moves up by one with probability \( \frac{\lambda}{\lambda + \mu} \) and downward by 1 with probability \( \frac{\mu}{\lambda + \mu} \) at each step. Give the (very simple) relationship between the state \( X_n \) of the Markov chain and the state \( S_n \) of the stopped random walk for each \( n \geq 0 \).

Solution: Consider a random walk where \( S_n = Y_1 + \cdots + Y_n \) and the sequence \( \{Y_n; n \geq 0 \} \) is IID with the PMF

\[
p_{Y_n}(1) = \frac{\lambda}{\lambda + \mu} \quad \text{and} \quad p_{Y_n}(-1) = \frac{\mu}{\lambda + \mu}.
\]

We note that \( X_n \) in the Markov chain is the sum of 1 (for the initial organism) and \( \pm 1 \) for each subsequent birth or death transition. Since each of those birth and death transitions in the Markov chain, is an IID random variable so long as the number of organisms is positive, we see that \( S_n = X_n - 1 \) so long as the number of organisms is positive. When the number of organisms goes to 0, the number of deaths exceeds the number of births by 1, so \( S_n = -1 \). The stopped process, stopping when \( S_n = -1 \), is then identical to the Markov chain, which stays in state 0 after it is once entered.

b) Find the probability that the population eventually dies out as a function of \( \lambda \) and \( \mu \). Be sure to consider all three cases \( \lambda > \mu \), \( \lambda < \mu \), and \( \lambda = \mu \).

Solution: The probability \( P_d \) that the population dies out is the probability that the random walk in part a) crosses a threshold at \(-1\). If \( \lambda < \mu \), the random walk has a negative drift \( (i.e., \text{the rv}'s Y_n \text{ of part a) have a negative mean}) \), so the walk crosses the threshold eventually with probability 1 and \( P_d = 1 \).

For \( \lambda > \mu \), the random walk has a positive drift and the problem is identical to that of the negative of the random walk (where the step \(-Y \) has a negative drift) crossing a positive threshold at \(+1\). The probability of threshold crossing, (using the Wald identity and Corollary 7.1) is upper bounded by \( P_d \leq e^{-r^*} \) where \( r^* \) is the \( r \) where \( \gamma_{(-Y)}(r^*) = 0 \). Solving for \( r^* \), we get \( r^* = \ln(\mu/\lambda) \), so \( P_d \leq \mu/\lambda \). Since the random walk changes by \( \pm 1 \) at each step, and the threshold is an integer, this inequality is satisfied with equality, so \( P_d = \mu/\lambda \). This same analysis works for \( \lambda = \mu \), so \( P_d = 1 \) in that case.

An alternative approach is to note that this negative random walk with a threshold at \(+1\) is the same as Example 5.5.2 (Stop when you’re ahead in coin tossing) in the text.
Exercise 10.1: a) Consider the joint probability density \( f_{X,Z}(x,z) = e^{-z} \) for \( 0 \leq x \leq z \) and \( f_{X,Z}(x,z) = 0 \) otherwise. Find the pair \( x,z \) of values that maximize this density. Find the marginal density \( f_z(z) \) and find the value of \( z \) that maximizes this.

Solution: \( e^{-z} \) has value 1 when \( x = z = 0 \), and the joint density is smaller whenever \( z > 0 \), and is zero when \( z < 0 \), so \( p_{X,Z}(x,z) \) is maximized by \( x = z = 0 \). The marginal density is found by integrating \( p_{X,Z}(x,z) = e^{-z} \) over \( x \) in the range 0 to \( z \), so \( p_z(z) = ze^{-z} \) for \( z \geq 0 \). This is maximized at \( z = 1 \).

b) Let \( f_{X,Z}(x,z,y) \) be \( y^2e^{-yz} \) for \( 0 \leq x \leq z, 1 \leq y \leq 2 \) and be 0 otherwise. Conditional on an observation \( Y = y \), find the joint MAP estimate of \( X, Z \). Find \( f_{Z|Y}(z|y) \), the marginal density of \( Z \) conditional on \( Y = y \), and find the MAP estimate of \( Z \) conditional on \( Y = y \).

Solution: The joint MAP estimate is the value of \( x, z \), in the range \( 0 \leq x \leq z \), that maximizes \( p_{X,Z|Y}(x,z|y) = p_{X,Z}(x,z,y)/p_Y(y) = (y^2e^{-yz})/p_Y(y) \). For any given \( y, 0 \leq y \leq 1 \), this is maximized, as in part a, for \( x = z = 0 \). Next, integrating \( p_{X,Z,Y}(x,z,y) \) over \( x \) from 0 to \( z \), we get \( p_{Z,Y}(z,y) = y^2ze^{-yz} \). This, and thus \( p_{Z|Y}(z|y) \) is maximized by \( z = 1/y \), which is thus the MAP estimate for \( Z \) alone.

Exercise 10.3: a) Let \( X, Z_1, Z_2, \ldots, Z_n \) be independent zero-mean Gaussian rv’s with variances \( \sigma_X^2, \sigma_{Z_1}^2, \ldots, \sigma_{Z_n}^2 \), respectively. Let \( Y_j = h_jX + Z_j \) for \( j \geq 1 \) and let \( Y = (Y_1, \ldots, Y_n)^T \). Use (10.9) to show that the MMSE estimate of \( X \) conditional on \( Y = y = (y_1, \ldots, y_n)^T \), is given by

\[
\hat{x}(y) = \sum_{j=1}^{n} g_j y_j; \quad \text{where} \quad g_j = \frac{h_j/\sigma_{Z_j}^2}{1/\sigma_X^2 + \sum_{i=1}^{n} h_i^2/\sigma_{Z_i}^2},
\]

(A.44)

Hint: Let the row vector \( g^T \) be \([K_{X,Y}] [K_Y^{-1}] \) and multiply \( g^T \) by \( K_Y \) to solve for \( g^T \).

Solution: From (10.9), we see that \( \hat{x}(y) = g^T y \) where \( g^T = [K_{X,Y}] [K_Y^{-1}] \). Since \( Y = hX + Z \), we see that \([K_Y] = h\sigma_X^2 h^T + [K_Z] \) and \([K_{X,Y}] = \sigma_X^2 h^T \). Thus we want to solve the vector equation \( g^T h \sigma_X^2 h^T + g^T [K_Z] = \sigma_X^2 h^T \). Since \( g^T h \) is a scalar, we can rewrite this as \((1 - g^T h) \sigma_X^2 h^T = g^T [K_Z] \). The \( j \)th component of this equation is

\[
g_j = \frac{(1 - g^T h) \sigma_X^2 h_j}{\sigma_{Z_j}^2}
\]

(A.45)

This shows that the weighting factors \( g_j \) in \( \hat{x}(y) \) depend on \( j \) only through \( h_j/\sigma_{Z_j} \), which is reasonable. We still must determine the unknown constant \( 1 - g^T h \). To do this, multiply (A.45) by \( h_j \) and sum over \( j \), getting

\[
g^T h = (1 - g^T h) \sum_j \frac{\sigma_X^2 h_j^2}{\sigma_{Z_j}^2}.
\]

Solving for \( g^T h \), from this,

\[
g^T h = \frac{\sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}; \quad 1 - g^T h = \frac{1}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}.
\]

(A.46)

Substituting the expression for \( 1 - g^T h \) into (A.45) yields (A.44).
b) Let $\xi = X - \hat{X}(Y)$ and show that (10.29) is valid, i.e., that

$$1/\sigma_\xi^2 = 1/\sigma_X^2 + \sum_{i=1}^{n} \frac{h_i^2}{\sigma_{Z_i}^2}.$$ 

**Solution:** Using (10.6) in one dimension, $\sigma_\xi^2 = E[\xi X] = \sigma_X^2 - E[\hat{X}(Y)X]$. Since $\hat{X}(Y) = \sum_j g_j Y_j$ from (A.44), we have

$$\sigma_\xi^2 = \sigma_X^2 - \sum_{i=1}^{n} g_i E[Y_iX] = \sigma_X^2 \left(1 - \sum_{i=1}^{n} g_i h_i\right)$$

$$= \sigma_X^2 \left(1 - \mathbf{g}^T \mathbf{h}\right) = \frac{1}{1 + \sum_j \sigma_X^2 h_j^2 / \sigma_{Z_j}^2} = \frac{1}{1/\sigma_X^2 + \sum_j \sigma_X^2 h_j^2 / \sigma_{Z_j}^2},$$

where we have used (A.46). This is equivalent to (10.29).

e) Show that (10.28), i.e., $\hat{x}(y) = \sigma_\xi^2 \sum_{i=1}^{n} h_i y_i / \sigma_{Z_i}^2$, is valid.

**Solution:** Substitute the expression for $\sigma_\xi^2$ in part b) into (A.44).

d) Show that the expression in (10.29) is equivalent to the iterative expression in (10.27).

**Solution:** First, we show that (10.29) implies (10.27). We use $\xi_n$ to refer to the error for $n$ observations and $\xi_{n-1}$ for the error using the first $n-1$ of those observations. Using (10.29),

$$\frac{1}{\sigma_{\xi_n}^2} = \frac{1}{\sigma_X^2} + \sum_{i=1}^{n} \frac{h_i^2}{\sigma_{Z_i}^2} = \frac{1}{\sigma_X^2} + \sum_{i=1}^{n-1} \frac{h_i^2}{\sigma_{Z_i}^2} + \frac{h_n^2}{\sigma_{Z_n}^2}$$

$$= \frac{1}{\sigma_{\xi_{n-1}}^2} + \frac{h_n^2}{\sigma_{Z_n}^2},$$

(A.47)

which is (10.27). This holds for all $n$, so (10.27) for all $n$ also implies (10.29).

e) Show that the expression in (10.28) is equivalent to the iterative expression in (10.25).

**Solution:** Breaking (10.28) into the first $n-1$ terms followed by the term for $n$, we get

$$\hat{x}(y_1^n) = \sigma_{\xi_n}^2 \sum_{j=1}^{n-1} \frac{h_j y_j}{\sigma_{Z_j}^2} + \sigma_{\xi_n}^2 \frac{h_n y_n}{\sigma_{Z_n}^2} = \frac{\sigma_{\xi_n}^2}{\sigma_{\xi_{n-1}}^2} \hat{x}(y_1^{n-1}) + \sigma_{\xi_n}^2 \frac{h_n y_n}{\sigma_{Z_n}^2},$$

(A.48)

where we used (10.28) for $\hat{y}_1^{n-1}$. We can solve for $\sigma_{\xi_n}/\sigma_{\xi_{n-1}}$ by multiplying (A.47) by $\sigma_{\xi_n}^2$, getting

$$\frac{\sigma_{\xi_n}^2}{\sigma_{\xi_{n-1}}^2} = 1 - \frac{h_n^2}{\sigma_{Z_n}^2},$$

Substituting this into (A.48) yields

$$\hat{x}(y_1^n) = \hat{x}(y_1^{n-1}) + \sigma_{\xi_n}^2 \frac{h_n y_n - h_n^2 \hat{x}(y_1^{n-1})}{\sigma_{Z_n}^2}.$$
Finally, if we invert (A.47), we get
\[
\sigma^2_{n-1} = \frac{\sigma^2_{n} - \sigma^2_{Z_n}}{h^2_{n} \sigma^2_{Z_{n-1}} + \sigma^2_{Z_n}}.
\]
Substituting this into (A.48), we get (10.27).

**Exercise 10.8:** For a real inner-product space, show that \( n \) vectors, \( Y_1, \ldots, Y_n \) are linearly dependent if and only if the matrix of inner products, \( \{ Y_j, Y_k \}, 1 \leq j, k \leq n \), is singular.

**Solution:** Let \( K \) be the matrix of inner products, \( \{ Y_j, Y_i \}, 1 \leq i, j \leq n \), and let \( K_i = (Y_1, Y_i, \ldots, Y_n, Y_i)^T \) be the \( i \)th column of \( K \). Then \( K \) is singular iff there is a non-zero vector \( a = (a_1, \ldots, a_n)^T \) such that \( \sum_i a_i K_i = 0 \). This vector equation, written out by components, is \( \sum_i a_i (Y_j, Y_i) = 0 \) for \( 1 \leq j \leq n \). This can be rewritten as \( \langle Y_j, \sum_i a_i Y_i \rangle = 0 \) for \( 1 \leq j \leq n \). Thus, if \( K \) is singular, this equation is satisfied for \( 1 \leq j \leq n \), and thus also \( \langle \sum_j a_j Y_j, \sum_i a_i Y_i \rangle = 0 \). This implies that \( \sum_i a_i Y_i = 0 \) and \( \{ Y_i; 1 \leq i \leq n \} \) is a linearly dependent set. Conversely, if \( \{ Y_i; 1 \leq i \leq n \} \) is a linearly dependent set, then there is a non-zero \( a \) such that \( \sum_i a_i Y_i = 0 \), so \( \langle Y_j, \sum_i a_i Y_i \rangle = 0 \) for \( 1 \leq j \leq n \) and \( K \) is singular.

**Exercise 10.9:** From (10.90),
\[
b = E[X] - \sum_j a_j E[Y_j].
\]  

Note that \( E[XY_i] = E[X] E[Y_i] + K_{XY_i} \) and \( E[Y_i Y_j] = E[Y_i] E[Y_j] + K_{Y_i Y_j} \). Substituting this in (10.91),
\[
E[X] E[Y_i] + K_{XY_i} = b E[Y_i] + \sum_j a_j E[Y_j] E[Y_j] + \sum_j a_j K_{Y_i Y_j}.
\]  

Multiplying (A.49) by \( E[Y_i] \) and subtracting from (A.50),
\[
K_{XY_i} = \sum_j a_j K_{Y_i Y_j}.
\]  

This implies that \( a^T = K_{XY} K_{Y}^{-1} \) so that \( \hat{X}(Y_1, \ldots, Y_n) = b + a^T Y \) in 10.89 agrees with (10.12).

**Exercise 10.12:** (Derivation of circularly symmetric Gaussian density) Let \( X = X_{re} + i X_{im} \) be a zero-mean circularly symmetric \( n \) dimensional Gaussian complex rv. Let \( U = (X_{re}^T, X_{im}^T)^T \) be the corresponding \( 2n \) dimensional real rv. Let \( K_{re} = E[X_{re} X_{re}^T] \) and \( K_{ri} = E[X_{re} X_{im}^T] \).

a) Show that
\[
K_{u} = \begin{bmatrix} K_{re} & K_{ri} \\ -K_{ri} & K_{re} \end{bmatrix}.
\]

**Solution:**
\[
E[X X^T] = E[(X_{re} + j X_{im})(X_{re} + j X_{im})^T] = E[X_{re} X_{re}^T] - E[X_{im} X_{im}^T] + j E[X_{im} X_{re}^T] + j E[X_{im} X_{re}^T].
\]

Since this must be zero from the circular symmetry constraint,
\[
K_{re} = E[X_{re} X_{re}^T] = E[X_{im} X_{im}^T],
K_{ri} = E[X_{re} X_{im}^T] = -E[X_{im} X_{re}^T].
\]
Note that since $K_U$ is a covariance matrix, $K_{ri}^T = -K_{ri}$.

b) Show that

$$K_{ri}^{-1} = \begin{bmatrix} B & C \\ -C & B \end{bmatrix}.$$  

and find the $B, C$ for which this is true.

Solution: From (2.35), the upper left portion of $K_U$ is $B = K_{re} + K_{ri}K_{re}^{-1}K_{ri}$. The same formula applies to the lower right matrix, so it is $B$ also. Similarly, from (2.36), the upper right matrix is $C = -BK_{ri}K_{re}^{-1}$. Similarly, the lower left matrix is $-C$.

c) Show that $K_X = 2(K_{re} - K_{ri})$.

Solution:

$$K_X = E \left[ (X_{re} + iX_{im})(X_{re} - iX_{im})^T \right]$$

$$= E \left[ X_{re}X_{re}^T + [X_{im}X_{im}^T] + i [X_{im}X_{re}^T] - [X_{re}X_{im}^T] \right]$$

$$= 2K_{re} - 2iK_{ri}.$$  

d) Show that $K_X^{-1} = \frac{1}{2}(B - iC)$.

Solution: We are to show that $K_X^{-1} = \frac{1}{2}(B - iC)$. This is equivalent to showing that $K_X \frac{1}{2}(B - iC) = I_n$, which is equivalent to showing that $(K_{re} - iK_{ri})(B - iC) = I_n$. To show this, consider

$$I_{2n} = K_U K_{ri}^{-1} = \begin{bmatrix} K_{re} & K_{ri} \\ -K_{ri} & K_{re} \end{bmatrix} \begin{bmatrix} B & C \\ -C & B \end{bmatrix}.$$  

Expanding this, we see that $I_n = K_{re}B - K_{ri}C$ and $0 = K_{re}C + K_{ri}B$. Since the matrices $K_{re}, K_{ri}, B,$ and $C$ are real, this is equivalent to $(K_{re} - iK_{ri})(B - iC) = I_n$.

e) Define $f_X(x) = f_U(u)$ for $u = (x_{re}, x_{im})^T$ and show that

$$f_X(x) = \frac{\exp -x^TK_{ri}^{-1}x^T}{(2\pi)^n \sqrt{\det[K_U]}}.$$  

Solution: Since $U$ is a real zero mean Gaussian rv, its density is

$$f_U(u) = \frac{\exp -\frac{1}{2} [u^TK_U^{-1}u]}{(2\pi)^n \sqrt{\det[K_U]}}.$$  

Expanding the term in brackets, we get

$$u^TK_U^{-1}u^T = (x_{re}, x_{im}) \begin{bmatrix} B & C \\ -C & B \end{bmatrix} \begin{bmatrix} x_{re} \\ x_{im} \end{bmatrix}$$

$$= x_{re}^TBx_{re} + x_{im}^TBx_{im} + x_{re}^TCx_{im} - x_{im}^TCx_{re}. \quad \text{(ii)}$$

Since $B = B^T$, the first two terms in $(ii)$ can be rewritten as $(x_{re} - jx_{im})^TB(x_{re} + jx_{im})$. Since $C = -C^T$, we have $x^TCx = 0$ for any real $x$ (and thus for both $x_{re}$ and $x_{im}$). Thus
the latter two terms in (ii) can be rewritten as \((x_{re} - ix_{im})^T(-iC)(x_{re} + ix_{im})\). Thus, we have
\[
u^T K^{-1}_U \nu = x^T(B - iC)x = 2x^T K^{-1}_X x.
\]
Substituting this back in (i), we have the desired expression.

f) Show that
\[
det[K_U] = \det \begin{bmatrix} K_{re} + iK_{ri} & K_{ri} - iK_{re} \\ -K_{ri} & K_{re} \end{bmatrix}.
\]

Hint: Recall that elementary row operations do not change the value of a determinant.

Solution: Start with the \(2n \times 2n\) matrix \(K_U\) and add \(-i\) times row \(n+1\) to row 1. Similarly, for each \(j, 2 \leq j \leq n\), add \(-i\) times row \(n + j\) to row \(j\). This yields
\[
det[K_U] = \det \begin{bmatrix} K_{re} + iK_{ri} & K_{ri} - iK_{re} \\ -K_{ri} & K_{re} \end{bmatrix}.
\]

g) Show that
\[
det[K_U] = \det \begin{bmatrix} K_{re} + iK_{ri} & 0 \\ -K_{ri} & K_{re} - iK_{ri} \end{bmatrix}.
\]

Hint: Recall that elementary column operations do not change the value of a determinant.

Solution: This is essentially the same as part f). Add \(i\) times each of the first \(n\) columns to the corresponding final \(n\) columns to yield
\[
det[K_U] = \det \begin{bmatrix} K_{re} + iK_{ri} & 0 \\ -K_{ri} & K_{re} - iK_{ri} \end{bmatrix}.
\]

h) Show that
\[
det[K_U] = 2^{-2n} (det[K_X])^2,
\]
and from this conclude that (3.108) is valid.

Solution: The determinant in part g) is just the determinant of the upper left block times the determinant of the lower right block. The upper left block is just \((1/2)K^2_X\) and the lower right block is \((1/2)K_X\). Since multiplying each element of an \(n \times n\) matrix by \(1/2\) multiplies the determinant by \((1/2)^n\), we have
\[
det[K_U] = 2^{-2n}(det[K_X]).
\]

Exercise 10.13: (Alternate derivation of circularly symmetric Gaussian density)

a) Let \(X\) be a circularly symmetric zero-mean complex Gaussian rv with covariance 1. Show that
\[
f_X(x) = \frac{\exp -x^*x}{\pi}.
\]
Recall that the real part and imaginary part each have variance $1/2$.

**Solution:** Since $X_{re}$ and $X_{im}$ are independent and $\sigma^2_{X_{re}} = \sigma^2_{X_{im}} = \frac{1}{2}$, we have

$$f_X(x) = \frac{1}{2\pi\sigma_{X_{re}}\sigma_{X_{im}}} \exp \left[ -\frac{x_{re}^2}{2\sigma^2_{X_{re}}} - \frac{x_{im}^2}{2\sigma^2_{X_{im}}} \right] = \frac{\exp -x^T x}{\pi^n}.$$  

b) Let $X$ be an $n$ dimensional circularly symmetric complex Gaussian zero-mean random vector with $K_X = I_n$. Show that

$$f_X(x) = \frac{\exp -x^T x}{\pi^n}.$$  

**Solution:** Since all real and imaginary components are independent, the joint probability over all $2n$ components is the product of $n$ terms with the form in a).

$$f_X(x) = \exp -\sum_{i=1}^n x_i^* x_i = \frac{\exp -x^T x}{\pi^n}.$$  

c) Let $Y = HX$ where $H$ is $n \times n$ and invertible. Show that

$$f_Y(y) = \frac{\exp \left[ -y^T (H^{-1})^\dagger H^{-1} y \right]}{v \pi^n},$$

where $v$ is $dy/dx$, the ratio of an incremental $2n$ dimensional volume element after being transformed by $H$ to that before being transformed.

**Solution:** The argument here is exactly the same as that in Section 3.3.4. Since $Y = HX$, we have $X = H^{-1} Y$, so for any $y$ and $x = H^{-1} y$,

$$f_Y(y) |dy| = f_X(x) |dx|.$$  

$$f_Y(y) = \frac{f_X(H^{-1} y)}{|dy|/|dx|}.$$  

Substituting the result of b) into this,

$$f_Y(y) = \frac{\exp \left[ -y^T (H^*H)^{-1} H^{-1} y \right]}{|dy|/|dx|}.$$  

d) Use this to show that that (3.108) is valid.

**Solution:** View $|dx|$ as an incremental volume in $2n$ dimensional space ($n$ real and $n$ imaginary components) and view $|dy|$ as the corresponding incremental volume for $y = Hx$, i.e.,

$$\begin{bmatrix} y_{re} \\ y_{im} \end{bmatrix} = \begin{bmatrix} H_{re} & -H_{im} \\ H_{im} & H_{re} \end{bmatrix} \begin{bmatrix} x_{re} \\ x_{im} \end{bmatrix}.$$  

We then have

$$\frac{|dy|}{|dx|} = \left| \det \begin{bmatrix} H_{re} & -iH_{im} \\ H_{im} & H_{re} \end{bmatrix} \right| = \left| \det \begin{bmatrix} H_{re} + iH_{im} & iH_{re} - H_{im} \\ H_{im} & H_{re} \end{bmatrix} \right| = \left| \det[H] \det[H^*] \right| = \det[K_Y].$$
Exercise 10.15: a) Assume that for each parameter value \( x \), \( Y \) is Gaussian, \( \mathcal{N}(x, \sigma^2) \). Show that \( V_x(y) \) as defined in (10.103) is equal to \( (y - x)/\sigma^2 \) and show that the Fisher information is equal to \( 1/\sigma^2 \).

Solution: Note that we can view \( Y \) as \( X + Z \) where \( Z \) is \( \mathcal{N}(0, \sigma^2) \). We have \( f(y|x) = (2\pi\sigma^2)^{-1/2} \exp((y-x)/2\sigma^2) \). Thus \( df(y|x)/dx = [(y-x)/\sigma^2](2\pi\sigma^2)^{-1/2} \exp((y-x)^2/(2\sigma^2)) \), so \( V_x(y) = (y - x)/\sigma^2 \). Then the random variable \( V_x(Y) \), conditional on \( x \), is \( (Y - x)/\sigma^2 \). Since \( Y \sim \mathcal{N}(x, \sigma^2) \), the variance of \( (Y - x)/\sigma^2 \), conditional on \( x \), is \( \sigma^2/\sigma^4 = \sigma^{-2} \).

b) Show that the Cramer-Rao bound is satisfied with equality for ML estimation for this example. Show that if \( X \) is \( \mathcal{N}(0, \sigma^2_x) \), the MMSE estimate satisfies the Cramer-Rao bound (including bias) with equality.

Solution: For the ML estimate, \( \hat{X}_{ML} = y \), and for a given \( x \), \( E_x[\hat{X}_{ML}] = Ex[Y] = x \). Thus \( b(x) = 0 \) and the estimate is unbiased. The estimation error is just \( Z \), so the mean square estimation error is \( \sigma^2 \) for each \( x \). The Cramer-Rao bound is \( 1/J(x) = \sigma^2 \), so the bound is met with equality.