Stochastic Processes, Theory for Applications

Solutions to Selected Exercises

R.G.Gallager
October 5, 2014

The complete set of solutions is available to instructors teaching this course. Contact Cambridge Press at www.Cambridge.org.

The solutions here occasionally refer to theorems, corollaries, and lemmas in the text. The numbering of theorems etc. in the text is slightly different from that of the draft of Chapters 1-3 on my web site. Theorems, corollaries, lemmas, definitions, and examples are numbered separately on the web site, but numbered collectively in the text. Thus students using the web site must use some care in finding the theorem, etc. that is being referred to.

The author gratefully acknowleges the help of Shan-Yuan Ho, who has edited many of these solutions, and of a number of teaching assistants, particulartly Natasha Blitvic and Mina Karzand, who wrote earlier drafts of solutions. Thanks are also due to Kluwer Academic Press for permission to use a number of exercises that also appeared in Gallager, 'Discrete Stochastic Processes,' Kluwer, 1995. The original solutions to those exercises were prepared by Shan-Yuan Ho, but changed substantially here by the author both to be consistent with the new text and occasionally for added clarity.

The author will greatly appreciate being notified of typos, errors, different approaches, and lapses in clarity at gallager@mit.edu.

A.1 Solutions for Chapter 1

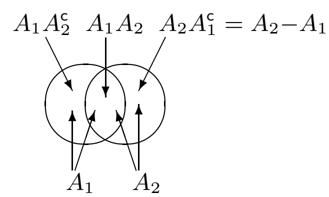
Exercise 1.2: This exercise derives the probability of an arbitrary (non-disjoint) union of events, derives the union bound, and derives some useful limit expressions.

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

$$A_1 \bigcup A_2 = A_1 \bigcup (A_2 - A_1), \tag{A.1}$$

where $A_2 - A_1 = A_2 A_1^{\mathsf{c}}$. 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 ω is in A_1 or A_1^{c}, but not both. Thus each ω is in exactly one of A_1, $A_1^{\mathsf{c}}A_2$ or $A_1^{\mathsf{c}}A_2^{\mathsf{c}}$. In the first two cases, ω 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,



b) For any $n \geq 2$ and arbitrary events $A_1, \dots, A_n$, define $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$. Show that $B_1, B_2, \dots$ 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 $B_1 = A_1$. From (a) B_1 and B_2 are disjoint and (from (A.1)), $A_1 \bigcup A_2 = B_1 \bigcup B_2$. 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 $C_2 = B_1 \bigcup B_2$, which forms the basis for the induction. We assume that $C_{n-1} = \bigcup_{i=1}^{n-1} B_i$ and prove that $C_n = \bigcup_{i=1}^{n} B_i$.

$$\begin{aligned} C_n &= C_{n-1} \bigcup A_n = C_{n-1} \bigcup A_n C_{n-1}^{\mathsf{c}} \\ &= C_{n-1} \bigcup B_n = \bigcup_{i-1}^{n} B_i. \end{aligned}$$

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 $B_1, \dots, B_{n-1}$.

c) Show that

$$\Pr\Big\{\bigcup_{n=1}^{\infty} A_n\Big\} = \Pr\Big\{\bigcup_{n=1}^{\infty} B_n\Big\} = \sum_{n=1}^{\infty} \Pr\{B_n\}.$$

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 the third axiom of probability.

d) Show that for each n, $\Pr\{B_n\} \leq \Pr\{A_n\}$. Use this to show that

$$\Pr\Big\{\bigcup_{n=1}^{\infty} A_n\Big\} \leq \sum_{n=1}^{\infty} \Pr\{A_n\}.$$

Solution: Since $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$, we see that $\omega \in B_n$ implies that $\omega \in A_n$, *i.e.*, that $B_n \subseteq A_n$. From (1.5), this implies that $\Pr\{B_n\} \leq \Pr\{A_n\}$ for each n. Thus

$$\Pr\Bigl\{\bigcup\nolimits_{n=1}^{\infty} A_n\Bigr\} \;=\; \sum\nolimits_{n=1}^{\infty} \Pr\{B_n\} \;\leq\; \sum\nolimits_{n=1}^{\infty} \Pr\{A_n\}\,.$$

e) Show that $\Pr\bigl\{\bigcup_{n=1}^{\infty} A_n\bigr\} = \lim_{n\to\infty} \Pr\bigl\{\bigcup_{i=1}^{n} A_i\bigr\}$. Hint: Combine (c) and (b). 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 (c),

$$\Pr\Bigl\{\bigcup\nolimits_{n=1}^{\infty} A_n\Bigr\} \;=\; \sum\nolimits_{n=1}^{\infty} \Pr\{B_n\} \;=\; \lim_{k\to\infty} \sum\nolimits_{n=1}^{k} \Pr\{B_n\}\,.$$

From (b), however,

$$\sum\nolimits_{n=1}^{k} \Pr\{B_n\} \;=\; \Pr\left\{\bigcup_{n=1}^{k} B_n\right\} \;=\; \Pr\left\{\bigcup_{n=1}^{k} A_n\right\}.$$

Combining the first equation with the limit in k of the second yields the desired result.

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

Solution: Using De Morgans equalities,

$$\begin{aligned}
\Pr\left\{\bigcap_{n=1}^{\infty} A_n\right\} \;&=\; 1 - \Pr\left\{\bigcup_{n=1}^{\infty} A_n^c\right\} \;=\; 1 - \lim_{k\to\infty} \Pr\left\{\bigcup_{n=1}^{k} A_n^{\mathsf{c}}\right\} \\
&=\; \lim_{k\to\infty} \Pr\left\{\bigcap_{n=1}^{k} A_n\right\}.
\end{aligned}$$

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 $\Pr\{A_1A_2A_3\} = \Pr\{A_1\}\Pr\{A_2\}\Pr\{A_3\}$.

a) Create an example where $\Pr\{A_1A_2\} = \Pr\{A_1A_3\} = \frac{1}{4}$ but $\Pr\{A_2A_3\} = \frac{1}{8}$. 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 one additional sample point that contains A_1 and A_2 but not A_3. Similarly, there must be one sample point that contains A_1 and A_3 but not A_2. These points give rise to the first three rows in the table below. There can be no additional sample point containing A_2 and A_3 since $\Pr\{A_2A_3\} = 1/8$. Thus each remaining sample point can be in at most 1 of the events A_1, A_2, and A_3. Since $\Pr\{A_i\} = 1/2$ for $1 \leq i \leq 3$ two sample points must contain A_2 alone, two must contain A_3 alone, and a single sample point must contain A_1 alone. This uniquely specifies the table below except for which sample point lies in each event.

Sample points	A_1	A_2	A_3
1	1	1	1
2	1	1	0
3	1	0	1
4	1	0	0
5	0	1	0
6	0	1	0
7	0	0	1
8	0	0	1

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 $\Pr\{A_1^c A_2^c A_3^c\} = 0 \neq \Pr\{A_1^c\}\Pr\{A_2^c\}\Pr\{A_3^c\}$, further reinforcing the conclusion that A_1, A_2, A_3 are not statistically independent. 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 1.9: (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

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}. \tag{A.2}$$

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.2) and from the above monotone property that

$$\binom{n}{k} < \sqrt{\frac{n}{2\pi k(n-k)}}\;\frac{n^n}{k^k(n{-}k)^{n-k}}.$$

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

Solution: Since the ratio of the first two terms of (A.2) is increasing in n, we have

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

Rearranging terms, we have the result in the hint. Applying the first inequality of (A.2) to $n-k$ and combining this with the result on $n!/k!$ yields the desired result.

b) Use the result of (a) to upper bound $\mathsf{p}_{S_n}(k)$ by

$$\mathsf{p}_{S_n}(k) < \sqrt{\frac{n}{2\pi k(n-k)}}\;\frac{p^k(1-p)^{n-k}n^n}{k^k(n{-}k)^{n-k}}.$$

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

Solution: Using the binomial equation and then (a),

$$\mathsf{p}_{S_n}(k) = \binom{n}{k}p^k(1-p)^{n-k} < \sqrt{\frac{n}{2\pi k(n-k)}}\;\frac{n^n}{k^k(n{-}k)^{n-k}}p^k(1-p)^{n-k}.$$

This is the the desired bound on $\mathsf{p}_{S_n}(k)$. Letting $\tilde{p} = k/n$, this becomes

$$\begin{aligned}
\mathsf{p}_{S_n}(\tilde{p}n) &< \sqrt{\frac{1}{2\pi n\tilde{p}(1-\tilde{p})}}\;\frac{p^{\tilde{p}n}(1-p)^{n(1-\tilde{p})}}{\tilde{p}^{\tilde{p}n}(1-\tilde{p})^{n(1-\tilde{p})}} \\
&= \sqrt{\frac{1}{2\pi n\tilde{p}(1-\tilde{p})}}\;\exp\left(n\left[\tilde{p}\ln\frac{p}{\tilde{p}} + \tilde{p}\ln\frac{1-p}{1-\tilde{p}}\right]\right),
\end{aligned}$$

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

c) Show that

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

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

$$\begin{aligned}
\binom{n}{k} &> \sqrt{\frac{n}{2\pi k(n-k)}}\;\frac{n^n}{k^k(n{-}k)^{n-k}}\exp\left(-\frac{1}{12k} - \frac{1}{12(n-k)}\right) \\
&> \sqrt{\frac{n}{2\pi k(n-k)}}\;\frac{n^n}{k^k(n{-}k)^{n-k}}\left[1 - \frac{n}{12k(n-k)}\right],
\end{aligned}$$

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 $\tilde{p}n$ for k in the solution to c) and substituting this in the binomial formula.

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

Solution: It is obvious that $\phi(p,\hat{p}) = 0$ for $\tilde{p} = p$. Taking the first two derivatives of $\phi(p,\tilde{p})$ with respect to $\tilde{p}$,

$$\frac{\partial\phi(p,\tilde{p})}{\partial\tilde{p}} = -\ln\left(\frac{p(1-\tilde{p})}{\tilde{p}(1-p)}\right) \qquad\qquad \frac{\partial f^2(p,\tilde{p})}{\partial\tilde{p}^2} = \frac{1}{\tilde{p}(1-\tilde{p})}.$$

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

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

Solution: We have seen in (1.34) that

$$\mathsf{E}[Y] = -\int_{y<0}\mathsf{F}_Y(y)\,dy + \int_{y\geq 0}\mathsf{F}^{\mathsf{c}}_Y(y)\,dy.$$

Since all negative values of Y become positive in $|Y|$,

$$\mathsf{E}\left[|Y|\right] = +\int_{y<0} \mathsf{F}_Y(y)\,dy + \int_{y\ge 0} \mathsf{F}^{\mathsf{c}}_Y(y)\,dy.$$

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 $|Y| = Y^+ - Y^- = Y^+ + |Y^-|$. Since $\mathsf{E}\left[Y^+\right] = \int_{y\ge 0} \mathsf{F}^{\mathsf{c}}_Y(y)\,dy$ and $\mathsf{E}\left[Y^-\right] = -\int_{y<0} \mathsf{F}_Y(y)\,dy$, the above results follow.

b) For some given rv X with $\mathsf{E}\left[|X|\right] < \infty$, let $Y = X - \alpha$. Using (a), show that

$$\mathsf{E}\left[|X-\alpha|\right] = \int_{-\infty}^{\alpha} \mathsf{F}_X(x)\,dx + \int_{\alpha}^{\infty} \mathsf{F}^{\mathsf{c}}_X(x)\,dx.$$

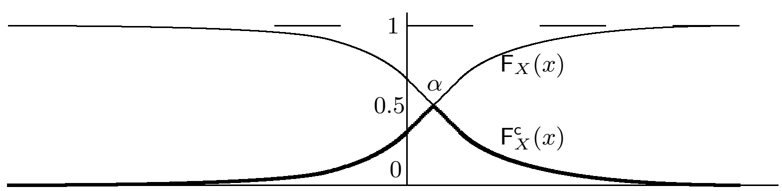
Solution: This follows by changing the variable of integration in (a). That is,

$$\begin{aligned}\mathsf{E}\left[|X-\alpha|\right] &= \mathsf{E}\left[|Y|\right] = +\int_{y<0} \mathsf{F}_Y(y)\,dy + \int_{y\ge 0} \mathsf{F}^{\mathsf{c}}_Y(y)\,dy \\ &= \int_{-\infty}^{\alpha} \mathsf{F}_X(x)\,dx + \int_{\alpha}^{\infty} \mathsf{F}^{\mathsf{c}}_X(x)\,dx,\end{aligned}$$

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

c) Show that $\mathsf{E}\left[|X-\alpha|\right]$ is minimized over α by choosing α 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 $\mathsf{F}_X(x)$ to $\mathsf{F}^{\mathsf{c}}_X(x)$ at $x = \alpha$. If $\mathsf{F}_X(x)$ is strictly increasing in x, then $\mathsf{F}^{\mathsf{c}}_X = 1 - \mathsf{F}_X$ is strictly decreasing. We then minimize the integrand over all x by choosing α to be the point where the curves cross, *i.e.*, where $\mathsf{F}_X(x) = .5$. 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 $\mathsf{F}_X(x) = .5$; all points on this interval are medians and also minimize the integral; Exercise 1.10 (c) gives an example where $\mathsf{F}_X(x) = 0.5$ over the interval $[1, 2)$. Finally, if $\mathsf{F}_X(\alpha) \ge 0.5$ and $\mathsf{F}_X(\alpha - \epsilon) < 0.5$ for some α and all $\epsilon > 0$ (as in parts (a) and (b) of Exercise 1.10), then the integral is minimized at that α and that α is also the median.

Exercise 1.12: Let X be a rv with CDF $\mathsf{F}_X(x)$. Find the CDF of the following rv's.

a) The maximum of n IID rv's, each with CDF $\mathsf{F}_X(x)$.

Solution: Let M_+ be the maximum of the n rv's $X_1, \dots, X_n$. Note that for any real x, M_+ is less than or equal to x if and only if $X_j \le x$ for each j, $1 \le j \le n$. Thus

$$\Pr\{M_+ \le x\} = \Pr\{X_1 \le x, X_2 \le x, \dots, X_n \le x\} = \prod_{j=1}^{n} \Pr\{X_j \le x\},$$

where we have used the independence of the X_j's. Finally, since $\Pr\{X_j \le x\} = \mathsf{F}_X(x)$ for each j, we have $\mathsf{F}_{M_+}(x) = \Pr\{M_+ \le x\} = \big(\mathsf{F}_X(x)\big)^n$.

b) The minimum of n IID rv's, each with CDF $\mathsf{F}_X(x)$.

Solution: Let M_- be the minimum of $X_1, \dots, X_n$. Then, in the same way as in ((a), $M_- > y$ if and only if $X_j > y$ for $1 \le j \le n$ 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,

$$\Pr\{M_- > y\} = \Pr\{X_1 > y, X_2 > y, \dots, X_n > y\} \;=\; \prod_{j=1}^{n} \Pr\{X_j > y\}\,,$$

It follows that $1 - \mathsf{F}_{M_-}(y) = \Big(1 - F_X(y)\Big)^n$.

c) The difference of the rv's defined in a) and b); assume X has a density $\mathsf{f}_X(x)$.

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 $X_j \le x$ for $2 \le j \le n$. Also, given $X_1 = M_+ = x$, we have $R = M_+ - M_- \le r$ if and only if $X_j > x - r$ for $2 \le j \le n$. Thus, since the X_j are IID,

$$\begin{aligned}
\Pr\{M_+{=}X_1, R \le r \mid X_1 = x\} \;&=\; \prod_{j=2}^{n} \Pr\{x{-}r < X_j \le x\} \\
&=\; [\Pr\{x{-}r < X \le x\}]^{n-1} \;=\; [\mathsf{F}_X(x) - \mathsf{F}_X(x-r)]^{n-1}\,.
\end{aligned}$$

We can now remove the conditioning by averaging over $X_1 = x$. Assuming that X has the density $\mathsf{f}_X(x)$,

$$\Pr\{X_1 = M_+,\ R \le r\} = \int_{-\infty}^{\infty} \mathsf{f}_X(x)\,[\mathsf{F}_X(x) - \mathsf{F}_X(x-r)]^{n-1}\,dx.$$

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 $\Pr\{X_j = M_+,\ R \le r\} = \Pr\{X_1 = M_+\, R \le r\}$ It follows that

$$\Pr\{R \le r\} = \int_{-\infty}^{\infty} n\mathsf{f}_X(x)\,[\mathsf{F}_X(x) - \mathsf{F}_X(x-r)]^{n-1}\,dx.$$

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 have a simple form if X is exponential with the PDF $\lambda e^{-\lambda x}$ for $x \ge 0$. Then

$$\Pr\{M_- \ge y\} = e^{-n\lambda y};\quad \Pr\{M_+ \le y\} = \big(1 - e^{-\lambda y}\big)^n;\quad \Pr\{R \le y\} = \big(1 - e^{-\lambda y}\big)^{n-1}.$$

We will see how to derive the above expression for $\Pr\{R \le y\}$ in Chapter 2.

Exercise 1.13: Let X and Y be rv's in some sample space Ω 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 ω for which $Z(\omega)$ is infinite or undefined has probability 0.

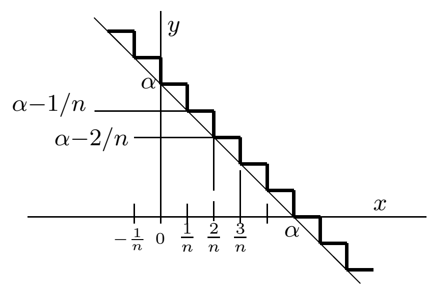
Solution: Note that Z can be infinite (either $\pm\infty$) or undefined only when either X or Y are infinite or undefined. Since these are events of zero probability, Z can be infinite or undefined only with probability 0.

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

Solution: We are trying to show that $\{Z \leq \alpha\}$ is an event for arbitrary α and doing this by first quantizing X and Y into intervals of size $1/n$ where k is used to number these quantized elements. Part (c) will make sense of how this is related to $\{Z \leq \alpha$, but for now we simply treat the sets as defined. Each set $B(n, k)$ is an intersection of two events, namely the event $\{\omega : X(\omega) \leq k/n\}$ and the event $\{\omega : Y(\omega) \leq \alpha + (1{-}k)/n\}$; these must be events since X and Y are rv's. For each n, $D(n)$ is a countable union (over k) of the sets $B(n, k)$, and thus $D(n)$ is an event for each n and each α

c) On a 2 dimensional sketch for a given α, 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 below the staircase function above. The kth step of the staircase, extended horizontally to the left and vertically down is the set $B(n, k)$. Thus we see that $D(n)$ is an upper bound to the set $\{Z \leq \alpha\}$, which is the straight line of slope -1 below the staircase.

d) Show that

$$\{\omega : X(\omega) + Y(\omega) \leq \alpha\} = \bigcap\nolimits_{n \geq 1} D(n). \tag{A.3}$$

Explain why this shows that $Z = X + Y$ is a rv.

Solution: The region $\{\omega : X(\omega) + Y(\omega) \leq \alpha\}$ is the region below the 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 ω for which $X(\omega) + Y(\omega) > \alpha$ is not contained in $D(n)$ for sufficiently large n. This verifies (A.3). 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 α. This, in conjunction with (a), shows that Z is a rv.

e) Explain why this implies that if $X_1, X_2, \dots, X_n$ are rv's, then $Y = X_1 + X_2 + \cdots + X_n$ is a rv. Hint: Only one or two lines of explanation are needed.

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

Exercise 1.15: (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 $d\mathsf{F}_X(x)$ as a unit impulse at $x = 0$.

Solution: The Riemann sum for this Stieltjes integral is $\sum_n h(x_n)[\mathsf{F}(y_n) - \mathsf{F}(y_{n-1})]$ 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 $h(x_n)[\mathsf{F}(y_n) - \mathsf{F}(y_{n-1})] = 0$ and in the second $h(x_n)[\mathsf{F}(y_n) - \mathsf{F}(y_{n-1})] = 1$. 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 $\int_{-1}^{1} h(x)d\mathsf{F}_X(x)$ 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 (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_k can be chosen to make $h(x_k)$ either 0 or 1, causing the integral to be undefined as in (a). If $a < b$, then for a sufficiently fine partion, $h(x_k) = 1$ 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, $F_X(y_n) - F_X(y_{n-1}) = 0$, 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 for infinite limits.

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 $\{\mathsf{p}(x_1), \dots, \mathsf{p}(x_K)\}$ and Y have the PMF $\{\mathsf{p}_Y(y_1), \dots, \mathsf{p}_Y(y_J)\}$. Then $\mathsf{F}_X(x) = \sum_{k=1}^{K} p(x_k)u(x - x_k)$ and $\mathsf{F}_Y(y) = \sum_{j=1}^{J} q(y_j)u(y - y_j)$. Then

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

From (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 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, $\mathsf{F}_X(x) = \Pr\{X \le x\}$ for any discrete rv X has jumps at the sample values of X and the value of $\mathsf{F}_X(x_k)$ 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 stipulation that Z be continuous on the right must be used in addition to the Stieltjes integral to define F_Z at its jumps.

Exercise 1.16: Let $X_1, X_2, \dots, X_n, \dots$ be a sequence of IID continuous rv's with the common probability density function $\mathsf{f}_X(x)$; note that $\Pr\{X{=}\alpha\} = 0$ for all α and that $\Pr\{X_i{=}X_j\} = 0$ for all $i \neq j$. For $n \ge 2$, 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 $n \ge 1$. Again use symmetry.

Solution: By the same symmetry argument, each X_i, $1 \le i \le n$ 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 and only if it is the largest of $X_1, \dots, X_n$, 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 $\mathbb{I}_n$ be 1 if X_n is a record-to-date and be 0 otherwise. Thus $\mathsf{E}[\mathbb{I}_i]$ is the expected value of the 'number' of records-to-date (either 1 or 0) on trial i. That is

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

Thus

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

This is the harmonic series, which goes to ∞ in the limit $m \to \infty$. If you are unfamiliar with this, note that $\sum_{n=1}^{\infty} 1/n \ge \int_1^{\infty} \frac{1}{x}\,dx = \infty$.

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 $\mathsf{p}_{XYZ}(x, y, z)$ 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 bnary 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

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

for a set of rv's $X_1, \dots, X_n$?

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

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 $\mathsf{g}(r)$ exists. Determine in each case whether $\mathsf{g}(r)$ exists at r_- and r_+. For parts a) and b) you should also find and sketch $\mathsf{g}(r)$. For parts c) and d), $\mathsf{g}(r)$ has no closed form.

a) Let λ, θ, be positive numbers and let X have the density.

$$\mathsf{f}_X(x) = \frac{1}{2}\lambda \exp(-\lambda x);\ x \geq 0; \quad \mathsf{f}_X(x) = \frac{1}{2}\theta \exp(\theta x);\ x < 0.$$

Solution: Integrating to find $\mathsf{g}_X(r)$ as a function of λ and θ, we get

$$\mathsf{g}_X(r) \;=\; \int_{-\infty}^{0} \frac{1}{2}\theta e^{\theta x + rx}\,dx + \int_{0}^{\infty} \frac{1}{2}\lambda e^{-\lambda x + rx}\,dx \;=\; \frac{\theta}{2(\theta + r)} + \frac{\lambda}{2(\lambda - r)}$$

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 σ^2.

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

$$\begin{aligned}
\mathsf{g}_Y(r) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-m)^2}{2\sigma^2} + ry\right) dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-m-r\sigma^2)^2}{2\sigma^2} + rm + \frac{r^2\sigma^2}{2}\right) dy \\
&= \exp\left(rm + \frac{r^2\sigma^2}{2}\right),
\end{aligned}$$

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 finite r so $r_- = -\infty$ and $r_+ = \infty$. Also $\mathsf{g}_Y(r)$ is infinite at each endpoint.

c) Let Z be a nonnegative random variable with density

$$\mathsf{f}_Z(z) = k(1+z)^{-2}\ \exp(-\lambda z); \quad z \geq 0.$$

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

Solution: Writing out the formula for $\mathsf{g}_Z(r)$, we have

$$\mathsf{g}_Z(r) = \int_{0}^{\infty} k(1+z)^{-2} \exp\big((r-\lambda)z\big)\,dz.$$

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 $z \le 1$ and goes to 0 as z^{-2} as $z \to \infty$, so the integral is finite. Thus r_+ belongs to the region where $\mathsf{g}_Z(r)$ 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 $\mathsf{g}_Z(\lambda)$ is finite depends on the coefficient of $e^{-\lambda z}$.

d) For the Z of (c), find the limit of $\gamma'(r)$ as r approaches λ from below. Then replace $(1+z)^2$ with $|1+z|^3$ in the definition of $\mathsf{f}_Z(z)$ and K and show whether the above limit is then finite or not. Hint: no integration is required.

Solution: Differentiating $\mathsf{g}_Z(r)$ with respect to r,

$$\mathsf{g}'_Z(r) = \int_0^\infty kz(1+z)^{-2}\exp\big((r-\lambda)z\big)\,dz.$$

For $r = \lambda$, the above integrand approaches 0 as $1/z$ and thus the integral does not converge. In other words, although $\mathsf{g}_Z(\lambda)$ is finite, the slope of $\mathsf{g}_Z(r)$ is unbounded as $r \to \lambda$ from below. If $(1+z)^{-2}$ is repaced 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 $\mathsf{g}_Z(r)$ remains bounded for $r < \lambda$.

Exercise 1.26: **a)** Assume that the random variable X has a moment generating function $\mathsf{g}_X(r)$ 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.*, $\mathsf{g}_{(X-c)}(r)$ in terms of the moment generating function of X. Show that $g''_{(X-c)}(r) \ge 0$.

Solution: Note that $\mathsf{g}_{(X-c)}(r) = \mathsf{E}\left[\exp(r(X-c))\right] = g_X(r)e^{-cr}$. Thus r_+ and r_- are the same for X and $X - c$. Thus (see Footnote 24), the derivatives of $\mathsf{g}_{(X-c)}(r)$ with respect to r are finite. The first two derivatives are then given by

$$g'_{(X-c)}(r) = \mathsf{E}\left[(X-c)\exp(r(X-c))\right],$$

$$g''_{(X-c)}(r) = \mathsf{E}\left[(X-c)^2\exp(r(X-c))\right] \;\ge\; 0,$$

since $(X-c)^2\exp(r(X-c)) \ge 0$ for all x.

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

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

$$\begin{aligned}
g''_{(X-c)}(r) &= \mathsf{E}\left[X^2\exp(r(X-c))\right] - 2c\mathsf{E}\left[X\exp(r(X-c))\right] + c^2\mathsf{E}\left[\exp(r(X-c))\right] \\
&= \left[\mathsf{E}\left[X^2\exp(rX)\right] - 2c\mathsf{E}\left[X\exp(rX)\right] + c^2\mathsf{E}\left[\exp(rX)\right]\right]\exp(-rc) \\
&= \left[g''_X(r) - 2cg'_X(r) + c^2\mathsf{g}(r)\right]\exp(-rc).
\end{aligned}$$

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

Solution: With the suggested choice for c,

$$\begin{aligned} g''_{(X-c)}(r) &= \left[g''_X(r) - 2\frac{(g'_X(r))^2}{\mathsf{g}_X(r)} + \frac{(g'_X(r))^2}{\mathsf{g}_X(r)}\right]\exp(-rc) \\ &= \left[\frac{\mathsf{g}_X(r)\mathsf{g}''_X(r) - [\mathsf{g}'_X(r)]^2}{\mathsf{g}_X(r)}\right]\exp(-cr). \end{aligned}$$

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

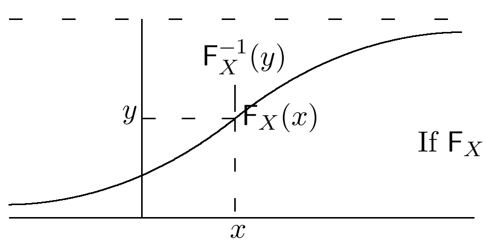
$$\gamma''_X(r) = \mathsf{g}_X(r)\mathsf{g}''_X(r) - [\mathsf{g}'_X(r)]^2 \geq 0.$$

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 (a), $\mathsf{g}''_{(X-c)}(r) > 0$. Thus the inequalities in parts b) and c) are strict also.

Exercise 1.28: Suppose the rv X is continuous and has the CDF $\mathsf{F}_X(x)$. Consider another rv $Y = \mathsf{F}_X(X)$. That is, for each sample point ω such that $X(\omega) = x$, we have $Y(\omega) = \mathsf{F}_X(x)$. Show that Y is uniformly distributed in the interval 0 to 1.

Solution: For simplicity, first assume that $\mathsf{F}_X(x)$ is strictly increasing in x, thus having the following appearance:



If $\mathsf{F}_X(x) = y$, then $\mathsf{F}_X^{-1}(y) = x$

Since $\mathsf{F}_X(x)$ is continuous in x and strictly increasing from 0 to 1, there must be an inverse function F_X^{-1} such that for each $y \in (0,1)$, $\mathsf{F}_X^{-1}(y) = x$ for that x such that $\mathsf{F}_X(x) = y$. For this y, then, the event $\{\mathsf{F}_X(X) \leq y\}$ is the same as the event $\{X \leq \mathsf{F}_X^{-1}(y)\}$. This is illustrated in the figure above. Using this equality for the given y,

$$\begin{aligned} \Pr\{Y \leq y\} &= \Pr\{\mathsf{F}_X(X) \leq y\} \;=\; \Pr\bigl\{X \leq \mathsf{F}_X^{-1}(y)\bigr\} \\ &= \mathsf{F}_X(\mathsf{F}_X^{-1}(y)) \;=\; y. \end{aligned}$$

where in the final equation, we have used the fact that F_X^{-1} is the inverse function of F_X. This relation, for all $y \in (0,1)$, shows that Y is uniformly distributed between 0 and 1.

If F_X is not strictly increasing, *i.e.*, if there is any interval over which $\mathsf{F}_X(x)$ has a constant value y, then we can define $\mathsf{F}_X^{-1}(y)$ to have any given value within that interval. The above argument then still holds, although F_X^{-1} is no longer the inverse of F_X.

If there is any discrete point, say z at which $\Pr\{X = z\} > 0$, then $\mathsf{F}_X(x)$ cannot take on values in the open interval between $\mathsf{F}_X(z) - a$ and $\mathsf{F}_X(z)$ where $a = \Pr\{X = z\}$. Thus F_X is uniformly distributed only for continuous rv's.

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 $\mathsf{F}_X(x)$ 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 $\{X_j \le x\}$ where $X_1, X_2, \dots, X_j, \dots$ are IID rv's with the CDF $\mathsf{F}_X(x)$. State the WLLN for the IID rv's $\{\mathbb{I}_1(x), \mathbb{I}_2(x), \dots\}$.

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

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(x) - \mathsf{F}_X(x)\right| > \epsilon\right\} = 0 \qquad \text{for all } x \text{ and } \epsilon > 0. \tag{A.4}$$

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 $\mathsf{F}_X(x)$ at each x, then the probability that the approximation error exceeds ϵ at any given x approaches 0 with increasing n.

b) Does the answer to (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 $X_1, X_2, \dots$. 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: Note that $(1/n)\sum_{j=1}^{n} \mathbb{I}_j(y)$ is a rv for each y. Any sample function $x_1, \dots, x_n$ of $X_1, \dots, X_n$ maps into a sample value of $(1/n)\sum_{j=1}^{n} \mathbb{I}_j(y)$ for each y. We can view this collection of sample values as a function of y. Any such sample function is non-decreasing in y, and as seen in (a) is an approximation to $F_X(y)$ at each y. This function of y has all the characteristics of a CDF itself, so we can let $\hat{\alpha}_n$ be the median of $(1/n)\sum_{j=1}^{n} \mathbb{I}_j(y)$ as a function of y. Let α be the true median of X and 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,

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

Because of the assumption of a nonzero density, there is some $\epsilon_1 > 0$ such that $\mathsf{F}_X(a - \delta) < .5 - \epsilon_1$ and some $\epsilon_2 > 0$ such that $\mathsf{F}_X(a - \delta) > .5 + \epsilon_1$. Thus,

$$\begin{aligned} \Pr\big\{\big|\hat{\alpha}_n - \alpha\big| \ge \delta\big\} \quad &\le \quad \Pr\left\{\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha - \delta) - \mathsf{F}_X(\alpha - \delta)\right| > \epsilon_1\right\} \\ &\quad + \quad \Pr\left\{\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha + \delta) - \mathsf{F}_X(\alpha + \delta)\right| > \epsilon_2\right\}. \end{aligned}$$

From (A.4), 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 α above to be any median in cases where the median is nonunique.

Exercise 1.35 a) Show that for any integers $0 < k < n$,

$$\binom{n}{k+1} \le \binom{n}{k}\frac{n-k}{k}.$$

Solution:

$$\begin{aligned}\binom{n}{k+1} &= \frac{n!}{(k+1)!(n-k-1)!} = \frac{n!}{k!(k+1)(n-k)!/(n-k)} \\ &= \binom{n}{k}\frac{n-k}{k+1} \le \binom{n}{k}\frac{n-k}{k}.\end{aligned} \tag{A.5}$$

b) Extend (a) to show that, for all $\ell \le n-k$,

$$\binom{n}{k+\ell} \le \binom{n}{k}\left[\frac{n-k}{k}\right]^{\ell}. \tag{A.6}$$

Solution: Using $k+\ell$ in place of $k+1$ in (A.5),

$$\binom{n}{k+\ell} \le \binom{n}{k+\ell-1}\left[\frac{n-k-(\ell-1)}{k+\ell-1}\right] \le \binom{n}{k+\ell-1}\left[\frac{n-k}{k}\right].$$

Applying recursion on ℓ, we get (A.6).

c) Let $\tilde{p} = k/n$ and $\tilde{q} = 1-\tilde{p}$. Let S_n be the sum of n binary IID rv's with $\mathsf{p}_X(0) = q$ and $\mathsf{p}_X(1) = p$. Show that for all $\ell \le n-k$,

$$\mathsf{p}_{S_n}(k+\ell) \le \mathsf{p}_{S_n}(k)\left(\frac{\tilde{q}p}{\tilde{p}q}\right)^{\ell}. \tag{A.7}$$

Solution: Using (b),

$$\mathsf{p}_{S_n}(k+\ell) = \binom{n}{k+\ell}p^{k+\ell}q^{n-k-\ell} \le \binom{n}{k}\left[\frac{n-k}{k}\right]^{\ell}\left[\frac{p}{q}\right]^{\ell}p^kq^{n-k} = \mathsf{p}_{S_n}(k)\left[\frac{\tilde{q}p}{\tilde{p}q}\right]^{\ell}.$$

d) For $k/n > p$, show that $\Pr\{S_n \ge k\} \le \frac{\tilde{p}q}{\tilde{p}-p}\,\mathsf{p}_{S_n}(k)$.

Solution: Using the bound in (A.7), we get

$$\begin{aligned}\Pr\{S_n \ge k\} &= \sum_{\ell=0}^{n-k}\mathsf{p}_{S_n}(k+\ell) \le \sum_{\ell=0}^{n-k}\mathsf{p}_{S_n}(k)\left(\frac{\tilde{q}p}{\tilde{p}q}\right)^{\ell} \\ &\le \mathsf{p}_{S_n}(k)\frac{1}{1-\tilde{q}p/\tilde{p}q} = \mathsf{p}_{S_n}(k)\frac{\tilde{p}q}{\tilde{p}q-\tilde{q}p} \\ &= \mathsf{p}_{S_n}(k)\frac{\tilde{p}q}{\tilde{p}(1-p)-(1-\tilde{p})p} = \mathsf{p}_{S_n}(k)\frac{\tilde{p}q}{\tilde{p}-p}.\end{aligned} \tag{A.8}$$

e) Now let ℓ be fixed and $k = \lceil n\tilde{p}\rceil$ for fixed $\tilde{p}$ such that $1 > \tilde{p} > p$. Argue that as $n \to \infty$,

$$\mathsf{p}_{S_n}(k+\ell) \sim \mathsf{p}_{S_n}(k)\left(\frac{\tilde{q}p}{\tilde{p}q}\right)^{\ell} \quad \text{and} \quad \Pr\{S_n \ge k\} \sim \frac{\tilde{p}q}{\tilde{p}-p}\,\mathsf{p}_{S_n}(k),$$

where $a(n) \sim b(n)$ means that $\lim_{n\to\infty} a(n)/b(n) = 1$.

Solution: Note that (A.6) provides an upper bound to $\binom{n}{k+\ell}$ and a slight modification of the same argument provides the lower bound $\binom{n}{k+\ell} \geq \binom{n}{k}\left(\frac{n-k-\ell}{k+\ell}\right)^{\ell}$. Taking the ratio of the upper to lower bound for fixed ℓ and $\tilde{p}$ as $n \to \infty$, we see that

$$\begin{aligned}
\binom{n}{k+\ell} \quad &\sim \quad \binom{n}{k}\left[\frac{n-k}{k}\right]^{\ell} \qquad \text{so that} \\
\mathsf{p}_{S_n}(k+\ell) \quad &\sim \quad \mathsf{p}_{S_n}(k)(\tilde{q}p/\tilde{p}q)^{\ell}
\end{aligned} \tag{A.9}$$

follows. Replacing the upper bound in (A.8) with the asymptotic equality in (A.9), and letting ℓ grow very slowly with n, we get $\Pr\{S_n \geq k\} \sim \frac{\tilde{p}q}{\tilde{p}-p}\,\mathsf{p}_{S_n}(k)$.

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:

$$\begin{aligned}
\sum_{i:\, n\delta-m\leq i\leq n\delta+m} \Pr\{S_n = i\} \quad &= \quad \Pr\{n\delta - m \leq S_n \leq n\delta + m\} \\
&= \quad \Pr\big\{-m \leq S_n - n\overline{X} \leq m\big\} \\
&= \quad \Pr\left\{\frac{-m}{\sigma\sqrt{n}} \leq \frac{S_n - n\overline{X}}{\sigma\sqrt{n}} \leq \frac{m}{\sigma\sqrt{n}}\right\},
\end{aligned}$$

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

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

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 (a). Arguing in the same way as in (a), we see that

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

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.

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

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 1.44: Let $X_1, X_2 \dots$ be a sequence of IID rv's each with mean 0 and variance σ^2. Let $S_n = X_1 + \cdots + X_n$ for all n and consider the random variable $S_n/\sigma\sqrt{n} - S_{2n}/\sigma\sqrt{2n}$. Find the limiting CDF for this sequence of rv's as $n \to \infty$. The point of this exercise is to see clearly that *the CDF* of $S_n/\sigma\sqrt{n} - S_{2n}/\sigma\sqrt{2n}$ is converging in n but that the sequence of rv's is not converging in any reasonable sense.

Solution: If we write out the above expression in terms of the X_i, we get

$$\frac{S_n}{\sigma\sqrt{n}} - \frac{S_{2n}}{\sigma\sqrt{2n}} = \sum_{i=1}^{n} X_i \left[\frac{1}{\sigma\sqrt{n}} - \frac{1}{\sigma\sqrt{2n}}\right] - \sum_{i=n+1}^{2n} \frac{X_i}{\sigma\sqrt{2n}}.$$

The first sum above approaches a Gaussian distribution of variance $(1 - 1/\sqrt{2})^2$ and the second sum approaches a Gaussian distribution of variance $1/2$. Since these two terms are independent, the difference approaches a Gaussian distribution of variance $1 + (1 - 1/\sqrt{2})^2$. This means that the distribution of the difference does converge to this Gaussian rv as $n \to \infty$. As n increases, however, this difference slowly changes, and each time n is doubled, the new difference is only weakly correlated with the old difference.

Note that $S_n/n\sigma_X$ behaves in this same way. The CDF converges as $n \to \infty$, but the rv's themselves do not approach each other in any reasonable way. The point of the problem was to emphasize this property.

Exercise 1.47: Consider a discrete rv X with the PMF

$$\begin{aligned} \mathsf{p}_X(-1) &= (1 - 10^{-10})/2, \\ \mathsf{p}_X(1) &= (1 - 10^{-10})/2, \\ \mathsf{p}_X(10^{12}) &= 10^{-10}. \end{aligned}$$

a) Find the mean and variance of X. Assuming that $\{X_m;\, m \ge 1\}$ 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: $\overline{X} = 100$ 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 $\{S_n \le 10^6\}$ in words. Find an exact expression for $\Pr\{S_n \le 10^6\} = F_{S_n}(10^6)$.

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

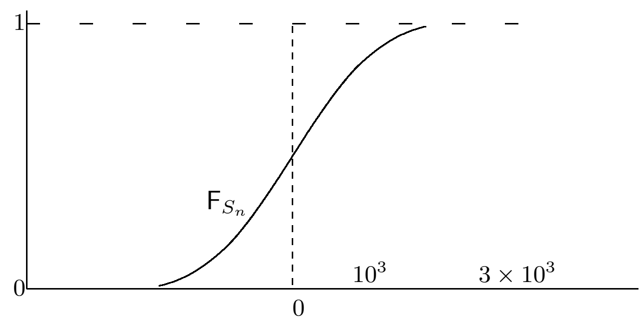
c) Find a way to use the union bound to get a simple upper bound and approximation of $1 - F_{S_n}(10^6)$.

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 of 10^6 terms, each equal to 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 -3×10^3 to 3×10^3 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, in the 4th decimal place out at 10^{12} which doesn't show up well in the sketch, but of course could be important for some purposes such as calculating σ^2.



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 2×10^{12}. Hint: First view S_n as conditioned on an appropriate rv.

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

$$\mathsf{p}_B(j) = \binom{10^{10}}{j} p^j (1-p)^{10^{10}-j} \qquad \text{where } p = 10^{-10}$$

$$\begin{aligned}
\mathsf{p}_B(0) &= (1-p)^{10^{10}} = \exp\{10^{10}\ln[1-p]\} \approx \exp(-10^{10}p) = e^{-1} \\
\mathsf{p}_B(1) &= 10^{10}p(1-p)^{10^{10}-1} = (1-p)^{10^{10}-1} \approx e^{-1} \\
\mathsf{p}_B(2) &= \binom{10^{10}}{2} p^2(1-p)^{10^{10}-2} \approx \frac{1}{2}e^{-1}.
\end{aligned}$$

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 -3×10^5 to $+3\times10^5$. 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 2×10^{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 $\lim_{n\to\infty} \mathsf{E}[|Y_n|] = 0$. 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

$$\Pr\{|Y_n| \geq \epsilon\} \leq \frac{\mathsf{E}[|Y_n|]}{\epsilon}.$$

Thus going to the limit $n \to \infty$ for the given ϵ,

$$\lim_{n\to\infty} \Pr\{|Y_n| \geq \epsilon\} = 0.$$

Since this is true for every $\epsilon > 0$, this satisfies the definition for convergence to 0 in probability.

A.2 Solutions for Chapter 2

Exercise 2.1: a) Find the Erlang density $\mathsf{f}_{S_n}(t)$ by convolving $\mathsf{f}_X(x) = \lambda \exp(-\lambda x)$, $x \geq 0$ with itself n times.

Solution: For $n = 2$, we convolve $\mathsf{f}_X(x)$ with itself.

$$\mathsf{f}_{S_2}(t) \;=\; \int_0^t \mathsf{f}_{X_1}(x)\mathsf{f}_{X_2}(t-x)\,dx \;=\; \int_0^t \lambda e^{-\lambda x}\lambda e^{-\lambda(t-x)}\,dx \;=\; \lambda^2 t e^{-\lambda t}.$$

For larger n, convolving $\mathsf{f}_X(x)$ with itself n times is found by taking the convolution $n-1$ times, *i.e.*, $\mathsf{f}_{S_{n-1}}(t)$, and convolving this with $\mathsf{f}_X(x)$. Starting with $n = 3$,

$$\begin{aligned}
\mathsf{f}_{S_3}(t) &= \int_0^t \mathsf{f}_{S_2}(x)\mathsf{f}_{X_3}(t-x)\,dx \;=\; \int_0^t \lambda^2 x e^{-\lambda x}\lambda e^{-\lambda(t-x)}\,dx \;=\; \frac{\lambda^3 t^2}{2}e^{-\lambda t} \\
\mathsf{f}_{S_4}(t) &= \int_0^t \frac{\lambda^3 x^2}{2}e^{-\lambda x}\cdot \lambda e^{\lambda(t-x)}\,dx \;=\; \frac{\lambda^4 t^3}{3!}e^{-\lambda t}.
\end{aligned}$$

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

$$\mathsf{g}_X(r) = \int_0^\infty \lambda e^{-\lambda x}e^{rx}\,dx \;=\; \frac{\lambda}{\lambda - r} \qquad \text{for } r < \lambda.$$

Since S_n is the sum of n IID rv's,

$$\mathsf{g}_{S_n}(r) \;=\; \big[\mathsf{g}_X(r)\big]^n \;=\; \left(\frac{\lambda}{\lambda - r}\right)^n.$$

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, $\mathsf{f}_{S_n}(s_n)$, 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

$$\begin{aligned}
\mathsf{f}_{S_2}(s_2) &= \lambda^2 e^{-\lambda s_2}\int_o^{s_2} ds_1 \;=\; \left(\lambda^2 e^{-\lambda s_2}\right) s_2 \\
\mathsf{f}_{S_3}(s_3) &= \lambda^3 e^{-\lambda s_3}\int_0^{s_3}\left[\int_0^{s_2} ds_1\right] ds_2 \;=\; \lambda^3 e^{-\lambda s_3}\int_0^{s_3} s_2\,ds_2 \;=\; \left(\lambda^3 e^{-\lambda s_3}\right)\frac{s_3^2}{2}.
\end{aligned}$$

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

$$\int_0^{s_n}\left[\int_0^{s_n-1}\cdots\int_0^{s_2} ds_1\dots ds_{n-2}\right] ds_{n-1} \;=\; \int_0^{s_n}\frac{s_{n-1}^{n-2}}{(n-2)!}ds_{n-1} \;=\; \frac{s_n^{n-1}}{(n-1)!}.$$

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 $(s_1, s_2, \dots, s_{n-1})$. Each component lies in $(0, s_n)$, but the cube doesn't have the ordering constraint $s_1 < s_2 < \cdots < s_{n-1}$. By symmetry, the volume of points in the cube satisfying this ordering constraint is the same as the volume in which the components $s_1, \dots s_{n-1}$ 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 constraints, 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 λ be the rate of the process.

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

Solution: The condition $S_n = \tau$ means that the epoch of the nth arrival is τ. Conditional on this, the event $\{N(t) = n\}$ for some $t > \tau$ means there have been no subsequent arrivals from τ 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

$$\Pr\{N(t) = n \mid S_n = \tau\} = \Pr\{X_{n+1} > t-\tau\} = e^{-\lambda(t-\tau)} \qquad \text{for } t > \tau. \tag{A.10}$$

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.10) over S_n.

$$\begin{aligned}
\Pr\{N(t){=}n\} &= \int_0^\infty \Pr\{N(t){=}n \mid S_n{=}\tau\}\, \mathsf{f}_{S_n}(\tau)\, d\tau \\
&= \int_0^t e^{-\lambda(t-\tau)}\frac{\lambda^n\tau^{n-1}e^{-\lambda\tau}}{(n-1)!}\, d\tau \\
&= \frac{\lambda^n e^{-\lambda t}}{(n-1)!}\int_0^t \tau^{n-1}\, d\tau \;=\; \frac{(\lambda t)^n e^{-\lambda t}}{n!}.
\end{aligned}$$

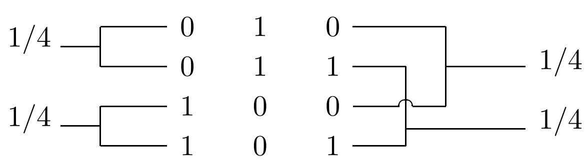
Exercise 2.5: The point of this exercise is to show that the sequence of PMF's for the counting process of a Bernoulli process does not specify the process. In other words, knowing that $N(t)$ satisfies the binomial distribution for all t does not mean that the process is Bernoulli. This helps us understand 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 $Y_1, Y_2, Y_3, \dots$, 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, $p_{N(t)}(k) = \binom{t}{k}2^{-t}$ for $t = 1, 2, 3$ and $0 \le k \le t$, 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 each have probability 1/8, and are 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.

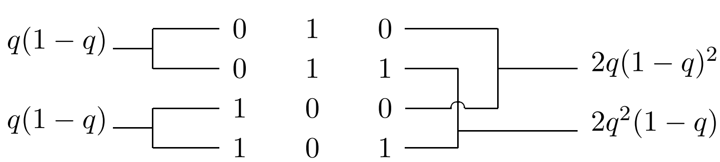
The four remaining 3-tuples are illustrated below, with the constraints imposed by $N(1)$ and $N(2)$ on the left and those imposed by $N(3)$ on the right.



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 satisfying the constraints: choose $(0, 1, 1)$ and $(1, 0, 0)$ to each have probability 1/4.

b) Generalize (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 (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.



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 $\Pr\{(1, 0, 0)\} = q(1 - q) - 2q^2(1 - q) = q(1 - q)(1 - 2q)$. This satisfies all the binomial constraints.

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 τ, 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. Note: Mathematically, a linear constraint specifies that a given weighted sum of components is 0. The type of constraint here, where the weighted sum is a nonzero constant, is more properly called a first-order constraint. Engineers often refer to first order constraints as linear, and we follow that practice here.

Since $\sum_{k=0}^{\tau} \binom{\tau}{k} p^k q^{\tau-k} = 1$, 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 (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 (c) that for each integer τ, $1 \le \tau \le t$, there are τ additional linear constraints on the PMF established by the binomial terms $N(\tau = k)$ for $0 \le k \le \tau$. 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 2.9: Consider a "shrinking Bernoulli" approximation $N_\delta(m\delta) = Y_1 + \cdots + Y_m$ to a Poisson process as described in Subsection 2.2.5.

a) Show that

$$\Pr\{N_\delta(m\delta) = n\} = \binom{m}{n} (\lambda\delta)^n (1 - \lambda\delta)^{m-n}.$$

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

$$\lim_{\delta\to 0} \Pr\{N_\delta(t) = n\} = \lim_{m\to\infty} \binom{m}{n} \left(\frac{\lambda t}{m}\right)^n \left(1 - \frac{\lambda t}{m}\right)^{m-n},$$

where the limit on the left is taken over values of δ that divide t.

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

c) Derive the following two equalities:

$$\lim_{m\to\infty} \binom{m}{n} \frac{1}{m^n} = \frac{1}{n!}; \qquad \text{and} \qquad \lim_{m\to\infty} \left(1 - \frac{\lambda t}{m}\right)^{m-n} = e^{-\lambda t}.$$

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

$$\binom{m}{n}\frac{1}{m^n} \;=\; \frac{1}{n!}\prod_{i=0}^{n-1}\frac{(m-i)}{m} \;=\; \frac{1}{n!}\prod_{i=0}^{n-1}\Big(1-\frac{i}{m}\Big). \tag{A.11}$$

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,

$$\begin{aligned}\Big(1-\frac{\lambda t}{m}\Big)^{m-n} &= \exp\Big[(m-n)\ln\Big(1-\frac{\lambda t}{m}\Big)\Big] \;=\; \exp\Big[(m-n)\Big(\frac{-\lambda t}{m}+\mathrm{o}(1/m)\Big)\Big]\\ &= \exp\Big[-\lambda t+\frac{n\lambda t}{m}+(m-n)\mathrm{o}(1/m)\Big].\end{aligned}$$

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.11) is approached, we have

$$\frac{1}{n!}\prod_{i=0}^{n-1}\Big(1-\frac{i}{m}\Big) \;=\; \frac{1}{n!}\exp\Big(\sum_{i=1}^{n-1}\ln\Big(1-\frac{i}{m}\Big)\Big) \;=\; \frac{1}{n!}\exp\Big(\frac{-n(n-1)}{2m}+\mathrm{o}(1/m)\Big).$$

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 λ.

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

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

This shows that the Poisson PMF is the limit of shrinking Bernoulli PMF's, but recall from Exercise 2.5 that this is not quite enough to show that a Poisson process is the limit of shrinking Bernoulli processes. It is also necessary to show that the stationary and independent increment properties hold in the limit $\delta \to 0$. It can be seen that the Bernoulli process has these properties at each increment δ, and it is intuitively clear that these properties should hold in the limit, but it seems that carrying out all the analytical details to show this precisely is neither warranted or interesting.

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

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 $\big(t,\, t+s\big)$. In order to find the joint distribution of $N(t)$ and $N(t+s)$, it makes sense to express $N(t+s)$ as $N(t)+\widetilde{N}(t,t+s)$ and to use the independent increment property to see that $\widetilde{N}\big(t,t+s)\big)$ is independent of $N(t)$. Thus for $m>n$,

$$\begin{aligned}\mathsf{p}_{N(t)N(t+s)}(n,m) &= \Pr\{N(t)=n\}\Pr\Big\{\widetilde{N}(t,t+s)=m-n)\Big\}\\ &= \frac{(\lambda t)^n e^{-\lambda t}}{n!}\times\frac{\big(\lambda s\big)^{m-n}e^{-\lambda s}}{(m-n)!},\end{aligned}$$

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

$$\mathsf{p}_{N(t)N(t+s)}(n, m) \;=\; \frac{\big(\lambda(t{+}s)\big)^m e^{-\lambda(s+t)}}{m!} \times \binom{m}{n}\Big(\frac{t}{t{+}s}\Big)^n\Big(\frac{s}{t{+}s}\Big)^{m-n},$$

where the first term is $\mathsf{p}_{N(t+s)}(m)$ (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 $\mathsf{E}\left[N(t) \cdot N(t+s)\right]$ for $s > 0$.

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

$$\begin{aligned}
\mathsf{E}\left[N(t) \cdot N(t{+}s)\right] &= \mathsf{E}\left[N^2(t)\right] + \mathsf{E}\left[N(t)\widetilde{N}(t, t{+}s)\right] \\
&= \mathsf{E}\left[N^2(t)\right] + \mathsf{E}\left[N(t)\right]\mathsf{E}\left[N(s)\right] \\
&= \lambda t + \lambda^2 t^2 + \lambda t \lambda s.
\end{aligned}$$

In the final step, we have used the fact (from Table 1.2 or a simple calculation) that the mean of a Poisson rv with PMF $(\lambda t)^n \exp(-\lambda t)/n!$ is λt and the variance is also λt (thus the second moment is $\lambda t + (\lambda t)^2$). This mean and variance was also derived in Exercise 2.2 and can also be calculated by looking at the limit of shrinking Bernoulli processes.

c) Find $\mathsf{E}\left[\widetilde{N}(t_1, t_3) \cdot \widetilde{N}(t_2, t_4)\right]$ where $\widetilde{N}(t, \tau)$ 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 (b). We break up $\widetilde{N}(t_1, t_3)$ as $\widetilde{N}(t_1, t_2) + \widetilde{N}(t_2, t_3)$ and break up $\widetilde{N}(t_2, t_4)$ as $\widetilde{N}(t_2, t_3) + \widetilde{N}(t_3, t_4)$. The interval $(t_2, t_3]$ is shared. Thus

$$\begin{aligned}
\mathsf{E}\left[\widetilde{N}(t_1, t_3)\widetilde{N}(t_2, t_4)\right] &= \mathsf{E}\left[\widetilde{N}(t_1, t_2)\widetilde{N}(t_2, t_4)\right] + \mathsf{E}\left[\widetilde{N}^2(t_2, t_3)\right] + \mathsf{E}\left[\widetilde{N}(t_2, t_3)\widetilde{N}(t_3, t_4)\right] \\
&= \lambda^2(t_2{-}t_1)(t_4{-}t_2) + \lambda^2(t_3{-}t_2)^2 + \lambda(t_3{-}t_2) + \lambda^2(t_3{-}t_2)(t_4{-}t_3) \\
&= \lambda^2(t_3{-}t_1)(t_4{-}t_2) + \lambda(t_3{-}t_2).
\end{aligned}$$

Exercise 2.11: An elementary experiment is independently performed N times where N is a Poisson rv of mean λ. Let $\{a_1, a_2, \dots, a_K\}$ be the set of sample points of the elementary experiment and let p_k, $1 \le k \le K$, denote the probability of a_k.

a) Let N_k denote the number of elementary experiments performed for which the output is a_k. Find the PMF for N_k $(1 \le k \le K)$. (Hint: no calculation is necessary.)

Solution: View the experiment as a combination of K Poisson processes where the kth has rate $\mathsf{p}_k\lambda$ and the combined process has rate λ. At $t = 1$, the total number of experiments is then Poisson with mean λ and the kth process is Poisson with mean $\mathsf{p}_k\lambda$. Thus $\mathsf{p}_{N_k}(n) = (\lambda\mathsf{p}_k)^n e^{-\lambda\mathsf{p}_k}/n!$.

b) Find the PMF for $N_1 + N_2$.

Solution: By the same argument,

$$\mathsf{p}_{N_1+N_2}(n) = \frac{[\lambda(\mathsf{p}_1 + \mathsf{p}_2)]^n e^{-\lambda(\mathsf{p}_1+\mathsf{p}_2)}}{n!}.$$

c) Find the conditional PMF for N_1 given that $N = n$.

Solution: Each of the n combined arrivals over $(0, 1]$ is then a_1 with probability p_1. Thus N_1 is binomial given that $N = n$,

$$\mathsf{p}_{N_1|N}(n_1|n) = \binom{n}{n_1}(\mathsf{p}_1)^{n_1}(1 - \mathsf{p}_1)^{n-n_1}.$$

d) Find the conditional PMF for $N_1 + N_2$ given that $N = n$.

Solution: Let the sample value of $N_1 + N_2$ be n_{12}. By the same argument in (c),

$$\mathsf{p}_{N_1+N_2|N}(n_{12}|n) = \binom{n}{n_{12}}(\mathsf{p}_1 + \mathsf{p}_2)^{n_{12}}(1 - \mathsf{p}_1 - \mathsf{p}_2)^{n-n_{12}}.$$

e) Find the conditional PMF for N given that $N_1 = n_1$.

Solution: Since N is then n_1 plus the number of arrivals from the other processes, and those additional arrivals are Poisson with mean $\lambda(1 - \mathsf{p}_1)$,

$$\mathsf{p}_{N|N_1}(n|n_1) = \frac{[\lambda(1 - \mathsf{p}_1)]^{n-n_1}e^{-\lambda(1-\mathsf{p}_1)}}{(n - n_1)!}.$$

Exercise 2.12: Starting from time 0, northbound buses arrive at 77 Mass. Avenue according to a Poisson process of rate λ. Customers arrive according to an independent Poisson process of rate μ. 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 mth 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 choices between bus or customer arrivals 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 $\big[\mu/(\lambda + \mu)\big]^n\lambda/(\lambda + \mu)$. This is the probability that n customers enter the mth bus, *i.e.*, defining N_m as the number of customers entering the mth bus, the PMF of N_m is

$$\mathsf{p}_{N_m}(n) = \left(\frac{\mu}{\lambda + \mu}\right)^n \frac{\lambda}{\lambda + \mu}. \tag{A.12}$$

b) Find the PMF for the number of customers entering the mth 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 μ. 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 that 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,

$$\mathsf{p}_{N_m|X_m}(n|x) = (\mu x)^n e^{-\mu x}/n!.$$

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. The number of customers entering bus m is still determined by the argument in (a) and has the PMF in (A.12). In other words, N_m is independent of the arrival time of bus $m-1$. From the formula in (A.12), 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.12).

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 (b), the number of customer arrivals from 10:30 to 11 is a Poisson rv, say N' with PMF $\mathsf{p}_{N'}(n) = (\mu/2)^n e^{-\mu/2}/n!$ (we are measuring time in hours so that μ is the customer arrival rate in arrivals 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 N'' arriving between 11 and the next bus arrival. By the argument in (a), N'' has the PMF in (A.12). 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.*,

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

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 ith 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 $1 \le m < n$, then n customers are waiting. Since the Z_i are IID, the PMF of the number N_{past} waiting at 2:30 is

$$\mathsf{p}_{N_{\text{past}}}(n) = \left(\frac{\mu}{\lambda+\mu}\right)^n \frac{\lambda}{\lambda+\mu}.$$

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 (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 $t \to \infty$. 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 neglibible 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 (a); look carefully at (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_{f} (found in (c)) until the next bus after 2:30. Note that N_{p} and N_{f} are IID. Convolving these PMF's, we get

$$\begin{aligned} \mathsf{p}_{N_{\mathrm{p}}+N_{\mathrm{f}}}(n) &= \sum_{m=0}^{n} \left(\frac{\mu}{\lambda+\mu}\right)^m \frac{\lambda}{\lambda+\mu} \left(\frac{\mu}{\lambda+\mu}\right)^{n-m} \frac{\lambda}{\lambda+\mu} \\ &= (n+1) \left(\frac{\mu}{\lambda+\mu}\right)^n \left(\frac{\lambda}{\lambda+\mu}\right)^2 . \end{aligned}$$

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 addition to the Poisson process of customers, and thus the number entering the next bus is $1 + N_{\mathrm{p}} + N_{\mathrm{f}}$. This has the sample value n if $N_{\mathrm{p}} + N_{\mathrm{f}}$ has the sample value $n-1$, so from (f),

$$\mathsf{p}_{1+N_{\mathrm{p}}+N_{\mathrm{f}}}(n) = n \left(\frac{\mu}{\lambda+\mu}\right)^{n-1} \left(\frac{\lambda}{\lambda+\mu}\right)^2 .$$

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 2.14: Equation (2.42) gives $\mathsf{f}_{S_i|N(t)}(s_i \mid n)$, which is the density of random variable S_i conditional on $N(t) = n$ for $n \geq i$. Multiply this expression by $\Pr\{N(t) = n\}$ and sum over n to find $\mathsf{f}_{S_i}(s_i)$; verify that your answer is indeed the Erlang density.

Solution: It is almost magical, but of course it has to work out.

$$\mathsf{f}_{S_i|N(t)}(s_i|n) = \frac{(s_i)^{i-1}}{(i-1)!}\frac{(t-s_i)^{n-i}}{(n-i)!}\frac{n!}{t^n}; \qquad \mathsf{p}_{N(t)}(n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

$$\begin{aligned}
\sum_{n=i}^{\infty} \mathsf{f}(s_i|n)\mathsf{p}(n) &= \frac{s_i^{i-1}}{(i-i)!}\sum_{n=i}^{\infty}\frac{(t-s_i)^{n-i}}{(n-i)!}\lambda^n e^{-\lambda t} \\
&= \frac{\lambda^i s_i^{i-1} e^{-\lambda s_i}}{(i-i)!}\sum_{n=i}^{\infty}\frac{\lambda^{n-i}(t-s_i)^{n-i}e^{-\lambda(t-s_i)}}{(n-i)!} \\
&= \frac{\lambda^i s_i^{i-1} e^{-\lambda s_i}}{(i-i)!}.
\end{aligned}$$

This is the Erlang distribution, and it follows because the preceding sum is the sum of terms in the PMF for the Poisson rv of rate $\lambda(t - s_i)$

Exercise 2.17: a) For a Poisson process of rate λ, find $\Pr\{N(t){=}n \mid S_1{=}\tau\}$ for $t > \tau$ and $n \geq 1$.

Solution: Given that $S_1 = \tau$, the number, $N(t)$, of arrivals in $(0, t]$ is 1 plus the number in $(\tau, t]$. This latter number, $\widetilde{N}(\tau, t)$ is Poisson with mean $\lambda(t - \tau)$. Thus,

$$\Pr\{N(t){=}n \mid S_1{=}\tau\} \;=\; \Pr\left\{\widetilde{N}(\tau,t) = n{-}1\right\} \;=\; \frac{[\lambda(t-\tau)]^{n-1}e^{-\lambda(t-\tau)}}{(n-1)!}.$$

b) Using this, find $\mathsf{f}_{S_1}(\tau) \mid N(t){=}n)$.

Solution: Using Bayes' law,

$$\mathsf{f}_{S_1|N(t)}(\tau|n) = \frac{n(t-\tau)^{n-1}}{t^n}.$$

c) Check your answer against (2.41).

Solution: Eq. (2.41) is $\Pr\{S_1 > \tau \mid N(t) = n\} = [(t-\tau)/t]^n$. The derivative of this with respect to τ is $-\mathsf{f}_{S_1|N(t)}(\tau|t)$, which clearly checks with (b).

Exercise 2.20: Suppose cars enter a one-way infinite length, infinite lane highway at a Poisson rate λ. The ith car to enter chooses a velocity V_i and travels at this velocity. Assume that the V_i's are independent positive rv's having a common CDF F. Derive the distribution of the number of cars that are located in an interval $(0, a)$ at time t.

Solution: This is a thinly disguised variation of an M/G/∞ queue. The arrival process is the Poisson process of cars entering the highway. We then view the service time of a car as the time interval until the car reaches or passes point a. All cars then have IID service times, and service always starts at the time of arrival (*i.e.*, this can be viewed as infinitely many independent and identical servers). To avoid distractions, assume initially that V is a continuous rv. The CDF $G(\tau)$ of the service time X is then given by the equation

$$G(\tau) = \Pr\{X \le \tau\} = \Pr\{a/V \le \tau\} = \Pr\{V \ge a/\tau\} = \mathsf{F}^{\mathsf{c}}_V(a/\tau).$$

The PMF of the number $N_1(t)$ of cars in service at time t is then given by (2.36) and (2.37) as

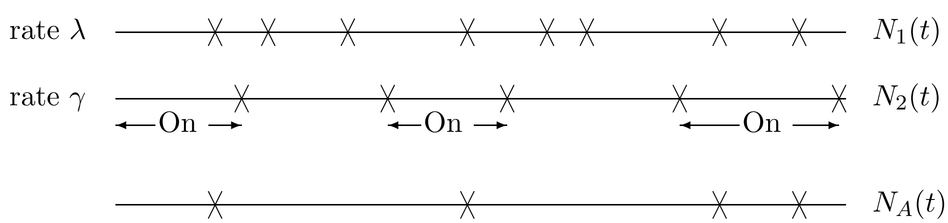
$$\mathsf{p}_{N_1(t)}(n) = \frac{m^n(t)\exp[-m(t)]}{n!},$$

where

$$m(t) = \lambda \int_0^t [1 - G(\tau)]\, d\tau = \lambda \int_0^t \mathsf{F}_V(a/\tau)\, d\tau.$$

Since this depends only on the CDF, it can be seen that the answer is the same if V is discrete or mixed.

Exercise 2.23: Let $\{N_1(t);\, t > 0\}$ be a Poisson counting process of rate λ. 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 γ.



Let $\{N_A(t);\, t{>}0\}$ 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 nth 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 $\mathsf{p}_{N_s}(n) = \big[\lambda/(\lambda{+}\gamma)\big]^n \big[\gamma/(\lambda{+}\gamma)\big]$ for integer $n \ge 0$.

b) Given that the first arrival for the second process occurs at epoch τ, find the conditional PMF for the number of arrivals N_a of the first process up to τ.

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 τ. 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 (a), we showed that $\mathsf{p}_{N_a}(n) = \left[\lambda/(\lambda{+}\gamma)\right]^n\left[\gamma/(\lambda{+}\gamma)\right]$ and in (b) we showed that $\mathsf{p}_{N_a|X_2}(n|\tau) = (\lambda\tau)^n e^{-\lambda\tau}/n!$. We can then use Bayes' law to find $\mathsf{f}_{X_2|N_a}(\tau \mid n)$, which is the desired solution. We have

$$\mathsf{f}_{X_2|N_a}(\tau \mid n) \;=\; \mathsf{f}_{X_2}(\tau)\frac{\mathsf{p}_{N_a|X_2}(n|\tau)}{\mathsf{p}_{N_a}(n)} \;=\; \frac{(\lambda{+}\gamma)^{n+1}\tau^n e^{-(\lambda+\gamma)\tau}}{n!},$$

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

$$\begin{aligned}
\mathsf{f}_{X_A}(x) &= \frac{\lambda}{\lambda{+}\gamma}\mathsf{f}_X(x) + \frac{\gamma}{\lambda{+}\gamma}\mathsf{f}_X(x) \otimes \mathsf{f}_{X_2}(x) \otimes \mathsf{f}_{X_A}(x) \\
&= \lambda\exp(-(\lambda{+}\gamma)x) + \gamma\exp(-(\lambda{+}\gamma)x) \otimes \gamma\exp(-\gamma x) \otimes \mathsf{f}_{X_A}(x).
\end{aligned}$$

where $\otimes$ 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

$$\mathsf{f}_{X_A}(x) = B\,\exp\left[-\frac{x}{2}\left(2\gamma{+}\lambda + \sqrt{4\gamma^2+\lambda^2}\right)\right] + C\,\exp\left[-\frac{x}{2}\left(2\gamma{+}\lambda - \sqrt{4\gamma^2+\lambda^2}\right)\right],$$

where

$$B = \frac{\lambda}{2}\left(1 + \frac{\lambda}{\sqrt{4\gamma^2+\lambda^2}}\right); \qquad C = \frac{\lambda}{2}\left(1 - \frac{\lambda}{\sqrt{4\gamma^2+\lambda^2}}\right).$$

Exercise 2.25: **a)** For $1 \leq i < n$, find the conditional density of S_{i+1}, conditional on $N(t) = n$ and $S_i = s_i$.

Solution: Recall from (2,41) that

$$\Pr\{S_1 > \tau \mid N(t) = n\} \;=\; \Pr\{X_1 > \tau \mid N(t) = n\} \;=\; \left(\frac{t-\tau}{t}\right)^n .$$

Given that $S_i = s_i$, we can apply this same formula to $\widetilde{N}(s_i, t)$ for the first arrival after s_i.

$$\Pr\{X_{i+1}{>}\tau \mid N(t){=}n, S_i{=}s_i\} \;=\; \Pr\Big\{X_{i+1}{>}\tau \mid \widetilde{N}(s_i, t){=}n{-}i, S_i{=}s_i\Big\} \;=\; \left(\frac{t{-}s_i{-}\tau}{t{-}s_i}\right)^{n-i} .$$

Since $S_{i+1} = S_i + X_{i+1}$, we get

$$\begin{aligned}
\Pr\{S_{i+1}{>}s_{i+1} \mid N(t){=}n, S_i{=}s_i\} &= \left(\frac{t{-}s_{i+1}}{t{-}s_i}\right)^{n-i} \\
\mathsf{f}_{S_{i+1}|N(t)S_i}\big(s_{i+1} \mid n, s_i\big) &= \frac{(n-i)(t-s_{i+1})^{n-i-1}}{(t-s_i)^{n-i}}. \qquad \text{(A.13)}
\end{aligned}$$

b) Use (a) to find the joint density of $S_1, \dots, S_n$ conditional on $N(t) = n$. Verify that your answer agrees with (2.38).

Solution: For each i, the conditional probability in (a) is clearly independent of $S_{i-2}, \dots, S_1$. Thus we can use the chain rule to multiply (A35) by itself for each value of i. We must also include $\mathsf{f}_{S_1|N(t)}\big(s_1 \mid n\big) = n(t-s_1)^{n-1}/t^n$. Thus

$$\begin{aligned}
\mathsf{f}_{\boldsymbol{S}^{(n)}|N(t)}\big(\boldsymbol{s}^{(n)}|n\big) &= \frac{n(t{-}s_1)^{n-1}}{t^n} \cdot \frac{(n{-}1)(t{-}s_2)^{n-2}}{(t-s_1)^{n-1}} \cdot \frac{(n{-}2)(t{-}s_3)^{n-3}}{(t-s_2)^{n-2}} \cdots \frac{(t{-}s_n)^0}{t{-}s_{n-1}} \\
&= \frac{n!}{t^n}.
\end{aligned}$$

Note: There is no great insight to be claimed from this exercise. it is useful, however, in providing some additional techniques for working with such problems.

Exercise 2.28: The purpose of this problem is to illustrate that for an arrival process with independent but not identically distributed interarrival intervals, $X_1, X_2, \dots$, 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 ith 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

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

Solution: Each X_i is an exponential rv, but the rate, α^{-i}, is rapidly increasing with i and the expected interarrival time, $\mathsf{E}[X_i] = \alpha^i$, is rapidly decreasing with i. Thus

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

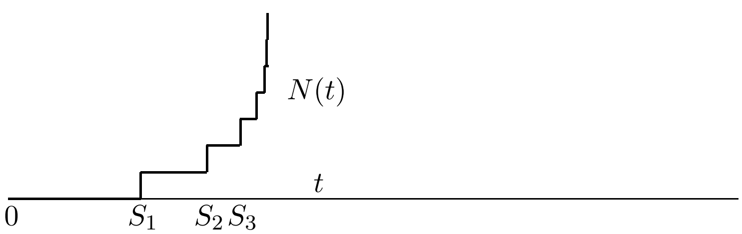
Recalling that $1 + \alpha + \alpha^2 + \cdots + \alpha^{n-1} = (1 - \alpha^n)/(1 - \alpha)$,

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

In other words, not only is $\mathsf{E}[X_i]$ decaying to 0 geometrically with increasing i, but $\mathsf{E}[S_n]$ 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:



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 $\sigma^2_{S_n}$.

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

$$\begin{aligned}
\sigma^2_{S_n} &= \sigma^2_{X_1} + \sigma^2_{X_2} \cdots + \sigma^2_{X_n} \\
&= \alpha^2 + \alpha^4 + \cdots + \alpha^{2n} \\
&= \alpha^2(1 + \alpha^2 + \cdots \alpha^{2(n-1)}) \\
&= \frac{\alpha^2(1 - \alpha^{2n})}{1 - \alpha^2} < \frac{\alpha^2}{1 - \alpha^2}.
\end{aligned}$$

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,

$$\begin{aligned}
\Pr\{S_n \geq t\} &\leq \frac{\overline{S}_n}{t} \leq \frac{\alpha}{t(1-\alpha)} \\
\Pr\{N(t) < n\} &= \Pr\{S_n > t\} \leq \Pr\{S_n \geq t\} \leq \frac{\alpha}{t(1-\alpha)}.
\end{aligned}$$

where we have used (2.3). Since this bound is independent of n, it also applies in the limit, *i.e.*,

$$\lim_{n\to\infty} \Pr\{N(t) \leq n\} \leq \frac{\alpha}{t(1-\alpha)}.$$

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 $\lim_{n\to\infty} \mathsf{E}[S_n]$.

Actually, by working 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 $t > 0$.

A.3 Solutions for Chapter 3

Exercise 3.1: **a)** Let X, Y be IID rv's, each with density $\mathsf{f}_X(x) = \alpha \exp(-x^2/2)$. In (b), we show that α must be $1/\sqrt{2\pi}$ in order for $\mathsf{f}_X(x)$ to integrate to 1, but in this part, we leave α undetermined. Let $S = X^2 + Y^2$. Find the probability density of S in terms of α.

Solution: First we find the CDF of S.

$$\begin{aligned} \mathsf{F}_S(s) &= \int\int_{x^2+y^2\le s} \alpha^2 e^{(-x^2-y^2)/2}\,dxdy \\ &= \int_0^{2\pi}\int_{r^2<s} \alpha^2 r e^{-r^2/2}\,dr d\theta \\ &= \int_{r^2\le s} 2\pi\alpha^2 e^{-r^2/2}\,d(r^2/2) \;=\; 2\pi\alpha^2(1-e^{-s/2}), \end{aligned} \tag{A.14}$$

where we first changed to polar coordinates and then integrated. The density is then

$$\mathsf{f}_S(s) = \pi\alpha^2 e^{-s/2}; \qquad \text{for } s \ge 0.$$

b) Prove from (a) that α must be $1/\sqrt{2\pi}$ in order for S, and thus X and Y, to be random variables. Show that $\mathsf{E}[X] = 0$ and that $\mathsf{E}\left[X^2\right] = 1$.

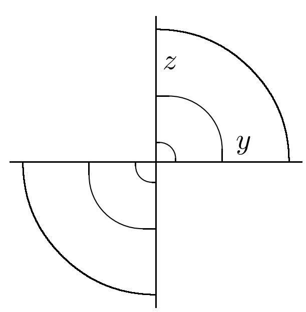
Solution: From (A.14) and the fact that $\lim_{s\to\infty} \mathsf{F}_S(s) = 1$, we see that $2\pi\alpha^2 = 1$, so $\alpha = 1/\sqrt{2\pi}$. From the fact that $\mathsf{f}_X(x) = \mathsf{f}_X(-x)$ for all x, we see that $\mathsf{E}[X] = 0$. Also, since S is exponential and is seen to have mean 2, and since X and Y must have the same second moment, we see that $\mathsf{E}\left[X^2\right] = 1$. This also follows by using integration by parts.

c) Find the probability density of $R = \sqrt{S}$. R is called a *Rayleigh* rv.

Solution: Since $S \ge 0$ and $\mathsf{F}_S(s) = 1 - e^{-s/2}$, we see that $R \ge 0$ and $\mathsf{F}_R(r) = 1 - e^{-r^2/2}$. Thus the density is given by $\mathsf{f}_R(r) = re^{-r^2/2}$.

Exercise 3.3: Let X and Z be IID normalized Gaussian random variables. Let $Y = |Z|\,\mathrm{Sgn}(X)$, where $\mathrm{Sgn}(X)$ is 1 if $X \ge 0$ and -1 otherwise. Show that X and Y are each Gaussian, but are not jointly Gaussian. Sketch the contours of equal joint probability density.

Solution: Note that Y has the magnitude of Z but the sign of X, so that X and Y are either both positive or both negative, *i.e.*, their joint density is nonzero only in the first and third quadrant of the X, Y plane. Conditional on a given X, the conditional density of Y is twice the conditional density of Z since both Z and $-Z$ are mapped into the same Y. Thus $\mathsf{f}_{XY}(x,y) = (1/\pi)\exp^{(-x^2-y^2)/2}$ for all x, y in the first or third quadrant.



Exercise 3.4: **a)** Let $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ and let $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ be independent of X_1. Convolve the density of X_1 with that of X_2 to show that $X_1 + X_2$ is Gaussian, $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

Solution: Let $Z = X_1 + X_2$. Since X_1 and X_2 are independent, the density of Z is the convolution of the X_1 and X_2 densities. For initial simplicity, assume $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$.

$$
\begin{aligned}
f_Z(z) &= f_{X_1}(z) * f_{X_2}(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)\,dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(z-x)^2/2}\,dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x^2 - xz + \frac{z^2}{2})}\,dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x^2 - xz + \frac{z^2}{4}) - \frac{z^2}{4}}\,dx \\
&= \frac{1}{2\sqrt{\pi}} e^{-z^2/4} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x - \frac{z}{2})^2}\,dx \\
&= \frac{1}{2\sqrt{\pi}} e^{-z^2/4},
\end{aligned}
$$

since the last integral integrates a Gaussian pdf with mean $z/2$ and variance $1/2$, which evaluates to 1. As expected, Z is Gaussian with zero mean and variance 2.

The 'trick' used here in the fourth equation above is called *completing the square.* The idea is to take a quadratic expression such as $x^2 + \alpha z + \beta z^2$ and to add and subtract $\alpha^2 z^2/4$. Then $x^2 + \alpha xz + \alpha z^2/4$ is $(x + \alpha z/2)^2$, which leads to a Gaussian form that can be integrated.

Repeating the same steps for arbitrary $\sigma_{X_1}^2$ and $\sigma_{X_2}^2$, we get the Gaussian density with mean 0 and variance $\sigma_{X_1}^2 + \sigma_{X_2}^2$.

b) Let W_1, W_2 be IID normalized Gaussian rv's . Show that $a_1W_1 + a_2W_2$ is Gaussian, $\mathcal{N}(0, a_1^2 + a_2^2)$. Hint: You could repeat all the equations of (a), but the insightful approach is to let $X_i = a_iW_i$ for $i = 1, 2$ and then use (a) directly.

Solution: Following the hint $\sigma_{X_i}^2 = \alpha_i^2$ for $i = 1, 2$, so $\alpha_1 W_1 + \alpha_2 W_2$ is $\mathcal{N}(0, \alpha_1^2 + \alpha_2^2)$.

c) Combine (b) with induction to show that all linear combinations of IID normalized Gaussian rv's are Gaussian.

Solution: The inductive hypothesis is that if $\{W_i;\, i \geq 1\}$ is a sequence of IID normal rv's, if $\{\alpha_i;\, i \geq 1\}$ is a sequence of numbers, and if $\sum_{i=1}^n \alpha_i W_i$ is $\mathcal{N}(0, \sum_{i=1}^n \alpha_i^2)$ for a given $n \geq 1$, then $\sum_{i=1}^{n+1} \alpha_i W_i$ is $\mathcal{N}(0, \sum_{i=1}^{n+1} \alpha_i^2)$. The basis for the induction was established in (b). For the inductive step, let $X = \sum_{i=1}^n \alpha_i W_i$. Now X is independent of W_{n+1} and by the inductive hypothesis $X \sim \mathcal{N}(0, \sum_{i=1}^n \alpha_i^2)$. From (a), $X + W_{n+1} \sim \mathcal{N}(0, \sum_{i=1}^{n+1} \alpha_i^2)$. This establishes the inductive step, so $\sum_{i=1}^n \alpha_i W_i \sim \mathcal{N}(0, \sum_{i=1}^n \alpha_i^2)$ for all n, *i.e.*, all linear combinations of IID normalized Gaussian rv's are Gaussian.

Exercise 3.7: Let $[Q]$ be an orthonormal matrix. Show that the squared distance between any two vectors $\boldsymbol{z}$ and $\boldsymbol{y}$ is equal to the squared distance between $[Q]\boldsymbol{z}$ and $[Q]\boldsymbol{y}$.

Solution: The squared distance, say d^2 between $\boldsymbol{z}$ and $\boldsymbol{y}$ is $d^2 = (\boldsymbol{z} - \boldsymbol{y})^\mathsf{T}(\boldsymbol{z} - \boldsymbol{y})$. Letting

$\boldsymbol{x} = \boldsymbol{z} - \boldsymbol{y}$, we have $d^2 = \boldsymbol{x}^\mathsf{T}\boldsymbol{x}$. The squared distance, say d_1^2, between $[Q]\boldsymbol{z}$ and $[Q]\boldsymbol{y}$ is

$$d_1^2 = \Big([Q]\boldsymbol{z} - [Q]\boldsymbol{y}\Big)^\mathsf{T}\Big([Q]\boldsymbol{z} - [Q]\boldsymbol{y}\Big) = \Big([Q]\boldsymbol{x}\Big)^\mathsf{T}\Big([Q]\boldsymbol{x}\Big) = \boldsymbol{x}^\mathsf{T}[Q]^\mathsf{T}[Q]\boldsymbol{x}.$$

From (3.26), $[Q^\mathsf{T}] = [Q^{-1}]$, so $[Q]^\mathsf{T}[Q] = [I]$ and $d_1^2 = d^2$.

Exercise 3.10: **a)** Let X and Y be zero-mean jointly Gaussian with variances σ_X^2, σ_Y^2, and normalized covariance ρ. Let $V = Y^3$. Find the conditional density $\mathsf{f}_{X|V}(x \mid v)$. Hint: This requires no computation.

Solution: Note that $v = y^3$ for $y \in \mathbb{R}$ is a one-to-one mapping. It follows that if $V = v$, then $Y = v^{1/3}$. Thus $\mathsf{f}_{X|V}(x|v)$ can be found directly from $\mathsf{f}_{X|Y}(x|y)$ as given in (3.37) by substituting $v^{1/3}$ for y, *i.e.*,

$$\mathsf{f}_{X|V}(x|v) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho^2)}}\exp\left[\frac{-\left[x - \rho(\sigma_X/\sigma_Y)v^{1/3}\right]^2}{2\sigma_X^2(1-\rho^2)}\right].$$

b) Let $U = Y^2$ and find the conditional density of $\mathsf{f}_{X|U}(x \mid u)$. Hint: first understand why this is harder than (a).

Solution: Note that $u = y^2$ for $y \in \mathbb{R}$ is not one-to-one. If $U = u$, then Y is either $u^{1/2}$ or $-u^{1/2}$. We then have

$$\mathsf{f}_{X|U}(x|u) \;=\; \frac{\mathsf{f}_{XU}(x,u)}{\mathsf{f}_U(u)} \;=\; \frac{\mathsf{f}_{XY}(x,\sqrt{u}) + \mathsf{f}_{XY}(x,-\sqrt{u})}{\mathsf{f}_Y(\sqrt{u}) + \mathsf{f}_Y(-\sqrt{u})}.$$

Since $\mathsf{f}_Y(\sqrt{u}) = \mathsf{f}_Y(-\sqrt{u})$, this can be rewritten as

$$\mathsf{f}_{X|U}(x|u) \;=\; \frac{\mathsf{f}_{X|Y}(x,\sqrt{y}) + \mathsf{f}_{X|Y}(x,-\sqrt{y})}{2}.$$

Substituting these terms into (3.37) gives a rather ugly answer. The point here is not the answer but rather the approach to finding a conditional probability for a Gaussian problem when the conditioning rv is a non-one-to one function of a Gaussian rv.

Exercise 3.11: **a)** Let $(\boldsymbol{X}^\mathsf{T}, \boldsymbol{Y}^\mathsf{T})$ have a non-singular covariance matrix $[K]$. Show that $[K_{\boldsymbol{X}}]$ and $[K_{\boldsymbol{Y}}]$ are positive definite, and thus non-singular.

Solution: Let $\boldsymbol{X}$ have dimension n and $\boldsymbol{Y}$ have dimension m. If $[K_{\boldsymbol{X}}]$ is not positive definite but only semi-definite, then there is a $\boldsymbol{b} \neq 0$ such that $\boldsymbol{b}^\mathsf{T}[k_{\boldsymbol{X}}]\boldsymbol{b} = 0$. Defining $\hat{\boldsymbol{b}}$ as the $m+n$ dimensional vector with $\boldsymbol{b}$ as the first n components and zeros as the last m components, we see that $\hat{\boldsymbol{b}}^\mathsf{T}[K]\hat{\boldsymbol{b}} = 0$, so that $[K]$ is not positive definite. The contrapositive of this is that if $[K]$ is positive definite, then $[K_{\boldsymbol{X}}]$ is also positive definite.

The same argument shows that if $[K]$ is positive definite, then $[K_{\boldsymbol{Y}}]$ is also positive definite. In fact, this argument shows that all submatrices of $[K]$ that are symmetric around the main diagonal are positive definite.

b) Show that the matrices $[B]$ and $[D]$ in (3.39) are also positive definite and thus non-singular.

Solution: We can represent $[K]$ in the form $[Q\Lambda Q^{-1}]$ where $[Q]$ is a matrix whose columns are orthonormal eigenvectors of $[K]$ and $[\Lambda]$ is the diagonal matrix of eigenvalues. Assuming that $[K]$ is positive definite, the eigenvalues are all strictly positive. It is easily seen that the inverse of $[K]$ can be represented as $[K^{-1}] = [Q\Lambda^{-1}Q^{-1}]$ where $[\Lambda^{-1}]$ is the diagonal matrix of the reciprocals of the eigenvalues. Since the eigenvalues are positive, their reciprocals are also, so $[K^{-1}]$ is also positive definite. From (a), then, the matrices $[B]$ and $[D]$ are also positive definite.

Exercise 3.13: **a)** Let $\boldsymbol{W}$ be a normalized IID Gaussian n-r**v** and let $\boldsymbol{Y}$ be a Gaussian m-r**v**. Suppose we would like to choose $\boldsymbol{Y}$ so that the joint covariance $\mathsf{E}\left[\boldsymbol{W}\boldsymbol{Y}^{\mathsf{T}}\right]$ is some arbitrary real-valued $n \times m$ matrix $[K]$. Find the matrix $[A]$ such that $\boldsymbol{Y} = [A]\boldsymbol{W}$ achieves the desired joint covariance. Note: this shows that any real-valued $n \times m$ matrix is the joint covariance matrix for some choice of random vectors.

Solution: $\boldsymbol{Y} = [A]\boldsymbol{W}$ means that $Y_i = \sum_{j=1}^{n} a_{ij}W_j$ for each i. Since the W_j are normalized and IID, $\mathsf{E}[Y_iW_j] = a_{ij}$. This is turn can be rewritten as $\mathsf{E}[YW^{\mathsf{T}}] = [A]$. Thus we choose $[A] = [K^{\mathsf{T}}]$. Note that the desired matrix $[K]$ also determines the covariance $[K_{\boldsymbol{Y}}]$, *i.e.*,

$$[K_{\boldsymbol{Y}}] = \mathsf{E}[\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}}] = \mathsf{E}[A\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}] = [AA^{\mathsf{T}}].$$

In other words, we can choose $[K]$ arbitrarily, but this also determines $[K_{\boldsymbol{Y}}]$

b) Let $\boldsymbol{Z}$ be a zero-mean Gaussian n-r**v** with non-singular covariance $[K_{\boldsymbol{Z}}]$, and let $\boldsymbol{Y}$ be a Gaussian m-r**v**. Suppose we would like the joint covariance $\mathsf{E}\left[\boldsymbol{Z}\boldsymbol{Y}^{\mathsf{T}}\right]$ to be some arbitrary $n \times m$ matrix $[K']$. Find the matrix $[B]$ such that $\boldsymbol{Y} = [B]\boldsymbol{Z}$ achieves the desired joint covariance. Note: this shows that any real valued $n \times m$ matrix is the joint covariance matrix for some choice of random vectors $\boldsymbol{Z}$ and $\boldsymbol{Y}$ where $[K_{\boldsymbol{Z}}]$ is given (and non-singular).

Solution: For any given m by n matrix $[B]$, if we choose $\boldsymbol{Y} = [B]\boldsymbol{Z}$, then

$$\mathsf{E}[\boldsymbol{Z}\boldsymbol{Y}^{\mathsf{T}}] \;=\; \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}[B^{\mathsf{T}}]] \;=\; [K_{\boldsymbol{Z}}][B^{\mathsf{T}}].$$

Thus if we want to set this equal to some given matrix $[K']$, it is sufficient to choose $[B] = [K']^{\mathsf{T}}[K_{\boldsymbol{Z}}^{-1}]$.

c) Now assume that $\boldsymbol{Z}$ has a singular covariance matrix in (b). Explain the constraints this places on possible choices for the joint covariance $\mathsf{E}\left[\boldsymbol{Z}\boldsymbol{Y}^{\mathsf{T}}\right]$. Hint: your solution should involve the eigenvectors of $[K_{\boldsymbol{Z}}]$.

Solution: If $[K_{\boldsymbol{Z}}]$ is singular, then there are one or more eigenvectors of $[K_{\boldsymbol{Z}}]$ of eigenvalue 0, *i.e.*, vectors $\boldsymbol{b}$ such that $\mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}]\,\boldsymbol{b} = 0$. For such $\boldsymbol{b}$, we must have $\mathsf{E}[(\boldsymbol{b}^{\mathsf{T}}\boldsymbol{Z})(\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{b})] = 0$, *i.e.*, the first and second moments of $\boldsymbol{b}^{\mathsf{T}}\boldsymbol{Z}$ must be 0. For each such $\boldsymbol{b}$, the transpose $\boldsymbol{b}^{\mathsf{T}}$ must satisfy $\boldsymbol{b}^{\mathsf{T}}\mathsf{E}[\boldsymbol{Z}\boldsymbol{Y}^{\mathsf{T}}] = 0$. This means that each linearly independent eigenvector $\boldsymbol{b}$ of eigenvalue 0 provides a linear constraint $\boldsymbol{b}^{\mathsf{T}}[K'] = 0$ on $[K']$.

Exercise 3.15: **a)** Solve directly for $[B]$, $[C]$, and $[D]$ in (3.39) for the one dimensional case where $n = m = 1$. Show that (3.40) agrees with (3.37)

Solution: For X and Y one-dimensional, the covariance matrix of $(X, Y)^{\mathsf{T}}$ is

$$[K] \;=\; \begin{bmatrix} [K_{\boldsymbol{X}}] & [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}] \\ [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}^{\mathsf{T}}] & [K_{\boldsymbol{Y}}] \end{bmatrix} \;=\; \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

With the assumption in (3.39) that $[K]$ is positive definite, $[K^{-1}]$ must exist and we can solve directly for B, C in

$$[K^{-1}] \;=\; \begin{bmatrix} B & C \\ C & D \end{bmatrix}$$

by multiplying out the left two terms in $[K][K^{-1}] = [I]$. We get $\sigma_X^2 B + \rho\sigma_X\sigma_Y C = 1$ and $\rho\sigma_X\sigma_Y B + \sigma_X^2 C = 0$. From the second equation, $C = -\rho\sigma_X B/\sigma_Y$. Substituting this into the first equation, $B = 1/[\sigma_X^2(1-\rho^2)]$. Substituting this into the equation for C, $C = -\rho/[\sigma_X\sigma_Y(1-\rho^2)]$. From symmetry or from the right two terms, we solve for D, getting

$$[K^{-1}] \;=\; \frac{1}{1-\rho^2}\begin{bmatrix} \sigma_X^{-2} & -\rho\sigma_X^{-1}\sigma_Y^{-1} \\ -\rho\sigma_X^{-1}\sigma_Y^{-1} & \sigma_Y^{-2} \end{bmatrix}.$$

Finally, noting that $B^{-1}C = \rho\sigma_X/\sigma_Y$, it is simple but slightly tedious to substitute these terms into (3.40) to get (3.37).

Exercise 3.16: **a)** Express $[B]$, $[C]$, and $[D]$, as defined in (3.39), in terms of $[K_{\boldsymbol{X}}]$, $[K_{\boldsymbol{Y}}]$ and $[K_{\boldsymbol{X}\cdot\boldsymbol{Y}}]$ 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 the left two of the four block equations in $[KK^{-1}]$. You can use the symmetry between $\boldsymbol{X}$ and $\boldsymbol{Y}$ 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 $[K_{\boldsymbol{X}}B] + [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}C^{\mathsf{T}}] = [I]$ and $[K_{X\cdot Y}^{\mathsf{T}}B] + [K_{\boldsymbol{Y}}C^{\mathsf{T}}] = 0$. These involve only two of the unknown terms, and we now solve for those terms. Recognizing that $[K_{\boldsymbol{X}}]$ and $[K_{\boldsymbol{Y}}]$ are invertible, we can rearrange the second equation as

$$[C^{\mathsf{T}}] \;=\; -[K_{\boldsymbol{Y}}^{-1}K_{X\cdot Y}^{\mathsf{T}}B].$$

Substituting this into the first equation, we get

$$\Big[[K_{\boldsymbol{X}}] - [K_{X\cdot Y}K_{\boldsymbol{Y}}^{-1}K_{X\cdot Y}^{\mathsf{T}}]\Big][B] = [I].$$

Now $[B]$ must be invertible (See Exercise 3.11c) Thus the matrix preceeding $[B]$ above is also invertible, so

$$[B] = \Big[[K_{\boldsymbol{X}}] - [K_{X\cdot Y}K_{\boldsymbol{Y}}^{-1}K_{X\cdot Y}^{\mathsf{T}}]\Big]^{-1}.$$

This agrees wth the solution (derived very differently) in (3.46). Next, to solve for $[C]$, we take the transpose of $[C^{\mathsf{T}}]$ above, leading to $[C] = -[BK_{\boldsymbol{X}\cdot\boldsymbol{Y}}K_{\boldsymbol{Y}}^{-1}]$. 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 (a) for $[C]$ plus the symmetry between $\boldsymbol{X}$ and $\boldsymbol{Y}$ to show that

$$[BK_{\boldsymbol{X}\cdot\boldsymbol{Y}}K_{\boldsymbol{Y}}^{-1}] = [K_{\boldsymbol{X}}^{-1}K_{\boldsymbol{X}\cdot\boldsymbol{Y}}D].$$

Solution: The quantity on the left above is $-[C]$ as derived in (a). By using the symmetry between $\boldsymbol{X}$ and $\boldsymbol{Y}$, we see that $[DK_{\boldsymbol{X}\cdot\boldsymbol{Y}}^{\mathsf{T}}K_{\boldsymbol{X}}^{-1}]$ is $-C^{\mathsf{T}}$, and taking the transpose completes the argument.

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

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

Exercise 3.21: **a)** Let $X(t) = R\cos(2\pi ft + \theta)$ where R is a Rayleigh rv and the rv θ is independent of R and uniformly distributed over the interval 0 to 2π. Show that $\mathsf{E}[X(t)] = 0$.

Solution: This can be done by standard (and quite tedious) manipulations, but if we first look at $t = 0$ and condition on a sample value of R, we are simply looking at $\cos(\theta)$, and since θ is uniform over $[0, 2\pi)$, it seems almost obvious that the mean should be 0. To capture this intuition, note that $\cos(\theta) = -\cos(\theta + \pi)$. Since θ is uniform between 0 and 2π, $\mathsf{E}[\cos(\theta)] = \mathsf{E}[\cos(\theta + \pi)]$, so that $\mathsf{E}[\cos(\theta)] = 0$. The same argument works for any t, so the result follows.

b) Show that $\mathsf{E}[X(t)X(t + \tau)] = \frac{1}{2}\mathsf{E}\left[R^2\right]\cos(2\pi f\tau)$.

Solution: Since θ and R are independent, we have

$$\begin{aligned}
\mathsf{E}[X(t)X(t+\tau)] &= \mathsf{E}\left[R^2\right]\mathsf{E}[\cos(2\pi ft + \theta)\cos(2\pi f(t+\tau) + \theta)] \\
&= \mathsf{E}\left[R^2\right]\frac{1}{2}\mathsf{E}[\cos(4\pi ft + 2\pi f\tau + 2\theta) + \cos(2\pi f\tau)] \\
&= \frac{\mathsf{E}\left[R^2\right]\cos(2\pi f\tau)}{2}.
\end{aligned}$$

where we used a standard trigonometric identity and then took the expectation over θ using the same argument as in (a).

c) Show that $\{X(t);\, t\in\mathbb{R}\}$ is a Gaussian process.

Solution: Let W_1, W_2 be IID normal Gaussian rv's. These can be expressed in polar coordinates as $W_1 = R\cos\theta$ and $W_2 = R\sin\theta$, where R is Rayleigh and θ is uniform. The rv $R\cos\theta$ is then $\mathcal{N}(0, 1)$. Similarly, $X(t)$ is a linear combination of W_1 and W_2 for each t, so each set $\{X(t_1), X(t_2), \ldots X(t_k)\}$ of rv's is jointly Gaussian. It follows that the process is Gaussian.

Exercise 3.22: Let $h(t)$ be a real square-integrable function whose Fourier transform is 0 for $|f| > B$

for some $B > 0$. Show that $\sum_n h^2(t - n/2B) = (1/2B)\int h^2(\tau)\,d\tau$ for all $t \in \Re$. Hint: find the sampling theorem expansion for a time shifted sinc function.

Solution: We use a slightly simpler approach than that of the hint. The sampling theorem expansion of $h(t)$ is given by

$$h(t) = \sum_n h\big(\frac{n}{2B}\big)\mathrm{sinc}(2Bt - n).$$

Since the functions $\{\sqrt{2B}\,\mathrm{sinc}(2Bt - n);\ n \in \mathbb{Z}\}$ are orthonormal, the energy equation, (3.64), says that $\int_{-\infty}^{\infty} h^2(t)dt = \sum_n 2Bh^2\big(n/2B)\big)$. For the special case of $t = 0$, this is the same as what is to be shown (summing over $-n$ instead of n). To generalize this, consider $h(t + \tau)$ as a function of t for fixed τ. Then using the same sampling expansion on this shifted function,

$$h(t + \tau) = \sum_n h\big(\frac{n}{2B} + \tau\big)\mathrm{sinc}(2Bt - n).$$

The Fourier transform of $h(t + \tau)$ (as a function of t for fixed τ) is still 0 for $f > B$, and $\int h^2(t)dt = \int h^2(t + \tau)\,dt$. Thus,

$$\int h^2(t)\,dt = \sum_n 2Bh^2\Big(\frac{n}{2B} + \tau\Big).$$

Replacing n by $-n$ in the sum over $\mathbb{Z}$ and interchanging t and τ, we have the desired result.

A.4 Solutions for Chapter 4

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

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

Solution: If there is no such state i_2, then i_1 is recurrent by definition. That state is distinct from i_1 since otherwise $i_1 \to i_2$ would imply $i_2 \to i_1$.

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

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

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

Solution: The argument why i_{k+1} exists with $i_k \to i_{k+1}$, $i_{k+1} \not\to i_k$ 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 $i_{k+1} \to i_{j+1} \to i_k$.

d) Show that for some $k \leq \mathsf{M}$, k is not transient, *i.e.*, it is recurrent, so a recurrent class exists.

Solution: For transient states $i_1, \dots, i_k$ generated in (c), state i_{k+1} found in (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 $k = \mathsf{M}$, a new distinct state $i_{\mathsf{M}+1}$ would be generated, which is impossible. Thus there must be some $k < \mathsf{M}$ 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 exactly one recurrent class $\mathcal{R}$ of states and state $1 \in \mathcal{R}$.

Solution: Since $j \to 1$ for each j, there can be no state j for which $1 \to j$ and $j \not\to 1$. Thus state 1 is recurrent. Next, for any given j, if $1 \not\to j$, then j must be transient since $j \to 1$. On the other hand, if $1 \to j$, 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 4.5: **(Proof of Theorem 4.2.11) a)** Show that an ergodic Markov chain with $\mathsf{M} > 1$ states must contain a cycle with $\tau < \mathsf{M}$ states. Hint: Use ergodicity to show that the smallest cycle cannot contain M states.

Solution: The states in any cycle (not counting the initial state) are distinct and thus the number of steps in a cycle is at most M. A recurrent chain must contain cycles, since for each pair of states $\ell \neq j$, there is a walk from ℓ to j and then back to ℓ; if any state i other than ℓ 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 $P_{ij} > 0$ for which (i, j) is not a transition on that cycle, then that transition can be added to the cycle and all the transitions between i and j on the existing cycle can be omitted, thus creating a cycle of fewer than M steps. If there are no nonzero transitions other than those in a cycle with M steps, then the Markov chain is periodic with period M and thus not ergodic.

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

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

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

$$\mathcal{T}(0) \subseteq \mathcal{T}(\tau) \subseteq \mathcal{T}(2\tau) \subseteq \cdots \subseteq \mathcal{T}(n\tau) \subseteq \cdots .$$

Solution: Since $\mathcal{T}(0) = \{\ell\}$ and $\ell \in \mathcal{T}(\tau)$, we see that $\mathcal{T}(0) \subseteq \mathcal{T}(\tau)$. Next, for each $n \geq 1$, use (b), with $m = n\tau$, to see that $\mathcal{T}(n\tau) \subseteq \mathcal{T}(n\tau + \tau)$. 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 $\mathcal{T}(n\tau) = \mathcal{T}((M-1)\tau)$ for all $n \geq M-1$.

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

$$\mathcal{T}(n\tau) = \mathcal{T}(k\tau) \qquad \text{for all } n \geq k.$$

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

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

Solution: For any t such that $P_{\ell\ell}^t > 0$, we can repeat the argument in part (b), replacing τ by t to see that for any $m \geq 1$, $\mathcal{T}(m) \subseteq \mathcal{T}(m+t)$. Thus we have

$$\mathcal{T}\big((\mathsf{M}{-}1)\tau\big) \subseteq \mathcal{T}\big((M{-}1)\tau + t\big) \subseteq \cdots \subseteq \mathcal{T}\big((M{-}1)\tau + t\tau\big) = \mathcal{T}\big((\mathsf{M}{-}1)\tau\big),$$

where (d) was used in the final equality. This shows that all the inclusions above are satisfied with equality and thus that $\mathcal{T}\big((\mathsf{M}{-}1)\tau\big) = \mathcal{T}\big((M{-}1)\tau + kt\big)$ for all $k \leq \tau$. Using t in place of τ in the argument in (d), this can be extended to

$$\mathcal{T}\big((\mathsf{M}{-}1)\tau\big) = \mathcal{T}\big((M{-}1)\tau + kt\big) \qquad \text{for all } k \geq 1.$$

Since the chain is ergodic, we can choose t so that both $P^t_{\ell\ell} > 0$ 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

$$\mathcal{T}\big((M-1)\tau\big) = \mathcal{T}\big((\mathsf{M}-1)\tau + kt\big) = \mathcal{T}\big((\mathsf{M}-1+j)\tau + 1\big) = \mathcal{T}\big((M-1)\tau+1\big). \qquad (4.5a)$$

As in (d), $\mathcal{T}\big((M-1)\tau+2\big)$ is the set of states reachable in one step from $\mathcal{T}\big((M-1)\tau+1\big)$. From (4.5a), this is the set of states reachable from $\mathcal{T}\big((M-1)\tau\big)$ in 1 step, *i.e.*,

$$\mathcal{T}\big((M-1)\tau+2\big) = \mathcal{T}\big((M-1)\tau+1\big) = \mathcal{T}\big((M-1)\tau\big).$$

Extending this,

$$\mathcal{T}\big((M-1)\tau\big) = \mathcal{T}\big((M-1)\tau+m\big) \qquad \text{for all } m \geq 1.$$

This means that $\mathcal{T}\big((M-1)\tau\big)$ contains all states that can ever occur from time $\big((M-1)\tau$ on, and thus must contain all states since the chain is recurrent.

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

Solution: We have shown that all states are accessible from state ℓ at all times $\tau(\mathsf{M}-1)$ or later, and since $\tau \leq \mathsf{M}-1$, all are accessible at all times $n \geq (\mathsf{M}-1)^2$. The same applies to any state on a cycle of length at most $\mathsf{M}-1$. 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 $\mathsf{M}-\tau$ steps. Using this path to reach a state on the cycle and following this by paths of length $\tau(\mathsf{M}-1)$, all states can reach all other states at all times greater than or equal to

$$\tau(\mathsf{M}-1) + \mathsf{M} - \tau \leq (\mathsf{M}-1)^2 + 1.$$

The above derivation assumed $M > 1$. The case $M = 1$ is obvious, so the theorem is proven.

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, each row sum and each column sum equals 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 $[P]\boldsymbol{e} = \boldsymbol{e}$. If the column sums are also equal to 1, then $\boldsymbol{e}^{\mathsf{T}}[P] = \boldsymbol{e}^{\mathsf{T}}$. Thus $\boldsymbol{e}^{\mathsf{T}}$ is a left eigenvector of $[P]$ with eigenvalue 1, and it is unique within a scale factor since the chain is ergodic. Scaling $\boldsymbol{e}^{\mathsf{T}}$ to be probabilities, $\boldsymbol{\pi} = (1/\mathsf{M}, 1/\mathsf{M}, \dots, 1/\mathsf{M})$.

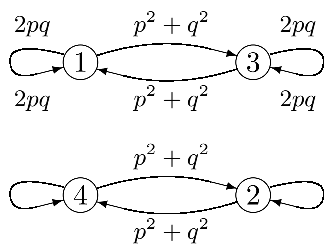
Exercise 4.10: **a)** Find the steady-state probabilities for each of the Markov chains in Figure 4.2. Assume that all clockwise probabilities in the first graph are the same, say p, and assume that $P_{4,5} = P_{4,1}$ in the second graph.

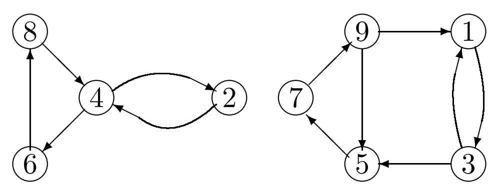
Solution: These probabilities can be found in a straightforward but tedious fashion by solving (4.8). Note that $\boldsymbol{\pi} = \boldsymbol{\pi}[P]$ is a set of M linear equations of which only $\mathsf{M}-1$ are linearly independent and $\sum \pi_i = 1$ provides the needed extra equation. The solutions are $\pi_i = 1/4$ for each state in the first graph and $\pi_i = 1/10$ for all but state 4 in the second graph; $\pi_4 = 1/5$.

One learns more by trying to find $\boldsymbol{\pi}$ by inspection. For the first graph, the π_i are clearly equal by symmetry. For the second graph, states 1 and 5 are immediately accessible only from state 4 and are thus equally likely and each has half the probability of 4. The probabilities on the states of each loop should be the same, leading to the answer above. It would be prudent to check this answer by (4.8), but that is certainly easier than solving (4.8).

b) Find the matrices $[P^2]$ for the same chains. Draw the graphs for the Markov chains represented by $[P^2]$, i.e., the graph of two step transitions for the original chains. Find the steady-state probabilities for these two-step chains. Explain why your steady-state probabilities are not unique.

Solution: Let $q = 1-p$ in the first graph. In the second graph, all transitions out of states 3, 4, and 9 have probability 1/2. All other transitions have probability 1.





One steady-state probability for the first chain is $\pi_1 = \pi_3 = 1/2$ and the other is $\pi_2 = \pi_4 = 1/2$. These are the steady-state probabilities for the two recurrent classes of $[P^2]$. The second chain also has two recurrent classes. The steady-state probabilities for the first are $\pi_2 = \pi_6 = \pi_8 = 0.2$ and $\pi_4 = .4$. Those for the second are $\pi_1 = \pi_3 = \pi_5 = \pi_7 = \pi_9 = 0.2$.

c) Find $\lim_{n\to\infty}[P^{2n}]$ for each of the chains.

Solution: The limit for each chain is block diagonal with one block being the even numbers and the other the odd numbers. Within a block, the rows are the same. For the first chain, the blocks are (1, 3) and (2, 4). We have $\lim_{n\to\infty} P_{ij}^{2n} = 1/2$ for i, j both odd or both even; it is 0 otherwise. For the second chain, within the even block, $\lim_{n\to\infty} P_{ij}^{n} = 0.2$ for $j \neq 4$ and 0.4 for $j = 4$. For the odd block, $\lim_{n\to\infty} P_{ij}^{n} = 0.2$ for all odd i, j.

Exercise 4.13: Consider a finite state Markov chain with matrix $[P]$ which has κ aperiodic recurrent classes, $\mathcal{R}_1, \dots, \mathcal{R}_\kappa$ and a set $\mathcal{T}$ of transient states. For any given recurrent class $\mathcal{R}_\ell$, consider a vector $\boldsymbol{\nu}$ such that $\nu_i = 1$ for each $i \in \mathcal{R}_\ell$, $\nu_i = \lim_{n\to\infty} \Pr\{X_n \in \mathcal{R}_\ell | X_0 = i\}$ for each $i \in \mathcal{T}$, and $\nu_i = 0$ otherwise. Show that $\boldsymbol{\nu}$ is a right eigenvector of $[P]$ with eigenvalue 1. Hint: Redraw Figure 4.5 for multiple recurrent classes and first show that $\boldsymbol{\nu}$ is an eigenvector of $[P^n]$ in the limit.

Solution: A simple example of this result is treated in Exercise 4.29 and a complete derivation (extended almost trivially to periodic as well as aperiodic recurrent classes) is given in the solution to Exercise 4.18. Thus we give a more intuitive and slightly less complete derivation here.

Number the states with the transient states first, followed by each recurent class in order.

Then $[P]$ has the following block structure

$$[P] = \begin{bmatrix} [P_{\mathcal{T}}] & [P_{\mathcal{T}\mathcal{R}_1}] & \ddots & \ddots & [P_{\mathcal{T}\mathcal{R}_\kappa}] \\ 0 & [P_{\mathcal{R}_1}] & 0 & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & 0 & [P_{\mathcal{R}_\kappa}] \end{bmatrix}.$$

The ℓth recurrent class has an $|\mathcal{R}_\ell|$ by $|\mathcal{R}_\ell|$ transition matrix which, viewed alone, has an eigenvalue $\lambda = 1$ of multiplicity 1, a corresponding unique (within a scale factor) left eigenvector, say $\boldsymbol{\pi}(\mathcal{R}_\ell)$, and a corresponding unique (within a scale factor) right eigenvector, $\boldsymbol{\nu}(\mathcal{R}_\ell) = (1, \dots, 1)^\mathsf{T}$ (see Theorem 4.4.2).

Let $\boldsymbol{\pi}^{(\ell)}$ be an M dimensional row vector whose components are equal to those of $\boldsymbol{\pi}(\mathcal{R}_\ell)$ over the states of $\mathcal{R}_\ell$ and equal to 0 otherwise. Then it can be seen by visualizing elementary row/matrix multiplication on the block structure of $[P]$ that $\boldsymbol{\pi}^{(\ell)}[P] = \boldsymbol{\pi}^{(\ell)}$. This gives us κ left eigenvectors of eigenvalue 1, one for each recurrent class $\mathcal{R}_\ell$.

These κ left eigenvectors are clearly linearly independent and span the κ dimensional space of left eigenvectors of eigenvalue 1 (see Exercise 4.18).

If there are no transient states, then a set of κ right eigenvectors can be chosen in the same way as the left eigenvectors. That is, for each ℓ, $1 \le \ell \le \kappa$, the components of $\boldsymbol{\nu}^{(\ell)}$ can be chosen to be 1 for each state in $\mathcal{R}_\ell$ and 0 for all other states. This doesn't satisfy the eigenvector equation if there are transient states, however. We now show, instead, that for each ℓ, $1 \le \ell \le \kappa$, there is a right eigenvector $\boldsymbol{\nu}^{(\ell)}$ of eigenvalue 1 that can be nonzero both on $\mathcal{R}_\ell$ and $\mathcal{T}$. such that $\nu_i^{(\ell)} = 0$ for all $i \in \mathcal{R}_m$, for each $m \ne \ell$. Finally we will show that these κ vectors are linearly independent and have the properties specified in the problem statement.

The right eigenvector equation that must be satisfied by $\boldsymbol{\nu}^{(\ell)}$ assuming that $\nu_i^{(\ell)} \ne 0$ only for $i \in \mathcal{R}_\ell \cup \mathcal{T}$ can be written out component by component, getting

$$\begin{aligned} \nu_i^{(\ell)} &= \sum_{j\in\mathcal{R}_\ell} P_{ij}\nu_j^{(\ell)} && \text{for } i \in \mathcal{R}_\ell \\ \nu_i^{(\ell)} &= \sum_{j\in\mathcal{T}} P_{ij}\nu_j^{(\ell)} + \sum_{j\in\mathcal{R}_\ell} P_{ij}\nu_j^{(\ell)} && \text{for } i \in \mathcal{T}. \end{aligned}$$

The first set of equations above are simply the usual right eigenvector equations for eigenvalue 1 over the recurrent submatrix $[P_{\mathcal{R}_\ell}]$. Thus $\nu_j^{(\ell)} = 1$ for $j \in \mathcal{R}_\ell$ and this solution (over $\mathcal{R}_\ell$) is unique within a scale factor. Substituting this solution into the second set of equations, we get

$$\nu_i^{(\ell)} = \sum_{j\in\mathcal{T}} P_{ij}\nu_j^{(\ell)} + \sum_{j\in\mathcal{R}_\ell} P_{ij} \qquad \text{for } i \in \mathcal{T}.$$

This has a unique solution for each ℓ (see Exercise 4.18). These eigenvectors, $\{\boldsymbol{\nu}^{(\ell)}, 1 \le \ell \le \kappa\}$ must be linearly independent since $\nu_i^{(\ell)} = 1$ for $i \in \mathcal{R}_\ell$ and $\nu_i^{(\ell)} = 0$ for $i \in \mathcal{R}_m$, $m \ne \ell$. They then form a basis for the κ dimensional space of eigenvectors of $[P]$ of eigenvalue 1.

These eigenvectors are also eigenvectors of eigenvalue 1 for $[P^n]$ for each $n > 1$. Thus

$$\nu_i^{(\ell)} = \sum_{j\in\mathcal{T}} P_{ij}^n \nu_j^{(\ell)} + \sum_{j\in\mathcal{R}_\ell} P_{ij}^n \qquad \text{for } i \in \mathcal{T}.$$

Now recall that for $i, j \in \mathcal{T}$, we have $\lim_{n\to\infty} P_{ij}^n = 0$. Also $\sum_{j\in\mathcal{R}_\ell} P_{ij}^n$ is the probability that $X_n \in \mathcal{R}_\ell$ given $X_0 \in \mathcal{T}$. Since there is no exit from $\mathcal{R}_\ell$, this quantity is non-decreasing in n and bounded by 1, so it has a limit. This limit is the probability of ever going from i to $\mathcal{R}_\ell$, completing the derivation.

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.

$$[P^2] = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/4 & 2/4 & 1/4 \\ 0 & 1/4 & 3/4 \\ 0 & 0 & 1 \end{bmatrix}.$$

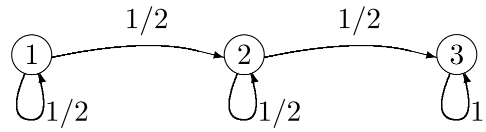
$$[P^3] = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/4 & 2/4 & 1/4 \\ 0 & 1/4 & 3/4 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/8 & 3/8 & 4/8 \\ 0 & 1/8 & 7/8 \\ 0 & 0 & 1 \end{bmatrix}.$$

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 it is $n2^{-n}$ and verify it by induction,

$$\begin{aligned} [P^{n+1}] &= [P]\,[P^n] = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2^{-n} & n2^{-n} & 1-(n{+}1)2^{-n} \\ 0 & 2^{-n} & 1-2^{-n} \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2^{-n-1} & (n{+}1)2^{-n-1} & 1-(n{+}2)2^{-n-1} \\ 0 & 2^{-n} & 1-2^{-n} \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

This solution is not very satisfying, first because it is tedious, second because it required a somewhat unmotivated guess, 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,



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 $1 \to 2$ 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_{i3}^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, $\boldsymbol{e} = (1, \dots, 1)^{\mathsf{T}}$ is a right eigenvector. For the given chain, this is unique within a scale factor, since $\lambda = 1$ has multiplicity 1. For $\boldsymbol{\nu}$ to be a right eigenvector of eigenvalue 1/2, it must satisfy

$$\begin{aligned} \frac{1}{2}\nu_1 + \frac{1}{2}\nu_2 + 0\nu_3 &= \frac{1}{2}\nu_1 \\ 0\nu_1 + \frac{1}{2}\nu_2 + \frac{1}{2}\nu_3 &= \frac{1}{2}\nu_2 \\ \nu_3 &= \frac{1}{2}\nu_3. \end{aligned}$$

From the first equation, $\nu_2 = 0$ and from the third $\nu_3 = 0$, so $\boldsymbol{\nu} = (1, 0, 0)$ is the right eigenvector of $\lambda = 1/2$, unique within a scale factor. Thus $\lambda = 1/2$ does not have 2 linearly independent eigenvectors.

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 $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3$ be the columns of an hypothesized matrix $[U]$, we see that $[P][U] = [U][\Lambda]$ can be written out as $[P]\boldsymbol{\nu}_i = \lambda_i \boldsymbol{\nu}_i$ for $i = 1, 2, 3$. For $[U]$ to be invertible, $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3$ must be linearly independent eigenvectors of $[P]$. However part (c) showed that 3 such eigenvectors do not exist.

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

Solution: If the $[U]$ and $[\Lambda]$ of (d) exist, then $[P^n] = [U][\Lambda^n][U^{-1}]$. Then, as in (4.30), $[P^n] = \sum_{i=1}^3 \lambda_i^n \boldsymbol{\nu}^{(i)} \boldsymbol{\pi}^{(i)}$ where $\boldsymbol{\nu}^{(i)}$ is the ith column of $[U]$ and $\boldsymbol{\pi}^{(i)}$ 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 λ_i^n, then the eigenvectors can't span the space and a Jordan form decomposition is required.

Exercise 4.16: **a)** Let λ be an eigenvalue of a matrix $[A]$, and let $\boldsymbol{\nu}$ and $\boldsymbol{\pi}$ be right and left eigenvectors respectively of λ, normalized so that $\boldsymbol{\pi}\boldsymbol{\nu} = 1$. Show that

$$[[A] - \lambda\boldsymbol{\nu}\boldsymbol{\pi}]^2 = [A^2] - \lambda^2\boldsymbol{\nu}\boldsymbol{\pi}.$$

Solution: We simply multiply out the original square,

$$\begin{aligned}
[[A] - \lambda\boldsymbol{\nu}\boldsymbol{\pi}]^2 &= [A^2] - \lambda\boldsymbol{\nu}\boldsymbol{\pi}[A] - \lambda[A]\boldsymbol{\nu}\boldsymbol{\pi} + \lambda^2\boldsymbol{\nu}\boldsymbol{\pi}\boldsymbol{\nu}\boldsymbol{\pi} \\
&= [A^2] - \lambda^2\boldsymbol{\nu}\boldsymbol{\pi} - \lambda^2\boldsymbol{\nu}\boldsymbol{\pi} + \lambda^2\boldsymbol{\nu}\boldsymbol{\pi} \;=\; [A^2] - \lambda^2\boldsymbol{\nu}\boldsymbol{\pi}.
\end{aligned}$$

b) Show that $\big[[A^n] - \lambda^n\boldsymbol{\nu}\boldsymbol{\pi}\big]\,\big[[A] - \lambda\boldsymbol{\nu}\boldsymbol{\pi}\big] = [A^{n+1}] - \lambda^{n+1}\boldsymbol{\nu}\boldsymbol{\pi}$.

Solution: This is essentially the same as (a)

$$\begin{aligned}
\big[[A^n] - \lambda^n\boldsymbol{\nu}\boldsymbol{\pi}\big]\,\big[[A] - \lambda\boldsymbol{\nu}\boldsymbol{\pi}\big] &= [A^{n+1}] - \lambda^n\boldsymbol{\nu}\boldsymbol{\pi}[A] - \lambda[A^n]\boldsymbol{\nu}\boldsymbol{\pi} + \;\lambda^{n+1}\boldsymbol{\nu}\boldsymbol{\pi}\boldsymbol{\nu}\boldsymbol{\pi} \\
&= [A^{n+1}] - \lambda^{n+1}\boldsymbol{\nu}\boldsymbol{\pi}.
\end{aligned}$$

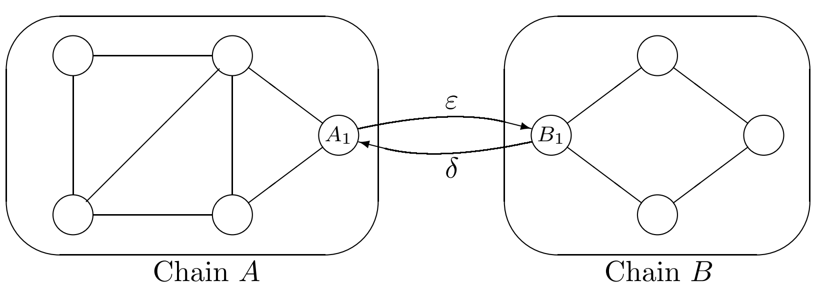
c) Use induction to show that $[[A] - \lambda\boldsymbol{\nu}\boldsymbol{\pi}]^n = [A^n] - \lambda^n\boldsymbol{\nu}\boldsymbol{\pi}$.

Solution: (a) gives the base of the induction and (b) gives the inductive step.

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

$$\begin{aligned}
P'_{A_1,B_1} &= \varepsilon, \quad P'_{A_1,A_j} = (1-\varepsilon)P_{A_1,A_j} \qquad \text{for all } A_j \\
P'_{B_1,A_1} &= \delta, \quad P'_{B_1,B_j} = (1-\delta)P_{B_1,B_j} \qquad \text{for all } B_j.
\end{aligned}$$

All other transition probabilities remain the same. Think intuitively of ε and δ as being small, but do not make any approximations in what follows. Give your answers to the following questions as functions of ε, δ, $\{\pi_{A_i}\}$ 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.

Solution: Conditional on the first transition from state A_1 being to a state $A_i \neq A_1$, 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 π_{A_1} so we might hypothesize that, within A, 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 simple 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 $(1, 2, \dots, \mathsf{M})$ where 1 stands for A_1. For $2 \leq i \leq \mathsf{M}$, let v_i be the expected time to first reach state 1 starting in state i. As in (4.31),

$$v_i = 1 + \sum_{j \neq 1} P_{ij} v_j; \qquad 2 \leq i \leq \mathsf{M}. \tag{A.15}$$

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. We define $v_1 = 0$ since if the first transition from 1 goes to 1, there is no remaining time required to return to state A_1. We then have

$$T = 1 + \sum_{j=1}^{\mathsf{M}} P_{ij} v_j. \tag{A.16}$$

Note that this is very different from (4.32) where $[P]$ is a Markov chain in which 1 is a trapping state. We can now combine (A.16) (for component 1) and (A.15) (for components 2 to M) into the following vector equation:

$$T\boldsymbol{e}_1 + \boldsymbol{v} = \boldsymbol{e} + [P]\boldsymbol{v},$$

where $\boldsymbol{e}_1 = (1, 0, 0, \dots, 0)^\mathsf{T}$ and $\boldsymbol{e} = (1, 1, \dots, 1)^\mathsf{T}$. Motivated by the hypothesis that $T = 1/\pi_1$, we premultiply this vector equation by the steady-state row vector $\boldsymbol{\pi}$, getting

$$T\pi_1 + \boldsymbol{\pi v} = 1 + \boldsymbol{\pi}[P]\boldsymbol{v} \;=\; 1 + \boldsymbol{\pi v}.$$

Cancelling $\boldsymbol{\pi v}$ from each side, 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 ϵ 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 ϵ. With probability $1 - \epsilon$, we wait for a return to A_1 and then have expected time $T_{A,B}$ remaining. Thus $T_{A,B} = \epsilon + (1-\epsilon)\big(\frac{1}{\pi_{A_1}} + T_{A,B}\big)$. Solving this equation,

$$T_{A,B} = 1 + \frac{1-\epsilon}{\epsilon \pi_{A_1}}.$$

c) Assume $\varepsilon > 0$, $\delta > 0$. Find $T_{B,A}$, the expected time to first reach state A_1, starting in state B_1. Your answer should depend only on δ 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 π_{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 $\varepsilon > 0$ 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 $0 < \epsilon < 1$ and $0 < \delta < 1$, the combined chain is ergodic. To see this, note that all states communicate with each other so the combined chain is recurrent. Also, all walks in A are still walks in the combined chain, so the gcd of their lengths is 1. Thus A, and consequently B, are still aperiodic.

We can thus use the steady state equations to find the unique steady-state vector $\boldsymbol{\pi}$. In parts (d), (e), and (f), we are interested in those probabilities in chain A. We denote those states, as before, as $(1, \dots, \mathsf{M})$ where 1 is state A_1. Steady-state probabilities for A in the combined chain for given ϵ, δ are denoted π'_j, whereas they are denoted as π_j in the original chain. We first find π'_1 and then π'_j for $2 \le j \le \mathsf{M}$.

As we saw in (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_1 to A_1. Given that the first transition from state 1 goes to a state in A, the expected first-return time is $1/\pi_1$ from (a). If the transition goes to B_1, the expected first-return time is $1+T_{BA}$, where T_{BA} is found in (c). Combining these with the a priori probabilities of going to A or B,

$$T_{AA} = (1-\epsilon)/\pi_1 + \epsilon\big[1 + T_{BA}\big] \;=\; \frac{1-\epsilon}{\pi_1} + 2\epsilon + \frac{\epsilon(1-\delta)}{\delta\pi_{B_1}}.$$

Thus

$$\pi'_1 = \left[\frac{1-\epsilon}{\pi_1} + 2\epsilon + \frac{\epsilon(1-\delta)}{\delta\pi_{B_1}}\right]^{-1}.$$

Next we find π'_j for the other states in A in terms of the π_j for the uncombined chains and π'_1. The original and the combined steady-state equations for $2 \le j \le \mathsf{M}$ are

$$\pi_j = \sum_{i \ne 1} \pi_i P_{ij} + \pi_1 P_{1j}; \qquad \pi'_j = \sum_{i\ne 1} \pi'_i P_{ij} + \pi'_1 (1-\epsilon) P_{1j}.$$

These equations, as $\mathsf{M}-1$ equations in the unknowns π_j; $j \ge 2$ for given π_1, uniquely specify $\pi_2, \dots, \pi_{\mathsf{M}}$ and they differ in π_1 being replaced by $(1-\epsilon)\pi'_1$. From this, we see that the second set of equations is satisfied if we choose

$$\pi'_j = \pi_j \frac{(1-\epsilon)\pi'_1}{\pi_1}.$$

We can now sum the steady-state probabilities in A to get

$$\Pr\{A\} = \sum_{j=1}^{\mathsf{M}} \pi'_j = \pi'_1 \left[\epsilon + \frac{1-\epsilon}{\pi_1}\right].$$

e) Assume $\varepsilon > 0$, $\delta = 0$. 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 ε and $\{\pi_{A_i}\}$.

Solution: We use a variation on the first passage time problem in (a) to find the expected number of visits to state j, $\mathsf{E}[N_j]$, in the original chain starting in state 1 before the first return to 1. Here we let $v_i(j)$ be the expected number of visits to j, starting in state i, before the first return to 1. The equations are

$$\mathsf{E}[N_j] = \sum_{k\neq 1} P_{1k} v_k(j) + P_{1j}; \qquad v_i(j) = \sum_{k\neq 1} P_{ik} v_k(j) + P_{ij} \quad \text{for } i \neq 1.$$

The first equation represents $\mathsf{E}[N(j)]$ in terms of the expected number of visits to j conditional on each first transition k from the initial state 1. in the special case where that first transition is to j, the expected number of visits to j includes both 1 for the initial visit plus $v_j(j)$ for the subsequent visits. The second set of equations are similar, but give the expected number of visits to j (before a return to 1) starting from each state other than 1.

Writing this as a vector equation, with $\boldsymbol{v}(j) = (0, v_2(j), v_3(j). \dots, v_{\mathsf{M}}(j))^{\mathsf{T}}$, we get

$$\mathsf{E}[N_j]\,\boldsymbol{e}_1 + \boldsymbol{v}(j) = [P]\boldsymbol{v}(j) + [P]\boldsymbol{e}_j.$$

Note that $[P]\boldsymbol{e}_j$ is the column vector $(p_{1j}, \dots, P_{\mathsf{M}j})^{\mathsf{T}}$ Premultiplying by $\boldsymbol{\pi}$, we see that $\mathsf{E}[N_j] = \pi_j/\pi_1$. Finally, to find v_{A_j}, the expected number of visits to state j before the first to B_1, we have

$$v_{A_j} = (1-\epsilon)[\mathsf{E}[N_j] + v_{A_j}] = \frac{(1-\epsilon)\mathsf{E}[N_j]}{\epsilon} = \frac{(1-\epsilon)\pi_j}{\epsilon\pi_i}.$$

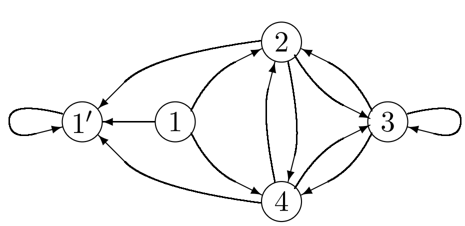
f) Assume $\varepsilon > 0$, $\delta > 0$. For each state A_j in A, find π'_{A_j}, the steady-state probability of being in state A_j in the combined chain. Hint: Be careful in your treatment of state A_1.

Solution: This was solved in (d). Readers might want to come back to this exercise later and re-solve it more simply using renewal theory.

Exercise 4.22: Section 4.5.1 showed how to find the expected first passage times to a fixed state, say 1, from all other states. It is often desirable to include the expected first recurrence time from state 1 to return to state 1. This can be done by splitting state 1 into 2 states, first an initial state with no transitions coming into it but the original transitions going out, and second, a final trapping state with the original transitions coming in.

a) For the chain on the left side of Figure 4.6, draw the graph for the modified chain with 5 states where state 1 has been split into 2 states.

Solution: We split state 1 into states $1'$ and 1, where $1'$ is the trapping state and 1 can be an initial state.



b) Suppose one has found the expected first-passage-times v_j for states $j = 2$ to 4 (or in general from 2 to M). Find an expression for v_1, the expected first recurrence time for state 1 in terms of $v_2, v_3, \dots v_{\mathsf{M}}$ and $P_{12}, \dots, P_{1\mathsf{M}}$.

Solution: Note that v_2, v_3, and v_4 are unchanged by the addition of state 1, since no transitions can go to 1 from states 2, 3, or 4. We then have

$$v_1 = 1 + \sum_{j=2}^{4} P_{1j} v_j .$$

Exercise 4.23: **a)** Assume throughout that $[P]$ is the transition matrix of a unichain (and thus the eigenvalue 1 has multiplicity 1). Show that a solution to the equation $[P]\boldsymbol{w} - \boldsymbol{w} = \boldsymbol{r} - g\boldsymbol{e}$ exists if and only if $\boldsymbol{r} - g\boldsymbol{e}$ lies in the column space of $[P - I]$ where $[I]$ is the identity matrix.

Solution: Let $\mathcal{C}[P - I]$ be the column space of $[P - I]$. A vector $\boldsymbol{x}$ is in $\mathcal{C}[P - I]$ if $\boldsymbol{x}$ is a linear combination of columns of $[P - I)$, *i.e.*, if $\boldsymbol{x}$ is w_1 times the first column of $[P - I]$ plus w_2 times the second column, etc. More succinctly, $\boldsymbol{x} \in \mathcal{C}[P - I]$ if and only if $\boldsymbol{x} = [P - I]\boldsymbol{w}$ for some vector $\boldsymbol{w}$. Thus $\boldsymbol{r} - g\boldsymbol{e} \in \mathcal{C}[P - I]$ if and only if $[P - I]\boldsymbol{w} = \boldsymbol{r} - g\boldsymbol{e}$ for some $\boldsymbol{w}$, which after rearrangement is what is to be shown.

b) Show that this column space is the set of vectors $\boldsymbol{x}$ for which $\boldsymbol{\pi}\boldsymbol{x} = 0$. Then show that $\boldsymbol{r} - g\boldsymbol{e}$ lies in this column space.

Solution: We know that $[P]$ has a single eigenvalue equal to 1. Thus $[P - I]$ is singular and the steady-state vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi}[P - I] = 0$. Thus for every $\boldsymbol{w}$, $\boldsymbol{\pi}[P - I]\boldsymbol{w} = 0$ so $\boldsymbol{\pi}\boldsymbol{x} = 0$ for every $\boldsymbol{x} \in \mathcal{C}[P - I]$. Furthermore, $\boldsymbol{\pi}$ (and its scalar multiples) are the only vectors to satisfy $\boldsymbol{\pi}[P - I] = 0$ and thus $\mathcal{C}[P - I]$ is an $\mathsf{M} - 1$ dimensional vector space. Since the vector space of vectors $\boldsymbol{x}$ that satisfy $\boldsymbol{\pi}\boldsymbol{x} = 0$ is $\mathsf{M} - 1$ dimensional, this must be the same as $\mathcal{C}[P - I]$. Finally, since g is defined as $\boldsymbol{\pi}\boldsymbol{r}$, we have $\boldsymbol{\pi}(\boldsymbol{r} - g\boldsymbol{e}) = 0$, so $\boldsymbol{r} - g\boldsymbol{e} \in \mathcal{C}[P - I]$.

c) Show that, with the extra constraint that $\boldsymbol{\pi}\boldsymbol{w} = 0$, the equation $[P]\boldsymbol{w} - \boldsymbol{w} = \boldsymbol{r} - g\boldsymbol{e}$ has a unique solution.

Solution: For any $\boldsymbol{w}'$ that satisfies $[P]\boldsymbol{w}' - \boldsymbol{w}' = \boldsymbol{r} - g\boldsymbol{e}$, it is easy to see that $\boldsymbol{w}' + \alpha\boldsymbol{e}$ also satisfies this equation for any real α. Furthermore, since the column space of $[P - I]$ is $\mathsf{M} - 1$ dimensional, this set of solutions, namely $\{\boldsymbol{w}' + \alpha\boldsymbol{e};\ \alpha \in \Re\}$ is the entire set of solutions. The additional constraint that $\boldsymbol{\pi}(\boldsymbol{w}' + \alpha\boldsymbol{e}) = \boldsymbol{\pi}\boldsymbol{w}' + \alpha = 0$ specifies a unique element in this set.

Exercise 4.24: For the Markov chain with rewards in Figure 4.8,

a) Find the solution to (4.37) and find the gain g.

Solution: The symmetry in the transition probabilities shows that $\boldsymbol{\pi} = (1/2, 1/2)^{\mathsf{T}}$, and thus $g = \boldsymbol{\pi}\boldsymbol{r} = 1/2$. The first component of (4.37), *i.e.*, of $\boldsymbol{w} + g\boldsymbol{e} = [P]\boldsymbol{w} + \boldsymbol{r}$ is then $w_1 + 1/2 = P_{11}w_1 + P_{12}w_2 + r_1$. This simplifies to $w_1 - w_2 = -50$. The second component is redundant, and $\boldsymbol{\pi}\boldsymbol{w} = 0$ simplifies to $w_1 + w_2 = 0$. Thus $w_1 = -25$ and $w_2 = 25$.

b) Modify Figure 4.8 by letting P_{12} be an arbitrary probability. Find g and $\boldsymbol{w}$ again and give an intuitive explanation of why P_{12} effects w_2.

Solution: With arbitrary P_{12}, the steady-state probabilities become

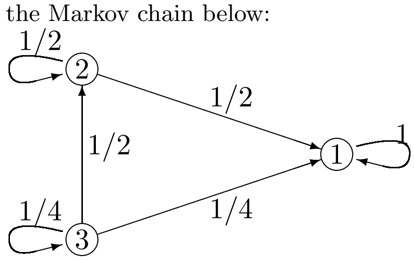
$$\pi_1 = \frac{0.01}{P_{12} + 0.01}; \qquad\qquad \pi_2 = \frac{P_{12}}{P_{12} + 0.01}.$$

The steady-state gain, $g = \boldsymbol{\pi r}$, then becomes $g = P_{12}/(P_{12}+0.01)$. Solving for $\boldsymbol{w}$ as before, we get

$$w_1 = \frac{-P_{12}}{(P_{12} + 0.01)^2} \qquad\qquad w_2 = \frac{0.01}{(P_{12} + 0.01)^2}.$$

As P_{12} increases, the mean duration of each visit to state 1 decreases so that the fraction of time spent in state 1 also decreases, thus increasing the expected gain. At the same time, the relative advantage of starting in state 2 decreases since the interruptions to having a reward on each transition become shorter. For example, if $P_{12} = 1/2$, the mean time to leave state 1, starting in state 1, is 2.

Exercise 4.26: Consider the Markov chain below:



a) Suppose the chain is started in state i and goes through n transitions; let $v_i(n)$ be the expected number of transitions (out of the total of n) until the chain enters the trapping state, state 1. Find an expression for $\boldsymbol{v}(n) = (v_1(n), v_2(n), v_3(n))^{\mathsf{T}}$ in terms of $\boldsymbol{v}(n-1)$ (take $v_1(n) = 0$ for all n). (Hint: view the system as a Markov reward system; what is the value of $\boldsymbol{r}$?)

Solution: We use essentially the same approach as in Section 4.5.1, but we are explicit here about the number of transitions. Starting in any state $i \neq 1$ with n transitions to go, the expected number of transitions until the first that either enters the trapping state or completes the nth transition is given by

$$v_i(n) = 1 + \sum_{j=1}^{3} P_{ij} v_j(n{-}1). \tag{A.17}$$

Since $i = 1$ is the trapping state, we can express $v_1(n) = 0$ for all n, since no transitions are required to enter the trapping state. Thus (A.17) is modified to $v_1(n) = \sum_{j=1}^{3} P_{1j} v_j(n{-}1)$ for $i = 1$. Viewing $\boldsymbol{r}$ as a reward vector, with one unit of reward for being in any state other than the trapping state, we have $\boldsymbol{r} = (0, 1, 1)^{\mathsf{T}}$. Thus (A.17) can be expressed in vector form as

$$\boldsymbol{v}(n) = \boldsymbol{r} + [P]\boldsymbol{v}(n-1).$$

b) Solve numerically for $\lim_{n\to\infty} \boldsymbol{v}(n)$. Interpret the meaning of the elements v_i in the solution of (4.32).

Solution: We have already seen that $v_1(n) = 0$ for all n, and thus, since $P_{23} = 0$, we have $v_2(n) = 1 + (1/2)v_2(n-1)$. Since $v_2(0) = 0$, this iterates to

$$v_2(n) \;=\; 1 + \frac{1}{2}\big(1 + v_2(n-2)\big) \;=\; 1 + \frac{1}{2} + \frac{1}{4}\big(1 + v_2(n-3)\big) \;=\; \cdots \;=\; 2 - 2^{-(n-1)}.$$

For $v_3(n)$, we use the same approach, but directly use the above solution for $v_2(n)$ for each n.

$$\begin{aligned}
v_3(n) &= 1 + P_{32}v_2(n-1) + P_{33}v_3(n-1) \;=\; 1 + \frac{1}{2}\big(1 - 2^{-(n-2)}\big) + \frac{1}{4}v_3(n-1)\\
&= (2 - 2^{n-1}) + \frac{1}{4}\Big[2 - 2^{n-2} + \frac{1}{4}v_3(n-2)\Big] \;=\; \cdots\\
&= (2 - 2^{n-1}) + \frac{1}{4}\Big[2 - 2^{n-2}\Big] + \frac{1}{16}\Big[2 - 2^{n-3}\Big] + \cdots + \frac{1}{4^{n-1}}[2 - 2^0]\\
&= 2\Big[1 + \frac{1}{4} + \frac{1}{4^2} + \cdots \frac{1}{4^{n-1}}\Big] - \Big[2^{-(n-1)} + 2^{-n} + \cdots + 2^{-2n+2}\Big]\\
&= 2\left[\frac{1 - 4^{-n}}{1 - 4^{-1}}\right] - 2^{-(n-1)}\left[\frac{1 - 2^{-n}}{1 - 2^{-1}}\right] \;=\; \frac{8}{3} - \frac{2^{-2n+3}}{3} - 2^{-n+2} + 2^{-2n+2}.
\end{aligned}$$

Taking the limit,

$$\lim_{n\to\infty} v_2(n) = 2; \qquad \lim_{n\to\infty} v_3(n) = \frac{8}{3}.$$

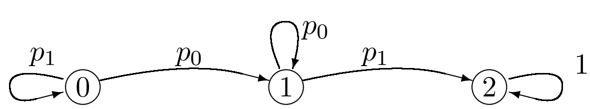
c) Give a direct argument why (4.32) provides the solution directly to the expected time from each state to enter the trapping state.

Solution: The limit as $n \to \infty$ gives the expected time to enter the trapping state with no limit on the required number of transitions. This limit exists in general since the transient states persist with probabilities decreasing exponentially with n.

Exercise 4.28: Consider finding the expected time until a given string appears in an 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 that string.

Solution:



Let v_i be the expected first-passage time from node i to node 2. Then

$$\begin{aligned}
v_0 &= 1 + p_1 v_0 + p_0 v_1; \qquad & v_1 = 1 + p_0 v_1 + p_1 v_2\\
v_0 &= 1/p_0 + v_1; & v_1 = 1/p_1 + v_2.
\end{aligned}$$

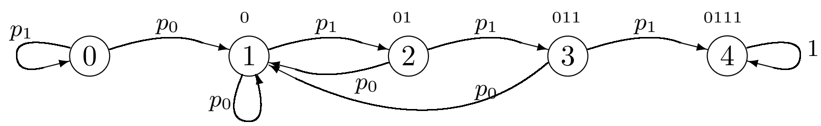
Combining these equations to eliminate v_1,

$$v_0 = 1/p_0 + 1/p_1 + v_2 = 1/p_0 p_1 + v_2.$$

Finally, the trapping state, 2, is reached when the string 01 first occurs, so $v_2 = 0$ and $v_0 = 1/p_0p_1$.

b) For parts b) to d), let $(a_1, a_2, a_3, \dots, a_k) = (0, 1, 1, \dots, 1)$, i.e., zero followed by $k-1$ ones. Draw the corresponding Markov chain for $k = 4$.

Solution:



c) Let v_i, $1 \le i \le k$ be the expected first-passage time from state i to state k. Note that $v_k = 0$. For each i, $1 \le i < k$, show that $v_i = \alpha_i + v_{i+1}$ and $v_0 = \beta_i + v_{i+1}$ where α_i and β_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 β_{i+1} as a function of β_i starting with $i = 1$ and using the equation $v_0 = 1/p_0 + v_1$.

Solution: (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 $i \ge 1$. Using the basic first-passage-time equation,

$$\begin{aligned} v_{i+1} &= 1 + p_0v_1 + p_1v_{i+2} \\ &= p_0v_0 + p_1v_{i+2} \\ &= p_0\beta_i + p_0v_{i+1} + p_1v_{i+2}. \end{aligned}$$

The second equality uses the basic equation for v_0, *i.e.*, $v_0 = 1 + p_1v_0 + p_0v_1$ which reduces to $p_0v_1 + 1 = p_0v_0$ and the third equality uses the inductive hypothesis $v_0 = \beta_i + v_{i+1}$. 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 this result plus the inductive hypothesis $v_0 = \beta_i + v_{i+1}$ to get

$$v_0 \;=\; \beta_i + v_{i+1} \;=\; \beta_i + \frac{p_0\beta_i}{p_1} + v_{i+2} \;=\; \frac{\beta_i}{p_1} + v_{i+2}.$$

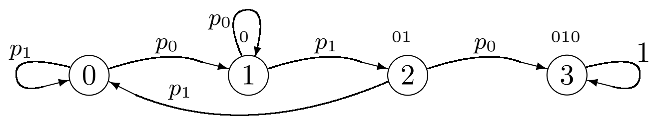
This completes the second half of the induction, showing that $\beta_{i+1} = \beta_i/p_1$. Iterating on these equations for β_i and α_i, we find the explicit expression

$$\alpha_i = \frac{1}{p_1^i}; \qquad \beta_i = \frac{1}{p_0p_1^i}.$$

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 view the occurrence of these strings as renewals in Exercise 5.35 and find a more intuitive way to derive the same answer.

d) Let $\boldsymbol{a} = (0,1,0)$. 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 (a). The basic equation for v_2 in terms of its outward transitions is

$$\begin{aligned} v_2 &= 1 + p_0 v_0 + p_1 v_3 \\ &= 1 + p_0\left[\frac{1}{p_0 p_1} + v_2\right]. \end{aligned}$$

Combining the terms in v_2, we get

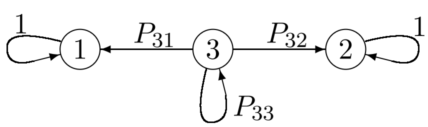
$$p_1 v_2 = 1 + 1/p_1.$$

Using $v_0 = 1/p_0 p_1 + v_2$,

$$v_0 = \frac{1}{p_1^2 p_0} + \frac{1}{p_1} + \frac{1}{p_1^2} = \frac{1}{p_1} + \frac{1}{p_0 p_1^2}.$$

This solution will become more transparent after doing Exercise 5.35.

Exercise 4.29: **a)** Find $\lim_{n\to\infty}[P^n]$ for the Markov chain below. Hint: Think in terms of the long term transition probabilities. Recall that the edges in the graph for a Markov chain correspond to the positive transition probabilities.



Solution: The chain has 2 recurrent states, each in its own class, and one transient state. Thus $\lim_{n\to\infty} P_{11}^n = 1$ and $\lim_{n\to\infty} P_{22}^n = 1$. Let $q_1 = \lim_{n\to\infty} P_{31}^n$ and $q_2 = \lim_{n\to\infty} P_{32}^n$. Since $q_1 + q_2 = 1$ and since in each transition starting in state 3, P_{31} and P_{32} give the probabilities of moving to 1 or 2, $q_1 = P_{31}/(P_{31}+P_{32})$ and $q_2 = P_{32}/(P_{31}+P_{32})$. The other long term transtion probabilities are zero, so

$$\lim_{n\to\infty}[P^n] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ q_1 & q_2 & 0 \end{bmatrix}.$$

The general case here, with an arbitrary set of transient states and an arbitrary number of recurrent classes is solved in Exercise 4.18.

b) Let $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ denote the first two rows of $\lim_{n\to\infty}[P^n]$ and let $\boldsymbol{\nu}^{(1)}$ and $\boldsymbol{\nu}^{(2)}$ denote the first two columns of $\lim_{n\to\infty}[P^n]$. Show that $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ are independent left eigenvectors of $[P]$, and that $\boldsymbol{\nu}^{(1)}$ and $\boldsymbol{\nu}^{(2)}$ are independent right eigenvectors of $[P]$. Find the eigenvalue for each eigenvector.

Solution: $\boldsymbol{\pi}^{(1)} = (1, 0, 0)$ and $\boldsymbol{\pi}^{(2)} = (0, 1, 0)$. Multiplying $\boldsymbol{\pi}^{(i)}$ by $\lim_{n\to\infty}[P^n]$ for $i = 1, 2$ we see that these are the left eigenvectors of eigenvalue 1 of the two recurrent classes. We also know this from Section 4.4. Similarly, $\boldsymbol{\nu}^{(1)} = (1, 0, q_1)^\mathsf{T}$ and $\boldsymbol{\nu}^{(2)} = (0, 1, q_2)^\mathsf{T}$ are right eigenvectors of eigenvalue 1, both of $\lim_{n\to\infty}[P^n]$ and also of $[P]$.

c) Let $\boldsymbol{r}$ be an arbitrary reward vector and consider the equation

$$\boldsymbol{w} + g^{(1)}\boldsymbol{\nu}^{(1)} + g^{(2)}\boldsymbol{\nu}^{(2)} = \boldsymbol{r} + [P]\boldsymbol{w}. \tag{A.18}$$

Determine what values $g^{(1)}$ and $g^{(2)}$ must have in order for (A.18) to have a solution. Argue that with the additional constraints $w_1 = w_2 = 0$, (A.18) has a unique solution for $\boldsymbol{w}$ and find that $\boldsymbol{w}$.

Solution: In order for this equation to have a solution, it is necessary, first, for a solution to exist when both sides are premultiplied by $\boldsymbol{\pi}^{(1)}$. This results in $g^{(1)} = r_1$. Similarly, premultiplying by $\boldsymbol{\pi}^{(2)}$ results in $g^{(2)} = r_2$. In other words, $g^{(1)}$ is the gain per transition when in state 1 or when the chain moves fom state 3 to state 1. Similarly $g^{(2)}$ is the gain per transition in state 2 or after a transition from 3 to 2. Since there are two recurrent sets of states, there is no common meaning to a gain per transition.

Setting $w_1 = w_2 = 0$ makes a certain amount of sense, since starting in state 1 or starting in state 2, the reward increases by r_1 or r_2 per transition with no initial transient. With this choice, the first two components of the vector equation in (A.18) are 0=0 and the third is

$$w_3 + r_3 q_1 + r_2 q_2 = r_3 + P_{33} w_3.$$

Solving for w_3,

$$w_3 = \frac{1}{P_{31} + P_{32}} \left[r_3 - r_1 \frac{P_{31}}{P_{31} + P_{32}} - r_2 \frac{P_{32}}{P_{31} + P_{32}} \right].$$

This can be interpreted as the relative gain of starting in state 3 relative to the 'average' of starting in 1 or 2. The interpretation is quite limited, since the gain per transition depends on which recurrent class is entered, and thus relative gain has nothing definitive for comparison.

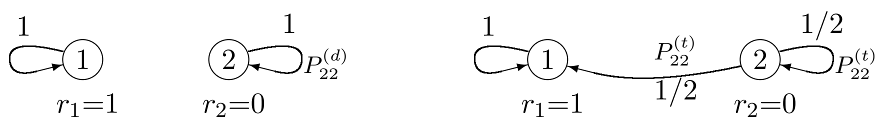
Exercise 4.32: George drives his car to the theater, which is at the end of a one-way street. There are parking places along the side of the street and a parking garage that costs \$5 at the theater. Each parking place is independently occupied or unoccupied with probability 1/2. If George parks n parking places away from the theater, it costs him n cents (in time and shoe leather) to walk the rest of the way. George is myopic and can only see the parking place he is currently passing. If George has not already parked by the time he reaches the nth place, he first decides whether or not he will park if the place is unoccupied, and then observes the place and acts according to his decision. George can never go back and must park in the parking garage if he has not parked before.

a) Model the above problem as a 2 state dynamic programming problem. In the "driving" state, state 2, there are two possible decisions: park if the current place is unoccupied or drive on whether or not the current place is unoccupied.

Solution: View the two states as walking (state 1) and driving (state 2). The state transitions correspond to passing successive possibly idle parking spaces, which are passed either by car or foot. Thus $r_1 = 1$ (the cost in shoe leather and time) and $r_2 = 0$. There is no choice of policy in state 1, but in state 2, policy t is to try to park and policy d is continue to drive.

First consider stage 1 where George is between the second and first parking space from the end. If George is in the walk state, there is one unit of cost going from 2 to 1 and 1 unit of final cost going from 1 to 0. If George is in the drive state (2) and uses the try policy, $r_2 = 0$ and with probability 1/2, he parks and has one unit final cost getting to the theatre. With probability 1/2, he doesn't park and the final cost is 500. For policy d, $r_2 = 0$ and with probability 1, the final cost is 500.

In summary, $\boldsymbol{u} = (1, 500)^\mathsf{T}$, $\boldsymbol{r} = (0, 1)^\mathsf{T}$ and the Markov chain with the t and d policies in state 2 is



b) Find $v_i^*(n, \boldsymbol{u})$, the *minimum* expected aggregate cost for n stages (i.e., immediately before observation of the nth parking place) starting in state $i = 1$ or 2; it is sufficient to express $v_i^*(n, \boldsymbol{u})$ in terms of $v_i^*(n-1)$. The final costs, in cents, at stage 0 should be $v_2(0) = 500$, $v_1(0) = 0$.

Solution: We start with stage 1.

$$\boldsymbol{v}^*(1, \boldsymbol{u}) = \boldsymbol{r} + \min_{\boldsymbol{k}} [P^{\boldsymbol{k}}]\boldsymbol{u} \;=\; (2,\, 250.5)^\mathsf{T},$$

where policy t is clearly optimal. At an arbitrary stage n, the cost is found iteratively from stage $n-1$ by (4.48), *i.e.*,

$$\boldsymbol{v}^*(n, \boldsymbol{u}) = \min_{\boldsymbol{k}} \left(\boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{v}^*(n-1, \boldsymbol{u}) \right). \tag{A.19}$$

c) For what values of n is the optimal decision the decision to drive on?

Solution: The straightforward approach (particularly with computional aids is to simply calculate (A.19). The cost drops sharply with increasing n and for $n \geq 8$, the optimal decision is to drive on, whereas for $n \leq 7$, the optimal decision is to park if possible.

As with many problems, the formalism of dymamic programming makes hand calculation and understanding more awkward than it need be. The simple approach, for any distance n away from the theatre, is to simply calculate the cost of parking if a place is available, namely n, and to calculate the expected cost if one drives on. This expected cost, E_n is easily seen to be

$$E_n = \frac{1}{2}(m-1) + \frac{1}{4}(m-2) + \cdots + \frac{1}{2^{m-1}}(1) + \frac{500}{2^{m-1}}.$$

With some patience, this is simplified to $E_n = m - 2 + (501)/2^{m-1}$, from which it is clear that one should drive for $n \geq 8$ and park for $n \leq 7$.

d) What is the probability that George will park in the garage, assuming that he follows the optimal policy?

Solution: George will park in the garage if the last 7 potential parking spaces are full, an event of probability 2^{-7}.

Exercise 4.33: (Proof of Corollary 4.6.9) **a)** Show that if two stationary policies $\boldsymbol{k}'$ and $\boldsymbol{k}$ have the same recurrent class $\mathcal{R}'$ and if $k_i' = k_i$ for all $i \in \mathcal{R}'$, then $w_i' = w_i$ for all $i \in \mathcal{R}'$. Hint: See the first part of the proof of Lemma 4.6.7.

Solution: For all $i \in \mathcal{R}'$, $P_{ij}^{(k_i)} = P_{ij}^{(k_i')}$. The steady-state vector $\boldsymbol{\pi}$ is determined solely by the transition probabilities in the recurrent class, so $\pi_i = \pi_i'$ for all $i \in \mathcal{R}'$. Since $r_i = r_i'$ for all $i \in \mathcal{R}'$, it also follows that $g = g'$. The equations for the components of the relative gain vector $\boldsymbol{w}$ for $i \in \mathcal{R}'$ are

$$w_i + g = r_i^{(k_i)} + \sum_j P_{ij}^{(k_i)} w_j; \qquad i \in \mathcal{R}.$$

These equations, along with $\boldsymbol{\pi w} = 0$ have a unique solution if we look at them only over the recurrent class $\mathcal{R}'$. Since all components are the same for $\boldsymbol{k}$ and $\boldsymbol{k}'$, there is a unique solution over $\mathcal{R}'$, *i.e.*, $w_i = w_i'$ for all $i \in \mathcal{R}'$.

b) Assume that $\boldsymbol{k}'$ satisfies 4.50 (*i.e.*, that it satisfies the termination condition of the policy improvement algorithm) and that $\boldsymbol{k}$ satisfies the conditions of (a). Show that (4.64) is satisfied for all states ℓ.

Solution: The implication from $\mathcal{R}'$ being *the* recurrent class of $\boldsymbol{k}$ and $\boldsymbol{k}'$ is that each of them are unichains. Thus $\boldsymbol{w} + g\boldsymbol{e} = \ \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}$ with $\boldsymbol{\pi w} = 0$ has a unique solution, and there is a unique solution for the primed case. Rewriting these equations for the primed and unprimed case, and recognizing from (a) that $g\boldsymbol{e} = g'\boldsymbol{e}$ and $\boldsymbol{\pi w} = \boldsymbol{\pi w}'$,

$$\boldsymbol{w} - \boldsymbol{w}' = [P^{\boldsymbol{k}}](\boldsymbol{w} - \boldsymbol{w}') + \left\{ \boldsymbol{r}^{\boldsymbol{k}} - \boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}}]\boldsymbol{w}' - [P^{\boldsymbol{k}'}]\boldsymbol{w}' \right\}. \tag{A.20}$$

This is the vector form of (4.64).

c) Show that $\boldsymbol{w} \le \boldsymbol{w}'$. Hint: Follow the reasoning at the end of the proof of Lemma 4.6.7.

Solution: The quantity in braces in (A.20) is non-negative because of the termination condition of the policy improvement algorithm. Also $\boldsymbol{w} = \boldsymbol{w}'$ over the recurrent components of $\boldsymbol{k}$. Viewing the term in braces as the non-negative difference between two reward vectors, Corollary 4.5.6 shows that $\boldsymbol{w} \le \boldsymbol{w}'$.

Exercise 4.35: Consider a Markov decision problem in which the stationary policies $\boldsymbol{k}$ and $\boldsymbol{k}'$ each satisfy (4.50) and each correspond to ergodic Markov chains.

a) Show that if $\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' \ge \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'$ is not satisfied with equality, then $g' > g$.

Solution: The solution is very similar to the proof of Lemma 4.6.5. Since $[P^{\boldsymbol{k}}]$ is ergodic, $\boldsymbol{\pi}^{\boldsymbol{k}}$ is strictly positive. Now $\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' \ge \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'$ must be satisfied because of (4.50), and if one or more components are not satisfied with equality, then

$$\boldsymbol{\pi}^{\boldsymbol{k}}\left(\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}'\right) \ > \ \boldsymbol{\pi}^{\boldsymbol{k}}\left(\boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'\right) \ = \ g + \boldsymbol{\pi}^{\boldsymbol{k}}\boldsymbol{w}', \tag{A.21}$$

where in the equality, we used the definition of g and the fact that $\boldsymbol{\pi}^{\boldsymbol{k}}$ is an eigenvector of $[P^{\boldsymbol{k}}]$. Since $\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' = \boldsymbol{w}' + g'\boldsymbol{e}$ from (4.37), (A.21) simplifies to

$$\boldsymbol{\pi}^{\boldsymbol{k}}\boldsymbol{w}' + g'\boldsymbol{\pi}^{\boldsymbol{k}}\boldsymbol{e} > g + \boldsymbol{\pi}^{\boldsymbol{k}}\boldsymbol{w}'.$$

After canceling $\boldsymbol{\pi}^{\boldsymbol{k}}\boldsymbol{w}'$ from each side, it is clear that $g' > g$.

b) Show that $\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' = \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'$ (Hint: use (a).

Solution: The ergodicity of $\boldsymbol{k}$ and $\boldsymbol{k}'$ assures the inherently reachable assumption of Theorem 4.6.8, and thus we know from the fact that $\boldsymbol{k}$ satisfies the termination condition of the policy improvement algorithm that $g \geq g'$. Thus $g' > g$ is impossible and none of the components of $\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' \geq \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'$ can be satisfied with inequality, *i.e.*,

$$\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' \geq \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'.$$

c) Find the relationship between the relative gain vector $\boldsymbol{w}$ for policy $\boldsymbol{k}$ and the relative-gain vector $\boldsymbol{w}'$ for policy $\boldsymbol{k}'$. (Hint: Show that $\boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}' = g\boldsymbol{e} + \boldsymbol{w}'$; what does this say about $\boldsymbol{w}$ and $\boldsymbol{w}'$?)

Solution: Using (4.37) and (b), we have

$$\boldsymbol{w}' + g'\boldsymbol{e} \;=\; \boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' \;=\; \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}'.$$

We also have

$$\boldsymbol{w} + g\boldsymbol{e} \;=\; \boldsymbol{r}^{\boldsymbol{k}} + [P^{\boldsymbol{k}}]\boldsymbol{w}.$$

Subtracting the second equation from the first, and remembering that $g' = g$, we get

$$\boldsymbol{w}' - \boldsymbol{w} = [P^{\boldsymbol{k}}](\boldsymbol{w}' - \boldsymbol{w}).$$

Thus $\boldsymbol{w}' - \boldsymbol{w}$ is a right eigenvector (of eigenvalue 1) of $[P^{\boldsymbol{k}}]$. This implies that $\boldsymbol{w}' - \boldsymbol{w} = \alpha\boldsymbol{e}$ for some $\alpha \in \mathbb{R}$. Note that this was seen in the special case of Exercise 4.34(d).

d) Suppose that policy $\boldsymbol{k}$ uses decision 1 in state 1 and policy $\boldsymbol{k}'$ uses decision 2 in state 1 (i.e., $k_1 = 1$ for policy $\boldsymbol{k}$ and $k_1 = 2$ for policy $\boldsymbol{k}'$). What is the relationship between $r_1^{(k)}, P_{11}^{(k)}, P_{12}^{(k)}, \dots P_{1J}^{(k)}$ for k equal to 1 and 2?

Solution: The first component of the equation $\boldsymbol{r}^{\boldsymbol{k}'} + [P^{\boldsymbol{k}'}]\boldsymbol{w}' = \boldsymbol{w}' + g\boldsymbol{e}$ is $r_1^{(2)} + \sum_{J=1}^{\mathsf{M}} P_{1j}^{(2)} w'_j = {w'}_1^{(1)} g$. From (c), $\boldsymbol{w}' = \boldsymbol{w} + \alpha\boldsymbol{e}$, so we have

$$r_1^{(2)} + \sum_{J=1}^{\mathsf{M}} P_{1j}^{(2)}(w_j + \alpha) = g(w_1 + \alpha).$$

Since α cancels out,

$$r_1^{(2)} + \sum_{J=1}^{\mathsf{M}} P_{1j}^{(2)} w_j = r_1^{(1)} + \sum_{J=1}^{\mathsf{M}} P_{1j}^{(1)} w_j.$$

e) Now suppose that policy $\boldsymbol{k}$ uses decision 1 in each state and policy $\boldsymbol{k}'$ uses decision 2 in each state. Is it possible that $r_i^{(1)} > r_i^{(2)}$ for all i? Explain carefully.

Solution: It seems a little surprising, since $\boldsymbol{\pi}^{\boldsymbol{k}}\boldsymbol{r}^{\boldsymbol{k}} = \boldsymbol{\pi}^{\boldsymbol{k}'}\boldsymbol{r}^{\boldsymbol{k}'}$, but Exercise 4.34 essentially provides an example. In that example, r_1 is fixed over both policies, but by providing a choice in state 1 like that provided in state 2, one can create a policy with $g = g'$ where $r_i^{(1)} > r_i^{(2)}$ for all i.

f) Now assume that $r_i^{(1)}$ is the same for all i. Does this change your answer to (e)? Explain.

Solution: If $r_i^{(1)}$ is the same, say r for all i, then $g = r$. The only way to achieve $g' = r$ for all i is for $r_i^{(2)} \geq r$ for at least one i. Thus we cannot have $r_i^{(1)} > r_i^{(2)}$ for all i.

A.5 Solutions for Chapter 5

Exercise 5.1: The purpose of this exercise is to show that for an arbitrary renewal process, $N(t)$, the number of renewals in $(0, t]$ is a (non-defective) random variable.

a) Let $X_1, X_2, \ldots$ be a sequence of IID inter-renewal rv's. Let $S_n = X_1 + \cdots + X_n$ be the corresponding renewal epochs for each $n \geq 1$. Assume that each X_i has a finite expectation $\overline{X} > 0$ and, for any given $t > 0$, use the weak law of large numbers to show that $\lim_{n\to\infty} \Pr\{S_n \leq t\} = 0$.

Solution: From the WLLN, (1.75),

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| > \epsilon\right\} = 0 \qquad \text{for every } \epsilon > 0.$$

Choosing $\epsilon = \overline{X}/2$, say, and looking only at the lower limit above, $\lim_{n\to\infty} \Pr\{S_n < n\overline{X}/2\} = 0$. For any given t and all large enough n, $t < n\overline{X}/2$, so $\lim_{n\to\infty} \Pr\{S_n \leq t\} = 0$.

b) Use (a) to show that $\lim_{n\to\infty} \Pr\{N(t) \geq n\} = 0$ for each $t > 0$ and explain why this means that $N(t)$ is a rv, *i.e.*, is not defective.

Solution: Since $\{S_n \leq t\} = \{N(t) \geq n\}$, we see that $\lim_{n\to\infty} \Pr\{N(t) \geq n\} = 0$ for each $t > 0$. Since $N(t)$ is nonnegative, this shows that it is a rv.

c) Now suppose that the X_i do not have a finite mean. Consider truncating each X_i to $\breve{X}_i$, where for any given $b > 0$, $\breve{X}_i = \min(X_i, b)$. Let $\breve{N}(t)$ be the renewal counting process for the inter-renewal intervals $\breve{X}_i$. Show that $\breve{N}(t)$ is non-defective for each $t > 0$. Show that $N(t) \leq \breve{N}(t)$ and thus that $N(t)$ is non-defective. Note: Large inter-renewal intervals create small values of $N(t)$, and thus $\mathsf{E}[X] = \infty$ has nothing to do with potentially large values of $N(t)$, so the argument here was purely technical.

Solution: Since $\Pr\left\{\breve{X} > b\right\} = 0$, we know that $\breve{X}$ has a finite mean, and consequently, from (b), $\breve{N}(t)$ is a rv for each $t > 0$. Since $\breve{X}_n \leq X_n$ for each n, we also have $\breve{S}_n \leq S_n$ for all $n \geq 1$. Thus if $\breve{S}_n > t$, we also have $S_n > t$. Consequently, if $\breve{N}(t) < n$, we also have $N(t) < n$. It follows that $\Pr\{N(t) \geq n\} \leq \Pr\left\{\breve{N}(t) \geq n\right\}$, so $N(t)$ is also a rv.

Exercise 5.2: This exercise shows 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 $\mathsf{F}_X(x)$, with, as usual, $\mathsf{F}_X(0) = 0$. Using whatever combination of mathematics and common sense is comfortable for you, show that for any ϵ, $0 < \epsilon < 1$, there is a $\delta > 0$ such that $\mathsf{F}_X(\delta) \leq 1 - \epsilon$. 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 δ 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 $X_i \leq \delta$ for each i, $1 \leq i \leq n$. Using the ϵ and δ of (a), $\Pr\{X_i \leq \delta\} \leq (1-\epsilon)$ for $1 \leq i \leq n$. Since the X_i are independent, we then have $\Pr\{S_n \leq \delta\} \leq (1-\epsilon)^n$.

c) Show that $\mathsf{E}[N(\delta)] \leq 1/\epsilon$.

Solution: Since $N(t)$ is nonnegative and integer,

$$\begin{aligned}
\mathsf{E}[N(\delta)] &= \sum_{n=1}^{\infty} \Pr\{N(\delta) \geq n\} \\
&= \sum_{n=1}^{\infty} \Pr\{S_n \leq \delta\} \\
&\leq \sum_{n=1}^{\infty} (1-\epsilon)^n \\
&= \frac{1-\epsilon}{1-(1-\epsilon)} = \frac{1-\epsilon}{\epsilon} \leq \frac{1}{\epsilon}.
\end{aligned}$$

d) For the ϵ, δ of (a), show that for every integer k, $\mathsf{E}[N(k\delta)] \leq k/\epsilon$ and thus that $\mathsf{E}[N(t)] \leq \frac{t+\delta}{\epsilon\delta}$ for any $t > 0$.

Solution: The solution of (c) suggests breaking the interval $(0, k\delta]$ into k intervals each of size δ. Letting $\widetilde{N}_i = N(i\delta) - N((i-1)\delta)$ be the number of arrivals in the ith of these intervals, we have $\mathsf{E}[N(\delta k)] = \sum_{i=1}^{k} \mathsf{E}\left[\widetilde{N}_i\right]$.

For the first of these intervals, we have shown that $\mathsf{E}\left[\widetilde{N}_1\right] \leq 1/\epsilon$, but that argument does not quite work for the subsequent intervals, since the first arrival in an interval $i > 1$ might correspond to an interarrival interval that starts early in the interval $i-1$ and ends late in interval i. Subsequent arrivals in interval i must correspond to interarrival intervals that both start and end in interval i and thus have duration at most δ. Thus if we let $S_n^{(i)}$ be the number of arrivals in the ith interval, we have

$$\Pr\left\{S_n^{(i)} \leq \delta\right\} \leq (1-\epsilon)^{n-1}.$$

Repeating the argument in (c) for $i > 1$, then,

$$\begin{aligned}
\mathsf{E}\left[\widetilde{N}_i\right] &= \sum_{n=1}^{\infty} \Pr\left\{\widetilde{N}_i \geq n\right\} = \sum_{n=1}^{\infty} \Pr\left\{S_n^{(i)} \leq \delta\right\} \\
&\leq \sum_{n=1}^{\infty} (1-\epsilon)^{n-1} = \frac{1}{1-(1-\epsilon)} = \frac{1}{\epsilon}.
\end{aligned}$$

Since $\mathsf{E}[N(\delta k)] = \sum_{i=1}^{k} \mathsf{E}\left[\widetilde{N}_i\right]$, we then have

$$\mathsf{E}[N(k\delta)] \le k/\epsilon.$$

Since $N(t)$ is non-decreasing in t, $N(t)$ can be upper bounded by replacing t by the smallest integer multiple of $1/\delta$ that is t or greater, *i.e.*,

$$\mathsf{E}[N(t)] \;\le\; \mathsf{E}[N(\delta\lceil t/\delta\rceil)] \;\le\; \frac{\lceil t/\delta\rceil}{\epsilon} \;\le\; \frac{(t/\delta)+1}{\epsilon}.$$

e) Use the result here to show that $N(t)$ is non-defective.

Solution: Since $N(t)$, for each t, is nonnegative and has finite expectation, it is obviously non-defective. One way to see this is that $\mathsf{E}[N(t)]$ is the integral of the complementary CDF, $\mathsf{F}^{\mathsf{c}}_{N(t)}(n)$ of $N(t)$. Since this integral is finite, $\mathsf{F}^{\mathsf{c}}_{N(t)}(n)$ 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) $N(t) \le n$ if and only if $S_n \ge t$?

c) $N(t) > n$ if and only if $S_n < t$?

Solution: Part (a) is true, as pointed out in (5.1) and more fully explained in (2.2) and (2.3).

Parts b) and c) are false, as seen by any situation where $S_n < t$ and $S_{n+1} > t$. In these cases, $N(t) = n$. In other words, one must be careful about strict versus non-strict inequalities when using $\{S_n \le t\} = \{N(t) \ge n\}$

Exercise 5.5: (This shows that convergence WP1 implies convergence in probability.) Let $\{Y_n;\, n \ge 1\}$ be a sequence of rv's that converges to 0 WP1. For any positive integers m and k, let

$$A(m,k) = \{\omega : |Y_n(\omega)| \le 1/k \quad \text{for all } n \ge m\}.$$

a) Show that if $\lim_{n\to\infty} Y_n(\omega) = 0$ for some given ω, then (for any given k) $\omega \in A(m,k)$ for some positive integer m.

Solution: Note that for a given ω, $\{Y_n(\omega);\, n \ge 1\}$ is simply a sequence of real numbers. The definition of convergence of such a sequence to 0 says that for any ϵ (or any $1/k$ where $k > 0$ is an integer), there must be an m large enough that $Y_n(\omega) \le 1/k$ for all $n \ge m$. In other words, the given ω is contained in $A(m,k)$ for that m.

b) Show that for all $k \ge 1$

$$\Pr\left\{\bigcup_{m=1}^{\infty} A(m,k)\right\} = 1.$$

Solution: The set of ω for which $\lim_{n\to\infty} Y_n(\omega) = 0$ has probability 1. For any such ω and any given k, part (a) showed that $\omega \in A(m,k)$ for some integer $m > 0$. Thus each such ω lies in the above union, implying that the union has probability 1.

c) Show that, for all $m \ge 1$, $A(m,k) \subseteq A(m{+}1, k)$. Use this (plus (1.9)) to show that

$$\lim_{m\to\infty} \Pr\{A(m,k)\} = 1.$$

Solution: Note that if $|Y_n(\omega)| \le 1/k$ for all $n \ge m$, then also $|Y_n(\omega)| \le 1/k$ for all $n \ge m+1$. This means that $A(m,k) \subseteq A(m{+}1,\, k)$. From (1.9) then

$$1 = \Pr\Big\{\bigcup_m A(m,k)\Big\} = \lim_{m\to\infty} \Pr\{A(m,k)\}\,.$$

d) Show that if $\omega \in A(m,k)$, then $|Y_m(\omega)| \le 1/k$. Use this (plus (c)) to show that

$$\lim_{m\to\infty} \Pr\{|Y_m| > 1/k\} = 0.$$

Since $k \ge 1$ is arbitrary, this shows that $\{Y_n;\, n \ge 1\}$ converges in probabiity.

Solution: Note that if $|Y_n(\omega)| \le 1/k$ for all $n \ge m$, then certainly $|Y_n(\omega)| \le 1/k$ for $n = m$. It then follows from (c) that $\lim_{m\to\infty} \Pr\{|Y_m| \le 1/k\} = 1$, which is equivalent to the desired statement. This shows that $\{Y_n;\, n \ge 1\}$ converges in probability.

Exercise 5.7: In this exercise, you will find an explicit expression for $\{\omega : \lim_n Y_n(\omega) = 0\}$. You need not be mathematically precise.

a) Let $\{Y_n;\, n \ge 1\}$ be a sequence of rv's. Using the definition of convergence for a sequence of numbers, justify the following set equivalences:

$$\begin{aligned}
\{\omega : \lim_n Y_n(\omega) = 0\} &= \bigcap_{k=1}^{\infty} \{\omega : \text{there exists an } m \text{ such that } |Y_n(\omega)| \le 1/k \text{ for all } n \ge m\} && \text{(A.22)}\\
&= \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \{\omega : Y_n(\omega) \le 1/k \text{ for all } n \ge m\} && \text{(A.23)}\\
&= \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega : Y_n(\omega) \le 1/k\}. && \text{(A.24)}
\end{aligned}$$

Solution: For any given sample point ω, $\{Y_n(\omega);\, n \ge 1\}$ is a sample sequence of the random sequence $\{Y_n;\, n \ge 1\}$ and is simply a sequence of real numbers. That sequence converges to 0 if for every integer $k > 0$, there is an $m(k)$ such that $Y_n(\omega) \le 1/k$ for $n \ge k$. The set of ω that satisfies this test is given by (A.22). A set theoretic way to express the existence of an $m(k)$ is to take the union over $m \ge 1$, giving us (A.23). Finally, the set theoretic way to express 'for all $n \ge m$' is to take the intersection over $n \ge m$, giving us (A.24).

b) Explain how this shows that $\{\omega : \lim_n Y_n(\omega) = 0\}$ must be an event.

Solution: Since Y_n is a rv, $\{\omega : Y_n(\omega) \le 1/k\}$ is an event for all n, k. The countable intersection over $n \ge m$ is then an event (by the axioms of probability), the countable union of these events is an event for each k, and the final intersection is an event.

c) Use De Morgan's laws to show that the complement of the above equivalence is

$$\{\omega : \lim_n Y_n(\omega) = 0\}^{\mathsf{c}} = \bigcup_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega : Y_n(\omega) > 1/k\}. \qquad \text{(A.25)}$$

Solution: For a sequence of sets $A_1, A_2, \dots$, de Morgan's laws say that $\bigcup_n A_n^{\mathsf{c}} = \{\bigcap_n A_n\}^{\mathsf{c}}$ and also that $\bigcap_n A_n^{\mathsf{c}} = \{\bigcup_n A_n\}^{\mathsf{c}}$. Applying the second form of de Morgan's laws to the right side of the complement of each side of (A.24), we get

$$\begin{aligned}
\{\omega : \lim_n Y_n(\omega) = 0\}^{\mathsf{c}} &= \bigcup_{k=1}^{\infty} \Big\{ \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega : Y_n(\omega) \le 1/k\} \Big\}^{\mathsf{c}}\\
&= \bigcup_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \Big\{ \bigcap_{n=m}^{\infty} \{\omega : Y_n(\omega) \le 1/k\} \Big\}^{\mathsf{c}}\\
&= \bigcup_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \Big\{\omega : Y_n(\omega) \le 1/k\Big\}^{\mathsf{c}}\\
&= \bigcup_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \Big\{\omega : Y_n(\omega) > 1/k\Big\}
\end{aligned}$$

In the second line, the first form of de Morgan's laws were applied to the complemented term in braces on the first line. The third line applied the second form to the complemented term in the second line.

d) Show that for $\{Y_n;\, n \geq 1\}$ to converge to zero WP1, it is necessary and sufficient to satisfy

$$\Pr\Big\{\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{Y_n > 1/k\}\Big\} = 0 \qquad \text{for all } k \geq 1. \tag{A.26}$$

Solution: Applying the union bound to (A.25),

$$\Pr\Big\{\{\omega : \lim_n Y_n(\omega) = 0\}^{\mathsf{c}}\Big\} \;\leq\; \sum_{k=1}^{\infty}\Pr\Big\{\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{Y_n > 1/k\}\Big\}.$$

If (A.26) is satisfied for all $k \geq 1$, then the above sum is 0 and $\Pr\{\{\omega : \lim_n Y_n(\omega) = 0\}^{\mathsf{c}}\} = 0$. This means that $\Pr\{\lim_{n\to\infty} A_n = 0\} = 1$, *i.e.*, that $\lim_{n\to\infty} A_n = 0$ WP1.

If (A.26) is not satisfied for some given k, on the other hand, then the probability of the event $\{\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{Y_n > 1/k\}\}$ must be positive, say $\epsilon > 0$, so, using (A.25) and lower bounding the probabiity of the union over k by the single k above,

$$\begin{aligned}\Pr\Big\{\{\omega : \lim_n Y_n(\omega) = 0\}^{\mathsf{c}}\Big\} &= \Pr\Big\{\bigcup_{k=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\omega : Y_n(\omega) > 1/k\}\Big\} \\ &\geq \Pr\Big\{\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\omega : Y_n(\omega) > 1/k\}\Big\} \;\geq\; \epsilon.\end{aligned}$$

Thus there is a set of ω, whose probability is at least ϵ, for which $\lim_n Y_n(\omega)$ either does not exist or does not equal 0.

e) Show that for $\{Y_n;\, n \geq 1\}$ to converge WP1, it is necessary and sufficient to satisfy

$$\lim_{m\to\infty}\Pr\Big\{\bigcup_{n=m}^{\infty}\{Y_n > 1/k\}\Big\} = 0 \qquad \text{for all } k \geq 1.$$

Hint: Use (a) of Exercise 5.8. Note: (e) provides an equivalent condition that is often useful in establishing convergence WP1. It also brings out quite clearly the difference between convergence WP1 and convergence in probability.

Solution: As in Exercise 5.8, let $B_m = \bigcup_{n\geq m}\{Y_n > 1/k\}$. Then the condition in (e) is $\lim_{m\to\infty}\Pr\{B_m\} = 0$ and the necessary and sufficient condition established in (d) is $\Pr\Big\{\bigcap_{m\geq 1} B_m\Big\} = 0$. The equivalence of these conditions is implied by (1.10).

The solution to this exercise given here is mathematically precise and has the added benefit of showing that the various sets used in establishing the SLLN are in fact events since they are expressed as countable intersections and unions of events.

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 $\mathsf{E}[X_i] = \infty$. Let $b > 0$ be an arbitrary number and $\breve{X}_i$ be a truncated random variable defined by $\breve{X}_i = X_i$ if $X_i \leq b$ and $\breve{X}_i = b$ otherwise.

a) Show that for any constant $M > 0$, there is a b sufficiently large so that $\mathsf{E}\big[\breve{X}_i\big] \geq M$.

Solution: Since $\mathsf{E}[X] = \int_0^{\infty}\mathsf{F}^{\mathsf{c}}_X(x)\,dx = \infty$, we know from the definition of an integral over an infinite limit that

$$\mathsf{E}[X] = \lim_{b\to\infty}\int_0^b \mathsf{F}^{\mathsf{c}}_X(x)\,dx = \infty.$$

For $\breve{X} = \min(X, b)$, we see that $\mathsf{F}_{\breve{X}}(x) = \mathsf{F}_X(x)$ for $x < b$ and $\mathsf{F}_{\breve{X}}(x) = 1$ for $x \geq b$. Thus $\mathsf{E}\left[\breve{X}\right] = \int_0^b \mathsf{F}^{\mathsf{c}}_X(x)\,dx$. We have just seen that the limit of this as $b \to \infty$ is ∞, so that for any $M > 0$, there is a b sufficiently large that $\mathsf{E}\left[\breve{X}\right] \geq M$.

b) Let $\{\breve{N}(t);\, t{\geq}0\}$ be the renewal counting process with inter-renewal intervals $\{\breve{X}_i; i \geq 1\}$ and show that for all $t > 0$, $\breve{N}(t) \geq N(t)$.

Solution: Note that $X - \breve{X}$ is a non-negative rv, *i.e.*, it is 0 for $X \leq b$ and greater than b otherwise. Thus $\breve{X} \leq X$. It follows then that for all $n \geq 1$,

$$\breve{S}_n = \breve{X}_1 + \breve{X}_2 + \cdots \breve{X}_n \leq X_1 + X_2 + \cdots X_n = S_n.$$

Since $\breve{S}_n \leq S_n$, it follows for all $t > 0$ that if $S_n \leq t$ then also $\breve{S}_n \leq t$. This then means that if $N(t) \geq n$, then also $\breve{N}(t) \geq n$. Since this is true for all n, $\breve{N}(t) \geq N(t)$, *i.e.*, the number of renewals after truncation is greater than or equal to the number before truncation.

c) Show that for all sample functions $N(t, \omega)$, except a set of probability 0, $N(t, \omega)/t \leq 2/M$ 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 $\mathsf{E}\left[\breve{X}\right] \geq M$ be fixed in what follows. Since $\breve{X} \leq b$, we see that $\mathsf{E}\left[\breve{X}\right] < \infty$, so we can apply Theorem 5.3.1, which asserts that

$$\Pr\left\{\omega : \lim_{t\to\infty} \frac{\breve{N}(t,\omega)}{t} = \frac{1}{\mathsf{E}\left[\breve{X}\right]}\right\} \;=\; 1.$$

Let $A(M)$ denote the set of sample points for which the above limit exists, *i.e.*, for which $\lim_{t\to\infty} \breve{N}(t,\omega)/t = 1/\mathsf{E}\left[\breve{X}\right]$; Thus $A(M)$ has probability 1 from Theorem 5.3.1. We will show that, for each $\omega \in A(M)$, $\lim_t N(t,\omega)/t \leq 1/(2M)$. We know that for every $\omega \in A(M)$, $\lim_t \breve{N}(t,\omega)/t = 1/\mathsf{E}\left[\breve{X}\right]$. The definition of the limit of a real-valued function states that for any $\epsilon > 0$, there is a $\tau(\epsilon)$ such that

$$\left|\frac{\breve{N}(t,\omega)}{t} - \frac{1}{\mathsf{E}\left[\breve{X}\right]}\right| \;\leq \epsilon \qquad \text{for all}\, t \geq \tau(\epsilon).$$

Note that $\tau(\epsilon)$ depends on b and ω as well as ϵ, so we denote it as $\tau(\epsilon, b, \omega)$. Using only one side of this inequality, $N(t,\omega)/t \leq \epsilon + 1/\mathsf{E}\left[\breve{X}\right]$ for all $t \geq \tau(\epsilon, b, \omega)$. Since we have seen that $N(t,\omega) \leq \breve{N}(t,\omega)$ and $\mathsf{E}\left[\breve{X}\right] \geq M$, we have

$$\frac{N(t,\omega)}{t} \leq \epsilon + \frac{1}{M} \qquad \text{for all}\, t \geq \tau(\epsilon, b, \omega).$$

Since ϵ is arbitrary, we can choose it as $1/M$, so $N(t,\omega)/t \leq 1/(2M)$ for all sufficiently large t for each $\omega \in A(M)$. Now consider the intersection $\bigcap_M A(M)$ over integer $M \geq 1$. Since each

$A(M)$ has probability 1, the intersection has probability 1 also. For all ω in this intersection, $\lim_{n\to\infty} \Pr\{N(t)/t\} \le 1/(2M)$ for all integer $M \ge 1$, so $\lim_{n\to\infty} \Pr\{N(t)/t\} = 0$ WP1.

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 λ. 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 $\mathsf{F}_Y(y)$.

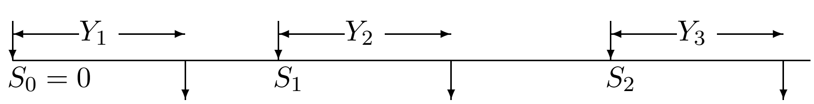
Successive service times (for those customers that are served) are IID and independent of arrival times. Assume that customer number 0 arrives and enters service at time $t = 0$.

a) Show that the sequence of times $S_1, S_2, \dots$ 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 ith service time; find the probability density of U_i.

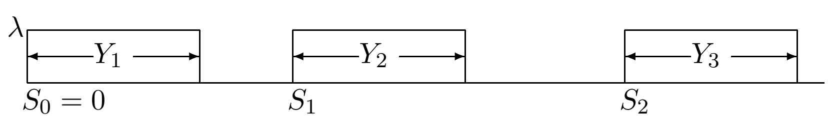
Solution: Let Y_1 be the first service time, *i.e.*, the time spent serving customer 0. Customers who arrive during $(0, Y_1]$ are lost, and, given that $Y_1 = y$, the residual time until the next customer arrives is memoryless and exponential with rate λ. Thus the time $X_1 = S_1$ at which the next customer enters service is $Y_1 + U_1$ where U_1 is exponential with rate λ, *i.e.*, $\mathsf{f}_{U_1}(u) = \lambda \exp(-\lambda u)$.

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 λ 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 $X_1, X_2, \dots$ are IID intervals between successive services. Thus $\{X_i;\, i \ge 1\}$ is a sequence of inter-renewal intervals for a renewal process and $S_1, S_2, \dots$ 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.



Solution: Customers are turned away at rate λ during the service times, so if we let $R(t)$ be the rate at which customers are turned away for a given sample path of services and arrivals, we have $R(t) = \lambda$ for t in a service interval and $R(t) = 0$ otherwise.



Note that the number of arrivals within a service interval is dependent on the length of the service interval but independent of arrivals outside of that interval and independent of other service intervals.

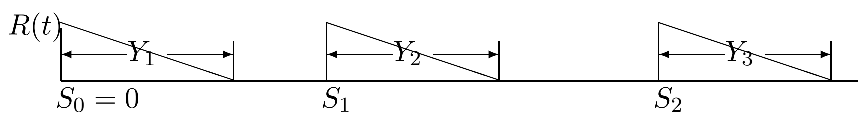
To be more systematic, we would define $\{Z_m;\, m \ge 1\}$ as the sequence of customer inter-arrival intervals. This sequence along with the service sequence $\{Y_n;\, n \ge 1\}$ specifies the sequences $\{S_n, Y_n;\, n \geq 1\}$. The discarded customer arrivals within each service interval is then Poisson of rate λ. The function $\{R(t);\, t > 0\}$ above would then better be described as $\mathsf{E}\left[R(t) \mid \{S_n, Y_n;\, n \geq 1\}\right]$ where $R(t)$ would then become a unit impulse at each discarded arrival.

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 nth inter-renewal interval (averaged over the discarded arrivals, but conditional on Y_n) is $R_n = \lambda Y_n$. Assuming that $\mathsf{E}[Y_n] < \infty$, Theorem 5.4.5 asserts that average reward, WP1, is $\frac{\lambda \mathsf{E}[Y]}{\mathsf{E}[Y]+1/\lambda}$. Thus the theorem asserts that the limit exists with probability 1. If $\mathsf{E}[Y_n] = \infty$, one could use a truncation argument to show that the average reward, WP1, is λ.

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



The reward over the nth inter-renewal interval is then $\lambda Y_n^2/2$ so the sample path average of the expected reward per unit time is

$$\frac{\mathsf{E}[R(t)]}{\overline{X}} = \frac{\lambda \mathsf{E}\left[Y^2\right]}{2(\overline{Y} + 1/\lambda)}.$$

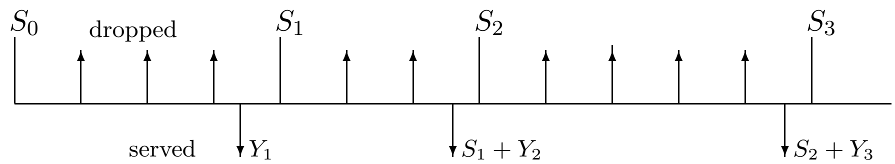
e) Now assume that the arrivals are deterministic, with the first arrival at time 0 and the nth arrival at time $n-1$. Does the sequence of times $S_1, S_2, \ldots$ at which subsequent customers start service still constitute the renewal times of a renewal process? Draw a sketch of arrivals, departures, 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 customer to be served at the completion of Y_1 is the customer arriving at $\lceil Y_1 \rceil$ (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 $\lceil Y_1 \rceil - 1$ are then dropped as illustrated below.

We see that the interval between the first and second service is $\lceil Y_2 \rceil$ and in general between service $n-1$ and n is $\lceil Y_n \rceil$. These intervals are IID and thus form a renewal process.

Finding the sample path average as before,

$$\frac{\mathsf{E}[R(t)]}{\overline{X}} = \frac{\mathsf{E}[\lceil Y \rceil - 1]}{\mathsf{E}[\lceil Y \rceil]}.$$



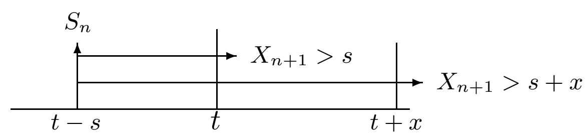
Exercise 5.13: Let $Z(t) = t - S_{N(t)}$ be the age of a renewal process and $Y(t) = S_{N(t)+1} - t$ be the residual life. Let $\mathsf{F}_X(x)$ be the CDF of the inter-renewal interval and find the following as a function of $\mathsf{F}_X(x)$:

a) $\Pr\{Y(t){>}x \mid Z(t){=}s\}$

Solution: First assume that X is discrete, and thus S_n is also discrete for each n. We first find $\Pr\{Z(t) = s\}$ for any given s, $0 < s \le t$. Since $\{Z(t) = s\} = \{S_{N(t)} = t - s\}$, we have

$$\begin{aligned}
\Pr\{Z(t){=}s\} &= \Pr\{S_{N(t)} = t{-}s\} = \sum_{n=0}^{\infty} \Pr\{N(t) = n,\, S_n = t{-}s\} \\
&= \sum_{n=0}^{\infty} \Pr\{S_n = t{-}s,\, X_{n+1} > s\} = \mathsf{F}^{\mathsf{c}}_X(s) \sum_{n=0}^{\infty} \Pr\{S_n = t{-}s\} \qquad \text{(A.27)}
\end{aligned}$$

where in the next to last step we noted that if the nth arrival comes at $t - s$ and $N(t) = n$, then no arrivals occur in $(t - s,\, t]$, so $X_{n+1} > t{-}s$. The last step used the fact that X_{n+1} is IID over n and statistically independent of S_n. The figure below illustrates these relationships and also illustrates how to find $\Pr\{Y(t){>}x,\ Z(t){=}s\}$.



$$\begin{aligned}
\Pr\{Y(t){>}x,\ Z(t){=}s\} &= \sum_{n=0}^{\infty} \Pr\{N(t){=}n,\ S_n{=}t{-}s,\ S_{n+1} > t + x\} \\
&= \sum_{n=0}^{\infty} \Pr\{N(t){=}n,\ S_n{=}t{-}s,\ X_{n+1} > s + x\} \\
&= \mathsf{F}^{\mathsf{c}}_X(s + x) \sum_{n=0}^{\infty} \Pr\{N(t){=}n,\ S_n{=}t{-}s\} \qquad \text{(A.28)}
\end{aligned}$$

We can now find $\Pr\{Y(t) > x \mid Z(t) = s\}$ by dividing (A.28) by (A.27) and cancelling out the common summation. Thus

$$\Pr\{Y(t) > x \mid Z(t) = s\} = \frac{\mathsf{F}^{\mathsf{c}}_X(s + x)}{\mathsf{F}^{\mathsf{c}}_X(s)}$$

We can see the intuitive reason for this answer from the figure above: $Z(t) = s$ specifies that there is an n such that $S_n = t{-}s$ and there are no arrivals in $(t{-}s, t]$; $Y(t) > x$ extends this interval of no arrivals to $(t - s, t + x]$, and these probabilities do not depend on n.

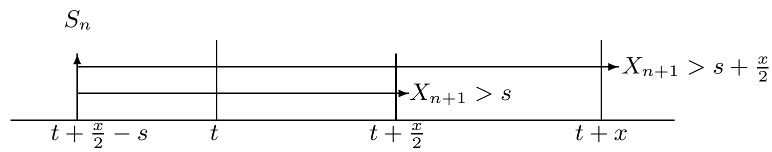
The argument above assumed that X is discrete, but this is a technicality that could be handled by a limiting argument. In this case, however, it is cleaner to modify the above argument by working directly with the conditional probabilities.

$$\begin{aligned}\Pr\{Y(t) > x \mid Z(t) = s\} &= \sum_{n=0}^{\infty} \Pr\{N(t){=}n\} \Pr\big\{Y(t) > x \mid S_{N(t)}{=}t{-}s,\ N(t){=}n\big\} \\ &= \sum_{n=0}^{\infty} \Pr\{N(t){=}n\} \Pr\{X_{n+1} > t{+}s \mid S_n{=}t{-}s,\ X_{n+1}{>}s\} \\ &= \frac{F^{\mathsf{c}}_X(s{+}x)}{F^{\mathsf{c}}_X(s)} \sum_{n=0}^{\infty} \Pr\{N(t){=}n\} \;=\; \frac{F^{\mathsf{c}}_X(s{+}x)}{F^{\mathsf{c}}_X(s)},\end{aligned}$$

where the second line again uses the fact that $\{S_n{=}t{-}s,\ N(t){=}n\} = \{S_n{=}t{-}s,\ X_{n+1}{>}s\}$.

b) $\Pr\{Y(t){>}x \,|\, Z(t{+}x/2){=}s\}$.

Solution: The conditioning event $\{Z(t + x/2) = s\}$ means that there is an arrival at $t + x/2 - s$ and no further arrival until after $t + x/2$. We must look at two cases, first where the arrival at $t + x/2 - s$ comes at or before t (*i.e.*, $s \geq x/2$) and second where it comes after t. The solution where $s \geq x/2$ is quite similar to (a), and we repeat the argument there for general X modifying the terms in (a) as needed. The diagram below will clarify the various terms, assuming $s \geq x/2$.



As in (a),

$$\Pr\Big\{Y(t){>}x \mid Z(t{+}\frac{x}{2}){=}s\Big\} = \sum_{n=0}^{\infty} \Pr\Big\{N(t{+}\frac{x}{2}){=}n\Big\} \Pr\Big\{Y(t){>}x \mid Z(t{+}\frac{x}{2}){=}s,\ N(t{+}\frac{x}{2}){=}n\Big\}.$$

The conditioning event above is

$$\Big\{Z\big(t + \frac{x}{2}\big) = s,\ N(t{+}\frac{x}{2}) = n\Big\} \;=\; \Big\{S_n = t + \frac{x}{2} - s,\ X_{n+1} > s\Big\}. \tag{A.29}$$

Given this conditioning event, the diagram shows that $Y(t) > x$ implies that $X_{n+1} > s{+}\frac{x}{2}$. Thus

$$\begin{aligned}\Pr\Big\{Y(t){>}x \mid Z(t{+}\frac{x}{2}){=}s\Big\} &= \sum_{n=0}^{\infty} \Pr\Big\{N(t{+}\frac{x}{2}){=}n\Big\} \Pr\Big\{X_{n+1}{>}s{+}\frac{x}{2} \mid S_n = t + \frac{x}{2} - s,\ X_{n+1} > s\Big\} \\ &= \frac{\mathsf{F}^{\mathsf{c}}_X(s{+}\frac{x}{2})}{\mathsf{F}^{\mathsf{c}}_X(s)} \qquad \text{for } s \geq x/2.\end{aligned} \tag{A.30}$$

For the other case, where $s < x/2$, the condition asserts that $S_{N(t+\frac{x}{2})} > t$. Thus there must be an arrival in the interval from t to $t{+}\frac{x}{2}$, so $\Pr\big\{Y(t) > x \mid Z(t{+}\frac{x}{2}){=}s\big\} = 0$.

c) $\Pr\{Y(t){>}x \,|\, Z(t{+}x){>}s\}$ for a Poisson process.

Solution: The event $\{Z(t+x) > s\}$ is the event that there are no arrivals in the interval $(t{+}x{-}s,\ t{+}x]$. We consider two cases separately, first $s \geq x$ and next $s < x$. If $s \geq x$, then there are no arrivals from t to $t+x$, so that $Y(t)$ must exceed x. Thus

$$\Pr\{Y(t) > x \mid Z(t+x) > s\} = 1; \qquad \text{for } s \geq x.$$

Alternatively, if $s < x$, any arrival between t and $t+x$ must arrive between t and $t+x-s$, so the probability of no arrival between t and $t+x$ given no arrival between $t+x-s$ and $t+x$ is the probability of no arrival between t and $t+x-s$.

$$\Pr\{Y(t) > x \mid Z(t+x) > s\} = \exp\big(-\lambda(x{-}s)\big); \qquad \text{for } s < x,$$

where λ is the rate of the Poisson process.

Exercise 5.14: Let $\mathsf{F}_Z(z)$ be the fraction of time (over the limiting interval $(0,\infty)$) that the age of a renewal process is at most z. Show that $\mathsf{F}_Z(z)$ satisfies

$$\mathsf{F}_Z(z) = \frac{1}{\overline{X}} \int_{x=0}^{z} \Pr\{X > x\}\, dx \qquad \text{WP1.}$$

Hint: Follow the argument in Example 5.4.7.

Solution: We want to find the time-average CDF of the age $Z(t)$ of an arbitrary renewal process, and do this by following Example 5.4.7. For any given $z > 0$, define the reward function $R(t)$ to be 1 for $Z(t) \leq 1$ and to be 0 otherwise, *i.e.*,

$$R(t) = \mathcal{R}(Z(t), \widetilde{X}(t)) = \begin{cases} 1; & Z(t) \leq z \\ 0; & \text{otherwise} \end{cases}.$$

Note that $R(t)$ is positive only over the first z units of a renewal interval. Thus $R_n = \min(z, X_n)$. It follows that

$$\mathsf{E}[R_n] = \mathsf{E}[\min(z, X_n)] = \int_0^\infty \Pr\{\min(X, z) > x\}\, dx \tag{A.31}$$

$$= \int_0^z \Pr\{X > x\}\, dx. \tag{A.32}$$

Let $\mathsf{F}_Z(z) = \lim_{t\to\infty} \int_0^t (1/t) R(\tau)\, d\tau$ denote the fraction of time that the age is less than or equal to z. From Theorem 5.4.5 and (A.32),

$$\mathsf{F}_Z(z) = \frac{\mathsf{E}[R_n]}{\overline{X}} = \frac{1}{\overline{X}} \int_0^z \Pr\{X > x\}\, dx \qquad \text{WP1.}$$

Exercise 5.16: **a)** Use Wald's equality to compute the expected number of trials of a Bernoulli process up to and including the kth success.

Solution: In a Bernoulli process $\{X_n; n \geq 1$, we call trial n a success if $X_n = 1$. Define a stopping trial J as the first trial n at which $\sum_{m=1}^n X_m = k$. This constitutes a stopping rule since $\mathbb{I}_{J=n}$ is a function of $X_1, \dots, X_n$. Given $J = n$, $S_J = X_1 + \cdots X_n = k$, and since this is true for all n, $S_j = k$ unconditionally. Thus $\mathsf{E}[S_J] = k$. From Wald's equality, $\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]$ so that $\mathsf{E}[J] = k/\overline{X}$.

We should have shown that $\mathsf{E}[J] < \infty$ to justify the use of Wald's equality, but we will show that in (b).

b) Use elementary means to find the expected number of trials up to and including the first success. Use this to find the expected number of trials up to and including the kth success. Compare with (a).

Solution: Let $\Pr\{X_n = 1\} = p$. Then the first success comes at trial 1 with probability p, and at trial n with probability $(1-p)^{n-1}p$. The expected time to the first success is then $1/p = 1/\overline{X}$. The expected time to the kth success is then $k/\overline{X}$, which agrees with the result in (a).

The reader might question the value of Wald's equality in this exercise, since the demonstration that $\mathsf{E}[J] < \infty$ was most easily accomplished by solving the entire problem by elementary means. In typical applications, however, the condition that $\mathsf{E}[J] < \infty$ is essentially trivial.

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 ith play. Assume that $\{X_i; i{=}1, 2, \dots\}$ is a set of IID random variables taking on integer values $\{-1, 0, 1, \dots\}$. Assume that $\mathsf{E}[X_i] < 0$. 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 $1 \le i \le n$. Thus $\Pr\{J > n\} \le \Pr\{S_n > -d\}$. Since $\mathsf{E}[S_n] = n\overline{X}$ and $\overline{X} < 0$, 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,

$$\Pr\{S_n > -d\} = \Pr\left\{\frac{S_n}{n} - \overline{X} > \frac{-d}{n} - \overline{X}\right\}.$$

The WLLN says that $\lim_{n\to\infty} \Pr\left\{|\frac{S_n}{n} - \overline{X}| > \epsilon\right\} = 0$ for all $\epsilon > 0$, and this implies the same statement with the absolute value signs removed, *i.e.*, $\Pr\left\{\frac{S_n}{n} - \overline{X} > \epsilon\right\} = 0$ for all $\epsilon > 0$. If we choose $\epsilon = -\overline{X}/2$ in the equation above, it becomes

$$\lim_{n\to\infty} \Pr\left\{S_n > \frac{n\overline{X}}{2}\right\} = 0.$$

Since $\overline{X} < 0$, we see that $-d > n\overline{X}/2$ for $n > 2\left|d/\overline{X}\right|$, and thus $\lim_{n\to\infty} \Pr\{S_n > -d\} = 0$.

b) Find $\mathsf{E}[J]$. Hint: The fact that there is only one possible negative outcome is important here.

Solution: One stops playing on trial $J = n$ if one's capital reaches 0 for the first time on the nth trial, *i.e.*, if $S_n = -d$ for the first time at trial n. This is clearly a function of $X_1, X_2, \dots, X_n$, 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, we then have

$$\mathsf{E}[J] = \frac{-d}{\overline{X}},$$

which is positive since $\overline{X}$ is negative. You should note from the exercises we have done with Wald's equality that it is often used to solve for $\mathsf{E}[J]$ after determining $\mathsf{E}[S_J]$.

For mathematical completeness, we should also verify that $\mathsf{E}[J] < \infty$ to use Wald's equality. If X is bounded, this can be done by using the Chernoff bound on $\Pr\{S_n > -d\}$. This is exponentially decreasing in n, thus verifying that $\mathsf{E}[J] < \infty$, and consequently that $\mathsf{E}[J] = -d/\overline{X}$. If X is not bounded, a truncation argument can be used. Letting $\breve{X}_b$ be X truncated to b, we see that $\mathsf{E}\left[\breve{X}_b\right]$ is increasing with b toward $\overline{X}$, and is less $\mathsf{E}[X]$ for all b. Thus the expected stopping time, say $\mathsf{E}[J_b]$ is upper bounded by $-d/\overline{X}$ for all b. It follows that $\lim_{b\to\infty} \mathsf{E}[J_b]$ is finite (and equal to $-d/\overline{X}$). Most students are ill-advised to worry too much about such details at first.

Exercise 5.19: Let $J = \min\{n \mid S_n \le b \text{ or } S_n \ge a\}$, where a is a positive integer, b is a negative integer, and $S_n = X_1 + X_2 + \cdots + X_n$. Assume that $\{X_i; i \ge 1\}$ 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, for each n, $\mathbb{I}_{J=n}$ must be a function of $X_1, \dots, X_n$. For the case here, $S_n = X_1 + \cdots + X_n$ is clearly a function of $X_1, \dots, X_n$, so the event that $S_n \ge a$ or $S_n \le b$ is a function of $X_1, \dots, X_n$. The first n at which this occurs is a function of $S_1, \dots, S_n$, which is a function of $X_1, \dots, X_n$. Thus $\mathbb{I}_{J=n}$ must be a function of $X_1, \dots, X_n$. For J to be a rv, we must show that $\lim_{n\to\infty} \Pr\{J \le n\} = 1$. 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 X is given to be zero-mean, $S_n/\sqrt{n}\sigma$ must approach normal. Now both $a/\sqrt{n}\sigma$ and $b/\sqrt{n}\sigma$ approach 0 as $n \to \infty$, so the probability that $\{S_n;\, n > 0\}$ (*i.e.*, the process in the absence of a stopping rule) remains 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 $\{S_n; n \ge 1\}$ 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 $\{S_n;\, n > 0\}$ can change only in integer steps, S_n cannot exceed a without some S_m, $m < n$ first equaling a and it cannot be less than b without some S_k, $k < n$ first equaling b. Thus S_J is only a or b.

c) Find an expression for $\mathsf{E}[S_J]$ in terms of p, a, and b, where $p = \Pr\{S_J \ge a\}$.

Solution: $\mathsf{E}[S_J] \;=\; a\Pr\{S_J = a\} + b\Pr\{S_J = b\} \;=\; pa + (1-p)b$

d) Find an expression for $\mathsf{E}[S_J]$ from Wald's equality. Use this to solve for p.

Solution: We have seen that J is a stopping trial for the IID rv's $\{X_i;\, i > 0\}$. Assuming for the moment that $\mathsf{E}[J] < \infty$, Wald's equality holds and $\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]$. Since $\overline{X} = 0$, we conclude that $\mathsf{E}[S_J] = 0$. Combining this with (c), we have $0 = pa + (1-p)b$, so $p = -b/(a-b)$. This is easier to interpret as

$$p = |b|/(a + |b|). \tag{A.33}$$

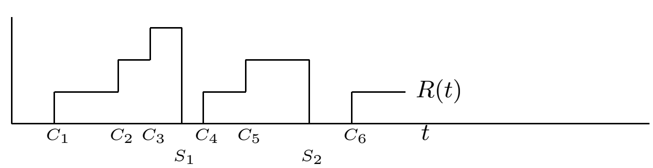
The assumption that $\mathsf{E}[J] < \infty$ is more than idle mathematical nitpicking, since we saw in Example 5.5.4 that $\mathsf{E}[J] = \infty$ for $b = -\infty$. The CLT does not resolve this issue, since the probability that S_n is outside the limits $[-b, a]$ approaches 0 with increasing n only as n^{-1}. Viewing the process as a finite-state Markov chain with a and b viewed as a common trapping state does resolve the issue, since, as seen in Theorem 4.5.4 applied to expected first passage times, the expected time to reach the trapping state is finite. As will be seen when we study Markov chains with countably many states, this result is no longer valid, as illustrated by Example 5.5.4. This issue will become more transparent when we study random walks in Chapter 9.

The solution in (A.33) applies only for $\overline{X} = 0$, but Chapter 9 shows how to solve the problem for an arbitrary distribution on X. We also note that the solution is independent of $\mathsf{p}_X(0)$, although $\mathsf{p}_X(0)$ is obviously involved in finding $\mathsf{E}[J]$. 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 catastophe if the gamblers capital is wiped out.

Exercise 5.31: Customers arrive at a bus stop according to a Poisson process of rate λ. Independently, buses arrive according to a renewal process with the inter-renewal interval CDF $\mathsf{F}_X(x)$. At the epoch of a bus arrival, all waiting passengers enter the bus and the bus leaves immediately. Let $R(t)$ be the number of customers waiting at time t.

a) Draw a sketch of a sample function of $R(t)$.

Solution: Let $S_n = X_1 + X_2 + \cdots$ be the epoch of the nth bus arrival and C_m the epoch of the mth customer arrival. Then $\{C_m; m \geq 1\}$ is the sequence of arrival times in a Poisson process of rate λ and $\{S_n; n \geq 1\}$ is the renewal time sequence of a renewal process.



b) Given that the first bus arrives at time $X_1 = x$, find the expected number of customers picked up; then find $\mathsf{E}\left[\int_0^x R(t)dt\right]$, again given the first bus arrival at $X_1 = x$.

Solution: The expected number of customers picked up, given $S_1 = x$ is the expected number of arrivals in the Poisson process by x, *i.e.*, λx. For $t < x$, $R(t)$ is the number of customers that have arrived by time t. $R(t)$, for $t < S_1$, is independent of S_1, so

$$\mathsf{E}\left[\int_0^x R(t)\,dt\right] \;=\; \int_0^x \lambda x\,dx \;=\; \frac{\lambda x^2}{2}.$$

c) Find $\lim_{t\to\infty} \frac{1}{t}\int_0^t R(\tau)d\tau$ (with probability 1). Assuming that F_X is a non-arithmetic distribution, find $\lim_{t\to\infty} \mathsf{E}[R(t)]$. Interpret what these quantities mean.

Solution: In (b), we found the expected reward in an inter-renewal interval of size x, *i.e.*, we took the expected value over the Poisson process given a specific value of X_1. Taking the expected value of this over the bus renewal interval, we get $(1/2)\lambda\mathsf{E}\left[X^2\right]$. This is the expected accumulated reward over the first bus inter-renewal period, denoted $\mathsf{E}[\mathrm{R}_1]$ in (5.22). It is the expected number of customers at each time, integrated over the time until the first bus arrival. This integrated value is not the expected number waiting for a bus, but rather will be used as a step in finding the time-average number waiting over time. Since the bus arrivals form a renewal process, this is equal to $\mathsf{E}[\mathrm{R}_n]$ for each inter-renewal period n. By (5.24),

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t R(\tau)d\tau \;=\; \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}} \;=\; \frac{\lambda\mathsf{E}\left[X^2\right]}{2\overline{X}} \qquad \text{WP1.} \tag{A.34}$$

Since the renewals are non-arithmetic, this is the same as $\lim_{t\to\infty}\mathsf{E}[R(t)]$.

Note that this is not the same as the expected number of customers per bus. It is the expected number of waiting customers, averaged over time. If a bus is highly delayed, a large number of customers accumulate, but also, because of the large interval waiting for the bus, the contribution to the time average grows as the square of the wait.

d) Use (c) to find the time-average expected wait per customer.

Solution: Little's theorem can be applied here since (A.34) gives the limiting time-average number of customers in the system, L. The time-average wait per customer W is $W = L/\lambda$, which from (A.34) is $\mathsf{E}\left[X^2\right]/2\overline{X}$. The system here does not satisfy the conditions of Little's theorem, but it is easy to check that the proof of the theorem applies in this case.

e) Find the fraction of time that there are no customers at the bus stop. (Hint: this part is independent of (a), (b), and (c); check your answer for $\mathsf{E}[X] << 1/\lambda$).

Solution: There are no customers at the bus stop at the beginning of each renewal period. Let U_n be the interval from the beginning of the nth renewal period until the first customer arrival. It is possible that no customer will arrive before the next bus arrives, so the interval within the nth inter-renewal period when there is no customer waiting is $\min(U_n, X_n)$. Consider a reward function $R(t)$ equal to 1 when no customer is in the system and 0 otherwise. The accumulated reward R_n within the nth inter-renewal period is then $\mathrm{R}_n = \min(U_n, X_n)$. Thus, using the independence of U_n and X_n,

$$\begin{aligned}
\mathsf{E}[\mathrm{R}_n] &= \int_0^\infty \Pr\{\min(U_n, X_n) > t\}\, dt \;=\; \int_0^\infty \Pr\{(U_n > t\}\Pr\{X_n > t\}\, dt \\
&= \int_0^t \Pr\{X_n > t\}\, e^{-\lambda t}\, dt.
\end{aligned}$$

Using (5.24), the limiting time-average fraction of time when no customers are waiting is $(1/\overline{X})\int_0^t \Pr\{X_n > t\}\, e^{-\lambda t}\, dt$. Checking when $\overline{X} << 1/\lambda$, we see that the above integral is close to $\overline{X}$. In particular, if X_n is exponential with rate μ, we see that the fraction above is $\mu/(\mu+\lambda)$.

Exercise 5.44: This is a very simple exercise designed to clarify confusion about the roles of past, present, and future in stopping rules. Let $\{X_n; n \geq 1\}$ be a sequence of IID binary rv's , each with the pmf $\mathsf{p}_X(1) = 1/2$, $\mathsf{p}_X(0) = 1/2$. Let J be a positive integer-valued rv that takes on the sample value n of the first trial for which $X_n = 1$. That is, for each $n \geq 1$,

$$\{J = n\} = \{X_1{=}0,\ X_2{=}0, \dots , X_{n-1}{=}0,\ X_n{=}1\}.$$

a) Use the definition of stopping trial, Definition 5.5.1, to show that J is a stopping trial for $\{X_n; n \geq 1\}$.

Solution: Note that $\mathbb{I}_{J=n} = 1$ if $X_n = 1$ and $X_i = 0$ for all $i < n$. $\mathbb{I}_{J=n} = 0$ otherwise. Thus $\mathbb{I}_{J=n}$ is a function of $X_1, \dots , X_n$. We also see that J is a positive integer-valued rv. That is, each sample sequence of $\{X_i; i \geq 1\}$ (except the zero probability sequence of all zeros) maps into a positive integer.

b) Show that for any given n, the rv's X_n and $\mathbb{I}_{J=n}$ are *statistically dependent*.

Solution: Note that $\Pr\{\mathbb{I}_{J=n} = 1\} = 2^{-n}$ and $\Pr\{X_n = 1\} = 1/2$. The product of these is $2^{-(n+1)}$, whereas $\Pr\{X_n = 1, \mathbb{I}_{J=n} = 1\} = 2^{-n}$, demonstrating statistical dependence. More intuitively $\Pr\{X_n = 1 \mid \mathbb{I}_{J=n} = 1\} = 1$ and $\Pr\{X_n = 1\} = 1/2$.

c) Show that for every $m > n$, X_n and $\mathbb{I}_{J=m}$ are *statistically dependent*.

Solution: For $m > n$, $\mathbb{I}_{J=m}{=}1$ implies that $X_i = 0$ for $i < m$, and thus that $X_n = 0$. Thus $\Pr\{X_n = 1 \mid \mathbb{I}_{J=m}{=}1\} = 0$. Since $\Pr\{X_n = 1\} = 1/2$, $\mathbb{I}_{J=m}$ and X_n are dependent.

d) Show that for every $m < n$, X_n and $\mathbb{I}_{J=m}$ are *statistically independent*.

Solution: Since $\mathbb{I}_{J=m}$ is a function of $X_1, \dots , X_m$ and X_n is independent of $X_1, \dots , X_m$, it is clear that X_n is independent of $\mathbb{I}_{J=m}$

e) Show that X_n and $\mathbb{I}_{J\geq n}$ are *statistically independent*. Give the simplest characterization you can of the event $\{J \geq n\}$.

Solution: For the same reason as $\mathbb{I}_{J=m}$ is independent of X_n for $m < n$, we see that $\mathbb{I}_{J<n}$ is independent of X_n. Now $\{J \geq n\} = \{J < n\}^{\mathsf{c}}$, so $\mathbb{I}_{J\geq n} = 1 - \mathbb{I}_{J<n}$. Thus $\{J \geq n\}$ is also independent of X_n. We give an intuitive explanation after (f).

f) Show that X_n and $\mathbb{I}_{J>n}$ are *statistically dependent*.

Solution: The event $\{J > n\}$ implies that $X_1, \dots , X_n$ are all 0, so $\Pr\{X_n = 1 \mid J > n\} = 0$. Since $\Pr\{X_n{=}1\} = 1/2$, it follows that X_n and $\mathbb{I}_{J>n}$ are dependent.

It is important (and essentially the central issue of the exercise) to understand why (e) and (f) are different. Note that $\{J \geq n\} = \{J = n\} \bigcup \{J > n\}$. Now $\{J = n\}$ implies that $X_n = 1$ whereas $\{J > n\}$ implies that $X_n = 0$. The union, however, implies nothing about X_n, and is independent of X_n. The union, $\mathbb{I}_{J\geq n}$ is the event that $X_1, \dots , X_{n-1}$ are all 0 and X_n then determines whether $J = n$ or $J > n$.

Note: The results here are characteristic of most sequences of IID rv's. For most people, this requires some realignment of intuition, since $\{J{\geq}n\}$ is the union of $\{J{=}m\}$ for all $m \geq n$, and each of these events are highly dependent on X_n. The right way to think of this is that $\{J{\geq}n\}$ is the complement of $\{J{<}n\}$, which is determined by $X_1, \dots , X_{n-1}$. Thus $\{J{\geq}n\}$ is also determined by $X_1, \dots , X_{n-1}$ and is thus independent of X_n. The moral of the story is that thinking of stopping rules as rv's independent of the future is very tricky, even in totally obvious cases such as this.

A.6 Solutions for Chapter 6

Exercise 6.1: Let $\{P_{ij}; i,j \geq 0\}$ be the set of transition probabilities for a countable-state Markov chain. For each i, j, let $\mathsf{F}_{ij}(n)$ be the probability that state j occurs sometime between time 1 and n inclusive, given $X_0 = i$. For some given j, assume that $\{x_i; i \geq 0\}$ 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 $x_i \geq \mathsf{F}_{ij}(n)$ for all n and i, and hence that $x_i \geq \mathsf{F}_{ij}(\infty)$ 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

$$F_{ij}(1) \;=\; P_{ij} \;\leq\; P_{ij} + \sum_{k\neq j} P_{ik}x_k \;=\; x_i.$$

For the inductive step, assume that $F_{ij}(n) \leq x_i$ for a given n and all i. Using (6.8),

$$\begin{aligned} F_{ij}(n+1) \;&=\; P_{ij} + \sum_{k\neq j} P_{ik}F_{kj}(n) \\ &\leq\; P_{ij} + \sum_{k\neq j} P_{ik}x_k \;=\; x_i \quad \text{for all } i. \end{aligned}$$

By induction, it then follows that $F_{ij}(n) \leq x_i$ for all i, n. From (6.7), $F_{ij}(n)$ is non-decreasing in n and upper bounded by 1. It thus has a limit, $F_{ij}(\infty)$, which satisfies $F_{ij}(\infty) \leq x_i$ for all i.

One solution to the set of equations $x_i = P_{ij} + \sum_{k\neq j} P_{ik}x_k$ is $x_i = 1$ for all i. Another (from 6.9)) is $x_i = F_{ij}(\infty)$. These solutions are different when $F_{ij}(\infty) < 1$, and this exercise then shows that the solution $\{F_{ij}(\infty); i \geq 0\}$ is the smallest of all possible solutions.

Exercise 6.2: **a)** For the Markov chain in Figure 6.2, show that, for $p \geq 1/2$, $\mathsf{F}_{00}(\infty) = 2(1-p)$ and show that $\mathsf{F}_{i0}(\infty) = [(1-p)/p]^i$ 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: Note that the diagram in Figure 6.2 implicitly assumes $p < 1$, and we assume that here, since the $p = 1$ case is trivial anyway.

The hint provides a straightforward algebraic solution, but provides little insight. It may be helpful, before carrying that solution out, to solve the problem by the more intuitive method used in the 'stop when you're ahead' example, Example 5,4.4. This method also derives the values of $F_{i0}(\infty)$ rather than just verifying that they form a solution.

Let $F_{10}(\infty)$ be denoted by α. This is the probability of ever reaching state 0 starting from state 1. This probability is unchanged by converting state 0 to a trapping state, since when starting in state 1, transitions from state 0 can occur only after reaching state 0.

In the same way, $F_{i,i-1}(\infty)$ for any $i > 1$ is unchanged by converting state $i-1$ into a trapping state, and the modified Markov chain, for states $k \geq i-1$ is then the same as the modified chain that uses state 0 as a trapping state. Thus $F_{i,i-1}(\infty) = \alpha$.

We can proceed from this to calculate $F_{i,i-2}$ for any $i > 2$. In order to access state $i-2$ from state i, it is necessary to first access state $i-1$. Given a first passage from i to $i-1$, (an event of probability α), it is necessary to have a subsequent first passage from $i-1$ to $i-2$, so $F_{i,i-2}(\infty) = \alpha^2$. Iterating on this, $F_{i0}(\infty) = \alpha^i$.

We can now solve for α by using (6.9) for $i = 1$, $j = 0$: $F_{10}(\infty) = q + pF_{20}(\infty)$. Thus $\alpha = q + p\alpha^2$. This is a quadratic equation in α with the solutions $\alpha = 1$ and $\alpha = q/p$. We then know that $F_{10}(\infty)$ has either the value 1 or q/p. For whichever value of α is correct, we also have $P_{i0}(\infty) = \alpha^i$ for $i \geq 1$. Finally, from (6.9) for $j = 0$, $i = 0$, $F_{00}(\infty) = q + p\alpha$.

From the reasoning above, we know that these two possible solutions are the only possibilities. If both of them satisfy (6.9), then, from Exercise 6.1, the one with $\alpha = q/p$ is the correct one since it is the smaller solution to (6.9). We now show that the solution with $\alpha = q/p$ satisfies (6.9). This is the first part of what the question asks, but it is now unnecessary to also show that this is the smallest solution.

If we substitute the hypothesized solution,

$$F_{00}(\infty) = 2q; \qquad F_{i0}(\infty) = (q/p)^i \quad \text{for } i > 0, \tag{A.35}$$

into (6.9) for $j = 0$, we get the hypothesized set of equalities,

$$\begin{aligned} 2q &= q + p(q/p) && \text{for } i = 0 \\ q/p &= q + p(q/p)^2 && \text{for } i = 1 \\ (q/p)^i &= q(q/p)^{i-1} + p(q/p)^{i+1} && \text{for all } i \geq 2. \end{aligned}$$

The first of these is clearly an identity, and the third is seen to be an identity by rewriting the right side as $p(q/p)^i + q(q/p)^i$. The second is an identity by almost the same argument. Thus (A.35) satisfies (6.9) and thus gives the correct solution.

For those who dutifully took the hint directly, it is still necessary to show that (A.35) is the smallest solution to (6.9). Let x_i abbreviate $F_{i0}(\infty)$ for $i \geq 0$. Then (6.9) for $j = 0$ can be rewritten as

$$\begin{aligned} x_0 &= q + px_1 \\ x_1 &= q + px_2 \\ x_i &= qx_{i-1} + px_{i+1} && \text{for } i \geq 2. \end{aligned}$$

The equation for each $i \geq 2$ can be rearranged to the alternate form,

$$x_{i+1} - (q/p)x_i = x_i - (q/p)x_{i-1}. \tag{A.36}$$

For $i = 1$, the similar rearrangement is $x_2 - (q/p)x_1 = x_1 - q/p$. Now consider the possibility of a solution to these equations with $x_1 < q/p$, say $x_1 = (q/p) - \delta$ with $\delta > 0$. Recursing on the equations (A.36), we have

$$x_{i+1} - (q/p)x_i = -\delta \qquad \text{for } i \geq 1. \tag{A.37}$$

For the case $q = p = 1/2$, this becomes $x_{i+1} - x_i = -\delta$. Thus for sufficiently large i, x_i becomes negative, so there can be no non-negative solution of (6.9) with $F_{10} < q/p$.

For the case $1/2 < p < 1$, (A.37) leads to

$$x_{i+1} = (q/p)x_i - \delta \le (q/p)^2 x_{i-1} - \delta \le \cdots \le (q/p)^i - \delta.$$

For large enough i, this shows that x_{i+1} is negative, showing that no non-negative solution exists with $x_1 < q/p$.

b) Under the same conditions as (a), show that $\mathsf{F}_{ij}(\infty)$ equals $2(1-p)$ for $j = i$, equals $[(1-p)/p]^{i-j}$ for $i > j$, and equals 1 for $i < j$.

Solution: In the first part of the solution to (a), we used a trapping state argument to show that $F_{i,i-1}(\infty) = F_{10}(\infty)$ for each $i > 1$. That same argument shows that $F_{ij} = F_{i-j,0}(\infty)$ for all $i > j$. Thus $F_{ij}(\infty) = (q/p)^{i-j}$ for $i > j$.

Next, for $i < j$, consider converting state j into a trapping state. This does not alter $F_{ij}(\infty)$ for $i < j$, but converts the states $0, 1, \dots, j$ into a finite-state Markov chain with a single recurrent state, j. Thus $F_{ij}(\infty) = 1$ for $i < j$.

Finally, for $i = j > 1$, (6.9) says that

$$F_{ii}(\infty) = qF_{i-1,i}(\infty) + pF_{i+1,i}(\infty) = q + pF_{10}(\infty) = 2q,$$

where we have used the earlier results for $i < j$ and $i > j$. The case for $F_{11}(\infty)$ is virtually the same.

Exercise 6.3 a): Show that the nth order transition probabilities, starting in state 0, for the Markov chain in Figure 6.2 satisfy

$$P_{0j}^n = pP_{0,j-1}^{n-1} + qP_{0,j+1}^{n-1}, \quad \text{for } j \ne 0; \qquad P_{00}^n = qP_{00}^{n-1} + qP_{01}^{n-1}.$$

Hint: Use the Chapman-Kolmogorov equality, (4.7).

Solution: This *is* the Chapman-Kolmogorov equality in the form $P_{ij}^n = \sum_k P_{ik}^{n-1} P_{kj}$ where $P_{00} = q$, $P_{k,k+1} = p$ for all $k \ge 0$ and $P_{k,k-1} = q$ for all $k \ge 1$; $P_{kj} = 0$ otherwise.

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 $p \ne 1/2$, so don't attempt this in general.

Solution: This is less tedious if organized as an array of terms. Each term (except P_{00}^n) for each n is then half the term diagonally above and to the left, plus half the term diagonally above and to the right. P_{00}^n is half the term above plus half the term diagonally above and to the right.

j	0	1	3	3	4
$P_{0,j}^1$	1/2	1/2			
$P_{0,j}^2$	1/2	1/4	1/4		
$P_{0,j}^3$	3/8	3/8	1/8	1/8	
$P_{0,j}^4$	3/8	1/4	1/4	1/16	1/16

terms are similar. We also see (with less work) that (6.3) is valid for $1 \leq n \leq 3$ for all $j, 0 \leq j \leq n$. We can avoid fussing with the constraint $j \leq n$ by following the convention that $\binom{n}{k} = 0$ for all $n \geq 1$ and $k > n$. We have then shown that (6.3) is valid for $1 \leq n \leq 4$ and all $j \geq 0$.

We next use induction to validate (6.3) in general. From the previous calculation, any $n, 1 \leq n \leq 4$ can be used as the basis of the induction and then, given the assumption that (6.3) is valid for any given $n > 0$ and all $j \geq 0$, we will prove that (6.3) is also valid for $n{+}1$ and all $j \geq 0$. Initially, for given n, we assume $j > 0$; the case $j = 0$ is a little different since it has self transitions, so we consider it later.

For the subcase where $n + j$ is even, we have

$$\begin{aligned} P_{0j}^{n+1} &= \frac{1}{2}\left[P_{0,j-1}^{n} + P_{0,j+1}^{n}\right] \\ &= \frac{1}{2}\left[\binom{n}{(j+n)/2}2^{-n} + \binom{n}{((j+n)/2)+1}2^{-n}\right] \\ &= \binom{n+1}{((j+n)/2)+1}2^{-(n+1)}. \end{aligned}$$

The first equality comes from part (a) with $p = q = 1/2$. The second equality uses (6.3) for the given n and uses the fact that $n{+}j{-}1$ and $n{+}j{+}1$ are odd. The final equality follows immediately from the combinatorial identity

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

This identity follows 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. The final result for P_{0j}^{n+1} above is for $n{+}1{+}j$ odd and agrees with the odd case in (6.3) after substituting $n{+}1$ for n in (6.3). This validates (6.3) for this case. The subcase where $n{+}j$ is odd is handled in the same way.

For the case $j = 0$, we use the second portion of part (a), namely $P_{00}^{n+1} = \frac{1}{2}P_{00}^{n} + \frac{1}{2}P_{01}^{n}$. For the subcase where n is even, we then have

$$\begin{aligned} P_{00}^{n+1} &= \frac{1}{2}\left[\binom{n}{n/2}2^{-n} + \binom{n}{(n/2)+1}2^{-n}\right] \\ &= \binom{n{+}1}{(n/2){+}1}2^{-(n+1)}. \end{aligned}$$

In the first equality, we have used (6.3) with $n{+}j$ even for $j{=}0$ and $n{+}j$ odd for $j{=}1$. The second equality uses the combinatorial identity above. This result agrees with the odd case in (6.3) after substituting $n{+}1$ for n. Finally, the subcase where n is odd is handled in almost the same way, except that one must recognize that $\binom{n}{(n+1)/2} = \binom{n}{(n-1)/2}$.

(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. Similarly, (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 meaning of the hint will be less obscure if we redraw Figure 6.1 in the following way:

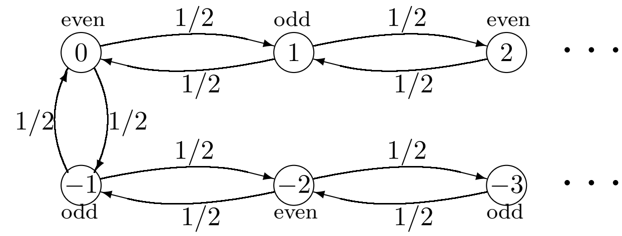


Figure 6.1 redrawn

Compare this with the discrete M/M/1 chain of Figure 6.2,

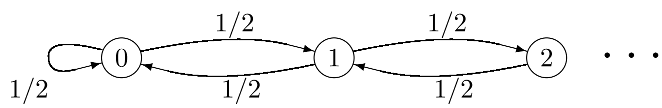


Figure 6.2 redrawn

Note that Figure 6.2 can be viewed as the result of combining each nonnegative state i in the redrawn Figure 6.1 with the state $-i-1$ lying directly beneath it. To be more specific, the transition probability from state 0 in Fig. 6.1 to the aggregated states 1 and -2 is 1/2, and the transition probability to the aggregated states 0 and -1 is 1/2. The same transition probabilities hold for state -1. Similarly, starting from any state i in Fig 6.1, there is a transition probablity 1/2 to the aggregated states $i+1, -i-2$ and 1/2 to the aggregated states $i-1$ and i. The same aggregated transition probabilities hold starting from state $-i-1$.

What this means is that the set of aggregated pairs forms a Markov chain in its own right, and this Markov chain is the M/M/1 chain of Fig 6.2. The nth order transition probabilites P^n_{0i} for the M/M/1 chain are thus the same as the nth order aggregate transition probabilities say $Q^n_{0i} + Q^n_{0,-i-1}$ for the Bernoulli chain. Since the Bernoulli chain is periodic with period 2, however, and each pair of states consists of one even and one odd term, only one of these aggregated terms is nonzero. This helps explain the strange looking difference in part (b) between $n+i$ even and odd.

Exercise 6.8: Let $\{X_n; n \geq 0\}$ be a branching process with $X_0 = 1$. Let $\overline{Y}$, σ^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 $\mathsf{F}_{10}(\infty)$) or the value ∞ (with probability $1 - \mathsf{F}_{10}(\infty)$).

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. You should be proud of yourself if you noticed the need for ruling this case out before constructing a proof.

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 $P_{jj} = p_1^j$ and thus $P_{jj}^m = p_1^{mj} \to 0$ as $m \to \infty$.

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 $1, 2, \dots$ are in a transient class. To see this, note that $P_{10} = p_0 > 0$, so $\mathsf{F}_{11}(\infty) \le 1 - p_0 < 1$, which means by definition that state 1 is transient. All states $i > 1$ communicate with state 1, so by Theorem 6.2.5, all states $j \ge 1$ are transient. Thus one can argue that the process has 'no place to go' other than 0 or ∞.

The following tedious analysis makes this precise. Each $j > 0$ is transient, so from Theorem 6.2.6 part 3,

$$\lim_{t\to\infty} \mathsf{E}\left[N_{jj}(t)\right] < \infty.$$

Note that $N_{1j}(t)$ is the number of visits to j in the interval $[1, t]$ starting from state 1 at time 0. This is one more than the number of returns to j after the first visit to j. The expected number of such returns is upper bounded by the number in t steps starting in j, so $\mathsf{E}\left[N_{1j}(t)\right] \le 1 + \mathsf{E}\left[N_{jj}(t)\right]$. It follows that

$$\lim_{t\to\infty} \mathsf{E}\left[N_{1j}(t)\right] < \infty \qquad \text{for each } j > 0.$$

Now note that the expected number of visits to j in $[1, t]$ can be rewritten as $\sum_{n=1}^{t} P_{ij}(n)$. Since this sum in the limit $t \to \infty$ is finite, the remainder in the sum from t to ∞ must approach 0 as $t \to \infty$, so

$$\lim_{t\to\infty} \sum_{n>t} P_{1j}^n = 0.$$

From this, we see that for every finite integer ℓ,

$$\lim_{t\to\infty} \sum_{n>t} \sum_{j=1}^{\ell} P_{1j}^n = 0.$$

This says that for every $\epsilon > 0$, there is a t sufficiently large that the probability of ever entering states 1 to ℓ on or after step t is less than ϵ. Since $\epsilon > 0$ is arbitrary, all sample paths (other than a set of probability 0) never enter states 1 to ℓ after some finite time. Since ℓ is arbitrary, $\lim_{n\to\infty} X_n$ exists WP1 and is either 0 or ∞. By definition, it is 0 with probability $F_{10}(\infty)$.

b) Show that $\mathsf{VAR}\left[X_n\right] = \sigma^2 \overline{Y}^{n-1}(\overline{Y}^n - 1)/(\overline{Y} - 1)$ for $\overline{Y} \ne 1$ and $\mathsf{VAR}\left[X_n\right] = n\sigma^2$ for $\overline{Y} = 1$.

Solution: We will demonstrate the case for $\overline{Y} \ne 1$ and $\overline{Y} = 1$ together by showing that

$$\mathsf{VAR}\left[X_n\right] = \sigma^2 \overline{Y}^{n-1} \Big[1 + \overline{Y} + \overline{Y}^2 + \cdots + \overline{Y}^{n-1}\Big]. \tag{A.38}$$

First express $\mathsf{E}\left[X_n^2\right]$ in terms of $\mathsf{E}\left[X_{n-1}^2\right]$. Note that, conditional on $X_{n-1} = \ell$, X_n is the sum of ℓ IID rv's each with mean $\overline{Y}$ and variance σ^2, so

$$\begin{aligned}
\mathsf{E}\left[X_n^2\right] &= \sum_\ell \Pr\{X_{n-1} = \ell\}\, \mathsf{E}\left[X_n^2 \mid X_{n-1} = \ell\right] \\
&= \sum_\ell \Pr\{X_{n-1} = \ell\} \left[\ell\sigma^2 + \ell^2\overline{Y}^2\right] \\
&= \sigma^2 \mathsf{E}\left[X_{n-1}\right] + \overline{Y}^2 \mathsf{E}\left[X_{n-1}^2\right] \\
&= \sigma^2 \overline{Y}^{n-1} + \overline{Y}^2 \mathsf{E}\left[X_{n-1}^2\right].
\end{aligned}$$

We also know from (6.50) (or by simple calculation) that $\mathsf{E}\left[X_n\right] = \overline{Y}\mathsf{E}\left[X_{n-1}\right]$. Thus,

$$\mathsf{VAR}\left[X_n\right] = \mathsf{E}\left[X_n^2\right] - \Big[\mathsf{E}\left[X_n\right]\Big]^2 \;=\; \sigma^2\overline{Y}^{n-1} + \overline{Y}^2 \mathsf{VAR}\left[X_{n-1}\right]. \tag{A.39}$$

We now use induction to derive (A.38) from (A.39). For the base of the induction, we see that X_1 is the number of progeny from the single element X_0, so $\mathsf{VAR}\left[X_1\right] = \sigma^2$. For the inductive step, we assume that

$$\mathsf{VAR}\left[X_{n-1}\right] = \sigma^2\overline{Y}^{n-2}\Big[1 + \overline{Y} + \overline{Y}^2 + \cdots + \overline{Y}^{n-2}\Big].$$

Substituting this into (A.39),

$$\begin{aligned}
\mathsf{VAR}\left[X_n\right] &= \sigma^2\overline{Y}^{n-1} + \sigma^2\overline{Y}^{n}\Big[1 + \overline{Y} + \cdots + \overline{Y}^{n-2}\Big] \\
&= \sigma^2\overline{Y}^{n-1}\Big[1 + \overline{Y} + \cdots + \overline{Y}^{n-1}\Big],
\end{aligned}$$

completing the induction.

Exercise 6.9: There are n states and for each pair of states i and j, a positive number $d_{ij} = d_{ji}$ is given. A particle moves from state to state in the following manner: Given that the particle is in any state i, it will next move to any $j \neq i$ with probability P_{ij} given by

$$P_{ij} = \frac{d_{ij}}{\sum_{k\neq i}\, d_{ik}}. \tag{A.40}$$

Assume that $P_{ii} = 0$ for all i. Show that the sequence of positions is a reversible Markov chain and find the limiting probabilities.

Solution: From Theorem 6.5.3, $\{P_{ij}\}$ is the set of transition probabilities and $\boldsymbol{\pi}$ is the steady state probability vector of a reversible chain if $\pi_i P_{ij} = \pi_j P_{ji}$ for all i, j. Thus, given $\{d_{ij}\}$, we attempt to find $\boldsymbol{\pi}$ to satisfy the equations

$$\pi_i P_{ij} = \frac{\pi_i d_{ij}}{\sum_k d_{ik}} = \frac{\pi_j d_{ji}}{\sum_k d_{jk}} = \pi_j P_{ji}; \qquad \text{for all } \; i, j.$$

We have taken $d_{ii} = 0$ for all i here. Since $d_{ij} = d_{ji}$ for all i, j, we can cancel d_{ij} and d_{ji} from the inner equations, getting

$$\frac{\pi_i}{\sum_k d_{ik}} = \frac{\pi_j}{\sum_k d_{jk}}; \qquad \text{for all } \; i, j.$$

Thus π_i must be proportional to $\sum_k d_{ik}$, so normalizing to make $\sum_i \pi_i = 1$, we get

$$\pi_i = \frac{\sum_k d_{ik}}{\sum_\ell \sum_k d_{\ell k}}; \qquad \text{for all } i.$$

If the chain has a countably infinite number of states, this still works if the sums exist.

This exercise is not quite as specialized as it sounds, since given any reversible Markov chain, we can definite $d_{ij} = \pi_i P_{ij}$ and get this same set of equations (normalized in a special way).

Exercise 6.10: Consider a reversible Markov chain with transition probabilities P_{ij} and limiting probabilities π_i. Also consider the same chain truncated to the states $0, 1, \dots, M$. That is, the transition probabilities $\{P'_{ij}\}$ of the truncated chain are

$$P'_{ij} = \begin{cases} \frac{P_{ij}}{\sum_{k=0}^{m} P_{ik}} & ; \quad 0 \le i, j \le M \\ 0 & ; \quad \text{elsewhere.} \end{cases}.$$

Show that the truncated chain is also reversible and has limiting probabilities given by

$$\overline{\pi}_i = \frac{\pi_i \sum_{j=0}^{M} P_{ij}}{\sum_{k=0}^{M} \left(\pi_k \sum_{m=0}^{M} P_{km}\right)}. \tag{A.41}$$

Solution: The steady state probabilities $\{\pi_i;\, i \ge 0\}$ of the original chain must be positive. Thus $\overline{\pi}_i > 0$ for each i. By summing $\overline{\pi}_i$ over i in (A.41), it is seen that the numerator is the same as the denominator, so $\sum_{i=0}^{\mathsf{M}} \overline{\pi}_i = 1$. Finally,

$$\overline{\pi}_i P'_{ij} = \frac{\pi_i P_{ij}}{\sum_{k=0}^{M} \left(\pi_k \sum_{m=0}^{M} P_{km}\right)}.$$

Since $\pi_i P_{ij} = \pi_j P_{ji}$ for each i, j, it is clear that $\overline{\pi}_i P'_{ij} = \overline{\pi}_j P'_{ji}$ for each $i, j \le \mathsf{M}$. Thus, by Theorem 6.5.3, the truncated chain is reversible. It seems like this should be obvious intuititvely, but the rules for truncation are sufficiently complicated that it doesn't seem obvious at all.

Exercise 6.12: **a)** Use the birth and death model described in figure 6.4 to find the steady-state probability mass function for the number of customers in the system (queue plus service facility) for the following queues:

i) M/M/1 with arrival probability $\lambda\delta$, service completion probability $\mu\delta$.

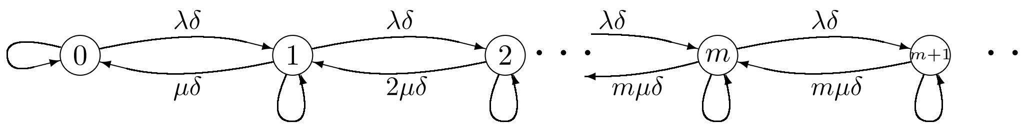
ii) M/M/m with arrival probability $\lambda\delta$, service completion probability $i\mu\delta$ for i servers busy, $1 \le i \le m$.

iii) M/M/∞ with arrival probability $\lambda\delta$, service probability $i\mu\delta$ for i servers. Assume δ so small that $i\mu\delta < 1$ for all i of interest.

Assume the system is positive recurrent.

Solution: **M/M/1:** This is worked out in detail in section 6.6. From (6.45), $\pi_i = (1-\rho)\rho^i$ where $\rho = \lambda/\mu < 1$.

M/M/m: The Markov chain for this is



This is a birth-death chain and the steady state probabilitiers are given by (6.33) where $\rho_i = \lambda/((i{+}1)\mu)$ for $i < m$ and $\rho_i = \lambda/m\mu$ for $i \geq m$. Evaluating this,

$$\pi_i = \frac{\pi_0(\lambda/\mu)^i}{i!} \quad \text{for } i < m; \qquad \pi_i = \frac{\pi_0(\lambda/\mu)^i}{m!(m^{i-m})} \quad \text{for } i \geq m,$$

where π_0 is given by

$$\pi_0^{-1} = 1 + \sum_{i=1}^{m-1} \frac{(\lambda/\mu)^i}{i!} + \frac{(m\rho_m)^m}{m!(1-\rho_m)}. \tag{A.42}$$

M/M/∞: The assumption that δ is so small that $i\mu\delta < 1$ for all 'i of interest' is rather strange, since we don't know what is of interest. When we look at the solution to the M/M/1 and M/M/m sample-time queues above, however, we see that they do not depend on δ. It is necessary for $m\mu\delta \leq 1$ for the Markov chain to be defined, and $m\mu\delta << 1$ for it to be a reasonable approximation to the behavior of the continuous time queue, but the results turn out to be the same as the continuous time queue in any case, as will be seen in Chapter 7. What was intended here was to look at the limit of the M/M/m queue in the limit $m \to \infty$. When we do this,

$$\pi_i = \frac{\pi_0(\lambda/\mu)^i}{i!} \quad \text{for all } i,$$

where

$$\pi_0^{-1} = 1 + \sum_{i=1}^{\infty} \frac{(\lambda/\mu)^i}{i!}.$$

Recognizing this as the power series of an exponential, $\pi_0 = e^{-\lambda/\mu}$.

b) For each of the queues above give necessary conditions (if any) for the states in the chain to be i) transient, ii) null recurrent, iii) positive recurrent.

Solution: The M/M/1 queue is transient if $\lambda/\mu > 1$, null recurrent if $\lambda/\mu = 1$, and positive recurrent if $\lambda/\mu < 1$ (see Section 6.6). In the same way, the M/M/m queue is transient if $\lambda/m\mu > 1$, null recurrent if $\lambda/m\mu = 1$ and positive recurrent if $\lambda/m\mu < 1$. The M/M/∞ queue is positive recurrent in all cases. As the arrival rate speeds up, the number of servers in use increases accordingly.

c) For each of the queues find:

L = (steady-state) mean number of customers in the system.

L^q = (steady-state) mean number of customers in the queue.

W = (steady-state) mean waiting time in the system.

W^q = (steady-state) mean waiting time in the queue.

Solution: The above parameters are related in a common way for the three types of queues. First, applying Little's law first to the system and then to the queue, we get

$$W = L/\lambda; \qquad W^q = L^q/\lambda. \tag{A.43}$$

Next, define L^v as the steady-state mean number of customers in service and W^v as the steady state waiting time per customer in service. We see that for each system, $W^v = 1/\mu$, since when a customer enters service the mean time to completion is $1/\mu$. From Little's law applied to service, $L^v = \lambda/\mu$. Now $L = L^q + L^v$ and $W = W^q + W^v$. Thus

$$L = L^q + \lambda/\mu; \qquad W = W^q + 1/\mu. \tag{A.44}$$

Thus for each of the above queue types, we need only compute one of these four quantities and the others are trivially determined. We assume positive recurrence in all cases.

For M/M/1, we compute

$$L \;=\; \sum_i i\pi_i \;=\; (1-\rho)\sum_{i=0}^{\infty} i\rho^i \;=\; \frac{\rho}{1-\rho} \;=\; \frac{\lambda}{\mu-\lambda}.$$

For M/M/m, we compute L^q since queueing occurs only in states $i > m$.

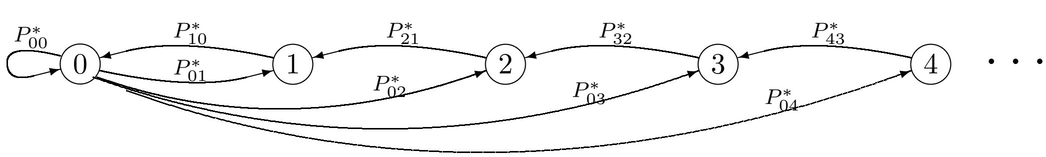
$$\begin{aligned} L^q \;&=\; \sum_{i>m}(i-m)\pi_i \;=\; \sum_{i>m}\frac{(i{-}m)\pi_0(\lambda/\mu)^i}{m!(m^{i-m})} \\ &=\; \sum_{j>0}\frac{j\pi_0\rho_m^j(\lambda/\mu)^m}{m!} \;=\; \frac{\pi_0\rho_m(\lambda/\mu)^m}{(1-\rho_m)^2 m!}, \end{aligned}$$

where π_0 is given in (A.42).

Finally, for M/M/∞, there is no queueing, so $L^q = 0$.

Exercise 6.14: Find the backward transition probabilities for the Markov chain model of age in Figure 6.3. Draw the graph for the backward Markov chain, and interpret it as a model for residual life.

Solution: The backward transition probabilities are by definition given by $P^*_{ij} = \pi_j P_{ji}/\pi_i$. Since the steady state probabilities π_i are all positive, $P^*_{ij} > 0$ if and only if $P_{ji} > 0$. Thus the graph for the backward chain is the same as the forward chain except that all the arrows are reversed, and, of course, the labels are changed accordingly. The graph shows that there is only one transition coming out of each positive numbered state in the backward chain. Thus $P^*_{i,i-1} = 1$ for all $i > 0$. This can be easily verified algebraically from 6.27. It is also seen that $P^*_{i0} = \Pr\{W = i+1\}$.



We can interpret this Markov chain as representing residual life for an integer renewal process. The number of the state represents the residual life immediately before a transition, *i.e.*, state 0 means that a renewal will happen immediately, state 1 means that it will happen in one time unit, etc. When a renewal occurs, the next state will indicate the residual life at the end of that unit interval. Thus when a transition from 0 to i occurs, the corresponding renewal interval is $i + 1$.

For those confused by both comparing the chain above as the backward chain of age in Figure 6.3 and as the residual life chain of the same renewal process, consider a sample path for the renewal process in which an inter-renewal interval of 5 is followed by one of 2. The sample path for age in Figure 6.3 is then (0, 1, 2, 3, 4, 0, 1, 2, 0) The sample path for residual life for the same sample path of inter-renewals, for the interpretation above, is (0, 4, 3, 2, 1, 0, 2, 1, 0). On the other hand, the backward sample path for age is (0, 2, 1, 0, 4, 3, 2, 1, 0). In other words, the residual life sample path runs backward from age within each renewal interval, but forward between renewals.

A.7 Solutions for Chapter 7

Exercise 7.1: Consider an M/M/1 queue as represented in Figure 7.4. Assume throughout that $X_0 = i$ where $i > 0$. The purpose of this exercise is to understand the relationship between the holding interval until the next state transition and the interval until the next arrival to the M/M/1 queue. Your explanations in the following parts can and should be very brief.

a) Explain why the expected holding interval $\mathsf{E}[U_1|X_0 = i]$ until the next state transition is $1/(\lambda + \mu)$.

Solution: By definition of a countable state Markov process, U_n, conditional on X_{n-1} is an expontial. For the M/M/1 queue, the rate of the exponential out of state $i > 0$ is $\lambda + \mu$, and thus the expected interval U_1 is $1/(\lambda{+}\mu)$.

b) Explain why the expected holding interval U_1, conditional on $X_0 = i$ and $X_1 = i + 1$, is

$$\mathsf{E}[U_1|X_0 = i,\ X_1 = i + 1] = 1/(\lambda + \mu).$$

Show that $\mathsf{E}[U_1|X_0 = i, X_1 = i - 1]$ is the same.

Solution: The holding interval U_1, again by definition of a countable state Markov process, conditional on $X_0 = i > 0$, is independent of the next state X_1 (see Figure 7.1). Thus

$$\mathsf{E}[U_1|X_0 = i,\ X_1 = i + 1] = \mathsf{E}[U_1|X_0 = i] = \frac{1}{\lambda + \mu}.$$

This can be visualized by viewing the arrival process and the departure process (while the queue is busy) as independent Poisson processes of rate λ and μ respectively. From Section 2.3, on splitting and combining of Poisson processes, U_1 (the time until the first occurrence in the combined Poisson process) is independent of which split process (arrival or departure) that first occurrence comes from.

Since this result is quite unintuitive for most people, we explain it in yet another way. Quantizing time into very small increments of size δ, the probability of a customer arrival in each increment is $\lambda\delta$ and the probability of a customer departure (assuming the server is busy) is $\mu\delta$. This is the same for every increment and is independent of previous increments (so long as the server is busy). Thus X_1 (which is $X_0 + 1$ for an arrival and $X_0 - 1$ for a departure) is independent of the time of that first occurrence, Thus given $X_0 > 0$, the time of the next occurrence (U_1) is independent of X_1.

c) Let V be the time of the first arrival after time 0 (this may occur either before or after the time W of the first departure.) Show that

$$\mathsf{E}[V|X_0 = i,\ X_1 = i + 1] = \frac{1}{\lambda + \mu}. \tag{A.45}$$

$$\mathsf{E}[V|X_0 = i,\ X_1 = i - 1] = \frac{1}{\lambda + \mu} + \frac{1}{\lambda}. \tag{A.46}$$

Hint: In the second equation, use the memorylessness of the exponential rv and the fact that V under this condition is the time to the first departure plus the remaining time to an arrival.

Solution: Given $X_0 = i > 0$ and $X_1 = i{+}1$, the first transtion is an arrival, and from (b), the interval until that arrival is $1/(\lambda{+}\mu)$, verifying (A.45). Given $X_0 = i > 0$ and $X_1 = i{-}1$, the first transition is a departure, and from (b) the conditional expected time until this first transition is $1/(\lambda{+}\mu)$. The expected time after this first transition to an arrival is simply the expected interval until an arrival, starting at that first transition, verifying (A.46).

d) Use your solution to (c) plus the probability of moving up or down in the Markov chain to show that $\mathsf{E}[V] = 1/\lambda$. (Note: you already know that $\mathsf{E}[V] = 1/\lambda$. The purpose here is to show that your solution to (c) is consistent with that fact.)

Solution: This is a sanity check on (A.45) and (A.46). For $i > 0$,

$$\Pr\{X_1{=}i{+}1 \mid X_0{=}i\} = \frac{\lambda}{\lambda{+}\mu}; \qquad \Pr\{X_1{=}i{-}1 \mid X_0{=}i\} = \frac{\mu}{\lambda{+}\mu}.$$

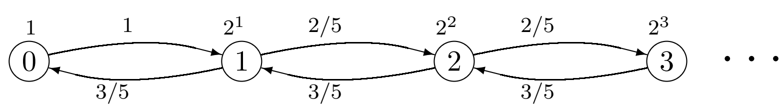
Using this with (A.45) and (A.46),

$$\begin{aligned}
\mathsf{E}[V \mid X_0{=}i] &= \frac{\lambda}{\lambda{+}\mu}\cdot\frac{1}{\lambda{+}\mu} + \frac{\mu}{\lambda{+}\mu}\left[\frac{1}{\lambda{+}\mu} + \frac{1}{\lambda}\right] \\
&= \frac{1}{\lambda{+}\mu} + \frac{\mu}{\lambda(\lambda{+}\mu)} \;=\; \frac{1}{\lambda}.
\end{aligned}$$

Exercise 7.2: Consider a Markov process for which the embedded Markov chain is a birth-death chain with transition probabilities $P_{i,i+1} = 2/5$ for all $i \geq 1$, $P_{i,i-1} = 3/5$ for all $i \geq 1$, $P_{01} = 1$, and $P_{ij} = 0$ otherwise.

a) Find the steady-state probabilities $\{\pi_i;\, i \geq 0\}$ for the embedded chain.

Solution: The embedded chain is given below. Note that it is the same embedded chain as in Example 7.2.8 and Exercise 7.3, but the process behavior will be very different when the holding times are brought into consideration.



The embedded chain is a birth/death chain, and thus the steady-state probabilities are related by $(2/5)\pi_i = (3/5)\pi_{i+1}$ for $i \geq 1$. Thus, as we have done many times, $\pi_i = \pi_1(2/3)^{i-1}$. The transitions between state 0 and 1 are different, and $\pi_1 = (5/3)\pi_0$. Thus for $i \geq 1$, we have

$$\pi_i \;=\; \frac{5}{3}\cdot\left(\frac{2}{3}\right)^{i-1}\pi_0 \;=\; \frac{5}{2}\cdot\left(\frac{2}{3}\right)^{i}\pi_0.$$

Setting the sum of π_0 plus $\sum_{i\geq 1}\pi_i$ to 1, we get $\pi_0 = 1/6$. Thus

$$\pi_i = (5/12)(2/3)^i \qquad \text{for } i \geq 1; \qquad \pi_0 = \frac{1}{6}. \tag{A.47}$$

b) Assume that the transition rate ν_i out of state i, for $i \geq 0$, is given by $\nu_i = 2^i$. Find the transition rates $\{q_{ij}\}$ between states and find the steady-state probabilities $\{p_i\}$ for the Markov process. Explain heuristically why $\pi_i \neq p_i$.

Solution: The transition rate q_{ij} is given by $\nu_i P_{ij}$. Thus for $i > 0$,

$$q_{i,i+1} = \frac{2}{5}\cdot 2^i; \qquad q_{i,i-1} = \frac{3}{5}\cdot 2^i; \qquad q_{01} = 1.$$

The simplest way to evaluate p_i is by (7.7), *i.e.*, $p_i = \pi_i/\big[\nu_i \sum_{j\geq 0} \pi_j/\nu_j\big]$.

$$\sum_{j\geq 0} \frac{\pi_j}{2^j} \;=\; \frac{1}{6} + \frac{5}{12}\sum_{i>0}\left(\frac{1}{3}\right)^i \;=\; \frac{3}{8}.$$

Thus $p_0 = 4/9$ and, for $i > 0$, $p_i = (10/9)3^{-i}$. Note that π_i is the steady-state fraction of *transitions* going into state i and p_i is the steady-state fraction of *time* in state i. Since $\nu_i^{-1} = 2^{-i}$ is the expected time-interval in state i per transition into state i, we would think (correctly) that p_i would approach 0 much more rapidly as $i \to \infty$ than π_i does.

c) Explain why there is no sampled-time approximation for this process. Then truncate the embedded chain to states 0 to m and find the steady-state probabilities for the sampled-time approximation to the truncated process.

Solution: In order for a sampled-time approximation to exist, one must be able to choose the time-increment δ small enough so that the conditional probability of a transition out of the state in that increment is small. Since the transition rates are unbounded, this cannot be done here. If we truncate the chain to states 0 to m, then the ν_i are bounded by 2^m, so choosing a time-increment less than 2^{-m} will allow a sampled-time approximation to exist.

There are two sensible ways to do this truncation. In both, $q_{m,m+1}$ must be changed to 0. but then $q_{m,m-1}$ can either be kept the same (thus reducing ν_m) or ν_m can be kept the same (thus increasing $q_{m,m-1}$)). We keep $q_{m,m-1}$ the same since it simplifies the answer sightly. Let $\{p_i^{(m)}; 0 \leq i \leq m\}$ be the steady-state process PMF in the truncated chain. Since these truncated chains (as well as the untruncated chain) are birth-death chains, we can use (7.38), $p_i^{(m)} = p_0^{(m)} \prod_{j<i} \rho_j$, where $\rho_j = q_{j,j+1}/q_{j+1,j}$. Thus, $\rho_0 = 5/6$ and $\rho_i = 1/3$ for $1 < i < m$. Thus

$$p_i^{(m)} = p_0^{(m)} \frac{5}{6}\left(\frac{1}{3}\right)^{i-1} \qquad \text{for } \; i \leq m.$$

Since $p_0^{(m)} + p_1^{(m)} + \cdots + p_m^{(m)} = 1$, we can solve for $p_0^{(m)}$ from

$$1 = p_0^{(m)}\left[1 + \frac{5}{6}\left(1 + 3^{-1} + 3^{-2} + \cdots 3^{-m+1}\right)\right] \;=\; p_0^{(m)}\left[1 + \frac{5}{4}(1 - 3^{-m})\right].$$

Combining these equations.

$$p_0^{(m)} \;=\; \frac{4}{9 - 5(3^{-m})}; \qquad p_i^{(m)} \;=\; \frac{10(3^{-i})}{9 - 5(3^{-m})} \quad \text{for } 1 \leq i \leq m. \tag{A.48}$$

For $\delta < 2^{-m}$, these are also the sampled-time 'approximation' to the truncated chain.

d) Show that as $m \to \infty$, the steady-state probabilities for the sequence of sampled-time approximations approach the probabilities p_i in (b).

Solution: For each i, we see from (A.48) that $\lim_{m\to\infty} p_i^{(m)} \;=\; p_i$. It is important to recognize that the convergence is not uniform in i since (A.48) is defined only for $i \leq m$. Similarly, for each i, the approximation is close only when both $m \geq i$ and δ is small relative to 3^{-m}.

Exercise 7.3: Consider a Markov process for which the embedded Markov chain is a birth-death chain with transition probabilities $P_{i,i+1} = 2/5$ for all $i \geq 1$, $P_{i,i-1} = 3/5$ for all $i \geq 1$, $P_{01} = 1$, and $P_{ij} = 0$ otherwise.

a) Find the steady-state probabilities $\{\pi_i;\, i \geq 0\}$ for the embedded chain.

Solution: This is the same embedded chain as in Exercise 7.2. The steady-state embedded probabilities were calculated there in (A.47) to be

$$\pi_i = (5/12)(2/3)^i \qquad \text{for } i \geq 1; \qquad \pi_0 = \frac{1}{6}.$$

b) Assume that the transition rate out of state i, for $i \geq 0$, is given by $\nu_i = 2^{-i}$. Find the transition rates $\{q_{ij}\}$ between states and show that there is no probability vector solution $\{p_i;\, i \geq 0\}$ to (7.23).

Solution: This particular Markov process was discussed in Example 7.2.8. The transition rates q_{ij} are given by $q_{01} = 1$ and, for all $i > 0$,

$$q_{i,i+1} = P_{i,i+1}\nu_i = \frac{2}{5} \cdot 2^{-i}; \qquad P_{i,i-1}\nu_i = \frac{3}{5} \cdot 2^{-i}.$$

The other transition rates are all 0. We are to show that there is no probability vector $\{p_i :\, i \geq 0\}$ solution to (7.23), *i.e.*, no solution to

$$p_j\nu_j = \sum_i p_i q_{ij} \quad \text{for all } j \geq 0 \quad \text{with } \sum_i p_i = 1. \tag{7.23$'$}$$

Let $\alpha_j = p_j\nu_j$, so that any hypothesized solution to (7.23$'$) becomes $\alpha_j = \sum_i \alpha_i P_{ij}$ for all j and $\sum_i \alpha_i/\nu_i = 1$.

Since the embedded chain is positive recurrent, there is a unique solution to $\{\pi_i;\, i \geq 0\}$ such that $\pi_j = \sum_i \pi_i P_{ij}$ for all j and $\sum_i \pi_i = 1$. Thus there must be some $0 < \beta < \infty$ such that any hypothesized solution to (7.23$'$) satisfies $\alpha_j = \beta\pi_j$ for all $j \geq 0$. Thus $\sum_i \pi_j/\nu_j < \infty$ for this hypothesized solution.

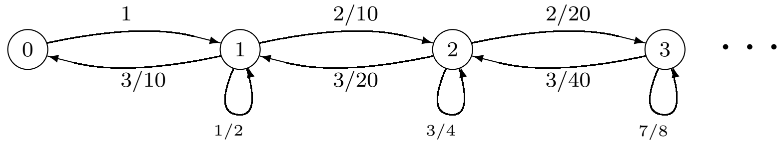
Now note that $\pi_j/\nu_j = (5/12)(4/3)^j$. Since this increases exponentially with j, we must have $\sum_j \pi_j/\nu_j = \infty$. Thus there cannot be a probability vector solution to (7.23). We discuss this further after (c) and (d).

c) Argue that the expected time between visits to any given state i is infinite. Find the expected number of transitions between visits to any given state i. Argue that, starting from any state i, an eventual return to state i occurs with probability 1.

Solution: From Theorem 7.2.6 (which applies since the embedded chain is positive recurrent), the expected time between returns to state i is $\overline{W}(i) = (1/\pi_i)\sum_k \pi_k/\nu_k$. Since $\sum_k \pi_k/\nu_k$ is infinite and π_i is positive for all i, $\overline{W}(i) = \infty$. As explained in Section 7.2, $W(i)$ is a rv(*i.e.*, non-defective), and is thus finite with probability 1 (guaranteeing an eventual return with probability 1). Under the circumstances here, however, it is a rv with infinite expectation for each i. The expected number of transitions between returns to state i is finite (since the embedded chain is positive recurrent), but the increasingly long intervals spent in high numbered states causes the expected renewal time from i to i to be infinite.

d) Consider the sampled-time approximation of this process with $\delta = 1$. Draw the graph of the resulting Markov chain and argue why it must be null recurrent.

Solution:



Note that the approximation is not very good at the low-numbered states, but it gets better and better for higher numbered states since ν_i becomes increasingly negligible compared to $\delta = 1$ and thus there is negligible probability of more than one arrival or departure in a unit increment. Thus it is intuitively convincing that the mean interval between returns to any given state i must be infinite, but that a return must happen eventually with probability 1. This is a convincing argument why the chain is null recurrent.

At the same time, the probability of an up-transition from i to $i+1$ is 4/3 of the probability of a down-transition from $i+1$ to i, makng the chain look like the classical example of a transient chain in Figure 6.2. Thus there is a need for a more convincing argument.

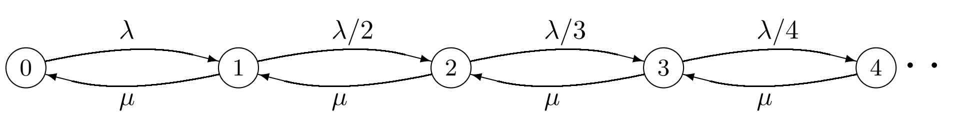
One could truncate both the process and the sample-time chain to states 0 to m and then go to the limit as $m \to \infty$, but this would be very tedious.

We next outline a procedure that is mathematically rigorous to show that the chain above is null recurrent; the procedure can also be useful in other circumstances. Each sample sequence of states for the chain above consists of runs of the same state separated on each side by a state either one larger or one smaller. Consider creating a new Markov chain by replacing each such run of repeated states with a single copy of that state. It is easy to see that this new chain is Markov. Also since there can be no repetition of a state, the new transition probabilities, say $Q_{i,i+1}$ and $Q_{i,i-1}$ for $i > 0$ satisfy $Q_{i,i+1} = P_{i,i+1}/(P_{i,i+1} + P_{i,i-1}) = 2/5$ and $Q_{i,i-1} = 3/5$. Thus this new chain is the same as the embedded chain of the original process, which is already known to be positive recurrent.

At this point, we can repeat the argument in Section 7.2.2, merely replacing the exponentially distributed reward interval in Figure 7.7 with a geometrically distributed interval. The expected first passage time from i to i (in the approximation chain) is then infinite as before, and the return occurs eventually with probability 1. Thus the approximation chain is null recurrent.

The nice thing about the above procedure is that it can be applied to any birth death chain with self transitions.

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.



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 Markov process are related by $p_{i+1}q_{i+1,i} = p_i q_{i,i+1}$, which in this case is $p_{i+1}\mu = p_i\lambda/(i+1)$. Iterating this equation,

$$p_i = p_{i-1}\frac{\lambda}{\mu i} = p_{i-2}\frac{\lambda^2}{\mu^2 i(i-1)} = \cdots = p_0\frac{\lambda^i}{\mu^i i!}.$$

Denoting λ/μ by ρ,

$$1 = \sum_{i=0}^{\infty} p_i = p_0\left[\sum_{i=0}^{\infty}\frac{\rho^i}{i!}\right] = p_0 e^{\rho}.$$

Thus,

$$p_0 = e^{-\rho}; \qquad p_i = \frac{\rho^i e^{-\rho}}{i!}.$$

b) Find a closed form solution to $\sum_i p_i\nu_i$ where ν_i is the rate at which transitions out of state i occur. Show that the embedded chain is positive recurrent for all choices of $\lambda > 0$ and $\mu > 0$ and explain intuitively why this must be so.

Solution: The embedded chain steady-state probabilities π_i can be found from the steady-state process probabilities by (7.11), *i.e.*,

$$\pi_j = \frac{p_j\nu_j}{\sum_i \pi_i\nu_i} \qquad (7.11').$$

We start by finding $\sum_i p_i\nu_i$. The departure rate from state i is

$$\nu_0 = \lambda; \qquad \nu_i = \mu + \frac{\lambda}{i+1} \quad \text{for all } i > 0.$$

We now calculate $\sum_i p_i\nu_i$ by separating out the $i = 0$ term and then, for $i \geq 1$, sum separately over the two terms, μ and $\lambda/(i+1)$, of ν_i.

$$\sum_{i=0}^{\infty} p_i\nu_i = e^{-\rho}\lambda + \sum_{i=1}^{\infty} e^{-\rho}\frac{\rho^i\mu}{i!} + \sum_{i=1}^{\infty} e^{-\rho}\frac{\rho^i\lambda}{i!(i+1)}.$$

Substituting $\mu\rho$ for λ and combining the first and third term,

$$\begin{aligned}\sum_{i=0}^{\infty} p_i\nu_i &= \sum_{i=1}^{\infty} e^{-\rho}\frac{\rho^i\mu}{i!} + \sum_{i=0}^{\infty} e^{-\rho}\frac{\rho^{i+1}\mu}{(i+1)!} \\ &= 2\sum_{i=1}^{\infty} e^{-\rho}\frac{\rho^i\mu}{i!} = 2\mu(1-e^{-\rho}).\end{aligned}$$

Since $\sum_i p_i\nu_i < \infty$, we see from (7.11′) that each π_i is strictly positive and that $\sum_i \pi_i = 1$. Thus the embedded chain is positive recurrent. Intuitively, $\{\pi_i;\, i \geq 0\}$ is found from $\{p_i;\, i \geq 0\}$ by finding $\pi_i' = p_i\nu_i$ and then normalizing $\{\pi_i';\, i \geq 0\}$ to sum to 1. Since $\lambda \leq \nu_i \leq \lambda + \mu$, *i.e.*, the ν_i all lie within positive bounds, this normalization must work (*i.e.*, $\sum_i \pi_i\nu_i < \infty$).

c) For the embedded Markov chain corresponding to this process, find the steady-state probabilities π_i for each $i \geq 0$ and the transition probabilities P_{ij} for each i, j.

Solution: Since $\pi_j = p_j\nu_j / \big[\sum_i p_i\nu_i\big]$, we simply plug in the values for p_i, ν_i, and $\sum_i p_i\nu_i$ found in (a) and (b) to get

$$\pi_0 = \frac{\rho}{2(e^\rho - 1)}; \qquad \pi_i = \frac{\rho^i}{2i!(e^\rho - 1)}\left(\frac{\rho}{i+1} + 1\right); \quad \text{for } i > 1.$$

There are many forms for this answer. One sanity check is to observe that the embedded chain probabilities do not change if λ and μ are both multiplied by the same constant, and thus the π_i must be a function of ρ 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 i is $\overline{W}_i = 1/(p_i\nu_i)$. This is explained in detail in Section 7.2.6, but the essence of the result is that for renewals at successive entries to state i, p_i must be the ratio of the expected time $1/\nu_i$ spent in state i to the overall expected renewal interval $\overline{W}_i$. Thus $\overline{W}_i = 1/(\nu_i\rho_i)$.

$$\overline{W}_0 = \frac{e^\rho}{\lambda}; \qquad \overline{W}_i = \frac{(i+1)!\, e^\rho}{\rho^i[\lambda + (i+1)\mu]}; \qquad \text{for } i \geq 1.$$

The time-average number of state transitions per visit to state i is $\overline{T}_{ii} = 1/\pi_i$. This is proven in Theorem 6.3.8.

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 $i \geq 0$.

Solution: The process is a birth/death process, so we can find the steady-state probabilities (if they exist) from the equations $p_i q_{i,i+1} = p_{i+1} q_{i+1,i}$ for $i \geq 0$. Thus $p_{i+1} = p_i/2$. Normalizing to $\sum_i p_i = 1$, we get $p_i = 2^{-i-1}$.

b) Find the transition probabilities for the embedded Markov chain and show that the chain is null recurrent.

Solution: First assume, for the purpose of finding a contradiction, that the embedded Markov chain is positive recurrent. Then by (7.21), $\pi_j = p_j\nu_j / \sum_i p_i\nu_i$. Note that $\nu_i = q_{i,i+1} + q_{i,i-1} = 2^i$ for all $i \geq 1$. Thus $p_i\nu_i = 1/2$ for $i \geq 1$ and $\sum_i p_i\nu_i = \infty$. Thus $\pi_j = 0$ for $j \geq 0$ and the chain must be either null recurrent or transient. We show that the chain is null recurrent in (c)

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.

Solution: We use the same argument as in Section 7.2.2 with unit reward when the process is in state i. The limiting fraction of time in state i given that $X_0 = i$ is then $p_i(i) =$

$1/(\nu_i\overline{W}(i))$ where $\overline{W}(i)$ is the mean renewal time between entries to state i. This is the same for all starting states, and by Blackwell's theorem it is also $\lim_{t\to\infty}\Pr\{X(t)=i\}$. This $p_i(i)$ must also satisfy the steady-state process equations and thus be equal to $p_i = 2^{-i-1}$. Since $\nu_i = 2^i$, we have $\overline{W}(i) = 2$ for all $i \geq 1$. Finally, this means that a return to i must happen in a finite number of transitions. Since the embedded chain cannot be positive recurrent, it thus must be null recurrent.

The argument here has been a little tricky since the development in the text usually assumes that the embedded chain is positive recurrent, but the use of renewal theory above gets around that.

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 (b) and (c) that the embedded chain is null-recurrent. This means that, given $X_0 = i$, for any given i, a return to i must happen in a finite number of transitions (i.e., $\lim_{n\to\infty}\mathsf{F}_{ii}(n) = 1$), but the expected number of such transitions is infinite. We have seen many rv's that have an infinite expectation, but, being rv's, have a finite sample value WP1.

Exercise 7.10: **a)** Consider the two state Markov process of Example 7.3.1 with $q_{01} = \lambda$ and $q_{10} = \mu$. Find the eigenvalues and eigenvectors of the transition rate matrix $[Q]$.

Solution: The matrix $[Q]$ is $\begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$. For any transition rate matrix, the rows all sum to 0, and thus $[Q]\boldsymbol{e} = 0$, establishing that 0 is an eigenvalue with the right eigenvector $\boldsymbol{e} = (1,1)^{\mathsf{T}}$. The left eigenvector, normalized to be a probability vector, is the steady-state vector $\boldsymbol{p} = (\mu/(\lambda+\mu), \lambda/(\lambda+\mu)$. The other eigenvalue is then easily calculated as $-(\lambda+\mu)$ with left eigenvector $(-1,1)$ and right eigenvector $(-\lambda/(\lambda+\mu), \mu/(\lambda+\mu))$. This answer is not unique, since the eigenvectors can be scaled differently while still maintaining $\boldsymbol{p}_i\boldsymbol{v}_j^{\mathsf{T}} = \delta_{ij}$.

b) If $[Q]$ has M distinct eigenvalues, the differential equation $d[P(t)]/dt = [Q][P(t)]$ can be solved by the equation

$$[P(t)] = \sum_{i=1}^{\mathsf{M}} \boldsymbol{\nu}_i e^{t\lambda_i}\boldsymbol{p}_i^{\mathsf{T}},$$

where $\boldsymbol{p}_i$ and $\boldsymbol{\nu}_i$ are the left and right eigenvectors of eigenvalue λ_i. Show that this equation gives the same solution as that given for Example 7.3.1.

Solution: This follows from substituting the values above. Note that the eigenvectors above are the same as those for the sample time approximations and the eigenvalues are related as explained in Example 7.3.1.

Exercise 7.13: **a)** Consider an M/M/1 queue in steady state. Assume $\rho = \lambda/\mu < 1$. Find the probability $Q(i,j)$ for $i \geq j > 0$ that the system is in state i at time t and that $i-j$ departures occur before the next arrival.

Solution: The probability that the queue is in state i at time t is $p_i = (1-\rho)\rho^i$ (see (7.40)). Given state $i > 0$, successive departures (so long as the state remains positive) are Poisson at rate μ and arrivals are Poisson at rate λ. Thus (conditional on $X(t) = i$) the probability of exactly $i-j$ departures before the next arrival is $\big(\mu/(\lambda+\mu)\big)^{i-j}\lambda/(\lambda+\mu)$. Thus

$$Q(i,j) = (1-\rho)\rho^i\left(\frac{\mu}{\lambda+\mu}\right)^{i-j}\frac{\lambda}{\lambda+\mu}.$$

b) Find the PMF of the state immediately before the first arrival after time t.

Solution: $Q(i,j)$ is the probability that $X(t)=i$ and that $X(\tau^-)=j$ where τ is the time of the next arrival after t and $j>0$. Thus for $j>0$,

$$\begin{aligned}
\Pr\big\{X(\tau^-){=}j\big\} &= \sum_{i\geq j} Q(i,j) \;=\; \sum_{i\geq j}(1-\rho)\left(\frac{\lambda}{\lambda{+}\mu}\right)^i\left(\frac{\mu}{\lambda{+}\mu}\right)^{-j}\frac{\lambda}{\lambda{+}\mu}\\
&= (1-\rho)\left(\frac{\lambda}{\mu}\right)^j\left(\frac{1}{1-\lambda/(\lambda{+}\mu)}\right)\left(\frac{\lambda}{\lambda{+}\mu}\right) \;=\; (1-\rho)\rho^{j+1}.
\end{aligned}$$

For $j=0$, a similar but simpler calculation leads to $\Pr\{X(\tau^-)=0\}=(1-\rho)(1+\rho)$. In other words, the system is not in steady state immediately before the next arrival. This is not surprising, since customers can depart but not arrive in the interval (t,τ)

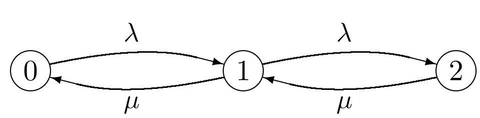
c) There is a well-known queueing principle called PASTA, standing for "Poisson arrivals see time averages". Given your results above, give a more precise statement of what that principle means in the case of the M/M/1 queue.

Solution: The PASTA principle requires being so careful about a precise statement of the conditions under which it holds that it is better treated as an hypothesis to be considered rather than a principle. One plausible meaning for what arrivals 'see' is given in (b), and that is not steady state. Another plausible meaning is to look at the fraction of arrivals that arise from each state; that fraction is the steady-state probability. Perhaps the simplest visualization of PASTA is to look at the discrete-time model of the M/M/1 queue. There the steady-state fraction of arrivals that come from state i is equal to the steady-state probability p_i.

Exercise 7.14: A small bookie shop has room for at most two customers. Potential customers arrive at a Poisson rate of 10 customers per hour; they enter if there is room and are turned away, never to return, otherwise. The bookie serves the admitted customers in order, requiring an exponentially distributed time of mean 4 minutes per customer.

a) Find the steady-state distribution of number of customers in the shop.

Solution: The system can be modeled as a Markov process with 3 states, representing 0, 1, or 2 customers in the system. Arrivals in state 2 are turned away, so they do not change the state and are not shown. In transitions per hour, $\lambda=10$ and $\mu=15$. Since the process is a birth-death process, the steady-state equations are $\lambda p_i=\mu p_{i+1}$ for $i=0,1$. Thus $p_1=(2/3)p_0$ and $p_2=4/9p_0$. Normalizing, $p_0=9/19$, $p_1=6/19$, $p_2=4/19$.



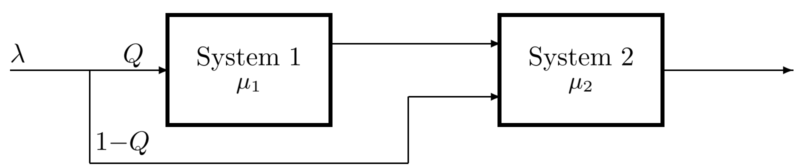
b) Find the rate at which potential customers are turned away.

Solution: All arrivals in state 2 are turned away, so the average rate at which customers are turned away is 40/19 per hour.

c) Suppose the bookie hires an assistant; the bookie and assistant, working together, now serve each customer in an exponentially distributed time of mean 2 minutes, but there is only room for one customer (*i.e.*, the customer being served) in the shop. Find the new rate at which customers are turned away.

Solution: Now $\lambda = 10$ and $\mu = 30$. There are two states, with $p_0 = 3/4$ and $p_1 = 1/4$. Customers are turned away at rate 10/4, which is somewhat higher than the rate without the assistant.

Exercise 7.16: Consider the job sharing computer system illustrated below. Incoming jobs arrive from the left in a Poisson stream. Each job, independently of other jobs, requires pre-processing in system 1 with probability Q. Jobs in system 1 are served FCFS and the service times for successive jobs entering system 1 are IID with an exponential distribution of mean $1/\mu_1$. The jobs entering system 2 are also served FCFS and successive service times are IID with an exponential distribution of mean $1/\mu_2$. The service times in the two systems are independent of each other and of the arrival times. Assume that $\mu_1 > \lambda Q$ and that $\mu_2 > \lambda$. Assume that the combined system is in steady state.



a) Is the input to system 1 Poisson? Explain.

Solution: The input to system 1 is the splitting of a Poisson process and is thus Poisson of rate λQ.

b) Are each of the two input processes coming into system 2 Poisson? Explain.

Solution: The output of system 1 is Poisson of rate λQ by Burke's theorem. The other process entering system 2 is also Poisson with rate $(1 - Q)\lambda$, since it is a splitting of the original Poisson process of rate λ. This input to system 2 is independent of the departures from system 1 (since it is independent both of the arrivals and departures from system 1.) Thus the combined process into system 2 is Poisson with rate λ

c) Give the joint steady-state PMF of the number of jobs in the two systems. Explain briefly.

Solution: The state $X_2(t)$ of system 2 at time t is dependent on the inputs to system 2 up to time t and the services in system 2 up to time t. By Burke's theorem, the outputs from system 1 at times up to time t are independent of the state $X_1(t)$ of system 1 at time t, and thus the inputs to system 2 from system 1 at times up to t are independent of $X_1(t)$. The other input to system 2 is also independent entirely of system 1. Thus $X_1(t)$ and $X_2(t)$ are independent. Thus $X_1(t)$ and $X_2(t)$ are independent and are the states of M/M/1 queues in steady state at time t.

$$\Pr\{X_1(t), X_2(t) = i, j\} = (1 - \rho_1)\rho_1^i(1 - \rho_2)\rho_2^j,$$

where $\rho_1 = \lambda Q/\mu_1$ and $\rho_2 = \lambda/\mu_2$. Note that this independence applies to the states of system 1 and 2 at the same instant t. The processes are not independent, and, for example, $X_1(t)$ and $X_2(t + \tau)$ for $\tau > 0$ are not independent.

d) What is the probability that the first job to leave system 1 after time t is the same as the first job that entered the entire system after time t?

Solution: The first job out of system 1 after t is the same as the first job into system 1 after t if and only if system 1 is empty at time t. This is also the same as the first job into the overall system if the first arrival to the entire system goes into system 1. The probability of both of these events (which are independent) is thus $P_1 = Q(1-\rho_1)$.

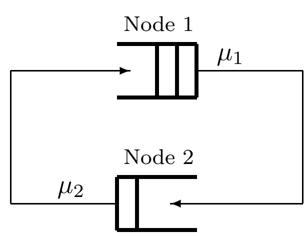
e) What is the probability that the first job to leave system 2 after time t both passed through system 1 and arrived at system 1 after time t.

Solution: Let P_2 be the probability of this event. This requires, first, that the event in (d) is satisfied, second that system 2 is empty at time t, and, third, that the first job to bypass sytem 1 after the first arrival to system 1 occurs after the service of that first arrival.

This third event is the event that an exponential rv of rate μ_1 has a smaller sample value than one of rate $\lambda(1-Q)$. The probability of this is $\mu_1/\big(\mu_1 + \lambda(1-Q)\big)$. Thus

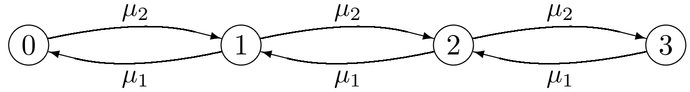
$$P_2 = P_1(1-\rho_2)\frac{\mu_1}{\mu_1 + \lambda(1-Q)}.$$

Exercise 7.19: Consider the closed queueing network in the figure below. There are three customers who are doomed forever to cycle between node 1 and node 2. Both nodes use FCFS service and have exponentially distributed IID service times. The service times at one node are also independent of those at the other node and are independent of the customer being served. The server at node i has mean service time $1/\mu_i$, $i = 1, 2$. Assume to be specific that $\mu_2 < \mu_1$.



a) The system can be represented by a four state Markov process. Draw its graphical representation and label it with the individual states and the transition rates between them.

Solution: The three customers are statistically identical, so we can take the number of customers in node 1 (which can be 0, 1, 2, 3) to be the state. In states 1, 2, 3, departures from node 1 take place at rate μ_1. In states 0, 1, 2, departures occur from node 2 at rate μ_2, and these serve as arrivals to node 1. Thus the process has the following graphical representation.



b) Find the steady-state probability of each state.

Solution: Perhaps surprisingly, this is the same as an M/M/1 queue in which arrivals are turned away when $X(t) = 3$. In the queue representation, we lose the identity of the three customers, but we need not keep track of them since they are statistically identical. As we see in the rest of the exercise, the customer identities can be tracked supplementally, since the arrivals to node 1 rotate from customer 1 to 2 to 3. Thus each third arrival in the queue representation corresponds to the same customer.

Let $\rho = \mu_2/\mu_1$. Then the steady-state process probabilities are $p_1 = p_0\rho$; $p_2 = \rho^2 p_0$, $p_3 = \rho^3 p_0$, where $p_0 = 1/(1+\rho+\rho^2+\rho^3)$.

c) Find the time-average rate at which customers leave node 1.

Solution: Each customer departure from node 1 corresponds to a downward transition from the queue representation, and thus occurs at rate μ_1 from each state except 0 of the queue. Thus in steady state, customers leave node 1 at rate $r = (1-p_0)\mu_1$. This is also the time-average rate at which customers leave node 1.

d) Find the time-average rate at which a given customer cycles through the system.

Solution: Each third departure from node 1 is a departure of the same customer, and this corresponds to each third downward transition in the queue. Thus the departure rate of a given customer from queue 1, which is the same as the rotation rate of that customer, is $r/3$.

e) Is the Markov process reversible? Suppose that the backward Markov process is interpreted as a closed queueing network. What does a departure from node 1 in the forward process correspond to in the backward process? Can the transitions of a single customer in the forward process be associated with transitions of a single customer in the backward process?

Solution: The queueing process is reversible since it is a birth-death process. The backward process from the two node system, however, views each forward departure from node 1 as an arrival to node 1, *i.e.*, a departure from node 2. Customers 1, 2, 3 then rotate in backward order, 1, 3, 2, in the backward system. Thus if one wants to maintain the customer order, the system is not reversible. The system, viewed as a closed queueing system in the sense of Section 7.7.1, is in essence the same as the queueing network here and also abstracts away the individual customer identities, so it is again reversible.

A.8 Solutions for Chapter 8

Exercise 8.1: In this exercise, we evaluate $\Pr\{e_\eta \mid X = \boldsymbol{a}\}$ and $\Pr\{e_\eta \mid X = \boldsymbol{b}\}$ for binary detection from vector signals in Gaussian noise directly from (8.40) and(8.41).

a) By using (8.40) for each sample value $\boldsymbol{y}$ of $\boldsymbol{Y}$, show that

$$\mathsf{E}\left[\mathrm{LLR}(\boldsymbol{Y}) \mid X{=}\boldsymbol{a}\right] = \frac{-(\boldsymbol{b}-\boldsymbol{a})^{\mathsf{T}}(\boldsymbol{b}-\boldsymbol{a})}{2\sigma^2}.$$

Hint: Note that, given $X = \boldsymbol{a}$, $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{Z}$.

Solution: Taking the expectation of (8.40) conditional on $X = \boldsymbol{a}$,

$$\begin{aligned}\mathsf{E}\left[\mathrm{LLR}(\boldsymbol{Y}) \mid X{=}\boldsymbol{a}\right] &= \frac{(\boldsymbol{b}-\boldsymbol{a})^{\mathsf{T}}}{\sigma^2}\mathsf{E}\left[\boldsymbol{Y} - \frac{\boldsymbol{b}+\boldsymbol{a}}{2}\right] \\ &= \frac{(\boldsymbol{b}-\boldsymbol{a})^{\mathsf{T}}}{\sigma^2}\left(\boldsymbol{a} - \frac{\boldsymbol{b}+\boldsymbol{a}}{2}\right),\end{aligned}$$

from which the desired result is obvious.

b) Defining $\gamma = \|\boldsymbol{b}-\boldsymbol{a}\|/(2\sigma)$, show that

$$\mathsf{E}\left[\mathrm{LLR}(\boldsymbol{Y}) \mid X{=}\boldsymbol{a}\right] = -2\gamma^2.$$

Solution: The result in (a) can be expressed as $\mathsf{E}\left[\mathrm{LLR}(\boldsymbol{Y})\right] = -\|\boldsymbol{b}-\boldsymbol{a}\|^2/2\sigma^2$, from which the result follows.

c) Show that

$$\mathsf{VAR}\left[\mathrm{LLR}(\boldsymbol{Y}) \,|\, X{=}\boldsymbol{a}\right] = 4\gamma^2.$$

Hint: Note that the fluctuation of $\mathrm{LLR}(\boldsymbol{Y})$ conditional on $X = \boldsymbol{a}$ is $(1/\sigma^2)(\boldsymbol{b}-\boldsymbol{a})^{\mathsf{T}}\boldsymbol{Z}$.

Solution: Using the hint,

$$\begin{aligned}\mathsf{VAR}\left[\mathrm{LLR}(\boldsymbol{Y})\right] &= \frac{1}{\sigma^2}(\boldsymbol{b}-\boldsymbol{a})^{\mathsf{T}}\mathsf{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}\right](\boldsymbol{b}-\boldsymbol{a})\frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2}(\boldsymbol{b}-\boldsymbol{a})^{\mathsf{T}}[I](\boldsymbol{b}-\boldsymbol{a}) \;=\; \frac{1}{\sigma^2}\|\boldsymbol{b}-\boldsymbol{a}\|^2,\end{aligned}$$

from which the result follows.

d) Show that, conditional on $X = \boldsymbol{a}$, $\mathrm{LLR}(\boldsymbol{Y}) \sim \mathcal{N}(-2\gamma^2, 4\gamma^2)$. Show that, conditional on $X = \boldsymbol{a}$, $\mathrm{LLR}(\boldsymbol{Y})/2\gamma \sim \mathcal{N}(-\gamma, 1)$.

Solution: Conditional on $X = \boldsymbol{a}$, we see that $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{Z}$ is Gaussian and thus $\mathrm{LLR}(\boldsymbol{Y})$ is also Gaussian conditional on $X = \boldsymbol{a}$. Using the conditional mean and variance of $\mathrm{LLR}(\boldsymbol{Y})$ found in (b) and (c), $\mathrm{LLR}(\boldsymbol{Y}) \sim \mathcal{N}(-2\gamma^2, 4\gamma^2)$.

When the rv $\mathrm{LLR}(\boldsymbol{Y})$ is divided by 2γ, the conditional mean is also divided by 2γ, and the variance is divided by $(2\gamma)^2$, leading to the desired result.

Note that $\mathrm{LLR}(\boldsymbol{Y})/2\gamma$ is very different from $\mathrm{LLR}(\boldsymbol{Y}/2\gamma)$. The first scales the LLR and the second scales the observation $\boldsymbol{Y}$. If the observation itself is scaled, the result is a sufficient statistic and the LLR is unchanged by the scaling.

e) Show that the first half of (8.44) is valid, *i.e.*, that

$$\Pr\{e_\eta \mid X{=}\boldsymbol{a}\} = \Pr\{\mathrm{LLR}(\boldsymbol{Y}) \geq \ln\eta \mid X{=}\boldsymbol{a}\} = Q\left(\frac{\ln\eta}{2\gamma} + \gamma\right).$$

Solution: The first equality above is simply the result of a threshold test with the threshold η. The second uses the fact in (d) that $\mathrm{LLR}(\boldsymbol{Y})/2\gamma$, conditional on $X = \boldsymbol{a}$, is $\mathcal{N}(-\gamma, 1$. This is a unit variance Gaussian rv with mean $-\gamma$. The probability that it exceeds $\ln\eta/2\gamma$ is then $Q\big(\ln(\eta/2\gamma + \gamma\big)$.

f) By essentially repeating (a) through (e), show that the second half of (8.44) is valid, *i.e.*, that

$$\Pr\{e_\eta \mid X{=}\boldsymbol{b}\} = Q\left(\frac{-\ln\eta}{2\gamma} + \gamma\right).$$

Solution: One can simply rewrite each equation above, but care is needed in observing that the likelihood ratio requires a convention for which hypothesis goes on top of the fraction. Thus, here the sign of the LLR is opposite to that in parts (a) to (e). This also means that the error event occurs on the opposite side of the threshold.

Exercise 8.3: **a)** Let $\boldsymbol{Y}$ be the observation rv for a binary detection problem, let $\boldsymbol{y}$ be the observed sample value. Let $v = v(\boldsymbol{y})$ be a sufficient statistic and let V be the corresponding random variable. Show that $\Lambda(\boldsymbol{y})$ is equal to $\mathsf{p}_{V|X}(v(\boldsymbol{y}) \mid \boldsymbol{b})/\mathsf{p}_{V|X}(v(\boldsymbol{y}) \mid \boldsymbol{a})$. In other words, show that the likelihood ratio of a sufficient statistic is the same as the likelihood ratio of the original observation.

Solution: This is the third statement of Theorem 8.2.8, and we don't see any way of improving on that proof. It relies on the second statement of the theorem, which is further investigated in Exercise 8.4

b) Show that this also holds for the ratio of probability densities if V is a continuous rv or random vector.

Solution: Repeating the proof of the third statement of Theorem 8.2.8 for the case in which $\boldsymbol{Y}$ and V have densities, we start with the second statement, *i.e.*, $\mathsf{p}_{X|Y}(x|\boldsymbol{y}) = \mathsf{p}_{X|V}(x|v(\boldsymbol{y}))$.

$$\begin{aligned}
\frac{\mathsf{p}_{X|\boldsymbol{Y}}(1|\boldsymbol{y})}{\mathsf{p}_{X|\boldsymbol{Y}}(0|\boldsymbol{y})} &= \frac{\mathsf{p}_{X|V}(1|v(\boldsymbol{y}))}{\mathsf{p}_{X|V}(0|v(\boldsymbol{y}))} \\
\frac{\mathsf{f}_{\boldsymbol{Y}|X}(\boldsymbol{y}|1)\mathsf{p}_X(1)/\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y})}{\mathsf{f}_{\boldsymbol{Y}|X}(\boldsymbol{y}|0)\mathsf{p}_X(0)/\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y}))} &= \frac{\mathsf{f}_{V|X}(v(\boldsymbol{y})|1)\mathsf{p}_X(1)/\mathsf{f}_V(v(\boldsymbol{y})}{\mathsf{f}_{V|X}(v(\boldsymbol{y})|0)\mathsf{p}_X(0)/\mathsf{f}_V(v(\boldsymbol{y})},
\end{aligned}$$

Where we have used Bayes' rule on each term. Cancelling terms, we get

$$\frac{\mathsf{f}_{\boldsymbol{Y}|X}(\boldsymbol{y}|1)}{\mathsf{f}_{\boldsymbol{Y}|X}(\boldsymbol{y}|0)} = \frac{\mathsf{f}_{V|X}(v(\boldsymbol{y})|1)}{\mathsf{f}_{V|X}(v(\boldsymbol{y})|0)}.$$

The left side of this is $\Lambda(\boldsymbol{y})$, so this is the desired result. We see that it is derived simply by replacing PMF's with PDF's, but there are some assumptions made about the densities being positive. Turning this into a mathematical theorem with precise conditions would require measure theory, and without that, it is better to rely on common sense applied to simple models.

Exercise 8.4:

a) Show that if $v(\boldsymbol{y})$ is a sufficient statistic according to condition 1 of Theorem 8.2.8, then

$$p_{X|YV}(x \mid \boldsymbol{y}, v(\boldsymbol{y})) = p_{X|Y}(x \mid \boldsymbol{y}). \tag{A.49}$$

Solution: Let $V(\boldsymbol{Y})$ be the rv with sample values $v(\boldsymbol{y})$. Note that $\{\boldsymbol{Y} = \boldsymbol{y}\}$ is the same event as $\{\boldsymbol{Y} = \boldsymbol{y}\} \bigcap \{V = v(\boldsymbol{y})\}$. If $\boldsymbol{Y}$ is discrete and this event has positive probability, then (A.49) is obvious since the condition on both sides is the same and has positive probability. If Y is a rv with a positive density, then (A.49) is true if the condition $Y = y$ is replaced with $y - \delta < Y \le y$. Then (A.49) holds if $\lim_{\delta\to 0}(1/\delta)\Pr\{y - \delta < Y \le y\} > 0$, which is valid since Y has a positive density. This type of argument can be extended to the case where Y is a random vector and holds whether or not $V(\boldsymbol{Y})$ is a sufficient statistic. This says that $X \to \boldsymbol{Y} \to V$ is Markov, which is not surprising since V is simply a function of $\boldsymbol{Y}$.

b) Consider the subspace of events conditional on $V(\boldsymbol{y}) = v$ for a given v. Show that for $\boldsymbol{y}$ such that $v(\boldsymbol{y}) = v$,

$$p_{X|YV}(x \mid \boldsymbol{y}, v(\boldsymbol{y})) = p_{X|V}(x \mid v). \tag{A.50}$$

Solution: We must assume (as a natural extension of (a)), that $V(\boldsymbol{Y})$ is a sufficient statistic, *i.e.*, that there is a function u such that for each v, $u(v) = \Lambda(\boldsymbol{y})$ for all $\boldsymbol{y}$ such that $v(\boldsymbol{y}) = v$. We also assume that $\boldsymbol{Y} = \boldsymbol{y}$ has positive probabiity or probability density, since the conditional probabilities don't have much meaning otherwise. Then

$$\begin{aligned}
\frac{p_{X|YV}\big(1 \mid \boldsymbol{y}, v(\boldsymbol{y})\big)}{p_{X|YV}\big(0 \mid \boldsymbol{y}, v(\boldsymbol{y})\big)} \;&=\; \frac{p_{X|Y}(1 \mid \boldsymbol{y})}{p_{X|Y}(0 \mid \boldsymbol{y})} \;=\; \frac{\mathsf{p}_1}{\mathsf{p}_0}\Lambda(\boldsymbol{y}) \\
&=\; \frac{\mathsf{p}_1}{\mathsf{p}_0}u\big(v(\boldsymbol{y})\big),
\end{aligned}$$

where we first used (A.49), then Bayes' law, and then the assumption that $u(v)$ is a sufficient statistic. Since this ratio is the same for all $\boldsymbol{y}$ for which $v(\boldsymbol{y})$ has the same value, the ratio is a function of v alone,

$$\frac{p_{X|YV}\big(1 \mid \boldsymbol{y}, v(\boldsymbol{y})\big)}{p_{X|YV}\big(0 \mid \boldsymbol{y}, v(\boldsymbol{y})\big)} \;=\; \frac{p_{X|V}\big(1 \mid v\big)}{p_{X|V}\big(0 \mid v\big)}. \tag{A.51}$$

Finally, since $\mathsf{p}_{X|V}(0|v) = 1 - \mathsf{p}_{X|V}(1|v)$ and $\mathsf{p}_{X|\boldsymbol{Y}V}(0|\boldsymbol{y}, v(\boldsymbol{y})) = 1 - \mathsf{p}_{X|\boldsymbol{Y}V}(1|\boldsymbol{y}, v(\boldsymbol{y}))$, we see that (A.51) implies (A.50). Note that this says that $X \to V \to \boldsymbol{Y}$ is Markov.

c) Explain why this argument is valid whether $\boldsymbol{Y}$ is a discrete or continuous random vector and whether V is discrete, continuous or part discrete and part continuous.

Solution: The argument above essentially makes no assumptions about either $\boldsymbol{Y}$ nor V being discrete or having a density. The argument does depend on the assumption that the given conditional probabilities are defined.

Exercise 8.5: **a)** Let $\boldsymbol{Y}$ be a discrete observation random vector and let $v(\boldsymbol{y})$ be a function of the sample values of $\boldsymbol{Y}$. Show that

$$\mathsf{p}_{\boldsymbol{Y}|VX}\big(\boldsymbol{y} \mid v(\boldsymbol{y}), x\big) = \frac{\mathsf{p}_{\boldsymbol{Y}|X}(\boldsymbol{y} \mid x)}{\mathsf{p}_{V|X}(v(\boldsymbol{y}) \mid x)}. \tag{A.52}$$

Solution: We must assume that $\mathsf{p}_{VX}(v(\boldsymbol{y}), x) > 0$ so that the expression on the left is defined. Then, using Bayes' law on $\boldsymbol{Y}$ and V for fixed x on the left side of (A.52),

$$\mathsf{p}_{\boldsymbol{Y}|VX}\big(\boldsymbol{y} \mid v(\boldsymbol{y}), x\big) = \frac{\mathsf{p}_{V|\boldsymbol{Y}X}\big(v(\boldsymbol{y}) \mid \boldsymbol{y}x\big)\mathsf{p}_{\boldsymbol{Y}|X}\big(\boldsymbol{y} \mid x\big)}{\mathsf{p}_{V|X}\big(v(\boldsymbol{y}) \mid x\big)}. \tag{A.53}$$

Since V is a deterministic function of $\boldsymbol{Y}$, the first term in the numerator above is 1, so this is equivalent to (A.52),

b) Using Theorem 8.2.8, show that the above fraction is independent of X if and only if $v(\boldsymbol{y})$ is a sufficient statistic.

Solution: Using Bayes' law on the numerator and denominator of (A.52),

$$\frac{\mathsf{p}_{\boldsymbol{Y}|X}(\boldsymbol{y} \mid x)}{\mathsf{p}_{\boldsymbol{V}|X}\big(v(\boldsymbol{y}) \mid x\big)} = \frac{\mathsf{p}_{X|\boldsymbol{Y}}(x \mid \boldsymbol{y})\, \mathsf{p}_{\boldsymbol{Y}}(\boldsymbol{y})}{\mathsf{p}_{X|V}\big(x \mid v(\boldsymbol{y})\big)\, \mathsf{p}_{V}\big(v(\boldsymbol{y})\big)}. \tag{A.54}$$

If $v(\boldsymbol{y})$ is a sufficient statistic, then the second equivalent statement of Theorem 8.2.8, *i.e.*,

$$\mathsf{p}_{X|\boldsymbol{Y}}(x \mid \boldsymbol{y}) = \mathsf{p}_{X|V}(x \mid v(\boldsymbol{y})),$$

shows that the first terms on the right side of (A.54) cancel, showing that the fraction is independent of x. Conversely, if the fraction is independent of x, then the ratio of $\mathsf{p}_{X|\boldsymbol{Y}}(x|\boldsymbol{y})$ to $\mathsf{p}_{X|V}(x|v(\boldsymbol{y}))$, is a function only of $\boldsymbol{y}$. Since X is binary, this fraction must be 1, establishing the second statement of Theorem 8.2.8.

c) Now assume that $\boldsymbol{Y}$ is a continuous observation random vector, that $v(\boldsymbol{y})$ is a given function, and $V = v(\boldsymbol{Y})$ has a probability density. Define

$$\mathsf{f}_{\boldsymbol{Y}|VX}(\boldsymbol{y} \mid v(\boldsymbol{y}), x) = \frac{\mathsf{f}_{\boldsymbol{Y}|X}(\boldsymbol{y} \mid x)}{\mathsf{f}_{V|X}\big(v(\boldsymbol{y}) \mid x\big)}. \tag{A.55}$$

One can interpret this as a strange kind of probability density on a conditional sample space, but it is more straightforward to regard it simply as a fraction. Show that $v(\boldsymbol{y})$ is a sufficient statistic if and only if this fraction is independent of x. Hint: Model your derivation on that in (b), modifying (b) as necessary to do this.

Solution: The trouble with (A.53) can be seen by looking at (A.54). When this is converted to densities, the joint density of $\boldsymbol{Y}, V(\boldsymbol{Y})$ has the same problem as the densities in Example 8.2.10. The numerator and denominator of (A.52) are well defined as densities, however, and the argument in (b) carries through as before.

Exercise 8.9: A disease has two strains, 0 and 1, which occur with *a priori* probabilities p_0 and $p_1 = 1-p_0$ respectively.

a) Initially, a rather noisy test was developed to find which strain is present for patients with the disease. The output of the test is the sample value y_1 of a random variable Y_1. Given strain 0 (X=0), $Y_1 = 5 + Z_1$, and given strain 1 (X=1), $Y_1 = 1 + Z_1$. The measurement noise Z_1 is independent of X and is Gaussian, $Z_1 \sim \mathcal{N}(0, \sigma^2)$. Give the MAP decision rule, *i.e.*, determine the set of observations y_1 for which the decision is $\hat{x}$=1. Give $\Pr\{e \mid X{=}0\}$ and $\Pr\{e \mid X{=}1\}$ in terms of the function $Q(x)$.

Solution: This is simply a case of binary detection with an additive Gaussian noise rv. To prevent simply copying the answer from Example 8.2.3, the signal a associated with $X = 0$

is 5 and the signal b associated with $X = 1$ is 1. Thus $b < a$, contrary to the assumption in Example 8.2.3. Looking at that example, we see that (8.27), repeated below, is still valid.

$$\mathrm{LLR}(y) \;=\; \left[\left(\frac{b-a}{\sigma^2}\right)\left(y - \frac{b+a}{2}\right)\right] \;\mathop{\gtrless}_{\hat{x}(y)=a}^{\hat{x}(y)=b}\; \ln(\eta).$$

We can get a threshold test on y directly by first taking the negative of this expression and then dividing both sides by the positive term $(a-b)/\sigma^2$ to get

$$y \;\mathop{\lessgtr}_{\hat{x}(y)=a}^{\hat{x}(y)=b}\; \frac{-\sigma^2\ln(\eta)}{a-b} + \frac{b+a}{2}.$$

We get the same equation by switching the association of $X = 1$ and $X = 0$, which also changes the sign of the log threshold.

b) A budding medical researcher determines that the test is making too many errors. A new measurement procedure is devised with two observation random variables Y_1 and Y_2. Y_1 is the same as in (a). Y_2, under hypothesis 0, is given by $Y_2 = 5 + Z_1 + Z_2$, and, under hypothesis 1, is given by $Y_2 = 1 + Z_1 + Z_2$. Assume that Z_2 is independent of both Z_1 and X, and that $Z_2 \sim \mathcal{N}(0, \sigma^2)$. Find the MAP decision rule for $\hat{x}$ in terms of the joint observation (y_1, y_2), and find $\Pr\{e \mid X{=}0\}$ and $\Pr\{e \mid X{=}1\}$. Hint: Find $\mathsf{f}_{Y_2|Y_1,X}(y_2 \mid y_1, 0)$ and $\mathsf{f}_{Y_2|Y_1,X}(y_2 \mid y_1, 1)$.

Solution: Note that Y_2 is simply Y_1 plus the noise term Z_2, and that Z_2 is independent of X and Y_1. Thus, Y_2, conditional on Y_1 and X is simply $\mathcal{N}(Y_1, \sigma^2)$, which is independent of X. Thus Y_1 is a sufficient statistic and Y_2 is irrelevant. Including Y_2 does not change the probability of error.

c) Explain in laymen's terms why the medical researcher should learn more about probability.

Solution: It should have been clear intuitively that adding an additional observation that is only a noisy version of what has already been observed will not help in the decision, but knowledge of probability sharpens one's intuition so that something like this becomes self evident even without mathematical proof.

d) Now suppose that Z_2, in (b), is uniformly distributed between 0 and 1 rather than being Gaussian. We are still given that Z_2 is independent of both Z_1 and X. Find the MAP decision rule for $\hat{x}$ in terms of the joint observation (y_1, y_2) and find $\Pr(e \mid X{=}0)$ and $\Pr(e \mid X{=}1)$.

Solution: The same argument as in (b) shows that Y_2, conditional on Y_1, is independent of X, and thus the decision rule and error probability do not change.

e) Finally, suppose that Z_1 is also uniformly distributed between 0 and 1. Again find the MAP decision rule and error probabilities.

Solution: By the same argument as before, Y_2, conditional on Y_1 is independent of X, so Y_1 is a sufficient statistic and Y_2 is irrelevant. Since Z_1 is uniformly distributed between 0 and 1, then Y_1 lies between 5 and 6 for $X = 0$ and between 1 and 2 for $X = 1$. There is thus no possibility of error in this case.

Exercise 8.10: **a)** Consider a binary hypothesis testing problem, and denote the hypotheses as $X = 1$ and $X = -1$. Let $\boldsymbol{a} = (a_1, a_2, \dots, a_n)^{\mathsf{T}}$ be an arbitrary real n-vector and let the observation be a sample value $\boldsymbol{y}$ of the random vector $\boldsymbol{Y} = X\boldsymbol{a} + \boldsymbol{Z}$ where $Z \sim \mathcal{N}(0, \sigma^2 I_n)$ and I_n is the $n \times n$ identity matrix. Assume that Z and X are independent. Find the maximum likelihood decision rule and find the probabilities of error $\Pr(e \mid X{=}{-1})$ and $\Pr(e \mid X{=}1)$ in terms of the function $Q(x)$.

Solution: This is a minor notational variation on Example 8.2.4. Since we are interested in maximum likelihood, $\ln \eta = 0$. The ML test, from (8.41), is then

$$\mathrm{LLR}(\boldsymbol{y}) \;=\; \frac{2\boldsymbol{a}^\mathsf{T}\boldsymbol{y}}{\sigma^2} \;\; \mathop{\gtrless}_{\hat{x}(\boldsymbol{y})=-1}^{\hat{x}(\boldsymbol{y})=1} \;\; 0.$$

The error probabilities, from (8.44), are then

$$\Pr\{e \mid \boldsymbol{X}{=}1\} = Q\left(\gamma\right) \qquad \Pr\{e \mid \boldsymbol{X}{=}{-1}\} = Q\left(\gamma\right),$$

where $\gamma = \|2\boldsymbol{a}\|/(2\sigma)$.

b) Now suppose a third hypothesis, $X = 0$, is added to the situation of (a). Again the observation random vector is $\boldsymbol{Y} = X\boldsymbol{a} + \boldsymbol{Z}$, but here X can take on values $-1, 0$, or $+1$. Find a one dimensional sufficient statistic for this problem (i.e., a one dimensional function of y from which the likelihood ratios

$$\Lambda_1(\boldsymbol{y}) = \frac{p_{Y|X}(\boldsymbol{y} \mid 1)}{p_{Y|X}(\boldsymbol{y} \mid 0)} \quad \text{and} \quad \Lambda_{-1}(\boldsymbol{y}) = \frac{p_{Y|X}(\boldsymbol{y} \mid -1)}{p_{Y|X}(\boldsymbol{y} \mid 0)}$$

can be calculated).

Solution: We have seen that the likelihood ratio for each of these binary decisions depends only on the noise in the direction of the difference between the vectors. Since each difference is $\boldsymbol{a}$, we conclude that $\boldsymbol{a}^\mathsf{T}\boldsymbol{y}$ is a sufficient statistic. One can verify this easily by calculating $\Lambda_1(\boldsymbol{y})$ and $\Lambda_{-1}(\boldsymbol{y})$.

c) Find the maximum likelihood decision rule for the situation in (b) and find the probabilities of error, $\Pr(e \mid X{=}x)$ for $x = -1, 0, +1$.

Solution: For $X = 1$, an error is made if $\Lambda_1(\boldsymbol{y})$ is less than 1. This occurs if $\boldsymbol{a}^\mathsf{T}\boldsymbol{y} < \|\boldsymbol{a}\|^2/2$ and has probability $\Pr\{e \mid X = 1\} = Q(\|a\|/2\sigma)$. For $X = 0$, an error occurs if $\boldsymbol{a}^\mathsf{T}\boldsymbol{y} \geq \|\boldsymbol{a}\|^2/2$ or if $\boldsymbol{a}^\mathsf{T}\boldsymbol{y} < -\|\boldsymbol{a}\|^2/2$. Thus $\Pr\{e|X = 0\} = 2Q(\|a\|/2\sigma)$. Finally, for $X = -1$, an error occurs if $\boldsymbol{a}^\mathsf{T}\boldsymbol{y} \geq -\|\boldsymbol{a}\|^2/2$. $\Pr\{e|X = -1\} = Q(\|a\|/2\sigma)$.

d) Now suppose that $Z_1, \dots, Z_n$ in (a) are IID and each is uniformly distributed over the interval -2 to $+2$. Also assume that $\boldsymbol{a} = (1, 1, \dots, 1)^\mathsf{T}$. Find the maximum likelihood decision rule for this situation.

Solution: If $X = 1$, then each $Y_i, 1 \leq i \leq n$ lies between -1 and 3 and the conditional probability density of each such point is $(1/4)^n$. Similarly, for $X = -1$, each Y_i lies between -3 and 1. If all Y_i are between -1 and +1, then the LLR is 0. If any are above 1, the LLR is ∞, and if any are below -1, the LLR is $-\infty$. Thus the ML rule is $\hat{x} = 1$ if any $Y_i > 1$ and $\hat{x} = -1$ if any $Y_i < 1$. Everything else (which has aggregate probability 2^{-n}) is 'don't care,' which by convention is detected as $X = 1$.

Exercise 8.11: A sales executive hears that one of his salespeople is routing half of his incoming sales to a competitor. In particular, arriving sales are known to be Poisson at rate one per hour. According to the report (which we view as hypothesis $X{=}1$), each second arrival is routed to the competition; thus under hypothesis 1 the interarrival density for successful sales is $f(y|X{=}1) = ye^{-y}$; $y \geq 0$. The alternate hypothesis

(X=0) is that the rumor is false and the interarrival density for successful sales is $f(y|X{=}0) = e^{-y}$; $y \geq 0$. Assume that, *a priori*, the hypotheses are equally likely. The executive, a recent student of stochastic processes, explores various alternatives for choosing between the hypotheses; he can only observe the times of successful sales however.

a) Starting with a successful sale at time 0, let S_i be the arrival time of the i^{th} subsequent successful sale. The executive observes $S_1, S_2, \dots, S_n (n \geq 1)$ and chooses the maximum aposteriori probability hypothesis given this data. Find the joint probability density $f(S_1, S_2, \dots, S_n|X{=}1)$ and $f(S_1, \dots, S_n|X{=}0)$ and give the decision rule.

Solution: The interarrival times are independent conditional each on $X = 1$ and $X = 0$. The density of an interarrival interval given $X = 1$ is Erlang of order 2, with density xe^{-x}, so

$$\mathsf{f}_{\boldsymbol{S}|X}(s_1, \dots, s_n|1) \;=\; \prod_{i=1}^{n} \Big[(s_i - s_{i-1}) \exp -(s_i - s_{i-1})\Big] \;=\; e^{-s_n} \prod_{i=1}^{n} \big(s_i - s_{i-1}\big).$$

The density of an interarrival interval given $X = 0$ is exponential, so

$$\mathsf{f}_{\boldsymbol{S}|X}(s_1, \dots, s_n|0) \;=\; e^{-s_n}.$$

The MAP rule, with $p_0 = p_1$ is then

$$\mathrm{LLR}(\boldsymbol{y}) \;=\; \ln(s_0) + \sum_{i=2}^{n} \ln(s_i - s_{i-1}) \;\;\underset{\hat{x}(\boldsymbol{y})=0}{\overset{\hat{x}(\boldsymbol{y})=1}{\gtrless}}\;\; 0. \tag{A.56}$$

The executive might have based a decision only on the aggregate time for n sales to take place, but this would not have been a sufficient statistic for the sequence of sale times, so this would have yielded a higher probability of error. It is also interesting to note that a very short interarrival interval is weighed very heavily in (A.56), and this is not surprising since very short intervals are very improbable under $X = 1$. The sales person, if both fraudulent and a master of stochastic processes, would recognize that randomizing the sales to the competitor would make the fraud much more difficult to detect.

b) This is the same as (a) except that the system is in steady state at time 0 (rather than starting with a successful sale). Find the density of S_1 (the time of the first arrival after time 0) conditional on X=0 and on X=1. What is the decision rule now after observing $S_1, \dots, S_n$.

Solution: Under $X = 1$, the last arrival before 0 was successful with probability 1/2 and routed away with probability 1/2. Thus $\mathsf{f}_{S_1|X}(s_1|1) = (1/2)e^{-s_1} + (1/2)s_1e^{-s_1}$. This could also be derived as a residual life probability. With this modification, the first term in the LLR of (A.56) would be changed from $\ln(s_1)$ to $\ln((s_1 + 1)/2)$.

c) This is the same as (b), except rather than observing n successful sales, the successful sales up to some given time t are observed. Find the probability, under each hypothesis, that the first successful sale occurs in $(s_1, s_1 + \Delta]$, the second in $(s_2, s_2 + \Delta], \dots,$ and the last in $(s_{N(t)}, s_{N(t)} + \Delta]$ (assume Δ very small). What is the decision rule now?

Solution: The somewhat artificial use of Δ here is to avoid dealing with the discrete rv $N(t)$ and the density of a random number of rv's. One can eliminate this artificiality after understanding the solution. Under $X = 0$, the probability that $N(t) = n$ and $S_i \in [s_i, s_i + \Delta)$ for $1 \leq i \leq n$ and $s_n \leq t$ (using an approximation for Δ very small) is

$$\Delta^n e^{-s_1} \left[\prod_{i=2}^{n} e^{-(s_i - s_{i-1})} \right] e^{-(t-s_n)} \;=\; \Delta^n e^{-t}. \tag{A.57}$$

The term $e^{-(t-s_n)}$ on the left side is the probability of no arrivals in $(s_n, t]$, which along with the other arrival times specifies that $N(t) = n$.

Under $X = 1$, the term e^{-s_1} in (A.57) is changed to $(1/2)(s_1 + 1)e^{-s_1}$ as we saw in (b). The final term, $e^{-(t-s_n)}$ in (A.57) must be changed to the probability of no successful sales in $(s_n, t]$ under $X = 1$. This is $(t - s_n + 1)e^{-(t-s_n)}$. The term Δ^n cancels out in the likelihood ratio, so the decision rule is

$$\mathrm{LLR}(\boldsymbol{y}) \;=\; \ln\left(\frac{1+s_1}{2}\right) + \ln(t - s_n + 1) + \sum_{i=2}^{n} \ln(s_i - s_{i-1}) \;\;\underset{\hat{x}(\boldsymbol{y})=0}{\overset{\hat{x}(\boldsymbol{y})=1}{\gtrless}}\;\; 0.$$

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 $\mathsf{f}_{Y|X}(y \mid 0)$, $\mathsf{f}_{Y|X}(y \mid 1)$ and the likelihood ratio $\Lambda(y)$.

Solution: Note that $f_{Y|X}$ is simply the density of U shifted by X, *i.e.*,

$$\mathsf{f}_{Y|X}(y \mid 0) \;=\; \begin{cases} 1/2; & -1 \leq y \leq 1 \\ 0; & \text{elsewhere} \end{cases} \qquad \mathsf{f}_{Y|X}(y \mid 1) \;=\; \begin{cases} 1/2; & 0 \leq y \leq 2 \\ 0; & \text{elsewhere} \end{cases}.$$

The likelihood ratio $\Lambda(y)$ is defined only for $-1 \leq y \leq 2$ since neither conditional density is non-zero outside this range.

$$\Lambda(y) = \frac{\mathsf{f}_{Y|X}(y \mid 1)}{\mathsf{f}_{Y|X}(y \mid 0)} \;=\; \begin{cases} 0; & -1 \leq y < 0 \\ 1; & 0 \leq y \leq 1 \\ \infty; & 1 < y \leq 2 \end{cases}.$$

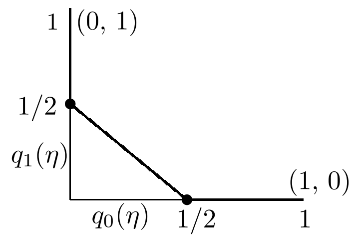
b) Find the threshold test at η for each η, $0 < \eta < \infty$ and evaluate the conditional error probabilities, $q_0(\eta)$ and $q_1(\eta)$.

Solution: Since $\Lambda(y)$ has finitely many (3) possible values, all values of η between any adjacent pair lead to the same threshold test. Thus, for $\eta > 1$, $\Lambda(y) \geq \eta$, if and only if (iff) $\Lambda(y) = \infty$. Thus $\hat{x} = 1$ iff $1 < y \leq 2$. For $\eta = 1$, $\hat{x} = 1$ iff $\Lambda(y) \geq 1$, *i.e.*, iff $\Lambda(y)$ is 1 or ∞. Thus $\hat{x} = 1$ iff $0 \leq y \leq 2$. For $\eta < 1$, $\lambda(y) \geq \eta$ iff $\Lambda(y)$ is 1 or ∞. Thus $\hat{x} = 1$ iff $0 \leq y \leq 2$. Note that the MAP test is the same for $\eta = 1$ and $\eta < 1$, in both cases choosing $\hat{x} = 1$ for $0 \leq y \leq 2$.

Consider $q_1(\eta)$ (the error probabiity using a threshold test at η conditional of $X = 1$). For $\eta > 1$, we have seen that $\hat{x} = 1$ (no error) for $1 < y \leq 2$. This occurs with probability 1/2 given $X = 1$. Thus $q_1(\eta) = 1/2$ for $\eta > 1$. Also, for $\eta > 1$, $\hat{x} = 0$ for $-1 \leq y \leq 1$. Thus $q_0(\eta) = 0$. Reasoning in the same way for $\eta \leq 1$, we have $q_1(\eta) = 0$ and $q_0(\eta) = 1/2$.

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: Each $\eta > 1$ maps into the pair of error probabilities $(q_0(\eta), q_1(\eta)) = (0,\ 1/2)$. Similarly, each $\eta \leq 1$ maps into the pair of error probabilities $(q_0(\eta), q_1(\eta)) = (1/2,\ 0)$. The error curve contains these points and also contains the supremum of the straight lines of each slope $-\eta$ around $(0, 1/2)$ for $\eta > 1$ and around $(1/2,\ 0)$ for $\eta \leq 1$. The resulting curve is given below.



Another approach (perhaps more insightful) is to repeat (a) and (b) for the alternative threshold tests that choose $\hat{x} = 0$ in the don't care cases, *i.e.*, the cases for $\eta = 1$ and $0 \leq y \leq 1$. It can be seen that Lemma 8.4.1 and Theorem 8.4.2 apply to these alternative threshold tests also. The points on the straight line between (0, 1/2) and (1/2, 0) can then be achieved by randomizing the choice between the threshold tests and the alternative threshold tests.

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 more easily, by mapping $y > 1/2$ into $\hat{x} = 1$ and $y \leq 1/2$ into $\hat{x} = 0$.

A.9 Solutions for Chapter 9

Exercise 9.1: Consider the simple random walk $\{S_n;\, n \geq 1\}$ of Section 9.1.1 with $S_n = X_1 + \cdots + X_n$ and $\Pr\{X_i = 1\} = p$; $\Pr\{X_i = -1\} = 1 - p$; assume that $p \leq 1/2$.

a) Show that $\Pr\Big\{\bigcup_{n\geq 1} \{S_n \geq k\}\Big\} = \Big[\Pr\Big\{\bigcup_{n\geq 1} \{S_n \geq 1\}\Big\}\Big]^k$ for any positive integer k. Hint: Given that the random walk ever reaches the value 1, consider a new random walk starting at that time and explore the probability that the new walk ever reaches a value 1 greater than its starting point.

Solution: Since $\{S_n; n \geq 1\}$ changes only in increments of 1 or -1, the only way that the walk can reach a threshold at integer $k > 1$ (*i.e.*, the only way that the event $\bigcup_{n\geq 1} \{S_n \geq k\}$ can occur), is if the walk first eventually reaches the value 1, and then starting from the first time that 1 is reached, goes on to eventually reach 2, and so forth onto k.

The probability of eventually reaching 2 given that 1 is reached is the same as the probability of eventually reaching 1 starting from 0; this is most clearly seen from the Markov chain depiction of the simple random walk given in Figure 6.1. Similarly, the probability of eventually reaching any j starting from $j - 1$ is again the same, so (using induction if one insists on being formal), we get the desired relationship. This relationship also holds for $p > 1/2$.

b) Find a quadratic equation for $y = \Pr\Big\{\bigcup_{n\geq 1} \{S_n \geq 1\}\Big\}$. Hint: explore each of the two possibilities immediately after the first trial.

Solution: As explained in Example 5.5.4, $y = p + (1 - p)y^2$.

c) For $p < 1/2$, show that the two roots of this quadratic equation are $p/(1 - p)$ and 1. Argue that $\Pr\Big\{\bigcup_{n\geq 1} \{S_n \geq 1\}\Big\}$ cannot be 1 and thus must be $p/(1 - p)$.

Solution: This is also explained in Example 5.5.4.

d) For $p = 1/2$, show that the quadratic equation in (c) has a double root at 1, and thus $\Pr\Big\{\bigcup_{n\geq 1} \{S_n \geq 1\}\Big\} = 1$. Note: this is the very peculiar case explained in Section 5.5.1.

Solution: As explained in Example 5.5.4, the fact that the roots are both 1 means that $\Pr\Big\{\bigcup_{n\geq 1} \{S_n \geq 1\}\Big\} = 1$.

e) For $p < 1/2$, show that $p/(1 - p) = \exp(-r^*)$ where r^* is the unique positive root of $g(r) = 1$ where $g(r) = \mathsf{E}\left[e^{rX}\right]$.

Solution: Note that $\mathsf{g}(r) = \mathsf{E}\left[e^{rX}\right] = pe^r + (1 - p)e^{-r}$. The positive root of $\mathsf{g}(r) = 1$ is the $r^* > 0$ for which

$$1 = pe^{r^*} + (1 - p)e^{-r^*}.$$

This is quadratic in e^{r^*} (and also in e^{-r^*}) and is the same equation as in (b) if we substitute y for e^{-r^*}. The two solutions are e^{-r^*}=1 (r^*=0) and $e^{-r^*} = p/(1 - p)$ ($r^* = \ln[(1 - p)/p]$). Thus the unique positive solution for r^* is $\ln[(1-p)/p]$. Thus the optimized Chernoff bound for crossing a threshold is satisfied with equality by the simple combinatorial solution in (c).

Exercise 9.2: Consider a G/G/1 queue with IID arrivals $\{X_i;\, i \geq 1\}$, IID FCFS service times $\{Y_i;\, i \geq 0\}$, and an initial arrival to an empty system at time 0. Define $U_i = Y_{i-1} - X_i$ for $i \geq 1$. (Note:

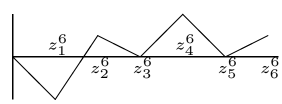
the problem statement in the text incorrectly defined this is $U_i = X_i - Y_{i-1}$, which is inconsistent with the treatment in Section 9.2) Consider a sample path where $(u_1, \ldots, u_6) = (1, -2, 2, -1, 3, -2)$.

a) Let $Z_i^6 = U_6 + U_{6-1} + \ldots + U_{6-i+1}$. Find the queueing delay for customer 6 as the maximum of the 'backward' random walk with elements $0, Z_1^6, Z_2^6, \ldots, Z_6^6$; sketch this random walk.

Solution:

$$\begin{aligned}
Z_1^6 &= & U_6 & &= -2 \\
Z_2^6 &= & U_6 + U_5 & &= 1 \\
Z_3^6 &= & U_6 + U_5 + U_4 & &= 0 \\
Z_4^6 &= & U_6 + U_5 + U_4 + U_3 & &= 2 \\
Z_5^6 &= & U_6 + U_5 + U_4 + U_3 + U_2 & &= 0 \\
Z_6^6 &= & U_6 + U_5 + U_4 + U_3 + U_2 + U_1 & &= 1
\end{aligned}$$

Note that the equation for Z_i^n is $\sum_{j=0}^{i-1} U_{n-j}$, rather than the slight typo given for Z_i^n in the equation between (9.5) and (9.6) in the text. The queueing delay W_6, for customer 6 is the maximum of $(0, Z_1^6, \ldots, Z_6^6)$, *i.e.*, 2.



b) Find the queueing delay for customers 1 to 5.

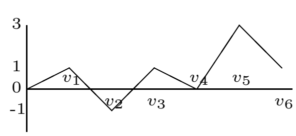
Solution: In the same way, $(W_1, \ldots, W_5) = (1, 0, 2, 1, 4)$

c) Which customers start a busy period (*i.e.*, arrive when the queue and server are both empty)? Verify that if Z_i^6 maximizes the random walk in (a), then a busy period starts with arrival $6 - i$.

Solution: Customer 2 starts a busy period (*i.e.*, $W_2 = 0$) and Z_{6-i}^6 is maximized in the random walk in (a) by $i = 2$.

d) Now consider a forward random walk $V_n = U_1 + \cdots + U_n$. Sketch this walk for the sample path above and show that the queueing delay for each customer is the difference between two appropriately chosen values of this walk.

Solution: $(v_1, \ldots, v_6) = (1, -1, 1, 0, 3, 1)$. The corresponding sample path segment is then



Note that $Z_i^n = U_n + \cdots + U_{n-i+1}$ and this can be rewritten as $V_n - V_{n-i}$. Thus $W_n = \max_{i \le n} V_n - V_{n-i} = V_n - \min_{i \le n} V_i$. Thus the queue delay at any n is the difference between the V random walk at n and the lowest previous point on that random walk. To understand this, consider the random walk $\{V_n;\, n \ge 1\}$. The first arrival $i > 0$ for which $v_i < 0$ is the first $i > 0$ to see an already empty system, and thus $W_i = 0$ rather than $W_i = V_i < 0$. Subsequent delays are then determined as if the random walk is reset to 0 at the ith arrival. The next arrival j that sees an already empty system is the first j for which $V_j < V_i$. and so forth at each arrival to see an already empty system. Thus the smallest V_j for $j \le n$ is the last $j \le n$ to start a new busy period. This is somewhat more intuitive than the backward random walk, but the backward random walk is the one that connects G/G/1 queue-waiting with random walks.

Exercise 9.3: A G/G/1 queue has a deterministic service time of 2 and interarrival times that are 3 with probability $p < 1/2$ and 1 with probability $1 - p$.

a) Find the distribution of W_1, the wait in queue of the first arrival after the beginning of a busy period.

Solution: The service time is always 2 units, so the first arrival after the beginning of a busy period must arrive either after 1 unit of the service is completed or 1 unit after the service is completed. Thus $\Pr\{W_1 = 0\} = p$ and $\Pr\{W_1 = 1\} = 1 - p$.

b) Find the distribution of W_∞, the steady-state wait in queue.

Solution: The rv $U_i = Y_{i+1} - X_i$ is binary (either 2-1 or 2-3), so $\Pr\{U_i = 1\} = p$ and $\Pr\{U_i = -1\} = 1-p$. The backward random walk is thus a simple random walk. From Theorem 9.2.1, $\Pr\{W \ge k\}$ is the probability that the random walk based on $\{U_i; i \ge 1\}$ ever crosses k. From (9.2), recognizing that the simple random walk is integer valued,

$$\Pr\{W \ge k\} = \left(\frac{p}{1-p}\right)^k .$$

c) Repeat (a) and (b) assuming the service times and interarrival times are exponentially distributed with rates μ and λ respectively.

Solution: This is an M/M/1 queue. With probability $\mu/(\lambda + \mu)$, the first arrival appears after the initial departure and thus has no wait in the queue. With probability $\lambda/(\lambda+\mu)$ the first arrival appears before the initial departure, and then waits in queue for an exponentially distributed time of rate μ. Thus $\Pr\{W_1 > w\} = \lambda/(\lambda + \mu)\exp(-\mu w)$ for $w \ge 0$.

In steady state, *i.e.*, to find $\Pr\{W > w\} = \lim_{n\to\infty} \Pr\{W_n > w\}$, we recall that the steady-state number N in the system is geometric with $\mathsf{p}_N(n) = (1-\lambda/\mu)(\lambda/\mu)^n$ for $n \ge 0$. Given $N = 0$ (an event of probability $1 - \lambda/\mu$), the wait is 0. Given $N = n$, the wait in the queue is $\sum_{i=1}^n Y_i^*$ where each Y_i^* is exponential with rate μ. Thus for $N > 0$ (an event of probability λ/μ the wait is a geometrically distributed sum of exponential rv's. From Example 2.3.3, this is equivalent to a sum of two Poisson processes, one of rate λ and one of rate $\mu - \lambda$ where the first arrival from the $\mu - \lambda$ process is geometrically distributed in the combined μ process. Thus,

$$\Pr\{W \ge w\} = \frac{\lambda}{\mu}\exp\big(-(\mu-\lambda)w\big); \qquad \text{for } w \ge 0.$$

Exercise 9.6: Define $\gamma(r)$ as $\ln[\mathsf{g}(r)]$ where $\mathsf{g}(r) = \mathsf{E}[\exp(rX)]$. Assume that X is discrete with possible outcomes $\{a_i;\ i \ge 1\}$, let p_i denote $\Pr\{X = a_i\}$, and assume that $\mathsf{g}(r)$ exists in some open interval (r_-, r_+) containing $r = 0$. For any given r, $r_- < r < r_+$, define a random variable X_r with the same set of possible outcomes $\{a_i;\ i \ge 1\}$ as X, but with a probability mass function $q_i = \Pr\{X_r = a_i\} = p_i \exp[a_i r - \gamma(r)]$. X_r is not a function of X, and is not even to be viewed as in the same probability space as X; it is of interest simply because of the behavior of its defined probability mass function. It is called a tilted random variable relative to X, and this exercise, along with Exercise 9.11 will justify our interest in it.

a) Verify that $\sum_i q_i = 1$.

Solution: Note that these tilted probabilities are described in Section 9.3.2.

$$\sum_i q_i \;=\; \sum_i p_i \exp[a_i r - \gamma(r)] \;=\; \frac{1}{\mathsf{g}(r)} \sum_i p_i \exp[a_i r] \;=\; \frac{\mathsf{g}(r)}{\mathsf{g}(r)} \;=\; 1.$$

b) Verify that $\mathsf{E}[X_r] = \sum_i a_i q_i$ is equal to $\gamma'(r)$.

Solution:

$$\begin{aligned} \mathsf{E}[X_r] &= \sum_i a_i p_i \exp[a_i r - \gamma(r)] \;=\; \frac{1}{\mathsf{g}(r)} \sum_i a_i p_i \exp[a_i r] \\ &= \frac{1}{\mathsf{g}(r)} \frac{d}{dr} \sum_i p_i e^{r a_i} \;=\; \frac{\mathsf{g}'(r)}{\mathsf{g}(r)} \;=\; \gamma'(r). \end{aligned} \tag{A.58}$$

c) Verify that $\mathsf{VAR}[X_r] = \sum_i a_i^2 q_i - (\mathsf{E}[X_r])^2$ is equal to $\gamma''(r)$.

Solution: We first calculate the second moment, $\mathsf{E}\left[X_r^2\right]$.

$$\begin{aligned} \mathsf{E}\left[X_r^2\right] &= \sum_i a_i^2 p_i \exp[a_i r - \gamma(r)] \;=\; \frac{1}{\mathsf{g}(r)} \sum_i a_i^2 p_i \exp[a_i r] \\ &= \frac{1}{\mathsf{g}(r)} \frac{d^2}{dr^2} \sum_i p_i e^{r a_i} \;=\; \frac{\mathsf{g}''(r)}{\mathsf{g}(r)}. \end{aligned} \tag{A.59}$$

Using (A.58) and (A.59),

$$\mathsf{VAR}[X_r] \;=\; \frac{\mathsf{g}''(r)}{\mathsf{g}(r)} - \frac{[\mathsf{g}'(r)]^2}{[\mathsf{g}(r)]^2} \;=\; \frac{d^2}{dr^2} \ln(g(r)) \;=\; \gamma''(r).$$

d) Argue that $\gamma''(r) \geq 0$ for all r such that $\mathsf{g}(r)$ exists, and that $\gamma''(r) > 0$ if $\gamma''(0) > 0$.

Solution: Since $\gamma''(r)$ is the variance of a rv, it is nonnegative. If $\mathsf{VAR}[X] > 0$, then X is non-atomic, which shows that X_r is non-atomic and thus has a positive variance wherever it exists.

e) Give a similar definition of X_r for a random variable X with a density, and modify (a) to (d) accordingly.

Solution: If X has a density, $\mathsf{f}_X(x)$, and also has an MGF over some region of r, then the tilted variable X_r is defined to have the density $\mathsf{f}_{X_r}(x) = \mathsf{f}_X(x)\exp(xr - \gamma(r))$. The derivations above follow as before except for the need of more care about convergence.

Exercise 9.9: [Details in proof of Theorem 9.3.3] **a)** Show that the two appearances of ϵ in (9.24) can be replaced with two independent arbitrary positive quantities ϵ_1 and ϵ_2, getting

$$\Pr\{S_n \geq n(\gamma'(r) - \epsilon_1)\} \geq (1-\delta)\exp[-n(r\gamma'(r) + r\epsilon_2 - \gamma(r))]. \tag{A.60}$$

Show that if this equation is valid for ϵ_1 and ϵ_2, then it is valid for all larger values of ϵ_1 and ϵ_2. Hint: Note that the left side of (9.24) is increasing in ϵ and the right side is decreasing.

Solution: For an arbitrary ϵ_1 and ϵ_2, let $\epsilon = \min(\epsilon_1, \epsilon_2)$. For that ϵ and any $\delta > 0$, there is an n_o such that (9.24) is satisfied for all $n \geq n_o$. Then, using the hint to replace ϵ by $\epsilon_1 \geq \epsilon$ on the left of (9.24),

$$\begin{aligned}
\Pr\{S_n \geq n(\gamma'(r) - \epsilon_1)\} &\geq \Pr\{S_n \geq n(\gamma'(r) - \epsilon)\} \\
&\geq (1-\delta)\exp[-n(r\gamma'(r) + r\epsilon - \gamma(r))] \\
&\geq (1-\delta)\exp[-n(r\gamma'(r) + r\epsilon_2 - \gamma(r))].
\end{aligned}$$

This also shows that if (A.60) is satisfied for a given ϵ_1, ϵ_2 and n, then it is satisfied for all larger ϵ_1, ϵ_2 for that n.

b) Show that by increasing the required value of n_o, the factor of $(1-\delta)$ can be eliminated in (A.60).

Solution: We can rewrite (A.60) as

$$\Pr\{S_n \geq n(\gamma'(r) - \epsilon_1)\} \geq (1-\delta)e^{nr\epsilon_2}\exp[-n(r\gamma'(r) + 2r\epsilon_2 - \gamma(r))].$$

This applies for $n \geq n_o$ for the n_o in part (a). Now choose n_o' large enough so that both $(1-\delta)e^{n_o r\epsilon_2'} \geq 1$ and (A.60) (for the given (ϵ_1, ϵ_2)) is satisfied. Then for $n \geq n_o'$,

$$\begin{aligned}
\Pr\{S_n \geq n(\gamma'(r) - \epsilon_1)\} &\geq (1-\delta)e^{nr\epsilon'}\exp[-n(r\gamma'(r) + 2r\epsilon_2 - \gamma(r))] \\
&\geq \exp[-n(r\gamma'(r) + 2r\epsilon_2 - \gamma(r))] \\
&= \exp[-n(r\gamma'(r) + r\epsilon_2' - \gamma(r))], \qquad \text{(A.61)}
\end{aligned}$$

where $\epsilon_2' = 2\epsilon_2$. Since $\epsilon_2 > 0$ is arbitrary, $\epsilon_2' > 0$ is also arbitrary, with the corresponding change to n_o'.

c) For any $r \in (0, r_+)$, let δ_1 be an arbitrary number in $(0, r_+ - r)$, let $r_1 = r + \delta_1$, and let $\epsilon_1 = \gamma'(r_1) - \gamma'(r)$. Show that there is an m such that for all $n \geq m$,

$$\Pr\{S_n \geq n\gamma'(r)\} \geq \exp\{-n[(r+\delta_1)\gamma'(r+\delta_1) + (r+\delta_1)\epsilon_2 - \gamma(r+\delta_1)]\}. \qquad \text{(A.62)}$$

Using the continuity of γ and its derivatives, show that for any $\epsilon > 0$, there is a $\delta_1 > 0$ so that the right side of (A.62) is greater than or equal to $\exp[-n(\gamma'(r) - r\gamma(r) + r\epsilon)]$.

Solution: Since $\gamma'(r)$ is increasing in r, $\epsilon_1 = \gamma'(r_1) - \gamma'(r) > 0$. Now (A.61) applies for arbitrary $r \in (0, r_+)$ and ϵ_1, ϵ_2 (using the appropriate n_o' for those parameters). Thus we can apply (A.61) to r_1 and ϵ_1 as defined above and with ϵ_2 replacing ϵ_2'.

$$\Pr\{S_n \geq n(\gamma'(r_1) - \epsilon_1)\} \geq \exp[-n(r_1\gamma'(r_1) + r_1\epsilon_2 - \gamma(r_1))]$$

Replacing $\gamma'(r_1) - \epsilon_1$ on the left with $\gamma'(r)$ and r_1 on the right by $r + \delta_1$, we get (A.62). This is valid for $n \geq m$ where m is the value of n_o' for these modifications of r, ϵ_1, and ϵ_2.

Let $\epsilon > 0$ be arbitrary. Since $r\gamma'(r)$ is continuous in r, we see that for all small enough δ_1, $(r+\delta_1)\gamma'(r+\delta_1) \leq r\gamma'(r) + r\epsilon/3$. Similarly, $-\gamma(r+\delta_1) \leq -\gamma(r) + r\epsilon/3$ for all small enough δ_1. Finally, for any such small enough δ_1, we can choose $\epsilon_2 > 0$ small enough that

$(r+\delta_1)\epsilon_2 \leq r\epsilon/3$. Substituting these inequalities into (A.62), there is an m such that for $n \geq m$,

$$\Pr\{S_n \geq n\gamma'(r)\} \geq \exp\left\{-n\left[r\gamma'(r) + r\epsilon - \gamma(r)\right]\right\},$$

thus completing the final details of the proof of Theorem 9.3.3.

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 $\mathsf{f}(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. In other words, the random walk is the sequence of arrival epochs in a Poisson process.

a) Show that for $\lambda a > 1$, 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 $\mathsf{g}(r) = \mathsf{E}\left[e^{Xr}\right] = \lambda/(\lambda - r)$ for $r < \lambda$. Thus $\gamma(r) = \ln \mathsf{g}(r) = \ln(\lambda/(\lambda - r))$ and $\gamma'(r) = 1/(\lambda - r)$. 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,

$$\Pr\{S_n \geq na\} \;\leq\; \exp\left[n \ln(\frac{\lambda}{\lambda - r}) - nra\right] \;=\; \exp[n\ln(a\lambda) - n(a\lambda - 1)],$$

which is equivalent to the desired expression.

b) Show that the exact value of $\Pr\{S_n \geq na\}$ is given by

$$\Pr\{S_n \geq na\} = \sum_{i=0}^{n-1} \frac{(na\lambda)^i\, e^{-na\lambda}}{i!}. \tag{A.63}$$

Solution: For a Poisson counting process $\{N(t); t > 0\}$, the event $\{S_n > na\}$ is the same as $\{N(na) < n\} = \bigcup_{i=0}^{n-1}\{N(na) = i\}$. Since this is a union of disjoint events,

$$\Pr\{S_n > na\} = \sum_{i=0}^{n-1} \Pr\{N(na) = i\}$$

Using the Poisson PMF, the right side of this is equal to the right side of (A.63). Since S_n is continuous, $\Pr\{S_n > na\} = \Pr\{S_n \geq na\}$.

c) By upper and lower bounding the quantity on the right of (A.63), 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 (A.63). For the upper bound, rewrite the sum in (b) as

$$\begin{aligned}
\sum_{i=0}^{n-1} \frac{(na\lambda)^i\, e^{-na\lambda}}{i!} &= \frac{(na\lambda)^n\, e^{-na\lambda}}{n!}\left[\frac{n}{na\lambda} + \frac{n(n-1)}{(na\lambda)^2} + \cdots\right] \\
&\leq \frac{(na\lambda)^n\, e^{-na\lambda}}{n!}\left[\frac{1}{a\lambda} + \frac{1}{(a\lambda)^2} + \cdots\right] \;=\; \frac{(na\lambda)^n e^{-na\lambda}}{n!(a\lambda - 1)}.
\end{aligned}$$

d) Use the Stirling bounds on $n!$ to show that

$$\frac{(a\lambda)^n\, e^{-n(a\lambda-1)}}{\sqrt{2\pi n}\, a\lambda \exp(1/12n)} \;\le\; \Pr\{S_n \ge na\} \;\le\; \frac{(a\lambda)^n e^{-n(a\lambda-1)}}{\sqrt{2\pi n}\,(a\lambda-1)}.$$

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 $n!$ in (c) 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 $\Pr\{S_J \ge \alpha\}$ in the absence of a lower threshold. Use the upper bound in (9.46) for the probability that the random walk crosses α before β.

a) Given that the random walk crosses β first, find an upper bound to the probability that α is now crossed before a yet lower threshold at 2β is crossed.

Solution: Let J_1 be the stopping trial at which the walk first crosses either α or β. Let J_2 be the stopping trial at which the random walk first crosses either α or 2β (assuming the random walk continues forever rather than actually stopping at any stopping trial. Note that if $S_{J_1} \ge \alpha$, then $S_{J_2} = S_{J_1}$, but if $S_{J_1} \le \beta$, then it is still possible to have $S_{J_2} \ge \alpha$. 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,

$$\begin{aligned}
\Pr\{S_{J_2} \ge \alpha\} &= \Pr\{S_{J_1} \ge \alpha\} + \Pr\{S_{J_2} \ge \alpha \mid S_{J_1} \le \beta\} \Pr\{S_{J_1} \le \beta\} \\
&\le \Pr\{S_{J_1} \ge \alpha\} + \Pr\{S_{J_2} \ge \alpha \mid S_{J_1} \le \beta\}.
\end{aligned}$$

The second term above 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 $S_{J_1} \le \beta$. Since $\alpha - S_{J_1} \ge \alpha - \beta$,

$$\Pr\{S_{J_2} \ge \alpha \mid S_{J_1} \le \beta\} \le \exp[-r^*(\alpha - \beta)],$$

Thus,

$$\Pr\{S_{J_2} \ge \alpha\} \le \exp(-r^*\alpha) + \exp[-r^*(\alpha - \beta)].$$

Note that it is conceivable that $S_{J_1} = S_{J_2}$, *i.e.*, that β and 2β are crossed at the same time. The argument above includes this case, and no further note about it is necessary.

b) Given that 2β is crossed before α, upper bound the probability that α is crossed before a threshold at 3β. 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 α is crossed. By observing that β is arbitrary, show that (9.46) is valid with no lower threshold.

Solution: Let J_k for each $k \ge 1$ be the stopping trial at which α or $k\beta$ is crossed. By the same argument as above,

$$\begin{aligned}
\Pr\left\{S_{J_{k+1}} \ge \alpha\right\} &= \Pr\{S_{J_k} \ge \alpha\} + \Pr\left\{S_{J_{k+1}} \ge \alpha \mid S_{J_k} \le k\beta\right\} \Pr\{S_{J_k} \le k\beta\} \\
&\le \Pr\{S_{J_k} \ge \alpha\} + \exp[-r^*(\alpha - k\beta)],
\end{aligned}$$

Finally, let J_∞ be the defective stopping time at which α 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

$$\begin{aligned} \Pr\{S_{J_\infty} \geq \alpha\} &\leq \Pr\{S_{J_1} \geq \alpha\} + \sum_{k=1}^{\infty} \Pr\{S_{J_{k+1}} \geq \alpha \mid S_{J_k} \leq k\beta\} \\ &\leq \exp[-r^*\alpha] \frac{1}{1 - \exp[-r^*\beta]}. \end{aligned}$$

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 non-defective stopping rule. The approach here is to define the limit as a union of non-defective stopping rules.

Exercise 9.16: **a)** Use Wald's equality to show that if $\overline{X} = 0$, then $\mathsf{E}[S_J] = 0$ 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 $\mathsf{E}[|J|] < \infty$, which follows from Lemma 9.4.1, which says that J has an MGF in an interval around 0. Thus $\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]$. Since $\overline{X} = 0$, it follows that $\mathsf{E}[S_J] = 0$.

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 $\mathsf{E}[S_J] = 0$ in terms of conditional expectations,

$$\begin{aligned} \mathsf{E}[S_J] &= \Pr\{S_J \geq \alpha\}\, \mathsf{E}[S_J \mid S_J \geq \alpha] + \Pr\{S_J \leq \beta\}\, \mathsf{E}[S_J \mid S_J \leq \beta] \\ &= \Pr\{S_J \geq \alpha\}\, \mathsf{E}[S_J \mid S_J \geq \alpha] + [1 - \Pr\{S_J \geq \alpha\}]\mathsf{E}[S_J \mid S_J \leq \beta]. \end{aligned}$$

Using $\mathsf{E}[S_J] = 0$, we can solve this for $\Pr\{S_J \geq \alpha\}$,

$$\Pr\{S_J \geq \alpha\} = \frac{\mathsf{E}[-S_J \mid S_J \leq \beta]}{\mathsf{E}[-S_J \mid S_J \leq \beta] + \mathsf{E}[S_J \mid S_J \geq \alpha]}.$$

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 α and β are integers, there can be no overshoot when a threshold is crossed. Thus $\mathsf{E}[S_J \mid S_J \geq \alpha] = \alpha$ and $\mathsf{E}[S_J \mid S_J \leq \beta] = \beta$. Thus $\Pr\{S_J \geq \alpha\} = \frac{|\beta|}{|\beta|+\alpha}$. If α is non-integer, then a positive threshold crossing occurs at $\lceil \alpha \rceil$ and a lower threshold crossing at $\lfloor \beta \rfloor$. Thus, in this general case $\Pr\{S_J \geq \alpha\} = \frac{|\lfloor \beta \rfloor|}{|\lfloor \beta \rfloor| + \lceil \alpha \rceil}$.

d) Evaluate your expression when X has an exponential density, $\mathsf{f}_X(x) = a_1 e^{-\lambda x}$ for $x \geq 0$ and $\mathsf{f}_X(x) = a_2 e^{\mu x}$ for $x < 0$ and where a_1 and a_2 are chosen so that $\overline{X} = 0$.

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 $\mathsf{f}_V(v) = \lambda e^{-\lambda v}$ 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 $\mathsf{E}[S_J \mid J \geq \alpha] = \alpha + 1/\lambda$. In the same way, $\mathsf{E}[S_J \mid S_J \leq \beta] = \beta - 1/\mu$. Thus

$$\Pr\{S_J \geq \alpha\} = \frac{|\beta| + \mu^{-1}}{\alpha + \lambda^{-1} + |\beta| + \mu^{-1}}.$$

Note that it is not necessary to calculate a_1 or a_2.

Exercise 9.23: Suppose $\{Z_n;\, n \geq 1\}$ is a martingale. Show that

$$\mathsf{E}\left[Z_m \mid Z_{n_i}, Z_{n_{i-1}}, \dots, Z_{n_1}\right] = Z_{n_i} \quad \text{for all } 0 < n_1 < n_2 < \dots < n_i < m.$$

Solution: First observe from Lemma 9.6.5 that for $n_i < m$,

$$\mathsf{E}[Z_m \mid Z_{n_i}, Z_{n_i-1}, Z_{n_i-2}, \dots, Z_1] = Z_{n_i}.$$

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}, the expected value of this conditional expectation over Z_{n_i-1} is $\mathsf{E}[Z_m \mid Z_{n_i}, Z_{n_i-2}, \dots, Z_1] = Z_{n_i}$. 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 $\mathsf{E}[J]$ until some fixed pattern, $a_1, a_2, \dots, a_k$, of successive binary digits occurs within a sequence of IID binary random variables $X_1, X_2, \dots$ (see Example 4.5.1 and Exercise 5.35 for alternative approaches). We take the stopping time J to be the smallest n for which $(X_{n-k+1}, \dots, X_n) = (a_1, \dots, a_k)$. A mythical casino and sequence of gamblers who follow a prescribed strategy will be used to determine $\mathsf{E}[J]$. 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 $\mathsf{E}[Y_n] = 0$. 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 $\mathsf{E}[Y_n | Y_{n-1} \dots, Y_1] = 0$. As seen in Example 9.6.3, this implies that $\{Z_n;\, n \geq 1\}$ is a martingale.

b) In order to determine $\mathsf{E}[J]$ for a given pattern $(a_1, a_2, \dots, a_k)$, 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 $1/[p_{a_1} \dots p_{a_k}]$).

ii) Gambler ℓ, for each $\ell > 1$, follows the same strategy, but starts at trial ℓ. Note that if the string $(a_1, \dots, a_k)$ appears for the first time at trials $n{-}k{+}1$, $n{-}k{+}2, \dots, n$, *i.e.*, if $J = n$, then gambler $n - k + 1$ leaves at time n with capital $1/[p(a_1) \dots p(a_k)]$ 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 $X_1 = 1$, $X_2 = 0$, $X_3 = 1$, ... 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 pattern (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 $\mathsf{E}[Z_J]$ from (a). Use this plus (b) to find $\mathsf{E}[J]$. Hint: This uses the special form of the solution in (b), not the Wald equality.

Solution: The casino's expected gain up to each time n is $\mathsf{E}[Z_n] = 0$, so it follows that $\mathsf{E}[Z_J] = 0$ (It is easy to verify that the condition in 9.104 is satisfied in this case). We saw in (b) that $\mathsf{E}[Z_n \mid J = n] = n - 1/p_0p_1$, so $\mathsf{E}[Z_J] = \mathsf{E}[J] - 1/p_0p_1$. Thus $\mathsf{E}[J] = 1/p_0p_1$. 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 $(a_1, \dots, a_k) = (1, 1)$ and initially assuming $(X_1X_2X_3) = (011)$. Be careful about gambler 3 for $J = 3$. Show that $\mathsf{E}[J] = 1/p_1^2 + 1/p_1$.

Solution: This is almost the same as (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 $\mathsf{E}[J] = 1/p_1^2 + 1/p_1$.

e) Repeat parts b) and c) for $(a_1, \dots, a_k) = (1, 1, 1, 0, 1, 1)$.

Solution: Given that $J = n$, we know that $(X_{n-5}, \dots, X_n) = (111011)$ 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 $\mathsf{E}[J] = 1/(p_1^5p_0) + 1/p_1 + 1/p_1^2$. 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 $a_1, \dots, a_k$ and condition on $J = n$ for some $n \geq k$. Show that the sample capital of gambler ℓ is then equal to

- 0 for $\ell < n - k$.
- $1/[p_{a_1}p_{a_2} \cdots p_{a_k}]$ for $\ell = n - k + 1$.
- $1/[p_{a_1}p_{a_2} \cdots p_{a_i}]$ for $\ell = n - i + 1,\ 1 \leq i < k$ if $(a_1, \dots, a_i) = (a_{k-i+1}, \dots, a_k)$.
- 0 for $\ell = n - i + 1,\ 1 \leq i < k$ if $(a_1, \dots, a_i) \neq (a_{k-i+1}, \dots, a_k)$.

Verify that this general formula agrees with parts (b), (d), and (e).

Solution: Gambler ℓ for $\ell \le n-k$ bets (until losing a bet) on $a_1, a_2, \dots, a_k$. Since the first occurrence of $(a_1, \dots, a_k)$ occurs at n, we see that each of these gamblers loses at some point and thus is reduced to 0 capital at that point and stays there. Gambler $n-k+1$ bets on $a_1, \dots, a_k$ at times $n-k+1, \dots, n$ and thus wins each bet for $J = n$. Finally, gambler $\ell = n-i+1$ bets (until losing) on $a_1, a_2, \dots, a_i$ at times $n-i+1$ to n. Since $J = n$ implies that $X_{n-k+1}, \dots, X_n = a_1, \dots a_k$, gambler $n-i+1$ is successful on all i bets if and only if $(a_1, \dots, a_i) = (a_{k-i+1}, \dots, a_k)$.

For (b), gambler n is unsuccessful and in (d), gambler n is successful. In (e), gamblers $n-1$ and n are each successful. It might be slightly mystifing at first that conditioning on J is enough to specify what happens to each gambler 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 gamblers are also specified.

g) For a given binary string $(a_1, \dots, a_k)$, and each j, $1 \le j \le k$ let $\mathbb{I}_j = 1$ if $(a_1, \dots, a_j) = (a_{k-j+1}, \dots, a_k)$ and let $\mathbb{I}_j = 0$ otherwise. Show that

$$\mathsf{E}[J] = \sum_{i=1}^{k} \frac{\mathbb{I}_i}{\prod_{\ell=1}^{i} p_{a_\ell}}.$$

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 ith term in the above expansion is the capital of gambler $n-i+1$ at time n. The final term at $i = k$ corresponds to the gambler who retires to Maui and $\mathbb{I}_k = 1$ 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 $\mathsf{E}[|Z_J|] < \infty$ is required in Lemma 9.8.4. Let $Z_1 = -2$ and, for $n \ge 1$, let $Z_{n+1} = Z_n[1 + X_n(3n+1)/(n+1)]$ where $X_1, X_2, \dots$ are IID and take on the values $+1$ and -1 with probability 1/2 each. Show that $\{Z_n;\, n \ge 1\}$ is a martingale.

Solution: From the definition of Z_n above,

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] = \mathsf{E}[Z_{n-1}[1 + X_{n-1}(3n-2)/n] \mid Z_{n-1}, \dots, Z_1].$$

Since the X_n are zero mean and IID, this is just $\mathsf{E}[Z_{n-1} \mid Z_{n-1} \dots, Z_1]$, which is Z_{n-1}. We also must show that $\mathsf{E}[|Z_n|] < \infty$ for each n. Note that

$$|Z_n| \le |Z_{n-1}|[1 + (3n-2)/n] \le 4|Z_{n-1}|$$

Thus for each n, $|Z_n|$ is bounded (although the bound is rapidly increasing in n), so $\mathsf{E}[|Z_n|] < \infty$ for each n. It follows that $\{Z_n;\, n \ge 1\}$ 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 $n \ge 2$ for which $X_{n-1} = 1$. Thus for $n < J$, we must have $X_i = -1$ for

$1 \le i < n$. 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 $X_{J-1} = 1$,

$$Z_J \;=\; Z_{J-1}\left[1 + \frac{3(J-1)+1}{J}\right] \;=\; \frac{(-2)^{J-1}}{J-1}\,\frac{4J-2}{J} \;=\; \frac{-(-2)^J(2J-1)}{J(J-1)}.$$

c) Show that $\mathsf{E}[|Z_J|]$ is infinite, so that $\mathsf{E}[Z_J]$ does not exist according to the definition of expectation, and show that $\lim_{n\to\infty} \mathsf{E}[Z_n|J > n]\Pr\{J > n\} = 0$.

Solution: We have seen that $J = n$ if and only if $X_i = -1$ for $1 \le i \le n-2$ and $X_{J-1} = 1$. Thus $\Pr\{J = n\} = 2^{-n+1}$ so

$$\mathsf{E}[|Z_J|] \;=\; \sum_{n=2}^{\infty} 2^{-n+1}\cdot\frac{2^n(2n-1)}{n(n-1)} \;=\; \sum_{n=2}^{\infty}\frac{2(2n-1)}{n(n-1)} \;\ge\; \sum_{n=2}^{\infty}\frac{4}{n} \;=\; \infty, \qquad \text{(i)}$$

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 $1 \le i < n$). Thus

$$\mathsf{E}[Z_n \mid J > n]\Pr\{J > n\} \;=\; \frac{2^{-n+1}(-2)^n}{n} \;=\; \frac{2(-1)^n}{n} \;\to 0\,.$$

This martingale and stopping rule gives an example where the condition $\mathsf{E}[|Y_J|]$ is needed in Lemma 9.8.4. Note that $\mathsf{E}[Y_J]$, if it exists, is $\lim_{n\to\infty}\sum_{i=1}^{\infty}\mathsf{E}[Z_J{=}i]\Pr\{J{=}i\}$. For this example, this is the alternating version of the harmonic series in (i), which converges in the Cauchy sense but not in the usual sense. Exercise 1.8 provides a reminder of why this series is defined to be nonexistant, *i.e.*, of why the condition $\mathsf{E}[|Y_J|] < \infty$ is needed.

Exercise 9.31: Prove Corollaries 9.9.3 to 9.9.5, *i.e.*, prove the following three statements:

a) Let $\{Z_n;\, n \ge 1\}$ be a martingale with $\mathsf{E}\left[Z_n^2\right] < \infty$ for all $n \ge 1$. Then

$$\Pr\left\{\max_{1\le n\le m}|Z_n| \ge b\right\} \le \frac{\mathsf{E}\left[Z_m^2\right]}{b^2};\ \text{ for all integer } m \ge 2, \text{ and all } b > 0. \qquad \text{(A.64)}$$

Hint: First show that $\{Z_n^2;\, n \ge 1\}$ is a submartingale.

Solution: Z_n^2 is a convex function of Z_n and $\mathsf{E}\left[Z_n^2\right] < \infty$ for all $n \ge 1$. Thus, from Theorem 9.7.4, $\{Z_n^2;\, n \ge 1\}$ is a submartingale. Applying the submartingale inequality to $\{Z_n^2;\, n \ge 1\}$,

$$\Pr\left\{\max_{1\le n\le m} Z_n^2 \ge a\right\} \le \frac{\mathsf{E}\left[Z_m^2\right]}{a}.$$

Substituting b^2 for a, this becomes (A.64).

b) [Kolmogorov's random walk inequality] Let $\{S_n;\, n \geq 1\}$ be a random walk with $S_n = X_1 + \cdots + X_n$ where $\{X_i;\, i \geq i\}$ is a set of IID random variables with mean $\overline{X}$ and variance σ^2. Then for any positive integer m and any $\epsilon > 0$,

$$\Pr\left\{\max_{1\leq n\leq m} |S_n - n\overline{X}| \geq m\epsilon\right\} \leq \frac{\sigma^2}{m\epsilon^2}. \tag{A.65}$$

Hint: Firt show that $\{S_n - n\overline{X};\, n \geq 1\}$ is a martingale.

Solution: Let $Z_n = S_n - n\overline{X}$. Then $\{Z_n;\, n \geq 1\}$ is a martingale as shown in Example 9.6.2. $\mathsf{E}\left[Z_m^2\right] = m\sigma^2$, so using (A.64) with $m\epsilon$ substituted for b, we get (A.65).

c) Let $\{S_n;\, n \geq 1\}$ be a random walk, $S_n = X_1 + \cdots + X_n$, where each X_i has mean $\overline{X} < 0$ and semi-invariant moment generating function $\gamma(r)$. For any $r > 0$ such that $0 < \gamma(r) < \infty$, and for any $\alpha > 0$. show that

$$\Pr\left\{\max_{1\leq i\leq n} S_i \geq \alpha\right\} \leq \exp\{-r\alpha + n\gamma(r)\}. \tag{A.66}$$

Hint: First show that $\{e^{rS_n};\, n \geq 1\}$ is a submartingale.

Solution: Again let $Z_n = S_n - n\overline{X}$, so $\{Z_n,\, n \geq 1\}$ is a martingale. Then $e^{rS_n} = e^{rZ_n + rn\overline{X}}$ and we see by differentiating twice with respect to Z_n that e^{rS_n} is a convex function of Z_n. Thus by Theorem 9.7.4, $\{e^{rS_n};\, n \geq 1\}$ is a submartingale. Thus by the Kolmogorov submartingale inequality,

$$\Pr\left\{\max_{1\leq i\leq n} e^{rS_i} \geq a\right\} \leq \frac{\mathsf{E}\left[e^{rS_n}\right]}{a}. \tag{A.67}$$

Now $\mathsf{E}\left[e^{rS_n}\right] = [\mathsf{g}_X(r)]^n = e^{n\gamma(r)}$. Substituting $e^{r\alpha}$ for a, (A.67) becomes (A.66)

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 has a nonzero sample value in any race.

a) For any given allocation $\boldsymbol{\lambda}$ 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 $W_1(\boldsymbol{\lambda})$ is $\lambda(j)r(j) + \lambda(\ell)$. Thus

$$\begin{aligned}
\mathsf{E}\left[W_1(\boldsymbol{\lambda})\right] &= \sum_{j=1}^{\ell-1} p(j)\left[\lambda(j)r(j) + \lambda(\ell)\right], \\
\mathsf{E}\left[L_1(\boldsymbol{\lambda})\right] &= \sum_{j=1}^{\ell-1} p(j)\ln\left[\lambda(j)r(j) + \lambda(\ell)\right].
\end{aligned}$$

b) Assume that a statistically identical sequence of races are run, *i.e.*, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots,$ are IID where each $\boldsymbol{X}_n = \left(X_n(1), \ldots, X_n(\ell)\right)^{\mathsf{T}}$. Assuming unit initial wealth and a constant allocation $\boldsymbol{\lambda}$ on each race, find the expected log-wealth $\mathsf{E}\left[L_n(\boldsymbol{\lambda})\right]$ at the end of the nth race and express it as $n\mathsf{E}\left[Y(\boldsymbol{\lambda})\right]$.

Solution: Using (9.128) to express $\mathsf{E}[L_n(\boldsymbol{\lambda})]$ as $n\mathsf{E}[Y(\boldsymbol{\lambda})]$, we have

$$\mathsf{E}[Y(\boldsymbol{\lambda})] = \sum_{j=1}^{\ell-1} p(j) \ln \big[\lambda(j) r(j) + \lambda(\ell)\big]. \tag{A.68}$$

c) Let $\boldsymbol{\lambda}^*$ maximize $\mathsf{E}[Y(\boldsymbol{\lambda})]$. Use the necessary and sufficient condition (9.136) on $\boldsymbol{\lambda}^*$ for horse j, $1 \le j < \ell$ to show that $\lambda^*(j)$ can be expressed in the following two equivalent ways; each uniquely specifies each $\lambda^*(j)$ in terms of $\lambda^*(\ell)$.

$$\lambda^*(j) \;\ge\; p(j) - \frac{\lambda^*(\ell)}{r(j)}; \qquad \text{with equality if } \lambda^*(j) > 0 \tag{A.69}$$

$$\lambda^*(j) \;=\; \max\Big\{p(j) - \frac{\lambda^*(\ell)}{r(j)},\ 0\Big\}. \tag{A.70}$$

Solution: The necessary and sufficient condition for $\boldsymbol{\lambda}^*$ in (9.136) for horse j is

$$\mathsf{E}\left[\frac{X(j)}{\sum_k \lambda^*(k) X(k)}\right] \le 1; \qquad \text{with equality if } \lambda^*(j) > 0.$$

In the event that horse j wins, $X(j) = r(j)$ while $X(k) = 0$ for horses $k \ne j$. Also $X(\ell) = 1$. Thus in the event that horse j wins, $\frac{X(j)}{\sum_k \lambda^*(k)X(k)} = \frac{r(j)}{\lambda^*(j)r(j)+\lambda^*(\ell)}$. If any other horse wins, $\frac{X(j)}{\sum_k \lambda^*(k)X(k)} = 0$. Thus, since j wins with probability $p(j)$,

$$\mathsf{E}\left[\frac{X(j)}{\sum_k \lambda^*(k) X(k)}\right] \;=\; \frac{p(j)\, r(j)}{\lambda^*(j) r(j) + \lambda^*(\ell)} \;\le\; 1; \qquad \text{with equality if } \lambda^*(j) > 0. \tag{A.71}$$

Rearranging this inequality yields (A.69); (A.70) 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 $\boldsymbol{\lambda}^*$) breaks into 3 special cases which are treated below in parts d), e), and f) respectively. The first case, in (d), shows that if $\sum_{j<\ell} 1/r(j) < 1$, then $\lambda^*(\ell) = 0$. The second case, in (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.73). The third case, in (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.69) over j to show that if $\lambda^*(\ell) > 0$, then $\sum_{j<\ell} 1/r(j) \ge 1$. 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.69) over $j < \ell$ and using the fact that $\sum_{j<\ell} \lambda^*(j) = 1 - \lambda^*(\ell)$, we get

$$1 - \lambda^*(\ell) \ge 1 - \lambda^*(\ell) \sum_{j<\ell} 1/r(j).$$

If $\lambda^*(\ell) > 0$, this shows that $\sum_j 1/r(j) \ge 1$. 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.69), 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 (g).

e) In (c), $\lambda^*(j)$ 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:

$$\sum_{j<\ell} \frac{p(j)}{\lambda^*(j)r(j) + \lambda^*(\ell)} \quad \leq \quad 1; \qquad \text{with equality if } \lambda^*(\ell) > 0 \tag{A.72}$$

$$\sum_{j<\ell} \frac{p(j)}{\max\left[p(j)r(j),\ \lambda^*(\ell)\right]} \quad \leq \quad 1; \qquad \text{with equality if } \lambda^*(\ell) > 0. \tag{A.73}$$

Show from (A.73) 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.73) 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 ℓ) is

$$\mathsf{E}\left[\frac{X(\ell)}{\sum_k \lambda^*(k)X(k)}\right] \leq 1; \qquad \text{with equality if } \lambda^*(\ell) > 0. \tag{A.74}$$

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.74) reduces to (A.72). Now if we multiply both sides of (A.70) 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.72) into (A.73). 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.73) is simply $p(j)r(j)$ for each j and (A.73) becomes $\sum_{j<\ell} 1/r(j) \leq 1$. The logical obverse is that $\sum_{j<\ell} 1/r(j) > 1$ implies that $\lambda^*(\ell) > \min_j \big(p(j)r(j)\big)$, as was to be shown.

Finally, we must show that if $\sum_{j<\ell} 1/r(j) > 1$, then (A.73) has a unique solution for $\lambda^*(\ell)$. The left side of (A.73), viewed as a function of $\lambda^*(\ell)$, is $\sum_{j<\ell} 1/r(j) > 1$ for $\lambda^*(\ell) = \min_j \big(p(j)r(j)\big)$. This function is continuous and strictly decreasing with further increases in $\lambda^*(\ell)$ 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.73) is satisfied.

It is interesting to observe from (A.70) 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 $\mathsf{E}\left[X(j)\right] < \lambda^*(\ell)$. This is in marked contrast to the case in (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.73) 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.73) is $\sum_{j<\ell} 1/r(j) = 1$ over this range. Thus the inequality in (A.73) is satisfied for all $\lambda^*(\ell)$ in this range. From (A.204), $\lambda^*(j) = p(j) - \lambda^*(\ell)/r(j)$ can be used for each $j < \ell$, to see that all the necessary and sufficient conditions are satisfied for maximizing $\mathsf{E}\left[Y(\boldsymbol{\lambda})\right]$.

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 $\mathsf{E}[Y(\boldsymbol{\lambda})]$. With $\lambda(3) = 0$, we have

$$\mathsf{E}[Y(\boldsymbol{\lambda})] = \frac{1}{2}\ln[\lambda(1)r(1)] + \frac{1}{2}\ln[\lambda(2)r(2)] = \frac{1}{2}\ln[\lambda(1)r(1)\big(1-\lambda(1)\big)r(2)]. \tag{A.75}$$

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.75) 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

$$W_n = \big(\lambda(1)\big)^{n/2}\big(1-\lambda(1)\big)^{n/2}\big(r(1)\big)^{n/2}\big(r(2)\big)^{n/2},$$

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 $\boldsymbol{\lambda}^*$ 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 $\mathsf{E}[X(2)] = p(2)r(2) < 1$, *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 $\boldsymbol{\lambda}^* = (1/2, 1/2, 0)^\mathsf{T}$ and then $(1/4, 0, 3/4)^\mathsf{T}$ into the necessary and sufficient conditions; each satisfies those conditions. Approach 2: Note that $1/r(1) + 1/r(2) = 1$. Thus, from (f), each of these values is satisfied.

Both choices of $\boldsymbol{\lambda}^*$ here lead to the same rv, *i.e.*, $Y(\boldsymbol{\lambda}^*) = \ln[3/2]$ for the event that horse 1 wins and $Y(\boldsymbol{\lambda}^*) = \ln[3/4]$ for the event that horse 2 wins. In other words, the maximizing rv $Y(\boldsymbol{\lambda}^*)$ is uniquely specified, even though $\boldsymbol{\lambda}^*$ is not unique. All points on the straight line between these two choices of $\boldsymbol{\lambda}^*$, *i.e.*, $(1/2 - \alpha, 1/2 - 2\alpha, 3\alpha)^\mathsf{T}$ for $0 \le \alpha \le 1/4$ also lead to the same optimizing $Y(\boldsymbol{\lambda}^*)$.

For $r(2) > 3/2$, we have $1/r(1) + 1/r(2) < 1$, so from (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 $\boldsymbol{\lambda}^*$ into the necessary and sufficient conditions, first with $r(2) > 3/2$ and then with $r(2) < 3/2$.

From (e), the choice $\boldsymbol{\lambda}^* = (1/4, 0, 3/4)$ 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 $\mathsf{E}[Y(\lambda)]$ 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 $\{1, \dots, \ell - 1\}$. Show that $\mathsf{E}[Y(\boldsymbol{\lambda}^*)] = D(\mathsf{p} \,\|\, \mathsf{q})$ for the conditions of (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 $\{p(j);\ 1 \le j < \ell\}$, then has 'inside information' if $p(j) \ne q(j)$, and can establish a positive rate of return equal to $D(\mathsf{p} \,\|\, \mathsf{q})$.

Solution: From (f), $\Big(p(1), \dots, p(\ell-1), 0\Big)^\mathsf{T}$ is one choice for the optimizing $\boldsymbol{\lambda}^*$. Using this choice,

$$\mathsf{E}[Y(\boldsymbol{\lambda}^*)] = \sum_{j<\ell} p(j)\ln[p(j)r(j)] = D(\mathsf{p} \,\|\, \mathsf{q}).$$

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 expected return from the betting.

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 $\mathsf{E}\left[Y_n \mid Z_{n+m-1} = z_{n+m-1}, Z_{n+m-2} = z_{n+m-2}, \dots, Z_m = z_m, \dots, Z_1 = z_1\right] = z_{n+m-1} - z_m$.

Solution: This is more straightforward if the desired result is written in the more abbreviated form

$$\mathsf{E}\left[Y_n \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_m, \dots, Z_1\right] = Z_{n+m-1} - Z_m.$$

Since $Y_n = Z_{n+m} - Z_m$, the left side above is

$$\mathsf{E}\left[Z_{n+m} - Z_m | Z_{n+m-1}, \dots, Z_1\right] = Z_{n+m-1} - \mathsf{E}\left[Z_m \mid Z_{n+m-1}, \dots, Z_m, \dots, Z_1\right].$$

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 $\mathsf{E}\left[Y_n \mid Z_{n+m-1}, \dots, Z_1\right] = Z_{n+m-1} - Z_m$.

b) Show that $\mathsf{E}\left[Y_n \mid Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1\right] = y_{n-1}$.

Solution: In abbreviated form, we want to show that $\mathsf{E}\left[Y_n \mid Y_{n-1}, \dots, Y_1\right] = Y_{n-1}$. We showed in (a) that $\mathsf{E}\left[Y_n \mid Z_{n+m-1}, \dots, Z_1\right] = Y_{n-1}$. For each sample point ω, $Y_{n-1}(\omega), \dots, Y_1(\omega)$ is a function of $Z_{n+m-1}(\omega), \dots, Z_1(\omega)$. Thus, the rv $\mathsf{E}\left[Y_n \mid Z_{n+m-1}, \dots, Z_1\right]$ specifies the rv $\mathsf{E}\left[Y_n \mid Y_{n-1}, \dots, Y_1\right]$, which then must be Y_{n-1}.

c) Show that $\mathsf{E}\left[|Y_n|\right] < \infty$. 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, $\mathsf{E}\left[|Z_n|\right] < \infty$ for each n so

$$\mathsf{E}\left[|Y_n|\right] \leq \mathsf{E}\left[|Z_{n+m}|\right] + \mathsf{E}\left[|Z_m|\right] < \infty.$$

A.10 Solutions for Chapter 10

Exercise 10.1: **a)** Consider the joint probability density $\mathsf{f}_{X,Z}(x,z) = e^{-z}$ for $0 \le x \le z$ and $\mathsf{f}_{X,Z}(x,z) = 0$ otherwise. Find the pair x, z of values that maximize this density. Find the marginal density $\mathsf{f}_Z(z)$ 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 $z \ge 0$. This is maximized at $z = 1$.

b) Let $\mathsf{f}_{X,Z,Y}(x,z,y)$ be y^2e^{-yz} for $0 \le x \le z$, $1 \le y \le 2$ and be 0 otherwise. Conditional on an observation $Y = y$, find the joint MAP estimate of X, Z. Find $\mathsf{f}_{Z|Y}(z|y)$, 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 $0 \le x \le z$, that maximizes $\mathsf{f}_{X,Z|Y}(x,z|y) = \mathsf{f}_{X,Z,Y}(x,z,y)/\mathsf{f}_Y(y) = (y^2e^{-yz})/\mathsf{f}_Y(y)$. For any given y, $0 < y \le 1$, this is maximized, as in (a, for $x = z = 0$. Next, integrating $\mathsf{f}_{X,Z,Y}(x,z,y)$ over x from 0 to z, we get $\mathsf{f}_{Z,Y}(z,y) = y^2ze^{-yz}$. This, and thus $\mathsf{f}_{Z|Y}(z|y)$ is maximized by $z = 1/y$, which is thus the MAP estimate for Z alone.

This shows that MAP estimation on joint rv's does not necessarily agree with the MAP estimates of the individual rv's. This indicates that MAP estimates do not necessarily have the kinds of properties that one would expect in trying to estimate something from an observation.

Exercise 10.2: Let $Y = X + Z$ where X and Z are IID and $\mathcal{N}(0,1)$. Let $U = Z - X$.

a) Explain (using the results of Chapter 3) why Y and U are jointly Gaussian and why they are statistically independent.

Solution: Since Y and U are both linear combinations of X, Z, they are jointly Gaussian by definition 3.3.1. Since $\mathsf{E}[YU] = 0$, *i.e.*, Y and U are uncorrelated, and since they are jointly Gaussian, they are independent

b) Without using any matrices, write down the joint probability density of Y and U. Verify your answer from (3.22).

Solution: Since Y and U are independent and are each $\mathcal{N}(0, \sqrt{2})$, the joint density is

$$\mathsf{f}_{YU}(y,u) = \frac{1}{4\pi}\exp\left[\frac{-y^2 - u^2}{4}\right].$$

Since the covariance matrix of Y and U is $[K] = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$, this is the same as (3.22).

c) Find the MMSE estimate $\hat{x}(y)$ of X conditional on a given sample value y of Y. You can derive this from first principles, or use (10.9) or Example 10.2.2.

Solution: From first principles, $\hat{x}(y) = \mathsf{E}[X|Y=y]$. To find this, we first find $\mathsf{f}_{X|Y}(x|y)$.

$$\mathsf{f}_{X|Y}(x|y) = \frac{\mathsf{f}_{Y|X}(y|x)\mathsf{f}_X(x)}{\mathsf{f}_Y(y)} = \frac{1}{\sqrt{\pi}}\exp\left(-\frac{(y-x)^2}{2} - \frac{x^2}{2} + \frac{-y^2}{4}\right)$$
$$= \frac{1}{\sqrt{\pi}}\exp\big(-(x-y/2)^2\big).$$

Thus, given $Y=y$, $X \sim \mathcal{N}(y/2, 1/\sqrt{2})$. It follows that $\hat{x}(y) = \mathsf{E}[X|Y=y] = y/2$.

d) Show that the estimation error $\xi = X - \widehat{X}(Y)$ is equal to $-U/2$.

Solution: From (c), $\widehat{X}(Y) = Y/2$ so

$$\xi = X - \widehat{X}(Y) = X - (X+Z)/2 = (X-Z)/2 = -U/2.$$

e) Write down the probability density of U conditional on $Y=y$ and that of ξ conditional on $Y=y$.

Solution: Since U and Y are statistically independent,

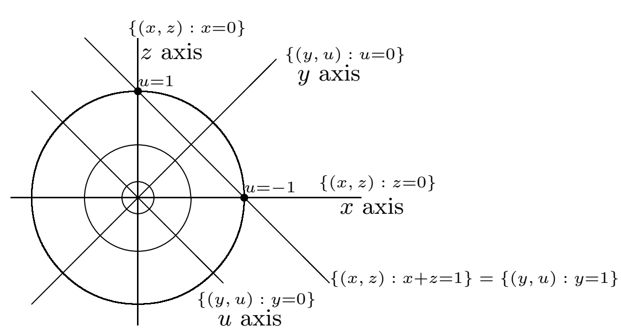
$$\mathsf{f}_{U|Y}(u|y) = \mathsf{f}_U(u) = \frac{1}{\sqrt{4\pi}}\exp\big(-u^2/4\big).$$

Since $\xi = U/2$ (or since ξ, conditional on $Y=y$, is the fluctuation of X, conditional on $Y=y$),

$$\mathsf{f}_{\xi|Y}(\xi|y) = \mathsf{f}_\xi(\xi) = \frac{1}{\sqrt{\pi}}\exp\big(-\xi^2\big).$$

f) Draw a sketch, in the x, z plane of the equiprobability contours of X and Z. Explain why these are also equiprobability contours for Y and U. For some given sample value of Y, say $Y=1$, illustrate the set of points for which $x+z=1$. For a given point on this line, illustrate the sample value of U.

Solution: The circles below are equiprobability contours of X, Z. Since $y^2+u^2 = 2(x^2+z^2)$, they are also equiprobable contours of Y, U.



The point of the problem, in associating the estimation error with $-u/2$, is to give a graphical explanation of why the estimation error is independent of the estimate. Y and U are independent since the y axis and the u axis are at 45° rotations from the the x and y axes.

Exercise 10.3: **a)** Let $X, Z_1, Z_2, \dots, Z_n$ be independent zero-mean Gaussian rv's with variances $\sigma_X^2, \sigma_{Z_1}^2, \dots \sigma_{Z_n}^2$ respectively. Let $Y_j = h_j X + Z_j$ for $j \geq 1$ and let $\boldsymbol{Y} = (Y_1, \dots Y_n)^\mathsf{T}$. Use (10.9) to show that the MMSE estimate of X conditional on $\boldsymbol{Y} = \boldsymbol{y} = (y_1, \dots, y_n)^\mathsf{T}$, is given by

$$\hat{x}(\boldsymbol{y}) = \sum_{j=1}^{n} g_j y_j; \quad \text{where} \quad g_j = \frac{h_j/\sigma_{Z_j}^2}{(1/\sigma_X^2) + \sum_{i=1}^n h_i^2/\sigma_{Z_i}^2}. \tag{A.76}$$

Hint: Let the row vector $\boldsymbol{g}^\mathsf{T}$ be $[K_{X\cdot\boldsymbol{Y}}][K_{\boldsymbol{Y}}^{-1}]$ and multiply $\boldsymbol{g}^\mathsf{T}$ by $K_{\boldsymbol{Y}}$ to solve for $\boldsymbol{g}^\mathsf{T}$.

Solution: From (10.9), we see that $\hat{x}(\boldsymbol{y}) = \boldsymbol{g}^\mathsf{T}\boldsymbol{y}$ where $\boldsymbol{g}^\mathsf{T} = [K_{X\cdot\boldsymbol{Y}}][K_{\boldsymbol{Y}}^{-1}]$. Since $\boldsymbol{Y} = \boldsymbol{h}X + \boldsymbol{Z}$, we see that $[K_{\boldsymbol{Y}}] = \boldsymbol{h}\sigma_X^2\boldsymbol{h}^\mathsf{T} + [K_{\boldsymbol{Z}}]$ and $[K_{X\cdot\boldsymbol{Y}}] = \sigma_X^2\boldsymbol{h}^\mathsf{T}$. Thus we want to solve the vector equation $\boldsymbol{g}^\mathsf{T}\boldsymbol{h}\sigma_X^2\boldsymbol{h}^\mathsf{T} + \boldsymbol{g}^\mathsf{T}[K_{\boldsymbol{Z}}] = \sigma_X^2\boldsymbol{h}^\mathsf{T}$. Since $\boldsymbol{g}^\mathsf{T}\boldsymbol{h}$ is a scalar, we can rewrite this as $(1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{h})\sigma_X^2\boldsymbol{h}^\mathsf{T} = \boldsymbol{g}^\mathsf{T}[K_{\boldsymbol{Z}}]$. The jth component of this equation is

$$g_j = \frac{(1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{h})\sigma_X^2 h_j}{\sigma_{Z_j}^2}. \tag{A.77}$$

This shows that the weighting factors g_j in $\hat{x}(\boldsymbol{y})$ depend on j only through h_j/σ_{Z_j}, which is reasonable. We still must determine the unknown constant $1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{h}$. To do this, multiply (A.77) by h_j and sum over j, getting

$$\boldsymbol{g}^\mathsf{T}\boldsymbol{h} = (1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{h}) \sum_j \frac{\sigma_X^2 h_j^2}{\sigma_{Z_j}^2}.$$

Solving for $\boldsymbol{g}^\mathsf{T}\boldsymbol{h}$, from this,

$$\boldsymbol{g}^\mathsf{T}\boldsymbol{h} = \frac{\sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}; \qquad 1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{h} = \frac{1}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2}. \tag{A.78}$$

Substituting the expression for $1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{y}$ into (A.77) yields (A.76).

b) Let $\xi = X - \widehat{X}(\boldsymbol{Y})$ 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, $\sigma_\xi^2 = \mathsf{E}[\xi X] = \sigma_X^2 - \mathsf{E}\left[\widehat{X}(\boldsymbol{Y})X\right]$. Since $\widehat{X}(\boldsymbol{Y}) = \sum_j g_j Y_j$ from (A.76), we have

$$\begin{aligned}
\sigma_\xi^2 &= \sigma_X^2 - \sum_{i=1}^n g_i \mathsf{E}[Y_i X] \;=\; \sigma_X^2\left(1 - \sum_{i=1}^n g_i h_i\right) \\
&= \sigma_X^2(1 - \boldsymbol{g}^\mathsf{T}\boldsymbol{h}) \;=\; \frac{\sigma_X^2}{1 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2} \;=\; \frac{1}{1/\sigma_X^2 + \sum_j \sigma_X^2 h_j^2/\sigma_{Z_j}^2},
\end{aligned}$$

where we have used (A.78). This is equivalent to (10.29).

c) Show that (10.28), *i.e.*, $\hat{x}(\boldsymbol{y}) = \sigma_\xi^2 \sum_{j=1}^n h_j y_j/\sigma_{Z_j}^2$, is valid.

Solution: Substitute the expression for σ_ξ^2 in (b) into (A.76).

d) Show that the expression in (10.29) is equivalent to the iterative expression in (10.25).

Solution: First, we show that (10.29) implies (10.27). We use ξ_n to refer to the error for n observations and ξ_{n-1} for the error using the first $n-1$ of those observations. Using (10.29),

$$\begin{aligned} \frac{1}{\sigma^2_{\xi_n}} &= \frac{1}{\sigma^2_X} + \sum_{i=1}^{n} \frac{h_i^2}{\sigma^2_{Z_i}} = \frac{1}{\sigma^2_X} + \sum_{i=1}^{n-1} \frac{h_i^2}{\sigma^2_{Z_i}} + \frac{h_n^2}{\sigma^2_{Z_n}} \\ &= \frac{1}{\sigma^2_{\xi_{n-1}}} + \frac{h_n^2}{\sigma^2_{Z_n}},, \end{aligned} \tag{A.79}$$

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

$$\hat{x}(y_1^n) = \sigma^2_{\xi_n} \sum_{j=1}^{n-1} \frac{h_j y_j}{\sigma^2_{Z_j}} + \sigma^2_{\xi_n} \frac{h_n y_n}{\sigma^2_{Zn}} = \frac{\sigma^2_{\xi_n}}{\sigma^2_{\xi_{n-1}}} \hat{x}(y_1^{n-1}) + \sigma^2_{\xi_n} \frac{h_n y_n}{\sigma^2_{Z_n}}, \tag{A.80}$$

where we used (10.28) for $\hat{y}_1^{n-1}$. We can solve for $\sigma^2_{\xi_n}/\sigma^2_{\xi_{n-1}}$ by multiplying (A.79) by $\sigma^2_{\xi_n}$, getting

$$\frac{\sigma^2_{\xi_n}}{\sigma^2_{\xi_{n-1}}} = 1 - \frac{\sigma^2_{\xi_n} h_n^2}{\sigma^2_{Z_n}}.$$

Substituting this into (A.80) yields

$$\hat{x}(y_1^n) = \hat{x}(y_1^{n-1}) + \sigma^2_{\xi_n} \frac{h_n y_n - h_n^2 \hat{x}(y_1^{n-1})}{\sigma^2_{Z_n}}.$$

Finally, if we invert (A.79), we get

$$\sigma^2_{\xi_n} = \frac{\sigma^2_{\xi_{n-1}} \sigma^2_{Z_n}}{h_n^2 \sigma^2_{\xi_{n-1}} + \sigma^2_{Z_n}}.$$

Substituting this into (A.80), we get (10.27).

Exercise 10.5: **a)** Assume that $X_1 \sim \mathcal{N}(\overline{X}_1, \sigma^2_{X_1})$ and that for each $n \geq 1$, $X_{n+1} = \alpha X_n + W_n$ where $0 < \alpha < 1$, $W_n \sim \mathcal{N}(0, \sigma^2_W)$, and $X_1, W_1, W_2, \dots$, are independent. Show that for each $n \geq 1$,

$$\mathsf{E}[X_n] = \alpha^{n-1}\overline{X}_1; \qquad \sigma^2_{X_n} = \frac{(1-\alpha^{2(n-1)})\sigma^2_W}{1-\alpha^2} + \alpha^{2(n-1)}\sigma^2_{X_1}.$$

Solution: For each $n > 1$, $\mathsf{E}[X_n] = \alpha \mathsf{E}[X_{n-1}]$, so by iteration, $\mathsf{E}[X_n] = \alpha^{n-1}\overline{X}_1$. Similarly,

$$\begin{aligned} \sigma^2_{X_n} &= \alpha^2 \sigma^2_{X_{n-1}} + \sigma^2_W = \alpha^2 \Big[\alpha^2 \sigma^2_{X_{n-2}} + \sigma^2_W \Big] + \sigma^2_W && \text{(A.81)} \\ &= \alpha^4 \sigma^2_{X_{n-2}} + (1+\alpha^2)\sigma^2_W = \cdots \\ &= \alpha^{2(n-1}\sigma^2_{X_1} + \big(1 + \alpha^2 + \alpha^4 + \cdots \alpha^{2(n-2)}\big)\sigma^2_W \\ &= \frac{(1-\alpha^{2(n-1)})\sigma^2_W}{1-\alpha^2} + \alpha^{2(n-1)}\sigma^2_{X_1}. && \text{(A.82)} \end{aligned}$$

b) Show directly, by comparing the equation $\sigma^2_{X_n} = \alpha^2\sigma^2_{X_{n-1}} + \sigma^2_W$ for each two adjoining values of n that $\sigma^2_{X_n}$ moves monotonically from $\sigma^2_{X_1}$ to $\sigma^2_W/(1-\alpha^2)$ as $n \to \infty$.

Solution: View $\sigma^2_{X_n} = \alpha^2\sigma^2_{X_{n-1}} + \sigma^2_W$ as a function of $\sigma^2_{X_{n-1}}$ for fixed α^2 and σ^2_W. This is clearly a monotonic increasing function since $\alpha^2 > 0$ and $\sigma^2_W > 0$. Thus if $\sigma^2_{X_{n-1}}$ is replaced by $\sigma^2_{X_n}$ and $\sigma^2_{X_{n-1}} < \sigma^2_{X_n}$, then $\sigma^2_{X_n} < \sigma^2_{X_{n+1}}$. By induction, then, $\sigma^2_{X_n}$ is increasing in n. Similarly, if $\sigma^2_{X_{n-1}} > \sigma^2_{X_n}$, then $\sigma^2_{X_n}$ is decreasing in n, so either way $\sigma^2_{X_n}$ is monotonic in n.

It is easy to see from (A.82) that $\sigma^2_{X_n}$ approaches $\sigma^2_W/(1-\alpha^2)$ as $n \to \infty$. We can then conclude that $\sigma^2_{X_n}$ is increasing in n to $\lim \sigma^2_W/(1-\alpha^2)$ if $\sigma^2_{X_1} < \sigma^2_W/(1-\alpha^2)$ and decreasing to that limit if the inequality is reversed.

c) Assume that sample values of $Y_1, Y_2, \dots,$ are observed, where $Y_n = hX_n + Z_n$ and where $Z_1, Z_2, \dots,$ are IID zero-mean Gaussian rv's with variance σ^2_Z. Assume that $Z_1, \dots, W_1, \dots, X_1$ are independent and assume $h \geq 0$. Rewrite the recursion for the variance of the estimation error in (10.41) for this special case. Show that if $h/\sigma_Z = 0$, then $\sigma^2_{\xi_n} = \sigma^2_{X_n}$ for each $n \geq 1$. Hint: Compare the recursion in (b) to that for $\sigma^2_{\xi_n}$.

Solution: Rewriting (10.41) for this special case,

$$\frac{1}{\sigma^2_{\xi_n}} = \frac{1}{\alpha^2\sigma^2_{\xi_{n-1}} + \sigma^2_W} + \frac{h^2}{\sigma^2_Z}. \tag{A.83}$$

If $h/\sigma_Z = 0$, this simplifies to

$$\sigma^2_{\xi_n} = \alpha^2\sigma^2_{\xi_{n-1}} + \sigma^2_W. \tag{A.84}$$

This is the same recursion as (A.81) with $\sigma^2_{X_n}$ replaced by $\sigma^2_{\xi_n}$. Now for $n = 1$, $\sigma^2_{\xi_1}$ is the variance of the error in the MMSE estimate of X_1 with no measurement, *i.e.*, $\sigma^2_{X_1} = \sigma^2_{\xi_1}$ (this is also clear from (10.33)). Thus from the recursion in (A.84), $\sigma^2_{\xi_n} = \sigma^2_{X_n}$ for all $n \geq 1$.

d) Show from the recursion that $\sigma^2_{\xi_n}$ is a decreasing function of h/σ_Z for each $n \geq 2$. Use this to show that $\sigma^2_{\xi_n} \leq \sigma^2_{X_n}$ for each n. Explain (without equations) why this result must be true.

Solution: From (10.33), $\sigma^2_{\xi_1}$ is decreasing in h. We use this as the basis for induction on (A.83), *i.e.*, as h/σ_z increases, $\sigma^2_{\xi_{n-1}}$ decreases by the inductive hypothesis, and thus from (A.83), $\sigma^2_{\xi_n}$ also decreases. Since $\sigma^2_{\xi_n} = \sigma^2_{X_n}$ for $h = 0$, we must then have $\sigma^2_{\xi_n} \leq \sigma^2_{X_n}$ for $h > 0$. This must be true because one possible estimate for X_n is $\mathsf{E}[X_n]$, *i.e.*, the mean of X_n in the absence of measurements. The error variance is then $\sigma^2_{X_n}$, which must be greater than or equal to the variance with a MMSE estimate using the measurements.

e) Show that the sequence $\{\sigma^2_{\xi_n}; n \geq 1\}$ is monotonic in n. Hint: Use the same technique as in (b). From this and (d), show that $\lambda = \lim_{n\to\infty} \sigma^2_{\xi_n}$ exists. Show that the limit satisfies (10.42) (note that (10.42) must have 2 roots, one positive and one negative, so the limit must be the positive root).

Solution: View (A.83) as a monotonic increasing function of $\sigma^2_{\xi_{n-1}}$, viewing all other quantities as constants. Thus if $\sigma^2_{\xi_{n-1}}$ is replaced by $\sigma^2_{\xi_n}$ and $\sigma^2_{\xi_{n-1}} < \sigma^2_{\xi_n}$, then $\sigma^2_{\xi_n} < \sigma^2_{\xi_{n+1}}$. By induction, then, $\sigma^2_{\xi_n}$ is increasing in n. Similarly, if $\sigma^2_{\xi_{n-1}} > \sigma^2_{\xi_n}$, then $\sigma^2_{X_n}$ is decreasing in n, so either way $\sigma^2_{\xi_n}$ is monotonic in n. From (b), $\sigma^2_{X_n}$ is also bounded independent of n, so $\lambda = \lim_{n\to\infty} \sigma^2_{\xi_n}$ must exist. Eq. (10.42) is simply a rearrangement of (A.83) replacing $\sigma^2_{\xi_n}$ and $\sigma^2_{\xi_{n-1}}$ by λ, yielding

$$\alpha^2 h^2 \sigma_Z^{-2} \lambda^2 + \big[h^2\sigma_Z^{-2})\sigma_W^2 + (1-\alpha^2\big]\lambda - \sigma_W^2 = 0. \tag{A.85}$$

This must be satisfied by $\lambda = \lim_{n\to\infty} \sigma^2_{\xi_n}$. Since each term within the brackets above is positive, the coefficient of λ is positive. Since the constant term is negative, this quadratic equation must have two solutions, one positive and one negative. The limiting error variance is, of course, the positive term, and thus can be found uniquely simply by solving this equation.

f) Show that for each $n \geq 1$, $\sigma^2_{\xi_n}$ is increasing in σ^2_W and increasing in α. Note: This increase with α is surprising, since when α is close to one, X_n changes slowly, so we would expect to be able to track X_n well. The problem is that $\lim_n \sigma^2_{X_n} = \sigma^2_W/(1-\alpha^2)$ so the variance of the untracked X_n is increasing without limit as α approaches 1. (g) is somewhat messy, but resolves this issue.

Solution: The argument that $\sigma^2_{\xi_n}$ is increasing in α and increasing in σ^2_W is the same as the argument in (d).

g) Show that if the recursion is expressed in terms of $\beta = \sigma^2_W/(1-\alpha^2)$ and α, then λ is decreasing in α for constant β.

Solution: If make the suggested substitution in (A.85 and rearrange it, we get

$$\frac{\alpha^2\lambda^2}{1-\alpha^2} + \left(\beta + \frac{\sigma_Z^2}{h^2}\right)\lambda - \frac{\beta\sigma_Z^2}{h^2} = 0.$$

Since the quadratic term is increasing in α over the range $0 < \alpha < 1$), the positive root is decreasing.

Exercise 10.7: **a)** Write out $\mathsf{E}\left[(X - \boldsymbol{g}^{\mathsf{T}}\boldsymbol{Y})^2\right] = \sigma_X^2 - 2[K_{X\cdot\boldsymbol{Y}}]\boldsymbol{g} + \boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}]\boldsymbol{g}$ as a function of $\boldsymbol{g} = (g_1, g_2, \dots, g_n)^{\mathsf{T}}$ and take the partial derivative of the function with respect to g_i for each i, $1 \leq i \leq n$. Show that the vector of these partial derivatives is $-2[K_{X\cdot\boldsymbol{Y}}] + 2\boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}]$.

Solution: Note that $\frac{\partial}{\partial \boldsymbol{g}}[K_{X\cdot\boldsymbol{Y}}]\boldsymbol{g} = [K_{X\cdot\boldsymbol{Y}}]$. For the second term, we take the partial derivative with respect to g_i for any given i, getting

$$\frac{\partial}{\partial g_i}\boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}]\boldsymbol{g} \;=\; [K_{\boldsymbol{Y}}]_{i\cdot}\,\boldsymbol{g} + \boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}]_{\cdot i} \;=\; 2\boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}]_{\cdot i},$$

where $[K_{\boldsymbol{Y}}]_{\cdot i}$ denotes the ith column of $[K_{\boldsymbol{Y}}]$ and $[K_{\boldsymbol{Y}}]_{i\cdot}$ denotes the ith row. We have used the fact that $[K_{\boldsymbol{Y}}]$ is symmetric for the latter equality. Putting these equations into vector form,

$$\frac{\partial}{\partial \boldsymbol{g}}\left[\sigma_X^2 - 2[K_{X\cdot\boldsymbol{Y}}]\boldsymbol{g} + \boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}]\boldsymbol{g}\right] = -2[K_{X\cdot\boldsymbol{Y}}] + 2\boldsymbol{g}^{\mathsf{T}}[K_{\boldsymbol{Y}}].$$

b) Explain why the stationary point here is actually a minimum.

Solution: The stationary point, say $\tilde{\boldsymbol{g}}$, *i.e.*, the point at which this vector partial derivative is 0, is $\tilde{\boldsymbol{g}}^{\mathsf{T}} = [K_{X\cdot\boldsymbol{Y}}K_{\boldsymbol{Y}}^{-1}]$. Taking the transpose, $\tilde{\boldsymbol{g}} = [K_{\boldsymbol{Y}}^{-1}][K_{X\cdot\boldsymbol{Y}}]$. There are two ways of showing that $\tilde{\boldsymbol{g}}$ minimizes $\mathsf{E}\left[(X - \boldsymbol{g}^\mathsf{T}\boldsymbol{Y})^2\right]$ over $\boldsymbol{g}$. For the first of these, recall that $\hat{x}_{\text{MMSE}}(\boldsymbol{y}) = \tilde{\boldsymbol{g}}^\mathsf{T}\boldsymbol{y}$, and thus this stationary point is the MMSE estimate. Thus $\hat{x}_{\text{MMSE}}(\boldsymbol{y})$ is linear in $\boldsymbol{y}$ and thus is the same as the linear least-square estimate, *i.e.*, the minimum over $\boldsymbol{g}$ of $\mathsf{E}\left[(X - \boldsymbol{g}^\mathsf{T}\boldsymbol{Y})^2\right]$.

The second way to show that the stationary point $\tilde{\boldsymbol{g}}$ minimizes $\mathsf{E}\left[(X - \boldsymbol{g}^\mathsf{T}\boldsymbol{Y})^2\right]$ is to note that $\mathsf{E}\left[(X - \boldsymbol{g}^\mathsf{T}\boldsymbol{Y})^2\right]$ is convex in $\boldsymbol{g}$, and thus must be minimized by a stationary point if a stationary point exists.

Exercise 10.10: Let $\boldsymbol{X} = X_1, \dots, X_n)^\mathsf{T}$ be a zero-mean complex r**v** with real and imaginary components $X_{\text{re},j}, X_{\text{im},j}$, $1 \le j \le n$ respectively. Express $\mathsf{E}[X_{\text{re},j}X_{\text{re},k}]$, $\mathsf{E}[X_{\text{re},j}X_{\text{im},k}]$, $\mathsf{E}[X_{\text{im},j}X_{\text{im},k}]$, $\mathsf{E}[X_{\text{im},j}X_{\text{re},k}]$ as functions of the components of $[K_{\boldsymbol{X}}]$ and $\mathsf{E}\left[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\right]$.

Solution: Note that $\mathsf{E}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}]$ is the pseudo-covariance matrix as treated in Section 3.7.2 and the desired covariances of real and imaginary parts are given in (3.100). We simply rewrite those results in the notation of the problem statement..

$$\begin{aligned}
\mathsf{E}[X_{\text{re},j}X_{\text{re},k}] &= \frac{1}{2}\Re\left([K_{\boldsymbol{X}}]_{jk} + [M_{\boldsymbol{X}}]_{jk}\right), \\
\mathsf{E}[X_{\text{re},j}X_{\text{im},k}] &= \frac{1}{2}\Im\left(-[K_{\boldsymbol{X}}]_{jk} + [M_{\boldsymbol{X}}]_{jk}\right), \\
\mathsf{E}[X_{\text{im},j}X_{\text{im},k}] &= \frac{1}{2}\Re\left([K_{\boldsymbol{X}}]_{jk} - [M_{\boldsymbol{X}}]_{jk}\right), \\
\mathsf{E}[X_{\text{im},j}X_{\text{re},k}] &= \frac{1}{2}\Im\left([K_{\boldsymbol{X}}]_{jk} + [M_{\boldsymbol{X}}]_{jk}\right).
\end{aligned}$$

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

$$\mathsf{f}_X(x) = \frac{\exp -x^*x}{\pi}.$$

Recall that the real part and imaginary part each have variance 1/2.

Solution: Since X_{re} and X_{im} are independent and $\sigma^2_{X_{\text{re}}} = \sigma^2_{X_{\text{im}}} = \frac{1}{2}$, we have

$$\mathsf{f}_X(x) = \frac{1}{2\pi\sigma_{X_{\text{re}}}\sigma_{X_{\text{im}}}}\exp\left[\frac{-x^2_{\text{re}}}{2\sigma^2_{X_{\text{re}}}} - \frac{x^2_{\text{im}}}{2\sigma^2_{X_{\text{im}}}}\right] = \frac{\exp -x^*x}{\pi}.$$

b) Let $\boldsymbol{X}$ be an n dimensional circularly symmetric complex Gaussian zero-mean random vector with $K_{\boldsymbol{X}} = I_n$. Show that

$$\mathsf{f}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp -\boldsymbol{x}^\dagger\boldsymbol{x}}{\pi^n}.$$

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).

$$\mathsf{f}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp -\sum_{i=1}^n x_i^*x_i}{\pi^n} = \frac{\exp -\boldsymbol{x}^{*T}\boldsymbol{x}}{\pi^n}.$$

c) Let $\boldsymbol{Y} = H\boldsymbol{X}$ where H is $n \times n$ and invertible. Show that

$$\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\exp\left[-\boldsymbol{y}^{\dagger}(H^{-1})^{\dagger}H^{-1}\boldsymbol{y}\right]}{v\pi^n},$$

where v is $d\boldsymbol{y}/d\boldsymbol{x}$, 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 $\boldsymbol{Y} = H\boldsymbol{X}$, we have $\boldsymbol{X} = H^{-1}\boldsymbol{Y}$, so for any $\boldsymbol{y}$ and $\boldsymbol{x} = H^{-1}\boldsymbol{y}$,

$$\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y})|d\boldsymbol{y}| = \mathsf{f}_{\boldsymbol{X}}(\boldsymbol{x})|d\boldsymbol{x}|.$$

$$\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\mathsf{f}_{\boldsymbol{X}}(H^{-1}\boldsymbol{y})}{|d\boldsymbol{y}|/|d\boldsymbol{x}|}.$$

Substituting the result of b) into this,

$$\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\exp\left[-\boldsymbol{y}^{*T}(H^{*T})^{-1}H^{-1}\boldsymbol{y}\right]}{|d\boldsymbol{y}|/|d\boldsymbol{x}|}.$$

d) Use this to show that that (3.108) is valid.

Solution: View $|d\boldsymbol{x}|$ as an incremental volume in $2n$ dimensional space (n real and n imaginary components) and view $|d\boldsymbol{y}|$ as the corresponding incremental volume for $\boldsymbol{y} = H\boldsymbol{x}$, i.e.,

$$\begin{bmatrix} \boldsymbol{y}_{\mathrm{re}} \\ \boldsymbol{y}_{\mathrm{im}} \end{bmatrix} = \begin{bmatrix} H_{\mathrm{re}} & -H_{\mathrm{im}} \\ H_{\mathrm{im}} & H_{\mathrm{re}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{\mathrm{re}} \\ \boldsymbol{x}_{\mathrm{im}} \end{bmatrix}.$$

We then have

$$\begin{aligned} \frac{|d\boldsymbol{y}|}{|d\boldsymbol{x}|} &= \left|\det\begin{bmatrix} H_{\mathrm{re}} & -H_{\mathrm{im}} \\ H_{\mathrm{im}} & H_{\mathrm{re}} \end{bmatrix}\right| = \left|\det\begin{bmatrix} H_{\mathrm{re}} + iH_{\mathrm{im}} & iH_{\mathrm{re}} - H_{\mathrm{im}} \\ H_{\mathrm{im}} & H_{\mathrm{re}} \end{bmatrix}\right| \\ &= \left|\det\begin{bmatrix} H_{\mathrm{re}} + iH_{\mathrm{im}} & 0 \\ H_{\mathrm{im}} & H_{\mathrm{re}} - iH_{\mathrm{im}} \end{bmatrix}\right| = |\det[H]\det[H^*]| = \det[K_{\boldsymbol{Y}}]. \end{aligned}$$

Exercise 10.14: **a)** Let $Y = X^2 + Z$ where Z is a zero-mean unit variance Gaussian random variable. Show that no unbiased estimate of X exists from observation of Y. Hint. Consider any $x > 0$ and compare with $-x$.

Solution: Let $x > 0$ be arbitrary. Then $\mathsf{f}_{Y|X}(y|x) = \mathsf{f}_{Z|X}(y - x^2)$. Similarly, $\mathsf{f}_{Y|X}(y|-x) = \mathsf{f}_{Z|X}(y - x^2)$. Thus for all choices of parameter x, $\mathsf{f}_{Y|X}(y|x) = \mathsf{f}_{Y|X}(y|-x)$. It follows that

$$\mathsf{E}\left[\hat{X}(Y)|X = x\right] \;=\; \int \mathsf{f}_{Y|X}(y|x)\hat{x}(y)\,dy \;=\; \mathsf{E}\left[\hat{X}(Y)|X = -x\right].$$

As a consequence,

$$\begin{aligned} b_{\hat{x}}(x) &= \mathsf{E}\left[\hat{X}(Y){-}X|X{=}x\right] \;=\; \mathsf{E}\left[\hat{X}(Y)|X{=}x\right]-x \\ b_{\hat{x}}(-x) &= \mathsf{E}\left[\hat{X}(Y)|X{=}x\right]+x \;=\; b_{\hat{x}}(x)+2x. \end{aligned}$$

Thus for each $x \neq 0$, if $\hat{(}x)$ is unbiased at $X = x$, then it must be biased at $X = -x$ and vice-versa.

This is not a bizarre type of example; it is unusual only in its extreme simplicity. The underlying problem here occurs whether or not one assumes an *a priori* distribution for X, but it is easiest to understand if X has a symmetric distribution around 0. In this case, Y is independent of the sign of X, so one should not expect any success in estimating the sign of X from Y. The point of interest here, however, is that bias is not a very fundamental quantity. Estimation usually requires some sort of tradeoff between successful estimation of different values of x, and focussing on bias hides this tradeoff.

b) Let $Y = X + Z$ where Z is uniform over $(-1, 1)$ and X is a parameter lying in (-1, 1). Show that $\hat{x}(y) = y$ is an unbiased estimate of x. Find a biased estimate $\hat{x}_1(y)$ for which $|\hat{x}_1(y) - x| \leq |\hat{x}(y) - x|$ for all x and y with strict inequality with positive probability for all $x \in (-1, 1)$.

Solution: Choosing $\hat{x}(y) = y$ implies that $\mathsf{E}\left[\hat{X}(Y)|X{=}x\right] = \mathsf{E}\,[X + Z)|X = x] = x$ since $\overline{Z} = 0$. Thus $b_{\hat{x}}(x) = 0$ for all x, $-1 < x < 1$ and $\hat{x}(y)$ is unbiased. This is clearly a stupid estimation rule, since whenever $y > 1$, the rule chooses $\hat{x}$ to be larger than any possible x. Reducing the estimate to 1 whenever $y > 1$ clearly reduces the error $|\hat{x}(y) - x|$ for all x when $y > 1$. Increasing the estimate to -1 when $y < -1$ has the same advantage. Thus

$$b_1\hat{x}(y) \;=\; \begin{cases} \;\;1; & y \geq 1 \\ \;\;y; & -1 < y < 1 \\ -1; & y \leq -1 \end{cases}.$$

satisfies the desired conditions.

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 $N(0, \sigma^2)$. We have $f(y|x) = (2\pi\sigma^2)^{-1/2}\exp(-(y{-}x)^2/(2\sigma^2))$. Thus

$$\partial f(y|x)/\partial x = [(y - x)/\sigma^2](2\pi\sigma^2)^{-1/2}\exp(-(y - x)^2/(2\sigma^2)).$$

Dividing by $f(y|x)$, we see that $V_x(y) = (y - x)/\sigma^2$. Then the random variable $V_x(Y)$, conditional on x, is $(Y - x)/\sigma^2$. Since $Y \sim N(x, \sigma^2)$, the variance of $(Y - x)/\sigma^2$, conditional on x, is $\sigma^2/\sigma^4 = \sigma^{-2}$. By definition (see (10.107)), this is the Fisher information, *i.e.*, $J(X) = 1/\sigma^2$.

b) Show that for ML estimation, the bias is 0 for all x. Show that the Cramer-Rao bound is satisfied with equality for all x in this case.

Solution: For the ML estimate, $\hat{x}_{\mathrm{ML}}(y) = y$, and for a given x,

$$\mathsf{E}\left[\hat{X}_{\mathrm{ML}}(Y) \mid X = x\right] = \mathsf{E}\left[Y \mid X = x\right] = x.$$

Thus $b_{\hat{x}_{\mathrm{ML}}}(x) = 0$ and the the estimate is unbiased. The estimation error is just Z, so the mean-square estimation error, $\mathsf{E}\left[\left(\hat{X}_{\mathrm{ML}}(Y) - x\right)^2 | X = x\right]$, is σ^2 for each x. Combining this with (a),

$$\mathsf{E}\left[\left(\hat{X}(Y) - x\right)^2 | X = x\right] = 1/J(x) \qquad \text{for each } x.$$

Comparing with (10.113), the Cramer-Rao bound for an unbiased estimator is met with equality for all x by using the ML estimate.

c)Consider using the MMSE estimator for the *a priori* distribution $X \sim \mathcal{N}(0, \sigma_X^2)$. Show that the bias satisfies $b_{\hat{x}_{\mathrm{MMSE}}}(x) = -x\sigma^2/(\sigma^2 + \sigma_X^2)$.

Solution: For the MMSE estimator, $\hat{x}(y) = \frac{y\,\sigma_X^2}{\sigma^2+\sigma^2}$. Thus, since $\mathsf{E}\left[Y|X{=}x\right] = x + Z$, we have $\mathsf{E}\left[\hat{X}(Y) \mid X{=}x\right] = \frac{x\,\sigma_X^2}{\sigma^2+\sigma_Z^2}$ and

$$b_{\hat{x}_{\mathrm{MMSE}}}(x) = -x\sigma^2/(\sigma^2 + \sigma_X^2).$$

d) Show that the MMSE estimator in (c) satisfies the Cramer-Rao bound with equality for each x. Note that the mean-squared error, as a function of x is smaller than that for the ML case for small x and larger for large x.

Solution: From (10.109), the Cramer-Rao bound for a biased estimate is

$$\begin{aligned}
\mathsf{VAR}\left[\widehat{X}(Y)|X{=}x\right] \quad &\geq \quad \frac{\left[1 + \frac{\partial b_{\hat{X}}(x)}{\partial x}\right]^2}{J(x)} \quad = \quad \frac{\left[1 - \sigma^2/(\sigma^2 + \sigma_X^2)\right]^2}{1/\sigma^2} \\
&= \quad \frac{\left[\sigma_X^2/(\sigma^2 + \sigma_X^2)\right]^2}{1/\sigma^2} \quad = \quad \frac{\sigma^2\sigma_X^4}{(\sigma^2 + \sigma_X^2)^2}. \qquad (A.86)
\end{aligned}$$

Calculating $\mathsf{VAR}\left[\widehat{X}(Y)|X{=}x\right]$ directly,

$$\begin{aligned}
\mathsf{VAR}\left[\widehat{X}(Y)|X{=}x\right] \quad &= \quad \mathsf{VAR}\left[\frac{Y\sigma_X^2}{\sigma^2 + \sigma_X^2} \mid X{=}x\right] \\
&= \quad \mathsf{VAR}\left[\frac{(x + Z)\sigma_X^2}{\sigma^2 + \sigma_X^2} \mid X{=}x\right] \\
&= \quad \mathsf{VAR}\left[\frac{Z\sigma_X^2}{\sigma^2 + \sigma_X^2}\right] \quad = \quad \frac{\sigma^2\sigma_X^4}{(\sigma^2 + \sigma_X^2)^2}. \qquad (A.87)
\end{aligned}$$

Comparing (A.86) and (A.87), we see that the Cramer Rao bound is met with equality for all x. Note also that the variance of the estimation error is smaller for all σ_X^2 than that for the ML estimate, and that this variance shrinks with decreasing σ^2 to 0 as $\sigma_X^2 \to 0$.

This seems surprising until we recognize that the estimate is shrinking toward 0, and even though the variance of the error is small, the magnitude is large for large x.

If we use the Cramer-Rao bound to look at the mean-square error given x rather than the variance given x (see (10.112)), we get

$$\mathsf{E}\left[(\widehat{X}(Y) - x)^2 | X{=}x\right] \;\geq\; \frac{\sigma^2\sigma_X^4}{(\sigma^2+\sigma_X^2)^2} + \frac{\sigma^2 x^4}{(\sigma^2+\sigma_X^2)^2}.$$

We see then that the mean-square error using MMSE estimation is small when $|x|$ is small and large when $|x|$ is large. This is not surprising, since the MMSE estimate chooses $\hat{x}(y)$ to be a fraction smaller than 1 of the observation y.