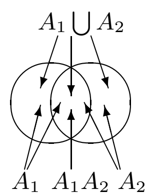
A.1 Solutions for Chapter 1

Exercise 1.1: Let A_1 and A_2 be arbitrary events and show that $\Pr\{A_1 \bigcup A_2\} + \Pr\{A_1A_2\} = \Pr\{A_1\} + \Pr\{A_2\}$. Explain which parts of the sample space are being double-counted on both sides of this equation and which parts are being counted once.

Solution: As shown in the figure below, A_1A_2 is part of $A_1 \bigcup A_2$ and is thus being double counted on the left side of the equation. It is also being double counted on the right (and is in fact the meaning of A_1A_2 as those sample points that are both in A_1 and in A_2).



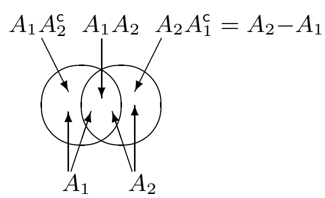
Exercise 1.2: This exercise derives the probability of an arbitrary (non-disjoint) union of events, derives the union bound, and derives some useful limit expressions.

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

Solution: Note that each sample point ω is in A_1 or A_1^c, but not both. Thus each ω is in exactly one of A_1, $A_1^c A_2$ or $A_1^c A_2^c$. 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 $A_1, \ldots, A_n$, define $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$. Show that $B_1, B_2, \ldots$ are disjoint events and show that for each $n \geq 2$, $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$. Hint: Use induction.

Solution: Let $B_1 = A_1$. 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 $C_n = \bigcup_{i=1}^{n} B_i$.

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

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

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

e) Show that $\Pr\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{n\to\infty} \Pr\{\bigcup_{i=1}^{n} A_i\}$. 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),

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

From (b), however,

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

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

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

Solution: Using De 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 1.3: Find the probability that a five card poker hand, chosen randomly from a 52 card deck, contains 4 aces. That is, if all 52! arrangements of a deck of cards are equally likely, what is the probability that all 4 aces are in the first 5 cards of the deck.

Solution: The ace of spades can be in any of the first 5 positions, the ace of hearts in any of the 4 remaining positions out of the first 5, and so forth for the other two aces. The remaining 48 cards can be in any of the remaining 48 positions. Thus there are $(5\cdot4\cdot3\cdot2)48!$ permutations of the 52 cards for which the first 5 cards contain 4 aces. Thus

$$\Pr\{4 \text{ aces}\} = = \frac{5!48!}{52!} = \frac{5\cdot 4\cdot 3\cdot 2}{52\cdot 51\cdot 50\cdot 49} = 1.847\times 10^{-5}.$$

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 $1 \le i \le 3$ 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 1.5: This exercise shows that for all rv's X, $\mathsf{F}_X(x)$ is continuous from the right.

a) For any given rv X, any real number x, and each integer $n \ge 1$, let $A_n = \{\omega : X > x + 1/n\}$, and show that $A_1 \subseteq A_2 \subseteq \cdots$. Use this and the corollaries to the axioms of probability to show that $\Pr\left\{\bigcup_{n\ge1} A_n\right\} = \lim_{n\to\infty}\Pr\{A_n\}$.

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

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

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

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

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

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

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

$$\int_0^\infty \Pr\{X > x\}\, dx = \int_0^\infty x \mathsf{f}_X(x)\, dx. \tag{A.2}$$

Hint 1: First rewrite $\Pr\{X > x\}$ on the left side of (A.2) as $\int_x^\infty \mathsf{f}_X(y)\, dy$. Then think through, to your level of comfort, how and why the order of integration can be interchnged in the resulting expression.

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

$$\begin{aligned}
\int_0^a \mathsf{F}^\mathsf{c}(x)\, dx &= \int_0^a \int_x^\infty \mathsf{f}_X(y)\, dy\, dx \\
&= \int_0^a \int_x^a \mathsf{f}_X(y)\, dy\, dx + \int_0^a \int_a^\infty \mathsf{f}_X(y)\, dy\, dx \\
&= \int_0^a \int_0^y \mathsf{f}_X(y)\, dx\, dy + a\mathsf{F}_X^\mathsf{c}(a).
\end{aligned}$$

We first broke the integral on the right into two parts, one for $y < x$ and the other for $y \geq x$. Since the limits of integration on the first part were finite, they could be interchanged. The inner integral of the first part is $y\mathsf{f}_X(y)$, so

$$\lim_{a\to\infty} \int_0^a \mathsf{F}_X^\mathsf{c}(x)\, dx = \lim_{a\to\infty} \int_0^a y\mathsf{f}_X(y)\, dy + \lim_{a\to\infty} a\mathsf{F}_X^c(a).$$

Assuming that $\mathsf{E}[X]$ exists, the integral on the left is nondecreasing in A and has the finite limit $\overline{X}$. The first integral on the right is also nondecreasing and upper bounded by the first integral, so it also has a limit. This means that $\lim_{a\to\infty} a\mathsf{F}_X^\mathsf{c}(a)$ must also have a limit, say β. Now if $\beta > 0$, then for any $\epsilon \in (0, a)$, $a\mathsf{F}_X(a) > \beta - \epsilon$ for all sufficiently large a. For all such a, then $\mathsf{F}^{\mathsf{c}}_X(a) > (\beta - \epsilon)/a$. This would imply that $\overline{X} = \int_0^\infty \mathsf{F}^{\mathsf{c}}_X(x)\, dx = \infty$, which is a contradiction. Thus $\beta = 0$, *i.e.*, $\lim_{a\to\infty} a\mathsf{F}^{\mathsf{c}}_X(a) = 0$, establishing (A.2) for the case where $\mathsf{E}[X]$ is finite. The case where $\mathsf{E}[X]$ is infinite is a minor perturbation.

The result that $\lim_{a\to\infty} aF^{\mathsf{c}}_X(a) = 0$ is also important and can be seen intuitively from Figure 1.3.

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

Solution: Using integration by parts and being less careful,

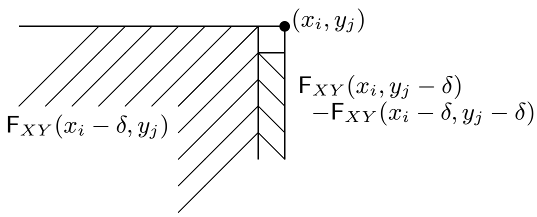
$$\int_0^\infty d\big(x\mathsf{F}^{\mathsf{c}}_X(x)\big) = -\int_0^\infty x\mathsf{f}_X(x)\, dx + \int_0^\infty \mathsf{F}^{\mathsf{c}}_X(x)\, dx.$$

The left side is $\lim_{a\to\infty} a\mathsf{F}^{\mathsf{c}}_X(a) - 0\mathsf{F}_X(0)$ so this shows the same thing, again requiring the fact that $\lim_{a\to\infty} a\mathsf{F}^{\mathsf{c}}_X(a) = 0$ when $\mathsf{E}[X]$ exists.

Exercise 1.7: Suppose X and Y are discrete rv's with the PMF $\mathsf{p}_{XY}(x_i, y_j)$. Show (a picture will help) that this is related to the joint CDF by

$$\mathsf{p}_{XY}(x_i, y_j) = \lim_{\delta>0, \delta\to 0} \left[\mathsf{F}(x_i, y_j) - \mathsf{F}(x_i - \delta, y_j) - \mathsf{F}(x_i, y_j - \delta) + \mathsf{F}(x_i - \delta, y_j - \delta)\right].$$

Solution: The picture below makes this equation obvious. Note that $\mathsf{F}(x_i, y_j)$ is the probability of the quadrant of joint sample values (x, y) that satisfy $(x, y) \le (x_i, y_j)$. The term $\mathsf{F}_{XY}(x_i - \delta, y_j)$ is the probability of the hatched region on the left and $\mathsf{F}_{XY}(x_i, y_j - \delta) - \mathsf{F}_{XY}(x_i - \delta, y_j - \delta)$ is the probability of the hatched region on the right. Thus the expression on the right of the equation is the probability of the δ by δ square to the left and below (x_i, y_j). When this square becomes too small to include any other sample point, the probability of the square is $\mathsf{p}_{XY}(x_i, y_j)$.



Exercise 1.8: A variation of Example 1.5.1 is to let M be a random variable that takes on both positive and negative values with the PMF

$$\mathsf{p}_M(m) = \frac{1}{2|m|\,(|m| + 1)}.$$

In other words, M is symmetric around 0 and $|M|$ has the same PMF as the nonnegative rv N of Example 1.5.1.

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

Solution:

$$\sum_{m\geq 0} m\mathsf{p}_M(m) = \sum_{m\geq 0} \frac{1}{2(|m|+1)} = \infty$$
$$\sum_{m<0} m\mathsf{p}_M(m) = \sum_{m<0} \frac{-1}{2(|m|+1)} = -\infty.$$

b) Suppose that the terms in $\sum_{m=-\infty}^{\infty} m\mathsf{p}_M(m)$ are summed in the order of 2 positive terms for each negative term (*i.e.*, in the order $1, 2, -1, 3, 4, -2, 5, \cdots$). Find the limiting value of the partial sums in this series. Hint: You may find it helpful to know that

$$\lim_{n\to\infty}\left[\sum_{i=1}^{n}\frac{1}{i} - \int_1^n \frac{1}{x}\,dx\right] = \gamma,$$

where γ is the Euler-Mascheroni constant, $\gamma = 0.57721\cdots$.

Solution: The sum after $3n$ terms is

$$\sum_{m=1}^{2n}\frac{1}{2(m+1)} - \sum_{m=1}^{n}\frac{1}{2(m+1)} = \frac{1}{2}\sum_{i=n+2}^{2n+1}\frac{1}{i}.$$

Taking the limit as $n\to\infty$, the Euler-Mascheroni constant cancels out and

$$\lim_{n\to\infty}\frac{1}{2}\sum_{i=n+2}^{2n+1}\frac{1}{i} = \lim_{n\to\infty}\frac{1}{2}\int_{n+2}^{2n+1}\frac{1}{x}\,dx = = \frac{1}{2}\ln 2.$$

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

Solution: This is done the same way, and the answer is $\frac{1}{2}\ln k$. What the exercise essentially shows is that in a sum for which both the positive terms sum to infinity and the negative terms sum to $-\infty$, one can get any desired limit by summing terms in an appropriate order. In fact, to reach any desired limit, one alternates between positive terms until exceeding the desired limit, then negative terms until falling below the desired limit, then positive terms again, etc.

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

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}. \tag{A.3}$$

The ratio, $\sqrt{2\pi n}(n/e)^n/n!$, of the first two terms is monotonically increasing with n toward the limit 1, and the ratio $\sqrt{2\pi n}(n/e)^n\exp(1/12n)/n!$ is monotonically decreasing toward 1. The upper bound is more accurate, but the lower bound is simpler and known as the Stirling approximation. See [8] for proofs and further discussion of the above facts.

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

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

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

a) $p_0 = 0.2,\ p_1 = 0.4,\ p_2 = 0.4$.

b) $p_0 = 0.2,\ p_1 = 0.2,\ p_2 = 0.6$.

c) $p_0 = 0.2,\ p_1 = 0.3,\ p_2 = 0.5$.

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

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

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

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

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

For each of the first 3 examples, there is no α for which $\Pr\{X < \alpha\} = \Pr\{X > \alpha\}$. One should then ask why there must always be an x such that $\Pr\{X \geq x\} \geq 1/2$ and $\Pr\{X \leq x\} \geq 1/2$. To see this, let $x_o = \inf\{x : \mathsf{F}_X(x) \geq 1/2\}$. We must have $\mathsf{F}_X(x_o) \geq 1/2$ since F_X is continuous on the right. Because of the infimum, we must have $\mathsf{F}_X(x_o - \epsilon) < 1/2$ for all $\epsilon > 0$, and therefore $\Pr\{X \geq x_o - \epsilon\} \geq 1/2$. But $\Pr\{X \geq x\}$ is continuous on the left for the same reason that $\mathsf{F}_X(x)$ is continuous on the right, and thus x_o is a median of X. This is the kind of argument that makes many people hate analysis.

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 $|Y|$,

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

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

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

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

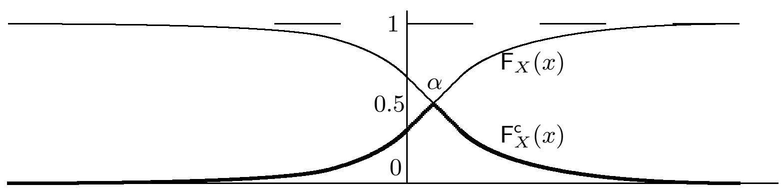
Solution: This follows by changing the variable of integration in (a). That is,

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

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

c) Show that $\mathsf{E}[|X-\alpha|]$ 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 $\mathsf{F}_X(\alpha) \geq 0.5$ 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 $X_j \leq x$ for each j, $1 \leq j \leq n$. Thus

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

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

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.

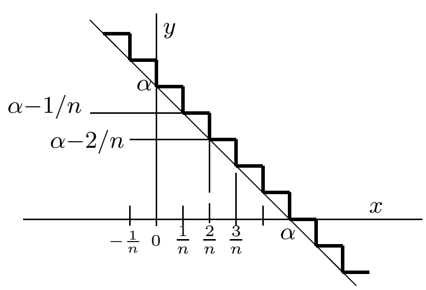
Solution: Note that Z can be infinite (either $\pm\infty$) or undefined only when either X or Y are infinite or undefined. Since these are events of zero probability, Z can be infinite or undefined only with probability 0.

b) We must show that $\{\omega \in \Omega : Z(\omega) \le \alpha\}$ 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 $\{\omega \in \Omega : X(\omega) + Y(\omega) \le \alpha\}$ is an event. Let $B(n,k) = \{\omega : X(\omega) \le k/n\} \bigcap \{Y(\omega) \le \alpha + (1-k)/n\}$ 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 $\{Z \le \alpha\}$ 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 $\{Z \le \alpha$, but for now we simply treat the sets as defined. Each set $B(n,k)$ is an intersection of two events, namely the event $\{\omega : X(\omega) \le k/n\}$ and the event $\{\omega : Y(\omega) \le \alpha + (1-k)/n\}$; 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 $\{Z \le \alpha\}$, which is the straight line of slope -1 below the staircase.

d) Show that

$$\{\omega : X(\omega) + Y(\omega) \le \alpha\} = \bigcap_{n\ge1} D(n). \tag{A.4}$$

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

Solution: The region $\{\omega : X(\omega) + Y(\omega) \le \alpha\}$ 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 $n \ge 1$ and is thus contained in $\bigcap_{n\ge1} D(n)$. 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.4). Since $D(n)$ is an event, the countable intersection is also an event, so $\{\omega : X(\omega) + Y(\omega) \le \alpha\}$ 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 1.14: **a)** Let $X_1, X_2, \dots, X_n$ be rv's with expected values $\overline{X}_1, \dots, \overline{X}_n$. Show that $\mathsf{E}[X_1 + \cdots + X_n] = \overline{X}_1 + \cdots + \overline{X}_n$. You may assume that the rv's have a joint density function, but do not assume that the rv's are independent.

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

$$\begin{aligned}
\mathsf{E}[X_1 + \cdots + X_n] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1 + \cdots + x_n) \mathsf{f}_{X_1 \cdots X_n}(x_1, \dots, x_n)\, dx_1 \cdots dx_n \\
&= \sum_{j=1}^{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_j\, \mathsf{f}_{X_1 \cdots X_n}(x_1, \dots, x_n)\, dx_1 \cdots dx_n \\
&= \sum_{j=1}^{n} \int_{-\infty}^{\infty} x_j \mathsf{f}_{X_j}(x_j)\, dx_j \;=\; \sum_{j=1}^{n} \mathsf{E}[X_j].
\end{aligned}$$

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

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

Solution: From the independence, $\mathsf{f}_{X_1 \cdots X_n}(x_1, \dots, x_n) = \prod_{j=1}^{n} \mathsf{f}_{X_j}(x_j)$. Thus

$$\begin{aligned}
\mathsf{E}[X_1 X_2 \cdots X_n] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{n} x_j \prod_{j=1}^{n} \mathsf{f}_{X_j}(x_j)\, dx_1 \cdots dx_n \\
&= \prod_{j=1}^{n} \int_{-\infty}^{\infty} x_j\, \mathsf{f}_{X_j}(x_j)\, dx_j \;=\; \prod_{j=1}^{n} \mathsf{E}[X_j].
\end{aligned}$$

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

Solution: Since (a) shows that $\mathsf{E}\left[\sum_j X_j\right] = \sum_j \overline{X}_j$, we have

$$\begin{aligned}
\mathsf{VAR}\left[\sum_{j=1}^{n} X_j\right] &= \mathsf{E}\left[\left(\sum_{j=1}^{n} X_j - \sum_{j=1}^{n} \overline{X}_j\right)^2\right] \\
&= \mathsf{E}\left[\sum_{j=1}^{n} \sum_{i=1}^{n} (X_j - \overline{X}_j)(X_i - \overline{X}_i)\right] \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} \mathsf{E}\left[(X_j - \overline{X}_j)(X_i - \overline{X}_i)\right], \qquad \text{(A.5)}
\end{aligned}$$

where we have again used (a). Now from (b) (which used the independence of the X_j), $\mathsf{E}\left[(X_j - \overline{X}_j)(X_i - \overline{X}_i)\right] = 0$ for $i \neq j$. Thus(A.5) simplifies to

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

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

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

b) Show that N_1 is a rv (*i.e.*, that N_1 is not defective).

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

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

Solution: Since N_1 is a nonnegative rv,

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

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

Solution: For any given n_1 and n_2, $2 \le n_1 \le n_2$, we start by finding $\Pr\{N_1 = n_1, N_2 > n_2\}$, which we show can be expressed as

$$\{N_1 = n_1, N_2 > n_2\} \;=\; \{X_{n_1}{=}\max(X_1, \dots, X_{n_2})\} \bigcap \{X_1{=}\max(X_1, \dots, X_{n_1-1})\}. \quad \text{(A.6)}$$

The first of these events, $\{X_{n_1} = \max(X_1, \dots, X_{n_2})\}$, means that X_{n_1} must be a record-to-date and also must be the last record to date up to and including n_2. The second event, $\{X_1{=}\max(X_1, \dots, X_{n_1-1})\}$ ensures that X_{n_1} is the first record-to-date. The first event has probability $1/n_2$. The second event has to do only with the ordering of the first $n_1{-}1$ terms, and thus is independent of the first event and has probability $1/n_1$. Thus,

$$\Pr\{\{N_1 = n_1, N_2 > n_2\}\} \;=\; \frac{1}{n_2(n_1{-}1)}.$$

The marginal probability $\Pr\{N_2 > n_2\}$ can now be found by summing over n_1 and including the event $\{N_1 > n_2, N_2 > n_2\}$, which has probability $1/n_2$.

$$\Pr\{N_2 > n_2\} = \frac{1}{n_2} + \sum_{n_1=2}^{n_2} \frac{1}{n_2(n_1 - 1)}$$

This approaches 0 with increasing n_2, so N_2 is a rv. More precisely, we see that for large n_2, $\Pr\{N_2 > n_2\} \approx (\ln n_2)/n_2$, so that this approaches 0 somewhat more slowly than $\Pr\{N_1 > n_2\}$, which is not surprising. These results do not depend on the CDF of X beyond the assumption that X is continuous since they are purely ordering results.

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
\mathsf{E}[\text{Number of pairs}] &= \sum_{n=2}^{\infty} \frac{1}{n(n+1)} = \sum_{n=2}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) \\
&= \sum_{n=2}^{\infty} \frac{1}{n} - \sum_{n=3}^{\infty} \frac{1}{n} = \frac{1}{2},
\end{aligned}$$

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

The intuition here is that records-to-date tend to become more rare with increasing n (X_n is a record with probability $1/n$). As we have seen, the expected number of records from 2 to m is on the order of $\ln m$, which grows very slowly with m. The probability of an adjacent pair of records, as we have seen, decreases as $1/n(n+1)$ with n, which means that if one does not occur for small n, it will probably not occur at all. It can be seen from this that the time until the first pair of records is a defective random variable.

Exercise 1.19: **a)** Assume that X is a nonnegative discrete rv taking on values $a_1, a_2, \dots$, and let $Y = h(X)$ for some nonnegative function h. Let $b_i = h(a_i)$, $i \geq 1$ be the i^{th} value taken on by Y. Show that $\mathsf{E}[Y] = \sum_i b_i \mathsf{p}_Y(b_i) = \sum_i h(a_i)\mathsf{p}_X(a_i)$. Find an example where $\mathsf{E}[X]$ exists but $\mathsf{E}[Y] = \infty$.

Solution: If we make the added assumption that $b_i \neq b_j$ for all $i \neq j$, then Y has the sample value b_i if and only if X has the sample value a_i; thus $\mathsf{p}_Y(b_i) = \mathsf{p}_X(a_i)$ for each i. It then follows that $\sum_i b_i \mathsf{p}_Y(b_i) = \sum_i h(a_i)\mathsf{p}_X(a_i)$. This must be $\mathsf{E}[Y]$ (which might be finite or infinite). The idea is the same without the assumption that $b_i \neq b_j$ for $i \neq j$, but now the more complicated notation $\Pr\{b\} = \sum_{i:h(a_i)=b} \Pr\{a_i\}$ must be used for each sample value b of Y.

A simple example where $\mathsf{E}[X]$ is finite and $\mathsf{E}[Y] = \infty$ is to choose $a_1, a_2, \dots$, to be 1, 2, $\dots$ and choose $\mathsf{p}_X(i) = 2^{-i}$. Then $\mathsf{E}[X] = 2$. Choosing $h(i) = 2^i$, we have $b_i = 2^i$ and $\mathsf{E}[Y] = \sum_i 2^i \cdot 2^{-i} = \infty$. Without the assumption that $b_i \neq b_j$, the set of sample points of Y is the set of distinct values of b_i.

b) Let X be a nonnegative continuous rv with density $\mathsf{f}_X(x)$ and let $h(x)$ be differentiable, nonnegative, and nondecreasing in x. Let $A(\delta) = \sum_{n\geq 1} h(n\delta)[\mathsf{F}(n\delta) - \mathsf{F}(n\delta - \delta)]$, *i.e.*, $A(\delta)$ is a δth order approximation to the Stieltjes integral $\int h(x)d\mathsf{F}(x)$. Show that if $A(1) < \infty$, then $A(2^{-k}) \leq A(2^{-(k-1)}) < \infty$ for $k \geq 1$. Show from this that $\int h(x)\, d\mathsf{F}(x)$ converges to a finite value. Note: this is a very special case, but it can be extended to many cases of interest. It seems better to consider these convergence questions as required rather than consider them in general.

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

$$
\begin{aligned}
A(2\delta) &= \sum_{n\geq 1} h(2n\delta)\Big[\mathsf{F}(2n\delta) - \mathsf{F}(2n\delta - 2\delta)\Big] \\
&= \sum_{n\geq 1} h(2n\delta)\Big[\mathsf{F}(2n\delta) - \mathsf{F}(2n\delta - \delta)\Big] + \sum_{n\geq 1} h(2n\delta)\Big[\mathsf{F}(2n\delta - \delta) - \mathsf{F}(2n\delta - 2\delta)\Big] \\
&\geq \sum_{n\geq 1} h(2n\delta)\Big[\mathsf{F}(2n\delta) - \mathsf{F}(2n\delta - \delta)\Big] + \sum_{n\geq 1} h(2n\delta - \delta)\Big[\mathsf{F}(2n\delta - \delta) - \mathsf{F}(2n\delta - 2\delta)\Big] \\
&= \sum_{m\geq 1} h(m\delta)\Big[\mathsf{F}(m\delta) - \mathsf{F}(m\delta - \delta)\Big] \;=\; A(\delta),
\end{aligned}
$$

where to get the final line, we substituted m for $2n$ in the first term of the preceding line and m for $2n-1$ in the second term, resulting in $A(\delta)$. Thus $A(2^{-k})$ is nonnegative and nonincreasing in k. Since $A(1)$is finite y definition, $\lim_{k\to\infty} A(2^{-k})$ has a limit, which must be the value of the Stieltjes integral.

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

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

$$\mathsf{F}_Y(y) = 1 - \frac{2}{(y+1)(y+2)} \qquad \text{for integer } \; y \geq 0.$$

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

Solution: Combining (1.31) with the hint, we have

$$\begin{aligned}\mathsf{E}[Y] &= \sum_{y\geq 0} F_Y^{\mathsf{c}}(y) = \sum_{y\geq 0} \frac{2}{y+1} - \sum_{y\geq 0} \frac{2}{y+2} \\ &= \sum_{y\geq 0} \frac{2}{y+1} - \sum_{y\geq 1} \frac{2}{y+1} = 2,\end{aligned}$$

where the second sum in the second line eliminates all but the first term of the first sum.

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

Solution: For $y = 0$, $\mathsf{p}_Y(0) = \mathsf{F}_Y(0) = 0$. For integer $y \geq 1$, $\mathsf{p}_Y(y) = \mathsf{F}_Y(y) - \mathsf{F}_Y(y-1)$. Thus for $y \geq 1$,

$$\mathsf{p}_Y(y) = \frac{2}{y(y+1)} - \frac{2}{(y+1)(y+2)} = \frac{4}{y(y+1)(y+2)}.$$

Finding $\mathsf{E}[Y]$ from the PMF, we have

$$\begin{aligned}\mathsf{E}[Y] &= \sum_{y=1}^{\infty} y\,\mathsf{p}_Y(y) = \sum_{y=1}^{\infty} \frac{4}{(y+1)(y+2)} \\ &= \sum_{y=1}^{\infty} \frac{4}{y+1} - \sum_{y=2}^{\infty} \frac{4}{y+1} = 2.\end{aligned}$$

c) Let X be another positive, integer-valued rv. Assume its conditional PMF is given by

$$\mathsf{p}_{X|Y}(x|y) = \frac{1}{y} \qquad \text{for } 1 \leq x \leq y.$$

Find $\mathsf{E}[X \mid Y = y]$ and use it to show that $\mathsf{E}[X] = 3/2$. Explore finding $\mathsf{p}_X(x)$ until you are convinced that using the conditional expectation to calculate $\mathsf{E}[X]$ is considerably easier than using $\mathsf{p}_X(x)$.

Solution: Conditioned on $Y = y$, X is uniform over $\{1, 2, \ldots, y\}$ and thus has the conditional mean $(y+1)/2$. If you are unfamiliar with this fact, think of adding $\{1+2+\cdots+y\}$ to $\{y+(y-1)+\cdots+1\}$, getting $y(y+1)$. Thus $\{1+2+\cdots+y\} = y(y+1)/2$. It follows that

$$\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X|Y]] = \mathsf{E}\left[\frac{Y+1}{2}\right] = 3/2.$$

Calculating this expectation in the conventional way would require first calculating $\mathsf{p}_X(x)$ and then calculating the expectation. Calculating $p_X(x)$,

$$\mathsf{p}_X(x) = \sum_{y=x}^{\infty} \mathsf{p}_Y(y)\mathsf{p}_{X|Y}(x|y) = \sum_{y=x}^{\infty} \frac{4}{y(y+1)(y+2)}\,\frac{1}{y}.$$

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

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

$$\mathsf{p}_{Z|Y}(z|y) = \frac{1}{y^2} \qquad \text{for } 1 \le z \le y^2.$$

Find $\mathsf{E}[Z \mid Y = y]$ for each integer $y \ge 1$ and find $\mathsf{E}[Z]$.

Solution: As in (c), $\mathsf{E}[Z|Y] = \frac{Y^2+1}{2}$. Since $\mathsf{p}_Y(y)$ approaches 0 as y^{-3}, we see that $\mathsf{E}\left[Y^2\right]$ is infinite and thus $\mathsf{E}[Z] = \infty$.

Exercise 1.21: **a)** Show that, for uncorrelated rv's, the expected value of the product is equal to the product of the expected values (by definition, X and Y are uncorrelated if $\mathsf{E}\left[(X - \overline{X})(Y - \overline{Y})\right] = 0$).

Solution: This results from a straightforward computation.

$$\begin{aligned} 0 &= \mathsf{E}\left[(X - \overline{X})(Y - \overline{Y})\right] = \mathsf{E}[XY] - \mathsf{E}\left[X\overline{Y}\right] - \mathsf{E}\left[\overline{X}Y\right] + \mathsf{E}\left[\overline{X}\,\overline{Y}\right] \\ &= \mathsf{E}[XY] - \overline{X}\,\overline{Y}. \end{aligned}$$

b) Show that if X and Y are uncorrelated, then the variance of $X + Y$ is equal to the variance of X plus the variance of Y.

Solution: This is also straightforward, but a bit more tedious.

$$\begin{aligned} \mathsf{VAR}[X+Y] &= \mathsf{E}\left[(X+Y)^2\right] - (\mathsf{E}[X+Y])^2 \\ &= \mathsf{E}\left[X^2\right] + 2\mathsf{E}[XY] + \mathsf{E}\left[Y^2\right] - \overline{X}^2 - 2\overline{XY} - \overline{Y}^2 \\ &= \mathsf{VAR}[X] + \mathsf{VAR}[Y] + 2\mathsf{E}[XY] - 2\overline{X}\,\overline{Y} \;=\; \mathsf{VAR}[X] + \mathsf{VAR}[Y], \end{aligned}$$

where we used (a) in the last step.

c) Show that if $X_1, \dots, X_n$ are uncorrelated, then the variance of the sum is equal to the sum of the variances.

Solution: If $X_1, \dots, X_n$ are uncorrelated, then $X_1 + \dots, X_{n-1}$ is uncorrelated from X_n. Thus, from (b),

$$\mathsf{VAR}[X_1 + \cdots + X_n] = \mathsf{VAR}[X_1 + \cdots + X_{n-1}] + \mathsf{VAR}[X_n].$$

Using this equation for $n = 2$ as the basis for induction, we see that $\mathsf{VAR}[X_1 + \cdots + X_n] = \sum_{i=1}^n \mathsf{VAR}[X_i]$.

d) Show that independent rv's are uncorrelated.

Solution: If X and Y are independent, then $\mathsf{E}\left[(X - \overline{X})(Y - \overline{Y})\right] = 0$, so X and Y are uncorrelated

e) Let X, Y be identically distributed ternary valued random variables with the PMF $\mathsf{p}_X(-1) = \mathsf{p}_X(1) = 1/4$; $\mathsf{p}_X(0) = 1/2$. Find a simple joint probability assignment such that X and Y are uncorrelated but dependent.

Solution: This becomes easier to understand if we express X as $|X| \cdot X_s$ where $|X|$ is the magnitude of X and X_s is ± 1 with equal probability. From the given PMF, we see that $|X|$ and X_s are independent. Similarly, let $Y = |Y| \cdot Y_s$. For reasons soon to be apparent, we construct a joint PMF by taking $|Y| = |X|$ and by taking, X_s, Y_s, and $|X|$ to be statistically independent.

Since X and Y are zero mean, they are uncorrelated if $\mathsf{E}[XY] = 0$. For the given joint PMF, we have

$$\mathsf{E}[XY] \;=\; \mathsf{E}[|X| \cdot X_s \cdot |Y| \cdot Y_s] \;=\; \mathsf{E}\left[|X|^2\right] \mathsf{E}[X_s Y_s] = 0,$$

where we have used the fact that X_s and Y_s are zero mean and independent. Thus X and Y are uncorrelated. On the other hand, X and Y are certainly not independent since $|X|$ and $|Y|$ are dependent and in fact identical. This should not be surprising since the correlation of two rv's is a single number, whereas many numbers are needed to specify the joint distribution. This example was constructed through the realization that the type of symmetry here between X and Y is sufficient to guarantee uncorrelatedness, but not enough to cause independence.

f) You have seen that the moment generating function of a sum of independent rv's is equal to the product of the individual moment generating functions. Give an example where this is false if the variables are uncorrelated but dependent.

Solution: We know (although we have not proven) that the MGF of a rv specifies the distribution, and thus the MGF of a dependent sum cannot be the same as the MGF of an independent sum with the same marginals. Thus any will work, but the one in (e) is quite simple.

$$\begin{aligned}
\mathsf{g}_{X+Y}(r) &= \frac{1}{8}e^{-2r} + \frac{3}{4} + \frac{1}{8}e^{2r} \\
\mathsf{g}_X(r)\mathsf{g}_Y(r) &= \left(\frac{1}{4}e^{-r} + \frac{1}{2} + \frac{1}{4}e^{r}\right)^2 = \frac{1}{16}e^{-2r} + \frac{1}{4}e^{-r} + \frac{3}{8} + \frac{1}{4}e^{r} + \frac{1}{16}e^{2r}.
\end{aligned}$$

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

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

$$\begin{aligned}
\mathsf{p}_Z(n) &= \sum_{k=0}^{n} \mathsf{p}_X(k)\mathsf{p}_Y(n-k) \;=\; \sum_{k=0}^{n} \frac{\lambda^k e^{-\lambda}}{k!} \times \frac{\mu^{n-k}e^{-\mu}}{(n-k)!} \\
&= e^{-(\lambda+\mu)} \sum_{k=0}^{n} \frac{\lambda^k \mu^{n-k}}{k!(n-k)!}.
\end{aligned}$$

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