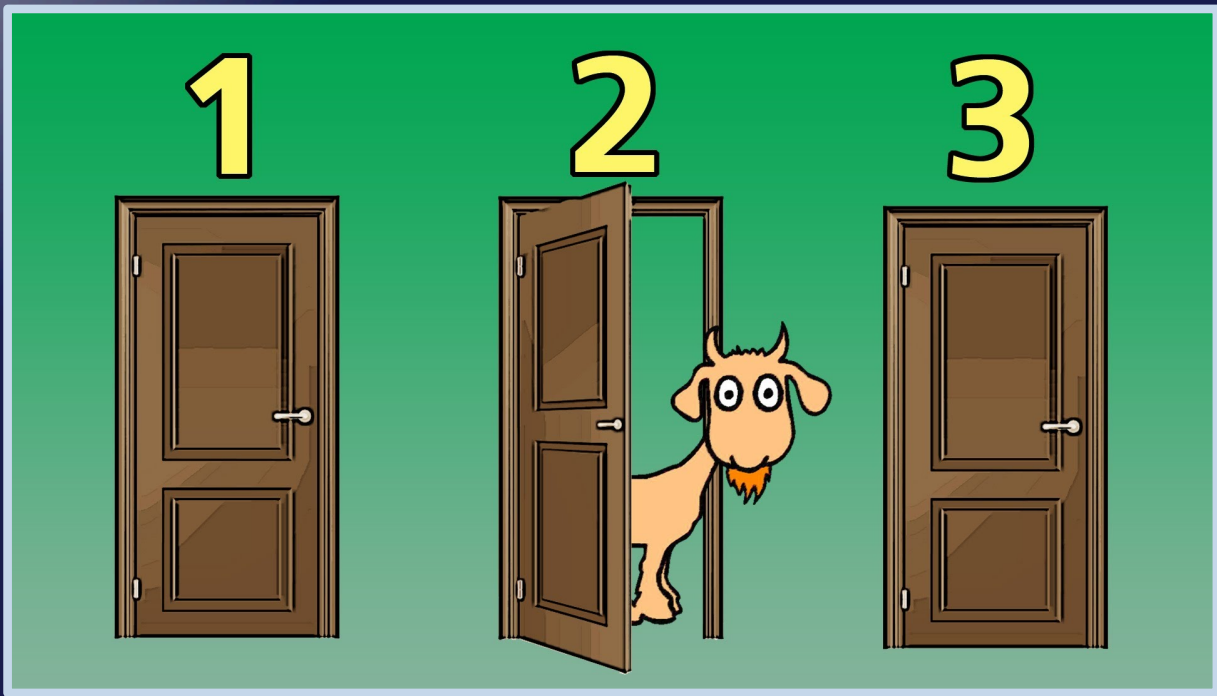


DOUGLAS RUBIN PhD



# A Complete Solutions Guide to Pishro-Nik's:

Introduction to Probability,  
Statistics and Random Processes



# Contents

<b>1</b>	<b>Basic Concepts</b>	<b>5</b>
<b>2</b>	<b>Combinatorics: Counting Methods</b>	<b>27</b>
<b>3</b>	<b>Discrete Random Variables</b>	<b>35</b>
<b>4</b>	<b>Continuous and Mixed Random Variables</b>	<b>53</b>
<b>5</b>	<b>Joint Distributions</b>	<b>71</b>
<b>6</b>	<b>Methods for More Than Two Random Variables</b>	<b>105</b>
<b>7</b>	<b>Limit Theorems and Convergence of Random Variables</b>	<b>123</b>
<b>8</b>	<b>Statistical Inference I: Classical Methods</b>	<b>135</b>
<b>9</b>	<b>Statistical Inference II: Bayesian Inference</b>	<b>153</b>
<b>10</b>	<b>Introduction to Simulation Using Python</b>	<b>169</b>
<b>11</b>	<b>Recursive Methods</b>	<b>195</b>



## Chapter 1

# Basic Concepts

**Problem 1.**

(a)

$$A \cup B = \{1, 2, 3\} \cup \{2, 3, 4, 5, 6, 7\} = \{1, 2, 3, 4, 5, 6, 7\}$$

(b)

$$\begin{aligned} (A \cup C) - B &= \{1, 2, 3\} \cup \{7, 8, 9, 10\} - \{2, 3, 4, 5, 6, 7\} \\ &= \{1, 2, 3, 7, 8, 9, 10\} - \{2, 3, 4, 5, 6, 7\} \\ &= \{1, 8, 9, 10\} \end{aligned}$$

(c)

$$\begin{aligned} \bar{A} \cup (B - C) &= (\{1, 2, \dots, 10\} - A) \cup (\{2, 3, 4, 5, 6, 7\} - \{7, 8, 9, 10\}) \\ &= \{4, 5, 6, 7, 8, 9, 10\} \cup \{2, 3, 4, 5, 6\} \\ &= \{2, 3, \dots, 10\} \end{aligned}$$

(d) No,  $A$ ,  $B$  and  $C$  do not partition  $S$  since  $2, 3 \in A$  as well as in  $B$ . 7 is also in both  $B$  and  $C$ .**Problem 2.**

(a)

$$[6, 8] \cup [2, 7] = [2, 8]$$

(b)

$$[6, 8] \cap [2, 7] = [6, 7]$$

(c)

$$[0, 1]^c = (-\infty, 0) \cup (1, \infty)$$

(d)

$$[6, 8] - (2, 7) = [7, 8]$$

**Problem 3.**

(a)

$$\begin{aligned} (A \cup B) - (A \cap B) &= \\ &= (A \cup B) \cap (A \cap B)^c \\ &= (A \cup B) \cap (A^c \cup B^c), \end{aligned}$$

where I have used De Morgan.

(b)

$$B - C = B \cap C^c$$

(c)

$$(A \cap C) \cup (A \cap B)$$

(d)

$$(C - A - B) \cup ((A \cap B) - C)$$

**Problem 4.**

(a)

$$A = \{(H, H), (H, T)\}$$

(b)

$$B = \{(H, T), (T, H), (T, T)\}$$

(c)

$$C = \{(H, T), (T, H)\}$$

**Problem 5.**

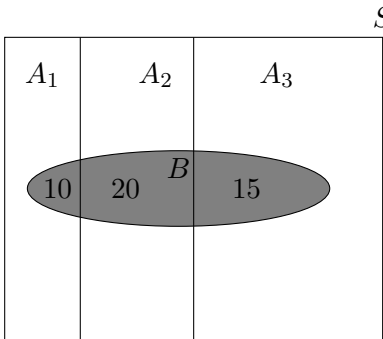
(a)  $|A_2|$  is half of the numbers from 1 to 100, so  $|A_2| = 50$ . To solve for  $|A_3|$  note that there are 2 numbers between each pair of elements in  $A_3$  where  $A_3$  is assumed to be pre-sorted (e.g., 4, 5 are between 3 and 6). There are also  $|A_3| - 1$  of these pairs, and thus  $|A_3| + 2(|A_3| - 1) + 3 = 100$ , where I have added 3 to account for the numbers at the beginning and end of the sequence which are not divisible by 3 (1, 2 and 100). Thus, I find that  $|A_3| = 33$ .  $|A_4|$  is exactly half of  $|A_2|$ , and thus  $|A_4| = 25$ . Finally, to solve for  $|A_5|$  we may use the same method we used to solve for  $|A_3|$ :  $|A_5| + 4(|A_5| - 1) + 4 = 100$ , from which we find that  $|A_5| = 20$ .

(b) By inclusion-exclusion:

$$|A_2 \cup A_3 \cup A_5| = |A_2| + |A_3| + |A_5| - |A_2 \cap A_3| - |A_2 \cap A_5| - |A_3 \cap A_5| + |A_2 \cap A_3 \cap A_5|.$$

Note that  $|A_2 \cap A_3| = |A_6| = 16$ ,  $|A_2 \cap A_5| = |A_{10}| = 10$ ,  $|A_3 \cap A_5| = |A_{15}| = 6$ , where  $|A_{10}|$  and  $|A_{15}|$  were found by counting (since there are very few elements in these sets), and  $|A_6|$  was found by the same method I used to compute  $|A_3|$ . Lastly, the intersection of all 3 sets is given by the set of multiples of 30, so that  $|A_2 \cap A_3 \cap A_5| = |\{30, 60, 90\}| = 3$ . Therefore:  $|A_2 \cup A_3 \cup A_5| = 50 + 33 + 20 - 16 - 10 - 6 + 3 = 74$ .

**Problem 6.** From the following figure, it is clear that  $|B| = 10 + 20 + 15 = 45$ .

**Problem 7.**

(a)  $A$  is a subset of a countable set,  $\mathbb{N}$ , and is thus countable.

(b) As shown in the book, if we can write any set,  $S$  in the form:

$$S = \bigcup_{i \in \mathcal{B}} \bigcup_{j \in \mathcal{C}} \{q_{ij}\}, \quad (1.1)$$

where  $\mathcal{B}$  and  $\mathcal{C}$  are countable sets, then  $S$  is a countable set. It is easy to see that we may re-write  $B$  as:

$$B = \bigcup_{i \in \mathbb{Q}} \bigcup_{j \in \mathbb{Q}} \{a_i + b_j \sqrt{2}\}, \quad (1.2)$$

where  $q_{ij} \equiv a_i + b_j \sqrt{2}$ , and thus  $B$  is countable.

- (c)  $C$  is uncountable. One way to prove this is to note that for all  $x \in [0, 1]$ ,  $(x, 0) \in C$ , so that  $C \supset [0, 1]$ , i.e,  $C$  is a superset of an uncountable set and is thus uncountable.

**Problem 8.** I first prove that  $A_n \subset A_{n+1}$  (a proper subset) for all  $n = 1, 2, \dots$ . To do this, it suffices to prove that  $(n-1)/n < n/(n+1)$ , which I do with proof by contradiction. By assuming  $(n-1)/n \geq n/(n+1)$ , after a little algebra, one concludes that  $-1 \geq 0$ , which is clearly a contradiction and therefore  $(n-1)/n < n/(n+1)$ . Thus the union of all the  $A_n$ s is given by the largest set in the sequence, which is  $A_\infty (= \lim_{n \rightarrow \infty} [0, \frac{n-1}{n}])$ . After applying L'hopitals rule, one can show that  $A_\infty = [0, 1)$ , and thus:

$$A = \bigcup_{n=1}^{\infty} A_n = [0, 1).$$

**Problem 9.** As with the previous problem, one may show that  $A_{n+1} \subset A_n$  for all  $n = 1, 2, \dots$  by proving that  $1/(n+1) < 1/n$ . This is somewhat obvious, but if you really want to be formal, you can prove it with a proof by contradiction. Therefore, the intersection of all the  $A_n$ s is given by the smallest set,  $A_\infty$ , which is  $\lim_{n \rightarrow \infty} [0, \frac{1}{n}] = [0, 0) = \{0\}$ , and thus:

$$A = \bigcap_{n=1}^{\infty} A_n = \{0\}.$$

**Problem 10.**

- (a) To motivate the bijection (the one-to-one mapping between  $2^{\mathbb{N}}$  and  $C$ ) we are about to construct, note that for every set in  $2^{\mathbb{N}}$ , a natural number  $n$  will either appear once, or not at all. Therefore, it is convenient to indicate its presence in the set with a 1 and its absence with a 0. For example  $\{1, 3, 6\}$  will get mapped to the sequence 101001000... (this is implicitly assuming that we have pre-ordered the elements in the particular set from  $2^{\mathbb{N}}$ ). In general, the bijective mapping we use  $f: 2^{\mathbb{N}} \rightarrow C$ , is given by:

$$f(x) = \mathbb{1}(1 \in x) \mathbb{1}(2 \in x) \dots,$$

where  $\mathbb{1}$  is the so-called indicator function which is 1 if its argument evaluates to true and 0 otherwise. To prove that this mapping is bijective, we must prove it is both injective and surjective.

To prove it is injective, I use a proof by contradiction. Assume it is not injective. Under this assumption there exists  $x, x' \in 2^{\mathbb{N}}$ , where  $x \neq x'$  such that  $f(x) = f(x')$ .  $x$  and  $x'$  can either have the same cardinality, or they can be different. Without loss of generality, if they are different, let us call  $x$  the one with the larger cardinality. Since  $x \neq x'$  there exists at least 1 natural number  $n$  in  $x$  which is not in  $x'$ . Therefore in the sequences  $f(x)$  and  $f(x')$ , there is at least one value in the sequences which does not match up, namely the value at position  $n$ , and therefore  $f(x) \neq f(x')$ , which violates our original assumption.

The proof of surjectivity is also straightforward.



- (b) Any number in  $x \in [0, 1)$  always has a unique binary expansion given by  $x = b_1/2 + b_2/2^2 + \dots$ , and therefore we can construct a bijective mapping between  $x \in [0, 1)$  and  $C$  by computing  $b_1/2 + b_2/2^2 + \dots$ , and then by dropping the 0. at the beginning of the sequence. Since there is a bijection between  $2^{\mathbb{N}}$  and  $C$  and a bijection between  $C$  and  $[0, 1)$  (and given the fact that the composition of 2 bijections is a bijection) there is thus a bijection between  $2^{\mathbb{N}}$  and  $[0, 1)$ . Assuming (correctly so) that the interval  $[0, 1)$  is uncountable, then so too is  $2^{\mathbb{N}}$ .

**Problem 11.** As shown in the previous problem, there is a bijection between  $[0, 1)$  and  $C$ . Therefore, if  $C$  is uncountable, then so too is  $[0, 1)$ . We can use what is known as Cantor's diagonal argument to prove that  $C$  is uncountable.

Let us try to search for a bijective mapping between  $C$  and  $\mathbb{N}$ . Suppose, for example, that the first few mappings are given by:

1	→	<b>0000000</b> ...
2	→	<b>1111111</b> ...
3	→	0 <b>101010</b> ...
4	→	10 <b>10101</b> ...
5	→	1101 <b>1011</b> ...
6	→	00110 <b>11</b> ...
7	→	10001 <b>00</b> ...
		⋮

Let us now construct a new sequence,  $s \in C$  by enumerating the complement of the elements along the diagonal of the mapping (which I have highlighted in boldface above),  $s = 1011101\dots$ . By construction,  $s$  differs from every proposed mapping since the  $n^{th}$  digit in  $s$  is different than the  $n^{th}$  digits in all of the mappings. Thus, no natural number gets mapped to  $s$ , and hence the proposed mapping is not surjective. The mappings I chose for illustration in this example for  $1, \dots, 7$  were arbitrary, and this argument applies to any potential mapping. Therefore, there is no bijective mapping between  $\mathbb{N}$  and  $C$ , and hence no bijection between  $[0, 1)$  and  $\mathbb{N}$ . Thus, the interval  $[0, 1)$  is uncountable.

**Problem 12.**

- (a) The domain is  $\{H, T\}^3$  and the codomain is  $\mathbb{N} \cup \{0\}$
- (b)  $range(f) = \{0, 1, 2, 3\}$
- (c)  $x$  can be all triplets that contain exactly 2 heads:  $(H, H, T)$ ,  $(H, T, H)$  or  $(T, H, H)$ .

**Problem 13.**

- (a) The universal set is partitioned by the events  $a$ ,  $b$ ,  $d$ , and thus  $P(b) = 1 - P(a) - P(d) = 1 - 0.5 - 0.25 = 0.25$ .
- (b) Since the events  $b$  and  $d$  are disjoint, by the 3rd axiom of probability,  $P(b \cup d) = P(b) + P(d) = 0.25 + 0.25 = 0.5$ .

**Problem 14.**

- (a) By inclusion-exclusion:  $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.7 - 0.9 = 0.2$ .

(b)

$$\begin{aligned}
P(A^c \cap B) &= P(B - A) \\
&= P(B) - P(A \cap B) \\
&= 0.7 - 0.2 \\
&= 0.5
\end{aligned}$$

(c)

$$\begin{aligned}
P(A - B) &= P(A) - P(A \cap B) \\
&= 0.4 - 0.2 \\
&= 0.2
\end{aligned}$$

(d) By drawing the Venn diagram, one can see that

$$\begin{aligned}
P(A^c - B) &= P(S) - P(A \cup B) \\
&= 1 - 0.9 \\
&= 0.1,
\end{aligned}$$

where  $S$  is the universal set.

(e) By drawing the Venn diagram, one can see that

$$\begin{aligned}
P(A^c \cup B) &= P(S) - P(A - B) \\
&= 1 - 0.2 \\
&= 0.8.
\end{aligned}$$

(f)

$$\begin{aligned}
P(A \cap (B \cup A^c)) &= P((A \cap B) \cup (A \cap A^c)) \\
&= P((A \cap B) \cup \emptyset) \\
&= P(A \cap B) \\
&= 0.2.
\end{aligned}$$

**Problem 15.**

- (a) The second roll is independent of the first, so we only need to consider the second roll, in which case  $P(X_2 = 4) = 1/6$  since this is a finite sample space with equal probabilities for all outcomes.
- (b) The sample space is  $\{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$ , which has a cardinality of 36, and the possible outcomes corresponding to the event that  $X_1 + X_2 = 7$  are given by the set  $\{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}$ , which has a cardinality of 6, and therefore  $P(X_1 + X_2 = 7) = 6/36 = 1/6$ .
- (c) Listing out the tuples that satisfy the second condition in a matrix-like representation, we have:

$$\begin{array}{ccc}
(1, 4) & (1, 5) & (1, 6) \\
(2, 4) & (2, 5) & (2, 6) \\
& \vdots & \\
(6, 4) & (6, 5) & (6, 6)
\end{array},$$

of which there are  $3 \times 6$  elements. However, the first condition does not allow the elements  $(2, 4)$ ,  $(2, 5)$ ,  $(2, 6)$ , and thus the total size of the event space is  $3 \times 6 - 3 = 15$ . Thus  $P(X_1 \neq 2 \cap X_2 \geq 4) = 15/36 = 5/12$ .

**Problem 16.**

- (a) The formula for a geometric series will be useful here:  $\sum_{k=0}^{\infty} cr^k = a/(1-r)$  for  $|r| < 1$ . To solve for  $c$ , we can use the normalization constraint:

$$\begin{aligned}
1 &= \sum_{k=1}^{\infty} P(k) \\
&= -c + \sum_{k=0}^{\infty} c \left(\frac{1}{3}\right)^k \\
&= -c + \frac{c}{1 - 1/3},
\end{aligned}$$

and therefore  $c = 2$ .

(b)

$$\begin{aligned}
P(\{2, 4, 6\}) &= P(\{2\} \cup \{4\} \cup \{6\}) \\
&= P(2) + P(4) + P(6) \\
&= 2 \left[ \frac{1}{3^2} + \frac{1}{3^4} + \frac{1}{3^6} \right] \\
&= \frac{182}{729}
\end{aligned}$$

(c)

$$\begin{aligned}
P(\{3, 4, 5, \dots\}) &= 2 \sum_{k=3}^{\infty} \left(\frac{1}{3}\right)^k \\
&= -2 \left[ 1 + \frac{1}{3} + \frac{1}{9} \right] + 2 \sum_{k=0}^{\infty} \left(\frac{1}{3}\right)^k \\
&= -2 \left(\frac{13}{9}\right) + 2 \left(\frac{3}{2}\right) \\
&= \frac{1}{9}
\end{aligned}$$

This answer may also have been computed  $1 - P(1) - P(2)$ .

**Problem 17.** Let us write down what we know in equations. Let  $a$ ,  $b$ ,  $c$ ,  $d$  represent the events that teams, A, B, C and D win the tournament respectively. Then as stated in the problem,

$P(a) = P(b)$ ,  $P(c) = 2P(d)$  and  $P(a \cup c) = 0.6$ . Since the events partition the sample space,  $P(a \cup c) = P(a) + P(c)$ . We know one more equation, which is that the probabilities must sum to one:  $P(a) + P(b) + P(c) + P(d) = 1$ . We therefore have a linear system with 4 equations and 4 unknowns, and it will thus be convenient to write this in matrix notation in order to solve for the probabilities:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} P(a) \\ P(b) \\ P(c) \\ P(d) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.6 \\ 1 \end{bmatrix}.$$

$\implies$

$$\begin{aligned} \begin{bmatrix} P(a) \\ P(b) \\ P(c) \\ P(d) \end{bmatrix} &= \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0.6 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 1 & -3 & 2 \\ 1 & 1 & -3 & 2 \\ -2 & -1 & 4 & 2 \\ -1 & -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0.6 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0.2 \\ 0.2 \\ 0.4 \\ 0.2 \end{bmatrix}. \end{aligned}$$

Notice that, as required, the probabilities sum to 1.

**Problem 18.**

(a)  $P(T \leq 1) = \frac{1}{16}$

(b)

$$\begin{aligned} P(T > 2) &= 1 - P(T \leq 2) \\ &= 1 - \frac{4}{16} \\ &= \frac{3}{4} \end{aligned}$$

(c)

$$\begin{aligned} P(1 \leq T \leq 3) &= P(T \leq 3) - P(T < 1) \\ &= \frac{9}{16} - \frac{1}{16} \\ &= \frac{1}{2} \end{aligned}$$

**Problem 19.** The solutions to the quadratic are given by the quadratic formula:

$$X = \frac{-1 \pm \sqrt{1 - 4AB}}{2A}, \tag{1.3}$$

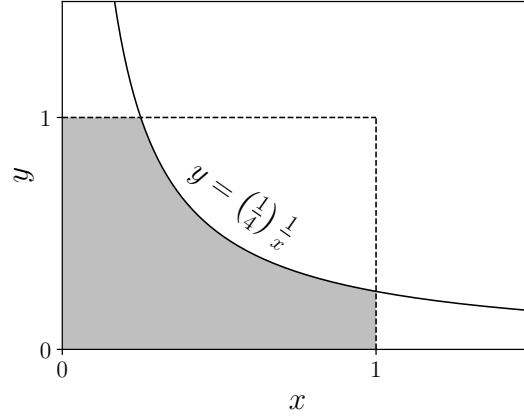


Figure 1.1: area of unit square resulting in real solutions

which has real solutions iff the condition  $1 - 4AB \geq 0$  is satisfied. We therefore seek the probability that  $P(1 - 4AB \geq 0)$  (in the unit square), which, since the point  $(A, B)$  is picked uniformly, is the fraction of area in the unit square which satisfies this constraint. Therefore points which satisfy the following inequalities contribute to this probability:

$$\frac{1}{4} \left( \frac{1}{x} \right) \geq y,$$

$$x \leq 1$$

and

$$y \leq 1,$$

where the last 2 inequalities follow since the randomly drawn points must lie within the unit square. The area in the unit square which satisfies these constraints is shown in Fig. 6.2.

It is clear from the figure that the area is given by:

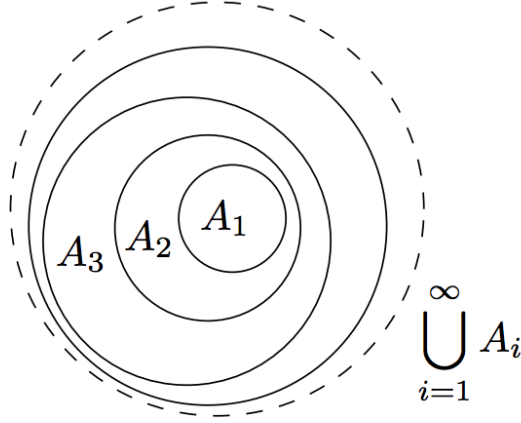
$$\begin{aligned} P(\text{real solns.}) &= \frac{1}{4} + \frac{1}{4} \int_{1/4}^1 \frac{1}{x} dx \\ &= \frac{1}{4} + \frac{1}{4} \ln 4 \\ &\approx 0.60. \end{aligned}$$

**Problem 20.**

(a) To solve this problem, note that:

$$A \equiv \bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \dots,$$

where in the figure,  $A_1$  is the innermost circle,  $(A_2 - A_1)$  is the “annulus” around  $A_1$ ,  $(A_3 - A_2)$  is the next “annulus” and so-forth. It is clear that the union of  $A_1$  and all of the annuli results in  $A$ , and that each of these regions are also disjoint. I utilize the previous equation in the desired proof:

Figure 1.2: Venn diagram of events  $A_1, A_2, \dots$ 

*Proof.*

$$\begin{aligned}
 P(A) &= P(A_1) + \sum_{i=2}^{\infty} P(A_i - A_{i-1}) \\
 &= P(A_1) + \lim_{n \rightarrow \infty} \sum_{i=2}^n P(A_i - A_{i-1}) \\
 &= P(A_1) + \lim_{n \rightarrow \infty} \sum_{i=2}^n [P(A_i) - P(A_{i-1})] \\
 &= P(A_1) + \lim_{n \rightarrow \infty} \{[P(A_2) - P(A_1)] + [P(A_3) - P(A_2)] + [P(A_4) - P(A_3)] \\
 &\quad + \dots + [P(A_n) - P(A_{n-1})]\} \\
 &= P(A_1) + \lim_{n \rightarrow \infty} [P(A_n) - P(A_1)] \\
 &= \lim_{n \rightarrow \infty} P(A_n)
 \end{aligned}$$

■

- (b) Redefining  $A$ :  $A \equiv \bigcap_{i=1}^{\infty} A_i$ , we seek to find  $P(A)$ . If  $A_1, A_2, \dots$  is a series of decreasing events, then  $A_1^c, A_2^c, \dots$  must be a series of increasing events, and we can therefore utilize the results of the part (a) on sequence of the complements (as well as De Morgan):

$$P(A^c) = P\left(\left(\bigcap_{i=1}^{\infty} A_i\right)^c\right) = P\left(\bigcup_{i=1}^{\infty} A_i^c\right) = \lim_{n \rightarrow \infty} P(A_n^c).$$

A few more steps completes the proof:

*Proof.*

$$\begin{aligned}
 P(A) &= 1 - P(A^c) \\
 &= 1 - \lim_{n \rightarrow \infty} P(A_n^c) \\
 &= \lim_{n \rightarrow \infty} [1 - P(A_n^c)] \\
 &= \lim_{n \rightarrow \infty} P(A_n)
 \end{aligned}$$

■

**Problem 21.**

- (a) Let us define new events,  $B_i$ , such that  $B_1 = A_1$ ,  $B_2 = A_2 - A_1$ ,  $B_3 = A_3 - A_2 - A_1, \dots$ . Note that the  $B_i$ s are disjoint. Also note that:

$$\begin{aligned} \bigcup_{i=1}^n B_i &= A_1 \cup (A_2 - A_1) \cup (A_3 - A_2 - A_1) \cup \dots \cup (A_n - A_{n-1} - \dots - A_1) \\ &= A_1 \cup A_2 \cup A_3 \dots \cup A_n \\ &= \bigcup_{i=1}^n A_i, \end{aligned}$$

and for the same reason  $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ . Using these facts, the proof is now straightforward:

*Proof.*

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} P(B_i) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) \end{aligned} \quad \blacksquare$$

- (b) To prove this second result I use the previous result as well as De Morgan (twice):

*Proof.*

$$\begin{aligned} P\left(\bigcap_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i^c\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i^c\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned} \quad \blacksquare$$

**Problem 22.** Let  $A_{coffee}$  be the event that a customer purchases coffee and  $A_{cake}$  be the event that a customer purchases cake. We know that  $P(A_{coffee}) = 0.7$ ,  $P(A_{cake}) = 0.4$  and  $P(A_{coffee}, A_{cake}) = 0.2$ . Thus, the conditional probability we seek is:

$$P(A_{coffee} | A_{cake}) = \frac{P(A_{coffee}, A_{cake})}{P(A_{cake})} = \frac{0.2}{0.4} = 0.5.$$

**Problem 23.**

(a)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1 + 0.1}{0.1 + 0.1 + 0.1 + 0.05} \approx 0.57$$

(b)

$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{0.1 + 0.05}{0.1 + 0.1 + 0.1 + 0.05} \approx 0.43$$

(c)

$$P(B|A \cup C) = \frac{P(B \cap (A \cup C))}{P(A \cup C)} = \frac{0.1 + 0.1 + 0.05}{0.1 + 0.2 + 0.1 + 0.1 + 0.05 + 0.15} \approx 0.36$$

(d)

$$P(B|A, C) = \frac{P(B \cap (A \cap C))}{P(A \cap C)} = \frac{0.1}{0.1 + 0.1} = 0.5$$

**Problem 24.**

(a)

$$P(2 \leq X \leq 5) = \frac{3}{10} = 0.3$$

(b)

$$P(X \leq 2|X \leq 5) = \frac{2}{5} = 0.4$$

(c)

$$P(3 \leq X \leq 8|X \geq 4) = \frac{P(3 \leq X \leq 8 \cap X \geq 4)}{P(X \geq 4)} = \frac{4}{6} = \frac{2}{3}$$

**Problem 25.** Let  $ON$  denote the event that a student lives on campus,  $OFF$  denote the event that a student lives off campus and  $A$  denote the event that a student receives an A. Given the data I compute the following probabilities:

$$P(ON) \approx \frac{200}{600} = \frac{1}{3}$$

$$P(A) \approx \frac{120}{600} = \frac{1}{5}$$

$$P(A \cap ON) = P(A) - P(A \cap OFF) \approx \frac{1}{5} - \frac{80}{600} = \frac{1}{15}$$

If the events  $ON$  and  $A$  are independent, then  $P(A \cap ON) = P(A)P(ON)$ . Looking at the probabilities above, we see that the data suggests this relationship, and thus the data suggests that getting an A and living on campus are independent.

**Problem 26.** Let  $N_1$  be the number of times out of  $n$  that a 1 is rolled,  $N_6$  be the number of times out of  $n$  that a 6 is rolled and  $X_i$  be the value of the  $i^{th}$  roll. Then:



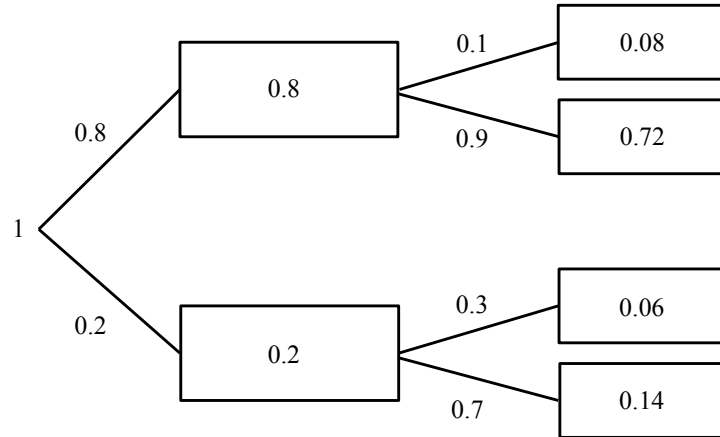


Figure 1.3: Tree diagram for problem 27.

$$\begin{aligned}
 P(N_1 \geq 1 \cap N_6 \geq 1) &= 1 - P((N_1 \geq 1 \cap N_6 \geq 1)^c) \\
 &= 1 - P(N_1 = 0 \cup N_6 = 0) \\
 &= 1 - [P(X_1 \neq 1, X_2 \neq 1, \dots, X_n \neq 1) + P(X_1 \neq 6, X_2 \neq 6, \dots, X_n \neq 6) \\
 &\quad - P((X_1 \neq 1, X_2 \neq 1, \dots, X_n \neq 1) \cap (X_1 \neq 6, X_2 \neq 6, \dots, X_n \neq 6))] \\
 &= 1 - \left[ \left( \frac{5}{6} \right)^n + \left( \frac{5}{6} \right)^n - P(X_1 \neq 1, X_1 \neq 6, X_2 \neq 1, X_2 \neq 6, \dots, X_n \neq 1, X_n \neq 6) \right] \\
 &= 1 - \left[ 2 \left( \frac{5}{6} \right)^n - P(X_1 \neq 1, X_1 \neq 6)^n \right] \\
 &= 1 - \left[ 2 \left( \frac{5}{6} \right)^n - \left( \frac{4}{6} \right)^n \right] \\
 &= 1 - \frac{2(5^n) - 4^n}{6^n}.
 \end{aligned}$$

In the second line I have used De Morgan, and I have also used the fact, several times, that the outcome of roll  $i$  is independent of the outcome of roll  $j$ . Testing for a for values of  $n$ , I find that, when  $n = 1$ , the probability is 0, which makes sense because at the very minimum we would need at least one 1 and one 6, which cannot happen if we have only rolled once. The probability then monotonically increases, which also makes sense because it becomes more and more likely that we roll at least one 1 and at least one 6 the more rolls we throw. Note that as a sanity check, one can show that  $\lim_{n \rightarrow \infty} 1 - (2(5^n) - 4^n)/(6^n)$  is 1, so that our formula for the probability is bounded between 0 and 1. Also note that this formula can also be obtained more easily with combinatorics, which will be introduced in Chapter 2.

### Problem 27.

(a) Refer to Fig. 5.4

(b)  $P(E) = P(E \cap G) + P(E \cap G^c) = 0.08 + 0.06 = 0.14$

(c)  $P(G|E^c) = \frac{P(G \cap E^c)}{P(E^c)} = \frac{0.72}{1-0.14} \approx 0.84.$

**Problem 28.** Let  $A_i$  be the event that the  $i^{th}$  ( $i = 1, 2, 3$ ) unit of the 3 picks is defective, while the other 2 are not defective. Note that  $A_1$ ,  $A_2$  and  $A_3$  are all disjoint, since it is impossible for any unit to be both defective and not defective simultaneously. The probability we seek is therefore:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &= \left(\frac{5}{100}\right) \left(\frac{95}{99}\right) \left(\frac{94}{98}\right) + \left(\frac{95}{100}\right) \left(\frac{5}{99}\right) \left(\frac{94}{98}\right) + \left(\frac{95}{100}\right) \left(\frac{94}{99}\right) \left(\frac{5}{98}\right) \\ &\approx 0.14. \end{aligned}$$

**Problem 29.** Let  $F$  be the event that the system is functional, and  $C_i$  be the event that component  $i$  is functional.

(a)  $P(F) = P(C_1, C_2, C_3) = P_1 P_2 P_3$

(b) By inclusion-exclusion:

$$\begin{aligned} P(F) &= P(C_1 \cup C_2 \cup C_3) \\ &= P_1 + P_2 + P_3 - P_1 P_2 - P_1 P_3 - P_2 P_3 + P_1 P_2 P_3 \end{aligned}$$

(c)

$$\begin{aligned} P(F) &= P((C_1, C_3) \cup (C_2, C_3)) \\ &= P(C_1, C_3) + P(C_2, C_3) - P((C_1 \cap C_3) \cap (C_2 \cap C_3)) \\ &= P_1 P_3 + P_2 P_3 - P(C_1, C_2, C_3) \\ &= P_1 P_3 + P_2 P_3 - P_1 P_2 P_3 \end{aligned}$$

(d)  $P(F) = P(C_1, C_2) + P(C_3) = P_1 P_2 + P_3 - P_1 P_2 P_3$

(e)  $P(F) = P(C_1, C_2, C_5) + P(C_3, C_4, C_5) - P(C_1, C_2, C_3, C_4, C_5) = P_1 P_2 P_5 + P_3 P_4 P_5 - P_1 P_2 P_3 P_4 P_5$

**Problem 30.**

(a) The region in the unit square corresponding to set  $A$  can be made more clear if we write the absolute value as a piecewise function:

$$|x - y| \leq \frac{1}{2} \implies \begin{cases} x - y \leq \frac{1}{2} & \text{if } x \geq y \\ y - x \leq \frac{1}{2} & \text{if } x < y \end{cases} \implies \begin{cases} y \geq -\frac{1}{2} + x & \text{if } y \leq x \\ y \leq \frac{1}{2} + x & \text{if } y > x \end{cases}.$$

This piecewise function, along with the fact that  $A$  must be bounded in the unit square leads to hashed-in region in Fig. 5.6. The region corresponding to set  $B$  is just the are in the unit square above the  $45^\circ$  line (corresponding to the gray shaded region in Fig. 5.6).

(b) Using a little geometry, I find:  $P(A) = 1 - 2 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \frac{3}{4}$  and  $P(B) = \frac{1}{2}$ .

(c) Again, using some geometry, I find:  $P(A \cap B) = \frac{1}{2} - \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \frac{3}{8}$ . Since  $P(A)P(B) = \left(\frac{3}{4}\right) \left(\frac{1}{2}\right) = \frac{3}{8}$ , the 2 events are indeed independent.

**Problem 31.**

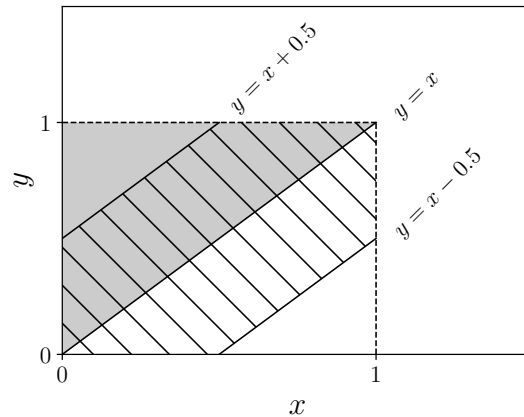


Figure 1.4: The unit square for Problem 30. The shaded region represents the set  $B$  and the hashed region represents the set  $A$

- (a) Let  $s$  be the event that the received email is spam and  $r$  be the event that the received email contains the word refinance. From the problem statement, we know that  $P(s) = 0.5$  (so that  $P(s^c) = 0.5$ ),  $P(r|s) = 0.01$  and  $P(r|s^c) = 0.00001$ . Using Bayes' rule:

$$\begin{aligned}
 P(s|r) &= \frac{P(r|s)P(s)}{P(r)} \\
 &= \frac{P(r|s)P(s)}{P(r|s)P(s) + P(r|s^c)P(s^c)} \\
 &= \frac{(0.01)(0.5)}{(0.01)(0.5) + (0.00001)(0.5)} \\
 &\approx 0.999
 \end{aligned}$$

### Problem 32.

- (a) There are 4 possible paths from  $A$  to  $B$ : 1 to 4 (path 1), 2 to 5 (path 2), 1 to 3 to 5 (path 3), 2 to 3 to 4 (path 4). Let  $\mathcal{P}_i$  be the event that path  $i$  is open. Only 1 path needs to be open for event  $A$  to occur, so the probability of  $A$  is given by the probability of  $\mathcal{P}_1$  or  $\mathcal{P}_2$  or  $\mathcal{P}_3$  or  $\mathcal{P}_4$ . We expand this probability with inclusion-exclusion, making sure to enumerate all unique pairs and all unique triplets:

$$\begin{aligned}
P(A) &= P(\mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3 \cup \mathcal{P}_4) \\
&= P(\mathcal{P}_1) + P(\mathcal{P}_2) + P(\mathcal{P}_3) + P(\mathcal{P}_4) \\
&\quad - P(\mathcal{P}_1 \cap \mathcal{P}_2) - P(\mathcal{P}_1 \cap \mathcal{P}_3) - P(\mathcal{P}_1 \cap \mathcal{P}_4) - P(\mathcal{P}_2 \cap \mathcal{P}_3) - P(\mathcal{P}_2 \cap \mathcal{P}_4) - P(\mathcal{P}_3 \cap \mathcal{P}_4) \\
&\quad + P(\mathcal{P}_1 \cap \mathcal{P}_2 \cap \mathcal{P}_3) + P(\mathcal{P}_1 \cap \mathcal{P}_2 \cap \mathcal{P}_4) + P(\mathcal{P}_1 \cap \mathcal{P}_3 \cap \mathcal{P}_4) + P(\mathcal{P}_2 \cap \mathcal{P}_3 \cap \mathcal{P}_4) \\
&\quad - P(\mathcal{P}_1 \cap \mathcal{P}_2 \cap \mathcal{P}_3 \cap \mathcal{P}_4) \\
&= P(B_1 \cap B_4) + P(B_2 \cap B_5) + P(B_1 \cap B_3 \cap B_5) + P(B_2 \cap B_3 \cap B_4) \\
&\quad - P((B_1 \cap B_4) \cap (B_2 \cap B_5)) - P((B_1 \cap B_4) \cap (B_1 \cap B_3 \cap B_5)) - P((B_1 \cap B_4) \cap (B_2 \cap B_3 \cap B_4)) \\
&\quad - P((B_2 \cap B_5) \cap (B_1 \cap B_3 \cap B_5)) - P((B_2 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&\quad - P((B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&\quad + P((B_1 \cap B_4) \cap (B_2 \cap B_5) \cap (B_1 \cap B_3 \cap B_5)) + P((B_1 \cap B_4) \cap (B_2 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&\quad + P((B_1 \cap B_4) \cap (B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&\quad + P((B_2 \cap B_5) \cap (B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&\quad - P((B_1 \cap B_4) \cap (B_2 \cap B_5) \cap (B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&= P(B_1 \cap B_4) + P(B_2 \cap B_5) + P(B_1 \cap B_3 \cap B_5) + P(B_2 \cap B_3 \cap B_4) \\
&\quad - P(B_1 \cap B_4 \cap B_2 \cap B_5) - P(B_1 \cap B_4 \cap B_3 \cap B_5) - P(B_1 \cap B_4 \cap B_2 \cap B_3) \\
&\quad - P(B_2 \cap B_5 \cap B_1 \cap B_3) - P(B_2 \cap B_5 \cap B_3 \cap B_4) - P(B_1 \cap B_3 \cap B_5 \cap B_2 \cap B_4) \\
&\quad + P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) + P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&\quad + P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) + P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&\quad - P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&= P_1 P_4 + P_2 P_5 + P_1 P_3 P_5 + P_2 P_3 P_4 - P_1 P_4 P_2 P_5 - P_1 P_4 P_3 P_5 - P_1 P_4 P_2 P_3 \\
&\quad - P_2 P_5 P_1 P_3 - P_2 P_5 P_3 P_4 + 2P_1 P_2 P_3 P_4 P_5 \\
&= P_1 P_4 (1 - P_2 P_5 - P_3 P_5 - P_2 P_3) + P_2 P_5 + P_1 P_3 P_5 + P_2 P_3 P_4 \\
&\quad - P_2 P_5 P_1 P_3 - P_2 P_5 P_3 P_4 + 2P_1 P_2 P_3 P_4 P_5
\end{aligned}$$

As a sanity check, if bridge 3 does not exist (i.e., if  $P_3 = 0$ ), then there are only 2 paths and by inclusion-exclusions,  $P(A) = P_1 P_4 + P_2 P_5 - P_1 P_4 P_2 P_5$ . In the limit that  $P_3 = 0$ , we see that, indeed, the above formula matches this probability.

(b) To solve for  $P(B_3|A)$  I use Bayes' rule:

$$P(B_3|A) = \frac{P(A|B_3)P_3}{P(A)}.$$

$P(A)$  has already been calculated. To solve for the probability of  $A$  conditioned on  $B_3$  we need only to condition each probability term in  $P(A)$  on  $B_3$ , which effectively turns all the  $P_3$  terms in the formula for  $P(A)$  to unity. Therefore,

$$P(A|B_3) = P_1 P_4 (1 - P_2 P_5 - P_5 - P_2) + P_2 P_5 + P_1 P_5 + P_2 P_4 - P_2 P_5 P_1 - P_2 P_5 P_4 + 2P_1 P_2 P_4 P_5,$$

and we can insert  $P(A|B_3)$  and  $P(A)$  into Bayes' rule to obtain the answer:

$$\begin{aligned}
P(A|B_3) &= \\
&\frac{P_1 P_4 P_3 (1 - P_2 P_5 - P_5 - P_2) + P_3 P_2 P_5 + P_3 P_1 P_5 + P_3 P_2 P_4 - P_3 P_2 P_5 P_1 - P_3 P_2 P_5 P_4 + 2P_1 P_2 P_3 P_4 P_5}{P_1 P_4 (1 - P_2 P_5 - P_3 P_5 - P_2 P_3) + P_2 P_5 + P_1 P_3 P_5 + P_2 P_3 P_4 - P_2 P_5 P_1 P_3 - P_2 P_5 P_3 P_4 + 2P_1 P_2 P_3 P_4 P_5}.
\end{aligned}$$

**Problem 33.** Without loss of generality, let us call the door that you picked door 1, and let us arbitrarily denote the remaining doors by 2 and 3. Let  $C_i$  denote the event that the car is behind door  $i$  and  $H_i$  denote the event that the host opens door  $i$ . The original probability that you guessed the door with the car is  $P(C_1) = 1/3$ . Since the host will not open door 1, and he will also not open the door with the car behind it, we have the following probabilities:

$$P(H_1|C_1) = 0$$

$$P(H_2|C_1) = \frac{1}{2}$$

$$P(H_3|C_1) = \frac{1}{2}$$

$$P(H_1|C_2) = 0$$

$$P(H_2|C_2) = 0$$

$$P(H_3|C_2) = 1$$

$$P(H_1|C_3) = 0$$

$$P(H_2|C_3) = 1$$

$$P(H_3|C_3) = 0$$

If the host opens door 3, we would like to know  $P(C_2|H_3)$ , because if this value is higher than  $1/3$ , it is in our interest to switch to door 2. Likewise if the host opens door 2, we would like to know  $P(C_3|H_2)$  to know if we should switch to door 3. Given the symmetry of the problem  $P(C_2|H_3) = P(C_3|H_2)$ , so I only need to compute the probability once, which I do using Baye's rule:

$$\begin{aligned} P(C_2|H_3) &= \frac{P(H_3|C_2)P(C_2)}{P(H_3|C_1)P(C_1) + P(H_3|C_2)P(C_2) + P(H_3|C_3)P(C_3)} \\ &= \frac{1 \left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right) \left(\frac{1}{3}\right) + 1 \left(\frac{1}{3}\right) + 0} \\ &= \frac{2}{3}. \end{aligned}$$

It is therefore in your interest to switch to door 2 if the host opens door 3 or to switch to door 3 if the host opens door 2.

**Problem 34.**

- (a)  $P(A) = 1/6$ ,  $P(B) = |\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}|/36 = 1/6$ ,  $P(A, B) = 1/36$ . Since  $P(A)P(B) = 1/36 = P(A, B)$ , the events are indeed independent.
- (b)  $P(C) = 1/6$ , so that  $P(A, C) = 1/36$ , and  $P(A, C) = 1/36$  and therefore they are independent.
- (c)  $P(B)P(C) = 1/36$ ,  $P(B, C) = 1/36$ , so yes, they are independent.
- (d) The events  $A$ ,  $B$  and  $A$ ,  $C$  and  $B$ ,  $C$  are pairwise independent. We also need to check if  $P(A, B, C) = P(A)P(B)P(C)$ . The probability of  $P(A, B, C)$  equals 0 since those events cannot all occur at once, whereas  $P(A)P(B)P(C) \neq 0$ . Therefore, the events  $A$ ,  $B$  and  $C$  are not independent.

**Problem 35.** Let  $X_1$  denote the outcome of the first roll,  $W$  denote the event that I win, and let the probability of tails be  $q$  ( $= 1 - p$ ). From Bayes', the probability that the first roll was heads is given that I won the game is:

$$P(X_1 = H|W) = \frac{P(W|X_1 = H)P(X_1 = H)}{P(W)}.$$

I first calculate the probability of winning:

$$\begin{aligned} P(W) &= P(HH) + P(THH) + P(HTHH) + P(THTHH) + \dots \\ &= P(HH) + P(HTHH) + \dots + P(THH) + P(THTHH) + \dots \\ &= q^0p^2 + qp^3 + \dots + qp^2 + q^2p^3 + \dots \\ &= (q^0p + qp^2 + \dots)p + (q^0p^2 + qp^3 + \dots)q \end{aligned}$$

Note that since  $P(W) = P(W|X_1 = H)p + P(W|X_1 = T)q$ , the first term in the parentheses in the above equation represents  $P(W|X_1 = H)$  while the second term in the parentheses represents  $P(W|X_1 = T)$ . I solve for both of these separately:

$$\begin{aligned} P(W|X_1 = H) &= q^0p + qp^2 + \dots \\ &= (qp)^0p + (qp)^1p + \dots \\ &= p \sum_{k=0}^{\infty} (qp)^k \\ &= \frac{p}{1 - qp}, \end{aligned}$$

while

$$\begin{aligned} P(W|X_1 = T) &= q^0p^2 + qp^3 + \dots \\ &= (qp)^0p^2 + (qp)^1p^2 + \dots \\ &= p^2 \sum_{k=0}^{\infty} (qp)^k \\ &= \frac{p^2}{1 - qp}, \end{aligned}$$

where I have used the formula for a geometric series. Thus we can compute the probability of winning as

$$\begin{aligned} P(W) &= P(W|X_1 = H)p + P(W|X_1 = T)q \\ &= \frac{p^2}{1 - qp} + \frac{p^2q}{1 - qp} \end{aligned}$$

Finally, we can plug all of these formulas into Bayes' equation:

$$\begin{aligned}
P(X_1 = H|W) &= \frac{P(W|X_1 = H)P(X_1 = H)}{P(W)} \\
&= \frac{\frac{p^2}{1-qp}}{\frac{p^2}{1-qp} + \frac{p^2q}{1-qp}} \\
&= \frac{1}{1+q} \\
&= \frac{1}{2-p}.
\end{aligned}$$

**Problem 36.** Let  $H_{n+1}$  denote the event that the  $(n+1)^{th}$  flip is a head,  $H \dots H$  denote the event of observing  $n$  heads and  $F$  denote the event that we pick the fair coin. We would like to find  $P(H_{n+1}|H \dots H)$ , and we know that from the law of total probability  $P(H_{n+1}) = P(H_{n+1}|F)P(F) + P(H_{n+1}|F^c)P(F^c)$ . By conditioning all of the probabilities on  $H \dots H$ , this equation gives a formula for the probability we desire:

$$P(H_{n+1}|H \dots H) = P(H_{n+1}|F, H \dots H)P(F|H \dots H) + P(H_{n+1}|F^c, H \dots H)P(F^c|H \dots H). \quad (1.4)$$

The terms  $P(H_{n+1}|F, H \dots H)$  and  $P(H_{n+1}|F^c, H \dots H)$  are conditionally independent of  $H \dots H$  given  $F$  (or  $F^c$ ) and therefore these probabilities are  $1/2$  and  $1$  respectively. We may obtain  $P(F|H \dots H)$  from Bayes' rule:

$$\begin{aligned}
P(F|H \dots H) &= \frac{P(H \dots H|F)P(F)}{P(H \dots H|F)P(F) + P(H \dots H|F^c)P(F^c)} \\
&= \frac{\left(\frac{1}{2}\right)^n \frac{1}{2}}{\left(\frac{1}{2}\right)^n \frac{1}{2} + \frac{1}{2}} \\
&= \frac{1}{1+2^n}.
\end{aligned}$$

Thus:

$$\begin{aligned}
P(H_{n+1}|H \dots H) &= \left(\frac{1}{2}\right) \frac{1}{1+2^n} + \left(1 - \frac{1}{1+2^n}\right) \\
&= 1 - \frac{1}{2(1+2^n)}.
\end{aligned}$$

We can check this formula for the extremes that  $n = 0$  and  $n \rightarrow \infty$ . In the first case, if  $n = 0$ , we can calculate the probability of heads directly:  $P(H) = (1/2)(1/2) + 1(1/2) = 3/4$ , which matches what the formula predicts when  $n = 0$ . When  $n \rightarrow \infty$ , we would expect that the coin is probably unfair, so that the probability of the next flip landing heads is 1. Indeed, this is what the formula predicts in the limit that  $n \rightarrow \infty$ .

**Problem 37.** Let  $X_i$  denote the number of girls for the  $i^{th}$  child. Note that  $X_i$  can only take on values 0 or 1. We seek the probability  $P(X_1 = 1, \dots, X_n = 1 | X_1 + \dots + X_n \geq 1)$ . This can be re-written with Bayes' rule:

$$P(X_1 = 1, \dots, X_n = 1 | X_1 + \dots + X_n \geq 1) = \frac{P(X_1 + \dots + X_n \geq 1 | X_1 = 1, \dots, X_n = 1)P(X_1 = 1, \dots, X_n = 1)}{P(X_1 + \dots + X_n \geq 1)}. \quad (1.5)$$

In the numerator, the first term is just 1 since  $X_1 + \dots + X_n$  is guaranteed to be at least 1 if all  $X_1, \dots, X_n$  are equal to 1. The second term is  $(1/2)^n$  since all boy/girl events are independent. To calculate the denominator, it is easier to consider its complement:  $P(X_1 + \dots + X_n \geq 1) = 1 - P(X_1 + \dots + X_n = 0) = 1 - P(X_1 = 0, \dots, X_n = 0) = 1 - (1/2)^n$ . Putting all of these into Bayes' rule, I obtain:

$$P(X_1 = 1, \dots, X_n = 1 | X_1 + \dots + X_n \geq 1) = \frac{1}{2^n - 1}.$$

We can test this formula for low values of  $n$ . For  $n = 1$ ,  $P(X_1 = 1 | X_1 = 1) = 1$ , which the above formula predicts. For  $n = 2$ , by listing out the boy/girl event space, it is not difficult to determine that  $P(X_1 = 1, X_2 = 1 | X_1 + X_2 \geq 1) = 1/3$ , which also what the above formula predicts.

**Problem 38.** Let  $L$  be the event that the family has at least 1 daughter named Lilia,  $G \dots G$  be the event that the  $n$  children are girls,  $BG \dots G$  be the event that the first child is a boy and the following  $n - 1$  children are girls,  $GBG \dots G$  be the event that the first child is a girl, the second is a boy and following  $n - 2$  children are girls, etc.... We are interested in  $P(G \dots G | L)$  which we can obtain with Bayes' rule:

$$P(G \dots G | L) = \frac{P(L | G \dots G)P(G \dots G)}{P(L)}. \quad (1.6)$$

The denominator is given by:

$$\begin{aligned} P(L) &= 1 - P(L^c) \\ &= 1 - P(G \dots G)[P(L^c | G \dots G) + P(L^c | BG \dots G) + P(L^c | GBG \dots G) + \dots + P(L^c | B \dots B)] \\ &= 1 - P(G \dots G)[(1 - \alpha)^n + (1 - \alpha)^{n-1} + (1 - \alpha)^{n-1} + \dots + (1 - \alpha)^0] \\ &= 1 - P(G \dots G) \sum_{k=0}^n \frac{n!}{k!(n-k)!} (1 - \alpha)^k \\ &= 1 - P(G \dots G)(2 - \alpha)^n. \end{aligned}$$

In the third line I used the fact that to have no daughters named Lilia given  $n$  daughters, the daughters need to be not be named Lilia  $n$  times, which occurs with probability  $(1 - \alpha)^n$ . In the fourth line, I used the fact that the total number of sequences of  $BG \dots G$ ,  $GBG \dots G$ , ..., is given by the number of permutations of  $n$  elements ( $n!$ ) divided by the number of repeats for any element (which for the  $G$ s is  $k!$ , where  $k$  is the number of  $G$ s in the sequence, and which for the  $B$ s is  $(n - k)!$ ). This is a simple combinatorics problem which will be discussed in the following chapter. Finally, to solve the summation, I used the binomial theorem.

We need one more probability, which is  $P(L | G \dots G) = 1 - P(L^c | G \dots G) = 1 - (1 - \alpha)^n$ . Sticking in all of these probabilities into Baye's rule, I obtain:



$$\begin{aligned}
P(G \dots G|L) &= \frac{[1 - (1 - \alpha)^n] \left(\frac{1}{2}\right)^n}{1 - \left(\frac{1}{2}\right)^n (2 - \alpha)^n} \\
&= \frac{1 - (1 - \alpha)^n}{2^n - (2 - \alpha)^n} \\
&\approx \frac{1 - (1 - n\alpha)}{2^n - 2^n(1 - \frac{n\alpha}{2})} \\
&= \frac{1}{2^{n-1}},
\end{aligned}$$

where I have used a Taylor expansion to simplify the polynomial terms since  $\alpha \ll 1$ . The case of  $n = 2$  corresponds to Problem 7 of section 1.4.5 in the book. Evaluating my formula with  $n = 2$  I find that  $P(GG|L) = (2 - \alpha)/(4 - \alpha) \approx 1/2$  which is the same formula as the answer to Problem 7 of section 1.4.5.

**Problem 39.** Let  $R$  be the event that a randomly chosen child is a girl, and  $G \dots G$  be the event that the family has  $n$  girls. We seek to find  $P(G \dots G|R)$ , which we can get from Bayes' rule:

$$P(G \dots G|R) = \frac{P(R|G \dots G)P(G \dots G)}{P(R)}.$$

$R$  is certain to happen conditioned on  $G \dots G$ ,  $P(G \dots G)$  is simply  $(1/2)^n$  and the probability of randomly choosing a girl from a family without any prior information about genders is  $1/2$  as shown for the  $n = 1$  ( $S = \{G, B\}$ ) and  $n = 2$  ( $S = \{BB, BG, GB, GG\}$ ) cases below:

$n = 1$  :

$$P(R) = P(R|G)P(G) + P(R|B)P(B) = 1 \left(\frac{1}{2}\right) + 0 \left(\frac{1}{2}\right) = \frac{1}{2},$$

$n = 2$  :

$$\begin{aligned}
P(R) &= P(R|BB)P(BB) + P(R|BG)P(BG) + P(R|GB)P(GB) + P(R|GG)P(GG) \\
&= 0 \left(\frac{1}{4}\right) + \left(\frac{1}{2}\right) \left(\frac{1}{4}\right) + \left(\frac{1}{2}\right) \left(\frac{1}{4}\right) + 1 \left(\frac{1}{4}\right) \\
&= \frac{1}{2}.
\end{aligned}$$

Therefore  $P(G \dots G|R) = 1/2^{n-1}$ .



## Chapter 2

# Combinatorics: Counting Methods

**Problem 1.** We can use the multiplication principle, making sure to enumerate all the cream/-sugar/milk possibilities:

$$4 \cdot 3 \cdot \left[ \binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} \right] = 4 \cdot 3 \cdot 8 = 96. \quad (2.1)$$

**Problem 2.** Let  $N$  be the number of unique permutations of the 8 people in the 12 chairs. The 4 empty chairs are indistinguishable, so, for any given unique permutation, the permutations amongst those 4 chairs do not count toward the number of unique permutations,  $N$ . We know that the total number of permutations (including the non-unique permutations is  $12!$ ), and therefore  $12! = N4!$ , so that

$$\frac{12!}{4!} = 19958400. \quad (2.2)$$

**Problem 3.**

- (a) Let  $B$  represent the set of the 20 black cell phones,  $\{b_1, b_2, \dots, b_{20}\}$ , and  $W$  represent the set of the 30 white cell phones,  $\{w_1, w_2, \dots, w_{30}\}$ . Let  $\mathcal{B}$  be the set containing all possible sets of the 4 distinct black cell phones that were chosen (without replacement) from the 20 black cell phones,  $\mathcal{B} = \{\{b_1, b_2, b_3, b_4\}, \{b_1, b_2, b_3, b_5\} \dots \{b_{17}, b_{18}, b_{19}, b_{20}\}\}$ , and  $\mathcal{W}$  be the corresponding set for the 6 white cell phones. Therefore, the sets of sets representing all unique ways to obtain 4 black cell phones and 6 white cell phones is given by  $\{\mathcal{B}_1 \cup \mathcal{W}_1, \mathcal{B}_1 \cup \mathcal{W}_2 \dots \mathcal{B}_{|\mathcal{B}|} \cup \mathcal{W}_{|\mathcal{W}|}\}$ , whose total cardinality can be seen to be  $|\mathcal{B}||\mathcal{W}|$ .  $|\mathcal{B}|$  is clearly  $\binom{20}{4}$ , and  $|\mathcal{W}|$  is clearly  $\binom{30}{6}$ , so the size of this set is  $\binom{20}{4}\binom{30}{6}$ . The sample space for this experiment is all possible unique sets of size 10 that can be chosen from  $B \cup W$ . Therefore, the probability of obtaining exactly 4 black cell phones is given by:

$$P(4 \text{ black phones}) = \frac{\binom{20}{4}\binom{30}{6}}{\binom{50}{10}} \approx 0.28. \quad (2.3)$$

In this problem I somewhat laboriously spelled out how to obtain the proper number of sets from the sample space with exactly 4 black cell phones. I did this for the purpose of illustration since this type of situation arises commonly in combinatorics problems. In the future I will typically be typically be more terse.

(b)

$$\begin{aligned} P(N_B < 3) &= P(N_B = 0 \cup N_B = 1 \cup N_B = 2) \\ &= P(N_B = 0) + P(N_B = 1) + P(N_B = 2) \\ &= \frac{\binom{20}{0}\binom{30}{10}}{\binom{50}{10}} + \frac{\binom{20}{1}\binom{30}{9}}{\binom{50}{10}} + \frac{\binom{20}{2}\binom{30}{8}}{\binom{50}{10}} \\ &\approx 0.14 \end{aligned}$$

**Problem 4.**

- (a) The sample space is all possible sets of length 5 chosen from the 52 cards, and the events we are interested in are all possible sets of size 5, one of which is certainly an ace. Therefore:

$$P(N_A = 1) = \frac{\binom{4}{1}\binom{48}{4}}{\binom{52}{5}} \approx 0.30.$$

- (b) Let  $N_A \geq 1$  be the event that the hand contains at least 1 ace. It will be easier to consider the complement of this event:

$$\begin{aligned} P(N_A \geq 1) &= 1 - P(N_A = 0) \\ &= 1 - \frac{\binom{48}{5}}{\binom{52}{5}} \\ &\approx 0.34. \end{aligned}$$

**Problem 5.** It will be convenient to use Bayes' rule so that we can move  $N_A \geq 1$  to the first slot of  $P(\cdot|\cdot)$ :

$$P(N_A = 2|N_A \geq 1) = \frac{P(N_A \geq 1|N_A = 2)P(N_A = 2)}{P(N_A \geq 1)}. \quad (2.4)$$

We have already computed the denominator in the previous problem. In the numerator  $P(N_A \geq 1|N_A = 2) = 1$  since the probability of obtaining at least 1 ace is unity if we already know there are 2 aces. The probability of obtaining exactly 2 aces is

$$P(N_A = 2) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} \approx 0.04,$$

and therefore:

$$P(N_A = 2|N_A \geq 1) \approx \frac{1 \cdot 0.04}{0.34} \approx 0.12. \quad (2.5)$$

**Problem 6.** Let  $C_4$  be the event that  $C$  receives exactly 4 spades. Each player has 13 cards, and between players  $A$  and  $B$ , we know there are 7 spades, and 19 non-spades. This leaves 6 spades and 20 non-spades to be chosen amongst players  $C$  and  $D$ . If the 26 cards are first dealt to  $A$  and  $B$ , and another 13 are dealt to  $C$ , then the probability that  $C$  obtains exactly 4 spades is:

$$P(C_4) = \frac{\binom{6}{4}\binom{20}{9}}{\binom{26}{13}} \approx 0.24.$$

**Problem 7.** Let  $J$  be the event that Joe is chosen and  $Y$  be the event that you are chosen. By inclusion-exclusion:

$$P(J \cup Y) = P(J) + P(Y) - P(J, Y).$$

There are  $\binom{1}{1}\binom{49}{14}$  different ways Joe can be chosen and the same number of ways you can be chosen. There are  $\binom{2}{2}\binom{48}{13}$  different ways both you and Joe can be chosen, and thus:

$$P(J \cup Y) = 2 \frac{\binom{49}{14}}{\binom{50}{15}} - \frac{\binom{48}{13}}{\binom{50}{15}} \approx 0.51.$$

**Problem 8.** In general, for a sequence with  $n$  elements,  $r$  of which are unique, the number of unique permutations is given by:

$$N = \frac{n!}{n_1!n_2!\dots n_r!}, \quad (2.6)$$

where  $n_i$  is the number of repeats of the  $i^{th}$  unique element in the original sequence. This can easily be shown, since the number of total permutations must be equal to the sum of each unique permutation, times the number of times each element in the unique permutation can be permuted amongst themselves,  $n! = Nn_1!n_2!\dots n_r!$ . For example, one unique permutation of the

word “Massachusetts” is Massachusetts itself. We see that the “a”s can be permuted  $2!$  ways amongst second and fifth position, while still forming the word Massachusetts. Likewise, the “s”s can be permuted  $4!$  ways and the “t”s  $2!$  ways, resulting in  $2!4!2!$  permutations of all letters which result in this unique permutation. Thus, the total number of ways of arranging the word “Massachusetts” is:

$$N = \frac{n!}{n_a!n_s!n_t!} = \frac{13!}{2!4!2!} = 64864800. \quad (2.7)$$

**Problem 9.**

(a) Using the formula for the binomial distribution, I find:

$$P(k = 8) = \binom{20}{8} p^8 (1 - p)^{20-8}.$$

(b) Since both the number of heads and number of tails must be  $> 8$ , the possible observed number of heads (tails) can be 9 (11) or 10 (10) or 11 (9). These are disjoint events, so the total probability we are interested in is

$$\begin{aligned} P(\{k = 9, k = 10, k = 11\}) &= P(k = 9) + P(k = 10) + P(k = 11) \\ &= \binom{20}{9} p^9 (1 - p)^{20-9} + \binom{20}{10} p^{10} (1 - p)^{20-10} + \binom{20}{11} p^{11} (1 - p)^{20-11} \\ &= \binom{20}{9} p^9 (1 - p)^9 [1 - 2p + 2p^2] + \binom{20}{10} p^{10} (1 - p)^{10}. \end{aligned}$$

**Problem 10.** Let  $u$  denote a move up and  $r$  denote a move to the right. A path from  $(0, 0)$  to  $(20, 10)$  can be represented by a sequence of  $us$  and  $rs$ . Note that in every possible sequence, there must be 10  $us$  and 20  $rs$  because we always need to travel 10 units up and 20 units to the right regardless of the path. Therefore, the problem reduces to ascertaining the number of unique sequences with 10  $us$  and 20  $rs$ , which, from Problem 8 we can see to be:

$$\frac{30!}{20!10!} = 30045015.$$

**Problem 11.** Let  $A$  denote the event that the message passes through  $(10, 5)$  on its way to  $(20, 10)$ . To reach the point  $(10, 5)$  on the way from  $(0, 0)$  to  $(20, 10)$ , the first 15 entries of the sequence must have exactly 5  $us$  and 10  $rs$ . This may occur in any number of the unique permutations  $15!/(10!5!)$ . To reach  $(20, 10)$ , the remaining entries must also contain exactly 5  $us$  and 10  $rs$ , again giving  $15!/(10!5!)$  unique permutations from  $(10, 5)$  to  $(20, 10)$ . The total number of unique permutations starting at  $(0, 0)$  and going through  $(10, 5)$  on its way to  $(20, 10)$  is therefore  $(15!/(10!5!))^2$ , so that the probability that the message goes through  $(10, 5)$  is:

$$P(A) = \frac{\left(\frac{15!}{10!5!}\right)^2}{\frac{30!}{10!20!}} \approx 0.30$$

**Problem 12.** Let  $A$  denote the event that the message passes through  $(10, 5)$ . This occurs if, out of the first 15 entries of the sequence there are exactly 5  $us$  and 10  $rs$  in any order. For a binary outcome experiment, the probability of obtaining 5  $us$  with probability  $p_a$  is given by the binomial distribution:

$$P(A) = \binom{15}{5} p_a^5 (1 - p_a)^{10}.$$

**Problem 13.** Let  $p_i$  be the probability of flipping a heads for coin  $i$  ( $i \in \{1, 2\}$ ), let  $C_i$  be the event that coin  $i$  is chosen. Using the law of total probability and the binomial distribution, I find:

(a)

$$\begin{aligned} P(N_H \geq 3) &= P(N_H = 3 \cup N_H = 4 \cup N_H = 5) \\ &= P(N_H = 3) + P(N_H = 4) + P(N_H = 5) \\ &= \sum_{i=1}^2 [P(N_H = 3|C_i)P(C_i) + P(N_H = 4|C_i)P(C_i) + P(N_H = 5|C_i)P(C_i)] \\ &= \frac{1}{2} \sum_{i=1}^2 \left[ \binom{5}{3} p_i^3 (1 - p_i)^2 + \binom{5}{4} p_i^4 (1 - p_i) + \binom{5}{5} p_i^5 \right] \\ &\approx 0.35. \end{aligned}$$

(b) From Bayes':

$$P(C_2|N_H \geq 3) = \frac{P(N_H \geq 3|C_2)P(C_2)}{P(N_H \geq 3)},$$

where  $P(N_H \geq 3)$  has already been solved,  $P(C_2) = 0.5$  and

$$P(N_H \geq 3|C_2) = \sum_{j=3}^5 \binom{5}{j} p_2^j (1 - p_2)^{5-j} \approx 0.21.$$

Therefore, the probably we are interested in is

$$P(C_2|N_H \geq 3) \approx \frac{0.21 \cdot 0.5}{0.35} = 0.3.$$

The fact that this probability is less than 0.5 makes sense, since more heads were observed than tails, and so it is more probable that coin 1 was chosen since the probability that it lands heads is higher.

**Problem 14.** There are  $\binom{13}{3,5,5}$  different ways Hannah and Sarah can be arranged on the same team. However, we do not care about players being assigned to a particular team name, we just care about the number of possible divisions. Therefore, to avoid over-counting, we must divide this value by  $2!$ . Likewise, there are  $\binom{15}{5,5,5}$  total ways to construct 3 teams of 5 each, and we must divide by  $3!$  since we only care about the number of possible divisions:

$$P(\text{H and S in same division}) = \frac{\frac{1}{2!} \binom{13}{3,5,5}}{\frac{1}{3!} \binom{15}{5,5,5}} \approx 0.29.$$

**Problem 15.** We would like to find  $P(N_1 > 1 \cup N_2 > 1 \cup \dots \cup N_6 > 1) = 1 - P(N_1 \leq 1, N_2 \leq 1, \dots, N_6 \leq 1)$ . For the first roll, we therefore have 6 allowable options, for the second 5 allowable options, .... Therefore the probability is:

$$P(N_1 > 1 \cup N_2 > 1 \cup \dots \cup N_6 > 1) = 1 - \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{6^5} \approx 0.91.$$

**Problem 16.**

- (a) Let  $A$  be the desired event. If the first 15 cards are to have 10 red cards, then there are  $\binom{10}{5}\binom{10}{10}$  different possible groups for the first 15 cards, of which we can arrange in  $15!$  possible ways. There are  $\binom{5}{5}$  possible groups for the remaining 5 cards, of which we can arrange  $5!$  possible ways. Finally, the total number of permutations of the 20 cards is  $20!$ , and therefore:

$$P(A) = \frac{\binom{10}{5}\binom{10}{10}15!\binom{5}{5}5!}{20!} \approx 0.02.$$

- (b) Let  $A'$  be the event we desire. This problem is almost identical to the first:

$$P(A') = \frac{\binom{10}{7}\binom{10}{8}15!\binom{3}{3}\binom{2}{2}5!}{20!} \approx 0.35.$$

**Problem 17.** Let  $B_i$  be the event that I choose bag  $i$  ( $i \in \{1, 2\}$ ) and  $N_r = 2$  be the event that I choose exactly 2 red marbles out of the 5. Using Bayes'

$$P(B_1|N_r = 2) = \frac{P(N_r = 2|B_1)P(B_1)}{P(N_r = 2|B_1)P(B_1) + P(N_r = 2|B_2)P(B_2)}. \quad (2.8)$$

The probability of choosing either bag is  $1/2$  and the probability of  $N_r = 2$  conditioned on choosing bag 1 is

$$P(N_r = 2|B_1) = \frac{\binom{6}{2}\binom{10}{3}}{\binom{16}{5}} \approx 0.41,$$

while the probability of  $N_r = 2$  conditioned on choosing bag 2 is

$$P(N_r = 2|B_2) = \frac{\binom{6}{2}\binom{15}{3}}{\binom{21}{5}} \approx 0.34.$$

Sticking all relevant probabilities into Bayes', I arrive at

$$P(B_1|N_r = 2) = \frac{0.41 \cdot 0.5}{0.41 \cdot 0.5 + 0.34 \cdot 0.5} \approx 0.55. \quad (2.9)$$

**Problem 18.** Let  $E^c$  denote the event that an error has not occurred for the  $X_i^{th}$  trial. We seek the probability of all sequences of length  $n$ , ending in  $E^c$ , where the 1st  $n - 1$  entries can be any sequence, provided they contain exactly  $k - 1$   $E^c$ s, which I denote by  $A_{n-1}$ . The probability we desire is  $P(A_{n-1}, X_n = E^c) = P(A_{n-1})P(X_n = E^c)$  by independence. I note that  $P(A_{n-1})$  is a binomial distribution, and therefore:

$$P(A_{n-1}, X_n = E^c) = \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} p = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

**Problem 19.** Let  $y_i \equiv x_i - 1$  for  $i = 1, \dots, 5$ , and therefore all  $y_i$ s can take on values  $\{0, 1, 2, \dots\}$ . The equation for which we are trying to find the number of distinct integer solutions then becomes:

$$y_1 + y_2 + y_3 + y_4 + y_5 = 95, \quad (2.10)$$

which has  $\binom{5+95-1}{95} = \binom{99}{95}$  integer solutions.



**Problem 20.** It is not difficult to explicitly to enumerate the total number of solutions when  $x_1 = 0, 1, \dots, 10$ . The total number of integer valued solutions is thus the number of solutions for when  $x_1 = 0$ , plus the number of solutions for when  $x_1 = 1$ , ... plus the total number of solutions for when  $x_1 = 10$ . In each one of these instances, we must find the number of integer solutions for the equation

$$x_2 + x_3 + x_4 = 100 - i,$$

(where  $x_2, x_3, x_4 \in \{0, 1, 2, \dots\}$ ) which has  $\binom{3+100-i-1}{100-i}$  solutions. Therefore, the total number of integer solutions for this equation,  $N$ , with  $x_1 \in \{0, 1, \dots, 10\}$  is:

$$\begin{aligned} N &= \sum_{i=0}^{10} \binom{3+100-i-1}{100-i} \\ &= \frac{1}{2} \sum_{i=0}^{10} [10302 - 203i + i^2] \\ &= \frac{11 \cdot 10302}{2} - \frac{203}{2}(0 + 1 + 2 + \dots + 10) + \frac{1}{2}(0 + 1 + 4 + \dots + 100) \\ &= 51271. \end{aligned}$$

**Problem 21.** Let  $A_1 = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 100, x_1 \in \{41, 42, \dots\}, x_2, x_3 \in \{0, 1, 2, \dots\}\}$ , and let  $A_2$  and  $A_3$  be defined analogously. By inclusion-exclusion, the total number of possible unique integer solutions to this problem is then:

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| &= |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3| \\ &= 3|A_1| - 3|A_1 \cap A_2| + |A_1 \cap A_2 \cap A_3|, \end{aligned}$$

where the second equality follows from symmetry. The cardinality of  $|A_1 \cap A_2 \cap A_3|$  is 0 since it is impossible to have all  $x_i$ s  $> 40$  and constrained to add to 100.

The cardinality of  $|A_1|$  can be found by letting  $y_1 \equiv x_1 - 41$ , so that  $y_1, x_2, x_3 \in \{0, 1, 2, \dots\}$ :  $y_1 + x_2 + x_3 = 59$ , which has  $\binom{3+59-1}{59} = \binom{61}{59}$  solutions. The cardinality of  $|A_1 \cap A_2|$  can be found by letting  $y_1 \equiv x_1 - 41$  and  $y_2 \equiv x_2 - 41$  so that  $y_1, y_2, x_3 \in \{0, 1, 2, \dots\}$ :  $y_1 + y_2 + x_3 = 18$ , which has  $\binom{3+18-1}{18} = \binom{20}{18}$  solutions. Therefore, the total number of solutions to this problem is:

$$|A_1 \cup A_2 \cup A_3| = 3 \binom{61}{59} - 3 \binom{20}{18} = 4920.$$

The following bit of python code confirms what we derived theoretically:

```
In [1]: i=range(101); j=range(101); k=range(101)

In [2]: tups = [(x, y, z) for x in i for y in j for z in k]

In [3]: len([x for x in tups if x[0]+x[1]+x[2]==100
...: and (x[0]>40 or x[1]>40 or x[2]>40)])
Out[3]: 4920
```



## Chapter 3

# Discrete Random Variables

**Problem 1.**

- (a)  $R_X = \{0, 1, 2\}$
- (b)  $P(X \geq 1.5) = P(X = 2) = 1/6$
- (c)  $P(0 < X < 2) = P(X = 1) = 1/3$
- (d)

$$\begin{aligned}
 P(X = 0 | X < 2) &= \frac{P(X = 0 \cap X < 2)}{P(X < 2)} \\
 &= \frac{P(X = 0)}{P(X < 2)} \\
 &= \frac{1/2}{1/2 + 1/3} \\
 &= \frac{3}{5}
 \end{aligned}$$

**Problem 2.** From the set-up of the problem, we have that:

$$P_X(x) = \begin{cases} p' & \text{for } x = 0 \\ p & \text{for } x = 1 \\ p & \text{for } x = 2 \\ p' & \text{for } x = 3 \\ 0 & \text{otherwise,} \end{cases}$$

as well as the following equation  $P(X = 1 \text{ or } X = 2) = (1/2)P(X = 0 \text{ or } X = 3)$ , so that  $p = p'/2$ . Finally, we know that the PMF must be normalized, leading to the following coupled equations:

$$\begin{cases} 2p' + 2p = 1 \\ p = \frac{1}{2}p', \end{cases}$$

which, when solved for results in  $p = 1/6$  and  $p' = 1/3$ . Thus, the PMF for this problem is:

$$P_X(x) = \begin{cases} \frac{1}{3} & \text{for } x = 0 \\ \frac{1}{6} & \text{for } x = 1 \\ \frac{1}{6} & \text{for } x = 2 \\ \frac{1}{3} & \text{for } x = 3 \\ 0 & \text{otherwise,} \end{cases}$$

which can easily be verified to be normalized.

**Problem 3.** The range of both  $X$  and  $Y$  is  $\{1, 2, \dots, 6\}$ , so that  $R_Z = \{-5, 4, \dots, 4, 5\}$ . We may find the PMF by conditioning and using the law of total probability:

$$\begin{aligned}
P(Z = k) &= P(X - Y = k) \\
&= \sum_{y=1}^6 P(X = k + Y | Y = y) P(Y = y) \\
&= \sum_{y=1}^6 P(X = k + y | Y = y) P(Y = y) \\
&= \sum_{y=1}^6 P(X = k + y) P(Y = y) \\
&= \sum_{y=1}^6 P(X = k + y) \frac{1}{6} \\
&= \frac{1}{6} \sum_{y=1}^6 \frac{1}{6} \mathbb{1}\{1 \leq k + y \leq 6\},
\end{aligned}$$

where the fourth equality follows since  $X$  and  $Y$  are independent, and where  $\mathbb{1}\{\cdot\}$  is the so called indicator function which is equal to 1 if its argument evaluates to true, and 0 otherwise. By explicitly evaluating the sum for all  $k$ , I find that  $P(Z = -5) = 1/36, P(Z = -4) = 2/36, \dots, P(Z = 0) = 6/36, P(Z = 1) = 5/36, \dots, P(Z = 5) = 1/36$ , which can be conveniently written as:

$$P(Z = k) = \frac{6 - |k|}{36}, \quad (3.1)$$

and which can explicitly be checked to be normalized.

**Problem 4.**

(a) Since  $X$  and  $Y$  are independent:

$$\begin{aligned}
P(X \leq 2, Y \leq 2) &= P(X \leq 2)P(Y \leq 2) \\
&= [P_X(1) + P_X(2)][P_Y(1) + P_Y(2)] \\
&= \left(\frac{1}{4} + \frac{1}{8}\right) \left(\frac{1}{6} + \frac{1}{6}\right) \\
&= \frac{1}{8}.
\end{aligned}$$

(b) By inclusion-exclusion (and also using independence):

$$\begin{aligned}
P(X > 2 \cup Y > 2) &= P(X > 2) + P(Y > 2) - P(X > 2, Y > 2) \\
&= P(X > 2) + P(Y > 2) - P(X > 2)P(Y > 2) \\
&= [P_X(3) + P_X(4)] + [P_Y(3) + P_Y(4)] - [P_X(3) + P_X(4)][P_Y(3) + P_Y(4)] \\
&= \left(\frac{1}{8} + \frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{3}\right) - \left(\frac{1}{8} + \frac{1}{2}\right) \left(\frac{1}{3} + \frac{1}{3}\right) \\
&= \frac{7}{8}.
\end{aligned}$$

(c) Since  $X$  and  $Y$  are independent,  $P(X > 2 | Y > 2) = P(X > 2) = 1/8 + 1/2 = 5/8$ .

(d) I use conditioning, the law of total probability and independence to solve for this:

$$\begin{aligned}
P(X < Y) &= \sum_{y=1}^4 P(X < Y|Y = y)P(Y = y) \\
&= \sum_{y=1}^4 P(X < y|Y = y)P(Y = y) \\
&= \sum_{y=1}^4 P(X < y)P(Y = y) \\
&= P(X < 1)P(Y = 1) + P(X < 2)P(Y = 2) + P(X < 3)P(Y = 3) + P(X < 4)P(Y = 4) \\
&= P(X = 1)P(Y = 2) + [P(X = 1) + P(X = 2)]P(Y = 3) + [P(X = 1) \\
&\quad + P(X = 2) + P(X = 3)]P(Y = 4) \\
&= \frac{1}{4} \cdot \frac{1}{6} + \left(\frac{1}{4} + \frac{1}{8}\right) \frac{1}{3} + \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{8}\right) \frac{1}{3} \\
&= \frac{1}{3}.
\end{aligned}$$

**Problem 5.** Let  $X_i$  denote the number of cars that student  $i$  owns (which can be either 0 or 1). We then seek the probability  $P(X_1 + X_2 + \dots + X_{50} > 30)$ , where the probability that  $X_i = 1$  is  $1/2$  for all  $i$ . In other words, we seek the probability of obtaining at least 31 successes out of 50 Bernoulli trials. We may obtain the probability of 31 successes out of 50 Bernoulli trials by evaluating a Binomial(50, 0.5) distribution at 31, the probability of 32 successes out of 50 Bernoulli trials by evaluating a Binomial(50, 0.5) distribution at 32, etc ... . Therefore:

$$\begin{aligned}
P(X_1 + X_2 + \dots + X_{50} > 30) &= \sum_{k=31}^{50} \binom{50}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{50-k} \\
&= \left(\frac{1}{2}\right)^{50} \sum_{k=31}^{50} \binom{50}{k} \\
&\approx 0.06,
\end{aligned}$$

where the summation has been evaluated numerically.

**Problem 6.** The formula for  $P(X_N = 0)$  was derived in the book and is given by:

$$P(X_N = 0) = \frac{1}{2!} - \frac{1}{3!} + \dots (-1)^N \frac{1}{N!}.$$

I will need this formula in my answer below. Let  $A_i$  ( $i = 1, \dots, N$ ) be the event that the  $i^{th}$  person receives their hat. Therefore, for  $X_N = 1$ :

$$\begin{aligned}
P(X_N = 1) &= P(A_1, A_2^c, A_3^c, \dots, A_N^c) + P(A_1^c, A_2, A_3^c, \dots, A_N^c) + \dots + P(A_1^c, A_2^c, A_3^c, \dots, A_N) \\
&= N \cdot P(A_1, A_2^c, A_3^c, \dots, A_N^c) \\
&= NP(A_1)P(A_2^c, A_3^c, \dots, A_N^c) \\
&= N \frac{1}{N} P(X_{N-1} = 0) \\
&= P(X_{N-1} = 0),
\end{aligned}$$

where I have used symmetry in the second equality, independence in the third and the fact that the probability that person 1 gets their hat out of  $N$  hats is  $1/N$ .

For  $X_N = 2$ :

$$\begin{aligned} P(X_N = 2) &= \sum_{i < j} P(A_i, A_j) P(X_{N-2} = 0) \\ &= \binom{N}{2} P(A_1, A_2) P(X_{N-2} = 0) \\ &= \binom{N}{2} \frac{1}{N} \frac{1}{N-1} P(X_{N-2} = 0) \\ &= \frac{1}{2!} P(X_{N-2} = 0), \end{aligned}$$

where I am summing over all  $N$  choose 2 unordered pairs of people who get their hats. The probability that the first person gets their hat is  $1/N$  while the probability that the second person gets their hat is  $1/(N-1)$ .

Continuing in this fashion one can see the general formula for the PMF we would like to derive:

$$P(X_N = k) = \frac{1}{k!} P(X_{N-k} = 0) \quad \text{for } k = 0, 1, 2, \dots, N,$$

where

$$P(X_{N-k} = 0) = \frac{1}{2!} - \frac{1}{3!} + \dots (-1)^{N-k} \frac{1}{(N-k)!}.$$

**Problem 7.** Computing the probabilities will be simplified by noting that  $P(X > 5) = 1 - P(X \leq 5)$ , and  $P(X > 5 | X < 8) = P(5 < X < 8) / P(X < 8)$ . Note, that I do not explicitly evaluate the formulas to obtain a numerical answer, but this can easily be done numerically on a computer.

(a)  $X \sim \text{Geom}(1/5)$

(i)

$$P(X > 5) = 1 - \sum_{k=1}^5 \frac{1}{5} \left(1 - \frac{1}{5}\right)^{k-1}$$

(ii)

$$P(2 < X \leq 6) = \sum_{k=3}^6 \frac{1}{5} \left(1 - \frac{1}{5}\right)^{k-1}$$

(iii)

$$\begin{aligned} P(X > 5 | X < 8) &= \frac{P(5 < X < 8)}{P(X < 8)} \\ &= \frac{\sum_{k=6}^7 \frac{1}{5} \left(1 - \frac{1}{5}\right)^{k-1}}{\sum_{k=1}^7 \frac{1}{5} \left(1 - \frac{1}{5}\right)^{k-1}} \end{aligned}$$

(b)  $X \sim \text{Bin}(10, 1/3)$

(i)

$$P(X > 5) = 1 - \sum_{k=0}^5 \binom{10}{k} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{10-k}$$

(ii)

$$P(2 < X \leq 6) = \sum_{k=3}^6 \binom{10}{k} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{10-k}$$

(iii)

$$P(X > 5 | X < 8) = \frac{\sum_{k=6}^7 \binom{10}{k} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{10-k}}{\sum_{k=0}^7 \binom{10}{k} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{10-k}}$$

(c)  $X \sim \text{Pascal}(3, 1/2)$ 

(i)

$$P(X > 5) = 1 - \sum_{k=3}^5 \binom{k-1}{2} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{k-3}$$

(ii)

$$P(2 < X \leq 6) = \sum_{k=3}^6 \binom{k-1}{2} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{k-3}$$

(iii)

$$P(X > 5 | X < 8) = \frac{\sum_{k=6}^7 \binom{k-1}{2} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{k-3}}{\sum_{k=3}^7 \binom{k-1}{2} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{k-3}}$$

(d)  $X \sim \text{Hypergeom}(10, 10, 12)$ 

(i)

$$P(X > 5) = 1 - \sum_{j=2}^5 \frac{\binom{10}{j} \binom{10}{12-j}}{\binom{20}{12}}$$

(ii)

$$P(2 < X \leq 6) = \sum_{j=3}^6 \frac{\binom{10}{j} \binom{10}{12-j}}{\binom{20}{12}}$$

(iii)

$$P(X > 5 | X < 8) = \frac{\sum_{j=6}^7 \frac{\binom{10}{j} \binom{10}{12-j}}{\binom{20}{12}}}{\sum_{j=2}^7 \frac{\binom{10}{j} \binom{10}{12-j}}{\binom{20}{12}}}$$

(e)  $X \sim \text{Pois}(5)$ 

(i)

$$P(X > 5) = 1 - \sum_{k=0}^5 \frac{e^{-5} 5^k}{k!}$$

(ii)

$$P(2 < X \leq 6) = \sum_{k=3}^6 \frac{e^{-5} 5^k}{k!}$$

(iii)

$$P(X > 5 | X < 8) = \frac{\sum_{k=6}^7 \frac{e^{-5} 5^k}{k!}}{\sum_{k=0}^7 \frac{e^{-5} 5^k}{k!}}$$



**Problem 8.**

- (a) In general, for this problem  $P(X = x) = P(F_1, F_2, \dots, F_{x-1}, S_x) = P(F_1)P(F_2) \dots P(F_{x-1})P(S_x)$ , where I have used independence. Therefore  $P(X = 1) = P(S_1) = 1/2$ ,

$$\begin{aligned} P(X = 2) &= P(F_1)P(S_2) \\ &= \frac{1}{2} \left[ 1 - \left( \frac{1}{2} \right)^2 \right] \\ &= \frac{3}{8} \end{aligned}$$

and

$$\begin{aligned} P(X = 3) &= P(F_1)P(F_2)P(S_3) \\ &= \left( \frac{1}{2} \right) \left( \frac{1}{2} \right)^2 \left[ 1 - \left( \frac{1}{2} \right)^3 \right] \\ &= \frac{7}{64}. \end{aligned}$$

- (b) By inspection, one can determine that the general formula for  $P(X = k)$  for  $k = 1, 2, \dots$  is:

$$P(X = k) = \left[ \prod_{j=0}^{k-1} \left( \frac{1}{2} \right)^j \right] \left[ 1 - \left( \frac{1}{2} \right)^k \right].$$

- (c)

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - [P(X = 1) + P(X = 2)] \\ &= 1 - \left[ \frac{1}{2} + \frac{3}{8} \right] \\ &= \frac{1}{8} \end{aligned}$$

- (d)

$$\begin{aligned} P(X = 2 | X > 1) &= \frac{P(X = 2, X > 1)}{P(X > 1)} \\ &= \frac{P(X = 2)}{1 - P(X = 1)} \\ &= \frac{3/8}{1 - 1/2} \\ &= \frac{3}{4} \end{aligned}$$

**Problem 9.** To prove this equation, I will work on the RHS and LHS of the equation separately. Let me first simplify the LHS:

$$\begin{aligned}
P(X > m + l | X > m) &= \frac{P(X > m + l, X > m)}{P(X > m)} \\
&= \frac{P(X > m + l)}{P(X > m)} \\
&= \frac{\sum_{k=m+l+1}^{\infty} p(1-p)^{k-1}}{\sum_{j=m+1}^{\infty} p(1-p)^{j-1}} \\
&= \frac{p(1-p)^{m+l}[1 + (1-p) + (1-p)^2 + \dots]}{p(1-p)^m[1 + (1-p) + (1-p)^2 + \dots]} \\
&= (1-p)^l.
\end{aligned}$$

The RHS can also be simplified:

$$\begin{aligned}
P(X > l) &= \sum_{k=l+1}^{\infty} p(1-p)^{k-1} \\
&= p(1-p)^l[1 + (1-p) + (1-p)^2 + \dots] \\
&= p(1-p)^l \frac{1}{1 - (1-p)} \\
&= (1-p)^l,
\end{aligned}$$

where in the third equality I summed the geometric series. This result is exactly what I obtained when I simplified the LHS of the equation, and thus the equation is proved.

**Problem 10.**

- (a) We have dealt with this type of problem extensively in the combinatorics chapter. The probability we seek is  $|A|/|S|$ , where  $A$  is the set of all possible ways we can pick exactly 4 red balls out of 10, and  $S$  is the set of all possible ways to pick 10 balls out of the 50. Let  $X_r$  be total number of red balls drawn. We thus we:

$$P(X_r = 4) = \frac{\binom{20}{4}\binom{30}{6}}{\binom{50}{10}} \approx 0.28. \quad (3.2)$$

- (b) We seek to find  $P(X_r = 4 | X_r \geq 3)$ . The inequality will be easier to deal with if I put it in the first slot of  $P(\cdot | \cdot)$ , and thus I start by employing Bayes' rule:

$$\begin{aligned}
P(X_r = 4 | X_r \geq 3) &= \frac{P(X_r \geq 3 | X_r = 4)P(X_r = 4)}{P(X_r \geq 3)} \\
&= \frac{P(X_r = 4)}{1 - P(X_r = 0) - P(X_r = 1) - P(X_r = 2)} \\
&= \frac{\frac{\binom{20}{4}\binom{30}{6}}{\binom{50}{10}}}{1 - \frac{\binom{20}{0}\binom{30}{10}}{\binom{50}{10}} - \frac{\binom{20}{1}\binom{30}{9}}{\binom{50}{10}} - \frac{\binom{20}{2}\binom{30}{8}}{\binom{50}{10}}} \\
&= \frac{\binom{20}{4}\binom{30}{6}}{\binom{50}{10} - \binom{30}{10} - \binom{20}{1}\binom{30}{9} - \binom{20}{2}\binom{30}{8}} \\
&\approx 0.33,
\end{aligned}$$

where in the second equality I have used the fact that  $P(X_r \geq 3|X_r = 4) = 1$  and in the third equality I used what was derived in the previous part of this problem.

**Problem 11.**

- (a) The average number of emails received on the weekend is 2 per hour or 8 per 4 hours. Since we are modeling this process with a Poisson distribution, the probability that you receive 0 emails on the weekend per 4 hour interval is:

$$P(k=0) = \frac{e^{-8} \cdot 8^0}{0!} \approx 3.4 \times 10^{-4}. \quad (3.3)$$

- (b) The average number of emails received on the weekend is 2 per hour and 10 per hour on any weekday. Let  $A_{wd}$  be the event that a weekday was chosen and  $A_{we}$  be the event that a weekend was chosen. This problem can be solved using Baye's rule

$$\begin{aligned} P(A_{wd}|k=0) &= \frac{P(k=0|A_{wd})P(A_{wd})}{P(k=0|A_{wd})P(A_{wd}) + P(k=0|A_{we})P(A_{we})} \\ &= \frac{e^{-10} \cdot \frac{5}{7}}{e^{-10} \cdot \frac{5}{7} + e^{-2} \cdot \frac{2}{7}} \\ &\approx 8.4 \times 10^{-4}. \end{aligned}$$

**Problem 12.** The CDF can easily be computed from the PMF:

$$F_X(x) = \begin{cases} 0 & \text{for } x < -2 \\ 0.2 & \text{for } -2 \leq x < -1 \\ 0.2 + 0.3 & \text{for } -1 \leq x < 0 \\ 0.2 + 0.3 + 0.2 & \text{for } 0 \leq x < 1 \\ 0.2 + 0.3 + 0.2 + 0.2 & \text{for } 1 \leq x < 2 \\ 0.2 + 0.3 + 0.2 + 0.2 + 0.1 & \text{for } x \geq 2 \end{cases} = \begin{cases} 0 & \text{for } x < -2 \\ 0.2 & \text{for } -2 \leq x < -1 \\ 0.5 & \text{for } -1 \leq x < 0 \\ 0.7 & \text{for } 0 \leq x < 1 \\ 0.9 & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2. \end{cases}$$

See Fig. 3.1 for a plot of this function.

**Problem 13.** Whenever there is a jump in the CDF at a value of  $x$ , this indicates that that value of  $x$  is in the range of  $X$ . Therefore,  $R_X = \{0, 1, 2, 3\}$ . The probability at  $x$  can be found by subtracting out the probabilities at values  $< x$  from  $F_X(x)$ . Therefore, the following equations give the probabilities we need:

$$\begin{cases} P(0) = F_X(0) \\ P(1) = F_X(1) - P(0) \\ P(2) = F_X(2) - P(1) - P(0) \\ P(3) = F_X(3) - P(2) - P(1) - P(0), \end{cases}$$

and when plugging in the values for  $F_X(x)$ , this leads to:

$$P_X(x) = \begin{cases} \frac{1}{6} & \text{for } x = 0 \\ \frac{1}{3} & \text{for } x = 1 \\ \frac{1}{4} & \text{for } x = 2 \\ \frac{1}{4} & \text{for } x = 3. \end{cases}$$

As a sanity check, these probabilities do indeed add up to 1.

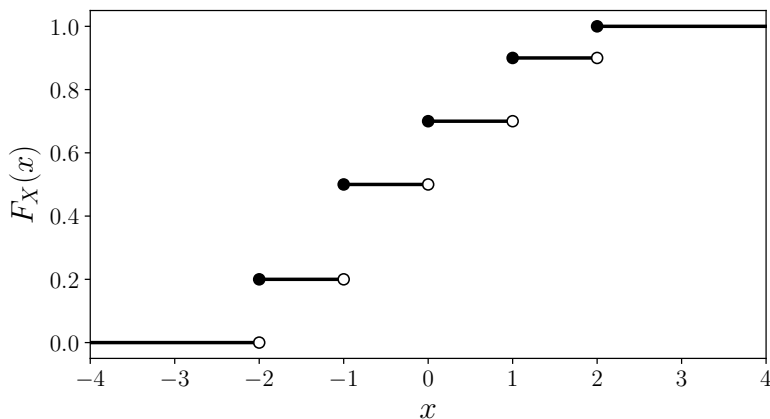


Figure 3.1: The associated CDF for the PMF of problem 12.

**Problem 14.**

(a)

$$E[X] = 1 \cdot 0.5 + 2 \cdot 0.3 + 3 \cdot 0.2 = 1.7$$

(b)

$$E[X^2] = 1 \cdot 0.5 + 4 \cdot 0.3 + 9 \cdot 0.2 = 3.5$$

 $\Rightarrow$ 

$$\text{Var}[X] = E[X^2] - E[X]^2 = 3.5 - 1.7^2 = 0.61$$

 $\Rightarrow$ 

$$\text{SD}[X] = \sqrt{\text{Var}[X]} \approx 0.78$$

(c) Using LOTUS:

$$\begin{aligned} E[Y] &= \sum_{x \in R_X} \frac{2}{x} P_X(x) \\ &= \frac{2}{1} \cdot 0.5 + \frac{2}{2} \cdot 0.3 + \frac{2}{3} \cdot 0.2 \approx 1.43. \end{aligned}$$

**Problem 15.** The range of  $X$  is  $\{1, 2, 3, \dots\}$ . For  $x \geq 5$ , these values get mapped to  $0, 1, 2, \dots$ . The values  $x = 1, 2, 3, 4$  get mapped to  $4, 3, 2, 1$ , and thus  $R_Y = \{0, 1, 2, \dots\}$ . To solve for the the corresponding PMF, note that  $P(X = k) = (1/3)(2/3)^{k-1}$ , and that  $P_Y(y = k) = P(Y = k) = P(|X - 5| = k)$ . We therefore have:

$$\begin{aligned}
P_Y(y=0) &= P(X=5) = \frac{1}{3} \left(\frac{2}{3}\right)^4 \\
P_Y(y=1) &= P(X=4 \text{ or } X=6) = \frac{1}{3} \left(\frac{2}{3}\right)^3 + \frac{1}{3} \left(\frac{2}{3}\right)^5 \\
P_Y(y=2) &= P(X=3 \text{ or } X=7) = \frac{1}{3} \left(\frac{2}{3}\right)^2 + \frac{1}{3} \left(\frac{2}{3}\right)^6 \\
P_Y(y=3) &= P(X=2 \text{ or } X=8) = \frac{1}{3} \left(\frac{2}{3}\right)^1 + \frac{1}{3} \left(\frac{2}{3}\right)^7 \\
P_Y(y=4) &= P(X=1 \text{ or } X=9) = \frac{1}{3} \left(\frac{2}{3}\right)^0 + \frac{1}{3} \left(\frac{2}{3}\right)^8 \\
P_Y(y=5) &= P(X=10) = \frac{1}{3} \left(\frac{2}{3}\right)^9 \\
P_Y(y=6) &= P(X=11) = \frac{1}{3} \left(\frac{2}{3}\right)^{10} \\
&\vdots
\end{aligned}$$

One can easily check that this distribution is normalized by summing all terms on the RHS of the equations:  $(1/3) \sum_{k=0}^{\infty} (2/3)^k = 1$ , where I have summed the geometric series.

**Problem 16.** I first note that the range of  $Y$  is  $\{0, 1, 2, 3, 4, 5\}$ , so that its PMF is

$$\begin{aligned}
P_Y(y=0) &= P(X=-10 \text{ or } X=-9 \text{ or } \dots \text{ or } X=0) = \frac{11}{21} \\
P_Y(y=1) &= P(X=1) = \frac{1}{21} \\
P_Y(y=2) &= P(X=2) = \frac{1}{21} \\
P_Y(y=3) &= P(X=3) = \frac{1}{21} \\
P_Y(y=4) &= P(X=4) = \frac{1}{21} \\
P_Y(y=5) &= P(X=5 \text{ or } X=6 \text{ or } \dots \text{ or } X=10) = \frac{6}{21},
\end{aligned}$$

which indeed sums to 1.

**Problem 17.** Since  $E[X]$  was found to be  $1/p$  for the geometric distribution from Example 3.12 in the book, if we can solve for  $E[X^2]$  then we can compute the variance with  $Var[X] = E[X^2] - E[X]^2$ . To do this, we will need a few formulas involving the geometric series. I claim that:

$$\begin{aligned}
\sum_{k=0}^{\infty} x^k &= \frac{1}{1-x} \quad |x| < 1, \\
\sum_{k=0}^{\infty} kx^{k-1} &= \frac{1}{(1-x)^2} \quad |x| < 1,
\end{aligned}$$

and

$$\sum_{k=0}^{\infty} k^2 x^{k-1} = \frac{1+x}{(1-x)^3} \quad |x| < 1.$$

The first formula is simply the sum of a geometric series, the second was already proved in the book in Example 3.12. I now prove the third formula.

*Proof.* We can take derivatives of the LHS and RHS of the second equation above to prove the third. Differentiating the LHS results in:

$$\begin{aligned} \frac{d}{dx} \sum_{k=0}^{\infty} kx^{k-1} &= \sum_{k=0}^{\infty} k(k-1)x^{k-2} \\ &= \sum_{k=1}^{\infty} k^2 x^{k-2} - \sum_{k=1}^{\infty} kx^{k-2} \\ &= \sum_{j=0}^{\infty} (j+1)^2 x^{j-1} - \sum_{k=0}^{\infty} (j+1)x^{j-1} \\ &= \sum_{j=0}^{\infty} j^2 x^{j-1} + 2 \sum_{j=0}^{\infty} jx^{j-1} + \sum_{j=0}^{\infty} x^{j-1} - \sum_{j=0}^{\infty} jx^{j-1} - \sum_{j=0}^{\infty} x^{j-1} \\ &= \sum_{j=0}^{\infty} j^2 x^{j-1} + \sum_{j=0}^{\infty} jx^{j-1} \\ &= \sum_{j=0}^{\infty} j^2 x^{j-1} + \frac{1}{(1-x)^2}, \end{aligned}$$

where I have made the substitution  $j = k - 1$ . Differentiating the RHS results in:

$$\frac{d}{dx} \frac{1}{(1-x)^2} = \frac{2}{(1-x)^3},$$

and putting the two together completes the proof. ■

I may now solve for  $E[X^2]$

$$\begin{aligned} E[X^2] &= \sum_{k=1}^{\infty} x^2 p(1-p)^{k-1} \\ &= p \sum_{k=0}^{\infty} x^2 (1-p)^{k-1} \\ &= \frac{p[1 + (1-p)]}{[1 - (1-p)]^3} \\ &= \frac{2-p}{p^2}, \end{aligned}$$

so that the variance is:

$$\text{Var}[X] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

**Problem 18.** In Problem 5 from 3.1.6 of the book, we showed that if  $X_1, X_2, \dots, X_m \sim \text{Geom}(p) = \text{Pascal}(1, p)$  (iid), then  $X = X_1 + X_2 + \dots + X_m \sim \text{Pascal}(m, p)$ . Therefore  $\text{Var}[X] = \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_m] = m(1-p)/(p^2)$ , by linearity in variance of independent random variables.

**Problem 19.** I use LOTUS repeatedly in this problem and linearity of expectation.

$$\begin{aligned} E[X] &= E\left[-\frac{Y}{2} + \frac{3}{2}\right] \\ &= -\frac{1}{2}E[Y] + \frac{3}{2} \\ &= 1 \end{aligned}$$

$$\begin{aligned} Var[X] &= E[X^2] - E[X]^2 \\ &= E\left[\frac{Y^2}{4} - \frac{3Y}{2} + \frac{9}{4}\right] - 1 \\ &= \frac{1}{4}E[Y^2] - \frac{3}{2}E[Y] + \frac{9}{4} - 1 \\ &= \frac{9}{4} - \frac{3}{2} + \frac{9}{4} - 1 \\ &= 2 \end{aligned}$$

**Problem 20.**

- (a) The range of  $X$  is  $\{1, 2, 3, 4, 5, 6\}$  and the probability for any of these values,  $x$ , is simply  $N_x/1000$ , where  $N_x$  is the number of households with  $x$  people. Therefore:

$$P_X(x) = \begin{cases} 0.1 & \text{for } x = 1 \\ 0.2 & \text{for } x = 2 \\ 0.3 & \text{for } x = 3 \\ 0.2 & \text{for } x = 4 \\ 0.1 & \text{for } x = 5 \\ 0.1 & \text{for } x = 6. \end{cases}$$

The expected value of  $X$  is:  $E[X] = 1 \cdot 0.1 + 2 \cdot 0.2 + 3 \cdot 0.3 + 4 \cdot 0.2 + 5 \cdot 0.1 + 6 \cdot 0.1 = 3.3$ .

- (b) The probability of picking a person from a household with  $k$  people is equal to the total number of people in households with  $k$  people divided by the total number of people in the town. In other words,  $P(Y = k) = (k \cdot N_k)/3300$ , so that:

$$P_Y(y) = \begin{cases} \frac{1}{33} & \text{for } y = 1 \\ \frac{4}{33} & \text{for } y = 2 \\ \frac{9}{33} & \text{for } y = 3 \\ \frac{8}{33} & \text{for } y = 4 \\ \frac{5}{33} & \text{for } y = 5 \\ \frac{6}{33} & \text{for } y = 6, \end{cases}$$

and  $E[Y] = 1 \cdot (1/33) + 2 \cdot (4/33) + 3 \cdot (9/33) + 4 \cdot (8/33) + 5 \cdot (5/33) + 6 \cdot (6/33) = 43/11$ .

**Problem 21.**

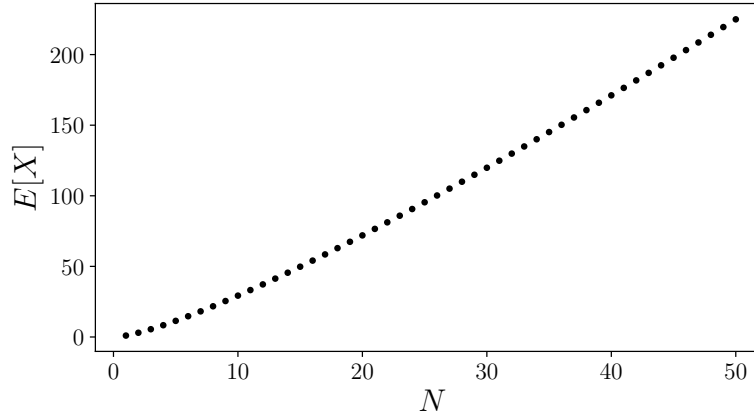


Figure 3.2: The expected number of tries to observe all unique coupons at least once.

- (a) It takes 1 try to observe the first unique coupon. Let this first coupon be called type  $C_1$ . Let the random variable,  $X_1$ , be the number of times it takes to observe a coupon different than type  $C_1$ . Call this type  $C_2$ . Let the random variable,  $X_2$ , be the number of times it takes to observe a coupon different than type  $C_1$  and  $C_2$ . Call this type  $C_3$ . Let us proceed in this fashion until we observe  $N - 1$  unique coupons. Finally, let the random variable  $X_{N-1}$  be the number of times it takes to observe a coupon different than type  $C_1, C_2, \dots, C_{N-1}$ , and call this coupon type  $C_N$ . Therefore, the total number of times it takes to observe all unique coupons at least once is  $X = 1 + X_1 + X_2 + \dots + X_{N-1}$ .

For each  $X_i$ , if we consider choosing  $C_1, C_2, \dots$  or,  $C_{i-1}$  as a failure and  $C_i$  as a success, we see that this is nothing more than a geometric random variable with probability  $(N - i)/N$  of success (since there are  $N - i$  un-observed coupons left). Therefore,  $X_1, X_2, \dots, X_{N-1} \sim \text{Geom}(\frac{N-i}{N})$ . Further let  $X_0 \sim \text{Geom}(\frac{N-0}{N})$ , and note that the probability of observing  $X_0 = 1$  for this distribution is unity since we are sure to have a success on the first trial. Thus, if we desire, we can replace 1 in  $X = 1 + X_1 + X_2 + \dots + X_{N-1}$  with the “random variable”  $X_0$ .

- (b) The expected number of tries it takes to observe all unique coupons at least once is:

$$\begin{aligned}
 E[X] &= E[1 + X_1 + X_2 + \dots + X_{N-1}] \\
 &= 1 + E[X_1] + E[X_2] + \dots + E[X_{N-1}] \\
 &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \dots + \frac{N}{N-(N-1)} \\
 &= N \sum_{i=0}^{N-1} \frac{1}{N-i}.
 \end{aligned}$$

The summation can be written in terms of a special function (called the digamma function), but I believe it is more illustrative to plot the actual function itself. In Fig. 3.2, I show  $E[X]$  for  $N = 1$  to 50 which I calculated numerically with the summation formula I derived above.

### Problem 22.

- (a) Let  $X'$  be the number of tosses until the game ends. We recognize that  $X'$  is distributed as a geometric random variable with  $p = q = 1/2$  since the coin is fair. The range of  $X'$  is  $R_{X'} = \{1, 2, 3, 4, \dots\}$ . Let the random variable  $X$  denote the amount of money won from



the game which has range  $R_X = \{1, 2, 4, 8, \dots\}$ . The function  $f : R_{X'} \rightarrow R_X$  is given by the bijective mapping  $2^{X'-1}$ . Thus, the PMF of  $X$  is given by  $P(X = x) = P(X' = x')$ , where  $x'$  is the pre-image of  $x$  under  $f$ . That is, the PMF of  $X$  is given by:  $P(X = 1) = P(X' = 1) = p$ ,  $P(X = 2) = P(X' = 2) = p^2$ ,  $P(X = 4) = P(X' = 3) = p^3$ ,  $P(X = 8) = P(X' = 4) = p^4, \dots$ . Thus the expected value of  $X$  is given by the following summation, which we see diverges:

$$\begin{aligned} E[X] &= \sum_{x \in R_X} xP(X = x) = \sum_{k=1}^{\infty} 2^{k-1}P(X' = k) \\ &= \sum_{k=1}^{\infty} 2^{k-1} \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^{k-1} \\ &= \infty. \end{aligned}$$

Thus, only considering your expected winnings (and ignoring issues like the variance of your winnings and your particular risk tolerance) you would be willing to pay any amount of money to play this game.

- (b) By noting that  $\dots, 2^{6-1}, 2^{7-1}, 2^{8-1} \dots = \dots, 32, 64, 128, \dots$ , one sees that when  $X' = 8$ ,  $X = 2^7 = 128$  which is the first time that  $X$  takes on a value greater than 65. Therefore, the probability we desire is:

$$\begin{aligned} P(X > 65) &= \sum_{k=8}^{\infty} P(X' = k) \\ &= 1 - \sum_{k=1}^7 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^{k-1} \\ &= 1 - \left[ \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + \dots + \left(\frac{1}{2}\right)^7 \right] \\ &= \frac{1}{128}. \end{aligned}$$

- (c) This problem is very similar to part a, except that the summation is truncated when  $x$  takes on the value  $2^{30}$ , which occurs when  $k = 31$ , thereafter, the payout remains  $2^{30}$ . Therefore, the expected value of  $Y$  is:

$$\begin{aligned} E[Y] &= \sum_{k=1}^{31} 2^{k-1}P(X' = k) + \sum_{k=32}^{\infty} 2^{30}P(X' = k) \\ &= \sum_{k=1}^{31} \frac{1}{2} + 2^{30} \sum_{k=32}^{\infty} \left(\frac{1}{2}\right)^k \\ &= \frac{31}{2} + 2^{30} 2^{-32} \sum_{k'=0}^{\infty} \left(\frac{1}{2}\right)^{k'} \\ &= \frac{31}{2} + 2^{30} 2^{-32} \frac{1}{1 - 1/2} \\ &= 16. \end{aligned}$$

We therefore see that in part a, the majority of the summation that contributes to the expectation value of  $X$  occurs much later in the series. This is called a “paradox” since, in

the first part the expectation value was infinite, but in the second part, even though  $2^{30}$  is a very large number, the expected winnings is much lower than what one would have guessed.

**Problem 23.** The goal is to find:

$$\alpha^* = \arg \min_{\alpha \in \mathbb{R}} f(\alpha),$$

where

$$\begin{aligned} f(\alpha) &= E[(X - \alpha)^2] \\ &= E[X^2 - 2\alpha X + \alpha^2] \\ &= E[X^2] - 2\alpha\mu + \alpha^2. \end{aligned}$$

Therefore:

$$\alpha^* = \arg \min_{\alpha \in \mathbb{R}} \{E[X^2] - 2\alpha\mu + \alpha^2\} = \arg \min_{\alpha \in \mathbb{R}} \{-2\alpha\mu + \alpha^2\},$$

which we can be found by setting the derivative of this equation,

$$\frac{d}{d\alpha}(-2\alpha\mu + \alpha^2) = -2\mu + 2\alpha,$$

equal to zero, and solving for  $\alpha^*$ . This results in  $\alpha^* = \mu$ .

**Problem 24.** If you choose to roll the die for a second time, your expected winnings is  $E[Y] = 3.5$ . Therefore, if you roll less than 3.5 on the first roll (i.e., 1, 2 or 3) you should roll again because you expect to do better on the second roll. However, if you roll a 4, 5 or 6, you will expect to do worse on the second roll, so you should not roll again.

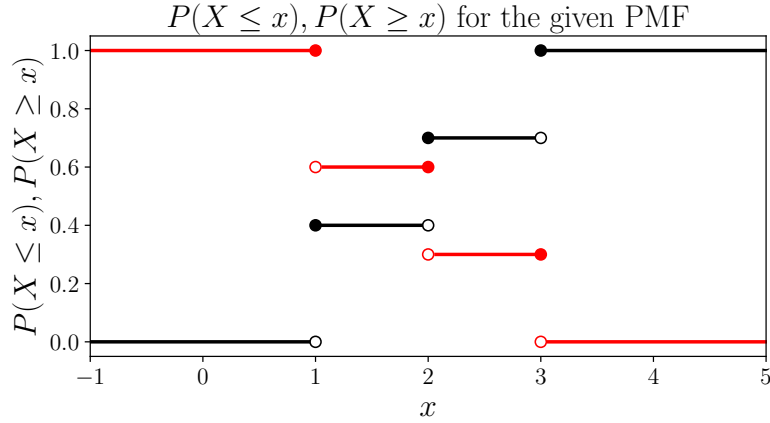
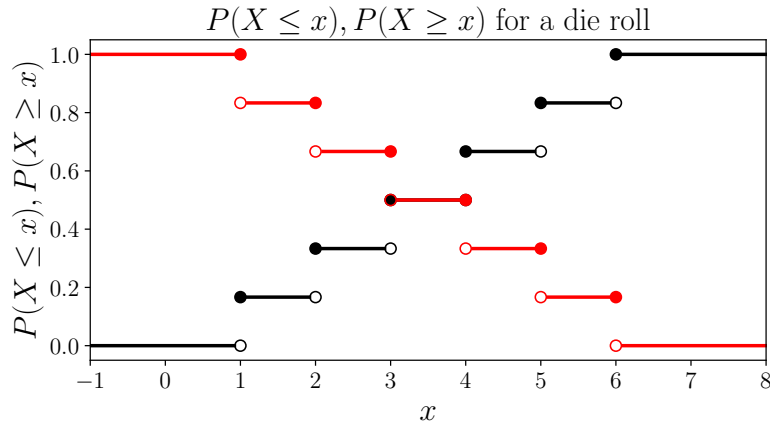
Given this strategy, your expected winnings is:

$$\begin{aligned} E[W] &= E[X\mathbb{1}\{X > 3\}] + E[Y\mathbb{1}\{X \leq 3\}] \\ &= E[X\mathbb{1}\{X > 3\}] + E[Y]E[\mathbb{1}\{X \leq 3\}] \\ &= \frac{1}{6} \sum_{x=1}^6 x\mathbb{1}\{x > 3\} + \left(\frac{7}{2}\right) \left(\frac{1}{6}\right) \sum_{x=1}^6 \mathbb{1}\{X \geq 3\} \\ &= \frac{1}{6}(4 + 5 + 6) + \left(\frac{7}{2}\right) \left(\frac{1}{6}\right)(1 + 1 + 1) \\ &= \frac{17}{4} \\ &= 4.25, \end{aligned}$$

where, in the second equality,  $E[Y\mathbb{1}\{X \leq 3\}] = E[Y]E[\mathbb{1}\{X \leq 3\}]$  since  $X$  and  $Y$  are independent (given the set strategy).

**Problem 25.**

- (a) In Fig. 3.3 I have plotted both  $P(X \geq x)$  and  $P(X \leq x)$  for this PMF. It is clear from this figure that in the range  $[2, \infty)$ ,  $P(X \leq x) \geq 1/2$ , and that in the range  $(-\infty, 2]$   $P(X \geq x) \geq 1/2$ . The only value that these ranges share in common is 2, and this is therefore the median for this PMF.

Figure 3.3:  $P(X \geq x)$  (red) and  $P(X \leq x)$  (black) for Problem 25a.Figure 3.4:  $P(X \geq x)$  (red) and  $P(X \leq x)$  (black) for Problem 25b.

- (b) In Fig. 3.4 I have plotted both  $P(X \geq x)$  and  $P(X \leq x)$  for a die roll. It is clear from this figure that in the range  $[3, \infty)$ ,  $P(X \leq x) \geq 1/2$ , and that in the range  $(-\infty, 4]$   $P(X \geq x) \geq 1/2$ . The (not unique) medians for this distribution are the intersection of these 2 sets, which is the interval  $[3, 4]$ .

- (c) We can compute  $P(X \leq x)$  with the geometric distribution explicitly with:

$$P(X \leq x) = p \sum_{k=1}^{k_x^u} q^{k-1},$$

where  $q = 1 - p$  and  $k_x^u$  is the appropriate upper integer bound which depends on the (not necessarily integer) value of  $x$ . By considering the staircase shape of  $P(X \leq x)$ , one can realize that for any  $x$ ,  $P(X \leq x) = P(X \leq \lfloor x \rfloor)$ , which holds up until  $\lceil x \rceil$  (where  $\lfloor \cdot \rfloor$  and  $\lceil \cdot \rceil$  are defined as rounding down and up to the nearest integer respectively). Therefore, if we want to find the lowest value of  $x$ ,  $x'$  for  $P(X \leq x')$  still equals  $P(X \leq x)$ , this occurs at the integer value  $x' = \lfloor x \rfloor$ .  $\lfloor x \rfloor$  is therefore the appropriate value to use for  $k_x^u$ , and we can

explicitly compute a formula for  $P(X \leq x)$ :

$$\begin{aligned}
P(X \leq x) &= P(X \leq \lfloor x \rfloor) \\
&= pq^{-1} \sum_{k=1}^{\lfloor x \rfloor} q^k \\
&= pq^{-1} \left[ -q^0 + \sum_{k=0}^{\lfloor x \rfloor} q^k \right] \\
&= pq^{-1} \left[ -1 + \sum_{k=0}^{\infty} q^k - \sum_{k=\lfloor m \rfloor+1}^{\infty} q^k \right] \\
&= pq^{-1} \left[ -1 + \sum_{k=0}^{\infty} q^k - q^{\lfloor x \rfloor+1} \sum_{k'=0}^{\infty} q^{k'} \right] \\
&= pq^{-1} \left[ -1 + \frac{1}{1-q} - \frac{q^{\lfloor x \rfloor+1}}{1-q} \right] \\
&= 1 - q^{\lfloor x \rfloor}.
\end{aligned}$$

Any value  $m$ , for which  $P(X \leq m) \geq 1/2$  is a potential candidate for the median (but of course we still have to consider the values of  $x$  for which  $P(X \geq x) \geq 1/2$ ) and the lowest value for which this occurs, call it  $\lfloor m_L \rfloor$ , can now be found by setting  $P(X \leq \lfloor m_L \rfloor) = 1/2$ , resulting in:

$$\lfloor m_L \rfloor = \frac{1}{\log_2 1/q}.$$

Similarly:

$$\begin{aligned}
P(X \geq x) &= P(X \geq \lceil x \rceil) \\
&= pq^{-1} \sum_{k=\lceil x \rceil}^{\infty} q^k \\
&= pq^{\lceil x \rceil-1} \sum_{k'=0}^{\infty} q^{k'} \\
&= \frac{pq^{\lceil x \rceil-1}}{1-q} \\
&= q^{\lceil x \rceil-1}.
\end{aligned}$$

Thus, the highest value for which  $P(X \geq x) \geq 1/2$ , call it  $\lceil m_U \rceil$  is found when this equation equals  $1/2$ , resulting in:

$$\lceil m_U \rceil = \frac{1}{\log_2 1/q} + 1.$$

Therefore, for the geometric distribution,  $P(X \leq m) \geq 1/2$  and  $P(X \geq m) \geq 1/2$  for  $x \in [\lfloor m_L \rfloor, \lceil m_U \rceil]$ . This interval thus gives the (not unique) medians for the geometric distribution.

## Chapter 4

# Continuous and Mixed Random Variables

**Problem 1.**

- (a) We recognize that this is a uniform random variable, so its CDF is:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 2 \\ \frac{x-2}{4} & \text{for } 2 \leq x \leq 6 \\ 1 & \text{for } x > 6. \end{cases}$$

- (b) For a uniform random variable, the expectation value is at the midpoint:  $E[X] = 2 + [(6 - 2)]/2 = 4$ .

**Problem 2.**

- (a) By normalization,  $1 = c \int_0^\infty \exp(-4x)dx$ , which leads to  $c = 4$ .

- (b)

$$F_X(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 4 \int_0^x e^{-4x'} dx' & \text{for } x > 0 \end{cases} = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-4x} & \text{for } x > 0. \end{cases}$$

- (c)

$$P(2 < X < 5) = 4 \int_2^5 e^{-4x} dx = e^{-8} - e^{-20}.$$

- (d) To find the expectation value, I use integration by parts:

$$\begin{aligned} E[X] &= 4 \int_0^\infty x e^{-4x} dx \\ &= - (x e^{-4x})_0^\infty + \int_0^\infty e^{-4x} dx \\ &= \frac{1}{4}, \end{aligned}$$

where the limits in the first term were evaluated using L'hospital's rule.

**Problem 3.**

- (a) Using LOTUS:

$$\begin{aligned} E \left[ X^n \left( X^2 + \frac{2}{3} \right) \right] &= \int_0^1 x^n \left( x^2 + \frac{2}{3} \right) dx \\ &= \frac{1}{3+n} x^{3+n} \Big|_0^1 + \left( \frac{2}{3} \right) \frac{1}{n+1} x^{n+1} \Big|_0^1 \\ &= \frac{1}{3+n} + \left( \frac{2}{3} \right) \frac{1}{n+1} \quad \text{for } n = 1, 2, 3, \dots \end{aligned}$$

- (b) We have already found  $E[X]$  and  $E[X^2]$  in the first part and thus:

$$\begin{aligned} \text{Var}[X] &= E[X^2] - E[X]^2 \\ &= \frac{1}{3+2} + \left( \frac{2}{3} \right) \frac{1}{2+1} - \left[ \frac{1}{3+1} + \left( \frac{2}{3} \right) \frac{1}{1+1} \right]^2 \\ &= \frac{59}{720}. \end{aligned}$$

**Problem 4.**

- (a) For this problem, we have  $R_X = [0, 1]$  and  $R_Y = [1/e, 1]$ . Thus in range of  $x = 0$  to  $1$ , the CDF of  $Y$  is given by:

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= P(e^{-X} \leq y) \\
 &= P(X \geq -\ln y) \\
 &= 1 - F_X(-\ln y) \\
 &= 1 + \ln y,
 \end{aligned}$$

where I used the fact that for  $x \in [0, 1]$  for a uniform 0, 1 distribution,  $F_X(x) = x$ . Therefore:

$$F_Y(y) = \begin{cases} 0 & \text{for } y < \frac{1}{e} \\ 1 + \ln(y) & \text{for } \frac{1}{e} \leq y \leq 1 \\ 1 & \text{for } y > 1. \end{cases}$$

- (b)

$$f_Y(y) = \frac{dF_Y}{dy} = \begin{cases} 0 & \text{for } y < \frac{1}{e} \\ \frac{1}{y} & \text{for } \frac{1}{e} \leq x \leq 1 \\ 0 & \text{for } x > 1 \end{cases}$$

- (c)

$$E[Y] = \int_{1/e}^1 y \frac{1}{y} dy = 1 - \frac{1}{e}$$

**Problem 5.**

- (a) The range of  $X$  and  $Y$  are  $R_X = (0, 2]$  and  $R_Y = (0, 4]$ , so that for  $y \in R_Y$ , we have

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= P(X^2 \leq y) \\
 &= P(0 < X < \sqrt{y}) \\
 &= \int_0^{\sqrt{y}} \frac{5}{32} x^4 dx \\
 &= \frac{1}{32} y^{5/2},
 \end{aligned}$$

and therefore:

$$F_Y(y) = \begin{cases} 0 & \text{for } y \leq 0 \\ \frac{1}{32} y^{5/2} & \text{for } 0 < y \leq 4 \\ 1 & \text{for } y > 4. \end{cases}$$

- (b)

$$f_Y(y) = \frac{dF_Y}{dy} = \begin{cases} 0 & \text{for } y \leq 0 \\ \frac{5}{64} y^{3/2} & \text{for } 0 < y \leq 4 \\ 0 & \text{for } y > 4 \end{cases}$$

(c)

$$E[Y] = \frac{5}{64} \int_0^4 y \cdot y^{3/2} \approx 2.9$$

**Problem 6.** We can convert the PDF for  $X$  to the PDF for  $Y$  using the method of transformations:

$$f_Y(y) = f_X\left(\frac{y}{\alpha}\right) \left| \frac{d(y/\alpha)}{dy} \right| = \begin{cases} \frac{\lambda}{\alpha} e^{-\frac{\lambda}{\alpha}y} & \text{for } y > 0 \\ 0 & \text{otherwise} \end{cases},$$

and since both  $\alpha, \lambda > 0$ ,  $Y \sim \text{Exp}(\lambda/\alpha)$ .

**Problem 7.**

(a) We can prove this relation using integration by parts:

$$\begin{aligned} E[X^n] &= \int_0^\infty x^n \lambda e^{-\lambda x} dx \\ &= -x^n e^{-\lambda x} \Big|_0^\infty + \frac{n}{\lambda} \int_0^\infty x^{n-1} \lambda e^{-\lambda x} dx \\ &= \frac{n}{\lambda} E[X^{n-1}], \end{aligned}$$

where the first term evaluated to zero by repeated application of L'Hopital's rule.

(b) We can use several properties of the Gamma function to prove this relation:

$$\begin{aligned} E[X^n] &= \int_0^\infty x^n \lambda e^{-\lambda x} dx \\ &= \frac{\lambda}{\lambda^{n+1}} \Gamma(n+1) \\ &= \frac{n!}{\lambda^n}, \end{aligned}$$

where in the second equality I used the second property of the Gamma function given in the book, and in the third equality I used the fourth property of the Gamma function given in the book.

**Problem 8.**

(a)

$$P(X > 0) = 1 - \Phi\left(\frac{0-3}{3}\right) \approx 0.84$$

(b)

$$P(-3 < X < 8) = \Phi\left(\frac{8-3}{3}\right) - \Phi\left(\frac{-3-3}{3}\right) \approx 0.93$$



(c)

$$\begin{aligned}
P(X > 5|X > 3) &= \frac{P(X > 5, X > 3)}{P(X > 3)} \\
&= \frac{P(X > 5)}{P(X > 3)} \\
&= \frac{1 - \Phi\left(\frac{5-3}{3}\right)}{1 - \Phi\left(\frac{3-3}{3}\right)} \\
&\approx 0.50
\end{aligned}$$

**Problem 9.** By Theorem 4.3 in the book, if  $X \sim \mathcal{N}(3, 9)$ , and  $Y = 5 - X$ , then  $Y \sim \mathcal{N}(-3+5, 9) = \mathcal{N}(2, 9)$ .

(a)

$$P(X > 2) = 1 - \Phi\left(\frac{2-3}{3}\right) \approx 0.63$$

(b)

$$P(-1 < Y < 3) = \Phi\left(\frac{3-2}{3}\right) - \Phi\left(\frac{-1-2}{3}\right) \approx 0.47$$

(c)

$$\begin{aligned}
P(X > 4|Y < 2) &= P(X > 4|5 - X < 2) \\
&= \frac{P(X > 4, X > 3)}{P(X > 3)} \\
&= \frac{P(X > 4)}{P(X > 3)} \\
&= \frac{1 - \Phi\left(\frac{4-3}{3}\right)}{1 - \Phi\left(\frac{3-3}{3}\right)} \\
&\approx 0.74
\end{aligned}$$

**Problem 10.** I first note that  $R_X = \mathbb{R}$ , and  $R_Y = [0, \infty)$ . The range of  $X$  can be partitioned into 2 regions,  $X \leq 0$  and  $X > 0$  which are strictly decreasing, and increasing respectively, where the corresponding inverse transformation back to  $X$  for both of these regions is:

$$X = \begin{cases} -Y^2 & \text{for } X \leq 0 \\ Y^2 & \text{for } X > 0. \end{cases}$$

Therefore:

$$\begin{aligned}
f_Y(y) &= f_X(y^2) \left| \frac{d(y^2)}{dy} \right| + f_X(-y^2) \left| \frac{d(-y^2)}{dy} \right| \\
&= \frac{4}{\sqrt{2\pi}} y e^{-\frac{y^4}{2}} \quad \text{for } y \geq 0,
\end{aligned}$$

which, as a sanity check, I made sure analytically integrates to unity over the range 0 to infinity.

**Problem 11.**

(a)

$$P(X > 2) = 2 \int_2^{\infty} e^{-2x} dx = e^{-4}$$

(b) I calculate  $E[Y]$  using LOTUS:

$$\begin{aligned} E[Y] &= \int_0^{\infty} (2 + 3x)2e^{-2x} dx \\ &= 2 \int_0^{\infty} 2e^{-2x} dx + 3 \int_0^{\infty} 2xe^{-2x} dx \\ &= 2 + 3E[X] \\ &= 2 + \frac{3}{2} \\ &= \frac{7}{2}, \end{aligned}$$

where I have used the fact that  $E[X]$  for  $Exp(\lambda)$  is  $1/\lambda$  (as computed in the book). To compute  $Var[Y]$ , I first must compute  $E[Y^2]$ , which I do using LOTUS:

$$\begin{aligned} E[Y^2] &= \int_0^{\infty} (2 + 3x)^2 2e^{-2x} dx \\ &= 4 \int_0^{\infty} 2e^{-2x} dx + 12 \int_0^{\infty} 2xe^{-2x} dx + 9 \int_0^{\infty} 2x^2 e^{-2x} dx \\ &= 4 + 12E[X] + 9E[X^2] \\ &= 4 + \frac{12}{2} + \frac{9 \cdot 2}{4} \\ &= \frac{29}{2}, \end{aligned}$$

where I have used the fact that  $E[X^2]$  for  $Exp(\lambda)$  is  $2/(\lambda^2)$  (as computed in the book). Finally, the variance is:

$$Var[Y] = E[Y^2] - E[Y]^2 = \frac{9}{4}.$$

(c)

$$\begin{aligned} P(X > 2 | Y < 11) &= P(X > 2 | 2 + 3X < 11) \\ &= P(X > 2 | X < 3) \\ &= \frac{P(2 < X < 3)}{P(X < 3)} \\ &= \frac{2 \int_2^3 e^{-2x} dx}{2 \int_0^3 e^{-2x} dx} \\ &= \frac{e^2 - 1}{e^6 - 1} \\ &\approx 1.6 \times 10^{-2} \end{aligned}$$

**Problem 12.** The equations defining the median for a continuous variable,  $P(X < m) = 1/2$  and  $P(X \geq m) = 1/2$ , are actually equivalent. That is,  $P(X < m) = 1/2 \Leftrightarrow P(X \geq m) = 1/2$  (which can easily be verified), so we can use whichever is convenient. Since we know the CDFs for the desired distributions, so using the condition that  $P(X < m) = 1/2$  will be most convenient.

(a) The CDF for the  $Unif(a, b)$  is:

$$P(X < x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b, \end{cases}$$

where I have ignored the equality in the argument of  $P$ , since  $P(X = x) = 0$  for a continuous random variable. Setting this equation equal to  $1/2$  and solving for  $m$ , I find:

$$m = \frac{b+a}{2},$$

which is the mean of the uniform distribution, which was expected since the uniform distribution is symmetric about its mean.

(b) The CDF for the  $Exp(\lambda)$  is:

$$P(Y < y) = \begin{cases} 0 & \text{for } y < 0 \\ 1 - e^{-\lambda y} & \text{for } y \geq 0, \end{cases}$$

where again, I have ignored the equality in the argument of  $P$ , since  $P(Y = y) = 0$ . Thus we see that  $1/2 = 1 - \exp(-\lambda m) \implies m = \ln 2 / \lambda$ .

(c) For the  $\mathcal{N}(\mu, \sigma^2)$ ,

$$\begin{aligned} P(W < m) &= P(W \leq m) \\ &= \Phi\left(\frac{m - \mu}{\sigma}\right) \\ &= \frac{1}{2}. \end{aligned}$$

Since the standard normal is symmetric about 0, this implies that  $\Phi(0) = 1/2$ , and therefore  $(m - \mu)/\sigma = 0$ , and thus  $m = \mu$ . Since we knew that a Gaussian is symmetric about its mean, this is what we expected.

**Problem 13.**

(a) See Fig. 4.1 for a plot of the CDF.  $X$  is a mixed random variable because there is a jump in the CDF at  $x = 1/4$  (indicating a probability “point mass” at  $1/4$  of  $P(X = 1/4) = 1/2$ ) and the CDF does not exhibit the staircase shape associated with only discrete random variables.

(b)

$$\begin{aligned} P\left(X \leq \frac{1}{3}\right) &= F_X\left(\frac{1}{3}\right) \\ &= \frac{1}{3} + \frac{1}{2} \\ &= \frac{5}{6} \end{aligned}$$

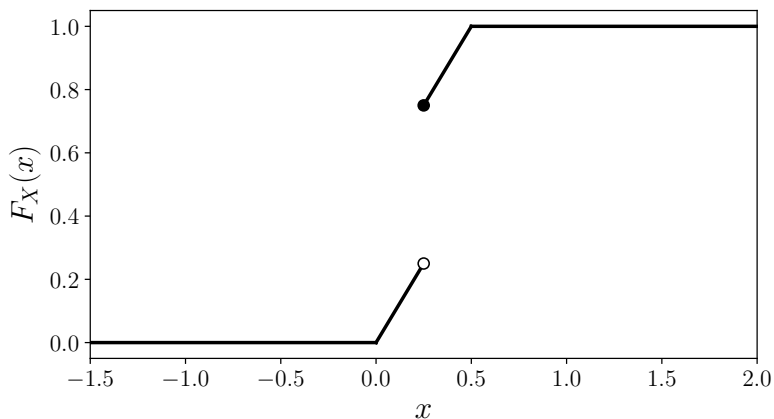


Figure 4.1: CDF for Problem 13

(c)

$$\begin{aligned}
 P\left(X \geq \frac{1}{4}\right) &= 1 - P\left(X < \frac{1}{4}\right) \\
 &= 1 - \left[P\left(X \leq \frac{1}{4}\right) - P\left(X = \frac{1}{4}\right)\right] \\
 &= 1 - \left[F_X\left(\frac{1}{4}\right) - P\left(X = \frac{1}{4}\right)\right] \\
 &= 1 - \left(\frac{1}{4} + \frac{1}{2} - \frac{1}{2}\right) \\
 &= \frac{3}{4}
 \end{aligned}$$

(d) The CDFs for both the discrete and continuous contributions can be written piece-wise as:

$$D(x) = \begin{cases} 0 & \text{for } x < \frac{1}{4} \\ \frac{1}{2} & \text{for } x \geq \frac{1}{4}, \end{cases}$$

and

$$C(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } \frac{1}{2} \leq x \leq \frac{1}{2} \\ \frac{1}{2} & \text{for } x > \frac{1}{2}. \end{cases}$$

These functions can be re-written using the unit step function:

$$D(x) = \frac{1}{2}u\left(x - \frac{1}{4}\right),$$

and

$$C(x) = xu(x) - xu\left(x - \frac{1}{2}\right) + \frac{1}{2}u\left(x - \frac{1}{2}\right),$$

where, in  $C(x)$ , I have started subtracting off the linear equation at  $x = 1/2$ , and adding a constant  $1/2$  at  $x = 1/2$  so as to keep the function flat at  $1/2$  after  $x = 1/2$ .

- (e) Since  $C(x)$  increases linearly from 0 to  $1/2$ , and has total probability mass  $1/2$ , we expect  $c(x)$  to be uniform with height 1 over the range 0 to  $1/2$ . Differentiating  $C(x)$

$$\begin{aligned}
 c(x) &= \frac{d}{dx}C(x) \\
 &= u(x) + x \frac{du(x)}{dx} - u\left(x - \frac{1}{2}\right) - x \frac{d}{dx}u\left(x - \frac{1}{2}\right) + \frac{1}{2} \frac{d}{dx}u\left(x - \frac{1}{2}\right) \\
 &= u(x) + x\delta(x) - u\left(x - \frac{1}{2}\right) - x\delta\left(x - \frac{1}{2}\right) + \frac{1}{2}\delta\left(x - \frac{1}{2}\right) \\
 &= u(x) - u\left(x - \frac{1}{2}\right),
 \end{aligned}$$

which is exactly what we had anticipated. Here I used the fact that when  $x \neq 1/2$ ,  $x\delta(x-1/2)$  and  $(-1/2)\delta(x-1/2)$  equal 0, while for  $x = 1/2$ ,  $(-1/2)\delta(x-1/2) = (-1/2)\delta(0)$  and  $x\delta(x-1/2) = (1/2)\delta(0)$ . Also, in either case,  $x\delta(x) = 0$ .

(f)

$$\begin{aligned}
 E[X] &= \int_{-\infty}^{\infty} xc(x)dx + \sum_k x_k a_k \\
 &= \int_0^{1/2} xdx + \frac{1}{4}P\left(X = \frac{1}{4}\right) \\
 &= \frac{1}{2}x^2 \Big|_0^{1/2} + \frac{1}{4} \cdot \frac{1}{2} \\
 &= \frac{1}{4}
 \end{aligned}$$

**Problem 14.**

- (a) The generalized PDF of  $X$  is:

$$\begin{aligned}
 f_X(x) &= \frac{d}{dx}F_X(x) \\
 &= \frac{d}{dx}D(x) + \frac{d}{dx}C(x) \\
 &= \frac{1}{2}\delta\left(x - \frac{1}{4}\right) + c(x) \\
 &= \frac{1}{2}\delta\left(x - \frac{1}{4}\right) + \left[u(x) - u\left(x - \frac{1}{2}\right)\right],
 \end{aligned}$$

where  $c(x)$  was found in the previous problem.

(b)

$$\begin{aligned}
 E[X] &= \frac{1}{2} \int_{-\infty}^{\infty} x\delta\left(x - \frac{1}{4}\right) dx + \int_{-\infty}^{\infty} xu(x)dx - \int_{-\infty}^{\infty} xu\left(x - \frac{1}{2}\right) dx \\
 &= \frac{1}{2} \int_{-\infty}^{\infty} x\delta\left(x - \frac{1}{4}\right) dx + \int_0^{1/2} xdx \\
 &= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{8} \\
 &= \frac{1}{4}
 \end{aligned}$$

(c)

$$\begin{aligned}
E[X^2] &= \frac{1}{2} \int_{-\infty}^{\infty} x^2 \delta\left(x - \frac{1}{4}\right) dx + \int_0^{1/2} x^2 dx \\
&= \frac{1}{2} \left(\frac{1}{4}\right)^2 + \frac{1}{3} x^3 \Big|_0^{1/2} \\
&= \frac{7}{96}
\end{aligned}$$

 $\Rightarrow$ 

$$\begin{aligned}
Var[X] &= E[X^2] - E[X]^2 \\
&= \frac{7}{96} - \left(\frac{1}{4}\right)^2 \\
&= \frac{1}{96}
\end{aligned}$$

**Problem 15.**

(a) From the form of the given generalized PDF, it is clear that there are 2 probability point masses at  $x = 1$  and  $x = -2$  (with  $P(X = 1) = 1/6$  and  $P(X = -2) = 1/3$ ), as well as a continuous random variable contribution from a Gaussian PDF. Since the continuous PDF contributes 0 probability at specific points,  $P(X = 1) = 1/6$  and  $P(X = -2) = 1/3$ .

(b)

$$\begin{aligned}
P(X \geq 1) &= \int_1^{\infty} \left[ \frac{1}{3} \delta(x+2) + \frac{1}{6} \delta(x-1) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right] dx \\
&= \frac{1}{6} + \frac{1}{2} \int_1^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{6} + \frac{1}{2} [1 - \Phi(1)] \\
&\approx 0.25
\end{aligned}$$

(c)

$$\begin{aligned}
P(X = 1 | X \geq 1) &= \frac{P(X = 1, X \geq 1)}{P(X \geq 1)} \\
&= \frac{P(X = 1)}{P(X \geq 1)} \\
&= \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{2} [1 - \Phi(1)]} \\
&\approx 0.68
\end{aligned}$$

(d) We can calculate  $E[X]$  by explicitly integrating over the generalized PDF:

$$\begin{aligned}
 E[X] &= \int_{-\infty}^{\infty} x \left[ \frac{1}{3} \delta(x+2) + \frac{1}{6} \delta(x-1) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right] dx \\
 &= \frac{1}{3}(-2) + \frac{1}{6}(1) + \frac{1}{2} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
 &= \frac{-2}{3} + \frac{1}{6} + \frac{1}{2} \cdot 0 \\
 &= -\frac{1}{2},
 \end{aligned}$$

where the integral in the second line is equal to zero, since this is just the mean of a standard normal distribution.

We can also calculate  $E[X^2]$  by explicitly integrating over the generalized PDF:

$$\begin{aligned}
 E[X^2] &= \int_{-\infty}^{\infty} x^2 \left[ \frac{1}{3} \delta(x+2) + \frac{1}{6} \delta(x-1) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right] dx \\
 &= \frac{1}{3}(4) + \frac{1}{6}(1) + \frac{1}{2} \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
 &= \frac{4}{3} + \frac{1}{6} + \frac{1}{2}(1) \\
 &= 2,
 \end{aligned}$$

where the integral in the second line is equal to 1, since this is just the variance of a standard normal distribution.

Thus:

$$\begin{aligned}
 Var[X] &= E[X^2] - E[X]^2 \\
 &= 2 - \left(-\frac{1}{2}\right)^2 \\
 &= \frac{7}{4}.
 \end{aligned}$$

### Problem 16.

(a) Let  $D$  denote the event that the device is defective, and let  $P(D) = p_d = 0.02$ . By the law of total probability, we have:

$$\begin{aligned}
 F_X(x) &= P(X \leq x) \\
 &= P(X \leq x|D)P(D) + P(X \leq x|D^c)P(D^c) \\
 &= u(x)p_d + (1 - e^{-\lambda x})(1 - p_d)u(x).
 \end{aligned}$$

By differentiating, we can find the generalized PDF:

$$\begin{aligned}
 f_X(x) &= \frac{d}{dx} F_X(x) \\
 &= p_d \delta(x) + (1 - p_d) \left[ (1 - e^{-\lambda x}) \delta(x) + u(x) \lambda e^{-\lambda x} \right] \\
 &= p_d \delta(x) + (1 - p_d) u(x) \lambda e^{-\lambda x},
 \end{aligned}$$

where I have used the fact that at  $x \neq 0$ ,  $\delta(x) = 0$ , and at  $x = 0$ ,  $(1 - \exp(-\lambda x)) = 0$ , so that the term,  $(1 - e^{-\lambda x})\delta(x)$  is 0 for all  $x$ . We also could have written this PDF down immediately by realizing that there is a probability point mass at  $x = 0$  with total probability 0.02, and there is a continuous probability contribution from the exponential distribution which must integrate to  $1 - 0.02 = 0.98$ .

(b)

$$\begin{aligned}
 P(X \geq 1) &= \int_1^\infty \left[ p_d \delta(x) + (1 - p_d)u(x)\lambda e^{-\lambda x} \right] dx \\
 &= (1 - p_d) \int_1^\infty \lambda e^{-\lambda x} dx \\
 &= (1 - p_d)e^{-\lambda} \\
 &= (0.98)e^{-2} \\
 &\approx 0.133
 \end{aligned}$$

(c)

$$\begin{aligned}
 P(X > 2 | X \geq 1) &= \frac{P(X > 2, X \geq 1)}{P(X \geq 1)} \\
 &= \frac{P(X > 2)}{P(X \geq 1)} \\
 &= \frac{\int_2^\infty e^{-\lambda x} dx}{\int_1^\infty e^{-\lambda x} dx} \\
 &= e^{-\lambda} \\
 &= e^{-2} \\
 &\approx 0.135
 \end{aligned}$$

(d) The expectation value of  $X$  is:

$$\begin{aligned}
 E[X] &= \int_{-\infty}^\infty x \left[ p_d \delta(x) + (1 - p_d)u(x)\lambda e^{-\lambda x} \right] dx \\
 &= (1 - p_d) \int_0^\infty x \lambda e^{-\lambda x} dx \\
 &= (1 - p_d) \frac{1}{\lambda} \\
 &= (0.98) \frac{1}{2} \\
 &= 0.49,
 \end{aligned}$$

where I have used the fact that  $E[X] = 1/\lambda$  for an exponential distribution.



The expectation value of  $X^2$  is:

$$\begin{aligned}
 E[X^2] &= \int_0^\infty x^2 \left[ p_d \delta(x) + (1 - p_d) u(x) \lambda e^{-\lambda x} \right] dx \\
 &= (1 - p_d) \int_0^\infty x^2 \lambda e^{-\lambda x} dx \\
 &= (1 - p_d) \frac{2}{\lambda^2} \\
 &= (0.98) \frac{1}{2} \\
 &= 0.49,
 \end{aligned}$$

where I have used the fact that  $E[X^2] = 2/\lambda^2$  for an exponential distribution. Therefore, the variance is:

$$\begin{aligned}
 Var[X] &= 0.49 - 0.49^2 \\
 &\approx 0.25.
 \end{aligned}$$

**Problem 17.**

- (a) We realize that for  $Lap(0, 1)$ ,  $f_X(x)$  is an even function, while  $x$  is an odd function and therefore  $E[X] = 0$ . Also,

$$E[X^2] = 2 \int_0^\infty x^2 \frac{1}{2} e^{-x} dx = 2,$$

where I have used the fact that since we are integrating an even function times an even function we need only integrate from 0 to  $\infty$  and multiply by 2. I have also used the fact that the integrand is  $E[X^2]$  of an  $Exp(1)$  distribution, and we know this integral evaluates to  $2/\lambda^2$ . Therefore  $Var[X] = 2$ .

- (b) We can solve for  $f_Y(y)$  using the method of transformations:

$$\begin{aligned}
 f_Y(y) &= f_X\left(\frac{y-\mu}{b}\right) \left| \frac{d}{dy} \left(\frac{y-\mu}{b}\right) \right| \\
 &= \begin{cases} \frac{1}{2} \exp\left(\frac{y-\mu}{b}\right) \frac{1}{b} & \text{for } \frac{y-\mu}{b} < 0 \\ \frac{1}{2} \exp\left[-\left(\frac{y-\mu}{b}\right)\right] \frac{1}{b} & \text{for } \frac{y-\mu}{b} \geq 0 \end{cases} \\
 &= \begin{cases} \frac{1}{2b} \exp\left(\frac{y-\mu}{b}\right) & \text{for } y < \mu \\ \frac{1}{2b} \exp\left[-\left(\frac{y-\mu}{b}\right)\right] & \text{for } y \geq \mu, \end{cases}
 \end{aligned}$$

which is  $Lap(\mu, b)$ .

- (c) Since

$$\begin{aligned}
 E[Y] &= E[bX + \mu] \\
 &= bE[X] + \mu \\
 &= \mu,
 \end{aligned}$$

and

$$\begin{aligned} E[Y^2] &= E[(bX + \mu)^2] \\ &= b^2 E[X^2] + 2b\mu E[X] + \mu^2 \\ &= 2b^2 + \mu^2, \end{aligned}$$

the variance is  $Var[Y] = E[Y^2] - E[Y]^2 = 2b^2$ .

**Problem 18.** We see firstly, that  $R_X = \mathbb{R}$  and  $R_Y = [0, \infty)$ . Also note that  $Y = |X| = -X$  for  $X < 0$  and  $X$  for  $X \geq 0$ . I use the method of transformations, breaking  $f_X(x)$  into 2 strictly monotonic regions. Let  $f_X^{(1)}(x) = (1/2b) \exp(x/b)$  and  $f_X^{(2)}(x) = (1/2b) \exp(-x/b)$ , then:

$$\begin{aligned} f_Y(y) &= f_X^{(1)}(-y) \left| \frac{dy}{dy} \right| + f_X^{(2)}(y) \left| \frac{d(-y)}{dy} \right| \\ &= \frac{1}{2b} \exp\left(-\frac{y}{b}\right) + \frac{1}{2b} \exp\left(-\frac{y}{b}\right) \\ &= \frac{1}{b} \exp\left(-\frac{y}{b}\right), \end{aligned}$$

which is  $Exp(1/b)$ .

**Problem 19.** Letting  $u \equiv 1 + x^2$ , the expectation value becomes:

$$\begin{aligned} E[X] &= \int_{-\infty}^0 \frac{1}{\pi} \frac{x}{1+x^2} dx + \int_0^{\infty} \frac{1}{\pi} \frac{x}{1+x^2} dx \\ &= \int_{\infty}^1 \frac{1}{2\pi} \frac{du}{u} + \int_1^{\infty} \frac{1}{2\pi} \frac{du}{u} \\ &= \frac{1}{2\pi} \ln(1+x^2) \Big|_{\infty}^1 + \frac{1}{2\pi} \ln(1+x^2) \Big|_1^{\infty} \\ &= -\infty + \infty, \end{aligned}$$

which is not well defined.

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{x^2}{1+x^2} dx \\ &= \frac{2}{\pi} \int_0^{\infty} \frac{x^2}{1+x^2} dx \\ &= \frac{2}{\pi} [x - \arctan(x)]_0^{\infty} \\ &= \frac{2}{\pi} \lim_{x \rightarrow \infty} [x - \arctan(x)] \\ &= \frac{2}{\pi} \lim_{x \rightarrow \infty} x - \frac{2}{\pi} \cdot \frac{\pi}{2} \\ &= \infty \end{aligned}$$

**Problem 20.**

(a) The expectation value is:

$$\begin{aligned}
 E[X] &= \frac{1}{\sigma^2} \int_0^\infty x^2 e^{-x^2/2\sigma^2} dx \\
 &= \frac{\sqrt{2\pi}}{2\sigma} \left[ \frac{2}{\sqrt{2\pi}\sigma} \int_0^\infty x^2 e^{-x^2/2\sigma^2} dx \right] \\
 &= \frac{\sqrt{2\pi}}{2\sigma} \sigma^2 \\
 &= \sigma \sqrt{\frac{\pi}{2}},
 \end{aligned}$$

where I have use the fact that the term in the brackets is the same integral one must compute to find the variance of a  $0, \sigma^2$  normal distribution.

(b) The integral we must calculate is

$$F_X(x) = \int_0^x \frac{x'}{\sigma^2} e^{-\frac{x'^2}{2\sigma^2}} dx',$$

which can be computed by a simple substitution,  $u \equiv x'^2/(2\sigma^2)$ , so that:

$$F_X(x) = \int_0^{\frac{x^2}{2\sigma^2}} e^{-u} du = 1 - e^{-\frac{x^2}{2\sigma^2}}.$$

(c) The range of both  $X$  and  $Y$  is  $[0, \infty)$ . Therefore, for all  $y \in [0, \infty)$ :

$$\begin{aligned}
 f_Y(y) &= f_X\left(\frac{y^2}{2\sigma^2}\right) \left| \frac{d}{dy} \left(\frac{y^2}{2\sigma^2}\right) \right| \\
 &= \begin{cases} \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}} & \text{for } y \geq 0 \\ 0 & \text{for } y < 0, \end{cases}
 \end{aligned}$$

which is *Rayleigh*( $\sigma$ ).

### Problem 21.

(a) The CDF is given by:

$$\begin{aligned}
 F_X(x) &= \alpha x_m^\alpha \int_{x_m}^x x'^{-\alpha-1} dx' \\
 &= 1 - \left(\frac{x_m}{x}\right)^\alpha
 \end{aligned}$$

for  $x \geq x_m$  and 0 otherwise.

(b)

$$\begin{aligned}
P(X > 3x_m | X > 2x_m) &= \frac{P(X > 3x_m, X > 2x_m)}{P(X > 2x_m)} \\
&= \frac{P(X > 3x_m)}{P(X > 2x_m)} \\
&= \frac{1 - P(X \leq 3x_m)}{1 - P(X \leq 2x_m)} \\
&= \frac{1 - F_X(3x_m)}{1 - F_X(2x_m)} \\
&= \frac{1 - \left[1 - \left(\frac{x_m}{3x_m}\right)^\alpha\right]}{1 - \left[1 - \left(\frac{x_m}{2x_m}\right)^\alpha\right]} \\
&= \left(\frac{2}{3}\right)^\alpha
\end{aligned}$$

(c) The expectation value is:

$$\begin{aligned}
E[X] &= \alpha x_m^\alpha \int_{x_m}^{\infty} x^{-\alpha-1} x dx \\
&= \frac{\alpha x_m^\alpha}{1 - \alpha} \left[ \frac{x}{x^\alpha} \right]_{x_m}^{\infty} \\
&= \frac{\alpha x_m}{\alpha - 1},
\end{aligned}$$

where the  $\lim_{x \rightarrow \infty} (x/x^\alpha)$  term evaluates to 0 since  $\alpha > 2$ . The expectation value of  $X^2$  is:

$$\begin{aligned}
E[X^2] &= \alpha x_m^\alpha \int_{x_m}^{\infty} x^{1-\alpha} x dx \\
&= \frac{\alpha x_m^2}{\alpha - 2},
\end{aligned}$$

where the  $\lim_{x \rightarrow \infty} (x^2/x^\alpha)$  term evaluates to 0 since  $\alpha > 2$ . Thus, the variance is

$$\begin{aligned}
Var[X] &= E[X^2] - E[X]^2 \\
&= \frac{\alpha x_m^2}{\alpha - 2} - \left( \frac{\alpha x_m}{\alpha - 1} \right)^2 \\
&= \frac{\alpha x_m^2}{(\alpha - 1)^2 (\alpha - 2)}.
\end{aligned}$$

**Problem 22.**

(a)

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(e^{\sigma Z + \mu} \leq x) \\
&= P\left(Z \leq \frac{\ln x - \mu}{\sigma}\right) \\
&= \Phi\left(\frac{\ln x - \mu}{\sigma}\right)
\end{aligned}$$

(b) I first find the PDF for the log-normal distribution:

$$\begin{aligned}
 f_X(x) &= \frac{d}{dx} F_X(x) \\
 &= \frac{d}{dx} \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \\
 &= \Phi'\left(\frac{\ln x - \mu}{\sigma}\right) \frac{1}{x\sigma} \\
 &= f_Z\left(\frac{\ln x - \mu}{\sigma}\right) \frac{1}{x\sigma} \\
 &= \frac{1}{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right],
 \end{aligned}$$

for all  $x \in (0, \infty)$ . The expectation value of  $X$  is thus

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] dx,$$

which can be simplified with the following substitution:

$$u \equiv \ln x - \mu,$$

$\Rightarrow$

$$du \equiv \frac{1}{x} dx = e^{-(u+\mu)} dx,$$

so that:

$$\begin{aligned}
 E[X] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty \exp\left(-\frac{1}{2} \cdot \frac{u^2}{\sigma^2}\right) \exp(u + \mu) du \\
 &= \exp\left(\mu + \frac{\sigma^2}{2}\right) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty \exp\left[-\frac{1}{2\sigma^2}(u - \sigma^2)^2\right] du \\
 &= \exp\left(\mu + \frac{\sigma^2}{2}\right).
 \end{aligned}$$

To go from the first equality to the second, I used the “completing the squares” trick to make the exponent in the form of a Gaussian for easy integration. In going from the second equality to the third, I used the fact that  $1/(\sqrt{2\pi}\sigma)$  times the integral evaluates to 1 since this is just a  $\sigma^2, \sigma^2$  Gaussian.

The expectation value of  $X^2$  is

$$E[X^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty x \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] dx,$$

and we can make the same substitutions, resulting in

$$\begin{aligned}
 E[X^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty \exp\left(-\frac{1}{2} \cdot \frac{u^2}{\sigma^2}\right) \exp(2u + 2\mu) du \\
 &= \exp(2\mu + 2\sigma^2) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty \exp\left[-\frac{1}{2\sigma^2}(u - 2\sigma^2)^2\right] du \\
 &= \exp(2\mu + 2\sigma^2),
 \end{aligned}$$

where, again, I used the completing the squares trick, and where the  $1/(\sqrt{2\pi}\sigma)$  times the integral evaluates to 1 since this is just a  $2\sigma^2, \sigma^2$  Gaussian.

Therefore:

$$\begin{aligned} \text{Var}[X] &= E[X^2] - E[X]^2 \\ &= \exp(2\mu + \sigma^2) (\exp \sigma^2 - 1). \end{aligned}$$

**Problem 23.** The expectation value is:

$$E[Y] = E[X_1 + X_2 + \dots + X_n] = nE[X_1] = \frac{n}{\lambda},$$

since the expectation value for an  $\text{Exp}(\lambda)$  distribution is  $1/\lambda$ .

Since  $X_1, X_2, \dots, X_n$  are iid, the variance is linear, and thus:

$$\text{Var}[Y] = \text{Var}[X_1 + X_2 + \dots + X_n] = n\text{Var}[X_1] = \frac{n}{\lambda^2},$$

since the variance for an  $\text{Exp}(\lambda)$  distribution is  $1/\lambda^2$ .

## Chapter 5

# Joint Distributions

**Problem 1.**

(a)

$$P(X \leq 2, Y > 1) = 0 + \frac{1}{12} = \frac{1}{12}$$

(b)

$$P_X(x) = \begin{cases} \frac{1}{3} + \frac{1}{12} & \text{for } x = 1 \\ \frac{1}{6} + 0 & \text{for } x = 2 \\ \frac{1}{12} + \frac{1}{3} & \text{for } x = 4 \end{cases} = \begin{cases} \frac{5}{12} & \text{for } x = 1 \\ \frac{1}{6} & \text{for } x = 2 \\ \frac{5}{12} & \text{for } x = 4 \end{cases}$$

$$P_Y(y) = \begin{cases} \frac{1}{3} + \frac{1}{6} + \frac{1}{12} & \text{for } y = 1 \\ \frac{1}{12} + 0 + \frac{1}{3} & \text{for } y = 2 \end{cases} = \begin{cases} \frac{7}{12} & \text{for } y = 1 \\ \frac{5}{12} & \text{for } y = 2 \end{cases}$$

(c)

$$P(Y = 2|X = 1) = \frac{P(Y = 2, X = 1)}{P(X = 1)} = \frac{\frac{1}{12}}{\frac{5}{12}} = \frac{1}{5}$$

$$P(Y = 2|X = 1) = \frac{1}{5} \neq P(Y = 2) = \frac{5}{12}$$

 $\implies$  not independent**Problem 2.**

- (a) The ranges of  $X$ ,  $Y$  and  $Z$  are:  $R_X = \{1, 2, 4\}$ ,  $R_Y = \{1, 2\}$  and  $R_Z = \{-3, -2, -1, 0, 2\}$ .  
The mapping  $g(x, y) = x - 2y$  (where  $g : R_X \times R_Y \rightarrow R_Z$ ) is explicitly given by:

$$(1, 1) \rightarrow -1$$

$$(1, 2) \rightarrow -3$$

$$(2, 1) \rightarrow 0$$

$$(2, 2) \rightarrow -2$$

$$(4, 1) \rightarrow 2$$

$$(4, 2) \rightarrow 0.$$

Thus we have the following:



$$\begin{aligned}
P_Z(z) &= \begin{cases} P(X - 2Y = -3) & \text{for } z = -3 \\ P(X - 2Y = -2) & \text{for } z = -2 \\ P(X - 2Y = -1) & \text{for } z = -1 \\ P(X - 2Y = 0) & \text{for } z = 0 \\ P(X - 2Y = 2) & \text{for } z = 2 \end{cases} \\
&= \begin{cases} P_{XY}(1, 2) & \text{for } z = -3 \\ P_{XY}(2, 2) & \text{for } z = -2 \\ P_{XY}(1, 1) & \text{for } z = -1 \\ P_{XY}(2, 1) + P_{XY}(4, 2) & \text{for } z = 0 \\ P_{XY}(4, 1) & \text{for } z = 2 \end{cases} \\
&= \begin{cases} \frac{1}{12} & \text{for } z = -3 \\ 0 & \text{for } z = -2 \\ \frac{1}{3} & \text{for } z = -1 \\ \frac{1}{2} & \text{for } z = 0 \\ \frac{1}{12} & \text{for } z = 2, \end{cases}
\end{aligned}$$

which, as a sanity check, add up to 1.

(b)

$$\begin{aligned}
P(X = 2|Z = 0) &= P(X = 2|X - 2Y = 0) \\
&= P(X = 2|X = 2Y) \\
&= \frac{P(X = 2, X = 2Y)}{P(X = 2Y)} \\
&= \frac{P(X = 2, Y = 1)}{P(X = 2, Y = 1) + P(X = 4, Y = 2)} \\
&= \frac{1}{3}
\end{aligned}$$

**Problem 3.** Let  $A$  be the event that the first coin we pick is the fair coin. We can find the joint PMF by conditioning on this event, and realizing that once conditioned,  $X$  and  $Y$  are independent (i.e.,  $X$  and  $Y$  are conditionally independent given  $A$ ):

$$\begin{aligned}
P_{XY}(x, y) &= P(X = x, Y = y|A)P(A) + P(X = x, Y = y|A^c)P(A^c) \\
&= P_{1/2}(x)P_{2/3}(y) \left(\frac{1}{2}\right) + P_{2/3}(x)P_{1/2}(y) \left(\frac{1}{2}\right),
\end{aligned}$$

where  $P_p(z)$  is the PMF associated with a  $Bern(p)$  trial. This PMF can be written conveniently as  $P_p(z) = p^z(1-p)^{1-z}$ , so that the joint PMF is

$$P_{XY}(x, y) = \left(\frac{1}{2}\right)^2 \left(\frac{2}{3}\right)^y \left(\frac{1}{3}\right)^{1-y} + \left(\frac{1}{2}\right)^2 \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{1-x}.$$

This can also be written in tabular form as:

	$Y = 0$	$Y = 1$
$X = 0$	$\frac{1}{6}$	$\frac{1}{4}$
$X = 1$	$\frac{1}{4}$	$\frac{1}{3}$

To check if  $X$  and  $Y$  are independent I first check  $(x, y) = (0, 0)$ . Adding the table horizontally and vertically, the marginalized PMFs at these values are  $P_X(0) = 5/12$  and  $P_Y(0) = 5/12$ , and thus  $P_X(0)P_Y(0) = 25/144 \neq 1/6$ , so  $X$  and  $Y$  are not independent.

**Problem 4.**

(a) I first find the marginalized PMFs:

$$\begin{aligned}
 P_X(k) &= \sum_{l=1}^{\infty} \frac{1}{2^{k+l}} \\
 &= \frac{1}{2^k} \left[ -1 + \sum_{l=0}^{\infty} \left( \frac{1}{2} \right)^l \right] \\
 &= \frac{1}{2^k},
 \end{aligned}$$

and by symmetry

$$P_Y(l) = \frac{1}{2^l},$$

We then have that:

$$\begin{aligned}
 P_{XY}(k, l) &= \frac{1}{2^{k+l}} \\
 &= \frac{1}{2^k 2^l} \\
 &= \frac{1}{2^k} \cdot \frac{1}{2^l} \\
 &= P_X(k)P_Y(l) \quad \forall (k, l) \in \mathbb{N} \times \mathbb{N},
 \end{aligned}$$

and thus  $X$  and  $Y$  are independent.

(b) We can easily enumerate all pairs of  $(x, y)$  that satisfy this inequality:

$$\begin{aligned}
 P(X^2 + Y^2 \leq 10) &= P_{XY}(1, 1) + P_{XY}(1, 2) + P_{XY}(1, 3) + P_{XY}(2, 1) + P_{XY}(2, 2) + P_{XY}(3, 1) \\
 &= \frac{1}{2 \cdot 2} + \frac{1}{2 \cdot 2^2} + \frac{1}{2 \cdot 2^3} + \frac{1}{2^2 \cdot 2} + \frac{1}{2^2 \cdot 2^2} + \frac{1}{2^3 \cdot 2} \\
 &= \frac{11}{16}.
 \end{aligned}$$

**Problem 5.**

(a)

$$P_{X|Y}(x|1) = \begin{cases} \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6} + \frac{1}{12}} & \text{for } x = 1 \\ \frac{\frac{1}{6}}{\frac{1}{3} + \frac{1}{6} + \frac{1}{12}} & \text{for } x = 2 \\ \frac{\frac{1}{12}}{\frac{1}{3} + \frac{1}{6} + \frac{1}{12}} & \text{for } x = 4 \end{cases} = \begin{cases} \frac{4}{7} & \text{for } x = 1 \\ \frac{2}{7} & \text{for } x = 2 \\ \frac{1}{7} & \text{for } x = 4 \end{cases}$$

(b)

$$E[X|Y = 1] = (1) \cdot \frac{4}{7} + (2) \cdot \frac{2}{7} + (4) \cdot \frac{1}{7} = \frac{12}{7}$$

(c)

$$Var[X|Y = 1] = \left(1 - \frac{12}{7}\right)^2 \cdot \frac{4}{7} + \left(2 - \frac{12}{7}\right)^2 \cdot \frac{2}{7} + \left(4 - \frac{12}{7}\right)^2 \cdot \frac{1}{7} = \frac{52}{49}$$

**Problem 6.** We know that  $X \sim Pois(10)$  and since each customer is female independent of the other customers, if the total number of customers in an hour is  $n$ , then the total number of female customers in an hour is the sum of  $n$  independent Bernoulli random variables. In other words,  $Y|X = n \sim Bin(n, 3/4)$ . Therefore, the joint PMF is:

$$\begin{aligned} P(X = n, Y = y) &= P(Y = y|X = n)P(X = n) \\ &= \binom{n}{y} \left(\frac{3}{4}\right)^y \left(\frac{1}{4}\right)^{n-y} \frac{10^n e^{-10}}{n!}. \end{aligned}$$

**Problem 7.** We know that for a  $Geom(p)$  distribution the mean is  $1/p$  and the variance is  $(1 - p)/p^2$ , so we should expect these answers. We can find the mean by conditioning on the first “toss”:

$$\begin{aligned} E[X] &= E[X|H]p(H) + E[X|H^c]p(H^c) \\ &= 1 \cdot p + (1 + E[X])(1 - p), \end{aligned}$$

where  $E[X|H] = 1$  since if we know the first toss is a heads, the experiment is done so that the mean is 1, and  $E[X|H^c] = (1 + E[X])$  since if the first toss is a tails, then we’ve wasted 1 toss, and since the geometric distribution is memoryless, it starts over at the next toss. Solving this equation for  $E[X]$  we find that  $E[X] = 1/p$ , which is what we expected.

We can solve for  $E[X^2]$  in a similar manner:

$$\begin{aligned} E[X^2] &= E[X^2|H]p(H) + E[X^2|H^c]p(H^c) \\ &= 1 \cdot p + E[(1 + X)^2](1 - p) \\ &= p + E[X^2](1 - p) + 2E[X](1 - p) + (1 - p), \end{aligned}$$

where  $E[X^2|H] = 1$  for the same reason as above, and  $E[X^2|H^c] = E[(1 + X)^2]$  since, as above, we’ve wasted 1 toss on the first toss, and then the experiment starts over on the second. Solving this equation, I find  $E[X^2] = (2 - p)/p^2$ . The variance is thus:  $Var[X] = E[X^2] - E[X]^2 = (1 - p)/p^2$ , which is what we expected.

**Problem 8.** If  $X, Y \stackrel{iid}{\sim} Geom(p)$ , then we can easily find the joint PMF and use LOTUS to solve for the expectation. The joint PMF is:

$$P_{XY}(x, y) = P_X(x)P_Y(y) = p(1 - p)^{x-1}p(1 - p)^{y-1} \quad \text{for } x, y = 1, 2, \dots, \quad (5.1)$$

where I have multiplied the marginal PMFs since  $X$  and  $Y$  are independent. Using LOTUS:

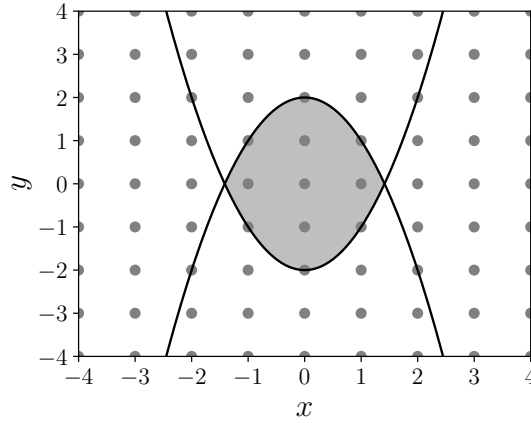


Figure 5.1: A visual representation of the set  $C$  for Problems 9 and 10.

$$\begin{aligned}
 E \left[ \frac{X^2 + Y^2}{XY} \right] &= \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \left( \frac{x^2 + y^2}{xy} \right) p(1-p)^{x-1} p(1-p)^{y-1} \\
 &= \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \frac{x}{y} p(1-p)^{x-1} p(1-p)^{y-1} + \sum_{y=1}^{\infty} \sum_{x=1}^{\infty} \frac{y}{x} p(1-p)^{x-1} p(1-p)^{y-1} \\
 &= 2 \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} \frac{x}{y} p(1-p)^{x-1} p(1-p)^{y-1} \\
 &= 2 \sum_{x=1}^{\infty} x p(1-p)^{x-1} \sum_{y=1}^{\infty} \frac{1}{y} p(1-p)^{y-1},
 \end{aligned}$$

where going from the second to third line we realize that due to the symmetry, both of the sums are the same. In the last line the first sum is just the mean of a  $\text{Geom}(p)$  distribution ( $1/p$ ). We can simplify the second sum by utilizing the following Taylor expansion:

$$-\ln(1-x) = \sum_{k=1}^{\infty} \frac{x^k}{k} \quad \text{for } |x| < 1.$$

Thus, we arrive at:

$$\begin{aligned}
 E \left[ \frac{X^2 + Y^2}{XY} \right] &= \frac{2}{p} p(1-p)^{-1} (-\ln p) \\
 &= \frac{2}{1-p} \ln \left( \frac{1}{p} \right).
 \end{aligned}$$

**Problem 9.** To better understand what is in the set  $C$ , note that  $C$  is the set of  $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ , such that  $y \leq 2 - x^2$  for  $y \geq 0$  and  $y \geq x^2 - 2$  for  $y < 0$ . To visualize  $C$ , I plot the set  $\mathbb{Z} \times \mathbb{Z}$  as grey points in Fig. 5.1 as well as the lines  $y = 2 - x^2$  and  $y = x^2 - 2$ . The shaded grey region (and the lines themselves) represents the region satisfying the 2 conditions, and thus any grey point in this region (or on the lines) is in  $C$ . Therefore, more explicitly,  $C = \{(0, 0), (1, 0), (0, 1), (1, 1), (0, 2), (0, -1), (1, -1), (0, -2), (-1, 0), (-1, 1), (-1, -1)\}$ .

(a) The joint PMF is:

$$P_{XY}(x, y) = \begin{cases} \frac{1}{11} & \text{for } (x, y) \in C \\ 0 & \text{otherwise.} \end{cases}$$

By looking at Fig. 5.1 and adding vertically and horizontally, we can easily determine that the marginal PMFs are:

$$P_X(x) = \begin{cases} \frac{3}{11} & \text{for } x = -1 \\ \frac{5}{11} & \text{for } x = 0 \\ \frac{3}{11} & \text{for } x = 1 \end{cases}$$

and

$$P_Y(y) = \begin{cases} \frac{1}{11} & \text{for } y = -2 \\ \frac{3}{11} & \text{for } y = -1 \\ \frac{3}{11} & \text{for } y = 0 \\ \frac{3}{11} & \text{for } y = 1 \\ \frac{1}{11} & \text{for } y = 2. \end{cases}$$

(b) Since there are 3 points at  $Y = 1$ , and each point is equally as likely, the total probability mass at  $Y = 1$  is  $3/11$ , while the total probability mass at  $(-1, 1)$ ,  $(0, 1)$ ,  $(1, 1)$  is  $1/11$  respectively. Therefore:

$$P_{X|Y}(x|1) = \begin{cases} \frac{1}{3} & \text{for } x = -1 \\ \frac{1}{3} & \text{for } x = 0 \\ \frac{1}{3} & \text{for } x = 1. \end{cases}$$

(c)  $X$  and  $Y$  are not independent since, for example, at  $X = -1$ :  $P(X = -1|Y = 1) = 1/3 \neq P(X = -1) = 3/11$ .

(d) Using LOTUS, we have

$$E[XY^2] = \sum_{(x,y) \in C} xy^2 P_{XY}(x, y) = \frac{1}{11} \sum_{(x,y) \in C} xy^2 = \frac{1}{11} [1 \cdot 1^2 + 1 \cdot (-1)^2 - 1 \cdot 1^2 - 1 \cdot (-1)^2] = 0,$$

where only 4 points contribute to the sum (since the rest have zeros).

**Problem 10.**

(a)

$$E[X|Y = 1] = \sum_{x \in R_{X|Y=1}} x P_{X|Y}(x|1) = (-1) \cdot \frac{1}{3} + (0) \cdot \frac{1}{3} + (1) \cdot \frac{1}{3} = 0$$

(b)

$$Var[X|Y = 1] = \sum_{x \in R_{X|Y=1}} x^2 P_{X|Y}(x|1) = (1) \cdot \frac{1}{3} + (0) \cdot \frac{1}{3} + (1) \cdot \frac{1}{3} = \frac{2}{3}$$

(c) One can easily see that the PMF,  $P_{X||Y| \leq 1}(x)$  is exactly the same as the PMF for  $P_{X|Y}(x|1)$ , and therefore the expectation and variance will be the same, thus  $E[X||Y| \leq 1] = 0$ .

(d) For the same reason as part c of this problem  $E[X^2||Y| \leq 1] = 2/3$ .

**Problem 11.** If there are  $n$  cars in the shop, then  $X = X_1 + X_2 + \dots + X_n$ , where  $X_i$  is a  $Bern(3/4)$  random variable (as specified in the problem), and where  $X_1, X_2, \dots, X_n$  are all independent (as specified in the problem). Thus we have that  $X|N = n \sim Bin(n, 3/4)$  and for the same reason,  $Y|N = n \sim Bin(n, 1/4)$ .

- (a) Noting that  $R_X = R_Y = \{0, 1, 2, 3\}$ , we can use the law of total probability to find both of the marginal PMFs, which are:

$$\begin{aligned} P_X(x) &= \sum_{n=0}^3 P(X = x|N = n)P_N(n) \\ &= \sum_{n=0}^3 \binom{n}{x} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{n-x} P_N(n), \end{aligned}$$

and

$$\begin{aligned} P_Y(y) &= \sum_{n=0}^3 P(Y = y|N = n)P_N(n) \\ &= \sum_{n=0}^3 \binom{n}{y} \left(\frac{1}{4}\right)^y \left(\frac{3}{4}\right)^{n-y} P_N(n). \end{aligned}$$

I compute both of these PMFs numerically to find:

$$P_X(x) \approx \begin{cases} 0.180 & \text{for } x = 0 \\ 0.258 & \text{for } x = 1 \\ 0.352 & \text{for } x = 2 \\ 0.211 & \text{for } x = 3 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$P_Y(y) \approx \begin{cases} 0.570 & \text{for } y = 0 \\ 0.336 & \text{for } y = 1 \\ 0.086 & \text{for } y = 2 \\ 0.008 & \text{for } y = 3 \\ 0 & \text{otherwise,} \end{cases}$$

which, as a sanity, both add up to approximately 1. We see that since a 4 door car is more likely than a 2 door car, the marginalized PMF for the 4 door cars skews towards high numbers, while the marginalized PMF for the 2 door cars skews towards lower numbers.

- (b) We can find the joint PMF for  $X$  and  $Y$  by conditioning on  $N$  and using the law of total probability:

$$P_{XY}(x, y) = \sum_{n=0}^3 P(X = x, Y = y|N = n)P_N(n),$$

where we can get rid of the sum because the probability is 0 if  $x + y \neq n$ :

$$\begin{aligned} P_{XY}(x, y) &= P(X = x, Y = y|N = x + y)P_N(x + y) \\ &= P(X = x|Y = y, N = x + y)P(Y = y|N = x + y)P_N(x + y) \\ &= P(Y = y|N = x + y)P_N(x + y) \\ &= \binom{x + y}{y} \left(\frac{1}{4}\right)^y \left(\frac{3}{4}\right)^x P_N(x + y), \end{aligned}$$

where in the second line I have used the chain rule of probability, in the third I have used the fact that given  $Y = y$  and  $N = x + y$ , we are sure that  $X = x$ , and in the fourth line I have used the fact that  $Y|N = x + y \sim \text{Bin}(x + y, 1/4)$ . I compute the joint PMF numerically and present the results in the following table:

$$P_{XY}(x, y) \approx$$

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$
$X = 0$	0.125	0.031	0.016	0.008
$X = 1$	0.094	0.094	0.070	0
$X = 2$	0.141	0.211	0	0
$X = 3$	0.211	0	0	0

As a check, I made sure that the above PMF sums to approximately 1.

- (c)  $X$  and  $Y$  are not independent since  $P_{XY}(x, y) \neq P_X(x)P_Y(y) \forall x, y$ . For example  $P_{XY}(0, 0) = 0.125$ , while  $P_X(0)P_Y(0) \approx 0.180 \cdot 0.570 = 0.103$ .

**Problem 12.** I first note that  $R_X = R_Y = \{1, 2, 3, 4, 5\}$  and  $R_Z = \{-4, -3, \dots, 3, 4\}$ . I can find  $P_Z(z)$  by conditioning on either  $X$  or  $Y$  and by using independence:

$$\begin{aligned}
 P_Z(z) &= P(Z = z) \\
 &= P(Y = X - z) \\
 &= \sum_{x=1}^5 P(Y = X - z | X = x) P_X(x) \\
 &= \frac{1}{5} \sum_{x=1}^5 P(Y = x - z | X = x) \\
 &= \frac{1}{5} \sum_{x=1}^5 P(Y = x - z) \\
 &= \frac{1}{25} \sum_{x=1}^5 \mathbb{1}\{x - z \in R_Y\},
 \end{aligned}$$

where in going from the fourth to fifth line I used independence. Thus we have:

$$P_Z(z) = \begin{cases} \frac{1}{25} & \text{for } z = -4 \\ \frac{2}{25} & \text{for } z = -3 \\ \frac{3}{25} & \text{for } z = -2 \\ \frac{4}{25} & \text{for } z = -1 \\ \frac{5}{25} & \text{for } z = 0 \\ \frac{4}{25} & \text{for } z = 1 \\ \frac{3}{25} & \text{for } z = 2 \\ \frac{2}{25} & \text{for } z = 3 \\ \frac{1}{25} & \text{for } z = 4. \end{cases}$$

**Problem 13.**

(a)

$$P_X(x) = \begin{cases} \frac{1}{6} + \frac{1}{6} + \frac{1}{8} & \text{for } x = 0 \\ \frac{1}{8} + \frac{1}{6} + \frac{1}{4} & \text{for } x = 1 \end{cases} = \begin{cases} \frac{11}{24} & \text{for } x = 0 \\ \frac{13}{24} & \text{for } x = 1 \end{cases}$$

$$P_Y(y) = \begin{cases} \frac{1}{6} + \frac{1}{8} & \text{for } y = 0 \\ \frac{1}{6} + \frac{1}{6} & \text{for } y = 1 \\ \frac{1}{8} + \frac{1}{4} & \text{for } y = 2 \end{cases} = \begin{cases} \frac{7}{24} & \text{for } y = 0 \\ \frac{1}{3} & \text{for } y = 1 \\ \frac{3}{8} & \text{for } y = 2 \end{cases}$$

(b)

$$P_{X|Y}(x|0) = \begin{cases} \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{8}} & \text{for } x = 0 \\ \frac{\frac{1}{8}}{\frac{1}{6} + \frac{1}{8}} & \text{for } x = 1 \end{cases} = \begin{cases} \frac{4}{7} & \text{for } x = 0 \\ \frac{3}{7} & \text{for } x = 1 \end{cases}$$

$$P_{X|Y}(x|1) = \begin{cases} \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{6}} & \text{for } x = 0 \\ \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{6}} & \text{for } x = 1 \end{cases} = \begin{cases} \frac{1}{2} & \text{for } x = 0 \\ \frac{1}{2} & \text{for } x = 1 \end{cases}$$

$$P_{X|Y}(x|2) = \begin{cases} \frac{\frac{1}{8}}{\frac{1}{8} + \frac{1}{4}} & \text{for } x = 0 \\ \frac{\frac{1}{4}}{\frac{1}{8} + \frac{1}{4}} & \text{for } x = 1 \end{cases} = \begin{cases} \frac{1}{3} & \text{for } x = 0 \\ \frac{2}{3} & \text{for } x = 1 \end{cases}$$

(c) We know that

$$Z = \begin{cases} E[X|Y = 0] & \text{with probability } P_Y(0) \\ E[X|Y = 1] & \text{with probability } P_Y(1) \\ E[X|Y = 2] & \text{with probability } P_Y(2), \end{cases}$$

or in other words:

$$P_Z(z) = \begin{cases} P_Y(0) & \text{for } z = E[X|Y = 0] \\ P_Y(1) & \text{for } z = E[X|Y = 1] \\ P_Y(2) & \text{for } z = E[X|Y = 2]. \end{cases}$$



We already know the marginal PMF of  $Y$ , and thus what is left to calculate is  $E[X|Y = y]$  for all  $y \in R_Y$ :

$$\begin{aligned} E[X|Y = 0] &= \sum_{x \in R_X} x P_{X|Y}(x|0) \\ &= 0 \left( \frac{4}{7} \right) + 1 \left( \frac{3}{7} \right) \\ &= \frac{3}{7}, \end{aligned}$$

$$\begin{aligned} E[X|Y = 1] &= \sum_{x \in R_X} x P_{X|Y}(x|1) \\ &= 0 \left( \frac{1}{2} \right) + 1 \left( \frac{1}{2} \right) \\ &= \frac{1}{2}, \end{aligned}$$

$$\begin{aligned} E[X|Y = 2] &= \sum_{x \in R_X} x P_{X|Y}(x|2) \\ &= 0 \left( \frac{1}{3} \right) + 1 \left( \frac{2}{3} \right) \\ &= \frac{2}{3}. \end{aligned}$$

Finally, we have that:

$$P_Z(z) = \begin{cases} \frac{7}{24} & \text{for } z = \frac{3}{7} \\ \frac{1}{3} & \text{for } z = \frac{1}{2} \\ \frac{3}{8} & \text{for } z = \frac{2}{3}. \end{cases}$$

- (d) For this problem we are checking that the law of iterated expectations holds. That is, we need to check explicitly that  $E[X] = E[E[X|Y]]$ , where the outer expectation on the RHS is over  $Y$ . Computing the LHS I have:

$$E[X] = \sum_{x \in R_X} x P_X(x) = 0 \left( \frac{11}{24} \right) + 1 \left( \frac{13}{24} \right) = \frac{13}{24}.$$

Computing the RHS I have:

$$\begin{aligned} E[Z] &= E_Y[E[X|Y]] \\ &= \sum_{y \in R_Y} E[X|Y = y] P_Y(y) \\ &= \left( \frac{3}{7} \right) \left( \frac{7}{24} \right) + \left( \frac{1}{2} \right) \left( \frac{1}{3} \right) + \left( \frac{2}{3} \right) \left( \frac{3}{8} \right) \\ &= \frac{13}{24}. \end{aligned}$$

The LHS and RHS agree, and thus the law of iterated expectations holds.

(e)

$$\begin{aligned}
E[Z^2] &= \sum_{z \in R_Z} z^2 P_Z(z) \\
&= \left(\frac{3}{7}\right)^2 \left(\frac{7}{24}\right) + \left(\frac{1}{2}\right)^2 \left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^2 \left(\frac{3}{8}\right) \\
&= \frac{17}{56}
\end{aligned}$$

 $\Rightarrow$ 

$$Var[Z] = E[Z^2] - E[Z]^2 = \frac{17}{56} - \left(\frac{13}{24}\right)^2 = \frac{41}{4032}$$

**Problem 14.**

(a) As with the previous problem, we know that

$$V = \begin{cases} Var[X|Y=0] & \text{with probability } P_Y(0) \\ Var[X|Y=1] & \text{with probability } P_Y(1) \\ Var[X|Y=2] & \text{with probability } P_Y(2), \end{cases}$$

or in other words:

$$P_V(v) = \begin{cases} P_Y(0) & \text{for } v = Var[X|Y=0] \\ P_Y(1) & \text{for } v = Var[X|Y=1] \\ P_Y(2) & \text{for } v = Var[X|Y=2]. \end{cases}$$

Thus, we must compute  $Var[X|Y=y] = E[X^2|Y=y] - E[X|Y=y]^2$  for all  $y \in R_Y$ .  $E[X|Y=y]$  was already computed in the previous problem, and, since to compute  $E[X^2|Y=y]$ , both terms in the summation are the same as the two terms in the summation to compute  $E[X|Y=y]$  (since  $0^2 = 0$  and  $1^2 = 1$ ), we have that  $E[X^2|Y=y] = E[X|Y=y]$  (this can be seen more clearly if one explicitly writes out the summation for  $E[X^2|Y=y]$ ). Thus we have that:

$$Var[X|Y=0] = E[X^2|Y=0] - E[X|Y=0]^2 = \frac{3}{7} - \left(\frac{3}{7}\right)^2 = \frac{12}{49},$$

$$Var[X|Y=1] = E[X^2|Y=1] - E[X|Y=1]^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4},$$

and

$$Var[X|Y=2] = E[X^2|Y=2] - E[X|Y=2]^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}.$$

The PMF for  $V$  is thus

$$P_V(v) = \begin{cases} \frac{7}{24} & \text{for } v = \frac{12}{49} \\ \frac{1}{3} & \text{for } v = \frac{1}{4} \\ \frac{3}{8} & \text{for } v = \frac{2}{9}. \end{cases}$$

(b)

$$E[V] = \sum_{v \in R_V} v P_V(v) = \frac{12}{49} \cdot \frac{7}{24} + \frac{1}{4} \cdot \frac{1}{3} + \frac{2}{9} \cdot \frac{3}{8} = \frac{5}{21}$$

- (c) In this problem we are checking that the law of total variance,  $Var[X] = E_Y[Var[X|Y]] + Var_Y[E[X|Y]]$ , holds (where the subscript  $Y$  on the expectation and variance denotes with respect to the random variable  $Y$ .) Computing the LHS:

$$\begin{aligned}
 Var[X] &= E[X^2] - E[X]^2 \\
 &= \sum_{x \in R_X} x^2 P_X(x) - \left[ \sum_{x \in R_X} x P_X(x) \right]^2 \\
 &= 0^2 \cdot \frac{11}{24} + 1^2 \cdot \frac{13}{24} - \left[ 0 \cdot \frac{11}{24} + 1 \cdot \frac{13}{24} \right]^2 \\
 &= \frac{143}{576}.
 \end{aligned}$$

Computing the RHS:

$$\begin{aligned}
 E_Y[Var[X|Y]] + Var_Y[E[X|Y]] &= E[V] + Var[Z] \\
 &= \frac{5}{21} + \frac{41}{4032} \\
 &= \frac{143}{576},
 \end{aligned}$$

which is in agreement with the LHS of the equation. Note that  $E[V]$  and  $Var[Z]$  were computed in this problem and the previous problem.

**Problem 15.** The law of total expectation gives:

$$\begin{aligned}
 E[Y] &= \sum_{n=0}^{\infty} E[Y|N=n] P_N(n) \\
 &= \sum_{n=0}^{\infty} E \left[ \sum_{i=1}^N X_i | N=n \right] P_N(n) \\
 &= \sum_{n=0}^{\infty} E \left[ \sum_{i=1}^n X_i \right] P_N(n) \\
 &= \sum_{n=0}^{\infty} \sum_{i=1}^n E[X_i] P_N(n) \\
 &= \sum_{n=0}^{\infty} \sum_{i=1}^n E[X_i] \frac{e^{-\beta} \beta^n}{n!} \\
 &= \sum_{n=0}^{\infty} \sum_{i=1}^n \frac{1}{\lambda} \frac{e^{-\beta} \beta^n}{n!} \\
 &= \frac{e^{-\beta}}{\lambda} \sum_{n=0}^{\infty} \frac{n \beta^n}{n!},
 \end{aligned}$$

where in going from the second to the third line I have used the fact that  $X_i$  and  $N$  are independent (for all  $i$ ), in going from the third to the fourth line I have used the linearity of expectation, and in going from the fifth to sixth line I have used the fact that for an  $Exp(\lambda)$  distribution,  $E[X] = 1/\lambda$ . This summation can be computed by considering the Taylor expansion of the

exponential,  $\exp(x) = \sum_{n=0}^{\infty} (x^n)/n!$ . Taking the derivative of both sides of this formula with respect to  $x$ , we find that the desired sum is:

$$\sum_{n=0}^{\infty} \frac{nx^n}{n!} = xe^x,$$

and hence

$$E[Y] = \frac{\beta}{\lambda}.$$

The calculation for  $Var[Y]$  is similar. For this calculation, we will need  $\sum_{n=0}^{\infty} (n^2 x^n)/n!$ , which can be found with the same differentiation strategy. I differentiate the equation for the previous summation with respect to  $x$  once more and solve for the desired summation to find

$$\sum_{n=0}^{\infty} \frac{n^2 x^n}{n!} = xe^x + x^2 e^x,$$

where I have used the chain rule in differentiating.

To find  $Var[Y]$  I now solve for  $E[Y^2]$ , which is only moderating more complicated than solving for  $E[Y]$ . The law of total expectation gives:

$$\begin{aligned} E[Y^2] &= \sum_{n=0}^{\infty} E[Y^2|N=n]P_N(n) \\ &= \sum_{n=0}^{\infty} E\left[\left(\sum_{i=1}^N X_i\right)^2 \mid N=n\right]P_N(n) \\ &= \sum_{n=0}^{\infty} E\left[\left(\sum_{i=1}^n X_i\right)^2\right]P_N(n) \\ &= \sum_{n=0}^{\infty} E\left[\sum_{i=1}^n X_i^2 + \sum_{j,k:j \neq k} X_j X_k\right]P_N(n) \\ &= \sum_{n=0}^{\infty} \left\{ \sum_{i=1}^n E[X_i^2] + \sum_{j,k:j \neq k} E[X_j X_k] \right\} P_N(n) \\ &= \sum_{n=0}^{\infty} \left\{ \sum_{i=1}^n E[X_i^2] + \sum_{j,k:j \neq k} E[X_j]E[X_k] \right\} P_N(n) \\ &= \sum_{n=0}^{\infty} \left\{ \sum_{i=1}^n \frac{2}{\lambda^2} + \sum_{j,k:j \neq k} \frac{1}{\lambda^2} \right\} P_N(n) \\ &= \sum_{n=0}^{\infty} \left[ \frac{2n}{\lambda^2} + \frac{n^2 - n}{\lambda^2} \right] \frac{e^{-\beta} \beta^n}{n!} \\ &= \frac{e^{-\beta}}{\lambda^2} \sum_{n=0}^{\infty} \frac{n\beta^n}{n!} + \frac{e^{-\beta}}{\lambda^2} \sum_{n=0}^{\infty} \frac{n^2 \beta^n}{n!}, \end{aligned}$$

where in going from the second to third line I have used the fact that  $X_i$  and  $N$  are independent (for all  $i$ ), in going from the third to fourth line I have broken the square of the summation into the summation of the squares plus the summation of the cross-terms. The notation  $\sum_{j,k:j \neq k}$  denotes

a sum over all possible tuples of  $(j, k)$ , where  $j, k = 1, 2, \dots, n$ , except the tuples where  $j = k$ . In going from the fourth to fifth line I have used the linearity of expectation, in going from the fifth to sixth line I have used the independence of all  $X_i$ s, and in going from the sixth to seventh line I have used the fact that for an  $Exp(\lambda)$  distribution,  $E[X^2] = 2/\lambda^2$  (as calculated in the book). The first summation summation has already been solved for, and to solve the second summation, I use the formula for  $\sum_{n=0}^{\infty} \frac{n^2 x^n}{n!}$  as derived above. Thus I have that

$$\begin{aligned} E[Y^2] &= \frac{1}{\lambda} \cdot \frac{\beta}{\lambda} + \frac{e^{-\beta}}{\lambda^2} (\beta e^{\beta} + \beta^2 e^{\beta}) \\ &= \frac{\beta^2 + 2\beta}{\lambda^2}. \end{aligned}$$

Finally we have that:

$$\begin{aligned} Var[Y] &= E[Y^2] - E[Y]^2 \\ &= \frac{\beta^2 + 2\beta}{\lambda^2} - \frac{\beta^2}{\lambda^2} \\ &= \frac{2\beta}{\lambda^2}. \end{aligned}$$

#### Problem 16.

(a)

$$\begin{aligned} 1 &= \int_0^1 \int_0^{\infty} f_{XY}(x, y) dx dy \\ &= \int_0^1 \int_0^{\infty} \frac{1}{2} e^{-x} dx dy + c \int_0^1 \int_0^{\infty} \frac{y}{(1+x)^2} dx dy \\ &= \int_0^1 \int_0^{\infty} \frac{1}{2} e^{-x} dx dy + c \int_0^1 \int_1^{\infty} \frac{y}{u^2} du dy \\ &= \frac{1}{2} + \frac{c}{2} \end{aligned}$$

$\Rightarrow$

$$c = 1$$

(b)

$$\begin{aligned} P\left(0 \leq X \leq 1, 0 \leq Y \leq \frac{1}{2}\right) &= \int_0^{1/2} \int_0^1 f_{XY}(x, y) dx dy \\ &= \int_0^{1/2} \int_0^1 \left[ \frac{1}{2} e^{-x} + \frac{y}{(1+x)^2} \right] dx dy \\ &= \frac{1}{4}(1 - e^{-1}) + \frac{1}{16} \\ &\approx 0.22 \end{aligned}$$

(c)

$$\begin{aligned}
P(0 \leq X \leq 1) &= \int_0^1 \int_0^1 f_{XY}(x, y) dx dy \\
&= \int_0^1 \int_0^1 \left[ \frac{1}{2} e^{-x} + \frac{y}{(1+x)^2} \right] dx dy \\
&= \frac{1}{2} (1 - e^{-1}) + \frac{1}{4} \\
&\approx 0.57
\end{aligned}$$

**Problem 17.**

(a)

$$\begin{aligned}
f_X(x) &= \int_{R_Y} f_{XY}(x, y) dy \\
&= \int_0^\infty e^{-xy} dy \\
&= - \left( \frac{e^{-xy}}{x} \right)_0^\infty \\
&= \frac{1}{x}
\end{aligned}$$

 $\Rightarrow$ 

$$f_X(x) = \begin{cases} \frac{1}{x} & \text{for } 1 \leq x < e \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned}
f_Y(y) &= \int_{R_X} f_{XY}(x, y) dx \\
&= \int_1^e e^{-xy} dx \\
&= - \left( \frac{e^{-xy}}{y} \right)_1^e \\
&= \frac{1}{y} \left( \frac{1}{e^y} - \frac{1}{e^{ey}} \right)
\end{aligned}$$

 $\Rightarrow$ 

$$f_Y(y) = \begin{cases} \frac{1}{y} \left( \frac{1}{e^y} - \frac{1}{e^{ey}} \right) & \text{for } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(b)

$$P(0 \leq Y \leq 1, 1 \leq X \leq \sqrt{e}) = \int_0^1 \int_1^{\sqrt{e}} e^{-xy} dx dy$$

**Problem 18.**

(a)

$$\begin{aligned}
 f_X(x) &= \int_{R_Y} f_{XY}(x, y) dy \\
 &= \int_0^2 \left( \frac{1}{4}x^2 + \frac{1}{6}y \right) dy \\
 &= \frac{1}{2}x^2 + \frac{1}{3}
 \end{aligned}$$

 $\Rightarrow$ 

$$f_X(x) = \begin{cases} \frac{1}{2}x^2 + \frac{1}{3} & \text{for } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned}
 f_Y(y) &= \int_{R_X} f_{XY}(x, y) dx \\
 &= \int_{-1}^1 \left( \frac{1}{4}x^2 + \frac{1}{6}y \right) dx \\
 &= \frac{1}{3}y + \frac{1}{6}
 \end{aligned}$$

 $\Rightarrow$ 

$$f_Y(y) = \begin{cases} \frac{1}{3}y + \frac{1}{6} & \text{for } 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(b)

$$\begin{aligned}
 P(X > 0, Y < 1) &= \int_0^1 \int_0^1 \left( \frac{1}{4}x^2 + \frac{1}{6}y \right) dx dy \\
 &= \frac{1}{6}
 \end{aligned}$$

(c) Using inclusion-exclusion:

$$\begin{aligned}
 P(X > 0 \cup Y < 1) &= P(X > 0) + P(Y < 1) - P(X > 0, Y < 1) \\
 &= \int_0^1 \int_0^2 \left( \frac{1}{4}x^2 + \frac{1}{6}y \right) dy dx + \int_{-1}^1 \int_0^1 \left( \frac{1}{4}x^2 + \frac{1}{6}y \right) dy dx - \frac{1}{6} \\
 &= \frac{1}{2} + \frac{1}{3} - \frac{1}{6} \\
 &= \frac{2}{3}.
 \end{aligned}$$

(d)

$$\begin{aligned}
 P(X > 0 | Y < 1) &= \frac{P(X > 0, Y < 1)}{P(Y < 1)} \\
 &= \frac{\frac{1}{6}}{\frac{1}{3}} \\
 &= \frac{1}{2}.
 \end{aligned}$$

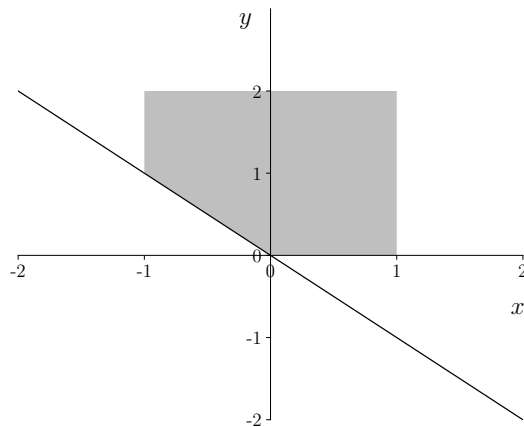


Figure 5.2: The region of integration for Problem 18 (e) (shaded region).

- (e) We must be slightly care in choosing the bounds of integration for this problem. The upper bound of the  $y$  integral is the upper bound of  $R_Y$ , and the lower bound of the  $y$  integral is  $\max\{0, -x\}$ , and not simply  $-x$ . This is because for  $x > 0$ ,  $-x < 0$ , but the lower bound of the range of  $Y$  is 0. An illustration of the domain of the double integral is shown in Fig. 6.1. The probability we seek is thus:

$$\begin{aligned}
 P(X + Y > 0) &= P(Y > -X) \\
 &= \int_{-1}^1 \int_{\max\{0, -x\}}^2 \left( \frac{1}{4}x^2 + \frac{1}{6}y \right) dy dx \\
 &= \int_{-1}^1 \left[ \frac{1}{4}x^2 y + \frac{1}{12}y^2 \right]_{\max\{0, -x\}}^2 dx \\
 &= \frac{1}{4} \int_{-1}^1 x^2 (2 - \max\{0, -x\}) dx + \frac{1}{12} \int_{-1}^1 (4 - \max\{0, -x\}^2) dx \\
 &= \frac{1}{4} \int_{-1}^0 x^2 (2 + x) dx + \frac{1}{4} \int_0^1 2x^2 dx + \frac{1}{12} \int_{-1}^0 (4 - x^2) dx + \frac{1}{12} \int_0^1 4 dx \\
 &= \frac{131}{144}.
 \end{aligned}$$

**Problem 19.**

(a)

$$\begin{aligned}
 f_{XY}(x, y) &= \frac{\partial F_{XY}}{\partial x \partial y} \\
 &= e^{-x} 2e^{-2y}
 \end{aligned}$$

$\Rightarrow$

$$f_{XY}(x, y) = \begin{cases} e^{-x} 2e^{-2y} & \text{for } x, y > 0 \\ 0 & \text{otherwise} \end{cases}$$



(b)

$$\begin{aligned}
P\left(Y > \frac{1}{2}X\right) &= \int_0^\infty \int_0^{x/2} e^{-x} 2e^{-2y} dy dx \\
&= \int_0^\infty (e^{-x} - e^{-2x}) dx \\
&= \frac{1}{2}
\end{aligned}$$

- (c)  $X$  and  $Y$  are independent because the joint PDF can be factored into the product of 2 marginal PDFs. Specifically, the joint PDF can be factored into  $f_X(x)f_Y(y)$  where  $X \sim \text{Exp}(1)$  and where  $Y \sim \text{Exp}(2)$ .

**Problem 20.**

- (a) To calculate the PDF, we simply need to condition on  $X > 0$ , and since a  $\mathcal{N}(0, 1)$  distribution is symmetric about zero, we know that  $P(X > 0) = 1/2$ :

$$\begin{aligned}
f_{X|X>0}(x) &= \frac{f_X(x)}{P(X > 0)} \\
&= \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

To find the conditional CDF, we need only to integrate the Gaussian:

$$\begin{aligned}
F_{X|X>0}(x) &= 2 \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x'^2}{2}} dx' \\
&= 2 \left[ \Phi(x) - \frac{1}{2} \right],
\end{aligned}$$

so that

$$F_{X|X>0}(x) = \begin{cases} 2\Phi(x) - 1 & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(b)

$$\begin{aligned}
E[X|X > 0] &= \int_0^\infty x f_{X|X>0}(x) dx \\
&= \frac{2}{\sqrt{2\pi}} \int_0^\infty x e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{u}{2}} du \\
&= \frac{2}{\sqrt{2\pi}}
\end{aligned}$$

- (c) We can compute  $E[X^2|X > 0]$  by noting that if  $Y \sim \mathcal{N}(0, 1)$ , then:

$$\begin{aligned}
1 &= E[Y^2] \\
&= \frac{2}{\sqrt{2\pi}} \int_0^\infty y^2 e^{-\frac{y^2}{2}} dy,
\end{aligned}$$

where I have used the fact that  $y^2$  times  $\exp(-y^2/2)$  is an even function, so I need only integrate from 0 to infinity and multiply by 2. Thus we have that

$$\begin{aligned} E[X^2|X > 0] &= \int_0^\infty x^2 f_{X|X>0}(x) dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^\infty x^2 e^{-\frac{x^2}{2}} dx \\ &= 1. \end{aligned}$$

Finally, we have that:

$$\begin{aligned} Var[X|X > 0] &= E[X^2|X > 0] - (E[X|X > 0])^2 \\ &= 1 - \left(\frac{2}{\sqrt{2\pi}}\right)^2 \\ &= \frac{\pi - 2}{\pi}. \end{aligned}$$

**Problem 21.**

(a) I first find the marginal PDF of  $Y$ :

$$\begin{aligned} f_Y(y) &= \int_{-1}^1 \left(x^2 + \frac{1}{3}y\right) dx \\ &= \frac{2}{3}(1 + y), \end{aligned}$$

so that we have

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{XY}(x, y)}{f_Y(y)} \\ &= \frac{x^2 + \frac{1}{3}y}{\frac{2}{3}(1 + y)}, \end{aligned}$$

and therefore:

$$f_{X|Y}(x|y) = \begin{cases} \frac{3x^2 + y}{2(1 + y)} & \text{for } -1 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(b)

$$\begin{aligned} P(X > 0|Y = y) &= \int_0^1 f_{X|Y}(x|y) dx \\ &= \frac{1}{2(1 + y)} \int_0^1 (3x^2 + y) dx \\ &= \frac{1}{2}. \end{aligned}$$

Notice that the probability,  $P(X > 0|Y = y)$ , does not depend on  $y$ .

(c) We have already found the marginal PDF of  $Y$ , and now I find the marginal PDF of  $X$ :

$$\begin{aligned} f_X(x) &= \int_0^1 \left(x^2 + \frac{1}{3}y\right) dy \\ &= x^2 + \frac{1}{6}. \end{aligned}$$

We thus see that  $f_X(x)f_Y(y) = 2x^2/3 + y/9 + 2yx^2/3 + 1/9 \neq f_{XY}(x, y)$ , and so  $X$  and  $Y$  are not independent.

**Problem 22.** I start by first finding the marginal PDF of  $X$ :

$$\begin{aligned} f_X(x) &= \int_0^1 \left( \frac{1}{2}x^2 + \frac{2}{3}y \right) dy \\ &= \frac{1}{2}x^2 + \frac{1}{3}, \end{aligned}$$

which is valid for  $-1 \leq x \leq 1$ , otherwise  $f_X(x) = 0$ .

I now find the PDF of  $Y$  conditioned on  $X = 0$ :

$$\begin{aligned} f_{Y|X}(y|0) &= \frac{f_{XY}(0, y)}{f_X(0)} \\ &= \frac{\frac{2}{3}y}{\frac{1}{3}} \\ &= 2y \end{aligned}$$

valid for  $0 \leq y \leq 1$ . I may now calculate  $E[Y|X = 0]$ ,

$$\begin{aligned} E[Y|X = 0] &= \int_{R_{Y|X=0}} y f_{Y|X}(y|0) dy \\ &= \int_0^1 2y^2 dy \\ &= \frac{2}{3}, \end{aligned}$$

and  $E[Y^2|X = 0]$ ,

$$\begin{aligned} E[Y^2|X = 0] &= \int_{R_{Y|X=0}} y^2 f_{Y|X}(y|0) dy \\ &= \int_0^1 2y^3 dy \\ &= \frac{1}{2}. \end{aligned}$$

Therefore, the variance is:  $Var[Y|X = 0] = E[Y^2|X = 0] - (E[Y|X = 0])^2 = 1/2 - (2/3)^2 = 1/18$ .

**Problem 23.**

(a) The set  $E$  is a diamond shaped region in  $\mathbb{R}^2$ , upper-bounded by  $1 - |x|$  and lower-bounded by  $|x| - 1$ , as shown in Fig. 5.3. The area of the region is thus 4 times the area of a triangle with a base length of 1 and height length of 1:  $4 \cdot (1/2) \cdot (1) \cdot (1) = 2$ . Since the total probability must integrate to unity, we thus have  $c = 1/2$ .

(b) The marginal PDF of  $X$  is given by

$$\begin{aligned} f_X(x) &= 2 \int_0^{1-|x|} \frac{1}{2} dy \\ &= 1 - |x|, \end{aligned}$$

so, that

$$f_X(x) = \begin{cases} 1 - |x| & \text{for } -1 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

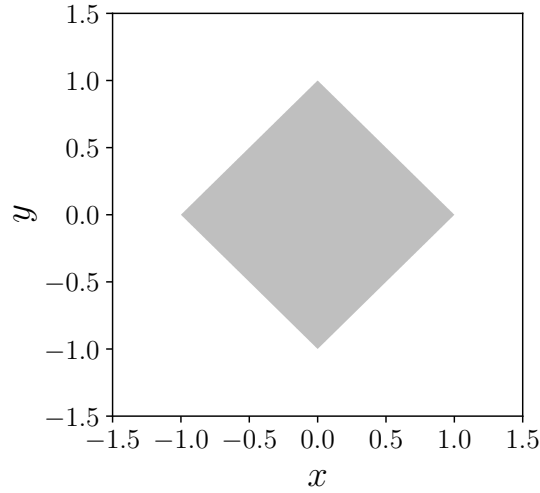


Figure 5.3: A visual representation of the set  $E$  for Problem 23.

By symmetry, we also know that

$$f_Y(y) = \begin{cases} 1 - |y| & \text{for } -1 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) The conditional PDF is given by

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{XY}(x, y)}{f_Y(y)} \\ &= \begin{cases} \frac{1}{2(1-|y|)} & \text{for } x, y \in E \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(d)  $X$  and  $Y$  are not independent, as it is clear that  $f_{X|Y}(x|y) \neq f_X(x)$ .

**Problem 24.** The marginal PDFs for  $X$  and  $Y$  are given by

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } 0 \leq x \leq 2 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f_Y(y) = \begin{cases} \frac{1}{2} & \text{for } 0 \leq y \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

I solve for the desired probability by conditioning on  $Y$  and using the fact that  $X$  and  $Y$  are

independent:

$$\begin{aligned}
 P(XY < 1) &= \int_0^2 P(XY < 1|Y = y)f_Y(y)dy \\
 &= \int_0^2 P(Xy < 1)f_Y(y)dy \\
 &= \int_0^2 P\left(X < \frac{1}{y}\right)f_Y(y)dy \\
 &= \int_0^2 \int_0^{\min\{2, 1/y\}} f_X(x)f_Y(y)dx dy \\
 &= \frac{1}{4} \int_0^2 \min\left\{2, \frac{1}{y}\right\} dy \\
 &= \frac{1}{4} \left( \int_0^{1/2} 2dy + \int_{1/2}^2 \frac{1}{y} dy \right) \\
 &= \frac{1}{4} + \frac{\ln 2}{2} \\
 &\approx 0.60.
 \end{aligned}$$

**Problem 25.** The easiest way to solve this problem will be to use the law of iterated expectations and the law of total variance. The following information will be useful: for  $X \sim \text{Exp}(1)$ ,  $E[X] = 1$ ,  $\text{Var}[X] = 1$  and  $E[X^2] = 2$ , and for  $Y|X \sim \text{Unif}(0, X)$ ,  $E[Y|X] = X/2$  and  $\text{Var}[Y|X] = X^2/12$ .

- (a) I use the law of iterated expectations, where the subscript on the first expectation denotes an expectation over  $X$ :

$$\begin{aligned}
 E[Y] &= E_X[E[Y|X]] \\
 &= E_X\left[\frac{X}{2}\right] \\
 &= \frac{1}{2}.
 \end{aligned}$$

- (b) I use the law of total variance, where the subscripts denote expectation and variance over  $X$ :

$$\begin{aligned}
 \text{Var}[Y] &= E_X[\text{Var}[Y|X]] + \text{Var}_X[E[Y|X]] \\
 &= E_X\left[\frac{X^2}{12}\right] + \text{Var}_X\left[\frac{X}{2}\right] \\
 &= \frac{5}{12}.
 \end{aligned}$$

**Problem 26.** For  $X \sim \text{Unif}(0, 1)$  we have:  $E[X] = 1/2$  and  $E[X^2] = 1/3$ .

- (a) Since  $X$  and  $Y$  are independent, the expectation of the product is the product of the expectations:

$$\begin{aligned}
 E[XY] &= E[X]E[Y] \\
 &= \frac{1}{4}.
 \end{aligned}$$

(b) Since  $X$  and  $Y$  are independent  $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ :

$$\begin{aligned} E[e^{X+Y}] &= E[e^X e^Y] \\ &= E[e^X] E[e^Y]. \end{aligned}$$

I can compute  $E[e^X]$  using LOTUS:

$$\begin{aligned} E[e^X] &= \int_0^1 e^x dx \\ &= e - 1, \end{aligned}$$

and plugging into the previous equation I find that

$$E[e^{X+Y}] = (e - 1)^2.$$

(c)

$$\begin{aligned} E[X^2 + Y^2 + XY] &= E[X^2] + E[Y^2] + E[XY] \\ &= \frac{1}{3} + \frac{1}{3} + \frac{1}{4} \\ &= \frac{11}{12} \end{aligned}$$

(d) We can compute this expectation with a 2D LOTUS over the joint distribution of  $X$  and  $Y$ . Since  $X$  and  $Y$  are independent,  $f_{XY}(x, y) = f_X(x)f_Y(y) = 1$  for  $x, y \in [0, 1]$ :

$$\begin{aligned} E[Y e^{XY}] &= \int_0^1 \int_0^1 y e^{xy} dx dy \\ &= \int_0^1 (1 - e^y) dy \\ &= e - 2. \end{aligned}$$

**Problem 27.** I first note that  $R_X = R_Y = [0, 1]$  and that  $R_Z = [0, \infty)$ . I solve for the CDF of  $Z$  by conditioning on  $X$  and using the fact that  $X$  and  $Y$  are independent:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P\left(\frac{X}{Y} \leq z\right) \\ &= P\left(Y \geq \frac{X}{z}\right) \\ &= \int_{-\infty}^{\infty} P\left(Y \geq \frac{x}{z} | X = x\right) f_X(x) dx \\ &= \int_0^1 P\left(Y \geq \frac{x}{z} | X = x\right) dx \\ &= \int_0^1 P\left(Y \geq \frac{x}{z}\right) dx, \end{aligned}$$

where in the last line I have used the fact that  $X$  and  $Y$  are independent. To solve for  $P(Y \geq X/z)$  by integrating  $F_Y(y)$  over  $y$ , some care must be taken in the limits of integration. Since  $x \in [0, 1]$

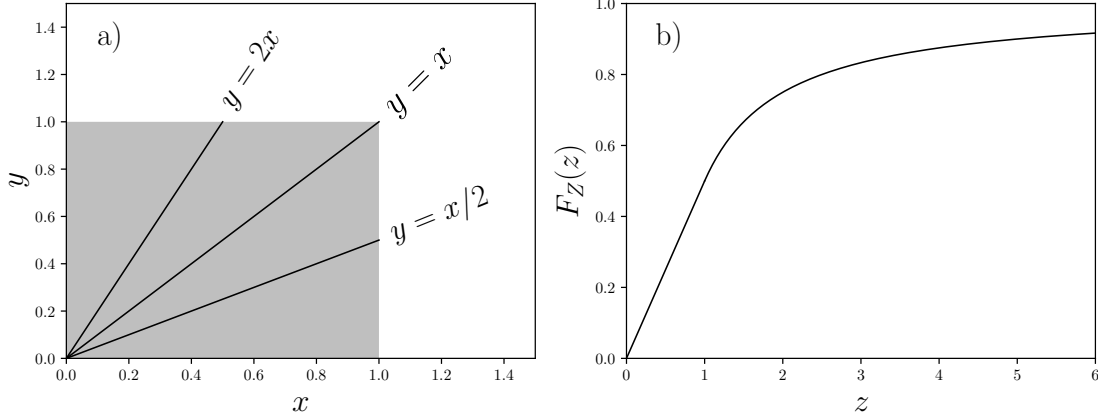


Figure 5.4: a) The  $x - y$  plane where the shaded region denotes the region of non-zero probability for the PDF at hand. b) A plot of the CDF of  $Z$ .

and  $z \in [0, \infty)$ , this implies that  $x/z \in [0, \infty)$ . However, we know that  $f_Y(y) = 0$  for  $y > 1$  and thus the lower bound of integration is not simply  $x/z$ , but is  $\min\{1, x/z\}$ . This can be seen pictorially in Fig. 5.4 a) (from a point of view of integrating over the joint PDF to solve the problem, rather than the conditional PDF), which is the  $x - y$  plane, where the grey region corresponds to the region of non-zero joint probability density. For any given  $z$ ,  $P(Y \geq X/z)$  represents the total probability mass above the the line defined by  $x/z$ . For example, I have, I have drawn 3 different lines in 5.4 a) corresponding to  $z = 1/2$  (highest line),  $z = 1$  (middle line) and  $z = 2$  (lowest line). For each of these  $z$  values,  $P(Y \geq X/z)$  is the fraction of the grey box above that line. We see that when  $z$  increases from 0 to 1 (corresponding to the vertical line along the  $y$ -axis and the line  $y = x$ ), the total probability mass above the line increases smoothly. However, due to the edge of the box, there is a kink in the total probability mass above the line when transitioning from  $z < 1$  to  $z > 1$ , which is the same reason we will get a kink in the function  $F_Z(z)$  due to  $\min\{1, x/z\}$ .

Continuing with the calculation:

$$\begin{aligned}
 F_Z(z) &= \int_0^1 P\left(Y \geq \frac{x}{z}\right) dx, \\
 &= \int_0^1 \int_{\min\{1, x/z\}}^1 f_Y(y) dy dx, \\
 &= \int_0^1 \left[1 - \min\left\{1, \frac{x}{z}\right\}\right] dx \\
 &= 1 - \int_0^1 \min\left\{1, \frac{x}{z}\right\} dx \\
 &= 1 - \left(\int_0^1 \mathbb{1}\{x \geq z\} dx + \int_0^1 \frac{x}{z} \mathbb{1}\{x < z\} dx\right),
 \end{aligned}$$

where I have picked out the proper value of the min function by utilizing an indicator function (which is much nicer to use in an integral since it “kills” the integral whenever the logical condition evaluates to false).

Thus, for  $z > 1$  we have that:

$$\begin{aligned} F_Z(z) &= 1 - \int_0^1 \frac{x}{z} dx \\ &= 1 - \frac{1}{2z}, \end{aligned}$$

while for  $z \in [0, 1]$  we have that:

$$\begin{aligned} F_Z(z) &= 1 - \left( \int_0^z \frac{x}{z} dx + \int_z^1 dx \right) \\ &= \frac{z}{2}. \end{aligned}$$

In summary we have:

$$F_Z(z) = \begin{cases} \frac{z}{2} & \text{for } 0 \leq z \leq 1 \\ 1 - \frac{1}{2z} & \text{for } z > 1 \\ 0 & \text{otherwise,} \end{cases}$$

which I plot in Fig. 5.4 b). Notice that even though this is a piecewise function, it appears very smooth because, at the transition ( $z = 1$ ), both the actual function and the first derivative match between the two piecewise regions.

To find the PDF we need only to differentiate:

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz} \\ &= \begin{cases} \frac{1}{2} & \text{for } 0 \leq z \leq 1 \\ \frac{1}{2z^2} & \text{for } z > 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

### Problem 28.

(a) To find  $f_{U|X}(x|u)$  I first solve for the conditional CDF then differentiate with respect to  $u$ :

$$\begin{aligned} F_{U|X}(u|x) &= P(U \leq u|X = x) \\ &= P(X + Y \leq u|X = x) \\ &= P(x + Y \leq u|X = x) \\ &= P(Y \leq u - x) \\ &= \Phi(u - x), \end{aligned}$$

where in the fourth line I have used the fact that  $X$  and  $Y$  are independent. To find the conditional PDF I now differentiate:

$$\begin{aligned} \frac{\partial F_{U|X}}{\partial u} &= \Phi'(u - x) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-x)^2}{2}}, \end{aligned}$$

and thus we see that:

$$U|X = x \sim \mathcal{N}(x, 1).$$



- (b) If  $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$  and  $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$  are independent, then, as shown in the book by the method of convolution,  $X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ , and thus:

$$U \sim \mathcal{N}(0, 2).$$

- (c) To find  $f_{X|U}(x|u)$ , I use Baye's rule for PDFs:

$$\begin{aligned} f_{X|U}(x|u) &= \frac{f_{U|X}(u|x)f_X(x)}{f_U(u)} \\ &= \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{(u-x)^2}{2}} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}{\frac{1}{2\sqrt{\pi}}e^{-\frac{u^2}{4}}} \\ &= \frac{1}{\sqrt{\pi}}e^{-\frac{(u-x)^2}{2} - \frac{x^2}{2} + \frac{u^2}{4}} \\ &= \frac{1}{\sqrt{\pi}}e^{-\left(x - \frac{u}{2}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}(1/\sqrt{2})}e^{-\frac{\left(x - \frac{u}{2}\right)^2}{2 \cdot \frac{1}{2}}}, \end{aligned}$$

where I have used the “completing the square” trick in the exponential to make it more Gaussian. We recognize this distribution as a normal,  $u/2$ ,  $1/2$  distribution:

$$X|U = u \sim \mathcal{N}\left(\frac{u}{2}, \frac{1}{2}\right).$$

- (d) Since  $X|U = u \sim \mathcal{N}\left(\frac{u}{2}, \frac{1}{2}\right)$ , we have:

$$E[X|U = u] = \frac{u}{2},$$

and

$$\text{Var}[X|U = u] = \frac{1}{2}.$$

**Problem 29.** This problem can be solved using the method of transformations. Since  $X$  and  $Y$  are independent, we have an axis-aligned 2D Gaussian for the joint distribution:

$$\begin{aligned} f_{XY}(x, y) &= f_X(x)f_Y(y) \\ &= \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}. \end{aligned}$$

I define the functions  $h_1$  and  $h_2$  as:

$$\begin{cases} X = h_1(R, \Theta) = R \cos \Theta \\ Y = h_2(R, \Theta) = R \sin \Theta, \end{cases}$$

so that, according to the method of transformations:

$$f_{R\Theta}(r, \theta) = f_{XY}(h_1(r, \theta), h_2(r, \theta)) \left| \begin{array}{cc} \frac{\partial h_1}{\partial r} & \frac{\partial h_1}{\partial \theta} \\ \frac{\partial h_2}{\partial r} & \frac{\partial h_2}{\partial \theta} \end{array} \right|.$$

The Jacobian is easy to calculate:

$$\begin{aligned} J &= \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} \\ &= r(\cos^2 \theta + \sin^2 \theta) \\ &= r, \end{aligned}$$

where I have used the Pythagorean trigonometric identity. Thus, we have that:

$$\begin{aligned} f_{R\Theta}(r, \theta) &= \frac{1}{2\pi} e^{-\frac{(r \cos \theta)^2}{2}} e^{-\frac{(r \sin \theta)^2}{2}} r \\ &= \frac{r}{2\pi} e^{-\frac{1}{2}r^2}, \end{aligned}$$

where I have again used the Pythagorean trigonometric identity.

If  $R$  and  $\Theta$  are independent, then we can factor  $f_{R\Theta}(r, \theta)$  into  $f_R(r)f_\Theta(\theta)$ . To help determine what these 2 functions are, I integrate the joint distribution:

$$\begin{aligned} 1 &= \int_0^\infty \int_{-\pi}^\pi f_{R\Theta}(r, \theta) dr d\theta \\ &= \int_0^\infty \int_{-\pi}^\pi \frac{r}{2\pi} e^{-\frac{1}{2}r^2} dr d\theta \\ &= \int_0^\infty r e^{-\frac{1}{2}r^2} dr. \end{aligned}$$

Thus, we see that the function  $re^{-\frac{1}{2}r^2}$ , which only depends on  $r$ , is always positive (for  $r \geq 0$ ) and integrates to 1. This is the marginal distribution of  $R$ . The function  $1/(2\pi)$  is always positive and integrates to 1 for  $\theta \in [-\pi, \pi]$ , and this is thus the marginal distribution of  $\Theta$ . Therefore we see that  $f_{R\Theta}(r, \theta)$  can be factored into  $f_R(r)f_\Theta(\theta)$ , and thus  $R$  and  $\Theta$  are independent.

**Problem 30.** If  $X, Y \stackrel{iid}{\sim} Unif(0, 1)$ , then:

$$F_{XY}(x, y) = \begin{cases} 1 & \text{for } x, y \in [0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

The Jacobian has already been calculated in the previous problem ( $J = r$ ), so that

$$\begin{aligned} f_{R\Theta}(r, \theta) &= rF_{XY}(r \cos \theta, r \sin \theta) \\ &= \begin{cases} r & \text{for } r \cos \theta, r \sin \theta \in [0, 1] \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where in Fig. 5.5, I have indicated where in the  $r - \theta$  plane  $f_{R\Theta}(r, \theta)$  is non-zero.

We can further examine the constraints  $r \cos \theta, r \sin \theta \in [0, 1]$  to gain more insight. Satisfying these conditions is equivalent to simultaneously satisfying the following four conditions:

$$\begin{cases} r \cos \theta \leq 1 \\ r \sin \theta \leq 1 \\ r \cos \theta \geq 0 \\ r \sin \theta \geq 0. \end{cases}$$

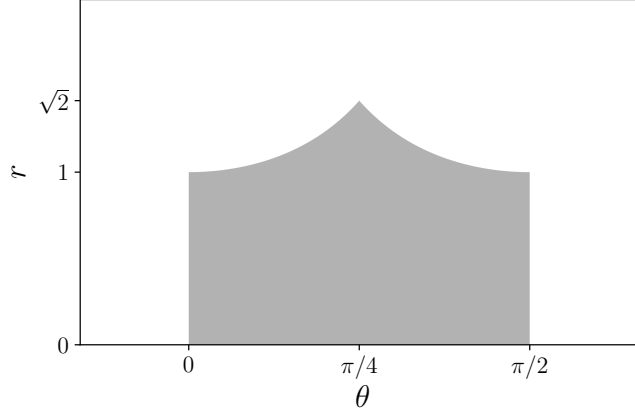


Figure 5.5: The  $r - \theta$  plane for Problem 30. The grey region denotes the region in the plane where  $f_{R\Theta}(r, \theta)$  is non-zero.

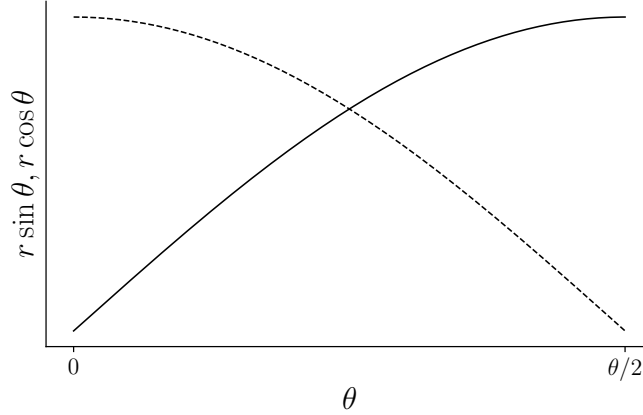


Figure 5.6:  $r \sin \theta$  (solid line) and  $r \cos \theta$  (dashed line) for  $0 \leq \theta \leq \pi/2$

Since  $r$  is always positive, the last 2 conditions yield  $\cos \theta \geq 0$  and  $\sin \theta \geq 0$ , which only happens in the first quadrant, i.e.,  $0 \leq \theta \leq \pi/2$ . If we plot the first 2 conditions for  $0 \leq \theta \leq \pi/2$ , as in Fig. 5.6, we see that when  $0 \leq \theta \leq \pi/4$ ,  $r \cos \theta \leq 1 \implies r \sin \theta \leq 1$  and when  $\pi/4 \leq \theta \leq \pi/2$ ,  $r \sin \theta \leq 1 \implies r \cos \theta \leq 1$ .

Thus, we can re-write  $f_{R\Theta}(r, \theta)$  with the constraints specified a little more explicitly:

$$f_{R\Theta}(r, \theta) = \begin{cases} r & \text{for } 0 \leq \theta \leq \frac{\pi}{4} \text{ and } r \leq \frac{1}{\cos \theta} \\ r & \text{for } \frac{\pi}{4} \leq \theta \leq \frac{\pi}{2} \text{ and } r \leq \frac{1}{\sin \theta} \\ 0 & \text{otherwise,} \end{cases}$$

(where the inequalities did not flip when I divided by  $\cos \theta$  and  $\sin \theta$  because these are both positive in the first quadrant). Note that these constraints imply that  $f_{R\Theta}(r, \theta) > 0$  in the unit square and  $f_{R\Theta}(r, \theta) = 0$  outside of the unit square, as we would expect for  $X, Y \sim \text{Unif}(0, 1)$  as shown in Fig. 5.7. The figure shows the unit square in the  $x - y$  plane (where the probability is non-zero), and it shows that for  $\theta$  less than  $\pi/4$ ,  $r$  is constrained by  $0 \leq r \leq 1/\cos \theta$  (for values of  $r$  within the unit square). One can similarly show in this figure that for  $\theta$  greater than  $\pi/4$ ,  $0 \leq r \leq 1/\sin \theta$ .

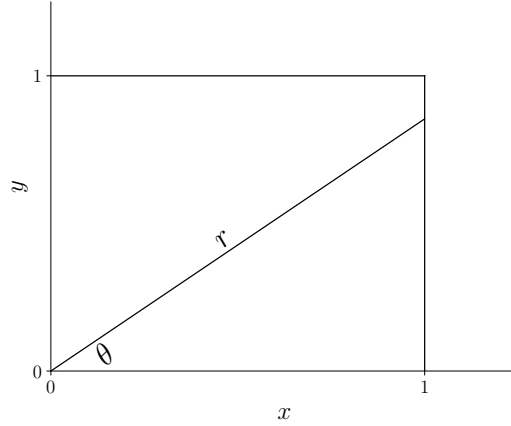


Figure 5.7: The unit square in the  $x - y$  plane. In polar coordinates, to be within the unit square, it can be seen geometrically that  $r$  is constrained by  $0 \leq r \leq 1/\cos \theta$  for  $0 \leq \theta \leq \pi/4$ , and by  $0 \leq r \leq 1/\sin \theta$  for  $\pi/4 < \theta \leq \pi/2$ .

We can check explicitly that this PDF integrates to 1:

$$\begin{aligned} \int_0^{\pi/4} \int_0^{\frac{1}{\cos \theta}} r dr d\theta + \int_{\pi/4}^{\pi/2} \int_0^{\frac{1}{\sin \theta}} r dr d\theta &= \frac{1}{2} \int_0^{\pi/4} \frac{1}{\cos^2 \theta} d\theta + \frac{1}{2} \int_{\pi/4}^{\pi/2} \frac{1}{\sin^2 \theta} d\theta \\ &= \frac{1}{2} \tan \theta \Big|_0^{\pi/4} - \frac{1}{2} \left( \frac{\cos \theta}{\sin \theta} \right) \Big|_{\pi/4}^{\pi/2} \\ &= 1. \end{aligned}$$

For this problem  $R$  and  $\Theta$  are not independent, because there is no way to factor  $f_{R\Theta}(r, \theta)$  into an equation of just  $r$  times an equation of just  $\theta$ , since the values of  $r$  over which the PDF is non-zero explicitly depend on the values of  $\theta$ .

**Problem 31.** The covariance can be computed straight from its definition:

$$\begin{aligned} \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] \\ &= \sum_{x=0}^1 \sum_{y=0}^2 xy P_{XY}(x, y) - \left( \sum_{x=0}^1 x P_X(x) \right) \left( \sum_{y=0}^2 y P_Y(y) \right) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} - \left[ 1 \cdot \left( \frac{1}{8} + \frac{1}{6} + \frac{1}{6} \right) \right] \left[ 1 \cdot \left( \frac{1}{4} + \frac{1}{6} \right) + 2 \cdot \left( \frac{1}{8} + \frac{1}{6} \right) \right] \\ &= \frac{1}{24}. \end{aligned}$$

To calculate  $\rho_{XY}$ , I first calculate the variances. For  $X$  we have

$$\begin{aligned} E[X] &= \sum_{x=0}^1 x P_X(x) = 1 \cdot \left( \frac{1}{8} + \frac{1}{6} + \frac{1}{6} \right) = \frac{11}{24}, \\ E[X^2] &= \sum_{x=0}^1 x^2 P_X(x) = 1 \cdot \left( \frac{1}{8} + \frac{1}{6} + \frac{1}{6} \right) = \frac{11}{24} \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} \text{Var}[X] &= E[X^2] - E[X]^2 \\ &= \frac{11}{24} - \left(\frac{11}{24}\right)^2 \\ &= \frac{143}{576}, \end{aligned}$$

and for  $Y$  we have

$$\begin{aligned} E[Y] &= \sum_{y=0}^2 y P_Y(y) = 1 \cdot \left(\frac{1}{4} + \frac{1}{6}\right) + 2 \cdot \left(\frac{1}{8} + \frac{1}{6}\right) = 1, \\ E[Y^2] &= \sum_{y=0}^2 y^2 P_Y(y) = 1 \cdot \left(\frac{1}{4} + \frac{1}{6}\right) + 2^2 \cdot \left(\frac{1}{8} + \frac{1}{6}\right) = \frac{19}{12} \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} \text{Var}[Y] &= E[Y^2] - E[Y]^2 \\ &= \frac{19}{12} - 1^2 \\ &= \frac{7}{12}. \end{aligned}$$

Finally, the correlation is:

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}} = \frac{\frac{1}{24}}{\sqrt{\frac{143}{576} \cdot \frac{7}{12}}} \approx 0.11. \quad (5.2)$$

Thus, there is a weak, positive correlation between  $X$  and  $Y$ .

**Problem 32.** We can use several of the items in Lemma 5.3 in the book to solve this problem:

$$\begin{aligned} \text{Cov}[Z, W] &= \text{Cov}[11 - X + X^2Y, 3 - Y] \\ &= \text{Cov}[-X + X^2Y, -Y] \\ &= \text{Cov}[X, Y] - \text{Cov}[X^2Y, Y] \\ &= -\text{Cov}[X^2Y, Y] \\ &= -(E[X^2Y^2] - E[X^2Y]E[Y]) \\ &= -(E[X^2]E[Y^2] - E[X^2]E[Y]^2) \\ &= -(1 - 0) \\ &= -1, \end{aligned}$$

where in the second line I have used item 5 of Lemma 5.3, in the third item 7, in the fourth item 2 in the sixth I have used the fact that  $X$  and  $Y$  are independent and in the seventh I have used the fact that  $X, Y \sim \mathcal{N}(0, 1)$ .

**Problem 33.** To solve this problem I use several of the items in Lemma 5.3 in the book. Since  $Z$  and  $W$  are independent, we have that:

$$\begin{aligned} 0 &= \text{Cov}[Z, W] \\ &= \text{Cov}[2X - Y, X + Y] \\ &= 2\text{Cov}[X, X] + 2\text{Cov}[X, Y] - \text{Cov}[Y, X] - \text{Cov}[Y, Y] \\ &= 2\text{Var}[X] + \text{Cov}[X, Y] - \text{Var}[Y] \\ &= 2 \cdot 4 + \text{Cov}[X, Y] - 9 \end{aligned}$$

$\Rightarrow$

$$\text{Cov}[X, Y] = 1.$$

The correlation is now straightforward to calculate:

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}} = \frac{1}{\sqrt{4 \cdot 9}} = \frac{1}{6}.$$

**Problem 34.** We know that  $X \sim \text{Unif}(1, 3)$  (so that  $E[X] = 2$ ) and  $Y|X = x \sim \text{Exp}(x)$  (so that  $E[Y|X] = 1/x$ ). Since  $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$ , and since we know the distribution of  $Y|X = x$  we can probably solve most of the expectations by conditioning on  $X$  and using the law of iterated expectations. To solve for  $E[Y]$  I use the law of iterated expectations (where the subscript  $X$  denotes an expectation over the random variable  $X$ ):

$$\begin{aligned} E[Y] &= E_X[E[Y|X]] \\ &= E_X\left[\frac{1}{X}\right] \\ &= \frac{1}{2} \int_1^3 \frac{1}{x} dx \\ &= \frac{1}{2} \ln 3. \end{aligned}$$

To solve for  $E[XY]$  I also condition on  $X$  and “take out what is known”:

$$\begin{aligned} E[XY] &= E_X[E[XY|X]] \\ &= E_X[XE[Y|X]] \\ &= E_X\left[X \frac{1}{X}\right] \\ &= 1. \end{aligned}$$

Thus, we have that the covariance is:

$$\begin{aligned} \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] \\ &= 1 - 2 \cdot \frac{1}{2} \ln 3 \\ &= 1 - \ln 3. \end{aligned}$$

**Problem 35.** The covariance is:

$$\begin{aligned} \text{Cov}[Z, W] &= \text{Cov}[7 + X + Y, 1 + Y] \\ &= \text{Cov}[X + Y, Y] \\ &= \text{Cov}[X, Y] + \text{Var}[Y] \\ &= 0 + 1 \\ &= 1, \end{aligned}$$

where I have used the fact that  $\text{Cov}[X, Y] = 0$  since  $X$  and  $Y$  are independent. Calculating the variances is easy as well:

$$\text{Var}[Z] = \text{Var}[7 + X + Y] = \text{Var}[X] + \text{Var}[Y] = 2,$$

and

$$\text{Var}[W] = \text{Var}[1 + Y] = \text{Var}[Y] = 1.$$

Thus we have that the correlation is:

$$\rho_{ZW} = \frac{\text{Cov}[Z, W]}{\sqrt{\text{Var}[Z]\text{Var}[W]}} = \frac{1}{\sqrt{2}} \approx 0.71.$$

**Problem 36.**

(a)

$$X + 2Y \sim \mathcal{N}(\mu_X + 2\mu_Y, \sigma_X^2 + 4\sigma_Y^2 + 2 \cdot 2\rho\sigma_X\sigma_Y) = \mathcal{N}(1, 4)$$

$\Rightarrow$

$$P(X + 2Y \leq 3) = \Phi\left(\frac{3-1}{2}\right) = \Phi(1) \approx 0.84.$$

(b)

$$\begin{aligned} \text{Cov}[X - Y, X + 2Y] &= \text{Cov}[X, X] + 2\text{Cov}[X, Y] - \text{Cov}[X, Y] - 2\text{Cov}[Y, Y] \\ &= \sigma_X^2 + \rho\sigma_X\sigma_Y - 2\sigma_Y^2 \\ &= 1 \end{aligned}$$

**Problem 37.**

(a)

$$X + 2Y \sim \mathcal{N}(\mu_X + 2\mu_Y, \sigma_X^2 + 4\sigma_Y^2 + 2 \cdot 2\rho\sigma_X\sigma_Y) = \mathcal{N}(3, 8)$$

$\Rightarrow$

$$\begin{aligned} P(X + 2Y > 4) &= 1 - P(X + 2Y \leq 4) \\ &= 1 - \Phi\left(\frac{4-3}{\sqrt{8}}\right) \\ &= 1 - \Phi\left(\frac{1}{2\sqrt{2}}\right) \\ &\approx 0.36. \end{aligned}$$

(b) Since  $X$  and  $Y$  are uncorrelated, jointly normal random variables they are independent, and thus:

$$\begin{aligned} E[X^2Y^2] &= E[X^2]E[Y^2] \\ &= (\text{Var}[X] + E[X]^2)(\text{Var}[Y] + E[Y]^2) \\ &= (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) \\ &= 10. \end{aligned}$$

**Problem 38.**

(a)  $X$  and  $Y$  are jointly normal random variables, and thus by Theorem 5.4 in the book:

$$\begin{aligned} Y|X = x &\sim \mathcal{N}\left(\mu_Y + \rho\sigma_Y \frac{x - \mu_X}{\sigma_X}, (1 - \rho^2)\sigma_Y^2\right) \\ &= \mathcal{N}\left(1 - \frac{3(x-2)}{4}, \frac{27}{4}\right). \end{aligned}$$

We therefore can immediately read off that

$$E[Y|X = 3] = 1 - \frac{3(3-2)}{4} = \frac{1}{4}.$$

(b) Using the same distribution:

$$\text{Var}[Y|X = 2] = \frac{27}{4}.$$

(c) To solve for this problem, I define the random variables  $U = X + 2Y$  and  $V = X + Y$  and I solve for the distribution of  $U|V$ . Since  $X$  and  $Y$  are jointly normal random variables, so too are  $U$  and  $V$  (since  $aU + bV = aX + 2aY + bX + bY = (a+b) \cdot X + (2a+b) \cdot Y$  which we know is normal for all  $a, b$ ) and thus Theorem 5.4 in the book gives an equation for the distribution of  $U|V$ . But first to use this formula, we will need to explicitly compute the distributions of  $U$  and  $V$ . The distribution of  $U$  is

$$\begin{aligned} U &\sim \mathcal{N}(\mu_X + 2\mu_Y, \sigma_X^2 + 4\sigma_Y^2 + 2 \cdot 2\rho\sigma_X\sigma_Y) \\ &= \mathcal{N}(4, 28), \end{aligned}$$

and the distribution of  $V$  is

$$\begin{aligned} V &\sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y) \\ &= \mathcal{N}(3, 7). \end{aligned}$$

I will also need to compute  $\rho_{UV}$ , so I here solve for  $\text{Cov}[U, V]$ :

$$\begin{aligned} \text{Cov}[U, V] &= \text{Cov}[X + 2Y, X + Y] \\ &= \text{Cov}[X, X] + \text{Cov}[X, Y] + 2\text{Cov}[Y, X] + 2\text{Cov}[Y, Y] \\ &= \sigma_X^2 + 3\rho\sigma_X\sigma_Y + 2\sigma_Y^2 \\ &= 13. \end{aligned}$$

Thus, we have that:

$$\rho_{UV} = \frac{\text{Cov}[U, V]}{\sqrt{\text{Var}[U]\text{Var}[V]}} = \frac{13}{\sqrt{28 \cdot 7}} = \frac{13}{\sqrt{196}},$$

which, as a sanity check is in between -1 and 1.

Putting everything together into Theorem 5.4 in the book we have:

$$\begin{aligned} U|V = 3 &\sim \mathcal{N}\left(\mu_U + \rho_{UV}\sigma_U \frac{3 - \mu_V}{\sigma_V}, (1 - \rho_{UV}^2)\sigma_U^2\right) \\ &= \mathcal{N}\left(4, \left(1 - \frac{13^2}{196}\right) \cdot 28\right) \\ &= \mathcal{N}\left(4, \frac{27}{7}\right). \end{aligned}$$

Finally, now that I have the distribution, I can compute the desired probability:

$$\begin{aligned} P(X + 2Y \leq 5 | X + Y = 3) &= P(U \leq 5 | V = 3) \\ &= \Phi\left(\frac{5 - 4}{\sqrt{\frac{27}{7}}}\right) \\ &= \Phi\left(\sqrt{\frac{7}{27}}\right) \\ &\approx 0.69. \end{aligned}$$



## Chapter 6

# Methods for More Than Two Random Variables

**Problem 1.**

(a)

$$\begin{aligned}
f_{XY}(x, y) &= \int f_{XYZ}(x, y, z) dz \\
&= \begin{cases} \int_0^1 (x + y) dz & \text{for } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} x + y & \text{for } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

(b)

$$\begin{aligned}
f_X(x) &= \int f_{XY}(x, y) dy \\
&= \begin{cases} \int_0^1 (x + y) dy & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} x + \frac{1}{2} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

(c) First note that:

$$\begin{aligned}
f_Z(z) &= \int f_{XYZ}(x, y, z) dx dy \\
&= \begin{cases} \int_0^1 \int_0^1 (x + y) dx dy & \text{for } 0 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} 1 & \text{for } 0 \leq z \leq 1 \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}$$

so that

$$f_{XY|Z}(x, y|z) = \frac{f_{XYZ}(x, y, z)}{f_Z(z)} = f_{XYZ}(x, y, z).$$

(d)

$$f_{XY|Z}(x, y|z) = f_{XYZ}(x, y, z) = f_{XY}(x, y)$$

 $\implies X$  and  $Y$  are independent  $Z$ **Problem 2.**Since  $X, Y, Z$  are independent,  $f_{XY|Z}(x, y|1) = f_{XY}(x, y) = f_X(x)f_Y(y)$ , so that:

$$E[XY|Z = 1] = E[XY] = E[X]E[Y] = 0,$$

and

$$E[X^2Y^2Z^2|Z = 1] = E[X^2Y^2] = E[X^2]E[Y^2] = 1.$$

**Problem 3.** To solve this problem, I first state a general result for a multivariate normal. Suppose that

$$\begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{bmatrix} \right),$$

where  $\mathbf{X}_A, \boldsymbol{\mu}_A \in \mathbb{R}^m$ ,  $\mathbf{X}_B, \boldsymbol{\mu}_B \in \mathbb{R}^n$ ,  $\boldsymbol{\Sigma}_{AA} \in \mathbb{R}^{m \times m}$ ,  $\boldsymbol{\Sigma}_{BB} \in \mathbb{R}^{n \times n}$ ,  $\boldsymbol{\Sigma}_{AB} \in \mathbb{R}^{m \times n}$  and  $\boldsymbol{\Sigma}_{BA} = \boldsymbol{\Sigma}_{AB}^T$ . (Note that, here, I have written this vector equation out in the so-called “partitioned form” for convenience.) Then, it is not difficult to show that<sup>1</sup>:

$$\mathbf{X}_A | \mathbf{X}_B = \mathbf{x}_B \sim \mathcal{N}(\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{x}_B - \boldsymbol{\mu}_B), \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} \boldsymbol{\Sigma}_{BA}).$$

I now define the random variable,  $U = Y + Z$ , solve for the joint PDF of  $X, Y, U$ , then condition on  $U$  using the formula above so that I may compute  $E[XY | Y + Z = 1]$ . Since  $U = Y + Z$ , and  $Y$  and  $Z$  are 2 independent  $\mathcal{N}(1, 1)$  distributions, then  $U \sim \mathcal{N}(2, 2)$ . Recall that for 3 marginally normal distributions, the joint distribution is:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} E[X_1] \\ E[X_2] \\ E[X_3] \end{bmatrix}, \begin{bmatrix} \text{Var}[X_1] & \text{Cov}[X_1, X_2] & \text{Cov}[X_1, X_3] \\ \text{Cov}[X_2, X_1] & \text{Var}[X_2] & \text{Cov}[X_2, X_3] \\ \text{Cov}[X_3, X_1] & \text{Cov}[X_3, X_2] & \text{Var}[X_3] \end{bmatrix} \right).$$

Thus, to solve for the joint distribution of  $X, Y, U$ , all that is left to do is to calculate the covariance terms involving  $U$ ,  $\text{Cov}[U, X] = \text{Cov}[Y + Z, X] = 0$  and  $\text{Cov}[U, Y] = \text{Cov}[Y + Z, Y] = \text{Var}[Y]$ , so the the joint distribution is:

$$\begin{bmatrix} X \\ Y \\ U \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \right).$$

To solve for the distribution of  $X, Y | U = 1$ , it is not difficult to identify that, here,  $\boldsymbol{\Sigma}_{AA} = I_2$  (the  $2 \times 2$  identity matrix),  $\boldsymbol{\Sigma}_{AB} = [0, 1]^T$ ,  $\boldsymbol{\Sigma}_{BA} = [0, 1]$ , and  $\boldsymbol{\Sigma}_{BB} = 2$ , where I identify  $\mathbf{X}_A$  with  $[X, Y]^T$  and  $\mathbf{X}_B$  with  $U$ . The mean of the conditional distribution is therefore given by

$$\boldsymbol{\mu}_{A|B} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{2}(1 - 2) = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix},$$

and the covariance matrix of the conditional distribution is given by:

$$\boldsymbol{\Sigma}_{A|B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{2} [0, 1] = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Finally, I have the conditional distribution I desire:

$$X, Y | U = 1 \sim \mathcal{N} \left( \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \right).$$

We would like to solve for  $E[XY | U = 1]$ , which we can get easily from the covariance term of the above distribution:  $E[XY | U = 1] = \text{Cov}[X, Y | U = 1] + E[X | U = 1]E[Y | U = 1] = 0 + (1) \cdot (1/2) = 1/2$ .

**Problem 4.** Due to the symmetry of the problem,  $Y_1, Y_2, \dots, Y_n$  are all identically distributed, and thus:

$$E[Y] = E[Y_1 + Y_2 + \dots + Y_n] = E[Y_1] + E[Y_2] + \dots + E[Y_n] = nE[Y_1],$$

<sup>1</sup>for example, see: Bishop, Christopher M. Pattern Recognition and Machine Learning. Springer, 2006.

and

$$Var[Y] = Var[Y_1 + Y_2 + \dots + Y_n] = \sum_{i=1}^n Var[Y_i] + 2 \sum_{i < j} Cov[Y_i, Y_j] = nVar[Y_1] + 2nCov[Y_1, Y_2].$$

The reason there is a factor of  $n$  in front of the covariance term can be seen from the matrix below (an  $n = 5$  example). The summation over  $i < j$  means that we can only consider pairs of  $i, j$  above the diagonal (shaded cells). Further, it is evident that  $Y_i, Y_j$  pairs that are 2 or more apart are independent. For example,  $Y_2 = X_2X_3$ ,  $Y_3 = X_3X_4$ ,  $Y_4 = X_4X_5$ , so that  $Y_2$  and  $Y_3$  are not independent because they share the  $X_3$  random variable, but  $Y_2$  and  $Y_4$  are independent because they share no random variables (and all the  $X_i$ s are independent). Since independence  $\implies$  no covariance, the only  $Y_i, Y_j$  pairs that contribute to the sum are the  $n - 1$  terms right above the diagonal (as indicated by the spades in the figure). We also cannot forget that the  $Y_n, Y_1$  pair is not independent because they share the  $X_1$  random variable (the spade in the upper right hand corner). We thus see that there are  $n$  pairs that contribute to the sum.

	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$
$Y_1$					♠
$Y_2$			♠		
$Y_3$				♠	
$Y_4$					♠
$Y_5$					

It remains to compute  $E[Y_1]$ ,  $Var[Y_1]$  and  $Cov[Y_1, Y_2]$ . Since  $Y_1 = X_1X_2$ , and  $X_1, X_2 \stackrel{iid}{\sim} Bern(p)$ , the range of  $Y_1$  is  $\{0, 1\}$ , with probability  $p^2$  of obtaining 1 ( $X_1 = 1$  and  $X_2 = 1$ ). In other words,  $Y_1 \sim Bern(p^2)$ , so that  $E[Y_1] = p^2$  and  $Var[Y_1] = p^2(1 - p^2)$ . All that is left to do is to compute the covariance:

$$\begin{aligned}
 Cov[Y_1, Y_2] &= E[Y_1Y_2] - E[Y_1]E[Y_2] \\
 &= E[X_1X_2X_2X_3] - E[X_1X_2]E[X_2X_3] \\
 &= E[X_1]E[X_2^2]E[X_3] - E[X_1]E[X_2]^2E[X_3] \\
 &= p \cdot p \cdot p - p \cdot p^2 \cdot p \\
 &= p^3(1 - p),
 \end{aligned}$$

where in the second line I have used the fact that all the  $X$ s are independent, and in the fourth line I have used the fact that for a  $Bern(p)$  distribution,  $p(1 - p) = E[X^2] - p^2$ .

Thus, we have that:

$$E[Y] = np^2$$

and

$$\begin{aligned}
 Var[Y] &= np^2(1 - p^2) + 2np^3(1 - p) \\
 &= np^2(3p + 1)(1 - p).
 \end{aligned}$$

**Problem 5.**

(a) To solve for the expectation, note that:

$$\begin{aligned}
 E[X] &= E[X_1 + X_2 + \dots + X_k] \\
 &= E[X_1] + E[X_2] + \dots + E[X_k] \\
 &= \sum_{j=0}^1 jP(X_1 = j) + \sum_{j=0}^1 jP(X_2 = j) + \dots + \sum_{j=0}^1 jP(X_k = j) \\
 &= P(X_1 = 1) + P(X_2 = 1) + \dots + P(X_k = 1),
 \end{aligned}$$

where in the second line I have used the linearity of expectation. To solve this problem, we therefore need to solve for  $P(X_i = 1)$  for all  $i = 1, 2, \dots, k$ .

To solve for  $P(X_i = 1)$ , first suppose that we draw all  $b+r$  balls and create a specific sequence of blues and reds. Note that all possible sequences are equally likely to occur. To see this, as an example, suppose  $r = 3$  and  $b = 2$  and we draw the sequence  $RRBRB$ . The probability that this occurs is

$$\frac{3}{3+2} \cdot \frac{3-1}{3+2-1} \cdot \frac{2}{3+2-2} \cdot \frac{3-2}{3+2-3} \cdot \frac{2-1}{3+2-4} = \frac{3!2!}{(3+2)!}.$$

Suppose instead we had drawn the sequence  $BRRBR$ . The probability that this sequence occurs is:

$$\frac{2}{3+2} \cdot \frac{3}{3+2-1} \cdot \frac{3-1}{3+2-2} \cdot \frac{2-1}{3+2-3} \cdot \frac{3-2}{3+2-4} = \frac{3!2!}{(3+2)!}.$$

Notice that since the probability of all possible sequences with  $r = 3$  and  $b = 2$  is simply the product of the same terms in the numerators and the same terms in the denominators, but in a different order, the probability of any possible sequence is the same product. Thus, in general, the probability of any specific sequence occurring is  $r!b!/(b+r)!$ . As a check that all possible sequences are equally as likely and the probability of each sequence is  $r!b!/(b+r)!$ , if we multiply this probability by the total number of distinct sequences the result should be 1. Taking into account the indistinguishability of all red balls and all blue balls, the total number of distinct sequences is  $(b+r)!/(r!b!)$ . Multiplying these values together indeed results in 1.

Now that we know that all outcomes are equally as likely and that there is a finite sample space, we can use combinatorics to find the probabilities we are after. Let us concentrate on the  $i^{th}$  draw and compute the probability that the  $i^{th}$  draw is blue. Since all sequences are equally likely to occur, to compute this probability, we need only to divide the total number of unique sequences with a blue ball in the  $i^{th}$  spot by the total number of possible unique sequences. Since all red balls are indistinguishable and all blue balls are indistinguishable, as above, the total number of unique sequences is thus  $(b+r)!/(r!b!)$ . The total number of unique sequences with a blue ball in the  $i^{th}$  spot is  $(b+r-1)!/[r!(b-1)!]$ , and therefore the probability of obtaining a blue ball in the  $i^{th}$  spot,  $P(X_i = 1)$ , is:

$$P(X_i = 1) = \frac{\frac{(b+r-1)!}{r!(b-1)!}}{\frac{(b+r)!}{r!b!}} = \frac{b}{b+r}.$$

The desired expectation value is therefore

$$\begin{aligned}
 E[X] &= P(X_1 = 1) + P(X_2 = 1) + \dots + P(X_k = 1) \\
 &= \underbrace{\frac{b}{b+r} + \frac{b}{b+r} + \dots + \frac{b}{b+r}}_{k \text{ times}} \\
 &= \frac{kb}{b+r}.
 \end{aligned}$$

(b) To solve for  $Var[X]$ , I already have  $E[X]$ , so I just need to solve for  $E[X^2]$ :

$$\begin{aligned} E[X^2] &= E[(X_1 + X_2 + \dots + X_k)^2] \\ &= E\left[\sum_{i=1}^k X_i^2 + \sum_{i,j:i \neq j} X_i X_j\right] \\ &= \sum_{i=1}^k E[X_i^2] + \sum_{i,j:i \neq j} E[X_i X_j], \end{aligned}$$

where in the third line I have used the linearity of expectation, and where the notation  $\sum_{i,j:i \neq j}$  refers to a summation over all  $i, j$  pairs ( $i, j=1, 2, \dots, k$ ) except the pairs for which  $i = j$  (which was accounted for in the first summation).

We have already pretty much solved for  $E[X_i^2]$ :

$$\begin{aligned} E[X_i^2] &= \sum_{j=0}^1 j^2 P(X_i = j) \\ &= P(X_i = 1) \\ &= \frac{b}{b+r}, \end{aligned}$$

so that the first summation is:

$$\sum_{i=1}^k E[X_i^2] = \frac{kb}{b+r}.$$

The second summation is slightly more difficult. The strategy I take is to condition on one of the random variables and to use the law of total expectation (since  $X_i, X_j \in \{0, 1\}$ , so that one of the terms will go to zero):

$$\begin{aligned} E[X_i X_j] &= E[X_i X_j | X_j = 0]P(X_j = 0) + E[X_i X_j | X_j = 1]P(X_j = 1) \\ &= E[X_i | X_j = 1]P(X_j = 1) \\ &= \left[ \sum_{l=0}^1 l P(X_i = l | X_j = 1) \right] P(X_j = 1) \\ &= P(X_i = 1 | X_j = 1)P(X_j = 1), \end{aligned}$$

I thus need to solve for  $P(X_i = 1 | X_j = 1)$ , which I can do in a very similar combinatorial fashion as I did for  $P(X_i = 1)$ . For this probability, the total sample space is all sequences of size  $k$  with  $b$  blue balls and  $r$  red balls, with a blue ball in the  $j^{th}$  spot. The size of the sample space is thus  $(b+r-1)!/[(b-1)!r!]$ . The number of unique sequences with a blue ball in the  $j^{th}$  spot and a blue ball in the  $i^{th}$  spot is  $(b+r-2)!/[r!(b-2)!]$ , so that

$$P(X_i = 1 | X_j = 1) = \frac{\frac{(b+r-2)!}{r!(b-2)!}}{\frac{(b+r-1)!}{(b-1)!r!}} = \frac{b-1}{b+r-1}.$$

Finally, the second summation is

$$\begin{aligned}
 \sum_{i,j:i \neq j} E[X_i X_j] &= \sum_{i,j:i \neq j} P(X_i = 1 | X_j = 1) P(X_j = 1) \\
 &= \sum_{i,j:i \neq j} \frac{b-1}{b+r-1} \frac{b}{b+r} \\
 &= \frac{(k^2 - k)(b-1)}{b+r-1} \frac{b}{b+r},
 \end{aligned}$$

so that

$$\begin{aligned}
 E[X^2] &= \frac{kb}{b+r} + \frac{(k^2 - k)(b-1)}{b+r-1} \frac{b}{b+r} \\
 &= \frac{kbr + k^2 b^2 - bk^2}{(b+r)(b+r-1)}.
 \end{aligned}$$

The variance is thus:

$$\begin{aligned}
 Var[X] &= E[X^2] - E[X]^2 \\
 &= \frac{kbr + k^2 b^2 - bk^2}{(b+r)(b+r-1)} - \left( \frac{kb}{b+r} \right)^2 \\
 &= \frac{kbr(b+r-k)}{(b+r)^2(b+r-1)}.
 \end{aligned}$$

**Problem 6.** I start by writing out the definition of the MGF:

$$\begin{aligned}
 M_X(s) &= E[e^{sX}] \\
 &= \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{sk} \\
 &= \frac{p}{1-p} \left\{ \sum_{k=0}^{\infty} [(1-p)e^s]^k - 1 \right\},
 \end{aligned}$$

where we recognize that the summation is a geometric series, and is finite provided  $(1-p)e^s < 1$ . Using the formula for a geometric series, and simplifying, I have that:

$$M_X(s) = \frac{pe^s}{1 + (p-1)e^s},$$

for  $s < -\ln(1-p)$ .

**Problem 7.** We can solve this problem by realizing that the  $k^{th}$  derivative of the MGF evaluated at  $s = 0$  gives the  $k^{th}$  moment of the distribution:

$$\begin{aligned}
 E[X] &= \left. \frac{dM_X}{ds} \right|_{s=0} \\
 &= \left. \frac{d}{ds} \left[ \frac{1}{4} + \frac{1}{2}e^s + \frac{1}{4}e^{2s} \right] \right|_{s=0} \\
 &= 1,
 \end{aligned}$$

$$\begin{aligned}
E[X^2] &= \frac{d^2 M_X}{ds^2} \Big|_{s=0} \\
&= \frac{d^2}{ds^2} \left[ \frac{1}{4} + \frac{1}{2}e^s + \frac{1}{4}e^{2s} \right] \Big|_{s=0} \\
&= \frac{3}{2}.
\end{aligned}$$

We therefore have that  $Var[X] = 3/2 - 1^2 = 1/2$ .

**Problem 8.** We already know from Problem 5 in section 6.1.6 of the book that the MGF for a  $\mathcal{N}(\mu, \sigma^2)$  distribution is  $M(s) = \exp(s\mu + \sigma^2 s^2/2)$ , and since the MGF of the sum of independent random variables is the product of the MGFs of the random variables, we have that:

$$\begin{aligned}
M_{X+Y}(s) &= M_X(s)M_Y(s) \\
&= \exp\left(s\mu_X + \frac{\sigma_X^2 s^2}{2}\right) \exp\left(s\mu_Y + \frac{\sigma_Y^2 s^2}{2}\right) \\
&= \exp\left[s(\mu_X + \mu_Y) + \frac{s^2}{2}(\sigma_X^2 + \sigma_Y^2)\right].
\end{aligned}$$

We recognize this as the MGF of a  $\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$  distribution. Further, by Theorem 6.1 in the book, the MGF of a random variable uniquely determines its distribution, so that indeed  $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ .

**Problem 9.** As a note, for the Laplace distribution,  $\lambda > 0$ .

$$\begin{aligned}
M_X(s) &= E[e^{sX}] \\
&= \frac{\lambda}{2} \int_{-\infty}^{\infty} e^{-\lambda|x|+sx} dx \\
&= \frac{\lambda}{2} \left[ \int_{-\infty}^0 e^{x(\lambda+s)} dx + \int_0^{\infty} e^{x(s-\lambda)} dx \right]
\end{aligned}$$

Notice that, for both integrals to be finite, we have the conditions that  $\lambda + s > 0$  (for the first integral) and  $s - \lambda < 0$  (for the second), or in other words  $|s| < \lambda$ . Assuming these two conditions, the integral can easily be evaluated, and is:

$$M_X(s) = \frac{\lambda^2}{\lambda^2 - s^2}.$$

**Problem 10.**

$$\begin{aligned}
M_X(s) &= E[e^{sX}] \\
&= \int_0^{\infty} \frac{e^{sx} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(\lambda-s)x} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda-s)^\alpha} \quad \text{for } s < \lambda \\
&= \left( \frac{\lambda}{\lambda-s} \right)^\alpha
\end{aligned}$$



**Problem 11.** For  $X_i \sim \text{Exp}(\lambda)$ , from Example 6.5 in the book, we have that  $M_{X_i} = \lambda/(\lambda - s)$  for  $s < \lambda$ . Moreover since the MGF of the sum of independent random variables is the product of the MGFs of the random variables, we have that:

$$\begin{aligned} M_Y(s) &= M_{X_1}(s)M_{X_2}(s) \dots M_{X_n}(s) \\ &= \left( \frac{\lambda}{\lambda - s} \right)^n, \end{aligned}$$

which, from the previous problem, we notice is the MGF of a  $\text{Gamma}(n, \lambda)$  random variable. By Theorem 6.1 in the book, the MGF of a random variable uniquely determines its distribution, so  $Y \sim \text{Gamma}(n, \lambda)$ .

**Problem 12.** By the definition of the characteristic function, we have that:

$$\begin{aligned} \phi_Y(\omega) &= E[e^{i\omega Y}] \\ &= E[e^{i\omega(aX+b)}] \\ &= e^{i\omega b} E[e^{i(a\omega)X}] \\ &= e^{i\omega b} \phi_X(a\omega). \end{aligned}$$

**Problem 13.**

(a) To solve for  $E[\mathbf{U}]$ , I first find the marginal PDFs:

$$f_X(x) = \begin{cases} \int_0^1 \frac{1}{2}(3x+y)dy & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{3}{2}x + \frac{1}{4} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases},$$

and

$$f_Y(y) = \begin{cases} \int_0^1 \frac{1}{2}(3x+y)dx & \text{for } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{2}y + \frac{3}{4} & \text{for } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Thus,  $E[X] = \int_0^1 x(3x/2 + 1/4)dx = 5/8$  and  $E[Y] = \int_0^1 y(y/2 + 3/4)dy = 13/24$ , so that

$$E[\mathbf{U}] = \begin{bmatrix} E[X] \\ E[Y] \end{bmatrix} = \begin{bmatrix} \frac{5}{8} \\ \frac{13}{24} \end{bmatrix}.$$

(b) In order to solve for  $\mathbf{R}_U$ , I will first need to compute  $E[X^2]$ ,  $E[Y^2]$  and  $E[XY]$ :

$$E[X^2] = \int_0^1 x^2 \left( \frac{3}{2}x + \frac{1}{4} \right) dx = \frac{11}{24},$$

$$E[Y^2] = \int_0^1 y^2 \left( \frac{1}{2}y + \frac{3}{4} \right) dy = \frac{3}{8},$$

and

$$E[XY] = \frac{1}{2} \int_0^1 \int_0^1 xy(3x+y)dx dy = \frac{1}{3}.$$

I can now immediately write down the correlation matrix:

$$\begin{aligned} \mathbf{R}_U &= E[\mathbf{U}\mathbf{U}^T] \\ &= \begin{bmatrix} E[X^2] & E[XY] \\ E[YX] & E[Y^2] \end{bmatrix} \\ &= \begin{bmatrix} \frac{11}{24} & \frac{1}{3} \\ \frac{1}{3} & \frac{3}{8} \end{bmatrix}. \end{aligned}$$

(c) The covariance matrix is:

$$\begin{aligned}
 \mathbf{C}_U &= E[\mathbf{U}\mathbf{U}^T] - E[\mathbf{U}]E[\mathbf{U}]^T \\
 &= \mathbf{R}_U - \begin{bmatrix} E[X]^2 & E[X]E[Y] \\ E[Y]E[X] & E[Y]^2 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{11}{24} & \frac{1}{3} \\ \frac{1}{3} & \frac{3}{8} \end{bmatrix} - \begin{bmatrix} \left(\frac{5}{8}\right)^2 & \left(\frac{5}{8}\right)\left(\frac{13}{24}\right) \\ \left(\frac{13}{24}\right)\left(\frac{5}{8}\right) & \left(\frac{13}{24}\right)^2 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{13}{192} & -\frac{1}{47} \\ -\frac{1}{192} & \frac{576}{576} \end{bmatrix}.
 \end{aligned}$$

**Problem 14.**

(a) First note that the range of  $Y$  is  $[0, 1]$ . Since we know the distribution of  $Y|X = x$  and the distribution of  $X$ , the law of total probability for PDFs will probably be useful. Note that  $f_{Y|X}(y|x) = 1/x$  for  $0 \leq y \leq x$  and 0 otherwise, which can be written as  $(1/x)\mathbb{1}\{0 \leq y \leq x\}$ , which will be helpful in the integral to get the bounds of integration correct:

$$\begin{aligned}
 f_Y(y) &= \int_0^1 f_{Y|X}(y|x)f_X(x)dx \\
 &= \int_0^1 \frac{1}{x} \mathbb{1}\{0 \leq y \leq x\} dx \\
 &= \int_0^1 \frac{1}{x} \mathbb{1}\{y \leq x\} dx \quad \text{for } y > 0 \\
 &= \int_y^1 \frac{1}{x} dx \\
 &= -\ln y,
 \end{aligned}$$

and thus

$$f_Y(y) = \begin{cases} -\ln y & \text{for } 0 \leq y \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

which I checked integrates to 1.

Finding the PDF of  $Z$  is very similar to that of finding the PDF for  $Y$ . Firstly, the range of  $Z$  is  $[0, 2]$ . Note that in this case  $f_{Z|X}(z|x) = 1/2x$  for  $0 \leq z \leq 2x$  and 0 otherwise, which can be written as  $(1/2x)\mathbb{1}\{0 \leq z \leq 2x\}$ , so that the integral above becomes:

$$\begin{aligned}
 f_Z(z) &= \int_0^1 \frac{1}{2x} \mathbb{1}\{z \leq 2x\} dx \quad \text{for } z > 0 \\
 &= \int_{z/2}^1 \frac{1}{2x} dx \\
 &= \frac{\ln 2}{2} - \frac{\ln z}{2},
 \end{aligned}$$

and thus

$$f_Z(z) = \begin{cases} \frac{\ln 2}{2} - \frac{\ln z}{2} & \text{for } 0 \leq z \leq 2 \\ 0 & \text{otherwise,} \end{cases}$$

which I checked integrates to 1.

(b) Using the chain rule of probability, we have that:

$$\begin{aligned} f_{XYZ}(x, y, z) &= f_{Z|XY}(z|x, y) f_{Y|X}(y|x) f_X(x) \\ &= f_{Z|X}(z|x) f_{Y|X}(y|x) f_X(x), \end{aligned}$$

where in the second line I used the fact that  $Z$  and  $Y$  are conditionally independent given  $X$ . We thus have that

$$f_{XYZ}(x, y, z) = \begin{cases} \frac{1}{2x^2} & \text{for } 0 \leq x \leq 1, 0 \leq y \leq x, 0 \leq z \leq 2x \\ 0 & \text{otherwise,} \end{cases}$$

which I again checked integrates to 1.

**Problem 15.**

(a) As stated in the problem, we have that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \right),$$

and for a bivariate normal, we know that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}, \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] \\ Cov[X_2, X_1] & Var[X_2] \end{bmatrix} \right).$$

Thus, I have that  $X_2 \sim \mathcal{N}(2, 1)$ , so that:

$$\begin{aligned} P(X_2 > 0) &= 1 - P(X_2 \leq 0) \\ &= 1 - \Phi \left( \frac{0 - 2}{1} \right) \\ &= \Phi(2) \\ &\approx 0.98. \end{aligned}$$

(b)

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

$$\begin{aligned} &= \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 2X_1 + X_2 - 1 \\ -X_1 + X_2 \\ X_1 + 3X_2 + 1 \end{bmatrix} \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} E[\mathbf{Y}] &= \begin{bmatrix} 2E[X_1] + E[X_2] - 1 \\ -E[X_1] + E[X_2] \\ E[X_1] + 3E[X_2] + 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 \cdot 1 + 2 - 1 \\ -1 + 2 \\ 1 + 3 \cdot 2 + 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 1 \\ 8 \end{bmatrix} \end{aligned}$$

- (c) We know that a linear combination of a multivariate Gaussian random variable is also Gaussian. Specifically,  $\mathbf{Y}$  is distributed as  $\mathbf{Y} \sim \mathcal{N}(\mathbf{A}E[\mathbf{X}] + \mathbf{b}, \mathbf{A}\mathbf{C}_X\mathbf{A}^T)$ , and thus the covariance matrix of  $\mathbf{Y}$  is

$$\begin{aligned} \mathbf{C}_Y &= \mathbf{A}\mathbf{C}_X\mathbf{A}^T \\ &= \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 1 \\ 1 & 1 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 21 & -6 & 18 \\ -6 & 3 & -13 \\ 18 & -3 & 19 \end{bmatrix}. \end{aligned}$$

Notice that, as it should be,  $\mathbf{C}_Y$  is symmetric.

- (d) As with the first part of this problem, we know that

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} E[Y_1] \\ E[Y_2] \\ E[Y_3] \end{bmatrix}, \begin{bmatrix} \text{Var}[Y_1] & \text{Cov}[Y_1, Y_2] & \text{Cov}[Y_1, Y_3] \\ \text{Cov}[Y_2, Y_1] & \text{Var}[Y_2] & \text{Cov}[Y_2, Y_3] \\ \text{Cov}[Y_3, Y_1] & \text{Cov}[Y_3, Y_2] & \text{Var}[Y_3] \end{bmatrix} \right),$$

so that  $Y_2 \sim \mathcal{N}(1, 3)$ , and therefore:

$$P(Y_2 \leq 2) = \Phi\left(\frac{2-1}{\sqrt{3}}\right) \approx 0.72.$$

**Problem 16.** To solve this problem, I first review how to “complete the square” for matrices. For  $a \in \mathbb{R}$ ,  $\mathbf{x}, \mathbf{b} \in \mathbb{R}^m$  and  $\mathbf{C} \in \mathbb{R}^{m \times m}$  (and symmetric), a quadratic of the form

$$a + \mathbf{b}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{C} \mathbf{x}$$

can be factored into the form

$$\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{M}(\mathbf{x} - \mathbf{m}) + v,$$

where

$$\mathbf{M} = \mathbf{C},$$

$$\mathbf{m} = -\mathbf{C}^{-1}\mathbf{b},$$

and

$$v = a - \frac{1}{2} \mathbf{b}^T \mathbf{C}^{-1} \mathbf{b}.$$

I now explicitly write out the MGF of  $\mathbf{X}$ :

$$\begin{aligned} M_{\mathbf{X}}(s, t, r) &= E[e^{sX_1 + tX_2 + rX_3}] \\ &= \int_{\mathbb{R}^3} \exp\{\mathbf{s}^T \mathbf{x}\} \frac{1}{(2\pi)^{3/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} d^3x \\ &= \frac{1}{(2\pi)^{3/2} |\boldsymbol{\Sigma}|^{1/2}} \int_{\mathbb{R}^3} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) + \mathbf{s}^T \mathbf{x}\right\} d^3x, \end{aligned}$$

where  $\mathbf{x}^T \equiv [x_1, x_2, x_3]$  and  $\mathbf{s}^T \equiv [s, t, r]$ . To make the exponent more Gaussian looking, I now expand the exponent out and complete the square (note that since  $\Sigma$  is symmetric, then so too is  $\Sigma^{-1}$ ):

$$\begin{aligned} -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) + \mathbf{s}^T \mathbf{x} &= -\frac{1}{2} [\mathbf{x}^T \Sigma^{-1} \mathbf{x} - \mathbf{x}^T \Sigma^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{x} + \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}] + \mathbf{s}^T \mathbf{x} \\ &= -\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} + \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \mathbf{s}^T \mathbf{x} \\ &= -\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} + (\mathbf{s}^T + \boldsymbol{\mu}^T \Sigma^{-1}) \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}, \end{aligned}$$

where I have used the fact that  $(\mathbf{x}^T \Sigma^{-1} \boldsymbol{\mu})^T = \mathbf{x}^T \Sigma^{-1} \boldsymbol{\mu}$  since this is just a real number. I can now read off  $a$ ,  $\mathbf{b}$  and  $\mathbf{C}$ :

$$a = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu},$$

$$\mathbf{b}^T = \mathbf{s}^T + \boldsymbol{\mu}^T \Sigma^{-1}$$

and

$$\mathbf{C} = -\Sigma^{-1},$$

so that

$$\mathbf{b} = (\mathbf{b}^T)^T = \mathbf{s} + \Sigma^{-1} \boldsymbol{\mu}$$

and

$$\mathbf{C}^{-1} = (-\Sigma^{-1})^{-1} = -\Sigma.$$

Finally, the exponent can be re-expressed as

$$-\frac{1}{2}(\mathbf{x} - \tilde{\mathbf{m}})^T \tilde{\mathbf{M}}(\mathbf{x} - \tilde{\mathbf{m}}) + \tilde{v},$$

where

$$\tilde{\mathbf{m}} = \Sigma(\mathbf{s} + \Sigma^{-1} \boldsymbol{\mu}),$$

$$\tilde{\mathbf{M}} = \Sigma^{-1}$$

and

$$\begin{aligned} \tilde{v} &= -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \frac{1}{2} (\mathbf{s}^T + \boldsymbol{\mu}^T \Sigma^{-1}) \Sigma (\mathbf{s} + \Sigma^{-1} \boldsymbol{\mu}) \\ &= \mathbf{s}^T \boldsymbol{\mu} + \frac{\mathbf{s}^T \Sigma \mathbf{s}}{2}, \end{aligned}$$

so that the integral becomes:

$$\begin{aligned} M_{\mathbf{X}}(s, t, r) &= \exp(\tilde{v}) \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2}} \int_{\mathbb{R}^3} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \tilde{\mathbf{m}})^T \Sigma^{-1} (\mathbf{x} - \tilde{\mathbf{m}}) \right\} d^3 x \\ &= \exp \left( \mathbf{s}^T \boldsymbol{\mu} + \frac{\mathbf{s}^T \Sigma \mathbf{s}}{2} \right) \quad \forall \mathbf{s} \in \mathbb{R}^3. \end{aligned}$$

I have used the fact that the integral is that of a Gaussian integrated over its entire domain, so that the integral evaluates to 1. Note that we probably could have guessed this form of the MGF of  $\mathbf{X}$ , since it is the vector analogue of the 1 dimensional case:  $M_X(s) = \exp\{s\mu + \sigma^2 s^2/2\}$ , as found in Problem 5 of 6.1.6 in the book.

The specified values of the mean vector and covariance matrix are:

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix},$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 9 & 1 & -1 \\ 1 & 4 & 2 \\ -1 & 2 & 4 \end{bmatrix},$$

and plugging in these specific values into the equation I derived above, and multiplying the matrices, I finally arrive at:

$$M_{\mathbf{X}}(s, t, r) = \exp \left\{ s \left( \frac{9}{2}s + 1 \right) + 2t(t + 1) + 2r^2 - rs + st + 2rt \right\}.$$

**Problem 17.** Let  $A_i$ ,  $i = 1, 2, 3, 4$ , be the event that the  $i^{th}$  component fails, so that the a failure occurs under the event  $\cup_{i=1}^4 A_i$ . We can thus obtain an upper limit on the event that failure occurs using the union bound:

$$\begin{aligned} P \left( \bigcup_{i=1}^4 A_i \right) &\leq \sum_{i=1}^4 P(A_i) \\ &\leq 4p_f \\ &= \frac{1}{25}. \end{aligned}$$

**Problem 18.**

- (a) First note that for the random position of a node  $(X_i, Y_i)$ ,  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n \stackrel{iid}{\sim} \text{Unif}(0, 1)$ . Let us call the node under consideration node  $j$ , and let the set  $\mathcal{S}$  be defined as  $\mathcal{S} \equiv \{1, 2, \dots, n\} - \{j\}$ . The probability that the node is isolated,  $p_d$ , is:

$$\begin{aligned} p_d &= P \left( \bigcap_{i \in \mathcal{S}} [(X_j - X_i)^2 + (Y_j - Y_i)^2 > r^2] \right) \\ &= \int_0^1 \int_0^1 P \left( \bigcap_{i \in \mathcal{S}} [(X_i - X_j)^2 + (Y_i - Y_j)^2 > r^2] \mid X_j = x_j, Y_j = y_j \right) f_{X_j Y_j}(x_j, y_j) dx_j dy_j \\ &= \int_0^1 \int_0^1 P \left( \bigcap_{i \in \mathcal{S}} [(X_i - x_j)^2 + (Y_i - y_j)^2 > r^2] \right) f_{X_j}(x_j) f_{Y_j}(y_j) dx_j dy_j \\ &= \int_0^1 \int_0^1 \prod_{i \in \mathcal{S}} P((X_i - x_j)^2 + (Y_i - y_j)^2 > r^2) f_{X_j}(x_j) f_{Y_j}(y_j) dx_j dy_j \\ &= \int_0^1 \int_0^1 \prod_{i \in \mathcal{S}} [1 - P((X_i - x_j)^2 + (Y_i - y_j)^2 \leq r^2)] f_{X_j}(x_j) f_{Y_j}(y_j) dx_j dy_j \\ &= \int_0^1 \int_0^1 [1 - P((X_1 - x_j)^2 + (Y_1 - y_j)^2 \leq r^2)]^{n-1} f_{X_j}(x_j) f_{Y_j}(y_j) dx_j dy_j, \end{aligned}$$

where in the third and fourth lines I have used the fact that the random variables are independent, and in the last line, I have used symmetry. I now must compute  $P((X_1 - x_j)^2 +$

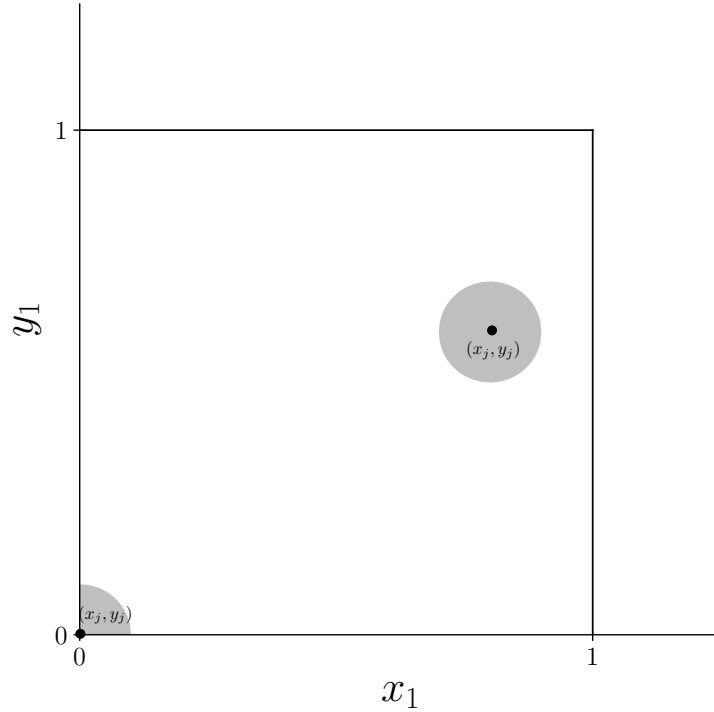


Figure 6.1: Two example nodes at  $(x_j, y_j)$  for Problem 18.

$(Y_1 - y_j)^2 \leq r^2$ ). If the given point  $(x_j, y_j)$  is near the middle of the square (as in the upper point in Fig. 6.1) then the probability of this event is simply the area of this circle (shaded grey region). However, we notice that if  $(x_j, y_j)$  is near the edge of the square, part of the shaded circle will get cutoff. In fact, if  $(x_j, y_j)$  is exactly at one of the corners of the unit square, for example, at  $(0, 0)$  as in Fig. 6.1, then the amount of shaded area is minimized at  $\pi r^2/4$ . Thus, for any given  $(x_j, y_j)$ ,  $P((X_1 - x_j)^2 + (Y_1 - y_j)^2 \leq r^2) \geq \pi r^2/4$ . Therefore,

$$\begin{aligned} p_d &\leq \int_0^1 \int_0^1 \left(1 - \frac{\pi r^2}{4}\right)^{n-1} f_{X_j}(x_j) f_{Y_j}(y_j) dx_j dy_j \\ &= \left(1 - \frac{\pi r^2}{4}\right)^{n-1} \end{aligned}$$

(b) Let  $A_i$  be the event that the  $i^{th}$  node is isolated. Then the probability we seek is:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &\leq \sum_{i=1}^n P(A_i) \\ &= \sum_{i=1}^n p_d \\ &= n \left(1 - \frac{\pi r^2}{4}\right)^{n-1} \end{aligned}$$

**Problem 19.** For  $X \sim \text{Geom}(p)$ ,  $E[X] = 1/p$ , so that the Markov inequality is:

$$P(X \geq a) \leq \frac{E[X]}{a} = \frac{1}{pa}.$$

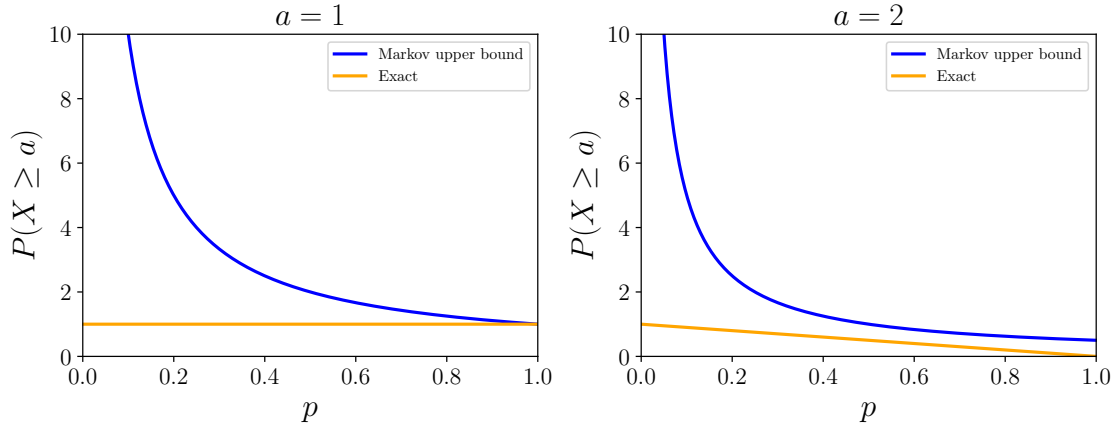


Figure 6.2: Comparison of Markov upper bound to exact probability for Problem 19.

The exact probability is:

$$\begin{aligned}
 P(X \geq a) &= \sum_{k=a}^{\infty} p(1-p)^{k-1} \\
 &= \frac{p}{1-p} \left( \sum_{k=0}^{\infty} (1-p)^k - \sum_{k=0}^{a-1} (1-p)^k \right) \\
 &= \frac{p}{1-p} \left( \frac{1}{1-(1-p)} - \frac{1-(1-p)^a}{1-(1-p)} \right) \\
 &= (1-p)^{a-1}.
 \end{aligned}$$

The Markov upper bound is greater than or equal to the exact probability for  $a \geq 1$  and  $0 < p < 1$  as shown for a few values of  $a$  in Fig. 6.2

**Problem 20.**

$$P(|X - E[X]| \geq b) \leq \frac{\text{Var}[X]}{b^2} = \frac{1-p}{pb^2}$$

**Problem 21.**

$$\begin{aligned}
 P(X \geq a) &= P\left(X + \frac{\sigma^2}{a} \geq a + \frac{\sigma^2}{a}\right) \\
 &= P\left(\left(X + \frac{\sigma^2}{a}\right)^2 \geq \left(a + \frac{\sigma^2}{a}\right)^2\right) \\
 &\leq \frac{E\left[\left(X + \frac{\sigma^2}{a}\right)^2\right]}{\left(a + \frac{\sigma^2}{a}\right)^2} \quad \text{Markov's Inequality} \\
 &= \frac{\sigma^2 + \frac{\sigma^4}{a^2}}{a^2\left(1 + \frac{\sigma^2}{a^2}\right)^2} \\
 &= \frac{\sigma^2}{a^2 + \sigma^2}
 \end{aligned}$$

**Problem 22.**



(a)

$$\begin{aligned}
P(X \leq 80 \text{ or } X \geq 120) &= P(|X - 100| \geq 20) \\
&= P(|X - E[X]| \geq 20) \\
&\leq \frac{225}{20^2} \text{ by Chebyshev} \\
&= \frac{9}{16}
\end{aligned}$$

(b)

$$P(X \geq 120) \leq \frac{225}{225 + 120^2} = \frac{45}{293}$$

**Problem 23.** We know from Problem 11 that if  $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \text{Exp}(\lambda)$ , then  $Y \equiv X_1 + X_2 + \dots + X_n \sim \text{Gamma}(n, \lambda)$ . The relevant Chernoff bound is given by:

$$P(Y \geq a) \leq \min_{s>0} \{e^{-sa} M_X(s)\},$$

where, in this case the  $M_X(s)$  is the MGF for a  $\text{Gamma}(n, \lambda)$  distribution. This MGF was solved for in Problem 10, and is given by

$$M_X(s) = \left( \frac{\lambda}{\lambda - s} \right)^n \text{ for } s < \lambda,$$

and therefore we must minimize the objective function over  $0 < s < \lambda$ . Let the optimal value be called  $s^*$ . I solve for  $s^*$  in the standard calculus manner by setting the derivative equal 0. I then check to make sure that this optimal value is within the interval  $(0, \lambda)$ . The derivative of the objective can be found easily with the chain rule:

$$\frac{d}{ds} e^{-sa} \left( \frac{\lambda}{\lambda - s} \right)^n = -ae^{-sa} \left( \frac{\lambda}{\lambda - s} \right)^n + ne^{-sa} \left( \frac{\lambda}{\lambda - s} \right)^n \frac{1}{\lambda - s},$$

and setting this equal to zero and solve for  $s^*$  results in

$$s^* = \lambda - \frac{n}{a}.$$

As stipulated in the problem  $a > n/\lambda$ , which means that  $n/a$  is positive and less than  $\lambda$ . Thus we have that  $s^* \in (0, \lambda)$  as required.

The desired bound is therefore

$$\begin{aligned}
P(Y \geq a) &\leq e^{-s^*a} M_X(s^*) \\
&= e^{-\lambda a + n} \left( \frac{\lambda a}{n} \right)^n.
\end{aligned}$$

We can understand the behavior of this function as  $n \rightarrow \infty$  by expanding the exponential in powers of  $1/n$ :

$$\begin{aligned}
e^{-\lambda a + n} \left( \frac{\lambda a}{n} \right)^n &= e^{-\lambda a} \left[ 1 + \mathcal{O}\left(\frac{1}{n}\right) \right] \left( \frac{1}{n} \right)^n (\lambda a)^n \\
&= e^{-\lambda a} \left( \frac{\lambda a}{n} \right)^n + \mathcal{O}\left( \left( \frac{1}{n} \right)^{n+1} \right),
\end{aligned}$$

and we thus see that the upper bound goes to 0 exponentially fast as  $n$  goes to infinity.

**Problem 24.** Using some properties of absolute values I have that:

$$\begin{aligned}
 E[|X + Y|^{p-1}|X|] &= E[(X + Y)^{p-1}|X|] \\
 &= E[(X + Y)^{p-1}X] \\
 &\leq E[(X + Y)^{p-1}]^{\frac{p-1}{p}} E[|X|^p]^{\frac{1}{p}} \\
 &= E[|X + Y|^p]^{\frac{p-1}{p}} E[|X|^p]^{\frac{1}{p}},
 \end{aligned}$$

where in the third line I have used Hölder's inequality,  $E[|UV|] \leq E[|U|^\alpha]^{1/\alpha} E[|V|^\beta]^{1/\beta}$ , with  $1 < \alpha, \beta < \infty$  and  $1/\alpha + 1/\beta = 1$ , and where I have specifically chosen  $\alpha = p/(p-1)$  and  $\beta = p$ .

Using the inequality provided in the book,

$$\begin{aligned}
 E[|X + Y|^p] &\leq E[|X + Y|^{p-1}|X|] + E[|X + Y|^{p-1}|Y|] \\
 &\leq E[|X + Y|^p]^{\frac{p-1}{p}} (E[|X|^p]^{\frac{1}{p}} + E[|Y|^p]^{\frac{1}{p}}),
 \end{aligned}$$

and multiplying both sides of this equation by  $E[|X + Y|^p]^{(1-p/p)}$  (which we can do without flipping the inequality sign since we know that this quantity is positive) yields the desired result:

$$E[|X + Y|^p]^{\frac{1}{p}} \leq E[|X|^p]^{\frac{1}{p}} + E[|Y|^p]^{\frac{1}{p}}.$$

**Problem 25.**

(a)

$$\frac{d^2}{dx^2}(x - x^3) = -6x$$

$\implies$

$$g''(x) = \begin{cases} + (\text{convex}) & \text{for } x < 0 \\ - (\text{concave}) & \text{for } x > 0, \end{cases}$$

Since we know  $X$  is a positive random variable, by Jensen's inequality, we have:

$$E[X - X^3] \leq E[X] - E[X]^3 = -990.$$

(b)

$$\frac{d^2}{dx^2}(x \ln \sqrt{x}) = \frac{1}{2x}$$

$$\implies g''(x) > 0 \text{ (convex) for } x > 0 \implies$$

$$E[X \ln \sqrt{X}] \geq E[X] \ln \sqrt{E[X]} = 10 \ln \sqrt{10}$$

(c) The function is a typical absolute value function, an upward v shape hitting  $y = 0$  at  $x = 2$ , which is clearly convex, since a straight line drawn from any 2 points on the graph is always above the graph. Therefore, I have that  $E[|2 - X|] \geq |2 - E[X]| = 8$ .

**Problem 26.** Taking the second derivative, we have that  $d^2/dx^2(x^3 - 6x^2) = 6x - 12$ . Setting to zero and solving for  $x$ , I find that the second derivative is negative for  $x < 2$  and positive for  $x > 2$ . Since the range of  $X$  is  $(0, 2)$ , we have that  $g(x) = x^3 - 6x^2$  is concave in this interval. By Jensen's inequality, this implies that  $E[Y] = E[g(X)] \leq g(E[X]) = E[X]^3 - 6E[X]^2 = 1 - 6 = -5$ .

## Chapter 7

# Limit Theorems and Convergence of Random Variables

**Problem 1.**

(a)

$$\begin{aligned}
E[M_n] &= \frac{1}{n} E[X_1 + \dots + X_n] \\
&= \frac{1}{n} n E[X_1] \\
&= \frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
Var[M_n] &= \frac{1}{n^2} Var[X_1 + \dots + X_n] \\
&= \frac{1}{n^2} n Var[X_1] \quad (\text{independence}) \\
&= \frac{1}{12n}
\end{aligned}$$

(b)

$$\begin{aligned}
P\left(\left|M_n - \frac{1}{2}\right| \geq \frac{1}{100}\right) &= P\left(|M_n - E[M_n]| \geq \frac{1}{100}\right) \\
&\leq \frac{Var[M_n]}{\left(\frac{1}{100}\right)^2} \\
&= \frac{2500}{3n}
\end{aligned}$$

(c)

$$\lim_{n \rightarrow \infty} P\left(|M_n - E[M_n]| \geq \frac{1}{100}\right) \leq \lim_{n \rightarrow \infty} \frac{2500}{3n} = 0$$

 $\Rightarrow$ 

$$\lim_{n \rightarrow \infty} P\left(\left|M_n - \frac{1}{2}\right| \geq \frac{1}{100}\right) = 0$$

**Problem 2.** Let  $X_1, X_2, \dots, X_{365}$  be the number of accidents on day 1, day 2,  $\dots$ , day 365, so that the total number of accidents in the year is  $Y = X_1 + \dots + X_{365}$ . We know that  $X_1, X_2, \dots, X_{365} \stackrel{iid}{\sim} Poiss(\lambda)$ , where  $\lambda = 10$  accidents/day, so that  $\mu = E[X_i] = \lambda = 10$  and  $\sigma = \sqrt{Var[X_i]} = \sqrt{\lambda} = \sqrt{10}$ . Using the central limit theorem, I have

$$\begin{aligned}
P(Y > 3800) &= P\left(\frac{Y - 365 \cdot 10}{\sqrt{365 \cdot 10}} > \frac{3800 - 365 \cdot 10}{\sqrt{365 \cdot 10}}\right) \\
&= P\left(Z_{365} > \frac{3800 - 365 \cdot 10}{\sqrt{365 \cdot 10}}\right) \\
&= 1 - P\left(Z_{365} \leq \frac{3800 - 365 \cdot 10}{\sqrt{365 \cdot 10}}\right) \\
&\approx 1 - \Phi\left(\frac{3800 - 365 \cdot 10}{\sqrt{365 \cdot 10}}\right) \\
&\approx 6.5 \times 10^{-3}.
\end{aligned}$$

**Problem 3.** Let the random variable,  $X_i$  be 0 if the  $i^{th}$  bit is not received in error and 1 if it is. Notice that  $X_1, X_2, \dots, X_{1000} \stackrel{iid}{\sim} \text{Bern}(0.1)$ , and let the total number of errors,  $Y$ , be  $Y = X_1 + \dots + X_{1000}$ . Note that  $\mu = E[X_i] = p = 0.1$  and  $\sigma = \sqrt{\text{Var}[X_i]} = \sqrt{p(1-p)} = \sqrt{0.09}$ . We seek the probability of decoding failure, in other words  $P(Y > 125)$ :

$$\begin{aligned}
 P(Y > 125) &= 1 - P(Y \leq 125) \\
 &= 1 - P\left(\frac{Y - 1000 \cdot 0.1}{\sqrt{0.09 \cdot 1000}} \leq \frac{125 - 1000 \cdot 0.1}{\sqrt{0.09 \cdot 1000}}\right) \\
 &= 1 - P\left(Z_{1000} \leq \frac{125 - 1000 \cdot 0.1}{\sqrt{0.09 \cdot 1000}}\right) \\
 &\approx 1 - \Phi\left(\frac{25}{3\sqrt{10}}\right) \\
 &= 4.2 \times 10^{-3}.
 \end{aligned}$$

**Problem 4.** Let the random variable,  $X_i$  be 0 if the  $i^{th}$  student does not have a car and 1 if the  $i^{th}$  student does have a car. Notice that  $X_1, X_2, \dots, X_{50} \stackrel{iid}{\sim} \text{Bern}(0.5)$ , and let the total number of cars,  $Y$ , be  $Y = X_1 + \dots + X_{50}$ . Note that  $\mu = E[X_i] = p = 0.5$  and  $\sigma = \sqrt{\text{Var}[X_i]} = \sqrt{p(1-p)} = 0.5$ . We seek the probability that there are not enough car spaces, in other words  $P(Y > 30)$ :

$$\begin{aligned}
 P(Y > 30) &= P(Y > 29.5) \quad (\text{using the continuity correction}) \\
 &= 1 - P(Y \leq 29.5) \\
 &= 1 - P\left(\frac{Y - 50 \cdot 0.5}{\sqrt{50 \cdot 0.5}} \leq \frac{29.5 - 50 \cdot 0.5}{\sqrt{50 \cdot 0.5}}\right) \\
 &= 1 - P\left(Z_{50} \leq \frac{29.5 - 50 \cdot 0.5}{\sqrt{50 \cdot 0.5}}\right) \\
 &\approx 1 - \Phi(1.27) \\
 &\approx 0.10.
 \end{aligned}$$

**Problem 5.** Let  $N$ , a random variable, be the number of jobs processed in 7 hours (420 mins). We seek the probability that the number of jobs processed in 7 hours is less than or equal to 40,  $P(N \leq 40)$ . This can be rephrased as the probability that the total time to processes 40 jobs is greater than or equal to 7 hours:

$$\begin{aligned}
 P(N \leq 40) &= P(X_1 + \dots + X_{40} \geq 420) \\
 &= P\left(\frac{X_1 + \dots + X_{40} - 40 \cdot 10}{\sqrt{40} \cdot \sqrt{2}} \geq \frac{420 - 40 \cdot 10}{\sqrt{40} \cdot \sqrt{2}}\right) \\
 &= P\left(Z_{40} \geq \frac{5}{\sqrt{5}}\right) \\
 &= 1 - P\left(Z_{40} < \frac{5}{\sqrt{5}}\right) \\
 &\approx 1 - \Phi\left(\frac{5}{\sqrt{5}}\right) \\
 &\approx 1.3 \times 10^{-2}.
 \end{aligned}$$

**Problem 6.** Let  $X_i$  be the number of heads flipped on toss  $i$ , so that the total proportion of heads out of  $n$  tosses,  $X$ , is  $X = (X_1 + \dots + X_n)/n$ . Notice that  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bern}(0.5)$ , so

that  $\mu = E[X_i] = p = 0.5$  and  $\sigma = \sqrt{\text{Var}[X_i]} = \sqrt{0.5^2} = 0.5$ . To be at least 95% sure that  $0.45 \leq X \leq 0.55$ , we have that:

$$\begin{aligned}
0.95 &\leq P(0.45 \leq X \leq 0.55) \\
&= P\left(\frac{0.45n - n \cdot 0.5}{\sqrt{n} \cdot 0.5} \leq \frac{X_1 + \dots + X_n - n \cdot 0.5}{\sqrt{n} \cdot 0.5} \leq \frac{0.55n - n \cdot 0.5}{\sqrt{n} \cdot 0.5}\right) \\
&= P(-0.1\sqrt{n} \leq Z_n \leq 0.1\sqrt{n}) \\
&\approx \Phi(0.1\sqrt{n}) - \Phi(-0.1\sqrt{n}) \\
&= 2\Phi(0.1\sqrt{n}) - 1.
\end{aligned}$$

Thus, I have that  $0.95 \lesssim 2\Phi(0.1\sqrt{n}) - 1$ . Applying the inverse normal CDF function to this inequality, I arrive at:

$$n \gtrsim \left\lceil 100 \left[ \Phi^{-1}\left(\frac{1.95}{2}\right) \right]^2 \right\rceil = 385. \quad (7.1)$$

**Problem 7.** Note that  $X_1, X_2, \dots, X_n$  are *iid* with  $\mu = E[X_i] = 0$  and  $\sigma = \sqrt{\text{Var}[X_i]} = 2$ , so we can use the CLT. To be at least 95% sure that the final estimate is within 0.1 units of  $q$ , we require:

$$\begin{aligned}
0.95 &\leq P(q - 0.1 \leq M_n \leq q + 0.1) \\
&= P\left(q - 0.1 \leq \frac{X_1 + \dots + X_n + nq}{n} \leq q + 0.1\right) \\
&= P((q - 0.1)n - nq \leq X_1 + \dots + X_n \leq (q + 0.1)n - nq) \\
&= P\left(\frac{(q - 0.1)n - nq}{2\sqrt{n}} \leq \frac{X_1 + \dots + X_n}{2\sqrt{n}} \leq \frac{(q + 0.1)n - nq}{2\sqrt{n}}\right) \\
&= P\left(\frac{-0.1\sqrt{n}}{2} \leq Z_n \leq \frac{0.1\sqrt{n}}{2}\right) \\
&\approx \Phi\left(\frac{0.1\sqrt{n}}{2}\right) - \Phi\left(\frac{-0.1\sqrt{n}}{2}\right) \\
&= 2\Phi\left(\frac{0.1\sqrt{n}}{2}\right) - 1.
\end{aligned}$$

We therefore have that  $0.95 \lesssim 2\Phi(0.1\sqrt{n}/2) - 1$ . Applying the inverse normal CDF function to this inequality, I arrive at:

$$n \gtrsim \left\lceil 400 \left[ \Phi^{-1}\left(\frac{1.95}{2}\right) \right]^2 \right\rceil = 1537. \quad (7.2)$$

**Problem 8.** To solve this problem, I first compute the limit of  $\exp[n(x-1)]/\{1 + \exp[n(x-1)]\}$  for  $x > 0$  as  $n$  goes to  $\infty$ . Notice that this function has different behavior for  $x = 1$  (in which case the limit evaluates easily to  $1/2$ ),  $0 < x < 1$  (in which case the limit evaluates easily to 0) and for  $x > 1$  (in which case the numerator and denominator evaluate to infinity). Using L'hopital's rule in this case I find that the limit evaluates to 1. Therefore, I have that:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} \lim_{n \rightarrow \infty} \frac{e^{n(x-1)}}{1 + e^{n(x-1)}} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 0 & \text{for } -\infty < x < 1 \\ \frac{1}{2} & \text{for } x = 1 \\ 1 & \text{for } x > 1. \end{cases}$$

For the “random variable”,  $X$ , that takes on a value of 1 with probability 1, the CDF is:

$$F_X(x) = \begin{cases} 0 & \text{for } -\infty < x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

Thus, we see that  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  everywhere  $F_X(x)$  is continuous (i.e.,  $\mathbb{R} - \{1\}$ ), and hence  $X_n \xrightarrow{d} X$ .

**Problem 9.** To solve this problem, I first state without proof the following 2 limits:

$$\lim_{n \rightarrow \infty} \frac{e^{nx} + xe^{nx}}{1 + \left(\frac{n+1}{n}\right)e^n} = x \quad \text{for } 0 \leq x \leq 1,$$

and

$$\lim_{n \rightarrow \infty} \frac{e^{nx} + e^{nx}}{1 + \left(\frac{n+1}{n}\right)e^n} = 1 \quad \text{for } x > 1.$$

I therefore have that:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{for } x < 0 \\ \lim_{n \rightarrow \infty} \frac{e^{nx} + xe^{nx}}{1 + \left(\frac{n+1}{n}\right)e^n} & \text{for } 0 \leq x \leq 1 \\ \lim_{n \rightarrow \infty} \frac{e^{nx} + e^{nx}}{1 + \left(\frac{n+1}{n}\right)e^n} & \text{for } x > 1 \end{cases} = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1, \end{cases}$$

which is the same CDF as a  $Unif(0, 1)$  distribution. Hence,  $X_n \xrightarrow{d} X$  for  $X \sim Unif(0, 1)$ .

**Problem 10.**

(a)

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|X_n - 0| \geq \epsilon) &= \lim_{n \rightarrow \infty} P(X_n \geq \epsilon) \quad (\text{since } X_n \geq 0) \\ &= \begin{cases} \frac{1}{n^2} & \text{for } \epsilon \leq n \\ 0 & \text{for } \epsilon > n \end{cases} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \\ &= 0 \end{aligned}$$

$$\implies X_n \xrightarrow{p} 0$$

(b)

$$\begin{aligned} \lim_{n \rightarrow \infty} E[|X_n - 0|^r] &= \lim_{n \rightarrow \infty} E[X_n^r] \quad (\text{since } X_n \geq 0) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^2} n^r \\ &= \lim_{n \rightarrow \infty} n^{r-2} \\ &= 0 \quad (\text{for } 1 \leq r < 2) \end{aligned}$$

$$\implies X_n \xrightarrow{L^r} 0 \quad (\text{for } 1 \leq r < 2)$$

(c) For  $r \geq 2$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} E[|X_n - 0|^r] &= \lim_{n \rightarrow \infty} E[X_n^r] \quad (\text{since } X_n \geq 0) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^2} n^r \\ &= \lim_{n \rightarrow \infty} n^{r-2}, \end{aligned}$$

which converges to 1 for  $r = 2$  and diverges for  $r > 2$ . Therefore,  $X_n$  does not converge to 0 in the  $r^{th}$  mean for  $r \geq 2$ .

- (d) To solve this problem I use Theorem 7.5 in the book, and must thus show that  $\sum_{n=1}^{\infty} P(|X_n| > \epsilon)$  (for all  $\epsilon > 0$ ) is finite:

$$\begin{aligned}
 \sum_{n=1}^{\infty} P(|X_n| > \epsilon) &= \sum_{n=1}^{\infty} P(X_n > \epsilon) \\
 &= \sum_{n=\lceil \epsilon \rceil}^{\infty} \frac{1}{n^2} \\
 &\leq \sum_{n=1}^{\infty} \frac{1}{n^2} \\
 &= \frac{\pi^2}{6} \\
 &< \infty,
 \end{aligned}$$

where in the first line I have used the fact that  $X_n$  is always greater than or equal to zero.

**Problem 11.** This is a hypergeometric experiment with  $b = n + \delta$  (with  $\delta = 0, 1, 2, \dots$ ),  $r = n$  and  $k = 10$ , so that the PMF for  $X_{10}, X_{11}, \dots$  is given by:

$$P_{X_n}(x) = \frac{\binom{n+\delta}{x} \binom{n}{10-x}}{\binom{2n+\delta}{10}},$$

for  $x = 0, 1, \dots, 10$  (and 0 otherwise). Since  $X, X_{10}, X_{11}, \dots$  are non-negative random integers (for  $X \sim \text{Bin}(10, 0.5)$ ), by Theorem 7.1 in the book, we need only prove that  $\lim_{n \rightarrow \infty} P_{X_n}(x) = P_X(x)$  to prove convergence in distribution. As for the RHS of this equation, for  $X \sim \text{Bin}(10, 0.5)$ , the PMF is given by:

$$P_X(x) = \binom{10}{x} (0.5)^{10},$$

for  $x = 0, 1, \dots, 10$  (and 0 otherwise).

As for the LHS of this equation, taking the limit of  $P_{X_n}(x)$  as  $n \rightarrow \infty$  I have that:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P_{X_n}(x) &= \lim_{n \rightarrow \infty} \frac{\binom{n+\delta}{x} \binom{n}{10-x}}{\binom{2n+\delta}{10}} \\
 &= \lim_{n \rightarrow \infty} \binom{n+\delta}{x} \lim_{n \rightarrow \infty} \binom{n}{10-x} \left[ \lim_{n \rightarrow \infty} \binom{2n+\delta}{10} \right]^{-1}.
 \end{aligned}$$

The first limit can be found easily by expanding the factorial in the numerator:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \binom{n+\delta}{x} &= \frac{1}{x!} \lim_{n \rightarrow \infty} (n+\delta)(n+\delta-1) \dots (n+\delta-x+1) \\
 &= \frac{1}{x!} \lim_{n \rightarrow \infty} n^x + \mathcal{O}(n^{x-1}) \\
 &= \lim_{n \rightarrow \infty} \frac{n^x}{x!},
 \end{aligned}$$



and the remaining 2 limits can be worked out similarly. Plugging these limits in, I have that:

$$\begin{aligned}\lim_{n \rightarrow \infty} P_{X_n}(x) &= \lim_{n \rightarrow \infty} \left\{ \frac{n^x}{x!} \right\} \lim_{n \rightarrow \infty} \left\{ \frac{n^{10-x}}{(10-x)!} \right\} \lim_{n \rightarrow \infty} \left\{ \left[ \frac{(2n)^{10}}{10!} \right]^{-1} \right\} \\ &= \lim_{n \rightarrow \infty} \left\{ \frac{n^x}{x!} \cdot \frac{n^{10-x}}{(10-x)!} \cdot \left[ \frac{(2n)^{10}}{10!} \right]^{-1} \right\} \\ &= \binom{10}{x} (0.5)^{10}.\end{aligned}$$

Thus  $\lim_{n \rightarrow \infty} P_{X_n}(x) = P_X(x)$ , and by Theorem 7.1  $X_n \xrightarrow{d} X$ .

**Problem 12.** Let

$$X_n = \frac{X_1 + \dots + X_n - n\mu_X}{\sigma_X \sqrt{n}},$$

where  $X_1, \dots, X_n$  are *iid* from any distribution with finite mean  $\mu_X$  and finite variance  $\sigma_X^2$ . Also, let  $Y_n$  be defined analogously (and let all  $X_i$ s be independent from all  $Y_i$ s, so that  $X_n$  is independent of  $Y_n$ ). Moreover, let  $X \sim \mathcal{N}(0, 1)$  and let  $Y = X$ . Now, from the CLT, we know that  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$ .

From the CLT, we also know that in the limit that  $n \rightarrow \infty$ ,  $X_n + Y_n$  is simply the sum of two independent standard normal random variables, so that in this limit  $X_n + Y_n \sim \mathcal{N}(0, 2)$ . Also, since  $X + Y = 2X$ , we have that  $X + Y \sim \mathcal{N}(0, 4)$ , since for  $\mathcal{X} \sim \mathcal{N}(E[\mathcal{X}], \text{Var}[\mathcal{X}])$ ,  $\mathcal{Y} = a\mathcal{X} + b \sim \mathcal{N}(aE[\mathcal{X}] + b, a^2\text{Var}[\mathcal{X}])$  (see Sec. 6.1.5 from the book). Thus, in this example, I have that  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$ , but  $X_n + Y_n$  does not converge in distribution to  $X + Y$ .

**Problem 13.**  $X_n \xrightarrow{d} 0$  since:

$$\begin{aligned}\lim_{n \rightarrow \infty} P(|X_n| \geq \epsilon) &= \lim_{n \rightarrow \infty} 2 \int_{\epsilon}^{\infty} \frac{n}{2} e^{-nx} dx \quad (\text{by symmetry}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{e^{n\epsilon}} \\ &= 0.\end{aligned}$$

**Problem 14.** This can easily be proven by realizing that  $X_n$  is never negative (so  $|X_n| = X_n$ ), and by re-expressing the integral over the PDF in terms of an indicator functions depending on whether  $\epsilon > 1/n$  (in which case the lower bound of the integral is  $\epsilon$ ) or whether  $\epsilon \leq 1/n$  (in which case the integral evaluates to 1):

$$\begin{aligned}\lim_{n \rightarrow \infty} P(|X_n| \geq \epsilon) &= \lim_{n \rightarrow \infty} P(X_n \geq \epsilon) \\ &= \lim_{n \rightarrow \infty} \left( \mathbb{1} \left\{ \epsilon > \frac{1}{n} \right\} \int_{\epsilon}^{\infty} \frac{1}{n} x^{-2} dx + \mathbb{1} \left\{ \epsilon \leq \frac{1}{n} \right\} \right) \\ &= \int_{\epsilon}^{\infty} x^{-2} dx \lim_{n \rightarrow \infty} \left( \mathbb{1} \left\{ \epsilon > \frac{1}{n} \right\} \right) \lim_{n \rightarrow \infty} \left( \frac{1}{n} \right) + \lim_{n \rightarrow \infty} \left( \mathbb{1} \left\{ \epsilon \leq \frac{1}{n} \right\} \right) \\ &= \int_{\epsilon}^{\infty} x^{-2} dx \cdot 1 \cdot 0 + 0 \\ &= 0.\end{aligned}$$

**Problem 15.** For convenience, I first write  $X_n$  in summation notation:

$$X_n = \frac{1}{n} \left( \sum_{i=1}^{n-1} Y_i Y_{i+1} + Y_n Y_1 \right).$$

To solve this problem, I will use Chebyshev's inequality, and will thus need to compute  $E[X_n]$  and  $Var[X_n]$ . Computing  $E[X_n]$ :

$$\begin{aligned} E[X_n] &= \frac{1}{n} \left( \sum_{i=1}^{n-1} E[Y_i Y_{i+1}] + E[Y_n Y_1] \right) \\ &= \frac{1}{n} \left( \sum_{i=1}^{n-1} E[Y_i] E[Y_{i+1}] + E[Y_n] E[Y_1] \right) \\ &= \frac{1}{n} \left( \sum_{i=1}^{n-1} \mu^2 + \mu^2 \right) \\ &= \mu^2, \end{aligned}$$

where in the first line I have used the linearity of expectation and in the second I have used the fact that all  $Y$ s are independent.

Solving for  $Var[X_n]$  is slightly more tricky. To do this, I will first need to compute  $Cov[Y_i Y_{i+1}, Y_{i+1} Y_{i+2}]$  for  $i = 1, 2, \dots, n-2$  (I will also need to compute  $Cov[Y_{n-1} Y_n, Y_n Y_1]$  and  $Cov[Y_n Y_1, Y_1 Y_2]$ , but it is not difficult to show that the following computation gives the same answer for these 2 covariances) and  $Var[Y_i Y_{i+1}]$  for  $i = 1, 2, \dots, n-1$  (I will also need to compute  $Var[Y_n Y_1]$  but, again, it is not difficult to show that the following computation gives the same answer for this variance). Computing the covariance:

$$\begin{aligned} Cov[Y_i Y_{i+1}, Y_{i+1} Y_{i+2}] &= E[Y_i Y_{i+1} Y_{i+1} Y_{i+2}] - E[Y_i Y_{i+1}] E[Y_{i+1} Y_{i+2}] \\ &= E[Y_i] E[Y_{i+1}^2] E[Y_{i+2}] - E[Y_i] E[Y_{i+1}]^2 E[Y_{i+2}] \\ &= \mu^2 (E[Y_{i+1}^2] - E[Y_{i+1}]^2) \\ &= \mu^2 \sigma^2, \end{aligned}$$

where in the second line I have used independence. Now I compute the variance:

$$\begin{aligned} Var[Y_i Y_{i+1}] &= E[(Y_i Y_{i+1})^2] - (E[Y_i Y_{i+1}])^2 \\ &= E[Y_i^2 Y_{i+1}^2] - E[Y_i]^2 E[Y_{i+1}]^2 \\ &= E[Y_i^2] E[Y_{i+1}^2] - E[Y_i]^2 E[Y_{i+1}]^2 \\ &= (\sigma^2 + E[Y_i]^2)(\sigma^2 + E[Y_{i+1}]^2) - E[Y_i]^2 E[Y_{i+1}]^2 \\ &= (\sigma^2 + \mu^2)(\sigma^2 + \mu^2) - \mu^2 \mu^2 \\ &= \sigma^4 + 2\sigma^2 \mu^2, \end{aligned}$$

where in the second and third lines I have used independence. I now compute  $Var[X_n]$ :

$$\begin{aligned}
 Var[X_n] &= \frac{1}{n^2} Var \left[ \sum_{i=1}^{n-1} Y_i Y_{i+1} + Y_n Y_1 \right] \\
 &= \frac{1}{n^2} \left( \sum_{i=1}^{n-1} Var[Y_i Y_{i+1}] + Var[Y_n Y_1] + 2 \sum_{i=1}^{n-2} Cov[Y_i Y_{i+1}, Y_{i+1} Y_{i+2}] + 2Cov[Y_{n-1} Y_n, Y_n Y_1] \right. \\
 &\quad \left. + 2Cov[Y_n Y_1, Y_1 Y_2] \right) \\
 &= \frac{1}{n^2} [n(\sigma^4 + 2\sigma^2 \mu^2) + 2n(\mu^2 \sigma^2)] \\
 &= \frac{\sigma^2}{n} (\sigma^2 + 4\mu^2).
 \end{aligned}$$

For the summation of the covariances, I have only summed over the covariances of adjacent pairs of  $Y_i Y_{i+1}$ , since pairs that are 2 or more away from each other have zero covariance since they are independent (since they do not share any  $Y$  random variables). To see why this is the proper summation over the covariances, I illustrate the summation in a matrix form for  $X_5$  below. We must sum all off diagonal terms, however, only adjacent pairs contribute non zero covariance, indicated by the spades in the figure. It is not difficult to see that my summation corresponds exactly to adding the cells containing spades in this figure.

	$Y_1 Y_2$	$Y_2 Y_3$	$Y_3 Y_4$	$Y_4 Y_5$	$Y_5 Y_1$
$Y_1 Y_2$					
$Y_2 Y_3$					
$Y_3 Y_4$					
$Y_4 Y_5$					
$Y_5 Y_1$					

Finally, I complete the problem using Chebyshev's inequality:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(|X_n - \mu^2| \geq \epsilon) &= \lim_{n \rightarrow \infty} P(|X_n - E[X_n]| \geq \epsilon) \\
 &\leq \lim_{n \rightarrow \infty} \frac{Var[X_n]}{\epsilon^2} \\
 &= \lim_{n \rightarrow \infty} \frac{\sigma^2(\sigma^2 + 4\mu^2)}{n\epsilon^2} \\
 &= 0.
 \end{aligned}$$

Since probabilities cannot be less than 0, I conclude that  $\lim_{n \rightarrow \infty} P(|X_n - \mu^2| \geq \epsilon) = 0$ , so that  $X_n \xrightarrow{p} \mu^2$ .

**Problem 16.** Using some simple algebra, since  $X_n = (\prod_{i=1}^n Y_i)^{1/n}$ , I have that  $\ln X_n = \frac{1}{n} \sum_{i=1}^n \ln Y_i$ .

Using the WLLN, I therefore have that:

$$\begin{aligned}\lim_{n \rightarrow \infty} P(|\ln X_n - \gamma| \geq \epsilon) &= \lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \ln Y_i - E[\ln Y_i]\right| \geq \epsilon\right) \\ &= 0.\end{aligned}$$

This therefore implies that  $\ln X_n \xrightarrow{P} \gamma$ . Now, by the Continuous Mapping Theorem (Theorem 7.7 in the book), since  $\exp(\cdot)$  is a continuous function,  $\exp(\ln X_n) \xrightarrow{P} \exp(\gamma)$ , or in other words:  $X_n \xrightarrow{P} e^\gamma$ .

**Problem 17.** To solve this problem, I compute  $E[|Y_n - \lambda|^2]$ , keeping in mind that for a *Poiss*( $\lambda$ ) distribution,  $E[X] = \text{Var}[X] = \lambda$ :

$$\begin{aligned}E[|Y_n - \lambda|^2] &= E[(Y_n - \lambda)^2] \\ &= E\left[\left(\frac{1}{n}X_n - \lambda\right)^2\right] \\ &= E\left[\frac{1}{n^2}(X_n - \lambda n)^2\right] \\ &= \frac{1}{n^2}E[(X_n - E[X_n])^2] \\ &= \frac{1}{n^2}\text{Var}[X_n] \\ &= \frac{1}{n^2}n\lambda \\ &= \frac{\lambda}{n}.\end{aligned}$$

I thus have that  $\lim_{n \rightarrow \infty} E[|Y_n - \lambda|^2] = 0$ , so that  $Y_n \xrightarrow{m.s.} \lambda$ .

**Problem 18.** Using Minkowski's inequality, I have

$$\begin{aligned}E[|X_n + Y_n - (X + Y)|^r] &= E[|(X_n - X) + (Y_n - Y)|^r] \\ &\leq E[|(X_n - X)|^r]^{1/r} + E[|(Y_n - Y)|^r]^{1/r},\end{aligned}$$

so that:

$$\begin{aligned}\lim_{n \rightarrow \infty} E[|X_n + Y_n - (X + Y)|^r] &\leq \lim_{n \rightarrow \infty} E[|(X_n - X)|^r]^{1/r} + \lim_{n \rightarrow \infty} E[|(Y_n - Y)|^r]^{1/r} \\ &= \left(\lim_{n \rightarrow \infty} E[|(X_n - X)|^r]\right)^{1/r} + \left(\lim_{n \rightarrow \infty} E[|(Y_n - Y)|^r]\right)^{1/r} \\ &= 0,\end{aligned}$$

where the last line follows since  $X_n \xrightarrow{L^r} X$  and  $Y_n \xrightarrow{L^r} Y$ . Since,  $|X_n + Y_n - (X + Y)|^r \geq 0$ ,  $E[|X_n + Y_n - (X + Y)|^r] \geq 0$ , so that the inequality must hold with equality, and thus  $X_n + Y_n \xrightarrow{L^r} X + Y$ .

**Problem 19.** To solve this problem, I utilize Theorem 7.5 in the book and show that  $\sum_{n=1}^{\infty} P(|X_n| > \epsilon)$  (for all  $\epsilon > 0$ ) is finite:

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|X_n| > \epsilon) &= \sum_{n=1}^{\infty} P(X_n > \epsilon) \\
&= \sum_{n=1}^{\infty} n^2 \int_{\epsilon}^{\infty} x e^{-\frac{n^2 x^2}{2}} dx \\
&= \sum_{n=1}^{\infty} e^{-\frac{n^2 \epsilon^2}{2}},
\end{aligned}$$

where in the first line I have used the fact that the random variable  $X_n$  is never negative and in the third line I have solved the integral with a substitution of  $u = n^2 \epsilon^2 / 2$ .

Now, it is not difficult to show that for positive  $\mu$  and  $n \geq 1$ , as we have in this case, that  $e^{-x^2 \mu} \leq e^{-x \mu}$ , and therefore, each term in the above summation is  $\leq e^{-n \epsilon^2 / 2}$ :

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|X_n| > \epsilon) &\leq \sum_{n=1}^{\infty} e^{-\frac{n \epsilon^2}{2}} \\
&= \frac{e^{-\frac{\epsilon^2}{2}}}{1 - e^{-\frac{\epsilon^2}{2}}} \\
&< \infty \quad (\text{for } \epsilon > 0).
\end{aligned}$$

**Problem 20.** Note that for  $X_2 = Y_1, X_3 = Y_1 Y_2, X_4 = Y_1 Y_2 Y_3, \dots$ , with  $Y_n \sim \text{Bern}(n/(n+1))$ ,  $R_{X_n} = \{0, 1\}$ . It is therefore not difficult to show that for this sequence, for  $0 < \epsilon < 1$ :

$$\begin{aligned}
\sum_{n=2}^{\infty} P(|X_n| > \epsilon) &= \sum_{k=1}^{\infty} \frac{1}{k+1} \\
&= \infty.
\end{aligned}$$

Therefore, we cannot simply appeal to Theorem 7.5 and must thus use Theorem 7.6. That is, we must show that for any  $\epsilon > 0$ ,  $\lim_{m \rightarrow \infty} P(A_m) = 1$ , where the set  $A_m$  is defined in the book. I show this for  $0 < \epsilon < 1$ . For this interval, from the definition of  $A_m$ :

$$\begin{aligned}
A_m &= \{|X_n| < \epsilon, \forall n \geq m\} \\
&= \{X_n < \epsilon, \forall n \geq m\} \\
&= \{X_n = 0, \forall n \geq m\}.
\end{aligned}$$

Evaluating the probability of this event, I have that:

$$\begin{aligned}
P(A_m) &= P(\{X_n = 0, \forall n \geq m\}) \\
&= P(\{X_m = 0, X_{m+1} = 0, \dots\}) \\
&= P(Y_{m-1} Y_{m-2} \dots Y_1 = 0, Y_m Y_{m-1} \dots Y_1 = 0, \dots) \\
&= P(Y_{m-1} Y_{m-2} \dots Y_1 = 0) P(Y_m Y_{m-1} \dots Y_1 = 0, Y_{m+1} Y_m \dots Y_1 = 0, \dots | Y_{m-1} Y_{m-2} \dots Y_1 = 0) \\
&= P(Y_{m-1} Y_{m-2} \dots Y_1 = 0) \\
&= 1 - P((Y_{m-1} = 0 \cup Y_{m-2} = 0 \cup \dots \cup Y_1 = 0)^c) \\
&= 1 - P(Y_{m-1} = 1, Y_{m-2} = 1, \dots, Y_1 = 1) \\
&= 1 - \prod_{k=1}^{m-1} \frac{k}{k+1},
\end{aligned}$$

where in the fifth line I have used the fact that if  $Y_{m-1}Y_{m-2}\dots Y_1 = 0$ , then at least 1  $Y_i$  ( $i = 1, \dots, m-1$ ) is 0. Since the random variables  $Y_m Y_{m-1} \dots Y_1, Y_{m+1} Y_m \dots Y_1, \dots$  are all products of  $Y_{m-1} Y_{m-2} \dots Y_1$ , given that  $Y_{m-1} Y_{m-2} \dots Y_1 = 0$ , we know for sure that all random variables  $Y_m Y_{m-1} \dots Y_1, Y_{m+1} Y_m \dots Y_1, \dots$  are 0. In the seventh line I have used De Morgan's law, and in the eighth I have used independence.

The product is easily solved:

$$\prod_{k=1}^{m-1} \frac{k}{k+1} = \frac{1}{2} \cdot \frac{2}{3} \cdots \frac{m-2}{m-1} \cdot \frac{m-1}{m} = \frac{1}{m},$$

where all denominators have cancelled out with the next numerator except for the last one. I therefore have that

$$\begin{aligned} \lim_{m \rightarrow \infty} P(A_m) &= \lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right) \\ &= 1, \end{aligned}$$

and therefore, by Theorem 7.6  $X_n \xrightarrow{a.s.} 0$ .

## Chapter 8

# Statistical Inference I: Classical Methods

**Problem 1.**

- (a) Using the formulas for the sample mean, sample variance and sample standard deviation, I find that:

$$\bar{X} \approx 164.3 \text{ lbs},$$

$$S^2 \approx 383.7 \text{ lbs}^2,$$

and

$$S \approx 19.59 \text{ lbs}.$$

**Problem 2.** To calculate the bias of this estimator, I first compute its expectation:

$$\begin{aligned} E[\hat{\Theta}] &= E \left[ \left( \frac{1}{n} \sum_{k=1}^n X_k \right)^2 \right] \\ &= \frac{1}{n^2} E \left[ \sum_{k=1}^n X_k^2 + \sum_{i,j:i \neq j} X_i X_j \right] \\ &= \frac{1}{n^2} \left( \sum_{k=1}^n E[X_k^2] + \sum_{i,j:i \neq j} E[X_i] E[X_j] \right) \\ &= \frac{1}{n^2} (n(\sigma^2 + E[X_k]^2) + (n^2 - n)E[X_k]^2) \\ &= \frac{1}{n^2} (n(\sigma^2 + \mu^2) + (n^2 - n)\mu^2) \\ &= \frac{\sigma^2}{n} + \mu^2, \end{aligned}$$

where the notation  $\sum_{i,j:i \neq j}$  refers to a sum over all pairs of  $i, j$  ( $i, j = 1, \dots, n$ ) except for the pairs where  $i = j$ . In the third line I have used the linearity of expectation and independence. The bias is thus:

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta = E[\hat{\Theta}] - \mu^2 = \frac{\sigma^2}{n},$$

and since  $B(\hat{\Theta}) \neq 0$ ,  $\hat{\Theta}$  is a biased estimator of  $\theta$ .

**Problem 3.**

- (a) To solve this problem, I first compute the expectation of  $X_i$ :

$$\begin{aligned} E[X_i] &= \int_0^1 \left[ \theta \left( x - \frac{1}{2} \right) + 1 \right] x dx \\ &= \frac{\theta}{3} - \frac{\theta}{4} + \frac{1}{2}. \end{aligned}$$



I now compute the expectation of the estimator:

$$\begin{aligned}
 E[\hat{\Theta}_n] &= E[12\bar{X} - 6] \\
 &= E\left[\frac{12}{n} \sum_{i=1}^n X_i - 6\right] \\
 &= \frac{12}{n} \sum_{i=1}^n E[X_i] - 6 \\
 &= \frac{12}{n} \sum_{i=1}^n \left(\frac{\theta}{3} - \frac{\theta}{4} + \frac{1}{2}\right) - 6 \\
 &= 12 \left(\frac{\theta}{3} - \frac{\theta}{4} + \frac{1}{2}\right) - 6 \\
 &= \theta.
 \end{aligned}$$

I therefore have that  $B(\hat{\Theta}_n) = E[\hat{\Theta}_n] - \theta = 0$ , and so  $\hat{\Theta}_n$  is an unbiased estimator of  $\theta$ .

- (b) I will use Chebyshev's inequality to show that this is a consistent estimator and I will therefore need to compute  $Var[\hat{\Theta}_n]$ . To do this, I first compute  $E[X_i^2]$ :

$$\begin{aligned}
 E[X_i^2] &= \int_0^1 \left[ \theta \left(x - \frac{1}{2}\right) + 1 \right] x^2 dx \\
 &= \frac{\theta}{4} - \frac{\theta}{6} + \frac{1}{3}.
 \end{aligned}$$

Now I compute  $E[\hat{\Theta}_n^2]$ :

$$\begin{aligned}
 E[\hat{\Theta}_n^2] &= E[(12\bar{X} - 6)^2] \\
 &= E\left[\frac{12^2}{n^2} \left(\sum_{i=1}^n X_i\right)^2 - 6 \cdot 2 \cdot \frac{12}{n} \sum_{i=1}^n X_i + 36\right] \\
 &= \frac{144}{n^2} \left( \sum_{i=1}^n E[X_i^2] + \sum_{i,j:i \neq j} E[X_i]E[X_j] \right) - \frac{144}{n} \sum_{i=1}^n E[X_i] + 36 \\
 &= \frac{144}{n^2} (nE[X_i^2] + (n^2 - n)E[X_i]^2) - 144E[X_i] + 36,
 \end{aligned}$$

where the notation  $\sum_{i,j:i \neq j}$  refers to a sum over all pairs of  $i, j$  ( $i, j = 1, \dots, n$ ) except for the pairs where  $i = j$ . In this derivation, I have used the linearity of expectation and independence. Plugging in  $E[X_i^2]$  and  $E[X_i]$  and simplifying, I find that

$$E[\hat{\Theta}_n^2] = \frac{12}{n} + \theta^2 \left(1 - \frac{1}{n}\right),$$

so that

$$\begin{aligned}
 Var[\hat{\Theta}_n] &= E[\hat{\Theta}_n^2] - E[\hat{\Theta}_n]^2 \\
 &= \frac{12}{n} + \theta^2 \left(1 - \frac{1}{n}\right) - \theta^2 \\
 &= \frac{12 - \theta^2}{n}.
 \end{aligned}$$

To show that  $\hat{\Theta}_n$  is a consistent estimator, I must show that  $\hat{\Theta}_n \xrightarrow{P} \theta$ . Using Chebyshev's inequality, I have:

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) &= \lim_{n \rightarrow \infty} P(|\hat{\Theta}_n - E[\hat{\Theta}_n]| \geq \epsilon) \\ &\leq \lim_{n \rightarrow \infty} \frac{\text{Var}[\hat{\Theta}_n]}{\epsilon^2} \\ &= \lim_{n \rightarrow \infty} \frac{12 - \theta^2}{n\epsilon^2} \\ &= 0. \end{aligned}$$

Since probabilities cannot be negative, I have that  $\lim_{n \rightarrow \infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0$  for all  $\epsilon > 0$ , and thus  $\hat{\Theta}_n$  is consistent.

- (c) Since I have already computed the variance and bias of  $\hat{\Theta}_n$ , computing the mean squared error is easy:

$$\text{MSE}(\hat{\Theta}_n) = \text{Var}[\hat{\Theta}_n] + B[\hat{\Theta}_n]^2 = \frac{12 - \theta^2}{n}.$$

**Problem 4.**

$$\begin{aligned} L(x_1, \dots, x_4; p) &= \prod_{i=1}^4 P_{X_i}(x_i; p) \\ &= \prod_{i=1}^4 p(1-p)^{x_i-1} \\ &= p^4(1-p)^{2-1}(1-p)^{3-1}(1-p)^{3-1}(1-p)^{5-1} \\ &= p^4(1-p)^9 \end{aligned}$$

**Problem 5.**

$$\begin{aligned} L(x_1, \dots, x_4; \theta) &= \prod_{i=1}^4 f_{X_i}(x_i; \theta) \\ &= \prod_{i=1}^4 \theta e^{-\theta x_i} \\ &= \theta^4 e^{-2.35\theta} e^{-1.55\theta} e^{-3.25\theta} e^{-2.65\theta} \\ &= \theta^4 e^{-9.8\theta} \end{aligned}$$

**Problem 6.** Since  $\log(\cdot)$  is a monotonic increasing function on  $\mathbb{R}$ ,  $\text{argmax}_{\theta \in \mathbb{R}} L(\mathbf{x}; \theta) = \text{argmax}_{\theta \in \mathbb{R}} \log L(\mathbf{x}; \theta)$ . This can easily be proven by considering the definition of a strictly monotonic function.

**Problem 7.**

- (a) For a single data point,  $X$ , and our estimator  $\hat{\Theta}$  (which is a function,  $f$ , of  $X$ ) of  $\sigma^2$ , we have that  $E[\hat{\Theta}] = E[f(X)] = \sigma^2$ , where the last equality follows because we want the estimator to be unbiased. Therefore, we are searching for a function such that:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} f(x) e^{-\frac{x^2}{2\sigma^2}} dx = \sigma^2.$$

Since the PDF is that of  $\mathcal{N}(0, \sigma^2)$ , (i.e., it has mean zero), it is clear that the function that satisfies this equation is  $f(x) = x^2$ , and therefore  $\hat{\Theta} = X^2$ .

(b)

$$\ln L(x; \sigma^2) = -\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{x^2}{2\sigma^2}$$

(c) Taking the derivative of the above equation with respect to  $\sigma$ ,

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{1}{\sigma} + \frac{x^2}{\sigma^3},$$

setting equal to zero,

$$0 = -\frac{1}{\hat{\sigma}_{ML}} + \frac{x^2}{\hat{\sigma}_{ML}^3},$$

and solving for  $\hat{\sigma}_{ML}$ , I find that  $\hat{\sigma}_{ML} = |x|$ .**Problem 8.**

(a)

$$\begin{aligned} L(x_1, \dots, x_n; \lambda) &= \prod_{i=1}^n P_{X_i}(x_i; \lambda) \\ &= \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!} \end{aligned}$$

(b) The log-likelihood,  $\ell$ , is:

$$\begin{aligned} \ell(\lambda) &= \ln L(x_1, \dots, x_n; \lambda) \\ &= \ln e^{-n\lambda} + \ln \lambda^{\sum_{i=1}^n x_i} + \sum_{i=1}^n \ln x_i!^{-1} \\ &= -n\lambda + \sum_{i=1}^n x_i \ln \lambda - \sum_{i=1}^n \ln x_i!. \end{aligned}$$

Differentiating this respect to  $\lambda$ , and setting equal to zero, I have:

$$0 = -n + \frac{1}{\hat{\lambda}_{ML}} \sum_{i=1}^n x_i.$$

Solving for the maximum likelihood estimate, I have that:

$$\hat{\lambda}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i,$$

that is, the maximum likelihood estimate of  $\lambda$  is simply the sample mean.

**Problem 9.** To solve for the CDF for the  $i^{th}$  order statistic, let us assume that  $X_1, X_2, \dots, X_n$  are a random sample from a continuous distribution with CDF,  $F_X(x)$ . I fix a value  $x \in \mathbb{R}$ , and define the indicator random variable,  $I_j$ , by

$$I_j(X_j) = \begin{cases} 1 & \text{if } X_j \leq x \\ 0 & \text{if } X_j > x, \end{cases}$$

where  $I_j = 1$  is a “success” and  $I_j = 0$  is a “failure.” Note that, since all  $X_j$ s are *iid*, the probability of a success,  $P(X_j \leq x)$ , is the same for each trial and is given by  $F_X(x)$ . Therefore, I have that  $I_j \stackrel{iid}{\sim} \text{Bern}(F_X(x))$ . I now define the random variable,  $Y = \sum_{j=1}^n I_j$ , and since this is the sum of  $n$  independent Bernoulli random trials, it has a distribution:  $Y \sim \text{Bin}(n, F_X(x))$ .

Now, given that  $Y \sim \text{Bin}(n, F_X(x))$ , the quantity,  $P(Y \geq i)$  is therefore the probability that there are at least  $i$  successes out of  $n$  trials. Given our definition of “success”, and given that the number of trials  $n$  is simply the number of observations, this can be re-phrased as the probability that there are at least  $i$  observations out of  $n$  with values less than or equal to  $x$ .

We desire to find  $P(X_{(i)} \leq x)$ , the probability that the  $i^{\text{th}}$  biggest observation out of  $n$  observations has a value less than or equal to  $x$ . In other words, we desire to find the probability that there are at least  $i$  observations out of  $n$  with a value less than or equal to  $x$ . Notice that this is exactly  $P(Y \geq i)$ , so that:

$$\begin{aligned} F_{X_{(i)}}(x) &= P(X_{(i)} \leq x) \\ &= P(Y \geq i) \\ &= \sum_{k=i}^n \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k}. \end{aligned}$$

**Problem 10.** Let region 1 be defined as the interval  $(-\infty, x]$ , region 2 as the interval  $(x, x + \delta]$ , (where  $\delta$  is a small positive number) and region 3 as the interval  $(x + \delta, \infty)$ . By the definition of the PDF, the probability that the  $i^{\text{th}}$  order statistic is in region 2 is given by  $P(x < X_{(i)} \leq x + \delta) \approx f_{X_{(i)}}(x)\delta$ . In other words, for  $\delta$  small enough,  $P(x < X_{(i)} \leq x + \delta)$  is the probability that, out of  $n$  samples, there are  $i - 1$  samples in region 1, one in region 2 and  $n - i$  in region 3.

Now, since all samples are *iid* from a distribution with PDF  $f_X(x)$  and CDF  $F_X(x)$ , the probability that a sample lands in region 1, is

$$p_1 = P(X \leq x) = F_X(x),$$

in region 2 is

$$p_2 = P(x \leq X \leq x + \delta) \approx f_X(x)\delta,$$

and in region 3 is

$$p_3 = P(X > x + \delta) = 1 - F_X(x + \delta) \approx 1 - F_X(x).$$

Notice that if we define  $s_i$  as the event that a sample, out of  $n$  samples, lands in region  $i$  (with associated probability,  $p_i$ ), this is precisely a multinomial experiment with 3 possible outcomes. Thus the probability that out of  $n$  samples (trials), there are  $i - 1$  in region 1, one in region 2 and  $n - i$  in region 3 is given by:

$$\frac{n!}{(i-1)!(n-1)!} p_1^{i-1} p_2 p_3^{n-1}.$$

However, this is precisely the probability  $f_{X_{(i)}}(x)\delta$ . Therefore, I have that:

$$\begin{aligned} f_{X_{(i)}}(x)\delta &= \frac{n!}{(i-1)!(n-1)!} p_1^{i-1} p_2 p_3^{n-1} \\ &= \frac{n!}{(i-1)!(n-1)!} [F_X(x)]^{i-1} f_X(x)\delta [1 - F_X(x)]^{n-1}. \end{aligned}$$

Canceling the  $\delta$  from both sides of the equation gives the desired result.

**Problem 11.** Since  $n$  is relatively large, the variance is known, and we would like an approximate confidence interval for  $\theta = E[X_i]$ , we can calculate the confidence interval by employing the CLT and by using  $\sqrt{n}(\bar{X} - \theta)$  as the pivotal quantity. This computation is done in the book and the interval is given by:

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

The quantities in this interval are  $\bar{X} = 50.1$ ,  $\sigma = 9$ ,  $n = 100$  and  $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = 1.96$ . Using these values, I find that the 95% confidence interval is given by:  $[48.3, 51.9]$ . Note that  $z_{\frac{\alpha}{2}}$  can be computed in Python with `scipy.stats.norm.ppf(1-alpha/2)`.

**Problem 12.** In this problem, we choose a random sample of size  $n$  from a population,  $X_1, X_2, \dots, X_n$ , where these random variables are *iid*  $Bern(\theta)$ , where  $X_i$  is 1 if the  $i^{th}$  voter intends to vote for Candidate A, and 0 otherwise.

- (a) We require there to be at least a 90% probability that the sample proportion,  $\bar{X}$  is within 3 percentage points of the actual proportion,  $\theta$ . In math, this is:

$$P(\theta - 0.03 \leq \bar{X} \leq \theta + 0.03) \geq 0.9,$$

and algebraically manipulating the argument, it is easy to show that the 90% confidence interval we require is:

$$[\bar{X} - 0.03, \bar{X} + 0.03].$$

Following along Example 8.18 in the book, utilizing the CLT, and obtaining a conservative estimate for the interval by using  $\sigma_{max}$ , which for a Bernoulli distribution is  $1/2$  (since we do not actually know  $\sigma$ ), the proper interval is given by:

$$\left[ \bar{X} - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}, \bar{X} + \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}} \right].$$

Comparing this interval with the one above, we see that:

$$\frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}} = 0.03.$$

Thus, we require  $n$  to be at least:

$$\left\lceil \left( \frac{z_{\frac{\alpha}{2}}}{2 \cdot 0.03} \right)^2 \right\rceil = 748,$$

where I have used  $z_{\frac{\alpha}{2}} = z_{\frac{0.1}{2}} \approx 1.64$ .

- (b) Using the same formula as above, but with  $z_{\frac{\alpha}{2}} = z_{\frac{0.01}{2}} \approx 2.58$ , I find that  $n$  must be at least 1849.

**Problem 13.** For this problem, since  $n$  is relatively large, I use the standard approximate confidence interval derived using the CLT. The variance, however, is unknown, but since  $n$  is large should be well approximated by the sample variance,  $S^2$ . The proper confidence interval is thus:

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right],$$

and using  $n = 100$ ,  $\bar{X} = 110.5$ ,  $S^2 = 45.6$ , and  $z_{\frac{0.05}{2}} = 1.96$ , I find the 95% confidence interval for the distribution mean to be approximately be:  $[109.2, 111.8]$ .

**Problem 14.**

- (a) For an  $n = 36$  random sample from  $\mathcal{N}(\mu, \sigma^2)$ , with  $\mu$  and  $\sigma^2$  unknown, the proper pivotal quantity to use to estimate  $\mu$  is  $T = (\bar{X} - \mu)/(S/\sqrt{n})$ , which because it has a  $T$  distribution, results in a confidence interval of:

$$\left[ \bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \right],$$

as shown in the book. For the desired confidence levels (90%, 95%, 99%), the appropriate  $t$  values are:  $t_{0.1, 35} \approx 1.69$ ,  $t_{0.05, 35} \approx 2.03$ ,  $t_{0.01, 35} \approx 2.72$ , and the corresponding confidence intervals are:  $[34.8, 36.8]$ ,  $[34.6, 37.0]$  and  $[34.2, 37.4]$ . We see that as the confidence level increases, the width of the interval gets wider since we desire more confidence that the actual value of  $\mu$  is encompassed by that random interval. Note that  $t_{\frac{\alpha}{2}, n-1}$  can be computed in Python with `scipy.stats.t.ppf(1-alpha/2, n-1)`.

- (b) The proper pivotal quantity to use to estimate  $\sigma^2$  is  $Q = (n-1)S^2/\sigma^2$ , which because it has a  $\chi^2$  distribution, results in a confidence interval of:

$$\left[ \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2} \right],$$

as shown in the book. Computing the proper  $\chi_{\frac{\alpha}{2}, n-1}^2$  and  $\chi_{1-\frac{\alpha}{2}, n-1}^2$  values, I find the following 90%, 95% and 99% confidence intervals for  $\sigma^2$ :  $[8.78, 19.47]$ ,  $[8.22, 21.3]$  and  $[7.26, 25.4]$ . Again, we see that as the confidence level increases, the width of the interval gets wider since we desire more confidence that the actual value of  $\sigma^2$  is encompassed by that random interval. Note that  $\chi_{\frac{\alpha}{2}, n-1}^2$  can be computed in Python with `scipy.stats.chi2.ppf(1-alpha/2, n-1)`.

**Problem 15.**

- (a) We recognize that since the data are drawn *iid* from a normal distribution, since  $\sigma^2$  is known, and since the hypotheses are of the form  $H_o : \mu = \mu_o$  and  $H_A : \mu \neq \mu_o$ , this is a 2-sided  $z$ -test, as outlined in Table 8.2 in the book. Thus, if the statistic  $W = (\bar{X} - \mu_o)/(\sigma/\sqrt{n})$  satisfies  $|W| \leq z_{\frac{\alpha}{2}}$  then we fail to reject the null hypothesis, otherwise we reject it in favor of the alternative hypothesis

Computing  $\bar{X}$  and  $W$ , I find:

$$\bar{X} \approx 5.96,$$

and

$$W \approx 2.15.$$

At a level of  $\alpha = 0.05$ , the proper threshold is  $z_{0.025} \approx 1.96$ . Since  $W > z_{\frac{\alpha}{2}}$ , we reject  $H_o$  in favor of  $H_A$  at a significance level of 0.05.

- (b) In this case, since the data are drawn *iid* from a normal distribution with known variance, the proper  $(1 - \alpha)100\%$  confidence interval to use as shown in Section 8.3.3 of the book is:

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right],$$

which, when plugging in the particular values for this problem results in a 95% confidence interval of approximately  $[5.08, 6.84]$ .

The value  $\mu_o = 5$  is not within this interval. As shown in Section 8.4.3 of this book, for this type of hypothesis test, since we accept  $H_o$  at a level of  $\alpha$  if  $|(\bar{X} - \mu_o)/(\sigma/\sqrt{n})| \leq z_{\frac{\alpha}{2}}$ , this results in the condition that we accept  $H_o$  if:

$$\mu_o \in \left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

I.e., for this test, if  $\mu_o$  is in the  $(1 - \alpha)100\%$  confidence interval, we accept  $H_o$  at a level of  $\alpha$ , otherwise we do not. Since  $\mu_o = 5$  is not in the calculated confidence interval, this corresponds to rejecting  $H_o$  in favor of  $H_A$ , which is indeed what we found above.

**Problem 16.**

- (a) As with the previous problem, since the data are drawn *iid* from a normal distribution with known variance, the proper  $(1 - \alpha)100\%$  confidence interval to use as shown in Section 8.3.3 of the book is:

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

For this problem  $\bar{X} \approx 17.0$ ,  $z_{\frac{\alpha}{2}} = z_{0.05} \approx 1.64$ , so that the 90% confidence interval is approximately  $[16.45, 17.55]$ . The value of  $\mu_o$  is not included in this interval, which, as explained above, means that we reject the null hypothesis at a significance level of  $\alpha = 0.1$ .

- (b) As shown in Section 8.4.3 of the book, the proper test statistic to use is  $W = (\bar{X} - \mu_o)/(\sigma/\sqrt{n})$ , and if  $|W| \leq z_{\frac{\alpha}{2}}$  (see Table 8.2) we cannot reject the null hypothesis. For this problem,  $W \approx 3$ ,  $z_{\frac{\alpha}{2}} = z_{0.05} \approx 1.64$ , and therefore we reject  $H_o$  at a significance level of  $\alpha = 0.1$ .

**Problem 17.** In this problem, the random sample comes from an unknown distribution with unknown variance and with a rather large  $n$  ( $n = 150$ ). Since the hypotheses we would like to test correspond to  $H_o : \mu = 50$  and  $H_A : \mu > 50$ , this will most likely be a 1-sided  $z$ -test (using the sample variance). Here I work out the test explicitly. Using  $W$  as my statistic,  $(\bar{X} - \mu_o)/(S/\sqrt{n})$ , If  $H_o$  is true, then we would expect  $\bar{X} \approx \mu_o$  and  $W \approx 0$ . On the other hand, if  $H_A$  is true we expect  $\bar{X} > \mu_o$  and  $W > 0$ . Therefore I employ the following test: if  $W \leq c$  I fail to reject  $H_o$ , while if  $W > c$  I reject  $H_o$  in favor of  $H_A$ .

To solve for  $c$  I must bound the probability of making a Type I error:

$$\begin{aligned} P(\text{Type I error}) &= P(\text{reject } H_o | H_o) \\ &= P(W > c | H_o) \\ &= 1 - \Phi(c) \quad (\text{since } W \sim \mathcal{N}(0, 1) \text{ under } H_o) \\ &\leq \alpha \end{aligned}$$

Therefore, the critical value,  $c$ , occurs at equality:  $1 - \Phi(c) = \alpha$ , or in other words  $c = z_\alpha$ . Thus, if  $W \leq z_\alpha$  I fail to reject  $H_o$ , while if  $W > z_\alpha$  I reject  $H_o$  in favor of  $H_A$ .

For this problem  $\bar{X} = 52.28$ ,  $S^2 = 30.9$ , and so  $W \approx 5.02$ , while  $z_\alpha = z_{0.05} \approx 1.64$ , so that I reject  $H_o$  in favor of  $H_A$  at a significance level of 0.05.

**Problem 18.** In this problem, the random sample comes from a normal distribution with unknown variance, and the hypotheses we are testing are of the form  $H_o : \mu \geq \mu_o$  and  $H_A : \mu < \mu_o$ , and I therefore use a 1-sided  $t$ -test. As indicated in Table 8.4, we fail to reject  $H_o$  if  $W \geq -t_{\alpha, n-1}$ . For this problem,

$$\bar{X} = \frac{27.72 + 22.24 + 32.86 + 19.66 + 35.34}{5} \approx 27.56,$$

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \approx 44.84,$$

so that

$$W = \frac{\bar{X} - \mu}{S/\sqrt{n}} \approx \frac{27.56 - 30}{\sqrt{44.84}/\sqrt{5}} \approx -0.81.$$

Also, for this problem  $-t_{\alpha, n-1} = -t_{0.05, 4} \approx -2.13$ . Since  $W \geq -t_{\alpha, n-1}$ , we fail to reject the null hypothesis at a level of 0.05.

**Problem 19.** Since the random sample is drawn from an unknown distribution with unknown variance, but the sample number is relatively large ( $n = 121$ ), I can use a  $z$ -test with the sample variance. Moreover, the hypotheses we are testing are of the form  $H_o : \mu = \mu_o$  and  $H_A : \mu < \mu_o$ . I therefore use a 1-sided  $z$ -test, where if  $W < -z_\alpha$  with  $W = (\bar{X} - \mu_o)/(S/\sqrt{n})$ , then I reject the null hypothesis at a significance level of  $\alpha$  (see table 8.4).

The  $p$ -value is the probability of making a Type I error when the statistic threshold is set to that which was observed ( $w_1$ ), in this case  $w_1 \approx -0.81$ . Thus, the  $p$ -value for this problem is:

$$\begin{aligned} p - \text{value} &= P(\text{Type I error when } c = w_1 | H_o) \\ &= P(W < w_1 | H_o) \\ &= \Phi(w_1) \\ &\approx 0.035, \end{aligned}$$

where, because of the CLT, I have used the CDF of a Gaussian.

**Problem 20.**

- (a) We would like to test the hypotheses that  $H_o : \theta \geq 0.1$  and  $H_A : \theta < 0.1$ , which, since equality gives us a worse case scenario (as shown in Section 8.4.3 of the book), can be simplified to:

$$\begin{aligned} H_o : \theta &= \theta_o = 0.1 \\ H_A : \theta &< \theta_o. \end{aligned}$$

- (b) If we let  $X_i = 1$  if the  $i^{\text{th}}$  student has allergies, and 0 otherwise, we see that  $X_i \sim \text{Bern}(\theta)$ , so that  $E[X_i] = \theta$  and  $\text{Var}[X_i] = \theta(1 - \theta)$ . Now, under the null hypothesis, and under the CLT (since  $n$  is large), I have that

$$\frac{\bar{X}n - \theta_o n}{\sqrt{n\theta_o(1 - \theta_o)}} \sim \mathcal{N}(0, 1).$$

This is a convenient test statistic to use since I have its distribution, and since if the alternative hypothesis is true,  $\bar{X}$  will be small (and so will the statistic), while if the null hypothesis is true,  $\bar{X}$  will be large (and so will the statistic). This suggests the following test: if

$$\frac{\bar{X}n - \theta_o n}{\sqrt{n\theta_o(1 - \theta_o)}} < c$$

then reject the null hypothesis in favor of the alternative hypothesis, while if

$$\frac{\bar{X}n - \theta_o n}{\sqrt{n\theta_o(1 - \theta_o)}} \geq c,$$



fail to reject the null hypothesis.

Calculating the value of the statistic for the particular instance of the data that we have collected:

$$w_1 = \frac{21 - 0.1 \cdot 225}{\sqrt{225 \cdot 0.1 \cdot (1 - 0.1)}} \approx -0.33.$$

Now, the  $p$ -value is the probability of making a Type I error when the test threshold,  $c$ , is set to be  $w_1$ :

$$\begin{aligned} p - \text{value} &= P(\text{Type I error with } c = w_1) \\ &= P(\text{reject } H_o \text{ with } c = w_1 | H_o) \\ &= P\left(\frac{\bar{X}n - \theta_o}{\sqrt{n\theta_o(1 - \theta_o)}} < w_1 | H_o\right) \\ &= \Phi(w_1) \\ &\approx 0.37. \end{aligned}$$

- (c) Since the  $p$ -value is the lowest significance level  $\alpha$  that results in rejecting the null hypothesis, at a level of  $\alpha = 0.05$ , we cannot reject the null hypothesis.

**Problem 21.**

- (a) Using the equations for simple linear regression I have the following:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{-5 - 3 + 0 + 2 + 1}{5} = -1,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{-2 + 1 + 4 + 6 + 3}{5} = 2.4,$$

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = (-5 + 1)^2 + (-3 + 1)^2 + (0 + 1)^2 + (2 + 1)^2 + (1 + 1)^2 = 34,$$

$$\begin{aligned} s_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = (-5 + 1)(-2 - 2.4) + (-3 + 1)(1 - 2.4) + (0 + 1)(4 - 2.4) \\ &\quad + (2 + 1)(6 - 2.4) + (1 + 1)(3 - 2.4) = 34, \end{aligned}$$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{34}{34} = 1,$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2.4 - 1(-1) = 3.4.$$

The regression line is given by  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ , and therefore

$$\hat{y} = 3.4 + x.$$

(b) The regression predictions for the training data are:

$$\hat{y}_1 = 3.4 - 5 = -1.6$$

$$\hat{y}_2 = 3.4 - 3 = 0.4$$

$$\hat{y}_3 = 3.4 + 0 = 3.4$$

$$\hat{y}_4 = 3.4 + 2 = 5.4$$

$$\hat{y}_5 = 3.4 + 1 = 4.4.$$

(c) The residuals are:

$$e_1 = y_1 - \hat{y}_1 = -2 + 1.6 = -0.4$$

$$e_2 = y_2 - \hat{y}_2 = 1 - 0.4 = 0.6$$

$$e_3 = y_3 - \hat{y}_3 = 4 - 3.4 = 0.6$$

$$e_4 = y_4 - \hat{y}_4 = 6 - 5.4 = 0.6$$

$$e_5 = y_5 - \hat{y}_5 = 3 - 4.4 = -1.4.$$

As a check, we know that  $\sum_{i=1}^n e_i = 0$ , which is indeed the case.

(d) To calculate the coefficient of determination, I first need to compute  $s_{yy}$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = (-2 - 2.4)^2 + (1 - 2.4)^2 + (4 - 2.4)^2 + (6 - 2.4)^2 + (3 - 2.4)^2 = 37.2,$$

so that

$$r^2 = \frac{s_{xy}^2}{s_{xx}s_{yy}} = \frac{34^2}{34 \cdot 37.2} \approx 0.91.$$

### Problem 22.

(a) Using the equations for simple linear regression I have the following:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1+3}{2} = 2,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{3+7}{2} = 5,$$

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = (1-2)^2 + (3-2)^2 = 2,$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = (1-2)(3-5) + (3-2)(7-5) = 4$$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{4}{2} = 2,$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 5 - 2 \cdot 2 = 1.$$

The regression line is given by  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ , and therefore

$$\hat{y} = 1 + 2x.$$

(b) The regression predictions for the training data are:

$$\hat{y}_1 = 1 + 2 \cdot 1 = 3$$

$$\hat{y}_2 = 1 + 2 \cdot 3 = 7.$$

(c) The residuals are:

$$e_1 = y_1 - \hat{y}_1 = 3 - 3 = 0$$

$$e_2 = y_2 - \hat{y}_2 = 7 - 7 = 0.$$

As a check, we know that  $\sum_{i=1}^n e_i = 0$ , which is indeed the case.

(d) To calculate the coefficient of determination, I first need to compute  $s_{yy}$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = (3 - 5)^2 + (7 - 5)^2 = 8,$$

so that

$$r^2 = \frac{s_{xy}^2}{s_{xx}s_{yy}} = \frac{4^2}{2 \cdot 8} = 1.$$

(e) Since there are only 2 data points in the training set, the regression line that minimizes the sum of squared errors goes exactly through those 2 points, and thus  $r^2 = 1$ . This is a good fit to the training data, however, it will probably not generalize well to new, unseen data, and is probably therefore a poor predictive model.

### Problem 23.

(a) According to this model,  $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ . To solve for the distribution of  $\hat{\beta}_1$ , note that it is fairly easy to find the distribution of a sum (or a linear combination) of independent normal random variables. However, due to the fact that each  $Y_i$  is in the sum that comprises  $\bar{Y}$ , clearly the term,  $S_{xy}$  is not a sum of independent random variables. In order to express the formula for  $\hat{\beta}_1$  as a linear combination of the  $Y_i$ s (which are independent), I expand the

formula for  $\hat{\beta}_1$  and group each  $Y_i$  term. With  $c_i \equiv (x_i - \bar{x})$ , I have:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{S_{xy}}{s_{xx}} \\
&= \frac{1}{s_{xx}} \sum_{i=1}^n c_i (Y_i - \bar{Y}) \\
&= \frac{1}{s_{xx}} \left\{ c_1 \left[ Y_1 - \frac{1}{n} (Y_1 + Y_2 + \dots + Y_n) \right] + c_2 \left[ Y_2 - \frac{1}{n} (Y_1 + Y_2 + \dots + Y_n) \right] \right. \\
&\quad \left. + \dots + c_n \left[ Y_n - \frac{1}{n} (Y_1 + Y_2 + \dots + Y_n) \right] \right\} \\
&= \frac{1}{s_{xx}} \left\{ \left[ Y_1 \left( c_1 - \frac{c_1}{n} - \frac{c_2}{n} - \dots - \frac{c_n}{n} \right) \right] + \left[ Y_2 \left( c_2 - \frac{c_1}{n} - \frac{c_2}{n} - \dots - \frac{c_n}{n} \right) \right] \right. \\
&\quad \left. + \dots + \left[ Y_n \left( c_n - \frac{c_1}{n} - \frac{c_2}{n} - \dots - \frac{c_n}{n} \right) \right] \right\} \\
&= Y_1 \left\{ \frac{1}{s_{xx}} \left[ (x_1 - \bar{x}) - \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}) \right] \right\} + Y_2 \left\{ \frac{1}{s_{xx}} \left[ (x_1 - \bar{x}) - \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}) \right] \right\} \\
&\quad + \dots + Y_n \left\{ \frac{1}{s_{xx}} \left[ (x_1 - \bar{x}) - \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}) \right] \right\} \\
&= \sum_{i=1}^n Y_i \left\{ \frac{1}{s_{xx}} \left[ (x_i - \bar{x}) - \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}) \right] \right\} \\
&= \sum_{i=1}^n Y_i \frac{1}{s_{xx}} (x_i - \bar{x}) \\
&= \sum_{i=1}^n U_i.
\end{aligned}$$

Now, since each  $U_i$  is a normal random variable ( $Y_i$ ) multiplied by a constant, the distribution for each  $U_i$  is given by:

$$U_i \sim \mathcal{N} \left( (\beta_o + \beta_1 x_i) \frac{1}{s_{xx}} (x_i - \bar{x}), \sigma^2 \frac{1}{s_{xx}^2} (x_i - \bar{x})^2 \right).$$

Note that now,  $\hat{\beta}_1$  is a sum of independent, normal random variables, so the distribution for  $\hat{\beta}_1$  is simply normal, where the mean is the sum of the means and where the variance is the sum of the variances:

$$\hat{\beta}_1 \sim \mathcal{N} \left( \sum_{i=1}^n (\beta_o + \beta_1 x_i) \frac{1}{s_{xx}} (x_i - \bar{x}), \sum_{i=1}^n \sigma^2 \frac{1}{s_{xx}^2} (x_i - \bar{x})^2 \right).$$

(b) Before I show that  $\hat{\beta}_1$  is unbiased, first note the following:

$$\begin{aligned}
 s_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\
 &= \sum_{i=1}^n x_i^2 - 2\bar{x}n \frac{1}{n} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \\
 &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \\
 &= \sum_{i=1}^n x_i^2 - n\bar{x} \frac{1}{n} \sum_{i=1}^n x_i \\
 &= \sum_{i=1}^n (x_i^2 - \bar{x}x_i) \\
 &= \sum_{i=1}^n x_i(x_i - \bar{x}).
 \end{aligned}$$

Now, it can be shown that  $\hat{\beta}_1$  is unbiased by simplifying the expectation of  $\hat{\beta}_1$  as given above:

$$\begin{aligned}
 E[\hat{\beta}_1] &= \sum_{i=1}^n (\beta_o + \beta_1 x_i) \frac{1}{s_{xx}} (x_i - \bar{x}) \\
 &= \frac{\beta_o}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) + \frac{\beta_1}{s_{xx}} \sum_{i=1}^n x_i(x_i - \bar{x}) \\
 &= \frac{\beta_o}{s_{xx}} (n\bar{x} - n\bar{x}) + \beta_1 \\
 &= \beta_1.
 \end{aligned}$$

(c) The variance can be further simplified by canceling out a factor of  $s_{xx}$ :

$$Var[\hat{\beta}_1] = \frac{\sigma^2}{s_{xx}^2} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sigma^2}{s_{xx}}.$$

#### Problem 24.

(a) This problem can be solved in a very similar manner to that of the previous problem. I first expand out  $\hat{\beta}_o$ , use the fact that,  $\hat{\beta}_1 = \sum_{i=1}^n Y_i \frac{1}{s_{xx}} (x_i - \bar{x})$  (as found above), and group each

$Y_i$  term:

$$\begin{aligned}
 \hat{\beta}_o &= \bar{Y} - \hat{\beta}_1 \bar{x} \\
 &= \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n) - \bar{x} \left[ Y_1 \frac{1}{s_{xx}}(x_1 - \bar{x}) + Y_2 \frac{1}{s_{xx}}(x_2 - \bar{x}) + \dots + Y_n \frac{1}{s_{xx}}(x_n - \bar{x}) \right] \\
 &= Y_1 \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_1 - \bar{x}) \right] + Y_2 \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_2 - \bar{x}) \right] + \dots + Y_n \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_n - \bar{x}) \right] \\
 &= \sum_{i=1}^n Y_i \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_i - \bar{x}) \right] \\
 &= \sum_{i=1}^n U_i.
 \end{aligned}$$

As in the previous problem, since each  $U_i$  is a normal random variable ( $Y_i$ ) multiplied by a constant, the distribution for each  $U_i$  is given by:

$$U_i \sim \mathcal{N} \left( (\beta_o + \beta_1 x_i) \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_i - \bar{x}) \right], \sigma^2 \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_i - \bar{x}) \right]^2 \right).$$

Also, as above,  $\hat{\beta}_o$  is now a sum of independent, normal random variables, so the distribution for  $\hat{\beta}_o$  is simply normal, where the mean is the sum of the means and where the variance is the sum of the variances:

$$\hat{\beta}_o \sim \mathcal{N} \left( \sum_{i=1}^n (\beta_o + \beta_1 x_i) \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_i - \bar{x}) \right], \sum_{i=1}^n \sigma^2 \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_i - \bar{x}) \right]^2 \right).$$

- (b) To show that  $\hat{\beta}_o$  is unbiased, I can simplify  $E[\hat{\beta}_o]$  as found above (using  $s_{xx} = \sum_{i=1}^n x_i(x_i - \bar{x})$  as found in the previous problem):

$$\begin{aligned}
 E[\hat{\beta}_o] &= \sum_{i=1}^n (\beta_o + \beta_1 x_i) \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}}(x_i - \bar{x}) \right] \\
 &= \sum_{i=1}^n \frac{\beta_o}{n} - \frac{\beta_o \bar{x}}{s_{xx}} \sum_{i=1}^n (x_i - \bar{x}) + \frac{\beta_1}{n} \sum_{i=1}^n x_i - \frac{\beta_1 \bar{x}}{s_{xx}} \sum_{i=1}^n x_i(x_i - \bar{x}) \\
 &= \beta_o - \frac{\beta_o \bar{x}}{s_{xx}}(n\bar{x} - n\bar{x}) + \beta_1 \bar{x} - \beta_1 \bar{x} \\
 &= \beta_o.
 \end{aligned}$$

- (c) For any  $i = 1, 2, \dots, n$ , using  $\hat{\beta}_1 = \sum_{i=1}^n Y_i \frac{1}{s_{xx}}(x_i - \bar{x})$  (as derived in the previous problem),

I have that:

$$\begin{aligned}
 Cov[\hat{\beta}_1, Y_i] &= Cov\left[\sum_{j=1}^n Y_j \frac{1}{s_{xx}}(x_j - \bar{x}), Y_i\right] \\
 &= \sum_{j=1}^n Cov\left[Y_j \frac{1}{s_{xx}}(x_j - \bar{x}), Y_i\right] \\
 &= \sum_{j=1}^n \frac{1}{s_{xx}}(x_j - \bar{x}) Cov[Y_j, Y_i] \\
 &= \frac{1}{s_{xx}}(x_i - \bar{x}) Var[Y_i, Y_i] \\
 &= \frac{(x_i - \bar{x})\sigma^2}{s_{xx}},
 \end{aligned}$$

where in the second to last line I have used the fact that all  $Y_i$ s are independent, so that  $Cov[Y_i, Y_j] = 0$  for  $i \neq j$ .

(d) Again, using  $\hat{\beta}_1 = \sum_{i=1}^n Y_i \frac{1}{s_{xx}}(x_i - \bar{x})$ , I have that:

$$\begin{aligned}
 Cov[\hat{\beta}_1, \bar{Y}] &= Cov\left[\sum_{i=1}^n Y_i \frac{1}{s_{xx}}(x_i - \bar{x}), \frac{1}{n} \sum_{j=1}^n Y_j\right] \\
 &= \sum_{i,j} Cov\left[Y_i \frac{1}{s_{xx}}(x_i - \bar{x}), \frac{1}{n} Y_j\right] \\
 &= \sum_{i,j} \frac{(x_i - \bar{x})}{ns_{xx}} Cov[Y_i, Y_j] \\
 &= \sum_{i=1}^n \frac{(x_i - \bar{x})}{ns_{xx}} Var[Y_i] \\
 &= \frac{\sigma^2}{ns_{xx}} \left( \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} \right) \\
 &= \frac{\sigma^2}{ns_{xx}} (n\bar{x} - n\bar{x}) \\
 &= 0.
 \end{aligned}$$

Again, I have used the fact that all  $Y_i$ s are independent, so that  $Cov[Y_i, Y_j] = 0$  for  $i \neq j$ .

(e) The variance of  $\hat{\beta}_o$  can be further simplified to give the desired result:

$$\begin{aligned}
 Var[\hat{\beta}_o] &= \sum_{i=1}^n \sigma^2 \left[ \frac{1}{n} - \bar{x} \frac{1}{s_{xx}} (x_i - \bar{x}) \right]^2 \\
 &= \sigma^2 \sum_{i=1}^n \left[ \frac{1}{n^2} - \frac{2\bar{x}}{ns_{xx}} (x_i - \bar{x}) + \frac{\bar{x}^2}{s_{xx}^2} (x_i - \bar{x})^2 \right] \\
 &= \sigma^2 \left[ \frac{1}{n} - \frac{2\bar{x}}{ns_{xx}} \sum_{i=1}^n (x_i - \bar{x}) + \frac{\bar{x}^2}{s_{xx}^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right] \\
 &= \sigma^2 \left[ \frac{1}{n} - \frac{2\bar{x}}{ns_{xx}} (n\bar{x} - n\bar{x}) + \frac{\bar{x}^2}{s_{xx}} \right] \\
 &= \sigma^2 \left[ \frac{s_{xx} + n\bar{x}^2}{ns_{xx}} \right] \\
 &= \sigma^2 \left[ \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n\bar{x}^2}{ns_{xx}} \right] \\
 &= \sigma^2 \left[ \frac{\sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) + n\bar{x}^2}{ns_{xx}} \right] \\
 &= \sigma^2 \left( \frac{\sum_{i=1}^n x_i^2}{ns_{xx}} \right).
 \end{aligned}$$



## Chapter 9

# Statistical Inference II: Bayesian Inference

**Problem 1.** The posterior density can be found with Bayes' rule:

$$f_{X|Y}(x|2) = \frac{P_{Y|X}(2|x)f_X(x)}{\int P_{Y|X}(2|x)f_X(x)dx},$$

where

$$P_{Y|X}(2|x) = x(1-x).$$

I therefore have that:

$$\begin{aligned} f_{X|Y}(x|2) &= \frac{x^2(1-x)^2}{\int_0^1 x^2(1-x)^2 dx} \\ &= 30x^2(1-x)^2 \end{aligned}$$

for  $0 \leq x \leq 1$  and 0 otherwise. As a sanity check, I made sure the posterior integrates to 1.

**Problem 2.** From Baye's rule we know that the posterior density for  $0 \leq x \leq 1$  is:

$$f_{X|Y}(x|5) \propto P_{Y|X}(5|x)f_X(x) = 3x^3(1-x)^4,$$

(and 0 otherwise) where the symbol  $\propto$  means proportional to as a function of  $x$ . Therefore, the MAP estimate is given by

$$\hat{x}_{MAP} = \arg \max_{x \in [0,1]} \{3x^3(1-x)^4\},$$

which can be found setting the derivative of the argument equal to zero and solving for  $x$ :

$$0 = 9\hat{x}_{MAP}^2(1 - \hat{x}_{MAP})^4 - 12\hat{x}_{MAP}^3(1 - \hat{x}_{MAP})^3$$

$$\implies \hat{x}_{MAP} = 3/7, \text{ which is indeed in the interval } [0, 1]$$

**Problem 3.** The conditional PDF of  $X|Y$  is given by

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \propto f_{XY}(x,y),$$

so that the MAP estimate of  $x$  is:

$$\hat{x}_{MAP} = \arg \max_{x \in [0,1]} \{f_{XY}(x,y)\}.$$

On the other hand, the conditional PDF of  $Y|X$  is given by

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)dy},$$

so that the ML estimate of  $x$  is:

$$\hat{x}_{ML} = \arg \max_{x \in [0,1]} \left\{ \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)dy} \right\}.$$

Since the joint PDF is

$$f_{XY}(x,y) = \begin{cases} x + \frac{3}{2}y^2 & \text{for } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

we see that  $f_{XY}(x, y)$  is maximized when  $x = 1$ , and therefore  $\hat{x}_{MAP} = 1$ .

From the equation above:

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{x + \frac{3}{2}y^2}{\int x + \frac{3}{2}y^2 dy} \\ &= \frac{2x + 3y^2}{2x + 1} \\ &= 1 + \frac{3y^2 - 1}{2x + 1}, \end{aligned}$$

and thus we see that, to maximize this expression, if  $3y^2 - 1 \leq 0$  we need to minimize the second term, while if  $3y^2 - 1 > 0$  we need to maximize it. Therefore:

$$\hat{x}_{ML} = \begin{cases} 1 & \text{for } y \leq \frac{1}{\sqrt{3}} \\ 0 & \text{otherwise.} \end{cases}$$

**Problem 4.** The posterior distribution is:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x)dx} \\ &= \frac{(xy - \frac{x}{2} + 1)(2x^2 + \frac{1}{3})}{\int_0^1 (xy - \frac{x}{2} + 1)(2x^2 + \frac{1}{3})dx} \\ &= \frac{2x^3y + \frac{xy}{3} - x^3 - \frac{x}{6} + 2x^2 + \frac{1}{3}}{\frac{2}{3}y + \frac{2}{3}}, \end{aligned}$$

so that

$$\begin{aligned} \hat{x}_M &= E[X|Y = y] \\ &= \frac{1}{\frac{2}{3}y + \frac{2}{3}} \int_0^1 \left( 2x^4y + \frac{x^2y}{3} - x^4 - \frac{x^2}{6} + 2x^3 + \frac{x}{3} \right) dx \\ &= \frac{46y - 37}{60(y + 1)}. \end{aligned}$$

**Problem 5.**

- (a) First note that since  $X \sim \mathcal{N}(0, 1)$ ,  $W \sim \mathcal{N}(0, 1)$  (where  $X$  and  $W$  are independent), we have that  $Y = 2X + W \sim \mathcal{N}(0, 5)$ . Now,  $aY + bX = a(2X + W) + bX = (2a + b)X + aW$  which is normal for all  $a, b \in \mathbb{R}$ , and thus  $X$  and  $Y$  are jointly normal. Thus, by Theorem 5.4 in the book, I have that:

$$\hat{X}_M = E[X|Y] = \mu_X + \rho\sigma_X \frac{Y - \mu_Y}{\sigma_Y}.$$

From the distributions of  $X$  and  $Y$ , I have that  $\mu_X = \mu_Y = 0$ ,  $\sigma_Y = \sqrt{5}$ ,  $\sigma_X = 1$  and:

$$\begin{aligned} \text{Cov}[X, Y] &= \text{Cov}[X, 2X + W] \\ &= 2\text{Var}[X] + \text{Cov}[X, W] \\ &= 2\text{Var}[X] \quad (\text{since } X, W \text{ are independent}) \\ &= 2. \end{aligned}$$

Plugging in the values, I find that

$$\begin{aligned}\hat{X}_M &= E[X|Y] \\ &= \frac{2}{\sqrt{5}} \cdot \frac{Y}{\sqrt{5}} \\ &= \frac{2Y}{5}.\end{aligned}$$

(b) To solve this problem, I use the facts that  $X$  and  $W$  are independent and  $E[X^2] = E[W^2] = 1$ :

$$\begin{aligned}MSE &= E[(X - \hat{X}_M)^2] \\ &= E[X^2] - 2E[X\hat{X}_M] + E[\hat{X}_M^2] \\ &= 1 - 2E\left[X\frac{2Y}{5}\right] + E\left[\left(\frac{2Y}{5}\right)^2\right] \\ &= 1 - 2E\left[\frac{2}{5}(2X^2 + XW)\right] + E\left[\frac{4}{25}(4X^2 + 4XW + W^2)\right] \\ &= 1 - \frac{4}{5}(E[2X^2] + E[X]E[W]) + \frac{4}{25}(E[(4X^2) + 4E[X]E[W] + E[W^2])) \\ &= 1 - \frac{4}{5}(2 \cdot 1) + \frac{4}{25}(4 \cdot 1 + 1) \\ &= \frac{1}{5}.\end{aligned}$$

(c) We know that  $E[X^2] = 1$ , and from the derivation above we can see that  $E[\hat{X}_M^2] = E[(2Y/5)^2] = 20/25 = 4/5$ . Also, from the derivation above  $E[\tilde{X}^2] = MSE = 1/5$ . Therefore  $E[\hat{X}_M^2] + E[\tilde{X}^2] = 4/5 + 1/5 = 1$ , and so the relation is verified.

### Problem 6.

(a) The linear MMSE estimator is given by

$$\hat{X}_L = \frac{Cov[X, Y]}{Var[Y]}(Y - E[Y]) + E[X],$$

and therefore I must find these values. First note that since  $X \sim Unif(0, 1)$ ,  $E[X] = 1/2$  and  $Var[X] = 1/12$ , which implies that  $E[X^2] = 1/3$ . Also note that since  $Y|X = x \sim Exp(1/(2x))$ ,  $E[Y|X] = 2X$  and  $Var[Y|X] = 4X^2$ . Now, since I know the distribution of  $Y|X$ , to find  $E[Y]$ , I use the law of iterated expectation:

$$\begin{aligned}E[Y] &= E[E[Y|X]] \\ &= E[2X] \\ &= 1.\end{aligned}$$

Similarly, to find  $Var[Y]$ , I use the law of total variance:

$$\begin{aligned}Var[Y] &= E[Var[Y|X]] + Var[E[Y|X]] \\ &= E[4X^2] + Var[2X] \\ &= 4E[X^2] + 4Var[X] \\ &= 4 \cdot \frac{1}{3} + 4 \cdot \frac{1}{12} \\ &= \frac{5}{3}.\end{aligned}$$

To find  $Cov[X, Y]$ , I first solve for  $E[XY]$ , again using the law of iterated expectation:

$$\begin{aligned} E[XY] &= E[E[XY|X]] \\ &= E[XE[Y|X]] \\ &= E[X^2] \\ &= 2E[X^2] \\ &= \frac{2}{3}. \end{aligned}$$

Thus I have that:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= \frac{2}{3} - \frac{1}{2} \cdot 1 \\ &= \frac{1}{6}. \end{aligned}$$

Plugging all these values into the formula for  $\hat{X}_L$ , I find that:

$$\hat{X}_L = \frac{1}{10}Y + \frac{2}{5}.$$

(b) The MSE of  $\hat{X}_L$  is given by:

$$\begin{aligned} MSE &= (1 - \rho^2)Var[X] \\ &= \left(1 - \frac{Cov[X, Y]^2}{Var[X]Var[Y]}\right) Var[X] \\ &= \left(1 - \frac{\left(\frac{1}{6}\right)^2}{\frac{1}{12} \cdot \frac{5}{3}}\right) \cdot \frac{1}{12} \\ &= \frac{1}{15}. \end{aligned}$$

(c) First, note that  $E[Y^2] = Var[Y] + (E[Y])^2 = 5/3 + 1 = 8/3$ . Now, I have that:

$$\begin{aligned} E[\tilde{X}Y] &= E[(X - \hat{X}_L)Y] \\ &= E\left[XY - \left(\frac{1}{10}Y + \frac{2}{5}\right)Y\right] \\ &= E[XY] - \left(\frac{1}{10}E[Y^2] + \frac{2}{5}E[Y]\right) \\ &= \frac{2}{3} - \left(\frac{1}{10} \cdot \frac{8}{3} + \frac{2}{5}\right) \\ &= 0. \end{aligned}$$

### Problem 7.

(a) First note that  $Y \sim (0, \sigma_X^2 + \sigma_W^2)$ . Also note that  $aX + bY = (a+b)X + bW$  which is normally distributed for all  $a, b \in \mathbb{R}$ , and thus  $X$  and  $Y$  are jointly normal, and so by Theorem 5.4 in the book:

$$\hat{X}_M = E[X|Y] = \mu_X + \rho\sigma_X \frac{Y - \mu_Y}{\sigma_Y}.$$

Solving for the covariance is easy:  $Cov[X, Y] = Cov[X, X + W] = Var[X] + Cov[X, W] = Var[X] = \sigma_X^2$ , where  $Cov[X, W] = 0$  since  $X$  and  $W$  are independent. I thus have that:

$$\begin{aligned}\hat{X}_M &= \mu_X + \frac{Cov[X, Y]}{\sigma_X \sigma_Y} \sigma_X \frac{Y - \mu_Y}{\sigma_Y} \\ &= \frac{\sigma_X^2}{\sigma_X \sqrt{\sigma_X^2 + \sigma_W^2}} \cdot \sigma_X \cdot \frac{Y}{\sqrt{\sigma_X^2 + \sigma_W^2}} \\ &= \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} Y.\end{aligned}$$

(b) The MSE is:

$$\begin{aligned}MSE(\hat{X}_M) &= E[(X - \hat{X}_M)^2] \\ &= E[X^2] - 2E[X\hat{X}_M] + E[\hat{X}_M^2] \\ &= E[X^2] - 2\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} E[XY] + \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}\right)^2 E[Y^2] \\ &= E[X^2] - 2\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} E[X(X + W)] + \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}\right)^2 E[(X + W)^2] \\ &= E[X^2] - 2\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} (E[X^2] + E[X]E[W]) \\ &\quad + \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}\right)^2 (E[X^2] + 2E[X]E[W] + E[W^2]) \\ &= \frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2},\end{aligned}$$

where in the second to last line I have used that  $X$  and  $W$  are independent, and where I have used that:  $E[X^2] = \sigma_X^2$ ,  $E[W^2] = \sigma_W^2$  and  $E[X] = E[W] = 0$ .

**Problem 8.** Following along Example 9.9 of the book, I use the principle of orthogonality to solve for  $\hat{X}_L$ . I first note that since  $E[X] = E[W_1] = E[W_2] = 0$ ,  $E[Y_1] = E[Y_2] = 0$ . We would like the linear MMSE estimator to be of the form:

$$\hat{X}_L = aY_1 + bY_2 + c.$$

Now, using the first of the two orthogonality principle equations I can solve for  $c$ :

$$0 = E[\tilde{X}] = E[X - \hat{X}_L] = E[X] - aE[Y_1] - bE[Y_2] - c = -c,$$

so that  $c = 0$ .

Using the second of the two orthogonality principle equations I may solve for the remaining constants by noting that

$$Cov[\tilde{X}, Y_j] = Cov[X - \hat{X}_L, Y_j] = 0 \quad \text{for } j = 1, 2,$$

and thus:

$$Cov[X, Y_j] = Cov[\hat{X}_L, Y_j] \quad \text{for } j = 1, 2.$$

I now compute the four covariances, and plug them into the above equations (one for  $j = 1$  and one for  $j = 2$ ) to obtain a coupled set of equations in  $a$  and  $b$ .

For the  $Cov[X, Y_j]$  covariances, I have

$$Cov[X, Y_1] = Cov[X, 2X + W_1] = 2Var[X] + Cov[X, W_1] = 2Var[X] = 10,$$

where I have used the fact that since  $X, W_1$  are independent, their covariance is 0. Also,

$$Cov[X, Y_2] = Cov[X, X + W_2] = Var[X] + Cov[X, W_2] = Var[X] = 5,$$

where again I have used independence.

I now solve for the other two covariances:

$$\begin{aligned} Cov[\hat{X}_L, Y_1] &= Cov[aY_1 + bY_2, Y_1] \\ &= aCov[Y_1, Y_1] + bCov[Y_2, Y_1] \\ &= aCov[2X + W_1, 2X + W_1] + bCov[X + W_2, 2X + W_1] \\ &= a(4Var[X] + 4Cov[X, W_1] + Var[W_1]) \\ &\quad + b(2Var[X] + Cov[X, W_1] + 2Cov[W_2, X] + Cov[W_2, W_1]) \\ &= a(4Var[X] + Var[W_1]) + 2bVar[X] \\ &= a(4 \cdot 5 + 2) + 2 \cdot b \cdot 5 \\ &= 22a + 10b, \end{aligned}$$

and

$$\begin{aligned} Cov[\hat{X}_L, Y_2] &= Cov[aY_1 + bY_2, Y_2] \\ &= aCov[Y_1, Y_2] + bCov[Y_2, Y_2] \\ &= aCov[2X + W_1, X + W_2] + bCov[X + W_2, X + W_2] \\ &= a(2Var[X] + 2Cov[X, W_2] + Cov[W_1, X] + Cov[W_1, W_2]) \\ &\quad + b(Var[X] + 2Cov[X, W_2] + Var[W_2]) \\ &= 2aVar[X] + b(Var[X] + Var[W_2]) \\ &= 2 \cdot a \cdot 5 + b(5 + 5) \\ &= 10a + 10b \end{aligned}$$

where in both computations, I have used the independence of  $X, W_1, W_2$ .

Putting these 4 equations together, I get a coupled set of algebraic equations in  $a$  and  $b$ :

$$\begin{cases} 10 = 22a + 10b \\ 5 = 10a + 10b, \end{cases}$$

resulting in  $a = 5/12$  and  $b = 1/12$ . The linear MMSE estimator of  $X$  is thus:

$$\hat{X}_L = \frac{5}{12}Y_1 + \frac{1}{12}Y_2.$$

**Problem 9.** In order to use the vector formula

$$\hat{X}_L = \mathbf{C}_{\mathbf{X}\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1}(\mathbf{Y} - E[\mathbf{Y}]) + E[X],$$

to solve for the linear MMSE estimator, I need to compute the vectors/matrices that go into this formula. Firstly, it can easily be shown that

$$E[\mathbf{Y}] = \begin{bmatrix} E[Y_1] \\ E[Y_2] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The covariance matrix of  $\mathbf{Y}$  is:

$$\begin{aligned}
 \mathbf{C}_{\mathbf{Y}} &= \begin{bmatrix} E[(Y_1 - E[Y_1])^2] & E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])] \\ E[(Y_2 - E[Y_2])(Y_1 - E[Y_1])] & E[(Y_2 - E[Y_2])^2] \end{bmatrix} \\
 &= \begin{bmatrix} E[Y_1^2] & E[Y_1 Y_2] \\ E[Y_2 Y_1] & E[Y_2^2] \end{bmatrix} \\
 &= \begin{bmatrix} E[(2X + W_1)^2] & E[(2X + W_1)(X + W_2)] \\ E[(X + W_2)(2X + W_1)] & E[(X + W_2)^2] \end{bmatrix} \\
 &= \begin{bmatrix} E[4X^2 + 4XW_1 + W_1^2] & E[2X^2 + 3XW_1 + W_1W_2] \\ E[2X^2 + 3XW_1 + W_1W_2] & E[X^2 + 2XW_2 + W_2^2] \end{bmatrix} \\
 &= \begin{bmatrix} 4E[X^2] + E[W_1^2] & 2E[X^2] \\ 2E[X^2] & E[X^2] + E[W_2^2] \end{bmatrix} \\
 &= \begin{bmatrix} 4 \cdot 5 + 2 & 2 \cdot 5 \\ 2 \cdot 5 & 5 + 5 \end{bmatrix} \\
 &= \begin{bmatrix} 22 & 10 \\ 10 & 10 \end{bmatrix},
 \end{aligned}$$

where I have used the fact that  $X, W_1, W_2$  are independent.

The cross-covariance matrix of  $X$  and  $\mathbf{Y}$  is:

$$\begin{aligned}
 \mathbf{C}_{\mathbf{XY}} &= [E[(X - E[X])(Y_1 - E[Y_1])] \quad E[(X - E[X])(Y_2 - E[Y_2])]] \\
 &= [E[XY_1] \quad E[XY_2]] \\
 &= [E[X(2X + W_1)] \quad E[X(X + W_2)]] \\
 &= [2E[X^2] + E[XW_1] \quad E[X^2] + E[XW_2]] \\
 &= [2E[X^2] \quad E[X^2]] \\
 &= [10 \quad 5],
 \end{aligned}$$

where, again, I have used the fact that  $X, W_1, W_2$  are independent.

Plugging these matrices into the formula for  $\hat{X}_L$ , I finally have:

$$\begin{aligned}
 \hat{X}_L &= [10 \quad 5] \begin{bmatrix} 22 & 10 \\ 10 & 10 \end{bmatrix}^{-1} \mathbf{Y} \\
 &= \frac{5}{12}Y_1 + \frac{1}{12}Y_2,
 \end{aligned}$$

which matches exactly with the answer found in the previous problem which utilized the orthogonality principle.

**Problem 10.** To solve this problem, I use the vector formula approach as in the previous problem. It can easily be shown that

$$E[\mathbf{Y}] = \begin{bmatrix} E[Y_1] \\ E[Y_2] \\ E[Y_3] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Also, since  $E[Y_i] = 0$  (for  $i = 1, 2, 3$ ), the covariance matrix of  $\mathbf{Y}$  is

$$\mathbf{C}_{\mathbf{Y}} = \begin{bmatrix} E[Y_1^2] & E[Y_1 Y_2] & E[Y_1 Y_3] \\ E[Y_2 Y_1] & E[Y_2^2] & E[Y_2 Y_3] \\ E[Y_3 Y_1] & E[Y_3 Y_2] & E[Y_3^2] \end{bmatrix}.$$



Therefore, I need to compute all of the pairwise expectations along the diagonal and in the upper right triangle of the matrix. Notice that all of the pairwise expectations are of the form  $E[Y_i Y_j]$ , where  $Y_i = a_i X + b_i W_i$ , so I derive a general formula for all 6 pairs for easy computation of the expectations:

$$\begin{aligned} E[Y_i Y_j] &= E[(a_i X + b_i W_i)(a_j X + b_j W_j)] \\ &= a_i a_j E[X^2] + a_i b_j E[X W_j] + b_i a_j E[W_i X] + b_i b_j E[W_i W_j] \\ &= a_i a_j E[X^2] + b_i b_j E[W_i^2] \delta_{i,j} \\ &= 5a_i a_j + b_i^2 E[W_i^2] \delta_{i,j}, \end{aligned}$$

where  $\delta_{i,j}$  is the Kronecker-delta function, and where I used independence several times. Using this formula, it is easy to find that

$$\mathbf{C}_Y = \begin{bmatrix} 22 & 10 & 10 \\ 10 & 10 & 5 \\ 10 & 5 & 17 \end{bmatrix}$$

Since  $E[X] = 0$  and  $E[Y_i] = 0$  (for  $i = 1, 2, 3$ ), the cross-covariance matrix is also easy to compute

$$\begin{aligned} \mathbf{C}_{XY} &= \begin{bmatrix} E[XY_1] & E[XY_2] & E[XY_3] \end{bmatrix} \\ &= \begin{bmatrix} E[X(2X + W_1)] & E[X(X + W_2)] & E[X(X + W_3)] \end{bmatrix} \\ &= \begin{bmatrix} 2E[X^2] & E[X^2] & E[X^2] \end{bmatrix} \\ &= \begin{bmatrix} 10 & 5 & 5 \end{bmatrix} \end{aligned}$$

Plugging these matrices into the vector formula for  $\hat{X}_L$ , I finally have:

$$\begin{aligned} \hat{X}_L &= \begin{bmatrix} 10 & 5 & 5 \end{bmatrix} \begin{bmatrix} 22 & 10 & 10 \\ 10 & 10 & 5 \\ 10 & 5 & 17 \end{bmatrix}^{-1} \mathbf{Y} \\ &= \frac{60}{149} Y_1 + \frac{12}{149} Y_2 + \frac{5}{149} Y_3. \end{aligned}$$

### Problem 11.

(a) The linear MMSE estimator is given by:

$$\hat{X}_L = \frac{Cov[X, Y]}{Var[Y]}(Y - E[Y]) + E[X],$$

and so I must compute the various terms in this equation.

The expectations are

$$E[X] = E[Y] = \frac{3}{7},$$

while the covariance is

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= 0 - \left(\frac{3}{7}\right)^2 \\ &= -\frac{9}{49}. \end{aligned}$$

Finally, the variance of  $Y$  is:

$$\begin{aligned} \text{Var}[Y] &= \left(0 - \frac{3}{7}\right)^2 \cdot \left(\frac{1}{7} + \frac{3}{7}\right) + \left(1 - \frac{3}{7}\right)^2 \cdot \frac{3}{7} \\ &= \frac{12}{49}. \end{aligned}$$

Plugging these values into the above equation for  $\hat{X}_L$ , I find that:

$$\hat{X}_L = -\frac{3}{4}Y + \frac{3}{4}.$$

(b) The MMSE estimator is given by

$$\hat{X}_M = E[X|Y],$$

and the conditional PMFs can easily be found to be

$$P_{X|Y}(x|0) = \begin{cases} \frac{1}{4} & \text{for } x = 0 \\ \frac{3}{4} & \text{for } x = 1 \end{cases}$$

and

$$P_{X|Y}(x|1) = \begin{cases} 1 & \text{for } x = 0 \\ 0 & \text{for } x = 1. \end{cases}$$

This results in the conditional expectations,  $E[X|Y = 0] = 3/4$  and  $E[X|Y = 1] = 0$ , which can be combined into one expression using an indicator random variable:

$$\hat{X}_M = \frac{3}{4}\mathbb{1}\{Y = 0\}.$$

(c) The MSE of  $\hat{X}_M$  is given by:

$$\begin{aligned} \text{MSE}(\hat{X}_M) &= E[(X - \hat{X}_M)^2] \\ &= \sum_{x,y} \left(x - \frac{3}{4}\mathbb{1}\{y = 0\}\right)^2 P_{XY}(x, y) \\ &= \left(0 - \frac{3}{4}\mathbb{1}\{0 = 0\}\right)^2 \cdot \frac{1}{7} + \left(1 - \frac{3}{4}\mathbb{1}\{0 = 0\}\right)^2 \cdot \frac{3}{7} + \left(0 - \frac{3}{4}\mathbb{1}\{1 = 0\}\right)^2 \cdot \frac{3}{7} \\ &= \left(-\frac{3}{4}\right)^2 \cdot \frac{1}{7} + \left(1 - \frac{3}{4}\right)^2 \cdot \frac{3}{7} \\ &= \frac{3}{28}. \end{aligned}$$

### Problem 12.

(a) Following along with the previous problem, I must calculate the terms that go into the formula for  $\hat{X}_L$ . The expectations are:

$$E[X] = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$$

and

$$E[Y] = \frac{1}{3} + 2 \cdot \frac{1}{6} = \frac{2}{3},$$

while the covariance is:

$$\begin{aligned} \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] \\ &= 2 \cdot \frac{1}{6} - \frac{2}{3} \cdot \frac{1}{2} \\ &= 0. \end{aligned}$$

Therefore, the linear MMSE estimator is simply a constant:

$$\hat{X}_L = E[X] = \frac{1}{2}.$$

(b) The MSE for  $\hat{X}_L$  is given by:

$$\begin{aligned} \text{MSE}(\hat{X}_L) &= E[(X - \hat{X}_L)^2] \\ &= E[(X - \frac{1}{2})^2] \\ &= E[X^2] - E[X] + \frac{1}{4} \\ &= \frac{1}{2} - \frac{1}{2} + \frac{1}{4} \\ &= \frac{1}{4}. \end{aligned}$$

(c) To find  $\hat{X}_M$ , I must first find the conditional PMF of  $X$ :

$$P_{X|Y}(x|0) = \begin{cases} \frac{1}{3} & \text{for } x = 0 \\ \frac{2}{3} & \text{for } x = 1, \end{cases}$$

$$P_{X|Y}(x|1) = \begin{cases} 1 & \text{for } x = 0 \\ 0 & \text{for } x = 1. \end{cases}$$

and

$$P_{X|Y}(x|2) = \begin{cases} 0 & \text{for } x = 0 \\ 1 & \text{for } x = 1. \end{cases}$$

Thus, I have the conditional expectations,  $E[X|Y = 0] = 2/3$ ,  $E[X|Y = 1] = 0$ , and  $E[X|Y = 2] = 1$ , which can be combined into one expression using an indicator random variable:

$$\begin{aligned} \hat{X}_M &= E[X|Y] \\ &= \frac{2}{3}\mathbb{1}\{Y = 0\} + \mathbb{1}\{Y = 2\}. \end{aligned}$$

(d) The MSE of  $\hat{X}_M$  is given by:

$$\begin{aligned}
 \text{MSE}(\hat{X}_M) &= E[(X - \hat{X}_M)^2] \\
 &= \sum_{x,y} \left( x - \frac{2}{3} \mathbb{1}\{y=0\} - \mathbb{1}\{y=2\} \right)^2 P_{XY}(x,y) \\
 &= \left( 0 - \frac{2}{3} \mathbb{1}\{0=0\} - \mathbb{1}\{0=2\} \right)^2 \cdot \frac{1}{6} + \left( 1 - \frac{2}{3} \mathbb{1}\{0=0\} - \mathbb{1}\{0=2\} \right)^2 \cdot \frac{1}{3} \\
 &\quad + \left( 0 - \frac{2}{3} \mathbb{1}\{1=0\} - \mathbb{1}\{1=2\} \right)^2 \cdot \frac{1}{3} + \left( 1 - \frac{2}{3} \mathbb{1}\{2=0\} - \mathbb{1}\{2=2\} \right)^2 \cdot \frac{1}{6} \\
 &= \left( 0 - \frac{2}{3} \right)^2 \cdot \frac{1}{6} + \left( 1 - \frac{2}{3} \right)^2 \cdot \frac{1}{3} + (0-0)^2 \cdot \frac{1}{3} + (1-1)^2 \cdot \frac{1}{6} \\
 &= \frac{1}{9}.
 \end{aligned}$$

**Problem 13.** In this problem, the two hypotheses are:

$$\begin{aligned}
 H_o &: X = 1 \\
 H_1 &: X = -1,
 \end{aligned}$$

where the priors are  $P(H_o) = p$  and  $P(H_1) = 1 - p$ . Note that given  $H_o$ ,  $Y = 2 + W$  so that  $Y|H_o \sim \mathcal{N}(2, \sigma^2)$ , and that given  $H_1$ ,  $Y = -2 + W$  so that  $Y|H_1 \sim \mathcal{N}(-2, \sigma^2)$ . The posterior probability of  $H_o$  is thus:

$$P(H_o|Y = y) \propto f_Y(y|H_o)P(H_o) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-2)^2}{2\sigma^2}} p,$$

while the posterior probability of  $H_1$  is thus:

$$P(H_1|Y = y) \propto f_Y(y|H_1)P(H_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+2)^2}{2\sigma^2}} (1 - p).$$

The MAP decision rule for this problem is to accept  $H_o$  if the posterior probability under  $H_o$  is greater than or equal to the posterior probability under  $H_1$ . In other words, we accept  $H_o$  if:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-2)^2}{2\sigma^2}} p \geq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+2)^2}{2\sigma^2}} (1 - p),$$

otherwise we accept  $H_1$ .

Using some algebra, this rule can be re-written as: if

$$y \geq \frac{\sigma^2}{2} \ln \left( \frac{1-p}{p} \right),$$

then accept  $H_o$ , otherwise accept  $H_1$ .

**Problem 14.** The error probability is given by:

$$P_e = P(\text{choose } H_1|H_o)P(H_o) + P(\text{choose } H_o|H_1)P(H_1).$$

We can replace “choose  $H_1$ ” and “choose  $H_o$ ” with our decision rule to compute the conditional probabilities:

$$\begin{aligned}
 P(\text{choose } H_1|H_o) &= P\left(Y < \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right) \middle| X = 1\right) \\
 &= P\left(2X + W < \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right) \middle| X = 1\right) \\
 &= P\left(2 + W < \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right)\right) \\
 &= \Phi\left(\frac{\sigma}{2} \ln\left(\frac{1-p}{p}\right) - \frac{2}{\sigma}\right).
 \end{aligned}$$

Likewise, the other conditional probability is given by

$$\begin{aligned}
 P(\text{choose } H_o|H_1) &= P\left(Y \geq \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right) \middle| X = -1\right) \\
 &= P\left(2X + W \geq \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right) \middle| X = -1\right) \\
 &= P\left(-2 + W \geq \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right)\right) \\
 &= 1 - \Phi\left(\frac{\sigma}{2} \ln\left(\frac{1-p}{p}\right) + \frac{2}{\sigma}\right).
 \end{aligned}$$

Plugging these conditional probabilities into the formula for the error probability, I find that

$$P_e = \Phi\left(\frac{\sigma}{2} \ln\left(\frac{1-p}{p}\right) - \frac{2}{\sigma}\right)p + \left[1 - \Phi\left(\frac{\sigma}{2} \ln\left(\frac{1-p}{p}\right) + \frac{2}{\sigma}\right)\right](1-p).$$

**Problem 15.** Using the minimum cost decision method, we should accept  $H_o$  if  $P(H_o|y)C_{10} \geq P(H_1|y)C_{01}$ . Note that  $C_{01}$  is the cost of choosing  $H_o$  (there is no malfunction) given  $H_1$  is true (there is a malfunction). That is,  $C_{01}$  is the cost of missing a malfunction, so that, as specified in the problem  $C_{01} = 30C_{10}$ .

The left hand side of the inequality decision rule is therefore:

$$\begin{aligned}
 P(H_o|y)C_{10} &= (1 - P(H_1|y))C_{10} \\
 &= 0.9C_{10},
 \end{aligned}$$

while the right hand side of the inequality is

$$\begin{aligned}
 P(H_1|y)C_{01} &= 0.1 \cdot 30C_{10} \\
 &= 3C_{10}.
 \end{aligned}$$

Since the costs are usually not negative, we see that  $P(H_1|y)C_{01} > P(H_o|y)C_{10}$  and we thus should accept  $H_1$ , the hypothesis that there *is* a malfunction.

**Problem 16.** For  $X, Y$  jointly normal, we know from Theorem 5.4 in the book that:

$$X|Y = y \sim \mathcal{N}\left(\mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}, (1 - \rho^2)\sigma_X^2\right).$$

Since  $X \sim \mathcal{N}(2, 1)$  and  $Y \sim \mathcal{N}(1, 5)$ , using the above formula, I have that the posterior distribution is  $X|Y = 1 \sim \mathcal{N}(2, 15/16)$ .

When choosing a confidence interval, to keep things symmetric, we usually choose the confidence interval such that  $\alpha/2$  of the probability is in the left tail of the distribution (i.e.,  $P(X < a|Y = 1) = \alpha/2$ ) and  $\alpha/2$  of the probability is in the right tail of the distribution (i.e.,  $P(X > b|Y = 1) = \alpha/2$ ). Therefore to find the 90% credible interval,  $[a, b]$ , I have the following equations:

$$0.05 = \Phi\left(\frac{a-2}{\sqrt{15/4}}\right),$$

and

$$1 - 0.05 = \Phi\left(\frac{b-2}{\sqrt{15/4}}\right).$$

Solving for  $a$  and  $b$  by using the inverse Gaussian CDF, I find that the 90% credible interval is approximately  $[0.41, 3.6]$ .

**Problem 17.**

- (a) The posterior distribution can be found using Bayes' rule. For  $x > 0$ :

$$\begin{aligned} f_{X|Y}(x|y) &\propto P_{Y|X}(y|x)f_X(x) \\ &\propto e^{-x}x^y x^{\alpha-1}e^{-\beta x} \\ &= e^{-x(\beta+1)}x^{\alpha+y-1} \\ &\propto (\beta+1)^{\alpha+y}e^{-x(\beta+1)}x^{\alpha+y-1}, \end{aligned}$$

while for  $x \leq 0$ ,  $f_{X|Y}(x|y) = 0$ . The notation  $\propto$  means proportional to as a function of  $x$ . Notice that since  $\alpha > 0$  and  $y \geq 0$ ,  $y + \alpha$  is therefore greater than 0. Also, since  $\beta > 0$ ,  $1 + \beta > 0$ . Therefore, this function is exactly of the form of a  $Gamma(\alpha+y, \beta+1)$  distribution, and so we know that the normalizing constant is  $\Gamma(\alpha+y)$ .

- (b) In the previous part of the problem, I showed that  $X|Y = y \sim Gamma(\alpha+y, \beta+1)$ , and therefore:

$$f_{X|Y}(x|y) = \begin{cases} \frac{(\beta+1)^{\alpha+y}x^{\alpha+y-1}e^{-x(\beta+1)}}{\Gamma(\alpha+y)} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- (c) For  $U \sim Gamma(\alpha, \lambda)$ , as shown in Section 4.2.4 of the book,  $E[U] = \alpha/\lambda$  and  $Var[X] = \alpha/\lambda^2$ . Therefore, I have

$$E[X|Y] = \frac{\alpha + Y}{\beta + 1}$$

and

$$Var[X|Y] = \frac{\alpha + Y}{(\beta + 1)^2}.$$

**Problem 18.**

- (a) The posterior distribution can be found using Bayes' rule. For  $0 \leq x \leq 1$ :

$$\begin{aligned} f_{X|Y}(x|y) &\propto P_{Y|X}(y|x)f_X(x) \\ &\propto x^y(1-x)^{n-y}x^{\alpha-1}(1-x)^{\beta-1} \\ &= x^{\alpha+y-1}(1-x)^{\beta+n-y-1}, \end{aligned}$$

while for  $x < 0$  and  $x > 1$ ,  $f_{X|Y}(x|y) = 0$ . Now, since  $\alpha > 0$  and  $y \geq 0$ ,  $\alpha + y > 0$ . Also, since  $\beta > 0$  and  $n \geq y$ ,  $\beta + n - y > 0$ . Thus, since this equation, up to a normalization constant, has the exact same functional form as a  $Beta(\alpha+y, \beta+n-y)$  distribution, the posterior is given by this distribution.

(b) Since the posterior is given by a  $Beta(\alpha + y, \beta + n - y)$  distribution, I have that:

$$f_{X|Y}(x|y) = \begin{cases} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)} x^{\alpha+y-1} (1-x)^{\beta+n-y-1} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Plugging in my values for  $\alpha$  and  $\beta$  into the formulas for the expectation and variance of a Beta distribution, I find:

$$E[X|Y] = \frac{\alpha + Y}{\alpha + \beta + n}$$

and

$$Var[X|Y] = \frac{(\alpha + Y)(\beta + n - Y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}.$$

### Problem 19.

(a) Since  $Y|X = x \sim Geom(x)$  I have that:

$$P_{Y|X}(y|x) = x(1-x)^{y-1} \quad y = 1, 2, 3, \dots$$

Now, the posterior distribution can be found using Bayes' rule. For  $0 \leq x \leq 1$ :

$$\begin{aligned} f_{X|Y}(x|y) &\propto P_{Y|X}(y|x)f_X(x) \\ &\propto x(1-x)^{y-1}x^{\alpha-1}(1-x)^{\beta-1} \\ &= x^{(\alpha+1)-1}(1-x)^{(\beta+y-1)-1}, \end{aligned}$$

while for  $x < 0$  and  $x > 1$ ,  $f_{X|Y}(x|y) = 0$ . Since  $\alpha > 0$ ,  $\alpha + 1 > 0$ . Also, since  $\beta > 0$ ,  $y - 1 \geq 0$ ,  $\beta + y - 1 > 0$ . We thus see that up to a normalizing constant, this is the PDF for a  $Beta(\alpha + 1, \beta + y - 1)$  distribution, and hence  $X|Y = y \sim Beta(\alpha + 1, \beta + y - 1)$ .

(b) Since  $X|Y = y \sim Beta(\alpha + 1, \beta + y - 1)$ , the posterior distribution is:

$$f_{X|Y}(x|y) = \begin{cases} \frac{\Gamma(\alpha+\beta+y-1)}{\Gamma(\alpha+1)\Gamma(\beta+y-1)} x^{(\alpha+1)-1} (1-x)^{(\beta+y-1)-1} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Plugging in my values for  $\alpha$  and  $\beta$  into the formulas for the expectation and variance of a Beta distribution, I find:

$$E[X|Y] = \frac{\alpha + 1}{\alpha + \beta + Y}$$

and

$$Var[X|Y] = \frac{(\alpha + 1)(\beta + Y - 1)}{(\alpha + \beta + Y)^2(\alpha + \beta + Y + 1)}.$$

### Problem 20.

(a) Since  $Y_i|X = x \stackrel{iid}{\sim} Exp(x)$ , I have that

$$f_{Y_i|X}(y|x) = \begin{cases} xe^{-xy} & \text{for } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The likelihood function is thus:

$$\begin{aligned}
 L(\mathbf{Y}; x) &= f_{Y_1, Y_2, \dots, Y_n | X}(y_1, y_2, \dots, y_n | x) \\
 &= \prod_{i=1}^n f_{Y_i | X}(y_i | x) \quad (\text{by independence}) \\
 &= \prod_{i=1}^n x e^{-x y_i} \\
 &= x^n e^{-x \sum_{i=1}^n y_i}.
 \end{aligned}$$

- (b) Since  $X \sim \text{Gamma}(\alpha, \beta)$ , I have that  $f_X(x) \propto x^{\alpha-1} e^{-\beta x}$  for  $x > 0$  and  $f_X(x) = 0$  otherwise. Therefore, for  $x > 0$ , the posterior is:

$$\begin{aligned}
 f_{X|Y_1, Y_2, \dots, Y_n}(x | y_1, y_2, \dots, y_n) &\propto f_{Y_1, Y_2, \dots, Y_n | X}(y_1, y_2, \dots, y_n | x) f_X(x) \\
 &= L(\mathbf{Y}; x) f_X(x) \\
 &\propto x^n e^{-x \sum_{i=1}^n y_i} x^{\alpha-1} e^{-\beta x} \\
 &= x^{\alpha+n-1} e^{-x(\sum_{i=1}^n y_i + \beta)},
 \end{aligned}$$

while for  $x \leq 0$ ,  $f_{X|Y_1, Y_2, \dots, Y_n}(x | y_1, y_2, \dots, y_n) = 0$ . Now, since  $\alpha$  and  $n$  are greater than 0, so too is  $\alpha + n$ . Further since it is assumed that  $Y_i | X \sim \text{Exp}(X)$ , it is implicit that all  $y_i$ s are greater than 0, and since  $\beta > 0$ , so too is  $\sum_{i=1}^n y_i + \beta$ . Therefore, we see that, up to a normalizing constant, the posterior has the functional form of a  $\text{Gamma}(\alpha + n, \sum_{i=1}^n y_i + \beta)$  distribution, so that  $X | \mathbf{Y} = \mathbf{y} \sim \text{Gamma}(\alpha + n, \sum_{i=1}^n y_i + \beta)$ .

- (c) The posterior PDF is given by:

$$f_{X|Y_1, Y_2, \dots, Y_n}(x | y_1, y_2, \dots, y_n) = \begin{cases} \frac{(\sum_{i=1}^n y_i + \beta)^{\alpha+n} x^{\alpha+n-1} e^{-x(\sum_{i=1}^n y_i + \beta)}}{\Gamma(\alpha+n)} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- (d) For  $U \sim \text{Gamma}(\alpha, \lambda)$ , as shown in Section 4.2.4 of the book,  $E[U] = \alpha/\lambda$  and  $\text{Var}[X] = \alpha/\lambda^2$ . Therefore, I have

$$E[X | \mathbf{Y}] = \frac{\alpha + n}{\sum_{i=1}^n Y_i + \beta}$$

and

$$\text{Var}[X | \mathbf{Y}] = \frac{\alpha + n}{(\sum_{i=1}^n Y_i + \beta)^2}.$$



## Chapter 10

# Introduction to Simulation Using Python

The Python programming language is one of the most popular languages in both academia and industry. It is heavily used in data science for simple data analysis and complex machine learning. By most accounts, in the last few years, Python has eclipsed the R programming language in popularity for scientific/statistical computation. Its popularity is due to intuitive and readable syntax that can be implemented in a powerful object oriented programming paradigm, if so desired, as well as being open source. It is for these reasons that I decided to transcribe the Introduction to Simulation chapter in Pishro-Nik's Introduction to Probability, Statistics and Random Processes book into Python.

This entire chapter was written in a Jupyter notebook, an interactive programming environment, primarily for Python, that can be run locally in a web browser. Jupyter notebooks are ideal for quick and interactive data analysis, incorporating markdown functionality for clean presentations and code sharing. If you are a fan of RStudio, you will most likely be fond of Jupyter notebooks. This entire notebook is available freely at [https://github.com/dsrub/solutions\\_to\\_\\_probability\\_statistics](https://github.com/dsrub/solutions_to__probability_statistics).

Additionally, much of this code was written using the Numpy/SciPy library, Python's main library for scientific computation and numerical methods. Numpy has a relatively clear and well documented API (<https://docs.scipy.org/doc/numpy/reference/index.html>), a reference which I utilize almost daily.

I start with a few basic imports, and define several functions I will use throughout the rest of this chapter.

```
#define html style element for notebook formatting
from IPython.core.display import HTML
```

```
with open('style.txt', 'r') as myfile:
    notebook_style = myfile.read().replace('\n', '')

HTML(notebook_style)
```

```
#import some relevant packages and plot inline
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
```

```
#define a few functions I will be using throughout the rest of the notebook
```

```
#function to print several of the RNGs to the screen
def print_vals(RNG_function, *args):
    for i in range(5):
        print('X_' + str(i) + ' = ', RNG_function(*args))
```

```
#plotting function
def plot_results(x, y, xlim=None, ylim=None, xlabel=None, ylabel=None, \
                title=None, labels=None):

    plt.figure(1, figsize = (6, 4))
    plt.rc('text', usetex=True)
    plt.rc('font', family = 'serif')
```

```

if labels:
    plt.plot(x[0], y[0], label=labels[0], linewidth = 2)
    plt.plot(x[1], y[1], label=labels[1], linewidth = 2)
    plt.legend(loc='upper right')
else:
    plt.plot(x, y, linewidth = 2)
if xlim:
    plt.xlim(xlim)
if ylim:
    plt.ylim(ylim)
if xlabel:
    plt.xlabel(xlabel, size = 15)
if ylabel:
    plt.ylabel(ylabel, size = 15)
if title:
    plt.title(title, size=15)

plt.xticks(fontsize = 15)
plt.yticks(fontsize = 15);

```

**Example 1.** (Bernoulli) Simulate tossing a coin with probability of heads  $p$ .

*Solution:* We can utilize the algorithm presented in the book, which uses random variables drawn from a  $Unif(0,1)$  distribution. The following function implements this algorithm in Python to generate a  $Bern(p)$  (pseudo) random variable.

```

def draw_bern(p, N):
    """
    A  $Bern(p)$  pseudo-RNG
    """
    U = np.random.uniform(size = N)
    if N == 1: U = U[0]
    X = (U < p) + 0

    return X

#print a few examples of the RNGs to the screen
p = 0.5
print_vals(draw_bern, p, 1)

X_0 = 0
X_1 = 0
X_2 = 0
X_3 = 0
X_4 = 1

```

Note that we can directly sample from a  $Bern(p)$  distribution with Numpy's binomial random number generator (RNG) by setting  $n = 1$  with: `np.random.binomial(1, p)`.

**Example 2.** (Coin Toss Simulation) Write code to simulate tossing a fair coin to see how the law of large numbers works.

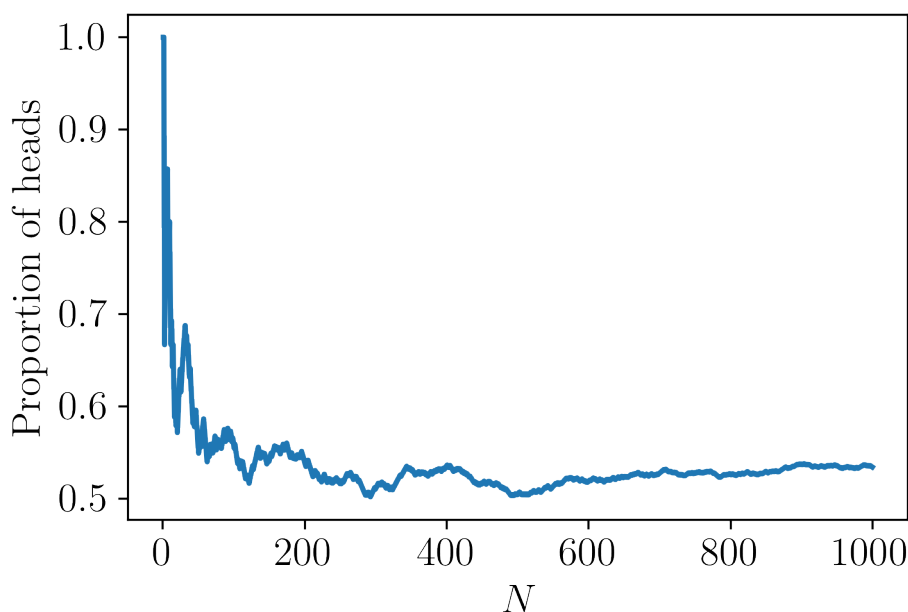
*Solution:* I draw 1000  $Bern(0.5)$  random variables and compute the cumulative average.

```
#generate data, compute proportion of heads and plot

#set a seed for reproducibility
np.random.seed(2)

X = draw_bern(0.5, 1000)
avg = np.cumsum(X)/(np.arange(1000) + 1)
plot_results(np.arange(1000) + 1, avg, xlabel='$N$', ylabel='Proportion of heads')

#reset seed
np.random.seed(0)
```



**Example 3.** (Binomial) Generate a  $Bin(50, 0.2)$  random variable.

*Solution:* If  $X_1, X_2, \dots, X_n$  are drawn iid from a  $Bern(p)$  distribution, then we can express a  $Bin(n, p)$  random variable as  $X = X_1 + X_2 + \dots + X_n$ . Therefore we can utilize the code we have already written for drawing a  $Bern(p)$  random variable to draw a  $Bin(n, p)$  random variable.

```
def draw_bin(n, p, N):
    """
    A Bin(n, p) pseudo-RNG
    """
    if N > 1:
        U = np.random.uniform(0, 1, (N, n))
        X = np.sum(U < p, axis = 1)
    else:
        U = np.random.uniform(0, 1, n)
        X = np.sum(U < p)
```

```

return X

#print a few examples of the RGNs to the screen
n = 50
p = 0.2
print_vals(draw_bin, n, p, 1)

X_0 = 8
X_1 = 17
X_2 = 3
X_3 = 13
X_4 = 10

```

Note that we can directly sample from a  $\text{Bin}(n, p)$  distribution with Numpy's binomial RNG with: `np.random.binomial(n, p)`.

**Example 4.** Write an algorithm to simulate the value of a random variable  $X$  such that:

$$P_X(x) = \begin{cases} 0.35 & \text{for } x = 1 \\ 0.15 & \text{for } x = 2 \\ 0.4 & \text{for } x = 3 \\ 0.1 & \text{for } x = 4. \end{cases}$$

*Solution:* We can utilize the algorithm presented in the book which divides the unit interval into 4 partitioned sets and uses a uniformly drawn random variable.

```

def draw_general_discrete(P, R_X, N):
    """
    A pseudo-RNG for any arbitrary discrete PMF specified by R_X and
    corresponding probabilities P
    """
    F_X = np.cumsum([0] + P)

    X_arr = []
    U_arr = np.random.uniform(0, 1, size = N)
    for U in U_arr:
        X = R_X[np.sum(U > F_X)-1]

        #take care of edge case where U = 0
        if U == 0:
            X = R_X[0]
        X_arr.append(X)
    if N == 1: X_arr = X_arr[0]

    return X_arr

#print a few examples of the RGNs to the screen
P = [0.35, 0.15, 0.4, 0.1]
R_X = [1, 2, 3, 4]
print_vals(draw_general_discrete, P, R_X, 1)

```

```

X_0 = 2
X_1 = 4
X_2 = 3
X_3 = 3
X_4 = 4

```

Note that we can directly sample from a discrete PMF using Numpy's multinomial RNG. A multinomial distribution is the  $k$  dimensional analogue of a binomial distribution, where  $k > 2$ . The multinomial distribution is a distribution over random vectors,  $\mathbf{X}$  (of size  $k$ ), where the entries in the vectors can take on values from  $0, 1, \dots, n$ , subject to  $X_1 + X_2 + \dots + X_k = n$ , where  $X_i$  represents the  $i^{th}$  component of  $\mathbf{X}$ .

If a binomial random variable represents the number of heads we flip out of  $n$  coin tosses (where the probability of heads is  $p$ ), then a multinomial random variable represents the number of times we roll a 1, the number of times we roll a 2, ..., the number of times we roll a  $k$ , when rolling a  $k$  sided die  $n$  times. For each roll, the probability of rolling the  $i^{th}$  face of the die is  $p_i$  (where  $\sum_{i=1}^k p_i = 1$ ). We store the value for the number times we roll the  $i^{th}$  face of the die in  $X_i$ . To denote a random vector drawn from a multinomial distribution, the notation,  $\mathbf{X} \sim Mult(n, \mathbf{p})$ , is typical, where  $\mathbf{p}$  denotes the  $k$  dimensional vector with the  $i^{th}$  component of  $\mathbf{p}$  given by  $p_i$ .

To directly sample from a discrete PMF with (ordered) range array  $\mathbf{R\_X}$  and associated probability array  $\mathbf{P}$  we can use Numpy's multinomial RNG function by setting  $n = 1$  (one roll). To sample one time we can use the code: `X = R_X[np.argmax(np.random.multinomial(1, pvals=P))]`, and to sample  $N$  times, we can use the code: `X = [R_X[np.argmax(x)] for x in np.random.multinomial(1, pvals=P, size=N)]`.

Additionally, to sample from an arbitrary discrete PMF, we can also use Numpy's choice function, which samples randomly from a specified list, where each entry in the list is sampled according to a specified probability. To sample  $N$  values from an array  $\mathbf{R\_X}$ , with corresponding probability array  $\mathbf{P}$ , we can use the code: `X = np.random.choice(R_X, size=N, replace=True, p=P)`. Make sure to specify `replace=True` to sample with replacement.

**Example 5.** (Exponential) Generate an  $Exp(1)$  random variable.

*Solution:* Using the method of inverse transformation, as shown in the book, for a strictly increasing CDF,  $F$ , the random variable  $X = F^{-1}(U)$ , where  $U \sim Unif(0,1)$ , has distribution  $X \sim F$ . Therefore, it is not difficult to show that,

$$-\frac{1}{\lambda} \ln(U) \sim Exp(\lambda),$$

where the fact that  $1 - U \sim Unif(0,1)$  has been used.

```

def draw_exp(lam, N):
    """
    An Exp(lambda) pseudo-RNG using the method of inverse transformation
    """
    U = np.random.uniform(0, 1, size = N)
    if N == 1:
        U = U[0]
    X = (-1/lam)*np.log(U)

    return X

```

```
#print a few examples of the RGNs to the screen
lam = 1
print_vals(draw_exp, lam, 1)
```

```
X_0 = 2.4838379957
X_1 = 0.593858616083
X_2 = 0.53703944167
X_3 = 0.0388069650697
X_4 = 1.23049637556
```

Note that we can directly sample from an  $Exp(\lambda)$  distribution with Numpy's exponential RNG with: `np.random.exponential(lam)`.

**Example 6.** (Gamma) Generate a  $Gamma(20, 1)$  random variable.

*Solution:* If  $X_1, X_2, \dots, X_n$  are drawn iid from an  $Exp(\lambda)$  distribution, then  $Y = X_1 + X_2 + \dots + X_n \sim Gamma(n, \lambda)$ . Therefore, to generate a  $Gamma(n, \lambda)$  random variable, we need only to generate  $n$  independent  $Exp(\lambda)$  random variables and add them.

```
def draw_gamma(alpha, lam, N):
    """
    A Gamma(n, lambda) pseudo-RNG using the method of inverse transformation
    """
    n = alpha
    if N > 1:
        U = np.random.uniform(0, 1, size = (N, n))
        X = np.sum((-1/lam)*np.log(U), axis = 1)

    else:
        U = np.random.uniform(0, 1, size = n)
        X = np.sum((-1/lam)*np.log(U))

    return X
```

```
#print a few examples of the RGNs to the screen
alpha = 20
lam = 1
print_vals(draw_gamma, alpha, lam, 1)
```

```
X_0 = 17.4925086879
X_1 = 20.6155480241
X_2 = 26.9115218192
X_3 = 22.3654600391
X_4 = 22.331744631
```

Note that we can directly sample from a  $Gamma(n, \lambda)$  distribution with Numpy's gamma RNG with: `np.random.gamma(shape, scale)`.

**Example 7.** (Poisson) Generate a Poisson random variable. Hint: In this example, use the fact that the number of events in the interval  $[0, t]$  has Poisson distribution when the elapsed times between the events are Exponential.

*Solution:* As shown in the book, we need only to continuously generate  $Exp(\lambda)$  variables

and count the number of draws it takes for the sum to be greater than 1. The Poisson random variable is then the count minus 1.

```
def draw_poiss(lam, N):
    """
    A Poiss(lambda) pseudo-RNG
    """
    X_list = []

    for _ in range(N):
        summ = 0
        count = 0
        while summ <= 1:
            summ += draw_exp(lam, 1)
            count += 1
        X_list.append(count-1)

    if N == 1:
        return X_list[0]
    else:
        return X_list
```

```
#print a few examples of the RGNs to the screen
lam = 1
print_vals(draw_poiss, lam, 1)
```

```
X_0 = 0
X_1 = 2
X_2 = 2
X_3 = 1
X_4 = 2
```

Note that we can directly sample from a  $Poiss(\lambda)$  distributions with Numpy's: `np.random.poisson(lam)` function.

**Example 8.** (Box-Muller) Generate 5000 pairs of normal random variables and plot both histograms.

*Solution:* Using the Box-Muller transformation as described in the book:

```
def draw_gaus_pairs(N):
    """
    An  $N(0, 1)$  pseudo-RNG to draw  $N$  pairs of independent using the Box-Muller
    transformation
    """
    U1 = np.random.uniform(size = N)
    U2 = np.random.uniform(size = N)

    Z1 = np.sqrt(-2*np.log(U1))*np.cos(2*np.pi*U2)
    Z2 = np.sqrt(-2*np.log(U1))*np.sin(2*np.pi*U2)
```



```

return (Z1, Z2)

#print a few examples of the RGNs to the screen
Z1_arr, Z2_arr = draw_gaus_pairs(5)

for i, (Z1, Z2) in enumerate(zip(Z1_arr, Z2_arr)):
    print('(Z_1, Z_2)_' + str(i) + ' = (', Z1, Z2, ')')

```

```

(Z_1, Z_2)_0 = ( 0.722134435205 -0.189448731182 )
(Z_1, Z_2)_1 = ( -0.918558147113 0.247330492682 )
(Z_1, Z_2)_2 = ( -1.42078058592 -0.914027516141 )
(Z_1, Z_2)_3 = ( 1.19799155228 -1.49105841693 )
(Z_1, Z_2)_4 = ( -0.65055423687 0.179187077215 )

```

In addition to plotting the histograms (plot in the first panel below) I also make a scatter plot of the 2 Gaussian random variables. The Box-Muller method produces pairs of independent random variables, and indeed, in the plot we see a bivariate Normal distribution with no correlation, i.e., it is axis-aligned (recall that independence  $\implies \rho = 0$ ). I further compute the correlation coefficient between  $Z_1$  and  $Z_2$  and it is indeed very close to 0.

```

#plot the histograms and scatter plot

#set seed for reproducibility
np.random.seed(8)

#generate data
Z1_arr, Z2_arr = draw_gaus_pairs(5000)

#plot histograms
f, (ax1, ax2) = plt.subplots(1, 2, figsize=(10, 4))

bins = np.linspace(-5, 5, 50)
ax1.hist(Z1_arr, bins, alpha=0.5, normed=1, label='$Z_1$', edgecolor = 'black')
ax1.hist(Z2_arr, bins, alpha=0.5, normed=1, label='$Z_2$', edgecolor = 'black')
ax1.legend(loc='upper right')
ax1.set_xlabel('$Z$', size = 15)
ax1.set_ylabel('Probability Density', size = 15)
ax1.tick_params(labelsize=15)

#plot scatter plot
ax2.scatter(Z1_arr, Z2_arr, s=2)
ax2.set_xlabel('$Z_1$', size = 15)
ax2.set_ylabel('$Z_2$', size = 15)
ax2.set_ylim((-4, 4))
ax2.set_xlim((-4, 4))
ax2.tick_params(labelsize=15)

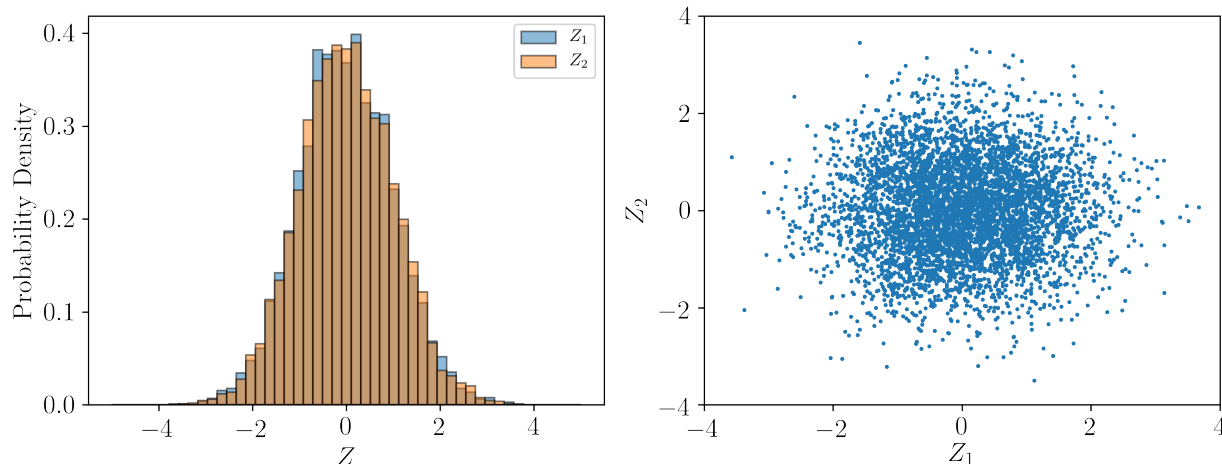
print('correlation coefficient = ', np.corrcoef(Z1_arr, Z2_arr)[0, 1])

#reset seed

```

```
np.random.seed(0)
```

```
correlation coefficient = 0.0177349514518
```



Note that we can directly sample from a  $\mathcal{N}(0, 1)$  distribution with Numpy's normal RNG with: `np.random.randn(d0, d1, ..., dn)`, where `d0`, `d1`, ..., `dn` are the dimensions of the desired output array.

**Exercise 1.** Write Python programs to generate  $Geom(p)$  and  $Pascal(m, p)$  random variables.

*Solution:* As in the book, I generate  $Bern(p)$  random variables until the first success and count the number of draws to generate a  $Geom(p)$  random variable. To generate a  $Pascal(m, p)$  random variable, I generate  $Bern(p)$  random variables until I obtain  $m$  successes and count the number of draws.

```
def draw_geom(p, N):
    """
    A Geom(p) pseudo-RNG
    """
    X_list = []
    for _ in range(N):
        count = 0
        X = 0
        while X == 0:
            X = draw_bern(p, 1)
            count += 1
        X_list.append(count)

    if N == 1:
        return X_list[0]
    else:
        return X_list

# print a few examples of the RNGs to the screen
p = 0.2
print_vals(draw_geom, p, 1)
```

```

X_0 = 15
X_1 = 1
X_2 = 1
X_3 = 8
X_4 = 2

```

```

def draw_pascal(m, p, N):
    """
    A Pascal(m, p) pseudo-RNG
    """
    X_list = []
    for _ in range(N):
        count_succ = 0
        count = 0
        while count_succ < m:
            X = draw_bern(p, 1)
            count_succ += X
            count += 1
        X_list.append(count)

    if N == 1:
        return X_list[0]
    else:
        return X_list

```

```

#print a few examples of the RNGs to the screen
p = 0.2
m = 2
print_vals(draw_pascal, m, p, 1)

```

```

X_0 = 17
X_1 = 10
X_2 = 7
X_3 = 3
X_4 = 4

```

Note that we can directly sample from  $Geom(p)$  and  $Pascal(m, p)$  distributions with Numpy's `np.random.geometric(p)` and `np.random.negative_binomial(n, p)` functions respectively.

**Exercise 2.** (Poisson) Use the algorithm for generating discrete random variables to obtain a Poisson random variable with parameter  $\lambda = 2$ .

*Solution:*

```

from scipy.misc import factorial

def draw_poiss2(lam, N):
    """
    A Poiss(lambda) pseudo-RNG using the method to generate an
    arbitrary discrete random variable
    """
    X_list = []

```

```

for _ in range(N):
    P = np.exp(-lam)
    i = 0
    U = np.random.uniform()
    while U >= P:
        i += 1
        P += np.exp(-lam)*lam**i/(factorial(i)+0)

    X_list.append(i)

if N == 1:
    return X_list[0]
else:
    return X_list

#print a few examples of the RGNs to the screen
lam = 2
print_vals(draw_pois2, 2, 1)

```

```

X_0 = 3
X_1 = 0
X_2 = 5
X_3 = 2
X_4 = 5

```

**Exercise 3.** Explain how to generate a random variable with the density

$$f(x) = 2.5x\sqrt{x}$$

for  $0 < x < 1$ .

*Solution:* The CDF is given by  $F_X(x) = 2.5 \int_0^x x'^{3/2} dx' = x^{5/2}$ , and therefore  $F_X^{-1}(x) = x^{2/5}$ . Using the method of inverse transformation, if  $U \sim \text{Unif}(0, 1)$ , then  $F_X^{-1}(U)$  is distributed according to the desired distribution.

```

def draw_dist3():
    """
    A pseudo-RNG for the distribution in Exercise 3
    """
    U = np.random.uniform()
    return U**(0.4)

#print a few examples of the RGNs to the screen
print_vals(draw_dist3)

```

```

X_0 = 0.8178201131579468
X_1 = 0.8861754700680049
X_2 = 0.27369087549414306
X_3 = 0.6033871249144047
X_4 = 0.4285059109745954

```

**Exercise 4.** Use the inverse transformation method to generate a random variable having distribution function

$$F_X(x) = \frac{x^2 + x}{2},$$

for  $0 \leq x \leq 1$ .

*Solution:* By inverting the CDF, we have that.

$$F_X^{-1}(x) = -\frac{1}{2} + \sqrt{\frac{1}{4} + 2x},$$

for  $0 \leq x \leq 1$ .

```
def draw_dist4():
    """
    A pseudo-RNG for the distribution in Exercise 4
    """
    U = np.random.uniform()
    return -0.5 + np.sqrt(0.25 + 2*U)
```

```
#print a few examples of the RNGs to the screen
print_vals(draw_dist4)
```

```
X_0 = 0.417758353296
X_1 = 0.198180089883
X_2 = 0.441257859881
X_3 = 0.538521058539
X_4 = 0.115056902
```

**Exercise 5.** Let  $X$  have a standard Cauchy distribution. function

$$F_X(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}.$$

Assuming you have  $U \sim Unif(0, 1)$ , explain how to generate  $X$ . Then, use this result to produce 1000 samples of  $X$  and compute the sample mean. Repeat the experiment 100 times. What do you observe and why?

*Solution:* The inverse CDF is given by  $F_X^{-1}(x) = \tan[\pi(x - 1/2)]$ .

```
def draw_stand_cauchy(N):
    """
    A standard Cauchy pseudo-RNG using the method of inverse transformation
    """
    U = np.random.uniform(size = N)
    X = np.tan(np.pi*(U - 1/2))

    if N == 1: return X[0]
    else: return X
```

```

#print a few examples of the RGNs to the screen
print_vals(draw_stand_cauchy, 1)

X_0 = 0.691013110859
X_1 = 0.212342443875
X_2 = -0.907695727473
X_3 = 0.0731660554841
X_4 = -3.28946953204

#plot means for 100 trials

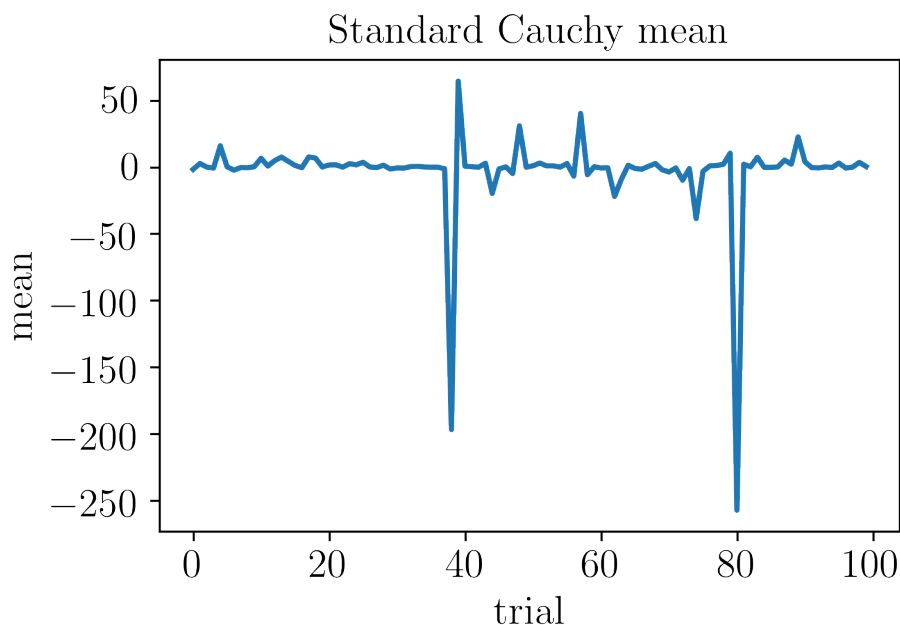
#set seed for reproducibility
np.random.seed(5)

#compute means and plot
means = [np.mean(np.array(draw_stand_cauchy(1000))) for _ in range(100)]

plot_results(range(100), means, xlabel='trial', ylabel='mean', \
             title='Standard Cauchy mean')

#reset seed
np.random.seed(0)

```



We see that the means for each trial vary wildly. This is because the Cauchy distribution actually has no mean.

**Exercise 6.** (The Rejection Method) When we use the Inverse Transformation Method, we need a simple form of the CDF,  $F(x)$ , that allows direct computation of  $X = F^{-1}(U)$ . When  $F(x)$  doesn't have a simple form but the PDF,  $f(x)$ , is available, random variables with density  $f(x)$  can be generated by the rejection method. Suppose you have a method for generating a random variable having density function  $g(x)$ . Now, assume you want to generate a random variable having

density function  $f(x)$ . Let  $c$  be a constant such that  $f(y)/g(y) \leq c$  (for all  $y$ ). Show that the following method generates a random variable,  $X$ , with density function  $f(x)$ .

```

1) initialize  $U$  and  $Y$  such that  $U > \frac{f(Y)}{cg(Y)}$ 

repeat until  $U \leq \frac{f(Y)}{cg(Y)}$  {
    2) Generate  $Y$  having density  $g$ 
    3) Generate a random number  $U$  from  $Unif(0,1)$ 
}

4) Set  $X = Y$ 

```

*Solution:*

Firstly, as a technical matter, note that  $c \geq 1$ , which can be shown by integrating both sides of  $f(y) \leq cg(y)$ .

We see that this algorithm keeps iterating until it outputs a random variable  $Y$ , given that we know that  $U \leq \frac{f(Y)}{cg(Y)}$ . Therefore, the goal is to show that the random variable  $Y|U \leq \frac{f(Y)}{cg(Y)}$  has PDF  $f(y)$  (or equivalently CDF  $F(y)$ ). In other words, we must show that  $P(Y \leq y | U \leq \frac{f(Y)}{cg(Y)}) = F(y)$ . I show this with Baye's rule:

$$\begin{aligned}
 P\left(Y \leq y | U \leq \frac{f(Y)}{cg(Y)}\right) &= \frac{P\left(U \leq \frac{f(Y)}{cg(Y)} | Y \leq y\right) P(Y \leq y)}{P\left(U \leq \frac{f(Y)}{cg(Y)}\right)} \\
 &= \frac{P\left(U \leq \frac{f(Y)}{cg(Y)} | Y \leq y\right) G(y)}{P\left(U \leq \frac{f(Y)}{cg(Y)}\right)}.
 \end{aligned}$$

Thus, we must calculate the quantities:  $P\left(U \leq \frac{f(Y)}{cg(Y)} | Y \leq y\right)$  and  $P\left(U \leq \frac{f(Y)}{cg(Y)}\right)$ .

As an intermediate step, note that

$$\begin{aligned}
 P\left(U \leq \frac{f(Y)}{cg(Y)} | Y = y\right) &= P\left(U \leq \frac{f(y)}{cg(y)} | Y = y\right) \\
 &= P\left(U \leq \frac{f(y)}{cg(y)}\right) \\
 &= F_U\left(\frac{f(y)}{cg(y)}\right) \\
 &= \frac{f(y)}{cg(y)},
 \end{aligned}$$

where in the second line I have used that  $U$  and  $Y$  are independent and in the fourth I have used the fact that for a uniform distribution  $F_U(u) = u$ . Notice that the requirement that  $f(y)/g(y) \leq c$  (for all  $y$ ) is crucial at this step. This is because  $f(y)/g(y) \leq c \implies c > 0$  (since  $f(y)$  and  $g(y)$  are positive), so that  $0 < f(y)/cg(y) \leq 1$ . If this condition did not hold, then the above expression

would be  $\min\{1, \frac{f(y)}{cg(y)}\}$ , for positive  $c$  and 0 for negative  $c$ , which would interfere with the rest of the derivation.

I may now calculate  $P\left(U \leq \frac{f(Y)}{cg(Y)}\right)$ :

$$\begin{aligned} P\left(U \leq \frac{f(Y)}{cg(Y)}\right) &= \int_{-\infty}^{\infty} P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y = y\right) g(y) dy \\ &= \int_{-\infty}^{\infty} \frac{f(y)}{cg(y)} g(y) dy \\ &= \frac{1}{c} \int_{-\infty}^{\infty} f(y) dy \\ &= \frac{1}{c}. \end{aligned}$$

I now calculate the remaining quantity:

$$\begin{aligned} P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y \leq y\right) &= \frac{P\left(U \leq \frac{f(Y)}{cg(Y)}, Y \leq y\right)}{G(y)} \\ &= \frac{\int_{-\infty}^{\infty} P\left(U \leq \frac{f(Y)}{cg(Y)}, Y \leq y \mid Y = v\right) g(v) dv}{G(y)} \\ &= \frac{\int_{-\infty}^{\infty} P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y \leq y, Y = v\right) P(Y \leq y \mid Y = v) g(v) dv}{G(y)}, \end{aligned}$$

where in the second line I have used the law of total probability, and in the third line I have used the definition of conditional probability. Note that:

$$P(Y \leq y \mid Y = v) = \begin{cases} 1 & \text{for } v \leq y \\ 0 & \text{for } v > y, \end{cases}$$

and thus

$$\begin{aligned} P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y \leq y\right) &= \frac{\int_{-\infty}^y P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y \leq y, Y = v\right) g(v) dv}{G(y)} \\ &= \frac{\int_{-\infty}^y P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y = v\right) g(v) dv}{G(y)} \\ &= \frac{\int_{-\infty}^y \frac{f(v)}{cg(v)} g(v) dv}{G(y)} \\ &= \frac{\frac{1}{c} F(y)}{G(y)}, \end{aligned}$$

where in the second line I have used the fact that conditioning on  $Y = v$  already implies that  $Y \leq y$  since we only consider values of  $v$  less than or equal to  $y$  in the integration. In the third line I have used the expression for  $P\left(U \leq \frac{f(Y)}{cg(Y)} \mid Y = y\right)$  that we derived above.



Inserting these quantities into Baye's rule:

$$\begin{aligned}
 P\left(Y \leq y | U \leq \frac{f(Y)}{cg(Y)}\right) &= \frac{P\left(U \leq \frac{f(Y)}{cg(Y)} | Y \leq y\right) G(y)}{P\left(U \leq \frac{f(Y)}{cg(Y)}\right)} \\
 &= \frac{\frac{\frac{1}{c} F(y)}{G(y)} G(y)}{\frac{1}{c}} \\
 &= F(y),
 \end{aligned}$$

which is what we set out to prove.

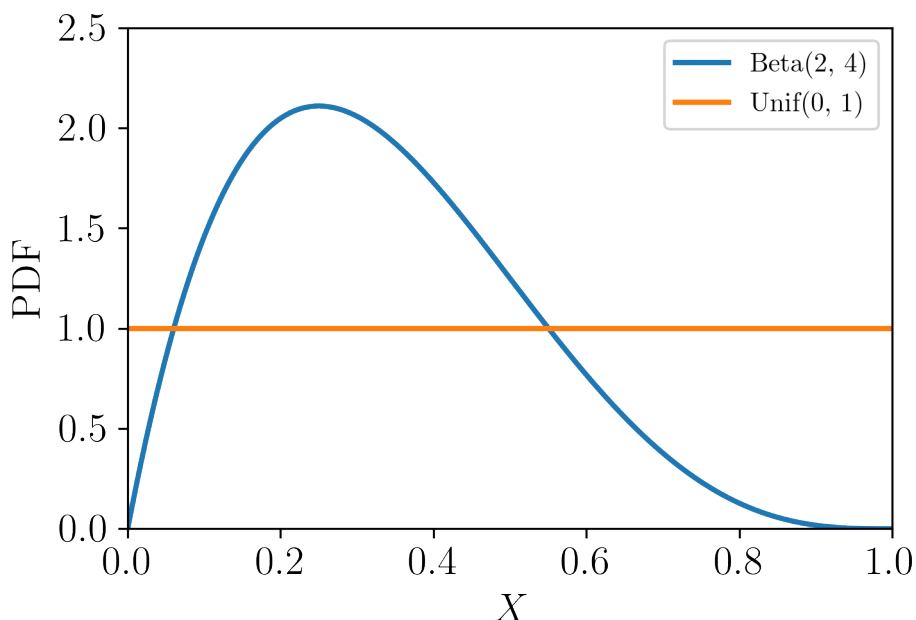
**Exercise 7.** Use the rejection method to generate a random variable having density function  $Beta(2, 4)$ . Hint: Assume  $g(x) = 1$  for  $0 < x < 1$ .

*Solution:* I first visualize these distributions so we can get a handle on what we are dealing with.

```
#plot Beta(2, 4) and Unif(0, 1)
from scipy.stats import beta

x1, x2 = np.linspace(0, 1, 1000), np.linspace(0, 1, 1000)
y1, y2 = beta.pdf(x1, 2, 4), x2*0+1
labels = ['Beta(2, 4)', 'Unif(0, 1)']

plot_results([x1, x2], [y1, y2], xlim=(0, 1), ylim=(0, 2.5), xlabel='$X$', \
             ylabel='PDF', labels=labels)
```



Since  $f(x)/g(x)$  (where  $f(x)$  is the PDF of the Beta and  $g(x)$  is the PDF of the uniform) needs to be smaller than  $c$  for all  $x$  in the support of these distributions, a fine value of  $c$  to use would be 2.5 since it is evident from the plot that this value satisfies the requirement. The book uses the

smallest possible value of  $c$ , i.e., the max of the  $Beta(2, 4)$  distribution, which it derives analytically and finds to be  $135/64 \approx 2.11$ . It is not necessary to use the smallest value of  $c$ , but will certainly help the speed of the algorithm since the algorithm only stops when  $U \leq f(Y)/cg(Y)$ . I will stick with the value of 2.5 just to illustrate that the algorithm works for this value as well.

```
def draw_beta_2_4(N):
    """
    A Beta(2, 4) pseudo-RNG using the rejection method
    """
    c = 2.5

    X_list = []
    for _ in range(N):

        U = 1
        f_Y = 0
        g_Y = 1

        while U > f_Y/(c*g_Y):
            Y = np.random.uniform()
            U = np.random.uniform()

            f_Y = 20*Y*(1-Y)**3
            g_Y = 1

        X_list.append(Y)

    if N == 1:
        return X_list[0]
    else:
        return X_list

#print a few examples of the RNGs to the screen
print_vals(draw_beta_2_4, 1)
```

```
X_0 = 0.4236547993389047
X_1 = 0.07103605819788694
X_2 = 0.11827442586893322
X_3 = 0.5218483217500717
X_4 = 0.26455561210462697
```

Note that we can directly sample from a  $Beta(\alpha, \beta)$  distribution with Numpy's beta RNG with: `np.random.beta(a, b)`.

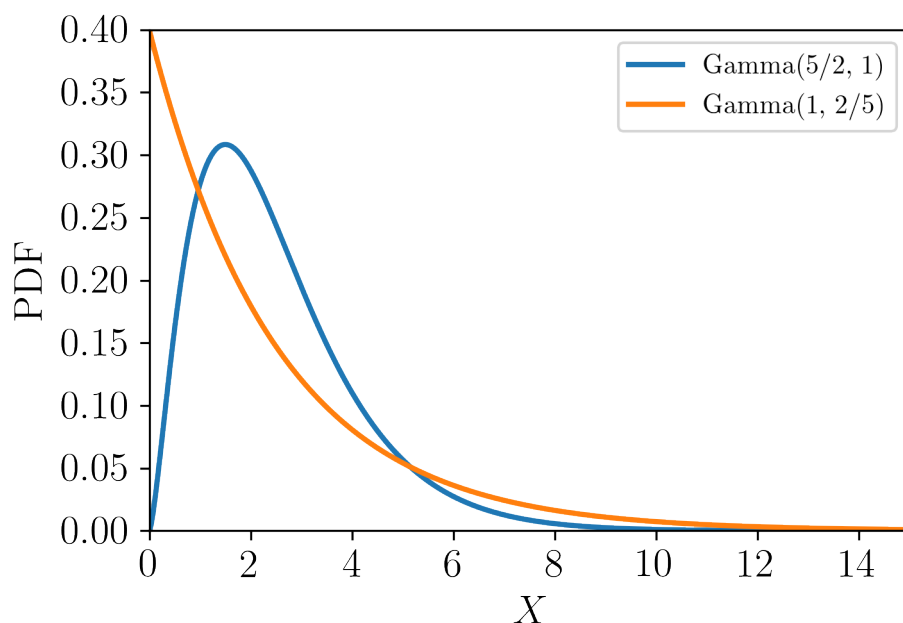
**Exercise 8.** Use the rejection method to generate a random variable having the  $Gamma(5/2, 1)$  density function. Hint: Assume  $g(x)$  is the PDF of the  $Gamma(1, 2/5)$ .

Solution: Note that there is a mistake in the phrasing of the question in the book. The PDF for  $g(x)$  should be  $Gamma(1, 2/5)$ , not  $Gamma(5/2, 1)$ . Also note that we cannot use the method that we used in **Example 6**, since in this case  $\alpha$  is not an integer (however, we can use that method to draw from  $g(x)$ ). I first visualize these distributions so we can get a handle on what we are dealing with.

```
#plot Gamma(5/2, 1) and Gamma(1, 2/5)

x1, x2 = np.linspace(0, 20, 1000), np.linspace(0, 20, 1000)
f, g = (4/(3*np.sqrt(np.pi)))*(x1**1.5)*np.exp(-x1), 0.4*np.exp(-0.4*x2)
labels = ['Gamma(5/2, 1)', 'Gamma(1, 2/5)']

plot_results([x1, x2], [f, g], xlim=(0, 15), ylim=(0, 0.4), xlabel='$X$', \
              ylabel='PDF', labels=labels)
```



The  $\max\{f(x)/g(x)\}$  for  $x > 0$  is approximately given by:

```
np.max(f/g)

1.6587150033103788
```

As a sanity check, this value is very close to the analytically derived value in the book, which is  $\frac{10}{3\sqrt{\pi}} \left(\frac{5}{2}\right)^{3/2} e^{-3/2} \approx 1.6587162$ . Therefore, I set the value of  $c$  to be 1.7, and use the function I wrote in **Example. 6**, `draw_gamma(alpha, lam, N)`, to draw from  $g(x)$ .

```
def draw_gamma_2(alpha, lam, N):
    """
    A Gamma(5/2, 1) pseudo-RNG using the rejection method
    """
    c = 1.7

    X_list = []
    for _ in range(N):

        U = 1
        f_Y = 0
```

```

g_Y = 1

while U > f_Y/(c*g_Y):
    Y = draw_gamma(1, 0.4, 1)
    U = np.random.uniform()

    f_Y = (4/(3*np.sqrt(np.pi)))*(Y**1.5)*np.exp(-Y)
    g_Y = 0.4*np.exp(-0.4*Y)

X_list.append(Y)

if N == 1:
    return X_list[0]
else:
    return X_list

#print a few examples of the RGNs to the screen
print_vals(draw_gamma_2, 5/2, 1, 1)

```

```

X_0 = 1.96233211971
X_1 = 1.22716649756
X_2 = 2.55754781375
X_3 = 0.900161721137
X_4 = 3.89706921546

```

**Exercise 9.** Use the rejection method to generate a standard normal random variable. Hint: Assume  $g(x)$  is the PDF of the exponential distribution with  $\lambda = 1$ .

*Solution* As in the book, to solve this problem, I use the rejection method to sample from a half Gaussian:

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

with range  $(0, \infty)$ , with an  $Exp(1)$  distribution for  $g(x)$ . The book analytically computes  $\max\{f(x)/g(x)\}$  to be  $\sqrt{2e/\pi} \approx 1.32$ , and I thus use  $c = 1.4$ . Once the algorithm is able to sample from the half Gaussian, to turn this distribution into a full Gaussian with range  $\mathbb{R}$ , one need only to randomly multiply by -1. I therefore sample  $Q (\in \{0, 1\})$  from a  $Bern(0.5)$  distribution and multiply by  $1-2Q (\in \{-1, 1\})$  in order to sample from the full Gaussian.

```

def draw_standard_normal(N):
    """
    A standard normal pseudo-RNG using the rejection method
    """
    c = 1.4

    X_list = []
    for _ in range(N):

        U = 1

```

```

f_Y = 0
g_Y = 1

while U > f_Y/(c*g_Y):
    Y = draw_exp(1, 1)
    U = np.random.uniform()

    f_Y = (2/np.sqrt(2*np.pi))*np.exp(-(Y**2)/2)
    g_Y = np.exp(-Y)

    # draw Bern(0.5) random variable for the sign
    Q = draw_bern(0.5, 1)

    X_list.append(Y*(1-2*Q))

if N == 1:
    return X_list[0]
else:
    return X_list

#print a few examples of the RGNs to the screen
print_vals(draw_standard_normal, 1)

```

```

X_0 = 1.1538237197
X_1 = -2.28234324111
X_2 = -0.426012274543
X_3 = -1.40884434358
X_4 = -0.421092193245

```

**Exercise 10.** Use the rejection method to generate a  $\text{Gamma}(2, 1)$  random variable conditional on its value being greater than 5. Hint: Assume  $g(x)$  is the density function of exponential distribution.

*Solution* As in the book, I use an  $\text{Exp}(0.5)$  conditioned on  $X > 5$  as the distribution for  $g(x)$ . It is not difficult to show by integrating the PDF of this distribution that  $G^{-1}(x) = 5 - 2\ln(1 - x)$  (where  $G$  is the CDF). I therefore use the method of inverse transformation to first draw a random variable from this distribution ( $Y$ ). Note that for  $U \sim \text{Unif}(0, 1)$ ,  $1 - U \sim \text{Unif}(0, 1)$ , and therefore the formula for  $G^{-1}(U)$  can be simplified to  $5 - 2\ln(U)$ . I then use the rejection method to sample from the desired distribution. By maximizing  $f(x)/g(x)$ , the book shows that  $c$  must be greater than  $5/3$ , and I therefore use  $c = 1.7$ .

```

def draw_gamma_2_1_cond_5(N):
    """
    A Gamma(2, 1) conditional on X>5 pseudo-RNG using the rejection method
    """
    c = 1.7

    X_list = []
    for _ in range(N):

        U = 1

```

```

f_Y = 0
g_Y = 1

while U > f_Y/(c*g_Y):
    Y = 5 - 2*np.log(np.random.uniform())
    U = np.random.uniform()

    f_Y = Y*np.exp(5-Y)/6
    g_Y = np.exp((5-Y)/2)/2

    X_list.append(Y)

if N == 1:
    return X_list[0]
else:
    return X_list

#print a few examples of the RNGs to the screen
print_vals(draw_gamma_2_1_cond_5, 1)

```

```

X_0 = 6.76250850879
X_1 = 5.73497460514
X_2 = 5.14665551227
X_3 = 5.8087003199
X_4 = 5.66723645483

```

Notice that, as required, the random variables are all  $> 5$ .

As a final check to close this chapter, I draw samples from most of the RNG functions that I implemented above, compute the corresponding PMFs/PDFs, and compare to the theoretical distributions. I first check the discrete distributions, and I start by writing a function that will compute the empirical PMFs. Note that the phrase, “empirical PMF”, (and “empirical PDF”) is standard terminology to refer to the probability distribution associated with a sample of data. Formally, for a collection of data,  $\{x_i\}_{i=1}^N$ , they are given by

$$P_X(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{I}(x = x_i)$$

for the empirical PMF, and by

$$f_X(x) = \frac{1}{N} \sum_{i=1}^N \delta(x = x_i)$$

for the empirical PDF (where  $\mathbf{I}(\cdot)$  is the indicator function, and  $\delta(\cdot)$  is the delta function).

```

def compute_PMFs(counts, xrange):
    """
    Compute empirical PMFs from a specified array of random variables,
    and a specified range
    """

```

```

count_arr = []
xrange2 = range(np.max([np.max(xrange), np.max(counts)])+1)
for i in xrange2:
    count_arr.append(np.sum(counts==i))
pmf = np.array(count_arr)/np.sum(np.array(count_arr))
return pmf[np.min(xrange):np.max(xrange)+1]

```

I now compute the theoretical distributions, generate the data and compute the empirical distributions.

```

from scipy.stats import bernoulli, binom, poisson, geom, nbinom
#set seed for reproducibility
np.random.seed(1984)

x_ranges = [range(2), range(26), range(9), range(1, 11), range(4, 26), range(9)]

#compute PMF arrays for the theoretical distributions
numpy_dists = [bernoulli, binom, poisson, geom, nbinom, poisson]
numpy_args = [[0.5], [50, 0.2], [1], [0.5], [4, 0.5, 4], [1]]
numpy_y = [np_dist.pmf(xrange, *np_args) for np_dist, xrange, np_args in \
            zip(numpy_dists, x_ranges, numpy_args)]

N = 1000 #number of points to sample

# draw random variables from my functions and compute corresponding PMFs
my_rngs = [draw_bern, draw_bin, draw_pois, draw_geom, draw_pascal, draw_pois2]
my_args = [[0.5, N], [50, 0.2, N], [1, N], [0.5, N], [4, 0.5, N], [1, N]]
my_counts = [rng(*args) for rng, args in zip(my_rngs, my_args)]
my_y = [compute_PMFs(np.array(counts), xrange) for counts, xrange in \
        zip(my_counts, x_ranges)]

```

Finally, I plot the results.

```

#plot theoretical and empirical PMFs
names = ['Bern(0.5)', 'Bin(50, 0.2)', 'Pois(1)', 'Geom(0.5)', 'Pascal(4, 0.5)', \
         'Pois(1) (discrete RV method)']
legend_loc = ['upper right']*6
legend_loc[0] = 'upper center'

f, [[ax1, ax2, ax3], [ax4, ax5, ax6]] = plt.subplots(2, 3, figsize=(15, 10))
ax_arr = [ax1, ax2, ax3, ax4, ax5, ax6]

for i, ax in enumerate(ax_arr):
    ax.plot(x_ranges[i], numpy_y[i], 'bo', ms=8, label='Theoretical Dist.', \
            alpha=.8)
    ax.vlines(x_ranges[i], 0, numpy_y[i], colors='b', lw=5, alpha=0.5)
    ax.plot(x_ranges[i], my_y[i], 'bo', ms=8, label='Empirical Dist.', \
            color='green', alpha=.8)

    ax.legend(loc=legend_loc[i], fontsize=11)

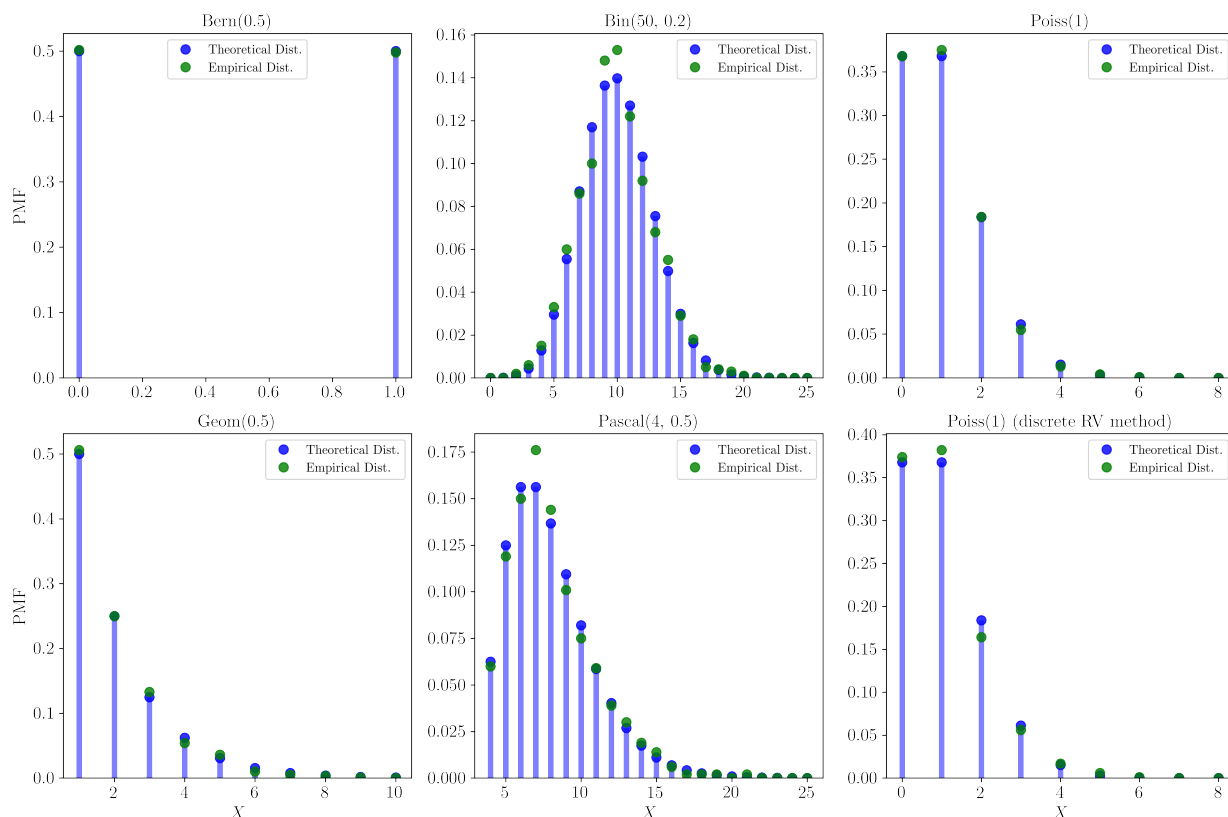
```

```

ax.set_title(names[i], size = 15)
ax.set_ylim(ymin=0)
ax.tick_params(labelsize=15)

if i in [3, 4, 5]:
    ax.set_xlabel('$X$', size = 15)
if i in [0, 3]:
    ax.set_ylabel('PMF', size = 15)

```



We see that the empirical distributions match almost perfectly with the theoretical distributions, with even better correspondence for larger  $N$ .

I now check some of the continuous RNG functions that I implemented in this chapter. I first start by computing the theoretical distributions and generating the data.

```

from scipy.stats import expon, gamma, cauchy, beta, norm

x_ranges = [np.linspace(0, 8, 1000), np.linspace(0, 50, 1000), \
            np.linspace(-20, 20, 1000), np.linspace(0, 1, 1000), \
            np.linspace(0, 15, 1000), np.linspace(-5, 5, 1000)]

#compute PDF arrays for the theoretical distributions
numpy_dists = [expon, gamma, cauchy, beta, gamma, norm]
numpy_args = [[0, 1], [20, 0, 1], [], [2, 4], [5/2, 0, 1], []]
numpy_y = [np_dist.pdf(xrange, *np_args) for np_dist, xrange, np_args in \
            zip(numpy_dists, x_ranges, numpy_args)]

```



```

N = 1000 #number of points to sample
# draw random variables from my functions to be plotted as histograms in next cell
my_rngs = [draw_exp, draw_gamma, draw_stand_cauchy, draw_beta_2_4, draw_gamma_2, \
            draw_standard_normal]
my_args = [[1, N], [20, 1, N], [N], [N], [5/2, 1, N], [N]]
my_rvs = [rng(*args) for rng, args in zip(my_rngs, my_args)]

#reset seed
np.random.seed(0)

```

I now plot normalized histograms of the data and compare to the theoretical distributions. Again, we see almost perfect correspondence between the empirical and theoretical distributions. The correspondence becomes even better with larger values of  $N$ .

```

#plot theoretical and empirical PDFs
names = ['Exp(1) (inverse trans.)', 'Gamma(20, 1) (inverse trans.)', \
         'Cauchy(0, 1) (inverse trans.)', 'Beta(2, 4) (rejection)', \
         'Gamma(5/2, 1) (rejection)', 'N(0, 1) (rejection)']
bin_arr = [50, 35, 60, 45, 45, 35]
xlims=[(0, 8), (0, 50), (-20, 20), (0, 1), (0, 15), (-5, 5)]
range_arr = [None]*6
range_arr[2] = (-20, 20)

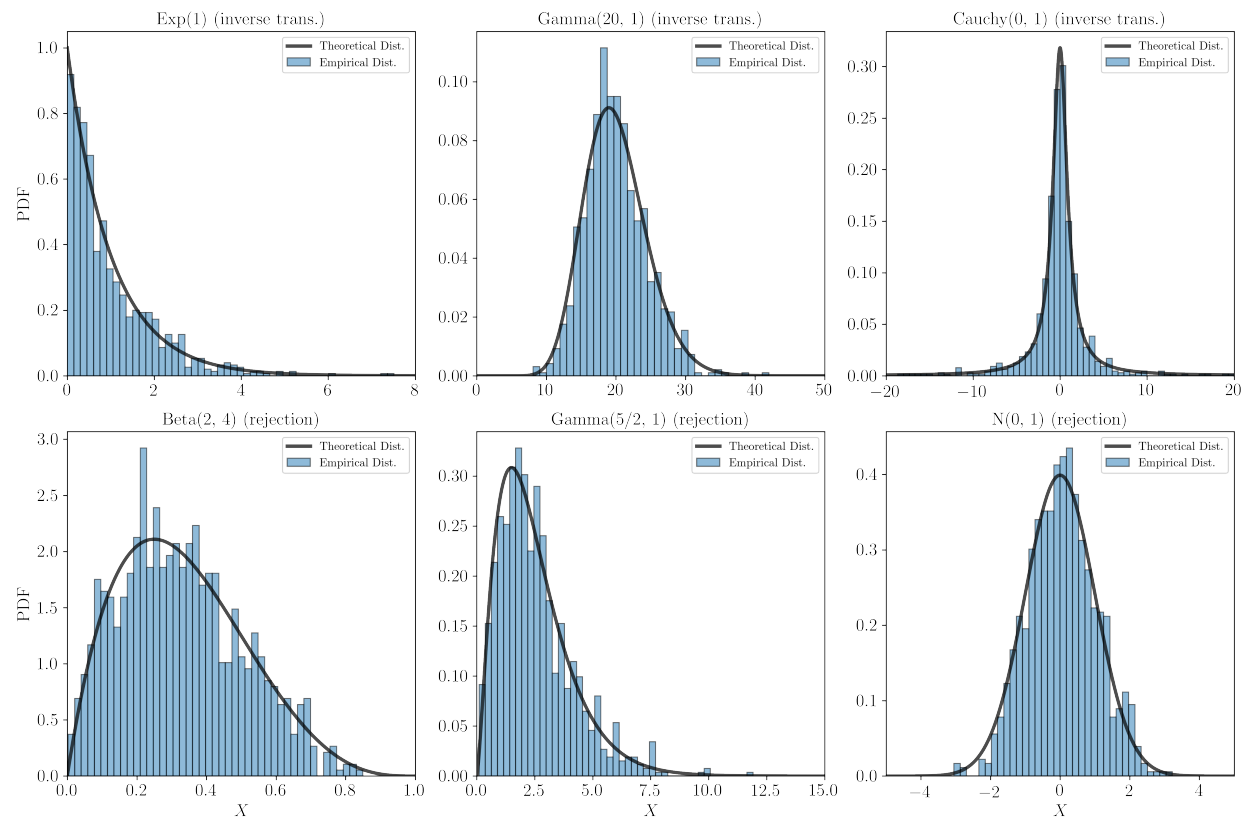
f, [[ax1, ax2, ax3], [ax4, ax5, ax6]] = plt.subplots(2, 3, figsize=(15, 10))
ax_arr = [ax1, ax2, ax3, ax4, ax5, ax6]

for i, ax in enumerate(ax_arr):
    ax.plot(x_ranges[i], numpy_y[i], label='Theoretical Dist.', color='black', \
            linewidth=3, alpha=.7)
    ax.hist(my_rvs[i], bins=bin_arr[i], alpha=.5, edgecolor='black', normed=True, \
            label='Empirical Dist.', range=range_arr[i])

    ax.set_title(names[i], size = 15)
    ax.legend(loc='upper right', fontsize=10)
    ax.set_xlim(xlims[i])
    ax.tick_params(labelsize=15)

    if i in [3, 4, 5]:
        ax.set_xlabel('$X$', size = 15)
    if i in [0, 3]:
        ax.set_ylabel('PDF', size = 15)

```



## Chapter 11

# Recursive Methods

**Problem 1.**

- (a) The characteristic polynomial for this recursive formula is

$$x^2 - 2x + \frac{3}{4} = 0,$$

which has roots  $1/2$  and  $3/2$ , and therefore:

$$a_n = \alpha \left(\frac{3}{2}\right)^n + \beta \left(\frac{1}{2}\right)^n.$$

Using the initial conditions  $a_0 = 1$  and  $a_1 = -1$  leads to  $\alpha = -1$  and  $\beta = 1$ . Thus, the solution to the recurrence equation is:

$$a_n = -\left(\frac{3}{2}\right)^n + \left(\frac{1}{2}\right)^n.$$

- (b) The characteristic polynomial for this recursive formula is

$$x^2 - 4x + 4 = 0,$$

which can be factored into  $(x - 2)^2 = 0$ . The polynomial thus has one root,  $x = 2$ , with a multiplicity of 2, and therefore:

$$a_n = \alpha 2^n + \beta n 2^n.$$

Using the initial conditions  $a_0 = 2$  and  $a_1 = 6$  leads to  $\alpha = 2$  and  $\beta = 1$ . Thus, the solution to the recurrence equation is:

$$a_n = 2n^{n+1} + n2^n.$$

**Problem 2.**

- (a) Let
- $A_{n,k}$
- be the event of observing exactly
- $k$
- heads out of
- $n$
- coin tosses, and let
- $H$
- denote the event that the last coin toss is a heads. By conditioning on the last coin toss I obtain:

$$\begin{aligned} P(A_{n,k}) &= P(A_{n,k}|H)p + P(A_{n,k}|H^c)(1-p) \\ &= P(A_{n-1,k-1})p + P(A_{n-1,k})(1-p), \end{aligned}$$

where the equality follows because if the last coin toss is heads, then we need exactly  $k - 1$  heads from the first  $n - 1$  tosses, and if the last coin toss is tails, then we need exactly  $k$  heads from the first  $n - 1$  tosses. Converting this to the notation used in the problem:

$$a_{n,k} = a_{n-1,k-1}p + a_{n-1,k}(1-p)$$

$\implies$

$$a_{n+1,k+1} = a_{n,k}p + a_{n,k+1}(1-p).$$

- (b) We recognize that this is precisely a binomial experiment, so the probability associated with exactly
- $k$
- heads out of
- $n$
- is given by
- $\binom{n}{k}p^k(1-p)^{n-k}$
- , and therefore, using the equation above,

$$\binom{n+1}{k+1}p^{k+1}(1-p)^{(n+1)-(k+1)} = p\binom{n}{k}p^k(1-p)^{n-k} + (1-p)\binom{n}{k+1}p^{k+1}(1-p)^{n-(k+1)},$$

which, when simplified results in:

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

We need the restriction that  $0 \leq k < n$  to hold for this equation to be true, since the original recursion relation does not hold if  $k = n$  (in that case the last flip cannot be a tails since we need all flips to be heads, so that the  $P(A_{n,k}|H^c)$  term should be 0).

**Problem 3.** Let  $A$  be the desired event and let  $q = 1 - p$  be the probability of tails. To solve this problem, I first condition on whether the first toss is a heads or tails:

$$P(A) = P(A|H)p + P(A|T)q.$$

To help solve for  $P(A|H)$ , I now condition on the second toss:

$$\begin{aligned} P(A|H) &= P(A|HH)p + P(A|HT)q \\ &= 1 \cdot p + P(A|T)q, \end{aligned}$$

where  $P(A|HH) = 1$  since if you flip 2 consecutive heads, the experiment is done and where  $P(A|HT) = P(A|T)$ , since the first heads does not matter because we are interested in 2 consecutive heads, and 1 isolated heads does not get us any closer to the event  $A$ .

To help solve for  $P(A|T)$ , I also condition on the second toss:

$$\begin{aligned} P(A|T) &= P(A|TH)p + P(A|TT)q \\ &= P(A|H)p + 0 \cdot q, \end{aligned}$$

where  $P(A|TT) = 0$  since if you flip 2 consecutive tails, the experiment is done (and the desired even did not occur) and where  $P(A|TH) = P(A|H)$ , for essentially the same reason that  $P(A|HT) = P(A|T)$  as described above.

I now re-express these 3 equations in slightly more readable notation:

$$\begin{cases} a = a^H p + a^T q \\ a^H = p + a^T q \\ a^T = a^H p, \end{cases}$$

and we see that we have a system of 3 equations with 3 unknowns. Solving for  $a$  and plugging  $q = 1 - p$  back in I find that:

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

As a check, we know that in the limit that  $p$  goes to 1, this expression should evaluate to unity (we definitely get  $HH$  before  $TT$ ) and in the limit that  $p$  goes to 0 this expression should evaluate to 0 (we definitely get  $TT$  before  $HH$ ). Indeed it is easily to check that this expression satisfies these 2 limits.

**Problem 4.** Let  $A_n$  be the event that the number of heads out of  $n$  tosses is divisible by 3,  $p$  be the probability of heads and  $q = 1 - p$  be the probability of tails. To solve this problem recursively, I first condition on whether the first toss is a heads or tails:

$$\begin{aligned} P(A_n) &= P(A_n|H)p + P(A_n|T)q \\ &= P(A_n|H)p + P(A_{n-1})q. \end{aligned}$$

Here,  $P(A_n|T) = P(A_{n-1})$  since if the first toss is a tails, as in the sequence below, we have observed no heads,

$$\frac{T}{n} \frac{-}{n-1} \frac{-}{n-2} \cdots \frac{-}{1}$$

so that the experiment just starts over at 1 less total flips ( $n-1$ ), and is equivalent to the sequence below:

$$\frac{-}{n-1} \frac{-}{n-2} \cdots \frac{-}{1}$$

The numbers below the sequence  $n, n-1, n-2, \dots$  show the number of flips remaining before you make that particular flip.

To solve for  $P(A_n|H)$ , I condition on the second toss:

$$\begin{aligned} P(A_n|H) &= P(A_n|HH)p + P(A_n|HT)q \\ &= P(A_n|HH)p + P(A_{n-1}|H)q. \end{aligned}$$

Here,  $P(A_n|HT) = P(A_{n-1}|H)$  since the the probability of  $A_n$  for the sequence below,

$$\frac{H}{n} \frac{T}{n-1} \frac{-}{n-2} \cdots \frac{-}{1}$$

is the same as the probability of  $A_n$  for a sequence that starts with 1 heads with  $n-1$  flips:

$$\frac{H}{n-1} \frac{-}{n-2} \cdots \frac{-}{1}$$

Finally, to solve for  $P(A_n|HH)$ , I condition on the third toss:

$$\begin{aligned} P(A_n|HH) &= P(A_n|HHH)p + P(A_n|HHT)q \\ &= P(A_n-3)p + P(A_{n-1}|HH)q. \end{aligned}$$

Here,  $P(A_n|HHT) = P(A_{n-1}|HH)$  since the the probability of  $A_n$  for the sequence below,

$$\frac{H}{n} \frac{H}{n-1} \frac{T}{n-2} \frac{-}{n-3} \cdots \frac{-}{1}$$

is the same as the probability of  $A_n$  for a sequence that starts with 2 heads with  $n-1$  flips:

$$\frac{H}{n-1} \frac{H}{n-2} \frac{-}{n-3} \cdots \frac{-}{1}$$

Also  $P(A_n|HHH) = P(A_n-3)$  since, if we have already gotten 3 heads in the first 3 flips then the probability that the number of heads flipped in the sequence is divisible by 3 is the same as if the remaining  $n-3$  flips is divisible by 3.

I summarize this set of recursive equations in somewhat more readable notation:

$$\begin{cases} a_n^{HH} = a_{n-3}p + a_{n-1}^{HH}q \\ a_n^H = a_n^{HH}p + a_{n-1}^Hq \\ a_n = a_n^H p + a_{n-1}q. \end{cases}$$

We see that the equations are a coupled set of recursive equations, so must be solved simultaneously by first solving for  $a_n^{HH}$ , then using this to solve for  $a_n^H$ , then using this to solve for  $a_n$ , iteratively until we reach the desired value of  $n$ . In order to do this, we will need several initial conditions which we can easily be compute by hand. For a sequence with  $n=1$ , the number of heads is divisible by 3 if we throw one tails (probability  $q$ ). For a sequence with  $n=2$ , the number of heads is divisible by 3 if we throw two tails (probability  $q^2$ ). For a sequence with  $n=3$ , the number of heads is divisible by 3 if we throw 3 tails or 3 heads (probability  $p^3 + q^3$ ):

$$\begin{aligned}a_1 &= q \\a_2 &= q^2 \\a_3 &= p^3 + q^3.\end{aligned}$$

For a sequence that starts with 1 head and with  $n = 1$ , the number of heads is never divisible by 3 (probability 0). For a sequence that starts with 1 head and with  $n = 2$ , the number of heads is never divisible by 3 (probability 0). For a sequence that starts with 1 head and with  $n = 3$ , the number of heads is only divisible by 3 if we throw 2 heads after the first (probability  $p^2$ ):

$$\begin{aligned}a_1^H &= 0 \\a_2^H &= 0 \\a_3^H &= p^2.\end{aligned}$$

Finally, for a sequence that starts with 2 head and with  $n = 1$ , the number of heads is never divisible by 3 (probability 0). For a sequence that starts with 2 head and with  $n = 2$ , the number of heads is never divisible by 3 (probability 0). For a sequence that starts with 2 head and with  $n = 3$ , the number of heads is only divisible by 3 if we throw 1 head after the first 2 (probability  $p$ ):

$$\begin{aligned}a_1^{HH} &= 0 \\a_2^{HH} &= 0 \\a_3^{HH} &= p.\end{aligned}$$

I can check this coupled set of recursive equations by recognizing that we can compute  $P(A_n)$  directly using the binomial distribution and only summing over the number of successes which are divisible by 3. This can be written as:

$$P(A_n) = \sum_{k=0}^{\lfloor \frac{n}{3} \rfloor} \binom{n}{3k} p^{3k} q^{n-3k}.$$

I wrote a python function (below) to compute  $P(A_n)$  using both methods. I compute  $P(A_n)$  for a range of  $n$ , for several values of  $p$ , and plot  $P(A_n)$  calculated recursively against  $P(A_n)$  calculated with the binomial distribution as well as the 45 degree line in Fig. 11.1. If there is perfect agreement between the 2 methods, the points should lie along this line, and indeed this is exactly what we see.

```
import numpy as np
from scipy.special import binom

def compute_binom_recur(p, N):
    """
    Compute probability that the number of heads out of N
    coin flips (each of probability p) is divisible by 3.
    Returns the probability computed from the binomial
    distribution and the probability computed from recursion.
    """
```

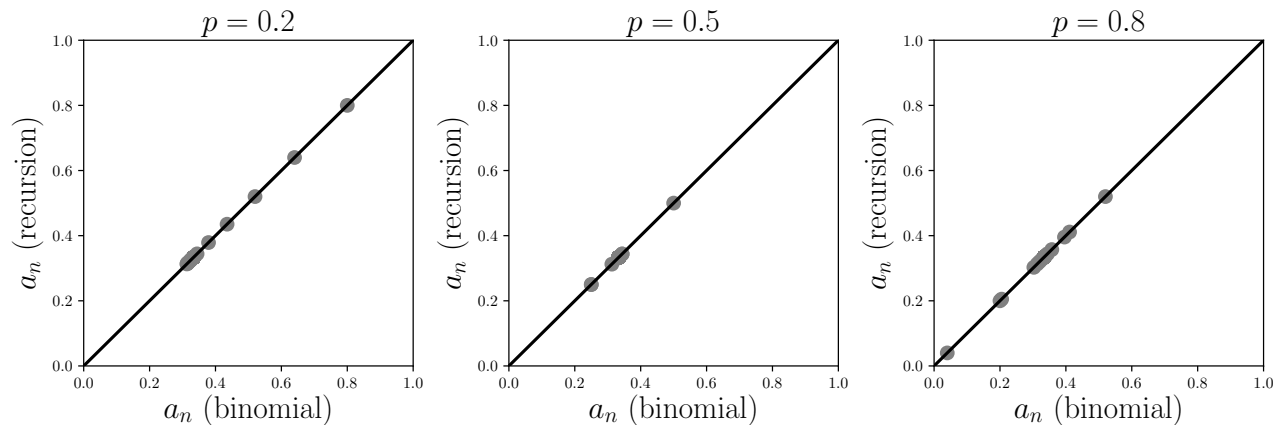


Figure 11.1: Comparison of  $P(A_n)$  calculated recursively against  $P(A_n)$  calculated with the binomial distribution (Problem 4).

```
#compute the probability from the binomial
P_arr=[]
for n in range(1, N):
    P = np.sum(np.array([binom(n, 3*k)*p**(3*k)*(1-p)**(n-3*k)\
                        for k in range(0, int(np.floor(n/3)+1))]))
    P_arr.append(P)

#compute the probability from recursion
q = 1-p

#initialize recursion
an_HH = [0, 0, p]
an_H = [0, 0, p**2]
an = [q, q**2, p**3+q**3]

for _ in range(N-4):
    #recursion update equations
    an_HH_new = an[-3]*p+an_HH[-1]*q
    an_H_new = an_HH_new*p+an_H[-1]*q
    an_new = an_H_new*p+an[-1]*q

    an_HH.append(an_HH_new)
    an_H.append(an_H_new)
    an.append(an_new)

return (P_arr, an)
```