# CHAPTER 2

**Solution to Problem 2.1.** Let $X$ be the number of points the MIT team earns over the weekend. We have

$$
\begin{aligned}
\mathbf{P}(X=0) &= 0.6\cdot 0.3 = 0.18,\\
\mathbf{P}(X=1) &= 0.4\cdot 0.5\cdot 0.3 + 0.6\cdot 0.5\cdot 0.7 = 0.27,\\
\mathbf{P}(X=2) &= 0.4\cdot 0.5\cdot 0.3 + 0.6\cdot 0.5\cdot 0.7 + 0.4\cdot 0.5\cdot 0.7\cdot 0.5 = 0.34,\\
\mathbf{P}(X=3) &= 0.4\cdot 0.5\cdot 0.7\cdot 0.5 + 0.4\cdot 0.5\cdot 0.7\cdot 0.5 = 0.14,\\
\mathbf{P}(X=4) &= 0.4\cdot 0.5\cdot 0.7\cdot 0.5 = 0.07,\\
\mathbf{P}(X>4) &= 0.
\end{aligned}
$$

**Solution to Problem 2.2.** The number of guests that have the same birthday as you is binomial with $p = 1/365$ and $n = 499$. Thus the probability that exactly one other guest has the same birthday is

$$\binom{499}{1}\frac{1}{365}\left(\frac{364}{365}\right)^{498} \approx 0.3486.$$

Let $\lambda = np = 499/365 \approx 1.367$. The Poisson approximation is $e^{-\lambda}\lambda = e^{-1.367}\cdot 1.367 \approx 0.3483$, which closely agrees with the correct probability based on the binomial.

**Solution to Problem 2.3.** (a) Let $L$ be the duration of the match. If Fischer wins a match consisting of $L$ games, then $L-1$ draws must first occur before he wins. Summing over all possible lengths, we obtain

$$\mathbf{P}(\text{Fischer wins}) = \sum_{l=1}^{10}(0.3)^{l-1}(0.4) = 0.571425.$$

(b) The match has length $L$ with $L < 10$, if and only if $(L-1)$ draws occur, followed by a win by either player. The match has length $L = 10$ if and only if 9 draws occur. The probability of a win by either player is 0.7. Thus

$$p_L(l) = \mathbf{P}(L=l) = \begin{cases} (0.3)^{l-1}(0.7), & l = 1,\ldots,9,\\ (0.3)^9, & l = 10,\\ 0, & \text{otherwise.}\end{cases}$$

**Solution to Problem 2.4.** (a) Let $X$ be the number of modems in use. For $k < 50$, the probability that $X = k$ is the same as the probability that $k$ out of 1000 customers need a connection:

$$p_X(k) = \binom{1000}{k}(0.01)^k(0.99)^{1000-k}, \qquad k = 0,1,\ldots,49.$$

The probability that $X = 50$, is the same as the probability that 50 or more out of 1000 customers need a connection:

$$p_X(50) = \sum_{k=50}^{1000} \binom{1000}{k} (0.01)^k (0.99)^{1000-k}.$$

(b) By approximating the binomial with a Poisson with parameter $\lambda = 1000 \cdot 0.01 = 10$, we have

$$p_X(k) = e^{-10} \frac{10^k}{k!}, \qquad k = 0, 1, \ldots, 49,$$

$$p_X(50) = \sum_{k=50}^{1000} e^{-10} \frac{10^k}{k!}.$$

(c) Let $A$ be the event that there are more customers needing a connection than there are modems. Then,

$$\mathbf{P}(A) = \sum_{k=51}^{1000} \binom{1000}{k} (0.01)^k (0.99)^{1000-k}.$$

With the Poisson approximation, $\mathbf{P}(A)$ is estimated by

$$\sum_{k=51}^{1000} e^{-10} \frac{10^k}{k!}.$$

**Solution to Problem 2.5.** (a) Let $X$ be the number of packets stored at the end of the first slot. For $k < b$, the probability that $X = k$ is the same as the probability that $k$ packets are generated by the source:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots, b-1,$$

while

$$p_X(b) = \sum_{k=b}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - \sum_{k=0}^{b-1} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Let $Y$ be the number of number of packets stored at the end of the second slot. Since $\min\{X, c\}$ is the number of packets transmitted in the second slot, we have $Y = X - \min\{X, c\}$. Thus,

$$p_Y(0) = \sum_{k=0}^{c} p_X(k) = \sum_{k=0}^{c} e^{-\lambda} \frac{\lambda^k}{k!},$$

$$p_Y(k) = p_X(k+c) = e^{-\lambda} \frac{\lambda^{k+c}}{(k+c)!}, \qquad k = 1, \ldots, b-c-1,$$

$$p_Y(b-c) = p_X(b) = 1 - \sum_{k=0}^{b-1} e^{-\lambda} \frac{\lambda^k}{k!}.$$

(b) The probability that some packets get discarded during the first slot is the same as the probability that more than $b$ packets are generated by the source, so it is equal to

$$\sum_{k=b+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!},$$

or

$$1 - \sum_{k=0}^{b} e^{-\lambda} \frac{\lambda^k}{k!}.$$

**Solution to Problem 2.6.** We consider the general case of part (b), and we show that $p > 1/2$ is a necessary and sufficient condition for $n = 2k+1$ games to be better than $n = 2k-1$ games. To prove this, let $N$ be the number of Celtics' wins in the first $2k-1$ games. If $A$ denotes the event that the Celtics win with $n = 2k+1$, and $B$ denotes the event that the Celtics win with $n = 2k-1$, then

$$\mathbf{P}(A) = \mathbf{P}(N \geq k+1) + \mathbf{P}(N = k) \cdot \big(1 - (1-p)^2\big) + \mathbf{P}(N = k-1) \cdot p^2,$$

$$\mathbf{P}(B) = \mathbf{P}(N \geq k) = \mathbf{P}(N = k) + \mathbf{P}(N \geq k+1),$$

and therefore

$$\begin{aligned} \mathbf{P}(A) - \mathbf{P}(B) &= \mathbf{P}(N = k-1) \cdot p^2 - \mathbf{P}(N = k) \cdot (1-p)^2 \\ &= \binom{2k-1}{k-1} p^{k-1} (1-p)^k p^2 - \binom{2k-1}{k} (1-p)^2 p^k (1-p)^{k-1} \\ &= \frac{(2k-1)!}{(k-1)!\, k!} p^k (1-p)^k (2p-1). \end{aligned}$$

It follows that $\mathbf{P}(A) > \mathbf{P}(B)$ if and only if $p > \frac{1}{2}$. Thus, a longer series is better for the better team.

**Solution to Problem 2.7.** Let random variable $X$ be the number of trials you need to open the door, and let $K_i$ be the event that the $i$th key selected opens the door.

(a) In case (1), we have

$$\begin{aligned} p_X(1) &= \mathbf{P}(K_1) = \frac{1}{5}, \\ p_X(2) &= \mathbf{P}(K_1^c)\mathbf{P}(K_2 \,|\, K_1^c) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5}, \\ p_X(3) &= \mathbf{P}(K_1^c)\mathbf{P}(K_2^c \,|\, K_1^c)\mathbf{P}(K_3 \,|\, K_1^c \cap K_2^c) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}. \end{aligned}$$

Proceeding similarly, we see that the PMF of $X$ is

$$p_X(x) = \frac{1}{5}, \qquad x = 1, 2, 3, 4, 5.$$

We can also view the problem as ordering the keys in advance and then trying them in succession, in which case the probability of any of the five keys being correct is 1/5.

In case (2), $X$ is a geometric random variable with $p = 1/5$, and its PMF is

$$p_X(k) = \frac{1}{5} \cdot \left(\frac{4}{5}\right)^{k-1}, \qquad k \geq 1.$$

(b) In case (1), we have

$$p_X(1) = \mathbf{P}(K_1) = \frac{2}{10},$$

$$p_X(2) = \mathbf{P}(K_1^c)\mathbf{P}(K_2 \,|\, K_1^c) = \frac{8}{10} \cdot \frac{2}{9},$$

$$p_X(3) = \mathbf{P}(K_1^c)\mathbf{P}(K_2^c \,|\, K_1^c)\mathbf{P}(K_3 \,|\, K_1^c \cap K_2^c) = \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{2}{8} = \frac{7}{10} \cdot \frac{2}{9}.$$

Proceeding similarly, we see that the PMF of $X$ is

$$p_X(x) = \frac{2 \cdot (10 - x)}{90}, \qquad x = 1, 2, \ldots, 10.$$

Consider now an alternative line of reasoning to derive the PMF of $X$. If we view the problem as ordering the keys in advance and then trying them in succession, the probability that the number of trials required is $x$ is the probability that the first $x - 1$ keys do not contain either of the two correct keys and the $x$th key is one of the correct keys. We can count the number of ways for this to happen and divide by the total number of ways to order the keys to determine $p_X(x)$. The total number of ways to order the keys is 10! For the $x$th key to be the first correct key, the other key must be among the last $10 - x$ keys, so there are $10 - x$ spots in which it can be located. There are 8! ways in which the other 8 keys can be in the other 8 locations. We must then multiply by two since either of the two correct keys could be in the $x$th position. We therefore have $2 \cdot 10 - x \cdot 8!$ ways for the $x$th key to be the first correct one and

$$p_X(x) = \frac{2 \cdot (10 - x)8!}{10!} = \frac{2 \cdot (10 - x)}{90}, \qquad x = 1, 2, \ldots, 10,$$

as before.

In case (2), $X$ is again a geometric random variable with $p = 1/5$.

**Solution to Problem 2.8.** For $k = 0, 1, \ldots, n - 1$, we have

$$\frac{p_X(k+1)}{p_X(k)} = \frac{\binom{n}{k+1} p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k} p^k(1-p)^{n-k}} = \frac{p}{1-p} \cdot \frac{n-k}{k+1}.$$

**Solution to Problem 2.9.** For $k = 1, \ldots, n$, we have

$$\frac{p_X(k)}{p_X(k-1)} = \frac{\binom{n}{k} p^k(1-p)^{n-k}}{\binom{n}{k-1} p^{k-1}(1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)} = \frac{(n+1)p - kp}{k - kp}.$$

If $k \leq k^*$, then $k \leq (n{+}1)p$, or equivalently $k{-}kp \leq (n{+}1)p{-}kp$, so that the above ratio is greater than or equal to 1. It follows that $p_X(k)$ is monotonically nondecreasing. If $k > k^*$, the ratio is less than one, and $p_X(k)$ is monotonically decreasing, as required.

**Solution to Problem 2.10.** Using the expression for the Poisson PMF, we have, for $k \geq 1$,

$$\frac{p_X(k)}{p_X(k-1)} = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \cdot \frac{(k-1)!}{\lambda^{k-1} \cdot e^{-\lambda}} = \frac{\lambda}{k}.$$

Thus if $k \leq \lambda$ the ratio is greater or equal to 1, and it follows that $p_X(k)$ is monotonically increasing. Otherwise, the ratio is less than one, and $p_X(k)$ is monotonically decreasing, as required.

**Solution to Problem 2.13.** We will use the PMF for the number of girls among the natural children together with the formula for the PMF of a function of a random variable. Let $N$ be the number of natural children that are girls. Then $N$ has a binomial PMF

$$p_N(k) = \begin{cases} \binom{5}{k} \cdot \left(\frac{1}{2}\right)^5, & \text{if } 0 \leq k \leq 5, \\ 0, & \text{otherwise.} \end{cases}$$

Let $G$ be the number of girls out of the 7 children, so that $G = N + 2$. By applying the formula for the PMF of a function of a random variable, we have

$$p_G(g) = \sum_{\{n \,|\, n+2=g\}} p_N(n) = p_N(g-2).$$

Thus

$$p_G(g) = \begin{cases} \binom{5}{g-2} \cdot \left(\frac{1}{2}\right)^5, & \text{if } 2 \leq g \leq 7, \\ 0, & \text{otherwise.} \end{cases}$$

**Solution to Problem 2.14.** (a) Using the formula $p_Y(y) = \sum_{\{x \,|\, x \bmod(3)=y\}} p_X(x)$, we obtain

$$\begin{aligned} p_Y(0) &= p_X(0) + p_X(3) + p_X(6) + p_X(9) = 4/10, \\ p_Y(1) &= p_X(1) + p_X(4) + p_X(7) = 3/10, \\ p_Y(2) &= p_X(2) + p_X(5) + p_X(8) = 3/10, \\ p_Y(y) &= 0, \qquad \text{if } y \notin \{0, 1, 2\}. \end{aligned}$$

(b) Similarly, using the formula $p_Y(y) = \sum_{\{x \,|\, 5 \bmod(x+1)=y\}} p_X(x)$, we obtain

$$p_Y(y) = \begin{cases} 2/10, & \text{if } y = 0, \\ 2/10, & \text{if } y = 1, \\ 1/10, & \text{if } y = 2, \\ 5/10, & \text{if } y = 5, \\ 0, & \text{otherwise.} \end{cases}$$

**Solution to Problem 2.15.** The random variable $Y$ takes the values $k \ln a$, where $k = 1, \ldots, n$, if and only if $X = a^k$ or $X = a^{-k}$. Furthermore, $Y$ takes the value 0, if and only if $X = 1$. Thus, we have

$$p_Y(y) = \begin{cases} \dfrac{2}{2n+1}, & \text{if } y = \ln a,\ 2\ln a, \ldots, k \ln a, \\ \dfrac{1}{2n+1}, & \text{if } y = 0, \\ 0, & \text{otherwise.} \end{cases}$$

**Solution to Problem 2.16.** (a) The scalar $a$ must satisfy

$$1 = \sum_x p_X(x) = \frac{1}{a} \sum_{x=-3}^{3} x^2,$$

so

$$a = \sum_{x=-3}^{3} x^2 = (-3)^2 + (-2)^2 + (-1)^2 + 1^2 + 2^2 + 3^2 = 28.$$

We also have $\mathbf{E}[X] = 0$ because the PMF is symmetric around 0.

(b) If $z \in \{1, 4, 9\}$, then

$$p_Z(z) = p_X(\sqrt{z}) + p_X(-\sqrt{z}) = \frac{z}{28} + \frac{z}{28} = \frac{z}{14}.$$

Otherwise $p_Z(z) = 0$.

(c) $\mathrm{var}(X) = \mathbf{E}[Z] = \sum_z z p_Z(z) = \sum_{z \in \{1,4,9\}} \frac{z^2}{14} = 7.$

(d) We have

$$\begin{aligned} \mathrm{var}(X) &= \sum_x (x - \mathbf{E}[X])^2 p_X(x) \\ &= 1^2 \cdot \big(p_X(-1) + p_X(1)\big) + 2^2 \cdot \big(p_X(-2) + p_X(2)\big) + 3^2 \cdot \big(p_X(-3) + p_X(3)\big) \\ &= 2 \cdot \frac{1}{28} + 8 \cdot \frac{4}{28} + 18 \cdot \frac{9}{28} \\ &= 7. \end{aligned}$$

**Solution to Problem 2.17.** If $X$ is the temperature in Celsius, the temperature in Fahrenheit is $Y = 32 + 9X/5$. Therefore,

$$\mathbf{E}[Y] = 32 + 9\mathbf{E}[X]/5 = 32 + 18 = 50.$$

Also

$$\mathrm{var}(Y) = (9/5)^2 \mathrm{var}(X),$$

where var($X$), the square of the given standard deviation of $X$, is equal to 100. Thus, the standard deviation of $Y$ is $(9/5) \cdot 10 = 18$. Hence a normal day in Fahrenheit is one for which the temperature is in the range $[32, 68]$.

**Solution to Problem 2.18.** We have

$$p_X(x) = \begin{cases} 1/(b-a+1), & \text{if } x = 2^k, \text{ where } a \le k \le b, \ k \text{ integer}, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\mathbf{E}[X] = \sum_{k=a}^{b} \frac{1}{b-a+1} 2^k = \frac{2^a}{b-a+1}(1 + 2 + \cdots + 2^{b-a}) = \frac{2^{b+1} - 2^a}{b-a+1}.$$

Similarly,

$$\mathbf{E}[X^2] = \sum_{k=a}^{b} \frac{1}{b-a+1} (2^k)^2 = \frac{4^{b+1} - 4^a}{3(b-a+1)},$$

and finally

$$\text{var}(X) = \frac{4^{b+1} - 4^a}{3(b-a+1)} - \left( \frac{2^{b+1} - 2^a}{b-a+1} \right)^2.$$

**Solution to Problem 2.19.** We will find the expected gain for each strategy, by computing the expected number of questions until we find the prize.

(a) With this strategy, the probability of finding the location of the prize with $i$ questions, where $i = 1, \ldots, 8$, is $1/10$. The probability of finding the location with 9 questions is $2/10$. Therefore, the expected number of questions is

$$\frac{2}{10} \cdot 9 + \frac{1}{10} \sum_{i=1}^{8} i = 5.4.$$

(b) It can be checked that for 4 of the 10 possible box numbers, exactly 4 questions will be needed, whereas for 6 of the 10 numbers, 3 questions will be needed. Therefore, with this strategy, the expected number of questions is

$$\frac{4}{10} \cdot 4 + \frac{6}{10} \cdot 3 = 3.4.$$

**Solution to Problem 2.20.** The number $C$ of candy bars you need to eat is a geometric random variable with parameter $p$. Thus the mean is $\mathbf{E}[C] = 1/p$, and the variance is $\text{var}(C) = (1-p)/p^2$.

**Solution to Problem 2.21.** The expected value of the gain for a single game is infinite since if $X$ is your gain, then

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k} = \sum_{k=1}^{\infty} 1 = \infty.$$

Thus if you are faced with the choice of playing for given fee $f$ or not playing at all, and your objective is to make the choice that maximizes your expected net gain, you would be willing to pay any value of $f$. However, this is in strong disagreement with the behavior of individuals. In fact experiments have shown that most people are willing to pay only about \$20 to \$30 to play the game. The discrepancy is due to a presumption that the amount one is willing to pay is determined by the expected gain. However, expected gain does not take into account a person's attitude towards risk taking.

**Solution to Problem 2.22.** (a) Let $X$ be the number of tosses until the game is over. Noting that $X$ is geometric with probability of success

$$\mathbf{P}\big(\{HT, TH\}\big) = p(1-q) + q(1-p),$$

we obtain

$$p_X(k) = \big(1 - p(1-q) - q(1-p)\big)^{k-1}\big(p(1-q) + q(1-p)\big), \qquad k = 1, 2, \ldots$$

Therefore

$$\mathbf{E}[X] = \frac{1}{p(1-q) + q(1-p)}$$

and

$$\text{var}(X) = \frac{pq + (1-p)(1-q)}{\big(p(1-q) + q(1-p)\big)^2}.$$

(b) The probability that the last toss of the first coin is a head is

$$\mathbf{P}\big(HT \,|\, \{HT, TH\}\big) = \frac{p(1-q)}{p(1-q) + (1-q)p}.$$

**Solution to Problem 2.23.** Let $X$ be the total number of tosses.

(a) For each toss after the first one, there is probability 1/2 that the result is the same as in the preceding toss. Thus, the random variable $X$ is of the form $X = Y + 1$, where $Y$ is a geometric random variable with parameter $p = 1/2$. It follows that

$$p_X(k) = \begin{cases} (1/2)^{k-1}, & \text{if } k \geq 2, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\mathbf{E}[X] = \mathbf{E}[Y] + 1 = \frac{1}{p} + 1 = 3.$$

We also have

$$\text{var}(X) = \text{var}(Y) = \frac{1-p}{p^2} = 2.$$

(b) If $k > 2$, there are $k-1$ sequences that lead to the event $\{X = k\}$. One such sequence is $H \cdots HT$, where $k-1$ heads are followed by a tail. The other $k-2$ possible sequences are of the form $T \cdots TH \cdots HT$, for various lengths of the initial $T \cdots T$

segment. For the case where $k = 2$, there is only one (hence $k-1$) possible sequence that leads to the event $\{X = k\}$, namely the sequence $HT$. Therefore, for any $k \geq 2$,

$$\mathbf{P}(X = k) = (k-1)(1/2)^k.$$

It follows that

$$p_X(k) = \begin{cases} (k-1)(1/2)^k, & \text{if } k \geq 2, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\mathbf{E}[X] = \sum_{k=2}^{\infty} k(k-1)(1/2)^k = \sum_{k=1}^{\infty} k(k-1)(1/2)^k = \sum_{k=1}^{\infty} k^2(1/2)^k - \sum_{k=1}^{\infty} k(1/2)^k = 6-2 = 4.$$

We have used here the equalities

$$\sum_{k=1}^{\infty} k(1/2)^k = \mathbf{E}[Y] = 2,$$

and

$$\sum_{k=1}^{\infty} k^2(1/2)^k = \mathbf{E}[Y^2] = \text{var}(Y) + \big(\mathbf{E}[Y]\big)^2 = 2 + 2^2 = 6,$$

where $Y$ is a geometric random variable with parameter $p = 1/2$.

**Solution to Problem 2.24.** (a) There are 21 integer pairs $(x, y)$ in the region

$$R = \Big\{(x, y) \mid -2 \leq x \leq 4,\ -1 \leq y - x \leq 1\Big\},$$

so that the joint PMF of $X$ and $Y$ is

$$p_{X,Y}(x, y) = \begin{cases} 1/21, & \text{if } (x, y) \text{ is in } R, \\ 0, & \text{otherwise.} \end{cases}$$

For each $x$ in the range $[-2, 4]$, there are three possible values of $Y$. Thus, we have

$$p_X(x) = \begin{cases} 3/21, & \text{if } x = -2, -1, 0, 1, 2, 3, 4, \\ 0, & \text{otherwise.} \end{cases}$$

The mean of $X$ is the midpoint of the range $[-2, 4]$:

$$\mathbf{E}[X] = 1.$$

The marginal PMF of $Y$ is obtained by using the tabular method. We have

$$p_Y(y) = \begin{cases} 1/21, & \text{if } y = -3, \\ 2/21, & \text{if } y = -2, \\ 3/21, & \text{if } y = -1, 0, 1, 2, 3, \\ 2/21, & \text{if } y = 4, \\ 1/21, & \text{if } y = 5, \\ 0, & \text{otherwise.} \end{cases}$$

The mean of $Y$ is

$$\mathbf{E}[Y] = \frac{1}{21} \cdot (-3+5) + \frac{2}{21} \cdot (-2+4) + \frac{3}{21} \cdot (-1+1+2+3) = 1.$$

(b) The profit is given by

$$P = 100X + 200Y,$$

so that

$$\mathbf{E}[P] = 100 \cdot \mathbf{E}[X] + 200 \cdot \mathbf{E}[Y] = 100 \cdot 1 + 200 \cdot 1 = 300.$$

**Solution to Problem 2.25.** (a) Since all possible values of $(I, J)$ are equally likely, we have

$$p_{I,J}(i,j) = \begin{cases} \frac{1}{\sum_{k=1}^{n} m_k}, & \text{if } j \le m_i, \\ 0, & \text{otherwise.} \end{cases}$$

The marginal PMFs are given by

$$p_I(i) = \sum_{j=1}^{m} p_{I,J}(i,j) = \frac{m_i}{\sum_{k=1}^{n} m_k}, \qquad i = 1, \ldots, n,$$

$$p_J(j) = \sum_{i=1}^{n} p_{I,J}(i,j) = \frac{l_j}{\sum_{k=1}^{n} m_k}, \qquad j = 1, \ldots, m,$$

where $l_j$ is the number of students that have answered question $j$, i.e., students $i$ with $j \le m_i$.

(b) The expected value of the score of student $i$ is the sum of the expected values $p_{ij}a + (1 - p_{ij})b$ of the scores on questions $j$ with $j = 1, \ldots, m_i$, i.e.,

$$\sum_{j=1}^{m_i} \big(p_{ij}a + (1 - p_{ij})b\big).$$

**Solution to Problem 2.26.** (a) The possible values of the random variable $X$ are the ten numbers $101, \ldots, 110$, and the PMF is given by

$$p_X(k) = \begin{cases} \mathbf{P}(X > k-1) - \mathbf{P}(X > k), & \text{if } k = 101, \ldots 110, \\ 0, & \text{otherwise.} \end{cases}$$

We have $\mathbf{P}(X > 100) = 1$ and for $k = 101, \ldots 110$,

$$\begin{aligned} \mathbf{P}(X > k) &= \mathbf{P}(X_1 > k, X_2 > k, X_3 > k) \\ &= \mathbf{P}(X_1 > k)\,\mathbf{P}(X_2 > k)\,\mathbf{P}(X_3 > k) \\ &= \frac{(110-k)^3}{10^3}. \end{aligned}$$

It follows that

$$p_X(k) = \begin{cases} \dfrac{(111-k)^3 - (110-k)^3}{10^3}, & \text{if } k = 101, \ldots 110, \\ 0, & \text{otherwise.} \end{cases}$$

(An alternative solution is based on the notion of a CDF, which will be introduced in Chapter 3.)

(b) Since $X_i$ is uniformly distributed over the integers in the range $[101, 110]$, we have $\mathbf{E}[X_i] = (101 + 110)/2 = 105.5$. The expected value of $X$ is

$$\mathbf{E}[X] = \sum_{k=-\infty}^{\infty} k \cdot p_X(k) = \sum_{k=101}^{110} k \cdot p_x(k) = \sum_{k=101}^{110} k \cdot \frac{(111-k)^3 - (110-k)^3}{10^3}.$$

The above expression can be evaluated to be equal to 103.025. The expected improvement is therefore 105.5 - 103.025 = 2.475.

**Solution to Problem 2.31.** The marginal PMF $p_Y$ is given by the binomial formula

$$p_Y(y) = \binom{4}{y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{4-y}, \qquad y = 0, 1, \ldots, 4.$$

To compute the conditional PMF $p_{X|Y}$, note that given that $Y = y$, $X$ is the number of 1's in the remaining $4 - y$ rolls, each of which can take the 5 values $1, 3, 4, 5, 6$ with equal probability $1/5$. Thus, the conditional PMF $p_{X|Y}$ is binomial with parameters $4 - y$ and $p = 1/5$:

$$p_{X|Y}(x \,|\, y) = \binom{4-y}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{4-y-x},$$

for all nonnegative integers $x$ and $y$ such that $0 \le x + y \le 4$. The joint PMF is now given by

$$\begin{aligned} p_{X,Y}(x,y) &= p_Y(y) p_{X|Y}(x \,|\, y) \\ &= \binom{4}{y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{4-y} \binom{4-y}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{4-y-x}, \end{aligned}$$

for all nonnegative integers $x$ and $y$ such that $0 \le x + y \le 4$. For other values of $x$ and $y$, we have $p_{X,Y}(x,y) = 0$.

**Solution to Problem 2.32.** Let $X_i$ be the random variable taking the value 1 or 0 depending on whether the first partner of the $i$th couple has survived or not. Let $Y_i$ be the corresponding random variable for the second partner of the $i$th couple. Then, we have $S = \sum_{i=1}^m X_i Y_i$, and by using the total expectation theorem,

$$\begin{aligned} \mathbf{E}[S \,|\, A = a] &= \sum_{i=1}^m \mathbf{E}[X_i Y_i \,|\, A = a] \\ &= m\mathbf{E}[X_1 Y_1 \,|\, A = a] \\ &= m\mathbf{E}[Y_1 = 1 \,|\, X_1 = 1,\ A = a]\mathbf{P}(X_1 = 1 \,|\, A = a) \\ &= m\mathbf{P}(Y_1 = 1 \,|\, X_1 = 1,\ A = a)\mathbf{P}(X_1 = 1 \,|\, A = a). \end{aligned}$$

We have

$$\mathbf{P}(Y_1 = 1 \,|\, X_1 = 1,\, A = a) = \frac{a-1}{2m-1}, \qquad \mathbf{P}(X_1 = 1 \,|\, A = a) = \frac{a}{2m}.$$

Thus

$$\mathbf{E}[S \,|\, A = a] = m\,\frac{a-1}{2m-1} \cdot \frac{a}{2m} = \frac{a(a-1)}{2(2m-1)}.$$

Note that $\mathbf{E}[S \,|\, A = a]$ does not depend on $p$.

**Solution to Problem 2.38.** (a) Let $X$ be the number of red lights that Alice encounters. The PMF of $X$ is binomial with $n = 4$ and $p = 1/2$. The mean and the variance of $X$ are $\mathbf{E}[X] = np = 2$ and $\text{var}(X) = np(1-p) = 4 \cdot (1/2) \cdot (1/2) = 1$.

(b) The variance of Alice's commuting time is the same as the variance of the time by which Alice is delayed by the red lights. This is equal to the variance of $2X$, which is $4\text{var}(X) = 4$.

**Solution to Problem 2.39.** Let $X_i$ be the number of eggs Harry eats on day $i$. Then, the $X_i$ are independent random variables, uniformly distributed over the set $\{1, \ldots, 6\}$. We have $X = \sum_{i=1}^{10} X_i$, and

$$\mathbf{E}[X] = \mathbf{E}\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} \mathbf{E}[X_i] = 35.$$

Similarly, we have

$$\text{var}(X) = \text{var}\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} \text{var}(X_i),$$

since the $X_i$ are independent. Using the formula of Example 2.6, we have

$$\text{var}(X_i) = \frac{(6-1)(6-1+2)}{12} \approx 2.9167,$$

so that $\text{var}(X) \approx 29.167$.

**Solution to Problem 2.40.** Associate a success with a paper that receives a grade that has not been received before. Let $X_i$ be the number of papers between the $i$th success and the $(i+1)$st success. Then we have $X = 1 + \sum_{i=1}^{5} X_i$ and hence

$$\mathbf{E}[X] = 1 + \sum_{i=1}^{5} \mathbf{E}[X_i].$$

After receiving $i-1$ different grades so far ($i-1$ successes), each subsequent paper has probability $(6-i)/6$ of receiving a grade that has not been received before. Therefore, the random variable $X_i$ is geometric with parameter $p_i = (6-i)/6$, so $\mathbf{E}[X_i] = 6/(6-i)$. It follows that

$$\mathbf{E}[X] = 1 + \sum_{i=1}^{5} \frac{6}{6-i} = 1 + 6\sum_{i=1}^{5} \frac{1}{i} = 14.7.$$

**Solution to Problem 2.41.** (a) The PMF of $X$ is the binomial PMF with parameters $p = 0.02$ and $n = 250$. The mean is $\mathbf{E}[X] = np = 250 \cdot 0.02 = 5$. The desired probability is

$$\mathbf{P}(X = 5) = \binom{250}{5}(0.02)^5(0.98)^{245} = 0.1773.$$

(b) The Poisson approximation has parameter $\lambda = np = 5$, so the probability in (a) is approximated by

$$e^{-\lambda}\frac{\lambda^5}{5!} = 0.1755.$$

(c) Let $Y$ be the amount of money you pay in traffic tickets during the year. Then

$$\mathbf{E}[Y] = \sum_{i=1}^{5} 50 \cdot \mathbf{E}[Y_i],$$

where $Y_i$ is the amount of money you pay on the $i$th day. The PMF of $Y_i$ is

$$\mathbf{P}(Y_i = y) = \begin{cases} 0.98, & \text{if } y = 0, \\ 0.01, & \text{if } y = 10, \\ 0.006, & \text{if } y = 20, \\ 0.004, & \text{if } y = 50. \end{cases}$$

The mean is

$$\mathbf{E}[Y_i] = 0.01 \cdot 10 + 0.006 \cdot 20 + 0.004 \cdot 50 = 0.42.$$

The variance is

$$\text{var}(Y_i) = \mathbf{E}[Y_i^2] - \big(\mathbf{E}[Y_i]\big)^2 = 0.01 \cdot (10)^2 + 0.006 \cdot (20)^2 + 0.004 \cdot (50)^2 - (0.42)^2 = 13.22.$$

The mean of $Y$ is

$$\mathbf{E}[Y] = 250 \cdot \mathbf{E}[Y_i] = 105,$$

and using the independence of the random variables $Y_i$, the variance of $Y$ is

$$\text{var}(Y) = 250 \cdot \text{var}(Y_i) = 3,305.$$

(d) The variance of the sample mean is

$$\frac{p(1-p)}{250}$$

so assuming that $|p - \hat{p}|$ is within 5 times the standard deviation, the possible values of $p$ are those that satisfy $p \in [0, 1]$ and

$$(p - 0.02)^2 \le \frac{25p(1-p)}{250}.$$

This is a quadratic inequality that can be solved for the interval of values of $p$. After some calculation, the inequality can be written as $275p^2 - 35p + 0.1 \leq 0$, which holds if and only if $p \in [0.0025, 0.1245]$.

**Solution to Problem 2.42.** (a) Noting that

$$\mathbf{P}(X_i = 1) = \frac{\text{Area}(S)}{\text{Area}\big([0,1] \times [0,1]\big)} = \text{Area}(S),$$

we obtain

$$\mathbf{E}[S_n] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] = \mathbf{E}[X_i] = \text{Area}(S),$$

and

$$\text{var}(S_n) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(X_i) = \frac{1}{n}\text{var}(X_i) = \frac{1}{n}\big(1 - \text{Area}(S)\big)\text{Area}(S),$$

which tends to zero as $n$ tends to infinity.

(b) We have

$$S_n = \frac{n-1}{n}S_{n-1} + \frac{1}{n}X_n.$$

(c) We can generate $S_{10000}$ (up to a certain precision) as follows :

1. Initialize $S$ to zero.
2. For $i = 1$ to 10000
3. Randomly select two real numbers $a$ and $b$ (up to a certain precision) independently and uniformly from the interval $[0, 1]$.
4. If $(a - 0.5)^2 + (b - 0.5)^2 < 0.25$, set $x$ to 1 else set $x$ to 0.
5. Set $S := (i - 1)S/i + x/i$ .
6. Return $S$.

By running the above algorithm, a value of $S_{10000}$ equal to 0.7783 was obtained (the exact number depends on the random number generator). We know from part (a) that the variance of $S_n$ tends to zero as $n$ tends to infinity, so the obtained value of $S_{10000}$ is an approximation of $\mathbf{E}[S_{10000}]$. But $\mathbf{E}[S_{10000}] = \text{Area}(S) = \pi/4$, this leads us to the following approximation of $\pi$:

$$4 \cdot 0.7783 = 3.1132.$$

(d) We only need to modify the test done at step 4. We have to test whether or not $0 \leq \cos \pi a + \sin \pi b \leq 1$. The obtained approximation of the area was 0.3755.