Solutions to Chapter 1 of Sheldon M. Ross' *Probability Models For Computer Science*

1. Let A and G denote the events that Al and George hit the targets respectively. Then, $P(A) = p_1$ and $P(G) = p_2$.

   (a)
$$P(A, G|A \cup G) = \frac{P(A, G)}{P(A \cup G)} = \frac{P(A)P(G)}{P(A) + P(G) - P(A, G)} = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2}$$

   (b)
$$P(A|A \cup G) = \frac{P(A)}{P(A \cup G)} = \frac{P(A)}{P(A) + P(G) - P(A, G)} = \frac{p_1}{p_1 + p_2 - p_1 p_2}$$

2.
$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - E[X])^2] \\
&= E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - E[X]^2 \\
\mathrm{Var}(aX + b) &= E[(aX + b)^2] - E[aX + b]^2 \\
&= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
&= a^2E[X^2] + 2abE[X] + b^2 - a^2E[X]^2 - 2abE[X] - b^2 \\
&= a^2(E[X^2] - E[X]^2) \\
&= a^2\,\mathrm{Var}(X)
\end{aligned}$$

3. Let X_i, $1 \le i \le r$, indicate that the ith urn does not contain any balls.
   The expected value of X_i is $P(X_i = 1)$, the probability that all r balls entered other urns;
$$E[X_i] = P(X_i = 1) = \left(\frac{r-1}{r}\right)^n.$$
   Additionally, since $X_i^2 = X_i$ because X_i is an indicator variable,
$$\mathrm{Var}(X_i) = E[X_i] - E[X_i]^2 = \left(\frac{r-1}{r}\right)^n - \left(\frac{r-1}{r}\right)^{2n}.$$
   Finally, for $1 \le j \le r$ and $j \ne i$,
$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= E[X_iX_j] - E[X_i]E[X_j] \\
&= P(X_i = 1, X_j = 1) - E[X_i]E[X_j] \\
&= \left(\frac{r-2}{r}\right)^n - \left(\frac{r-1}{r}\right)^{2n}.
\end{aligned}$$

Since X is the total number of urns that do not contain balls, $\sum_{i=1}^{r} X_i$, its expectation is

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^{r} X_i\right] \\ &= \sum_{i=1}^{r} E[X_i] \\ &= r\left(\frac{r-1}{r}\right)^n \end{aligned}$$

and its variance is

$$\begin{aligned} \operatorname{Var}(X) &= \operatorname{Var}\left(\sum_{i=1}^{r} X_i\right) \\ &= \sum_{i=1}^{r} \operatorname{Var}(X_i) + \sum_{i=1}^{r}\sum_{j\neq i} \operatorname{Cov}(X_i, X_j) \\ &= r\left(\left(\frac{r-1}{r}\right)^n - \left(\frac{r-1}{r}\right)^{2n}\right) + r(r-1)\left(\left(\frac{r-2}{r}\right)^n - \left(\frac{r-1}{r}\right)^{2n}\right) \\ &= r\left(\frac{r-1}{r}\right)^n - r^2\left(\frac{r-1}{r}\right)^{2n} + r(r-1)\left(\frac{r-2}{r}\right)^n. \end{aligned}$$

4. Let C denote the number of distinct coupon types collected, and $D = n - C$ denote the number of distinct coupon types not collected. Let D_i, $1 \leq i \leq n$, indicate that the ith type of coupon is not collected, noting that $E[D_i] = P(D_i = 1) = (1 - p_i)^k$.

$$E[D] = E\left[\sum_{i=1}^{n} D_i\right] = \sum_{i=1}^{n}(1-p_i)^k$$

$$E[C] = E[n - D] = n - E[D] = n - \sum_{i=1}^{n}(1-p_i)^k$$

5. Define D_i, C and D as in the previous solution.

   Using the fact that D_i is an indicator variable,

$$\operatorname{Var}(D_i) = E[D_i] - E[D_i]^2 = (1-p_i)^k - (1-p_i)^{2k}$$

   and, for $1 \leq j \leq n$ and $i \neq j$,

$$\begin{aligned} \operatorname{Cov}(D_i, D_j) &= E[D_i D_j] - E[D_i]E[D_j] \\ &= P(D_i = 1, D_j = 1) - E[D_i]E[D_j] \\ &= (1 - p_i - p_j)^k - (1-p_i)^k(1-p_j)^k. \end{aligned}$$

The variance of C is

$$\begin{aligned}
\operatorname{Var}(C) &= \operatorname{Var}(n-D)\\
&= \operatorname{Var}(D)\\
&= \operatorname{Var}\left(\sum_{i=1}^{n} D_i\right)\\
&= \sum_{i=1}^{n} \operatorname{Var}(D_i) + \sum_{i=1}^{n}\sum_{j\neq i} \operatorname{Cov}(D_i, D_j)\\
&= \sum_{i=1}^{n}((1-p_i)^k - (1-p_i)^{2k}) + \sum_{i=1}^{n}\sum_{j\neq i}((1-p_i-p_j)^k - (1-p_i)^k(1-p_j)^k).
\end{aligned}$$

6. Assume that h_i, $i \geq 1$, is a good hashing function, so that the probability of any record, r_j, being placed by h_i in any given location is $\frac{1}{m}$.

   Let T_i denote the number of trials required to find an empty location for record r_i. T_i is geometrically distributed with parameter $\frac{m-(i-1)}{m}$.

   Additionally, let $X_i = T_i - 1$ denote the number of collisions placing record r_i, and $X = \sum_{i=1}^{n} X_i$ the total number of collisions placing all records. We assume that each X_i is independent.

$$\begin{aligned}
E[X] &= \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} E[T_i - 1]\\
&= \sum_{i=1}^{n}\left(\frac{m}{m-(i-1)} - 1\right)\\
&= \sum_{j=0}^{n-1} \frac{j}{m-j}
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}(X) &= \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \operatorname{Var}\left(\sum_{i=1}^{n}(T_i - 1)\right) = \sum_{i=1}^{n} \operatorname{Var}(T_i)\\
&= \sum_{i=1}^{n} \frac{1 - \frac{m-(i-1)}{m}}{\left(\frac{m-(i-1)}{m}\right)^2} = \sum_{j=0}^{n-1} \frac{\frac{j}{m}}{\left(\frac{m-j}{m}\right)^2}\\
&= m\sum_{j=0}^{n-1} \frac{j}{(m-j)^2}
\end{aligned}$$

7. Define $\phi^{(0)}(t) = \lambda(\lambda - t)^{-1}$, the moment generating function of X. By induction, we can show $\phi^{(n)}(t) = n!\lambda(\lambda - t)^{-(n+1)}$, $n \in \mathbb{N}$.

   For $n = 0$,

$$\phi^{(0)}(t) = \lambda(\lambda - t)^{-1} = 0!\lambda(\lambda - t)^{-(0+1)}.$$

   Suppose $\phi^{(k)}(t) = k!\lambda(\lambda - t)^{-(k+1)}$. Then, by differentiation,

$$\begin{aligned}
\phi^{(k+1)}(t) &= -(-1)(k+1)k!\lambda(\lambda - t)^{-(k+1)-1}\\
&= (k+1)!\lambda(\lambda - t)^{-((k+1)+1)}.
\end{aligned}$$

Hence,

$$E[X^4] = \phi^{(4)}(0) = 4!\lambda(\lambda - 0)^{-(4+1)} = \frac{24}{\lambda^4}.$$

8. Define $\phi^{(0)}(t) = e^{t^2/2}$, the moment generating function of X. By induction, we can show $\phi^{(i)}(t) = t\phi^{(i-1)}(t) + (i-1)\phi^{(i-2)}(t)$, $i \in \mathbb{Z}^+$.

   For $n = 1$,

   $$\phi^{(1)} = \frac{d}{dt}e^{t^2/2} = te^{t^2/2} = t\phi^{(0)}(t) + (0)\phi^{(-1)}(t).$$

   Suppose $\phi^{(k)} = t\phi^{(k-1)}(t) + (k-1)\phi^{(k-2)}(t)$. Then, by differentiation,

   $$\begin{aligned}\phi^{(k+1)} &= t\phi^{(k)}(t) + \phi^{(k-1)}(t) + (k-1)\phi^{(k-1)}(t)\\ &= t\phi^{(k)}(t) + k\phi^{(k-1)}(t)\\ &= t\phi^{((k+1)-1)}(t) + ((k+1)-1)\phi^{((k+1)-2)}(t).\end{aligned}$$

   Hence,

   $$E[X^4] = \phi^{(4)}(0) = 3 \cdot \phi^{(2)} = 3 \cdot 1 \cdot \phi^{(0)}(0) = 3 \cdot 1 \cdot 1 = 3.$$

9. Let X_i, for $1 \leq i \leq r$, denote the number of trials for the ith success. X_i is geometrically distributed with parameter p, and $X = \sum_{i=1}^{r} X_i$.

   (a) If $r = 0$, X is certainly 0. Otherwise, the probability of r successes in exactly i trials is the probability that there are $r-1$ successes in the first $i-1$ trials and the ith trial is a success. For $0 < r \leq i$, the probability of $r-1$ successes in $i-1$ trials is $\binom{i-1}{r-1}p^{r-1}(1-p)^{(i-1)-(r-1)}$.

   $$P(X = i) = \begin{cases} 1 & \text{if } 0 = r = i \\ \binom{i-1}{r-1}p^r(1-p)^{i-r} & \text{for } 0 < r \leq i \\ 0 & \text{otherwise} \end{cases}$$

   (b)

   $$E[X] = \sum_{i=1}^{r} E[X_i] = \frac{r}{p}$$

   (c)

   $$\text{Var}(X) = \sum_{i=1}^{r} \text{Var}(X_i) = \frac{r(1-p)}{p^2}$$

10. Let T denote the number of days until the miner finds freedom and N denote the total number of doors the miner trials. N is geometrically distributed with parameter $\frac{1}{3}$, so $E[N] = 3$ and $\text{Var}(N) = 6$. Let D denote the miner's first choice of door, with $D = i$, $1 \leq i \leq 3$, if the ith door is chosen.

    Conditioning on D,

    $$\begin{aligned}E[T] = E[E[T|D]] &= \sum_{d} E[T|D = d]P(D = d)\\ &= \frac{1}{3}\left(E[T|D = 1] + E[T|D = 2] + E[T|D = 3]\right)\\ &= \frac{1}{3}\left((4 + E[T]) + (7 + E[T]) + (3)\right).\end{aligned}$$

Therefore, $3E[T] = 14 + 2E[T]$. $E[T]$ is finite, since $E[T] \le E[7N] = 21$. Hence, $E[T] = 14$. Again conditioning on D,

$$\begin{aligned} E[T^2] &= E[E[T^2|D]] \\ &= \frac{1}{3}\left(E[T^2|D=1] + E[T^2|D=2] + E[T^2|D=3]\right) \\ &= \frac{1}{3}\left(E[(T+4)^2] + E[(T+7)^2] + 3^2\right) \\ &= \frac{1}{3}\left(2E[T^2] + 22E[T] + 74\right) \\ 3E[T^2] &= 2E[T^2] + 382. \end{aligned}$$

$E[T^2]$ is finite, since

$$E[T^2] \le E[(7N)^2] = 49E[N^2] = 49(\mathrm{Var}(N) + E[N]^2) = 735.$$

Hence, $E[T^2] = 382$ and the variance of T is $E[T^2] - E[T]^2 = 186$.

11. Let N_i denote the number of trials needed to obtain $i \in \mathbb{N}$ consecutive successes.

   (a) Let F denote the number of trials until the first failure. Then, conditioning on F,

$$\begin{aligned} E[N_k] = E[E[N_k|F]] &= \sum_f E[N_k|F=f] \cdot P(F=f) \\ &= \sum_{1\le f\le k} E[N_k|F=f] \cdot P(F=f) \\ &\qquad + \sum_{f>k} E[N_k|F=f] \cdot P(F=f) \\ &= \sum_{1\le f\le k} (f + E[N_k]) \cdot p^{f-1}(1-p) \\ &\qquad + \sum_{f>k} k \cdot p^{f-1}(1-p) \\ &= (1-p)\left(\sum_{1\le f\le k} f p^{f-1}\right) + E[N_k](1-p)\left(\sum_{1\le f\le k} p^{f-1}\right) \\ &\qquad + k(1-p)\left(\sum_{f>k} p^{f-1}\right) \\ &= (1-p)\left(\frac{1-(k+1)p^k + kp^{k+1}}{(1-p)^2}\right) + E[N_k](1-p)\left(\frac{1-p^k}{1-p}\right) \\ &\qquad + k(1-p)\left(\frac{p^k}{1-p}\right) \\ &= \frac{1-(k+1)p^k + kp^{k+1}}{1-p} + E[N_k](1-p^k) + kp^k \\ &= \frac{1-p^k}{1-p} + E[N_k](1-p^k). \end{aligned}$$

Hence, if $E[N_k]$ is finite,

$$E[N_k] = \frac{1-p^k}{p^k(1-p)}.$$

(b) Suppose the time that it takes for the first $k-1$ consecutive successes is i. Then, if the $(i+1)$th trial is a success (with probability p), the expected time taken for k consecutive successes is $i+1$. Otherwise, if the $i+1$th trial is a failure (with probability $1-p$), the time it taken for k consecutive successes is the expected time taken for k consecutive successes starting again plus the $i+1$ time already spent. Hence,

$$\begin{aligned} E[N_k|N_{k-1}=i] &= p(i+1) + (1-p)(i+1+E[N_k]) \\ &= i+1+E[N_k] - pE[N_k]. \end{aligned}$$

Conditioning on N_{k-1},

$$\begin{aligned} E[N_k] = E[E[N_k|N_{k-1}]] &= \sum_i E[N_k|N_{k-1}=i] \cdot P(N_{k-1}=i) \\ &= \sum_i (i+1+E[N_k] - pE[N_k]) \cdot P(N_{k-1}=i) \\ &= \left(\sum_i iP(N_{k-1}=i)\right) + (1+E[N_k]-pE[N_k]) \sum_i P(N_{k-1}=i) \\ &= E[N_{k-1}] + (1+E[N_k]-pE[N_k]). \end{aligned}$$

Hence, $pE[N_k] = E[N_{k-1}] + 1$. Solving the recurrence relation

$$E[N_k] = \frac{E[N_{k-1}]}{p} + \frac{1}{p},$$

with base case

$$E[N_0] = 0,$$

it can been seen that, for $r \geq 0$,

$$E[N_r] = \sum_{i=1}^{r} \frac{1}{p^i} = \frac{1-p^r}{p^r(1-p)}.$$

12. For $1 \leq k \leq n$, $S \in \{A, B\}$, define indicators

$$S_k = \begin{cases} 1 & \text{if } k \in S \\ 0 & \text{otherwise} \end{cases}.$$

Since both A and B are likely to be any of the subsets of $\{1, ..., n\}$ with uniform probability, $P(S_k = 1) = P(S_k = 0) = \frac{1}{2}$, as exactly half of all subsets contain k. Then, given also that A_k and B_k are independent, by the independence of the values of A and B,

$$\begin{aligned} P(A_k < B_k) &= P(A_k = 0, B_k = 1) = \frac{1}{4}, \\ P(A_k > B_k) &= P(A_k = 1, B_k = 0) = \frac{1}{4}, \\ P(A_k = B_k) &= 1 - P(A_k < B_k) - P(A_k > B_k) = \frac{1}{2}. \end{aligned}$$

(a) Let E_k denote the event $A_k < B_k$. Define, for $r \in \mathbb{N}$,

$$M(r) \triangleq P(A_1 \leq B_1, ..., A_r \leq B_r, \bigcup_{k=1}^{r} E_k).$$

Conditioning on the ordering of A_i and B_i, for $i > 0$,

$$\begin{aligned}
M(i) = P(A_i < B_i) \cdot P(A_1 \leq B_1, ..., A_i \leq B_i, \bigcup_{k=1}^{i} E_k | A_i < B_i) \\
+ P(A_i = B_i) \cdot P(A_1 \leq B_1, ..., A_i \leq B_i, \bigcup_{k=1}^{i} E_k | A_i = B_i) \\
+ P(A_i > B_i) \cdot P(A_1 \leq B_1, ..., A_i \leq B_i, \bigcup_{k=1}^{i} E_k | A_i > B_i) \\
= \frac{1}{4} \cdot P(A_1 \leq B_1, ..., A_{i-1} \leq B_{i-1}) \\
+ \frac{1}{2} \cdot P(A_1 \leq B_1, ..., A_{i-1} \leq B_{i-1}, \bigcup_{k=1}^{i-1} E_k) \\
+ \frac{1}{4} \cdot 0 \\
= \frac{1}{4}\left(\frac{3}{4}\right)^{i-1} + \frac{1}{2} M(i-1).
\end{aligned}$$

The above recurrence, with base case

$$M(0) = P(\emptyset) = 0,$$

provides

$$M(r) = \sum_{i=1}^{r} \frac{3^{i-1} 2^{i-r}}{4^i} = \frac{3^r - 2^r}{4^r}.$$

Hence, $P(A \subset B) = M(n) = 4^{-n}(3^n - 2^n)$.

(b) Let C_i indicate that both A and B contain i; that is, $A_i = B_i = 1$. Then,

$$P(C_i = 0) = 1 - P(C_i = 1) = 1 - P(A_i = 1, B_i = 1) = \frac{3}{4}.$$

The probability that A and B are disjoint is $P(C_1 = 0, ..., C_n = 0) = P(C_1 = 0) \cdot ... \cdot P(C_n = 0) = \left(\frac{3}{4}\right)^n$.

13. (a)

$$\begin{aligned}
\mathrm{Var}(X|Y = y) &= E[(X - E[X|Y = y])^2 \, |Y = y] \\
&= E[X^2 - 2XE[X|Y = y] + E[X|Y = y]^2 | Y = y] \\
&= E[X^2|Y = y] - 2E[X|Y = y]E[X|Y = y] + E[X|Y = y]^2 \\
&= E[X^2|Y = y] - E[X|Y = y]^2
\end{aligned}$$

(b)

$$
\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\
&= E[X^2] - E[E[X|Y]^2] \\
&\qquad + E[E[X|Y]^2] - E[X]^2 \\
&= \sum_y E[X^2|Y=y]P(Y=y) - \sum_y E[X|Y=y]^2 P(Y=y) \\
&\qquad + E[E[X|Y]^2] - 2E[X]^2 + E[X]^2 \\
&= \sum_y (E[X^2|Y=y] - E[X|Y=y]^2)P(Y=y) \\
&\qquad + E[E[X|Y]^2 - 2E[X]E[X|Y] + E[X]^2] \\
&= \sum_y \mathrm{Var}(X|Y=y)P(Y=y) \\
&\qquad + E[(E[X|Y] - E[X])^2] \\
&= E[\mathrm{Var}(X|Y)] \\
&\qquad + E[(E[X|Y] - E[E[X|Y]])^2] \\
&= E[\mathrm{Var}(X|Y)] \\
&\qquad + \mathrm{Var}(E[X|Y])
\end{aligned}
$$

(c)

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(E[X|N]) + E[\mathrm{Var}(X|N)] \\
&= \mathrm{Var}(E[X|N]) + \sum_n \mathrm{Var}(X|N=n)P(N=n)) \\
&= \mathrm{Var}(E[X|N]) + \sum_n n\,\mathrm{Var}(X_1)P(N=n) \\
&= \mathrm{Var}(E[X|N]) + E[N]\,\mathrm{Var}(X_1) \\
&= E[E[X|N]^2] - E[E[X|N]]^2 + E[N]\,\mathrm{Var}(X_1) \\
&= E[E[X|N]^2] - \left(\sum_n E[X|N=n]P(N=n)\right)^2 + E[N]\,\mathrm{Var}(X_1) \\
&= E[E[X|N]^2] - \left(\sum_n nE[X_1]P(N=n)\right)^2 + E[N]\,\mathrm{Var}(X_1) \\
&= E[E[X|N]^2] - E[X_1]^2E[N]^2 + E[N]\,\mathrm{Var}(X_1) \\
&= \left(\sum_n E[X|N=n]^2P(N=n)\right) - E[X_1]^2E[N]^2 + E[N]\,\mathrm{Var}(X_1) \\
&= \left(\sum_n (nE[X_1])^2P(N=n)\right) - E[X_1]^2E[N]^2 + E[N]\,\mathrm{Var}(X_1) \\
&= E[X_1]^2E[N^2] - E[X_1]^2E[N]^2 + E[N]\,\mathrm{Var}(X_1) \\
&= E[X_1]^2\,\mathrm{Var}(N) + E[N]\,\mathrm{Var}(X_1)
\end{aligned}
$$

14. Let N denote the total number of rolls, M the number of rolls made after the initial roll,

and R the value of the first roll. Let $X = \{4, 5, 6, 8, 9, 10\}$, the point numbers.

Additionally, let r denote the probability mass function of the value of a roll, the sum of two independent discrete random variables uniformly distributed in $[1, 6]$. Then, for $2 \le i \le 7$,

$$r(i) = r(14 - i) = \frac{i-1}{36}.$$

(a) Note that the conditional probability distribution $M|R = i$ is geometrically distributed when $i \in X = \{4, 5, 6, 8, 9, 10\}$ with its parameter given by the probability of rolling either 7 or i, and otherwise always 0. Hence, given that the expectation of a geometric random variable with parameter $p > 0$ is p^{-1},

$$E[M|R = i] = \begin{cases} \frac{1}{r(7)+r(i)} & \text{if } i \in X \\ 0 & \text{otherwise} \end{cases}.$$

Conditioning on the value of the first roll,

$$\begin{aligned} E[M] = E[E[M|R]] &= \sum_r r(i)E[M|R = i] \\ &= \frac{r(4)}{r(7)+r(4)} + \frac{r(5)}{r(7)+r(5)} + \frac{r(6)}{r(7)+r(6)} \\ &\quad + \frac{r(8)}{r(7)+r(8)} + \frac{r(9)}{r(7)+r(9)} + \frac{r(10)}{r(7)+r(10)} \\ &= \frac{392}{165}. \end{aligned}$$

Finally, using $P(N = 1) = r(2) + r(3) + r(7) + r(11) + r(12) = \frac{12}{36}$ and $N = 1 + M$,

$$\begin{aligned} E[N|N > 1] &= \sum_i i \cdot P(N = i|N > 1) \\ &= \frac{\sum_i i \cdot P(N = i, N > 1)}{P(N > 1)} \\ &= \frac{(\sum_i i \cdot P(N = i)) - 1 \cdot P(N = 1)}{P(N > 1)} \\ &= \frac{E[N] - P(N = 1)}{1 - P(N = 1)} \\ &= \frac{1 + E[M] - P(N = 1)}{1 - P(N = 1)} \\ &= \frac{1 + \frac{392}{165} - \frac{12}{36}}{1 - \frac{12}{36}} \\ &= \frac{251}{55}. \end{aligned}$$

(b) Let W denote the event that the player wins.
Recalling $X = \{4, 5, 6, 8, 9, 10\}$, it should be plain to see that, for $i \in \mathbb{N}$ and $j \in \mathbb{N}$,

$$P(M = i, W|R = j) = \begin{cases} 1 & \text{if } i = 0 \text{ and } j \in \{7, 11\} \\ 0 & \text{if } i = 0 \text{ and } j \notin \{7, 11\} \\ r(j)\left(1 - r(7) - r(j)\right)^{i-1} & \text{if } i > 0 \text{ and } j \in X \\ 0 & \text{if } i > 0 \text{ and } j \notin X \end{cases}.$$

Then, using $P(M = i, W) = \sum_j P(M = i, W|R = j) \cdot r(j)$,

$$P(M = i, W) = \begin{cases} r(7) + r(11) & \text{if } i = 0 \\ \sum_{j \in X} r(j)^2 \left(1 - r(7) - r(j)\right)^{i-1} & \text{if } i > 0 \end{cases}.$$

Hence, using $P(W) = \sum_i P(M = i, W)$,

$$\begin{aligned} P(W) &= P(M = 0, W) + \sum_{i>0} P(M = i, W) \\ &= r(7) + r(11) + \sum_{i>0} \sum_{j \in X} r(j)^2 (1 - r(7) - r(j))^{i-1} \\ &= r(7) + r(11) + \sum_{j \in X} r(j)^2 \sum_{i>0} (1 - r(7) - r(j))^{i-1} \\ &= r(7) + r(11) + \sum_{j \in X} \frac{r(j)^2}{r(7) + r(j)} \\ &= \frac{244}{495}. \end{aligned}$$

Additionally, using $E[M|W] = \sum_i i \cdot P(M = i|W)$,

$$\begin{aligned} E[M|W] &= \sum_i i \cdot \frac{P(M = i, W)}{P(W)} \\ &= \frac{1}{P(W)} \sum_{i>0} i \cdot \sum_{j \in X} r(j)^2 (1 - r(7) - r(j))^{i-1} \\ &= \frac{1}{P(W)} \sum_{j \in X} r(j)^2 \sum_{i>0} i \cdot (1 - r(7) - r(j))^{i-1} \\ &= \frac{1}{P(W)} \sum_{j \in X} r(j)^2 \frac{1}{(r(7) + r(j))^2} \\ &= \frac{1}{P(W)} \sum_{j \in X} \left(\frac{r(j)}{r(7) + r(j)} \right)^2 \\ &= \frac{1}{P(W)} \frac{26012}{27225}. \end{aligned}$$

Finally, using $N = 1 + M$, the expected number of rolls in a game of craps given that the player wins, but not on the first roll is

$$
\begin{aligned}
E[N|N>1,W] &= E[M+1|M>0,W] \\
&= 1+E[M|M>0,W] \\
&= 1+\sum_i i\cdot P(M=i|M>0,W) \\
&= 1+\frac{\sum_i i\cdot P(M=i,M>0|W)}{P(M>0|W)} \\
&= 1+\frac{E[M|W]}{P(M>0|W)} \\
&= 1+\frac{E[M|W]}{1-\frac{P(M=0,W)}{P(W)}} \\
&= 1+\frac{P(W)\cdot E[M|W]}{P(W)-P(M=0,W)} \\
&= 1+\frac{\frac{26012}{27225}}{\frac{244}{495}-(r(7)+r(11))} \\
&= \frac{16691}{3685}.
\end{aligned}
$$

15. Using $\frac{d}{dt}P(X>t)=\frac{d}{dt}(1-P(X\le t))=-\frac{d}{dt}F_X(t)=-f_X(t)$ and integration by parts,

$$
\begin{aligned}
E[X] &= \int_0^\infty t\cdot f_X(t)\,dt \\
&= -tP(X>t)\Big|_0^\infty - \int_0^\infty -P(X>t)\,dt \\
&= \int_0^\infty P(X>t)\,dt.
\end{aligned}
$$

16. Assume there are a positive number of tokens, $n>0$. Let T_i, $1\le i\le n$, denote the clockwise distance from P to the ith nearest token when moving clockwise.

    (a) Let D_i, $1\le i\le n$, denote the distance from P to the ith token, which is the minimum over travelling either clockwise or counterclockwise from P to the token.
    Each token is uniformly located on a rim of circumference 1, so $0\le D_i\le\frac{1}{2}$. Then, the probability density that D_i is at a distance $0\le x\le\frac{1}{2}$ from P is $1/\frac{1}{2}=2$, and the distribution function is $P(D_i\le x)=\int_0^x 2\,dt=2x$.

Using $E[D] = \int_0^\infty P(D > t)\,dt$, as established in exercise 15,

$$\begin{aligned}
E[D] &= \int_0^{\frac{1}{2}} P(D > t)\,dt \\
&= \int_0^{\frac{1}{2}} P(D_1 > t, ..., D_n > t)\,dt \\
&= \int_0^{\frac{1}{2}} \prod_{i=1}^{n} P(D_i > t)\,dt, \text{ by independence of } D_i \\
&= \int_0^{\frac{1}{2}} \prod_{i=1}^{n} (1 - P(D_i \le t))\ dt \\
&= \int_0^{\frac{1}{2}} (1-2t)^n\,dt \\
&= \frac{1}{2(n+1)}.
\end{aligned}$$

(b) Suppose that the nearest token is in the clockwise direction from P. Then $0 < T_1 < \frac{1}{2}$ and, for $j > 1$, $T_1 < T_j < 1 - T_1$.
Using $E[T_n|T_1 = t] = \int_0^\infty P(T_n > u|T_1 = t)\,du$, the expected value of T_n, the distance of the furthest token moving clockwise from P, given T_1, the nearest token, is

$$\begin{aligned}
E[T_n|T_1 = t] &= \int_0^{1-t} P(T_n > u|T_1 = t)\,du \\
&= \int_0^{t} P(T_n > u|T_1 = t)\,du + \int_t^{1-t} P(T_n > u|T_1 = t)\,du \\
&= \int_0^{t} 1\,du + \int_t^{1-t} 1 - P(T_n \le u|T_1 = t)\,du \\
&= t + \int_t^{1-t} 1 - \left(\frac{u-t}{1-2t}\right)^{n-1} du \\
&= t + u\Big|_t^{1-t} - \frac{1-2t}{n}\left(\frac{u-t}{1-2t}\right)^{n}\Bigg|_t^{1-t} \\
&= t + 1 - 2t - \frac{1-2t}{n} \\
&= 1 - \frac{n-2}{n}t - \frac{1}{n}.
\end{aligned}$$

Let f denote the probability density function of T_1, the nearest token. Conditioning on T_1, the expected distance to the furthest token in the clockwise direction, T_n, is

$$\begin{aligned} E[T_n] &= \int_0^{\frac{1}{2}} f(t)E[T_n|T_1 = t]\,dt \\ &= \int_0^{\frac{1}{2}} f(t)\left(1 - \frac{n-2}{n}t - \frac{1}{n}\right)\,dt \\ &= \left(\int_0^{\frac{1}{2}} f(t)\,dt\right) - \frac{n-2}{n}\left(\int_0^{\frac{1}{2}} tf(t)\,dt\right) - \frac{1}{n}\left(\int_0^{\frac{1}{2}} f(t)\,dt\right) \\ &= 1 - \left(\frac{n-2}{n}E[T_1] + \frac{1}{n}\right). \end{aligned}$$

It is plain that the expected value of T_1 is equal to the expected value of D calculated earlier since, by the rim's symmetry and the uniform placement of tokens, the expected distance of the nearest token is independent of whether it is located clockwise or anticlockwise from P.

Given that $T_n = 1 - B$,

$$E[B] = \frac{n-2}{n}E[D] + \frac{1}{n} = \frac{3}{2(n+1)}.$$

(c) The value of X is the minimum distance required to pick up all the tokens travelling either clockwise, $1 - B$, or travelling anticlockwise, $1 - T_1$.

Suppose, again, that the nearest token is in the clockwise direction; $B > T_1$. The expected value of X is then

$$E[\min(1-B, 1-T_1)] = E[1 - \max(B, T_1)] = 1 - E[B] = 1 - \frac{3}{2(n+1)}.$$

A symmetric argument can be made supposing the nearest token is in the anticlockwise direction to yield the same expected value of X.

17. Suppose there are a positive number of elements, $n > 0$. For $1 \leq i \leq n$, let E_i denote a distinct element of S, and define $N_i \triangleq \min(k : E_i \notin S_k)$.

(a)

$$\begin{aligned} N &= \min(j : S_j = \emptyset) \\ &= \min(j : E_1 \notin S_j, ..., E_n \notin S_j) \\ &= \min(j : N_1 \leq j, ..., N_n \leq j) \\ &= \max(N_1, ..., N_n) \end{aligned}$$

Since each element, independent of others, has a fixed probability, p, of being removed at each stage, N_i is geometrically distributed with parameter p, and hence N is the maximum of a set of independent geometric random variables.

(b) Since each N_k is a geometric random variable with parameter p, the cumulative probability distribution is given by $P(N_k \leq i) = 1 - (1-p)^i$ for $i > 0$.

The cumulative probability distribution of N is given by

$$\begin{aligned} P(N \leq i) &= P(\max(N_1, ..., N_n) \leq i) \\ &= P(N_1 \leq i, ..., N_n \leq i) \\ &= \prod_{k=1}^{n} P(N_k \leq i), \text{ by independence} \\ &= \left(1 - (1-p)^i\right)^n . \end{aligned}$$

Hence, the probability mass function for $i > 0$ is

$$\begin{aligned} P(N = i) &= P(N \leq i) - P(N \leq i - 1) \\ &= \left(1 - (1-p)^i\right)^n - \left(1 - (1-p)^{i-1}\right)^n . \end{aligned}$$

18. (a) Let X_i, for $i \geq 1$, denote the time the miner spent journeying on selecting a door on the ith occasion.

(b) The probability mass function of X_i given N is

$$P(X_i = j | N = n) = \begin{cases} \frac{1}{2} & \text{if } i < n \text{ and } j \in \{4, 7\} \\ 1 & \text{if } i = n \text{ and } j = 3 \\ 0 & \text{otherwise} \end{cases} .$$

Hence,

$$\begin{aligned} E\left[\sum_{i=1}^{N} X_i \middle| N = n\right] &= \sum_{i=1}^{n} E[X_i | N = n] \\ &= E[X_n | N = n] + \sum_{i=1}^{n-1} E[X_i | N = n] \\ &= 3 + \sum_{i=1}^{n-1} \sum_{j} j P(X_i = j | N = n) \\ &= 3 + \sum_{i=1}^{n-1} (4 + 7) \frac{1}{2} \\ &= 3 + (n-1) \frac{11}{2}. \end{aligned}$$

(c) Recall from exercise 10 that N is geometrically distributed with parameter $\frac{1}{3}$, so $P(N = n) = \frac{1}{3}(\frac{2}{3})^{n-1}$ for $n > 0$.

The expected value of X is then

$$
\begin{aligned}
E[X] = E[E[X|N]] &= E\left[E\left[\sum_{i=1}^{N} X_i \middle| N\right]\right] \\
&= \sum_{n=1}^{\infty} P(N=n) E\left[\sum_{i=1}^{N} X_i \middle| N=n\right] \\
&= \sum_{n=1}^{\infty} \frac{1}{3}\left(\frac{2}{3}\right)^{n-1}\left(3+(n-1)\frac{11}{2}\right) \\
&= \sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^{n-1} + \frac{11}{6}\sum_{n=1}^{\infty}(n-1)\left(\frac{2}{3}\right)^{n-1} \\
&= \sum_{m=0}^{\infty} \left(\frac{2}{3}\right)^{m} + \frac{11}{6}\sum_{m=0}^{\infty} m\left(\frac{2}{3}\right)^{m} \\
&= 3 + \frac{11}{6}(6) \\
&= 14.
\end{aligned}
$$

19. Let S_1 and S_2 denote the time I spend at servers 1 and 2, respectively. Additionally, let C_A and C_B denote the times spent in service by customers A and B, respectively.

(a) Let f denote the probability density function of C_A. Then,

$$
\begin{aligned}
P_A = P(S_1 < C_A) &= \int_0^{\infty} f(c) P(S_1 < c)\, dc \\
&= \int_0^{\infty} 2e^{-2c}(1-e^{-c})\, dc \\
&= \left[-e^{-2c} + \frac{2}{3}e^{-3c}\right]_0^{\infty} \\
&= \frac{1}{3}.
\end{aligned}
$$

(b) Note that $P(S_1 \le C_B) = P(S_1 \le C_A) = P_A = \frac{1}{3}$, since C_A and C_B are identically distributed. Then,

$$
\begin{aligned}
P_B &= P(S_1 < C_A + C_B) \\
&= 1 - P(S_1 \ge C_A + C_B) \\
&= 1 - P(S_1 \ge C_A) P(S_1 \ge C_B) \\
&= 1 - [1 - P(S_1 \le C_A)][1 - P(S_1 \le C_B)] \\
&= 1 - \left(\frac{2}{3}\right)^2 \\
&= \frac{5}{9}.
\end{aligned}
$$

(c) The expected time I spend in the system is the total of the expected times I spend:

- with server 1, $E[S_1] = \frac{1}{1}$;

- waiting for customer A, $P_A \cdot E[C_A] = \frac{1}{3} \cdot \frac{1}{2}$;
- waiting for customer B, $P_B \cdot E[C_B] = \frac{5}{9} \cdot \frac{1}{2}$; and,
- with server 2, $E[S_2] = \frac{1}{2}$.

That is, $E[T] = \frac{35}{18}$ units of time.

20. Let $X_1, ..., X_n$ denote arbitrary independent exponential random variables with respective rates $\lambda_{X_1}, ..., \lambda_{X_n}$. The expected value of the minimum of these variables is

$$\begin{aligned} E[\min(X_1, ..., X_n)] &= \int_0^\infty P(\min(X_1, ..., X_n) > t)\, dt \\ &= \int_0^\infty P(X_1 > t, ..., X_n > t)\, dt \\ &= \int_0^\infty P(X_1 > t) \cdot ... \cdot P(X_n > t)\, dt \\ &= \int_0^\infty e^{-\lambda_{X_1} t} \cdot ... \cdot e^{-\lambda_{X_n} t}\, dt \\ &= \int_0^\infty e^{-(\lambda_{X_1} + ... + \lambda_{X_n})t}\, dt \\ &= -\frac{e^{-(\lambda_{X_1} + ... + \lambda_{X_n})t}}{\lambda_{X_1} + ... + \lambda_{X_n}}\Bigg|_0^\infty \\ &= \frac{1}{\lambda_{X_1} + ... + \lambda_{X_n}}. \end{aligned} \tag{1}$$

The probability that the minimum of these variables is X_1 is

$$\begin{aligned} P(\min(X_1, ..., X_n) = X_1) &= P(X_1 < X_2, ..., X_1 < X_n) \\ &= \int_0^\infty P(X_1 < X_2, ..., X_1 < X_n | X_1 = t) \cdot P(X_1 = t)\, dt \\ &= \int_0^\infty P(t < X_2, ..., t < X_n) \cdot \lambda_{X_1} e^{-\lambda_{X_1} t}\, dt \\ &= \int_0^\infty e^{-(\lambda_{X_2} + ... + \lambda_{X_n})t} \cdot \lambda_{X_1} e^{-\lambda_{X_1} t}\, dt \\ &= \lambda_{X_1} \int_0^\infty e^{-(\lambda_{X_1} + ... + \lambda_{X_n})t}\, dt \\ &= \frac{\lambda_{X_1}}{\lambda_{X_1} + ... + \lambda_{X_n}}. \end{aligned} \tag{2}$$

For $1 \leq i \leq 4$, let F_i denote the lifetime of fish i; let D_i denote the number of the ith dead fish; and let T_i denote the time of the ith death.

For convenience, let $\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$.

(a) The expected time until the first death occurs, consequent from (1), is

$$E[T_1] = E[\min(F_1, ..., F_4)] = \frac{1}{\lambda}.$$

(b) Consequent from (2), the probability that the first fish to die, D_1, is fish i is

$$P(D_1 = i) = P(\min(F_1, ..., F_4) = F_1) = \frac{\lambda_i}{\lambda}.$$

Hence, the probability that the first first to die is either fish 1 or fish 2 is

$$P(D_1 = 1) + P(D_1 = 2) = \frac{\lambda_1 + \lambda_2}{\lambda}.$$

(c) By the memorylessness of exponential random variables, the expected time between the first and second death, $T_2 - T_1$, is the expected minimum of the lifetimes of the remaining fish; independent of the time already elapsed. That is, weighting for each fish the probability it is first to die,

$$E[T_2 - T_1] = \sum_i P(D_1 = i) \cdot E[\min_{j \neq i}(F_j)].$$

Hence, using a number of previous results,

$$\begin{aligned} E[T_2] &= E[T_1] + \sum_i P(D_1 = i) \cdot E[\min_{j \neq i}(F_j)] \\ &= \frac{1}{\lambda} + \sum_{i=1}^{4} \frac{\lambda_i}{\lambda} \cdot \frac{1}{\lambda - \lambda_i} \\ &= \frac{1}{\lambda} + \frac{1}{\lambda} \sum_{i=1}^{4} \frac{\lambda_i}{\lambda - \lambda_i}. \end{aligned}$$

(d) By the memorylessness of exponential random variables, the probability of the first fish dying before any others that are also alive is independent of the elapsed time.

$$\begin{aligned} P(D_2 = 1) &= \sum_{i=1}^{4} P(D_2 = 1, D_1 = i) \\ &= \sum_{i=2}^{4} P(D_2 = 1, D_1 = i), \text{ since } P(D_2 = 1, D_1 = 1) = 0 \\ &= \sum_{i=2}^{4} P(D_2 = 1 | D_1 = i) \cdot P(D_1 = i) \\ &= \sum_{i=2}^{4} P(\min_{j \neq i}(F_j) = F_1) \cdot \frac{\lambda_i}{\lambda} \\ &= \sum_{i=2}^{4} \frac{\lambda_1}{\lambda - \lambda_i} \cdot \frac{\lambda_i}{\lambda} \\ &= \frac{\lambda_1}{\lambda} \sum_{i=2}^{4} \frac{\lambda_i}{\lambda - \lambda_i} \end{aligned}$$

21. Let T_A, T_B and T_C denote the times A, B and C spend being served, respectively.

(a) A is the first to depart if $T_A < T_B$.

$$\begin{aligned}
P(T_A < T_B) &= \int_0^\infty P(t < T_B)\mu_1 e^{-\mu_1 t}\,dt \\
&= \mu_1 \int_0^\infty e^{-\mu_2 t} e^{-\mu_1 t}\,dt \\
&= \mu_1 \int_0^\infty e^{-(\mu_1+\mu_2)t}\,dt \\
&= -\frac{\mu_1}{\mu_1+\mu_2} e^{-(\mu_1+\mu_2)t}\Big|_0^\infty \\
&= \frac{\mu_1}{\mu_1+\mu_2}.
\end{aligned}$$

(b) A is the last to depart if $T_A > T_B + T_C$. Noting that T_C is identically distributed to T_B when $T_A > T_B$, since C will be served by server 2,

$$\begin{aligned}
P(T_A > T_B + T_C) &= P(T_A > T_B + T_C, T_A > T_B) \\
&= P(T_A > T_B + T_C | T_A > T_B) \cdot P(T_A > T_B) \\
&= P(T_A > T_C | T_A > T_B) \cdot P(T_A > T_B), \text{ by memorylessness} \\
&= P(T_A > T_B)^2 \\
&= \left(1 - \frac{\mu_1}{\mu_1+\mu_2}\right)^2 \\
&= \left(\frac{\mu_2}{\mu_1+\mu_2}\right)^2.
\end{aligned}$$

(c) The expected value of T_C given that $T_A < T_B$ is $\frac{1}{\mu_1}$, since C would be served by server 1. Otherwise, the expected value of T_C given that $T_A \geq T_B$ is $\frac{1}{\mu_2}$, since C would be served by server 2. Hence,

$$\begin{aligned}
E[T_C] &= E[T_C | T_A < T_B] \cdot P(T_A < T_B) \\
&\quad + E[T_C | T_A \geq T_B] \cdot P(T_A \geq T_B) \\
&= \frac{1}{\mu_1} \cdot \frac{\mu_1}{\mu_1+\mu_2} + \frac{1}{\mu_2} \cdot \left(1 - \frac{\mu_1}{\mu_1+\mu_2}\right) \\
&= \frac{2}{\mu_1+\mu_2}.
\end{aligned}$$

Using (1) from previous work,

$$E[\min(T_A, T_B)] = \frac{1}{\mu_1+\mu_2}.$$

Then, the expected time until C departs is

$$E[\min(T_A, T_B) + T_C] = E[\min(T_A, T_B)] + E[T_C] = \frac{3}{\mu_1+\mu_2}.$$

22. This solution assumes that an item remaining to processed, if any, is immediately allocated to any idle machine. Let M_1 and M_2 denote exponential random variables with rates λ_1 and λ_2, representing the processing rates of machine 1 and machine 2, respectively.

If there is only a single item to be processed, $n = 1$, the processing time is determined by whether it is allocated to the first or second machine, with expected times of $E[M_1] = \lambda_1^{-1}$ and $E[M_2] = \lambda_2^{-1}$, respectively.

Otherwise, suppose $n > 1$, and let T_i for $1 < i \leq n$, denote the total time spent processing with i or fewer remaining unprocessed items.

With two remaining items, the remaining processing time, T_2, is the maximum of the times taken by machine 1 and machine 2 respectively to process their current item. This is independent of the prior time already spent processing their current items by the memoryless property. Then,

$$\begin{aligned}
E[T_2] &= E[\max(M_1, M_2)] \\
&= \int_0^\infty P(\max(M_1, M_2) > t)\, dt \\
&= \int_0^\infty 1 - P(\max(M_1, M_2) \leq t)\, dt \\
&= \int_0^\infty 1 - P(M_1 \leq t, M_2 \leq t)\, dt \\
&= \int_0^\infty 1 - (1 - e^{-\lambda_1 t})(1 - e^{-\lambda_2 t})\, dt \\
&= \int_0^\infty e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}\, dt \\
&= \left[\frac{e^{-\lambda_1 t}}{-\lambda_1} + \frac{e^{-\lambda_2 t}}{-\lambda_2} - \frac{e^{-(\lambda_1+\lambda_2)t}}{\lambda_1+\lambda_2} \right]_0^\infty \\
&= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1+\lambda_2}.
\end{aligned}$$

The total time spent processing $2 < j \leq n$ or fewer items, T_j, is the minimum of the times taken by machine 1 and machine 2 respectively to process their current item in addition to the time taken to process the afterwards remaining $j - 1$ items. Again, this is independent of the prior time already spent by the machines processing their current items. That is, for $2 < j \leq n$,

$$\begin{aligned}
E[T_j] &= E[\min(M_1, M_2) + T_{j-1}] \\
&= \frac{1}{\lambda_1+\lambda_2} + E[T_{j-1}], \text{ using (1)} \\
&= (j-2)\frac{1}{\lambda_1+\lambda_2} + E[T_2] \\
&= \frac{j-2}{\lambda_1+\lambda_2} + \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1+\lambda_2} \right) \\
&= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{j-3}{\lambda_1+\lambda_2}.
\end{aligned}$$

Hence, the expected time it takes to process a set of $n > 1$ items, T_n, is

$$E[T_n] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{n-3}{\lambda_1+\lambda_2}.$$

23. (a) Using, for $t \geq 0$, $P(V > t) = 1 - P(V \leq t) = 1 - (1 - e^{-rt}) = e^{-rt}$ for an exponential random variable V with rate r,

$$
\begin{aligned}
E[\min(X,Y)|X > c] &= \int_0^\infty P(\min(X,Y) > t|X > c)\, dt \\
&= \int_0^\infty P(X > t, Y > t|X > c)\, dt \\
&= \int_0^\infty P(X > t|X > c)P(Y > t|X > c)\, dt \\
&= \int_0^\infty P(X > t|X > c)P(Y > t)\, dt \\
&= \int_0^c P(X > t|X > c)P(Y > t)\, dt \\
&\quad + \int_c^\infty P(X > t|X > c)P(Y > t)\, dt \\
&= \int_0^c P(Y > t)\, dt + \int_c^\infty P(X > t - c)P(Y > t)\, dt \\
&= \int_0^c e^{-\mu t}\, dt + \int_c^\infty e^{-\lambda(t-c)} e^{-\mu t}\, dt \\
&= \left.\frac{e^{-\mu t}}{-\mu}\right|_0^c + e^{c\lambda} \int_c^\infty e^{-(\lambda+\mu)t}\, dt \\
&= \frac{1 - e^{-\mu c}}{\mu} + \frac{e^{c\lambda}}{\lambda + \mu} \left.(-e^{-(\lambda+\mu)t})\right|_c^\infty \\
&= \frac{1 - e^{-\mu c}}{\mu} + \frac{e^{c\lambda}}{\lambda + \mu} e^{-(\lambda+\mu)c} \\
&= \frac{1 - e^{-\mu c}}{\mu} + \frac{e^{-\mu c}}{\lambda + \mu}.
\end{aligned}
$$

(b) Let $f(x, y)$ denote the joint probability density function of X and Y.
Note that

$$
\begin{aligned}
P(X > Y + c) &= P(X > Y)P(X > c), \text{ by memorylessness and independence} \\
&= P(\min(X,Y) = Y)[1 - P(X \leq c)] \\
&= \frac{\mu}{\lambda + \mu} e^{-\lambda c}, \text{ using (2).}
\end{aligned}
$$

The joint probability density function of X and Y, conditioned on $X > Y + c$, is

$$
\begin{aligned}
g(x, y) &= f(x, y) \frac{\mathbb{I}_{x > y + c}}{P(X > Y + c)} \\
&= \lambda e^{-\lambda x} \mu e^{-\mu y} \frac{\mathbb{I}_{x > y + c}}{\frac{\mu}{\lambda+\mu} e^{-\lambda c}} \\
&= \begin{cases} (\lambda + \mu)\, \lambda e^{\lambda c} e^{-\lambda x} e^{-\mu y} & \text{if } x > y + c \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

$c \geq 0$ so $Y + c \geq Y$. Hence, given $X > Y + c$, $\min(X, Y)$ is Y and its expected value is

$$
\begin{aligned}
E\left[\min(X,Y)|X > Y + c\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min(x,y)g(x,y)\,dx\,dy \\
&= \int_{0}^{\infty}\int_{y+c}^{\infty} y\,(\lambda+\mu)\,\lambda e^{\lambda c}e^{-\lambda x}e^{-\mu y}\,dx\,dy \\
&= \lambda(\lambda+\mu)e^{\lambda c}\int_{0}^{\infty} ye^{-\mu y}\int_{y+c}^{\infty} e^{-\lambda x}\,dx\,dy \\
&= \lambda(\lambda+\mu)e^{\lambda c}\int_{0}^{\infty} ye^{-\mu y}\frac{e^{-\lambda(y+c)}}{\lambda}\,dy \\
&= (\lambda+\mu)\int_{0}^{\infty} ye^{-(\lambda+\mu)y}\,dy \\
&= (\lambda+\mu)\frac{1}{(\lambda+\mu)^2} \\
&= \frac{1}{\lambda+\mu}.
\end{aligned}
$$

(c) $X - Y|X > Y$, the lifetime of X beyond Y given X lives beyond Y, has the same distribution as X by the memoryless property; the remaining lifetime is independent of the time already passed, $Y|X > Y$, which is $\min(X,Y)|X > Y$. The following is a restatement of this intuition.

It can be shown that the probability distribution function of $X - Y|X > Y$ is equal to that of $X - Y|Y, X > Y$ from

$$
\begin{aligned}
\overline{F}_{X-Y|Y=y,X>Y}(z) &= P(X - Y > z|Y = y, X > Y) \\
&= P(X > y + z|X > y) \\
&= P(X > z) \\
&= P(X > z + Y|X > Y) \\
&= P(X - Y > z|X > Y) \\
&= \overline{F}_{X-Y|X>Y}(z).
\end{aligned}
$$

Therefore, $f_{X-Y|X>Y} = f_{X-Y|Y,X>Y}$, so, conditional on $X > Y$, $X - Y$ and $Y = \min(X, Y)$ are independent.

(d)

$$
\begin{aligned}
\frac{1}{\lambda+\mu} &= E[\min(X,Y)|X > Y + c] \\
&= E[\min(X,Y)|X > Y + c, X > Y] \\
&= E[\min(X,Y)|X - Y > c, X > Y] \\
&= E[\min(X,Y)|X > Y]
\end{aligned}
$$

24. Let D_i, for $i > 0$, denote the value of the ith roll of the die. Each D_i is mutually independent and has a discrete uniform distribution over $[1, 6]$ with mean $\mu = (6 + 1)/2 = 7/2$ and variance $\sigma^2 = \frac{1}{6}\sum_{i=1}^{6}(i-\mu)^2 = 35/12$.

The probability that at least 80 rolls are necessary for the total sum of all rolls to exceed 300 is the probability that the first 79 rolls do not exceeed 300. Letting Z denote the standard normal distribution, that is

$$\begin{aligned} P\left(\sum_{i=1}^{79} D_i \leq 300\right) &= P\left(\frac{\sum_{i=1}^{79} D_i - 79\mu}{\sqrt{79\sigma^2}} \leq \frac{300 - 79\mu}{\sqrt{79\sigma^2}}\right) \\ &\approx P\left(Z < \frac{300 - 79 \cdot \frac{7}{2}}{\sqrt{79 \cdot \frac{35}{12}}}\right) \\ &\approx P\left(Z < 1.5481\right) \\ &\approx 0.9392. \end{aligned}$$

25. Let E_i denote the individual round off error contributed by rounding the ith number, and assume it is independent of other errors. Each E_i has mean $\mu = (0.5 - 0.5)/2 = 0$ and variance $\sigma^2 = 1 \int_{-0.5}^{0.5} (x - \mu)^2 \, dx = 1/12$.

    The probability that the resultant sum differs from the exact sum by more than 3 is the probability the absolute error exceeds 3. Letting Z denote the standard normal distribution, that is

$$\begin{aligned} P\left(\left|\sum_{i=1}^{50} E_i\right| > 3\right) &= P\left(\sum_{i=1}^{50} E_i > 3\right) + P\left(\sum_{i=1}^{50} E_i < -3\right) \\ &\approx P\left(Z > \frac{3 - 50\mu}{\sqrt{50\sigma^2}}\right) + P\left(Z < \frac{-3 - 50\mu}{\sqrt{50\sigma^2}}\right) \\ &\approx P\left(Z > 1.4697\right) + P\left(Z < -1.4697\right) \\ &\approx 0.1416. \end{aligned}$$

26. Let I_i, for $i > 0$, indicate the ith coin flip comes up heads. Then, the proportion of the first n coin flips that are heads is

$$\frac{I_1 + ... + I_n}{n}.$$

    By the strong law of large numbers, since each I_i is independent and identically distributed with mean p,

$$P\left(\lim_{n \to \infty} \frac{I_1 + ... + I_n}{n} = p\right) = 1.$$

    Therefore, the long run proportion of flips that land on heads is p.

27. Let T_i, $1 \leq i$, denote the time spent wandering after making the ith selection of door, and N the number of doors chosen by the miner until she reaches safety. Each T_i has a mean of $(4 + 7 + 3)/3 = 14/3$, and N is geometrically distributed with parameter $1/3$ so has a mean of 3.

    Then, by Wald's equation, the expected time until the miner is free is

$$E\left[\sum_{i=1}^{N} T_i\right] = E[N]\frac{14}{3} = 14.$$

28. Given r, let S_r denote the number of required steps, and call the step where the person began the *start step* and the step r to-the-right the *goal step*.

For the case $r = 0$, the start step is the goal step, so clearly $S_0 = 0 = E[S_0]$.

Consider the case $r = 1$. Any sequence of steps that ends when first having landed on the goal step must:

- for some $i \in \mathbb{N}$, contain i steps to the left, and $i+1$ steps to the right, since the overall offset is 1 to the right;
- for every proper prefix of the sequence, the number of steps to the right must never exceed the number of steps to the left, otherwise the goal step will already have been visited prior to the end; and
- end in a right step, since the goal step is one to-the-right of the start step.

First we consider the number of sequences of steps of length $2j + 1$ that satisfy these properties. Notice that the last element of the sequence is fixed by the last property, but the first $2j$ elements can be any interleaving of j left steps and j right steps that satisfies the second property. The number of such interleavings is given by the jth Catalan number,

$$\frac{1}{j+1}\binom{2j}{j}.$$

An alternative view to derive the same result is to observe that the last two properties are equivalent to the property: every non-empty suffix of the sequence has more steps to the right than steps to the left. That is, the sequence when reversed always has the number of right steps strictly "ahead" of the number of left steps. Applying Bertrand's ballot theorem, the proportion of all permutations of a sequence with j left steps and $j+1$ right steps, $\binom{2j+1}{j}$, that satisfies that property is $\frac{(j+1)-j}{(j+1)+j}$. Hence, the number of sequences is

$$\frac{1}{2j+1}\binom{2j+1}{j}.$$

From the definitions of binomial coefficients, it should be easy to confirm this is equal to the jth Catalan number;

$$\frac{1}{2j+1}\binom{2j+1}{j} = \frac{1}{j+1}\binom{2j}{j}.$$

Now we consider the probability of a particular sequence of length $2j + 1$ that satisfies these properties occurring. From the first property, that is $0.6^{i+1}0.4^i$.

Putting the previous results together, the probability the number of steps it takes is $2i+1$, for $i \in \mathbb{N}$, is the product of the number of sequences of steps of that length and the probability of an individual sequence occurring,

$$P(S_1 = 2i+1) = \frac{1}{2i+1}\binom{2i+1}{i}0.6^{i+1}0.4^i.$$

Then, the expected number of steps is given by

$$\begin{aligned} E[S_1] &= \sum_{i=0}^{\infty}(2i+1)P(S_1 = 2i+1) \\ &= \sum_{i=0}^{\infty}(2i+1)\frac{1}{2i+1}\binom{2i+1}{i}0.6^{i+1}0.4^{i} \\ &= 0.6\sum_{i=0}^{\infty}\binom{2i+1}{i}0.24^{i}. \end{aligned}$$

Using

$$\sum_{n=0}^{\infty}\binom{2n+s}{n}x^n = \frac{2^s}{\left(\sqrt{1-4x}+1\right)^s\sqrt{1-4x}}$$

it can be seen that

$$\sum_{i=0}^{\infty}\binom{2i+1}{i}0.24^{i} = \frac{2^1}{\left(\sqrt{1-4(0.24)}+1\right)^1\sqrt{1-4(0.24)}} = \frac{25}{3}$$

so the expected number of steps is $E[S_1] = 0.6 \cdot \frac{25}{3} = 5$.

This result can be extended to cases where $r > 1$. The number of steps to move two right, S_2, for example, is the steps to move one right, S_1, plus the steps from the new location to move one more right. Noticing that the expected number of steps to move one right is independent of the current location, $E[S_2] = E[S_1] + E[S_1] = 2E[S_1]$, and in general $E[S_n] = nE[S_1] = 5n$.

Another argument for the generalisation can follow from the observation

$$E[S_n] = 0.4(1 + E[S_{n+1}]) + 0.6(1 + E[S_{n-1}]).$$

Rearrangement and substitution provides a non-homogenous linear recurrence relation

$$E[S_n] - \frac{5}{2}E[S_{n-1}] + \frac{3}{2}E[S_{n-2}] = -\frac{5}{2}$$

with a general solution

$$E[S_n] = a\left(\frac{3}{2}\right)^n + b + 5n$$

for arbitrary parameters a and b. With initial conditions $E[S_0] = 0$ and $E[S_1] = 5$, it follows $E[S_n] = 5n$.

29. Let I_i, for $i > 0$, indicate whether the ith flip of the coin is heads. Each I_i has an expected value of p.

    We look to find $E[N]$ given that

    $$\sum_{i=1}^{N} I_i = r.$$

    From Wald's equation

    $$\sum_{i=1}^{N} I_i = pE[N]$$

so

$$E[N] = \frac{r}{p}.$$

30. Let F_i denote the outcome of the ith toss of the coin, with value 1 if it is heads and -1 otherwise. Since the coin is fair, the expected value of F_i is 0.

    We are given that

    $$\sum_{i=0}^{N} F_i = 1.$$

    Suppose for the sake of contradiction that the expected value of N is finite. Then, from Wald's equation,

    $$\sum_{i=0}^{N} F_i = 1 = 0E[N]$$

    which is nonsense.

    We can conclude the expected value of N is infinite.