1. Let E be an event and $\overline{E}$ be its complement. Let $p = \mathrm{P}(E)$.

   (a) What is the covariance of the indicator random variables of E and $\overline{E}$?

   (b) If $0 < p < 1$ can E and $\overline{E}$ be independent? Justify your answer.

   Solution:

   (a) Let X_E and $X_{\bar{E}}$ denote the indicator functions of E and $\bar{E}$, that is, $X_E = 1$ if E occurs and 0 otherwise. Then $\mathrm{E}[X_E] = p$, $\mathrm{E}[X_{\overline{E}}] = 1 - p$, and $X_E X_{\bar{E}} = 0$ as the two are disjoint. The covariance is

   $$\mathrm{Cov}(X_E, X_{\bar{E}}) = \mathrm{E}[X_E X_{\bar{E}}] - \mathrm{E}[X_E]\,\mathrm{E}[X_{\bar{E}}] = 0 - p(1-p) = -p(1-p)$$

   (b) Since the covariance is nonzero E and $\overline{E}$ cannot be independent. We can also argue directly. Since $E \cap \bar{E} = \emptyset$, $\mathrm{P}(E \cap \bar{E})$, while by assumption $\mathrm{P}(E) = p > 0$ and $\mathrm{P}(\bar{E}) = 1 - p > 0$. Therefore $\mathrm{P}(E \cap \bar{E}) \neq \mathrm{P}(E)\,\mathrm{P}(\bar{E})$.

2. A fair 3-sided die has face values `1`, `1`, and `2`. Let X be the number of times the die is rolled until both a `1` and a `2` appears. What is $\mathrm{E}[X]$?

   Solution: Let N be the number of rolls and F be the first roll. Then $\mathrm{E}[N|F=1] = 1 + \mathrm{Geometric}(1/3)$ and $\mathrm{E}[N|F=2] = 1 + \mathrm{Geometric}(2/3)$. By the total expectation theorem

   $$\mathrm{E}[N] = \tfrac{2}{3}\cdot\mathrm{E}[1+\mathrm{Geometric}(1/3)] + \tfrac{1}{3}\cdot\mathrm{E}[1+\mathrm{Geometric}(2/3)] = \tfrac{2}{3}(1+3) + \tfrac{1}{3}(1+\tfrac{3}{2}) = 3\tfrac{1}{2}.$$

   Alternative solution Let N be the number of rolls. The event $N = n$ happens if the sequence of rolls consists of $n-1$ `1`s followed by a `2` or $n-1$ `2`s followed by a `1`, with $n \geq 2$. Therefore $\mathrm{P}(N = n) = (2/3)^{n-1}(1/3) + (1/3)^{n-1}(2/3)$. In general,

   $$\sum_{n=2}^{\infty} n(1-p)^{n-1}p = \sum_{n=2}^{\infty} n(1-p)^{n-1}p - p = \mathrm{E}[\mathrm{Geometric}(p)] - p = \frac{1}{1-p} - p.$$

   Therefore

   $$\mathrm{E}[N] = \sum_{n=2}^{\infty} n(2/3)^{n-1}(1/3) + \sum_{n=1}^{\infty} n(1/3)^{n-1}(2/3) = \frac{3}{2} - \frac{1}{3} + 3 - \frac{2}{3} = 3\tfrac{1}{2}.$$

3. The length of a class (in minutes) is a Normal random variable with mean 42 and variance 25.

   (a) What is the probability that the class ends within 40 minutes?

   (b) Ada attends 51 (independent) classes this month. Use the normal approximation to estimate the probability that at least half her classes end within 40 minutes.

Solution:

(a) $P(X < 40) = P\left(\frac{X-\mu}{\sigma} < \frac{40-42}{5}\right) = P(Z < -0.4) \approx 1 - 0.6554 = 0.3446$.

(b) Let W be the number of her classes that would be dismissed early this week. Then W is a Binomial random variable with parameters $n = 51$ and $p \approx 0.3446$. Its mean and standard deviation are $\mu\, E[W] = np \approx 17.575$ and $\sigma = \sqrt{\mathrm{Var}[W]} = \sqrt{np(1-p)} \approx 3.394$. By the Central Limit Theorem,

$$P(W \geq 51/2) \approx P(\mathrm{Normal}(17.575, 3.394^2) \geq 25.5) \approx P(\mathrm{Normal} \geq 2.335) \approx 0.0098.$$

4. Find $E[(X+Y)^2]$, where X is a Geometric random variable with parameter $p = 0.4$, and Y is an independent Poisson random variable with parameter $\lambda = 2$.

Solution: $E[X+Y] = E[X]+E[Y] = 1/0.4+2 = 4.5$. By independence, $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] = (1-0.4)/0.4^2 + 2 = 5.75$. Therefore $E[(X+Y)^2] = \mathrm{Var}[X+Y] + E[X+Y]^2 = 5.75 + 4.5^2 = 26$.

5. Let X and Y be two continuous random variables with joint PDF

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

(a) Find the marginal PDF of X.

(b) Find $P(Y > 2X)$.

Solution:

(a) Using the formula for the PDF of a marginal,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_x^{\infty} 2e^{x-2y}dy = e^x \int_x^{\infty} 2e^{-2y}dy = e^x e^{-2x} = e^{-x},$$

so X is an Exponential(1) random variable.

(b) The conditional PDF of Y given X is $f_{Y|X}(y|x) = f_{XY}(x,y)/f_X(x) = 2e^{-2(y-x)}$ when $0 < x < y$. In words, $Y - X$ is an Exponential(2) random variable independent of X. The conditional CDF of Y is

$$P(Y \leq y | X = x) = P(\mathrm{Exponential}(2) \leq y - x) = 1 - e^{-2(y-x)}$$

when $0 < x < y$. By the total probability theorem

$$P(Y > 2X) = \int_0^{\infty} P(Y > 2X | X = x) f_X(x)dx = \int_0^{\infty} e^{-2x}e^{-x}dx = \int_0^{\infty} e^{-3x}dx = \frac{1}{3}.$$

Alternative solution: We can integrate the joint PDF over the event $E = \{(x,y) : y > 2x\}$:

$$P(Y > 2X) = \int_0^{\infty}\int_{2x}^{\infty} 2e^{x-2y}dydx = \int_0^{\infty} e^x e^{-4x}dx = \frac{1}{3}.$$

6. Five runners are randomly positioned on a 1000m circular track. Let E be the event that (at least) two of the runners are within 10m of one another. Is $\mathrm{P}(E) > 1/4$? Justify your answer.

   Solution: No. We model the runners' positions as independent uniform random variables (relative to some origin). Let X be the number of pairs of runners that are within 10m of one another. Then X is a sum of $\binom{5}{2} = 10$ indicator random variables X_{ij} for the event "runners i and j are within 10m". Such an event happens if runner j is either 10m ahead of or 10m behind runner i, so $\mathrm{P}(X_{ij} = 1) = 20/1000 = 1/50$. By linearity of expectation $\mathrm{E}[X] = \sum \mathrm{E}[X_{ij}] = \sum \mathrm{P}(X_{ij} = 1) = 10/50 = 1/5$. By Markov's inequality, $\mathrm{P}(X > 1) \leq \mathrm{E}[X] = 1/5 < 1/4$.

   Alternative solution 1: Let Y_i be the indicator random variable for the event "some other runner is 0m to 10m ahead of runner i". Then $X = Y_1 + \cdots + Y_5$. By independence, $\mathrm{P}(Y_i = 0) = (1 - 10/1000)^4 = 0.99^4$. Therefore $\mathrm{E}[Y_i] = \mathrm{P}(Y_i = 1) = 1 - 0.99^4$. By linearity of expectation, $\mathrm{E}[X] = \mathrm{E}[Y_1] + \cdots + \mathrm{E}[Y_5] = 5(1 - 0.99^4) < 0.198$. By Markov's inequality, $\mathrm{P}(X > 1) \leq \mathrm{E}[X] < 1/4$.

   Alternative solution 2: Let A_i be the event "all runners among $2, \ldots, i$ are more than 10m apart". Then $\mathrm{P}(A_2) = 1 - 20/1000$ as above. While it is difficult to calculate $\mathrm{P}(A_3|A_2)$ exactly we can obtain a lower bound. Once the positions of the first two runners have been fixed, there are at most 40m of track that is within 10m of either, so $\mathrm{P}(A_3|A_2) \geq 1 - 40/1000$. By the same reasoning $\mathrm{P}(A_4|A_3) \geq 1 - 60/1000$ and $\mathrm{P}(A_5|A_4) \geq 1 - 80/1000$. By the multiplication rule

   $$\mathrm{P}(A_5) \geq \mathrm{P}(A_5|A_4)\,\mathrm{P}(A_4|A_3)\,\mathrm{P}(A_3|A_2)\,\mathrm{P}(A_2) \approx 0.81$$

   The event E is the complement of A_5 so $\mathrm{P}(E) < 1/5 < 1/4$.