# SM4202 Exercise 1

1. (a) In combinatorics, the *inclusion-exclusion principle* states that for finite sets $A_1, \ldots, A_n$,

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{1\le i<j\le n} |A_i \cap A_j| + \sum_{1\le i<j<k\le n} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1}|A_1 \cap \cdots \cap A_n|$$

   Use the above formula to deduce the probability of $\mathrm{P}(A_1 \cup A_2 \cup A_3)$.

   (b) $X$ is a number chosen at random from $\{1, 2, ..., 1000000\}$, so that each number is equally likely. Find the probability that $X$ is divisible by one or more of the numbers 4,10 or 25.

2. A fair coin will be tossed twice, the number $N$ of heads will be noted, and then the coin will be tossed $N$ more times. Let $X$ be the total number of heads obtained.

   (a) Decide on a probability space $\Omega$, and make a table with the heading $\omega$, $\mathrm{P}(\omega)$, and $X(\omega)$.

   (b) Calculate the expectation $\mathrm{E}(X)$.

3. In a multiple choice examination Freda knows the correct answer with probability $p$; otherwise she guesses by randomly selecting one of the $m$ possible answers. Given that Freda correctly answers a question, what is the probability that she guessed it?

4. If I keep tossing a fair coin, what is the probability I get (a) the pattern $HH$ before the pattern $HT$; (b) the pattern $HH$ before the pattern $TH$?

5. (a) For independent events $A_1, \ldots, A_n$, show that

$$\mathrm{P}(A_1 \cup \cdots \cup A_n) = 1 - \prod_{i=1}^{n} \big(1 - \mathrm{P}(A_i)\big).$$

   (b) A pair of dice is rolled $n$ times. How large must $n$ be so that the probability of rolling at least one double six is more than 1/2?

6. Consider a simple random walk on the integers $\{0, 1, ..., 9, 10\}$, with steps $\pm 1$ each with probability 1/2 , and stopped as soon as the walk reaches either 0 or 10. Let $T$ be the number of steps before the walk reaches either 0 or 10. Suppose that $0 \le a \le 10$ and set $m(a) = E(T|\text{walk starts at } a)$.

   (a) Explain why $m(0) = m(10) = 0$.

   (b) Argue that

$$m(a) = 1 + \frac{1}{2}m(a-1) + \frac{1}{2}m(a+1) \text{ for } 0 < a < 10.$$

(c) Show that $m(a) = (10 - a)a$ solves these equations.

(d) Is it the unique solution?

7. For a random variable $X$ with mean $\mu$ and variance $\mathrm{Var}(X)$ and any given constant $c \in \mathbb{R}$, prove that

   (a) $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2$.

   (b) $\mathrm{Var}(X) = \mathrm{E}\left(X(X-1)\right) + \mu - \mu^2$.

   (c) $\mathrm{E}\left((X-c)^2\right) = \mathrm{Var}(X) + (\mu - c)^2$ so that the minimum mean squared deviation occurs when $c = \mu$.

8. (a) By example, or otherwise, show that generally $\mathrm{E}[\phi(X)] \neq \phi(\mathrm{E}[X])$.

   (b) A game is presented to you as follows: Independent random variables $X_i$ whose values generated by a computer take on either 0.5 with probability $\frac{1}{i+1}$, or 1 otherwise, for $i = 1, \ldots, 5$. These values are then multiplied together to give $X = X_1 X_2 \cdots X_5$, and $Y = 1/X$ is calculated. You are returned B\$ $Y$ for playing this game, after paying a certain fee to play. What is the maximum fee you are willing to pay to play this game?

9. The number of insurance claims that will be made directly to your company in each of $n$ counties next month are modelled as $n$ independent random variables $X_i \sim \mathrm{Pois}(\theta_i)$, $i = 1, \ldots, n$. Write $\psi = \sum_{i=1}^{n} \theta_i$. The total monthly direct claims is modelled as the random variable $X = \sum_{i=1}^{n} X_i$.

   (a) Obtain the probability generating function of $X_i$, and hence of $X$. Deduce the distribution of $X$.

   (b) The number of indirect claims for next month is modelled as an independent random variable $W$, with PGF $G_W(s) = e^{\psi(s^2-1)}$. Obtain the PGF of the total claims $Y = X + W$, $\mathrm{E}(Y)$, and $\mathrm{Var}(Y)$.

10. Conditional upon an unknown scientific constant $\mu$, let $X_i \sim f$ be iid random variables representing the future outcomes of a series of experiments, with $\mathrm{E}(X_i) = \mu$ units and $\mathrm{Var}(X_i) = 400$ squared units. The estimator for $\mu$ will be $\bar{X} = \sum_{i=1}^{n} X_i/n$. Assuming that the CLT applied with sufficiently fast convergence,

    (a) What is the probability that the realisation of $\bar{X}$ will be within 1 unit of $\mu$ when

      i. $n = 1$?

      ii. $n = 4$?

      iii. $n = 16$?

      iv. $n = 100$?

    (b) If the experimenter asks you what is the least number of experiments that should be performed in order to have a probability of at least 0.95 that $\bar{X}$ will be within 2 units of $\mu$, what do you reply?