Practice questions

1. You toss a coin 100 times. Which of the following random variables are independent?

   (a) The number of consecutive heads `HH` and the number of consecutive tails `TT`.

   (b) The number of consecutive heads in the first 50 tosses and the number of consecutive tails in the last 50 tosses.

   (c) The random variables in part (b), conditioned on having exactly 50 heads in the 100 coin tosses.

   Solution: We denote the random variables by X and Y in each part.

   (a) Not independent. $\mathrm{P}(X = 99, Y = 99) = 0$ as there cannot be 99 consecutive heads and 99 consecutive tails. However, $\mathrm{P}(X = 99) > 0$ and $\mathrm{P}(Y = 99) > 0$ as each of these events may occur individually (and has probability 2^{-100}. Therefore $\mathrm{P}(X = 99, Y = 99) \neq \mathrm{P}(X = 99)\,\mathrm{P}(Y = 99)$.

   (b) Independent. It is not easy to calculate these numbers, but we can reason it out. The probability of the event $Y = y$ does not depend on what happens in the first 50 coin tosses, so all the conditional probabilities $\mathrm{P}(Y = y|X = 0), \mathrm{P}(Y = y|X = 1), \dots, \mathrm{P}(Y = y|X = 49)$ have the same value p. By the total probability theorem,

   $$\begin{aligned}\mathrm{P}(Y = y) &= \mathrm{P}(Y = y|X = 0)\,\mathrm{P}(X = 0) + \cdots + \mathrm{P}(Y = y|X = 49)\,\mathrm{P}(X = 49)\\ &= p\,\mathrm{P}(X = 0) + \cdots + p\,\mathrm{P}(X = 49)\\ &= p,\end{aligned}$$

   so $\mathrm{P}(Y = y|X = x)$ and $\mathrm{P}(Y = y)$ are always the same.

   (c) Not independent. Let E be the event we are conditioning on. Conditioned on A, all $\binom{100}{50}$ balanced sequences of heads and tails are equally likely. In particular, $\mathrm{P}(X = 49|A) = 1/\binom{100}{50}$, as $X = 49$ can occur in one possible way given A. For the same reason, $\mathrm{P}(Y = 49|A) = 1/\binom{100}{50}$. But $\mathrm{P}(X = 49, Y = 49|A)$ is also $1/\binom{100}{50}$. Therefore $\mathrm{P}(X = 49, Y = 49|A) \neq \mathrm{P}(X = 49|A)\,\mathrm{P}(Y = 49|A)$ and so the two are not independent.

2. A fair coin is tossed 100 times. What is the expected number of times T that three consecutive heads occur? For example, if the outcome is `HHHHTHHH` then $T = 3$.

   Solution: Let T_i be the random variable that takes value 1 if tosses i, $i + 1$, and $i + 2$ are all heads, and 0 if not. Then $T = T_1 + T_2 + \cdots + T_{98}$. By the linearity of expectation $\mathrm{E}[T] = \mathrm{E}[T_1] + \cdots + \mathrm{E}[T_{98}]$. Each T_i takes value 1 with probability $(1/2)^3 = 1/8$, therefore has expected value $1/8$. Therefore $\mathrm{E}[T] = 98 \cdot (1/8) = 12.25$.

3. In 2017 there were 0.848 men for every woman in Hong Kong. Men and women had life expectancies of 81.7 years and 87.7 years, respectively. What was the life expectancy of a random person?

   Solution: Suppose a random person in 2017 lives L years. Let M be the event that person is a man. The life expectancies of men is the expected value of L, conditioned on that person is male, i.e. $\mathrm{E}[L \mid M]$. Assume every person in Hong Kong has equal probability to get picked, then $P(M)$ is the ratio of men to the entire population. By the law of total expectation, $\mathrm{E}[L] = \mathrm{E}[L \mid M]P(M) + \mathrm{E}[L \mid M^C]P(M^C) = 81.7 \cdot 0.848/1.848 + 87.7 \cdot 1/1.848 \approx 84.9$.

4. Consider $2m$ persons forming m couples who live together at a given time. Suppose that at some later time, the probability of each person being alive is p, independent of other persons. At that later time, let A be the number of persons that are alive and let S be the number of couples in which both partners are alive. Find $\mathrm{E}[S \mid A]$. *(Textbook problem 2.32)*

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

$$\begin{aligned}
\mathrm{E}[S \mid A = a] &= \sum_{i=1}^{m} \mathrm{E}[X_i Y_i \mid A = a] && (1)\\
&= m\,\mathrm{E}[X_1 Y_1 \mid A = a] && (2)\\
&= m\,\mathrm{E}[Y_1 \mid X_1 = 1, A = a]P(X_1 = 1 \mid A = a) && (3)\\
&= mP(Y_1 = 1 \mid X_1 = 1, A = a)P(X_1 = 1 \mid A = a) && (4)
\end{aligned}$$

   Here, equation (2) is due to linearity of expectation. Equation (3) is due to total expectation theorem and the expectation $\mathrm{E}[X_1 Y_1 \mid X_1 = 0, A = a] = 0$. In equation (4) we replace the expectation of a Bernoulli (0-1) random variable with the probability that it takes value 1.

   We can calculate $P(Y_1 = 1 \mid X_1 = 1, A = a)$: This is the probability that my partner has survived, given that I have survived and a people have survived. As all $2m - 1$ people have the same probability to be among the $a - 1$ other survivors. the probability that my partner made it is $(a - 1)/(2m - 1)$. We can similarly calculate $P(X_1 = 1 \mid A = a)$ as $a/(2m)$, as everyone including me is equally likely to be among the a survivors. Therefore $\mathrm{E}[S \mid A = a] = a(a - 1)/2(2m - 1)$.

5. Charlie is conducting telephone surveys as a part time job at CCPOS of CUHK. He needs 2 more surveys before going home. However, on randomly dialed calls, only 15% of receivers would complete the survey. Let X be the number of dials Charlie needs to make before going home. Find the expected value and variance of X.

   Solution: Let X_1 be the number of calls Charlie made up to and including the first success, and X_2 be the extra number of calls until (and including) his second success. The random variable of interest is $X_1 + X_2$. Each of X_1 and X_2 is a Geometric(0.15) random variable. By linearity of expectation, $\mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 2/0.15 \approx 13.3$.

   Moreover, the random variables X_1 and X_2 are independent because after calling X_1 people, Charlie restarts the experiment from scratch, regardless of the number of calls he made. We can therefore use linearity of variance and conclude that $\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] = 2 \cdot (1 - 0.15)/0.15^2 \approx 75.6$.