DMMR Tutorial 9

Discrete Probability

1. Suppose a biased coin, which lands heads with probability $1/3$ (and tails with probability $2/3$) each time it is flipped, is flipped 10 times.

   What is the conditional probability that the number of times the coin lands heads is exactly 3, given that (i.e., conditioned on the event that) the number of times it lands heads is divisible by 3? (Give an expression for this probability. You do not need to compute its exact value.)

   Solution:
   We know that the number of time the coin lands heads is described by the binomial distribution, with parameters $p = 1/3$ and $n = 10$.

   The probability that the coin lands head exactly k times is $\binom{n}{k}p^k(1-p)^{n-k}$.

   Recall that the conditional probability of event A conditioned on event B is defined by $P(A \mid B) = P(A \cap B)/P(B)$, assuming that $P(B) > 0$. Let A be the event that the coin comes up heads exactly 3 times. Let B be the event that the number of times the coin comes up heads is divisible by 3. Note that since $A \subseteq B$, so we have $A \cap B = A$. thus $P(A \mid B) = P(A)/P(B)$. We thus need to compute the probabilities $P(A)$ and $P(B)$.

   The probability $P(A)$ that the coin comes up heads exactly 3 times is (by the binomial distribution), given by:

   $$\binom{10}{3}(1/3)^3(2/3)^7$$

   The probability $P(B)$ that the number of times the coin lands heads is divisible by 3 is:

   $$\sum_{i=0}^{3}\binom{10}{3\cdot i}(1/3)^{3\cdot i}(2/3)^{10-(3\cdot i)}$$

   Thus, the conditional probability $P(A|B) = P(A)/P(B)$ is given by:

   $$\frac{\binom{10}{3}(1/3)^3(2/3)^7}{\sum_{i=0}^{3}\binom{10}{3\cdot i}(1/3)^{3\cdot i}(2/3)^{10-(3\cdot i)}}$$

   □

2. Suppose that a fair coin is flipped three times consecutively.

   Let E_1 and E_2 denote the events that “the first flip comes up heads”, and “the second flip comes up heads”, respectively.

   Let B denote the event that exactly one of the first and second coin tosses comes up heads.

   Prove that E_1, E_2, and B are pairwise-independent events, but that they are NOT mutually independent events.

(Thus, observe that pairwise independence of a collection of events does not imply mutual independence.)

Solution:
We represent each outcome of flipping the coin three times as a triple where, for example, (h, t, t) represents the outcome in which the first flip comes up heads, and the other two tails. There are thus 8 such possible outcomes, and we know that all 8 outcomes are equally likely (each has probability $\frac{1}{8}$).

First, let's establish the fact that E_1 is independent of E_2. We need to show that $P(E_1 \cap E_2) = P(E_1)P(E_2)$. We obviously know that $P(E_i) = \frac{1}{2}$, for $i \in \{1, 2\}$. We can also alternatively see this from the fact that the event $E_1 = \{(h, t, t), (h, t, h), (h, h, t), (h, h, h)\}$ contains exactly 4 outcomes, and thus $P(E_1) = \frac{4}{8} = \frac{1}{2}$. Similarly, $E_2 = \{(t, h, t), (t, h, h), (h, h, t), (h, h, h)\}$, and thus $P(E_2) = \frac{1}{2}$. Finally note that $E_1 \cap E_2 = \{(h, h, t), (h, h, h)\}$, and thus $P(E_1 \cap E_2) = \frac{2}{8} = \frac{1}{4}$. Thus $P(E_1 \cap E_2) = \frac{1}{4} = P(E_1)P(E_2)$.

We next show that B is pairwise-independent of both E_1 and E_2. We need to show that $P(B \cap E_i) = P(B)P(E_i)$. First, let us calculate $P(B)$. There are 4 outcomes in which exactly one of the first two flips comes up heads: $B = \{(h, t, h), (h, t, t), (t, h, h), (t, h, t)\}$. Thus, $P(B) = 4/8 = 1/2$. Moreover, $B \cap E_1 = \{(h, t, h), (h, t, t)\}$, and $B \cap E_2 = \{(t, h, h), (t, h, t)\}$ thus $P(B \cap E_1) = P(B \cap E_2) = \frac{2}{8} = \frac{1}{4} = P(B)P(E_1)$. Thus B and E_1 are (pairwise) independent, and also B and E_2 are (pairwise) independent.

Now observe that $E_1 \cap E_2 \cap B = \emptyset$, because it can not be the case that both of the first two coin flips came up heads *and* exactly one of the first two coin flips came up heads. Thus $P(E_1 \cap E_2 \cap B) = 0 \neq (1/2)^3 = P(E_1)P(E_2)P(B)$. Thus, the three events are not mutually independent. $\square$

3. Suppose that a pharmaceutical company has developed a new non-invasive test for a type of cancer. Their studies show that this new test has the following properties.

   (a) If the test is performed on a (random) person who has this type of cancer, then there is an 88% chance that the test result will be positive.

   (b) If, on the other hand, the test is performed on a (random) person who does not have this cancer, then there is a 9% chance that the test result will be positive.

   (c) Approximately 1 in 1000 persons in the entire population have this type of cancer.

   Suppose that this new test is performed on a (random) person in the population. What is the probability that the person actually has this cancer, given that their test result was positive?

   Solution:
   We want to compute $P(C|T)$, where C represents the event that the person has cancer, and T represents the event that the test is positive. We use $\overline{C}$ to denote the event that the person does not have cancer.

   Bayes' theorem tells us that $P(C|T) = P(T|C)P(C)/P(T)$, and moreover,

$$\begin{aligned} P(T) &= P(T \cap C) + P(T \cap \overline{C}) \\ &= P(T|C)P(C) + P(T|\overline{C})P(\overline{C}) \end{aligned}$$

Therefore

$$\begin{aligned} P(C|T) &= \frac{P(T|C)P(C)}{P(T)} \\ &= \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\overline{C})P(\overline{C})} \\ &= \frac{(0.88)(0.001)}{(0.88)(0.001) + (0.09)(0.999)} = 0.00969 \end{aligned}$$

Thus, a random person who tests positive for this test is still very unlikely to have this type of cancer: namely they only have a chance of less than 1 in 100 of having this cancer.
Still, this is nearly 10 times more likely than the probability $(1/1000)$ that a random person in the general population has this cancer. So the test does tell us something.

However, if the follow-up tests are painful and cause substantial long term discomfort, it is tricky to justify causing so much discomfort to so many people in the population who do not have the cancer, in order to catch early on the rarer cases of those who do have the cancer. Furthermore, obviously in the real world, the costs of such tests also have to be taken into account.

Although this example is very simplistic, it already indicates some of the dilemmas and trade-offs faced regularly by policy makers in health care. □

4. Let us suppose that the number of crisp packets that the Walkers Crisp Company produces in a given day is a random variable. Suppose that the average number of crisp packets that Walkers produces in a day is 100,000.

   Prove that the probability that Walkers produces more than 1.5 million packets in a given day is at most $1/15$.

   Solution:
   This is a straightforward application of Markov's inequality, which tells us that for any non-negative random variable X, and any $a > 0$, $P(X \geq a) \leq \frac{E(X)}{a}$. Assuming $E(X) > 0$, this says that for any $k > 0$, $P(X \geq k \cdot E(X)) \leq \frac{1}{k}$.

   In our case, if X is the random variable which gives the number of crisp packets produced by Walkers in a given day, then $E(X) = 100000$, and $P(X \geq 1.5 \cdot 10^6) = P(X \geq 15 \cdot E(X)) \leq \frac{1}{15}$.
   □

5. (This question is harder than the others. I offer some hints below.)

   Suppose you are a big fan of Star Wars.

   Suppose that Kellogg's Corn Flakes has made a deal with Disney (the company that owns the rights to Star Wars), allowing Kellogg's to place inside each Corn Flakes cereal box a small replica action figure for one of 25 Star Wars characters.

   Suppose each of the 25 Star Wars action figures is equally likely to be placed in each Corn Flakes cereal box. (The box cover does not indicate which action figure is inside the cereal box.)

   Suppose your goal is to collect all $n = 25$ Star Wars action figures.

   What is the expected number of cereal boxes that you would have to buy in order to do this?

   Solve this for general n, and then plug in $n = 25$ to get the specific solution.

   (*Hints:* Consider the random variable X, which denotes the total number of boxes you bought until you obtained all n action figures. Consider also the random variables, X_i, $i = 1, \ldots, n$, denoting the number of boxes you bought after you had already collected $i - 1$ different action figures, up to and including the box which gave you the new i'th different action figure. Note that $X = \sum_{i=1}^{n} X_i$. Note also that X_i is a geometrically distributed random variable.)

Solution:

Note that X_i is a geometrically distributed random variable, with parameter $p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$. This is because, after having seen $i-1$ action figures, the probability that we will next see an action figure different from any of those is p_i.

Thus, since X_i is geometrically distributed, we have $E(X_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$.

Using linearity of expectation, we have:

$$\begin{aligned}
E(X) &= E(\sum_{i=1}^{n} X_i) \\
&= \sum_{i=1}^{n} E(X_i) \\
&= \sum_{i=1}^{n} \frac{n}{n-i+1} \\
&= n \sum_{i=1}^{n} \frac{1}{n-i+1} \\
&= n \sum_{i=1}^{n} \frac{1}{i}
\end{aligned}$$

Using the fact that $\ln n \leq \sum_{i=1}^{n} \frac{1}{i} \leq (\ln n) + 1$, we get that $n \ln n < E(X) \leq n \ln n + n$. So, $E(X) = n \ln n + O(n)$.

In particular, if there are $n = 25$ action figures, then the expected number of cereal boxes you would need to buy to collect all of them t is $25 \sum_{i=1}^{25} \frac{1}{i}$, which is between $25 \cdot \ln 25 = 80.47$ and $80.47 + 25 = 105.47$. A tighter approximation of the hormonic series $\sum_{i=1}^{n} \frac{1}{i}$ yields that $\sum_{i=1}^{25} \frac{1}{i} \approx 3.8159$ and thus that $E(X)$ is roughly 95.4, when $n = 25$.

So, the expected number of cereal boxes you would have to buy to collect all 25 Star Wars action figures is roughly 95.4. $\square$