Exercise sheet: Review of Probability

Solutions prepared by: Mr Chunchao Ma, Mr Areeb Sherwani. Supervised by M Álvarez

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time to solve compared to an exercise with (***), which is a more complex exercise.

1. (*) The table below gives details of symptoms that patients presented and whether they were suffering from meningitis Using this dataset, calculate the following probabilities

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

(a) $P(\text{Vomiting} = \text{true})$.

(b) $P(\text{Headache} = \text{false})$.

(c) $P(\text{Headache} = \text{true}, \text{Vomiting} = \text{false})$.

(d) $P(\text{Vomiting} = \text{false}|\text{Headache} = \text{true})$.

(e) $P(\text{Meningitis} = \text{true}|\text{Fever} = \text{true}, \text{Vomiting} = \text{false})$.

Solution:
(a) Vomiting = true corresponds to IDs 3,4,6,7,8 and 10 (6 samples). Since there is no conditioning, we consider the full set of data (10 samples). Therefore
$P(\text{Vomiting} = \text{true}) = 6/10 = 0.60$

(b) Headache = false corresponds to IDs 2, 5 and 9 (3 samples). Since there is no conditioning, we consider the full set of data (10 samples). Therefore
$P(\text{Headache} = \text{false}) = 3/10 = 0.30$

(c) Since we're taking the joint here, we're looking for the IDs for which both statements are correct. That's only ID 1 (1 sample).Since there is no conditioning, we consider the full set of data (10 samples). Therefore
$P(\text{Headache} = \text{true}, \text{Vomiting} = \text{false}) = 1/10 = 0.10$

(d) We're asked to find the probability of picking a symptom where Vomiting is false, given that they already have the symptom of Headache being true. Since we're conditioning on the statement Headache = true, we first identify the IDs for which this is correct. These are 1,3,4,6,7,8 and 10 (7

samples). This is the set we'll be working with. Next, of this set, we identify the IDs for which Vomiting = false. This is only ID 1 (1 sample). Therefore
$P(\text{Vomiting} = \text{false}|\text{Headache} = \text{true}) = 1/7 = 0.14$

(e) We're asked to find the probability of picking a symptom where Meningitis is true, given that they already have the symptom of Fever being true as well as Vomiting being false. Since we're conditioning on the statement Fever = true and Vomiting = false, we first identify the IDs for which this is correct. These are 1,2,5 and 9 (4 samples). This is the set we'll be working with. Next, of this set, we identify the IDs for which Meningitis = true. This is only ID 5 (1 sample). Therefore
$P(\text{Meningitis} = \text{true}|\text{Fever} = \text{true}, \text{Vomiting} = \text{false}) = 1/4 = 0.25$

2. (*) Consider the experiment of tossing a coin three times. Let X be the RV giving the number of heads obtained. We assume that the tosses are independent and the probability of a head is p. Find the probabilities $P(X = 0), P(X = 1), P(X = 2)$, and $P(X = 3)$.

Solution:
For this, we need to find the sum of the probabilities of each event that satisfies the condition. We can do this by finding the probability of a single event that satisfies the condition (for example getting HTT satisfies $X = 1$) and multiplying it by all the different combinations that satisfy it (HTT, THT and TTH gives us 3 combinations). The probability for getting n heads in 3 trials is $p^n(1-p)^(3-n)$. The number of combinations of n heads in 3 trials is given by $\binom{3}{n}$. Giving us a final formula of

$$P(X = n) = \binom{n}{x} p^n (1-p)^{3-n}$$

Therefore

$$P(X = 0) = \binom{3}{0} p^0 (1-p)^{3-0} = (1-p)^3$$

$$P(X = 1) = \binom{3}{1} p^1 (1-p)^{3-1} = 3(1-p)^2 p$$

$$P(X = 2) = \binom{3}{2} p^2 (1-p)^{3-2} = 3(1-p)p^2$$

$$P(X = 3) = \binom{3}{3} p^3 (1-p)^{3-3} = p^3$$

3. (***) Suppose that the two RVs X and Z are statistically independent. Show that the mean and variance of their sum satisfies

$$\begin{aligned} E\{X+Z\} &= E\{X\} + E\{Z\} \\ \mathrm{var}\{X+Z\} &= \mathrm{var}\{X\} + \mathrm{var}\{Z\}. \end{aligned}$$

Solution:

Since X and Z are independent, their joint distribution factorises $p(x,z) = p(x)p(z)$, and so

$$\begin{aligned} \mathbb{E}\{X+Z\} &= \int\int (x+z)p(x,z)\mathrm{d}x\mathrm{d}z = \int\int (x+z)p(x)p(z)\mathrm{d}x\mathrm{d}z \\ &= \int\int xp(x)p(z)\mathrm{d}x\mathrm{d}z + \int\int zp(x)p(z)\mathrm{d}z\mathrm{d}x \\ &= \int xp(x)\mathrm{d}x \int p(z)\mathrm{d}z + \int zp(z)\mathrm{d}z \int p(x)\mathrm{d}x \\ &= \int xp(x)\mathrm{d}x + \int zp(z)\mathrm{d}z \\ &= \mathbb{E}\{X\} + \mathbb{E}\{Z\} \end{aligned}$$

where we have used $\int p(z)\mathrm{d}z = 1$ and $\int p(x)\mathrm{d}x$.

Similarly for the variances, say we first define $W = X + Z$. We are want to compute

$$\mathrm{var}\{W\} = E\{W - E\{W\}\}^2.$$

By replacing W for $X + Z$ in the expression above, we get

$$\begin{aligned} \mathrm{var}\{X+Z\} &= E\{X+Z-E\{X+Z\}\}^2 = E\{(X-E\{X\}) + (Z-E\{Z\})\}^2 \\ &= E\{(X-\mathbb{E}\{X\})^2 + (Z-\mathbb{E}\{Z\})^2 + 2(X-\mathbb{E}\{X\})(Z-\mathbb{E}\{Z\})\}. \end{aligned}$$

Applying the definition of the expected value, we get

$$\begin{aligned} \mathrm{var}\{X+Z\} &= \int\int \left[(x-\mathbb{E}\{X\})^2 + (z-\mathbb{E}\{Z\})^2 + 2(x-\mathbb{E}\{X\})(z-\mathbb{E}\{Z\})\right] p(x,z)\mathrm{d}x\mathrm{d}z \\ &= \int\int (x-\mathbb{E}\{X\})^2 p(x,z)\mathrm{d}x\mathrm{d}z + \int\int (z-\mathbb{E}\{Z\})^2 p(x,z)\mathrm{d}x\mathrm{d}z \\ &+ 2\int\int (x-\mathbb{E}\{X\})(z-\mathbb{E}\{Z\})p(x,z)\mathrm{d}x\mathrm{d}z. \end{aligned}$$

Because $p(x,z) = p(x)p(z)$, due to independence, the three double integrals follow as

$$\begin{aligned}
\int\int (x - \mathbb{E}\{X\})^2 p(x,z)\mathrm{d}x\mathrm{d}z &= \int\int (x - \mathbb{E}\{X\})^2 p(x)p(z)\mathrm{d}x\mathrm{d}z \\
&= \int (x - \mathbb{E}\{X\})^2 p(x)\mathrm{d}x \int p(z)\mathrm{d}z = \mathrm{var}\{X\} \\
\int\int (z - \mathbb{E}\{Z\})^2 p(x,z)\mathrm{d}x\mathrm{d}z &= \int\int (z - \mathbb{E}\{Z\})^2 p(x)p(z)\mathrm{d}x\mathrm{d}z \\
&= \int p(x)\mathrm{d}x \int (z - \mathbb{E}\{Z\})^2 p(z)\mathrm{d}z = \mathrm{var}\{Z\} \\
\int\int (x - \mathbb{E}\{X\})(z - \mathbb{E}\{Z\} p(x,z)\mathrm{d}x\mathrm{d}z &= \int\int (x - \mathbb{E}\{X\})(z - \mathbb{E}\{Z\}) p(x)p(z)\mathrm{d}x\mathrm{d}z \\
&= \int (x - \mathbb{E}\{X\}) p(x)\mathrm{d}x \int (z - \mathbb{E}\{Z\}) p(z)\mathrm{d}z \\
&= \left[\int xp(x)\mathrm{d}x - \int \mathbb{E}\{X\} p(x)\mathrm{d}x\right]\left[\int zp(z)\mathrm{d}z - \int \mathbb{E}\{Z\} p(z)\mathrm{d}z\right] \\
&= \left[\mathbb{E}\{X\} - \mathbb{E}\{X\}\int p(x)\mathrm{d}x\right]\left[\mathbb{E}\{Z\} - \mathbb{E}\{Z\}\int p(z)\mathrm{d}z\right] \\
&= 0.
\end{aligned}$$

Putting together these results, we get

$$\mathrm{var}\{X+Z\} = \mathrm{var}\{X\} + \mathrm{var}\{Z\}.$$

For discrete variables the integrals are replaced by summations, and the same results are again obtained.

4. (*) Consider a discrete RV X whose pmf is given as

$$P(X) = \begin{cases} \frac{1}{3}, & \text{if } x = -1, 0, 1, \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and variance of X.

Solution:

The mean of X is

$$E(X) = \frac{1}{3}(-1 + 0 + 1) = 0$$

The variance of X is

$$\mathrm{Var}(X) = E\left[(X - E(X))^2\right] = E\left(X^2\right) = \frac{1}{3}\left[(-1)^2 + (0)^2 + (1)^2\right] = \frac{2}{3}$$

5. (**) The RV X can take values $x_1 = 1$ and $x_2 = 2$. Likewise, the RV Y can take values $y_1 = 1$ and $y_2 = 2$. The joint pmf of the RVs X and Y is given as

$$P(X,Y) = \begin{cases} k(2x_i + y_j), & \text{for } i = 1,2\,; j = 1,2 \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Find the value of k.

(b) Find the marginal pmf for X and Y.

(c) Are X and Y independent?

Solution:

(a) Find the value of k:

$$\sum_{x_i}\sum_{y_j} P\left(x_i, y_j\right) = \sum_{x_i=1}^{2}\sum_{y_j=1}^{2} k \times \left(2x_i + y_j\right)$$
$$= k \times [(2+1)+(2+2)+(4+1)+(4+2)] = k \times 18 = 1$$

We then obtain $k = \frac{1}{18}$.

(b) The marginal pmf for X is

$$\begin{aligned} P\left(X\right) &= \sum_{y_j} P\left(x_i, y_j\right) = \sum_{y_j=1}^{2} \frac{1}{18}\left(2x_i + y_j\right) \\ &= \frac{1}{18}\left(2x_i + 1\right) + \frac{1}{18}\left(2x_i + 2\right) = \frac{1}{18}\left(4x_i + 3\right) \quad x_i = 1, 2. \end{aligned}$$

We therefore obtain:

$$P(X) = \begin{cases} \frac{1}{18}\left(4x_i + 3\right), & \text{for } i = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

The marginal pmf for Y is

$$\begin{aligned} P\left(Y\right) &= \sum_{x_i} P\left(x_i, y_j\right) = \sum_{x_i=1}^{2} \frac{1}{18}\left(2x_i + y_j\right) \\ &= \frac{1}{18}\left(2 + y_j\right) + \frac{1}{18}\left(4 + y_j\right) = \frac{1}{18}\left(2y_j + 6\right) \quad y_j = 1, 2. \end{aligned}$$

We therefore obtain:

$$P(Y) = \begin{cases} \frac{1}{18}\left(2y_j + 6\right), & \text{for } j = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

(c) Now $P\left(X\right)P\left(Y\right) \neq P\left(X, Y\right)$; Hence X and Y are not independent.

6. (**) The joint pdf of the RVs X and Y is given by

$$p(x,y) = \begin{cases} k(x+y), & \text{for } 0 < x < 2, 0 < y < 2 \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Find the value of k.

(b) Find the marginal pdf for X and Y.

(c) Are X and Y independent?

Solution:

(a) Find the value of k:

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dxdy &= k\int_0^2\int_0^2 (x+y)dxdy \\ &= k\int_0^2 \left(\frac{x^2}{2}+xy\right)\Big|_{x=0}^{x=2} dy \\ &= k\int_0^2 (2+2y)dy = 8k = 1 \end{aligned}$$

We then obatin $k = \frac{1}{8}$

(b) The marginal pdf for X is

$$\begin{aligned} p(x) &= \int_{-\infty}^{\infty} p(x,y)dy = \frac{1}{8}\int_0^2 (x+y)dy \\ &= \frac{1}{8}\left(xy+\frac{y^2}{2}\right)\Big|_{y=0}^{y=2} = \left\{ \begin{array}{ll} \frac{1}{4}(x+1) & 0 < x < 2 \\ 0 & \text{otherwise} \end{array} \right. \end{aligned}$$

Since $p(x,y)$ is symmetric with respect to x and y, the marginal pdf of Y is

$$p(y) = \left\{ \begin{array}{ll} \frac{1}{4}(y+1) & 0 < y < 2 \\ 0 & \text{otherwise} \end{array} \right.$$

(c)Now $p(x)\,p(y) \neq p(x,y)$; Hence X and y are not independent.

7. (**) Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $P(r) = 0.2, P(b) = 0.2, P(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Solution

Based on the question, we know that $P(r) = 0.2, P(b) = 0.2, P(g) = 0.6$. We assume $P(\text{Apple}|r)$ is the probability of selecting an apple in the red colour box; $P(\text{Apple}|b)$ is the probability of selecting an apple in the blue colour box; $P(\text{Apple}|g)$ is the probability of selecting an apple in the green colour box. Then we get

$$P(\text{Apple} \mid r) = \frac{3}{3+4+3} = \frac{3}{10},$$
$$P(\text{Apple} \mid b) = \frac{1}{1+1+0} = \frac{1}{2},$$
$$P(\text{Apple} \mid g) = \frac{3}{3+3+4} = \frac{3}{10}.$$

The probability of selecting an apple is $P(\text{Apple})$:

$$\begin{aligned} P(\text{Apple}) =& P(\text{Apple} \mid r)P(r) + P(\text{Apple} \mid b)P(b) + P(\text{Apple} \mid g)P(g) \\ =& \frac{3}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6 = 0.34 \quad . \end{aligned}$$

For the second question, we have observed that the selected fruit is in fact an orange. We want to know $P(g \mid \text{Orange})$, that is, the probability that a selected orange coming comes from a green colour box.

Similarly, we assume $P(\text{Orange}|r)$ is the probability of selecting an orange in red colour box; $P(\text{Orange}|b)$ is the probability of selecting an orange in blue colour box; $P(\text{orange}|g)$ is the probability of selecting an orange in green colour box. Then we get

$$P(\text{Orange} \mid r) = \frac{4}{3+4+3} = \frac{4}{10},$$
$$P(\text{Orange} \mid b) = \frac{1}{1+1+0} = \frac{1}{2},$$
$$P(\text{Orange} \mid g) = \frac{3}{3+3+4} = \frac{3}{10}.$$

Based on Bayes' theorem,

$$P(g \mid \text{Orange}) = \frac{P(\text{Orange} \mid g)P(g)}{P(\text{Orange})}.$$

The probability of selecting an Orange is $P(\text{Orange})$:

$$\begin{aligned} P(\text{Orange}) =& P(\text{Orange} \mid r)P(r) + P(\text{Orange} \mid b)P(b) + P(\text{Orange} \mid g)P(g) \\ =& \frac{4}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6 = 0.36 \quad . \end{aligned}$$

We thus obtain:

$$P(g \mid \text{Orange}) = \frac{P(\text{Orange} \mid g)P(g)}{P(\text{Orange})} = \frac{\frac{3}{10} \times 0.6}{0.36} = 0.5 \quad .$$