MAST20004 Probability

Assignment Two: Solutions

Problem 1. A randomly selected family has n children with probability αp^n ($n \geqslant 1$) where $p \in (0,1)$ and $\alpha \leqslant (1-p)/p$.

(i) What is the probability that a randomly selected family does not have any children?
(ii) Suppose that each child is equally likely to be a boy or a girl (independently of each other), and the number of children is also independent of the sexes. What is the probability that a randomly selected family has one boy (and any number of girls)? [Hint: *Given that the family has n children, what is the distribution of the number of boys? You don't need to justify this mathematically.*]
(iii) Under the same assumption as in Part (ii), what is the probability that a randomly selected family has k boys (and any number of girls)? [Hint: *You may use the following identity which is a direct corollary of the extended binomial theorem in the lecture:*

$$(1-x)^{-(k+1)} = \sum_{m=0}^{\infty} \binom{m+k}{k} x^m, \quad |x| < 1.]$$

Solution. (i) We have

$$\mathbb{P}(\text{family has no children}) = 1 - \sum_{n=1}^{\infty} \alpha p^n = 1 - \frac{\alpha p}{1-p} = \frac{1-(\alpha+1)p}{1-p}.$$

(ii) For each $n \geqslant 1$, given that the family has n children, the number of boys is a binomial random variable with parameter n and $1/2$. Therefore,

$$\begin{aligned}
\mathbb{P}(\text{one boy}) &= \sum_{n=1}^{\infty} \mathbb{P}(\text{one boy}, n \text{ children in total}) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(\text{one boy}|n \text{ children in total}) \times \mathbb{P}(n \text{ children in total}) \\
&= \sum_{n=1}^{\infty} \binom{n}{1} \frac{1}{2} \times \big(\frac{1}{2}\big)^{n-1} \times \alpha p^n = \alpha \sum_{n=1}^{\infty} n \big(\frac{p}{2}\big)^n \\
&= \frac{2\alpha p}{(2-p)^2}.
\end{aligned}$$

(iii) We first consider the case when $k \geqslant 1$. In a similar way as in Part (ii), we have

$$\begin{aligned}
\mathbb{P}(k \text{ boys}) &= \sum_{n=k}^{\infty} \mathbb{P}(k \text{ boys}|n \text{ children}) \times \mathbb{P}(n \text{ children}) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} \frac{1}{2^n} \cdot \alpha p^n = \alpha \sum_{n=k}^{\infty} \binom{n}{k} \cdot \big(\frac{p}{2}\big)^n \\
&= \alpha \cdot \big(\frac{p}{2}\big)^k \cdot \sum_{m=0}^{\infty} \binom{m+k}{k} \big(\frac{p}{2}\big)^m.
\end{aligned}$$

To evaluate the series, recall from the extended binomial theorem that

$$(1+x)^\alpha = \sum_{m=0}^{\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-m+1)}{m!} x^m, \quad |x| < 1.$$

By taking $\alpha = -(k+1)$ and changing x to $-x$, we have

$$\begin{aligned}
(1-x)^{-(k+1)} &= \sum_{m=0}^{\infty} \frac{(-1)^m (k+1) \cdot \cdots \cdot (k+m)}{m!} (-x)^m \\
&= \sum_{m=0}^{\infty} \binom{m+k}{k} x^m.
\end{aligned}$$

Therefore, we conclude that

$$\mathbb{P}(k \text{ boys}) = \alpha \cdot \big(\frac{p}{2}\big)^k \cdot \big(1 - \frac{p}{2}\big)^{-(k+1)} = \frac{2\alpha p^k}{(2-p)^{k+1}}.$$

Now we consider the case when $k=0$. We have

$$\begin{aligned}
\mathbb{P}(\text{no boys}) &= \mathbb{P}(\text{no boys, no children}) + \sum_{n=1}^{\infty} \mathbb{P}(\text{no boys, } n \text{ children}) \\
&= \mathbb{P}(\text{no children}) + \sum_{n=1}^{\infty} \mathbb{P}(n \text{ children}) \mathbb{P}(\text{no boys} \mid n \text{ children}) \\
&= \frac{1-(\alpha+1)p}{1-p} + \sum_{n=1}^{\infty} \alpha p^n \times \frac{1}{2^n} \\
&= \frac{1-(\alpha+1)p}{1-p} + \frac{\alpha p}{2-p} \\
&= 1 - \frac{\alpha p}{(1-p)(2-p)}.
\end{aligned}$$

The result can also be obtained by using

$$\mathbb{P}(\text{no boys}) = 1 - \sum_{k=1}^{\infty} \mathbb{P}(k \text{ boys}).$$

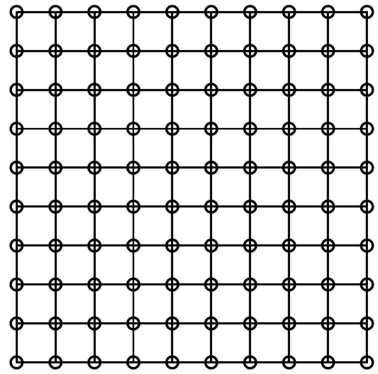
Remark. For those who wish to pursue full mathematical precision, in Part (ii) we are implicitly making use of the following structure. Let $\{Y_i : i \geqslant 1\}$ be a sequence of independent $B(1, 1/2)$-random variables, where $Y_i = 1$ if the i-th child is a boy. Let X be another random variable that is independent of the sequence $\{Y_i : i \geqslant 1\}$. X represents the total number of children in the family. Then the number of boys is given by the random variable

$$Z = Y_1 + \cdots + Y_X.$$

As a result, for any $k \leqslant n$ we have

$$\begin{aligned}
&\mathbb{P}(Z = k | X = n) \\
&= \mathbb{P}(Y_1 + \cdots + Y_X = k | X = n) \\
&= \mathbb{P}(Y_1 + \cdots + Y_n = k | X = n) \quad (\text{because we are given } X = n) \\
&= \mathbb{P}(Y_1 + \cdots + Y_n = k) \quad (\text{because } X \text{ and } \{Y_1, \cdots, Y_n\} \text{ are independent}) \\
&= \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}.
\end{aligned}$$

Problem 2. Suppose that a particular suburb of Melbourne is built on the following 10×10 grid.



At each intersection, a street light is placed (represented by a small circle). A street refers to a segment between two adjacent lights. We say that a street is *dark*, if the two lights at both ends are broken. Suppose that the lights on the boundary of the grid are lack of proper maintenance so that each has a probability of 0.2 being broken. All other lights have broken probability 0.1. All the lights are independent of each other. Let X be the number of dark streets in this suburb.

(i) How do you write X as a sum of Bernoulli random variables? State the corresponding Bernoulli trials precisely.
(ii) Use Poisson approximation to calculate the probability that there are at least three dark streets in the suburb. Explain heuristically why we can use Poisson approximation in this problem (you don't need to make any precise mathematical justification).

Solution. (i) There are $9 \times 10 \times 2 = 180$ streets in total. Each street corresponds to a Bernoulli trial (whether the street is dark or not). The associated Bernoulli random variable is defined to be 1 if this street is dark and 0 otherwise. Then we can write X as the sum of all these 180 Bernoulli random variables.

(ii) Note that not all the 180 Bernoulli random variables have the same parameter (darkness probability). In view of the assumption on the light broken probabilities, we need to divide these 180 streets into three groups:

Group A. Both lights of the street are on the boundary of the grid.
Group B. Precisely one of the lights of the street is on the boundary.
Group C. None of the lights of the street are on the boundary.

For a street in Group A, its darkness probability is $0.2^2 = 0.04$. There are $9 \times 4 = 36$ streets in this group. For a street in Group B, its darkness probability is $0.2\times 0.1 = 0.02$. There are $8\times 4 = 32$ streets in this group. For a street in Group C, its darkness probability is $0.1^2 = 0.01$. There are $180-36-32 = 112$ streets in this group. Since each street is at most dependent on 6 streets who may share one common light with the given street, we can say that the dependence among these streets (Bernoulli trials) is weak. Therefore, we can use Poisson approximation

$$X \overset{d}{\approx} \mathrm{Pn}(3.2 = 36\times 0.04 + 32\times 0.02 + 112\times 0.01)$$

to calculate

$$\begin{aligned}\mathbb{P}(X \geqslant 3) &= 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1) - \mathbb{P}(X=2)\\ &= 1 - e^{-3.2} - e^{-3.2}\cdot 3.2 - \frac{e^{-3.2}}{2!}3.2^2\\ &\approx 0.6201.\end{aligned}$$

Problem 3. Let X be the lifetime (measured in hours) of a particular type of electronic device, whose probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{C}{x^3}, & x > 10,\\ 0, & \text{otherwise}.\end{cases}$$

(i) Find the value of C.
(ii) Compute $\mathbb{P}(X > 20)$ and $\mathbb{E}[X]$. Does X have finite variance?
(iii) In a batch of 6 such devices, what is the probability that at least 3 of them will function for at least 20 hours? We assume that all the 6 devices are independent.

Solution. (i) We have

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_{10}^{\infty} \frac{C}{x^3}dx = \frac{C}{200} \implies C = 200.$$

(ii) We have

$$\mathbb{P}(X > 20) = \int_{20}^{\infty} \frac{200}{x^3}dx = \frac{1}{4} = 0.25,$$

and

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{10}^{\infty} x\cdot\frac{200}{x^3}dx = 20.$$

The random variable does not have a finite variance since

$$\mathbb{E}[X^2] = \int_{10}^{\infty} x^2\cdot\frac{200}{x^3}dx = 200\int_{10}^{\infty}\frac{dx}{x} = +\infty.$$

(iii) Let Y be the number of devices that function at least 20 hours. From Part (ii) we know that $Y \sim B(6, 0.25)$. Therefore,

$$\begin{aligned}\mathbb{P}(Y \geqslant 3) &= 1 - \mathbb{P}(Y=0) - \mathbb{P}(Y=1) - \mathbb{P}(Y=2)\\ &= 1 - 0.75^6 - \binom{6}{1}\cdot 0.25 \cdot 0.75^5 - \binom{6}{2} 0.25^2 \cdot 0.75^4\\ &\approx 0.1694.\end{aligned}$$

Problem 4. Let X be a normal random variable with mean 3 and variance 4.

(i) Find the following probabilities:

$$\mathbb{P}(2 < X \leqslant 5),\ \mathbb{P}(X > 3),\ \mathbb{P}(|X| > 2).$$

(ii) Find the values of a, b (where $a < b$) such that

$$\mathbb{P}(X < a) : \mathbb{P}(a < X < b) : \mathbb{P}(X > b) = 1 : 2 : 3.$$

You may use the fact that $\Phi(0.96741) \approx 0.83333$.

Solution. (i) We have

$$\begin{aligned}\mathbb{P}(2 < X \leqslant 5) &= \mathbb{P}\big(-0.5 < Z \leqslant 1\big) \quad (Z = \frac{X-3}{2})\\ &= \Phi(1) - \Phi(-0.5)\\ &= \Phi(1) + \Phi(0.5) - 1\\ &\approx 0.84134 + 0.69146 - 1\\ &= 0.5328.\end{aligned}$$

Similarly, we have

$$\mathbb{P}(X > 3) = 0.5$$

and

$$\begin{aligned}\mathbb{P}(|X| > 2) &= \mathbb{P}(X > 2) + \mathbb{P}(X < -2)\\ &= \mathbb{P}(Z > -0.5) + \mathbb{P}(Z < -2.5)\\ &= \Phi(0.5) + 1 - \Phi(2.5)\\ &\approx 0.69146 + 1 - 0.99379\\ &= 0.69767.\end{aligned}$$

(ii) Since the sum of the three probabilities must be 1, we have

$$\mathbb{P}(X < a) = \frac{1}{6},\ \mathbb{P}(a < X < b) = \frac{1}{3},\ \mathbb{P}(X > b) = \frac{1}{2}.$$

The last property implies that $b = 3$. The first property implies that $\Phi(\frac{a-3}{2}) = \frac{1}{6}$, or equivalently,

$$\Phi\big(\frac{3-a}{2}\big) = \frac{5}{6} \approx 0.83333.$$

Therefore, $\frac{3-a}{2} = 0.96741$, namely

$$a = 1.06518.$$

Problem 5. (1) A table tennis match is played between Horatio and Xi. The winner of the match is the one who first wins 4 games in total, and in any individual game the winner is the one who first scores 11 points. Note that in an individual game, if the score is 10 to 10, the game goes into extra play (called *deuce*) until one player has gained a lead of 2 points. Let p be the probability that Horatio wins a point in any single round of serve, and assume that different rounds in all games are independent.

(1-i) In an individual game, what is the probability that the game runs into the deuce stage?
(1-ii) Suppose that $p = 0.55$. What is the probability that Horatio wins the match?

(2) Let $n \geqslant 1$ and $p \in (0,1)$ and $q = 1-p$. By constructing suitable probability models or otherwise, show that

$$\binom{2n}{n} p^n q^n + \sum_{k=0}^{n-1} \binom{n+k}{k} (q^k p^{n+1} + p^k q^{n+1}) = 1.$$

[Hint: *There are at least two methods to show this. Method 1: Consider an artifical table tennis game which has the deuce rule when the score is n-n, and use two methods to calculate the probability of entering the deuce stage. Method 2: Consider the relation between binomial and negative binomial random variables.*]

Solution. (i) We provide two methods. Let $q = 1-p$.

Method 1. If the game runs into deuce, there are precisely 20 rounds played, in which Horatio scores 10 points and Xi scores 10 points. Therefore,

$$\mathbb{P}\,(\text{Deuce}) = \binom{20}{10} p^{10} \cdot q^{10}.$$

Method 2. We can first compute the probability that the game ends without entering the deuce stage, that is,

$$\begin{aligned}\mathbb{P}(\text{Deuce}) &= 1 - \mathbb{P}(\text{No Deuce}) \\ &= 1 - \sum_{k=0}^{9} \binom{10+k}{k} (q^k p^{11} + p^k q^{11}).\end{aligned}$$

(ii) We first compute the probability that Horatio wins an individual game. Let D (respectively, E) be the event that Horatio wins the game in the deuce stage (respectively, without deuce). Note that if the game runs into scores 10-10, an extra even number of rounds must be played before a winner appears. For $r \geqslant 1$, let D_r be the event that Horatio wins the game at the $2r$-th round after entering deuce. In particular, in each pair $(2i-1, 2i)$ of rounds at the deuce stage ($1 \leqslant i \leqslant r-1$), the scoring pattern must be either (Horatio, Xi) or (Xi, Horatio). In addition, the last pair $(2r-1, 2r)$ of rounds must be both scored by Horatio. Therefore, we have

$$\begin{aligned}&\mathbb{P}(\text{Horatio wins the game}) \\ &= \mathbb{P}(D) + \mathbb{P}(E) = \sum_{r=1}^{\infty} \mathbb{P}(D_r) + \mathbb{P}(E) \\ &= \binom{20}{10} p^{10} q^{10} \times \sum_{r=1}^{\infty} (2pq)^{r-1} p^2 + \sum_{k=0}^{9} \binom{10+k}{k} q^k p^{11} \\ &= \binom{20}{10} p^{12} \cdot q^{10} \cdot \frac{1}{1-2pq} + \sum_{k=0}^{9} \binom{10+k}{k} q^k p^{11} \\ &\approx 0.6868.\end{aligned}$$

We call the above probability p_1. A simple use of negative binomial distributions shows that

$$\mathbb{P}(\text{Horatio wins the match}) = \sum_{k=0}^{3} \binom{3+k}{k} (1-p_1)^k p_1^4 \approx 0.8562.$$

Here $k = 0, 1, 2, 3$ represents the number of games Xi wins.

(2) We show this identity by using two methods.

Method 1. Consider an artificial table tennis game with the deuce rule at scores n-n. Let p be Horatio's winning probability of each point. The identity follows immediate from the same argument as in Part (1-ii), by computing $\mathbb{P}(\text{Deuce})$ directly and computing $1 - \mathbb{P}(\text{No Deuce})$ using negative binomial distributions.

Method 2. Let us denote

$$A = \binom{2n}{n} p^n q^n, \ B = \sum_{k=0}^{n-1} \binom{n+k}{k} q^k p^{n+1}, \ C = \sum_{k=0}^{n-1} \binom{n+k}{k} p^k q^{n+1}.$$

It is not hard to see that B is a negative binomial probability. More precisely, if we consider a negative binomial random variable $X \sim NB(n+1, p)$, then

$$B = \mathbb{P}(X \leqslant n-1).$$

From the relation between negative binomial and binomial distributions, we know that B can also be expressed as

$$B = \mathbb{P}(Z_1 \geqslant n+1), \quad Z_1 \sim B(2n, p).$$

Similarly, C can be expressed as

$$C = \mathbb{P}(Z_2 \geqslant n+1), \quad Z_2 \sim B(2n, q).$$

Now a crucial observation is that, $Z_2 = 2n - Z_1$ (because $q = 1 - p$). Therefore,

$$\begin{aligned} B + C &= \mathbb{P}(Z_1 \geqslant n+1) + \mathbb{P}(2n - Z_1 \geqslant n+1) \\ &= \mathbb{P}(Z_1 \geqslant n+1) + \mathbb{P}(Z_1 \leqslant n-1) \\ &= 1 - \mathbb{P}(Z_1 = n) \\ &= 1 - \binom{2n}{n} p^n q^n \\ &= 1 - A. \end{aligned}$$

It follows that

$$A + B + C = 1.$$