Probability and Statistics 1

Sheet 3

29 January 2024

Every problem sheet in this course has two sections:

1. **Practice makes permanent**: problems that support you to develop a sound knowledge of and capacity to apply the definitions, ideas and methods introduced each week.
2. **The deeper thinking**: problems that provide challenge and intrigue, alongside a richer appreciation of the underlying ideas.

You should attempt all of the problems in the **practice makes permanent** section. Solutions to these problems will be provided at a later date.

The deeper thinking problems are optional - but encouraged! Full solutions to these problems will **not** usually be provided, though I may share hints, ideas and outline solutions for some of these.

Practice makes permanent

Problem 1. Kartik plays a game with his friends to identify three colas (labelled as colas 1, 2 and 3) after tasting one can of each of three different brands (Coca-cola, Pepsi and Karma Cola). Kartik has no idea which cola is which so assigns the brand names to the numbers completely at random. Let X denote the number of colas correctly identified by this random strategy. Define the distribution of X by confirming the value of $\frac{1}{2}$ in the table below and filling in the gaps:

k	0	1	2	3
$p_X(k)$		$\frac{1}{2}$		

Check that the function $p_X(k)$ defined by your table is a probability mass function, that is, it satisfies the conditions of Theorem 3.1.5.

Problem 2. Shenjun is throwing darts at a dartboard, and aims every dart at the bull's eye. A probability model for her throws is constructed from the following assumptions:

- The probability that Shenjun hits the bull's eye is 0.4 for each shot.

- Each throw is independent of all the others.

Let X be the number of Darts Shenjun throws until she his the bull's eye for the first time.

(a) Show that under the given probability assumptions, $X \sim \text{Geo}(0.4)$

(b) Calculate $\mathbb{P}(X = 4)$

(c) Calculate $\mathbb{P}(X \leq 10)$

(d) Using the formulas from Theorem 3.3.7, find $\mathbb{E}(X)$ and $\text{Var}(X)$.

(e) Shenjun enters a competition in which she wins £100 less £20 for each missed shot at the bull's eye (if this means her winnings are negative, she loses money!). Show that her winnings can be represented by the formula $120 - 20X$, and hence find the expectation and variance of her winnings by applying Theorem 3.3.5.

Problem 3. My tutorial class has 20 students in it, but not everyone attends every session (even though they should!). I construct a probability model for this situation using the following assumptions:

- The probability that any given student does not attend any given tutorial is the same and is equal to 0.2
- The students' attendance is independent of each other.

Using this probability model,

(a) Let X be the number of students (out of the 20 students in my class) who don't attend my tutorial this week. What is the distribution of X?

(b) Compute the probability that at least one student doesn't attend my tutorial this week.

(c) Compute the probability that at least two students don't attend my tutorial this week.

(d) Compute the probability that at least two students don't attend my tutorial this week given that at least one student doesn't attend my tutorial this week.

Problem 4. At the Hangar Lane Gyratory, accidents requiring an ambulance occur with a frequency, on average, of 1.8 per week. In this question we assume that such accidents occur randomly and independently, and at a uniform average rate.

(a) If X represents the number of accidents requiring an ambulance in a given week at the Hangar Lane Gyratory, what is the distribution of X under our assumptions?

(b) Compute the probability that there will not be an accident in a particular week.

(c) How many weeks in a row must there be before the chance of at least one accident across all those weeks exceeds 99.9%?

(d) Compute the probability that there are exactly 2 accidents in a particular week given that there are at least 2 accidents that week.

Problem 5. This question asks you to prove that the Geometric and Poisson distributions provide well-defined probability mass functions, thereby completing Exercise 3.2.1 from our lectures.

(a) Let $X \sim \text{Geo}(p)$, so $p_X(k) = p(1-p)^{k-1}$ for $k = 1, 2, 3, \cdots$. Show that p_X thus defined is a probability mass function.

(b) Let $X \sim \text{Po}(\lambda)$, so $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \cdots$. Show that p_X thus defined is a probability mass function.

Hints: In part (a), you should find a geometric progression that you can sum. In part (b), you will need to recall the series expansion $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

Problem 6. In this question we prove Theorem 3.3.5(ii) from lectures, which is that $\text{Var}(aX + b) = a^2 \text{Var}(X)$. In both parts (a) and (b) below you will need to:

- Use the definition of variance as $\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right]$
- Use Theorem3.3.5(i), which states that $\mathbb{E}(pX + q) = p\mathbb{E}(X) + q$ when $p, q \in \mathbb{R}$

Suppose that X is a discrete random variable such that $\mathbb{E}(X)$ and $\text{Var}(X)$ exist.

(a) For $a \in \mathbb{R}$, show that $\text{Var}(aX) = a^2 \text{Var}(x)$.

(b) For $b \in \mathbb{R}$, show that $\text{Var}(X + b) = \text{Var}(x)$.

(c) Combine parts (a) and (b) to show that $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

The deeper thinking

Problem 7. Bush House
A group of m students enter a single lift in Bush House, which has 8 floors (in addition to the ground floor). The building happens to be otherwise deserted. Each student chooses their destination floor independently of the others, and - from our point of view - completely at random, so that each student selects a floor with probability 1/8. Let S_m be the number of times the elevator stops.

In order to study S_m, we introduce for $i = 1, 2, \ldots, 8$ random variables R_i, given by

$$R_i = \begin{cases} 1 & \text{if the elevator stops at the } i\text{th floor} \\ 0 & \text{if the elevator does not stop at the } i\text{th floor} \end{cases}$$

(a) Show that each R_i has a Bernoulli distribution with parameter $p = 1 - \left(\frac{7}{8}\right)^m$.

(b) From the way we defined S_m, it follows that

$$S_m = R_1 + R_2 + \cdots R_8.$$

Can we conclude that S_m has a $Bin(8, p)$ distribution, with p as in (a)?
Why or why not?

(c) Clearly, if $m = 1$, one has that $\mathbb{P}(S_1 = 1) = 1$. Show that for $m = 2$,

$$\mathbb{P}(S_2 = 1) = \frac{1}{8} = 1 - \mathbb{P}(S_2 = 2),$$

and that S_3 has the following distribution:

a	1	2	3
$\mathbb{P}(S_3 = a)$	1/64	21/64	21/32

Problem 8. Derivation of Poisson probabilities

Let X be the number of electrons emitted by a radioactive source in 1 second, and suppose that the *average* number of emissions in 1 second is known to be λ. The aim of this question is to derive the probability mass function for a Poisson distribution from two assumptions:

- *Homogeneity*: The average arrival rate λ is constant over time: in an interval of time of length t, the expectation of the number of emissions is λt.
- *Independence*: The number of emissions in disjoint time intervals are independent random variables.

The derivation proceeds by constructing probabilities that approximate those we want, and then refines that approximation by taking a limit.

We begin by dividing the time period of 1 second into n sub-periods each lasting $1/n$ seconds.

(a) Considering a 'trial' to be 'is there an emission in the ith sub-period?', explain why X can be approximated by a $\mathrm{Bin}(n, p)$, where p is the probability that there is an emission in the ith sub-period. Your explanation should involve the *independence* assumption above.

(b) When n is large enough, every sub-period of time will contain either 0 or 1 emission, so the number of emissions in the ith sub-period will have a $\mathrm{Ber}(p)$ distribution. Use the *homogeneity* assumption to show that $p = \frac{\lambda}{n}$.

(c) Hence deduce that

$$\mathbb{P}(X = k) = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad \text{for } k = 0, 1, \ldots, n$$

(d) Show that

$$\lim_{n\to\infty} \binom{n}{k} \frac{1}{n^k} = \frac{1}{k!}.$$

(e) Now use the result from calculus that

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

to show that

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, \ldots$$