ST102/ST109 Exercise 7

In this exercise you will practise working with some common discrete distributions. Question 1 is quite standard and makes use of the binomial distribution. Question 2 makes use of the fact that the sum of n independent and identically distributed Bernoulli trials with success probability π has a $\text{Bin}(n, \pi)$ distribution (as noted in lectures), and also remember how to calculate a conditional probability. Question 3 uses one distribution (the Poisson) to determine the parameter of another distribution (the binomial) – this will require you to *think*! Question 4 requires the derivation of the moment generating function (mgf) of the binomial distribution – this was seen in lectures via the mgf of the Bernoulli distribution, but here try and use the binomial theorem. Finally, Question 5 uses the Poisson distribution – a subtle hint for part (a)! After this are some simple R examples (for reference) about sampling.

You should attempt these questions ahead of the corresponding class. It will be covered by your class teacher in your seventh class, which will take place in the week commencing Monday 18 November 2024.

1. A certain electronic system contains 15 components. Suppose that the probability that each individual component will fail is 0.4 and that the components fail independently of each other. Given that at least two of the components have failed, what is the probability that at least three of the components have failed?

2.* Suppose that the random variables $X_1, X_2, \ldots, X_n$ form n independent Bernoulli trials, each with parameter π. Determine the conditional probability that $X_1 = 1$, given that:

$$\sum_{i=1}^{n} X_i = k$$

for $k = 1, 2, \ldots, n$.

3.* Suppose that a book with n pages contains on average λ typographical errors, i.e. 'typos', per page. What is the probability that there will be at least m pages which contain more than k typos?

Hint: Try using a binomial distribution where the probability of success is determined from a Poisson distribution. Your answer should be a sum of binomial probabilities.

4. (a) Prove that the mgf of the binomial distribution $\text{Bin}(n, \pi)$ is:

$$M_X(t) = (\mathrm{e}^t \pi + (1 - \pi))^n.$$

Hint: Use the binomial theorem, which is given on page 74 of the lecture material course pack.

(b) Use the moment generating function to derive the mean and variance of the binomial distribution.

5. A glacier in Greenland 'calves' (lets fall off into the sea) an iceberg on average twice every five weeks. (Seasonal effects can be ignored for this question, and so the calving process can be thought of as random, i.e. the calving of icebergs can be assumed to be independent events.)

   (a) Explain which distribution you would use to estimate the probabilities of different numbers of icebergs being calved in different periods, justifying your selection.

   (b) What is the probability that no iceberg is calved in the next three weeks?

   (c) What is the probability that no iceberg is calved in the three weeks after the next three weeks?

   (d) What is the probability that exactly five icebergs are calved in the next four weeks?

   (e) If exactly five icebergs are calved in the next four weeks, what is the probability that exactly five more icebergs will be calved in the four-week period after the next four weeks?

   (f) Comment on the relationship between your answers to (d) and (e).

Discrete uniform distributions in R (for reference only)

We can use the function `sample` to generate samples from a vector. The default setting is sampling *without replacement*, for which (obviously) we cannot select more cases than the length of the vector!

For example, `sample(1:10,6)` draws 6 integers (the second argument) from the integers 1 to 10 (the first argument) at random without replacement, such as:

```
> sample(1:10,6)
[1]  4 10  8  6  9  7
```

In fact it would be sufficient to just use `sample(10,6)` as the first argument (here 10) is viewed as the length of a sequence of integers.

```
> sample(10,6)
[1] 2 8 6 1 9 4
```

Clearly, we may choose to sample *with replacement* which may result in a value being sampled more than once (or may not, but at least allows for the possibility). This is achieved by adding the argument '`replace=TRUE`', or more concisely simply '`replace=T`'. For example:

```
> sample(10,6,replace=T)
[1] 2 4 5 5 8 8
```

Under sampling with replacement, using '`sample`' we can simulate observations from any discrete uniform distribution where the first argument is the parameter k (see page 89).

Try experimenting with different values of k and different sample sizes, n. In each case plot a histogram and see how closely it resembles the probability function of the discrete uniform distribution. For example, if $k = 100$ and $n = 50{,}000$ use:

```
> hist(sample(100,50000,replace=T))
```