ST102 Exercise 11

In this exercise you will practise sampling distributions. Question 1 considers a random sample drawn from a Poisson distribution. Question 2 involves the sampling distribution of the sample mean. Question 3 concerns the use of the central limit theorem to approximate the sampling distribution of the sample mean. Question 4 requires you to derive the sampling distribution of the nth order statistic, i.e. the largest value in a random sample. Finally, Question 5 requires you to work with the normal distribution.

Your answers to this problem set should be submitted as a pdf file upload to Moodle. It will be covered by your class teacher in your eleventh class, which will take place in the week commencing Monday 23 January 2023.

1.* Suppose $\{X_1, X_2, \ldots, X_n\}$ is a random sample from the Pois(λ) distribution.

   (a) What is the sampling distribution of $\sum_{i=1}^{n} X_i$?

   (b) Write down the sampling distribution of $\bar{X} = \sum_{i=1}^{n} X_i/n$. In other words, write down the possible values of $\bar{X}$ and their probabilities. (Assume n is *not* large.)

   Hint: What are the possible values of $\sum_{i=1}^{n} X_i$ and their respective probabilities?

   (c) What are the mean and variance of the sampling distribution of $\bar{X}$ when $\lambda = 5$ and $n = 100$?

   (d) Use R to simulate 10 random samples of size $n = 100$ from a Poisson distribution with $\lambda = 5$. Calculate the sample mean for each of these observed samples, then calculate the sample mean and variance of these 10 values of $\bar{X}$. Are they reasonably close to the theoretical values you stated in (c)?

   For example:

```
> sample1<-rpois(100,5)
> mean(sample1)
[1] 4.88
```

2. A manufacturer of objects packages them in boxes of 200. It is known that, on average, the objects weigh 1 kg with a standard deviation of 0.03 kg. The manufacturer is interested in calculating:

$$P(\text{200 objects weigh more than 200.5 kg})$$

which would help detect whether too many objects are being put in a box. Explain how you would calculate the (approximate?) value of this probability. Mention any relevant theorems or assumptions needed.

3.* A country is about to hold a referendum about leaving the European Union. A survey of a random sample of adult citizens of the country is conducted. In the sample, n respondents say that they plan to vote in the referendum. These n respondents are then asked whether they plan to vote 'Yes' or 'No'. Define $X_i = 1$ if the ith person plans to vote 'Yes', and $X_i = 0$ if the ith person plans to vote 'No', for $i = 1, 2, \ldots, n$.

Suppose that in the whole population 49% of those people who plan to vote are currently planning to vote Yes, and hence the referendum result would show a (very small) majority *opposing* leaving the European Union.

(a) Let $\bar{X} = \sum_{i=1}^{n} X_i/n$ denote the *proportion* of the n voters in the sample who plan to vote Yes. What is the central limit theorem approximation of the sampling distribution of $\bar{X}$ here?

(b) If there are $n = 50$ likely voters in the sample, what is the probability that $\bar{X} > 0.50$? (Such an opinion poll would suggest a majority *supporting* leaving the European Union in the referendum.)

(c) How large should n be so that there is less than a 1% chance that $\bar{X} > 0.50$ in the random sample? (This means less than a 1% chance of the opinion poll *incorrectly* predicting a majority *supporting* leaving the European Union in the referendum.)

4.* Suppose that $\{X_1, X_2, \ldots, X_n\}$ is a random sample from a continuous distribution with probability density function $f_X(x)$ and cumulative distribution function $F_X(x)$. Here we consider the sampling distribution of the statistic:

$$Y = X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$$

i.e. the largest value of X_i in the random sample, for $i = 1, 2, \ldots, n$.

(a) Write down the formula for the cumulative distribution function $F_Y(y)$ of Y, i.e. for the probability that all observations in the sample are less than or equal to y.

(b) From the result in (a), derive the probability density function $f_Y(y)$ of Y.
Hint: The answer is on page 6 of the Lent term lecture notes! ;o)

5. The heights (in cm) of men aged over 16 in England are approximately normally distributed with a mean of 174.9 and a standard deviation of 7.39. What is the probability that in a random sample of 60 men from this population *at least one* man is more than 1.92 metres tall?

Hint: This relates to Question 4 (a) and think about using the $P(A) = 1 - P(A^c)$ trick.