Practical 2: The Poisson Process

This work should be submitted to NESS before ***4pm Friday 12th October 2018****.*

Marks for each question are indicated. These marks should be used as a rough guide to the relative weight of individual questions. The final marks for the practical may be slightly different from those indicated.

An ability to forecast patient arrivals can be very useful for hospital emergency departments in various decision making processes, for example, concerning the staffing of the unit.

1 Assessing the Suitability of a Poisson Process Model

In the first part of this practical you will consider whether the Poisson process is a good model for patient arrivals in a particular hospital emergency department. For this emergency department, the arrival times of patients, in hours after 9am, have been recorded over a 12 hour period. The data can be read into R as follows

```
url = "http://www.mas.ncl.ac.uk/~nseg4/teaching/MAS8380/practical2.dat"
arr.times = scan(url)
```

Question 1:

Examine the arrivals process graphically by plotting each arrival time against its position in the sequence, i.e. plot i vs t_i where t_i is the arrival time of the i-th patient. In a Poisson process, events occur randomly in time at a constant rate, say λ. Does this seem compatible with your plot?

[**5 marks**]

If the patients are arriving according to a Poisson process with rate λ then in an interval of length ℓ the number of arrivals will be a random variable X with

$$X \sim \text{Po}(\lambda\ell).$$

Note that $E[X] = \lambda\ell$ and $\text{Var}(X) = \lambda\ell$. Suppose that the period over which arrivals are observed has length L and we split it into k equal length intervals. If we count the number of arrivals in each of the k intervals, this gives rise to random variables $X_1, X_2, \ldots, X_k$. If the arrivals occur according to a Poisson process, the X_i will be independent with $X_i \sim \text{Po}(\lambda L/k)$ for each $i = 1, \ldots, k$. If $x_1, \ldots, x_k$ are the *observed* counts, then we should have

$$\begin{aligned} \text{mean}\{x_1, \ldots, x_k\} &\approx \lambda L/k, \text{ and} \\ \text{variance}\{x_1, \ldots, x_k\} &\approx \lambda L/k \end{aligned}$$

so

$$\boxed{\text{mean}\{x_1, \ldots, x_k\} \approx \text{variance}\{x_1, \ldots, x_k\}}$$

You can count the number of arrivals in an interval of length L/k using the `table` and `cut` commands.

```
x = table(cut(arr.times, breaks=k))
```

The vector `x` will then contain the observations $x_1, \ldots, x_k$.

Question 2:

For the patient arrivals data set, count the number of arrivals in $k = 10$ intervals of equal length. Record the mean and variance of the counts. Repeat this for $k = 25, 50, 75, 100, 150$. Prepare a graph of mean versus variance for these different values of k. Does this plot support the idea the arrivals follow a Poisson process?

Provide an estimate of the rate parameter λ along with your method for doing this.

Include your R code in your report for this question.

[**12 marks**]

2 Simulating a Poisson Process

A second emergency department believes it can model patient arrivals using a Poisson process with rate 10 arrivals per hour. It wants to simulate a sample of patient arrivals over a 12 hour period.

Question 3:

You need to simulate a collection of points from a Poisson process with rate 10 arrivals per hour. As we saw in lectures, the distance between times (i.e. the inter-arrival times) in a Poisson process with rate λ corresponds to an exponential Exp(λ) random variable. You can therefore simulate points from a Poisson process as follows.

1. Simulate a single random quantity from the exponential distribution with parameter $\lambda = 10$, and call it y_1. This will be the time of the first arrival $t_1 = y_1$.
2. Simulate a second Exp(10) random quantity, say y_2 and add it to t_1. This sum will be the time of the second arrival $t_2 = y_1 + y_2$.
3. Simulate a third Exp(10) quantity, call it y_3, and let $t_3 = y_1 + y_2 + y_3$.
4. Continue in the same way to obtain $t_4, t_5, \ldots$ until the times t exceeds $L = 12$.
5. Collect all the values $t_1, t_2, t_3, \ldots$ into a single vector.

Use the `rexp` command to generate exponential random numbers in R. The `cumsum` command provides a cumulative sum of a vector.

Produce a plot of the arrival times like that you produced in Question 1.

Verify graphically that your generated arrivals follow a Poisson process with rate $\lambda = 10$.

In writing up this question, simply submit a copy of this plot plus the R code you used to generate the data.

[**8 marks**]