EDRP Lab 3

Version: 2020-04-10

Problem 1 - Poisson distribution

Recall that there is a couple of commands for working with probability distributions in R. The commands take the *root name* of the probability distribution and prefix the root with some letters. The letters are d, p, q, or r. These give, respectively,

- continuous density or discrete probability mass function (d),
- cumulative distribution function (p),
- quantile (q),
- and random variable (r).

The root name of Poisson distribution is *pois*. Here you have the full syntax for all quantities that can be computed using R:

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

So, for example,

```
dpois(x=3,lambda=1)
```

compare it with

```
lambda<-1
k<-3
(exp(-lambda)*(lambda)^k)/factorial(k)
```

is just $\mathbb{P}(X = 3)$ when X has a Poisson distribution with $\lambda = 1$,

```
ppois(2, lambda=1)
```

compare it with

```
lambda=1

k<-0
s0<-(exp(-lambda)*(lambda)^k)/factorial(k)

k<-1
s1<-(exp(-lambda)*(lambda)^k)/factorial(k)

k<-2
s2<-(exp(-lambda)*(lambda)^k)/factorial(k)

print(s0+s1+s2)
```

is the value of the cumulative distribution function for X at 2

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2).$$

If you want to draw 100 random numbers from the set $\{0, 1, 2, \ldots\}$ according to a Poisson distribution with $\lambda = 7$, you just type

```
rpois(n=100, lambda=7)
```

Observe that the arithmetical average of such numbers is

```
mean(rpois(n=100, lambda=7))
```

what agrees with the fact we know from the lectures that $\mathbb{E}X = \lambda$ for $X \sim \text{Poisson}(\lambda)$. The ordinary arithmetical average is an *(empirical) estimate* of the *(theoretical) expected value.*

Or, the proportion of the value 6 in a sample of the size 10^6 is

```
length(which(rpois(n=1e6, lambda=7)==6))/1e6
```

what agrees with the probability $\mathbb{P}(X = 6)$ for $X \sim \text{Poiss}(7)$, which equals

```
dpois(6,lambda=7)
```

So the proportion of 6's is an *empirical estimate* of the *(theoretical) probability* $\mathbb{P}(X = 6)$.

1. Generate two vectors. The first vector should contain 10^6 random numbers from a Poisson distribution with $\lambda_1 = 3$, and the second should contain 10^6 random numbers from a Poisson distribution with $\lambda_2 = 2$. Add the two vectors, and treat it as a random sample with 10^6 elements. For $i = 0, 1, \ldots, 10$ find the proportion of the values i in the sample and compare it with the values of $\mathbb{P}(X = i)$ for $X \sim \text{Poisson}(3 + 2)$.

Hint: you can use something like

```
cat("var1 = ", var1, "var2 = ", var2, "\n")
```

to display a mix of some strings and values of some variables in single lines.

Problem 2 - Exponential distribution

In R, the root name for exponential distribution is *exp*.

1. Simulate 1000 observations from an exponential distribution. How can you use the observations to convince yourself that $\mathbb{E}(X) = 1/\lambda$ if $X \sim \text{Exp}(\lambda)$?
2. Recall the following

Theorem

Let $X_1, \ldots, X_n$ be independent exponential random variables with respective parameters $\lambda_1, \ldots, \lambda_n$. If $M = \min(X_1, \ldots, X_n)$, then:

- $M \sim \text{Exp}(\lambda_1 + \ldots + \lambda_n)$,
- for $k = 1, \ldots, n$,

$$\mathbb{P}(M = X_k) = \frac{\lambda_k}{\lambda_1 + \ldots + \lambda_n}.$$

Devise and perform a simulation illustrating this theorem.

Problem 3 - Simulation of a Poisson process

One of the theoretical results presented in the lectures gives the following direct method for simulating Poisson proces:

- Let $\tau_0 = 0$.

- Generate i.i.d. exponential random variables $\rho_1, \rho_2, \ldots$.
- Let $\tau_n = \rho_1 + \ldots + \rho_n$ for $n = 1, 2, \ldots$.
- For each $k = 0, 1, \ldots$, let $N_t = k$ for $\tau_k \leq t < \tau_{k+1}$.

1. Using this method, generate a realization of a Poisson process $(N_t)_t$ with $\lambda = 0.5$ on the interval $[0, 20]$. Plot the trajectory you generated.

Remark: You need to decide how many observations to generate from exponential distribution in order to be able to generate the trajectory of the Poisson process up to time 20. How can you estimate this number?

2. Generate 10000 realizations of a Poisson process $(N_t)_t$ with $\lambda = 0.5$ and use your results to estimate $\mathbb{P}(N_{10} = i)$, $i = 0, \ldots, 9$, and $\mathbb{E}N_{10}$. Compare the obtained estimates with the theoretical values.

Problem 4 - Another method to simulate a Poisson process

Some mathematical results for arrival times and uniform distribution offer another method for simulating a Poisson process with parameter λ on an interval $[0, t]$:

- Simulate the number of arrivals N in $[0, t]$ from a Poisson distribution with parameter λt.
- Generate N random numbers from the uniform distribution on $(0, t)$.
- Sort the variables in increasing order - the sorted list gives the Poisson arrival times.

1. Using this method, generate a realization of a Poisson process $(N_t)_t$ with $\lambda = 0.5$ on the interval $[0, 20]$. Plot the trajectory you generated.
2. Generate 10000 realizations of a Poisson process $(N_t)_t$ with $\lambda = 0.5$ and use your results to estimate $\mathbb{P}(N_{10} = i)$, $i = 0, \ldots, 9$, and $\mathbb{E}N_{10}$. Compare the obtained estimates with the theoretical values.

Problem 5 - Thinning of a Poisson process

1. Simulate 50 Poisson arrivals with rate λ. With each arrival associate a random label 1, 2 or 3, with the respective probabilities 0.5, 0.2, 0.3. One way to do it is to generate the values of $\tau_1, \ldots, \tau_{50}$ as one list, and a sample of size 50 from the distibution

1	2	3
0.5	0.2	0.3

as another list. Then you can apply a technique from EDRPLab0 tutorial to transform the two lists into a data frame with two columns and 51 rows (remember that $\tau_0 = 0$). You can do it in such way that the first column will contain the values of $\tau_0, \ldots, \tau_0$, while the second column will have anything (say, 0) as the first value, and the sample generated from the three-point distribution as the remaining 50 values.

2. Split the list of the arrivals into three sublists, looking at the values from the second column of your data frame. Treat the three resulting sublists as the lists of arrival times of three new Poisson processes $(N_t^{(i)})_t$, $i = 1, 2, 3$.
3. Repeat the steps from the previous points 100000 times. Thus, you should create 100000 realizations of the processes $(N_t^{(i)})_t$, $i = 1, 2, 3$. Your aim is to find, for each new process, the proportion of realizations for which $N_3^{(i)} = j$, with $i = 1, 2, 3$, $j = 0, 1, \ldots, 5$. Compare the proportions with the corresponding theoretical values.