Statistics with R
AIMS Senegal

Tabea Rebafka
October 2018

# Feuille de TP N° 3

## Comparison of estimators

## 1 Simulation of probability distributions

It is very useful to know to generate realizations of diverse probability distributions. In `R` this is done by the functions of the form `r`*func*(`n,param`) where *func* indicates the name of the probability distribution, `n` is the number of realizations to be generated and `param` are the parameters of the distribution. Here are some examples, where `n` denotes the number of independent realisations of the distribution:

| Distribution | `R` function |
|---|---|
| uniform $U[0,1]$ | `runif(n)` |
| uniform $U[a,b]$ | `runif(n, a, b)` |
| normal $\mathcal{N}(0,1)$ | `rnorm(n)` |
| normal $\mathcal{N}(m,s^2)$ | `rnorm(n, m, s)` |
| Poisson $\mathcal{P}(\lambda)$ | `rpois(n, lambda)` |
| binomial $\text{Bin}(k,p)$ | `rbinom(n, k, p)` |
| exponential $\mathcal{E}(\theta)$ | `rexp(n, theta)` |
| Gamma $\Gamma(a,b)$ | `rgamma(n, a, b)` |
| Cauchy$(\theta)$ | `rcauchy(n, 0, theta)` |

The functions of the form `r`*func* (with *func*=`norm` or `unif`...) all have little sisters of the form

- `d`*func*(`x,arguments`) : evaluates the probability density function at `x` when *func* is a continuous distribution. For discrete distributions, `d`*func*(`x,arguments`) returns the probability of the corresponding distribution to take the value `x`.
- `p`*func*(`x,arguments`) : returns the cumulative distribution function evaluated at `x`,
- `q`*func*(`a,arguments`) : returns the `a`-quantile of the corresponding distribution.

**Exercice 1.**

1. Compute the value of the density of the standard normal distribution at 0.
2. Compute the probability $\mathbb{P}(X \leq 0)$ for $X \sim \mathcal{N}(0,5)$.
3. Generate a sample named `data.exp` containing 300 observations of the exponential distribution $\mathcal{E}(4)$. Plot the histogram of this sample.
4. Use the function `curve()` to add the density function of the exponential distribution $\mathcal{E}(4)$ to the graph of the histogram of the sample `data.exp`. Interpret the result.
5. Generate a sample named `data.pois` from the Poisson distribution $\mathcal{P}(2)$. Plot the bar chart by the following instruction

```r
plot(table(data.pois)/300)
```

6. Evaluate the probabilities $\mathbb{P}(X = k)$ of a random variable $X \sim \mathcal{P}(2)$ for $k \in \{0, 1, \dots, 10\}$. Add these probabilities to the bar chart. Interpret the result.

# 2 Programming in `R`

Concerning programming in `R`, read the notebook `Programming in R.ipynb` (or the pdf *R_Programming_english.pdf*).

# 3 Convergence of the sample mean

The aim of the exercise is to approximate the value of the mean squared error *MSE* of a given estimator $\hat{\theta}$ of $\theta$ defined by

$$MSE = \mathbb{E}_\theta \left[ (\hat{\theta} - \theta)^2 \right].$$

To this end, we can simulate $s$ independant i.i.d. samples $\mathbf{x}_1 = (x_{1,1}, \dots, x_{1,n}), \dots, \mathbf{x}_s = (x_{s,1}, \dots, x_{s,n})$ all of size $n$, from the distribution $\mathbb{P}_\theta$. On every sample we evaluate the estimator $\hat{\theta}$. Denote $\hat{\theta}_k = \hat{\theta}(\mathbf{x}_k)$ for $k = 1, \dots, s$ the values of the estimator $\hat{\theta}$ on the different samples. An approximation of the mean squared error *MSE* is then given by

$$\widehat{MSE} = \frac{1}{s} \sum_{k=1}^{s} (\hat{\theta}_k - \theta)^2.$$

**Exercice 2.**

1. Write a function named `mean_exp` whose arguments are the sample size `nb` and the parameter `param` of the exponential distribution. The function generates 100 samples of the exponential distribution $\mathcal{E}$(`param`) and returns the vector of the sample means associated with these samples.

2. Write a function called `mse` whose arguments are a vector of estimates `estimates` and the true parameter value `param`. The function returns the mean squared error $\widehat{MSE}$ associated with the estimates.

3. Call function `mean_exp()` with arguments `param` $= 2$ and `nb` $= 100, 1000$ and $10000$. Save the results in different variables. Compute the mean squared error for the three results and compare.

4. Draw the boxplots of these result vectors in the same graphic.

5. Comment on the graphic. Which result from probability theory explain the phenomenon?

**Exercice 3.** Write a function `mean_cauchy` similar to the function `mean_exp` with the difference that samples from the Cauchy distribution are generated. Compare the values of the mean squared error and boxplots for different sample sizes `nb`. What do you observe? Explain the phenomenon.