# EDRP Lab 1

*Version: 2020-03-28*

## Simulations and estimation - a general idea

Given a random experiment and event $A$, a **Monte Carlo estimate** of the probability $\mathbb{P}(A)$ of $A$ is obtained by repeating the random experiment many times and taking the proportion of trials in which $A$ occurs as an approximation for $\mathbb{P}(A)$.

The relative frequency interpretation of probability says that the probability of an event is the long-term proportion of times that the event occurs in repeated trials. It is justified theoretically by the strong law of large numbers.

As a simple example, consider a random experiment of flipping a fair coin 100 times and counting the number $X$ of heads.

In R, the experiment can be simulated by the following code:

```
samplesize=100
s<-sample(c("H","T"),samplesize,replace=T)
```

Here is the random sample we obtained:

```
print(s)
```

The sample is *random* - if we run the same code again, we'll get another sample:

```
samplesize=100
s<-sample(c("H","T"),samplesize,replace=T)
s
```

Based on the sample we generated, we can **estimate** the probability that a head comes up in a single toss. In order to do so, we count the number of heads we've got in 100 flips, and divide it by 100:

```
ind<-which(s=="H")
ind
l<-length(ind)
l
```

Based on the current sample, the **estimated probability** that a head comes up is

```
l/samplesize
```

We should not be surprised that the number we've got differs from the true (theoretical) value 0.5. One of the foundamental results in probability theory, the Law of Large Numbers, states that the true value is obtained when the sample size goes to infinity.

Using a computer one cannot use infinite sample sizes, but in our example we can definitely simulate a much larger number of tosses:

```
samplesize=1e+6
s<-sample(c("H","T"),samplesize,replace=T)
length(which(s=="H"))/samplesize
```

Generally speaking, larger sample sizes lead to better (more precise) estimates.

One can follow basically the same path to estimate the expected value of a random variable. Say $X$ is $-1$ if a head comes up, and 1 if the coin comes up tail (we need this kind of translation to be able to speak about the expected value since the outcomes of the random experiment are not numbers). What is $\mathbb{E}X$? Clearly,

$$\mathbb{E}X = (-1) \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0.$$

On the other hand, the **estimated expected value** is

```
samplesize=1e+6
s<-sample(c("H","T"),samplesize,replace=T)
(-1)*(length(which(s=="H"))/samplesize)+1*(length(which(s=="T"))/samplesize)
```

## Problem 1

Simulate flipping three fair coins and counting the number $X$ of heads.

1. Use your simulation to estimate $\mathbb{P}(X = 1)$ and $\mathbb{E}X$. Compare the estimates with the true values, derived from theoretical computations.
2. Modify the above to allow for a biased coin where $\mathbb{P}$(heads)=3/4.

## Simulations of Markov chains

Simulation is a powerful tool for studying Markov chains. For many chains that arise in applications, state spaces are huge and matrix methods may not be practical or even possible to implement.

A Markov chain can be simulated from an initial distribution and transition matrix. To simulate a Markov sequence $X_0, X_1, \ldots$, simulate each random variable sequentially conditional on the outcome of the previous variable. That is, first simulate $X_0$ according to the initial distribution. If $X_0 = i$, then simulate $X_1$ from the $i$-th row of the transition matrix. If $X_1 = j$, then simulate $X_2$ from the $j$-th row of the transition matrix, and so on.

### Problem 2

Consider the Markov chain with state space $S = \{1, 2\}$, transition matrix

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1 \end{bmatrix},$$

and initial distribution $\alpha = (1/2, 1/2)$.

1. Simulate 5 steps (that is, simulate $X_0, X_1, \ldots, X_5$) of the Markov chain. Carry out 100 Monte Carlo simulations (that is, repeat the simulation of the first five steps of the chain 100 times). Use the results of your simulations to solve the following problems.

Hint: to handle $6 * 100$ simulated numbers in a convenient way, you may want to use a technique based on data frames, described in the end of EDRPLab0.pdf file.

- Estimate $\mathbb{P}(X_1 = 1|X_0 = 1)$. Compare your result with the exact probability.

Hint: recall the conditional probability formula - to estimate the above conditional probability, you need to estimate two probabilities:

$$\mathbb{P}(X_0 = 1, X_1 = 1) \text{ and } \mathbb{P}(X_0 = 1).$$

- Estimate $\mathbb{P}(X_2 = 1|X_0 = 1)$. Compare your result with the exact probability.
- Estimate $\mathbb{P}(X_5 = 2|X_0 = 1, X_2 = 1)$ and $\mathbb{P}(X_5 = 2|X_2 = 1)$. Compare your result with the exact probabilities.

- Estimate $\mathbb{E}X_2$. Compare your result with the exact value.
- Estimate $\mathbb{P}(X_1 = 1, X_3 = 1)$. Compare your result with the exact probability.

2. Change the initial distribution to $\alpha = (1, 0)$ and solve the above estimation problems (or their modified versions).
3. Change the initial distribution to $\alpha = (0, 1)$ and solve the above estimation problems (or their modified versions).