# TP 1 – AOS1

## Introduction to Bayesian inference

## 1 Introduction

The practical session will leave much place to manipulating probability distributions and generating random samples, from which parameters are to be estimated. The `numpy`, `scipy`, `scipy.stats` and `matplotlib.pyplot` packages will be useful for this purpose.

```
import numpy as np
import scipy as sp
import scipy.stats as spst
import matplotlib.pyplot as plt
import itertools
```

## 2 Visualization of a pdf, generation of a random sample

First, we are interested in generating random samples according to some specific (user-defined) distribution.

(1) For the binomial and Poisson distributions:

1. pick a particular parameter value,
2. generate a sample of desired size,
3. visualize the empirical distribution of the data (using `plt.bar`) and compare it to the actual distribution (using `distrib.pmf`).

(2) For the beta, gamma, inverse gamma, exponential, Gaussian distributions:

1. pick a particular (set of) parameter value(s),
2. generate a sample of desired size,
3. visualize the empirical distribution of the data (using `plt.hist`) and compare it to the actual distribution (using `distrib.pdf`).

## 3 Maximum likelihood estimation, Bayesian updating

### 3.1 Likelihood plot

(3) Program a function `loglike` which makes it possible to compute the log-likelihood of a parameter given a sample and a family of distributions; for instance, given a random vector $\boldsymbol{X} \sim \mathcal{N}(\mu, \sigma^2)$, we would compute the log-likelihood $\ln L(\mu = 0, \sigma = 2; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ by:

```
loglike((0,2), spst.norm, x)
```

where x contains the data sample. You will take care of the fact that for multivariate distributions, instances are to be stored row-wise in x.

4 Write a script which plots the likelihood for a set of parameter values. In the case of two parameters, the level curves will be displayed.

### 3.2 Bayesian prior and Bayesian updating

Eventually, consider a very simple case of a binomial random variable

$$X \sim \mathcal{B}(n, \theta),$$

with $n$ fixed, and with $\theta$ unknown and to be estimated from the observation $x = 40$. Assume that a Bayesian prior is available:

$$\pi_\theta(t|\alpha, \beta) = \mathrm{beta}(t; \alpha, \beta) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{\mathrm{B}(\alpha, \beta)},$$

with

$$\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The questions of why we choose the beta distribution, and how we select the prior parameters $\alpha$ and $\beta$, will not be addressed for now.

5 Show that the posterior distribution of $\theta|x$ is a beta distribution $\mathrm{beta}(\alpha + x, \beta + n - x)$. For this purpose, we recall that the Gamma function is the "generalization" of the factorial to complex numbers; we have

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} \exp(-t) dt, \qquad \Gamma(z+1) = z\Gamma(z), \quad \Gamma(n) = (n-1)! \forall n \in \mathbb{N}^*.$$

6 For various parameter values for $\alpha > 0$ and $\beta > 0$, plot the prior $\pi_\theta(t|\alpha, \beta)$, the likelihood function $L(t|x)$ and the posterior $\pi_\theta(t|x; \alpha, \beta)$. What can you say about the prior distribution influencing the inference on $\theta$ ?