TP 1 – AOS1

Introduction to Bayesian inference

1 Introduction

In this practical session, we will use the `numpy`, `scipy`, and `matplotlib` packages.

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

2 Maximum likelihood estimation

2.1 Random sample generation

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

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

(2) Do the same for the beta, gamma, exponential and Gaussian distributions. You may display the empirical distribution of the sample using using `plt.hist` and the actual (theoretical) distribution using using `distrib.pdf`.

2.2 Likelihood plot

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

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

where `x` contains the data sample. Beware that for multivariate distributions, instances are by convention stored row-wise in `x`.

(4) Write a script which plots the likelihood for a single parameter. Using some of the previous distributions, locate the maximum likelihood estimate of the parameter. What do you notice when you increase the size of the sample ?

5 Write a script which plots the level curves of the likelihood for a couple of parameter values. Using the Gaussian distribution, locate the maximum likelihood estimate of the parameter vector (μ, σ^2). What do you notice when you increase the size of the sample ?

3 Bayesian updating using conjugate priors

3.1 Beta-binomial distribution

6 Recall the expression of the beta distribution. What is its definition domain ? On which parameters does it depend ? What are the expectation, the mode, and the variance ? Which particular distribution can be retrieved as a special case of the beta distribution ?

7 Write a script which plots the prior distribution given a set of chosen hyper-parameter values.

8 Consider a random variable X following a binomial sampling distribution $\mathcal{B}(n, \theta)$, with n known and $\theta \sim \text{beta}(\alpha, \beta)$. Compute the posterior distribution of θ given x, α and β.

9 Plot $\pi_\theta(\cdot)$, $L(\cdot|x)$ and $\pi_\theta(\cdot|x; \alpha, \beta)$.

10 Let us now assume that we have previously observed $X = x$ positive outcomes out of n outcomes. What is the predictive distribution for the number X_0 of positive outcomes out of n_0 new experiments, given $X = x$ out of n, α and β ?

3.2 Gamma-Poisson distribution

11 Recall the expression of the gamma distribution, its definition domain, the parameters on which it depends. Recall its expectation, mode, and variance.
Plot the prior distribution given a set of chosen hyper-parameter values.

12 Consider a random variable $X \sim \mathcal{P}(\theta)$, with $\theta \sim \text{gamma}(\alpha, \beta)$. What is the posterior distribution of θ given an iid sample $x_1, \ldots, x_n$, α and β ?
Plot $\pi_\theta(\cdot)$, $L(\cdot|x_1, \ldots, x_n)$ and $\pi_\theta(\cdot|x_1, \ldots, x_n; \alpha, \beta)$, for various values of θ, α, β and n.

13 Assume that an iid sample $x_1, \ldots, x_n$ of realizations of $X \sim \mathcal{P}(\theta)$ has been observed. Show that the predictive distribution of a new outcome x_0 given the sample, α and β is a negative binomial (or Pólya) distribution. At some point, you may want to make a change of integration variable, by replacing t with $z = (\beta + n + 1)t$.

3.3 Normal-gamma Gaussian distribution

We now consider a Gaussian random variable $X \sim \mathcal{N}(\mu, \lambda^{-1})$, where the Gaussian distribution is parameterized using the expectation μ and the *precision* $\lambda = 1/(\sigma^2)$.
Classically, a normal-gamma prior is used for parameters μ and λ:

$$\pi_\lambda(\ell|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \ell^{\alpha-1} \exp(-\beta \ell), \qquad \text{for } \ell > 0;$$
$$\pi_{\mu|\lambda}(u|\nu, \lambda, \eta) = (2\,\pi)^{-1/2} \lambda^{1/2} \exp\left(-\frac{\eta\lambda}{2}(t-\nu)^2\right), \qquad \text{for } t \in \mathbb{R}.$$

The parameter η is called the *shrinkage* parameter of the normal prior.

(14) Compute the pdf of the normal-gamma prior, i.e. the joint pdf $\pi_{\mu,\lambda}(u, \ell)$.
Display the contour plot of the normal-gamma prior for various values of α, β, ν and η.

(15) Assume that we have observed an iid sample $x_1, \ldots, x_n$ of realizations of a random variable $X \sim \mathcal{N}(\mu, \lambda^{-1})$. Recall the expression for the likelihood function $L(\mu, \lambda)$.

(16) Show that the posterior distribution for (μ, σ^2) given the sample, λ, α and β is a normal-gamma distribution. You may drop the computation of the denominator (normalization constant). Display the prior, likelihood, and posterior contours, for various values of n.