HOLON INSTITUTE OF TECHNOLOGY, DEPARTMENT OF APPLIED MATHEMATICS

Sequential Models in Data Science
Bayesian Inference

Yirmeyahu Kaminski

INTRODUCTION

This exercise aims to give an overall understanding of Bayesian Inference to implement it in Python using a simple Monte-Carlo scheme.
The exercise is to be done with pairs of students. Each pair must present a jupyter notebook with both the theoretical answers and the Python computations and functions. The notebook will be tested as this. No modification should be necessary for making it work.

1 POISSON DISTRIBUTION

From Wikipedia:
In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.
The Poisson distribution is defined by the following expression for $X \sim \mathscr{P}(\lambda)$:

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Draw $n = 100$ samples from a Poisson distribution, using *numpy.random.poisson(lam=5,size=100)*.

2 CONJUGATE PRIOR

Given $X_1, \cdots, X_n \overset{iid}{\sim} \mathscr{P}(\lambda)$, one wants to define a conjugate prior for λ.

1. Write the Poisson distribution as an exponential family.
2. Find the conjugate prior relying on the presentation as an exponential family.
3. Check that the prior you found is a Gamma distribution $\Gamma(\alpha,\beta)$, defined by the following density function: $\pi(\lambda|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$, with $\alpha,\beta > 0$. The Gamma function is defined by $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$, for any complex number z with $\Re(z) > 0$.
4. Let $\mathbf{X} = (X_1,\cdots,X_n)$. Write the posterior distribution of λ as a function of $\mathbf{X}$ and the hyper parameters α,β.

3 Marginalizing the Hyper Parameters

In this section, we want to compute the posterior distribution of λ with no hyper parameters. For this purpose, we write:

$$f(\lambda|\mathbf{X}) = \int f(\lambda,\alpha,\beta|\mathbf{X})d\alpha d\beta = \int f(\lambda|\mathbf{X},\alpha,\beta)\pi_{\alpha,\beta}(\alpha,\beta)d\alpha d\beta. \tag{3.1}$$

1. Prove the Fisher information matrix of the Gamma distribution $\Gamma(\alpha,\beta)$ is

   $$I(\alpha,\beta) = \begin{bmatrix} \psi'(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix},$$

   where $\psi(x) = \frac{d}{dx}\ln(\Gamma(x))$ is the *digamma* function.
2. Compute the Jeffrey's prior for (α,β).
3. Write explicitly the integral (3.1).

4 Maximum Likelihood Type II

In this section, we will compute an estimate of α and β by maximum likelihood estimation.

1. Compute the likelihood of α and β, that is $f(\mathbf{X}|\alpha,\beta)$ by marginalization with respect to λ.
2. Find $\alpha^*,\beta^* = \underset{\alpha,\beta}{\operatorname{argmax}} f(\mathbf{X}|\alpha,\beta)$. For this purpose, show that $\beta^* = \frac{n}{s}\alpha^*$. where $s = \sum_{i=1}^n X_i$, and that α^* is solution of

   $$\ln\left(\frac{\alpha^*}{\alpha^*+s}\right) + \sum_{j=0}^{s-1}\frac{1}{\alpha^*+j} = 0, \tag{4.1}$$

   where $s = \sum_{i=1}^n X_i \in \mathbb{N}$.

   Hint: In your computation, you will have to use that $\Gamma(z+1) = z\Gamma(z)$.

   Considering the function $\phi(\alpha) = \ln\left(\frac{\alpha}{\alpha+s}\right) + \sum_{j=0}^{s-1}\frac{1}{\alpha+j}$, show that actually $\alpha^* = +\infty$.

   Therefore in the sequel, we shall take $\alpha^* = 1000$ and $\beta^* = \frac{n}{s}1000$.

5 Predictive Posterior Distribution

Compute the Predictive Posterior Distribution computed with the estimated values of α and β, that is with α^*, β^*:

$$f(x|\mathbf{X}) = \int f(x, \lambda|\mathbf{X}) d\lambda = \int f(x|\lambda) f(\lambda|\mathbf{X}) d\lambda = \int f(x|\lambda) f(\lambda|\mathbf{X}, \alpha^*, \beta^*) d\lambda$$

6 Monte-Carlo Computations

We want to compute the expectation of this distribution:

$$\mu = E(x) = \int x f(x|\lambda) f(\lambda|\mathbf{X}, \alpha^*, \beta^*) d\lambda dx,$$

which represents a good prediction of the value of x given the observations $X_1, \cdots, X_n$.
Since we also want an estimate of the uncertainty, we shall also compute the variance of it:

$$\sigma^2 = Var(x) = \int (x - \mu)^2 f(x|\lambda) f(\lambda|\mathbf{X}, \alpha^*, \beta^*) d\lambda dx.$$

In order to compute these two integrals, we shall use a Monte-Carlo approach.
Implement the MC algorithm and compute the estimations of both μ and σ^2.