Math 459: Bayesian Statistics – Homework 1

1. Suppose you have an i.i.d. sample $X_1, \dots, X_n$ from a Poisson distribution with parameter λ, i.e.

$$\Pr(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \lambda > 0, \quad x = 0, 1, \dots$$

(a) Find the MLE for λ.

(b) Find the posterior mean, assuming a gamma prior, Gamma(α, β), with hyperparameters α and β, i.e.

$$p(\lambda|\alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)}, \quad \lambda, \alpha, \beta > 0.$$

To be clear, you should view α and β as fixed; treat them as constants in your calculations.

(c) Using **R**, generate a sample of size 40 from a Poi$(\lambda = 6)$ distribution. Find the MLE using the `optim` function in the **R** package `stats`; submit the **R** code.

(d) Find the posterior mode for hyperparameter values $\alpha = 2$, $\beta = 3$, again using a sample of size 40 from a Poi$(\lambda = 6)$ distribution. Do not use a package which does this for you; either write your own function or use the `optim` function. Submit the **R** code.

Solution.

(a) Let $X_1, \dots, X_n$ be an i.i.d. collection of sample from Poisson(λ). The likelihood function will thus be

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^n P(x_i|\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}.$$

Then the log-likelihood is (ignoring the constant terms that do not depend on λ)

$$\ell(\lambda|\mathbf{x}) = \log(L(\lambda|\mathbf{x})) \propto \sum_{i=1}^n x_i \log(\lambda) - \lambda n.$$

The maximum likelihood estimate (MLE) of λ (called $\hat{\lambda}$ hereafter) can be found by the first order condition for the log-likelihood function,

$$0 = \frac{d\ell}{d\lambda} \propto \sum_{i=1}^n x_i \frac{1}{\lambda} - n.$$

$$\Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}.$$

Note, taking the second derivative gives

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\lambda^{-2} \sum_{i=1}^n x_i < 0,$$

hence we conclude that $\hat{\lambda}$ maximizes the likelihood function. Therefore, for a Poisson sample, the MLE for is just the sample mean. $\square$

(b) Assuming a gamma prior, by Bayes' Rule the posterior

$$\pi(\lambda|\mathbf{x}) \propto \left(\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}\right)\left(\frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)}\right) \propto \lambda^{\sum_{i=1}^n x_i+\alpha-1} e^{-\lambda(n+\beta)}.$$

This is the form of Gamma distribution. So we can conclude that the posterior distribution is Gamma$\left(\sum_{i=1}^n x_i + \alpha, n+\beta\right)$ (Conjugate). Since the mean of Gamma distribution Gamma(α, β) is α/β, we have the posterior mean,

$$\frac{\sum_{i=1}^n x_i + \alpha}{n+\beta} = \left[\frac{n}{n+\beta}\right]\left(\frac{\sum_{i=1}^n x_i}{n}\right) + \left[\frac{\beta}{n+\beta}\right]\left(\frac{\alpha}{\beta}\right).$$

Remark: in the posterior mean, the data get weighted more heavily as $n \to +\infty$. $\square$

(c)

```r
library(stats)
set.seed(1234)
sample <- rpois(40, 6)
poisson.lik<-function(mu,y){
  n<-length(y)
  logl<-sum(y)*log(mu)-n*mu
  return(-logl)
}
optim(1, poisson.lik, y=sample, method="BFGS")
```

```
## $par
## [1] 5.75
##
## $value
## [1] -172.316
##
## $counts
## function gradient
##       21        9
##
## $convergence
## [1] 0
##
## $message
## NULL
```

Figure 1: R Code for 1.(c)

(d) The mode for Gamma distribution Gamma(α, β) is $\frac{\alpha-1}{\beta}$, $\alpha \geq 1$. $\square$

2. Consider the Gaussian model from Lecture 3. Assume $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Find the marginal posterior for μ, assuming the prior

$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$$

where $p(\sigma^{-2}) = e^{-\sigma^{-2}}, \quad \sigma^{-2} > 0$. You will need to use a change of variables to find $p(\sigma^2)$. Also, assume that $p(\mu|\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\{-\mu^2/(2\sigma^2)\}$. You may ignore constants of proportionality. Can you identify which type of distribution this marginal posterior density represents?

Solution.

Since $p(\sigma^{-2}) = e^{-\sigma^{-2}}$, $\sigma^{-2} > 0$, we know $p(\sigma) = e^{-\sigma}$, $\sigma > 0$. By change-of-variables argument, we know $p(\sigma^2) = \sigma^{-4}e^{-\sigma^{-2}}$, $\sigma > 0$. Then, the prior

$$\begin{aligned} p(\mu, \sigma^2) &= p(\mu|\sigma^2)p(\sigma^2) \\ &= (2\pi\sigma^2)^{-1/2}\exp\{-\mu^2/(2\sigma^2)\}\frac{1}{\sigma^4}\exp\{-\frac{1}{\sigma^2}\} \\ &\propto \sigma^{-5}\exp\left\{-\frac{\mu^2+2}{2\sigma^2}\right\} \end{aligned}$$

We start with the most basic likelihood function for assumed normally distributed data, and rearrange it slightly:

$$\begin{aligned} L(\mu, \sigma^2|\mathbf{x}) &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\bar{x})-(\mu-\bar{x})]^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 - \underbrace{2\sum_{i=1}^{n}(x_i\mu - x_i\bar{x} - \bar{x}\mu + \bar{x}^2)}_{=0} + n(\mu-\bar{x})^2\right)\right\} \\ &\propto \sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\mu-\bar{x})^2\right)\right\} \end{aligned}$$

where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

is the sample variance. The resulting joint posterior distribution for (μ, σ^2) is then

$$\begin{aligned}\pi(\mu, \sigma^2|\mathbf{x}) &\propto L(\mu, \sigma^2|\mathbf{x})p(\mu, \sigma^2) \\ &\propto \sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\Big((n-1)s^2 + n(\mu-\bar{x})^2\Big)\right\}\sigma^{-5}\exp\left\{-\frac{\mu^2+2}{2\sigma^2}\right\} \\ &= (\sigma^2)^{-\frac{n+5}{2}}\exp\left\{-\frac{1}{2\sigma^2}\Big((n-1)s^2 + n(\mu-\bar{x})^2 - \mu^2 - 2\Big)\right\}\end{aligned}$$

The marginal posterior distribution for μ can be obtained by integrating σ^2 out of the joint posterior distribution:

$$\begin{aligned}\pi(\mu|\mathbf{x}) &= \int_0^\infty \pi(\mu, \sigma^2|\mathbf{x})d\sigma^2 \\ &\propto \int_0^\infty (\sigma^2)^{-\frac{n+5}{2}}\exp\left\{-\frac{1}{2\sigma^2}\Big((n-1)s^2 + n(\mu-\bar{x})^2 - \mu^2 - 2\Big)\right\}d\sigma^2\end{aligned}$$

This integral can be evaluated using the substitution

$$z = \frac{A}{2\sigma^2},$$

$$A = (n-1)s^2 + n(\mu-\bar{x})^2 - \mu^2 - 2.$$

Then,

$$\pi(\mu|\mathbf{x}) \propto A^{-\frac{n+3}{2}}\int_0^\infty z^{\frac{n+1}{2}}e^{-z}dz$$

This integrand is the inverse gamma density and thus the integral is a constant. Therefore,

$$\pi(\mu|\mathbf{x}) \propto A^{-\frac{n+3}{2}} = [(n-1)s^2 + n(\mu-\bar{x})^2 - \mu^2 - 2]^{-\frac{n+3}{2}}$$

which is a scaled non-central t distribution.

□

3. Suppose you agree with a Bayesian that we should quantify our prior beliefs by specifying a prior density on the unknown parameter, and a sampling model to describe the data generating process. However, you are not convinced that Bayes's Rule is the right way to update your beliefs after observing new information. Suggest an alternative rule, with a brief justification, and give a mathematical (probabilistic) expression for this rule.

Solution.

The joint probability mass or density function can be written as a product of two densities that are often referred to as the prior distribution $p(\theta)$ and the sampling distribution (or data distribution) $p(y|\theta)$, respectively:

$$p(\theta, y) = p(\theta)p(y|\theta).$$

Simply conditioning on the known value of the data y, using the basic property of conditional probability known as *Bayes' rule*, yields the posterior density:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}.$$

From this *Bayes' Rule*, we find the information updating process heavily depends on the prior distribution $p(\theta)$. And this prior belief is not necessarily a true reflection of the underlining *nature*. Hence, one reasonable alternative rule to update beliefs after observing new information is to consider multiple possible priors, $p_1(\theta)$,..., $p_n(\theta)$, with corresponding probability a_1,..., a_n. Then, the *Bayes' Rule* can be rewritten as

$$p(\theta|y) = \sum_{i=1}^{n} a_i \frac{p_i(\theta)p_i(y|\theta)}{p_i(y)}.$$

□