QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Practical 4

2023-2024

Conjugate priors: Exponential likelihood/gamma prior example

One of the datasets available in R is called `faithful`, containing data on eruptions of the Old Faithful geyser. The waiting time between successive eruptions can be accessed as

```
faithful$waiting
```

You could store this column of the dataset in a new vector to simplify the later code

```
t = faithful$waiting
```

We assume that these values follow an exponential distribution, with parameter λ. Calculate and display the MLE for λ. Recall from lectures week 1 the MLE for λ is

$$\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n} t_i},$$

where $t_1, \ldots t_n$, are the observed waiting time between successive eruptions. To calculate formulae such as this, if you have a vector `v`, then `sum(v)` is the sum of the elements, and `length(v)` is the number of elements.

We saw that a gamma distribution is conjugate to the exponential likelihood. The help files for the R commands for the gamma distribution may be found using

```
?dgamma
```

There are two common ways of specifying the parameters for the gamma distribution. The form used in this module is

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)},\ x > 0$$

The mean of the gamma distribution is α/β. In R, the pdf for this distribution is given by the following

```
dgamma(x, shape=alpha, rate=beta)
```

Calculate the parameters for the posterior gamma distribution when combining a gamma prior distribution with the likelihood based on the Old Faithful data. Do this for more than one prior distribution. For example, try $\alpha = 1, \beta = 1$ and then $\alpha = 200, \beta = 200$. These both have the same mean, but different variances.

Then calculate the posterior mean for λ using each prior distribution. Which is more similar to the MLE?

Generating a sample from the posterior density

Later in the module, we will look at Markov Chain Monte Carlo methods, the main computational methods for Bayesian analysis of data. These methods work by generating a sample of parameters from the posterior distribution. For now, we generate a sample even though we have an exact formula for the posterior, to understand how we might use such a sample.

In the exponential example the posterior is a gamma distribution. The following command generates a random sample of size NS from the gamma(α, β) distribution and stores the sample in a vector called post_lambda.

```
post_lambda = rgamma(NS, shape=alpha, rate=beta)
```

Run this command with the appropriate shape=alpha, rate=beta parameters (i.e those for the posterior distribution). Take NS to be a large integer such as $10,000$.

Now we have a sample from the posterior distribution, we can summarize it. The commands mean and median return the mean and median of a vector, in this case post_lambda. These will give approximate values for the posterior mean and median: they are approximate as they are based on a random sample, but since it is a large sample, they should be similar to the exact values you calculated earlier.

Suppose that we are interested in the posterior probability that $\lambda > 0.05$. Calculate this probability from the exact posterior distribution using the pgamma function. Also, if you can figure out how, calculate an approximation to this probability using the random sample post_lambda.