Jeffreys' prior: definition

Let us first give the general deffinition of the Jeffreys' prior $p_J(\theta)$:

$$p_J(\theta) = \sqrt{\mathcal{I}_n(\theta)}$$

where $\mathcal{I}_n(\theta)$ is the Fisher information of the sample. The Fisher information is defined as follows:

$$\mathcal{I}_n(\theta) = -E_\theta\left[\frac{\partial^2 ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta^2}\right]$$

or equivalently

$$\mathcal{I}_n(\theta) = var_\theta\left(\frac{\partial ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta}\right) = E_\theta\left[\left(\frac{\partial ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta}\right)^2\right]$$

The first derivative of the log-likelihood function with respect to the model parameter $\frac{\partial ln\mathcal{L}(\theta|\mathbf{x})}{\partial\theta}$ is sometimes referred to as the score function.

Likelihood and derived functions for a Poisson model

Likelihood and log-likelihood, score function and second derivative of the log-likelihood function for a Poisson model:

$$\mathcal{L}(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$

$$l(p \mid \mathbf{x}) = ln\big(\mathcal{L}(\theta \mid \mathbf{x})\big) = \sum_{i=1}^{n} x_i \; ln(\theta) + -n\theta - \sum_{i=1}^{n} log(x_i!)$$

$$\frac{\partial ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial \theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - n$$

$$\frac{\partial^2 ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial \theta^2} = -\frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

Fisher information and Jeffrey's prior for a Poisson model 1/2

So we want

$$\mathcal{I}_n(\theta) = -E\left[\frac{\partial^2 ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta^2}\right]$$

and we note that

$$E\left[\sum_{i=1}^{n} x_i\right] = E[n\overline{x}] = nE[\overline{x}] = n\theta$$

Fisher information and Jeffrey's prior for a Poisson model 2/2

Thus we have that

$$\begin{aligned}\mathcal{I}_n(\theta) &= \frac{E\left[\sum_{i=1}^{n} x_i\right]}{\theta^2} \\ &= \frac{n\theta}{\theta^2} \\ &= \frac{n}{\theta} \\ &\propto \theta^{-1}\end{aligned}$$

And we conclude that

$$p_J(\theta) = \sqrt{\mathcal{I}(\theta)} \propto \theta^{-1/2}$$

which is the Gamma distribution $Gamma(1/2, 0)$ (improper prior)

Example of application to water quality assessment

Suppose that we record the number of a specific bacteria present in 10 water samples taken in the Mekon delta so that we have the following data at hand

$$x_i = 2, 1, 8, 0, 1, 4, 2, 3, 0, 7$$

Assuming a Poisson likelihood for the data and using the Jeffreys' prior, what is the posterior mean and the 95% credible intreval for the model parameter ?

Example of application to water quality assessment

From the Bayes Theorem, we know that

$$p(\theta \mid x) = \frac{p(x \mid \theta)\ p(\theta)}{p(x)} \propto p(x \mid \theta)\ p(\theta)$$

In our case, we have that

$$\begin{aligned} p(\theta \mid x) &\propto \underbrace{\theta^{\sum_{i=1}^{10} x_i} e^{-10\theta}}_{\text{Poisson Likelihood}}\ \underbrace{p(\theta)}_{\text{Prior}} \\ &\propto \theta^{28} e^{-10\theta}\ \theta^{-1/5} \end{aligned}$$

and we recognize the functional form of a Gamma density, that is $Gamma(\alpha = 28.5, \beta = 10)$. The posterior mean is just given by α/β.

Full code in R language

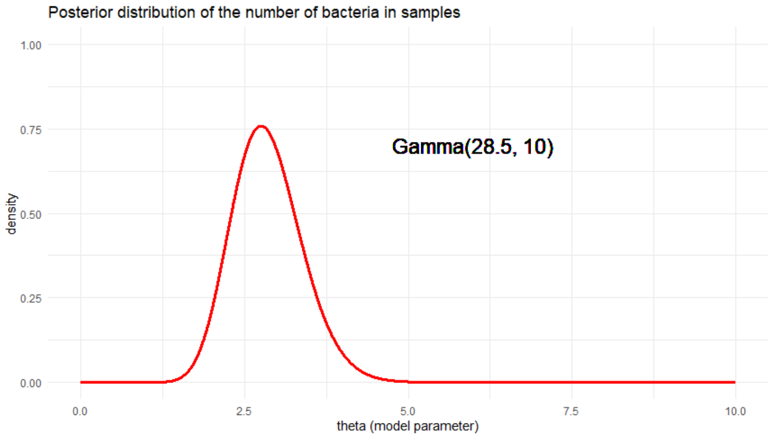
```r
# posterior mean (alpha/beta)
28.5/10
# [1] 2.85

# 95% credible interval
qgamma(c(0.025, 0.975), 28.5, 10)
# [1] 1.901337 3.987610

# plot of the posterior distribution
sequence.grid = seq(from = 0, to = 10, by = 0.001)
density = dgamma(sequence.grid, 28.5, 10)

library(ggplot2)
data.frame = data.frame(sequence.grid, density)
ggplot(data.frame, aes(sequence.grid)) +
  geom_line(aes(y = density), colour = 'red', size = 1.2) +
  xlab('theta (model parameter)') +
  ylab('density') +
  xlim(0, 10) + ylim(0, 1) +
  geom_text(x = 6, y = 0.7, label = 'Gamma(28.5, 10)', size = 6) +
  ggtitle('Posterior distribution of the number of bacteria in samples') +
  theme_minimal()
```

Plot of the posterior distribution



Bayesian Questionnaire: Question 4

Please give a probability of correctness (a number between 0 and 100%) to all of the four possible answers. If you are completely sure of your answer, you can give 100% to the correct answer and 0 everywhere else.

Suppose that we record the number of a specific bacteria present in 10 water samples taken in the Mekong delta so that we have the following data:

$$x_i = 2, 2, 5, 2, 5, 9, 8, 3, 0, 8$$

If we assume a Poisson likelihood for the data and use the Jeffreys' prior, what is the posterior distribution of the model parameter and a 95% credible interval for the model parameter ?

A: Gamma(44.5, 10) (3.2397, 5.8495)
B: Gamma(45.5, 10) (2.2397, 4.8495)
C: Poisson(10) (4.8247, 5.8532)
D: Poisson (10) (7.8247, 9.8532)