ST451 Bayesian Machine Learning
Kostas Kalogeropoulos

Exercises

1. A company produces big magnetic roll tapes and wants to examine the number of defect spots per meter of tape. An experiment was conducted in which one meter of the tape was selected randomly and the number of defect spots was recorded. The procedure was repeated 5 times independently to provide the data $y = (2, 2, 6, 0, 3)$. Assume that these observations are independent and distributed according to the Poisson(λ) distribution.

   (a) Find the Jeffrey's prior for this model and derive the corresponding posterior distribution. Provide the mean, variance and a 95% credible interval for the number of defect spots per tape meter taking into account the data from the experiment.

   Answer: The likelihood

   $$f(y|\lambda) = \prod_{i=1}^{n} \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!} \propto \exp(-n\lambda)\lambda^{\sum y_i}$$

   To find the Jeffreys prior we follow the steps below

   $$\log f(y|\lambda) = \log\left[\exp(-n\lambda)\lambda^{\sum y_i}\right] = -n\lambda + \log(\lambda)\sum_{i=1}^{n} y_i,$$

   $$\frac{\partial}{\partial\lambda}\log f(y|\lambda) = \frac{\sum_{i=1}^{n} y_i}{\lambda} - n.$$

   $$\mathcal{I}(\lambda|y) = -E_Y\left(\frac{\partial^2}{\partial\lambda^2}\log f(y|\lambda)\right) = -E_Y\left(-\frac{\sum_{i=1}^{n} y_i}{\lambda^2}\right) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

   Hence the Jeffreys prior: $\pi(\lambda) \propto 1/\sqrt{\lambda}$.
   The posterior is then

   $$\begin{aligned}\pi(\lambda|y) &\propto f(y|\lambda)\pi(\lambda) \propto \exp(-n\lambda)\lambda^{\sum y_i}\lambda^{-1/2} \\ &= \lambda^{1/2+\sum y_i - 1}\exp(-n\lambda) \overset{\mathcal{D}}{=} \text{Gamma}(1/2 + \sum y_i, n)\end{aligned}$$

   In our case $\sum y_i$ and $n = 5$ so the posterior distribution is the Gamma$(13.5, 5)$ distribution, implying that the posterior mean is 2.7 and the posterior variance is 0.54. A 95% credible interval is provided by the 2.5th and 97.5th percentiles of the Gamma$(13.5, 5)$ distribution that are 1.46 and 4.32 respectively.

   (b) From previous experience, the beliefs of the researcher about λ can be expressed by a Gamma distribution with mean and variance equal to 3. Repeat the previous part with this prior distribution and compare the results.

   Answer: Following the derivation of the lecture slides of week 1, we get a posterior Gamma$(\alpha + \sum y_i, n + \beta)$.

   We know that the prior mean and variance are equal to 3, meaning $\frac{\alpha}{\beta} = 3$, $\frac{\alpha}{\beta^2} = 3$, which implies that $\alpha = 3$, $\beta = 1$. Also $\sum y_i = 2+2+6+0+3 = 13$. Hence the posterior distribution is a Gamma$(16, 6)$ distribution, implying that the posterior mean is $8/3$. This is slightly lower than the prior mean of 3 but higher than what the data suggest as the MLE $\bar{y} = 2.6$ and quite close the result of the previous part. The posterior variance $4/9$ which is lower than the prior variance of 3 reflecting the fact that our uncertainty decreased in light of the observed data y.

The posterior variance is also smaller than that of the previous part as we incorporated prior information. A 95% credible interval is provided by the 2.5th and 97.5th percentiles of the Gamma(16, 6) distribution that are 1.51 and 4.10 respectively. Here the interval is narrower than that of the previous part.

2. Let $y = (y_1, \dots, y_n)$ be a random sample from the Exponential(λ) distribution

$$f(y_i|\lambda) = \lambda \exp(-\lambda y_i), \quad y_i > 0. \;\; \lambda > 0.$$

Set the prior for λ to be a Gamma(α, β).

(a) Derive the posterior distribution and find a Bayes estimator for λ. Also derive the predictive distribution for a new observation y_n (assuming that it also follows the same Exponential(λ) model).

Answer: The joint density (likelihood) can be written as

$$f(y|\lambda) = f(y_1, \dots, y_n|\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda y_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} y_i\right)$$

The prior is set to a Gamma(α, β), so we can write

$$\pi(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda)$$

The posterior is then proportional to

$$\begin{aligned}
\pi(\lambda|y) \quad &\propto f(y|\lambda)\pi(\lambda) \propto \lambda^n \exp\left(-\lambda \textstyle\sum_{i=1}^{n} y_i\right) \lambda^{\alpha-1} \exp(-\beta\lambda) \\
&= \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\beta + \textstyle\sum_{i=1}^{n} y_i\right)\right] \\
&\stackrel{\mathcal{D}}{=} \text{Gamma}\left(n+\alpha, \beta + \textstyle\sum_{i=1}^{n} y_i\right)
\end{aligned}$$

A Bayes estimator is provided by the posterior mean who in this case is equal to

$$\frac{n+\alpha}{\beta + \sum_{i=1}^{n} y_i}.$$

The predictive distribution (for $y > 0$) is

$$\begin{aligned}
f(y|x) &= \int \lambda \exp(-\lambda y) \frac{(n\bar{x}+\beta)^{n+\alpha}}{\Gamma(n+\alpha)} \lambda^{n+\alpha-1} \exp(-(n\bar{x}+\beta)\lambda) d\lambda \\
&= \frac{(n\bar{x}+\beta)^{n+\alpha}}{\Gamma(n+\alpha)} \int \lambda^{n+\alpha+1-1} \exp(-(n\bar{x}+\beta+y)\lambda) d\lambda \\
&= \frac{(n\bar{x}+\beta)^{n+\alpha}}{\Gamma(n+\alpha)} \frac{\Gamma(n+\alpha+1)}{(n\bar{x}+\beta+y)^{n+\alpha+1}}
\end{aligned}$$

(b) **[Computer Exercise]** The data y are given below

11.8, 6.7, 26.8, 5.0, 5.4, 9.7, 1.6, 8.6, 15.8, 10.8, 32.8, 12.9, 0.2, 0.1, 6.4, 0.3, 15.2, 3.6, 18.4, 2.1.

In a similar manner to activity 3 of the computer class assess the performance of the predictive distribution via the steps below

i. Split the data into a training set (first 10 observations) and a test set (remaining observations).

ii. Specify the predictive distribution via the training set and use it to obtain point forecasts for the test set.

iii. Repeat the previous step but with 95% prediction intervals rather than point forecasts.

Note that the Exponential(λ) is the same as the Gamma($1, \lambda$).

Answer: See jupyter notebook file 'Exercise3b.ipynb'