Bertil Wegmann
Div of Statistics and Machine Learning
Dept of Computer and Information Science
Linköping University

Bayesian Learning

Computer Lab 1

You are recommended to use R for solving the labs.
You work and submit your labs in pairs, but both of you should contribute equally and understand all parts of your solutions.
`It is not allowed to share exact solutions` with other student pairs.
The submitted lab reports will be verified through OURIGINAL and indications of plagiarism will be investigated by the Disciplinary Board.
Submit your solutions via LISAM, no later than April 21 at 23:59.

Please note the following about the format of the submitted lab report:
1. The lab report should include all solutions and plots to the stated problems with necessary comments.
2. Submit the lab report with your code attached to the solution of each sub-problem (1a), 1b),...) in **one** PDF document.
3. Submit a separate file containing all code.

---

1. *Daniel Bernoulli*
   Let $y_1, ..., y_n|\theta \sim \text{Bern}(\theta)$, and assume that you have obtained a sample with $f = 35$ failures in $n = 78$ trials. Assume a $\text{Beta}(\alpha_0, \beta_0)$ prior for θ and let $\alpha_0 = \beta_0 = 7$.

   (a) Draw 10000 random values (`nDraws = 10000`) from the posterior $\theta|y \sim \text{Beta}(\alpha_0+s, \beta_0 + f)$, where $y = (y_1, \ldots, y_n)$, and verify graphically that the posterior mean $E[\theta|y]$ and standard deviation $SD[\theta|y]$ converges to the true values as the number of random draws grows large. [Hint: `use rbeta()` to draw random values and make graphs of the sample means and standard deviations of θ as a function of the accumulating number of drawn values].

   (b) Draw 10000 random values from the posterior to compute the posterior probability
   $\Pr(\theta > 0.5|y)$ and compare with the exact value from the Beta posterior. [Hint: `use pbeta()` to calculate the exact value].

   (c) Draw 10000 random values from the posterior of the odds $\phi = \frac{\theta}{1-\theta}$ by using the previous random draws from the Beta posterior for θ and plot the posterior distribution of ϕ. [Hint: `hist()` and `density()` can be utilized].

2. *Log-normal distribution and the Gini coefficient.*
   Assume that you have asked 8 randomly selected persons about their monthly income (in thousands Swedish Krona) and obtained the following eight observations:

22, 33, 31, 49, 65, 78, 17, 24. A common model for non-negative continuous variables is the log-normal distribution. The log-normal distribution $\log \mathcal{N}(\mu, \sigma^2)$ has density function

$$p(y|\mu, \sigma^2) = \frac{1}{y \cdot \sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\log y - \mu)^2\right],$$

where $y > 0$, $-\infty < \mu < \infty$ and $\sigma^2 > 0$. The log-normal distribution is related to the normal distribution as follows: if $y \sim \log \mathcal{N}(\mu, \sigma^2)$ then $\log y \sim \mathcal{N}(\mu, \sigma^2)$. Let $y_1, ..., y_n|\mu, \sigma^2 \overset{iid}{\sim} \log \mathcal{N}(\mu, \sigma^2)$, where $\mu = 3.65$ is assumed to be known but σ^2 is unknown with non-informative prior $p(\sigma^2) \propto 1/\sigma^2$. The posterior for σ^2 is the scaled inverse chi-squared distribution, Scale-inv-$\chi^2(n, \tau^2)$, where

$$\tau^2 = \frac{\sum_{i=1}^{n}(\log y_i - \mu)^2}{n}.$$

(a) Draw 10000 random values from the posterior of σ^2 by assuming $\mu = 3.65$ and plot the posterior distribution.

(b) The most common measure of income inequality is the Gini coefficient, G, where $0 \leq G \leq 1$. $G = 0$ means a completely equal income distribution, whereas $G = 1$ means complete income inequality (see e.g. Wikipedia for more information about the Gini coefficient). It can be shown that $G = 2\Phi\left(\sigma/\sqrt{2}\right) - 1$ when incomes follow a $\log \mathcal{N}(\mu, \sigma^2)$ distribution. $\Phi(z)$ is the cumulative distribution function (CDF) for the standard normal distribution with mean zero and unit variance. Use the posterior draws in a) to compute the posterior distribution of the Gini coefficient G for the current data set.

(c) Use the posterior draws from b) to compute a 95% equal tail credible interval for G. A 95% equal tail credible interval (a, b) cuts off 2.5% percent of the posterior probability mass to the left of a, and 2.5% to the right of b.

(d) Use the posterior draws from b) to compute a 95% Highest Posterior Density Interval (HPDI) for G. Compare the two intervals in (c) and (d). [Hint: Use the `hdi()` function from the `bayestestR` package.]

3. *Bayesian inference for the rate parameter in the Poisson distribution.*

This exercise aims to show you that a grid approximation can be used to obtain information of the posterior distribution when the distributions for prior and posterior are not conjugate. In some cases, it is sufficient to evaluate the posterior pdf $p(\lambda|y)$ at a finite set of λ values to understand its structure and obtain important quantities.

The data points below represent the number of goals scored in each match of the opening week of the 2024 Swedish women's football league (Damallsvenskan):

$$y = (0, 2, 5, 5, 7, 1, 4).$$

We assume that these data points are independent observations from the Poisson distribution with rate parameter $\lambda > 0$ which tells us about the average goals rate in a match. Let the prior distribution of λ be the half-normal distribution with prior pdf

$$p(\lambda|\sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{\lambda^2}{2\sigma^2}\right), \ \lambda \geq 0,$$

and the scale parameter that is set to be $\sigma = 5$.

(a) Derive the expression for what the posterior pdf $p(\lambda|y,\sigma)$ is proportional to. Then, plot the posterior distribution of the average goals rate parameter λ over a fine grid of λ values. [Hint: you need to normalize the posterior pdf $p(\lambda|y,\sigma)$ so that it integrates to one.]

(b) Find the (approximate) posterior mode of λ from the information in a).

GOOD LUCK!