Matias Quiroz and Mattias Villani
Division of Statistics and Machine Learning
Dept. of Computer and Information Science
Linköpings universitet
matias.quiroz@liu.se

2016-04-05

Bayesian Learning, 6 hp

Computer lab 1

You can use any programming language for the labs, but my hints, help and solutions will be in R.
You are supposed to work and submit your labs in pairs, but do make sure that both of you are contributing. Submit your solutions by Lisam no later than **April** 19 at **midnight**.

1. *Bernoulli ... again.*
   Let $y_1, ..., y_n|\theta \sim \text{Bern}(\theta)$, and assume that you have obtained a sample with $s = 14$ successes in $n = 20$ trials. Assume a $\text{Beta}(\alpha_0, \beta_0)$ prior for θ and let $\alpha_0 = \beta_0 = 2$.

   (a) Draw random numbers from the posterior $\theta|y \sim \text{Beta}(\alpha_0 + s, \beta_0 + f)$, $y = (y_1, \ldots, y_n)$, and verify that the posterior mean and standard deviation converges to the true values as the number of random draws grows large.

   (b) Use simulation to compute the posterior probability $\Pr(\theta < 0.4|y)$ and compare with the exact value [Hint: `pbeta()`].

   (c) Compute the posterior distribution of the log-odds $\phi = \log \frac{\theta}{1-\theta}$ by simulation. [Hint: `hist()` and `density()` might come in handy]

2. *Log-normal distribution and the Gini coefficient.*
   Assume that you have asked 10 randomly selected persons about their monthly income (in thousands Swedish Krona) and obtained the following ten observations: 14, 25, 45, 25, 30, 33, 19, 50, 34 and 67. 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

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

   for $y > 0$, $\mu > 0$ 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 $\log y \sim \mathcal{N}(\mu, \sigma^2)$.

   (a) Let $y_1, ..., y_n|\mu, \sigma^2 \overset{iid}{\sim} \log \mathcal{N}(\mu, \sigma^2)$, where $\mu = 3.5$ is assumed to be known but σ^2 is unknown with non-informative prior $p(\sigma^2) \propto 1/\sigma^2$. Show analytically that the posterior for σ^2 is the $Inv - \chi^2(n, \tau^2)$ distribution, where

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

(b) Simulate 10,000 draws from the posterior of σ^2 (assuming $\mu = 3.5$) and compare with the results in a).

(c) 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 Wikipedia for more information. 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 b) to compute the posterior distribution of the Gini coefficient G for the current data set.

3. Bayesian inference for the concentration parameter in the von Mises distribution. This exercise is concerned with directional data. The data points are observed wind directions at a given location on ten different days. The data are recorded in degrees: $(40, 303, 326, 285, 296, 314, 20, 308, 299, 296)$, where North is located at zero degrees (see Figure 1 on the next page, where the angles are measured clockwise). To fit with Wikipedias description of probability distributions for circular data we convert the data into radians $-\pi \leq y \leq \pi$. The 10 observations in radians are $(-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02)$. Assume that these data points are independent observations following the von Mises distribution

$$p(y|\mu,\kappa) = \frac{\exp\left[\kappa \cdot \cos(y-\mu)\right]}{2\pi I_0(\kappa)}, \quad -\pi \leq y \leq \pi,$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind of order zero [see `?besselI` in R]. The parameter μ ($-\pi \leq \mu \leq \pi$) is the mean direction and $\kappa > 0$ is called the concentration parameter. Large κ gives a small variance around μ, and vice versa. Assume that μ is known to be 2.39. Let $\kappa \sim \text{Exponential}(\lambda = 1)$ a priori, where λ is the rate parameter of the exponential distribution (so that the mean is $1/\lambda$).

(a) Plot the posterior distribution of κ for the wind direction data over a fine grid of κ values.

(b) Find the posterior mode of κ.

4. The populations, n_i , and the number of cases, y_i , of a disease in a year in each of six districts ($m = 6$) are given in the table below.

Population n_i	Cases y_i
120342	2
235967	5
243745	3
197452	5
276935	3
157222	1

Suppose that the number y_i in a district with population n_i is a Poisson random variable with mean $n_i\lambda/100000$. The number of cases in each district is independent of the numbers in other districts, given the value of λ. Our prior distribution for λ is a Gamma distribution $\lambda \sim \text{Gamma}(a_0, b_0)$, where a_0 is its shape parameter and b_0 is the rate parameter, $a_0 > 0$, $b_0 > 0$.

(a) Show mathematically that the posterior distribution of λ is also a gamma distribution $\lambda \sim \text{Gamma}(a_m, b_m)$, where $a_m = a_0 + \sum_{i=1}^{m} y_i$, and $b_m = b_0 + \sum_{i=1}^{m} n_i/100000$.

(b) Suppose that we have prior information that λ can be expected to be 4 and that we are 50% certain that λ lies in the interval 3 to 5. Use trial and error to find parameter values for the prior that (approximately) satisfies this. [Hint: Search for a β, such that $\lambda \sim \text{Gamma}(4\beta, \beta)$ and $Pr(3 \leqslant \lambda \leqslant 5) \approx 0.5$. `pgamma()`]

(c) Use the data to update the prior information. Plot a graph showing the prior and posterior probability density functions of λ on the same axes. What is the posterior probability that λ is between 3 and 5?

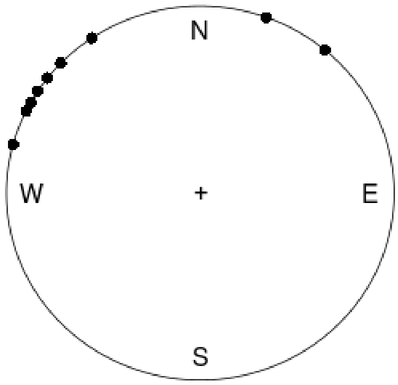

Figure 1: The wind direction data. Angles are measured clock-wise starting from North.

MAY BAYES BE WITH YOU!