Mattias Villani
Avdelning för statistik
Institutionen för datavetenskap
Linköpings universitet
mattias.villani@liu.se

2012-11-04

# 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. [Disclosure: I am actually a long time Matlab user, but I am moving more toward Free software like R and Python].
You are allowed to work and submit your labs in pairs, but do make sure that both of you are contributing.
The **deadline** for this lab is **November 13**.

1. *Bernoulli - again.*
   Let $x_1, ..., x_n|\theta \overset{iid}{\sim} Bern(\theta)$, and assume that you have obtained a sample with $s = 14$ successes in $n = 20$ trials. Assume a $Beta(\alpha, \beta)$ prior for $\theta$ and let $\alpha = \beta = 2$.

   (a) Draw random numbers from the posterior $\theta|\mathbf{x} \sim Beta(\alpha + s, \beta + f)$ 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 $Pr(\theta < 0.01|\mathbf{x})$ and compare with the exact value [Hint: `pbeta()`].

   (c) Compute the posterior distribution of the log-odds $\phi = \ln \frac{\theta}{1-\theta}$ using random draws from $\theta|\mathbf{x} \sim Beta(\alpha + s, \beta + f)$. [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 $LN(\mu, \sigma^2)$ has density function

$$p(x|\mu, \sigma^2) = \frac{1}{x \cdot \sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(\ln x - \mu)^2],$$

   where $x > 0$, $\mu > 0$ and $\sigma^2 > 0$. The log-normal distribution is related to the normal distribution as follows: if $x \sim LN(\mu, \sigma^2)$, then $\ln x \sim N(\mu, \sigma^2)$.

   (a) Let $x_1, x_2, ..., x_n|\mu, \sigma^2 \overset{iid}{\sim} LN(\mu, \sigma^2)$, where $\mu$ is assumed known and $\mu = 3.5$, but $\sigma^2$ is unknown. Assume the noninformative prior $p(\sigma^2) \propto 1/\sigma^2$. Show that the posterior for $\sigma^2$ is the $Inv - \chi^2(\nu, \tau^2)$ distribution [in Wikipedia's parametrization], where $\nu = n$ and

$$\tau^2 = \frac{\sum_{i=1}^{n} (\ln x_i - \mu)^2}{n}$$

(b) Simulate 10000 draws from the posterior of $\sigma^2$ (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 $G = 2\Phi(\sigma/\sqrt{2}) - 1$ for the special case of a $LN(\mu, \sigma)$ distribution. Use the posterior draws in b to compute the posterior distribution of the Gini coefficient $G = 2\Phi(\sigma/\sqrt{2}) - 1$ 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*: `c(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* $-\pi \leq x < \pi$. The 10 observations in radians are `c(-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

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

where $I_0(\kappa)$ is the modified Bessel function of the first kind of order 0 [see `?besselI` in R]. The parameter $\mu$ ($-\pi \leq \mu < \pi$) is the mean direction and $\kappa > 0$ is called the concentration parameter. Large $\kappa$ gives a small variance around $\mu$, and vice versa. Assume that $\mu$ is known to be 2.39. Let $\kappa \sim Exponential(\lambda = 1)$ a priori, where $\lambda$ is the rate parameter of the exponential distribution (so that the mean is $1/\lambda$). [Side remark: The packages `CircStats` and `circular` has some useful functions for circular data, but you will not need them here.]

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

(b) Find the posterior mode of $\kappa$.

MAY BAYES BE WITH YOU!

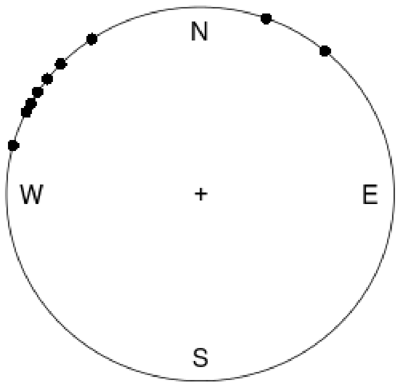


Figure 1: The wind direction data. Angles are measured clock-wise starting from North.