Bayesian Inference

Tutorial 4

1. (From May 2007 exam - number in brackets correspond to number of marks - the total was 50 marks.) Let $X_1, \dots, X_n$ be independent and identically distributed random variables such that $X_i \sim N(\mu, \omega^{-1})$ for $i = 1, \dots, n$, where ω^{-1} is known.

   (a) Define Jeffreys' prior and suggest why we may wish to specify such a prior, in general. [2]

   (b) Show that Jeffreys' prior for μ is of the form, $p(\mu) \propto 1$. [3]

   (c) Hence show that the posterior distribution for μ is also Normal, where the posterior mean and variance should be specified. [3]

   (d) Suppose that we observe data $x_1, \dots, x_{10}$ such that $\bar{x} = 10.1$. Assuming that $\omega = 1$, show that a 95% highest posterior density interval for μ is (9.480, 10.720). [4]

   (e) Now, suppose that a further observation, y, is independently generated from the same distribution as each X_i, so that,

   $$Y|\mu \sim N(\mu, \omega^{-1}).$$

   Calculate the posterior predictive distribution for Y. [5]

2. (a) Suppose an urn contains red and blue balls and we are interested in the proportion of red balls, θ. Our prior belief is that $\theta \sim Beta(9, 14)$. We extract 10 samples and 8 are red. We might ask whether the observed data is 'compatible' with the expressed prior distribution. One method is to calculate the predictive probability of observing such an extreme number of successes under this prior: this is a standard p-values but where the null hypothesis is a distribution. Using a `for` loop in `R` (and the `beta` function), find the probability of getting at least 8 red balls in 10 extractions. Are the data incompatible with the prior?

   (b) Suppose that 9 out of 10 experts suggest the $Beta(9, 14)$ prior but one expert disagrees and suggests instead that the proportion of red balls should have a prior distribution with mean 0.8 and standard deviation 0.1. What $Beta(a, b)$ distribution might represent the belief of this expert? Use the prior information from all the experts to set up a mixture prior and repeat the prior/data compatibility test. Use `R` to check whether the data are more compatible with this new prior.

   (*Note*: a mixture distribution has the form $f(x) = \sum_{i=1}^{k} w_k p_k(x)$ where $p_k(x)$ are pdfs or pmfs and w_k are the relative weights which add up to 1).

   (c) Use `R` to compute the posterior probability that θ is greater than 0.7.

3. Suppose $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma)^2$, given μ and σ, where both μ and σ^2 are unknown. The following priors were specified:

   $$\mu \sim N(0, s^2); \qquad \text{and } \sigma \sim U[0, T],$$

   where T is "large".

(a) Using the transformation of variables formula, calculate the corresponding prior on σ^2.

(b) Calculate the posterior conditional distributions $f(\mu|\boldsymbol{x}, \sigma)$ and $f(\sigma^2|\boldsymbol{x}, \mu)$, after observing data $\boldsymbol{x} = (x_1, \ldots, x_n)$.

4. Consider parameters $\theta = \{\theta_1, \ldots, \theta_p\}$ (for $p \geq 2$), with posterior density function $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$. We wish to obtain posterior summary statistics of interest using Monte Carlo integration. We have n sampled values of θ denoted $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^n$, such that $\boldsymbol{\theta}^i = \left\{\theta_1^i, \ldots, \theta_p^i\right\}$. Discuss how we would obtain estimates of the following summary statistics:

   (a) $\mathbb{E}_\pi(\theta_1)$;

   (b) $\mathrm{Var}_\pi(\theta_1)$

   (c) 95% highest posterior density interval of θ_1 (you can assume that the distribution is unimodal);

   (d) $\mathbb{P}_\pi(\theta_1 > 0)$;

   (e) Posterior correlation between θ_1 and θ_2.

5. Use Monte Carlo integration in `R` to compute the following integral:

$$\int_{-1}^{1} 2\sqrt{1 - x^2}dx = \pi$$

Run the code multiple times to see how the estimates improve with increasing sample size N (you could use $N = 10^i$ and `i=seq(2,6,by=0.05)`) and the error $= |\pi - \hat{\pi}|$ decreases. Produce a log-log plot of the error as a function of N and show that the data can be fit to a straight line of slope -1/2.