Bayesian Inference

Tutorial 7

1. (a) Consider the general hypothesis testing problem, where we have,

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1,$$

such that the union of Θ_0 and Θ_1 is the complete parameter space Θ and Θ_0 and Θ_1 are disjoint. Letting p_0 and p_1 denote the prior probabilities for the null hypothesis and alternative hypothesis, respectively, show that the posterior probability of H_0 is given by,

$$\mathbb{P}(H_0|\boldsymbol{x}) = \frac{p_0}{p_0 + p_1/B_{01}},$$

where B_{01} denotes the Bayes factor of H_0 to H_1.

(b) Now, suppose that we observe data $\boldsymbol{x} = \{x_1, \dots, x_n\}$, such that,

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2).$$

We wish to test,

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu = \mu_1.$$

Show that the Bayes factor is given by,

$$B_{01} = \exp\left(-\frac{n(\mu_0 - \mu_1)(\mu_0 + \mu_1 - 2\bar{x})}{2\sigma^2}\right).$$

Calculate the Bayes factor for H_0 against H_1, when $\mu_0 = 0$, $\mu_1 = 1$, $\sigma^2 = 1$, $n = 9$ and $\bar{x} = 0.645$. Interpret the result. What happens if we increase n?

(c) Alternatively, we wish to test,

$$H_0 : \mu < \theta \quad \text{vs} \quad H_1 : \mu > \theta.$$

Calculate the corresponding Bayes' factor, B_{01} for the general case, where we specify the prior,

$$\mu \sim N(\phi, \nu^2).$$

For $\theta = 0.25$, $\sigma^2 = 1$, $\phi = 0$, $\nu^2 = 1$, $n = 9$ and $\bar{x} = 0.645$ show that $B_{10} = 8.59$.

Note: You can use without proof the formula for $\pi(\mu|\boldsymbol{x})$ given a normal prior on μ and normal data. You will need R to compute the CDFs of a normal distribution to obtain the numerical result given.

(d) Finally, suppose we wish to test,

$$H_0 : \mu = \mu_o \quad \text{vs} \quad H_1 : \mu \neq \mu_0,$$

and we specify $\mu|H_1 \sim N(0, \tau^2)$. Calculate Bayes factor for H_0 against H_1. Comment on the limiting case as we make the prior on μ under H_1 increasingly vague (i.e. $\tau^2 \to \infty$).

2. Genetic linkage example (lecture notes section 2.5) - Write the R code to compute the posterior distribution of θ and Z using Gibbs sampling. Perform model checks to assess convergence and number of collected samples, report and comment on all your results. Finally, compute the posterior mean and 95%HPDI of θ (write a `for` loop to obtain the HPDI).

3. (A more challenging question!) Write NIMBLE code for the California's earthquakes example in Lecture 16 using (1) the hierarchical model suggested in the slides (slide 20, lecture 16) (2) a no-pooling model using a truncated $N(0, 10^5)$ prior on $\mathbb{R}^+$ for λ_i. Add a variable in both models to monitor the average time between events. Use data available in the `earthQ.R` file to run the two analyses. Run model checks and if problems are detected take appropriate actions to fix them. Once you are satisfied with your models, answer the following questions:

   (a) Compare the posterior mean and 95% CI of λ_i obtained with the two approaches. Do you notice anything of interest here? (The answer is obviously yes here - so what is it that is interesting?!)

   (b) Using only the posterior samples obtained from model 1 (i.e. without re-running the code), compute the probability that no earthquakes will be registered at fault 1 in the next 5 years.

   (c) Using model (1), produce a scatterplot for $\pi(a, b|y)$. What do you think might be the implications of this result?

   (d) Run again model (1) using the following prior for λ_i:

   $$\lambda_i \sim \Gamma(\phi, \kappa),$$

   where ϕ and κ are the Gamma mean and variance. Do you see any difference with the previous results of model 2? Compare the scatterplot of $\pi(\phi, \kappa|y)$ with the plot in (c).