ST4234: Bayesian Statistics

Tutorial 3, AY 19/20

Instructions: For each group who will present on Tuesday, 11 February, please upload **one** copy of your solutions to the folder "Tutorial Submission" by **6pm Tuesday, 11 February**.

1. Assume that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ is an i.i.d. sample generated from

$$\boldsymbol{x} = (x_1, x_2, x_3) \sim \text{Multinomial}(1; \theta_1, \theta_2, \theta_3),$$

which has a pdf

$$p(\boldsymbol{x}|\theta_1, \theta_2, \theta_3) = \frac{1!}{x_1!x_2!x_3!}\theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3},$$

$x_j \in 0, 1$ for $j = 1, 2, 3$, $x_1 + x_2 + x_3 = 1$ and $\theta_j \geq 0$ for $j = 1, 2, 3$, $\theta_1 + \theta_2 + \theta_3 = 1$. Assume a Dirichlet prior with known nonnegative parameters α_j, $j = 1, 2, 3$ for $(\theta_1, \theta_2, \theta_3)$:

$$p(\theta_1, \theta_2, \theta_3) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}\theta_1^{\alpha_1 - 1}\theta_2^{\alpha_2 - 1}\theta_3^{\alpha_3 - 1}.$$

This density is defined on the set $\{(\theta_1, \theta_2, \theta_3) : \ \theta_1 \geq 0, \theta_2 \geq 0, \theta_3 \geq 0, \theta_1 + \theta_2 + \theta_3 = 1\}$.

Denote $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3})$, $i = 1, 2, \ldots, n$ and set $x_{\cdot j} = \sum_{i=1}^{n} x_{ij}$, $j = 1, 2, 3$.

(a) Find the posterior pdf $p(\theta_1, \theta_2, \theta_3|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and show that the prior $p(\theta_1, \theta_2, \theta_3)$ is a conjugate prior for $(\theta_1, \theta_2, \theta_3)$ with respect to $p(\boldsymbol{x}|\theta_1, \theta_2, \theta_3)$.
(Hint: To show that a prior is conjugate, just show that the posterior belongs to the same family as the prior, i.e. Dirichlet distribution.)

(b) Let $\boldsymbol{x}_{n+1}$ be a new observation. Derive the predictive pdf $p(\boldsymbol{x}_{n+1}|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ and identify the distribution. Note that $\Gamma(a + 1) = a\Gamma(a)$, $a > 0$, $1! = 0! = 1$.
(Hint: $\boldsymbol{x}_{n+1} = (1, 0, 0)$, (0,1,0) or (0,0,1). Therefore the predictive pdf must be a multinomial distribution and we just need to find the parameters.)

2. Suppose that $\boldsymbol{x} = \{x_1, \ldots, x_m\}$ is an i.i.d. sample from $\mathrm{N}(\theta_1, \sigma^2)$. $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ is an i.i.d. sample from $\mathrm{N}(\theta_2, \sigma^2)$. $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent.

(a) Suppose that σ^2 is known. We assign the independent prior on (θ_1, θ_2): $p(\theta_1, \theta_2) = p(\theta_1)p(\theta_2)$, where $p(\theta_1)$ is the density of $\mathrm{N}(0, \tau_1^2)$, $p(\theta_2)$ is the density of $\mathrm{N}(0, \tau_2^2)$, and $\tau_1^2 > 0$, $\tau_2^2 > 0$ are fixed numbers. Derive the posterior distribution of $p(\theta_1, \theta_2 | \boldsymbol{x}, \boldsymbol{y})$. Are θ_1 and θ_2 independent in the posterior?

(b) Let $\delta = \theta_1 - \theta_2$. From (a), find the posterior distribution $p(\delta | \boldsymbol{x}, \boldsymbol{y})$.

(c) In the frequentist statistics, the following random variable is often used for testing hypothesis about $\delta = \theta_1 - \theta_2$:

$$T = \frac{(\bar{x} - \bar{y}) - (\theta_1 - \theta_2)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}}.$$

What is the distribution of T given θ_1 and θ_2? Comment on the relation between this distribution and $p(\delta | \boldsymbol{x}, \boldsymbol{y})$ from (b).

3. The following data represent the number of arrivals for 45 time intervals of length 2 minutes at a cashier's desk at a supermarket and are taken from Andersen (1980):

$$0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,$$
$$2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5$$

We can model such arrival counts by a Poisson model $p(x|\theta) = \theta^x e^{-\theta}/x!$, for $\theta > 0$ and $x = 0, 1, 2, \ldots$.

(a) For a Gamma(2,1) prior on θ, find the posterior distribution of θ. Plot the prior and the posterior. Find the mean, standard deviation and mode of the posterior. Is this prior sensible? What prior information does it add to the data?

(b) We consider an noninformative prior Gamma (1/2,0), with $p(\theta) \propto \theta^{-1/2}$. Repeat part (a) and compare the results.