Series 1

1. **Posterior predictive distribution**

   a. Assume the model $X \sim \mathcal{N}(\theta, \sigma^2)$ with the prior $\pi(\theta, \sigma^2) \propto \frac{1}{\sigma^2}$.[1] Consider $Y \sim \mathcal{N}(\rho X, \sigma^2)$ with ρ known and fixed. Derive the density $f(y \mid x)$ of the posterior predictive distribution of Y given X.

   b. Assume the model $X \sim$ Binomial(θ, n) with the prior $\theta \sim$ Beta(α, β). Further, assume that $Y \sim$ Binomial(θ, n) and that conditional on θ, Y is independent of X. Derive the density $f(y \mid x)$ of the posterior predictive distribution of Y given X.

Solution

a. The posterior predictive density is calculated as

$$
\begin{aligned}
f(y \mid x) &= \int f(y \mid x, \theta, \sigma^2) \pi(\theta, \sigma^2 | x) d\theta d\sigma^2 \\
&\propto \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\rho x)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) \frac{1}{\sigma^2} d\theta d\sigma^2 \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\rho x)^2}{2\sigma^2}\right) \frac{1}{\sigma^2} d\sigma^2 \\
&= \frac{1}{\sqrt{2}} \int \frac{1}{\Gamma(\alpha)} \exp\left(-\frac{\beta}{\sigma^2}\right) (\sigma^2)^{-\alpha-1} d\sigma^2
\end{aligned}
$$

where

$$
\beta = \frac{(y-\rho x)^2}{2}, \quad \alpha = 0.5,
$$

and we have used the fact that $\Gamma(0.5) = \sqrt{\pi}$. The last quantity is proportional to an inverse gamma distribution, thus

$$
\begin{aligned}
f(y \mid x) &\propto \frac{1}{\sqrt{2}} \frac{1}{\beta^\alpha} \\
&= \frac{1}{|y - \rho x|}.
\end{aligned}
$$

This is not integrable and, consequently, the posterior predictive distribution is not well defined.

b. We know that the posterior of θ is Beta$(\alpha + x, \beta + n - x)$. Further, for notational simplicity we write

$$
B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.
$$

[1] This is an example of a so called improper prior. As we will see later in the lecture, as long as $\pi(\theta) f(x \mid \theta)$ has finite total mass, one is allowed to use improper priors.

Using this, we obtain

$$
\begin{aligned}
f(y \mid x) &= \int \binom{n}{y} \theta^y (1-\theta)^{n-y} \frac{1}{B(\alpha+x, \beta+n-x)} \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1} d\theta \\
&= \binom{n}{y} \frac{1}{B(\alpha+x, \beta+n-x)} \int \theta^{\alpha+x+y-1}(1-\theta)^{\beta+2n-x-y-1} d\theta \\
&= \binom{n}{y} \frac{B(\alpha+x+y, \beta+2n-x-y)}{B(\alpha+x, \beta+n-x)}.
\end{aligned}
$$

This distribution is called the Beta-Binomial distribution.

2. **Bayesian decision theory**

   In the lecture, we saw the connection between Bayesian point estimates and Bayesian decision theory.

   a. Show that we obtian the posterior mean if we use a quadratic loss function $L(T, \theta) = (T - \theta)^2$.

   b. Show that we obtian the posterior median if we use $L(T, \theta) = |T - \theta|$.

   c. Show that we obtian the posterior mode if we use $L(T, \theta) = 1_{[-\varepsilon,\varepsilon]^c}(T - \theta)$ and we let ε go to zero.

   Hint:

   - For the median, use the Leibniz integral rule (see, e.g., https://en.wikipedia.org/wiki/Leibniz_integral_rule).

Solution

a. We have

$$\mathbb{E}((\theta - T)^2 \mid x) = \mathbb{E}((\theta)^2 \mid x) - 2\mathbb{E}(\theta \mid x)T + T^2.$$

This is minimized for $T = T(X) = \mathbb{E}(\theta \mid x)$.

b. We have

$$\mathbb{E}(|\theta - T| \mid x) = \int_{-\infty}^{T} (T - \theta)\pi(\theta \mid x) d\theta + \int_{T}^{\infty} (\theta - T)\pi(\theta \mid x) d\theta.$$

Using the Leibniz integral rule it follows that

$$\frac{\partial}{\partial T}\mathbb{E}(|\theta - T| \mid x) = \int_{-\infty}^{T} \pi(\theta \mid x) d\theta - \int_{T}^{\infty} \pi(\theta \mid x) d\theta.$$

This equals zero if $T = T(X) = \text{median } \pi(\theta \mid x)$.

c. We have

$$\mathbb{E}(1_{[-\varepsilon,\varepsilon]^c}(T-\theta) \mid x) = 1 - \int_{T-\varepsilon}^{T+\varepsilon} \pi(\theta \mid x) d\theta.$$

For small ε, we have

$$\int_{T-\varepsilon}^{T+\varepsilon} \pi(\theta \mid x) d\theta \approx 2\varepsilon\pi(\theta \mid x).$$

This is maximized, i.e., $\mathbb{E}(1_{[-\varepsilon,\varepsilon]^c}(T-\theta) \mid x)$ is minimized, for $T = T(X) = \text{mode } \pi(\theta \mid x)$.

3. **Bayesian testing and Bayes factor**

Assume the model $X \sim \mathcal{N}(\theta, 1)$ and for θ the prior $\pi(\theta) \propto 1$. Our goal is to test the hypothesis $H_0 : |\theta| \leq c$ versus $H_1 : |\theta| > c$.

a. Determine the maximal posterior probability of the null hypothesis $\max_x \pi(\Theta_0|x)$ as a function of c.

b. Determine the values of x and c for which $\pi(\Theta_0|x)$ equals 0.95 and calculate the Bayes factor.

Solution

a. The posterior of θ is $\mathcal{N}(x, 1)$. The posterior probability of the null hypothesis $\pi(\Theta_0|x)$ is given by

$$\begin{aligned}
\pi(\Theta_0|x) &= P(|\theta| \leq c|x) \\
&= P(-c \leq \theta \leq c|x) \\
&= P(-c - x \leq \theta - x \leq c - x|x) \\
&= \Phi(c - x) - \Phi(-c + x),
\end{aligned}$$

where $\Phi()$ is the cumulative distribution function of a stanadrd normal random variable.

It is easily seen that this is maximal for $x = 0$ and its values is then

$$\Phi(c) - \Phi(-c) = 2\Phi(c) - 1.$$

b. $\pi(\Theta_0|x) = \Phi(c - x) - \Phi(-c - x)$ equals 0.95 if

$$c - x = \Phi^{-1}(0.975) \approx 1.96.$$

The Bayes factor is given by

$$\frac{\pi(\Theta_0 \mid x)}{\pi(\Theta_1 \mid x)} \frac{\pi(\Theta_1)}{\pi(\Theta_0)}.$$

Since we use an improper prior, $\pi(\Theta_0)$ and $\pi(\Theta_1)$ are not well defined, and the Bayes factor is also not well defined neither.