

# Bayesian\_Statistics\_HW2

陈子睿 15220212202842

4/16/24

## 1 Calculate the Bayes Factor

We assume that the prior distribution for the two candidate models  $M_1$  and  $M_2$  are identical. Thus the Bayes Factor reduces to

$$BF = \frac{P((x, y)|M_1)}{P((x, y)|M_2)}.$$

Then the Bayes Factor is:

$$\begin{aligned} BF &= \frac{P((x, y)|M_1)}{P((x, y)|M_2)} \\ &= \frac{\prod_{i=1}^3 f((X_i, Y_i)|M_1)}{\prod_{i=1}^3 f((X_i, Y_i)|M_2)} \\ &= \frac{\prod_{i=1}^3 (2\pi \times 2)^{-1/2} \exp\left(-\frac{1}{2 \times 2} Y_i^2\right)}{\prod_{i=1}^3 (2\pi \times \sigma_i^2)^{-1/2} \exp\left(-\frac{1}{2 \times \sigma_i^2} Y_i^2\right)} \\ &= \prod_{i=1}^3 (\sigma_i^2)^{1/2} \times 2^{-3/2} \exp\left(\sum_{i=1}^3 \left(\frac{1}{2\sigma_i^2} - \frac{1}{4}\right) Y_i^2\right) \\ &= 0.789. \end{aligned}$$

## 2 Inference for time series model

(a) We can write the time series model as:

$$Y_t = \rho Y_{t-1} + \epsilon_t, \quad \{\epsilon_t\} \sim WN(0, \sigma^2).$$

while  $Y_1 \sim N(0, 10^2)$ . Now we consider priors  $\rho \sim N(0, \tau^2)$  and  $\sigma^2 \sim InvGamma(a, b)$ . Then we

can calculate the posterior as:

$$\begin{aligned}
p(\rho|\sigma^2, Y_1, \dots, Y_n) &\propto p(\rho)p(Y_1, \dots, Y_n|\rho, \sigma^2) \\
&= p(\rho)p(Y_1|\rho, \sigma^2)p(Y_2|Y_1, \rho, \sigma^2) \dots P(Y_n|Y_{n-1}, \dots, Y_1, \rho, \sigma^2) \\
&= p(\rho)p(Y_1|\rho, \sigma^2) \prod_{i=2}^n p(Y_i|Y_{i-1}, \rho, \sigma^2) \\
&\propto \exp\left(-\frac{\rho^2}{2\tau^2}\right) \exp\left(-\frac{y_1^2}{20}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=2}^n (y_i - \rho y_{i-1})^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2\tau^2}[(\tau^2 \sum y_{i-1}^2) + \sigma^2]\rho^2 - \tau^2(\sum y_i y_{i-1})\rho\right) \\
&\sim N(\mu_\rho, \sigma_\rho^2),
\end{aligned}$$

where

$$\mu_\rho = \frac{\sum_{i=2}^n y_i y_{i-1}}{\sum_{i=2}^n y_{i-1}^2 + \sigma^2/\tau^2}, \sigma_\rho^2 = \frac{\sigma^2}{\sum_{i=2}^n y_{i-1}^2 + \sigma^2/\tau^2}.$$

(b) In the same case as (a), we can compute the posteriori by:

$$\begin{aligned}
p(\sigma^2|\rho, Y_1, \dots, Y_n) &\propto p(\sigma^2)p(Y_1, \dots, Y_n|\rho, \sigma^2) \\
&= p(\sigma^2)p(Y_1|\rho, \sigma^2)p(Y_2|Y_1, \rho, \sigma^2) \dots P(Y_n|Y_{n-1}, \dots, Y_1, \rho, \sigma^2) \\
&= p(\sigma^2)p(Y_1|\rho, \sigma^2) \prod_{i=2}^n p(Y_i|Y_{i-1}, \rho, \sigma^2) \\
&\propto (\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right) \prod_{i=2}^n (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=2}^n (y_i - \rho y_{i-1})^2\right) \\
&= (\sigma^2)^{-(a+\frac{n-1}{2}+1)} \exp\left(-\frac{1}{\sigma^2} \left(b + \frac{1}{2} \sum_{i=2}^n (y_i - \rho y_{i-1})^2\right)\right) \\
&\sim \text{InvGamma}(a', b'),
\end{aligned}$$

where

$$a' = a + \frac{n-1}{2}, b' = b + \frac{1}{2} \sum_{i=2}^n (y_i - \rho y_{i-1})^2.$$

### 3 Importance Sampling

(a) To estimate  $\sigma^2 = E(X^2)$  using importance sampling with a standard normal proposal distribution, we need the weights  $\omega(x) = \frac{q(x)}{g(x)}$ . Where  $g(x)$  is the standard normal density, and we do not need the normalizing constant of  $q(x)$ . The weight function is given by:

$$\omega(x) = \frac{e^{-|x|^3/3}}{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}} = \sqrt{2\pi}e^{-|x|^3/3+x^2/2}$$

Then the estimator is given by:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \omega(X_i) X_i^2}{\sum_{i=1}^N \omega(X_i)}$$

where  $X_i \sim N(0, 1)$  are samples from the proposal distribution.

The R code is pasted below:

```
> # Number of samples
> N <- 10000
> # Sample from the standard normal distribution (proposal distribution)
> X <- rnorm(N)
> # Compute weights  $w(X) = \sqrt{2\pi} \cdot \exp(-\text{abs}(X)^3/3 + X^2/2)$ 
> weights <- sqrt(2*pi) * exp(-abs(X)^3/3 + X^2/2)
> # Compute weighted estimator for  $\sigma^2 = E(X^2)$ 
> sigma_squared_estimate <- sum(weights * X^2) / sum(weights)
> # Print the estimate of  $\sigma^2$ 
> print(sigma_squared_estimate)
[1] 0.7897861
```

(b) To use the rejection sampling to estimate  $\sigma^2$ , we can

```
# Number of samples we want
N <- 1000

# Samples and acceptance storage
samples <- numeric(N)
accepted <- 0

# Determining the M constant empirically or analytically
# For simplicity, we start with an assumption and adjust based on trial
M <- 10 # Start with an arbitrary value and adjust based on rejection rate

# Sampling loop
while (accepted < N) {
  X_proposed <- rnorm(1)
  U <- runif(1, 0, M * dnorm(X_proposed))
  q_X <- exp(-abs(X_proposed)^3 / 3)

  # Acceptance check
  if (U <= q_X) {
    accepted <- accepted + 1
    samples[accepted] <- X_proposed
  }
}

# Estimation of  $\sigma^2$ 
```

```
sigma_squared_estimate <- mean(samples^2)

# Print the estimate of sigma^2
print(sigma_squared_estimate)
[1] 0.763087
```

## 4 Bayesian hierarchical model

Given the problem setup, where  $y_{i,j}$  is drawn from a normal distribution with mean  $\mu_j$  and precision  $\tau$  (inverse of variance), and considering the exchangeability of the  $J$  samples, we can specify a fully Bayesian hierarchical model to analyze this data. The model includes specifying prior distributions for  $\mu_j$  and  $\tau$ , and possibly hyperpriors for parameters of these priors if deemed necessary.

First we specify our model as

$$y_{i,j} \sim \text{Normal}(\mu_j, \tau^{-1}), \quad \text{for } i = 1, \dots, n_j \text{ and } j = 1, \dots, J.$$

Since there is no additional information to distinguish between the  $J$  samples, we assume the  $\mu_j$  are drawn from a common distribution, typically a normal distribution:

$$\mu_j \sim \text{Normal}(\mu_0, \sigma_0^2).$$

Here,  $\mu_0$  and  $\sigma_0^2$  can either be set to known values or treated as unknowns and estimated from the data.

Also, we can choose a conjugate prior for the precision  $\tau$ , namely Gamma distribution

$$\tau \sim \text{Gamma}(a, b),$$

To complete the hierarchy, we specify priors for  $\mu_0$  and  $\sigma_0^2$  if needed:

$$\mu_0 \sim N(\mu_\mu, \sigma_\mu^2), \sigma_0^2 \sim \text{InvGamma}(\alpha, \beta)$$

The posterior distribution of the parameters given the data,  $p(\mu_1, \dots, \mu_J, \tau | y)$ , is proportional to the product of the likelihood and the priors:

$$p(\mu_1, \dots, \mu_J, \tau | y) \propto p(y | \mu_1, \dots, \mu_J, \tau) p(\mu_1, \dots, \mu_J | \mu_0, \sigma_0^2) p(\tau)$$

Given that  $\mu_j$  are conditionally independent given  $\tau$  and the hyperparameters, and  $\tau$  is independent of the hyperparameters, this results in:

$$\prod_{j=1}^J \left[ \prod_{i=1}^{n_j} \text{Normal}(y_{i,j} | \mu_j, \tau^{-1}) \right] \times \prod_{j=1}^J \text{Normal}(\mu_j | \mu_0, \sigma_0^2) \times \text{Gamma}(\tau | a, b).$$

To sample from the posterior distribution, we can use Gibbs sampling, which involves iteratively sampling each parameter from its conditional posterior distribution, given the current values of all other parameters. The steps are as follows:

1. Sample  $\mu_j$  given  $\mu_0, \sigma_0^2, \tau$ , and  $y$ :

$$\mu_j | \mu_0, \sigma_0^2, \tau, y \sim \text{Normal}(\text{updated mean}, \text{updated variance})$$

2. Sample  $\tau$  given  $\mu_1, \dots, \mu_J$ , and  $y$ :

$$\tau | \mu_1, \dots, \mu_J, y \sim \text{Gamma}(\text{updated shape}, \text{updated rate})$$

3. Sample  $\mu_0$  and  $\sigma_0^2$  (if they are treated as random variables):

- Update  $\mu_0$  based on the distribution of  $\mu_j$ .
- Update  $\sigma_0^2$  based on the variance among  $\mu_j$ .

These steps are repeated until convergence, at which point the samples can be used to make Bayesian inferences about the parameters and predictive distributions.