STA3105-01 Bayesian Statistics
Homework 1
DUE Friday, September 22

Copying homework solutions from others lead to 0 score. No late submission is allowed. Your solution should contain both code and corresponding explanation for the answer. Submit your HW through LearnUs. You should submit (1) a report file (pdf) and (2) a relevant code file.

1. Consider iid random variables $X_1, \cdots, X_5$ following normal distribution with mean μ and variance 5. We observe $X_1 = 10, X_2 = 15, X_3 = 13, X_4 = 9, X_5 = 11$ in our dataset. We set the prior of $\mu \sim N(3,3)$.

   (a) (25 points) Derive the posterior distribution in this problem. You should (1) specify the kernel of the likelihood, prior and (2) explain line-by-line equations for how to calculate the posterior (not just the final answer).

   (b) (25 points) Report the posterior mean, 95% symmetrical density interval, and 95% highest posterior density interval. Are there any differences between the symmetrical density interval and the highest posterior density interval?

2. Consider iid random variables $X_1, \cdots, X_5$ following normal distribution with mean 10 and variance σ. We observe $X_1 = 10, X_2 = 15, X_3 = 13, X_4 = 9, X_5 = 11$ in our dataset. We set the prior of $\sigma \sim IG(3,3)$. Here, inverse Gamma distribution with shape α, scale β is defined as $p(\sigma) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma)^{\alpha+1}\exp(-\beta/\sigma)$ for $\sigma \in (0,\infty)$ (https://en.wikipedia.org/wiki/Inverse-gamma_distribution).

   (a) (25 points) Derive the posterior distribution in this problem. You should (1) specify the kernel of the likelihood, prior and (2) explain line-by-line equations for how to calculate the posterior (not just the final answer).

   (b) (25 points) Report the posterior mean, 95% symmetrical density interval, and 95% highest posterior density interval. Are there any differences between the symmetrical density interval and the highest posterior density interval?

STA3105-01 Bayesian Statistics

Jeong Geonwoo

DUE Friday, September 22

1. Consider iid random variables $X_1, ..., X_5$ following normal distribution with mean μ and variance 5. We observe $X_1 = 10,\ X_2 = 15,\ X_3 = 13,\ X_4 = 9,\ X_5 = 11$ in our dataset. We set the prior of $\mu \sim N(3, 3)$.

   (a) (25 points) Derive the posterior distribution in this problem. You should (1) specify the kernel of the likelihood, prior and (2) explain line-by-line equations for how to calculate the posterior (not just the final answer).

   solution
   Likelihood :

$$
\begin{aligned}
f(X|\mu) &= \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sqrt{5}} exp(-\frac{(X_i - \mu)^2}{2 * 5}) \\
&\propto exp(-\frac{1}{10}\sum_{i=1}^{5}(X_i - \mu)^2) \\
&\propto exp(-\frac{1}{2}\mu^2 + \frac{1}{5}\mu\sum_{i=1}^{5} X_i)
\end{aligned}
$$

   Prior :

$$
\begin{aligned}
p(\mu) &= \frac{1}{\sqrt{2\pi}\sqrt{3}} exp(-\frac{(\mu - 3)^2}{2 * 3}) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{3}} exp(-\frac{9}{2 * 3}) exp(-\frac{\mu^2 - 6\mu}{2 * 3}) \\
&\propto exp(-\frac{1}{6}\mu^2 + \mu)
\end{aligned}
$$

   Posterior :

$$
\begin{aligned}
\pi(\mu|X) &\propto f(X|\mu)p(\mu) \\
&\propto exp(-\frac{1}{2}\mu^2 + \frac{1}{5}\mu\sum_{i=1}^{5} X_i) exp(-\frac{1}{6}\mu^2 + \mu) \\
&= exp(-\frac{2}{3}\mu^2 + \frac{63}{5}\mu) \\
&\propto exp(-\frac{(\mu - \frac{189}{20})^2}{2 * \frac{3}{4}})
\end{aligned}
$$

   Hence $\mu|X \sim N(\frac{189}{20}, \frac{3}{4})$

(b) (25 points) Report the posterior mean, 95% symmetrical density interval, and 95% highest posterior density interval. Are there any differences between the symmetrical density interval and the highest posterior density interval?

solution

```
# Setting and sampling
library(coda)
set.seed(2018122062)
mean <- 189/20
sd <- sqrt(3/4)
samples <- rnorm(1000000, mean = mean, sd = sd)

# Posterior mean
cat("Posterior mean :",
   mean(samples)
   )
```

```
## Posterior mean : 9.449772
```

The posterior mean is almost equal to $\frac{189}{20} = 9.45$

```
# Highest Posterior Density (HPD) Interval
mcmc_samples <- as.mcmc(samples)
hpd_interval <- HPDinterval(mcmc_samples, prob = 0.95)
cat("95% highest posterior density interval
   [",hpd_interval,"]"
   )
```

```
## 95% highest posterior density interval
##     [ 7.745128 11.13625 ]
```

```
# Symmetrical Density Interval
sorted_samples <- sort(samples)
lower_percentile <- 0.025
upper_percentile <- 0.975
lower_index <- floor(length(sorted_samples) * lower_percentile) + 1
upper_index <- ceiling(length(sorted_samples) * upper_percentile)
confidence_interval <- c(sorted_samples[lower_index],
                        sorted_samples[upper_index])
cat("95% symmetrical density interval
   [", confidence_interval[1], ", ", confidence_interval[2],"]"
   )
```

```
## 95% symmetrical density interval
##     [ 7.75663 ,  11.14804 ]
```

Since posterior distribution is Normal distribution (which is **symmetric**), Highest posterior density interval and Symmetrical density interval are almost identical.

2. Consider iid random variables $X_1, ..., X_5$ following normal distribution with mean 10 and variance σ. We observe $X_1 = 10, X_2 = 15, X_3 = 13, X_4 = 9, X_5 = 11$ in our dataset. We set the prior of $\sigma \sim IG(3,3)$. Here, inverse Gamma distribution with shape α, scale β is defined as $p(\sigma) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma)^{\alpha+1}exp(-\beta/\sigma)$ for $\sigma \in (0,\infty)$

   (a) (25 points) Derive the posterior distribution in this problem. You should (1) specify the kernel of the likelihood, prior and (2) explain line-by-line equations for how to calculate the posterior (not just the final answer).

   solution

   Likelihood :

$$
\begin{aligned}
f(X|\mu) &= \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sqrt{\sigma}} exp(-\frac{(X_i-10)^2}{2\sigma}) \\
&= (\frac{1}{\sqrt{2\pi}\sqrt{\sigma}})^5 exp(-\frac{1}{2\sigma}[(10-10)^2 + (15-10)^2 + (13-10)^2 + (9-10)^2 + (11-10)^2]) \\
&= (\frac{1}{\sqrt{2\pi}\sqrt{\sigma}})^5 exp(-\frac{18}{\sigma}) \\
&\propto (\frac{1}{\sqrt{\sigma}})^5 exp(-\frac{18}{\sigma})
\end{aligned}
$$

   Prior :

$$
\begin{aligned}
p(\sigma) &= \frac{3^3}{\Gamma(3)}(1/\sigma)^4 exp(-3/\sigma) \\
&\propto (1/\sigma)^4 exp(-3/\sigma)
\end{aligned}
$$

   Posterior :

$$
\begin{aligned}
\pi(\sigma|X) &\propto f(X|\sigma)p(\sigma) \\
&\propto (\frac{1}{\sqrt{\sigma}})^5 exp(-\frac{18}{\sigma})(1/\sigma)^4 exp(-3/\sigma) \\
&= (\frac{1}{\sigma})^{\frac{11}{2}+1} exp(-\frac{21}{\sigma})
\end{aligned}
$$

   Hence $\sigma|X \sim IG(\frac{11}{2}, 21)$

(b) (25 points) Report the posterior mean, 95% symmetrical density interval, and 95% highest posterior density interval. Are there any differences between the symmetrical density interval and the highest posterior density interval?

solution

```
# Setting and sampling
set.seed(2018122062)
shape <- 11/2
scale <- 21
gamma_samples <- rgamma(1000000, shape=shape, scale=1/scale)
ig_samples <- 1 / gamma_samples

# Posterior mean
cat("Posterior mean :",
   mean(ig_samples)
   )
```

```
## Posterior mean : 4.668862
```

```
# Highest Posterior Density (HPD) Interval
mcmc_ig_samples <- as.mcmc(ig_samples)
hpd_interval <- HPDinterval(mcmc_ig_samples, prob = 0.95)
cat("95% highest posterior density interval
   [",hpd_interval,"]"
   )
```

```
## 95% highest posterior density interval
##     [ 1.458828 9.316545 ]
```

```
# Symmetrical Density Interval
sorted_ig_samples <- sort(ig_samples)
lower_percentile <- 0.025
upper_percentile <- 0.975
lower_index <- floor(length(sorted_ig_samples) * lower_percentile) + 1
upper_index <- ceiling(length(sorted_ig_samples) * upper_percentile)
confidence_interval <- c(sorted_ig_samples[lower_index],
                        sorted_ig_samples[upper_index])
cat("95% symmetrical density interval
   [", confidence_interval[1], ", ", confidence_interval[2],"]"
   )
```

```
## 95% symmetrical density interval
##     [ 1.915132 ,  11.03349 ]
```

Since posterior distribution is Inverse Gamma distribution (which is **asymmetric**), Highest posterior density interval and Symmetrical density interval are not equal.