ST740 – Assignment 2 – Solution

TA: Jianing Chu

1. Assume $Y_1, ..., Y_n|\theta \sim \text{Normal}(0, \theta)$ and variance parameter has prior $\theta \sim \text{Gamma}(a, b)$.

(a). Derive the posterior distribution of θ, i.e., give a parametric family like $\theta|Y \sim \text{Beta}(Y, a^b)$.

Solution:

$$f(\boldsymbol{Y}|\theta) \propto \theta^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n Y_i^2}{2\theta}\right\}.$$

$$\pi(\theta) \propto \theta^{a-1} \exp(-b\theta).$$

$$p(\theta|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\theta)\pi(\theta) = \theta^{a-n/2-1} \exp\left\{-\frac{\sum_{i=1}^n Y_i^2}{2\theta} - b\theta\right\}.$$

The posterior distribution is a generalized inverse Gaussian distribution with parameters $p = a - n/2$, $\alpha = 2b$ and $\beta = \sum_{i=1}^n Y_i^2$.

(b). Would you say this prior is conjugate? Justify your answer.

Solution:
No. Since a generalized inverse Gaussian distribution requires the parameter $\beta > 0$, the prior is not a generalized inverse Gaussian distribution. Thus, it's not conjugate.

2. Say $\mathbf{Y} = (Y_1, ..., Y_p)|\theta \sim \text{Multinomial}(n; \theta)$ for $\theta = (\theta_1, ..., \theta_p)$ so that the likelihood is

$$f(\boldsymbol{Y}|\theta) = \frac{n!}{\prod_{j=1}^p Y_j!} \prod_{j=1}^p \theta_j^{Y_j}.$$

(a). Derive the Jeffreys prior for θ.

Solution:

$$\log f(\boldsymbol{Y}|\theta) = C + \sum_{j=1}^n Y_j \log(\theta_j),$$

where C is a constant which does not depend on θ.

$$\frac{\partial \log f(\boldsymbol{Y}|\theta)}{\partial \theta_j} = \frac{Y_j}{\theta_j}.$$

$$\frac{\partial^2 \log f(\boldsymbol{Y}|\theta)}{\partial \theta_i \theta_j} = \begin{cases} -\frac{Y_j}{\theta_j^2} & i = j, \\ 0 & i \neq j. \end{cases}$$

$$\mathcal{I}(\theta) = -E\left\{\frac{\partial^2 \log f(\boldsymbol{Y}|\theta)}{\partial\theta\partial\theta^T}\right\} = \text{diag}\left(\frac{n}{\theta_1}, \ldots, \frac{n}{\theta_p}\right).$$

The Jeffreys' prior is thus

$$\pi(\theta) \propto |\mathcal{I}(\theta)|^{1/2} \propto \Pi_{j=1}^{p} \theta_j^{-1/2}.$$

(b). Derive the posterior under the prior in (a).

Solution:

$$p(\theta|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\theta)\pi(\theta) \propto \Pi_{j=1}^{p} \theta_j^{Y_j - 1/2}.$$

The posterior is the Dirichlet$(Y_1 + 1/2, \dots, Y_p + 1/2)$.

(c). Assume that $\mathbf{Y} = (10, 20, 30)$ and summarize the posterior under the prior in (a) in a figure and table.

Solution:

```
Y <- c(10,20,30)
# posterior mean
pst_mean <- (Y+0.5) / sum(Y+0.5)
pst_mean
```

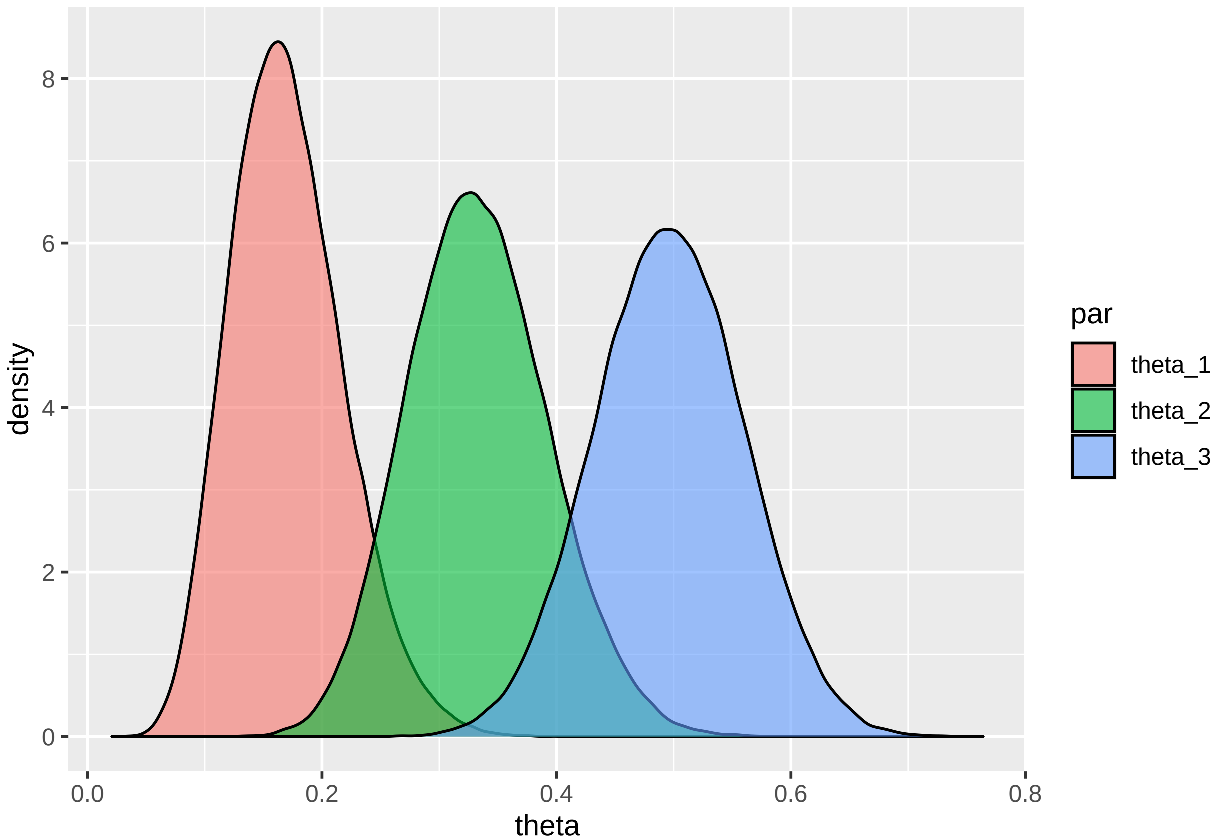
```
## [1] 0.1707317 0.3333333 0.4959350
```

```
# posterior variance
(diag(pst_mean) - pst_mean %*% t(pst_mean)) / (sum(Y+0.5)+1)
```

```
##               [,1]          [,2]         [,3]
## [1,]  0.0022653183 -0.0009105691 -0.001354749
## [2,] -0.0009105691  0.0035555556 -0.002644986
## [3,] -0.0013547492 -0.0026449864  0.003999736
```

posterior mean	posterior covariance matrix
$(0.1707, 0.3333, 0.4959)$	$\begin{pmatrix} 0.0023 & -0.0009 & -0.0014 \\ -0.0009 & 0.0036 & -0.0026 \\ -0.0014 & -0.0026 & 0.0040 \end{pmatrix}$

```
library(dirmult)
library(ggplot2)
theta <- rdirichlet(n = 10^5, alpha = Y+0.5)
dat <- data.frame(theta = c(theta[,1],theta[,2],theta[,3]),
                  par = rep(c("theta_1", "theta_2","theta_3"), each = 10^5))
ggplot(dat, aes(x = theta, fill = par)) + geom_density(alpha=0.6)
```



(d). Now apply the Bayesian Central Limit Theorem to obtain an approximate normal distribution for the posterior of θ given $\mathbf{Y} = (10, 20, 30)$. Summarize this approximate posterior in a figure and table. Are the results similar to the exact posterior? Is this a good approximation?

Solution:

$$\log p(\theta|\boldsymbol{Y}) = C + \sum_{j=1}^{p}\left(Y_j - \frac{1}{2}\right)\log(\theta_j),$$

where C is a constant which does not depend on θ.

$$\frac{\partial \log p(\theta|\boldsymbol{Y})}{\partial \theta_j} = \frac{Y_j - \frac{1}{2}}{\theta_j}.$$

$$\frac{\partial^2 \log p(\theta|\boldsymbol{Y})}{\partial \theta_i \theta_j} = \begin{cases} -\frac{Y_j - \frac{1}{2}}{\theta_j^2} & i = j, \\ 0 & i \neq j. \end{cases}$$

$$\mathcal{I}(\theta) = -E\left\{\frac{\partial^2 \log p(\theta|\boldsymbol{Y})}{\partial \theta \partial \theta^T}\right\} = \text{diag}\left(\frac{n\theta_1 - \frac{1}{2}}{\theta_1^2}, \ldots, \frac{n\theta_p - \frac{1}{2}}{\theta_p^2}\right).$$

$$\mathcal{I}^{-1}(\theta) = -E\left\{\frac{\partial^2 \log p(\theta|\boldsymbol{Y})}{\partial \theta \partial \theta^T}\right\} = \text{diag}\left(\frac{\theta_1^2}{n\theta_1 - \frac{1}{2}}, \ldots, \frac{\theta_p^2}{n\theta_p - \frac{1}{2}}\right).$$

The Lagrangian is

$$L(\theta, \lambda) = \sum_{j=1}^{p}\left(Y_j - \frac{1}{2}\right)\log(\theta_j) + \lambda\left(1 - \sum_{j=1}^{p}\theta_j\right).$$

$$\frac{\partial L(\theta, \lambda)}{\partial \theta_j} = \frac{Y_j - \frac{1}{2}}{\theta_j} - \lambda = 0.$$

$$\frac{\partial L(\theta,\lambda)}{\partial \lambda} = 1 - \sum_{j=1}^{p} \theta_j = 0.$$

$$\hat{\theta}_j = \frac{Y_j - \frac{1}{2}}{\sum_{j=1}^{p}\left(Y_j - \frac{1}{2}\right)}.$$

$$\hat{\theta}^T = \left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\right) = (0.1624, 0.3333, 0.5043).$$

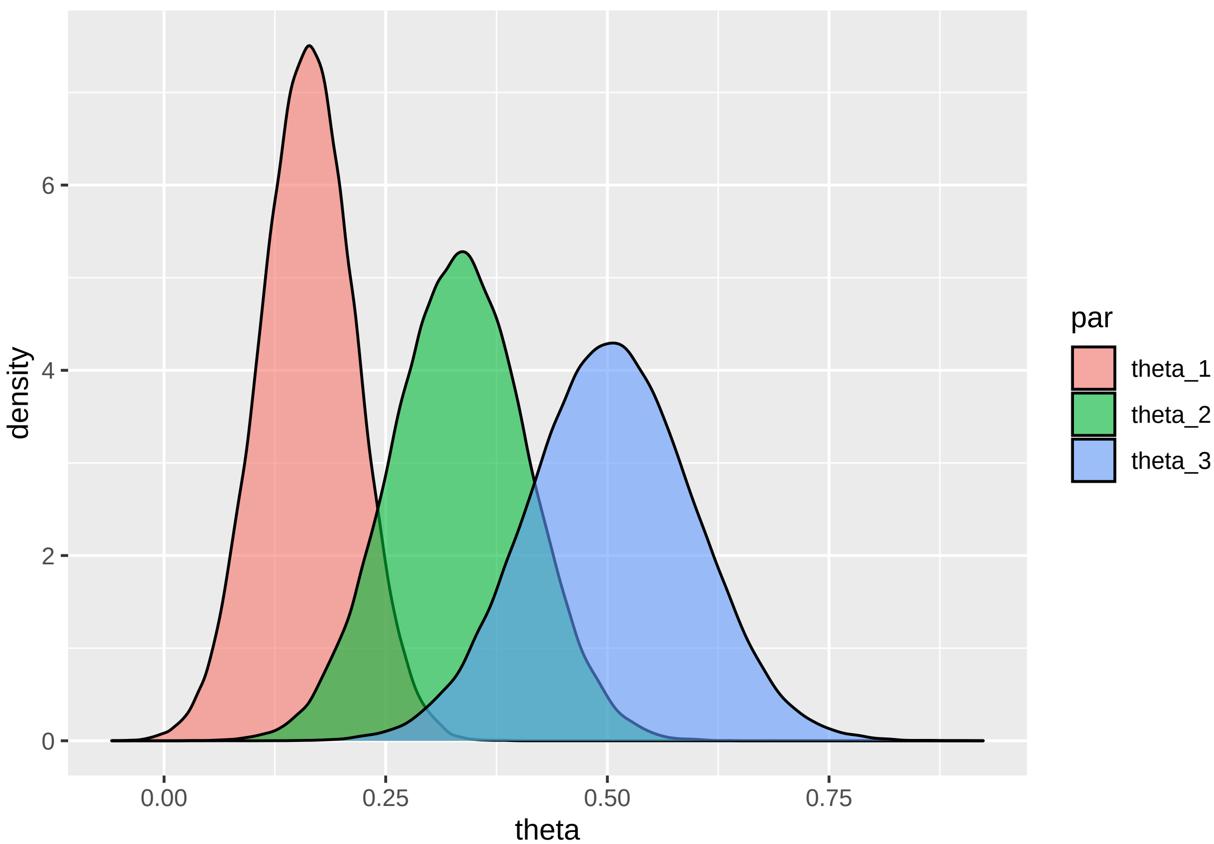
In this case, $n = \sum_{j=1}^{p} Y_j = 60$. Thus,

$$\mathcal{I}^{-1}(\hat{\theta}) = \text{diag}(0.0029, 0.0057, 0.0085).$$

By Bayesian CLT, $p(\theta|\boldsymbol{Y}) \to N(\hat{\theta}, \mathcal{I}^{-1}(\hat{\theta}))$.

posterior mean	posterior covariance matrix
$(0.1624, 0.3333, 0.5043)$	$\begin{pmatrix} 0.0029 & 0 & 0 \\ 0 & 0.0057 & 0 \\ 0 & 0 & 0.0085 \end{pmatrix}$

```
library(MASS)
Y <- c(10,20,30)
mu <- (Y-0.5) / sum(Y-0.5)
Sigma <- diag(mu^2 / (60*mu-0.5))
n <- 10^5
theta <- mvrnorm(n,mu,Sigma)
dat <- data.frame(theta = c(theta[,1],theta[,2],theta[,3]),
                  par = rep(c("theta_1", "theta_2","theta_3"), each = 10^5))
ggplot(dat, aes(x = theta, fill = par)) + geom_density(alpha=0.6)
```



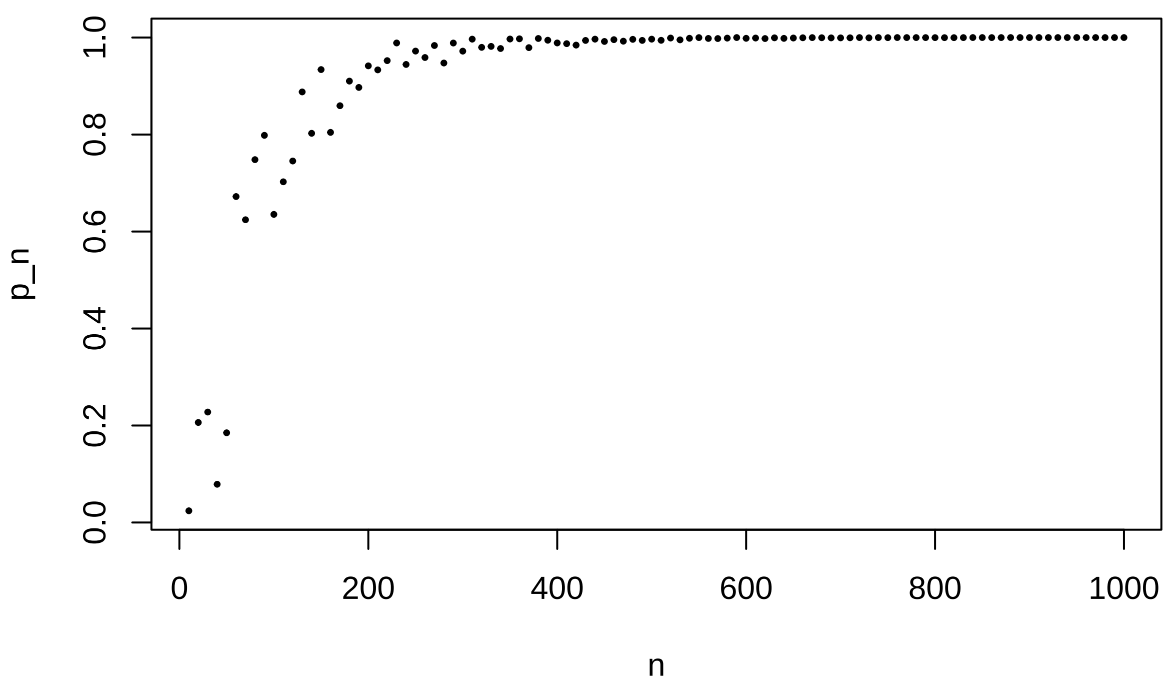
The means of the approximate posterior and the exact posterior are close to each other. However, the approximate posterior's variance is larger and there is no correlation between parameters.

4. Assume $Y_i|\theta \sim \text{Uniform}(0, \theta)$ independent for $i \in \{1, ..., n\}$ and prior $\theta \sim \text{Pareto}(\theta_0, \alpha)$ with support $\theta > \theta_0$ and CDF $\text{Prob}(\theta < t) = 1 - (\theta_0/t)^\alpha$.

(a). Say the true value of θ is $\theta^* = 10$ and the prior has $\theta_0 = \alpha = 1$. For a dataset of size n, $\mathbf{Y}_n = \{Y_1, ..., Y_n\}$, let $p_n = \text{E}_{\mathbf{Y}_n|\theta^*}\{\text{Prob}(\theta^* - \epsilon < \theta < \theta^* + \epsilon|\mathbf{Y}_n)\}$ for $\epsilon = 0.1$. Compute a Monte Carlo approximation to p_n for each $n \in \{10, 20, ..., 1000\}$. Does a plot of n versus p_n suggest posterior consistency? Why?

Solution: From Assignment 1, we know the posterior is $\text{Pareto}(\max\{Y_{(n)}, \theta_0\}), \alpha + n)$.

```
library(Pareto)
set.seed(2022)
n <- seq(10, 1000, by=10)
theta_0 <- 1
theta_true <- 10
alpha <- 1
epsilon <- 0.1
p_n <- rep(0,length(n))
for (i in 1:length(n)){
  Y <- runif(n[i], 0, theta_true)
  pst_sample <- rPareto(10^5, max(Y,theta_0), alpha+n[i])
  p_n[i] <- sum(pst_sample > theta_true-epsilon & pst_sample < theta_true+epsilon)/10^5
}
plot(x=n, y=p_n, pch=20, cex=0.5)
```



p_n converges to 1 as n increases. Thus, the posterior of θ converges to the true value, which suggests posterior consistency.

(b). Without evoking any general theorems discussed in class, derive $\lim_{n\to\infty} p_n$. Do you get the same conclusion about posterior consistency as the Monte Carlo study in (a)?

Solution:

$$\text{Prob}(\theta < \theta^* + \epsilon|\mathbf{Y}_n) = 1 - \left[\frac{\max\{\theta_0, Y_{(n)}\}}{\theta^* + \epsilon}\right]^{n+\alpha}.$$

$$\text{Prob}(\theta < \theta^* - \epsilon|\boldsymbol{Y}_n) = \left\{1 - \left[\frac{\max\{\theta_0, Y_{(n)}\}}{\theta^* - \epsilon}\right]^{n+\alpha}\right\} I\left[\theta^* - \epsilon > \max\{\theta_0, Y_{(n)}\}\right]$$

In practice, we set $\theta_0 \leq \theta^*$.

(1) If $\theta_0 = \theta^*$,

$$\begin{aligned}
&\text{Prob}(\theta^* - \epsilon < \theta < \theta^* + \epsilon|\boldsymbol{Y}_n) \\
=&\text{Prob}(\theta < \theta^* + \epsilon|\boldsymbol{Y}_n) - \text{Prob}(\theta < \theta^* - \epsilon|\boldsymbol{Y}_n) \\
=&1 - \left(\frac{\theta^*}{\theta^* + \epsilon}\right)^{n+\alpha}
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} p_n &= \lim_{n\to\infty} \text{E}_{\mathbf{Y}_n|\theta^*}\{\text{Prob}(\theta^* - \epsilon < \theta < \theta^* + \epsilon|\boldsymbol{Y}_n)\} \\
&= \lim_{n\to\infty}\left[1 - \left(\frac{\theta^*}{\theta^* + \epsilon}\right)^{n+\alpha}\right] \\
&=1
\end{aligned}$$

(2) If $\theta_0 < \theta^*$, for a sufficiently small ϵ, we have $\theta^* - \epsilon > \theta_0$. In our case, $\theta^* = 10$, $\epsilon = 0.1$, $\theta_0 = 1$. Therefore, we have

$$\text{Prob}(\theta < \theta^* - \epsilon|\boldsymbol{Y}_n) = \left\{1 - \left[\frac{\max\{\theta_0, Y_{(n)}\}}{\theta^* - \epsilon}\right]^{n+\alpha}\right\} I\left[\theta^* - \epsilon > Y_{(n)}\right].$$

$$\begin{aligned}
&\text{E}_{\mathbf{Y}_n|\theta^*}\left\{\text{Prob}(\theta < \theta^* - \epsilon|\boldsymbol{Y}_n)\right\} \\
=&\int_0^{\theta_0}\left[1 - \left(\frac{\theta_0}{\theta^* - \epsilon}\right)^{n+\alpha}\right]\frac{ny^{n-1}}{(\theta^*)^n}dy + \int_{\theta_0}^{\theta^*-\epsilon}\left[1 - \left(\frac{y}{\theta^* - \epsilon}\right)^{n+\alpha}\right]\frac{ny^{n-1}}{(\theta^*)^n}dy \\
=&\left(\frac{\theta^* - \epsilon}{\theta^*}\right)^n - \frac{\theta_0^{2n+\alpha}}{(\theta^* - \epsilon)^{n+\alpha}(\theta^*)^n} + \frac{n}{2n+\alpha}\frac{(\theta^* - \epsilon)^{2n+\alpha} - \theta_0^{2n+\alpha}}{(\theta^* - \epsilon)^{n+\alpha}(\theta^*)^n}
\end{aligned}$$

$$\begin{aligned}
&\text{E}_{\mathbf{Y}_n|\theta^*}\left\{\text{Prob}(\theta < \theta^* + \epsilon|\boldsymbol{Y}_n)\right\} \\
=&\int_0^{\theta_0}\left[1 - \left(\frac{\theta_0}{\theta^* + \epsilon}\right)^{n+\alpha}\right]\frac{ny^{n-1}}{(\theta^*)^n}dy + \int_{\theta_0}^{\theta^*}\left[1 - \left(\frac{y}{\theta^* + \epsilon}\right)^{n+\alpha}\right]\frac{ny^{n-1}}{(\theta^*)^n}dy \\
=&1 - \frac{\theta_0^{2n+\alpha}}{(\theta^* + \epsilon)^{n+\alpha}(\theta^*)^n} + \frac{n}{2n+\alpha}\frac{(\theta^*)^{2n+\alpha} - \theta_0^{2n+\alpha}}{(\theta^* + \epsilon)^{n+\alpha}(\theta^*)^n}
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} p_n &= \lim_{n\to\infty} \text{E}_{\mathbf{Y}_n|\theta^*}\{\text{Prob}(\theta^* - \epsilon < \theta < \theta^* + \epsilon|\mathbf{Y}_n)\} \\
&= \lim_{n\to\infty} \text{E}_{\mathbf{Y}_n|\theta^*}\{\text{Prob}(\theta < \theta^* + \epsilon|\boldsymbol{Y}_n) - \text{Prob}(\theta < \theta^* - \epsilon|\boldsymbol{Y}_n)\} \\
&= \lim_{n\to\infty} \text{E}_{\mathbf{Y}_n|\theta^*}\{\text{Prob}(\theta < \theta^* + \epsilon|\boldsymbol{Y}_n)\} - \lim_{n\to\infty} \text{E}_{\mathbf{Y}_n|\theta^*}\{\text{Prob}(\theta < \theta^* - \epsilon|\boldsymbol{Y}_n)\} \\
&= 1 - 0 = 1
\end{aligned}$$

We get the same conclusion about posterior consistency as the Monte Carlo study in (a).