Assignment 2

Jimmy Hickey

9/14/2022

1. Assume $Y_1, \dots Y_n \mid \theta \sim$ Normal$(0, \theta)$ and variance parameter has prior $\theta \mid$ Gamma(a, b).

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

$$
\begin{aligned}
p(\theta \mid Y_1, \dots, Y_n) &= f(\boldsymbol{Y} \mid \theta)\pi(\theta) \\
&\propto \theta^{a-1} \exp -b\theta \cdot \theta^{-\frac{n}{2}} \exp\Big\{ - \frac{1}{2\theta} \sum_{i=1}^{n} (Y_i^2) \Big\} \\
&= \theta^{(a-\frac{n}{2})-1} \exp\Big\{ - b\theta - \frac{1}{2\theta} \sum_{i=1}^{n} (Y_i^2) \Big\} \\
&= \theta^{(a-\frac{n}{2})-1} \exp\Big[- \frac{1}{2} \Big\{ 2b\theta + \frac{1}{\theta} \sum_{i=1}^{n} (Y_i^2) \Big\} \Big] \\
&\Rightarrow \theta \mid Y_1, \dots Y_n \sim \text{Generalized Inverse Gaussian}\Big(2b, \sum_{i=1}^{n} (Y_i)^2, a - \frac{n}{2} \Big)
\end{aligned}
$$

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

This is not conjugate because it is not a Gamma distribution.

2. Say $\boldsymbol{Y} = (Y_1, \dots Y_p) \mid \theta \sim$ Multinomial(n, θ) for $\theta = (\theta_1, \dots, \theta_p)$ so that the likelihood is

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

(a) Derive the Jeffreys prior for θ.

We need the Fisher information from the log likelihood function. Also note that we have the restriction that $\sum_{i=1}^{p} \theta_i = 1$,

$$\begin{aligned}\log\{f(\boldsymbol{Y} \mid \theta)\} &\propto \sum_{i=1}^{p} Y_i \log(\theta_i) \\ &=\propto \sum_{i=1}^{p-1} Y_i \log(\theta_i) + Y_p \log(1 - \sum_{i=1}^{p-1} \theta_i).\end{aligned}$$

Now we take derivatives,

$$\frac{\partial \ell}{\partial \theta_i} = \frac{Y_i}{\theta_i} - \frac{Y_p}{1 - \sum_{i=1}^{p-1} \theta_i}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta_i^2} &= -\frac{-Y_i}{\theta_i^2} - \frac{Y_p}{(1 - \sum_{i=1}^{p-1} \theta_i)^2} \\ \frac{\partial^2 \ell}{\partial \theta_i \theta_j} &= -\frac{Y_p}{(1 - \sum_{i=1}^{p-1} \theta_i)^2}.\end{aligned}$$

This gives an information matrix with diagonals

$$-\frac{n}{\theta_i} - \frac{n}{1 - \sum_{i=1}^{p-1} \theta_i}$$

and off diagonals

$$\frac{n}{1 - \sum_{i=1}^{p-1} \theta_i}.$$

This can also be expressed as $I(\theta) = \mathrm{diag}(1/p_i) + \mathbf{1}\mathbf{1}^\top/(1 - \sum_{i=1}^{p-1} \theta_i)$.

This has known determinant, giving a Jefferys prior of $\pi(\theta) = \prod_{i=1}^{p-1} \theta^{1-1/2}$. This is the PDF of a Dirichlet$(\alpha_i = 1/2)$ distribution.

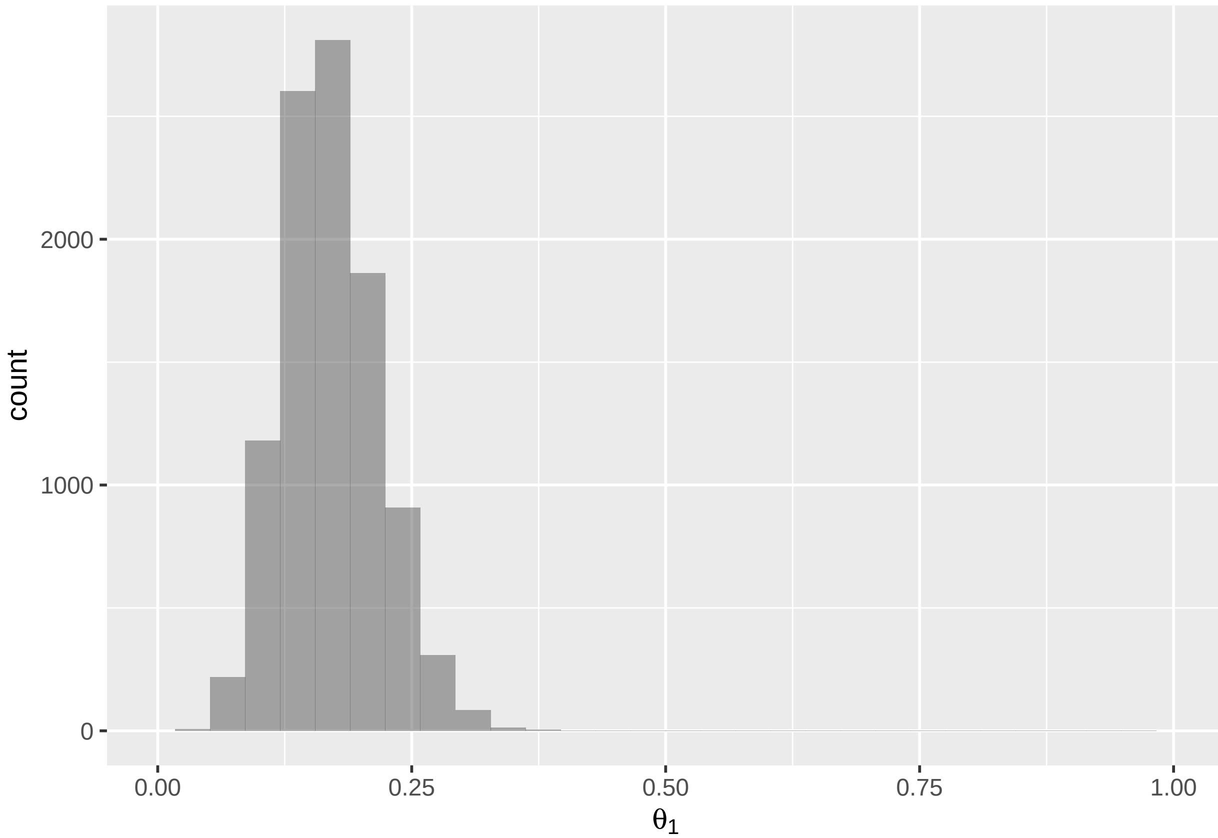
(b) Derive the posterior under the prior in (a).

$$\begin{aligned}p(\boldsymbol{\theta} \mid \boldsymbol{Y}) &\propto \Big(\prod_{i=1}^{p} \theta_i^{Y_i}\Big)\Big(\prod_{i=1}^{p} \theta_i^{-1/2}\Big) \\ &= \prod_{i=1}^{p} \theta_i^{Y_i - 1/2}\end{aligned}$$

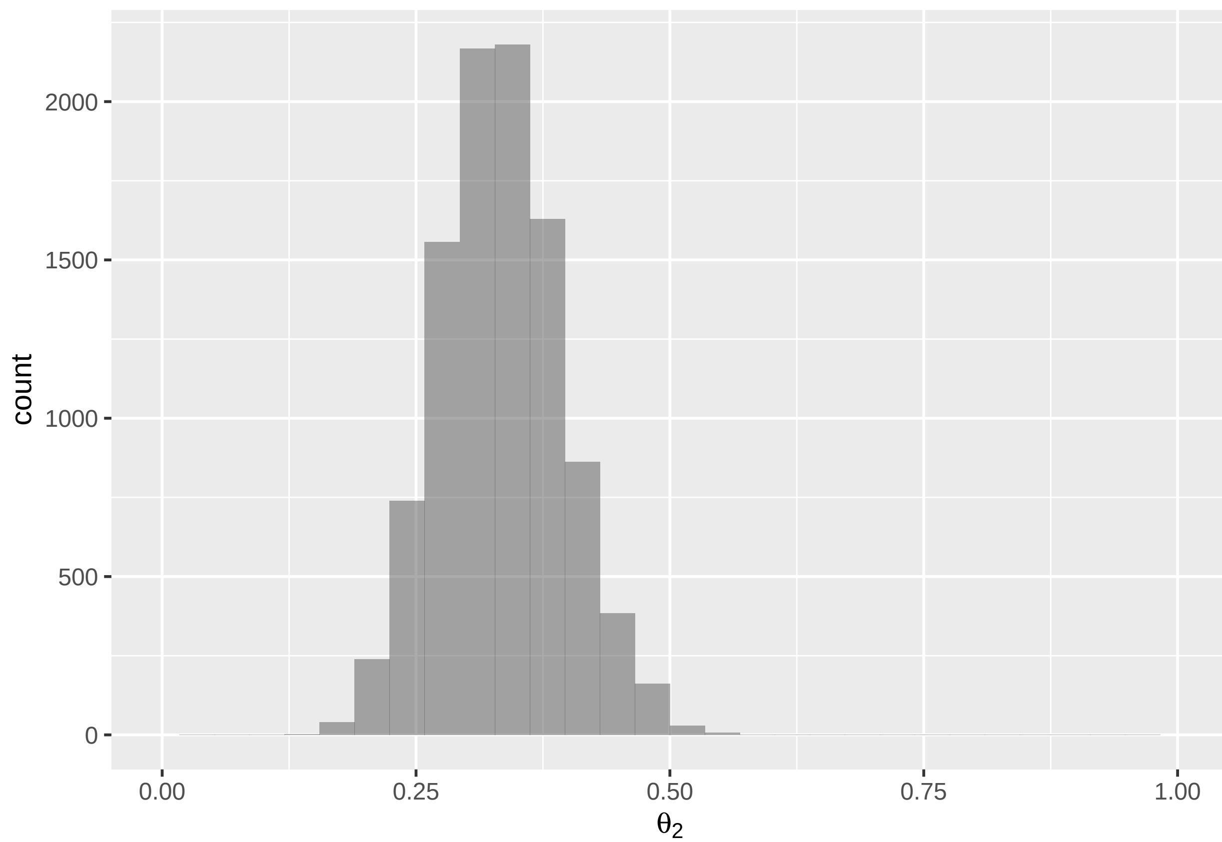
This is the PDF of a Dirichlet($\alpha_i = Y_i + 1/2$) distribution.

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

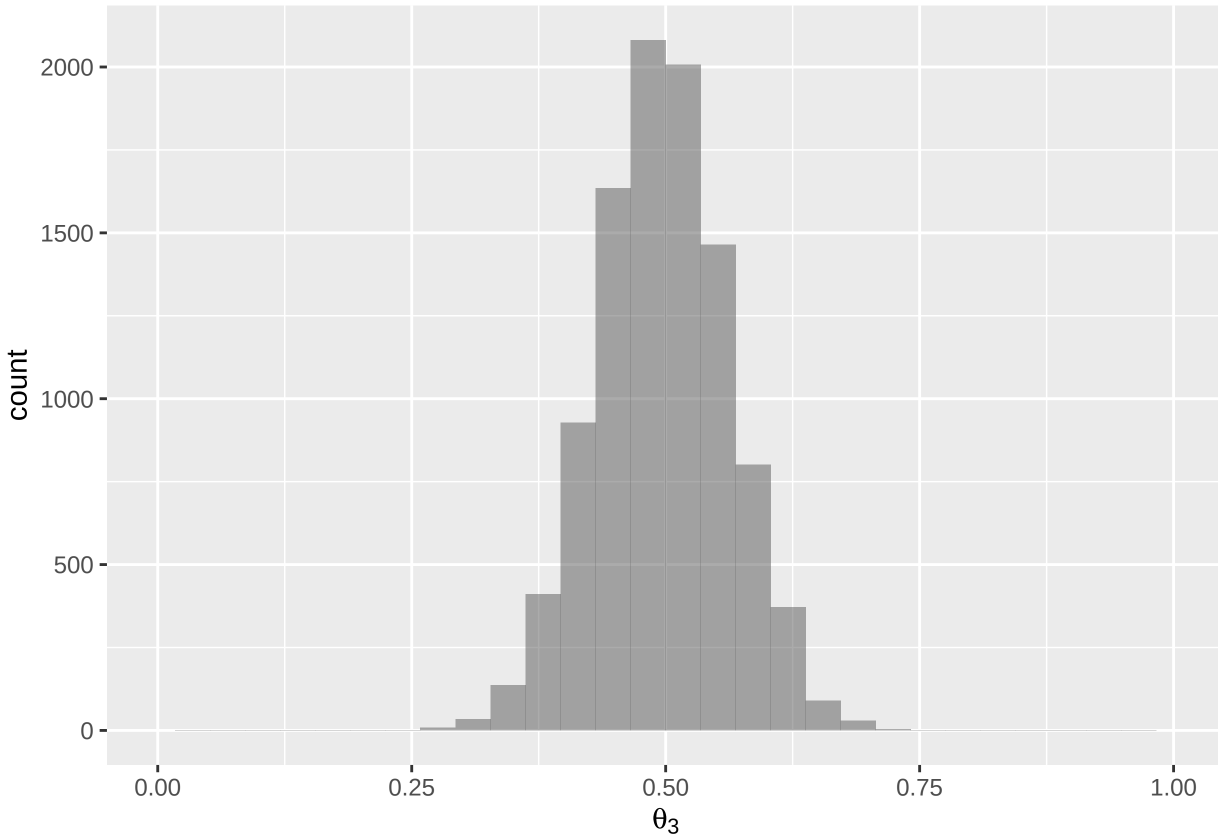
```
# use gtools for dirichlet draws
library(gtools)
library(ggplot2)
library(dplyr)
library(latex2exp)

set.seed(1978)
S = 10000

Y = c(10,20,30)
n = sum(Y)

alpha = Y + 1/2
posterior = rdirichlet(S, alpha)


posterior_df = data.frame(theta1 = posterior[,1],
                                                    theta2 = posterior[,2],
                                                    theta3 = posterior[,3])

ggplot(posterior_df) +
    geom_histogram(aes(x = theta1), alpha=0.5)  +
    xlim(0,1) +
    labs(x = TeX("$\\theta_{1}$"))
```



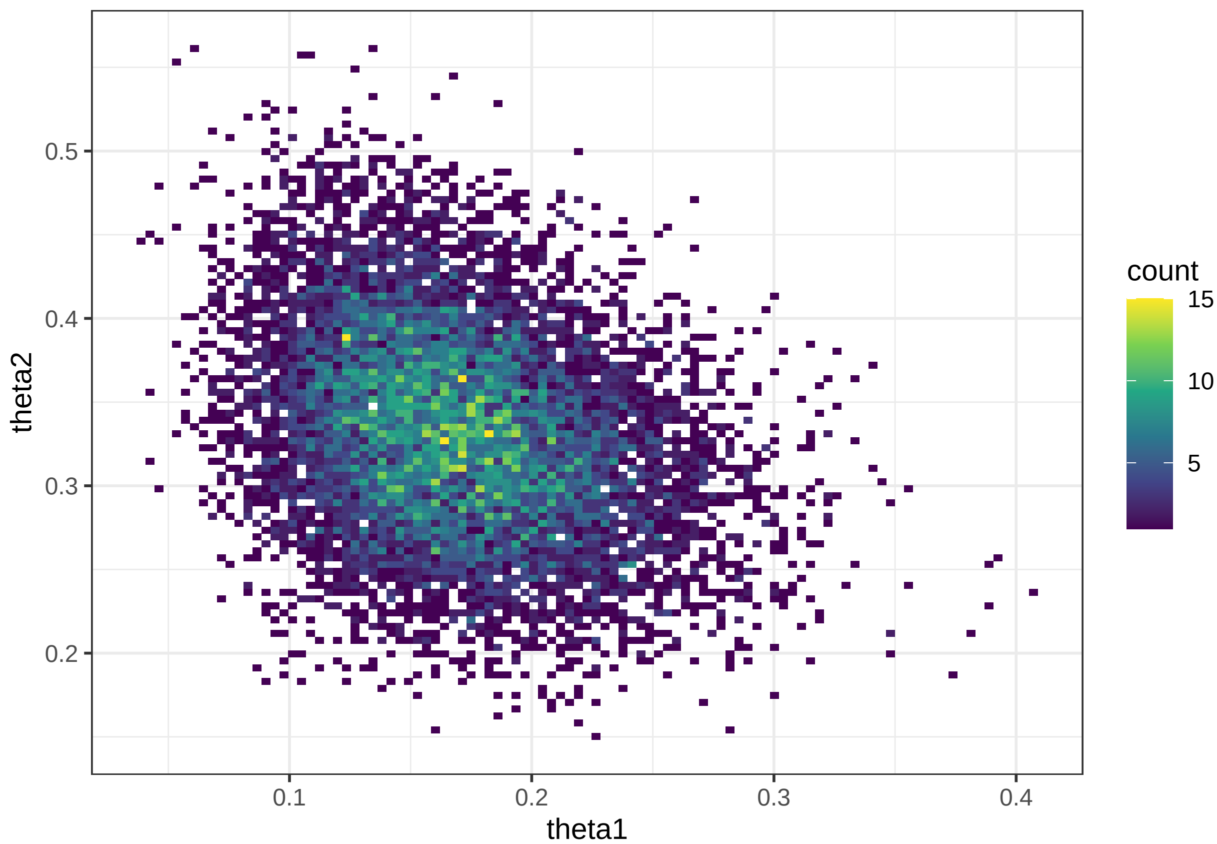

```
ggplot(posterior_df) +
    geom_histogram(aes(x = theta2), alpha=0.5)  +
    xlim(0,1) +
    labs(x = TeX("$\\theta_{2}$"))
```




```
ggplot(posterior_df) +
    geom_histogram(aes(x = theta3), alpha=0.5)  +
    xlim(0,1) +
    labs(x = TeX("$\\theta_{3}$"))
```




```
ggplot(posterior_df, aes(x=theta1, y=theta2) ) +
  geom_bin2d(bins = 100) +
  scale_fill_continuous(type = "viridis") +
  theme_bw()
```




```
summary_df = posterior_df  %>% summarise(
    theta1_mean = mean(theta1),
    theta1_sd = sd(theta1),
    theta1_lower = theta1_mean - qnorm(0.975) * theta1_sd,
    theta1_upper = theta1_mean + qnorm(0.975) * theta1_sd,
    theta2_mean = mean(theta2),
    theta2_sd = sd(theta2),
    theta2_lower = theta2_mean - qnorm(0.975) * theta2_sd,
    theta2_upper = theta2_mean + qnorm(0.975) * theta2_sd,
    theta3_mean = mean(theta3),
    theta3_sd = sd(theta3),
    theta3_lower = theta3_mean - qnorm(0.975) * theta3_sd,
    theta3_upper = theta3_mean + qnorm(0.975) * theta3_sd
)
```

	Posterior Mean	Posterior SD	95% credible set
θ_1	0.170	0.047	(0.077, 0.263)
θ_2	0.333	0.060	(0.216, 0.450)
θ_3	0.458	0.064	(0.371, 0.622)

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

We need to find the Fisher information of the posterior with the condition $\sum_{i=1}^{3} \theta_i = 1$. Take ℓ to the the log posterior,

$$\ell = (Y_1 - \frac{1}{2})\log(\theta_1) + (Y_2 - \frac{1}{2})\log(\theta_2) + (Y_3 - \frac{1}{2})\log(1-\theta_1-\theta_2)$$

$$\frac{\partial \ell}{\partial \theta_1} = \frac{Y_1 - \frac{1}{2}}{\theta_1} - \frac{Y_3 - \frac{1}{2}}{1-\theta_1-\theta_2}$$
$$\frac{\partial \ell}{\partial \theta_2} = \frac{Y_2 - \frac{1}{2}}{\theta_2} - \frac{Y_3 - \frac{1}{2}}{1-\theta_1-\theta_2}$$

$$\frac{\partial^2 \ell}{\partial \theta_1^2} = -\frac{Y_1 - \frac{1}{2}}{\theta_1^{2]}} - \frac{Y_3 - \frac{1}{2}}{(1-\theta_1-\theta_2)^2}$$
$$\frac{\partial^2 \ell}{\partial \theta_2^2} = -\frac{Y_2 - \frac{1}{2}}{\theta_2^{2}} - \frac{Y_3 - \frac{1}{2}}{(1-\theta_1-\theta_2)^2}$$
$$\frac{\partial^2 \ell}{\partial \theta_1 \theta_2} = -\frac{Y_3 - \frac{1}{2}}{(1-\theta_1-\theta_2)^2}.$$

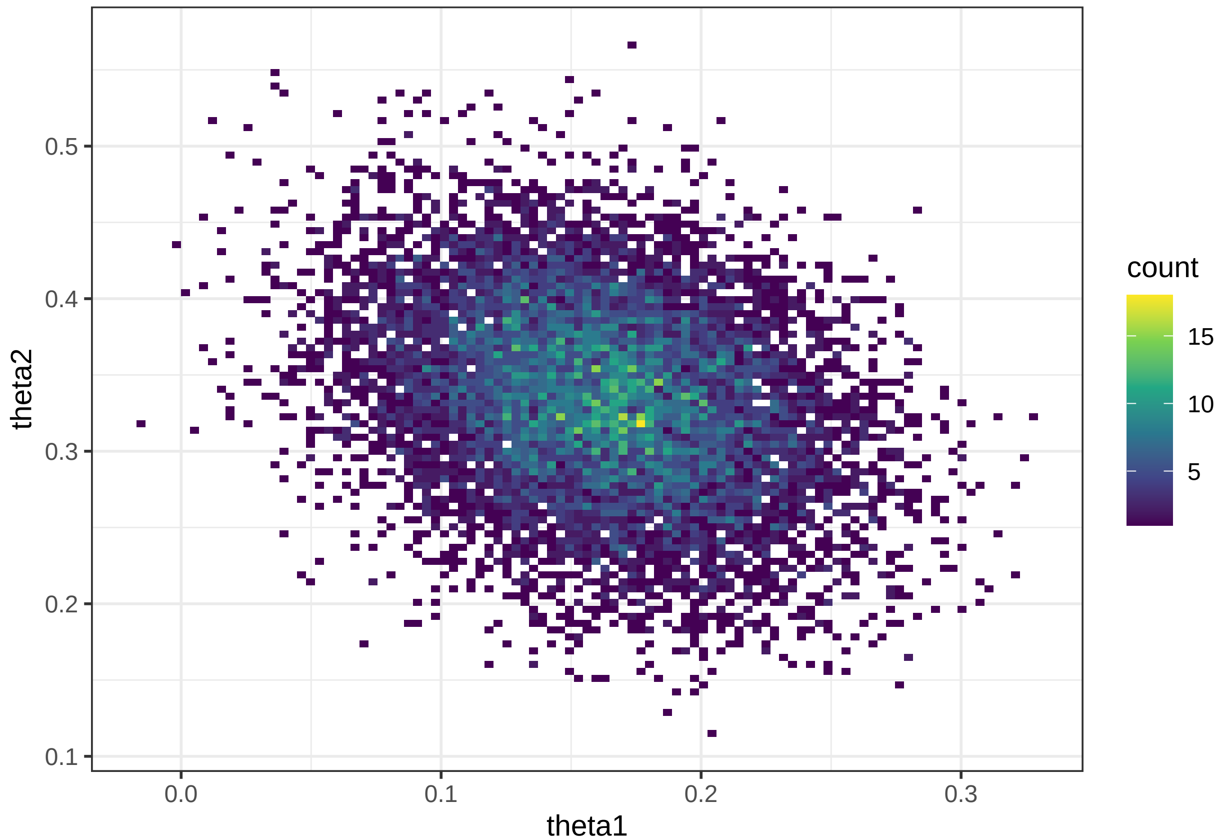
Then our Fisher information looks like

$$I = \begin{bmatrix} \frac{n\theta_1 - \frac{1}{2}}{\theta_1^2} + \frac{n(1-\theta_1-\theta_2)-\frac{1}{2}}{(1-\theta_1-\theta_2)^2} & -\frac{n(1-\theta_1-\theta_2)-\frac{1}{2}}{(1-\theta_1-\theta_2)^2} \\ \frac{n(1-\theta_1-\theta_2)-\frac{1}{2}}{(1-\theta_1-\theta_2)^2} & \frac{n\theta_2 - \frac{1}{2}}{\theta_2^2} + \frac{n(1-\theta_1-\theta_2)-\frac{1}{2}}{(1-\theta_1-\theta_2)^2} \end{bmatrix}.$$

Note we take $\theta_{1,0} = (Y_1 - 1/2)/(Y_1 + Y_2 - 1)$ and $\theta_{2,0} = (Y_2 - 1/2)/(Y_1 + Y_2 - 1)$.

```
library(MASS)
library(gtools)
library(ggplot2)
library(dplyr)
library(latex2exp)

set.seed(1978)
S = 10000

Y = c(10,20,30)
n = sum(Y)
theta10 = (Y[1] - 1/2) / (Y[1] + Y[2] +Y[3] -3/2)
theta20 = (Y[2] - 1/2) / (Y[1] + Y[2] +Y[3] -3/2)

mu = c( theta10, theta20)
I = matrix(c(
    (n * theta10 - 1/2) / (theta10^2), 0,
    0, (n * theta20 - 1/2) / (theta20^2)
), 2, 2) + ( n*( 1 - theta10 - theta20) - 1/2 ) / ((1 - theta10 - theta20)^2)


posterior = mvrnorm(S, mu, solve(I))
```

```r
posterior_df = data.frame(theta1 = posterior[,1], theta2 = posterior[,2])

ggplot(posterior_df, aes(x=theta1, y=theta2) ) +
  geom_bin2d(bins = 100) +
  scale_fill_continuous(type = "viridis") +
  theme_bw()
```




```r
summary_df = posterior_df  %>% summarise(
    theta1_mean = mean(theta1),
    theta1_sd = sd(theta1),
    theta1_lower = theta1_mean - qnorm(0.975) * theta1_sd,
    theta1_upper = theta1_mean + qnorm(0.975) * theta1_sd,
    theta2_mean = mean(theta2),
    theta2_sd = sd(theta2),
    theta2_lower = theta2_mean - qnorm(0.975) * theta2_sd,
    theta2_upper = theta2_mean + qnorm(0.975) * theta2_sd
)
```

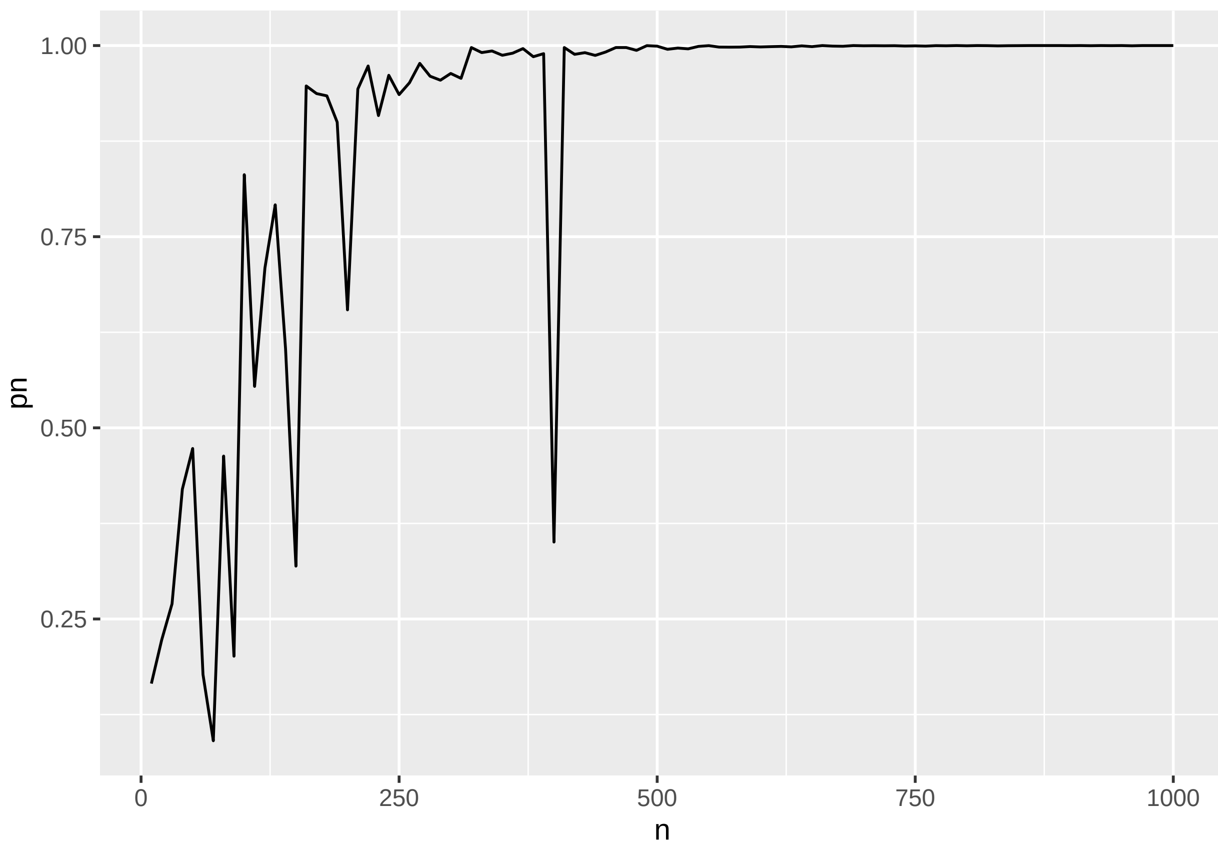
	Asymp Mean	Asymp SD	95% credible set
θ_1	0.169	0.048	(0.067, 0.257)
θ_2	0.335	0.061	(0.214, 0.455)

The asymptotic means and standard deviations are similar to that of the posterior without asymptotics. This is a good approximation!

3. Assume that $Y_i \mid \theta \sim$ Uniform$(0, \theta)$ independent for $i \in \{1, \ldots, n\}$ and prior $\theta \sim$ Pareto(θ_0, α) with support $\theta > \theta_0$ and CDF 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, $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$, let $p_n = E_{\boldsymbol{Y}_n|\theta^*}\{\text{Prob}(\theta^* - \epsilon < \theta < \theta^* + \epsilon \mid \boldsymbol{Y}_n)\}$ for $\epsilon = 0.1$. Compute a Monte Carlo approximation to p_n for each $n \in \{10, 20, \ldots, 1000\}$. Does a plot of n versus p_n suggest posterior consistency? Why?

```
# use for pareto distribution
library("EnvStats")
set.seed(1978)
theta_star = 10
theta0 = alpha = 1
epsilon = 0.1
S = 10000

n = seq(10, 1000, 10)
pn = sapply(n, function(n){
    Y = runif(n, 0, theta_star)
    theta_posterior = rpareto(S, max(max(Y), theta0), n + alpha)
    return(mean(theta_star - epsilon < theta_posterior & theta_posterior < theta_star + epsilon))
})

pn_df = data.frame(pn = pn, n = n)

ggplot(pn_df, aes(x=n, y=pn)) + geom_line()
```



(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)?

In homework 1 we found that the posterior distribution for this likelihood and prior is $\theta \mid Y_1, \dots Y_n \sim \text{Pareto}(\max(y_{(n)}, \theta_0), n + \alpha)$. Then the probability statement inside of the expectation looks like

$$
\begin{aligned}
&\text{Prob}(\theta^* - \varepsilon < \theta < \theta^* + \varepsilon \mid \boldsymbol{Y}_n) = F_{\theta|\boldsymbol{Y}}(\theta^* + \varepsilon) - F_{\theta|\boldsymbol{Y}}(\theta^* - \varepsilon)\\
&\quad = \Big(1 - \Big(\frac{\max(Y_{(n)}, \theta_0)}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big)\mathbb{I}(\theta^* + \varepsilon > \max(Y_{(n)}, \theta_0)) - \Big(1 - \Big(\frac{\max(Y_{(n)}, \theta_0)}{\theta^* - \varepsilon}\Big)^{n+\alpha}\Big)\mathbb{I}(\theta^* - \varepsilon > \max(Y_{(n)}, \theta_0))
\end{aligned}
$$

Now we take the expectation of the first part. Notice that the indicator is always 1.

$$
\begin{aligned}
&E\Big[\Big(1 - \frac{\max(Y_{(n)}, \theta_0)}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big]\\
&\quad = E\Big[1 - \Big(\frac{Y_{(n)}}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big]P(Y_{(n)} > \theta_0) + \Big[1 - \Big(\frac{\theta_0}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big]P(Y_{(n)} < \theta_0)\\
&\quad = \Big[1 - \frac{1}{(\theta^* + \varepsilon)^{n+\alpha}}\int_0^{\theta^*} y^{n+\alpha}\frac{ny^{n-1}}{\theta^{*n}}dy\Big]\Big[1 - \Big(\frac{\theta_0}{\theta_*}\Big)^n\Big] + \Big[1 - \Big(\frac{\theta_0}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big]\Big(\frac{\theta_0}{\theta^*}\Big)^n\\
&\quad = \Big[1 - \frac{\theta^{*(2n+\alpha)}}{\theta^{*n}(\theta^* + \varepsilon)^{n+\alpha}}\frac{n}{2n + \alpha}\Big]\Big[1 - \Big(\frac{\theta_0}{\theta_*}\Big)^n\Big] + \Big[1 - \Big(\frac{\theta_0}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big]\Big(\frac{\theta_0}{\theta^*}\Big)^n
\end{aligned}
$$

Taking the limit of this gives

$$
\begin{aligned}
&\lim_{n\to\infty} E\Big[\Big(1 - \frac{\max(Y_{(n)}, \theta_0)}{\theta^* + \varepsilon}\Big)^{n+\alpha}\Big]\\
&\quad = \Big[1 - 0 \cdot \frac{1}{2}\Big]\Big[1 - 0\Big] + \Big[1 - 0\Big]0\\
&\quad = 1.
\end{aligned}
$$

Now we focus on the second expectation

$$
\begin{aligned}
&E\Big[\Big(1 - \frac{\max(Y_{(n)}, \theta_0)}{\theta^* - \varepsilon}\Big)^{n+\alpha}\mathbb{I}(\theta^* - \varepsilon > \max(Y_{(n)}, \theta_0))\Big]\\
&\quad = E\Big[\Big(1 - \frac{\max(Y_{(n)}, \theta_0)}{\theta^* - \varepsilon}\Big)^{n+\alpha} \mid \theta^* - \varepsilon > \max(Y_{(n)}, \theta_0)\Big]P(\theta^* - \varepsilon > \max(Y_{(n)}, \theta_0))\\
&\quad + E\Big[\Big(1 - \frac{\max(Y_{(n)}, \theta_0)}{\theta^* - \varepsilon}\Big)^{n+\alpha} \mid \theta^* - \varepsilon < \max(Y_{(n)}, \theta_0)\Big]P(\theta^* - \varepsilon < \max(Y_{(n)}, \theta_0))\\
&\quad = E\Big[\Big(1 - \frac{\max(Y_{(n)}, \theta_0)}{\theta^* - \varepsilon}\Big)^{n+\alpha} \mid \theta^* - \varepsilon > \max(Y_{(n)}, \theta_0)\Big]P(\theta^* - \varepsilon > \max(Y_{(n)}, \theta_0)) + 0.
\end{aligned}
$$

Where the last simplification comes from the fact that the CDF of the Pareto distribution is 0 on that interval. Now we can integrate,

$$
\begin{aligned}
&E\Big[\Big(1-\frac{\max(Y_{(n)},\theta_0)}{\theta^*-\varepsilon}\Big)^{n+\alpha}\mathbb{I}(\theta^*-\varepsilon>\max(Y_{(n)},\theta_0))\Big]\\
&\quad=E\Big[1-\Big(\frac{Y_{(n)}}{\theta^*+\varepsilon}\Big)^{n+\alpha}\mid\theta^*-\varepsilon>Y_{(n)}\Big]P(Y_{(n)}>\theta_0)+\Big[1-\Big(\frac{\theta_0}{\theta^*-\varepsilon}\Big)^{n+\alpha}\Big]P(Y_{(n)}<\theta_0)\\
&\quad=\Big[1-\frac{(\theta^*-\varepsilon)^{(2n+\alpha)}}{\theta^{*n}(\theta^*+\varepsilon)^{n+\alpha}}\frac{n}{2n+\alpha}\Big]\Big[1-\Big(\frac{\theta_0}{\theta_*}\Big)^n\Big]+\Big[1-\Big(\frac{\theta_0}{\theta^*-\varepsilon}\Big)^{n+\alpha}\Big]\Big(\frac{\theta_0}{\theta^*}\Big)^n\\
&\quad\to\Big[1-0\cdot\frac{1}{2}\Big]\Big[1-0\Big]+\Big[1-0\Big]1
\end{aligned}
$$

Additionally, since $Y_{(n)}$ is consistent for θ^*, we know that $\lim_{n\to\infty}P(\frac{\max(Y_{(n)},\theta_0)}{\theta^*-\varepsilon}<1)=0$ because the definition of convergence states that for some N, $Y_{(n)}$ will be within ε of θ^*. Thus, $p_n\to 1-0=1$.