# Stat 133: Bayesian Statistical Inference
# Problem Set 2

Anne Amores
Miah Goti-ay
Justine Razon

March 27, 2025

## Problem 1

**Recall our example about the reports of number of sexual partners. Using the same dataset and assuming that the number of sexual partners are a random sample from Poisson($\lambda$), suppose we infer about the average number of sexual partners for women $\lambda$in the population. Use the four different vague priors for $\lambda$ used in that example.**

```r
# Creating a data frame with the counts of sex partners
srsp <- data.frame(
  Count = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 14),
  Men = c(44, 195, 20, 3, 3, 5, 3, 1, 1, 0, 0, 1),
  Women = c(102, 233, 18, 9, 2, 1, 0, 0, 0, 0, 0, 0)
)
print(srsp)
```

```
##    Count Men Women
## 1      0  44   102
## 2      1 195   233
## 3      2  20    18
## 4      3   3     9
## 5      4   3     2
## 6      5   5     1
## 7      6   3     0
## 8      7   1     0
## 9      8   1     0
## 10     9   0     0
## 11    10   0     0
## 12    14   1     0
```

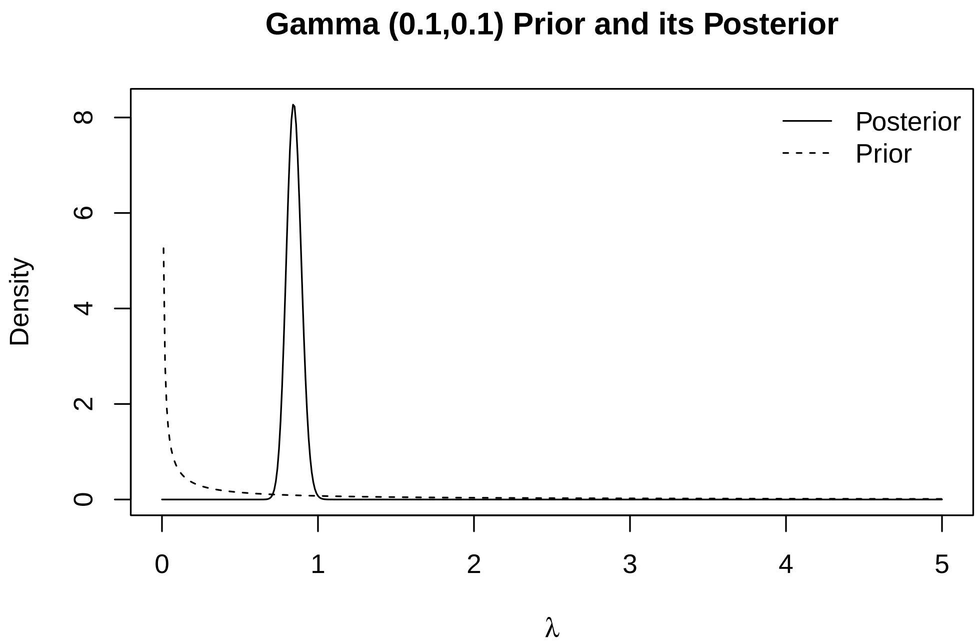
```r
# Creating a data frame with the prior hyperparameters
gamma <- data.frame(
  a = c(0.1, 0.5, 1, 2),
  b = c(0.1, 0.5, 1, 2)
)
# Updating the priors using the data
gamma <- gamma %>%
  mutate(total_count = sum(srsp$Count * srsp$Women),
         n = sum(srsp$Women),
```

```
        a.star = a + total_count,
        b.star = b + n)
print(gamma)
```

```
##     a   b total_count   n a.star b.star
## 1 0.1 0.1         309 365  309.1  365.1
## 2 0.5 0.5         309 365  309.5  365.5
## 3 1.0 1.0         309 365  310.0  366.0
## 4 2.0 2.0         309 365  311.0  367.0
```

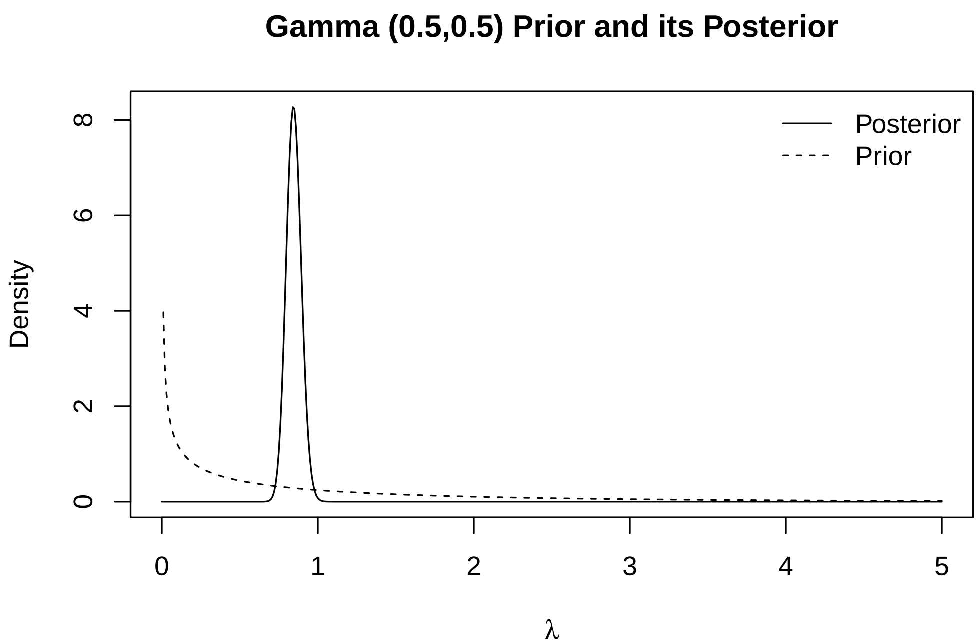
1. **Graph each prior and corresponding posterior distribution using R.**

```
# Plotting the Gamma (0.1,0.1) prior along with its posterior distribution
lambda <- seq(0.00,5.00,0.01)
prior1 <- dgamma(lambda, shape = 0.1, rate = 0.1)
post1 <- dgamma(lambda, shape = 309.1, rate = 365.1)
plot(lambda, post1, xlab = expression(lambda), ylab = "Density",
     type = "l", main = "Gamma (0.1,0.1) Prior and its Posterior")
lines(lambda, prior1, lty = 2)
legend("topright", legend = c("Posterior", "Prior"),
       lty = c(1, 2),
       col = c("black", "black"),
       bty = "n")
```

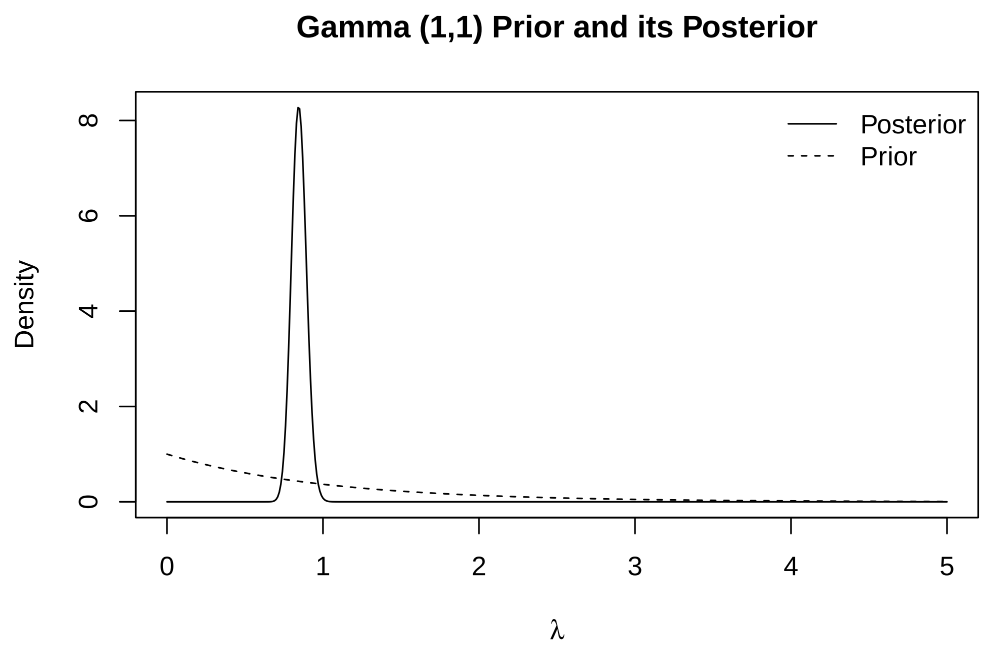



```
# Plotting the Gamma (0.5,0.5) prior along with its posterior distribution
lambda <- seq(0.00,5.00,0.01)
prior2 <- dgamma(lambda, shape = 0.5, rate = 0.5)
post2 <- dgamma(lambda, shape = 309.5, rate = 365.5)
plot(lambda, post2, xlab = expression(lambda), ylab = "Density",
     type = "l", main = "Gamma (0.5,0.5) Prior and its Posterior")
lines(lambda, prior2, lty = 2)
```

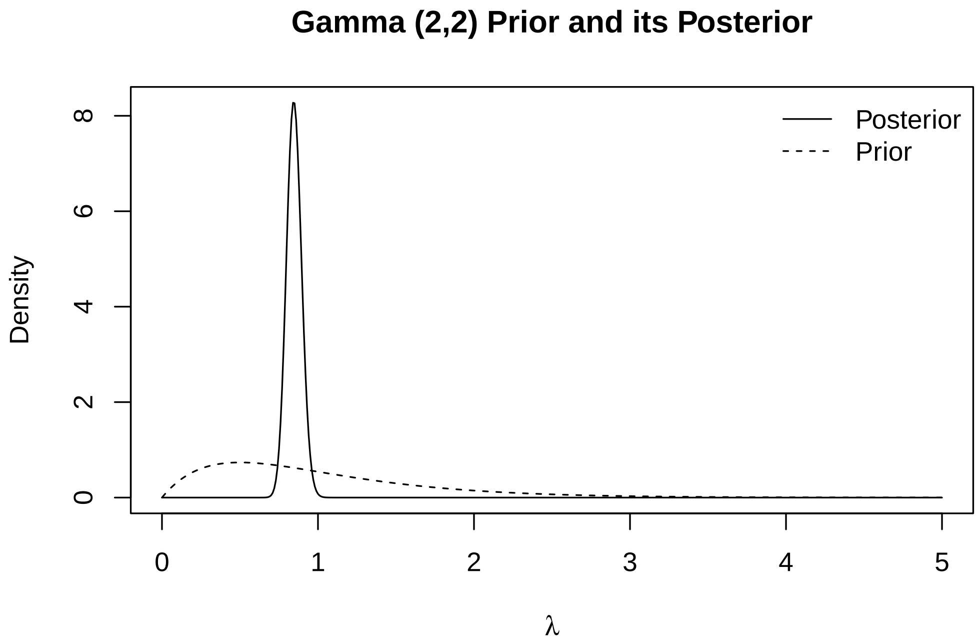
```r
legend("topright", legend = c("Posterior", "Prior"),
       lty = c(1, 2),
       col = c("black", "black"),
       bty = "n")
```




```r
# Plotting the Gamma (1,1) prior along with its posterior distribution
lambda <- seq(0.00,5.00,0.01)
prior3 <- dgamma(lambda, shape = 1, rate = 1)
post3 <- dgamma(lambda, shape = 310, rate = 366)
plot(lambda, post3, xlab = expression(lambda), ylab = "Density",
     type = "l", main = "Gamma (1,1) Prior and its Posterior")
lines(lambda, prior3, lty = 2)
legend("topright", legend = c("Posterior", "Prior"),
       lty = c(1, 2),
       col = c("black", "black"),
       bty = "n")
```




```r
# Plotting the Gamma (2,2) prior along with its posterior distribution
lambda <- seq(0.00,5.00,0.01)
prior4 <- dgamma(lambda, shape = 2, rate = 2)
post4 <- dgamma(lambda, shape = 311, rate = 367)
plot(lambda, post4, xlab = expression(lambda), ylab = "Density",
     type = "l", main = "Gamma (2,2) Prior and its Posterior")
lines(lambda, prior4, lty = 2)
legend("topright", legend = c("Posterior", "Prior"),
       lty = c(1, 2),
       col = c("black", "black"),
       bty = "n")
```




2. **Provide the Bayes estimates for $\lambda$ under the quadratic, symmetric linear, and binary losses.**

```
gamma <- gamma %>%
  mutate(mean.post = a.star / b.star,
         median.post = qgamma(p = 0.5, shape = a.star, rate = b.star),
         mode.post = (a.star - 1) / b.star) %>%
  # Rounding off values of decimals to 4 decimal places
  mutate(across(where(is.numeric), round, 4)) %>%
  print()
```

```
##     a   b total_count   n a.star b.star mean.post median.post mode.post
## 1 0.1 0.1         309 365  309.1  365.1    0.8466      0.8457    0.8439
## 2 0.5 0.5         309 365  309.5  365.5    0.8468      0.8459    0.8440
## 3 1.0 1.0         309 365  310.0  366.0    0.8470      0.8461    0.8443
## 4 2.0 2.0         309 365  311.0  367.0    0.8474      0.8465    0.8447
```

3. **Generate the 95% HPD interval for each posterior distribution.**

```
# HPD interval for posterior of Gamma (0.1,0.1) prior
hpd1 <- round(hpd(posterior.icdf = qgamma, shape = 309.1, rate = 365.1),
    digits = 4)

# HPD interval for posterior of Gamma (0.5,0.5) prior
hpd2 <- round(hpd(posterior.icdf = qgamma, shape = 309.5, rate = 365.5),
    digits = 4)

# HPD interval for posterior of Gamma (1,1) prior
hpd3 <- round(hpd(posterior.icdf = qgamma, shape = 310, rate = 366),
    digits = 4)

# HPD interval for posterior of Gamma (2,2) prior
```

```
hpd4 <- round(hpd(posterior.icdf = qgamma, shape = 311, rate = 367),
    digits = 4)

hpd_intervals <- data.frame(
  hpd.lower95 = c(hpd1[1], hpd2[1], hpd3[1], hpd4[1]),
  hpd.upper95 = c(hpd1[2], hpd2[2], hpd3[2], hpd4[2]))

gamma %>%
 bind_cols(hpd_intervals) %>%
  select(-mean.post, -median.post, -mode.post)
```

```
##     a   b total_count   n a.star b.star hpd.lower95 hpd.upper95
## 1 0.1 0.1         309 365  309.1  365.1      0.7531      0.9417
## 2 0.5 0.5         309 365  309.5  365.5      0.7533      0.9418
## 3 1.0 1.0         309 365  310.0  366.0      0.7536      0.9420
## 4 2.0 2.0         309 365  311.0  367.0      0.7541      0.9423
```

4. **Assess the evidence about $H_0 : \lambda \leq 1$ and $H_1 : \lambda > 1$ for each posterior distribution using Bayes factor.**

```
gamma %>%
  select(a, b, a.star, b.star) %>%
  mutate(
    p.h1.prior = 1 - pgamma(q = 1, shape = a, rate = b),
    p.h1.post = 1 - pgamma(q = 1, shape = a.star, rate = b.star)) %>%
  mutate(
    bf.10 = (p.h1.post)/(1-p.h1.post)/(p.h1.prior/(1-p.h1.prior)),
    bf.01 = 1/bf.10) %>%
  select(-p.h1.prior, -p.h1.post) %>%
    # Rounding off values of decimals to 4 decimal places
    mutate(across(where(is.numeric), round, 4)) %>%
  print()
```

```
##     a   b a.star b.star  bf.10    bf.01
## 1 0.1 0.1  309.1  365.1 0.0059 169.9986
## 2 0.5 0.5  309.5  365.5 0.0027 376.9806
## 3 1.0 1.0  310.0  366.0 0.0021 468.6151
## 4 2.0 2.0  311.0  367.0 0.0018 542.4913
```

For all four Gamma priors, the Bayes factor for $H_1$ is less than 1, meaning there is negative evidence for the alternative hypothesis, based on the scale suggested by Jeffreys (1961). Moreover, the Bayes factor for $H_0$, $B_{01} > 10^2$ for all four Gamma priors, indicating that there is **decisive evidence in support of the null hypothesis** that the average number of sexual partners for women is less than 1.

# Problem 2

**Recall our exercise about the proportion of voters who will support each of the 66 senatorial candidates. Suppose that we model the number of voters who will vote for the $i$th candidate from the February 2025 *Pulso ng Bayan* Pre-Electoral national survey using Binomial($n = 2400, \theta_i$). Use the posterior distribution for each candidate from our exercise as the prior distribution for each candidate.**

```r
# Importing the results of the PnB survey for Jan 2025
pnb.jan <- read_excel(
    path = "C:/Users/amore_6ou078y/OneDrive/Documents/Pulso ng Bayan.xlsx",
    sheet = "Jan 2025")
head(pnb.jan)
```

```
## # A tibble: 6 x 5
##   candidate                 party  aware  vote rank
##   <chr>                     <chr>  <dbl> <dbl> <chr>
## 1 TULFO, ERWIN              LAKAS     99  62.8 1
## 2 GO, BONG GO               PDPLBN    99  50.4 2-3
## 3 SOTTO, TITO               NPC       99  50.2 2-4
## 4 TULFO, BEN BITAG          IND       97  46.2 3-8
## 5 CAYETANO, PIA             NP        98  46.1 4-8
## 6 BONG REVILLA, RAMON, JR. LAKAS     98  46   4-8
```

```r
# Importing the results of the PnB survey for Feb 2025
pnb.feb <- read_excel(
    path = "C:/Users/amore_6ou078y/OneDrive/Documents/Pulso ng Bayan.xlsx",
    sheet = "Feb 2025")
head(pnb.feb)
```

```
## # A tibble: 6 x 5
##   candidate                 party  aware  vote rank
##   <chr>                     <chr>  <dbl> <dbl> <chr>
## 1 GO, BONG GO               PDPLBN   100  58.1 1-2
## 2 TULFO, ERWIN              LAKAS     98  56.6 1-2
## 3 SOTTO, TITO               NPC      100  49   3-4
## 4 BONG REVILLA, RAMON, JR. LAKAS    100  46.1 3-6
## 5 DELA ROSA, BATO           PDPLBN   100  44.3 4-7
## 6 REVILLAME, WILLIE WIL     IND       98  42.3 4-9
```

```r
# Updating pnb.jan with the posterior parameters computed in our class exercise
pnb.jan <- pnb.jan %>%
  mutate(count = ceiling(vote/100*2400),
         a.star = 1 + count,
         b.star = 1 + 2400 - count)
# Creating a new column in pnb.feb for the no. of votes per candidate
pnb.feb <- pnb.feb %>%
  mutate(count= ceiling(vote/100*2400))
head(pnb.jan)
```

```
## # A tibble: 6 x 8
##   candidate                 party  aware  vote rank  count a.star b.star
##   <chr>                     <chr>  <dbl> <dbl> <chr> <dbl>  <dbl>  <dbl>
## 1 TULFO, ERWIN              LAKAS     99  62.8 1      1508   1509    893
## 2 GO, BONG GO               PDPLBN    99  50.4 2-3    1210   1211   1191
## 3 SOTTO, TITO               NPC       99  50.2 2-4    1205   1206   1196
## 4 TULFO, BEN BITAG          IND       97  46.2 3-8    1109   1110   1292
## 5 CAYETANO, PIA             NP        98  46.1 4-8    1107   1108   1294
## 6 BONG REVILLA, RAMON, JR. LAKAS     98  46   4-8    1104   1105   1297
```

```r
head(pnb.feb)
```

```
## # A tibble: 6 x 6
##   candidate                 party  aware  vote rank  count
##   <chr>                     <chr>  <dbl> <dbl> <chr> <dbl>
```

```
## 1 GO, BONG GO              PDPLBN   100  58.1 1-2    1395
## 2 TULFO, ERWIN             LAKAS     98  56.6 1-2    1359
## 3 SOTTO, TITO              NPC      100  49   3-4    1176
## 4 BONG REVILLA, RAMON, JR. LAKAS    100  46.1 3-6    1107
## 5 DELA ROSA, BATO          PDPLBN   100  44.3 4-7    1064
## 6 REVILLAME, WILLIE WIL    IND       98  42.3 4-9    1016
```

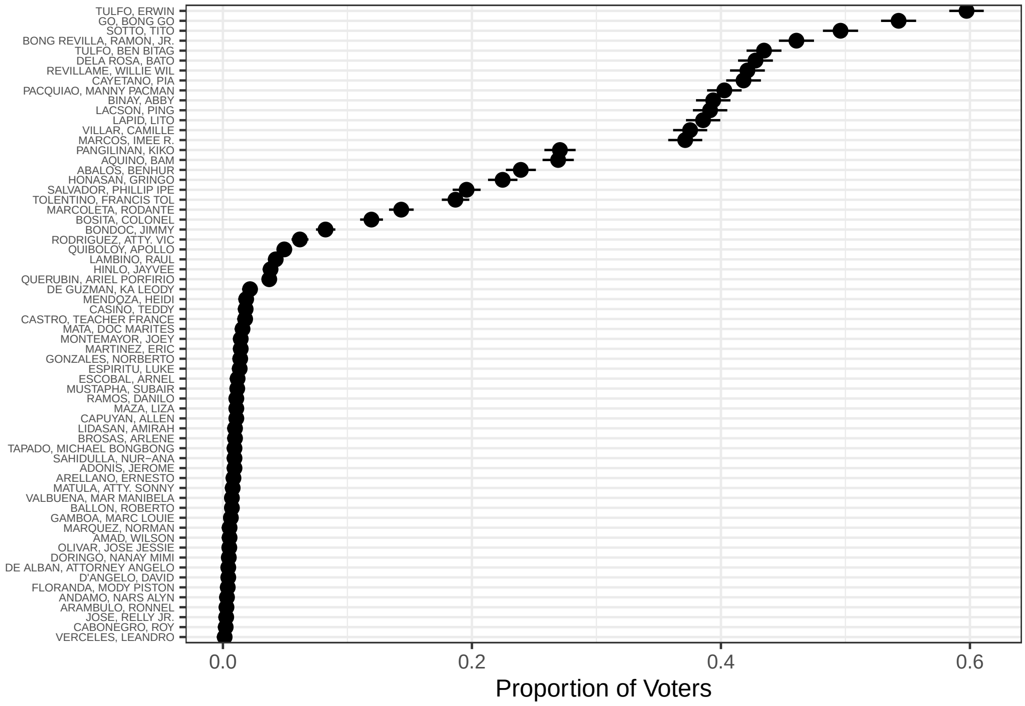
```r
# Merging pnb.jan and pnb.feb
pnb.feb <- pnb.feb %>%
  left_join(pnb.jan %>% select(candidate, a.star, b.star),
            by = "candidate",
            suffix = c("", "_jan")) %>%
  mutate(
    a_prior = ifelse(!is.na(a.star), a.star, 1),
    b_prior = ifelse(!is.na(b.star), b.star, 1),

    # Computing new a.star and b.star using priors
    a.star_new = a_prior + count,
    b.star_new = b_prior + (2400 - count)
  ) %>%
  select(-party, -aware, -vote, -rank, -a.star, -b.star)
head(pnb.feb)
```

```
## # A tibble: 64 x 6
##    candidate                count a_prior b_prior a.star_new b.star_new
##    <chr>                    <dbl>   <dbl>   <dbl>      <dbl>      <dbl>
##  1 GO, BONG GO               1395    1211    1191       2606       2196
##  2 TULFO, ERWIN              1359    1509     893       2868       1934
##  3 SOTTO, TITO               1176    1206    1196       2382       2420
##  4 BONG REVILLA, RAMON, JR.  1107    1105    1297       2212       2590
##  5 DELA ROSA, BATO           1064     990    1412       2054       2748
##  6 REVILLAME, WILLIE WIL     1016    1007    1395       2023       2779
##  7 TULFO, BEN BITAG           977    1110    1292       2087       2715
##  8 PACQUIAO, MANNY PACMAN     958     976    1426       1934       2868
##  9 LAPID, LITO                946     906    1496       1852       2950
## 10 BINAY, ABBY                903     988    1414       1891       2911
## # i 54 more rows
```

1. **Visualize the 95% HPD intervals for senatorial candidates with the corresponding posterior median using `geom_pointrange`. Order the candidates by increasing posterior median.**

```r
pnb.feb %>%
  mutate(median = qbeta(p = 0.5, shape1 = a.star_new, shape2 = b.star_new)) %>%
  rowwise() %>%
  mutate(hpd.lower95 = hpd(posterior.icdf = qbeta,
                           shape1 = a.star_new,
                           shape2 = b.star_new,
                           conf = 0.95)[1],
         hpd.upper95 = hpd(posterior.icdf = qbeta,
                           shape1 = a.star_new,
                           shape2 = b.star_new,
                           conf = 0.95)[2]) %>%
  ggplot() +
  geom_pointrange(aes(xmin = hpd.lower95,
                      xmax = hpd.upper95,
                      x = median,
```

```r
                         y = reorder(candidate, median))) +
  xlab("Proportion of Voters") +
  theme_bw() +
  theme(axis.title.y = element_blank())
```




2. **Who among the senatorial candidates is supported by the majority of voters? Support your answer using Bayes factor.**

```r
# Computing for Bayes factor
pnb.feb %>%
  mutate(
    p.h1.prior = 1 - pbeta(q = 0.5, shape1 = a_prior, shape2 = b_prior),
    p.h1.post = 1 - pbeta(q = 0.5, shape1 = a.star_new, shape2 = b.star_new),
    bf.10 = ifelse(p.h1.post == 0 | p.h1.post == 1, NA,
          (p.h1.post / (1 - p.h1.post)) / (p.h1.prior / (1 - p.h1.prior))),
    bf.01 = ifelse(is.na(bf.10), NA, 1 / bf.10)) %>%
  select(candidate, p.h1.prior, p.h1.post, bf.10, bf.01) %>%
  # Rounding off values of decimals to 4 decimal places
  mutate(across(where(is.numeric), round, 4)) %>%
  head()
```

```
## # A tibble: 6 x 5
##   candidate                 p.h1.prior p.h1.post    bf.10   bf.01
##   <chr>                          <dbl>     <dbl>    <dbl>   <dbl>
## 1 GO, BONG GO                    0.658     1      3.24e+8    0
## 2 TULFO, ERWIN                   1         1        NA      NA
## 3 SOTTO, TITO                    0.581     0.292  2.97e-1   3.36
## 4 BONG REVILLA, RAMON, JR.       0         0        5e-4    1838
## 5 DELA ROSA, BATO                0         0        NA      NA
## 6 REVILLAME, WILLIE WIL          0         0        NA      NA
```

According to the scale suggested by Jeffreys (1961) for interpreting the Bayes factor, it is only for **Bong Go** that there is decisive evidence that the majority of voters supports his candidacy, with a Bayes factor for $H_1$ of $B_{10} \approx 3.24e + 8$.

However, it is important to note that while the Bayes factor for $H_1$ for Erwin Tulfo cannot be assessed based on the scale, this is due to both his prior and posterior probability of $H_1$ being approximately equal to 1. As a result, the Bayes factor calculation leads to an indeterminate form (0/0), making it mathematically undefined and returning an NA. Despite this, the fact that the posterior probability remains virtually equal to 1 strongly suggests that **Erwin Tulfo** has decisive support from the majority of voters.

# Problem 3

**Suppose that MMDA is studying the number of traffic accidents per month occurring at a certain intersection. They collected data for the past two years, given as follows:**

2 4 3 1 1 3 2 2 4 0 5 2 5 2 4 4 3 1 3 8 4 2 1 1

**The researchers have prior information on the Poisson parameter and believe that the Poisson parameter is 3 on the average with a standard deviation of 3. A gamma conjugate prior is used to represent the prior information.**

1. **What is the 95% HPD interval for the Poisson parameter?**

$$y_1, y_2, ..., y_{24} | \lambda \sim Poisson(\lambda)$$

$$\lambda \sim Gamma(a, b), \quad b > 0$$

To compute the hyperparameters of the prior Gamma distribution, note that the mean of the gamma distribution is $E(\lambda) = \frac{a}{b}$ and its variance is $Var(\lambda) = \frac{a}{b^2}$. Thus, based on the given,

$$E(\lambda) = 3 = \frac{a}{b} \tag{1}$$

$$Var(\lambda) = 3^2 = \frac{a}{b^2} \tag{2}$$

From (1), we can express $a$ as $a = 3b$. Substituting this expression for $a$ into (2),

$$\frac{3b}{b^2} = 9$$

$$\rightarrow b = \frac{1}{3}$$

$$\xrightarrow{(1)} a = 3\left(\frac{1}{3}\right) = 1$$

```r
accidents <- c(2, 4, 3, 1, 1, 3, 2, 2, 4, 0, 5, 2, 5, 2, 4, 4, 3, 1, 3, 8,
               4, 2, 1, 1)
n <- length(accidents)

# Prior Parameters
a_prior <- 1
b_prior <- 1/3

# Posterior Parameters
a_post <- a_prior + sum(accidents)
b_post <- b_prior + n

# 95% HPD Interval
hpd_interval <- hpd(qgamma, shape = a_post, rate = b_post, conf = 0.95)
```

```
cat(sprintf("95%% HPD Interval: (%.4f, %.4f)", hpd_interval[1],
              hpd_interval[2]))
```

```
## 95% HPD Interval: (2.1450 3.4673)
```

2. **The researchers wanted to test whether the mean number of traffic accidents per month is equal to 3 or not at 5% level of significance. What is the conclusion when the hypothesis test is performed?**

$$H_o : \lambda = 3$$
$$H_1 : \lambda \neq 3$$

When dealing with a point null hypothesis, it suffices to check whether or not the hypothesized value of the parameter belongs in the HPD interval to decide if $H_0$ should be rejected. The computed 95% HPD interval for $\lambda$ in the previous item was (2.1450, 3.4673). Since the interval contains 3, the hypothesized value of $\lambda$, we **do not reject $H_0$.** There is insufficient evidence to say that the mean number of traffic accidents per month is not equal to 3 at 5% level of significance.

3. **Consequently, the researchers wanted to test whether the mean number of traffic accidents per month is greater than 3 or not.**

**(a) At 5% level of significance, what is the conclusion when the hypothesis test is performed?**

Testing $H_0 : \lambda \leq 3$ vs. $H_1 : \lambda > 3$ at $\alpha = 0.05$,

```
p.h0 = pgamma(q = 3, shape = a_post, rate = b_post)
reject.ho = ifelse(p.h0<0.05, "yes","no")
cat("reject.ho:", reject.ho)
```

```
## reject.ho: no
```

At 5% level of significance, we do not reject $H_o$. There is insufficient evidence to say that the mean number of traffic accidents per month is greater than 3.

**(b) Using Bayes factor, what is the conclusion when the hypothesis test is performed?**

```
# Computing for Bayes factor
p.h1.prior = 1 - pgamma(q = 3, shape = a_prior, rate = b_prior)
p.h1.post = 1 - pgamma(q = 3, shape = a_post, rate = b_post)
bf.10 = (p.h1.post / (1 - p.h1.post)) / (p.h1.prior / (1 - p.h1.prior))
cat("Bayes Factor:", round(bf.10, digits = 4))
```

```
## Bayes Factor: 0.6151
```

Since $B_{10} < 1$, we have negative evidence for the alternative hypothesis that $\lambda > 3$, i.e., **the null hypothesis is supported and should not be rejected.**