# Homework 3 — Bayesian Data Analysis

Kári Hlynsson
kah76@hi.is

Source codes are available at
https://github.com/lvthnn/STAE529M

All problems related to random-generation
were evaluated with seed set to 42.

# Exercise 1: Assembly lines

An assembly line relies on accurate measurements from an image recognition algorithm at the first stage of the process. It is known that the algorithm is unbiased, so assume that measurements follow a normal distribution with mean zero, $Y_i \mid \sigma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Some errors are permissible, but if $\sigma$ exceeds the threshold $c$ then the algorithm must be replaced. You make $n = 20$ measurements and observe $\sum_{i=1}^n Y_i = -2$ and $\sum_{i=1}^n Y_i^2 = 15$ and conduct a Bayesian analysis with InvGamma$(a, b)$ prior. Compute the posterior probability that $\sigma > c$ for

(a) $c = 1$ and $a = b = 0.1$

(b) $c = 1$ and $a = b = 1.0$

(c) $c = 2$ and $a = b = 0.1$

(d) $c = 2$ and $a = b = 1.0$

For each $c$, compute the ratio of probabilities for the two priors (i.e. $a = b = 0.1$ and $a = b = 1.0$). Which, if any, of the results are sensitive to the prior?

## Solution

Assume $\mathbf{Y} \mid \sigma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ where $\mathbf{Y} := (Y_1, \ldots, Y_n)$. By the independence of the $Y_i$'s, we have

$$
\begin{aligned}
\mathcal{L}(\mathbf{Y} \mid \sigma^2) &= \prod_{i=1}^n \mathcal{L}(Y_i \mid \sigma^2) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{Y_i^2}{2\sigma^2}\right] \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left[-\frac{\sum_{i=1}^n Y_i^2}{2\sigma^2}\right] \\
&\propto \frac{1}{\sigma^n} \exp\left[-\frac{\sum_{i=1}^n Y_i^2}{2\sigma^2}\right]
\end{aligned}
$$

Furthermore, we have $\pi(\sigma^2) \sim \text{InvGamma}(a, b)$, so that

$$
\begin{aligned}
p(\sigma^2 \mid \mathbf{Y}) &\propto \mathcal{L}(\mathbf{Y} \mid \sigma^2)\pi(\sigma^2) \\
&\propto \frac{1}{\sigma^n} \exp\left[-\frac{\sum_{i=1}^n Y_i^2}{2\sigma^2}\right] (\sigma^2)^{-a-1} \exp\left[-b/\sigma^2\right] \\
&\propto (\sigma^2)^{-n/2-a-1} \exp\left[-\left(\frac{\sum_{i=1}^n Y_i^2 + 2b}{2\sigma^2}\right)\right] \\
&\propto (\sigma^2)^{-n/2-a-1} \exp\left[-\left(\frac{\sum_{i=1}^n Y_i^2 + 2b}{2}\right)\frac{1}{\sigma^2}\right] \\
&\propto (\sigma^2)^{-\alpha-1} \exp\left[-\beta/\sigma^2\right],
\end{aligned}
$$

where $\alpha = n/2 - a$ and $\beta = (\sum_{i=1}^n Y_i^2 + 2b)/2$. Therefore $p(\sigma^2 \mid \mathbf{Y}) \sim \text{InvGamma}(\alpha, \beta)$. Thus given the data we have

$$
p(\sigma^2 \mid \mathbf{Y}) \sim \text{InvGamma}\left(10 + a, \frac{15 + 2b}{2}\right)
$$

We have that $\Pr(\sigma > c) = \Pr(\sigma^2 > c^2)$ due to the monotonicity of the quadratic transform on the positive interval $(0, \infty]$. Also, using $\sigma$ instead of $\sigma^2$ in the derivation yields the same result, so we take this as a given. We can evaluate this in R using:

```
> n <- 20
> ssy <- 15
> sy <- -2
> ct <- 2
> as <- c(0.1, 1.0)
> bs <- c(0.1, 1.0)
> cs <- rep(1:ct, each = 2)
> params <- cbind(data.frame(as, bs), cs)
> colnames(params) <- c("a", "b", "c")
> probs <- sapply(seq_len(nrow(params)), function(i) {
+   a <- params[i,"a"]
+   b <- params[i,"b"]
+   c <- params[i,"c"]
+   pinvgamma(q = c^2, shape = 0.5 * n + a, rate = 0.5 * 15 + b, lower.tail = FALSE)
+ })
```

Print the probabilities:

```
> probs
[1] 2.249838e-01 2.366380e-01 2.560058e-05 1.445810e-05
```

Additionally, we can evaluate the sensitivity by

```
> sens <- sapply(seq(1, nrow(params), by = 2), function(i) probs[i]/probs[i + 1])
> sens
[1] 0.9507509 1.7706743
```

We see that the sensitivity is greater for $c = 2$. In fact, the sensitivity increases proportionately to higher values of $c$. By running the entire code again but with `ct` (see code at top) set to 10, we can see the sensitivities for $c = 1, \ldots, 10$. The probabilities generated with `ct <- 10` are

```
> probs
 [1] 2.249838e-01 2.366380e-01 2.560058e-05 1.445810e-05 1.835070e-08
 [6] 5.636127e-09 7.675992e-11 1.465575e-11 9.883829e-13 1.287973e-13
[11] 2.704577e-14 2.565666e-15 1.264377e-15 9.146933e-17 8.806224e-17
[16] 5.030629e-18 8.343527e-18 3.866830e-19 1.009459e-18 3.878135e-20
```

Similarly, the sensitivities are

```
> sens
 [1]  0.9507509  1.7706743  3.2559063  5.2375282  7.6739407 10.5414203
 [7] 13.8229626 17.5052153 21.5771746 26.0294917
```

The ratio of probabilities as a function of $c$ is shown in Figure 1 on the next page.

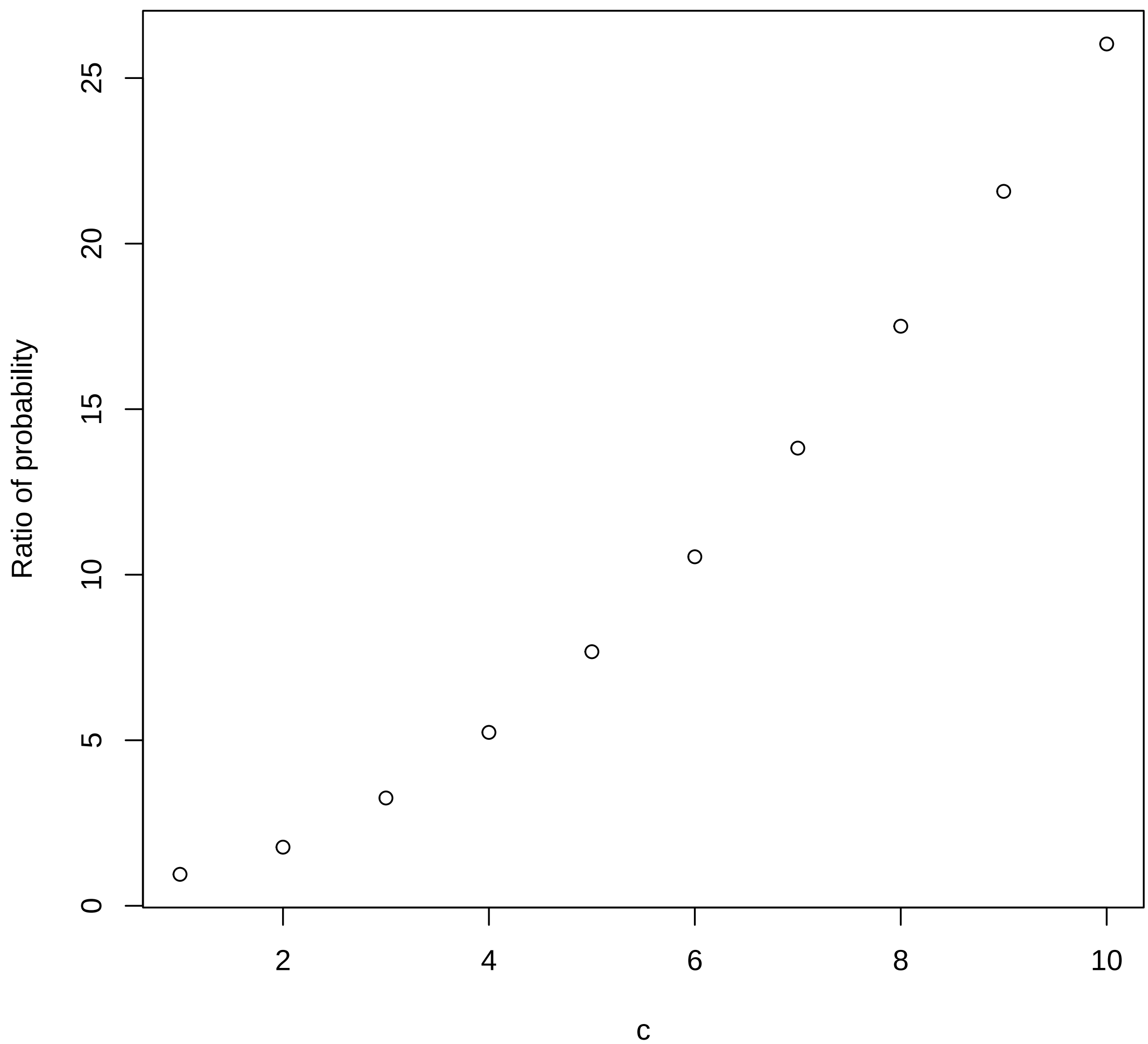


Figure 1: Ratio of probabilities as a function of $c$

# Exercise 2: Jeffreys' prior

Say $Y \mid \lambda \sim \text{Gamma}(1, \lambda)$.

(a) Derive and plot the Jeffreys' prior for $\lambda$.

(b) Is this prior proper?

(c) Derive the posterior and give conditions on $Y$ to ensure it is proper.

## Solution

### Part (a)

We set out to derive Jeffrey's prior for $\lambda$. We have that

$$\pi(\lambda) \propto \sqrt{I(\lambda)}$$

where

$$I(\lambda) = -E_{Y|\lambda}\left[\frac{\partial^2}{\partial\lambda^2}\log\mathcal{L}(Y \mid \lambda)\right] = -E_{Y|\lambda}\left[\frac{\partial^2}{\partial\lambda^2}\log(\lambda) - \lambda Y\right] = E_{Y|\lambda}\left[\frac{1}{\lambda^2}\right] = \frac{1}{\lambda^2},$$

so $\pi(\lambda) \propto \sqrt{1/\lambda^2} = 1/\lambda$. A plot of the prior is shown below in Figure 2.

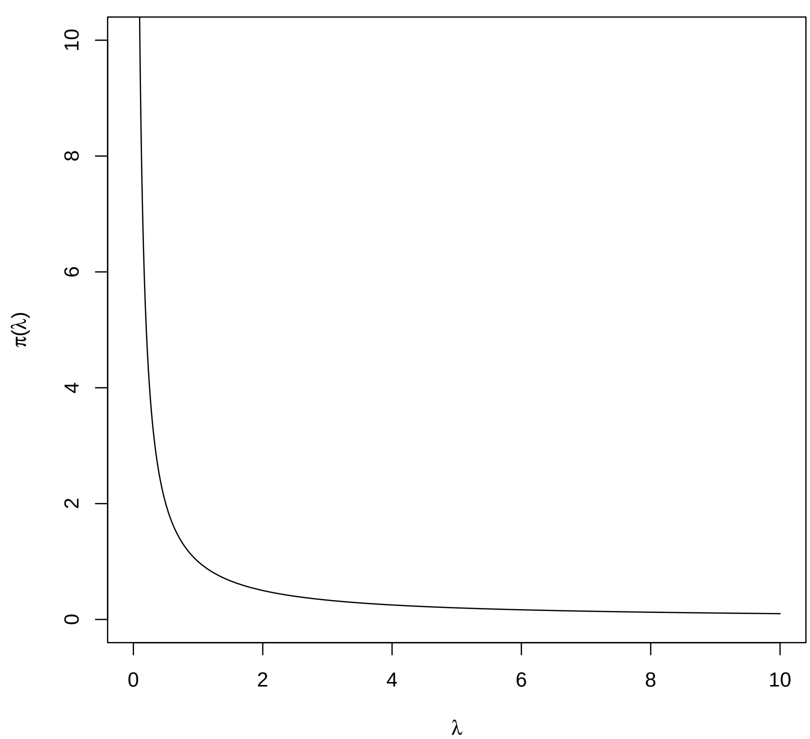


Figure 2: The prior $\pi(\lambda) = 1/\lambda$.

**Part (b)**

The prior $\pi(\lambda)$ is said to be proper if $\int_S \pi(\lambda)\mathrm{d}\lambda = 1$. We have that

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda}{\lambda} = \log(\lambda)\Big|_0^{\infty} &= \lim_{\lambda\to\infty} \left[\log(\lambda) - \log(1/\lambda)\right] \\
&= \lim_{\lambda\to\infty} \left[\log(\lambda) + \log(\lambda)\right] \\
&= \lim_{\lambda\to\infty} \log\left(\lambda^2\right) \\
&= \infty.
\end{aligned}$$

This shows that the prior is not proper.

**Part (c)**

Notice that

$$\begin{aligned}
\int_0^{\infty} p(\lambda \mid Y)\mathrm{d}\lambda = -e^{-\lambda Y}\Big|_0^{\infty} \\
&= e^0 - \lim_{\lambda\to\infty} e^{-\lambda Y} \\
&= 1 - \lim_{\lambda\to\infty} e^{-\lambda Y}.
\end{aligned}$$

The interval converges to 1 if and only if we have $Y > 0$, for else the quantity $-\lambda Y$ is positive or zero which would lead to the density being either zero or minus infinity. Therefore, $Y > 0$. □

# Exercise 3: Survey analysis

The data in the table below are the result of a survey for commuters in 10 countries likely to be affected by a proposed addition of a high occupancy vehicle (HOV) lane.

| County | Approve | Disapprove | County | Approve | Disapprove |
|---|---|---|---|---|---|
| 1 | 12 | 50 | 6 | 15 | 8 |
| 2 | 90 | 150 | 7 | 67 | 56 |
| 3 | 80 | 63 | 8 | 22 | 19 |
| 4 | 5 | 10 | 9 | 56 | 63 |
| 5 | 63 | 63 | 10 | 33 | 19 |

Table 1: Survey approval/disapproval rate by county

(a) Analyze the data in each country separately using the Jeffreys' prior distribution and report the posterior 95% credible set for each county.

(b) Let $\hat{p}_i$ be the sample proportion of commuters in county $i$ that approve of the HOV lane (e.g., $\hat{p}_1 = 12/(12+50) = 0.194$). Select $a$ and $b$ so that the mean and variance of the $\text{Beta}(a, b)$ distribution match the mean and variance of the sample proportions $\hat{p}_1, \dots, \hat{p}_{10}$.

(c) Conduct an empirical Bayesian analysis by computing the 95% posterior credible sets that results from analysing each county separately using the $\text{Beta}(a, b)$ prior you computed in (b).

(d) How do the results from (a) and (c) differ? What are the advantages and disadvantages of these two analyses?

## Solution

### Part (a)

Let $Y_1, \dots, Y_{10}$ denote the number of approving responses out of $n_1, \dots, n_{10}$ survey respondents for each county. Naturally we assume that for each of the $Y_i$'s,

$$Y_i \mid \theta \sim \text{Binomial}(n_i, \theta),$$

for which the Jeffrey's prior turns out to be $\pi(\theta) \sim \text{Beta}(1/2, 1/2)$. The posterior given the prior is

$$\begin{aligned} p(\theta \mid Y) &\propto \mathcal{L}(Y \mid \theta)\pi(\theta) \\ &\propto \theta^{Y_i}(1-\theta)^{n_i - Y_i}\theta^{1/2-1}(1-\theta)^{1/2-1} \\ &= \theta^{Y_i+1/2-1}(1-\theta)^{n_i-Y_i+1/2-1}, \end{aligned}$$

so $p(\theta \mid Y) \sim \text{Beta}(Y_i + 1/2, n_i - Y_i + 1/2)$. We now turn to the task of computing the 95% posterior credible sets for each county. Start by gathering up all the information we have into a data frame:

```
> cs <- 1:10 # counties
> ys <- c(12, 90, 80, 5, 63, 15, 67, 22, 56, 33) # approve
> ns <- c(50, 150, 63, 10, 63, 8, 56, 19, 63, 19) # disapprove
```

```
> dat <- data.frame(cs, ys, ns)
> colnames(dat) <- c("county", "approve", "disapprove")
> dat$prop_approve <- dat$approve / (dat$approve + dat$disapprove)
> dat$n <- dat$approve + dat$disapprove
```

We now have our dataframe, `dat`, containing all the information we need for the analysis:

```
> dat
   county approve disapprove prop_approve   n
1       1      12         50    0.1935484  62
2       2      90        150    0.3750000 240
3       3      80         63    0.5594406 143
4       4       5         10    0.3333333  15
5       5      63         63    0.5000000 126
6       6      15          8    0.6521739  23
7       7      67         56    0.5447154 123
8       8      22         19    0.5365854  41
9       9      56         63    0.4705882 119
10     10      33         19    0.6346154  52
```

Next, generate the posterior credible intervals using `lapply()`, and bind them into a data frame using `do.call()`, before declaring the credible sets as new columns:

```
> cis <- lapply(seq_len(nrow(dat)), function(i) {
+   yi <- dat[i,][[2]]
+   ni <- dat[i,][[5]]

+   l <- qbeta(p = 0.025, shape1 = yi + 1/2, shape2 = ni - yi + 1/2)
+   u <- qbeta(p = 0.975, shape1 = yi + 1/2, shape2 = ni - yi + 1/2)
+   return(c(l, u))
+ })
> cis <- do.call(rbind, cis)
> dat$ci_l <- cis[,1]
> dat$ci_u <- cis[,2]
```

Now let us view how the data frame has changed:

```
> dat
   county approve disapprove prop_approve   n      ci_l      ci_u
1       1      12         50    0.1935484  62 0.1103935 0.3045307
2       2      90        150    0.3750000 240 0.3155448 0.4374498
3       3      80         63    0.5594406 143 0.4775597 0.6389423
4       4       5         10    0.3333333  15 0.1402526 0.5841638
5       5      63         63    0.5000000 126 0.4135270 0.5864730
6       6      15          8    0.6521739  23 0.4488766 0.8197773
7       7      67         56    0.5447154 123 0.4565230 0.6308309
8       8      22         19    0.5365854  41 0.3858303 0.6823739
```

```
9        9        56           63     0.4705882 119 0.3825161 0.5600718
10       10       33           19     0.6346154  52 0.4993247 0.7553816
```

For example, we see that the credible set for county 1 is $[0.11033935, 0.3045307]$. We can verify that these are correct by Monte Carlo simulation:

```
> sapply(seq_len(nrow(dat)), function(i) {
  yi <- dat[i,][[2]]
+   s <- 1e+6
+   yi <- dat[i,][[2]]
+   ni <- dat[i,][[5]]
+   ci_l <- dat[i,][[6]]
+   ci_u <- dat[i,][[7]]
+
+   r <- rbeta(n = s, shape1 = yi + 1/2, shape2 = ni - yi + 1/2)
+   return(mean(r > ci_l & r < ci_u))
+ })
 [1] 0.949823 0.949970 0.949882 0.949951 0.949995 0.950181 0.949624
 [8] 0.950194 0.949838 0.950423
```

Indeed, all of these intervals constitute valid 95% posterior credible sets.

**Part (b)**

Recall that if $Y \sim \text{Beta}(a, b)$, then we have

$$\text{E}(Y) = \frac{a}{a+b}, \quad \text{and} \quad \text{Var}(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

For a given set of sample proportion estimates $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_n)$, let $\overline{p} = \sum_{i=1}^n \hat{p}_i$ and $s^2_{\hat{\mathbf{p}}} = \sum_{i=1}^n (\hat{p}_i - \overline{p})^2/(n-1)$. Now, we want to determine $a$ and $b$ sufficing

$$\text{E}(Y) = \frac{a}{a+b} = \overline{p} \quad \text{and} \quad \text{Var}(Y) = \frac{ab}{(a+b)^2(a+b+1)} = s^2_{\hat{\mathbf{p}}}$$

which is equivalent to solving

$$b = \frac{(\overline{p}-1)}{\overline{p}}a \quad \text{and} \quad s^2_{\hat{\mathbf{p}}}(a+b)^2(a+b+1) - ab = 0$$

letting $\gamma := (\overline{p}-1)/\overline{p}$ and substituting $b = \gamma a$, we see that we are solving

$$b = \gamma a \quad \text{and} \quad s^2_{\hat{\mathbf{p}}}\big((\gamma+1)a\big)^2\big((\gamma+1)a+1\big) - \gamma a^2 = 0,$$

which is equivalent to finding the minima of the function $\tau(a) := s^2_{\hat{\mathbf{p}}}\big((\gamma+1)a\big)^2\big((\gamma+1)a+1\big) - \gamma a^2$ along the domain $b = \gamma a$. Such a solution is difficult to obtain through manual calculation, but is easy to solve numerically in R using the `optim()` function. A solution is implemented below.

```
> prop_mean <- mean(dat$prop_approve)
> prop_var <- var(dat$prop_approve)
```

```
> cost <- function(pars) {
+   a <- pars[1]
+   b <- pars[2]
+
+   eq1 <- a / (a + b) - prop_mean
+   eq2 <- (a * b) / ((a + b)^2 * (a + b + 1)) - prop_var
+   return(eq1^2 + eq2^2)
+ }
> res <- optim(par = c(1, 1), cost)
> a <- res$par[1]
> b <- res$par[2]
```

We can now see what the values of $a$ and $b$ are which minimise $\tau(a)$:

```
> print(c(a, b))
[1] 5.434924 5.887903
```

It is straightforward to verify that these estimate do indeed provide sufficient approximations of $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$. We can see this by calculating the moments manually:

```
> # Verify mean
> a / (a + b)
[1] 0.4799971
> prop_mean
[1] 0.4800001
> # Verify variance
> (a * b) / ((a + b)^2 * (a + b + 1))
[1] 0.02025508
> prop_var
[1] 0.02025887
```

**Part (c)**

The procedure is identical to the one in (a), and so it warrants no further explanation than the code below:

```
> cis_emp_bayes <- lapply(seq_len(nrow(dat)), function(i) {
+   yi <- dat[i,][[2]]
+   ni <- dat[i,][[5]]
+
+   l <- qbeta(p = 0.025, shape = yi + a, shape2 = ni - yi + b)
+   u <- qbeta(p = 0.975, shape = yi + a, shape2 = ni - yi + b)
+   return(c(l, u))
+ })
> cis_emp_bayes <- do.call(rbind, cis_emp_bayes)
> dat$ci_emp_l <- cis_emp_bayes[,1]
> dat$ci_emp_u <- cis_emp_bayes[,2]
```

Figure 3 shows a scatter plot of the lower and upper bounds of the posterior credible intervals generated by means of Jeffreys' prior (JP) and Empirical Bayes (EB), respectively. Additionally, Figure 4 shows the 95% posterior credible sets for Jeffreys' prior and Empirical Bayes respectively:

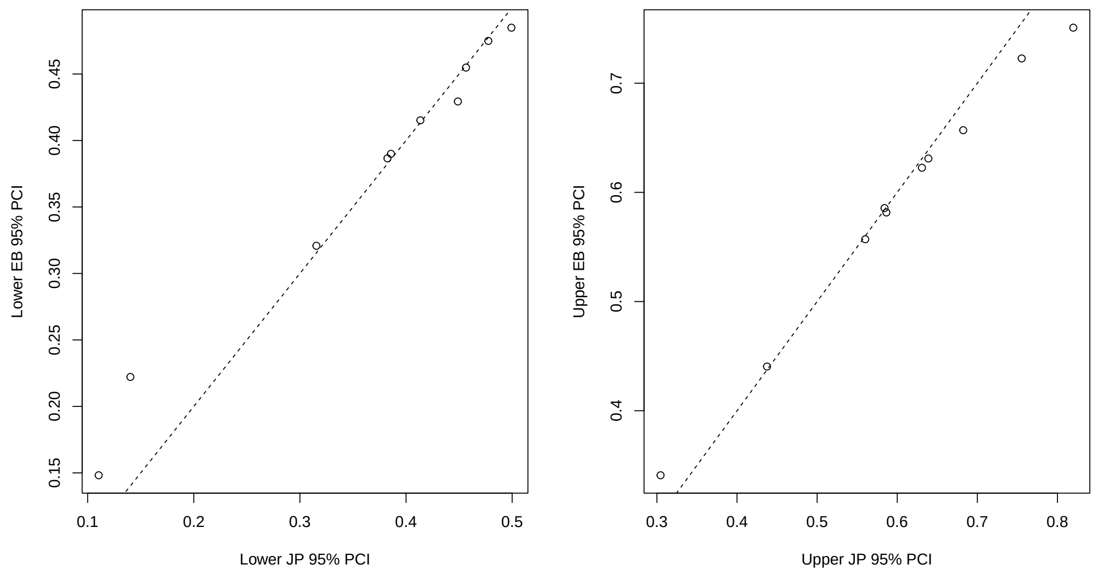


Figure 3: Lower and upper bounds of the posterior credible intervals generated by means of Jeffreys' prior (JP) and Empirical Bayes (EB)

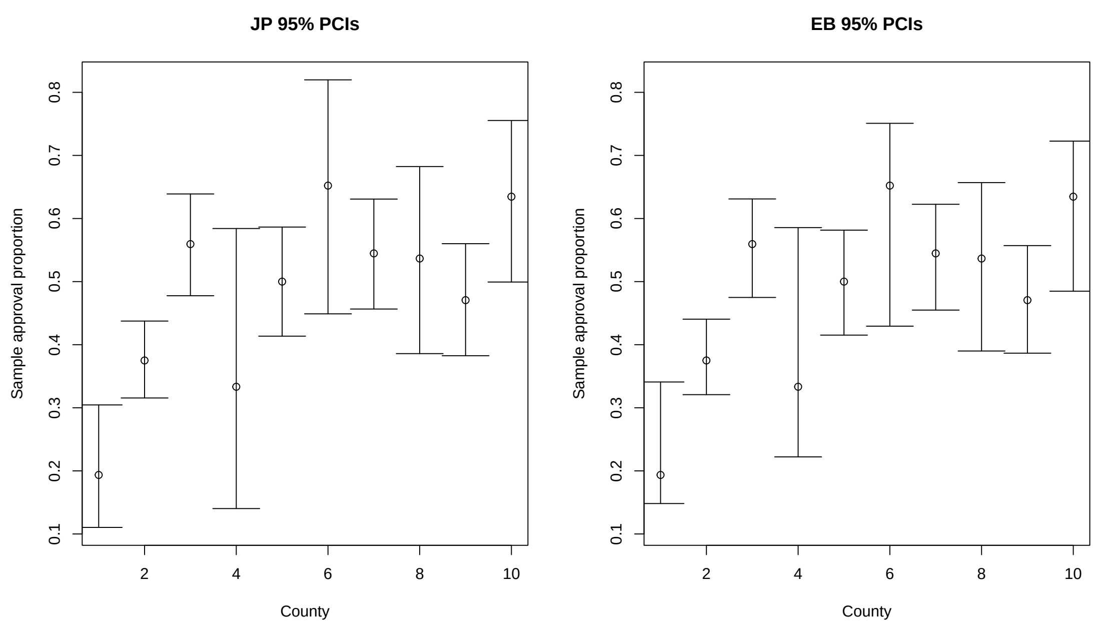


Figure 4: Posterior credible intervals generated by Jeffreys' prior (JP, left), and Empirical Bayes (EB, right).

Figures 3 and 4 show that the intervals are quite similar although some are less symmetric than their counterparts, such as the asymmetric credible EB PCI in county 1 compared to its relatively symmetric counterpart. To see why, recall that the posterior distributions arising from the methods discussed here for a given county $i$ are the following:

$$p_{\text{Jeffreys'}}(\theta \mid Y_i) \sim \text{Beta}(Y_i + 1/2, n_i - Y_i + 1/2) \quad \text{and} \quad p_{\text{Empirical}}(\theta \mid Y_i) \sim \text{Beta}(Y_i + a, n_i - Y_i + b)$$

where $a$ and $b$ are our estimated from part (b). If $Y_i$ and $N_i = n_i - Y_i$ are relatively similar, the resulting Beta distribution is relatively normal (or uniform in the case where $Y_i = N_i = 1$). However, the minute difference between $a$ and $b$ may be enough to cause the resulting Beta distribution to be skewed, and thus the confidence interval to be slightly asymmetric.

**Part (d)**

The credible intervals obtained by using Jeffreys' prior and Empirical Bayes are quite similar, as we previously discussed. In that regard, the results are not too different depending on method we used to derive the credible sets. This result is not surprising, since if these methods would yield wildly different results, it would suggest that one of them, if not both, are misrepresenting the data.

Jeffreys' prior is advantageous in that it provides an objective way with which one can derive a prior without the need for subjective elicitation of a prior distribution. However, when using the Jeffreys' prior, one needs to verify that the posterior distribution constitutes a valid probability density, which may not always be the case. Therefore, Jeffreys' prior is not guaranteed to yield valid results.

Empirical Bayes is a convenient method to obtain a prior as it uses the data itself to generate a prior. However, it is problematic since it uses the data to generate the prior, which is then used to create a posterior distribution meant to describe the data. In this regard, it has been criticised for "using the data twice", which can lead to underrepresentation of variance, and possibly debased results in cases where the sample does not accurately represent the population (although this would be a problem in statistic regardless of the setting).