Statistics 2 Unit 3 Team 8

Nikolaos Kornilakis   Rodrigo Viale   Jakub Trnan   Aleksandra Daneva
Luis Diego Pena Monge

2024-04-05

Exercise 37

```
bodytemp <- read.table("~/Documents/QFin/Q3/Statistics 2/Assignments/Data/bodytemp.txt",
                       sep = ",",
                       header = T)

means <- aggregate(bodytemp[,-2],
                   list(gender = bodytemp$gender),
                   mean)

means
```

```
##   gender temperature     rate
## 1      1    98.10462 73.36923
## 2      2    98.39385 74.15385
```

We compute confidence intervals:

```
alpha <- 0.05

n <- unlist(aggregate(bodytemp[,-2],
                      list(gender = bodytemp$gender),
                      length)[,-1])

S <- unlist(aggregate(bodytemp[,-2],
                      list(gender = bodytemp$gender),
                      sd)[,-1])

xBar <- unlist(means[,-1])

interval <- list(lower = xBar - (S/sqrt(n)) * qt(p = 1-alpha/2, df = n-1),
                 upper = xBar + (S/sqrt(n)) * qt(p = 1-alpha/2, df = n-1))
interval
```

```
## $lower
## temperature1 temperature2        rate1        rate2
##     97.93147     98.20962     71.91343     72.14547
##
## $upper
## temperature1 temperature2        rate1        rate2
##     98.27776     98.57807     74.82503     76.16222
```

We see that (at least for our sample) the folklore saying seems not to be the case (with a confidence of 95%), since the folklore sample does not fall within the confidence interval, either for men or for women.

Exercise 38

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} U(0, \theta)$ and define $X_{(n)} := max\{X_1, \ldots, X_n\}$.

a)

Note that:

$$\begin{aligned} P\left(X_{(n)} \leq y\right) &= P\left(max\{X_1, \ldots, X_n\} \leq y\right) \\ &= P\left(X_1 \leq y, \ldots, X_n \leq y\right) \\ &= \prod_{i=1}^{n} P(X_i \leq y) \\ &= \left(\frac{y}{\theta}\right)^n \end{aligned}$$

Therefore, when we look at the distribution of $X_{(n)}/\theta$, we have that:

$$F_{\frac{X_{(n)}}{\theta}}(y) = P\left(\frac{X_{(n)}}{\theta} \leq y\right) = P\left(X_{(n)} \leq \theta y\right) = \left(\frac{\theta y}{\theta}\right)^n = y^n$$

This does not depend on θ, so we indeed have a pivot.

b)

Notice that:

$$\begin{aligned} P\left(X_{(n)} \leq \theta \leq \alpha^{-1/n} X_{(n)}\right) &= P\left(1 \leq \frac{\theta}{X_{(n)}} \leq \alpha^{-1/n}\right) \\ &= P\left(\alpha^{1/n} \leq \frac{X_{(n)}}{\theta} \leq 1\right) \\ &= F_{\frac{X_{(n)}}{\theta}}(1) - F_{\frac{X_{(n)}}{\theta}}(\alpha^{1/n}) \\ &= 1^n - (\alpha^{1/n})^n \\ &= 1 - \alpha \end{aligned}$$

Which proves $\left(X_{(n)}, \alpha^{-1/n} X_{(n)}\right)$ is a $(1-\alpha) \cdot 100$ percent confidence interval for θ.

Exercise 39

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \alpha^2\mu^2)$ with α known. Note then that $\sum_{i=1}^n X_i \sim N(n\mu, n\alpha^2\mu^2)$. Then we have that:

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\alpha^2\mu^2}} = \frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha\mu} - \frac{\sqrt{n}}{\alpha} \sim N(0, 1)$$

We have then found a pivot for μ. Now, given a $100(1-\lambda)$ percent confidence, we can find a confidence interval for μ as follows:

$$\begin{aligned}
1-\lambda &= P\left(z_{\lambda/2} \leq \frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha\mu} - \frac{\sqrt{n}}{\alpha} \leq z_{1-\lambda/2}\right) \\
&= P\left(z_{\lambda/2} + \frac{\sqrt{n}}{\alpha} \leq \frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha\mu} \leq z_{1-\lambda/2} + \frac{\sqrt{n}}{\alpha}\right) \\
&= P\left(\left(z_{1-\lambda/2} + \frac{\sqrt{n}}{\alpha}\right)^{-1} \leq \frac{\sqrt{n}\alpha\mu}{\sum_{i=1}^n X_i} \leq \left(z_{\lambda/2} + \frac{\sqrt{n}}{\alpha}\right)^{-1}\right) \\
&= P\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha}\left(z_{1-\lambda/2} + \frac{\sqrt{n}}{\alpha}\right)^{-1} \leq \mu \leq \frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha}\left(z_{\lambda/2} + \frac{\sqrt{n}}{\alpha}\right)^{-1}\right)
\end{aligned}$$

We therefore have that a $100(1-\lambda)$ percent confidence interval for μ is given by:

$$\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha}\left(z_{1-\lambda/2} + \frac{\sqrt{n}}{\alpha}\right)^{-1}, \frac{\sum_{i=1}^n X_i}{\sqrt{n}\alpha}\left(z_{\lambda/2} + \frac{\sqrt{n}}{\alpha}\right)^{-1}\right)$$

Exercise 40

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ with σ known and let $0 < \beta < \alpha < 1$.

a)

$$\begin{aligned}
P\left(\bar{X} + z_{\alpha-\beta}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - z_\beta\frac{\sigma}{\sqrt{n}}\right) &= P\left(z_{\alpha-\beta} \leq \sqrt{n}\frac{\mu-\bar{X}}{\sigma} \leq -z_\beta\right) \\
&= P\left(z_\beta \leq \frac{\bar{X}-\mu}{\sigma} \leq -z_{\alpha-\beta}\right) \\
&= \Phi(-z_{\alpha-\beta}) - \Phi(z_\beta) \\
&= 1-(\alpha-\beta)-\beta \\
&= 1-\alpha
\end{aligned}$$

b)

Now, notice that the interval's length is not random and given by $\left(\bar{X} - z_\beta\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - z_{\alpha-\beta}\frac{\sigma}{\sqrt{n}}\right) = \frac{-\sigma}{\sqrt{n}}(z_\beta + z_{\alpha-\beta})$. We want to find the value of β that minimizes this length.

For this, first notice that, by definition:

$$\beta = \int_{-\infty}^{z_\beta}\phi(z)dz \qquad \alpha-\beta = \int_{-\infty}^{z_{\alpha-\beta}}\phi(z)dz$$

Differentiating with respect to β, we get that:

$$1 = \phi(z_\beta)\frac{d}{d\beta}z_\beta \qquad -1 = \phi(z_{\alpha-\beta})\frac{d}{d\beta}z_{\alpha-\beta}$$

Which means we can now solve for the differentials:

$$\frac{1}{\phi(z_\beta)} = \frac{d}{d\beta} z_\beta \qquad \frac{-1}{\phi(z_{\alpha-\beta})} = \frac{d}{d\beta} z_{\alpha-\beta}$$

Now we can minimize the length as follows (notice $\frac{-\sigma}{\sqrt{n}}$ is just a constant):

$$\begin{aligned} \frac{d}{d\beta}(z_\beta + z_{\alpha-\beta}) = 0 &\iff \frac{1}{\phi(z_\beta)} - \frac{1}{\phi(z_{\alpha-\beta})} = 0 \\ &\iff \phi(z_\beta) = \phi(z_{\alpha-\beta}) \end{aligned}$$

Now, recall ϕ is not a bijective function, but it is symmetric around 0. This means that we have two possible cases (note the case when either of them is 0, which is the only single-point level curve of the density, is actually contemplated in both cases):

$$\begin{aligned} &\text{Case 1: } z_\beta = z_{\alpha-\beta} \Rightarrow \beta = \alpha - \beta \Rightarrow \beta = \frac{\alpha}{2} \\ &\text{Case 2: } z_\beta = -z_{\alpha-\beta} \Rightarrow \beta = 1 - (\alpha - \beta) \Rightarrow \alpha = 1 \ \text{(!!)} \end{aligned}$$

We see then that only case 1 is possible, therefore the interval's length is minimized when the interval is symmetric.

Another way to prove that this minimizes, is to assume $\beta \neq \alpha/2$. Then we have two cases. If $\beta > \alpha/2$, since z_β is an increasing function in terms of β, then $z_\beta > z_{\alpha/2}$. For the same reason, $z_{\alpha-\beta} < z_{\alpha/2}$. Therefore, the interval's distance is given by:

$$\frac{-\sigma}{\sqrt{n}}(z_\beta + z_{\alpha-\beta}) > \frac{-\sigma}{\sqrt{n}}(z_{\alpha/2} + z_{\alpha/2})$$

Notice the right hand side of the inequality is exactly the length of the interval if we take $\beta = \alpha/2$. The second case is completely analogous, proving that the previous optimization minimizes the interval's length.

Exercise 41

Let $X_1, \ldots, X_n \sim logN(\mu, \sigma^2)$. Define $Y_i := \ln(X_i)$. By properties of the log-normal distribution, we know that $Y_i \sim N(\mu, \sigma^2)$. As seen in class, we know that $\frac{\sqrt{n}(\bar{Y}-\mu)}{S} \sim t_{n-1}$. This means that:

$$\begin{aligned} 1 - \alpha &= P\left(Q_{t_{n-1}}(\alpha/2) \leq \frac{\sqrt{n}(\bar{Y} - \mu)}{S} \leq Q_{t_{n-1}}(1 - \alpha/2)\right) \\ &= P\left(\frac{S}{\sqrt{n}} Q_{t_{n-1}}(\alpha/2) - \bar{Y} \leq -\mu \leq \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1 - \alpha/2) - \bar{Y}\right) \\ &= P\left(\bar{Y} - \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1 - \alpha/2) \leq \mu \leq \bar{Y} - \frac{S}{\sqrt{n}} Q_{t_{n-1}}(\alpha/2)\right) \\ &= P\left(\bar{Y} - \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1 - \alpha/2) \leq \mu \leq \bar{Y} + \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1 - \alpha/2)\right) \end{aligned}$$

Which yields the following $100(1 - \alpha)$ percent confidence interval for μ:

$$\left(\bar{Y} - \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1-\alpha/2), \bar{Y} + \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1-\alpha/2)\right)$$

To get a 95% confidence interval it suffices to plug in $\alpha = 0.05$, to get:

$$\left(\bar{Y} - \frac{S}{\sqrt{n}} Q_{t_{n-1}}(0.975), \bar{Y} + \frac{S}{\sqrt{n}} Q_{t_{n-1}}(0.975)\right)$$

We test this confidence interval through Monte Carlo simulation:

```r
set.seed(1234)

alpha <- 0.05
n <- 1000

coverage <- vapply(1:1e5, function(i){
  X <- rlnorm(n)
  Y <- log(X)
  S <- sd(Y)
  Y_bar <- mean(Y)

  interval <- c(Y_bar - S/sqrt(n)*qt(p = 1-alpha/2, df = n-1),
                Y_bar + S/sqrt(n)*qt(p = 1-alpha/2, df = n-1))

  (interval[1] <= 0) && (0 <= interval[2])
}, T)

sum(coverage)/length(coverage)
```

```
## [1] 0.94954
```

Exercise 42

First of all, recall the true mean of a Chi-squared distribution with n degrees of freedom is n. We use this to test the confidence interval through Monte Carlo simulation:

```r
set.seed(1234)

alpha <- 0.05
n <- 20
df <- 2

coverage <- vapply(1:1e5, function(i){
  X <- rchisq(n, df = df)
  S <- sd(X)
  X_bar <- mean(X)

  interval <- c(X_bar - S/sqrt(n)*qt(p = 1-alpha/2, df = n-1),
                X_bar + S/sqrt(n)*qt(p = 1-alpha/2, df = n-1))

  (interval[1] <= df) && (df <= interval[2])
}, T)

sum(coverage)/length(coverage)
```

```
## [1] 0.91807
```

We observe that the true coverage probability of this confidence interval is closer to 92% than 95%.

Exercise 43

Let $\theta = (\mu, \sigma^2)$. Recall that the normal distribution fulfills the smoothness criteria to calculate the Fisher information as:

$$I(\theta) = -E\left[\mathcal{H}(ln(f(X|\theta)))\right]$$

Let's start by noting that:

$$\begin{aligned} ln(f(X|\theta)) &= ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\right) \\ &= \frac{-1}{2}ln(\sigma^2) - \frac{1}{2}ln(2\pi) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} \end{aligned}$$

Let's calculate the first and second order partial derivatives to obtain the hessian $\mathcal{H}$:

$$\begin{aligned} \frac{\partial}{\partial\mu}ln(f(x|\theta)) = \frac{(x-\mu)^2}{\sigma^2} &\Rightarrow \frac{\partial^2}{\partial\mu^2}ln(f(x|\theta)) = \frac{-1}{\sigma^2} \\ &\Rightarrow \frac{\partial^2}{\partial\mu\partial\sigma}ln(f(x|\theta)) = \frac{-2(x-\mu)}{\sigma^3} \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial\sigma}ln(f(x|\theta)) = \frac{-1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3} &\Rightarrow \frac{\partial^2}{\partial\sigma^2}ln(f(x|\theta)) = \frac{1}{\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4} \\ &\Rightarrow \frac{\partial^2}{\partial\sigma^2\partial\mu}ln(f(x|\theta)) = \frac{-2(x-\mu)}{\sigma^3} \end{aligned}$$

Then we have that the hessian matrix is given by:

$$\mathcal{H} = \begin{pmatrix} \frac{-1}{\sigma^2} & \frac{-2(x-\mu)}{\sigma^3} \\ \frac{-2(x-\mu)}{\sigma^3} & \frac{1}{\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4} \end{pmatrix}$$

We now take negative expectations entry-wise:

$$-E\left[\frac{-1}{\sigma^2}\right] = \frac{1}{\sigma^2}$$

$$\begin{aligned} -E\left[\frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}\right] &= -\left(\frac{1}{\sigma^2} - \frac{3}{\sigma^4}\left[E[X^2] - 2\mu E[X] + \mu^2\right]\right) \\ &= -\left(\frac{1}{\sigma^2} - \frac{3}{\sigma^4}\left[E[X^2] - \mu^2\right]\right) \\ &= -\left(\frac{1}{\sigma^2} - \frac{3}{\sigma^2}\right) \\ &= \frac{2}{\sigma^2} \end{aligned}$$

$$-E\left[\frac{-2(X-\mu)}{\sigma^3}\right] = \frac{2(E[X]-\mu)}{\sigma^3} = \frac{2(\mu-\mu)}{\sigma^3} = 0$$

Therefore, we have that the Fisher information is given by:

$$I(\theta) = -E\left[\mathcal{H}(ln(f(X|\theta)))\right] = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

Note the Fisher information is a diagonal matrix. This means that its inverse is also a diagonal matrix, given by

$$I^{-1}(\theta) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix}$$

As we know, asymptotically the distribution of the MSE tends to a multivariate normal distribution, in which the covariance matrix is determined by the inverse of the Fisher information (times a constant related to the number of observations). Since we have a diagonal covariance matrix for a multivariate normal, this implies that asymptotically the distribution of the MLE for μ and σ are independent.

Exercise 44

Let X be multinomial, then its probability mass function is given by:

$$\frac{n!}{\prod_{i=1}^{k} x_i!}\prod_{i=1}^{k} p_i^{x_i} = \frac{n!}{\prod_{i=1}^{k} x_i!}\left(\prod_{i=1}^{k-1} p_i^{x_i}\right)\left(1-\sum_{i=1}^{k-1} p_i\right)^{x_k}$$

Then the log-likelihood is given by:

$$LL = ln(n!) - \sum_{i=1}^{k} ln(x_i!) + \sum_{i=1}^{k-1} x_i ln(p_i) + x_k ln\left(1-\sum_{i=1}^{k-1} p_i\right)$$

Then, we can calculate the score function as the gradient of the log-likelihood as follows:

$$\begin{aligned}
\nabla LL &= \left(\frac{x_1}{p_1} - \frac{x_k}{p_k}, \ldots, \frac{x_{k-1}}{p_{k-1}} - \frac{x_k}{p_k}\right)^T \\
&= \begin{pmatrix} \frac{1}{p_1} & 0 & \cdots & \frac{-1}{p_k} \\ 0 & \frac{1}{p_2} & \cdots & \frac{-1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{p_k} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \\
&:= AX
\end{aligned}$$

Define $P = (p_1, \ldots, p_k)^T$. Recall the Fisher information is the covariance of the score, in this case given by:

$$
\begin{aligned}
I(\theta) &= Var[AX] \\
&= An(\text{diag}(P) - PP^T)A^T \\
&= n\left[A\text{diag}(P)A^T - APP^TA^T\right] \\
&= n\left[A\text{diag}(P)A^T - AP(AP)^T\right]
\end{aligned}
$$

We'll now look at each matrix product individually:

$$
\begin{aligned}
A\text{diag}(P)A^T &= \begin{pmatrix} \frac{1}{p_1} & 0 & \cdots & \frac{-1}{p_k} \\ 0 & \frac{1}{p_2} & \cdots & \frac{-1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{p_k} \end{pmatrix} \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_k \end{pmatrix} \begin{pmatrix} \frac{1}{p_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{p_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{p_k} & \frac{-1}{p_k} & \cdots & \frac{-1}{p_k} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \cdots & -1 \\ 0 & 1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{p_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{p_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{p_k} & \frac{-1}{p_k} & \cdots & \frac{-1}{p_k} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \frac{1}{p_k} & \frac{1}{p_2} + \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}} + \frac{1}{p_k} \end{pmatrix} \\
&= \text{diag}\left(\frac{1}{p_1}, \dots, \frac{1}{p_{k-1}}\right) + 11^T \frac{1}{p_k}
\end{aligned}
$$

Now:

$$
AP = \begin{pmatrix} \frac{1}{p_1} & 0 & \cdots & \frac{-1}{p_k} \\ 0 & \frac{1}{p_2} & \cdots & \frac{-1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{p_k} \end{pmatrix} \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}
$$

Then $AP(AP)^T = 0$. Therefore:

$$
I(\theta) = \text{diag}\left(\frac{1}{p_1}, \dots, \frac{1}{p_{k-1}}\right) + 11^T \frac{1}{p_k}
$$

Now, let's check that $I(\theta)^{-1} = \frac{1}{n}(\text{diag}(\theta) - \theta\theta^T)$:

$$\mathrm{diag}(\theta) - \theta\theta^T = \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_{k-1} \end{pmatrix} \begin{pmatrix} p_1^2 & p_1 p_2 & \cdots & p_1 p_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k-1}p_1 & p_{k-1}p_2 & \cdots & p_{k-1}^2 \end{pmatrix}$$

$$= \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{k-1}p_1 & -p_{k-1}p_2 & \cdots & p_{k-1}(1-p_{k-1}) \end{pmatrix}$$

Multiplying our candidate inverse (call it B) on the left, we get that:

$$BI(\theta) = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_{k-1} \\ -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{k-1}p_1 & -p_{k-1}p_2 & \cdots & p_{k-1}(1-p_{k-1}) \end{pmatrix} \begin{pmatrix} \frac{1}{p_1}+\frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \frac{1}{p_k} & \frac{1}{p_2}+\frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}}+\frac{1}{p_k} \end{pmatrix}$$

Notice that for the ij-th entry of $BI(\theta)$ we have two cases. When $i = j$ (i.e. in the diagonal) and when $i \neq j$ (i.e. off-diagonal). For the first case, we have that:

$$\begin{aligned} (BI(\theta))_{ii} &= p_i(1-p_i)\left(\frac{1}{p_i}+\frac{1}{p_k}\right) - \frac{p_i}{p_k}\sum_{\substack{n\neq i \\ 1\leq n\leq k-1}} p_n \\ &= p_i(1-p_i)\left(\frac{1}{p_i}+\frac{1}{p_k}\right) - \frac{p_i}{p_k}(1-p_k-p_i) \\ &= \left(1+\frac{p_i}{p_k}-p_i-\frac{p_i^2}{p_k}\right) - \left(\frac{p_i}{p_k}-p_i-\frac{p_i^2}{p_k}\right) \\ &= 1 \end{aligned}$$

For the second case, when $i \neq j$, we have the following:

$$\begin{aligned} (BI(\theta))_{ij} &= -p_ip_j\left(\frac{1}{p_j}+\frac{1}{p_k}\right) + \frac{p_i(1-p_i)}{p_k} - \frac{p_i}{p_k}\sum_{\substack{n\neq i,j \\ 1\leq n\leq k-1}} p_n \\ &= \left(-p_i - \frac{p_ip_j}{p_k}\right) + \frac{p_i}{p_k} - \frac{p_i^2}{p_k} - \frac{p_i}{p_k}(1-p_k-p_i-p_j) \\ &= \left(-p_i - \frac{p_ip_j}{p_k}\right) + \frac{p_i}{p_k} - \frac{p_i^2}{p_k} - \frac{p_i}{p_k} + p_i + \frac{p_i^2}{p_k} + \frac{p_ip_j}{p_k} \\ &= 0 \end{aligned}$$

Therefore, $BI(\theta) = Id$. Now, notice that both B and $I(\theta)$ are symmetric matrices. This means that:

$$BI(\theta) = Id = (Id)^T = (BI(\theta))^T = I(\theta)^T B^T = I(\theta)B$$

Therefore,

$$B = I(\theta)^{-1} = \text{diag}(\theta) - \theta\theta^T$$

Exercise 53

Recall the MSE of an estimator p_i of p can be calculated as $MSE(p_i) = Var(p_i) + (E(p_i) - p)^2$. Then we have the following MSEs for the given estimators:

$$\begin{aligned}
MSE(p_1) &= Var(p_1) + (E(p_1) - p)^2 \\
&= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] + \left(E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - p\right)^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] + \left(\frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] - p\right)^2 \\
&= \frac{np(1-p)}{n^2} + \left(\frac{np}{n} - p\right)^2 \\
&= \frac{p(1-p)}{n}
\end{aligned}$$

For the second estimator:

$$\begin{aligned}
MSE(p_2) &= Var(p_2) + (E(p_2) - p)^2 \\
&= Var\left[\frac{1}{n+2}\left(\sum_{i=1}^{n} X_i + 1\right)\right] + \left(E\left[\frac{1}{n+2}\left(\sum_{i=1}^{n} X_i + 1\right)\right] - p\right)^2 \\
&= \frac{1}{(n+2)^2} Var\left[\sum_{i=1}^{n} X_i + 1\right] + \left(\frac{1}{n+2}\left(\sum_{i=1}^{n} E\left[X_i\right] + 1\right) - p\right)^2 \\
&= \frac{np(1-p)}{(n+2)^2} + \left(\frac{np+1}{n+2} - p\right)^2 \\
&= p(1-p)\frac{n-4}{(n+2)^2} + \frac{1}{(n+2)^2}
\end{aligned}$$

And finally:

$$\begin{aligned}
MSE(p_3) &= Var(p_3) + (E(p_3) - p)^2 \\
&= Var(X_1) + (E[X_1] - p)^2 \\
&= p(1-p)
\end{aligned}$$

Evidently, $MSE(p_1) \leq MSE(p_3) \forall n \in \mathbb{N}, p \in [0,1]$. However, notice that when we compare $MSE(p_1)$ with $MSE(p_2)$, we get the following:

$$
\begin{aligned}
MSE(p_2) - MSE(p_1) &= p(1-p)\frac{n-4}{(n+2)^2} + \frac{1}{(n+2)^2} - \frac{p(1-p)}{n} \\
&= p(1-p)\left(\frac{n-4}{(n+2)^2} - \frac{1}{n}\right) + \frac{1}{(n+2)^2} \\
&= \frac{1}{(n+2)^2}\left(1 - \frac{4p(1-p)}{n}\right)
\end{aligned}
$$

Note that the first factor is always positive as it's a square. So whether or not the difference is positive will be completely determined by the second factor. Then we have that the difference is positive if and only if $n \geq 4p(1-p)$. This means that, with a large enough sample (particularly as $n \to \infty$), p_1 is better than the other two estimators in terms of MSE. However, depending on the true value of p (which remember is unknown), if the sample size is too small (small enough that the previous inequality does not hold), then p_2 performs better than p_1 in terms of MSE. We can therefore not guarantee that one estimate is uniformly better than the others in terms of MSE.

Exercise 54

Suppose there exists an unbiased estimator $\hat{\theta}(X_1, \ldots, X_n)$ for $\frac{p}{1-p}$. Then it must hold that $E[\hat{\theta}(X_1, \ldots, X_n)] = \frac{p}{1-p}$. Let $\mathcal{B}$ represent all possible values of $(X_1, \ldots, X_n)$. Note that this set contains 2^n elements since $X_1, \ldots, X_n$ are Bernoulli. Then:

$$
\begin{aligned}
E[\hat{\theta}(X_1, \ldots, X_n)] &= \sum_{(x_1,\ldots,x_n)\in\mathcal{B}} \hat{\theta}(x_1, ..., x_n) P(X_1 = x_1, ..., X_n = x_n) \\
&= \sum_{(x_1,\ldots,x_n)\in\mathcal{B}} \hat{\theta}(x_1, ..., x_n) P(X_1 = x_1) \cdots P(X_n = x_n)
\end{aligned}
$$

Recall

$$
P(X_i = x_i) = \begin{cases} p & x_i = 1 \\ 1-p & x_i = 0 \end{cases}
$$

Without loss of generality assume there are exactly $0 \leq n_1 \leq n$ values in $(x_1, ..., x_n)$ such that $x_i = 1$. Then:

$$
E[\hat{\theta}(X_1, \ldots, X_n)] = \sum_{(x_1,\ldots,x_n)\in\mathcal{B}} \hat{\theta}(x_1, ..., x_n) p^{n_1}(1-p)^{n-n_1}
$$

Since $\hat{\theta} \in \mathbb{R}$ and it takes at most 2^n values, then the previous expression is a polynomial in p of at most degree n (degree n is attained when $n_1 = 0$ or $n_1 = n$). Now, under the assumption of unbiasedness of $\hat{\theta}$, we must have that:

$$
E[\hat{\theta}(X_1, \ldots, X_n)] = \frac{p}{1-p} \iff (1-p)E[\hat{\theta}(X_1, \ldots, X_n)] - p = 0
$$

But notice now that we have a polynomial of degree at most $n+1$, which means it has at most $n+1$ real roots. This does not cover the entire possible values of p, which is $[0, 1]$ (and therefore has infinite cardinality). This is a contradiction, and therefore there exists no unbiased estimator $\hat{\theta}$ for the odds ratio.

Exercise 56

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} f(x|\theta) = e^{-(x-\theta)}$ for $x \geq \theta$.

a)

Notice the likelihood and log-likelihood functions are given by:

$$L(\theta|X_1,\ldots,X_n)=\prod_{i=1}^{n}e^{-(X_i-\theta)}\Rightarrow LL(\theta|X_1,\ldots,X_n)=-\sum_{i=1}^{n}(X_i-\theta)$$
$$=n\theta-\sum_{i=1}^{n}X_i$$

Then, notice that:

$$\frac{\partial}{\partial\theta}LL(\theta|X_1,\ldots,X_n)=n>0$$

This means that the likelihood function is increasing with respect to θ. We therefore maximize the likelihood by finding the biggest possible value of θ. Recall that the density function is defined for $x\geq\theta$. This gives us an upper bound for our estimator, and so we maximize θ (and therefore also the likelihood) when $\theta=\min\{X_1,\ldots,X_n\}$.

$$\therefore\hat{\theta}_n=\min\{X_1,\ldots,X_n\}$$

c)

Note we will solve part c) first, since that will be useful for part b). Let's calculate the CDF of $n(\hat{\theta}_n-\theta)$:

$$\begin{aligned}
P(n(\hat{\theta}_n-\theta)\leq x)&=P\left(\hat{\theta}_n\leq\frac{x}{n}+\theta\right)\\
&=P\left(\min\{X_1,\ldots,X_n\}\leq\frac{x}{n}+\theta\right)\\
&=1-P\left(\min\{X_1,\ldots,X_n\}\geq\frac{x}{n}+\theta\right)\\
&=1-P\left(X_1\geq\frac{x}{n}+\theta,\ldots,X_n\geq\frac{x}{n}+\theta\right)\\
&=1-\prod_{i=1}^{n}P\left(X_i\geq\frac{x}{n}+\theta\right)\\
&=1-\prod_{i=1}^{n}\int_{x/n+\theta}^{\infty}e^{-(y-\theta)}dy\\
&\text{Substituting } u=y-\theta\\
&=1-\left(\int_{x/n}^{\infty}e^{-u}du\right)^n\\
&=1-\left(-e^{-u}|_{x/n}^{\infty}\right)^n\\
&=1-(e^{-x/n})^n\\
&=1-e^{-x}
\end{aligned}$$

Which is exactly the CDF for a standard exponential distribution. Having this in mind we can now solve part b):

b)

Let $\varepsilon > 0$. Notice that since $min\{X_1, \ldots, X_n\}$ is a realization of X, and we know that by definition $X \geq \theta$, then we must have that $\hat{\theta}_n - \theta \geq 0$. Then we have the following:

$$\begin{aligned}
P(|\hat{\theta}_n - \theta| > \varepsilon) &= P(\hat{\theta}_n - \theta > \varepsilon) \\
&= P(n(\hat{\theta}_n - \theta) > n\varepsilon) \\
&= 1 - P(n(\hat{\theta}_n - \theta) < n\varepsilon) \\
&= 1 - (1 - e^{-n\varepsilon}) \\
&= e^{-n\varepsilon} \\
&\to 0 \text{ when } n \to \infty
\end{aligned}$$

Therefore, $\hat{\theta}_n$ is consistent.

d)

First of all, note that as shown above $\hat{\theta}_n - \theta \geq 0 \Rightarrow \sqrt{n}(\hat{\theta}_n - \theta) \geq 0 \Rightarrow P(\sqrt{n}(\hat{\theta}_n - \theta) \leq x) = 0 \forall x < 0$. Then, let $x \geq 0$. We then have that:

$$\begin{aligned}
P(\sqrt{n}(\hat{\theta}_n - \theta) \leq x) &= P(n(\hat{\theta}_n - \theta) \leq x\sqrt{n}) \\
&= 1 - e^{-x\sqrt{n}} \\
&\to 1 \text{ when } n \to \infty
\end{aligned}$$

We see then that, in the limit, $P(\sqrt{n}(\hat{\theta}_n - \theta) \leq x) = 1$ for all $x \geq 0$. This means that, asymptotically, $\sqrt{n}(\hat{\theta}_n - \theta) \sim 0$.

e)

Note that in this case, our MLE is not obtained at a clear "smooth'' maximum of the likelihood function. This maximization is not bounded in terms of θ. We do not obtain a maxima of the parameter space itself, but rather we obtain a bound set by the observed realizations of the random variable. In other words, the necessary smoothness conditions necessary for asymptotic normality seem not to hold in this case.

Exercise 57

We will use a Bayesian approach to prove an therefore make sense of the rationale of the claim. Suppose $X_1, \ldots, X_n \sim Bernoulli(p)$, where we don't know p. We also assume the observations to be independent. For our Bayesian approach, let's assign a fairly uninformative prior to p, taking a standard uniform distribution. That is: $f_p(p) = 1_{[0,1]}$. We want to calculate the following:

$$
\begin{aligned}
P\left(X_{n+1}=1 \middle| \sum_{i=1}^{n} X_i = n\right) &= \int_0^1 P\left(X_{n+1}=1 \middle| \sum_{i=1}^{n} X_i = n, p=p\right) f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) dp \\
&= \int_0^1 P\left(X_{n+1}=1 \middle| p=p\right) f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) dp \\
&= \int_0^1 p f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) dp
\end{aligned}
$$

Note:

$$
f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) = \frac{P\left(\sum_{i=1}^{n} X_i = n | p=p\right) f_p(p)}{P\left(\sum_{i=1}^{n} X_i = n\right)} = \frac{p^n}{P\left(\sum_{i=1}^{n} X_i = n\right)}
$$

Also, since $f_p(\cdot| \sum_{i=1}^{n} X_i = n)$ is a distribution, it must hold that:

$$
\begin{aligned}
\int_0^1 f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) dp = 1 &\iff \int_0^1 \frac{p^n}{P\left(\sum_{i=1}^{n} X_i = n\right)} dp = 1 \\
&\iff \int_0^1 p^n dp = P\left(\sum_{i=1}^{n} X_i = n\right) \\
&\iff \left.\frac{p^{n+1}}{n+1}\right|_0^1 = P\left(\sum_{i=1}^{n} X_i = n\right) \\
&\iff \frac{1}{n+1} = P\left(\sum_{i=1}^{n} X_i = n\right)
\end{aligned}
$$

Then:

$$
f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) = (n+1)p^n
$$

Which means we can finally calculate what we wanted:

$$
\begin{aligned}
P\left(X_{n+1}=1 \middle| \sum_{i=1}^{n} X_i = n\right) &= \int_0^1 p f_p\left(p \middle| \sum_{i=1}^{n} X_i = n\right) dp \\
&= \int_0^1 p(n+1)p^n dp \\
&= (n+1)\int_0^1 p^{n+1} dp \\
&= (n+1)\left.\frac{p^{n+2}}{n+2}\right|_0^1 \\
&= \frac{n+1}{n+2}
\end{aligned}
$$

Exercise 58

Let $X \sim exp(\lambda)$ and assume for λ a gamma prior, that is $\lambda \sim \Gamma(\alpha, \beta)$. That means that:

$$f_\lambda(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Then the joint distribution is the product of the likelihood and the parameter's density. Let $X_1, \ldots, X_n$ be i.i.d. observations of X. Then:

$$f_{X,\lambda}(X, \lambda) = f_{X|\lambda}(X_1, ..., X_n|\lambda) f_\lambda(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n X_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

The posterior then is given by:

$$\begin{aligned} f_{\lambda|X}(\lambda|X_1, \ldots, X_n) &= \frac{f_{X,\lambda}(X,\lambda)}{\int f_{X,\lambda}(X,\lambda) dA} \\ &\propto \lambda^n e^{-\lambda \sum_{i=1}^n X_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &\propto \lambda^{n+\alpha-1} e^{-(\sum_{i=1}^n X_i + \beta)\lambda} \end{aligned}$$

Which is exactly the kernel of a gamma distribution with parameters $n + \alpha$ and $\sum_{i=1}^n X_i + \beta$. This proves the gamma distribution is a conjugate prior for the exponential distribution.

For the next part, let the waiting time $W \sim exp(\lambda)$, and we have that $\frac{1}{20} \sum_{i=1}^{20} W_i = 5.1$. This means that $\sum_{i=1}^{20} W_i = 102$. We now use what we found previously to find the posterior distributions for cases 1 and 2. Before that, recall that if $X \sim Gamma(\alpha, \beta)$ then $E[X] = \alpha/\beta$ and $Var[X] = \alpha/\beta^2$, then $sd(X) = \sqrt{\alpha}/\beta$.

Case 1

If we take as prior a gamma with mean 0.5 and s.d. 1, then we must solve:

$$\begin{cases} 0.5 = \frac{\alpha}{\beta} \\ 1 = \frac{\sqrt{\alpha}}{\beta} \end{cases} \iff \beta = \sqrt{\alpha} \Rightarrow 0.5 = \frac{\alpha}{\sqrt{\alpha}} \Rightarrow \alpha = 0.25 \Rightarrow \beta = 0.5$$

Then the prior is a gamma distribution with parameters $\alpha = 0.25$ and $\beta = 0.5$. As seen above, the posterior distribution is also a gamma, but with parameters $\alpha_{posterior} = \alpha + n = 20.25$ and $\beta_{posterior} = \sum_{i=1}^n X_i + \beta = 102.5$.

Case 2

Similarly to the previous case, now we need to solve the system:

$$\begin{cases} 10 = \frac{\alpha}{\beta} \\ 20 = \frac{\sqrt{\alpha}}{\beta} \end{cases} \iff 10\beta = \alpha \Rightarrow 20 = \frac{\sqrt{10\beta}}{\beta} \Rightarrow \beta = 0.025 \Rightarrow \alpha = 0.25$$

Then the prior is a gamma distribution with parameters $\alpha = 0.25$ and $\beta = 0.025$. As seen above, the posterior distribution is also a gamma, but with parameters $\alpha_{posterior} = \alpha + n = 20.25$ and $\beta_{posterior} = \sum_{i=1}^n X_i + \beta = 102.025$.

Plots and means

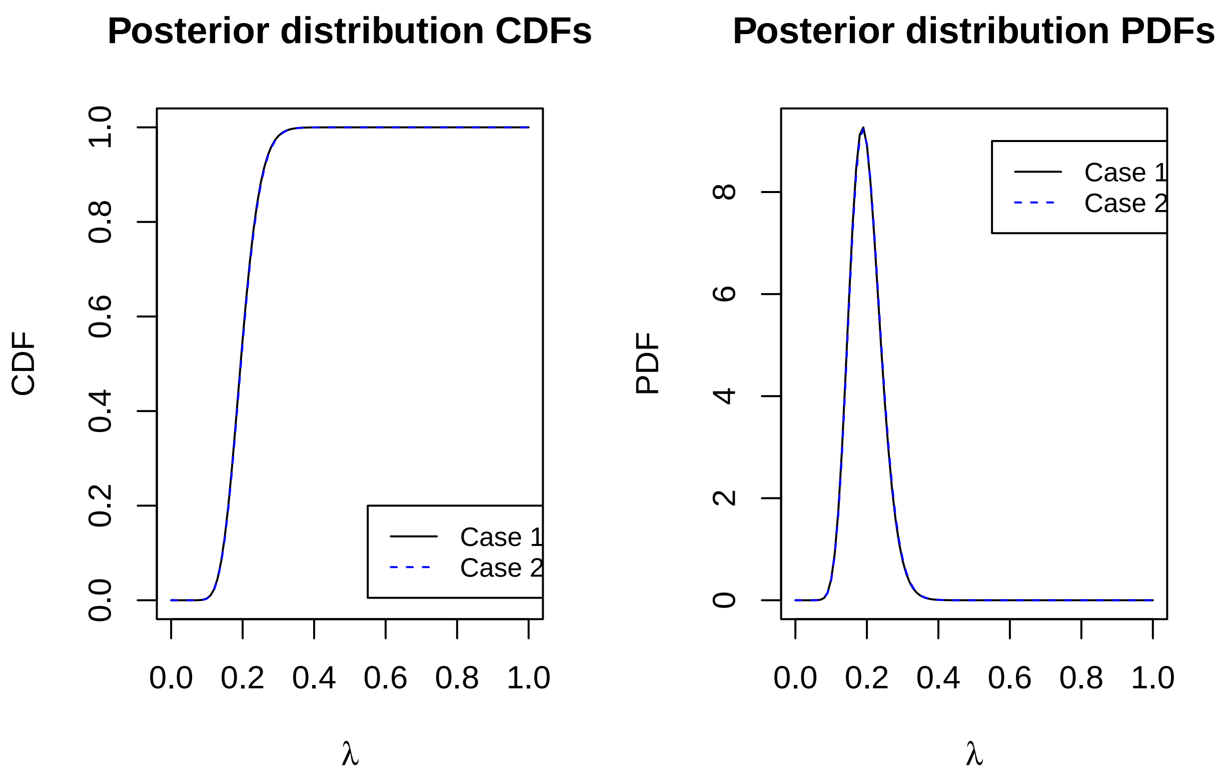
We proceed to plot the posteriors:

```r
alpha_1 <- 20.25
beta_1 <- 102.5

alpha_2 <- 20.25
beta_2 <- 102.025

par(mfrow = c(1,2))
curve(pgamma(x, shape = alpha_1, rate = beta_1),
      from = 0, to = 1,
      lty = 1, col = "black",
      main = "Posterior distribution CDFs",
      xlab = expression(lambda),
      ylab = "CDF")
curve(pgamma(x, shape = alpha_2, rate = beta_2),
      from = 0, to = 1,
      lty = 2, col = "blue", add = T)
legend(0.55, 0.2, legend=c("Case 1", "Case 2"),
       col=c("black", "blue"), lty=1:2, cex=0.8)

curve(dgamma(x, shape = alpha_1, rate = beta_1),
      from = 0, to = 1,
      lty = 1, col = "black",
      main = "Posterior distribution PDFs",
      xlab = expression(lambda),
      ylab = "PDF")
curve(dgamma(x, shape = alpha_2, rate = beta_2),
      from = 0, to = 1,
      lty = 2, col = "blue", add = T)
legend(0.55, 9, legend=c("Case 1", "Case 2"),
       col=c("black", "blue"), lty=1:2, cex=0.8)
```



We now compute their means. Once again, recall that if $X \sim Gamma(\alpha, \beta)$ then $E[X] = \alpha/\beta$. Then:

```
mean_1 <- alpha_1/beta_1
mean_2 <- alpha_2/beta_2

data.frame(Case = 1:2,
           Mean = c(mean_1, mean_2))
```

```
##   Case      Mean
## 1    1 0.1975610
## 2    2 0.1984808
```

As can be seen, both distributions are visually extremely similar. Not only that, but also the mean from both posteriors is fairly close. This is an interesting observation, since it illustrates that (at least for this example) having priors with different parameters results in posterior distributions that are relatively close to one another. It is important to note however that only the β parameter changed between both cases. Also, Case 2 results in a higher expected value than Case 1. This difference will become particularly apparent when calculating credibility intervals. Another important observation to note is the observation of only 20 realizations of the random variable. In other words, the difference comes exclusively from the discrepancy between the β parameter of both cases.

Exercise 59

a)

Let's assume we model the success of the shot via a $Bernoulli(\theta)$ random variable, and we have two independent observations of success. Assigning a uniform distribution prior to θ we have that:

$$f_\theta(\theta) = 1_{[0,1]} \qquad f(X_1 = 1, X_2 = 1|\theta) = \theta^2$$

Then:

$$
\begin{aligned}
f_{\theta|X} &= \frac{\theta^2}{\int_0^1 \theta^2 d\theta} \\
&= \frac{\theta^2}{\left(\frac{\theta^3}{3}\right)_0^1} \\
&= 3\theta^2
\end{aligned}
$$

b)

To estimate the probability of the next shot, we can use two approaches. In reality both approaches are the exact same thing. We can either calculate the expected value of the posterior (since the expected value of a Bernoulli is its probability):

$$E[\theta|X] = \int_0^1 3\theta^3 d\theta = 3\frac{\theta^4}{4}\bigg|_0^1 = \frac{3}{4}$$

Or we can use Laplace's rule of succession proven in exercise 57. Notice the previous approach is doing just that:

$$P(X_3 = 1|X_1 + X_2 = 2) = \frac{2+1}{2+2} = \frac{3}{4}$$

Exercise 60

a)

Recall that if $X \sim Beta(\alpha, \beta)$, then $f(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}1_{[0,1]}$, where B represents the Beta function.

Note that if we take $\alpha = \beta = 1$, we get that $f(x|\alpha,\beta) = 1_{[0,1]}$, since $B(1,1) = \frac{\Gamma(1)\Gamma(1)}{\Gamma(2)} = 1$.

Therefore, Realist's prior, the standard uniform distribution, belongs to the Beta family. Now, for Optimist, suppose n independent games have been played, resulting in $\sum_{i=1}^n X_i$ wins. Then:

$$
\begin{aligned}
f_{p|X}(p|X_1,\ldots,X_n) &= f_{X|p}(X_1,\ldots,X_n|p)f_p(p) \\
&\propto p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} f_p(p)
\end{aligned}
$$

Note that only if $f_p(p) \propto p^\alpha(1-p)^\beta$, then we have that:

$$f_{p|X}(p|X_1,\ldots,X_n) \propto p^{\alpha+\sum_{i=1}^n X_i}(1-p)^{\beta+n-\sum_{i=1}^n X_i}$$

Otherwise the kernel would not match between the prior and the posterior, which would not yield us the desired conjugation. Which means that the prior for optimist is also in the beta family of distributions.

Now let's find the prior parameters for both. For realist, as seen above, take $\alpha = \beta = 1$. For optimist, we need to solve the following system:

$$\begin{cases} \frac{\alpha}{\alpha+\beta} = \frac{1}{2} \\ \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{1}{36} \end{cases} \Rightarrow \alpha = \beta \Rightarrow \frac{\alpha^2}{(2\alpha)^2(2\alpha+1)} = \frac{1}{36} \Rightarrow \alpha = 4 = \beta$$

Therefore Optimist's prior is a beta distribution with parameters $\alpha = \beta = 4$.

b)

Note for a Bernoulli random variable we have that $f(x|p) = p^x(1-p)^{1-x}$. Then:

$$
\begin{aligned}
I(p) &= E\left[\left(\frac{d}{dp}\ln(f(x|p))\right)^2\right] \\
&= E\left[\left(\frac{d}{dp}xln(p) + (1-x)ln(1-p)\right)^2\right] \\
&= E\left[\left(\frac{x}{p} - \frac{1-x}{1-p}\right)^2\right] \\
&= E\left[\left(\frac{x}{p}\right)^2 - 2\frac{x(1-x)}{p(1-p)} + \left(\frac{1-x}{1-p}\right)^2\right] \\
&= \frac{1}{p^2}E[X^2] - \frac{2}{p(1-p)}(E[X] - E[X^2]) + \frac{1}{(1-p)^2}(1 - 2E[X] + E[X^2]) \\
&\text{Recall for a bernoulli(p) } E[X] = E[X^2] = p \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} \\
&= \frac{1}{p(1-p)}
\end{aligned}
$$

Then we have that:

$$
\pi(p) \propto \sqrt{\frac{1}{p(1-p)}} = p^{1/2-1}(1-p)^{1/2-1}
$$

Notice this is exactly the kernel for a beta distribution with parameters $\alpha = \beta = \frac{1}{2}$, which evidently belongs to the beta family.

c)

First of all, going back to what we did in a), suppose the prior has a $Beta(\alpha, \beta)$ distribution, and that n independent games are played, resulting in $\sum_{i=1}^n X_i$ wins. Then:

$$
\begin{aligned}
f_{p|X}(p|X) &\propto p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} p^{\alpha-1}(1-p)^{\beta-1} \\
&= p^{\sum_{i=1}^n X_i + \alpha - 1}(1-p)^{n-\sum_{i=1}^n X_i + \beta - 1}
\end{aligned}
$$

Then we see the posterior distribution is Beta with parameters $\alpha_{posterior} = \sum_{i=1}^n X_i + \alpha$ and $\beta_{posterior} = n - \sum_{i=1}^n X_i + \beta$. We then have that $\sum_{i=1}^n X_i = 12$ and $n = 25$. Therefore, the posteriors for each student are:

- Optimist: $Beta(16, 17)$
- Realist: $Beta(13, 14)$
- Pessimist: $Beta(12.5, 13.5)$

d)

We calculate the posterior means. Recall for a $Beta(\alpha, \beta)$, the mean is given by $\frac{\alpha}{\alpha+\beta}$:

```
post_alpha <- c(Optimist = 16, Realist = 13, Pessimist = 12.5)
post_beta <- c(Optimist = 17, Realist = 14, Pessimist = 13.5)

post_mean <- post_alpha/(post_alpha + post_beta)
post_mean
```

```
##  Optimist   Realist Pessimist
## 0.4848485 0.4814815 0.4807692
```

Now, for the credible intervals, let's use the fact that $\frac{p}{1-p}\frac{\beta}{\alpha} \sim F(2\alpha, 2\beta)$. Then, to get the $100(1-\lambda)$ percent interval:

$$
\begin{aligned}
1-\lambda &= P\left(Q_{F(2\alpha,2\beta)}(\lambda/2) \leq \frac{p}{1-p}\frac{\beta}{\alpha} \leq Q_{F(2\alpha,2\beta)}(1-\lambda/2)\right) \\
&= P\left(\frac{\alpha}{\beta}Q_{F(2\alpha,2\beta)}(\lambda/2) \leq \frac{p}{1-p} \leq \frac{\alpha}{\beta}Q_{F(2\alpha,2\beta)}(1-\lambda/2)\right)
\end{aligned}
$$

Now note that:

$$
\begin{aligned}
A \leq \frac{p}{1-p} \leq B &\iff \frac{1}{A} \geq \frac{1-p}{p} \geq \frac{1}{B} \\
&\iff \frac{1}{A} + 1 \geq \frac{1}{p} \geq \frac{1}{B} + 1 \\
&\iff \frac{A}{A+1} \leq p \leq \frac{B}{B+1}
\end{aligned}
$$

Using this in the previous equations we get that:

$$
1-\lambda = P\left(\frac{\frac{\alpha}{\beta}Q_{F(2\alpha,2\beta)}(\lambda/2)}{1+\frac{\alpha}{\beta}Q_{F(2\alpha,2\beta)}(\lambda/2)} \leq p \leq \frac{\frac{\alpha}{\beta}Q_{F(2\alpha,2\beta)}(1-\lambda/2)}{1+\frac{\alpha}{\beta}Q_{F(2\alpha,2\beta)}(1-\lambda/2)}\right)
$$

Then a 95% credible interval can be obtained by substituting $\lambda = 0.05$ and plugging in each student's posterior parameters:

```
post_params <- list(Optimist = c(alpha = 16, beta = 17),
                    Realist = c(alpha = 13, beta = 14),
                    Pessimist = c(alpha = 12.5, beta = 13.5))

lambda <- 0.05

lapply(post_params, function(par){
  a <- unname(par[1])
  b <- unname(par[2])

  A <- a/b * qf(p = lambda/2, df1 = 2*a, df2 = 2*b)
  B <- a/b * qf(p = 1-lambda/2, df1 = 2*a, df2 = 2*b)

  return(c(Lower = A/(A+1), Upper = B/(B+1) ))
})
```

```
## $Optimist
##     Lower     Upper
## 0.3188750 0.6525632
##
## $Realist
##     Lower     Upper
## 0.2992722 0.6662918
##
## $Pessimist
##     Lower     Upper
## 0.2953705 0.6689728
```

e)

For Skeptic, we will use MLE as his method of analysis. Recall for a Bernoulli distribution, the MLE for the parameter p is given by $\hat{p} = \bar{X}$. Also, as shown above, $I(p) = \frac{1}{p(1-p)}$, then we can estimate the Fisher information via $I(\hat{p}) = \frac{1}{\hat{p}(1-\hat{p})} = \frac{1}{\bar{X}(1-\bar{X})}$. Then, via asymptotic normality of the MLE, we can estimate a 95% confidence interval as:

$$\hat{p} \pm z_{1-0.05/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

```r
n <- 25
p_mle <- 12/n

c(Lower = p_mle - qnorm(p = 1-0.05/2)*sqrt( (p_mle*(1-p_mle))/n),
  Upper = p_mle + qnorm(p = 1-0.05/2)*sqrt( (p_mle*(1-p_mle))/n))
```

```
##     Lower     Upper
## 0.2841605 0.6758395
```