Maestría en Computo Estadístico
Inferencia Estadística
Tarea 8
30 de noviembre de 2020
Enrique Santibáñez Cortés
Repositorio de Git: Tarea 8, IE.

1. Sean $X_1, \cdots, X_n \sim Uniforme(0, \theta)$. Sea $f(\theta) \propto 1/\theta$. Calcule la densidad posterior.

RESPUESTA

> **Definición: 1** *Sea $X_1, \cdots, X_n$ con distribución a priori $f(\theta)$ entones la distribución posteriori cumple que*
>
> $$f(\theta|x) \propto L(\theta) f(\theta),$$
>
> *donde $L(\theta)$ representa la función de verosimilitud.*

Ocupando la definición (3) como sabemos que la función de verosimilitud para el caso en que la m.a es Uniforme es

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \frac{1}{\theta^n}_{\{\theta \leq \text{máx}(x_1, \cdots, x_n)\}} = \frac{1}{\theta^n}_{\{\theta \leq x_{(n)}\}}.$$

Entonces

$$f(\theta|x) \propto L(\theta) f(\theta) = \frac{1}{\theta^n}_{\{\theta \leq x_{(n)}\}} \frac{1}{\theta}$$
$$\propto \frac{1}{\theta^{n+1}}_{\{\theta \leq x_{(n)}\}}. \quad \blacksquare.$$

Ahora calculamos la constante de normalización,

$$\int_{x(n)}^{\infty} \frac{1}{\theta^{n+1}} = -\frac{1}{n\theta^n}\Big|_{x_{(n)}}^{\infty} = \frac{1}{n x_{(n)}^n}$$

Por lo tanto, la distribución a posteriori es

$$f(\theta|x) = \frac{n x_{(n)}^n}{\theta^{n+1}} 1_{\{\theta \leq x_{(n)}\}}.$$

2. Sean $X_1, \cdots, X_n \sim Normal(\mu, 1)$

a) Simule un conjunto de datos (use $\mu = 5$) de $n = 100$ observaciones.

RESPUESTA
Realizamos lo anterior con R.

```r
library(tidyverse)
set.seed(08081997) # fijamos la semilla
mu <- 5 # mu
n <- 100 # 100 observaciones
simulacion_normal <- rnorm(n, mu, 1) # simulación.
```

b) Tome $f(\mu) = 1$ y halle la densidad posteriori. Grafique la densidad.

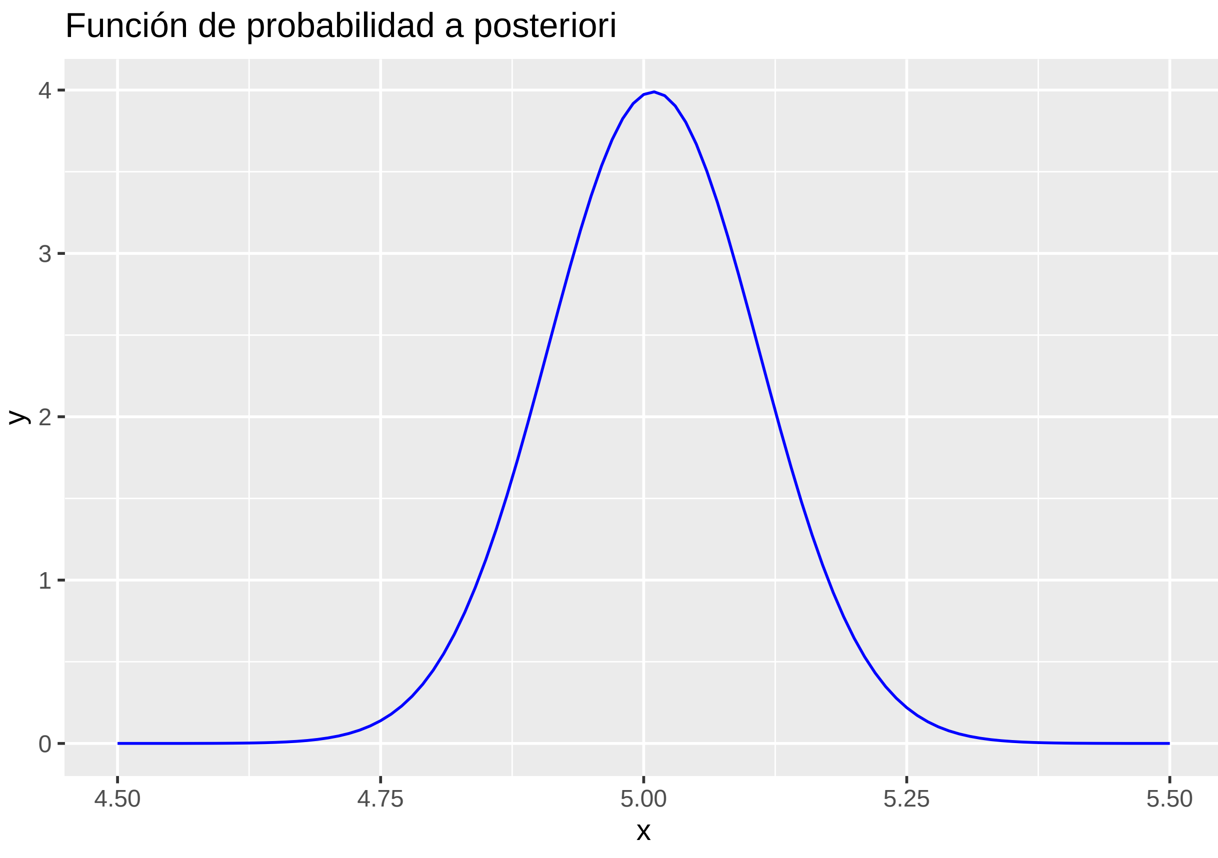
RESPUESTA
Por las clases sabemos que para el caso normal con desviación desconocida, la densidad posteriori es

$$f(\theta|x) \propto L(\theta)f(\theta) = e^{-\frac{(\mu-x)^2}{2/n}}.$$

Es decir, $\mu|X^n \sim Normal(\bar{x}, 1/\sqrt{n})$. Por lo que procedemos a graficar la densidad.

```
mu_posteriori <- mean(simulacion_normal) # parametros
std_posteriori <- sqrt(1/n)
df_posteriori <- data.frame(x=seq(4.5, 5.5, 0.01), # simulacion a posteriori
      y=dnorm(seq(4.5, 5.5, 0.01), mean=mu_posteriori, sd=std_posteriori))

ggplot(df_posteriori, aes(x,y))+ # graficamos
  geom_line(color="blue")+
  labs(title = "Función de probabilidad a posteriori")
```



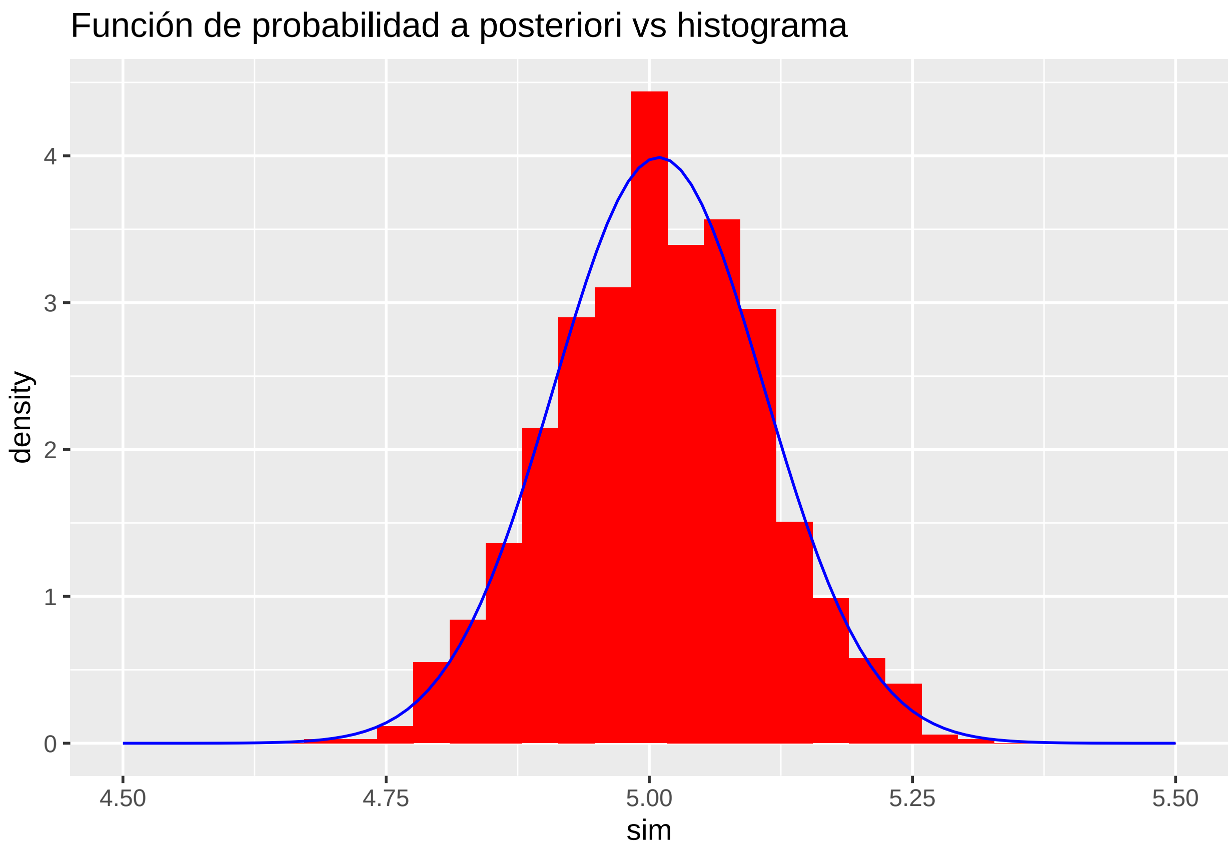
c) Simule 1000 observaciones de la posteriori. Grafique un histograma y compare con la densidad del punto anterior.

RESPUESTA
Simulamos 1000 observaciones de la a posteriori.

```
simulacion_aposteriori <- data.frame(sim=rnorm(1000, mu_posteriori, std_posteriori))
```

Comparamos el histograma con la función teorica.

```r
ggplot(simulacion_aposteriori, aes(x=sim))+
geom_histogram(aes(bins=40, y=..density..), fill="red")+
  geom_line(data=df_posteriori, aes(x,y), color="blue")+
  labs(title = "Función de probabilidad a posteriori vs histograma")
```



Observamos que el histograma y la densidad teórica son muy parecidas, por lo que podemos decir que si corresponde a lo que encontramos en el inciso b).

d) Sea $\theta = e^{\mu}$. Halle la densidad posteriori para θ de forma analítica y por simulación.

RESPUESTA
Ocupando algunas propiedades tenemos que

$$H_(\mu|x^n) = \mathbb{P}(e^{\mu} < z) = \mathbb{P}(\mu < \log(z)) = \int_0^{\log(x)} \frac{1}{\sqrt{2\pi}} e^{\frac{-n(\mu-\bar{x})}{2}} \Rightarrow$$
$$h(\mu|x^n) = \frac{1}{x\sqrt{2\pi}} e^{\frac{-n(\log(x)-\bar{x})}{2}}.$$

Por lo tanto, podemos concluir que

$$\theta = e^{\mu} \sim LogNormal(\bar{x}, 1/\sqrt{n}).$$

Ahora ocuparemos simulación para encontrar su densidad. En el punto c) tenemos la simulación entos tenemso que los nuevos datos son

```r
dlnorm=function(mu){ # programamos la función lognormal, ya que hay varias versiones
p<-(10/(sqrt(2*pi)*mu))*exp(-(n/2)*(log(mu)-mu_posteriori)**2 )
p
```

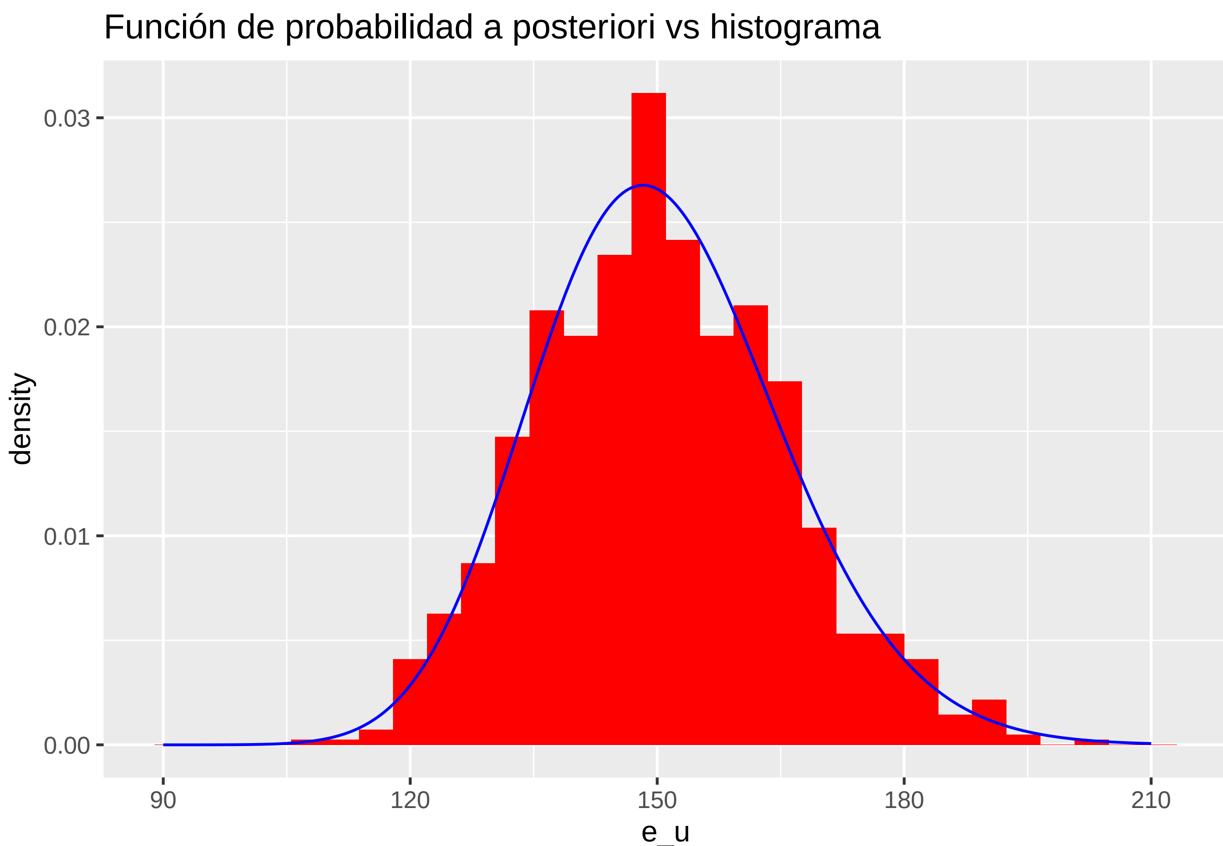
```
}

df_posteriori_d <- data.frame(x=seq(90,210,by=0.1),
              y=dlnorm(seq(90,210,by=0.1)))

simulacion_aposteriori$e_u <- exp(simulacion_aposteriori$sim)

ggplot(simulacion_aposteriori, aes(x=e_u))+
geom_histogram(aes(bins=40, y=..density..), fill="red")+
  geom_line(data=df_posteriori_d, aes(x,y), color="blue")+
  labs(title = "Función de probabilidad a posteriori vs histograma")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



e) Halle un intervalo posteriori del 95 % para θ.

RESPUESTA

Definición: 2 *Un intervalo posteriori del 95 % se puede obtener númericamente hallando a y b tal que*

$$\int_a^b f(\theta|x^n)d\theta = 0{,}95.$$

Entonces, normalmente se escogen a y b de manera que a deja probabilidad $(1-0{,}95)/2$ a su izquierda y b $(1-0{,}95)/2$ a la derecha. Por lo anterior, como sabemos que $\theta|x^n$ es Lognormal, entonces la calculamos númericamenta

```
a = qlnorm(0.025, mu_posteriori, 1/sqrt(n))
b = qlnorm(0.975, mu_posteriori, 1/sqrt(n))
```

Es decir, el intervalo a posterior del 95 % para θ

$$IC(\theta) = (123{,}1093104, 182{,}1927987)$$

f) Halle un intervalo de confianza del 95 % para θ.

RESPUESTA
Ocupando R, tenemos que el intervalo de confianza del 95 % para θ es

```
quantile(simulacion_aposteriori$e_u,c(0.025,0.975))
```

```
##     2.5%    97.5%
## 122.7885 181.7581
```

, es decir,

$$IC(\mu) = (122{,}7885, 181{,}7581). \quad \blacksquare.$$

Si comparamos ambos intervalos estos son muy parecidos, el por que son muy parecidos es por el tamaño de muestra.

3. Sean $X_1, \cdots, X_n \sim Poisson(\lambda)$. Sea $\lambda \sim Gamma(\alpha, \beta)$ la priori. Demuestre que la densidad posteriori es también una Gamma.

RESPUESTA

Definición: 3 *Sea $X_1, \cdots, X_n$ con distribución a priori $f(\theta)$ entones la distribución posteriori cumple que*

$$f(\theta|x) \propto L(\theta)f(\theta),$$

donde $L(\theta)$ representa la función de verosimilitud.

Ocupemos la definición (3), para el caso Poisson sabemos que la función de verosimilitud es

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!} \propto \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda},$$

ahora la función de densidad de una v.a Gamma(α, β) es

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0.$$

Entonces tenemos que

$$\begin{aligned} f(\theta) &\propto L(\lambda)f(\lambda) = \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &\propto \lambda^{\alpha^*-1} e^{-\beta*\lambda}, \end{aligned}$$

donde $\alpha^* = \alpha + \sum_{i=1}^{n} x_i$ y $\beta* = \beta + n$. Por lo que podemos concluir que la función a posteriori se distribuye como una $Gamma(\alpha^*, \beta^*)$. $\blacksquare$.

4. Suponga que a 50 personas e les da un placebo y a otras 50 un nuevo tratamiento. 30 de los pacientes con placebo muestran mejoría, mientras que 40 pacientes con el tratamiento nuevo muestran mejoría. Sea $\tau = p_2 - p_1$, donde p_2 es la probabilidad de mejorar bajo el tratamiento y p_1 es la probabilidad de mejorar bajo el placebo.

a) Calcule el EMV de τ. Halle el error estándar y un intervalo de confianza del 90 %.

RESPUESTA
Por el problema, podemos considerar que tenemos una m.a. $Binomial(50, p_1)$ y otra $Binomial(50, p_2)$. Entonces, como vimos rápidamente en clase, y como es probado en el libro de Casella G. y Berger R. L . (Statistical Inference) en la pág. 294 (en la sección de máxima verosimilitud), los estimadores de máxima verosimilitud son equivariantes independientemente de la dimensión del espacio de parámetros. Entonces si consideramos a τ como una función de los parámetros p_1 y p_2 tenemos:

$$\begin{aligned}\tau &= g(p_1, p_2) = p_1 - p_2 \\ &\Rightarrow \hat{\tau} = g(\hat{p_1} - \hat{p_2}) = \hat{p_1} - \hat{p_2} = \bar{X}_1 - \bar{X}_2\end{aligned}$$

Por lo tanto, $\tau_{MLE} = 40/50 - 30/50 = 0{,}2$.

Ahora encontremos un intervalo de confianza del 90 %, para ello encuentremos la matriz de información de Fisher $I(p_1, p_2)$. Como los parámetros p_1 y p_2 son independientes su función de distribución conjunta es el producto de sus distribuciones por separado consideremos la función de likelihood con una muestra de tamaño uno.

$$\begin{aligned}L(p_1, p_2) &= L(x; p_1, p_2) = L(x; p_1)L(x; p_2) \\ &\Rightarrow \ln L(p_1, p_2) = \ln L(x; p_1) + \ln L(x; p_2)\end{aligned}$$

Si derivamos lo anterior con respecto a p_i para obtener los scores $s(X; p_i)$:

$$\begin{aligned}s(X; p_i) = \frac{\partial \ln L(p_1, p_2)}{\partial p_i} &= \frac{\partial \ln \binom{n_i}{x} p_i^x (1 - p_i)^{n_i - x}}{\partial p_i} \\ &= \frac{\partial}{\partial p_i}((\ln(n_i!) - \ln((n_i - x)!) - \ln(x!)) + x \ln p_i + (n_i - x)\ln(1 - p_i)) \\ &= \frac{\partial}{\partial p_1}(C_i + x \ln p_1 + (n_i - x)\ln(1 - p_i)) \\ &= \frac{x}{p_i} - \frac{n_i - x}{1 - p_i} = \frac{x - n_i p_i}{p_i(1 - p_i)}\end{aligned}$$

Por un lado sabemos que la matriz de información de Fisher la construiremos como:

$$I_{(p_1, p_2)} = -\begin{pmatrix} \mathbb{E}\left(\frac{\partial}{\partial p_1} s(X; p_1)\right) & \mathbb{E}\left(\frac{\partial}{\partial p_1} s(X; p_2)\right) \\ \mathbb{E}\left(\frac{\partial}{\partial p_2} s(X; p_1)\right) & \mathbb{E}\left(\frac{\partial}{\partial p_2} s(X; p_2)\right) \end{pmatrix}$$

Si notamos que $\frac{\partial}{\partial p_j} s(X; p_i) = 0$ si $i \neq j$ entonces el calculo de $I_(p_i, p_2)$ se reduce a

$$I_{(p_1, p_2)} = -\begin{pmatrix} \mathbb{E}\left(\frac{\partial}{\partial p_1} s(X; p_1)\right) & 0 \\ 0 & \mathbb{E}\left(\frac{\partial}{\partial p_2} s(X; p_2)\right). \end{pmatrix}$$

Finalmente solo resta calcular las matrices de Inforción de Fisher considerando los parámetros por separado es decir $\mathbb{E}\left(\frac{\partial^2}{\partial p_j^2}\ln L(X;p_j)\right) = \mathbb{E}\left(\frac{\partial}{\partial p_j}s(X;p_j)\right)$. Por un resultado visto en clase, en el libro, y como ejercicio en esta tarea sabemos que $\mathbb{E}\left(\frac{\partial^2}{\partial p_j^2}\ln L(X;p_j)\right) = -\mathbb{E}\left(s(X;p_j)^2\right)$, y lo calculamos:

$$\begin{aligned}
\mathbb{E}\left(s(X;p_i)^2\right) &= \mathbb{E}\left(\left(\frac{x-n_ip_i}{p_i(1-p_i)}\right)^2\right)\\
&= \frac{1}{p_i^2(1-p_i)^2}\mathbb{E}\left(x^2-2xn_ip_i+n_i^2p_i^2\right)\\
&= \frac{1}{p_i^2(1-p_i)^2}\left(\mathbb{E}(x^2)-2\mathbb{E}(x)n_ip_i+n_i^2p_i^2\right)\\
&= \frac{1}{p_i^2(1-p_i)^2}\left(\mathbb{E}(x^2)-2n_i^2p_i^2+n_i^2p_i^2\right)\\
&= \frac{1}{p_i^2(1-p_i)^2}\left(\mathbb{E}(x^2)-n_i^2p_i^2\right)\\
&= \frac{1}{p_i^2(1-p_i)^2}\left(n_ip_i-n_ip_i^2-n_i^2p_i^2+n_i^2p_i^2\right)\\
&= \frac{n_ip_i(1-p_i)}{p_i^2(1-p_i)^2} = \frac{n_i}{p_i(1-p_i)}
\end{aligned}$$

Donde la penultima igualdad se debe a que $V(x) = \mathbb{E}(x^2)-\mathbb{E}(x)^2 \to \mathbb{E}(x^2) = V(X)+\mathbb{E}(x)^2 = n_ip_i(1-p_i) = n_i^2p_i^2 = n_i(p_i-p_i^2)+n_i^2p_i^2 = n_ip_i - n_ip_i^2+n_i^2p_i^2$. Por lo que la matriz queda

$$I_{(p_1,p_2)} = -\begin{pmatrix}\mathbb{E}\left(\frac{\partial}{\partial p_1}s(X;p_1)\right) & 0\\ 0 & \mathbb{E}\left(\frac{\partial}{\partial p_2}s(X;p_2)\right)\end{pmatrix} = \begin{pmatrix}\frac{n_1}{p_1(1-p_1)} & 0\\ 0 & \frac{n_2}{p_2(1-p_2)}\end{pmatrix}$$

Ahora, por el método Delta (de multiparámetros) sabemos que:

$$\hat{se}(\hat{\psi}) = \sqrt{(\hat{\nabla}\psi)^t\hat{J}_n(\hat{\nabla}\psi)}$$

Por un lado $(\hat{\nabla}\psi)^t = (1,-1)^t$ y por otro:

$$\hat{J}_n = \hat{I}^{-1}_{(p_1,p_2)} = \begin{pmatrix}\frac{n_1}{\hat{p}_1(1-\hat{p}_1)} & 0\\ 0 & \frac{n_2}{\hat{p}_2(1-\hat{p}_2)}\end{pmatrix}^{-1} = \begin{pmatrix}\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} & 0\\ 0 & \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\end{pmatrix}$$

Por lo que podemos escribir:

$$(\hat{\nabla}\psi)^t\hat{J}_n(\hat{\nabla}\psi) = (1,-1)^t\begin{pmatrix}\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} & 0\\ 0 & \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix} = (1,-1)^t\begin{pmatrix}\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\\ -\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\end{pmatrix} = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$$

Con lo que obtenemos $\hat{se}(\hat{\tau}) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$. Además por el método Delta sabemos que $\frac{\hat{\tau}-\tau}{\hat{se}(\hat{\tau})} \sim N(0,1)$ por lo que podemos construir un intervalo de confianza como sigue:

$$(\hat{\tau}-z_{,05}\hat{se}(\hat{\tau}), \hat{\tau}+z_{,05}\hat{se}(\hat{\tau}))$$

Por lo tanto, considerando $z_{0,10} = 1{,}64$ y $\hat{se}(\hat{\tau}) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0{,}8(0{,}2)}{50}+\frac{0{,}6(0{,}4)}{50}} = 0{,}089$ tenemos que un intervalo de confianza del 90 % es

$$(0{,}2-(1{,}64)(0{,}089), 0{,}2+(1{,}64)(0{,}089)) \Leftrightarrow (\mathbf{0{,}05404}, \mathbf{0{,}34596}).$$

b) Sea la priori $f(p_1, p_2) = 1$. Use simulación para hallar la media posterior y un intervalo posterior del 90 % para τ.

RESPUESTA

> **Lema: 1** *Sea $X_1, \cdots, X_n \sim Ben(p)$ con distribución a priori $f(p) = 1$ entonces la distribución posteriori cumple que*
>
> $$f(\theta|x) \propto L(\theta)f(\theta) = Beta(x+1, n+1-x).$$

Dado que $X_1 \sim Binomial(50, p_1), X_2 \sim Binomial(50, p_2)$ tenemos que

$$\begin{aligned} g(p_1, p_2|x_1, x_2) &\propto f(p_1, p_2)f(x_1|p_1)f(x_2|p_2) \\ &= p_1^{(x_1+1)-1}(1-p_1)^{(51-x_1)-1}p_2^{(x_2+1)-1}(1-p_2)^{(51-x_2)-1} \\ &= f(p_1|x_1)f(p_2|x_2). \end{aligned}$$

Ahora ocupando el lema (1) tenemos que

$$f(p_1|x_1) \sim Beta(X_1+1, 51-X_1) \quad \& \quad f(p_2|x_2) \sim Beta(X_2+1, 51-X_2)$$

Con lo anterior procedemso a realizar las simulaciones para hallar la media posterior y un intervalo posterior del 90 % para τ.

```
B <- 5000
sim.tau <- c()
x1 <- 30
x2 <- 40
for (i in 1:B) {
    sim.p1 <- rbeta(1, x1+1, 51-x1)
    sim.p2 <- rbeta(1, x2+1, 51-x2)
    sim.tau[i] <- sim.p2 - sim.p1
}

tau.post.mean <- mean(sim.tau)
print(paste0("La media aposteriori de tau es= ", round(tau.post.mean, 3)))
```

```
## [1] "La media aposteriori de tau es= 0.194"
```

```
tau.post.low <- quantile(sim.tau, 0.05)
tau.post.high <- quantile(sim.tau, 0.95)
print(paste0("Intervalo posteriori de tau es = (", round(tau.post.low, 3), ", ", round(tau.post.h
```

```
## [1] "Intervalo posteriori de tau es = (0.049, 0.338)"
```

c) Sea

$$\phi = \log\left(\left(\frac{p_1}{1-p_1}\right) / \left(\frac{p_2}{1-p_2}\right)\right)$$

el radio log-odds. Note que $\phi = 0$ si $p_1 = p_2$. Calcule el EMV de ϕ.

RESPUESTA

Utilizando el mismo razonamiento del inciso a) sabemos que los estimadores de máxima vrosimilitud son equivariantes independientemendete de la dimensión del espacio de parámetros. Entonces Si consideramos a ϕ como una función de los parámetros p_1 y p_2 tenemos que:

$$
\begin{aligned}
&\phi = g(p_1, p_2) = \log\left(\left(\frac{p_1}{1-p_1}\right) / \left(\frac{p_2}{1-p_2}\right)\right)\\
&\Rightarrow \hat{\phi} = g(\hat{p_1}, \hat{p_2}) = \log\left(\left(\frac{\hat{p_1}}{1-\hat{p_1}}\right) / \left(\frac{\hat{p_2}}{1-\hat{p_2}}\right)\right) = \log\left(\left(\frac{\bar{X}_1}{1-\bar{X}_1}\right) / \left(\frac{\bar{X}_2}{1-\bar{X}_2}\right)\right). \quad \blacksquare.
\end{aligned}
$$