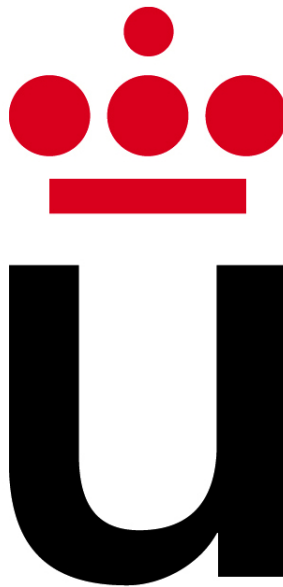


CASO PRÁCTICO III

Modelización y tratamiento de la incertidumbre

INFERENCIA BAYESIANA

José Ignacio Escribano



MÓSTOLES, 8 DE NOVIEMBRE DE 2015

Índice de figuras

1.	Función de densidad de la distribución empírica, a priori y a posteriori .	2
2.	Intervalo de probabilidad al 90 % sobre la distribución a posteriori . . .	3
3.	Función de densidad de la distribución predictiva	5
4.	Función de densidad de la distribución a priori	5
5.	Función de densidad de la distribución a posteriori	7
6.	Función de densidad de la distribución a priori y a posteriori	8
7.	Intervalo de probabilidad al 80 % sobre la distribución a posteriori . . .	8

Índice

1. Introducción	1
2. Estimando proporciones y predicción de futuras muestras	1
3. Estimando una media normal con una a priori discreta	4
4. Conclusiones	9
5. Código R	10

1. Introducción

Este caso práctico consta de dos partes relativas a dos modelos distintos: el primero es acerca del modelo beta-binomial, y el segundo, sobre el modelo normal-normal. En ambas partes se piden calcular intervalos de probabilidad y contrastes de hipótesis, entre otras cosas.

2. Estimando proporciones y predicción de futuras muestras

En esta primera parte, consideramos una población de 29 niños que tenían un contenido en plomo en los dientes de leche superior a 22.22 partes por millón, de los cuales, 22 terminaron la Educación Secundaria, y 7 que no lo hicieron. Consideraremos que la distribución a priori de la proporción p de niños que terminaron la Educación Secundaria sigue una distribución $p \sim \mathcal{Be}(1, 1)$.

Lo primero que tenemos que calcular es la función de verosimilitud $\mathcal{L}(p) = P(X = x|p)$. Sabemos que sólo hay dos resultados posibles: terminar o no terminar la Educación Secundaria, por lo que tenemos una distribución binomial con $n = 29$ y $x = 22$.

Por tanto, la función de verosimilitud es:

$$\begin{aligned}\mathcal{L}(p) &= P(X = 22|p) \\ &= \binom{29}{22} p^{22} (1-p)^{29-22} \\ &\propto p^{22} (1-p)^{29-22} \\ &= p^{22} (1-p)^7\end{aligned}$$

Por otro lado, tenemos como distribución a priori una $\mathcal{Be}(1, 1)$. La función de densidad de la distribución $\mathcal{Be}(\alpha, \beta)$ viene dada por

$$f(p|\alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}, & \text{si } 0 \leq p \leq 1 \\ 0, & \text{en otro caso} \end{cases}$$

donde $\Gamma(\cdot)$ denota función gamma definida como

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

Por tanto,

$$f(p|\alpha = 1, \beta = 1) = \frac{\Gamma(1+1)}{\Gamma(1) + \Gamma(1)} p^0 (1-p)^0 \\ \propto 1$$

Tenemos una distribución a priori que no nos da ninguna información sobre p .

La distribución a posteriori $f(p|x)$ viene dada por

$$f(p|x) \propto f(p) \cdot \mathcal{L}(p) \\ \propto 1 \cdot p^{22} (1-p)^7 \\ = p^{22} (1-p)^7$$

Comparando con la definición de función de densidad de una distribución beta, tenemos que la distribución a posteriori $p|x \sim \mathcal{Be}(23, 8)$.

La Figura 1 muestra las distribuciones a priori, a posteriori y la distribución empírica. Ésta última se ha generado generando con 1000 muestras de una distribución $\mathcal{Be}(23, 8)$ (nuestra distribución a posteriori) con el comando `rbeta` de R. Tanto la distribución a posteriori como la distribución empírica son bastante parecidas, por lo que el comando `rbeta` de R aproxima bien la función aún cuando el número de muestras es relativamente pequeño.

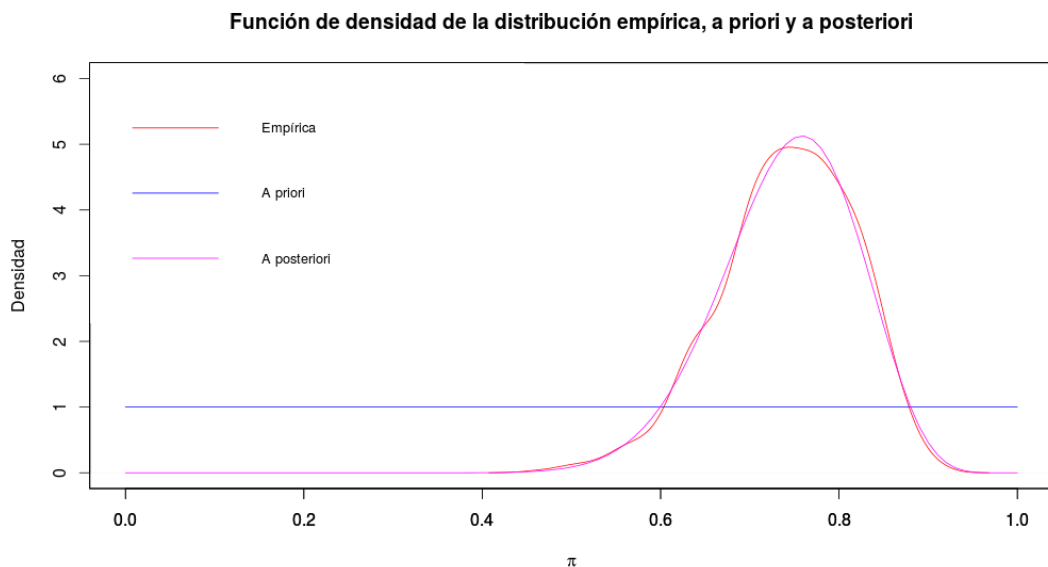


Figura 1: Función de densidad de la distribución empírica, a priori y a posteriori

Como tenemos nuestra distribución a posteriori, podemos calcular la media y la varianza a posteriori de esta distribución, ya que las fórmulas para la media y la varianza son conocidas:

$$E(x|p) = \frac{\alpha}{\alpha + \beta}$$

$$Var(x|p) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Sustituyendo en las fórmulas anteriores α por 23 y β por 8, se tiene que

$$E(x|p) = 0.7419$$

$$Var(x|p) = 0.0059$$

$$s(x|p) = \sqrt{Var(x|p)} = 0.0773$$

Para calcular el intervalo de probabilidad al 90 % calculamos los cuantiles que acumulan el 5 % y el 95 % de nuestra distribución a posteriori. Para ello aplicamos el comando qbeta de R. Así se tiene que el intervalo de probabilidad al 90 % es [0.6060, 0.8598]. La Figura 2 muestra el intervalo de probabilidad sobre la distribución a posteriori.

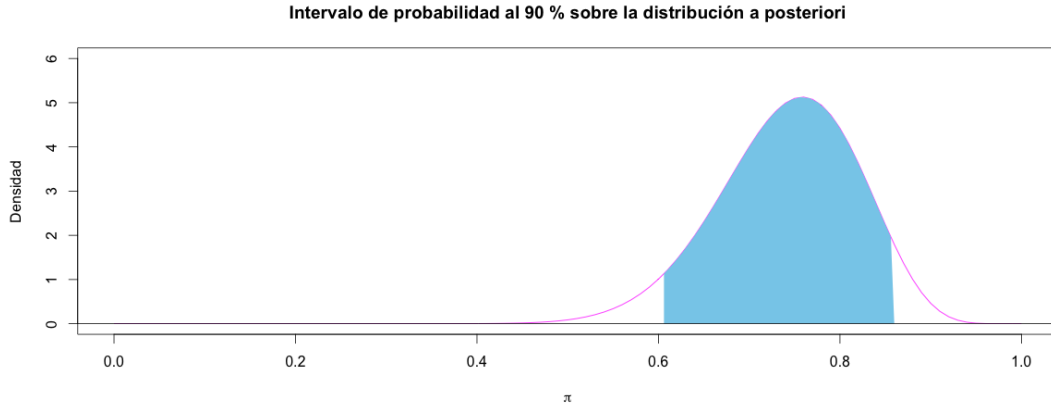


Figura 2: Intervalo de probabilidad al 90 % sobre la distribución a posteriori

También necesitamos hacer el siguiente contraste de hipótesis:

$$H_0 : p \leq 0.4$$

$$H_1 : p > 0.4$$

Para ello, calculamos la probabilidad de que nuestra distribución a posteriori sea menor que 0.4 y mayor que 0.4. Aplicamos el comando pbeta de R. Así tenemos que

$$P(H_0) = 4.9325 \cdot 10^{-5}$$

$$P(H_1) = 0.9999$$

Como $P(H_1) > P(H_0)$, rechazamos la hipótesis nula.

Por último, vamos a calcular la probabilidad predictiva del modelo beta-binomial de que haya k éxitos en los m siguientes ensayos, que viene dada por

$$P(k \text{ éxitos en } m|x) = \binom{m}{k} \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x)\Gamma(\beta + n - x)} \frac{\Gamma(\alpha + x + k)\Gamma(\beta + n - x + m - k)}{\Gamma(\alpha + \beta + m + n)}$$

donde $\Gamma(\cdot)$ es la función gamma definida como anteriormente, α y β son los parámetros de nuestra distribución beta a priori, x es el número de éxitos y n es el número de ensayos.

En nuestro caso, tenemos que $\alpha = \beta = 1$, $x = 22$ y $n = 29$. Hay que calcular la probabilidad de que al menos 9 de los 10 próximos niños terminen la Educación Secundaria.

Si denotamos con H al número de niños que terminan la Educación Secundaria entre 10 posibles, entonces

$$\begin{aligned} P(H \geq 9|x) &= 1 - P(H < 9|x) \\ &= 1 - P(H \leq 8|x) \\ &= 1 - \sum_{k=0}^8 P(k \text{ niños acaban la Educación Secundaria de } 10|x) \\ &= 1 - (0.00031 + 0.00209 + 0.00931 + 0.03026 + 0.07542 + 0.14666 \\ &\quad + 0.22095 + 0.24857 + 0.19026) \\ &= 1 - 0.92389 \\ &= 0.07610 \end{aligned}$$

Por tanto, la probabilidad predictiva de que al menos 9 de los 10 próximos niños acaben la Educación Secundaria es del 7.61 %.

La función de densidad de la probabilidad predictiva se muestra en la Figura 3.

3. Estimando una media normal con una a priori discreta

En esta segunda parte, queremos estimar la precipitaciones totales en forma de nieve. Los datos recogidos y_i provienen de una distribución normal con media μ y desviación típica $\sigma = 10$.

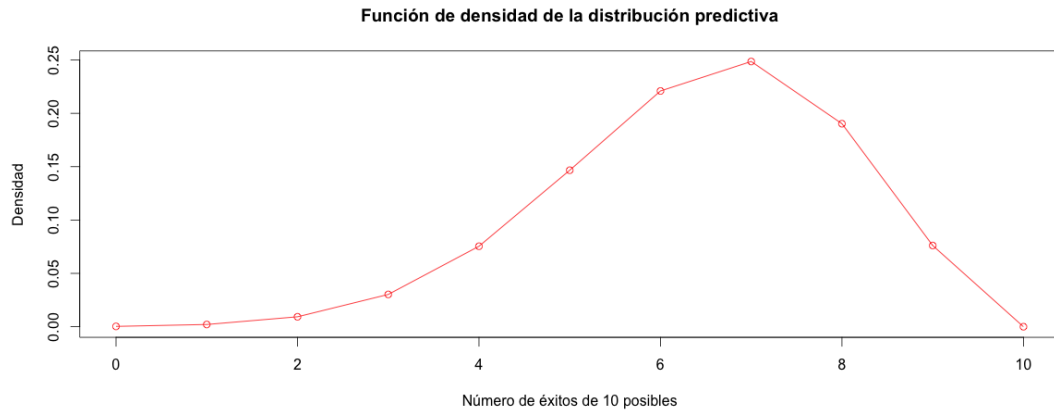


Figura 3: Función de densidad de la distribución predictiva

La distribución a priori viene dada por la siguiente función de densidad:

$$f(\mu) = \begin{cases} 0.1, & \text{si } \mu = 20 \\ 0.15, & \text{si } \mu = 30 \\ 0.25, & \text{si } \mu = 40 \\ 0.25, & \text{si } \mu = 50 \\ 0.15, & \text{si } \mu = 60 \\ 0.1, & \text{si } \mu = 70 \end{cases}$$

La Figura 4 muestra esta distribución.

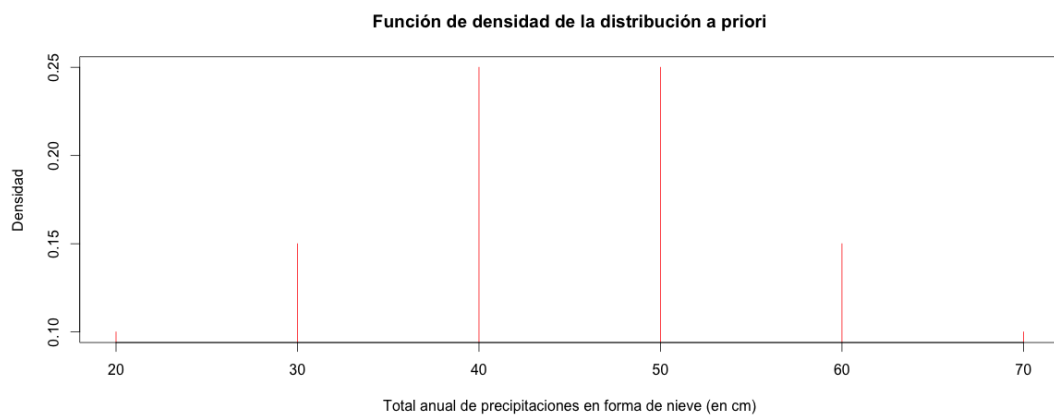


Figura 4: Función de densidad de la distribución a priori

La función de verosimilitud para este modelo viene dada por

$$\begin{aligned}
\mathcal{L}(\mu) &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \prod_{i=1}^n \exp \left[-\frac{1}{2} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right] \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[\sum_{i=1}^n -\frac{1}{2} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right] \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 + \mu^2 - 2y_i\mu) \right] \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[-\frac{1}{2\sigma^2} \left(n\mu^2 + \sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i \right) \right]
\end{aligned}$$

Los valores de y (precipitaciones de nieve en cm) son los siguientes:

y	38.6	42.4	57.5	40.5	51.7	67.1	33.4	60.9	64.1	40.1	40.7
-----	------	------	------	------	------	------	------	------	------	------	------

Calculamos la función de verosimilitud para cada uno de los valores de μ de la distribución a priori.

μ	20	30	40	50	60	70
$\mathcal{L}(\mu)$	$8.907 \cdot 10^{-41}$	$3.315 \cdot 10^{-30}$	$7.580 \cdot 10^{-25}$	$1.064 \cdot 10^{-24}$	$9.193 \cdot 10^{-30}$	$4.875 \cdot 10^{-40}$

Calculamos la distribución a posteriori para μ como $f(\mu|y) = \mathcal{L}(y|\mu)f(\mu)$.

$\mathcal{L}(\mu)$	$8.907 \cdot 10^{-41}$	$3.315 \cdot 10^{-30}$	$7.580 \cdot 10^{-25}$	$1.064 \cdot 10^{-24}$	$9.193 \cdot 10^{-30}$	$4.875 \cdot 10^{-40}$
$f(\mu)$	0.10	0.15	0.25	0.25	0.15	0.10
$f(\mu y)$	$9.807 \cdot 10^{-42}$	$4.972 \cdot 10^{-31}$	$1.895 \cdot 10^{-25}$	$2.662 \cdot 10^{-25}$	$1.378 \cdot 10^{-30}$	$4.875 \cdot 10^{-41}$

Si nos fijamos en los valores de $f(\mu|y)$ vemos que no tenemos una verdadera distribución, puesto que los valores calculados no suman 1. Para que sea una distribución, sumamos todos los valores calculados y dividimos cada valor por la suma mencionada anteriormente.

La suma de todos los valores es $4.557 \cdot 10^{-25}$, por lo que debemos dividir entre este valor cada uno de los $f(\mu|y)$. Por tanto, la distribución a posteriori (Figura 5) es la siguiente:

μ	20	30	40	50	60	70
$f(\mu y)$	$1.954 \cdot 10^{-17}$	$1.091 \cdot 10^{-6}$	$4.158 \cdot 10^{-1}$	$5.841 \cdot 10^{-1}$	$3.025 \cdot 10^{-6}$	$1.069 \cdot 10^{-16}$

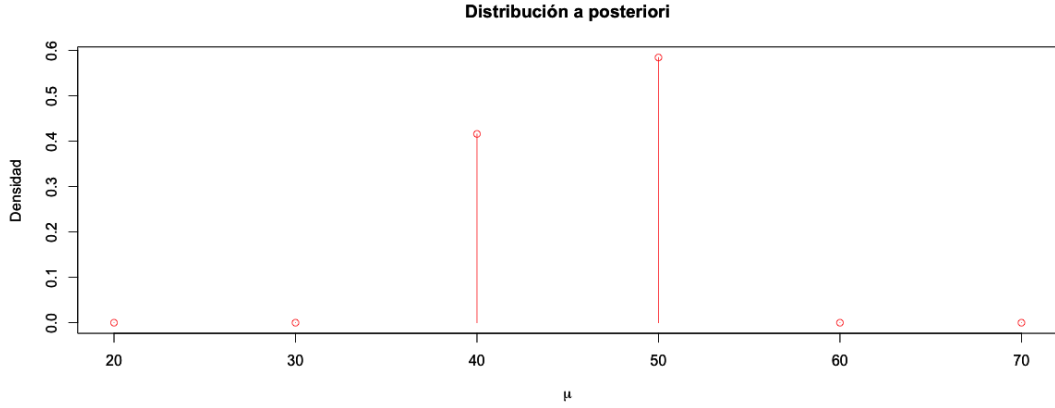


Figura 5: Función de densidad de la distribución a posteriori

Si queremos calcular un intervalo de probabilidad al 80 % para μ , usamos el comando `discint` de R. Esta función nos devuelve que el intervalo es $[40, 50]$, aunque la probabilidad de este intervalo es del 99.99 %, mayor que el pedido.

Si ahora consideramos que la distribución para $\mu \sim \mathcal{N}(\mu_0 = 60, \sigma_0 = 5)$, la distribución a posteriori viene dada por

$$\begin{aligned}
 f(\mu|y) &\propto f(\mu)\mathcal{L}(\mu) \\
 &\propto \exp \left[-\frac{1}{2} \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 \right] \left\{ \prod_{i=1}^n \exp \left[-\frac{1}{2} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right) \left[\mu - \left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \right) \right]^2 \right\} \\
 \text{Por tanto, } p|y &\sim \mathcal{N} \left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \sqrt{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}} \right).
 \end{aligned}$$

Sustituyendo, en la fórmula anterior se tiene que $p|y \sim \mathcal{N}(48.9625, 2.5)$. La Figura ?? muestra las funciones de densidad de las distribuciones a priori y a posteriori.

Para calcular el intervalo de probabilidad al 80 % para μ usamos el comando `qnorm` de R. Usamos los cuantiles 0.1 y 0.9 para obtener el intervalo pedido. R nos devuelve que el intervalo es $[44.850, 53.074]$. En la Figura 7 se muestra el intervalo de probabilidad al 80 % junto con la distribución a posteriori.

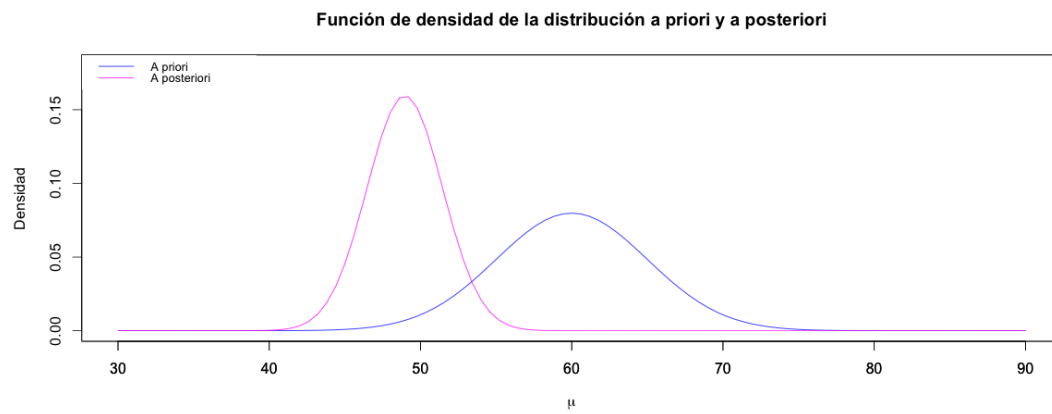


Figura 6: Función de densidad de la distribución a priori y a posteriori

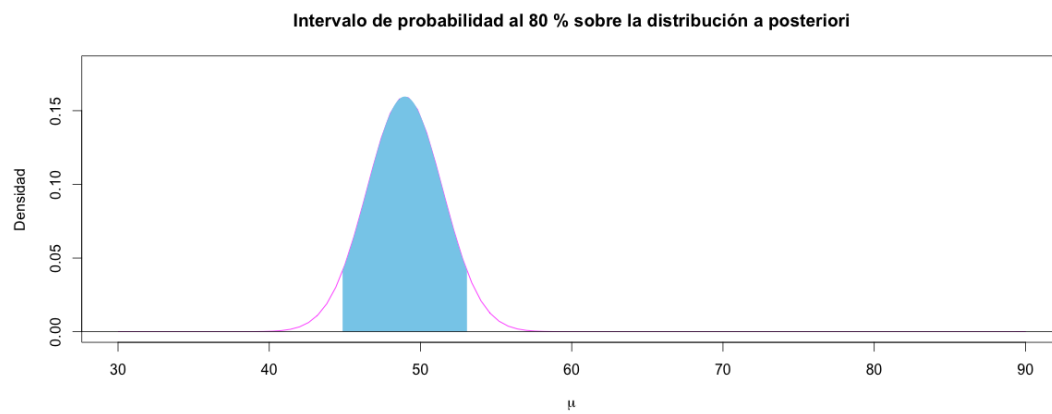


Figura 7: Intervalo de probabilidad al 80 % sobre la distribución a posteriori

4. Conclusiones

En este caso práctico hemos visto dos modelos clásicos de inferencia bayesiana: el modelo beta-binomial y el modelo normal-normal. Hemos puesto en práctica toda la teoría de ambos modelos para calcular probabilidades sobre la distribución a posteriori, intervalos de probabilidad, contrastes de hipótesis, etc. Para calcular todo lo anterior hemos hecho uso del poderoso software estadístico R, que nos ha ahorrado mucho tiempo evitando realizar cálculos tediosos manualmente.

5. Código R

1

```
#-----  
# Tarea III  
#-----  
  
# Definimos algunas variables de interés  
n <- 29 # Número de niños estudiados  
x <- 22 # Número de niños que terminaron la Educación Secundaria  
  
# Parámetros alfa y beta de la distribución beta a priori: Be(1,1)  
alpha <- 1  
beta <- 1  
  
# Parámetros alfa y beta de la distribución beta a posteriori: Be(23,8)  
  
alpha_post <- x + alpha  
beta_post <- n - x + beta  
  
# Generamos una muestra de tamaño 1000 para la distribución empírica  
samples <- rbeta(1000, alpha_post, beta_post)  
  
# Representamos la densidad empírica, la densidad a priori y la densidad a  
# posteriori  
plot(density(samples), xaxt='n', ann=FALSE, yaxt='n', xlab=, ylab=,  
     main = ,xlim = c(0,1), ylim = c(0,6), col=2)  
par(new = TRUE)  
curve(dbeta(x,alpha, beta), ann=FALSE, yaxt='n', xlab=, ylab=,  
      xlim = c(0,1), ylim = c(0,6), col=4)  
par(new = TRUE)  
curve(dbeta(x,alpha_post, beta_post), xlim = c(0,1), ylim = c(0,6), col=6,  
      xlab = expression(pi), ylab = "Densidad",  
      main="Función de densidad de la distribución empírica, a priori  
        y a posteriori")  
par(new = FALSE)  
legend(x= "topleft", y=0.2,legend = c("Empírica", "A priori", "A posteriori"),  
      col = c(2,4,6), lty=1, cex=0.8, box.lty=0)  
  
# Calculamos la media y desviación típica a posteriori  
post.mean <- (alpha + x)/(alpha + beta + n)  
post.var <- ((x + alpha)*(n - x + beta))  
            /((alpha + beta + n)^2*(alpha + beta + n + 1))
```

```

post.sd <- sqrt(post.var)

# Calculamos un intervalo de probabilidad al 90% para p

percentage <- 0.9
alpha_level <- (1- percentage)/2

lcb <- qbeta(alpha_level, alpha_post, beta_post)
ucb <- qbeta(1 - alpha_level, alpha_post, beta_post)

# Representamos el intervalo de probabilidad junto con la distribución
# a posteriori
curve(dbeta(x,alpha_post, beta_post), xlim = c(0,1), ylim = c(0,6),
      col=6, xlab = expression(pi), ylab = "Densidad",
      main="Intervalo de probabilidad al 90 % sobre la distribución
a posteriori")

cords2.x <- c(lcb, seq(lcb, ucb, 0.01), ucb)
cords2.y <- c(0, dbeta(seq(lcb, ucb, 0.01), alpha_post, beta_post), 0)
polygon(cords2.x,cords2.y,col='skyblue', border=NA)
abline(h=0)

# Hacemos el contraste de hipótesis
# H_0 : p <= 0.4
# H_1 : p > 0.4

x0 <- pbeta(0.4, alpha_post, beta_post)
x1 <- 1 - pbeta(0.4, alpha_post, beta_post)

# Calculamos la probabilidad predictiva de que al menos 9 de los 10 niños
# terminen la Educación Secundaria

k <- 9 # 9 niños acaban la Educación Secundaria
m <-10 # de 10
sum <- 0
x <- seq(0,m)
y <- c(rep(0, m+1))
for(k in 0:m){
  probability <- choose(m, k) * (gamma(alpha + beta + n)/
    (gamma(alpha + x)*gamma(beta + n - x)))*(gamma(alpha + x + k)*
    gamma(beta + n - x + m - k)/gamma(alpha + beta + m + n))
  y[k] = probability
}

```

```

# Representamos la función de densidad de la distribución predictiva
plot(x,y, main = "Función de densidad de la distribución predictiva",
     xlab = "Número de éxitos de 10 posibles", ylab = "Densidad", col = 2,
     type = "o")

#-----
# Tarea IV
#-----

# Importamos la librería LearnBayes
library("LearnBayes")

# Definimos una serie de variables que nos serán de utilidad
mu <- seq(20, 70,10)
g_mu <- c(0.1, 0.15, 0.25, 0.25, 0.15, 0.1)
y <- c(38.6, 42.4, 57.5, 40.5, 51.7, 67.1, 33.4, 60.9, 64.1, 40.1, 40.7, 6.4)

sigma <- 10

# Representamos la función de densidad de la distribución a priori
plot(mu, g_mu, col = 2, main = "Función de densidad de la distribución a priori",
     xlab = "Total anual de precipitaciones en forma de nieve (en cm)",
     ylab = "Densidad", type = "h")

# Calculamos la verosimilitud para cada uno de los valores de mu

# Definimos una función para la verosimilitud
likelihood <- function(mu){
  n <- length(y)
  temp <- 1/(sqrt(2*pi)*sigma)^n
  temp2 <- prod(exp((-1/2) * ((y-mu)/sigma)^2))
  return(temp*temp2)
}

likelihood2 <- function(mu, y, sigma){
  n <- length(y)
  temp <- 1/(sqrt(2*pi)*sigma)^n
  temp2 <- exp(-1/(2*sigma^2)*(n*mu^2 + sum(y^2) - 2*mu*sum(y)))
  return(temp*temp2)
}

```

```

likelihood <- likelihood2(mu,y,sigma)

# Calculamos la distribución a posteriori

post <- likelihood*g_mu # Sin normalizar

post <- post/sum(post) # Normalizada

# Representamos la distribución a posteriori
plot(mu, post, main="Distribución a posteriori", xlab=expression(mu),
      ylab="Densidad", type = "p", col = 2)
par(new = TRUE)
plot(mu, post, main="Distribución a posteriori", xlab=expression(mu),
      ylab="Densidad", type = "h", col = 2)

# Calculamos el intervalo de probabilidad al 80%

percentage <- 0.8
dist <- cbind(mu, post)
discint(dist,percentage)

# Ahora tenemos como distribución a priori, una distribución N(60, 5).

mu0 <- 60
sigma0 <- 5
n <- length(y)

# La distribución a posteriori tiene como parámetros

mu_post <- (mu0/sigma0^2 + n*mean(y)/sigma^2)/(1/sigma0^2 + n/sigma^2)
var_post <- (1/sigma0^2 + n/sigma^2)^-1
sd_post <- sqrt(var_post)

# Representamos las distribuciones a priori y posteriori
curve(xlim=c(30,90),dnorm(x,mu0, sigma0), ann=FALSE, yaxt='n', xlab=,
      ylab=, ylim = c(0,0.18), col=4)
par(new = TRUE)
curve(xlim=c(30,90),dnorm(x,mu_post, sd_post), ylim = c(0,0.18), col=6,
      xlab = expression(mu), ylab = "Densidad", main="Función de densidad de
      la distribución a priori y a posteriori")
par(new = FALSE)
legend(x= "topleft", y=0.2,legend = c("A priori", "A posteriori"),
      col = c(4,6), lty=1, cex=0.8, box.lty=0)

```



```

# Calculamos un intervalo de probabilidad al 80% para mu

percentage <- 0.8
alpha_level <- (1 - percentage)/2

lcb <- qnorm(alpha_level, mu_post, sd_post)
ucb <- qnorm(1 - alpha_level, mu_post, sd_post)

# Representamos el intervalo de probabilidad junto con la
# distribución a posteriori
curve(dnorm(x,mu_post, sd_post), xlim = c(30,90), ylim = c(0,0.18),
      col=6, xlab = expression(mu), ylab = "Densidad",
      main="Intervalo de probabilidad al 80 % sobre la distribución a posteriori")

cords2.x <- c(lcb, seq(lcb, ucb, 0.01), ucb)
cords2.y <- c(0, dnorm(seq(lcb, ucb, 0.01), mu_post, sd_post), 0)
polygon(cords2.x,cords2.y,col='skyblue', border=NA)
abline(h=0)

```