ANÁLISIS BAYESIANO DE DATOS EJERCICIOS TEMA 1

Nombre y apellidos: Alberto Juliana Quirós Fecha de entrega: 6 de marzo

EJERCICIO 1. Supongamos que 80 de cada 10.000 niños menores de 6 años padecen un trastorno del espectro autista y que una nueva prueba diagnóstica basada en neuroimagen permite la identificación del trastorno en un 87 por ciento de los casos. Como contrapartida, cuando el niño no padece el trastorno, la prueba da un resultado positivo en un 4 por ciento de los casos.

1. Utilice la ley de la probabilidad total para calcular la probabilidad de que un niño menor de 6 años obtenga un resultado positivo en la prueba diagnóstica.
2. Utilice el teorema de Bayes para calcular la probabilidad de que un niño menor de 6 años que haya obtenido un resultado positivo en la prueba diagnóstica padezca de hecho un trastorno del espectro autista.

EJERCICIO 2. Hemos medido el rendimiento en una tarea de memoria en una muestra aleatoria de 10 pacientes con depresión obteniendo los siguientes resultados:

$$X = 5, 4, 3, 6, 9, 10, 2, 6, 8, 7.$$

Asumiendo que la variable se distribuye de forma normal y que $\sigma = 3$, queremos estimar μ haciendo uso de dos distribuciones a priori basadas en estudios previos:

- $\mu \sim$ *normal* (4, 1).
- $\mu \sim$ *normal* (4, 8).

Para cada distribución a priori:

1. Obtenga el estimador EAP, la desviación típica a posteriori y el intervalo de probabilidad posterior (95%) de μ.
2. Represente gráficamente con R las distribuciones a priori y a posteriori de μ.
3. Finalmente, compare la influencia en los resultados de las dos distribuciones a priori.

EJERCICIO 3. Hemos registrado la variable número de crisis de pánico durante una semana en una muestra aleatoria de diez personas diagnosticadas de trastorno por pánico obteniendo los siguientes resultados:

$$X = 0, 2, 6, 1, 6, 3, 2, 7, 2, 1.$$

Asumiendo que la variable se distribuye de acuerdo con una distribución de Poisson, queremos estimar λ haciendo uso de dos distribuciones a priori:

- $\lambda \sim$ *gamma* (5, 1).
- $\lambda \sim$ *gamma* (20, 2).

Para cada distribución a priori:

1. Represente gráficamente con R las distribuciones a priori y a posteriori de λ.
2. Obtenga el valor esperado, la varianza y el intervalo de probabilidad (95%) de λ a priori y a posteriori.
3. Finalmente, compare la influencia en los resultados de las dos distribuciones a priori.

Resolución

1.1.

$$P(A) = \frac{80}{1 \cdot 10^4} = 8 \cdot 10^{-3}$$
$$P(P \mid A) = 0,87$$
$$P(P \mid \bar{A}) = 0,04$$

$$P(P \cap A) = P(A) \cdot P(P \mid A) =$$
$$8 \cdot 10^{-3} \cdot 0,87 = 6,96 \cdot 10^{-3}$$

$$P(P \cap \bar{A}) = P(\bar{A}) \cdot P(P \mid \bar{A}) =$$
$$(1 - 8 \cdot 10^{-3}) \cdot 0,04 =$$
$$0,04$$

$$P(P) = P(P \cap A) + P(P \cap \bar{A}) =$$
$$6,96 \cdot 10^{-3} + 0,04 =$$
$$0,05$$

1.2.

$$P(A \mid P) = \frac{P(P \cap A)}{P(P)} =$$
$$\frac{6,96 \cdot 10^{-3}}{0,05} = 0,14$$

2.1.

Distribución gamma (5, 1):

$$E(\mu \mid \boldsymbol{x}) = \mu_{a-\text{ posteriori}} = \frac{\frac{\theta}{\tau^2} + \frac{n\bar{X}}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} =$$
$$\frac{\frac{4}{1^2} + \frac{10 \cdot 6}{3^2}}{\frac{1}{1^2} + \frac{10}{3^2}} = 5,05$$

$$\text{Desv}(\mu \mid \boldsymbol{x}) = \sigma_{a-\text{ posteriori}} = \left(\frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\right)^{\frac{1}{2}} =$$

$$\left(\frac{1}{\frac{1}{1^2}+\frac{10}{3^2}}\right)^{1/2}=0,69$$

Con $\alpha = 0,05$, el intervalo de confianza es:

```
> qnorm(0.025, 5.05, 0.69)
[1] 3.697625
> qnorm(0.975, 5.05, 0.69)
[1] 6.402375
```

$$IC_{\mu-\text{post}} = (3,69, \quad 6,4)$$

Distribución gamma (20, 2):

$E(\mu \mid \boldsymbol{x}) = \mu_{a-\text{ posteriori}} = \dfrac{\frac{\theta}{\tau^2}+\frac{n\bar{X}}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} =$

$$\frac{\frac{4}{8^2}+\frac{10\cdot 6}{3^2}}{\frac{1}{8^2}+\frac{10}{3^2}}=5,97$$

$\mathrm{Desv}(\mu \mid \boldsymbol{x}) = \sigma_{a-\text{ posteriori}} = (\dfrac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}})^{\frac{1}{2}} =$

$$\left(\frac{1}{\frac{1}{8^2}+\frac{10}{3^2}}\right)^{1/2}=0,94$$

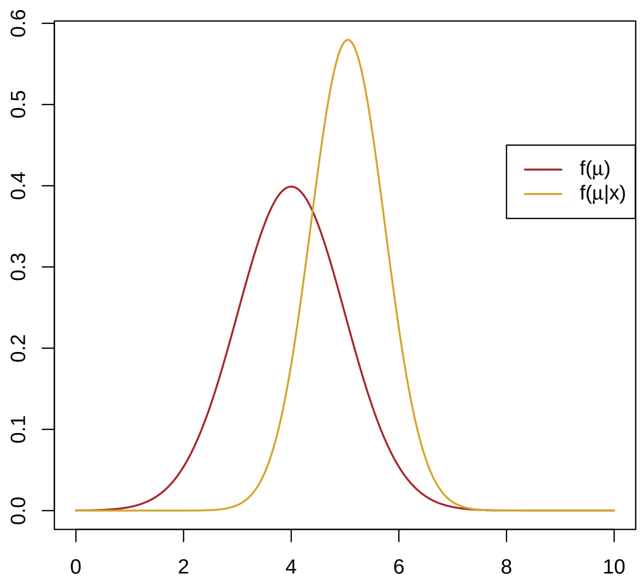
Con $\alpha = 0,05$, el intervalo de confianza es:

```
> qnorm(0.025, 5.97, 0.94)
[1] 4.127634
> qnorm(0.975, 5.97, 0.94)
[1] 7.812366
```

$$IC_{\mu-\text{post}} = (4,12, \quad 7,81)$$

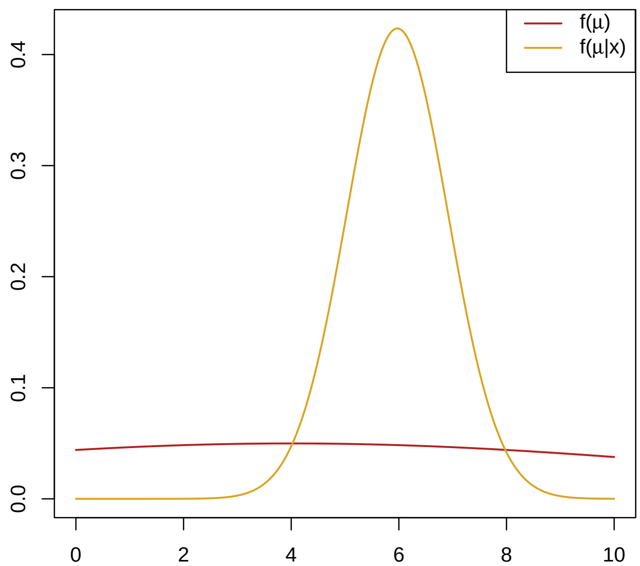
2.2.

```
> ## Distribución gamma (5, 1):
> # Datos del ejemplo
> x <- c(5, 4, 3, 6, 9, 10, 2, 6, 8, 7)
> n <- length(x)
> theta <- 4
> tau <- 1
> sigma <- 3
> # Estimador máximo-verosímil e intervalo de confianza
> media <- mean(x)
> se <- sigma/sqrt(n)
> li_ml <- qnorm(0.025, media, se)
> ls_ml <- qnorm(0.975, media, se)
> # Estimador EAP e intervalo de probabilidad posterior
> sigma2 <- sigma^2
> tau2 <- tau^2
> sigma2_post <- 1/((1/tau2)+(n/sigma2))
> sigma_post <- sqrt(sigma2_post)
> mu_post <- ((theta/tau^2) + (n*media/sigma2)) * sigma2_post
> li_post <- qnorm(0.025, mu_post, sigma_post)
> ls_post <- qnorm(0.975, mu_post, sigma_post)
> # Representación gráfica de f(mu) y f(mu|x)
> mu <- seq(0, 10, by=0.001)
> f1 <- dnorm(mu, theta, tau)
> f2 <- dnorm(mu, mu_post, sigma_post)
> plot(mu, f1, col="firebrick", lwd=1.5, type="l", ylim=c(0,max(f1,f2)), xlab="", ylab="")
> lines(mu, f2, col="goldenrod", lwd=1.5)
> legend(8,0.45,
+ c(expression(paste("f(",mu,")")), expression(paste("f(", mu, "|x)"))),
+ col=c("firebrick","goldenrod"), lwd=1.5)
```




```
> ## Distribución gamma (20, 2):
> # Datos del ejemplo
> x <- c(5, 4, 3, 6, 9, 10, 2, 6, 8, 7)
> n <- length(x)
> theta <- 4
> tau <- 8
> sigma <- 3
> # Estimador máximo-verosímil e intervalo de confianza
> media <- mean(x)
> se <- sigma/sqrt(n)
> li_ml <- qnorm(0.025, media, se)
> ls_ml <- qnorm(0.975, media, se)
> # Estimador EAP e intervalo de probabilidad posterior
> sigma2 <- sigma^2
> tau2 <- tau^2
> sigma2_post <- 1/((1/tau2)+(n/sigma2))
> sigma_post <- sqrt(sigma2_post)
> mu_post <- ((theta/tau^2) + (n*media/sigma2)) * sigma2_post
> li_post <- qnorm(0.025, mu_post, sigma_post)
> ls_post <- qnorm(0.975, mu_post, sigma_post)
> # Representación gráfica de f(mu) y f(mu|x)
> mu <- seq(0, 10, by=0.001)
```

```
> f1 <- dnorm(mu, theta, tau)
> f2 <- dnorm(mu, mu_post, sigma_post)
> plot(mu, f1, col="firebrick", lwd=1.5, type="l", ylim=c(0,max(f1,f2)), xlab="", ylab="")
> lines(mu, f2, col="goldenrod", lwd=1.5)
> legend(8,0.45,
+ c(expression(paste("f(",mu,")")), expression(paste("f(", mu, "|x)"))),
+ col=c("firebrick","goldenrod"), lwd=1.5)
```

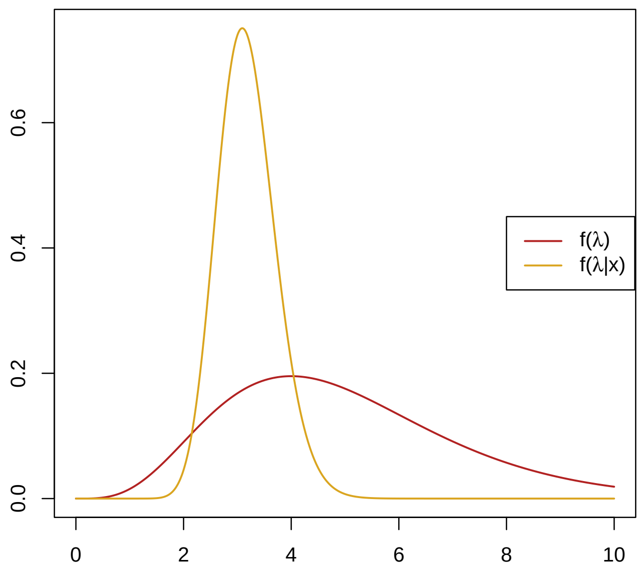


2.3.
Podemos observar cómo una mayor dispersión en la distribución a priori conlleva a mayores dispersiones a posteriori.
Por otra parte, las densidades de probabilidad a posteriori son proporcionales a las a priori, por lo que cuanta más certidumbre previa del valor del estadístico, mayor certidumbre tengo del nuevo valor que adquiere una vez dispongo de más información.

3.1.

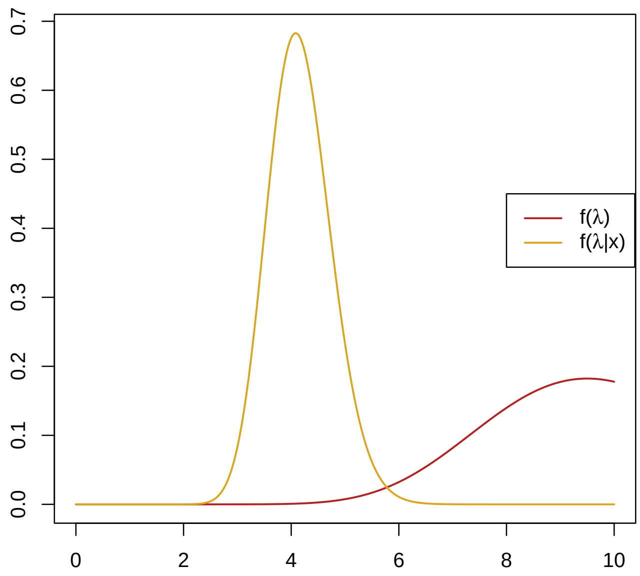
```
> x <- c(0, 2, 6, 1, 6, 3, 2, 7, 2, 1)
> n <- length(x)
> medx <- mean(x)
> ## Distribución gamma (5, 1):
```

```
> alpha <- 5
> beta <- 1
> lambda <- seq(0,10,by=0.01)
> ## funciones a priori y posteriori
>
> f_priori <- function(lambda){
+         ((beta^alpha)/gamma(alpha))*lambda^(alpha-1)*exp(-beta*lambda)
+ }
> alpha_post <- sum(x)+alpha
> beta_post <- n +beta
> f_posteriori <- dgamma(lambda, alpha_post, beta_post)
> plot(lambda, f_priori(lambda), col="firebrick", lwd=1.5, type="l",ylim=c(0,max(f_priori(la
> lines(lambda, f_posteriori, col="goldenrod", lwd=1.5)
> legend(8,0.45,
+ c(expression(paste("f(",lambda,")")), expression(paste("f(", lambda, "|x)"))),
+ col=c("firebrick","goldenrod"), lwd=1.5)
>
```




```
> ## Distribución gamma (20, 2):
> alpha <- 20
> beta <- 2
> lambda <- seq(0,10,by=0.01)
> f_priori <- function(lambda){
```

```
+          ((beta^alpha)/gamma(alpha))*(lambda^(alpha-1))*exp(-beta*lambda)
+ }
> alpha_post <- sum(x)+alpha
> beta_post <- n +beta
> f_posteriori <- dgamma(lambda, alpha_post, beta_post)
> plot(lambda, f_priori(lambda), col="firebrick", lwd=1.5, type="l",ylim=c(0,max(f_priori(la
> lines(lambda, f_posteriori, col="goldenrod", lwd=1.5)
> legend(8,0.45,
+ c(expression(paste("f(",lambda,")")), expression(paste("f(", lambda, "|x)"))),
+ col=c("firebrick","goldenrod"), lwd=1.5)
```



3.2

Distribución gamma (5, 1):

$$E(\lambda) = \frac{\alpha}{\beta} = \frac{5}{1} = 5$$

$$E(\lambda \mid \boldsymbol{x}) = \frac{n\bar{X} + \alpha}{n + \beta} = \frac{10 \cdot 3 + 5}{10 + 1} = 3,18$$

$$\mathrm{Var}(\lambda) = \frac{\alpha}{\beta^2} = 5$$

$$\mathrm{Var}(\lambda \mid \boldsymbol{x}) = \frac{n\bar{X} + \alpha}{(n+\beta)^2} =$$
$$\frac{10 \cdot 3 + 5}{(10+1)^2} = 0,289$$

$$\hat{\lambda} = \bar{X} = 3$$
$$\mathrm{Var}(\hat{\lambda}) = \frac{\lambda}{n} = \frac{3}{10} = 0,3$$
$$L_{\text{spriori}} = 3 + 1,96\sqrt{\frac{3}{10}} = 4,07$$
$$L_{i\text{priori}} = 3 - 1,96\sqrt{\frac{3}{10}} = 1,92$$
$$IC_{\lambda-\text{priori}} = (4,07, \quad 1,92)$$

```
> alpha <- 5
> beta <- 1
> alpha_post <- sum(x)+alpha
> beta_post <- n +beta
> qgamma(0.025,alpha_post,beta_post)

[1] 2.216253

> qgamma(0.975,alpha_post,beta_post)

[1] 4.319236
```

$$IC_{\lambda-\text{post}} = (2,22, \quad 4,32)$$

Distribución gamma (20, 2):

$$E(\lambda) = \frac{\alpha}{\beta} = \frac{20}{2} = 10$$

$$E(\lambda \mid \boldsymbol{x}) = \frac{n\bar{X} + \alpha}{n+\beta} = \frac{10 \cdot 3 + 20}{10+2} = 4,17$$

$$\mathrm{Var}(\lambda) = \frac{\alpha}{\beta^2} = 5$$

$$\mathrm{Var}(\lambda \mid \boldsymbol{x}) = \frac{n\bar{X} + \alpha}{(n+\beta)^2} =$$
$$\frac{10 \cdot 3 + 20}{(10+2)^2} = 0,35$$

$$\hat{\lambda} = \bar{X} = 3$$
$$\mathrm{Var}(\hat{\lambda}) = \frac{\lambda}{n} = \frac{3}{10} = 0,3$$
$$L_{\text{spriori}} = 3 + 1,96\sqrt{\frac{3}{10}} = 4,07$$
$$L_{i\text{priori}} = 3 - 1,96\sqrt{\frac{3}{10}} = 1,92$$
$$IC_{\lambda-\text{priori}} = (4,07, \quad 1,92)$$

```
> alpha <- 20
> beta <- 2
> alpha_post <- sum(x)+alpha
> beta_post <- n +beta
> qgamma(0.025,alpha_post,beta_post)

[1] 3.09258

> qgamma(0.975,alpha_post,beta_post)

[1] 5.398383
```

$$IC_{\lambda-\text{post}} = (3,09, \quad 5,4)$$

3.3.
Podemos observar cómo, a igualdad de dispersión entre las a priori, el desplazamiento de las curvas posteriori será proporcional a la distancia respecto al valor esperado de la muestra.