EstComp-Tarea11

Bruno Gonzalez

7/11/2019

11-Familias conjugadas

1. Modelo Beta-Binomial

Una compañía farmacéutica afirma que su nueva medicina incrementa la probabilidad de concebir un niño (sexo masculino), pero aún no publican estudios. Supón que conduces un experimento en el cual 50 parejas se seleccionan de manera aleatoria de la población, toman la medicina y conciben.

a) Quieres estimar la probabilidad de concebir un niño para parejas que toman la medicina. ¿Cuál es una inicial apropiada? No tiene que estar centrada en 0.5 pues esta corresponde a personas que no toman la medicina, y la inicial debe reflejar tu incertidumbre sobre el efecto de la droga.

Dado que tenemos el número de la muestra $n = 50$, podemos partir definiendo a m, la proporción de éxitos (concebir un niño) y establecerla como $m = 0.7$. De esta manera:

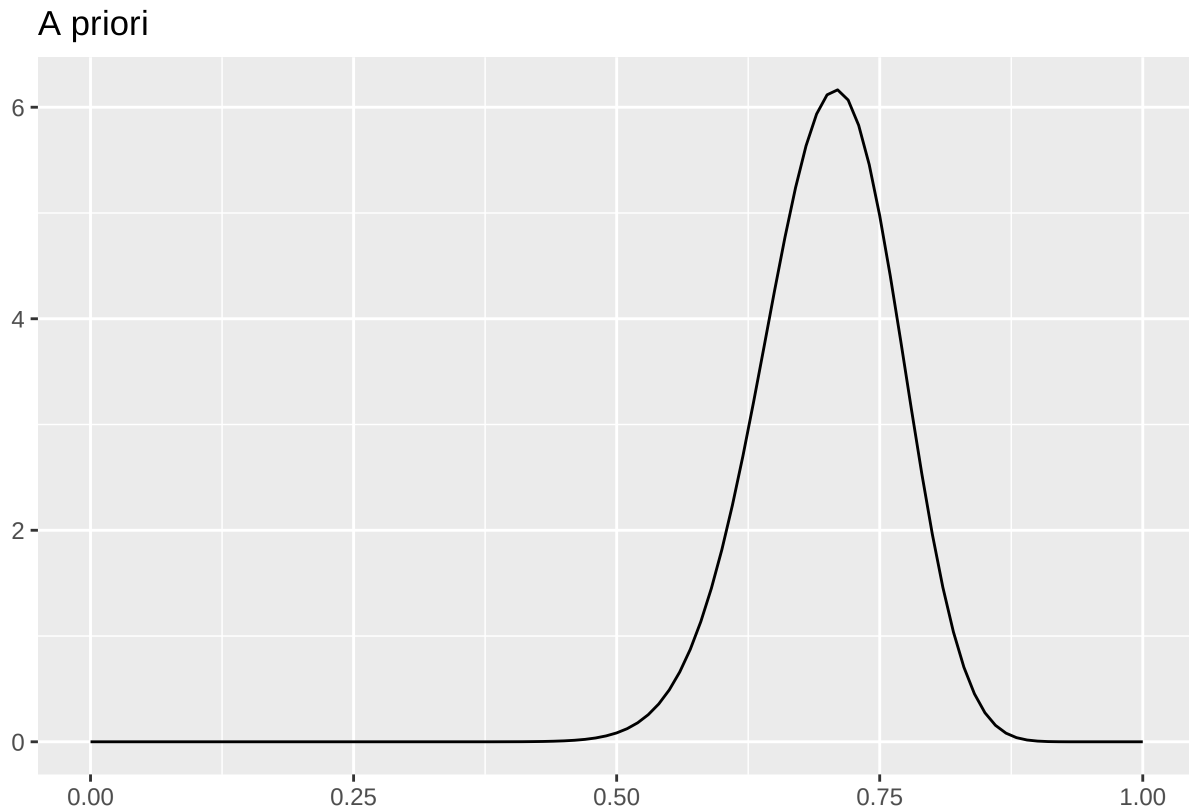
$$a = mn = 35$$
$$b = (1 - m)n = 15$$

Así, podemos definir la distribución a priori de la siguiente manera:

```
N <-  50
m <-  0.7
a <-  m*N
b <-  (1-m)*N

priori <- ggplot(tibble(x = c(0,1)), aes(x)) +
  stat_function(fun = dbeta, args = list(a,b)) +
  labs(title = "A priori",
    x = "",
    y = "")

priori
```



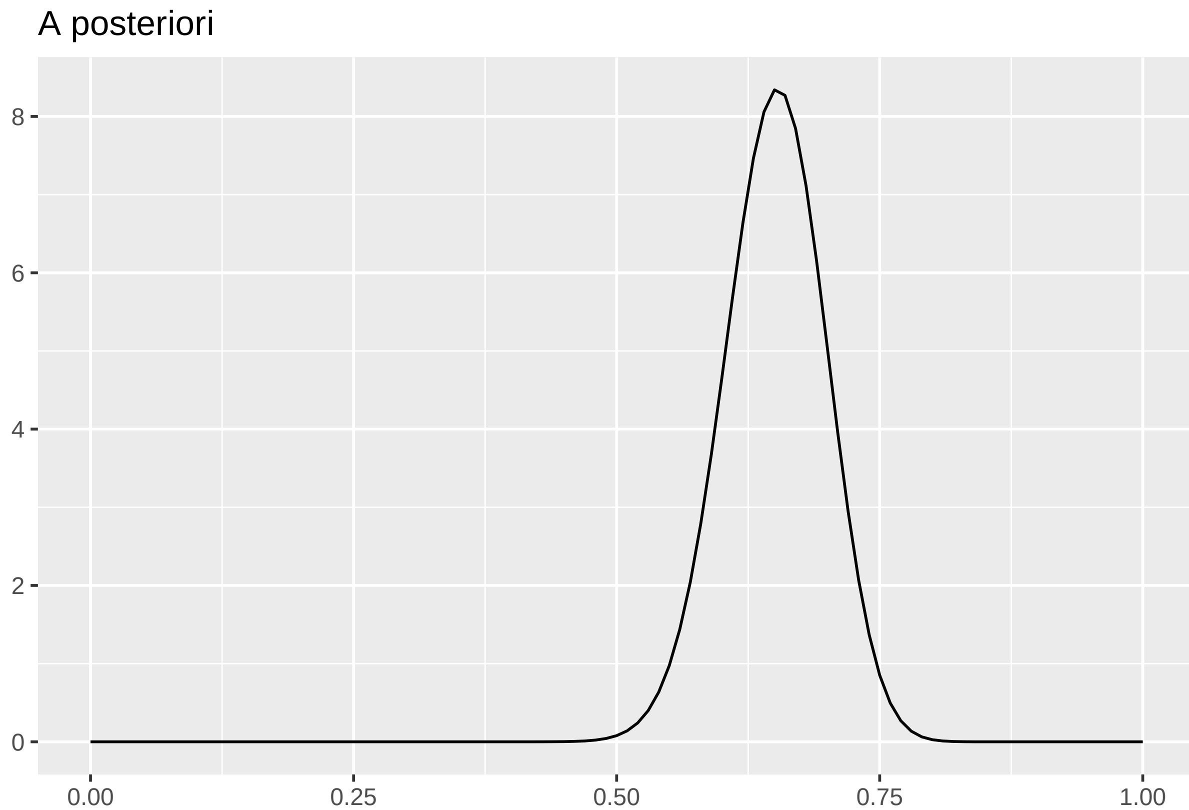
b) Usando tu inicial de a) grafica la posterior y decide si es creíble que las parejas que toman la medicina tienen una probabilidad de 0.5 de concebir un niño.

Sea z el número de éxitos de la prueba, en esta caso la gráfica posterior es de la siguiente manera:

```
# Supongamos z
z = 30

post <- ggplot(tibble(x = c(0,1)), aes(x)) +
  stat_function(fun = dbeta, args = list(z+a,N-z+b)) +
  labs(title = "A posteriori",
    x = "",
    y = "")

post
```



Dado el supuesto de $z = 60$, sí es probable que la medicina surja efecto ya la probabilidad de éxito es mayormente mayor a *50%*.

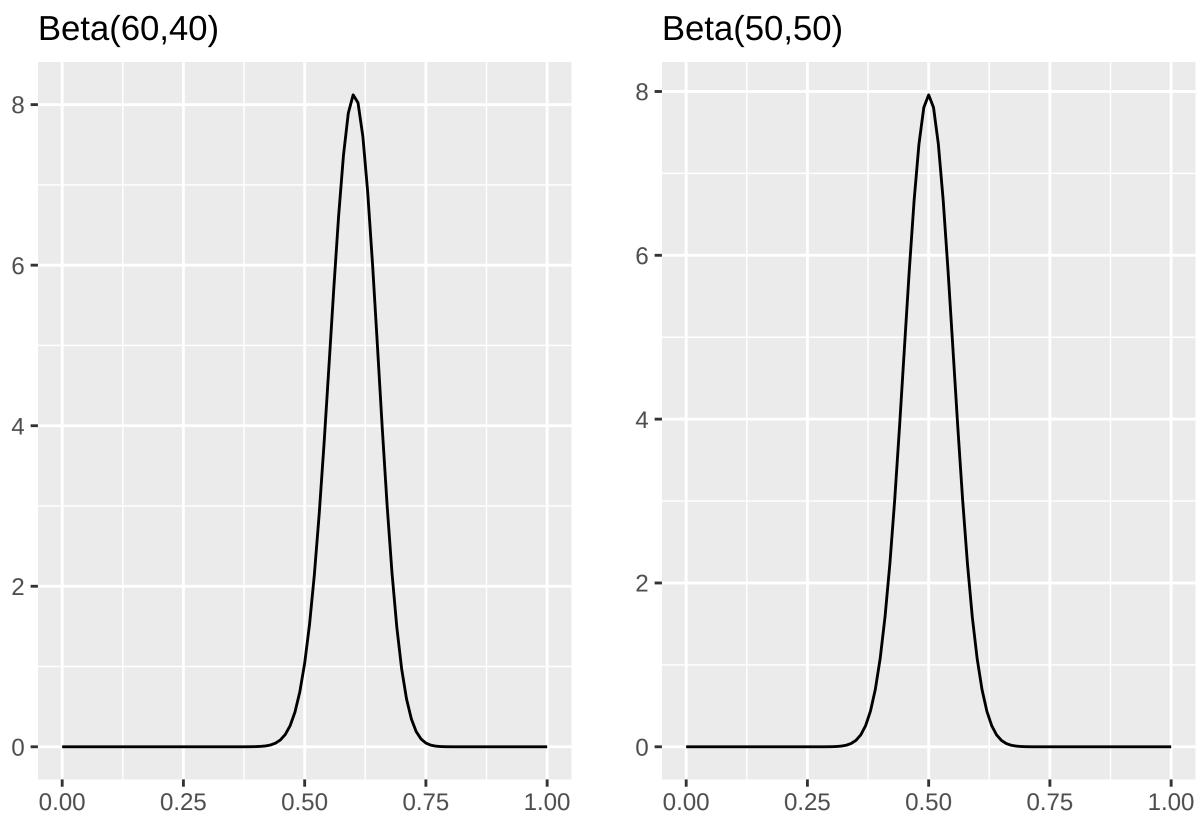
c) Supón que la farmacéutica asevera que la probabilidad de concebir un niño cuando se toma la medicina es cercana al 60% con alta certeza. Representa esta postura con una distribución inicial $Beta(60, 40)$. Comparala con la inicial de un escéptico que afirma que la medicina no hace diferencia, representa esta creencia con una inicial $Beta(50, 50)$. ¿Cómo se compara la probabilidad posterior de concebir un niño (usando las distintas iniciales)?

En esta caso, las distribuciones a priori se distribuyen de la siguiente manera:

```
priori_comp1 <- ggplot(tibble(x = c(0,1)), aes(x)) +
  stat_function(fun = dbeta, args = list(60,40)) +
  labs(title = "Beta(60,40)",
    x = "",
    y = "")

priori_comp2 <- ggplot(tibble(x = c(0,1)), aes(x)) +
  stat_function(fun = dbeta, args = list(50,50)) +
  labs(title = "Beta(50,50)",
    x = "",
    y = "")

grid.arrange(priori_comp1, priori_comp2, nrow = 1)
```

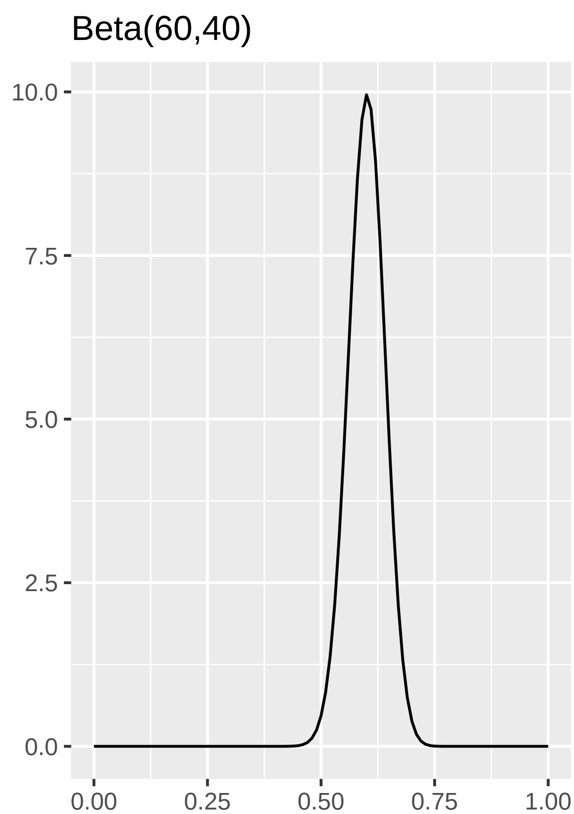


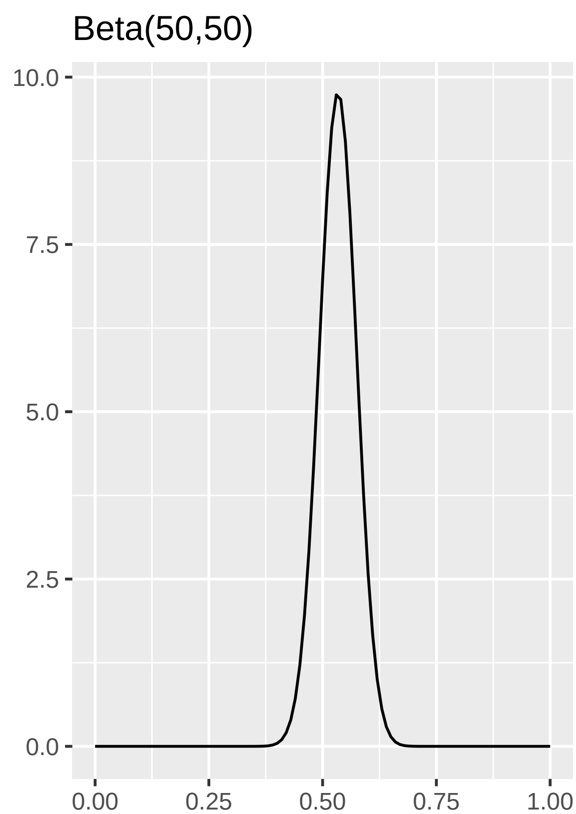
Usando el mismo número de éxitos del inciso b), tenemos las siguientes distribuciones a posteriori.

```
post_comp1 <- ggplot(tibble(x = c(0,1)), aes(x)) +
  stat_function(fun = dbeta, args = list(z + 60,N - z +40)) +
  labs(title = "Beta(60,40)",
    x = "",
    y = "")

post_comp2 <- ggplot(tibble(x = c(0,1)), aes(x)) +
  stat_function(fun = dbeta, args = list(z + 50,N - z +50)) +
  labs(title = "Beta(50,50)",
    x = "",
    y = "")

grid.arrange(post_comp1, post_comp2, nrow = 1)
```





La distribución posterior para un escéptico sigue representando su escepticismo, lo ilustra que las distribuicones a priori impactan las posteriori independientemente de la información disponible.

2. Otra familia conjugada

Supongamos que nos interesa analizar el IQ de una muestra de estudiantes del ITAM y suponemos que el IQ de un estudiante tiene una distribución normal $x \sim N(\theta, \sigma^2)$ **con σ^2 conocida.** Considera que observamos el IQ de un estudiante x. La verosimilitud del modelo es:

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right)$$

Realizaremos un análisis bayesiano por lo que hace falta establer una distribución inicial, elegimos $p(\theta)$ que se distribuya $N(\mu, \tau^2)$ donde elegimos los parámetros μ, τ que mejor describan nuestras creencias iniciales, por ejemplo si tengo mucha certeza de que el IQ promedio se ubica en 150, elegiría $\mu = 150$ y una desviación estándar chica por ejemplo $\tau = 5$. Entonces la distribución inicial es:

$$p(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} exp\left(-\frac{1}{2\tau^2}(\theta-\mu)^2\right)$$

Calcula la distribución posterior $p(\theta|x) \propto p(x|\theta)p(\theta)$, usando la inicial y verosimilitud que definimos arriba. Una vez que realices la multiplicación debes identificar el núcleo de una distribución Normal, ¿cuáles son sus parámetros (media y varianza)?

Dado que sabemos que la suma de normales es normal, podemos obtener la distribución fijándonos solo en el exponente.

$$
\begin{aligned}
p(\theta|x) &\propto p(x|\theta)p(\theta) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right) \frac{1}{\sqrt{2\pi\tau^2}} exp\left(-\frac{1}{2\tau^2}(\theta-\mu)^2\right) \\
&\propto exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right) \\
&= exp\left(\frac{-\tau^2x^2 + 2\tau^2x\theta - \tau^2\theta^2 - \sigma^2\theta^2 + 2\sigma^2\theta\mu - \sigma^2\mu^2}{2\sigma^2\tau^2}\right) \\
&= exp\left(-\theta^2\left(\frac{\tau^2+\sigma^2}{2\sigma^2\tau^2}\right) + \theta\left(\frac{2\tau^2x + 2\sigma^2\mu}{2\sigma^2\tau^2}\right) - \left(\frac{\tau^2x^2+\sigma^2\mu^2}{2\sigma^2\tau^2}\right)\right) \\
&= exp\left(-\frac{\theta^2}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right) + \frac{2\theta}{2}\left(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}\right) - \frac{1}{2}\left(\frac{x^2}{\sigma^2}+\frac{\mu^2}{\tau^2}\right)\right) \\
&= exp\left(-\frac{1}{2\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}}\left(\theta^2 - 2\theta\left(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1} + \left(\frac{x^2}{\sigma^2}+\frac{\mu^2}{\tau^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right)\right) \\
&= exp\left(-\frac{1}{2\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}}\left(\theta^2 + \left(\frac{x}{\sigma}+\frac{\mu}{\tau}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right)^2\right)
\end{aligned}
$$

De esta manera tenemos que la distribución posterior se distribuye normal con parámetros:

$$
p(\theta|x) \sim N\left(\left(\frac{x}{\sigma}+\frac{\mu}{\tau}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}, \left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right)
$$