3.11

Santiago Robatto y Sofia Terra

Exercise 3.11 (Regular bike ridership) A university wants to know what proportion of students are regular bike riders, π, so that they can install an appropriate number of bike racks. Since the university is in sunny Southern California, staff think that π has a mean of 1 in 4 students, and a mode of 5/22.

a. Specify and plot a Beta model that reflects the staff's prior ideas about π.
b. Among 50 surveyed students, 15 are regular bike riders. What is the posterior model for π?
c. What is the mean, mode, and standard deviation of the posterior model?
d. Does the posterior model more closely reflect the prior information or the data? Explain your reasoning.

Nuestra realidad: Una universidad quiere saber la proporción de alumnos que andan regularmente en bicicleta. Este será nuestro π. Quieren saber esto para poder instalar bicicleteros acordes a este número. La universidad cree que π tiene una media de $1/4$ y un modo de $5/22$.

a)Nos piden especificar el modelo Beta que refleja la priori de la universidad con respecto a π.

Tenemos que:

$$\pi \sim \text{Beta}(\alpha, \beta)$$

Donde α y β son los hiperparámetros.

Luego sabemos también que la esperanza y el modo de Beta se calculan de la siguiente manera:

$$\text{E}(\pi) = \frac{\alpha}{\alpha + \beta}$$
$$\text{Mode}(\pi) = \frac{\alpha - 1}{\alpha + \beta - 2} \quad \text{cuando } \alpha, \beta > 1.$$

Por dato de letra sabemos que $E(\pi) = 1/4$ y que $\text{Mode}(\pi) = 5/22$. Tenemos que resolver un pequeño sistema de ecuaciones:

$$\frac{1}{4} = \frac{\alpha}{\alpha + \beta}$$
$$\frac{5}{22} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Resolvemos y obtenemos:

Por un lado:

$$\frac{1}{4} = \frac{\alpha}{\alpha + \beta}$$

$$\implies (0.25)(\alpha + \beta) = \alpha$$

$$\implies 0.25\beta = 0.75\alpha$$

$$\implies \frac{0.25\beta}{0.75} = \alpha = \frac{1}{3}\beta$$

Por otro lado:

$$\frac{5}{22} = \frac{\frac{1}{3}\beta - 1}{\frac{1}{3}\beta + \beta - 2}$$

$$\implies \frac{5}{22} = \frac{\frac{\beta - 3}{3}}{\frac{4}{3}\beta - 2} = \frac{\frac{\beta - 3}{3}}{\frac{4\beta - 6}{3}} = \frac{\beta - 3}{4\beta - 6}$$

$$5(4\beta - 6) = 22(\beta - 3)$$

$$\implies -2\beta = -36$$

$$\implies \beta = 18$$

$$\implies \alpha = 6$$

Concluimos que nuestro modelo a priori sigue una distribución

$$\pi \sim \mathrm{Beta}(6, 18)$$

b) Ahora nos dicen que entre los 50 alumnos encuestados, 15 andan en bicicleta regularmente. Se nos pide encontrar el modelo a posteriori para π.

Para encontrar este partimos del modelo Beta-Binomial:

$$\begin{aligned} Y \mid \pi &\sim \text{Bin}(n, \pi) \\ \pi &\sim \text{Beta}(\alpha, \beta) \end{aligned}$$

Habiendo observado Y=y éxitos en n pruebas, el posterior de π puede ser descrito por una modelo Beta que refleja la influencia del prior (α y β) y la información (datos) (n e y). Esto se escribe como:

$$\pi \mid (Y = y) \sim \text{Beta}(\alpha + y, \beta + n - y)$$

En nuestro caso específico tenemos:

$$\pi \mid (Y = y) \sim \text{Beta}(6 + 15, 18 + 50 - 15)$$

$$\implies \pi \mid (Y = y) \sim \text{Beta}(21, 53)$$

Este es nuestro modelo posteriori.

c)Ahora nos interesa calcular la media, el modo y el desvío típico de nuestro posteriori definido anteriorimente.

Las fórmulas de lo que buscamos calcular son las siguientes:

$$\text{E}(\pi \mid Y = y) = \frac{\alpha + y}{\alpha + \beta + n}$$

$$\text{Mode}(\pi \mid Y = y) = \frac{\alpha + y - 1}{\alpha + \beta + n - 2}$$

$$\text{SD}(\pi \mid Y = y) = \sqrt{\frac{(\alpha + y)(\beta + n - y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}}$$

Ahora simplemente sustituimos los valores de α, β, n e y en las fórmulas y obtenemos:

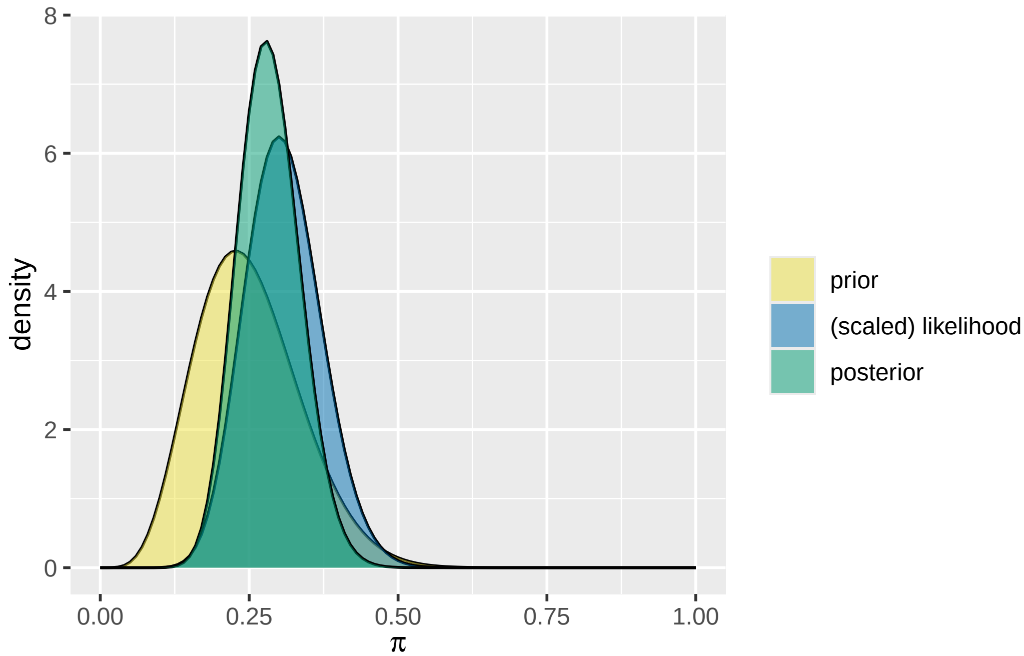
$$\text{E}(\pi \mid Y = y) = 21/74 = 0.283$$

$$\text{Mode}(\pi \mid Y = y) = 0.277$$

$$\mathrm{SD}(\pi \mid Y = y) = 0.052$$

Ahora podemos comparar con los resultados obtenidos al aplicar la funcion plot_beta_binomial del paquete BayesRules!

```
library(bayesrules)
plot_beta_binomial(alpha = 6, beta = 18, y = 15, n = 50)
```



Ahora aplicamos summarize a esto

```
summarize_beta_binomial(alpha = 6, beta = 18, y = 15, n = 50)
```

```
      model alpha beta      mean      mode         var         sd
1     prior     6   18 0.2500000 0.2272727 0.007500000 0.08660254
2 posterior    21   53 0.2837838 0.2777778 0.002710007 0.05205773
```

Comprobamos que ambas formas de cálculo dan lo mismo.

d) Finalmente, nos preguntan si el modelo a posteriori representa de forma cercana o parecida a la información del priori. La respuesta es que sí, ya que existe un buen balance entre la priori, la posteriori y la verosimilitud. El modelo Beta-Binomial parte del priori y se actualiza con la verosimilitud para generar el posteriori, donde se observa que el rango de plausibilidad disminuyó (esto se aprecia con la desviación típica, que disminuye en el modelo posteriori), pero la densidad aumentó. Esto muestra que el modelo priori no estaba muy alejado de la realidad, solo hubo que hacer ajustes menores. Esta conclusión es apreciada en los cálculos de la media para el priori y la posteriori, las cuales no son muy diferentes entre sí, están centradas en valores similares.