

# Ejercicios capitulos 4 y 5

Sofia Terra, Santiago Robatto

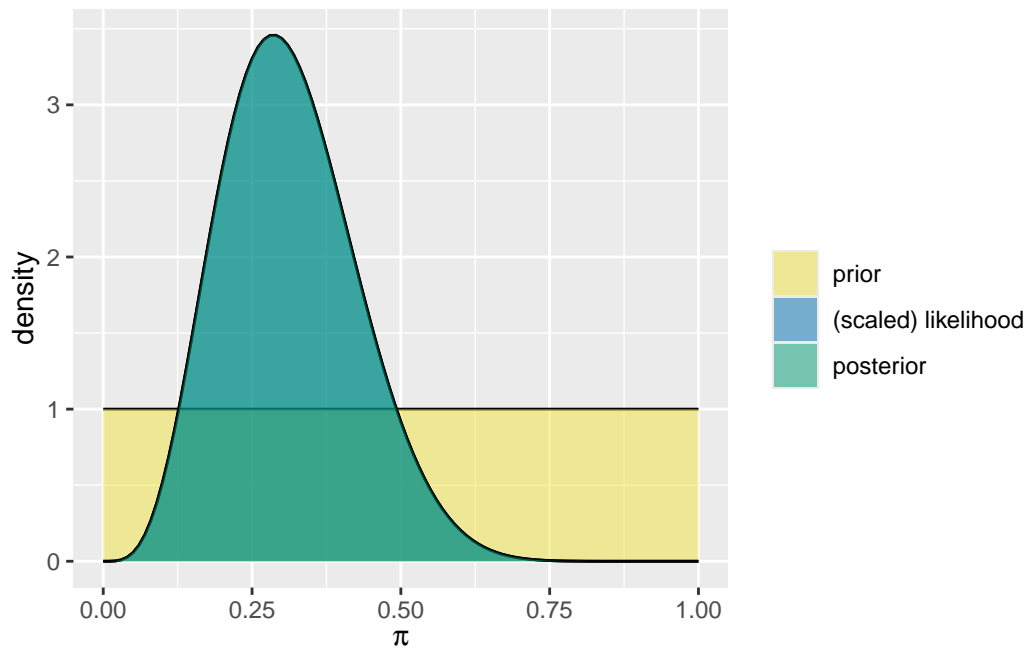
2025-09-17

## Ejercicio 4.19

Lo primero que se realizó fue cargar los datos y filtrarlos por el año de interés (1980).

binary	n	percent
FAIL	10	0.7142857
PASS	4	0.2857143
Total	14	1.0000000

Generamos el grafico de prior, verosimilitud y posterior con plot beta binomial



Se observa que la verosimilitud es igual al posterior. Esto ocurre porque el prior es una beta (1,1) que es uniforme en el recorrido. Interpretando esto de una manera bayesiana, significa que no aporta nada de información. Se utiliza para representar neutralidad absoluta. Siempre que partimos de una beta con tales parámetros, la verosimilitud será igual al posterior.

El posterior alcanza su moda entorno a 0.25 y deja prácticamente con probabilidad 0 a los valores superiores a 0.5. Esto nos da indicios que la varianza será considerablemente baja.

Para el **Cálculo teórico de la esperanza y modo** utilizaremos el cálculo del posterior del modelo beta binomial, que sabemos que distribuye Beta ( $\alpha+y, \beta+n-y$ ).

Dada dicha distribución, sabemos que la esperanza del posterior será:

$$\mathbb{E}[\pi] = \frac{\alpha + y}{\alpha + y + \beta + n - y} = \frac{\alpha + y}{\alpha + \beta + n} = \frac{1 + 4}{1 + 1 + 14} = \frac{5}{16} = 0.3125$$

Table 1: Medidas de resumen beta-binomial(1, 1, 4, 14)

model	alpha	beta	mean	mode	var	sd
prior	1	1	0.5000	NaN	0.0833	0.2887
posterior	5	11	0.3125	0.2857	0.0126	0.1124

De esta manera comprobamos que la Esperanza hallada de manera teórica es igual a la esperanza hallada por la función summarize.

Se refuerza la información que vimos en el gráfico del prior y posterior. Si bien comentamos que la moda del posterior estaba entorno a 0.25, ahora visualizamos que es exactamente 0.2857. Por otro lado se reafirma la baja varianza que comentamos.

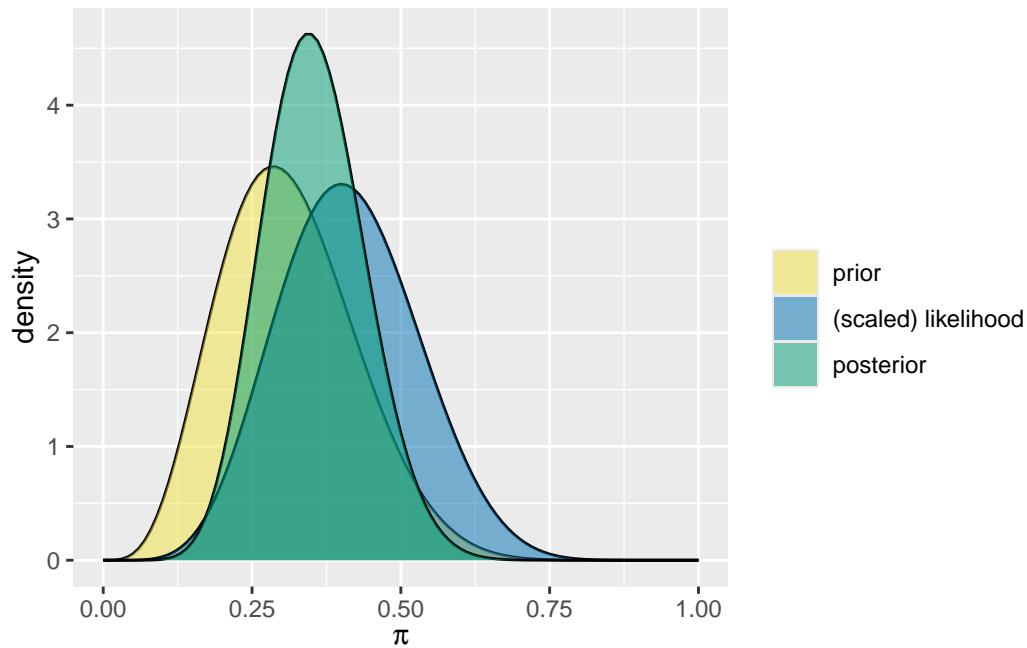
Finalmente, es correcto que no exista el modo del prior, dado que es una beta(1,1) y esta perfectamente equidistribuida.

## Parte dos: Cálculo para 1990

Busqueda de la información para 1990 (Verosimilitud)

binary	n	percent
FAIL	9	0.6
PASS	6	0.4
Total	15	1.0

Partimos del posterior anterior que ya sabemos como distribuye gracias a la formula del posterior en el modelo beta binomial y cargamos la verosimilitud buscada en el paso anterior. Dicho “posterior” se convirtio en nuestro nuevo prior y se agrega una nueva verosimilitud.



Se observa que el psoterior está en un punto medio entre priory verosimiitud. Entendemos que esto es correcto dado que el n dado no es lo suficientemente grande para volcar los datos hacia la verosimilitud, y además prior y verosimilitud son igual de concentrados entorno a un valor, ninguno es mucho más disperso ni concentrado que el otro.

Table 2: Medidas de resumen beta-binomial(5, 11, 6, 15)

model	alpha	beta	mean	mode	var	sd
prior	5	11	0.3125	0.2857	0.0126	0.1124
posterior	11	20	0.3548	0.3448	0.0072	0.0846

### Parte 3: Año 2000

Nuestro nuevo prior distribuirá Beta(11, 20)

```

binary n percent
FAIL 34 0.5396825
PASS 29 0.4603175
Total 63 1.0000000

```

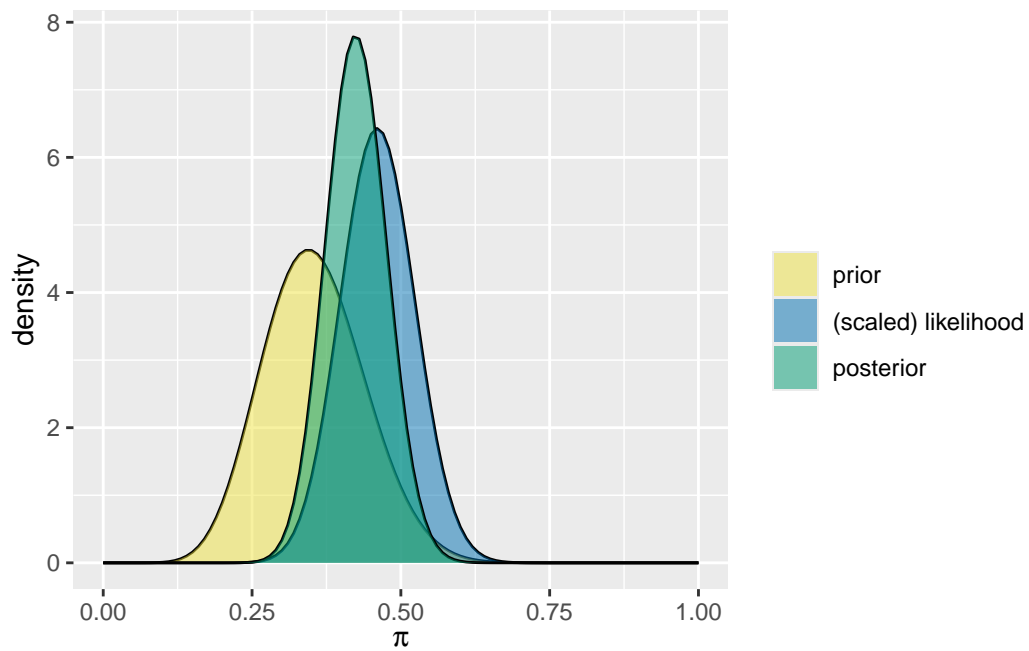


Table 3: Medidas de resumen beta-binomial(11, 20, 29, 63)

model	alpha	beta	mean	mode	var	sd
prior	11	20	0.3548	0.3448	0.0072	0.0846
posterior	40	54	0.4255	0.4239	0.0026	0.0507

#### Parte 4: Calculo de Jenna

```

binary  n  percent
FAIL   53  0.576087
PASS   39  0.423913
Total  92  1.000000

```

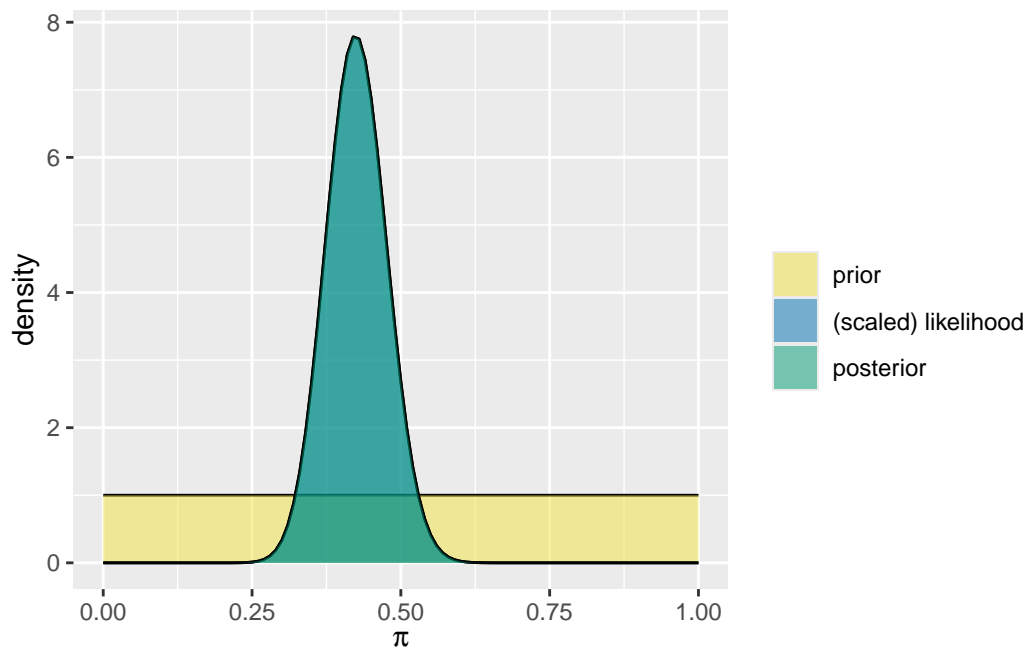


Table 4: Medidas de resumen beta-binomial(1, 1, 39, 92)

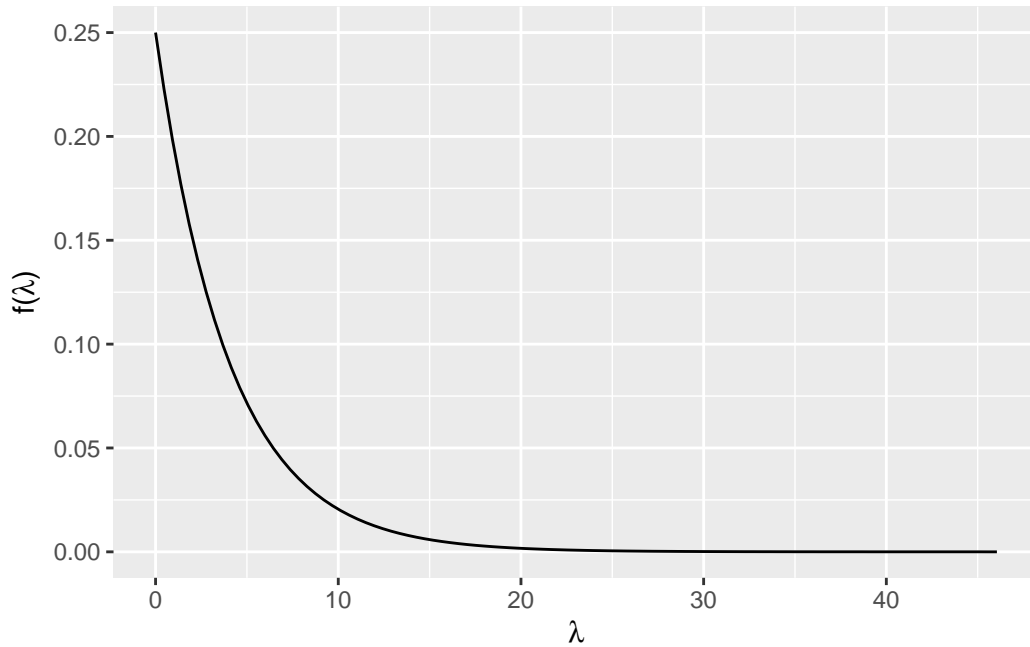
model	alpha	beta	mean	mode	var	sd
prior	1	1	0.5000	NaN	0.0833	0.2887
posterior	40	54	0.4255	0.4239	0.0026	0.0507

Se observa que Jenna llega a la misma posterior que John; probando que no importa el orden en el que mires la información, se llegará a los mismos resultados. John realiza su análisis en tres días (pasos), mientras que Jenna solo en uno. Pero al partir del mismo prior (Beta(1,1)) y utilizar las mismas muestras, llegan a los mismos posteriors.

Esta propiedad es denominada **propiedad de consistencia** o también **acumulación secuencial** y como comentamos anteriormente permite actualizar de una sola vez o en “capas”.

### Ejercicio 5.7: Womens world cup

Utilizaremos `plot_gamma(1, 0.25)` dado que es el prior dado.



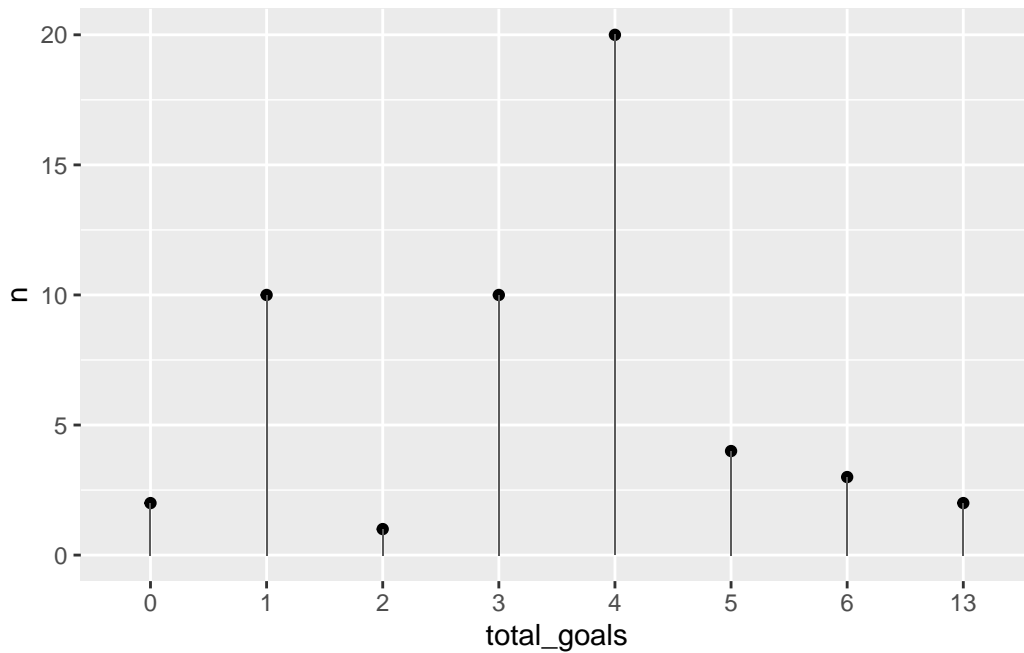
Se observa que  $\lambda$  es decreciente y parte desde cero, con asimetría a la derecha. Es sumamente coherente con el contexto de que  $\lambda$  representa la cantidad de goles promedio, sería extraño que 20 tenga probabilidad alta, por ejemplo. Sugiere que no habrá goles en la mayoría de partidos.

## Parte 2: $Y_i$ para modelo Poisson

Se utiliza el modelo Poisson para representar a  $Y$  dado que este es útil para los conteos.  $Y$  es una variable aleatoria para representar los goles.

Cada  $Y_i$  es una observación, es decir los goles de cada partido puntual. Por otro lado  $\lambda$  representa la tasa media de ocurrencia, o sea, la cantidad promedio de goles por partido.

### Parte 3: Total de goles por partido

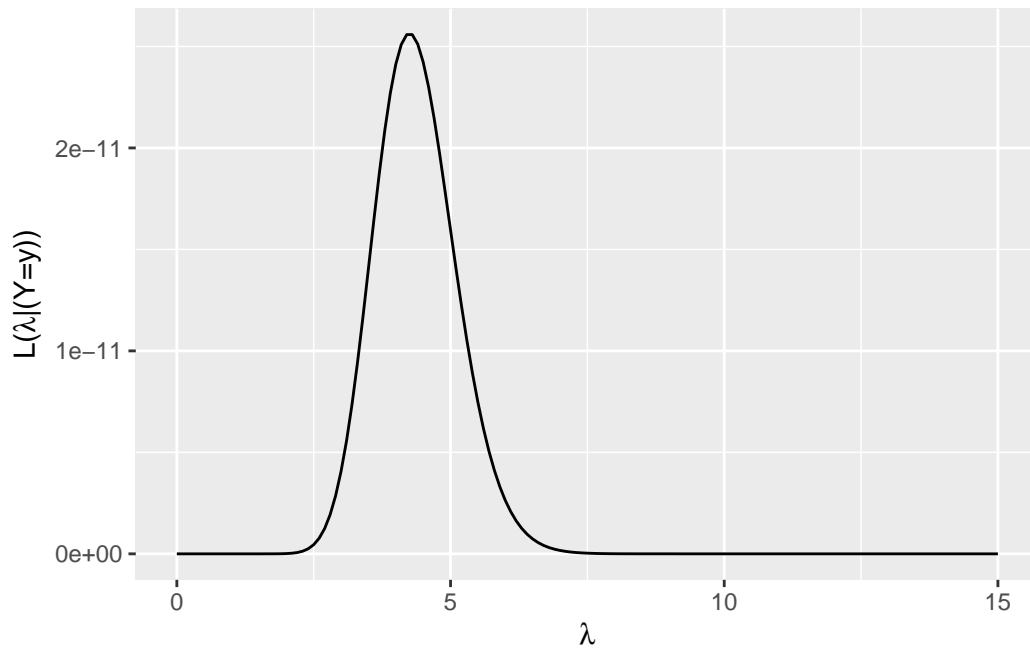


Se observan la cantidad de goles por partido y cuantas veces se repite ese resultado, es decir la frecuencia para cada resultado.

Para poder graficar el posterior y la verosimilitud necesitaremos calcular el total de goles que se convirtieron y el total de partidos. Este ultimo se observa que es 52 en `wwc_2019_matches`.

Para calcular el total de goles, debemos multiplicar la frecuencia de cada resultado por el total de goles de ese resultado y luego sumarlo.

“Sacamos” el valor NA que generamos antes al utilizar `as.numeric` (NO interesa graficarlo y alteraría la consistencia de nuestros datos)

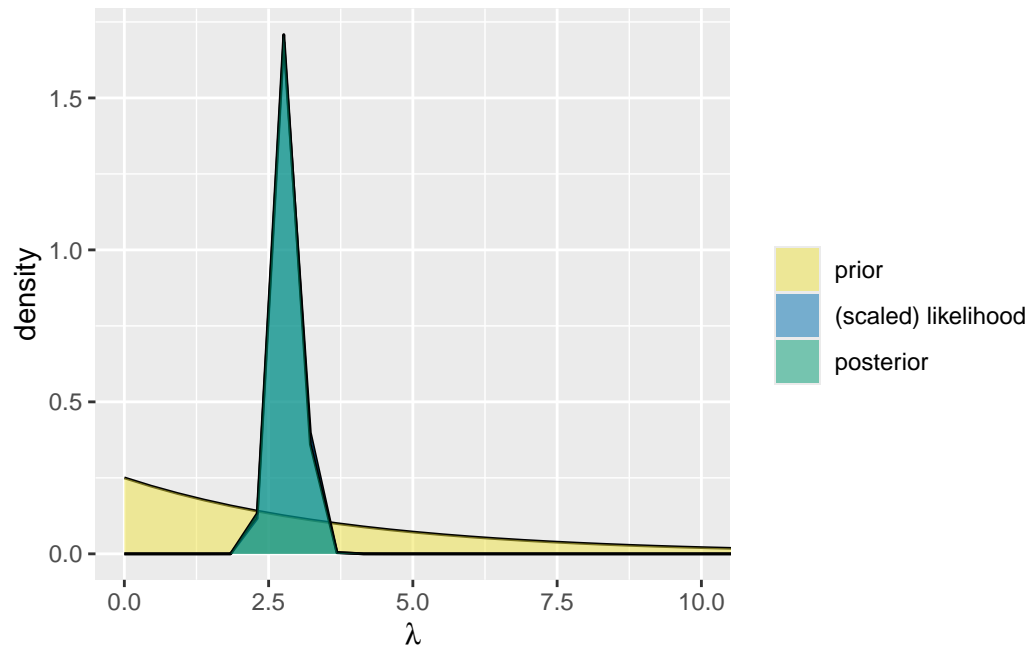


Se observa que los valores más verosímiles de  $\lambda$  rondan entre 3 y 5 goles de media por partido. Es una likelihood muy agresiva, en el sentido de que asigna probabilidad prácticamente nula a 0 y 1 gol.

### **Grafico del posterior**

Ahora interesa calcular el posterior. Para ello utilizaremos la funcion `plot_gamma_poisson`.





En este caso, el posterior y la likelihood son practicamente iguales. Recien al hacer zoom encontramos una pequena diferencia:

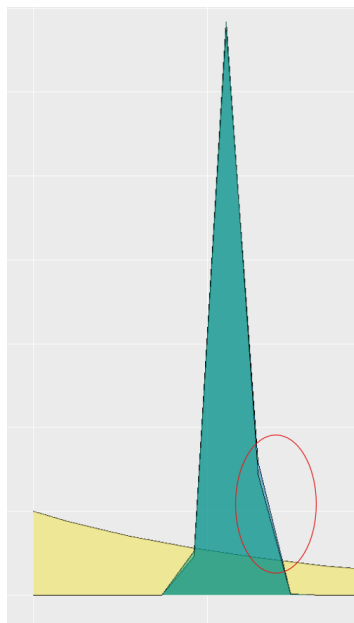


Figure 1: Diferencia entre Posterior y likelihood

Pensamos que esto ocurre por varios motivos: Tenemos un  $n$  considerablemente grande (52), la función de verosimilitud está sumamente concentrada entorno a un valor y el prior no está tan concentrado, sino que es más disperso.

De esta manera, hay un cambio radical en nuestro entendimiento de la media de goles por partido. De pasar de pensar que la media podía ser 0, 1 o 2, pasamos a poder afirmar que la media de goles por partido estará en el entorno de 3.

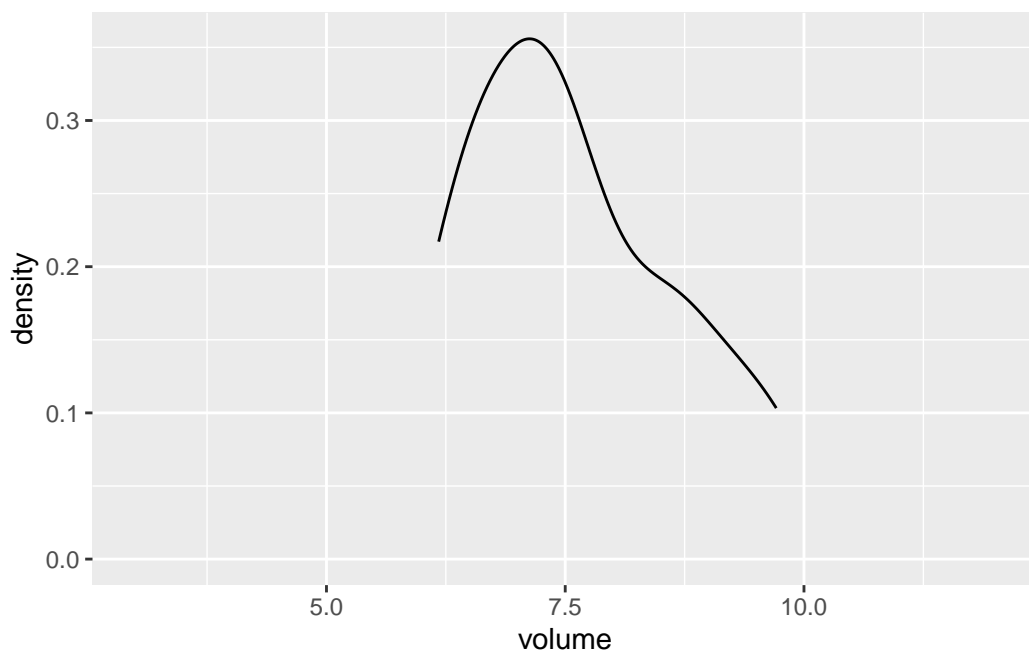
#### #Ejercicio 5.12: Control brains

Aclaración: Cuando la letra del ejercicio plantea “control subjects who have not been diagnosed with a concussion”, entendemos que refiere al grupo denominado control, pero puede haber habido un error de interpretación.

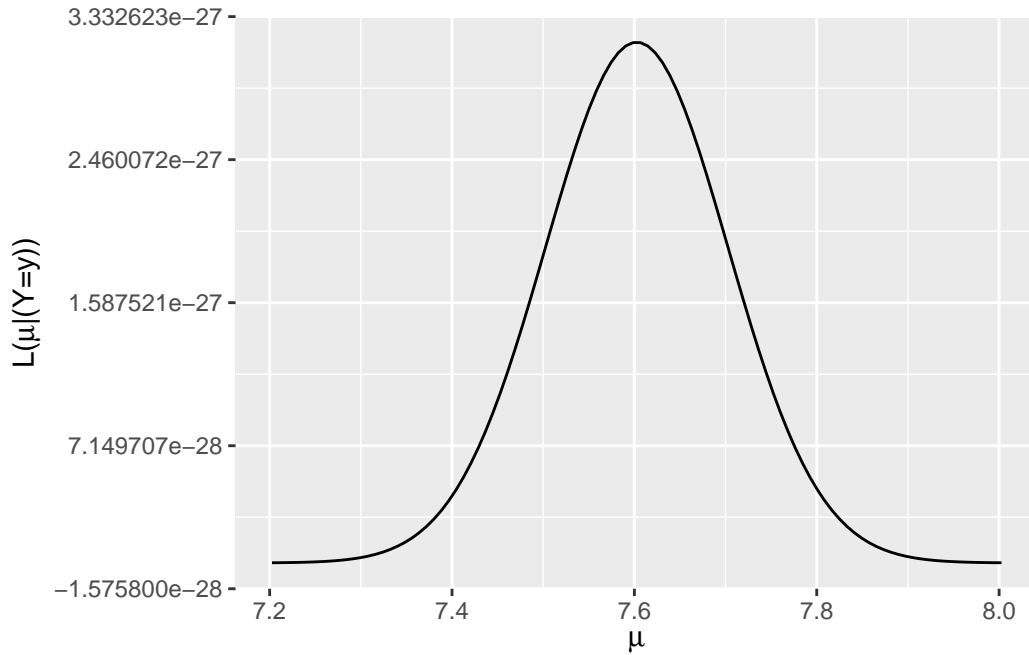
Primero procederemos con cargar los datos y filtramos para el grupo de control.

```
mean(volume)
1      7.6026
```

El promedio de los sujetos que no tuvieron contusiones es de 7.6026 cm<sup>3</sup> en una muestra de 25 personas.



Se observa que los valores van desde 6.175 hasta 9.71 y alcanzan el máximo de densidad entorno a 7.3, que es levemente menor al promedio (7.6026).



De esta manera se visualiza la verosimilitud de  $\mu$ . Los datos reflejan que la media, que calculamos anteriormente es el valor mas probable.

### Calculo del posterior

Ya tenemos todo lo necesario para identificar el posterior: El prior tiene media  $\theta=6.5$  y  $\tau=0.4$ . La muestra de los que no tuvieron contusión tiene media  $\bar{y}=7.6026$  y asumimos desvío conocido, pero por letra del ejercicio  $\sigma=0.5$ .

Sabemos que el Modelo Normal-Normal distribuye de la siguiente manera:

$$Y_i | \mu \stackrel{ind}{\sim} N(\mu, \sigma^2)$$

$$\mu \sim N(\theta, \tau^2)$$

$$\mu | \bar{y} \sim N \left( \theta \frac{\sigma^2}{n\tau^2 + \sigma^2} + \bar{y} \frac{n\tau^2}{n\tau^2 + \sigma^2}, \frac{\tau^2 \sigma^2}{n\tau^2 + \sigma^2} \right)$$

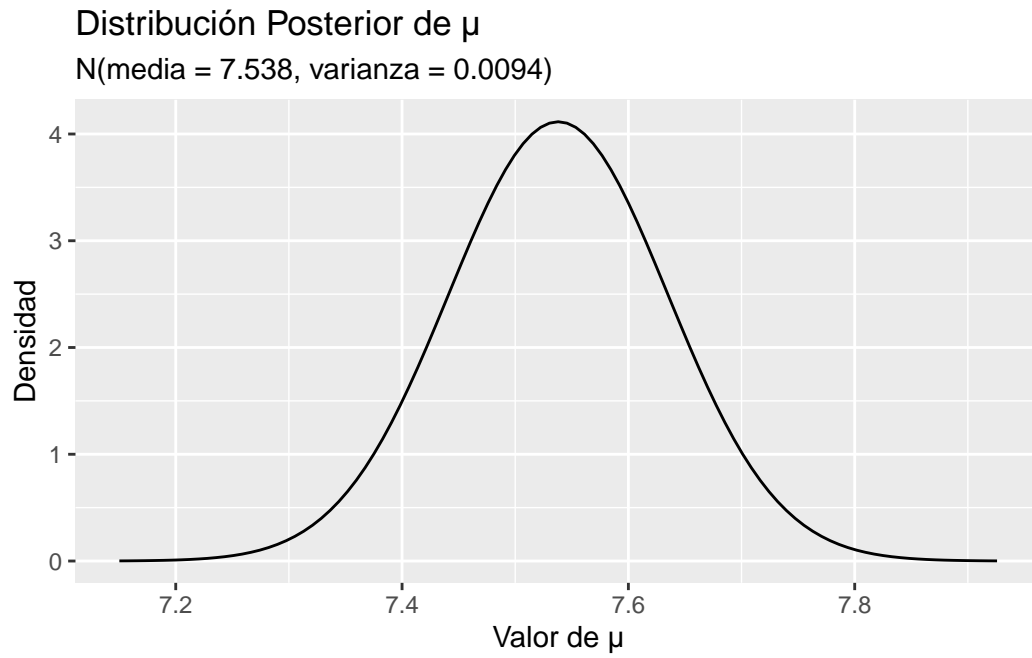
Por lo tanto, sustituyendo los valores que obtuvimos llegamos a que el modelo bayesiano conjugado Normal-Normal es el siguiente:

$$\mu | \bar{y} \sim N \left( 6.5 \frac{0.5^2}{25 * 0.4^2 + 0.5^2} + 7.6026 \frac{25 * 0.4^2}{25 * 0.4^2 + 0.5^2}, \frac{0.4^2 0.5^2}{25 * 0.4^2 + 0.5^2} \right)$$

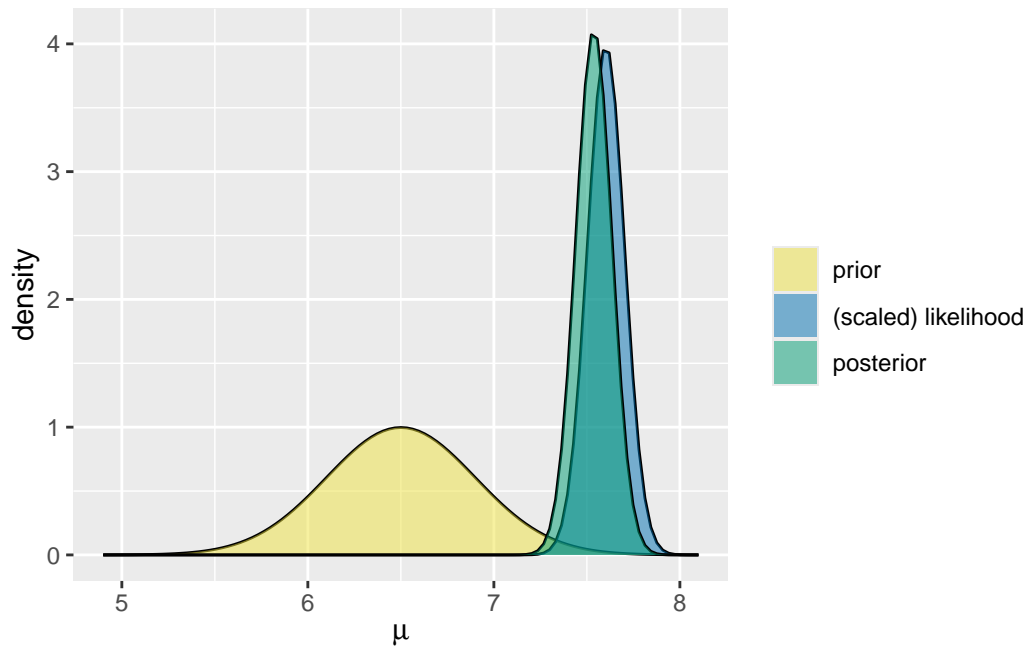
Entonces:

$$\mu|\vec{y} \sim N(7.538, 0.0094)$$

El grafico de nuestra distriubucion queda de la forma:



### Parte 3: Plot del prior, verosimilitud y posterior



Nuevamente, dado que tenemos una verosimilitud cuyos datos están tan concentrados, el posterior se pega mucho a la verosimilitud, siendo estos muy similares en forma y media.

Table 5: Medidas de resumen para Normal-Normal(6.5, 0.4, 0.5, 7.6026, 25)

model	mean	mode	var	sd
prior	6.5000	6.5000	0.1600	0.400
posterior	7.5377	7.5377	0.0094	0.097

De esta manera confirmamos que nuestro posterior calculado manualmente en la Parte 2 era correcto.

### Ejercicio Normal-Normal Conjuga

#### Consigna:

- Calcular la distribución a posteriori para  $\mu$ .
- Demostrar que la distribución a posteriori para  $\mu$  es normal con media igual a un promedio ponderado de la media a priori  $\theta$  y de la media muestral.

- (iii) Analizar la influencia del tamaño muestral  $n$  sobre la media y la varianza de la distribución a posteriori para  $\mu$ . Conclusión: la distribución a priori normal para  $\mu$  es conjugada con el modelo muestral de observaciones independientes de una población normal con media  $\mu$  y varianza conocida  $\sigma^2$

**Respuestas:** Nos basamos en la guía del libro para demostrar la consigna paso a paso. Vamos a probar que la posteriori del modelo Normal-Normal sigue también una distribución normal

Primero, tenemos que la función de densidad de  $\mu$  (posteriori) es proporcional al producto de la función de densidad priori de la Normal y la función de verosimilitud. Lo que queremos encontrar es la posteriori, por otro lado, el priori es lo que sabíamos de  $\mu$  antes de ver los datos (la verosimilitud). Para todo  $\mu$  perteneciente a los reales tenemos:

$$f(\mu|\vec{y}) \propto f(\mu)L(\mu|\vec{y}) \propto \exp\left[-\frac{(\mu - \theta)^2}{2\tau^2}\right] \cdot \exp\left[-\frac{(\bar{y} - \mu)^2}{2\sigma^2/n}\right]$$

De forma resumida:

$$f(\mu|\vec{y}) \propto \underbrace{\exp\left[-\frac{(\mu - \theta)^2}{2\tau^2}\right]}_{\text{Núcleo del Prior}} \cdot \underbrace{\exp\left[-\frac{(\bar{y} - \mu)^2}{2\sigma^2/n}\right]}_{\text{Núcleo de la Verosimilitud}}$$

El objetivo ahora es deshacernos de los paréntesis al cuadrado para poder agrupar los términos que contienen  $\mu$ . Para esto, se usa la fórmula del binomio conjugado:  $(a - b)^2 = a^2 - 2ab + b^2$ .

Al expandir los exponentes, obtenemos:

$$-(\mu - \theta)^2 = -(\mu^2 - 2\mu\theta + \theta^2) = -\mu^2 + 2\mu\theta - \theta^2$$

$$-(\bar{y} - \mu)^2 = -(\mu - \bar{y})^2 = -(\mu^2 - 2\mu\bar{y} + \bar{y}^2) = -\mu^2 + 2\mu\bar{y} - \bar{y}^2$$

$$f(\mu|\vec{y}) \propto \exp\left[\frac{-\mu^2 + 2\mu\theta - \theta^2}{2\tau^2}\right] \exp\left[\frac{-\mu^2 + 2\mu\bar{y} - \bar{y}^2}{2\sigma^2/n}\right]$$

Este siguiente paso es muy importante: “nos olvidaremos de las constantes”. Como estamos trabajando con proporcionalidad con respecto a  $\mu$  y los términos  $\theta^2$  y  $\bar{y}^2$  no contienen ningún  $\mu$ , consideramos a estos como constantes. Esto va a simplificar mucho las cuentas. Obtenemos:

$$\propto \exp\left[\frac{-\mu^2 + 2\mu\theta}{2\tau^2}\right] \exp\left[\frac{-\mu^2 + 2\mu\bar{y}}{2\sigma^2/n}\right]$$

A continuación, hacemos denominador común para luego poder aplicar las propiedades de potencia y sumar los exponentes:

$$\begin{aligned} &\propto \exp \left[ \frac{(-\mu^2 + 2\mu\theta)\sigma^2/n}{2\tau^2\sigma^2/n} \right] \exp \left[ \frac{(-\mu^2 + 2\mu\bar{y})n\tau^2}{2\tau^2\sigma^2/n} \right] \\ &\propto \exp \left[ \frac{(-\mu^2 + 2\mu\theta)\sigma^2 + (-\mu^2 + 2\mu\bar{y})n\tau^2}{2\tau^2\sigma^2} \right] \end{aligned}$$

Ahora, agrupamos los terminos  $\mu$  y  $\mu^2$  para simplificar:

$$f(\mu|\vec{y}) \propto \exp \left[ \frac{-\mu^2(n\tau^2 + \sigma^2) + 2\mu(\theta\sigma^2 + \bar{y}n\tau^2)}{2\tau^2\sigma^2} \right]$$

¿Cómo podemos hacer para demostrar que lo que obtuvimos es una distribución normal? Debemos lograr de alguna manera “hacer aparecer” el núcleo de la Normal. Un truco importante que se utiliza es el “complete the square” (completar el cuadrado).

$$\propto \exp \left[ \frac{-\mu^2 + 2\mu \left( \frac{\theta\sigma^2 + \bar{y}n\tau^2}{n\tau^2 + \sigma^2} \right)}{2(\tau^2\sigma^2)/(n\tau^2 + \sigma^2)} \right]$$

Ahora podemos factorizar:

$$f(\mu|\vec{y}) \propto \exp \left[ -\frac{\left( \mu - \frac{\theta\sigma^2 + \bar{y}n\tau^2}{n\tau^2 + \sigma^2} \right)^2}{2(\tau^2\sigma^2)/(n\tau^2 + \sigma^2)} \right]$$

Obtenemos el núcleo de una función de densidad de una Normal para  $\mu$ . Utilizando esto podemos concluir que la distribución queda de esta forma:

$$\mu|\vec{y} \sim N \left( \frac{\theta\sigma^2 + \bar{y}n\tau^2}{n\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2} \right)$$

Por lo tanto, queda demostrado que la posteriori de una Normal-Normal conjugada es también una Normal.

Podemos reorganizar la media de la posteriori para demostrar que es el promedio ponderado de la media a priori de  $\mu$  ( $E(\mu) = \theta$ ) y la media de la muestra de  $\bar{y}$ :

- El peso del prior depende de los datos
- El peso de los datos depende del prior y de la cantidad de datos

$$\frac{\theta\sigma^2 + \bar{y}n\tau^2}{n\tau^2 + \sigma^2} = \theta\frac{\sigma^2}{n\tau^2 + \sigma^2} + \bar{y}\frac{n\tau^2}{n\tau^2 + \sigma^2}$$

Concluimos también que la varianza a posteriori obtiene información de la variabilidad a priori de  $\tau$  y la variabilidad en los datos de  $\sigma$ . Ambos se ven afectados por el tamaño de la muestra,  $n$ .

En primer lugar, a medida que  $n$  aumente, la media a posteriori toma una ponderación menor en la media a priori y mayor ponderación en la media de la muestra  $\bar{y}$ . Tenemos entonces:

$$\frac{\sigma^2}{n\tau^2 + \sigma^2} \rightarrow 0 \quad \text{y} \quad \frac{n\tau^2}{n\tau^2 + \sigma^2} \rightarrow 1$$

Luego, cuando  $n$  aumenta, la varianza a posteriori disminuye:

$$\frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2} \rightarrow 0$$

Esto significa que, a medida que obtenemos mayor cantidad de datos, la certeza de la posterior con respecto a  $\mu$  aumenta y se vuelve más cercana a los valores de la información.