

Tarea 4

Juan M Karawacki

Bruno Pintos

2025-09-10

Ejercicio 3.11 (Uso regular de bicicleta)

Una universidad quiere saber qué proporción de estudiantes son ciclistas regulares, π , para poder instalar una cantidad apropiada de estacionamientos de bicicletas. Como la universidad está en la soleada California del Sur, el personal piensa que π tiene una media de 1 de cada 4 estudiantes, y una moda de 5/22.

1. Especifica y grafica un modelo Beta que refleje las ideas previas

La media de la Beta se calcula como $\mu = \frac{\alpha}{\alpha+\beta}$, y si α y β son mayores a 1, tenemos que la moda = $\frac{\alpha-1}{\alpha+\beta-2}$.

Cálculo de parámetros de la distribución Beta

Según el personal de la universidad, en promedio 1 de cada 4 estudiantes es ciclista regular, así que la media esperada es 0.25. Además, creen que el valor más “probable” es $\frac{5}{22}$.

$$\text{Media: } \frac{\alpha}{\alpha + \beta} = m, \quad \text{Moda: } \frac{\alpha - 1}{\alpha + \beta - 2} = M.$$

Sea ; $S = \alpha + \beta$. Entonces:

$$\alpha = mS, \quad \alpha = 1 + M(S - 2).$$

Igualando y despejando S :

$$mS = 1 + MS - 2M \Rightarrow S(m - M) = 1 - 2M \Rightarrow S = \frac{1 - 2M}{m - M}.$$

Luego:

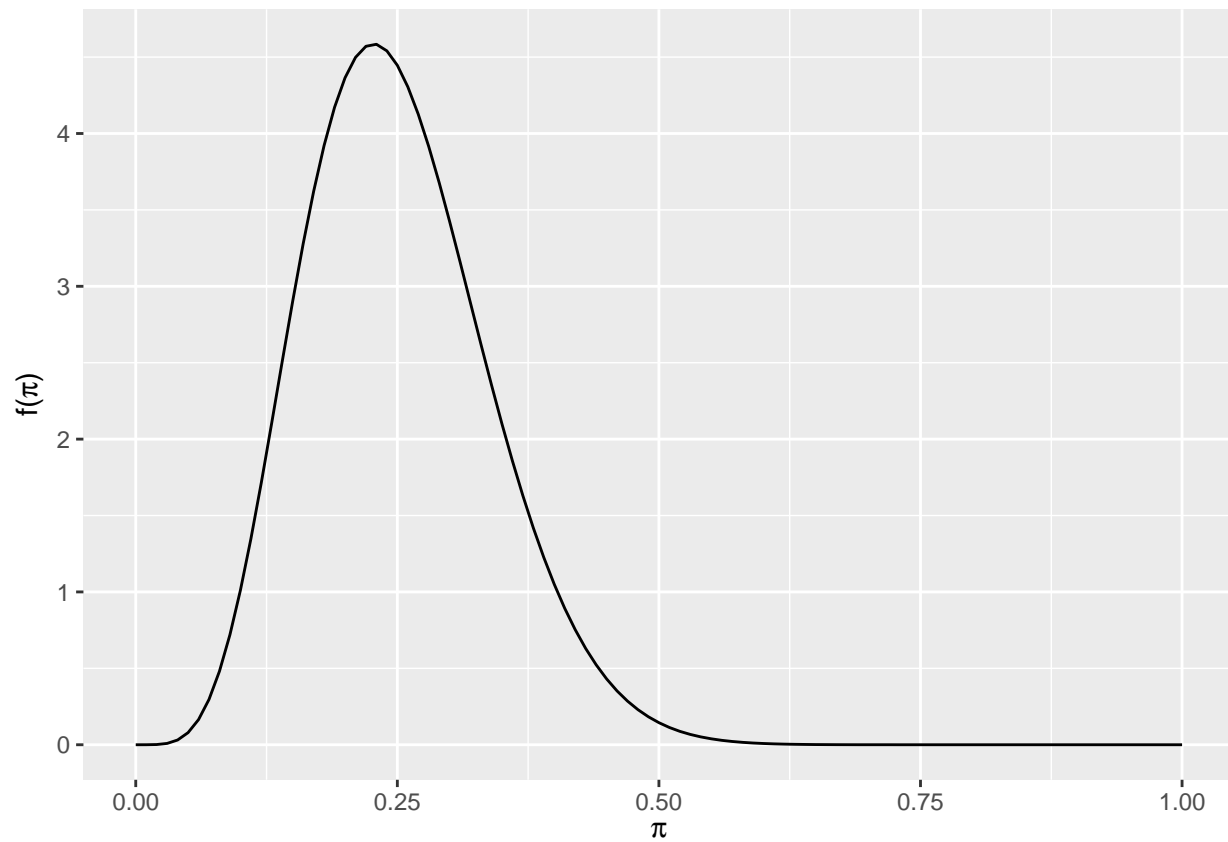
$$\alpha = mS, \quad \beta = S - \alpha.$$

Para reflejar estas ideas, igualamos la media y la moda a las fórmulas de la Beta. De la media obtenemos que $\beta = 3\alpha$, y al poner eso en la fórmula de la moda se resuelve que $\alpha = 6$ y $\beta = 18$.

Con esto, nuestro modelo previo queda como $\pi \sim \text{Beta}(6, 18)$. La media y la moda coinciden con lo que pensaba el personal, y la distribución tiene una cola hacia valores más altos de π , lo cual tiene sentido: la mayoría de los estudiantes no son ciclistas regulares, pero siempre hay algunos que sí lo son.

Para ver cómo queda la distribución, la graficamos en R con `plot_beta(alpha = 6, beta = 18)`. Así se nota claramente dónde se concentra la probabilidad y cuáles son los valores más plausibles según lo que sabía el personal.

```
library(bayesrules)
plot_beta(alpha = 6, beta = 18)
```

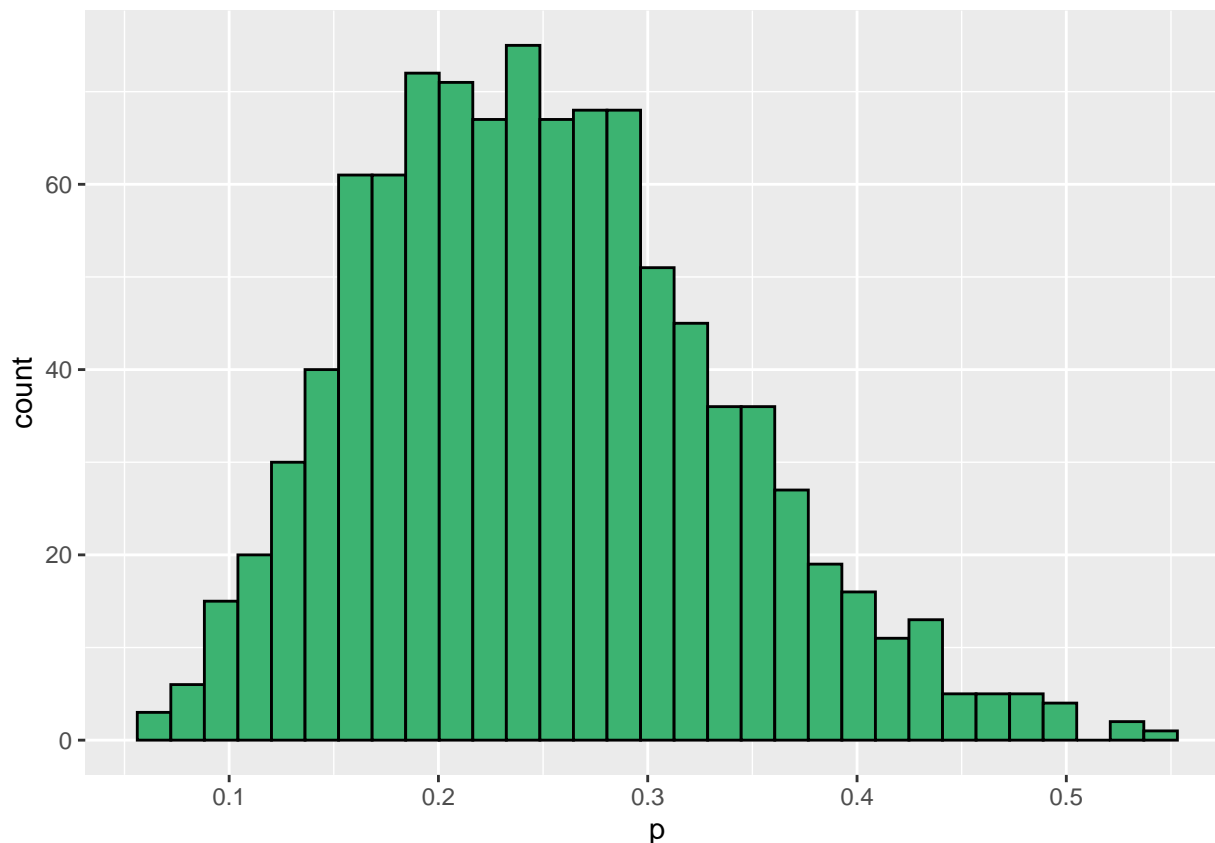


```
set.seed(1234)
n <- 1000
m <- 1/4
M <- 5/22
# k <- alpha + beta

k <- ((-2*M) + 1)/(m - M)
alpha <- m*k
bet <- k - alpha

p <- rbeta(n, alpha, bet, ncp = 0)
datos <- data.frame(p = p)

ggplot(data = datos, aes(x = p)) +
  geom_histogram(bins = trunc(sqrt(length(p))), fill = "mediumseagreen", color = "black")
```



2. Entre 50 estudiantes encuestados, 15 son ciclistas regulares. ¿Cuál es el modelo posterior para π ?

Tomando los parámetros del prior ($\alpha = 6$, $\beta = 18$) y los datos de la encuesta ($y = 15$, $n = 50$), podemos calcular el modelo posterior de π usando la regla de conjugación Beta-Binomial. La actualización se realiza sumando los éxitos observados al parámetro α y los fracasos al parámetro β :

$$\alpha_{\text{post}} = \alpha + y = 6 + 15 = 21$$

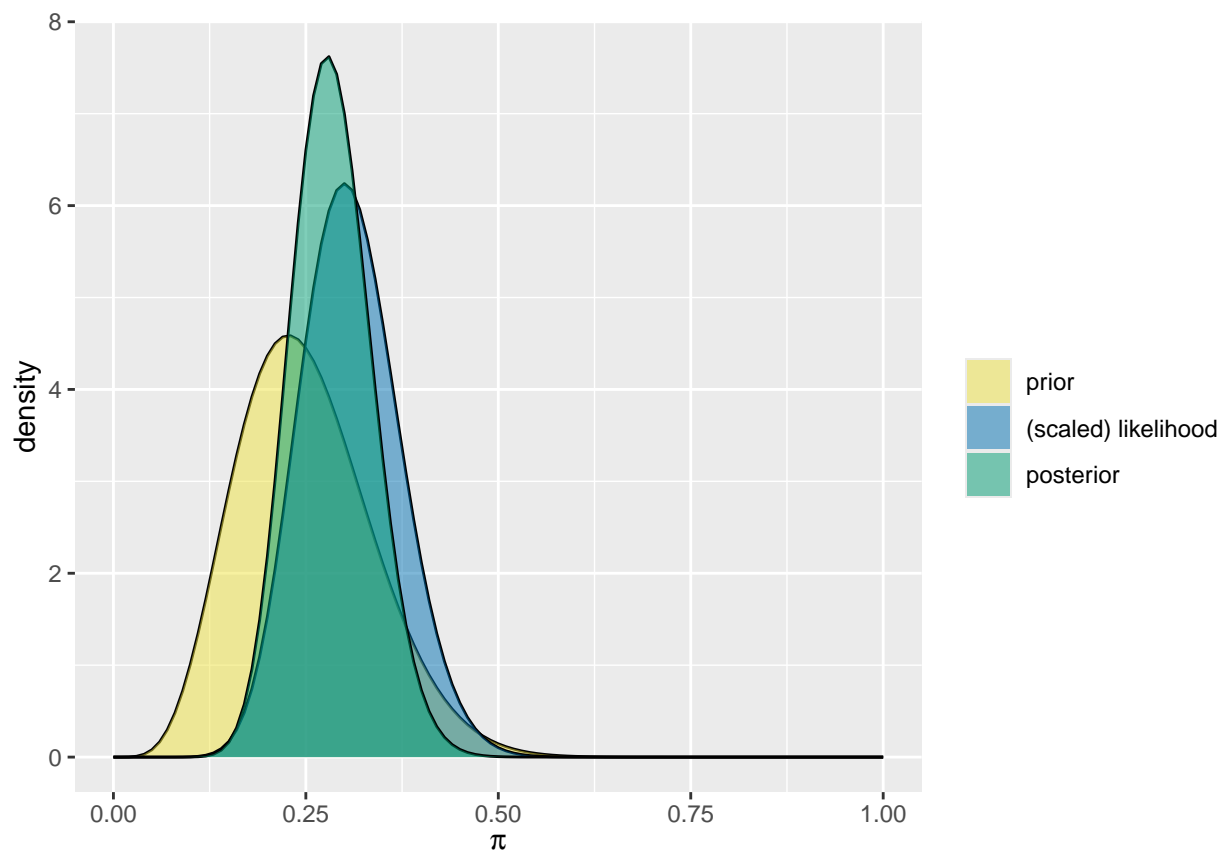
$$\beta_{\text{post}} = \beta + n - y = 18 + (50 - 15) = 53$$

Por lo tanto, el modelo posterior es:

$$\pi \mid (Y = 15) \sim \text{Beta}(21, 53)$$

Podemos visualizar este posterior y compararlo con el prior usando la función `plot_beta_binomial` del paquete **bayesrules**:

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



3. ¿Cuál es la media, la moda y la desviación estándar del modelo posterior?

```
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

Después de encuestar a 50 estudiantes y ver que 15 son ciclistas regulares, actualizamos nuestro modelo previo Beta(6,18) con los datos reales y obtuvimos la distribución posterior Beta(21,53). Esto nos da una media de alrededor de 0.28, una moda cerca de 0.28 también y una desviación estándar de 0.05. Comparando con el modelo previo, que tenía media 0.25, moda 0.23 y una desviación mayor, se nota que la nueva distribución se mueve un poco hacia los datos observados y además tiene menos incertidumbre. Es decir, ahora tenemos una idea más precisa de la proporción de ciclistas, todavía refleja las creencias previas del personal, pero incorpora de manera clara la información real que obtuvimos de la encuesta.

4. ¿El modelo posterior refleja más de cerca la información previa o los datos? Explica tu razonamiento.

Al analizar el modelo posterior, queda claro cómo se combinan la información previa y los datos reales. Nuestra prior Beta(6,18) reflejaba las creencias del personal, con una media de 0.25 y una moda de $\frac{5}{22}$, mostrando que esperaban que solo una minoría de estudiantes fueran ciclistas regulares. Luego, al observar que 15 de 50 estudiantes son ciclistas, el modelo posterior se actualiza a Beta(21,53), desplazando la media

y la moda a 0.28, con una desviación estándar más pequeña de 0.05. Esto indica que la información de los datos tuvo un efecto visible: la distribución posterior se mueve hacia la proporción observada, pero sin ignorar completamente las creencias previas.

Entonces podemos decir que el posterior representa un equilibrio entre lo que pensaba el personal y lo que mostraron los datos. La reducción en la dispersión refleja que ahora tenemos más confianza en la estimación de la proporción real de ciclistas. Podemos concluir que, aunque los datos influyeron para ajustar la estimación hacia la proporción observada, el modelo todavía conserva la forma y tendencia del prior. Así, la posterior ofrece una visión más completa y precisa de la proporción de ciclistas regulares, combinando la información previa y la evidencia empírica.

Ejercicio 3.14 (Resumiendo la Beta-Binomial: Primera Parte)

Escribe el código de entrada correspondiente para obtener la salida de `summarize_beta_binomial()` que se muestra a continuación.

model	alpha	beta	mean	mode	var	sd
prior	2	3	0.4000	0.3333	0.040000	0.20000
posterior	11	24	0.3143	0.3030	0.005986	0.07737

Mirando la tabla sabemos que nuestro modelo previo está definido como $\pi \sim \text{Beta}(\alpha, \beta)$, con $\alpha = 2$ y $\beta = 3$. Luego, observamos que el posterior tiene parámetros $\alpha_{\text{post}} = 11$ y $\beta_{\text{post}} = 24$. Para encontrar los datos que generaron esta actualización, usamos las reglas de conjugación Beta-Binomial, como hicimos en el inciso 2 del ejercicio anterior solo que ahora conocemos los α_{post} y β_{post} , nuestras incógnitas esta vez son y y n .

1. La actualización de α se realiza sumando los éxitos observados y al parámetro α del prior:

$$\alpha_{\text{post}} = \alpha + y$$

Despejando y :

$$y = \alpha_{\text{post}} - \alpha = 11 - 2 = 9$$

2. La actualización de β se realiza sumando los fracasos observados $(n - y)$ al parámetro β del prior:

$$\beta_{\text{post}} = \beta + n - y$$

Despejando n :

$$n = (\beta_{\text{post}} - \beta) + y = (24 - 3) + 9 = 30$$

Con estos cálculos, ahora tenemos todos los valores necesarios para correr `summarize_beta_binomial()`:

- Prior: $\alpha = 2$, $\beta = 3$
- Datos observados: $y = 9$, $n = 30$

```
summarize_beta_binomial(alpha = 2, beta = 3, y = 9, n = 30)
```

```
##      model alpha beta      mean      mode      var      sd
## 1    prior      2    3 0.4000000 0.3333333 0.040000000 0.20000000
## 2 posterior     11   24 0.3142857 0.3030303 0.005986395 0.07737179
```

Capítulo 4

La heladería local está abierta hasta que se le acaba el helado del día. Son las 2 p.m. y Chad quiere comprarse un cucurucho de helado. Él les pregunta a sus compañeros de trabajo sobre la probabilidad (π) de que la heladería todavía esté abierta. Los priors Beta que cada compañero tiene para π se muestran a continuación:

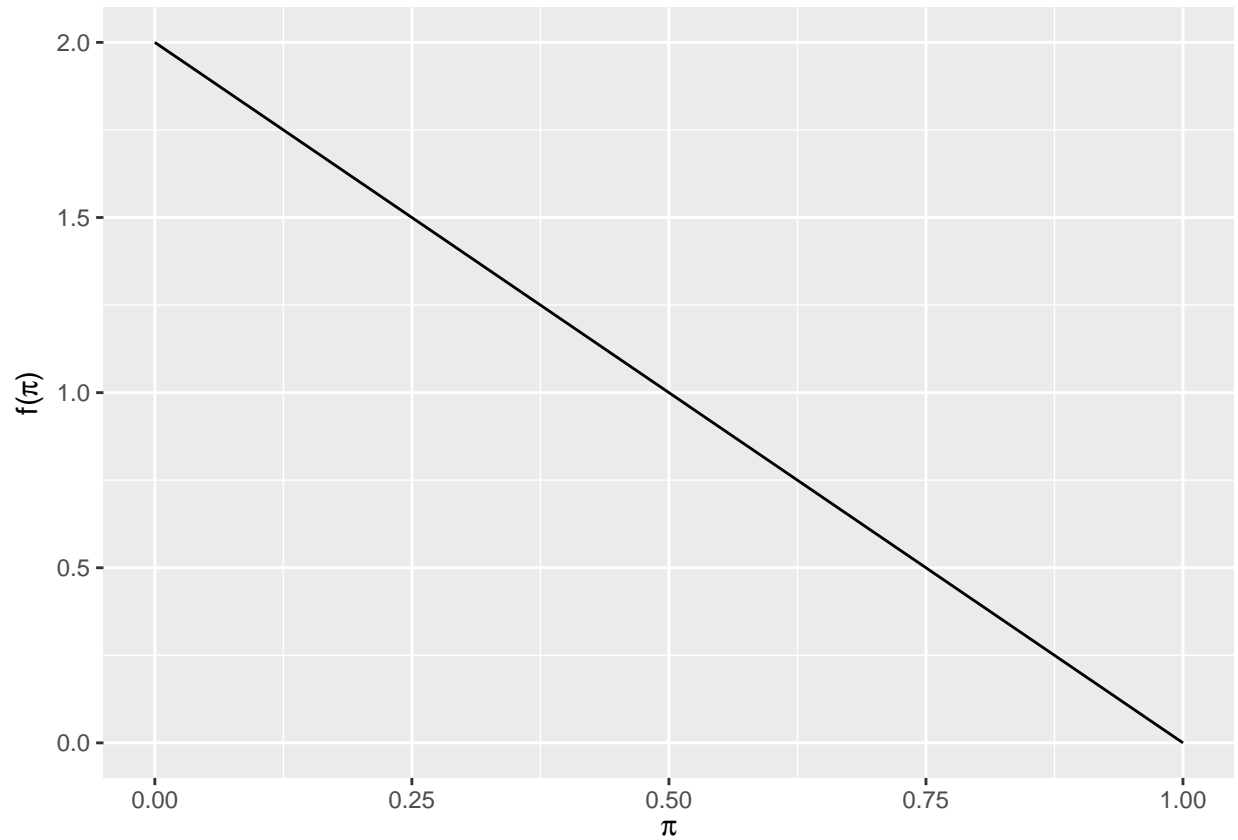
Compañero	Prior
Kimya	Beta(1, 2)
Fernando	Beta(0.5, 1)
Ciara	Beta(3, 10)
Taylor	Beta(2, 0.1)

Ejercicio 4.4 (Elección del prior)

Visualiza y resume (con palabras) la comprensión previa de cada compañero sobre las probabilidades de que Chad satisfaga su antojo de helado

Para un primer acercamiento graficamos las betas de las creencias de cada uno de los compañeros de Chad y además hacemos `summarize_beta` para tener la media y el desvío de cada distribución.

```
plot_beta(alpha = 1, beta = 2) # Kimya
```



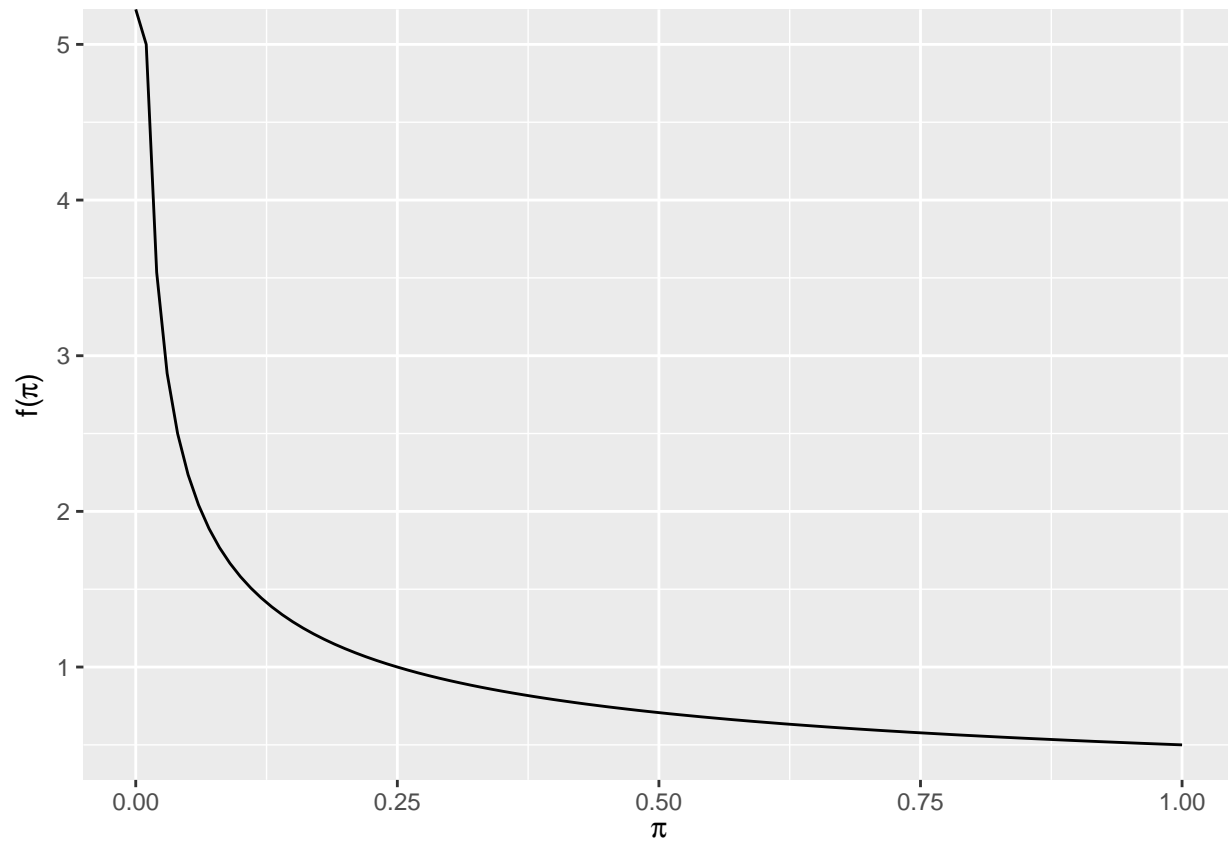
```
summarize_beta(alpha = 1, beta = 2)
```

```
##      mean mode      var      sd  
## 1 0.3333333    0 0.05555556 0.2357023
```

- **Kimya: Beta(1, 2)**

La distribución tiene forma decauyente hacia la derecha, con media de $1/3$. Esto indica que Kimya cree que Chad tiene pocas probabilidades de conseguir su helado, aunque no descarta totalmente que pueda lograrlo. La distribución no es muy concentrada ya que tiene un desvío de 0.24, mostrando cierta incertidumbre.

```
plot_beta(alpha = 0.5, beta = 1) # Fernando
```



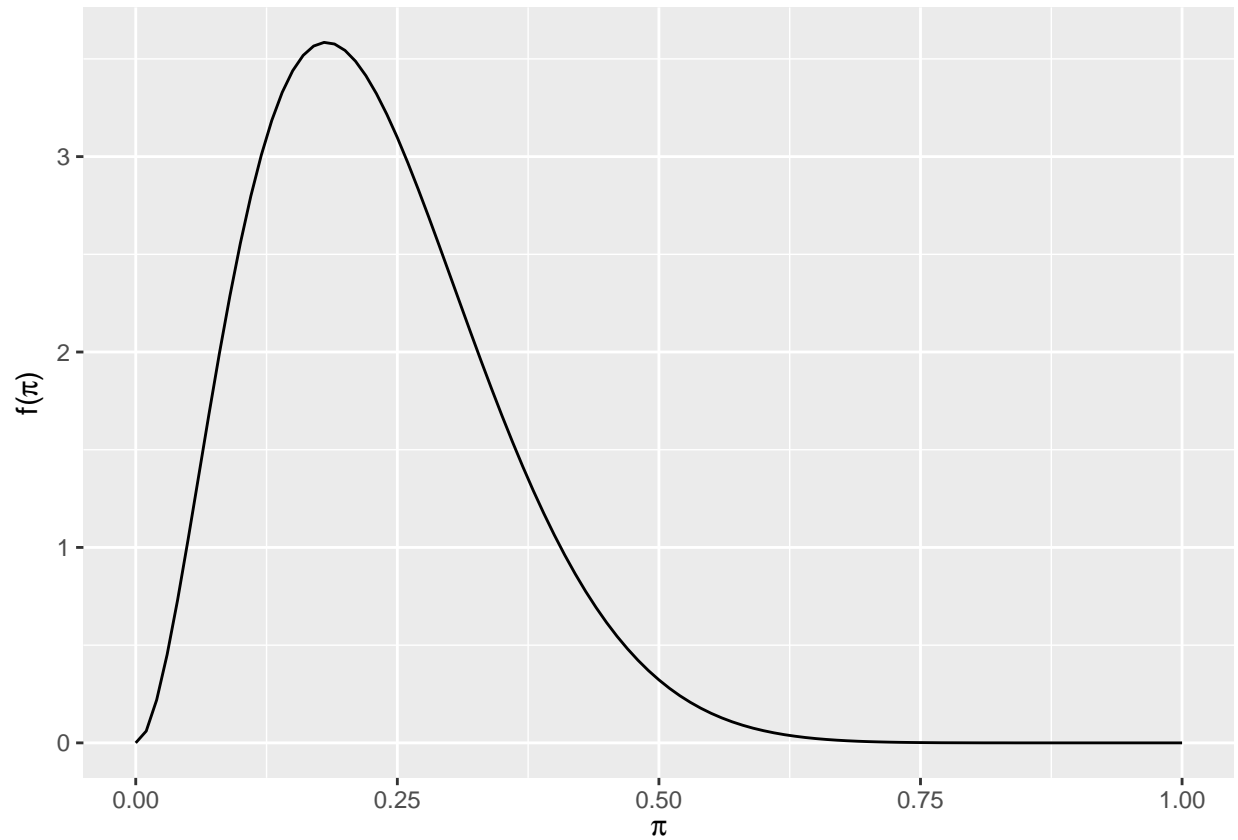
```
summarize_beta(alpha = 0.5, beta = 1)
```

```
##           mean mode           var           sd
## 1 0.3333333      1 0.08888889 0.2981424
```

- **Fernando: Beta(0.5, 1)**

La prior de Fernando también tiene media de $1/3$, pero la grafica por momentos es más “plana” y dispersa debido a los parámetros menores a 1. Esto refleja que Fernando cree que Chad tiene pocas probabilidades de éxito, pero con mucha incertidumbre: cualquier probabilidad entre 0 y 1 es relativamente plausible.


```
plot_beta(alpha = 3, beta = 10) # Ciara
```



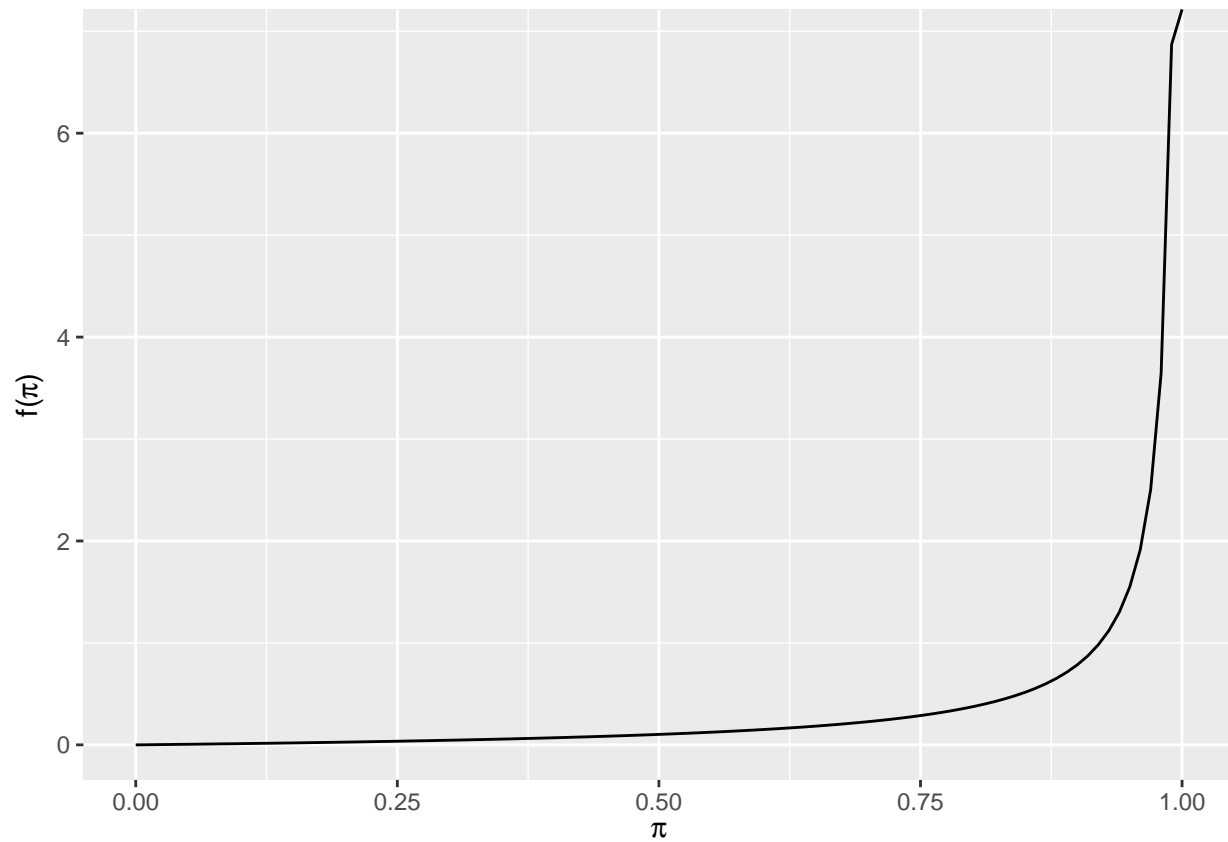
```
summarize_beta(alpha = 3, beta = 10)
```

```
##      mean      mode      var      sd
## 1 0.2307692 0.1818182 0.01267963 0.1126039
```

- **Ciara: Beta(3, 10)**

La media es $\frac{3}{10}$, indicando que Ciara es más pesimista que Kimya y Fernando, piensa que es muy poco probable que Chad consiga su helado. La distribución está concentrada hacia valores bajos, mostrando que Ciara tiene bastante confianza en su creencia.

```
plot_beta(alpha = 2, beta = 0.1) # Taylor
```



```
summarize_beta(alpha = 2, beta = 0.1)
```

```
##      mean mode      var      sd  
## 1 0.952381    1 0.01462951 0.1209525
```

- **Taylor: Beta(2, 0.1)**

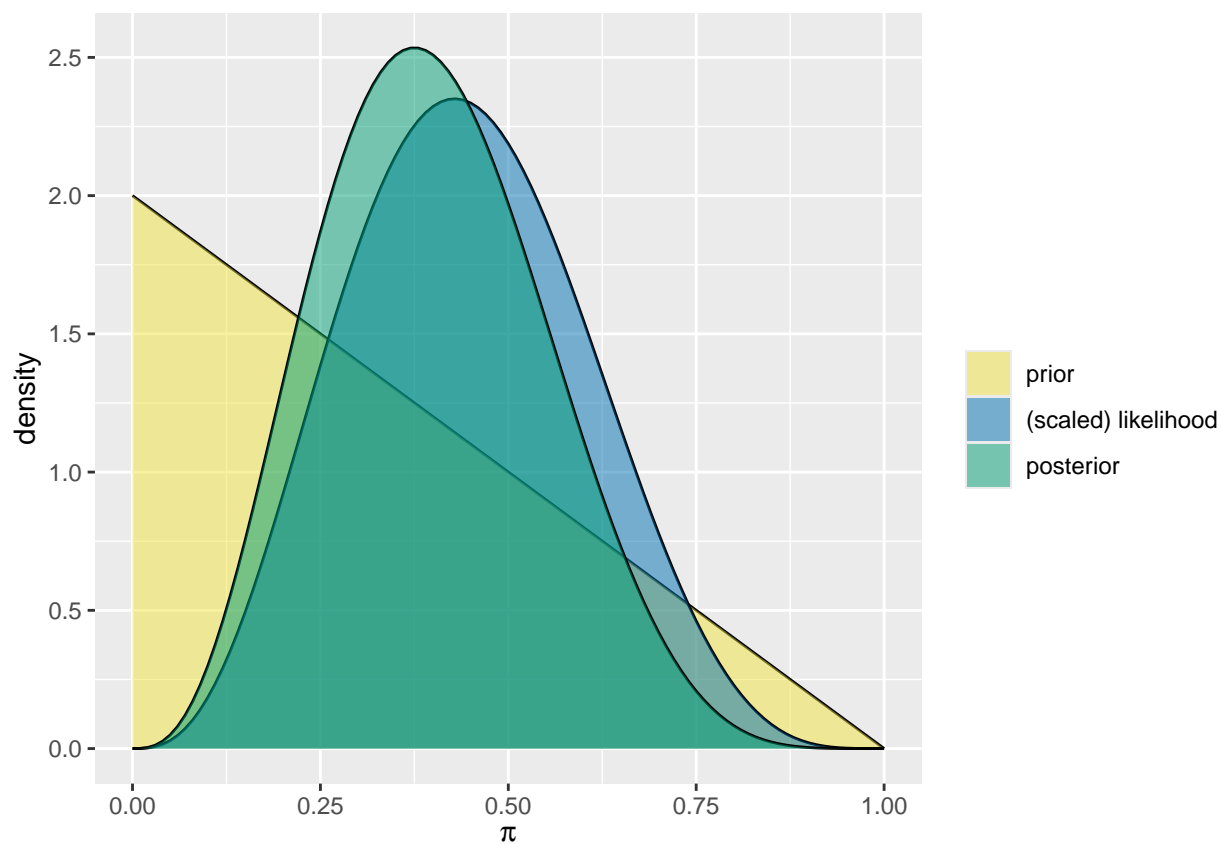
La media es ≈ 0.95 , y la distribución está fuertemente concentrada cerca de 1. Esto indica que Taylor es extremadamente optimista: está casi seguro de que Chad logrará satisfacer su antojo de helado.

Ejercicio 4.5 (Simulando el posterior)

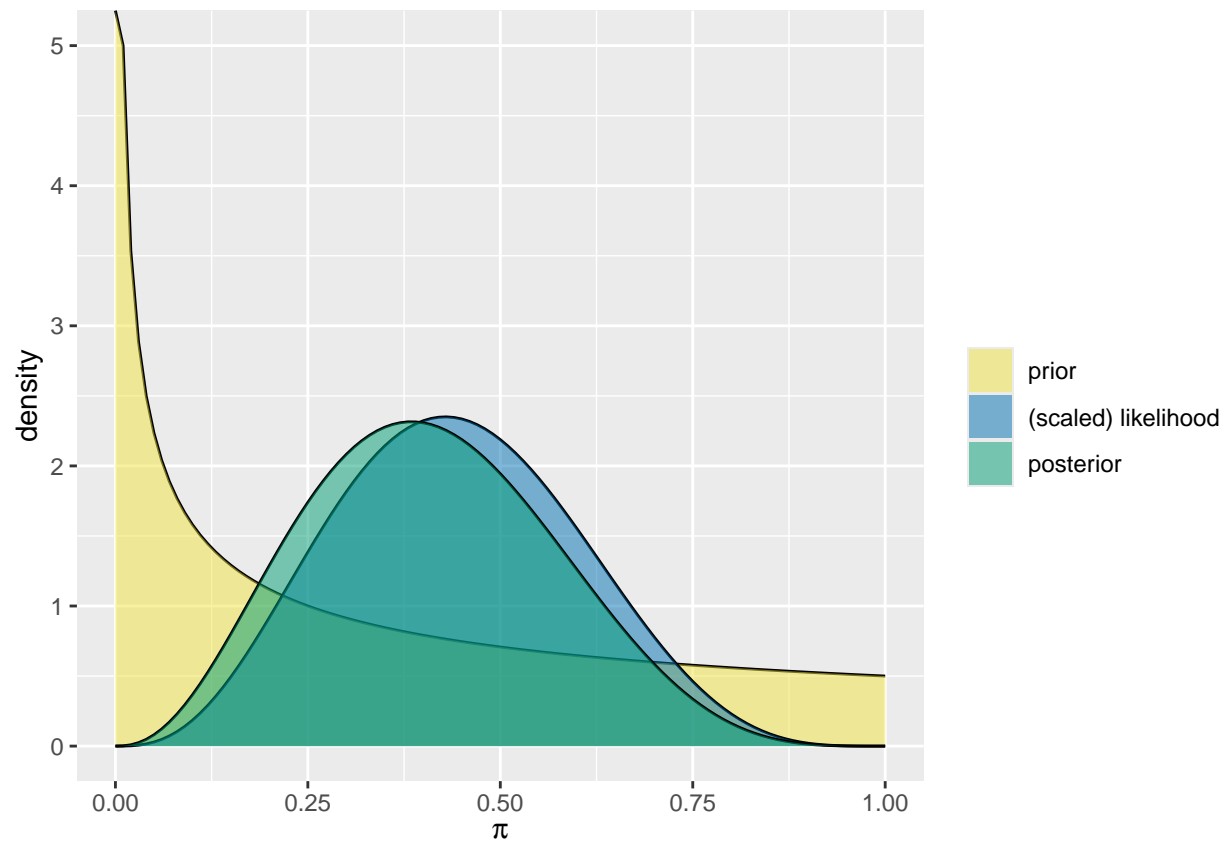
Chad revisa la página web de la tienda. En 3 de los últimos 7 días, todavía estaban abiertos a las 2 p.m. Completa lo siguiente para cada uno de los compañeros de Chad:

1. Simular su modelo posterior

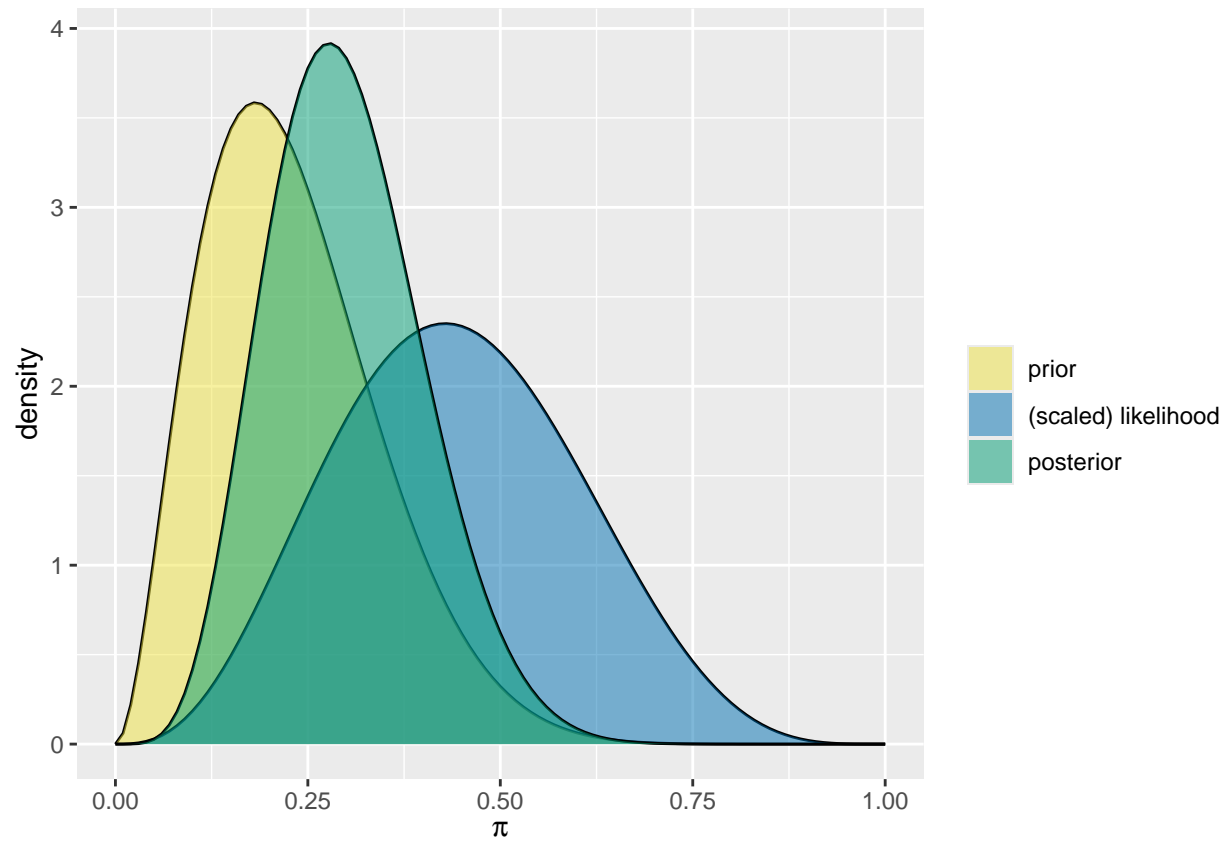
```
plot_beta_binomial(alpha = 1, beta = 2, y = 3, n = 7)
```



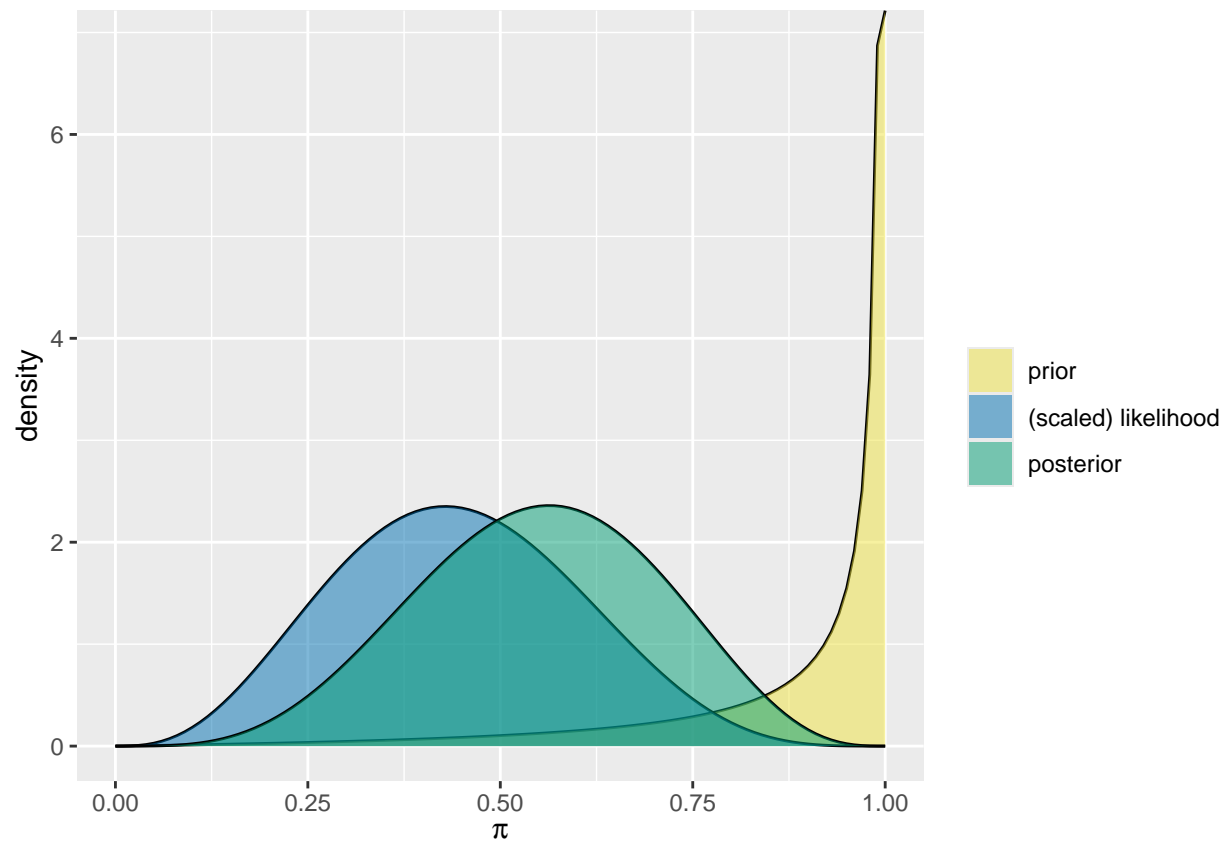
```
plot_beta_binomial(alpha = 0.5, beta = 1, y = 3, n = 7)
```



```
plot_beta_binomial(alpha = 3, beta = 10, y = 3, n = 7)
```



```
plot_beta_binomial(alpha = 2, beta = 0.1, y = 3, n = 7)
```



```
set.seed(1234)
datos <- data.frame(
  amigos = c("Kimya", "Fernando", "Ciara", "Taylor"),
  alpha = c(1, 0.5, 3, 2),
  beta = c(2, 1, 10, 0.1),
  n_obs = 7,
  y_obs = 3
)

simular <- function(tabla, n_simu = 5000) {
  out <- vector("list", nrow(tabla))
  for (i in 1:nrow(tabla)){
    a <- tabla$alpha[i] + tabla$y_obs[i]
    b <- tabla$beta[i] + tabla$n_obs[i] - tabla$y_obs[i]
    y <- rbeta(n_simu, a, b)
    out[[i]] <- data.frame(
      amigos = rep(tabla$amigos[i], length(y)),
      y = y
    )
  }
  reduce(out, rbind)
}

datos_1 <- simular(datos, n_simu = 5000)

datos_1$amigos <- as.factor(datos_1$amigos)
datos_1 %>%
  group_by(amigos) %>%
```

```

summarise(
  n      = n(),
  mean   = mean(y),
  sd     = sd(y),
  min    = min(y),
  q25    = quantile(y, 0.25),
  median = median(y),
  q75    = quantile(y, 0.75),
  max    = max(y),
  .groups = "drop"
)

```

```

## # A tibble: 4 x 9
##   amigos      n mean    sd   min  q25 median  q75   max
##   <fct>   <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Ciara     5000 0.298 0.0993 0.0595 0.226 0.292 0.365 0.772
## 2 Fernando  5000 0.414 0.158 0.0186 0.298 0.406 0.523 0.917
## 3 Kimya     5000 0.401 0.147 0.0383 0.291 0.394 0.504 0.886
## 4 Taylor    5000 0.549 0.159 0.0518 0.437 0.552 0.667 0.946

```

2. Crear un histograma del posterior simulado

```

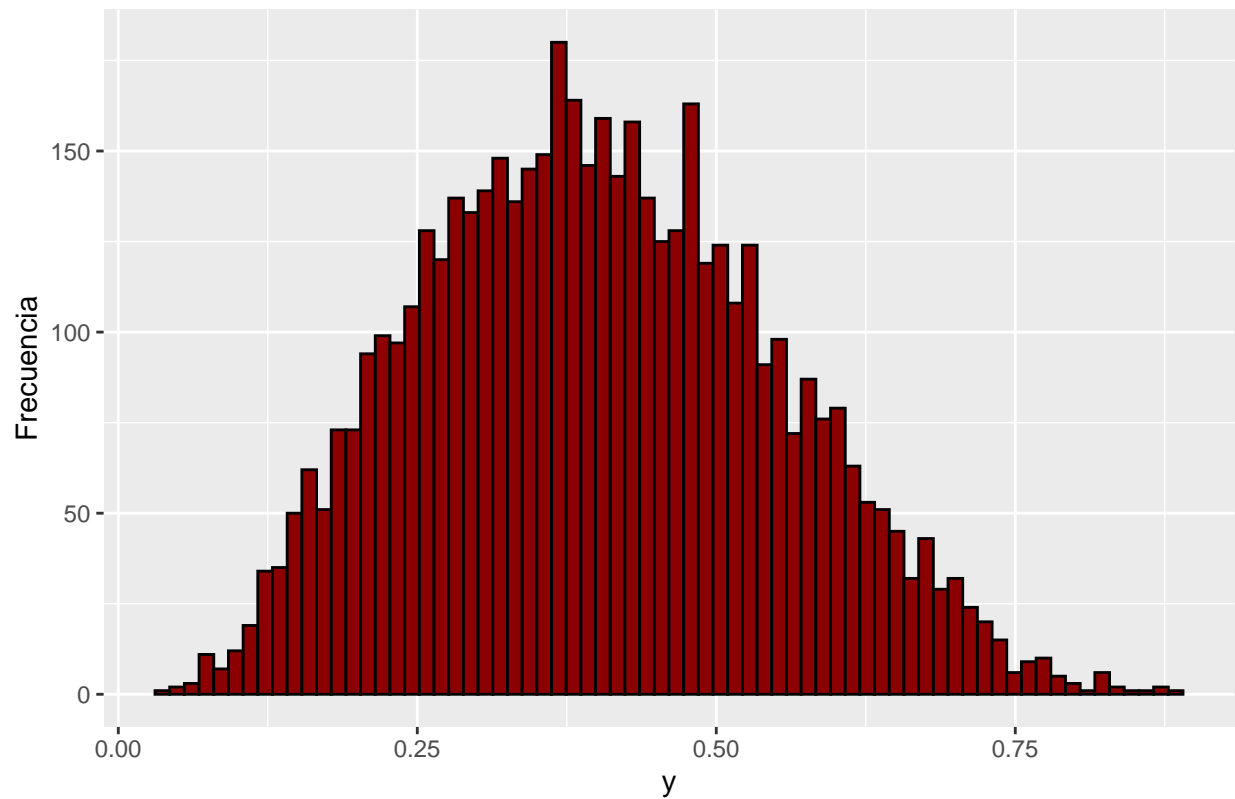
df_k <- datos_1 %>%
  filter(amigos == "Kimya")

bin <- trunc(sqrt(nrow(df_k)))

ggplot(data = df_k, aes(x = y)) +
  geom_histogram(bins = bin, fill = "darkred", color = "black") +
  labs(title = "Posterior para Kimya", x = "y", y = "Frecuencia")

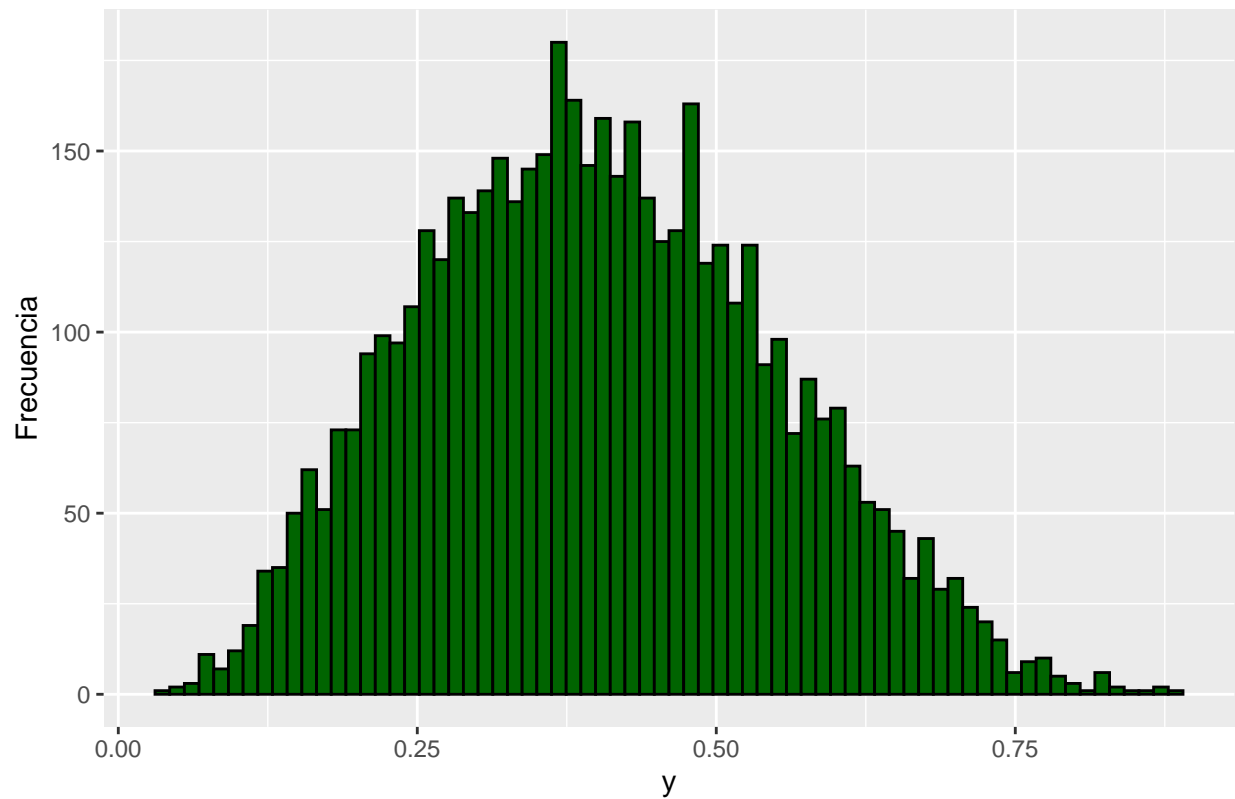
```

Posterior para Kimya



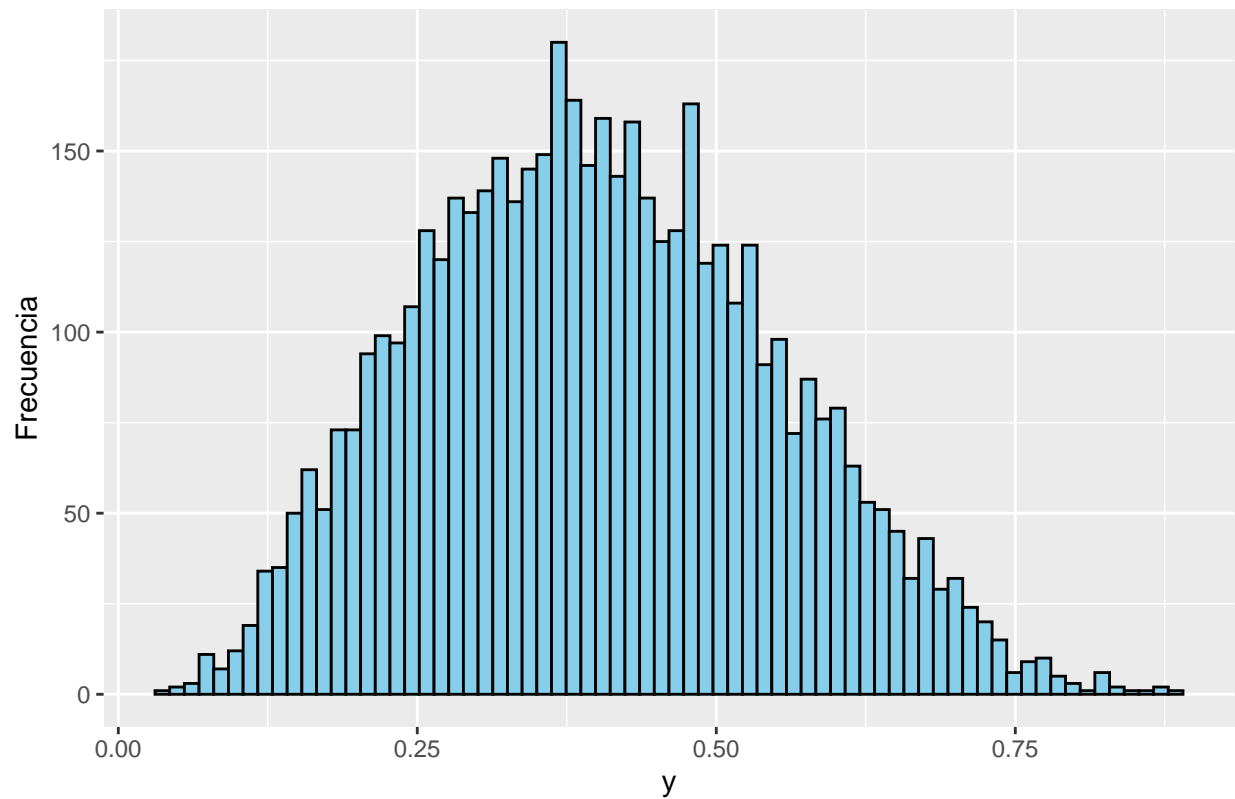
```
df_c <- datos_1 %>%  
  filter(amigos == "Ciara")  
  
bin <- trunc(sqrt(nrow(df_k)))  
  
ggplot(data = df_k, aes(x = y)) +  
  geom_histogram(bins = bin, fill = "darkgreen", color = "black") +  
  labs(title = "Posterior para Ciara", x = "y", y = "Frecuencia")
```


Posterior para Ciara



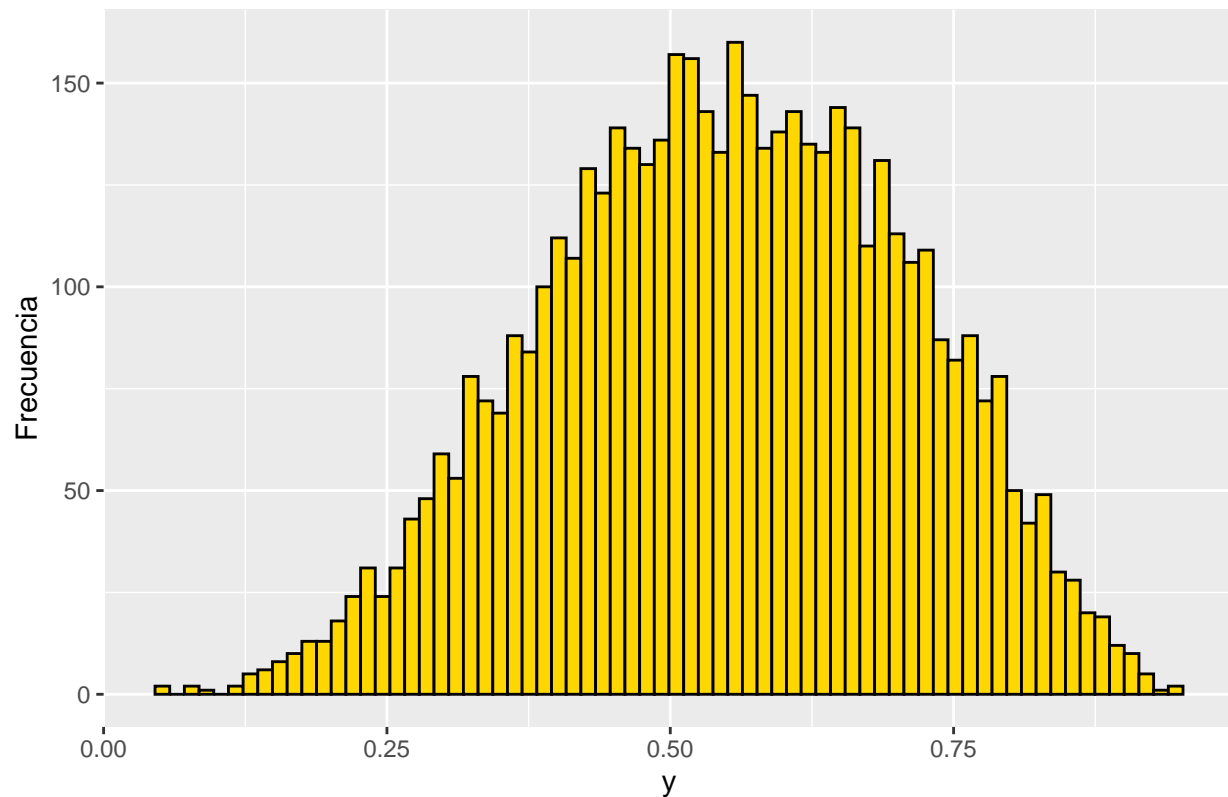
```
df_f <- datos_1 %>%  
  filter(amigos == "Fernando")  
  
bin <- trunc(sqrt(nrow(df_k)))  
  
ggplot(data = df_k, aes(x = y)) +  
  geom_histogram(bins = bin, fill = "skyblue", color = "black") +  
  labs(title = "Posterior para Fernando", x = "y", y = "Frecuencia")
```

Posterior para Fernando



```
df_t <- datos_1 %>%  
  filter(amigos == "Taylor")  
  
bin <- trunc(sqrt(nrow(df_k)))  
  
ggplot(data = df_t, aes(x = y)) +  
  geom_histogram(bins = bin, fill = "gold", color = "black") +  
  labs(title = "Posterior para Taylor", x = "y", y = "Frecuencia")
```

Posterior para Taylor



3. Y usar la simulación para aproximar el valor de la media posterior de π .

```
round(mean(df_k$y), 3)
```

```
## [1] 0.401
```

La media posterior de π de Kimya es 0.401

```
round(mean(df_c$y), 3)
```

```
## [1] 0.298
```

La media posterior de π de Ciara es 0.298

```
round(mean(df_f$y), 3)
```

```
## [1] 0.414
```

La media posterior de π de Fernando es 0.414

```
round(mean(df_t$y), 3)
```

```
## [1] 0.549
```

La media posterior de π de Taylor es 0.549

Ejercicio 4.6: Identificando la posterior —

Completa lo siguiente para cada uno de los compañeros de Chad:

1. Identificar el modelo posterior exacto de π .
2. Calcular la media posterior exacta de π .
3. Comparar estos resultados con los obtenidos por simulación en el ejercicio anterior.

1. Para Kimya su posterior exacta es

$$\pi \mid \text{datos} \sim \text{Beta}(4, 6)$$

```
summarize_beta_binomial(alpha = 1, beta = 2, y = 3, n = 7)
```

```
##      model alpha beta      mean  mode      var      sd
## 1   prior      1    2 0.3333333 0.000 0.05555556 0.2357023
## 2 posterior      4    6 0.4000000 0.375 0.02181818 0.1477098
```

Para Fernando su posterior exacta es

$$\pi \mid \text{datos} \sim \text{Beta}(3.5, 5)$$

```
summarize_beta_binomial(alpha = 0.5, beta = 1, y = 3, n = 7)
```

```
##      model alpha beta      mean  mode      var      sd
## 1   prior    0.5    1 0.3333333 1.0000000 0.08888889 0.2981424
## 2 posterior    3.5    5 0.4117647 0.3846154 0.02549627 0.1596755
```

Para Ciara su posterior exacta es

$$\pi \mid \text{datos} \sim \text{Beta}(6, 14)$$

```
summarize_beta_binomial(alpha = 3, beta = 10, y = 3, n = 7)
```

```
##      model alpha beta      mean  mode      var      sd
## 1   prior      3    10 0.2307692 0.1818182 0.01267963 0.1126039
## 2 posterior      6    14 0.3000000 0.2777778 0.01000000 0.1000000
```

Para Taylor su posterior exacta es

$$\pi \mid \text{datos} \sim \text{Beta}(5, 4.1)$$

```
summarize_beta_binomial(alpha = 2, beta = 0.1, y = 3, n = 7)
```

```
##      model alpha beta      mean  mode      var      sd
## 1   prior      2    0.1 0.9523810 1.0000000 0.01462951 0.1209525
## 2 posterior      5    4.1 0.5494505 0.5633803 0.02451036 0.1565579
```

2. Con respecto a sus medias la exactas para cada amigo es la siguiente: Kimya es 0.400 Fernando es 0.411 Ciara es 0.300 Taylor es 0.549

Amigo	Media simulada	Media exacta
Kimya	0.401	0.400
Fernando	0.298	0.411
Ciara	0.414	0.300
Taylor	0.549	0.549

3. Las medias simuladas resultan muy cercanas a las medias exactas, con diferencias mínimas atribuibles a la variabilidad de la simulación. Esto confirma que el enfoque de simulación reproduce de manera confiable los resultados analíticos

Ejercicio 4.14 (Desafío: modo posterior)

En el esquema **Beta–Binomial**, muestre que podemos escribir el **modo posterior** de π como el **promedio ponderado** entre el **modo previo** y la **tasa de éxitos observada** en la muestra:

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

Pregunta:

¿A qué valor **converge el modo posterior** cuando el tamaño muestral n aumenta indefinidamente? Fundamente su respuesta con evidencia.

¿A qué valor **converge el modo posterior** cuando $n \rightarrow \infty$?

Del resultado

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

se ve que el **peso** del modo previo es $\frac{\alpha + \beta - 2}{\alpha + \beta + n - 2} \rightarrow 0$ y el **peso** de la tasa muestral $\frac{y}{n}$ es $\frac{n}{\alpha + \beta + n - 2} \rightarrow 1$. Por lo tanto,

$$\text{Mode}(\pi \mid Y = y) \xrightarrow{n \rightarrow \infty} \frac{y}{n}.$$

Interpretación. A medida que crece el tamaño muestral, el modo posterior “olvida” el prior y se acerca a la **frecuencia de éxitos observada** y/n .

```
set.seed(123)

# Prior (ejemplo de la bici)
alpha <- 6; beta <- 18

# Verdadero pi para simular datos (solo para mostrar convergencia)
pi_true <- 0.32

# Secuencia de tamaños muestrales crecientes
n_seq <- c(20, 50, 100, 200, 500, 1000, 2000)

res <- lapply(n_seq, function(n){
  y <- rbinom(1, size = n, prob = pi_true)          # datos simulados
  a_post <- alpha + y
  b_post <- beta + n - y
  # modo de Beta(a_post, b_post), definido si a_post>1 y b_post>1
})
```

```

mode_post <- (a_post - 1) / (a_post + b_post - 2)
tibble(
  n = n,
  y = y,
  y_over_n = y/n,
  mode_post = mode_post,
  diff = mode_post - (y/n)
)
})

res_df <- bind_rows(res)
res_df

```

```

## # A tibble: 7 x 5
##       n      y y_over_n mode_post    diff
##   <dbl> <int>   <dbl>   <dbl>   <dbl>
## 1    20      5    0.25    0.238 -0.0119
## 2    50     19    0.38    0.333 -0.0467
## 3   100     27    0.27    0.262 -0.00770
## 4   200     69    0.345    0.333 -0.0117
## 5   500    152    0.304    0.301 -0.00323
## 6  1000    319    0.319    0.317 -0.00197
## 7  2000    649    0.324    0.323 -0.00106

```

```

ggplot(res_df, aes(x = n, y = abs(diff))) +
  geom_line() + geom_point() +
  labs(title = "Diferencia absoluta entre modo posterior y y/n",
        x = "n", y = "|modo posterior - y/n|") +
  theme_minimal()

```

