Entrega_3

Bruno Pintos Juan M Karawcki

2025-09-02

Ejercicio 2.14 (Autobús con retraso)
Li Qiang toma el autobús de las 8:30 a.m. para ir al trabajo todas las mañanas. Si el autobús se retrasa, Li Qiang llegará tarde al trabajo. Para conocer la probabilidad de que su autobús se retrase (π), Li Qiang primero encuesta a 20 compañeros de viaje:

- 3 creen que π es 0,15
- 3 creen que π es 0,25
- 8 creen que π es 0,50
- 3 creen que π es 0,75
- 3 creen que π es 0,85

a) Convierte la información de los 20 viajeros encuestados en un modelo previo para π.

```
n <- c(3,3,8,3,3)
N <- sum(n)

bus <- data.frame(pi = c(0.15, 0.25, 0.50, 0.75, 0.85))
prior <- n / N


 tabla_df <- data.frame(
  "pi"    = bus, #HAY QUE CAMBAIRLO
  "priori" = prior
 )

t(tabla_df)
```

```
##        [,1] [,2] [,3] [,4] [,5]
## pi     0.15 0.25  0.5 0.75 0.85
## priori 0.15 0.15  0.4 0.15 0.15
```

En esta primera parte lo que hicimos fue armar la distribución previa para la probabilidad de que el autobús se retrase. Para eso usamos los datos de la encuesta a los 20 compañeros: cada grupo de personas opinaba un valor distinto para π, y con esas frecuencias sacamos las proporciones. Eso nos permitió transformar la encuesta en una distribución discreta, donde cada valor posible de π tiene su peso según lo que contestó la gente. Básicamente, este paso nos sirvió para poder pasar las opiniones a un modelo previo que después podíamos usar en el análisis bayesiano.

b) Li Qiang quiere actualizar ese modelo previo con los datos que ella misma recopiló: en 13 días, el autobús de las 8:30 a. m. se retrasó en 3 ocasiones. Encuentra el modelo posterior para π.

```
set.seed(1234)
bus_sim <- sample_n(bus, size = 10000, weight = prior, replace = TRUE)

# Simulate 10000 match outcomes
bus_sim <- bus_sim %>%
  mutate(y = rbinom(10000, size = 13, prob = pi))

# Check it out
bus_sim %>%
  head(10)
```

```
##      pi  y
## 1  0.50  5
## 2  0.75 10
## 3  0.75  9
## 4  0.75  7
## 5  0.15  1
## 6  0.75 11
## 7  0.50  6
## 8  0.50  4
## 9  0.75 11
## 10 0.25  3
```

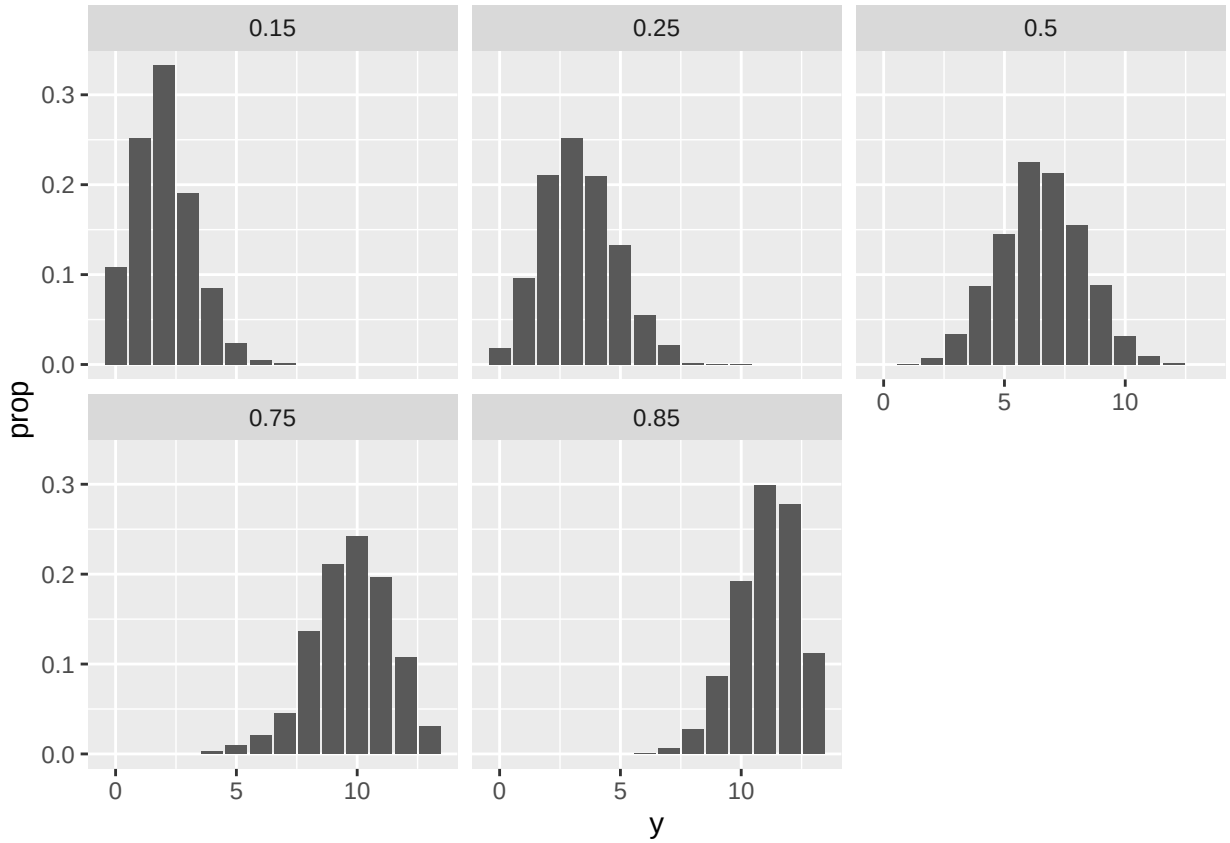
Después decidimos simular la prior, más que nada para asegurarnos de que los pesos que definimos se estaban respetando. Usamos tabyl y adorn_totals para resumir la simulación, y ahí vimos el conteo y porcentaje de cada valor de π. Al mirar los resultados, notamos que los porcentajes coincidían casi perfecto con lo que habíamos calculado antes, así que pudimos confirmar que la simulación estaba funcionando como debía.

```
# Summarize the prior
bus_sim %>%
  tabyl(pi) %>%
  adorn_totals("row")
```

```
##     pi     n percent
##   0.15  1470  0.1470
##   0.25  1552  0.1552
##    0.5  3960  0.3960
##   0.75  1529  0.1529
##   0.85  1489  0.1489
##  Total 10000  1.0000
```

```
ggplot(bus_sim, aes(x = y)) +
  stat_count(aes(y = ..prop..)) +
  facet_wrap(~ pi)
```

```
## Warning: The dot-dot notation (`..prop..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(prop)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

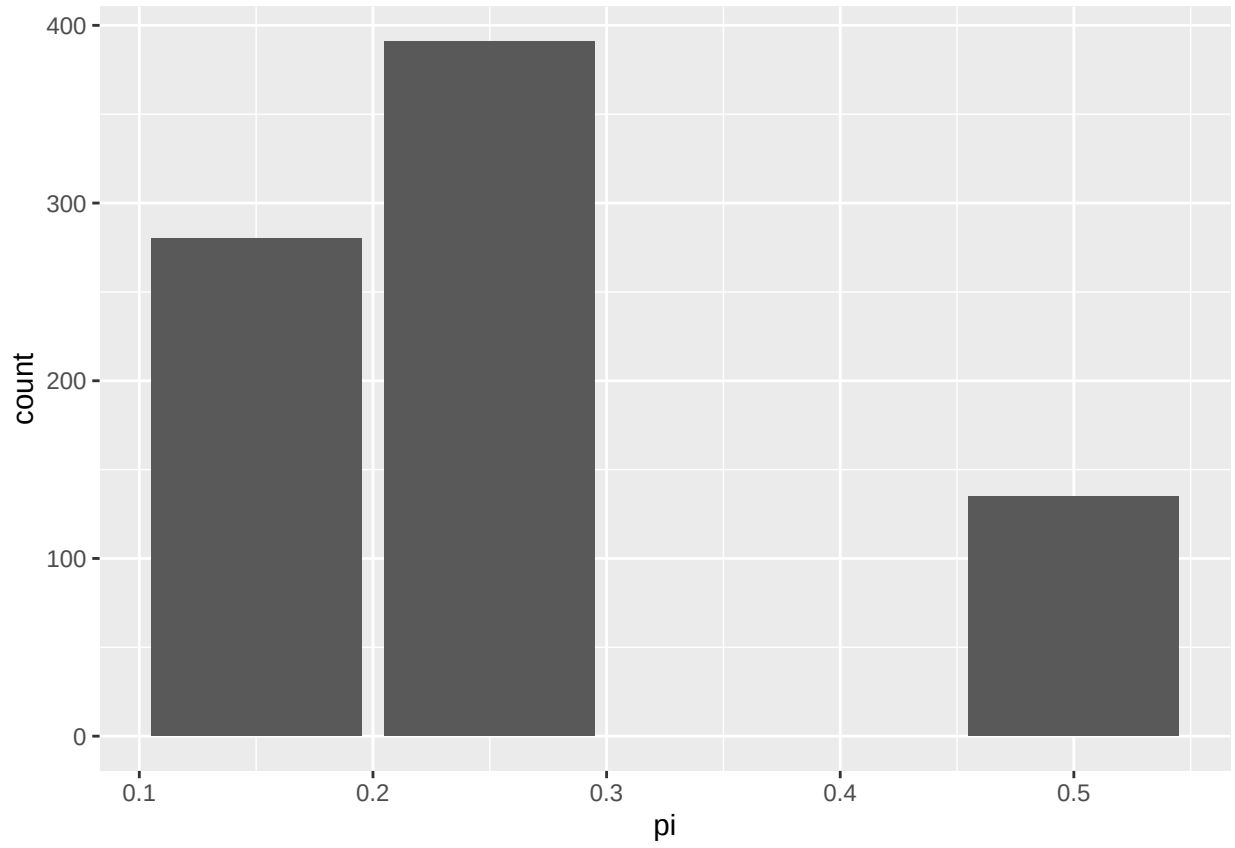


El paso siguiente fue visualizar todo eso. Armamos gráficos con ggplot para ver mejor cómo se distribuyen los valores simulados, y usamos facet_wrap para separar cada panel según el valor de π. De esa forma se puede mirar cada caso por separado y comparar más fácil. Queríamos ver cómo se reparten los valores y qué forma va tomando la prior cuando la simulamos. Con la nueva evidencia de 3 retrasos en 13 veces que fuimos a tomar el bus se aprecia en los gráficos de recién que probabilidades altas (mayores a 0,5) de que el bus se retrase son muy improbables. Por eso, al combinar la evidencia con la prior, construimos una función posterior que ajusta esas creencias iniciales.

```
lose_three <- bus_sim %>%
  filter(y == 3)

lose_three %>%
  tabyl(pi) %>%
  adorn_totals("row")
```

```
##     pi   n   percent
##   0.15 280 0.3473945
##   0.25 391 0.4851117
##    0.5 135 0.1674938
##  Total 806 1.0000000
```

```
ggplot(lose_three, aes(x = pi)) +
  geom_bar()
```

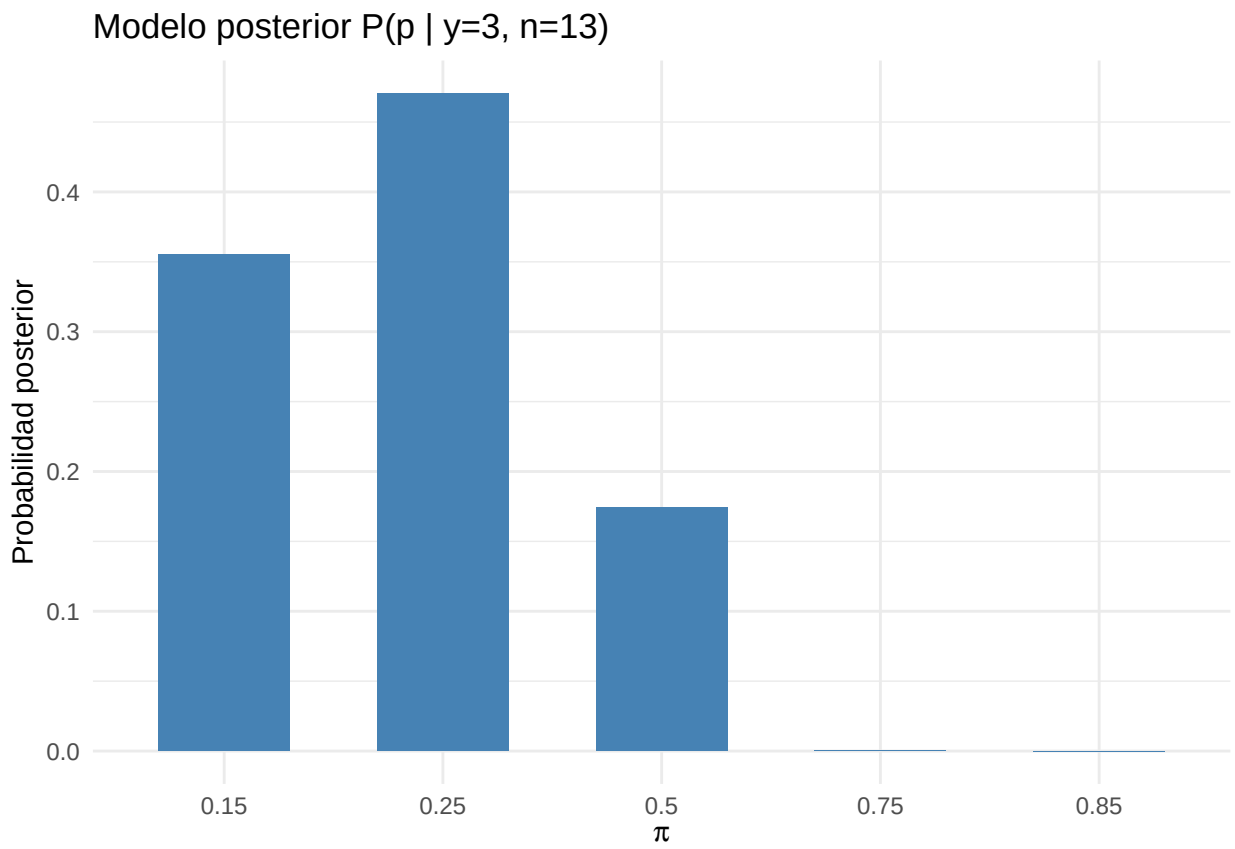



```
# --- Prior discreto desde la encuesta ---


bus <- tibble(
  pi    = c(0.15, 0.25, 0.50, 0.75, 0.85),
  prior = c(3,    3,    8,    3,    3)/20
)

# --- Evidencia ---
n <- 13; y_obs <- 3

# --- Posterior exacta: prior × likelihood y normalizar ---
posterior_df <- bus %>%
  mutate(likelihood = dbinom(y_obs, size = n, prob = pi),
         w          = prior * likelihood,
         posterior  = w / sum(w))

posterior_df
```

```
## # A tibble: 5 x 5
##      pi prior likelihood           w  posterior
##   <dbl> <dbl>      <dbl>       <dbl>      <dbl>
## 1  0.15  0.15 0.190      0.0285      0.355
## 2  0.25  0.15 0.252      0.0377      0.470
## 3  0.5   0.4  0.0349     0.0140      0.174
```

```
## 4  0.75  0.15 0.000115   0.0000173   0.000215
## 5  0.85  0.15 0.00000101 0.000000152 0.00000189
```

```r
# Gráfico de barras (posterior)
ggplot(posterior_df, aes(x = factor(pi), y = posterior)) +
  geom_col(width = 0.6, fill = "steelblue") +
  labs(title = "Modelo posterior P(  | y=3, n=13)",
       x = expression(pi), y = "Probabilidad posterior") +
  theme_minimal()
```



c) Finalmente, compara y comenta los modelos previo y posterior. ¿Qué aprendió Li Qiang sobre el autobús?

Resumen numérico (posterior discreta con $y = 3$ en $n = 13$):

- **Moda (valor más probable):**
  Previo: $\pi = 0{,}50$ (tenía la mayor masa previa, $0{,}40$).
  Posterior: $\pi = 0{,}25$ (pasa a ser el valor más probable).
- **Media (probabilidad esperada de retraso):**
  Previo: $\mathbb{E}[\pi] = 0{,}50$.
  Posterior: $\mathbb{E}[\pi \mid y] \approx 0{,}258$.
- **Masa en escenarios "altos" ($\pi \geq 0{,}5$):**
  Previo: $0{,}70 \rightarrow$ Posterior: $\approx 0{,}174 \Rightarrow$ *gran descenso*.

- **Masa en escenarios "bajos" ($\pi \leq 0{,}25$):**
  Previo: $0{,}30 \rightarrow$ Posterior: $\approx 0{,}826 \Rightarrow$ *gran aumento*.

Interpretación práctica (lo que aprendió Li Qiang):

- Los datos propios (3 retrasos en 13 días) desplazan fuertemente la creencia hacia valores *bajos* de π ($\approx 0{,}15$–$0{,}25$) y descartan en la práctica π *altos* (0,75; 0,85).
- El valor más plausible pasa a ser $\pi = 0{,}25$: el autobús se retrasa aproximadamente 1 de cada 4 días.
- **Predicción posterior:**
  - Probabilidad de retraso "mañana": $\mathbb{E}[\pi \mid y] \approx 0{,}258$.
  - Retrasos esperados en otros 13 días: $13 \times 0{,}258 \approx 3{,}36$ (coherente con lo observado).

En síntesis: la evidencia reduce drásticamente la probabilidad de escenarios con muchos retrasos y sugiere que el autobús se retrasa poco (en torno a un 25%).

- $P(\pi = 0{,}15 \mid y) \approx 0{,}3553$
- $P(\pi = 0{,}25 \mid y) \approx 0{,}4705$
- $P(\pi = 0{,}50 \mid y) \approx 0{,}1740$
- $P(\pi = 0{,}75 \mid y) \approx 0{,}0002$
- $P(\pi = 0{,}85 \mid y) \approx 0{,}000002$

- **Moda (valor más probable):**
  Previo: $\pi = 0{,}50$ (tenía la mayor masa previa, $0{,}40$).
  Posterior: $\pi = 0{,}25$ (pasa a ser el valor más probable).
- **Media (probabilidad esperada de retraso):**
  Previo: $\mathbb{E}[\pi] = 0{,}50$.
  Posterior: $\mathbb{E}[\pi \mid y] \approx 0{,}258$.
- **Masa en escenarios "altos" ($\pi \geq 0{,}5$):**
  Previo: $0{,}70 \rightarrow$ Posterior: $\approx 0{,}174 \Rightarrow$ *gran descenso*.
- **Masa en escenarios "bajos" ($\pi \leq 0{,}25$):**
  Previo: $0{,}30 \rightarrow$ Posterior: $\approx 0{,}826 \Rightarrow$ *gran aumento*.

Interpretación práctica (lo que aprendió Li Qiang):

- Los datos propios (3 retrasos en 13 días) desplazan fuertemente la creencia hacia valores *bajos* de π ($\approx 0{,}15$–$0{,}25$) y descartan en la práctica π *altos* (0,75; 0,85).
- El valor más plausible pasa a ser $\pi = 0{,}25$: el autobús se retrasa aproximadamente 1 de cada 4 días.
- **Predicción posterior:**
  - Probabilidad de retraso "mañana": $\mathbb{E}[\pi \mid y] \approx 0{,}258$.
  - Retrasos esperados en otros 13 días: $13 \times 0{,}258 \approx 3{,}36$ (coherente con lo observado).

π	0.6	0.65	0.7	0.75	**Total**
$f(\pi)$	0.3	0.4	0.2	0.1	**1**

En síntesis: la evidencia reduce drásticamente la probabilidad de escenarios con muchos retrasos y sugiere que el autobús se retrasa poco (en torno a un 25%).

Ejercicio 2.15 (Aves cuco) Las aves cuco son parásitos de cría, lo que significa que ponen sus huevos en los nidos de otras aves (hospedadoras), de modo que sean estas últimas las que críen a las crías del cuco. Lisa es una ornitóloga que estudia la tasa de éxito, π , de las crías de cuco que sobreviven al menos una semana. Ella está retomando el proyecto de un investigador anterior que, en sus notas, especuló el siguiente modelo previo para :

```
# --- Prior discreto desde la encuesta ---


sov <- tibble(
  pi    = c(0.6, 0.65, 0.7, 0.75),
  prior = c(0.3, 0.4, 0.2, 0.1)
)
```

Nuevamente iniciamos el ejercicio definiendo el modelo previo discreto: un vector con los posibles valores de π y otro con sus probabilidades previas asociadas. Con ambos construimos el tibble sov, que contiene las columnas pi (valores de π) y prior (probabilidades del previo).

```
set.seed(1234)
sov_sim <- sample_n(sov, size = 10000, weight = prior, replace = TRUE)

# Simulate 10000 match outcomes
sov_sim <- sov_sim %>%
  mutate(y = rbinom(10000, size = 15, prob = pi))

# Check it out
sov_sim %>%
  head(10)
```

```
## # A tibble: 10 x 3
##       pi prior     y
##    <dbl> <dbl> <int>
##  1  0.65   0.4    11
##  2  0.6    0.3     9
##  3  0.6    0.3     8
##  4  0.6    0.3     6
##  5  0.7    0.2    12
##  6  0.6    0.3    10
##  7  0.65   0.4    10
##  8  0.65   0.4    12
##  9  0.6    0.3    10
## 10  0.6    0.3     9
```

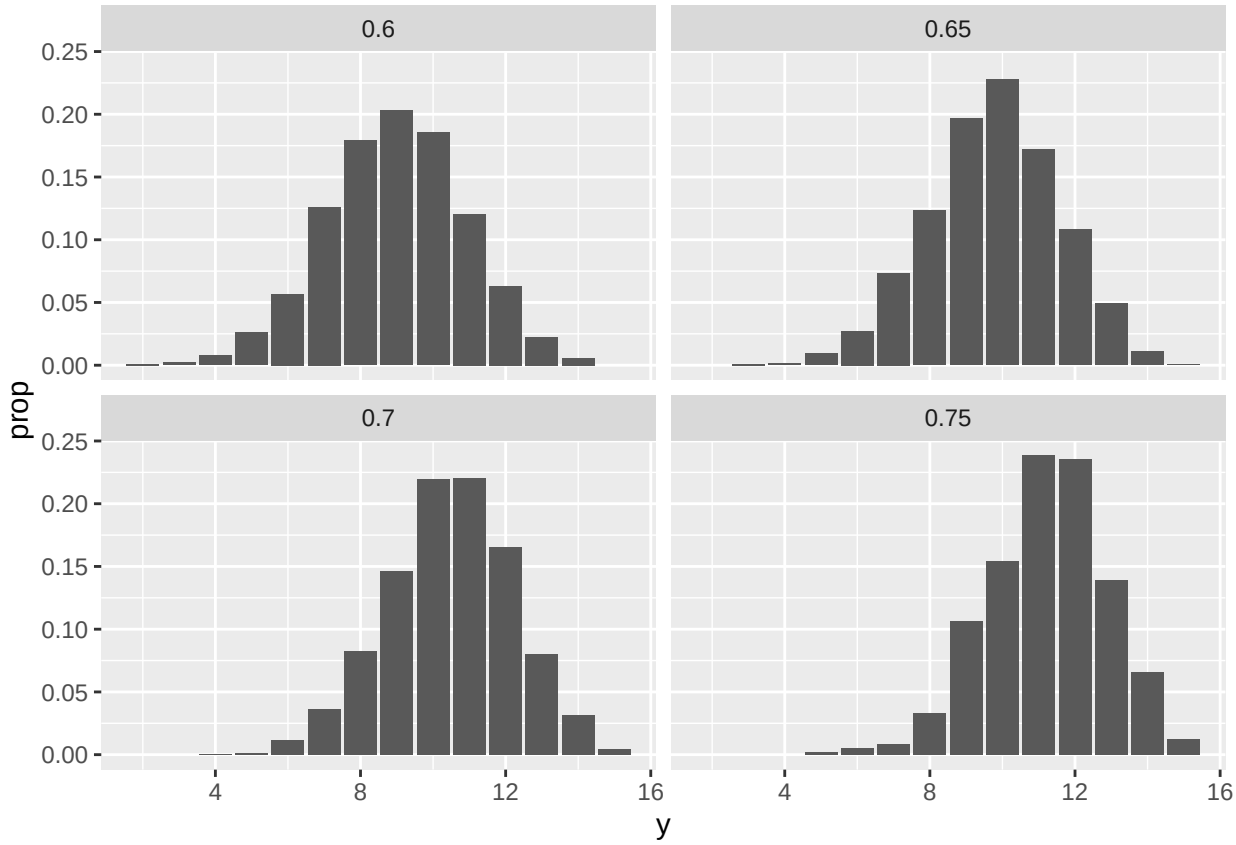
En este fragmento se simula la extracción de 10.000 valores mediante una muestra aleatoria con reposición, utilizando como pesos las probabilidades previas asignadas a cada valor discreto de π. De esta forma, se genera una distribución simulada.

Posteriormente, con mutate se incorpora al data frame una variable y, obtenida a partir de una distribución binomial con $n = 15$ ensayos y probabilidad de éxito π. Esto se repite 10.000 veces, de modo que cada observación contiene tanto el valor simulado de π como el correspondiente número de supervivientes (y) en 7 días.

```
# Summarize the prior
sov_sim %>%
  tabyl(pi) %>%
  adorn_totals("row")
```

```
##      pi     n percent
##     0.6  3081  0.3081
##    0.65  3960  0.3960
##     0.7  1973  0.1973
##    0.75   986  0.0986
##   Total 10000  1.0000
```

```
ggplot(sov_sim, aes(x = y)) +
  stat_count(aes(y = ..prop..)) +
  facet_wrap(~ pi)
```



Dado que el previo concentra masa en valores altos de π, si la evidencia empírica mostrara un conteo de supervivencias bajo en n nidos (por ejemplo, un y pequeño), esos π altos serían poco verosímiles y perderían peso en la posterior. Por el contrario, si el dato observado fuera un conteo alto de supervivencias, reforzaría los π cercanos a 0,70–0,75 y desplazaría la masa posterior hacia esos valores.
En ningún caso corresponde concluir "supervivencia nula"; lo correcto es hablar de mayor o menor verosimilitud del dato bajo cada π

Si el investigador hubiese estado *más seguro* de que una cría sobreviviría Su creencia previa tendría **mayor media** (más probabilidad en valores altos de π) y **menor dispersión** (más concentrada). Por ejemplo:

$f(\pi)$ más optimista y concentrado:

π	0.60	0.65	0.70	0.75	**Total**
$f(\pi)$	0.05	0.20	0.45	0.30	**1**

Este prior tiene media ≈ 0.70 (antes era 0.655) y menos varianza.

Si el investigador hubiese estado *menos seguro* de que una cría sobreviviría Hay dos interpretaciones razonables:

1. **Menos optimista (menor probabilidad de supervivencia):** se desplaza la masa hacia π más bajos (media menor).

π	0.60	0.65	0.70	0.75	**Total**
$f(\pi)$	0.50	0.30	0.15	0.05	**1**

   Media ≈ 0.638, claramente menor que la original 0.655.

2. **Menos confiado (más incertidumbre):** se hace el prior *más difuso* (mayor dispersión, menos preferencia entre valores).

π	0.60	0.65	0.70	0.75	**Total**
$f(\pi)$	0.25	0.25	0.25	0.25	**1**

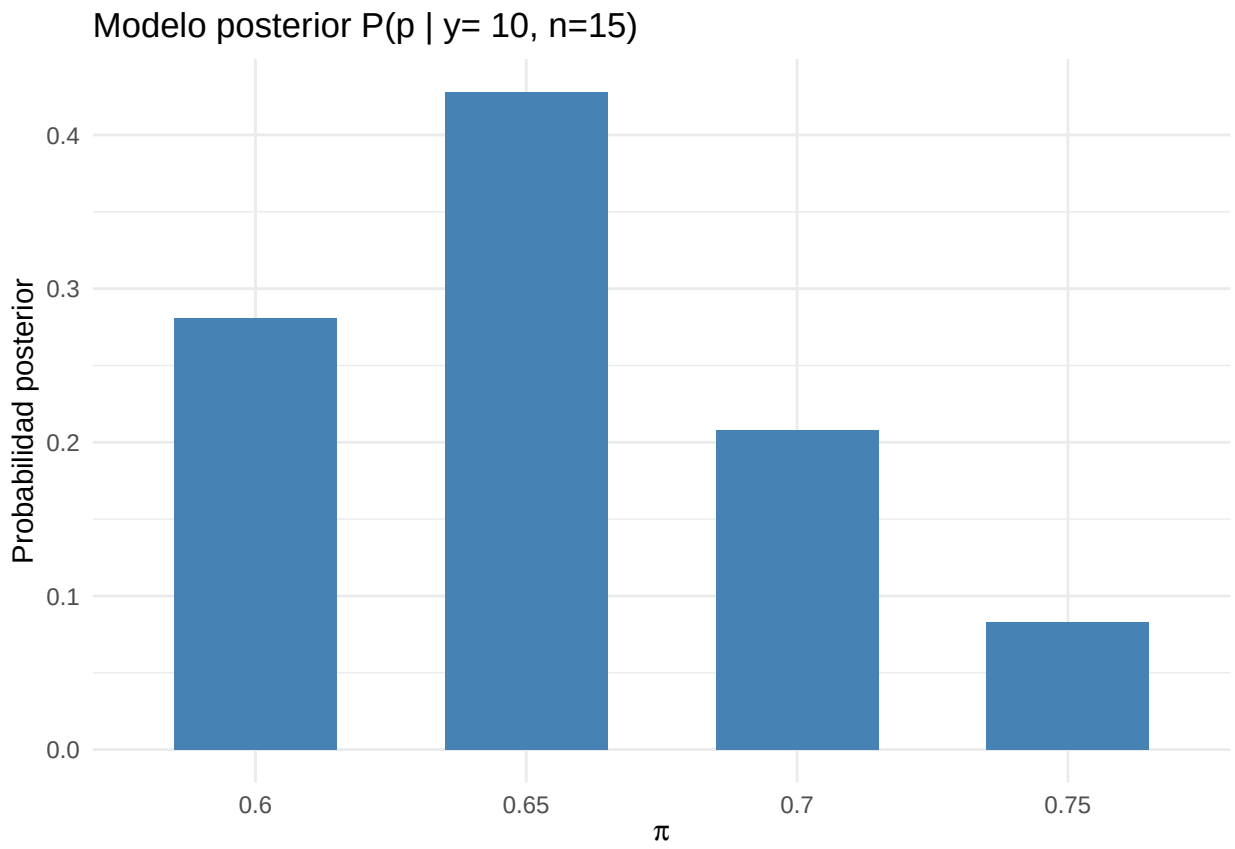
   Este prior expresa "no tengo una preferencia fuerte" entre los valores posibles (más plano).

c) Lisa recoge algunos datos. Entre las 15 crías que estudió, 10 sobrevivieron al menos una semana. ¿Cuál es el modelo posterior para ?

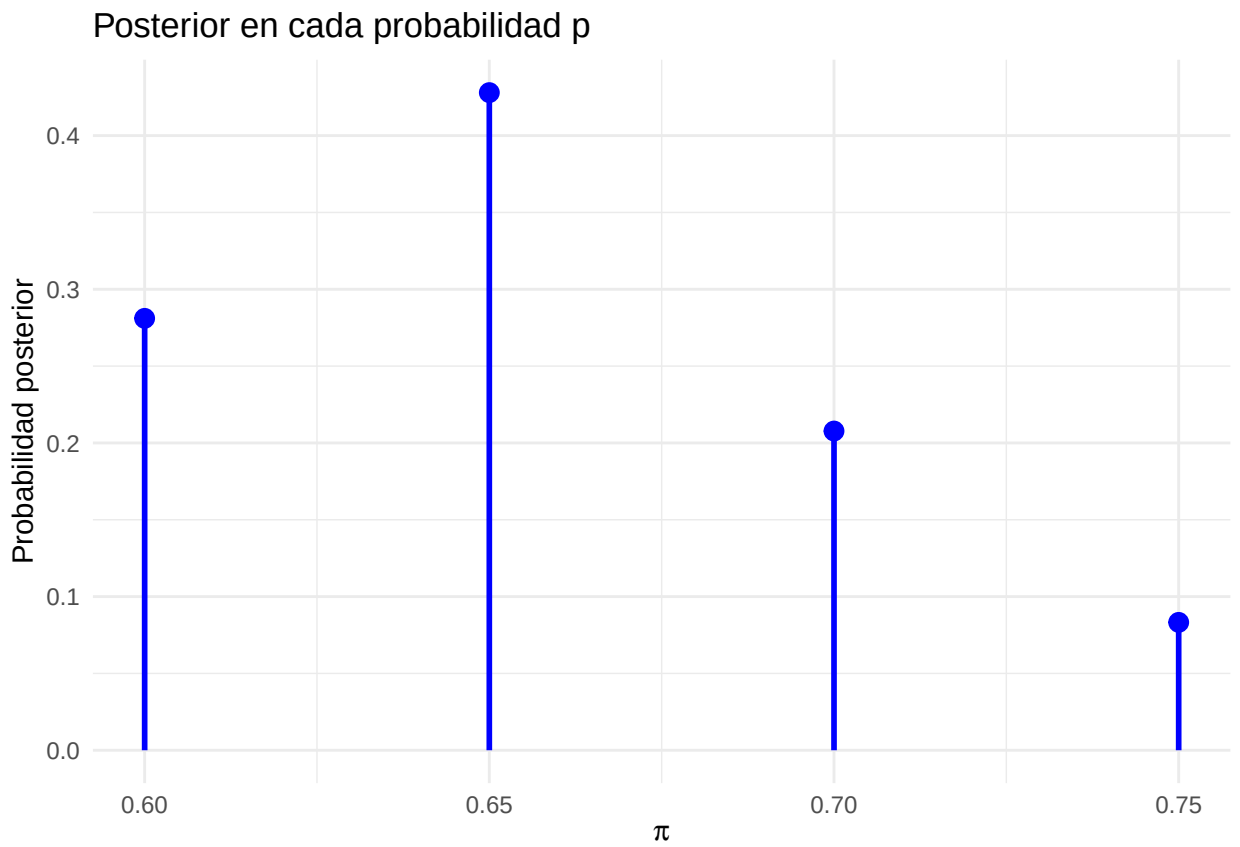
```
# Esta parte sirve tambien para el ejercicio 2.19

# --- Evidencia ---
n <- 15; y_obs <- 10

# --- Posterior exacta: prior × likelihood y normalizar ---
posterior_df <- sov %>%
  mutate(likelihood = dbinom(y_obs, size = n, prob = pi),
         w          = prior * likelihood,
         posterior  = w / sum(w))

posterior_df
```

```
## # A tibble: 4 x 5
##      pi prior likelihood      w posterior
##   <dbl> <dbl>      <dbl>  <dbl>     <dbl>
## 1  0.6    0.3      0.186 0.0558    0.281
## 2  0.65   0.4      0.212 0.0849    0.428
## 3  0.7    0.2      0.206 0.0412    0.208
## 4  0.75   0.1      0.165 0.0165    0.0832
```

```r
# Gráfico de barras (posterior)

ggplot(posterior_df, aes(x = factor(pi), y = posterior)) +
  geom_col(width = 0.6, fill = "steelblue") +
  labs(title = "Modelo posterior P(  | y= 10, n=15)",
       x = expression(pi), y = "Probabilidad posterior") +
  theme_minimal()
```




```r
ggplot(posterior_df, aes(x = pi, y = posterior)) +
  geom_point(size = 3, color = "blue") +
  geom_segment(aes(x = pi, xend = pi, y = 0, yend = posterior),
               color = "blue", linewidth = 1) +
  labs(title = "Posterior en cada probabilidad  ",
       x = expression(pi), y = "Probabilidad posterior") +
  theme_minimal()
```



Lo que estás haciendo es construir la distribución posterior de π a partir de un prior discreto y la verosimilitud binomial observando $y_{\text{obs}} = 10$ éxitos en ; $n = 15$ ensayos. Primero se calculan las probabilidades posteriores exactas como prior $\times$ likelihood, normalizadas. Luego se representan con dos gráficos: uno de barras que muestra la probabilidad posterior en cada valor discreto de π, y otro de segmentos verticales que resalta la masa posterior en cada π

d) Lisa necesita explicar el modelo posterior para en un artículo de investigación para ornitólogos, y no puede asumir que ellos entienden estadística bayesiana. Resume brevemente el modelo posterior en contexto.

Lo que hicimos fue arrancar con el prior que nos dejó el investigador anterior, donde ya había una idea previa de que la probabilidad de supervivencia de las crías de cuco estaba más o menos en 0.65, aunque también se consideraban 0.60, 0.70 y 0.75. Para entender mejor ese prior, lo transformamos en una simulación grande de valores posibles y generamos resultados ficticios de 15 crías, lo que nos permitió ver cómo se comportaba la distribución y asegurarnos de que realmente reflejaba lo que decía la tabla original. Después revisamos y graficamos esos resultados para tener una idea visual de qué tan probable era cada escenario antes de mirar los datos. Finalmente, metimos en juego la evidencia real que recolectó Lisa (10 sobrevivientes de 15), filtramos la simulación con ese resultado y obtuvimos así el posterior: una distribución actualizada que combina lo que pensábamos antes con lo que muestran los datos. Lo más fuerte que salió de todo este proceso es que el valor más probable de la tasa de supervivencia es 0.65, pero también hay bastante apoyo para 0.60 y 0.70, mientras que 0.75 queda como una opción más débil. En otras palabras, después de juntar todo, podemos decir que la supervivencia de las crías de cuco ronda el 65%, con algo de margen hacia abajo o hacia arriba.

Al final de todo este trabajo, lo que podemos decir es que la probabilidad de que una cría de cuco sobreviva al menos una semana está, en promedio, cerca del 65%. Eso fue lo que salió cuando juntamos lo que ya se sospechaba antes con los datos nuevos que recolectamos. Lo más probable es que el valor real esté entre 60% y 70%, y aunque existe una pequeña chance de que llegue al 75%, esa opción quedó bastante menos respaldada por la evidencia. Dicho de otra forma: los datos que conseguimos no contradicen lo que se pensaba antes, más bien lo refuerzan, porque tanto el prior como lo que vimos en el campo apuntan a una supervivencia típica en torno a dos de cada tres crías.

O sea, a modo de ejemplificar, podemos decir que si soltamos un grupo de pichones de cuco, lo más razonable sería esperar que sobrevivan unos 10 de cada 15, como le pasó a … en el estudio. Obviamente puede variar un poco, a veces sobreviven menos y a veces más, pero ese es el escenario más realista.

Lo que encontramos es que la idea previa de que las crías tienen bastantes chances de superar la primera semana se sostiene, aunque no es una garantía absoluta.