Tarea Corta Nº2

Estudiante Andreas Laffert

miércoles, 28 de agosto de 2024

La final del Mundial 2006 – Italia vs. Francia – se definió por penales. La tabla a continuación resume la información de la ronda de penales. La columna X registra los resultados de cada lanzamiento, donde $X_i = 1$ indica que el jugador i convirtió el penal y $X_i = 0$ indica que el penal fue atajado o perdido.

Jugador	Equipo	X: Resultado
Pirlo	Italia	1
Wiltord	Francia	1
Materazzi	Italia	1
Trezeguet	Francia	0
De Rossi	Italia	1
Abidal	Francia	1
Del Piero	Italia	1
Sagnol	Francia	1
Grosso	Italia	1

Asumiendo que el tiro de cada jugador no es afectado por los resultados en los tiros anteriores, es razonable sostener que el resultado de cada penal sigue una distribución Bernoulli. Formalmente: $X_i \sim \text{Bernoulli}(p_i)$, donde p_i es la probabilidad de que cada jugador marque su penal.

Preguntas:

1) Expresa la función de probabilidad de cada variable aleatoria X_i.

Una variable aleatoria es una variable cuyos resultados posibles son resultados númericos de un experimento aleatorio. Por su parte, una distribución de una variable aleatoria corresponde a una función de densidad que permite asignar un conjunto probabilidades de ocurrencia asociados a cada valor posible de la variable aleatoria. Aunque cada X_i es una variable aleatoria asociada a cada jugador i y su valor obtenido (1 = convertir gol, 0 = errar), podemos considerar a X_i como una única variable aleatoria con una distribución Bernoulli con probabilidad p_i.

La función de probabilidad de una variable aleatoria X_i que sigue una distribución Bernoulli corresponde a:

$$f(x) = p^x(1-p)^{x-1},$$

si $x = 1$ o $x = 0$

La función de verosimilitud ($\mathcal{L}$) se utiliza para encontrar el valor más plausible de un parámetro de una distribución teórica dado un conjunto de observaciones. En términos simples, permite maximizar la probabilidad de que una distribución teórica ajuste a los datos observados.

Primero, asumimos que la distribución subyacente a los resultados observados en la tanda de penales para ambos equipos sigue una distribución Bernoulli: $X_i \sim$ Bernoulli(p_i), donde p_i es la probabilidad de que cada jugador marque su penal.

Segundo, definimos la función de verosimilitud. En una distribución Bernoulli, esta función se expresa como:

$$\mathcal{L}(p \mid \text{Datos}) = p^k(1-p)^{n-k},$$

donde:

- k es el número de penales convertidos (resultados exitosos = 1).
- n es el número total de penales cobrados.

Tercero, expresamos esta función para q en el caso de Francia y para p en el caso de Italia, utilizando sus valores observados:

$$\mathcal{L}(q) = q^3(1-q)^{4-3} = q^3(1-q)$$

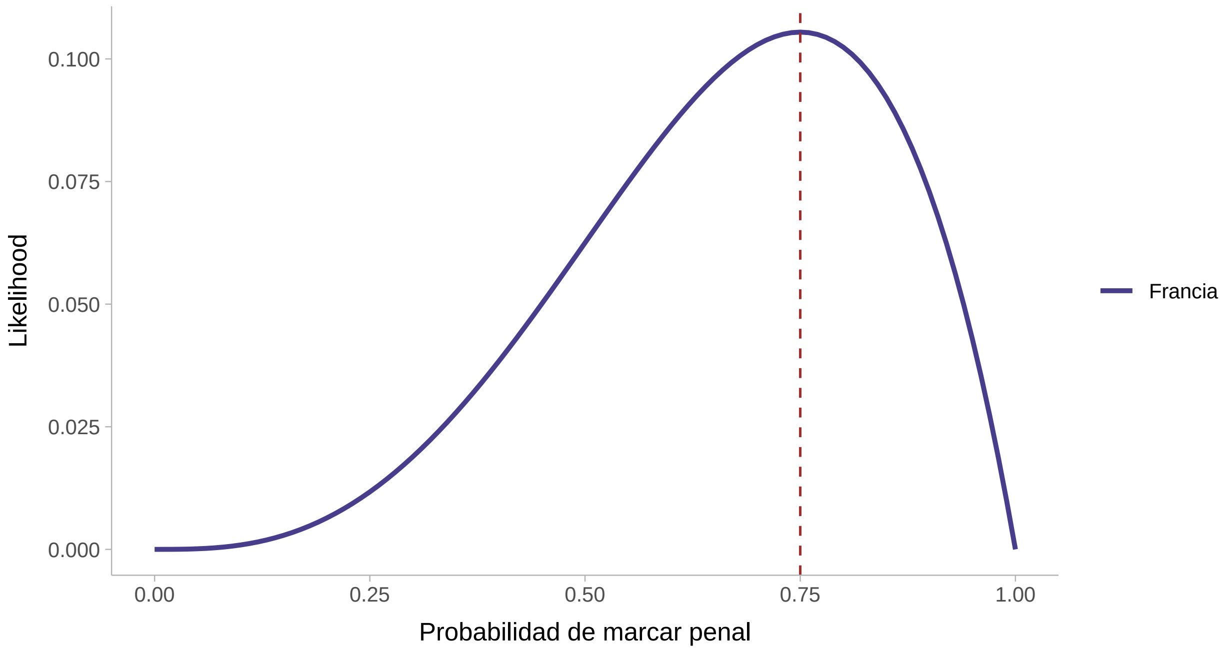
$$\mathcal{L}(p) = p^5(1-p)^{5-5} = p^5$$

En este caso, para Francia, $n = 4$ y $k = 3$, mientras que Italia cobró todos los $n = 5$ penales y los convirtió, es decir, $k = 5$.

3) Grafica la función $\mathcal{L}(q)$ (q en el eje-x y $\mathcal{L}(q)$ en el eje y).

4) Estima visualmente cuál es el valor de q que maximiza $\mathcal{L}(q)$. Explica el significado de éste número.

Figura 1: Likelihood function de q



De acuerdo con la Figura 1, podemos determinar que el valor de q que maximiza la verosimilitud para la probabilidad de marcar un gol del equipo francés es 0.75, ya que es el punto máximo de la distribución, donde la pendiente o derivada en el eje y es igual a 0. Este valor de 0.75 representa la probabilidad que maximiza la plausibilidad de ajustar una distribución Bernoulli a estos datos observados.

5) Supón la siguiente situación: Es sabido que $P(\text{marcar penal} \mid \text{Francia}) = 0.75$. Un fanático nervioso se resiste a ver la transmisión del primer penal y tampoco sabe qué equipo empieza pateando. Pasados unos segundos, escucha gritos de celebración indicando que el penal fue convertido. El pobre tifoso recuerda que puede usar el Teorema de Bayes para conocer la probabilidad de que el penal haya sido marcado por un jugador francés — $P(\text{Francia} \mid \text{marcar penal})$ — pero no recuerda cómo hacerlo. Realiza tú el cálculo requerido.

Recordando el Teorema de Bayes, tenemos que la $P(A \mid B)$ es:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Y que la $P(B \mid A)$ es:

$$P(B \mid A) = \frac{P(A \mid B) * P(B)}{P(A)}$$

Entonces, para calcular $P(\text{marcar penal} \mid \text{Francia}) = 0.75$ con el Teorema de Bayes, tenemos que:

$$P(\text{Francia} \mid \text{marcar penal}) = \frac{P(\text{marcar penal} \mid \text{Francia}) * P(\text{Francia})}{P(\text{marcar penal})}$$

Donde definimos que:

- $P(\text{marcar penal} \mid \text{Francia}) = 0.75$ es la probabilidad de que un jugador francés marque un penal.
- $P(\text{Francia}) = 0.5$ es la probabilidad de que el penal haya sido cobrado por Francia porque no sabemos cuál equipo comenzó según las instrucciones (50% chances Francia y 50% chances Italia).

Resolviendo tenemos que:

$$P(\text{Francia} \mid \text{marcar penal}) = \frac{0.75 * 0.5}{0.75} = 0.5$$

Así, la probabilidad de que el penal haya sido convertido por Francia es del 50%.

Código de R

```
knitr::opts_chunk$set(echo = F,
                      warning = F,
                      error = F,
                      message = F)
if (! require("pacman")) install.packages("pacman")

pacman::p_load(tidyverse,
               rio,
               here,
               ggdist)

options(scipen=999)
rm(list = ls())
datos <- data_frame(q = seq(from = 0, to = 1, by=0.01),
                    L_francia = q^3*(1-q))

max_francia <- datos %>%
  slice_max(L_francia) %>%
  pull(q)

datos %>%
  ggplot(aes(x = q, y = L_francia, color = "Francia")) +
  geom_line(linewidth = 1) +
  geom_vline(aes(xintercept = max_francia), linetype = "dashed", color = "brown")+
  scale_color_manual(values = "#483d8b") +
  labs(y = "Likelihood",
       x = "Probabilidad de marcar penal",
       colour = NULL) +
  theme_ggdist()
```