SOC3070 Análisis de Datos Categóricos

Tarea corta 3

Ponderación: 6% de la nota final del curso.

En esta pregunta aplicarán propiedades de distribuciones de probabilidad discreta y Maximum Likelihood Estimation.

Problema:

La última final de la Eurocopa – Italia vs. Inglaterra – 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
Berardi	Italia	1
Kane	Inglaterra	1
Belotti	Italia	0
Maguire	Inglaterra	1
Bonucci	Italia	1
Rashford	Inglaterra	0
Bernardeschi	Italia	1
Sancho	Inglaterra	0
Jorginho	Italia	0
Saka	Inglaterra	0

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) Exprese la función de probabilidad de cada variables aleatoria X_i.

$$X_i \sim p_i^{x_i}(1-p_i)^{1-x_i}$$

2) Asumiendo que todos los jugadores de un mismo equipo tienen la misma probabilidad de marcar su penal (p para Italia y q para Inglaterra), exprese la "Likelihood function" de p y q.

$$\mathcal{L}(p) = \mathbb{P}(x_{\text{Berardi}}, x_{\text{Belotti}}, x_{\text{Bonucci}}, x_{\text{Bernardeschi}}, x_{\text{Jorginho}} \mid p) = p^3(1-p)^2$$

y

$$\mathcal{L}(q) = \mathbb{P}(x_{\text{Kane}}, x_{\text{Maguire}}, x_{\text{Rashford}}, x_{\text{Sancho}}, x_{\text{Saka}} \mid q) = q^2(1-q)^3$$

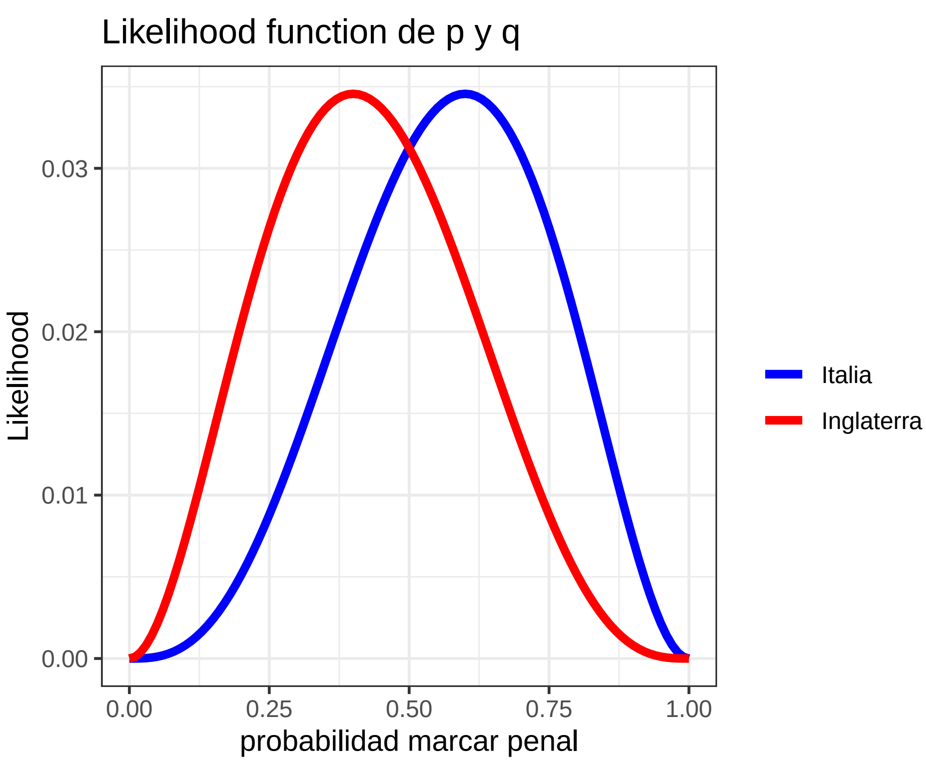
3) Grafique las funciónes $\mathcal{L}(p)$ y $\mathcal{L}(q)$ (p/q en el eje-x y $\mathcal{L}(p)/\mathcal{L}(q)$ en el eje y).

Para hacerlo, debes reemplazar las variables `L_italia` y `L_inglaterra` en el siguiente código por las respectivas expresiones obtenidas en la pregunta (2). Aquí `L_italia = sin(-3*p)` y `L_inglaterra = cos(-2*q)` sirven sólo a modo de ejemplo.

```
mydata <- data_frame(p = seq(from = 0, to = 1, by=0.01), q = seq(from = 0, to = 1, by=0.01),
                    L_italia = (p^3)*((1-p)^2),
                    L_inglaterra = (q^2)*((1-q)^3) )

plot <- ggplot(data = mydata) +
        geom_path(aes(x = p, y=L_italia, colour="Italia"), size=1.5) +
        geom_path(aes(x = q, y=L_inglaterra, colour="Inglaterra"), size=1.5) +
        labs(y="Likelihood", x="probabilidad marcar penal", title="Likelihood function de p y q") +
        scale_color_manual("",breaks=c("Italia","Inglaterra"), values=c("blue","red")) +
        theme_bw()

print(plot)
```



4) Estime visualmente cuál es el valor de p que maximiza $\mathcal{L}(p)$ y cuál es el valor de q que maximiza $\mathcal{L}(q)$. Explique el significado de estos números.

El valor de p que maximiza la "likelihood function" es $p = 0.6$. Es decir, estimamos que el escenario en que cada jugador italiano tiene una probabilidad de 0.6 de convertir su penal es el más "plausible" dado los datos de los que disponemos. Del mismo modo, el escenario en que cada jugador inglés tiene una probabilidad de 0.4 de convertir su penal es el más "plausible" dado el desenlace de la ronda de penales.