SOC3070 Análisis de Datos Categóricos

Tarea corta 3, respuestas

Ponderación: 6% de la nota final del curso.

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

Problema:

Tres personas pescan en un lago. El lago es grande y los peces están uniformemente distribuidos en el agua. Cada pescador está solo y alejado del resto, de tal manera que los resultados de cada pescador no afectan los resultados de otros. Cada pescador tira su caña tres veces, pero sus habilidades de pesca no cambian entre cada intento (para cada pescador, los resultados de un intento son independientes de los otros intentos).

La siguiente tabla resume los resultados de la jornada:

ID Pescador	X: Números de pescados
Pescador 1	2
Pescador 2	2
Pescador 3	1
Pescador 4	3

Dada la descripción de la situación es razonable sostener que, para cada pescador $i = 1, \ldots 4$, $X_i \sim \text{Binomial}(n = 3, p)$, donde p es la probabilidad de sacar un pescado en cada intento. p es desconocido y queremos estimarlo a partir de los datos.

Preguntas:

1) Exprese la "Likelihood function" de p, es decir:

$$\mathcal{L}(p) = \mathbb{P}(x_1, x_2, x_3, x_4 \mid p)$$

Respuesta:

$$\mathcal{L}(p) = \mathbb{P}(x_1, x_2, x_3, x_4 \mid p)$$

$$= \mathbb{P}(X_1 = 2)\mathbb{P}(X_2 = 2)\mathbb{P}(X_3 = 1)\mathbb{P}(X_4 = 3)$$

Dado que todas las X's son independientes y tienen distribución idéntica ($\text{Binomial}(n = 3, p)$), podremos re-escribir la expresión como:

$$
\begin{aligned}
\mathcal{L}(p) &= \mathbb{P}(X_1 = 2)\mathbb{P}(X_2 = 2)\mathbb{P}(X_3 = 1)\mathbb{P}(X_4 = 3) \\
&= \frac{3!}{2!1!}p^2(1-p) \cdot \frac{3!}{2!1!}p^2(1-p) \cdot \frac{3!}{2!1!}p(1-p)^2 \cdot \frac{3!}{0!3!}p^3 \\
&= 27p^8(1-p)^4
\end{aligned}
$$

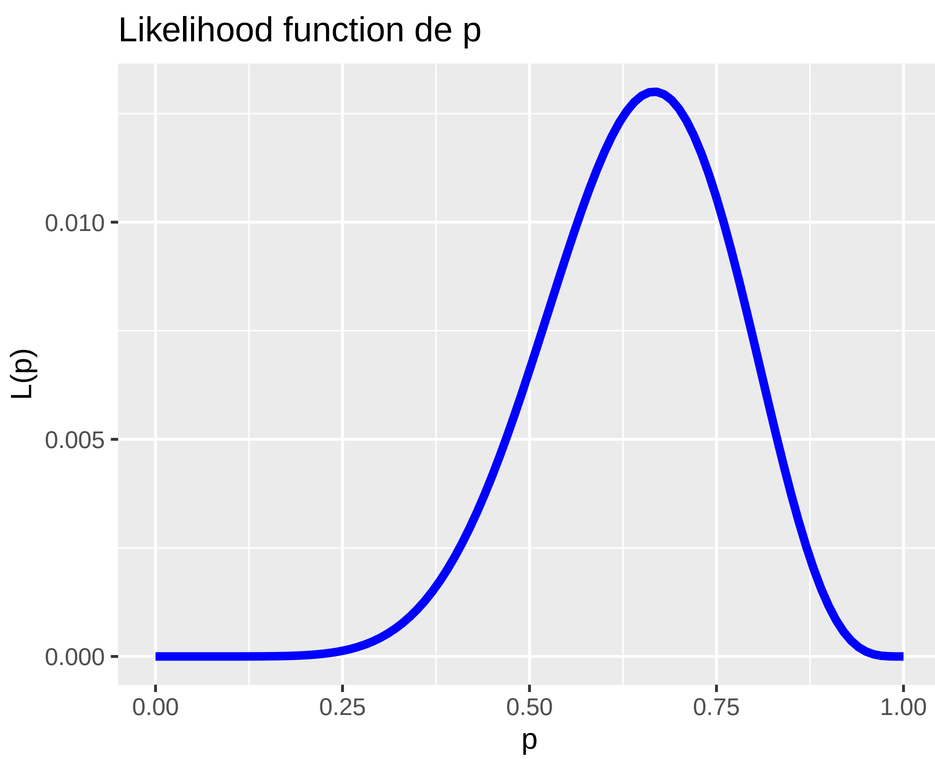
2) Grafique la función (p en el eje x y $\mathcal{L}(p)$ en el eje y) y estime visualmente cuál es el valor de p que maximiza $\mathcal{L}(p)$. Explique el significado de este número.

Para hacerlo, debes reemplazar la variable `mi_L` en el siguiente código por la expresión obtenida en la pregunta (1), usando `x` en lugar de p.

```
mydata <- data_frame(x = seq(from = 0, to = 1, by=0.01), mi_L = 27*(x^(8))*((1-x)^(4)) )

plot <- ggplot(data = mydata, mapping = aes(x = x)) +
        geom_path(aes(y=mi_L), size=1.5, colour="blue") +
        labs(y="L(p)", x="p", title="Likelihood function de p")

print(plot)
```



El valor de p que maximiza la "likelihood function" es aproximadamente $p = 0.67$. Es decir, estimamos que el escenario en que atrapar un pescado en cada tiro tiene una probabilidad de 0.67 es el más "plausible" dado los datos de los que disponemos.