

Examen anterior

Mayo PyE 2024

1 Primera pregunta

1.1 Enunciado

Un arquero realiza n disparos a una diana circular de radio desconocido θ . Sabemos que existe la misma probabilidad de acertar en todas los puntos que se encuentran a la misma distancia del centro, siendo ellas $r_1 \dots r_n$. Suponiendo que el arquero da al disco en cada disparo y que la densidad es proporcional a la distancia al centro, responde a las siguientes preguntas.

1.2 Planteamiento del problema

En primer lugar, al tratarse θ de un radio, ha de ser positivo. Por tanto, el espacio paramétrico vendrá dado por $\Theta = (0, +\infty)$.

Además, sabemos que la probabilidad de acertar puede reducirse simplemente a estudiar la distancia de cada disparo al centro de la diana. Por tanto, trabajaremos con la variable aleatoria R que mide la distancia de un disparo dado al centro de la diana. Veamos qué podemos deducir sobre ella a partir del enunciado.

Por hipótesis, el arquero siempre acierta, luego la probabilidad de $R > \theta$ ha de ser nula. Lo mismo ocurre para distancia $R < 0$, pues es físicamente imposible. Además, sabemos que la densidad es proporcional a la distancia; es decir, existe $k > 0$ tal que:

$$f_R(r) = \begin{cases} k r & \text{si } x \in [0, \theta) \\ 0 & \text{en otro caso} \end{cases}$$

Al ser f_R función de densidad es necesario que su integral en \mathbb{R} tenga valor 1. o, podemos hallar el valor de k como sigue:

$$1 = \int_{-\infty}^{\infty} f_R(r) dr = \int_0^{\theta} k r dr = k \left[\frac{r^2}{2} \right]_0^{\theta} = k \frac{\theta^2}{2} \Rightarrow k = \frac{2}{\theta^2}$$

Los resultados de los disparos r_1, \dots, r_n constituyen los valores muestrales de una muestra aleatoria simple R_1, \dots, R_n de esta variable aleatoria R .

1.3 Estimador máximo verosímil

Buscamos maximizar en función de θ la función de verosimilitud dada por:

$$L(r_1, \dots, r_n; \theta) = f_{R_1}(r_1; \theta) \cdot \dots \cdot f_{R_n}(r_n; \theta) = \left(\frac{2 r_1}{\theta^2}\right) \cdot \dots \cdot \left(\frac{2 r_n}{\theta^2}\right)$$

Como la función log es creciente en \mathbb{R} , basta maximizar el logaritmo neperiano de la función de verosimilitud, simplificando así el proceso. Aplicando las propiedades de los logaritmos, se tiene:

$$\log L(r_1, \dots, r_n; \theta) = \sum_{i=1}^n \log \left(\frac{2 r_i}{\theta^2}\right) \Rightarrow \frac{\partial}{\partial \theta} L(r_1, \dots, r_n; \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log \frac{2 r_i}{\theta^2} = \sum_{i=1}^n \frac{-2}{\theta}$$

No podemos maximizar la función de verosimilitud de este modo, pues el conjunto de valores que toma la variable R depende del parámetro.

Como ya hemos mencionado en el apartado anterior, el enunciado afirma que el arquero siempre atina al disco, luego $\forall i \in \{1, \dots, n\} : r_i \leq \theta$, lo que equivale a que el máximo de los r_i sea menor o igual que θ . De este modo, una vez conocida la muestra tenemos información sobre el espacio paramétrico:

$$\Theta|_{r_1, \dots, r_n} = [r_0, +\infty) \quad \text{para } r_0 := \max \{r_i : i \in \{1, \dots, n\}\}$$

Maximicemos ahora la función de verosimilitud gracias a esta información:

$$\log L(r_1, \dots, r_n; \theta) = \sum_{i=1}^n \log \left(\frac{2 r_i}{\theta^2}\right)$$

Basta maximizar los términos del sumatorio. Para ello es necesario tomar el menor valor posible de θ ; que, como sabemos ahora, vendrá dado por el máximo de los r_i . Por todo lo anterior, hemos hallado el estimador máximo verosímil:

$$\hat{\theta}(r_1, \dots, r_n) = \max \{r_i : i \in \{1, \dots, n\}\}$$

Se trata en efecto de un estimador, pues toma valores en $\Theta = (0, +\infty)$ y no depende de información desconocida.

1.4 ¿Es un estimador centrado?

Veamos qué distribución tiene el estimador a partir de la de R que calculamos en el primer apartado.

$$P(\hat{\theta} \leq t) = P(R_1 \leq t \cap \dots \cap R_n \leq t) = P(R_1 \leq t) \dots P(R_n \leq t) = (P(R \leq t))^n$$

$$F_{\hat{\theta}}(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{t^{2n}}{\theta^{2n}} & \text{si } 0 \leq t < \theta \\ 1 & \text{si } t \geq \theta \end{cases} \quad f_{\hat{\theta}}(t) = \begin{cases} \frac{2nt^{2n-1}}{\theta^{2n}} & \text{si } t \in [0, \theta) \\ 0 & \text{en otro caso} \end{cases}$$

Halleemos ahora la esperanza del estimador:

$$\begin{aligned} E(\hat{\theta}) &= \int_{\mathbb{R}} t f_{\hat{\theta}}(t) dt = \int_0^{\theta} \frac{2nt^{2n}}{\theta^{2n}} = \frac{2n}{\theta^{2n}} \int_0^{\theta} t^{2n} dt = \\ &= \frac{2n}{\theta^{2n}} \left[\frac{t^{2n+1}}{2n+1} \right]_0^{\theta} = \frac{2n\theta^{2n+1}}{\theta^{2n}(2n+1)} = \frac{2n\theta}{2n+1} \end{aligned}$$

El estimador no es centrado, pero es asintóticamente centrado.

1.5 ¿Es eficiente? ¿Es suficiente?

No tiene sentido hablar de eficiencia en este contexto, pues no se cumplen las condiciones de regularidad FCR al depender los valores que toma R (y por tanto el estimador $\hat{\theta}$) de θ .

Para comprobar la suficiencia, recurriremos a la factorización de Fisher-Neyman. En este caso, la función g que depende del estimador ha de ser la propia función de densidad, pues los valores del mismo dependen del parámetro. Será suficiente si y solo si se cumple la siguiente igualdad:

$$f(r_1, \dots, r_n; \theta) = f_T(r_1, \dots, r_n; \theta) \cdot h(r_1, \dots, r_n)$$

Por los apartados anteriores, conocemos las expresiones:

$$\begin{aligned} f(r_1, \dots, r_n; \theta) &= \prod_{i=1}^n \left(\frac{2}{\theta^2} \cdot 1_{(0, \theta)}(r_i) \right) \\ f_T(r_1, \dots, r_n; \theta) &= \frac{2n}{\theta^{2n}} \frac{T(r_1, \dots, r_n)^{2n-1}}{2n} \cdot 1_{(0, \theta)}(T(r_1, \dots, r_n)) \end{aligned}$$

Además, como ya hemos mencionado anteriormente, se cumple:

$$r_i \in (0, \theta) \quad \forall i \in \{1, \dots, n\} \iff \begin{cases} \max \{r_i : i \in \{1, \dots, n\}\} \in (0, \theta) \\ r_i \in (0, +\infty) \quad \forall i \in \{1, \dots, n\} \end{cases}$$

Basta tomar la siguiente función h no negativa para ver que es suficiente:

$$h(r_1, \dots, r_n) = \frac{1}{2n \max\{r_i : 1 \leq i \leq n\}^{2n-1}} \prod_{i=1}^n (2r_i \cdot 1_{(0, +\infty)}(r_i))$$

1.6 ¿Es asintóticamente insesgado o eficiente?

Hemos mencionado ya que no tiene sentido hablar de eficiencia en este contexto. Además, ya comentamos que el estimador es asintóticamente insesgado, pues se cumple:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \lim_{n \rightarrow \infty} \frac{2n\theta}{2n+1} = \theta$$