MA3402 Estadística
Profesor: Joaquín Fontbona
Auxiliares: Catalina Lizana, Claudio López, Fabián Ulloa
Fecha: 30 de septiembre de 2024

Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Auxiliar 10: Test de hipótesis paramétricos y no paramétricos

P1. (Antitrampas). En un casino hay un juego de la ruleta sencillo. La mitad de las casillas son negras y la otra mitad rojas. Gana el jugador que adivina donde caerá la bola. El casino se prepara para recibir a 1000 jugadores esta noche, por lo que usted es contactado dado su experiencia cursando MA3402. Usted es informado que algunos jugadores podrían ser tramposos. De alguna forma desconocida, pueden alterar los resultados de la ruleta a su favor y ganar más de lo esperado. El objetivo del casino es encontrarlos y arrestarlos. Usando estadística, proponga un método sensato para arrestar a todos los tramposos.

P2. (Todo es normal) Sean $x_1, ..., x_n$ muestras i.i.d. de una variable aleatoria X con función de distribución acumulada F. Kolmogorov-Smirnov busca testear si la distribución de F es la misma (o al menos cercana) a una distribución de referencia F_0, lo que se plantea como

$$H_0 : F = F_0 \quad \text{v/s} \quad H_1 : F \neq F_0$$

Definimos la función de distribución acumulada empírica como

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{x_k \leq x}$$

Y definimos el estadístico de Kolmogorov-Smirnov como

$$D_n = \sup_x |F_n(x) - F(x)|$$

Se demuestra que la distribución asintótica (en n) de $\sqrt{n}D_n$ no depende de F y es conocida como distribución K de Kolmogorov.

1. En base a lo anterior y a lo visto en clases para otros test, plantee la región crítica de rechazo del test de Kolmogorov-Smirnov y el p-value de este test.

Suponga que en un laboratorio están midiendo la discrepancia entre mediciones de la concentración esperada del principio activo de un fármaco. Y usted cuenta con tres mediciones: $-0{,}6$, $0{,}2$ y $0{,}8$ y busca testear que provengan de una $N(0, 1)$. Considere $1 - \alpha = 0{,}95$.

b) Construya la distribución acumulada empírica asociada a la muestra.

c) Sabiendo que $\Phi(-0{,}6) \approx 0{,}25$, $\Phi(0{,}2) \approx 0{,}55$ y $\Phi(0{,}8) \approx 0{,}75$, con Φ la distribución acumulada de una $N(0, 1)$, calcule D_n. Puede utilizar que si $Y \sim K$, entonces $\mathbb{P}(Y \geq 5\sqrt{3}/12) = 0{,}67$.

d) Concluya si se rechaza o no la hipótesis nula.

Recordemos

Definición 1. ***(Región de Confianza).*** *Sea $X_1, \ldots, X_n$ $(\in \mathbb{R}^d)$ m.a.s. de ley P_{θ^*} con $P_{\theta^*} \in \mathcal{M}_\Theta$, con $\mathcal{M}_\Theta$ un modelo estadístico paramétrico y sea $\alpha \in (0,1)$ un* ***nivel de riesgo****. Una región de confianza $R_\alpha = R_\alpha(X_1, \ldots, X_n) \subseteq \mathbb{R}^d$ de* ***nivel de confianza*** *$1-\alpha$ es un conjunto (aleatorio, pues depende de la m.a.s.) tal que*

$$\mathbb{P}_{\theta^*}(\theta^* \in R_\alpha) \geq 1-\alpha$$

Diremos que la región de confianza es de ***tamaño*** *$1-\alpha$ si*

$$\mathbb{P}_{\theta^*}(\theta^* \in R_\alpha) = 1-\alpha$$

Si $d=1$ y R_α es un intervalo, lo llamaremos ***intervalo de confianza*** *(de nivel o tamaño $1-\alpha$).*

Definición 2. ***(Función Pivote).*** *Dada $X_1, \ldots, X_n \sim P_\theta$ m.a.s. con $P_\theta \in \mathcal{M}_\Theta$, y $\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$ modelo estadístico. Decimos que un estadístico $S(X_1, \ldots, X_n, \theta)$ que* ***depende*** *de θ es una función pivote si su distribución* ***no depende*** *de θ. Decimos que su ley es libre del parámetro, y que $S(X_1, \ldots, X_n, \theta)$ es un* ***estadístico ancilario****.*

Teorema 1. ***(Pivotes normales)*** *Sea $\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$ un modelo estadístico paramétrico. Sea T_n un estimador de θ asintóticamente normal con varianza asintótica $V(\theta)$. Es decir, que satisfaga $\forall \theta \in \Theta$ que*

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} N(0, V(\theta))$$

Además, supongamos que $V(\theta) > 0$ y que $V : \theta \mapsto V(\theta)$ es continua. Entonces, se cumple que:

$$\frac{\sqrt{n}(T_n - \theta)}{\sqrt{V(\theta)}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1) \quad y \quad \frac{\sqrt{n}(T_n - \theta)}{\sqrt{V(T_n)}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1)$$

son funciones pivote asintóticas.

Teorema 2. ***(Desigualdad de Hoeffding)*** *Sean $X_1, \ldots, X_n$ variables aleatorias independientes, tales que $a_i \leq X_i \leq b_i$. Sea $S_n = X_1 + X_2 + ... + x_n$, entonces:*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \leq exp(-2t^2 \frac{1}{\sum_{i=1}^n (b_i - a_i)^2})$$