Inferencia Estadística

Apellidos: 6ª TG 1
Nombre: 2024-04-23

Demuestra que el estadígrafo de Kolmogórov y Smirnov,

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$$

donde

$$F_n = \text{ojiva empírica}$$

y

$$F_0 = \text{ojiva teórica completamente especificada de una distribución continua}$$

es de libre distribución bajo la hipótesis nula de que la población sigue la distribución F_0.

Sea $\vec{X} \stackrel{\text{def}}{=} (X_1, \ldots, X_n)$ una muestra aleatoria simple de la población X que sigue bajo la hipótesis nula la distribución dada por F_0. Se considera la siguiente cadena de igualdades:

- $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| = \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} I(X_i \leq x)}{n} - F_0(x) \right|$ por la definición de ojiva empírica, siendo I la función indicatriz.
- $D_n = \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} I(X_i \leq x)}{n} - F_0(x) \right| = \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} I\left(F_0(X_i) \leq F_0(x)\right)}{n} - F_0(x) \right|$ donde
  - se mantiene el sentido de la desigualdad porque F_0 es no decreciente por ser una ojiva;
  - se mantiene igual el sumatorio porque F_0 es estrictamente creciente (en el soporte de X) por ser X una distribución continua.
- $D_n = \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} I\left(F_0(X_i) \leq F_0(x)\right)}{n} - F_0(x) \right| = \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} I\left(U_i \leq F_0(x)\right)}{n} - F_0(x) \right|$ porque $U_i \stackrel{\text{def}}{=} F_0(X_i)$ es una variable aleatoria con distribución uniforme bajo la hipótesis nula, según la cual X_i sigue la distribución dada por F_0, y sabemos que una variable aleatoria continua transformada mediante su propia ojiva sigue una distribución uniforme típica.

- $D_n = \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} I\left(U_i \le F_0(x)\right)}{n} - F_0(x) \right| = \sup_{0<u<1} \left| \frac{\sum_{i=1}^{n} I\left(U_i \le u\right)}{n} - u \right|$ porque la trasformación $u = F_0(x)$ es biyectiva en el soporte de la variable aleatoria continua X y se recorre el intervalo $(0,1)$ en lugar de la recta real completa. Fuera de dicho soporte, la diferencia entre F_n y F_0 es nula, por lo que no afecta al supremo de cantidades positivas.

Finalmente, la expresión $D_n = \sup_{0<u<1} \left| \frac{\sum_{i=1}^{n} I\left(U_i \le u\right)}{n} - u \right|$ ya no depende ni de F_0 ni de las X_i , que son variables aleatorias con distribución F_0 . En consecuencia, la distribución de D_n no depende, bajo la hipótesis nula, de la distribución F_0 y, por tanto, es de libre distribución.