Un problema real en donde se usan estadísticas de orden,

- Durante la segunda guerra mundial los aliados se dieron cuenta que los alemanes ponían un número de serie en sus armas, en particular en los tanques. Si 30, 130, 78 y 51 fueran números de serie de algunos tanques destruidos por los aliados.

¿Cómo estimo el número total de tanques que tienen los alemanes?

Formulación

Sean $X_1, X_2, \ldots, X_n$ v.a. independientes con distribución uniforme discreta entre $\underline{1, 2, \ldots, N}$ → el objetivo es estimar $\underline{\underline{N}}$. De aquí es fácil ver que

$$\hat{N} = X_{(n)}$$

¿Qué tan bueno es este estimador?

Es obvio que siempre estaremos sub-estimando

Podemos obtener $\mathbb{E}(X_{(n)})$ y ver si podemos hacer alguna corrección para mejorarlo.

Resolver esto usando v.a. discretas es muy complicado, vamos a utilizar una aproximación continua

$X_1, X_2, \ldots, X_n$ v.a. independientes $U(0, N-1)$ todas con distribución

$$f_X(x) = \frac{1}{N-1} \mathbb{1}_{(0,N-1)}(x)$$

$$\Rightarrow F_X(x) = \frac{x}{N-1}$$

$$f_{X_{(n)}}(x) = n\, f_X(x)\, F_X^{n-1}(x)$$

$$= \frac{n}{(N-1)^n} x^{n-1}$$

$$\mathbb{E}(X_{(n)}) = \frac{n}{(N-1)^n} \int_0^{N-1} x^n \, dx$$

$$= \frac{n}{(N-1)^n} \frac{1}{n+1} \left[(N-1)^{n+1} - 0 \right]$$

$$= \frac{n}{n+1} (N-1)$$

$$\mathbb{E}(X_{(n)}) = \left(\frac{n}{n+1}\right)(N-1)$$

$$\Rightarrow \boxed{\mathbb{E}\left(\left(\frac{n+1}{n}\right)X_{(n)} + 1\right) = N}$$

¿Cómo estimo N?

$$\left(\frac{n+1}{n}\right) x_{(n)} + 1$$

Teorema

Sea $X_1, X_2, \dots, X_n$ v.a. independientes y todas con distribución $f_X(x)$ y distribución acumulada $F_X(x)$

$$\Rightarrow F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left(1 - F_X(x)\right)^{n-k}$$

y si X es una v.a. continua (de tarea demostrarán que)

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!}\, f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}$$

Dem

Vamos a crear nuevas v.a.

$$Z_i = \mathbb{1}_{(-\infty, x)}(X_i) = \begin{cases} 1, & \text{si } X_i \in (-\infty, x) \\ 0, & \text{eoc} \end{cases}$$

$$\Rightarrow P(Z_i = 1) = P(X_i \leq x) = F_{X_i}(x) = F_X(x)$$

$$Y = \sum_{i=1}^{n} Z_i \sim Bin(n, F_X(x))$$

$$P(Y = y) = \binom{n}{y} F_X^{y}(x) (1 - F_X(x))^{n-y}$$

$$F_{X_{(j)}}(x) = P(X_{(j)} \leq x)$$

$$= P(X_{(1)} \leq x, X_{(2)} \leq x, \dots X_{(j)} \leq x)$$

$\Rightarrow$ al menos j de los X_i, $i = 1, \dots, n$ son $\leq x$

$$= Y = \sum_{i=1}^{n} Z_i \geq j$$

$$\Rightarrow F_{X_{(j)}}(x) = P(Y \geq j) = \sum_{k=j}^{n} \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k}$$

$\dagger$

$$\boxed{f_{X_{(j)}}(x) = \frac{d}{dx} F_{X_{(j)}}(x)}$$