Clase práctica 19 06/11/18 Proba (C)

Ejercicio 1. **(German Tank Problem)** Suponga que se tiene una población finita con unidades indexadas $1, 2, 3, \ldots, N$ de la cual se desconoce su tamaño, N. Por ejemplo, la población es la de los tanques de guerra de un cierto modelo producidos por la Alemania nazi durante la guerra (en un cierto período, fábrica, etc) y los índices son sus números de serie de fabricación. Supongamos que elegimos al azar **sin reposición** n unidades de la población y observamos sus números de serie $X_1, ..., X_n$. Siguiendo con el ejemplo, los n tanques elegidos pueden ser aquellos capturados durante el combate. Se desea estimar la producción total de tanques N a partir de los números de serie de los tanques capturados.

a) Demuestre que la probabilidad conjunta del vector $\underline{X} = (X_1, \ldots, X_n)$ esta dada por

$$P(X_1 = x_1, \ldots, X_n = x_n) = \begin{cases} \frac{(N-n)!}{N!} & \text{si } \{x_1, ..., x_n\} \subseteq \{1, ..., N\} \text{ y } x_i \neq x_j \text{ si } i \neq j. \\ 0 & \text{en caso contrario.} \end{cases}$$

b) Probar que el estimador de máxima verosimilitud de N esta dado por $X_{\max} = \max(X_1, ..., X_n)$.

c) Probar que la distribución de $X_{\max}$ está dada por

$$P_N(X_{\max} = x) = \frac{\binom{x-1}{n-1}}{\binom{N}{n}} \mathbb{1}_{\{n,\ldots,N\}}(x)$$

d) Probar que $\mathbb{E}_N(X_{\max}) = \frac{n}{n+1}(N+1)$, para esto probar primero la **'hockey-stick identity'** $\sum_{x=n}^{N} \binom{x}{n} = \binom{N+1}{n+1}$.

e) Deducir que $\widehat{N} = X_{\max} + \frac{1}{n} X_{\max} - 1$ es un estimador insesgado de N.[1] [2]

f) (Opcional) Probar que $\text{Var}_N(X_{\max}) = \frac{(N+1)(N-n)}{n(n+2)}$.

g) (Tarea para el hogar) Escribir en **R** un código que simule para N y n la distribución conjunta, y que calcule el estadístico $\widehat{N}$. Realizar 100 simulaciones del estadístico y graficar los valores obtenidos, para distintos valores de N y n, por ejemplo $N = 300, n = 30$. Calcular el promedio de las 100 estimaciones y ver que es cercano a N, calcular el desvio estándar de las estimaciones y comparar con el desvio estándar teórico.

Ejercicio 2. Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución en $\mathcal{F} = \{f(x, \theta) : \theta > 0\}$ donde

$$f(x, \theta) = \frac{1}{\theta} e^{\frac{-x}{\theta}} \mathbb{1}_{(0,+\infty)(x)}$$

a) Hallar el estimador de máxima verosimilitud de θ.

b) Hallar un intervalo de confianza de nivel exacto de θ, para $\alpha = 0.05$ y $n = 10$.

c) Hallar intervalos de confianza asintóticos de θ para $\alpha = 0.05, n = 20$.

d) (Tarea para el hogar) Escribir en **R** un código que genere una muestra de tamaño $n = 20$ para $\theta = 1$, y que calcule los IC anteriores con $\alpha = 0.05$. Realizar 100 simulaciones y calcular el porcentaje de muestras para los que los IC contienen a θ.

---

[1] Puede demostrarse que el estimador $X_{\max}$ es suficiente y completo, y por lo tanto (por Lehmann-Scheffé) $\widehat{N}$ es el estimador insesgado de mínima varianza uniforme de N.

[2] El estimador $\widehat{N}$ fue usado efectivamente durante la Segunda Guerra Mundial para estimar el número de tanques (de ciertos modelos) y de otros artefactos (camiones, neumáticos, bombas voladoras V-1, misiles V-2) producidos por la Alemania nazi. Las estimaciones arrojadas por estos estimadores fueron luego contrastadas con los registros obtenidos de las fábricas alemanas luego de la victoria aliada quedando demostrado que las estimaciones habían sido muy precisas. De hecho, los errores en las estimaciones basadas en $\widehat{N}$ fueron ordenes de magnitud menores que los errores de cálculos no estadísticos realizados por la inteligencia aliada. Ver el artículo de R. Ruggles y H. Broadie.

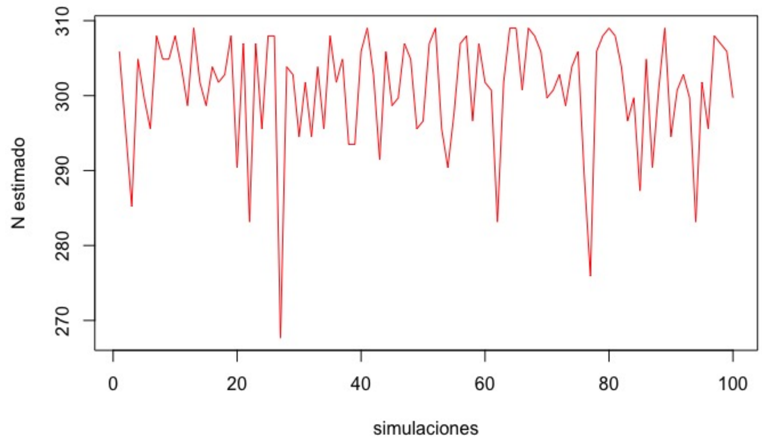

Figure 1: Simulaciones de $\widehat{N}$ para $N = 300, n = 30$.

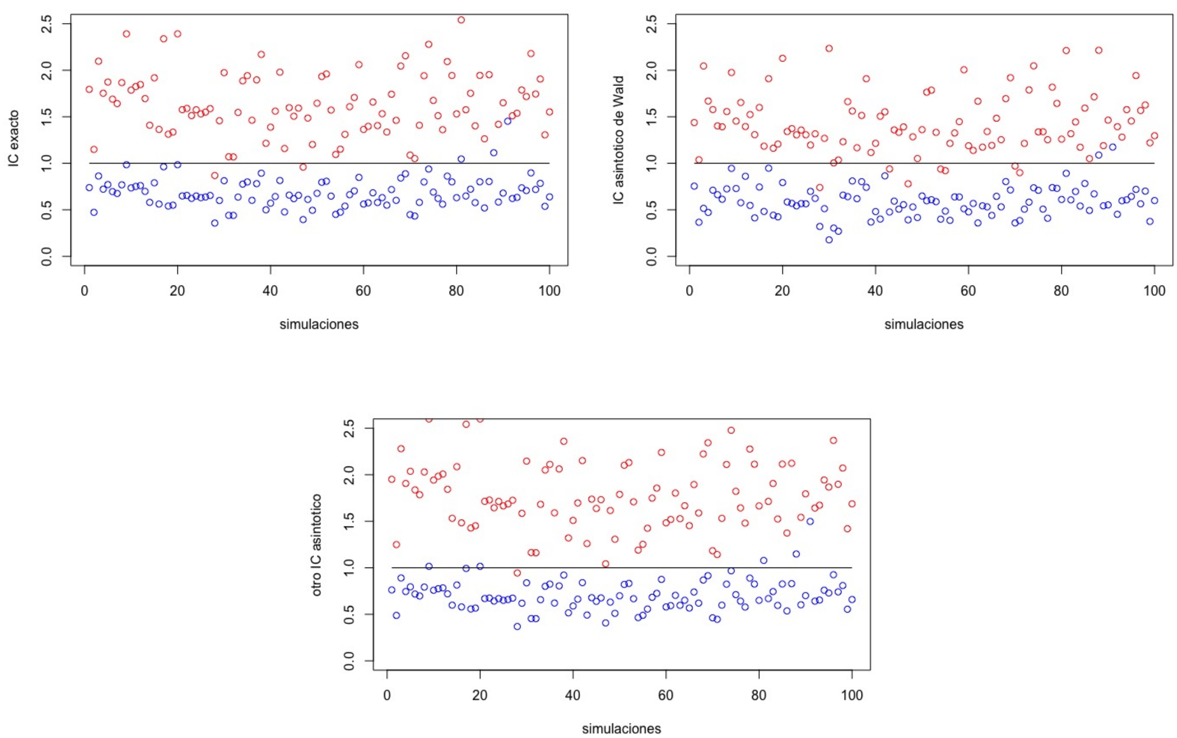

Figure 2: Simulaciones IC de nivel exacto, asintótico de Wald y otro asintótico.