Sea $(X_i)_{i\geq 1}$ una muestra aleatoria con distribución F. Denotemos con X a un elemento con misma distribución que X_i. Asuma que estamos interesados en estimar el la probabilidad de que X sea mayor a uno: $\theta(F) := \mathbb{P}_F(X > 1)$.

Estimación Puntual

Estimador 1:

1.1 Proponga un estimador $\widehat{\theta}_n$ consistente para $\theta(F) = \mathbb{P}_F(X > 1)$ (observar que aquí hacemos explícito que θ depende de la distribución F).

1.2 Implemente una función **estsombrero** que tenga por argumento un conjunto de datos $(x_1, \ldots, x_n)$ muestra y devuelva el valor de la estimación obtenida utilizando $\widehat{\theta}_n$.

1.3 Calcule el valor de $\widehat{\theta}_n$ en el siguiente conjunto de datos utilizando **estsombrero** :

$$12.23 \quad 6.37 \quad 6.10 \quad 0.70 \quad 3.48 \quad 2.82 \quad 9.55 \quad 2.21 \quad 0.72 \quad 9.09.$$

Mundo Exponencial: Calentando motores

1.4 Sea X una variable aleatoria con distribución F, exponencial de parámetro $\lambda = 0.2$: $X \sim \mathcal{E}(0.2)$. Indique el valor de

$$\mathbb{E}(X) = \ldots\ , \quad \mathbb{V}(X) = \ldots\ , \quad \mathbb{P}(X > 1) = \ldots\ , \quad \text{cuando } X \sim \mathcal{E}(0.2).$$

1.5 Sea ahora X una variable aleatoria con distribución F perteneciente a la familia exponencial: es decir, $X \sim \mathcal{E}(\lambda)$ con λ DESCONOCIDO. Exprese cada uno de los siguientes objetos en función de λ:

$$\mathbb{E}(X) = \ldots\ , \quad \mathbb{V}(X) = \ldots\ , \quad \mathbb{P}(X > 1) = \ldots\ , \quad \text{cuando } X \sim \mathcal{E}(\lambda).$$

Mundo Exponencial: Haciendo Estadística

Sean $(X_i)_{i\geq 1}$ i.i.d., con misma distribución que X. Asuma ahora que F pertenece a la familia exponencial; es decir, $X \sim \mathcal{E}(\lambda)$, con λ DESCONOCIDO.

1.6 Proponga un nuevo estimador $\widetilde{\theta}_n$ consistente para $\theta(F) = \mathbb{P}_F(X > 1)$ bajo este nuevo escenario. Es decir, defina $\widetilde{\theta}_n = f_n(X_1, \ldots, X_n)$ de forma tal que

$$\widetilde{\theta}_n = f_n(X_1, \ldots, X_n) \xrightarrow{p} e^{-\lambda} \text{ cuando } X_i \sim \mathcal{E}(\lambda) \ , \forall \lambda > 0.$$

1.7 Implemente una función **estrulo** que tenga por argumento un conjunto de datos $(x_1, \ldots, x_n)$ y devuelva el valor de la estimación obtenida utilizando $\widetilde{\theta}_n$.

1.8 Calcule el valor de $\widetilde{\theta}_n$ en el siguiente conjunto de datos utilizando **estrulo**:

12.23 6.37 6.10 0.70 3.48 2.82 9.55 2.21 0.72 9.09.

Simulación 1:. A lo largo de esta simulación generaremos variables con distribución exponencial de paramétro $\lambda = 0.2$.

1.9 Indique cuál es el verdadero valor que estamos queriendo estimar: $\theta = \mathbb{P}(X > 1)$, siendo $X \sim \mathcal{E}(0.2)$.

1.10 Genere una muestra de tamaño n=50 y calcule los dos estimadores.

1.11 Genere $Nrep$= 1000 muestras de tamaño n=50 y guarde los valores de cada uno de los dos estimadores calculados en cada uno de los $Nrep = 1000$ conjuntos de datos.

1.12 Realice un histograma de cada uno de los estimadores propuestos con los valores obtenidos en el ítem anterior. Comente los gráficos realizados. Indique que estimador prefiere en este escenario y explique a que atribuye sus bondades.

1.13 Represente en una tabla el error cuadrático medio (estimado) de los estimadores $\widehat{\theta}_n$ y $\widetilde{\theta}_n$ para muestras de tamaño n=150, n=200, n=500 y n=1000, utilizando Nrep=1000 replicaciones en cada caso. ¿Qué estimador prefiere bajo este escenario?

Mundo Normal: Ojo al Piojo! Considere ahora variables aleatorias X_i i.i.d. con distribución normal de media $\mu = 1/0.2$ y $\sigma^2 = 1/0.2^2$.

1.14 Calcule la probabilidad de que X_i supere el valor 1: $\mathbb{P}(X_i > 1)$

1.15 Calcule el valor de cada uno de los siguientes límites:

$$\lim_{n\to\infty} \widehat{\theta}_n(X_1, \ldots, X_n) = \ldots\ldots \ , \quad \lim_{n\to\infty} \widetilde{\theta}_n(X_1, \ldots, X_n) = \ldots\ldots$$

1.16 Proponga un nuevo estimador $\theta_n^* = \theta_n^*(X_1, \cdots, X_n)$ para $\theta(F) = \mathbb{P}_F(X_i > 1)$, asumiendo ahora que F pertenece a la normal: $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

Simulación 2: A lo largo de esta simulación generaremos variables con distribución normal de media $\mu = 1/0.2$ y $\sigma^2 = 1/0.2^2$. Represente en una tabla el error cuadrático medio (estimado) de los estimadores $\widehat{\theta}_n$, $\widetilde{\theta}_n$ y θ_n^* para muestras de tamaño n=150, n=200, n=500 y n=1000, utilizando Nrep=1000 replicaciones en cada caso. Analice los resultados obtenidos y explique que estimador elegiría bajo este escenario.

Estimación por Intervalos

El objetivo de este ejercicio es comparar empíricamente la performance de diferentes intervalos de confianza, basados en estimadores asintóticamente normales. Por un lado, utilizaremos intervalos construidos estimando la varianza asintótica mediante (i) boostrap y (ii) propagación de errores Finalmente, construiremos (iii) intervalos para una proporción, siendo que el parámetro de interés es una probabilidad. Estudiaremos el nivel de cubrimiento empírico y la longitud media de cada uno procedimientos mencionados.

Recordemos que el valor de interés es

$$\theta = \mathbb{P}(X > 1)\,.$$

Generaremos datos asumiendo que provienen de una familia exponencial $\mathcal{E}(\lambda)$:$X \sim \mathcal{E}(\lambda)$, co $\lambda = 0.2$

1. Obtenga una fórmula para el parámetro de interés en función de λ, que permita calcular la varianza asintótica del estimador $\widetilde{\theta}_n$ (propuesto en el punto 1.6) utilizando propagación de errores.

2. Construir los siguientes tres posibles intervalos de confianza asintóticos para el parámetro de interés, con nivel 0.95

   (a) Intervalos asintóticamente normal basado en $\widehat{\theta}_n$, propuesto en el item 1.1, considerando que se trata de un estimador para una probabilidad (es decir, *no deja de ser un* $\widehat{p}$).

   (b) Intervalos asintóticamente normal basado en $\widetilde{\theta}_n$, propuesto en el item 1.6, estimando la varianza asintótica con propagación de errores.

   (c) Intervalos asintóticamente normal basado en $\widetilde{\theta}_n$, propuesto en el item 1.6, estimando la varianza asintótica con Bootstrap, utilizando $Nboot = 2000$.

3. Genere datos utilizando $\lambda = 0.2$. Considere $n = 100, 200, 500$ y estudie el nivel de cubrimiento empírico en $Nrep = 1000$ replicaciones de los intervalos propuestos. Recuerde que el cubrimiento empírico calcula la proporción de veces que el intervalo simulado atrapa al valor de interés.