Bootstrap no parametrico

Sea $X_1, ..., X_n$ una m.a. de la distribución $poisson(\theta)$. Supongamos que el parametro de interés a estimar es $\tau(\theta) = e^{-\theta} = \mathbb{P}(X = 0)$. Se puede verificar que $\hat{\tau}(\theta) = (\frac{n-1}{n})^{\sum_{i=1}^{n} X_i}$ es el UNVUE de $\tau(\theta) = e^{\theta}$, sin embargo, no es tan facil encontrar la distribución de $\hat{\tau}$ o la expresión de $V(\hat{\tau})$

Metodo Monte Carlo

Primero realizaremos el método Monte Carlo.

Para utilizar este metodo, necesitamos conocer el valor de θ y el valor de n. Supongamos que $\theta = 1$, $n = 20$.

El numero de simulaciones o muestras aleatorias será de $B = 10,000$.

Los pasos a seguir del metodo Monte Carlo son los siguientes:

Generaremos $B = 10,000$ muestras aleatorias de tamaño $n = 20$ de una distribución $poisson(\theta = 1)$

Con cada muestra aleatoria $X_{1_i}, ..., X_{20_i}$ con $i = \{1, 2, ..., 10,000\}$, podemos generar numeros aleatorios de la distribución de $\hat{\tau}$, como:

$$\hat{\tau}_i(\theta) = \left(\frac{19}{20}\right)^{\sum_{j=1}^{20} X_{j_i}}$$

y asi obtenemos los valores de $\hat{\tau}_1, ..., \hat{\tau}_B$, que se consideran como los datos provenientes de la distribución de $\hat{\tau}$

Calculamos la varianza y la esperanza de $\hat{\tau}_1, ..., \hat{\tau}_B$

Cuadro 1: Esperanza y Varianza metodo Monte Carlo

Esperanza	Varianza
0.3696324	0.0070536

Para visualizar mejor los datos, graficamos el histograma de $\hat{\tau}_1, ..., \hat{\tau}_B$

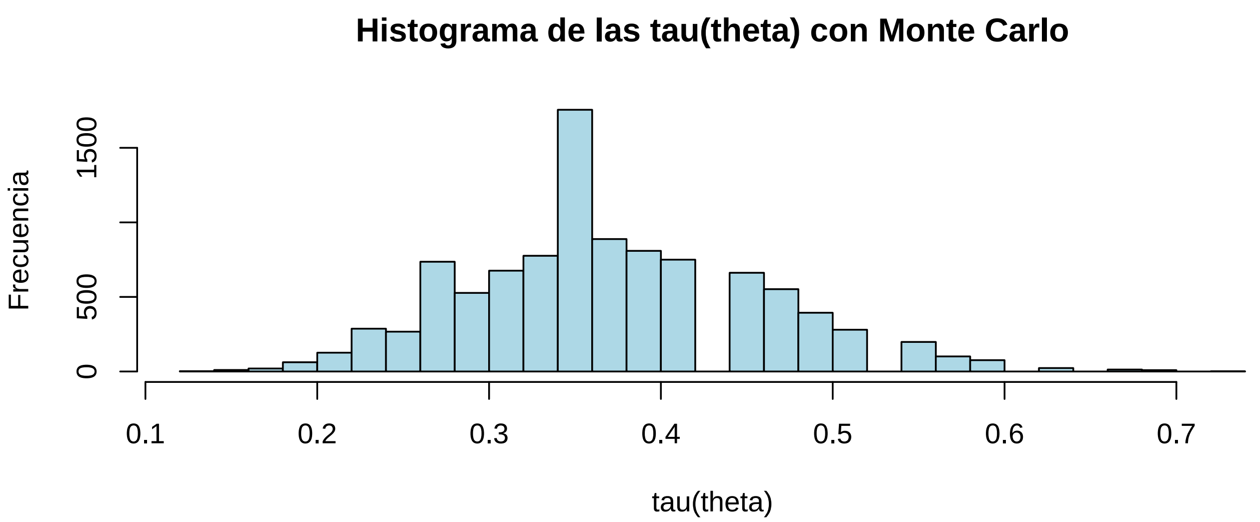

Figura 1: Histograma de tau estimada con el metodo Monte Carlo

Analizando la Figura 1, podemos ver que una gran parte de los datos se concentran aproximadamente en el intervalo $[3.4, 3.6]$,

Metodo Bootstrap

El método bootstrap se basa en construir una distribución muestral para un estadístico de interés mediante el remuestreo repetido de los datos disponibles.

La gran diferencia que tiene el Bootstrap con el metodo Monte Carlo es que en bootstrap solo tenemos una muestra aleatoria en en nuestro caso en particular es $S = \{X_1, ..., X_{20}\}$, es decir solo **una** m.a. de tamaño $n = 20$

La muestra con la que vamos a trabajar en particular será:

$S = \{0, 1, 3, 2, 1, 0, 1, 0, 2, 0, 1, 0, 2, 1, 2, 2, 1, 0, 0, 1\}$

A partir de esta muestra aleatoria S dada, generamos $B = 10,000$ muestras bootstrap, cada muestra se crea seleccionando elementos de la muestra S con la técnica de muestreo con reemplazo. Es necesario el muestreo con reemplazo para evitar reproducir simplemente la muestra original S, obteniendo asi la i-ésima muestra bootstrap $S^{*(i)} = \{X_1^{*(i)}, ..., X_{20}^{*(i)}\}$

Con cada muestra bootstrap $S^{*(i)}$ calculamos $\hat{\tau}$ de la siguiente forma:

$$\hat{\tau}_{S^{*(i)}}(\theta) = \left(\frac{n-1}{n}\right)^{\sum_{j=1}^{n} X_j^{*(i)}}$$

y obtenemos una estimación de $\hat{\tau}$ que está basada en las $B = 10,000$ replicaciones.

$$\hat{\tau}^*_{(1)}, ..., \hat{\tau}^*_{(B)}$$

Calculamos la varianza y la esperanza de $\hat{\tau}^*_{(1)}, ..., \hat{\tau}^*_{(B)}$

Cuadro 2: Esperanza y Varianza metodo Bootstrap

Esperanza	Varianza
0.3660667	0.0055919

Realizamos el histograma de las $\hat{\tau}^*_{(1)}, ..., \hat{\tau}^*_{(B)}$

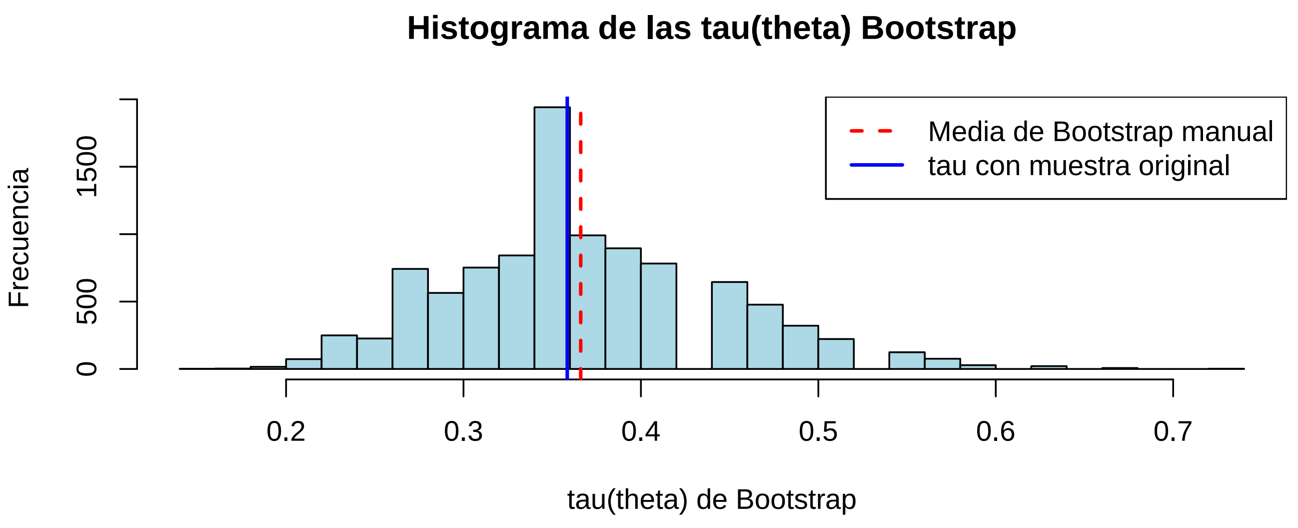

Figura 2: Histograma de tau estimada con bootstrap

Analizando la Figura 2, podemos ver que de igual forma que en la Figura 1 una gran parte de los datos se concentran aproximadamente en el intervalo $[3.4, 3.6]$,

Ahora analizaremos los datos y compararemos lo obtenido

Cuadro 3: Esperanza y Varianza

Metodo	Esperanza	Varianza
Monte Carlo	0.3696324	0.0070536
Bootstrap no parametrico	0.3660667	0.0055919

En la Tabla 3 se muestran los valores de la esperanza y varianza de $\hat{\tau}(\theta)$ para cada método.

Veamos como se ven los histogramas juntos

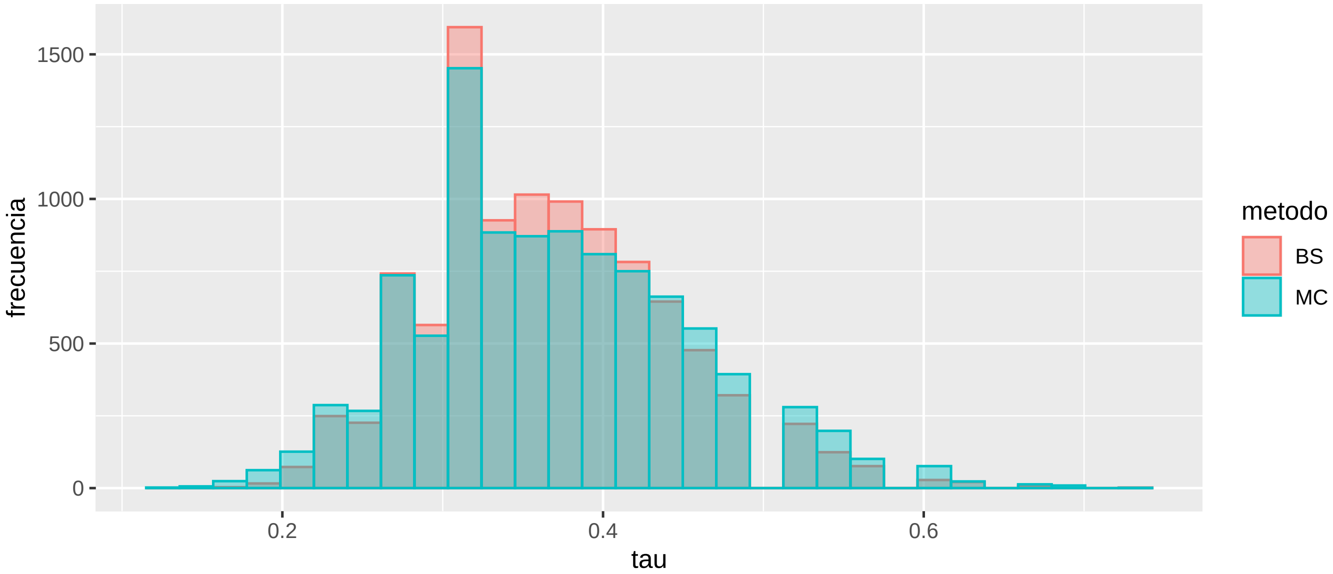

Figura 3: Histogramas sobrepuestos de tau con el metodo Monte Carlo y el metodo bootstrap

Observaciones generales:

Notemos que el metodo que más se acerca a la distribución de $\tau(\theta)$, es el metodo Monte Carlo, sin embargo necesitamos demasiadas muestras aleatorias cada una de tamaño 20, con el metodo bootstrap solo con una muestra aleatoria de tamaño 20 (para este caso en particular), podemos estimar la distribución de $\tau(\theta)$, siempre y cuando nuestros datos sean buenos, con base en los histogramas, podemos ver que deacuerdo a la Figura 3 el metodo bootstrap con la muestra dada, se asemeja al histograma del metodo Monte Carlo, además la varianza de $\hat{\tau}(\theta)$ con el metodo bootstrap es de 0.0055919 y con el metodo Monte Carlo es 0.0070536, estas tienen una diferencia de 0.0014617, notemos la varianza $\hat{\tau}(\theta)$ con el metodo de bootstrap es menor a la varianza de $\hat{\tau}(\theta)$ con el metodo Monte Carlo, sin embargo es pequeña la diferencia.

En conclusión el metodo bootstrap es una buena opción para estimar $\tau(\theta) = \mathbb{P}(X = 0)$ siempre y cuando la muestra aleatoria sea buena ya que con una solo muestra aleatoria podemos estimar la distribución de $\tau(\theta) = \mathbb{P}(X = 0)$