Reporte Ejecutivo: Análisis Bootstrap y Monte Carlo para la Estimación de $ au(heta) = P(X=0)$ en una Distribución Poisson

Omar Flores (Basado en el script R proporcionado)

21 de May de 2025

Resumen

Este reporte presenta los resultados de análisis estadísticos realizados para estimar el parámetro $\tau(\theta) = e^{-\theta} = P(X = 0)$ de una distribución Poisson, utilizando el estimador UMVUE $\hat{\tau} = \left(\frac{n-1}{n}\right)^{\sum X_i}$. Se aplicaron métodos de Monte Carlo y bootstrap no paramétrico según lo especificado, y se evaluó el desempeño de los intervalos de confianza bootstrap. Todos los análisis se implementaron en R, utilizando `set.seed(1234)` para reproducibilidad. Las cifras y tablas aquí presentadas se derivan de la ejecución de dicho script.

Introducción

El objetivo principal es la estimación del parámetro $\tau(\theta) = e^{-\theta}$, que representa la probabilidad de observar un cero en una variable aleatoria X que sigue una distribución de Poisson con parámetro θ. El estimador utilizado es $\hat{\tau} = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$, el cual se indica es el estimador insesgado de mínima varianza uniforme (UMVUE) de $\tau(\theta)$.

Este reporte se divide en tres partes principales:

1. **Simulación Monte Carlo (MC):** Aproximación del valor esperado $E(\hat{\tau})$ y la varianza $V(\hat{\tau})$ cuando θ es conocido.
2. **Bootstrap No Paramétrico:** Estimación de $\tau(\theta)$, $V(\hat{\tau})$ y construcción de un intervalo de confianza para $\tau(\theta)$ a partir de una muestra observada específica, sin asumir conocimiento de θ.
3. **Estudio de Simulación de Cobertura:** Evaluación de la probabilidad de cobertura de los intervalos de confianza bootstrap no paramétricos.

Para todos los análisis inferenciales no especificados, se utiliza un nivel de significancia $\alpha = 0.05$.

Parte a: Simulación Monte Carlo (Distribución Conocida)

En esta sección, se utiliza el método Monte Carlo para aproximar el valor esperado y la varianza del estimador $\hat{\tau}$. Se asume que los datos provienen de una distribución Poisson(θ) con $\theta = 1.3$, y se generan muestras de tamaño $n = 20$. Se realizaron $B = 10,000$ réplicas Monte Carlo.

Metodología

Para cada una de las $B = 10,000$ réplicas:

1. Se generó una muestra aleatoria $X_1, \ldots, X_{20}$ de una distribución Poisson($\theta = 1.3$).
2. Se calculó $\hat{\tau}_b = \left(\frac{20-1}{20}\right)^{\sum X_i}$ para la muestra b.

Luego, $E(\hat{\tau})$ se aproximó por $\frac{1}{B}\sum_{b=1}^{B}\hat{\tau}_b$ y $V(\hat{\tau})$ por $\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\tau}_b - \bar{\hat{\tau}})^2$.

Resultados

El valor verdadero de $\tau(\theta)$ para $\theta = 1.3$ es $\tau(1.3) = e^{-1.3}$. Los resultados obtenidos de la simulación Monte Carlo se resumen en la Tabla 1.

Cuadro 1. Resultados de la Simulación Monte Carlo para $\hat{\tau}$ ($\theta = 1.3$, $n = 20$, $B = 10,000$)

Métrica	Valor
Valor verdadero $\tau(1.3) = e^{-1.3}$	0.27253
Estimación MC de $E(\hat{\tau})$	0.27210
Estimación MC de $V(\hat{\tau})$	0.0050471

El histograma de las $10,000$ estimaciones $\hat{\tau}_b$ obtenidas se muestra en la Figura @ref(fig:hist-mc-a).

Parte b: Bootstrap No Paramétrico (Muestra Observada)

En esta sección, se utiliza el método bootstrap no paramétrico para analizar el estimador $\hat{\tau}$ a partir de una única muestra observada, sin suponer conocimiento previo de θ o la forma exacta de la distribución (aunque el estimador $\hat{\tau}$ se deriva bajo el supuesto Poisson).

Datos y Metodología

La muestra observada es $X = (1, 2, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0)$, con $n = 20$. Se generaron $B = 10,000$ muestras bootstrap:

1. Se calculó $\hat{\tau}_{obs}$ a partir de la muestra original.
2. Para cada una de las $B = 10,000$ réplicas bootstrap:
   (a) Se generó una muestra bootstrap $X_1^*, \ldots, X_{20}^*$ muestreando con reemplazo de la muestra observada X.

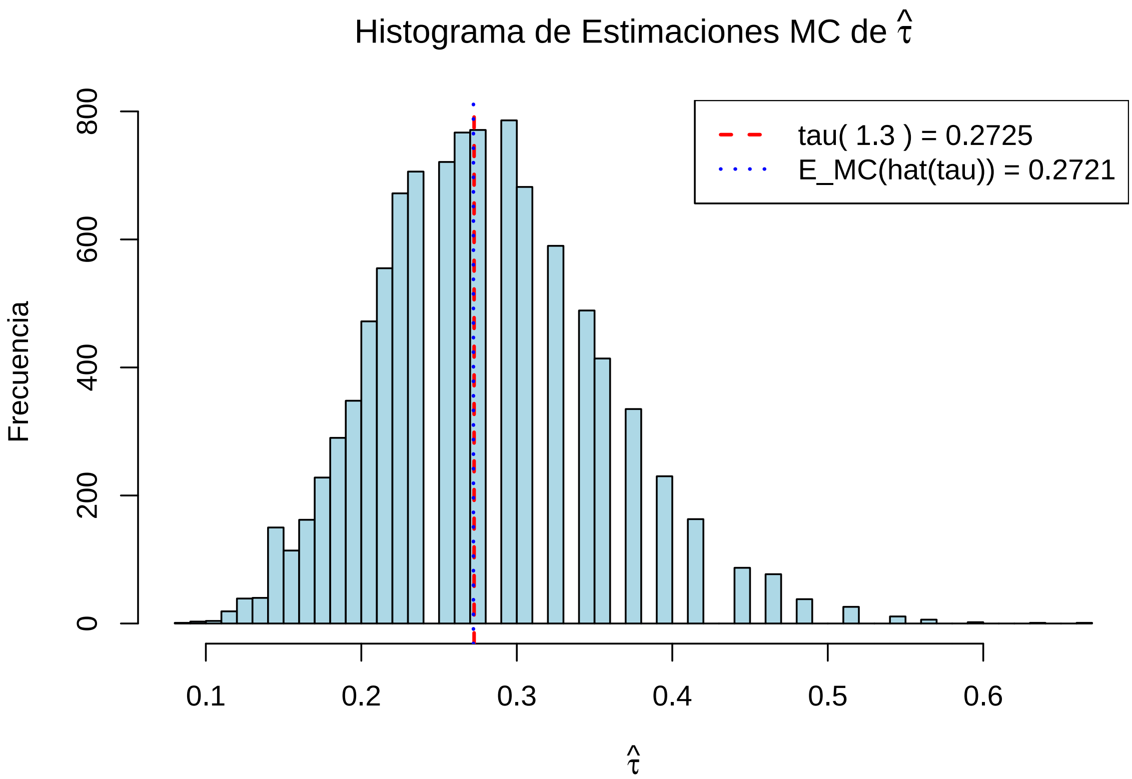

Figura 1. Histograma de las estimaciones $\hat{\tau}$ obtenidas mediante simulación Monte Carlo. La línea roja discontinua indica el valor verdadero $\tau(1.3)$ y la línea azul punteada indica la media de las estimaciones MC.

(b) Se calculó $\hat{\tau}_b^*$ para la muestra bootstrap b.

3. La varianza de $\hat{\tau}$ se estimó como la varianza de las $\hat{\tau}_b^*$.

4. Se construyó un intervalo de confianza (IC) del 95% para $\tau(\theta)$ utilizando el método de los percentiles sobre las $\hat{\tau}_b^*$.

Resultados

Los resultados del análisis bootstrap no paramétrico se resumen en la Tabla 2.

Cuadro 2. Resultados del Bootstrap No Paramétrico para $\hat{\tau}$ (Muestra observada, $n = 20$, $B = 10,000$)

Métrica	Valor
Estimación $\hat{\tau}_{obs}$ (de la muestra original)	0.54036
Estimación Bootstrap de $V(\hat{\tau})$	0.0053082
IC del 95% para $\tau(\theta)$ (percentiles)	(0.41812, 0.69834)

El histograma de las $10,000$ estimaciones bootstrap $\hat{\tau}_b^*$ se muestra en la Figura @ref(fig:hist-bootstrap-b).

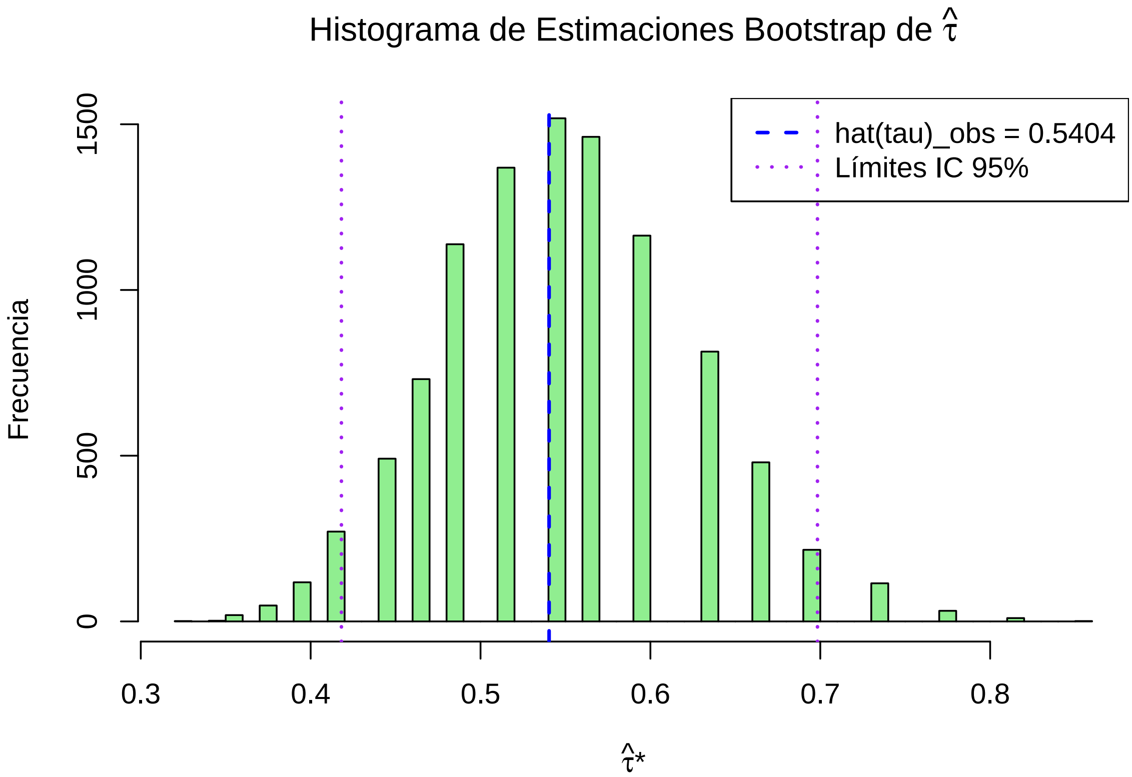

Figura 2. Histograma de las estimaciones bootstrap $\hat{\tau}^*$. La línea azul discontinua indica $\hat{\tau}_{obs}$ y las líneas púrpuras punteadas los límites del IC del 95%.

Comentario sobre los resultados si la muestra proviniera de Poisson($\theta = 1.3$): El valor verdadero $\tau(1.3) = e^{-1.3} \approx 0.27253$. La estimación observada $\hat{\tau}_{obs} = 0.54036$. El IC Bootstrap [0.41812, 0.69834] **no contiene** el valor $\tau(1.3)$.

Parte c: Estudio de Simulación para Cobertura del IC Bootstrap

Se realizó un estudio de simulación para evaluar el desempeño (probabilidad de cobertura) de los intervalos de confianza del 95% obtenidos mediante el método bootstrap no paramétrico (percentiles) para $\hat{\tau}$.

Metodología

Se repitió el siguiente procedimiento $M = 1,000$ veces:

1. Se generó una muestra aleatoria de tamaño $n = 20$ de una distribución Poisson($\theta = 1.3$). El verdadero valor del parámetro de interés es $\tau(1.3) = e^{-1.3}$.
2. Con esta muestra generada, se construyó un intervalo de confianza del 95% para $\tau(\theta)$ usando el método bootstrap no paramétrico con $B = 10,000$ réplicas (como en la Parte b).
3. Se verificó si el intervalo de confianza resultante contenía el valor verdadero $\tau(1.3)$.

La probabilidad de cobertura se estimó como la proporción de los M intervalos que contenían el valor verdadero.

Resultados

Los resultados del estudio de simulación de cobertura se presentan en la Tabla 3.

Cuadro 3. Resultados del Estudio de Simulación de Cobertura del IC Bootstrap ($M = 1,000$ sims, $\theta = 1.3$, $n = 20$, $B = 10,000$ por IC)

Métrica	Valor
Nivel de Confianza Nominal	0.95
Probabilidad de Cobertura Observada	0.9340
Número de ICs que cubrieron $\tau(1.3)$ (de $M = 1,000$)	934
p-valor (Prueba Binomial H_0: Cobertura = 0.95)	0.0242

Conclusión de la prueba binomial: Se rechaza H0. La cobertura observada (0.934) difiere significativamente de la nominal (0.95) al nivel alpha = 0.05.

“‘