Centro de Investigación en Matemáticas, A.C.
Inferencia Estadística

CIMAT

Ayudantías de Modelos Estadísticos

José Miguel Saavedra Aguilar

```
## Error in read_chunk("Asistencia.R"): no se pudo encontrar la función "read_chunk"
```

1. Introducción a Inferencia Estadística

1.1. Ejemplo 1

Para una v.a. $X \sim \exp$, graficamos la función de densidad para distintas tasas $\lambda = 5, 2, 0.5$.

Ahora, graficamos las respectivas funciones de distribución de X.

Para $\lambda = 0.5$, simulamos una muestra aleatoria de tamaño $n = 100$ datos de X.

Ordenamos los puntos simulados en $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. Les asociamos k_i definido por

$$k_i = \frac{i}{n+1} \tag{1}$$

Para conocer más sobre la función de distribución empírica, pueden consultar Wikipedia.

2. Ejemplos de Knitr

Se ejemplifica el uso de knitr. Es recomendable consultar el libro de Yihui Xie [2] para mayor información.

2.1. Ejemplo 2

Una variable aleatoria discreta X tiene función de masa de probabilidad:

x	0	1	2	3	4
$p(x)$	0.1	0.2	0.2	0.2	0.3

Utilicen el teorema de la transformación inversa para generar una muestra aleatoria de tamaño 1000 de la distribución de X. Construyan una tabla de frecuencias relativas y

comparen las probabilidades empíricas con las teóricas.
Repitan considerando la función de R sample.

2.2. Ejemplo 3

Obtengan una muestra de $10,000$ números de la siguiente distribución discreta:

$$p(x) = \frac{2x}{k(k+1)}, x = 1, 2, \ldots, k$$

para $k = 100$

2.3. Ejemplo 4

Una compañía de seguros tiene 1000 asegurados, cada uno de los cuales presentará de manera independiente una reclamación en el siguiente mes con probabilidad $p = 0.09245$. Suponiendo que las cantidades de los reclamos hechos son variables aleatorias Gamma(7000,1), hagan simulación para estimar la probabilidad de que la suma de los reclamos exceda $\$500,000$.

3. Tarea 3

3.1. Ejercicio 1

Simula las siguientes muestras Poisson, todas con $\lambda = 3$, pero de distintos tamaños, $n = 10, 20, 40, 80, 200$. Para cada muestra de estas tres calcula los tres estimadores de momentos dados en las notas en la pág. 3, λ_1, λ_2 y λ_3 .

3.2. Ejercicio 2

Simula una muestra de $n = 15$ variables aleatorias independientes $X_1, ..., X_n$, idénticamente distribuidas como normales con media $\mu = 60$ y parámetro de escala$\sigma = 5$.

Calcula los estimadores de momentos de μ y σ basados en ecuaciones de los primeros dos momentos, los primeros no centrados y los segundos momentos centrados. Denota a estos estimadores como $\hat{\mu}$y $\hat{\sigma}$.

En una misma figura, grafica la función de distribución teórica con línea continua, la distribución estimada con guiones y la función de distribución empírica, graficando puntos de las siguientes coordenadas

$$\left(x_i, \frac{i}{n+1}\right)$$

para $i = 1, ..., n$.

Repite lo mismo pero ahora para $n = 30$ y luego para $n = 100$.

4. Bootstrap no paramétrico

Para esta ayudantía, nos basamos en el libro [1]. Haremos bootstrap no paramétrico para estimar la desviación estándar de dos estimadores de la tasa λ de una v.a. exponencial. Sea $X \sim \exp(\lambda = 5)$. Recordemos

$$\mathbb{E}[X] = \frac{1}{\lambda}$$
$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Utilizaremos dos estimadores obtenidos por el método de momentos,

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}$$
$$\tilde{\lambda} = \sqrt{\frac{n-1}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}}$$

Iniciamos con una muestra aleatoria de $n = 300$ v.a. exponenciales con $\lambda = 5$.

Ahora, definimos las estadísticas que utilizaremos.

Para el bootstrap, tomaremos $k = 100$ muestras de tamaño $m = 30$ para cada una de las estadísticas $\hat{\lambda}$ y $\tilde{\lambda}$.

Finalmente, estimamos la desviación estándar de las estadísticas $\hat{\lambda}$ y $\tilde{\lambda}$.

5. Bootstrap paramétrico

Haremos bootstrap paramétrico para estimar la desviación estándar de dos estimadores de la tasa λ de una v.a. exponencial. Sea $X \sim \exp(\lambda = 5)$. Recordemos

$$\mathbb{E}[X] = \frac{1}{\lambda}\mathrm{Var}(X) \qquad = \frac{1}{\lambda^2}$$

Utilizaremos dos estimadores obtenidos por el método de momentos,

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}$$

$$\tilde{\lambda} = \sqrt{\frac{n-1}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}}$$

Iniciamos con una muestra aleatoria de $n = 300$ v.a. exponenciales con $\lambda = 5$.

Ahora, definimos las estadísticas que utilizaremos.

Para el bootstrap paramétrico, tomaremos $k = 100$ muestras de tamaño $m = 30$ para cada una de las estadísticas $\hat{\lambda}$ y $\tilde{\lambda}$, considerando la tasa $\lambda = \hat{\lambda}$.

Finalmente, estimamos la desviación estándar de las estadísticas $\hat{\lambda}$ y $\tilde{\lambda}$.

6. Bootstrap paramétrico con intervalos de confianza

Haremos bootstrap paramétrico para estimar la desviación estándar del cuantil muestral 0.975 de una muestra normal. Sea $X \sim \text{Normal}(\mu = 0.2, \sigma = 0.2)$.

Nuestro estimador será el cuantil muestral 0.975:

Para el bootstrap paramétrico, tomaremos $k = 100$ muestras de tamaño $m = 30$ para cada el cuantil muestral 0.975, considerando los EMVs de una Normal:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}}$$

Finalmente, estimamos la desviación estándar del cuantil muestral 0.975. Además, tomaremos intervalos del 96 % de confianza considerando el teorema del límite central:

```
S1 <- quantile(meanEst1, 0.02)

## Error in eval(expr, envir, enclos): objeto 'meanEst1' no encontrado

S2 <- quantile(meanEst1, 0.98)

## Error in eval(expr, envir, enclos): objeto 'meanEst1' no encontrado
```

```r
T1 <- mean(meanEst1) + sd(meanEst1) / sqrt(k) * qnorm(0.02, k - 1)

## Error in eval(expr, envir, enclos): objeto 'meanEst1' no encontrado

T2 <- mean(meanEst1) + sd(meanEst1) / sqrt(k) * qnorm(0.98, k - 1)

## Error in eval(expr, envir, enclos): objeto 'meanEst1' no encontrado

print(sprintf("El cuantil empírico 0.02 es %2.3f", S1))

## Error in eval(expr, envir, enclos): objeto 'S1' no encontrado

print(sprintf("El cuantil empírico 0.98 es %2.3f", S2))

## Error in eval(expr, envir, enclos): objeto 'S2' no encontrado

print(sprintf("El cuantil T 0.02 es %2.3f", T1))

## Error in eval(expr, envir, enclos): objeto 'T1' no encontrado

print(sprintf("El cuantil T 0.98 es %2.3f", T2))

## Error in eval(expr, envir, enclos): objeto 'T2' no encontrado
```

Referencias

[1] L. Wasserman, *All of Statistics*. Springer New York, 2004. [Online]. Available: https://doi.org/10.1007/978-0-387-21736-9

[2] Y. Xie, *Dynamic Documents with R and knitr, Second Edition*, 2nd ed., ser. Chapman & Hall/CRC The R Series. Philadelphia, PA: Chapman & Hall/CRC, Jun. 2015.