# Guía 5. Estimación puntual y por intervalo de un cuantil con Bootstrap paramétrico

Para trabajar en clases durante la semana del 19 de noviembre

Considera que tienes una muestra de $n$ variables independientes, $X_1, ..., X_n$, e idénticamente distribuidas como Gama con parámetro de forma $\alpha$ y de escala $\beta$. El valor esperado es $\mu = \alpha\beta$. Se desea estimar el cuantil $Q_{25}$ de probabilidad 0.25. Por máxima verosimilitud de la manera usual, puedes encontrar numéricamente los emv de $\alpha$ y de $\beta$, los cuales se denotarán como $\hat{\alpha}$ y $\hat{\beta}$. Con ellos puedes encontrar el emv del cuantil de interés como

$$\hat{Q}_{25} = F_X^{-1}\left(0.25; \hat{\alpha}, \hat{\beta}\right),$$

donde $F_X^{-1}$ es la función de distribución inversa Gama. Para los siguientes pasos te conviene reparametrizar la logverosimilitud en términos de $\alpha, \mu$.

En esta guía se explicarán los pasos a seguir para obtener un estimador puntual y un intervalo de estimación para este cuantil bajo un método Bootstrap paramétrico.

1. Con los emv de tu muestra original, simula $M = 10,000$ muestras independientes de tamaño $n$, idénticamente distribuidas como Gama con parámetros $\hat{\alpha}$ y $\hat{\beta}$.

2. Para cada muestra simulada, haz lo siguiente:

   (a) Encuentra numéricamente el emv de $\alpha$, a partir de la logverosimilitud perfil de $\alpha$ que tiene la expresión analítica cerrada,

   $$l_p(\alpha) = -n \ln \Gamma(\alpha) + n\alpha \left[\ln(\alpha) - \ln\left(\frac{t_1}{n}\right) + \frac{t_2}{n} - 1\right],$$

   donde $t_1 = \Sigma x_i,\ t_2 = \Sigma \ln x_i$ son las estadísticas suficientes para $(\alpha, \mu)$. Da como valor inicial el estimador de momentos

   $$\tilde{\alpha} = \left[\frac{nt_3}{t_1^2} - 1\right]^{-1},$$

   donde $t_3 = \Sigma x_i^2$.

   (b) Calcula los emv de $\mu$ como $\hat{\mu}^* = t_1/n$ y el de $\beta$ como $\hat{\beta}^* = \hat{\mu}/\hat{\alpha}$. Guarda estos tres emv de $\alpha, \beta, \mu$ denotados aquí como $\left(\hat{\alpha}_i^*, \hat{\beta}_i^*, \hat{\mu}_i^*\right)$, para $i = 1, ..., M$, en las primeras tres columnas de una matriz EMVS de dimensión $M$ por 4.

   (c) Ahora en la cuarta columna de la matriz EMVS guarda el emv del cuantil 0.25, $\hat{Q}_{25i}^*$, para $i = 1, ..., M$.

3. Calcula el emv Bootstrap puntual como el promedio de los valores en la cuarta columna:

$$\hat{Q}_{25}^* = \frac{1}{M}\sum_{i=1}^{M} \hat{Q}_{25i}^*.$$

4. Ahora para calcular el intervalo de estimación con probabilidad de cobertura del verdadero valor del parámetro de $(1-p) = 0.95$ realiza lo siguiente:

   (a) Ordena los valores de la cuarta columna de la matriz EMVS. La distribución Bootstrap del $Q_{25}$ se puede apreciar a través de estudiar el histograma de estos valores.

   (b) Obtén el cuantil empírico de ellos de probabilidad $p/2$ y el de probabilidad $(1-p/2)$. Estos cuantiles se denotarán aquí como $Q_1$ y $Q_2$ respectivamente.

   (c) El intervalo de estimación Bootstrap paramétrico del $(1-p)\,x100\%$ de confianza es $[Q_1, Q_2]$.

Usualmente el estimador Bootstrap y el EMV del cuantil son muy cercanos. Repórtalos juntos para poderlos comparar en el caso de la muestra original que hayas considerado.

**PARA CHECAR EL METODO:**

A manera de probar este método y compararlo contra otro estándar, se sugiere analizar los siguientes datos que son exponenciales, como si hubiesen sido Gama, para estimar el cuartil $Q_{25}$ con el método Bootstrap paramétrico descrito. Nota que una distribución exponencial con media $\theta$ se puede parametrizar en términos de $Q_{25}$ puesto que el parámetro $\theta$ y el cuartil guardan una relación uno a uno ya que:

$$0.25 = G(Q_{25}) = 1 - \exp\left(-\frac{Q_{25}}{\theta}\right).$$

De aquí notar que

$$\theta = \frac{Q_{25}}{-\ln(1-.25)} = \frac{Q_{25}}{-\ln(.75)},$$

por lo que podrás calcular la verosimilitud de $Q_{25}$ a partir de la densidad exponencial dada en términos de $Q_{25}$:

$$f(x;\theta) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right) I_{(0,\infty)}(x) = -\frac{\ln(.75)}{Q_{25}}\exp\left(\frac{x\ln(.75)}{Q_{25}}\right) I_{(0,\infty)}(x).$$

Calcula y grafica la verosimilitud relativa de $Q_{25}$ y encima sobre esta gráfica el intervalo de verosimilitud de nivel $c = 0.1465$ que es el intervalo asintótico de 95% de confianza. A la misma altura de $c$, marca ahora con $+$ los extremos del intervalo Bootstrap que hayas calculado antes para compararlos.

Considera como tu muestra original las siguientes $n = 20$ observaciones que fueron simuladas de una exponencial con tiempo medio de vida $\theta = 10$, donde el cuartil teórico fue $Q_{25} = 2.8768$. Las observaciones son:

1.4717; 3.5573; 12.8264; 3.9891; 21.6200; 22.3458; 8.6898; 2.5651; 3.6359; 4.5532;
1.0141; 43.8509; 3.8434; 2.0186; 1.6106; 10.2787; 17.4981; 9.0456; 4.1304; 2.5836.

Con ellas calcula el intervalo de verosimilitud para $Q_{25}$ de nivel $c = 0.1465$ y grafícalo sobre la verosimilitud perfil de $Q_{25}$. Lo deberás calcular numéricamente. En esta gráfica agrega una línea vertical en guiones ubicada en el emv de $Q_{25}$ y que vaya en $[0,1]$. A la altura de $c$ marca el intervalo que estimes con el método Bootstrap descrito y marca sobre el intervalo el estimador puntual Bootstrap con un asterisco.

Comenta sobre todos tus resultados y concluye sobre las bondades y/o desventajas del método Bootstrap que aplicaste.