

Estadística noparamétrica. Trabajo 1.

Máster en Técnicas Estadísticas. Curso 2009-2010

Intervalos Bootstrap

Vamos a describir el método pivotal bootstrap para construir intervalos para la media μ de una población. Sea X_1, \dots, X_n una muestra aleatoria simple de F y

$$R(X_1, \dots, X_n, F) = \bar{X} - \mu.$$

Si conociésemos c_l, c_u tal que

$$P(c_l \leq R \leq c_u) = 1 - \alpha$$

el intervalo

$$[\bar{X} - c_u, \bar{X} - c_l]$$

es un intervalo para μ de nivel $(1 - \alpha)$. Los valores c_l, c_u son desconocidos y dependen de la distribución F . Por ejemplo si F es la distribución normal y $\alpha = 0,05$ se tiene que $c_l = -1,96\sigma/\sqrt{n}$ y $c_u = 1,96\sigma/\sqrt{n}$, donde σ es la desviación típica poblacional. El remuestreo bootstrap nos permite aproximar estos valores para una distribución arbitraria. El mecanismo es muy sencillo:

1. Dada una muestra X_1, \dots, X_n se calcula la distribución empírica F_n
2. Se genera una muestra bootstrap X_1^*, \dots, X_n^* de F_n
3. Se evalúa $R^* = R(X_1^*, \dots, X_n^*, F_n)$ en la muestra bootstrap obteniendo

$$R^* = \bar{X}^* - \bar{X},$$

donde \bar{X}^* denota la media de la muestra bootstrap X_1^*, \dots, X_n^* .

4. Se repiten B veces los pasos 2 y 3, veces obteniendo los valores r_1^*, \dots, r_B^* (valores de R^* en las remuestras).

5. Finalmente se ordenan de menor a mayor los valores calculados de R^* y se toma el valor que ocupa la posición $\lfloor \alpha/2 * B \rfloor$ como estimación de c_l y el que ocupa la posición $\lfloor (1 - \alpha/2) * B \rfloor$ como estimación de c_u . Si denotamos estos valores mediante c_l^* y c_u^* entonces el intervalo bootstrap será

$$[\bar{x} - c_u^*, \bar{x} - c_l^*]$$

Ejercicio.

1. Comprueba el funcionamiento del método anterior cuando X_1, \dots, X_n es una muestra aleatoria simple de tamaño $n = 20$ de una distribución Exponencial de parámetro 1. Para ello construye $M = 1000$ intervalos de confianza de nivel 95% (toma $B = 500$) y calcula el porcentaje de veces en que $\mu = 1$ está en el intervalo. Este porcentaje debería de estar cerca del nivel de confianza del intervalo.
2. Compara el método pivotal (algoritmo anterior) con el bootstrap estudentizado y con el método basado en la aproximación normal.
3. ¿Qué método crees que es el mejor?
4. ¿Sucede lo mismo si la población sigue una distribución χ^2 con un grado de libertad?