

Corrección de sesgo en Bootstrap

María Fernanda Mora Alba

Noviembre, 2015

1 Explicación

Los métodos de Bootstrap se han vuelto muy populares en inferencia estadística y son aplicados en diversas áreas para obtener estimadores de errores estándar, intervalos de confianza, sesgos, etc [1].

Sea z_1, z_2, \dots, z_n una muestra *i.i.d.* de una distribución F . El objetivo es evaluar la precisión de un estimador $\hat{\theta} = t(z_1, \dots, z_n)$ al estimar el parámetro $\theta = t(F)$. Esto es, el sesgo de $\hat{\theta}$ dado por $Sesgo_F(\hat{\theta}) = E_F(\hat{\theta}) - \theta$. Sin embargo estas cantidades son desconocidas, así que las estimaremos de la siguiente manera:

Sean z_1^*, \dots, z_B^* muestras Bootstrap independientes. Sean ahora $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ los valores del estadístico $\hat{\theta}$ en cada una de las muestras Bootstrap. Usando estas cantidades es como vamos a estimar el sesgo del estimador $\hat{\theta}$. Finalmente sea \hat{F} la distribución no paramétrica que estima F .

En particular, la estimación Bootstrap del sesgo es como sigue: $Sesgo_{\hat{F}}(\hat{\theta}) = E_{\hat{F}}(\hat{\theta}^*) - \theta_{\hat{F}}$

En la práctica $E_{\hat{F}}(\hat{\theta}^*)$ se estima promediando $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ sobre un número grande B de muestras Bootstrap. Entonces el estimador queda: $\hat{Sesgo}_B(\hat{\theta}) = \overline{\hat{\theta}^*} - \theta_{\hat{F}}$.

Conforme $B \rightarrow \infty$ se tiene que $\hat{Sesgo}_B(\hat{\theta}) \rightarrow Sesgo_{\hat{F}}(\hat{\theta})$.

Una vez que se ha estimado el sesgo, se corrige el estimador inicial restando el sesgo que se acaba de estimar. De modo que el estimador de θ corregido por sesgo usando Bootstrap queda así:

$$\hat{\theta}^{BC} = \hat{\theta} - \hat{Sesgo}_B(\hat{\theta}) = 2\hat{\theta} - \overline{\hat{\theta}^*} \quad (1)$$

Existen otras correcciones, por ejemplo la multiplicativa, en donde el estimador de θ corregido por sesgo usando Bootstrap queda así:

$$\hat{\theta}^{BC} = \hat{\theta}^2 / \overline{\hat{\theta}^*} \quad (2)$$

2 Uso del Shiny

Para ver en acción la corrección del sesgo se hizo un Shiny.

Para hacer más interesante el ejemplo, primero se implementó el Ejemplo 9.3 del libro [2] para obtener estimaciones Bootstrap de $\theta = \beta_1 / \beta_0$.

Los datos observados nos dan una estimación de $\hat{\theta} = \hat{\beta}_1 / \hat{\beta}_0 = -0.185$.

Se van a remuestrear con reemplazo 13 elementos del conjunto de datos originales y se estima θ . Así para tantas simulaciones como el usuario haya escogido.

Posteriormente se dibuja el histograma de estos datos (gráfica blanca).

Después sigue la corrección del sesgo usando la fórmula (1) y Gráfica azul.

Claramente la gráfica azul está menos sesgada que la blanca, por lo que la corrección del sesgo fue exitosa.

Entonces, únicamente hay que seleccionar:

1. El tamaño de muestra deseado (grande).

El ejemplo que se implementó fue el del libro [2] de la Sección 9.2.4. y el Shiny grafica con sesgo y sin sesgo.

References

- [1] Corrêa, Solange. (2009). *Extended Bootstrap Bias Correction with Application to Multilevel Modelling of Survey Data under Informative Sampling*. Brazilian Institute of Geography and Statistics, Rio de Janeiro.
- [2] Geof Givens, J. Hoeting. (2013). *Computational Statistics*. Department of Statistics, Colorado State University, Fort Collins, CO.