Ejercicios: bootstrap

1. Se extrae una remuestra bootstrap de una muestra de n observaciones $X_1, \ldots, X_n$. Calcula la probabilidad de que una observación prefijada, X_j, no aparezca en la muestra bootstrap. Calcula el límite de esta probabilidad si $n \to \infty$.

2. Sea $X_1, \ldots, X_n$ una muestra de n observaciones iid de una distribución F con esperanza μ y varianza σ^2, y sea $X_1^*, \ldots, X_n^*$ una muestra de n observaciones iid de la distribución empírica de la muestra original F_n. Calcula las siguientes cantidades:

(a) $\mathrm{E}_{F_n}(\bar{X}_n^*) := \mathrm{E}(\bar{X}_n^* | X_1, \ldots, X_n)$.
(b) $\mathrm{E}_F(\bar{X}_n^*)$.
(c) $\mathrm{Var}_{F_n}(\bar{X}_n^*) := \mathrm{Var}(\bar{X}_n^* | X_1, \ldots, X_n)$.
(d) $\mathrm{Var}_F(\bar{X}_n^*)$.

3. Dada una muestra de n datos diferentes, calcula en función de n el número de remuestras bootstrap distintas que es posible obtener. Aplica la expresión obtenida al caso $n = 15$. ¿Qué implicación práctica tiene el resultado?

4. Consideremos la siguiente muestra de tamaño $n = 10$:

$$1 \quad 2 \quad 3.5 \quad 4 \quad 7 \quad 7.3 \quad 8.6 \quad 12.4 \quad 13.8 \quad 18.1$$

Sea $\hat{\theta}$ la media recortada al 40 % que se obtiene al eliminar los dos mayores y los dos menores datos y calcular el promedio de los 6 datos restantes. Sea $\hat{\sigma}_R$ el estimador bootstrap de la desviación típica de $\hat{\theta}$ basado en R remuestras. Calcula $\hat{\sigma}_R$ para $R = 10, 100, 1000, 2000$ y usando 10 conjuntos independientes de R remuestras. ¿A qué valor parecen converger los valores obtenidos? ¿Qué número de remuestras te parece suficiente?

5. Sea S^2 la varianza muestral de una muestra de vaiid $X_1, \ldots X_n$ de una distribución con varianza σ^2.

(a) Para la muestra del problema anterior se tiene $S^2 \approx 30{,}84$. Usa bootstrap para determinar el error típico de este estimador de σ^2.
(b) Compara el resultado con el error típico que darías si, por ejemplo, supieras que los datos proceden de una distribución normal.
(c) Calcula un intervalo de confianza para σ^2 usando el método bootstrap híbrido. Fija $1 - \alpha = 0{,}95$.

6. Sea $\hat{\sigma}_B^2$ el estimador bootstrap de la varianza de un estimador $\hat{\theta}$ basado en B remuestras y sea $\hat{\sigma}_n^2$ el estimador bootstrap ideal. Demuestra que $\mathrm{E}_F[\hat{\sigma}_B^2] = \mathrm{E}_F[\hat{\sigma}_n^2]$ pero $\mathrm{Var}_F[\hat{\sigma}_B^2] \geq \mathrm{Var}_F[\hat{\sigma}_n^2]$. (Indicación: condiciona a la muestra).

7. Sea F una distribución con media μ, varianza σ^2 y coeficiente de asimetría

$$\gamma = \mathrm{E}_F[(X-\mu)^3]/\sigma^3.$$

Genera $R = 1000$ muestras de observaciones iid $X_1, \ldots, X_n$ con $X_i \equiv \mathrm{N}(0,1)$ para $n = 100$. Para cada una de ellas, calcula tres intervalos de confianza bootstrap de nivel 95 % para γ usando el método híbrido, el método normal y el método percentil. Determina el porcentaje de intervalos que contienen al parámetro en cada caso. Repite el ejercicio con muestras procedentes de una distribución exponencial de parámetro $\lambda = 1$.