
Aclaraciones:

- El TP se puede hacer **individualmente o de a dos**.
- Los razonamientos y cálculos realizados se pueden presentar en una hoja (prolijamente) o en Latex. Se debe enviar también el script de R que usaron para resolver la parte de programación y algún archivo que contenga el gráfico del último item (con su interpretación).
- Se puede usar cualquier material, incluyendo lo visto en clase, Google, Wikipedia, el help de R, etc.
- La fecha límite de entrega es el **jueves 13 de octubre**.

Supongamos que tenemos $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $N(0, \sigma^2)$ con σ^2 desconocido. Nos interesa estimar σ^4.

1. Usando algún método visto en clase, hallar un estimador $\delta = \delta(X_1, \ldots, X_n)$ IMVU para σ^4.
2. Hallar la esperanza y la varianza de δ (en función de σ^2). Decidir si dicho estimador es consistente y, en caso afirmativo, decidir si esa consistencia es fuerte o débil.
   Sugerencia: Probar y usar que
   $$\sum_{i=1}^{n} X_i^2 \sim \sigma^2 \chi_n^2.$$
3. Hallar la distribución asintótica de δ.
4. Supongamos que $n = 12$ y que la muestra observada es

   3.87 4.37 -0.59 -2.31 2.74 -4.93 -4.41 0.63 2.89 14.70 -6.16 -5.56.

   Queremos tener distintas estimaciones para la varianza de δ (que de ahora en más llamaremos v) en este caso.

   (a) Estimar σ^2 (usando un estimador IMVU para este parámetro) y estimar v usando el item 2.

   (b) Estimar v usando la distribución asintótica de δ hallada en el punto 3. Si necesita estimar a una función $g(\sigma^2)$, hágalo tomando $g(\widehat{\sigma^2})$ donde $\widehat{\sigma^2}$ es el estimador IMVU de σ^2.

   (c) Estimar v haciendo un Bootstrap no paramétrico de los datos (es decir, remuestreando de la distribución empírica F_n). Antes de generar las muestras Bootstrap, fijar la semilla usando la función `set.seed()`. Fijar esta semilla fija la "aleatoriedad" con la que se generan los datos y de esta forma se puede replicar la simulación.

   (d) Estimar v haciendo un Bootstrap paramétrico (es decir, generando nuevas muestras de tamaño n con distribución $N(0, \widehat{\sigma^2})$). Al igual que en el item anterior, fije la semilla con las que genera las muestras Bootstrap.

5. Vamos a realizar el siguiente estudio de simulación: suponemos que el σ^2 real que genera los datos es igual a 1 y para cada uno de los valores de $n = 10, 30, 80, 150, 300$ hacemos lo siguiente:

   (a) Hallar la varianza exacta de δ cuando σ^2 es igual a 1.

   (b) Generar una muestra de tamaño n de una distribución $N(0, 1)$. Usando esta muestra, hacer las mismas cuatro estimaciones de v que hicimos en el item 4.

   (c) Repetir el item anterior 1000 veces, obteniendo así 1000 realizaciones para cada uno de los cuatro estimadores de v. Fijar la semilla que genera las 1000 muestras de tamaño n.

   (d) Para cada uno de los cuatro métodos, hacer un histograma de las 1000 realizaciones obtenidas. ¿Qué estima este histograma?

   (e) Usando las 1000 replicaciones para cada método y el valor exacto de v hallado en 5a, estimar el error cuadrático medio (ECM) para cada uno de los cuatro métodos.

   (f) Poner en un mismo gráfico y en función de n los valores estimados de los ECM de los cuatro métodos. En base al gráfico, decidir cual de los cuatro métodos usaría.

6. Queremos ver como funcionan los estimadores de v propuestos cuando el modelo especificado es incorrecto. Nos restringimos a un caso donde los elementos de la muestra siguen una distribución mezcla de normales

$$X = WY + (1 - W)Z$$

donde $W \sim Ber(\frac{1}{2})$, $Y \sim \mathcal{N}(1, 1)$ y $Z \sim \mathcal{N}(-1, 1)$ todas independientes.

(a) Calcular $\mathbb{E}(X)$. Realizar un histograma para $n = 2000$ observaciones de X.

(b) Se quiere obtener el valor exacto de v para muestras de tamaños $n = 10$, 20, 30, 80, 150, 300. Como la cuenta teórica es complicada vamos estimarlo con una simulación de Monte-Carlo. Para hacer esto calcular δ sobre $N = 10000$ muestras de tamaño n y obtener su varianza muestral.

(c) Haga las mismas estimaciones que en el punto 5 generando los datos bajo este modelo. Utilice como valor de v verdadero los obtenidos en el item anterior.

(d) Haga una tabla con los errores cuadraticos medios estimados para cada método y valor de n. Que método elegiría?