

## Aclaraciones:

- El TP se puede hacer en grupos de hasta 3 personas.
- Se debe presentar un informe explicando el desarrollo del trabajo (no hace falta explicar el código) y que incluya los gráficos y/o tablas pedidas. Las consignas que incluyan cuentas pueden ser incluídas en el informe o se pueden presentar escritas prolijamente en una hoja aparte. Se debe enviar también por separado el script de R que usaron para resolver la parte de programación.
- Se puede usar cualquier material, incluyendo lo visto en clase, Google, Wikipedia, el help de R, etc.
- La fecha límite de entrega es el **miércoles 18 de octubre**.

Supongamos que tenemos  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con densidad

$$f_X(x, \theta) = \theta x^{\theta-1} \mathbb{I}_{x \in (0,1)}.$$

con  $\theta > 0$  desconocido.

1. Usando algún método visto en clase, hallar un estimador IMVU para  $\theta$ . Llamamos  $\hat{\theta}_{IMVU}$  a este estimador.
2. Hallar el estimador de momentos para  $\theta$  basado en el segundo momento. Llamamos  $\hat{\theta}_{MO}$  a este estimador. Probar que es fuertemente consistente.
3. Hallar la distribución asintótica de  $\hat{\theta}_{MO}$ .
4. En muchas situaciones (como por ejemplo para construir intervalos de confianza), es importante conocer la varianza de un cierto estimador. En lo que sigue del trabajo, nos va a interesar estimar la cantidad  $v = \text{VAR}(\hat{\theta}_{MO})$ . Sin embargo, no podemos conocer el valor de  $v$  por dos motivos: primero, depende de  $\theta$ , que es un valor desconocido. Segundo, incluso si conociéramos  $\theta$ , no tenemos una fórmula explícita que permita calcular  $v$  a partir de  $\theta$ . Supongamos que observamos los siguientes datos:

0.342 0.841 0.807 0.626 0.861 0.738 0.897 0.215 0.776 0.315 0.851 0.441

A partir de esta muestra, se pide lo siguiente

- (a) Estimar  $v$  a partir de distribución asintótica de  $\hat{\theta}_{MO}$ . Si fuera necesario, usar  $\hat{\theta}_{IMVU}$  como estimador de  $\theta$ .
- (b) Estimar  $v$  a partir de distribución asintótica de  $\hat{\theta}_{MO}$ . Si fuera necesario, usar  $\hat{\theta}_{MO}$  como estimador de  $\theta$ .

- (c) Estimar  $v$  con Bootstrap paramétrico, considerando nuevas muestras obtenidas a partir de la densidad  $f(x, \hat{\theta}_{IMVU})$ .
- (d) Estimar  $v$  con Bootstrap paramétrico, considerando nuevas muestras obtenidas a partir de la densidad  $f(x, \hat{\theta}_{MO})$ .
- (e) Estimar  $v$  con Bootstrap no paramétrico, es decir, remuestreando de la distribución empírica  $F_n$ .
5. Programar una función que dado un valor de  $n$  y el valor verdadero de  $\theta$ , aproxime el valor verdadero de  $v$ . Para esto, generar  $m = 10000$  muestras de tamaño  $n$  de la densidad  $f(x, \theta)$ , calcular los  $m$  valores del estadístico  $\hat{\theta}_{MO}$  y luego tomar la varianza de este vector de  $m$  elementos. Al valor obtenido lo vamos a llamar “varianza oráculo”, porque la pudimos calcular suponiendo conocido el valor de  $\theta$ .
6. Vamos a realizar el siguiente estudio de simulación: suponemos que el  $\theta$  real que genera los datos es igual a 2 y para cada uno de los valores de  $n = 10, 30, 70, 120, 200$  hacemos lo siguiente:
- Calcular la varianza oráculo para  $n$  y  $\theta = 2$ .
  - Generar una muestra de tamaño  $n$  de una distribución con densidad  $f(x, \theta = 2)$ . Usando esta muestra, hacer las mismas cinco estimaciones de  $v$  que hicimos en el ítem 4.
  - Repetir el ítem anterior 1000 veces, obteniendo así 1000 realizaciones para cada uno de los cinco estimadores de  $v$ .
  - Para cada uno de los cinco métodos, hacer un histograma de las 1000 realizaciones obtenidas. ¿Qué estima este histograma?
  - Usando las 1000 replicaciones para cada método y el valor oráculo de  $v$  hallado en 6a, estimar el logaritmo del error cuadrático medio (ECM) para cada uno de los cinco métodos.
  - Poner en un mismo gráfico y en función de  $n$  los valores estimados de los logaritmos de los ECM de los cinco métodos. En base al gráfico, decidir cual de los cinco métodos usaría.
7. **(Opcional)** Extender los resultados de este trabajo en alguna de las muchas opciones posibles. Algunas ideas:
- Ver cómo varían los resultados del ítem 6 si se mueve el valor de  $\theta$ .
  - Ver cómo son afectados los diferentes métodos de estimación de  $v$  si las muestras tienen un porcentaje  $\epsilon$  de contaminación.
  - Inventar algún otro método para estimar  $v$ , en lo posible que le gane a los otros cinco métodos en el ítem 6.
  - Encontrar alguna otra distribución para las muestras que haga que la conclusión del trabajo sea distinta.