Introducción a la Estadística y Ciencia de Datos

Práctica 2 - Estimadores

Los ejercicios marcados con (*) son optativos, pueden dejarse para la etapa de repaso antes del parcial.

2.1. Estimadores de Momentos y de Máxima Verosimilitud

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución

   - $\mathcal{B}(1,\theta)\,,\ 0<\theta<1.$
   - $\mathcal{P}(\theta)\,,\ \theta>0.$
   - $\mathcal{E}(\theta)\,,\ \theta>0.$

   En cada uno de estos casos, encontrar:

   (a) Un estimador de θ basado únicamente en el primer momento.

   (b) Un estimador de θ basado únicamente en el segundo momento.

2. Sea $X_1, \ldots, X_n$ una muestra aleatoria, se busca un estimador basado en momentos para $q(\theta)$, donde θ es un parámetro desconocido de la distribución de la muestra y q es una función conocida. Se consideran dos estrategias:

   (i) Estimar $q(\theta)$ con $q(\hat{\theta})$, con $\hat{\theta}$ un estimador de momentos para θ (estimador plug-in).

   (ii) Buscar una función g tal que $\mathbf{E}[g(X_1)] = q(\theta)$ y luego estimar $\mathbf{E}[g(X_1)]$ con $\frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

   Este procedimiento se conoce como el **método generalizado de los momentos** (MGM).

   Aplicar las estrategias (i) y (ii) en los siguientes casos:

   (a) $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{E}(\theta)$ y $q(\theta) = P_\theta(X_1 \geq 1)$.

   (b) $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(0,\sigma^2)$, estimar σ.
   Sugerencia: Calcular $E(|X|)$.

3. (*) Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas. Hallar el estimador de los momentos de μ si

   (a) la densidad de X_1 está dada por:

   $$f(x) = \frac{1}{\mu^4} x I_{[0,\mu^2]}(x) + \frac{1}{\mu^4}(2\mu^2 - x) I_{[\mu^2, 2\mu^2]}(x) \qquad \mu > 0$$

   (b) la función de probabilidad de X_1 está dada por:

k	1	2	4
$p(k)$	μ^2	$2\mu(1-\mu)$	$(1-\mu)^2$

   $0<\mu<1$

4. Una moneda tiene una probabilidad de cara p, $p \in \{2/5, 4/5\}$. En 10 lanzamientos de la moneda se observaron exactamente 3 caras. Estimar por máxima verosimilitud la probabilidad de cara en base a la muestra observada.

5. Dada una muestra aleatoria $X_1, \ldots, X_n$ para cada una de las distribuciones consideradas en el ejercicio 1, encontrar los respectivos estimadores de máxima verosimilitud (EMV) para θ. Calcular el sesgo y la varianza para cada uno de los estimadores. ¿Coinciden con los estimadores basados en el primer momento? Calcular el sesgo y la varianza para cada uno de los estimadores de máxima verosimilitud.
   Sugerencias:
   - Recordar la siguiente propiedad de la distribución Gamma, si $Y_1 \sim \Gamma(\alpha_1, \lambda)$ e $Y_2 \sim \Gamma(\alpha_2, \lambda)$ independientes, entonces $Y_1 + Y_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.
   - Para calcular la $E(1/Y)$ y $E(1/Y^2)$ basta plantear la integral y darse cuenta de que queda la esperanza de una variable aleatoria conocida.

6. Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con función de densidad o función de probabilidad puntual f_θ con $\theta \in \Theta \subset \mathbb{R}^p$ y sea $\tau : \Theta \to \Lambda \subset \mathbb{R}^k$ con $k \leq p$. Siempre podemos asumir que la función τ es sobreyectiva, sino eso se puede conseguir tomando al conjunto Λ como la imagen de Θ por τ. Queremos probar la siguiente **propiedad de invarianza de los EMV**: Si $\widehat{\theta}$ es el EMV de θ, entonces, para cualquier función $\tau(\theta)$, el EMV de $\tau(\theta)$ es $\tau\left(\widehat{\theta}\right)$.
   Observemos que no se requiere que τ sea inyectiva. Para el caso en el que τ no es inyectiva para poder definir el EMV necesitamos introducir una generalización de la función de verosimilitud (ver Casella & Berger, *Statistical Inference*, 2nd. edition, sección 7.2)
   Para $\lambda = \tau(\theta)$ definimos la verosimilitud inducida L^* dada por

$$L^*(\lambda \,|X_1, \ldots, X_n) = \sup_{\{\theta:\tau(\theta)=\lambda\}} L(\theta \,|X_1, \ldots, X_n). \tag{1}$$

   donde $L(\theta \,|X_1, \ldots, X_n)$ es la función de verosimilitud correspondiente a la muestra $X_1, \ldots, X_n \sim f_\theta$. El valor $\widehat{\lambda}$ que maximiza a $L^*(\lambda \,|X_1, \ldots, X_n)$ se denominará el EMV de $\lambda = \tau(\theta)$. Observemos que de (1) se deduce que el máximo de L y L^* coinciden.

   (a) Sean $X_1, \ldots, X_n \sim Bi(1, \theta)$, queremos estimar por máxima verosimilitud a la varianza de X_i, $\lambda = \tau(\theta) = \theta(1-\theta)$. Observar que no se puede parametrizar a la familia de distribuciones $\mathcal{F} = \{Bi(1,\theta), \theta \in \Theta = (0,1)\}$ con el parámetro λ. Hallar la función de verosimilitud inducida correspondiente a λ basada en la muestra $X_1, \ldots, X_n \sim Bi(1,\theta)$. Concluir que cuando τ no es inyectiva, para un valor de λ puede haber más de un valor de θ que cumpla $\lambda = \tau(\theta)$.

   (b) Ahora sí probemos la invarianza en el caso τ general. Sea $\widehat{\lambda}$ el valor que maximiza a $L^*(\lambda \,|X_1, \ldots, X_n)$. Queremos ver que

$$L^*\left(\widehat{\lambda} \,|X_1, \ldots, X_n\right) = L^*\left(\tau\left(\widehat{\theta}\right) \,|X_1, \ldots, X_n\right) \tag{2}$$

   Lo hacemos en dos partes.

i. Justificar las siguientes igualdades, indicando las opciones A) - D) según sea la opción correcta,

$$\begin{aligned}
L^*\left(\widehat{\lambda}\,|X_1,\ldots,X_n\right) &= \sup_{\lambda} L^*\left(\lambda\,|X_1,\ldots,X_n\right) \\
&= \sup_{\lambda}\sup_{\{\theta:\tau(\theta)=\lambda\}} L\left(\theta\,|X_1,\ldots,X_n\right) \\
&= \sup_{\theta} L\left(\theta\,|X_1,\ldots,X_n\right) \\
&= L\left(\widehat{\theta}\,|X_1,\ldots,X_n\right).
\end{aligned}$$

A. por definición de $\widehat{\theta}$.

B. por definición de $\widehat{\lambda}$.

C. por definición de L^*.

D. porque la maximización iterada coincide con la maximización no condicional sobre θ que se alcanza en $\widehat{\theta}$.

ii. Justificar las siguientes igualdades, indicando las opciones A) - B) según sea la opción correcta,

$$\begin{aligned}
L\left(\widehat{\theta}\,|X_1,\ldots,X_n\right) &= \sup_{\left\{\theta:\tau(\theta)=\tau\left(\widehat{\theta}\right)\right\}} L\left(\theta\,|X_1,\ldots,X_n\right) \\
&= L^*\left(\tau\left(\widehat{\theta}\right)\,|X_1,\ldots,X_n\right).
\end{aligned}$$

A. porque $\widehat{\theta}$ es el EMV.

B. por definición de L^*.

iii. Observar que de las dos igualdades probadas en los items 6(b)i y 6(b)ii se obtiene que $\tau\left(\widehat{\theta}\right)$ es el EMV de $\tau(\theta)$.

7. Sea $X_1,\ldots,X_n$ una m.a. de una población con distribución $\mathcal{N}(\mu,\sigma^2)$. Sea el parámetro bivariado $\theta=(\mu,\sigma^2)$.

(a) Hallar el EMV de θ.

(b) Dado $\xi\in\mathbb{R}$, hallar el EMV de $p=P_\theta(X_1>\xi)$. Asimismo, también estimar esta probabilidad usando la estrategia del ejercicio 2 (ii) de esta práctica, es decir, usando el MGM que resulta ser la frecuencia relativa.

(c) Dado $p\in(0,1)$, hallar el EMV del valor ξ tal que $P_\theta(X_1>\xi)=p$. Notar que $\widehat{\xi}$ es un estimador paramétrico de ξ, es decir del cuantil p de la distribución.

(d) Hallar el EMV de μ cuando σ^2 es conocido. ¿Es razonable que no dependa de σ^2?

(e) Hallar el EMV de σ^2 cuando μ es conocido. ¿Es razonable que dependa de μ?

(f) (Para hacer en R) El índice de colesterol en cierta población es una variable aleatoria con distribución $\mathcal{N}(\mu,\sigma^2)$. Se hacen análisis a 25 personas elegidas al azar de la población y se obtienen los siguientes datos:

1.53	1.65	1.72	1.83	1.62	1.75	1.72	1.68	1.65	1.61
1.70	1.60	1.73	1.61	1.52	1.81	1.72	1.50	1.51	1.65
1.58	1.82	1.65	1.72	1.65					

i) Estimar μ y σ^2 por máxima verosimilitud basados en la muestra dada.

ii) Se considera que el índice es típico si es menor a 1.73. Estimar la proporción de la población con un índice típico usando los dos estimadores propuestos en b).

8. (*) El tiempo de respuesta del servidor A es una variable aleatoria X con densidad

$$f_X(x,\theta) = \theta^4 x e^{-\theta^2 x}\mathbf{I}_{(0,+\infty)}(x) \quad \theta > 0$$

Por otro lado, el tiempo de respuesta del servidor B es una variable aleatoria Y con densidad

$$f_Y(y,\theta) = \frac{2y}{\theta^2}\mathbf{I}_{(0,\theta)}(y) \quad \theta > 0$$

(a) Hallar el estimador de máxima verosimilitud de θ para las densidades de cada servidor.

(b) Se observaron 20 tiempos de respuesta para cada servidor:

A :1.49 1.12 1.08 0.65 0.98 0.860.37 0.41 0.91 0.85

0.25 0.78 0.30 1.41 0.18 0.52 0.40 0.690.73 0.81

B :1.32 0.59 1.41 1.15 0.34 1.16 1.29 1.25 1.18 0.27

1.31 0.96 0.57 0.66 0.67 1.34 0.47 1.07 1.13 0.25

Hallar, para cada servidor, la estimación de θ por el método de máxima verosimilitud y una estimación del tiempo esperado de respuesta basada en los estimadores de máxima verosimilitud.

9. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad

$$f(x) = e^{-(x-\theta)}\mathbf{I}\{x > \theta\}, \qquad \theta > 0.$$

(a) Encontrar el EMV de θ.

(b) Encontrar el estimador de los momentos de θ.

2.2. Error cuadrático medio y propiedades de los estimadores

10. Sea $X_1, \ldots, X_n$ es una muestra aleatoria de una población con distribución $\mathcal{U}[0,\theta]$

(a) Hallar el EMV de θ.

(b) Hallar el estimador de los momentos de θ.

(c) Calcular el sesgo para ambos estimadores. Probar que el estimador de momentos es insesgado y que el estimador de máxima verosimilitud es asintóticamente insesgado pero no insesgado.

(d) Corregir el EMV para que resulte insesgado. Llamemos $\widehat{\theta}_n^{\text{mod}}$ a este estimador modificado.

(e) Hallar la varianza para los tres estimadores propuestos.

(f) Calcular el ECM de los tres estimadores. Hallar el límite de los ECM cuando $n \to \infty$. ¿Cuál de ellos va más rápido a cero? Desde el punto de vista del mejor ECM, ¿cuál de los tres estimadores prefiere?

11. (*) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad de la forma

$$f_\theta(x) = 3\,\theta\,x^2\,e^{-\theta x^3}\,\mathbb{I}_{(0,+\infty)}(x), \qquad \theta > 0.$$

(a) Hallar el estimador de máxima verosimilitud para θ, que denominaremos $\hat{\theta}_{MV}$.

(b) Calcular el error cuadrático medio de $\hat{\theta}_{MV}$.
Sugerencia: ¿Qué distribución tiene X^3?

(c) Decidir si $\hat{\theta}_{MV}$ es un estimador fuertemente consistente. Justificar.

12. Sean $X_1, \ldots, X_n$ una m.a. de una distribución $N(\mu_1, \sigma^2)$ e $Y_1, \ldots, Y_m$ una m.a. de una distribución $N(\mu_2, \sigma^2)$, independientes entre sí.

(a) Hallar el EMV de $\theta = (\mu_1, \mu_2, \sigma^2)$. Observar que

$$\widehat{\sigma}^2 = \frac{n\,\widehat{\sigma}_X^2 + m\,\widehat{\sigma}_Y^2}{n+m}$$

donde $\widehat{\sigma}_X^2$ es el EMV de σ^2 basado en la muestra aleatoria $X_1, \ldots, X_n$ con distribución $N(\mu_1, \sigma^2)$, y $\widehat{\sigma}_Y^2$ es el EMV de σ^2 basado en $Y_1, \ldots, Y_m$ m.a. de una distribución $N(\mu_1, \sigma^2)$.

(b) Hallar el EMV de $\alpha = \mu_1 - \mu_2$.

13. Se tienen observaciones independientes $X_1, \ldots, X_n$ de poblaciones normales con la misma media μ pero con varianzas $\sigma_1^2, \ldots, \sigma_n^2$ respectivamente.

(a) ¿Es posible estimar todos los parámetros por máxima verosimilitud?

(b) Suponiendo $\sigma_1^2, \ldots, \sigma_n^2$ conocidos, hallar el EMV de μ. Interpretar.

14. *Para resolver con R*

(a) Generar $n = 100$ datos como si provinieran de una muestra aleatoria $X_1, ... X_n$ donde X_i es una variable aleatoria con distribución $\varepsilon(\lambda)$ con $\lambda = 1$ y graficar su función de distribución empírica junto con la verdadera función de distribución acumulada. En adelante, para simplificar diremos "*Generar una muestra de tamaño n de una distribución* $\varepsilon(\lambda)$". Nos referiremos a los valores como *observaciones* o *datos*.

(b) Para las 100 observaciones obtenidas en la parte a), calcular el estimador de λ por el método de los momentos y por máxima verosimilitud.

(c) Proponer un estimador de la función de distribución acumulada distinto al obtenido en a) y superponerlo en el gráfico anterior. ¿Qué observa? ¿Parece su propuesta mejorar la estimación dada en a)?

(d) Repetir los puntos anteriores utilizando $n = 10$ y $n = 1000$.

15. (Para hacer con R). Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathcal{U}(0, \theta)$, sea $\widehat{\theta}_n^{MV}$ el EMV de θ, $\widehat{\theta}_n^{\mathrm{m}}$ el estimador de θ basado en el primer momento y $\widehat{\theta}_n^{\mathrm{mod}}$ el estimador modificado calculados en el ejercicio 10.

   (a) Para $n = 6$ hacer lo siguiente:

      i. Generar una muestra de tamaño n de una distribución $U(0, \theta_0)$ con $\theta_0 = 3$.

      ii. Evaluar los tres estimadores $\widehat{\theta}_n^{MV}$, $\widehat{\theta}_n^{\mathrm{m}}$ y $\widehat{\theta}_n^{\mathrm{mod}}$ en dicha muestra.

      iii. Repetir los dos pasos anteriores $k = 1000$ veces, obteniendo así, para cada uno de los tres estimadores, replicaciones $\widehat{\theta}_{[1]}^*, \ldots, \widehat{\theta}_{[k]}^*$ y guardar estos valores en un vector.

      iv. Repetir lo anterior para $n = 10, 20, 40, 80, 200$.

      v. Observar que para cada uno de los estimadores y para cada n fijo, $\widehat{\theta}_{[1]}^*, \ldots, \widehat{\theta}_{[k]}^*$ es una muestra aleatoria de la distribución del estimador respectivo. Probar que el estimador basado en el primer momento del ECM $\left(\widehat{\theta}^*,\, \theta_0\right)$ cuando θ_0 es conocido es

$$\frac{1}{k}\sum_{i=1}^{k}(\theta_i^* - \theta_0)^2$$

      Calcularlo para cada estimador, y para cada n.

   (b) Finalmente, se pide realizar los cuatro gráficos siguientes.

      i. Los boxplots de los EMV simulados versus los tamaños de muestra n, los seis boxplots en un mismo gráfico. Agregarle al gráfico una línea (horizontal) a la altura del θ_0 con un color distinto.

      ii. Ídem para los estimadores de momentos.

      iii. Ídem para los estimadores modificados.

      iv. En un mismo gráfico hacer 3 curvas

         - el ECM estimado del EMV en el eje y, versus n en el eje x, en negro
         - el ECM estimado del estimador de momentos, versus n, en rojo
         - el ECM estimado del estimador modificado, versus n, en azul.

      En base a los gráficos realizados, ¿tienen sentido estos resultados respecto de los resultados teóricos calculados en el ejercicio 10?

16. La cantidad de lluvia (en milímetros) que cae en una tormenta en cierta región es una variable aleatoria con distribución $\Gamma(\alpha, \lambda)$. Se seleccionan al azar 200 tormentas en la región, la cantidad de lluvia observada en cada una de ellas está guardada en el archivo `tormentas.txt`.

(a) Hallar estimaciones basadas en el método de momentos y de máxima verosimilitud para α y λ. Para el estimador de MV, explorar el comando `fitdistr(x, densfun = "gamma")` del paquete MASS.

(b) Estimar por máxima verosimilitud la probabilidad de que la cantidad de lluvia en una tormenta en esta región supere los 20 milímetros. Hallar su valor observado en la muestra. Repetir haciendo "plug-in" con los estimadores de momentos. Y también estimar esta probabilidad usando la frecuencia relativa.

(c) Realizar un histograma de las observaciones, superponer una estimación no paramétrica de la densidad, y en otro color la función de densidad gamma con los parámetros estimados por el método de los momentos. También con un color diferente superponer la densidad usando los parámetros estimados por el método de máxima verosimilitud. ¿Qué observa?

2.3. Distribución asintótica de los estimadores

17. Para el ejercicio 2 de la Práctica 2.

(a) Sean $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{E}(\theta)$ y $q(\theta) = P_\theta(X_1 \geq 1)$. Hallar la distribución asintótica de los dos estimadores de $q(\theta) = P_\theta(X \geq 1)$ propuestos en el ejercicio 2a,

$$T_1 = e^{-1/\widetilde{X}_n} \qquad \text{y} \qquad T_2 = \frac{1}{n}\sum_{i=1}^{n} I_{[1,+\infty)}(X_i)\,.$$

Dado que calcular el ECM de T_1 no es una tarea sencilla, podemos comparar el desempeño de ambos estudiando las dos distribuciones asintóticas. Hallar las distribuciones asintóticas de ambos estimadores de $q(\theta)$ e indicar el que prefiere.

(b) Ídem con el ejercicio 2b donde $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$ y se desea estimar σ.

18. Sean $X_1, \dots, X_n \sim \mathcal{N}(\mu, 1)$, v.a.i.i.d. Queremos estudiar dos estimadores de $q(\mu) = P_\mu(X \leq 2)$.

(a) Obtener el estimador de máxima verosimilitud de $q(\mu)$. Hallar su distribución asintótica. Sugerencia: puede ser útil el EMV de σ^2 hallado en el ejercicio 7d.

(b) Dar el estimador generalizado de momentos de $q(\mu)$ aplicando la estrategia del ejercicio 2(ii). Derivar su distribución asintótica.

(c) Comparando ambas distribuciones asintóticas, indicar cuál de los dos estimadores es preferible.

Sugerencia: puede ser útil el siguiente resultado: sean ϕ y Φ las funciones de densidad y distribución acumulada de la normal standard, respectivamente y $H : \mathbb{R} \to (0, +\infty)$ dada por $H(u) = \frac{[\phi(u)]^2}{\Phi(u)(1-\Phi(u))}$. Entonces, $H(u) = H(-u)$ y se puede probar (aunque no es fácil) que H es decreciente en $[0, +\infty)$ (ver, por ejemplo, Samford, 1953, "Some inequalities on Mill's ratio and related functions", *Annals of Mathematical Statistics*, vol. 10, pag. 643-655).

19. A partir de lo obtenido en el ejercicio 10.

(a) Probar que el EMV de θ es consistente y hallar su distribución asintótica.
Sugerencia: Usar el ejercicio 20 de la Práctica 7 de Probabilidaes y Estadística (M) [1] Ahí se prueba que para una muestra de tamaño n de variables aleatorias con densidad $\mathcal{U}(a, b)$, resulta que

$$n\left(b - X^{(n)}\right) \xrightarrow{\mathcal{D}} W \sim \mathcal{E}\left(\frac{1}{b-a}\right)$$

(b) Probar que el estimador de momentos para θ es consistente y hallar su distribución asintótica.

(c) Haremos una simulación para comparar la performance de ambos estimadores desde otro punto de vista. Queremos aproximar la siguiente probabilidad:

$$P_\theta\left(\left|\widehat{\theta}_n - \theta\right| < \varepsilon\right) \tag{3}$$

para los dos estimadores.

i. Para $n = 10$.

A. Generar una muestra aleatoria de tamaño n con distribución $\mathcal{U}(0, \theta_0 = 1)$. Calcular $\widehat{\theta}_{EMV}$ y $\widehat{\theta}_{mom}$ para dicha muestra.

B. Para $\varepsilon = 0.01$ verificar si las estimaciones calculadas verifican la condición $\left\{\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon\right\}$, con $\theta_0 = 1$. Guardar un `TRUE` ó `FALSE`.

C. Repetir (A.) y (B.) un número $K = 1000$ veces, registrando los resultados para cada estimador, en los vectores de longitud K `pmom` y `pemv`.

D. A partir de `pmom` y `pemv` estimar (3) para ambos estimadores. Guardar las probabilidades estimadas por este medio.

ii. Repetir (i.) para $n = 100, 500, 1\,000, 2\,000, 5\,000, 10\,000$ y $100\,000$. Comparar las probabilidades estimadas para los distintos n y los dos métodos de estimación. ¿Cuál prefiere desde este punto de vista? ¿Le parece que esto guarda alguna relación con la velocidad a la que cada estimador converge a la distribución límite? ¿Coincide la elección realizada con lo obtenido en el ejercicio 10 al comparar los ECM de ambos?

iii. Para ambos estimadores, aproxime estas probabilidades para los distintos valores de n haciendo uso de las distribuciones asintóticas. Para ello, aproxime la probabilidad (3) evaluando la función de distribución límite (exponencial o normal según el caso) en puntos que dependerán de n. Compare con lo calculado en los ítems C. y D. para los distintos valores de n. ¿Resulta buena la aproximación obtenida de este modo?

20. Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra aleatoria de una población con densidad $\beta(\theta, 1)$, o sea de la forma

$$f_\theta(x) = \theta\, x^{\theta-1}\, \mathbb{I}_{(0,1)}(x), \qquad \theta > 0.$$

[1] En la página `https://cms.dm.uba.ar/academico/materias/1ercuat2021/probabilidades_y_estadistica_M/` están colgadas las prácticas de Proba(M) de ese cuatrimestre en forma pública, seguimos la numeración de los ejercicios de ahí.

(a) Hallar el estimador de θ basado en el primer momento, que denominaremos $\hat{\theta}_{Mom}(\mathbf{X})$.

(b) Probar que $\hat{\theta}_{Mom}$ es un estimador fuertemente consistente.

(c) Probar que la distribución asintótica de $\hat{\theta}_{Mom}$ es

$$\sqrt{n}\left(\hat{\theta}_{Mom}(\mathbf{X}) - \theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\theta\,(\theta+1)^2}{\theta+2}\right).$$

2.4. Estimación de la densidad

21. Considerar el conjunto de datos del archivo `cpu.txt`, que contiene tiempos, en segundos, de CPUs correspondientes a 1000 trabajos enviados por una consultora. Hallar la ventana de Silverman para estimar la densidad de los tiempos, utilizando el núcleo gaussiano, y comparar con la que brinda por defecto la función `density`. Graficar ambas densidades estimadas superpuestas.

22. Considerar el último ejercicio de la Práctica 1 (archivo `Debernardi.csv`).

    (a) Hallar las ventanas de Silverman para estimar las densidades de LYVE1 según los niveles de la variable factor DIAGNOSIS, usando el núcleo gaussiano.

    (b) Graficar superpuestas las densidades estimadas obtenidas a partir de las ventanas calculadas en el ítem anterior.

    (c) Repetir los ítems anteriores usando el núcleo de Epanechnikov.