

TÉCNICAS DE REMUESTREO (Parte I)

Programa Oficial de Postgrado en
Estadística e Investigación Operativa
(UDC/USC/UV) 2009-2011

JOSÉ MANUEL PRADA SÁNCHEZ
Catedrático de Estadística e Investigación Operativa
Universidad de Santiago

CONTENIDO DEL CURSO (Parte I)

- Preliminares: Órdenes de convergencia. Limitaciones de la inferencia estadística clásica. Simulación y métodos de Monte Carlo. Estimación no paramétrica de la densidad. **P1, P2, P3, P4 y P5.**
- Motivación del principio Bootstrap. El Bootstrap uniforme. Cálculo de la distribución Bootstrap: distribución exacta y distribución aproximada por Monte Carlo. Ejemplos.
- Aplicación del Bootstrap a la estimación de la precisión y el sesgo de un estimador. Ejemplos. **P6.**
- Motivación del método Jackknife. Estimación Jackknife de la precisión y el sesgo de un estimador. Relación Bootstrap/Jackknife en dicha estimación. Ejemplos. Estudios de simulación. **P7.**
- Modificaciones del Bootstrap uniforme: Bootstrap paramétrico, simetrizado, suavizado, ponderado y sesgado. Discusión y ejemplos. Validez de la aproximación Bootstrap. Ejemplos. **P8.**
- Aplicación del Bootstrap a la construcción de intervalos de confianza: Métodos percentil, percentil-t, percentil-t simetrizado. Ejemplos. Estudios de simulación. **P9.**
- Otras aplicaciones del Bootstrap: El Bootstrap de un modelo de regresión. Wild Bootstrap. El Bootstrap en la estimación del error de predicción. El Bootstrap en poblaciones finitas. Ejemplos. Estudios de simulación.
- Iteración del principio Bootstrap. Motivación y principales resultados. Aplicaciones del Bootstrap iterado: Corrección del sesgo de un estimador. Corrección del error de cobertura de un intervalo de confianza. Estudios de simulación. **P10.**

BIBLIOGRAFÍA GENERAL CURSO (Parte I)

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.

Efron, B. (1979). Bootstrap Methods: Another look at the Jackknife. *Ann. Statist.*, 7, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap and other Resampling Plans*. CBMS-NSF. Regional Conference series in applied mathematics.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

González-Manteiga, W., Prada-Sánchez, J.M. and Romo, J. (1994). The Bootstrap-A Review. *Computational Statistics*, 9, 165-205.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Verlag.

Mammen, E. (1992). *When does Bootstrap Work?*. Springer Verlag.

Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling*. Springer Verlag.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer Verlag.

BIBLIOGRAFÍA PRELIMINARES (Parte I)

Rubinstein, R.Y. (1981). "Simulation and the Monte Carlo Method". Wiley.

Silverman, B.W. (1986). "Density Estimation". Chapman and Hall.

REFERENCIAS CURSO (Parte I)

Beran, R. (1987). "Prepivoting to reduce level error of confidence sets". *Biometrika* 74, 457-468.

Bickel, P.J. and Freedman, D.A. (1981). "Some Asymptotic theory for the bootstrap". *Ann. Statist.* 12, 2, 470-482.

Bowman, A., Hall, P. and Prvan, T. (1998). "Bandwidth selection for the smoothing of distribution functions". *Biometrika*, 85, 4, 799-808.

Cao, R. (1990). "Aplicaciones y nuevos resultados del método bootstrap en la estimación no paramétrica de curvas". *Tesis doctoral*. Santiago.

Cao, R. and Prada-Sánchez, J.M. (1993). "Bootstrapping the mean of a symmetric population". *Statistics & Probability Letters* 17, 43-48.

Efron, B. (1983). "Estimating the error rate of a prediction rule: improvements on cross-validation". *J. Amer. Stat. Assoc.* 78, 316-331.

Efron, B. and Tibshirani, R. (1986). "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy". *Statistical Science* 1, 54-77.

Efron, B. (1987). "Better Bootstrap confidence intervals (with discussion)". *J. Amer. Statist. Assoc.* 82, 171-200.

Efron, B. (1990). "More Efficient Bootstrap Computations". *J. Amer. Statist. Assoc.* 85, 79-89.

Freedman, D.A. (1981). "Bootstrapping regression models". *Ann. Statist.* 9, 6, 118-1228.

González-Manteiga, W. y Prada-Sánchez, J.M. (1985). "Una aplicación de los métodos de suavización no paramétricos en la técnica bootstrap". *Proceedings Jornadas Hispano-Lusas de Matemáticas*. Murcia, 1985.

Hartigan, J.A. (1969). "Using subsample values as typical values". *J. Amer. Statist. Assoc.* 64, 1303-1317.

Hall, P. (1986). "On the bootstrap and confidence intervals". *Ann. Statist.* 14, 1431-1452.

Hall, P. (1988-a) "Theoretical comparison of bootstrap confidence intervals". *Ann. Statist.* 16, 927-953.

- Hall, P. (1988-b). "Rate of convergence in bootstrap approximations". *Ann. Probab.* 16, 4, 1665-1684.
- Hall, P. and Martin, M.A. (1988). "On bootstrap resampling and iteration". *Biometrika* 75, 661-671.
- Härdle, W. And Mammen, E. (1993). "Comparing nonparametric versus parametric regression fits". *Ann. Statist.* 21, 1926-1947.
- Härdle, W. and Marron, J.S. (1991). "Bootstrap simultaneous bars for nonparametric regression". *Ann. Statist.* 19, 778-796.
- Maritz, J.S. (1979). "A note on exact robust confidence intervals for location". *Biometrika*, 66, 163-166.
- Navidi, W. (1989). "Edgeworth expansions for bootstrapping regression models". *Ann. Statist.* 17, 4, 1472-1478.
- Prada-Sánchez, J.M. and Otero-Cepeda, X.L. (1989). "The use of smooth bootstrap techniques for estimating the error rate of a prediction rule". *Comm. Statist. -Simula.*, 18(3), 1169-1186.
- Prada-Sánchez, J.M. and Cotos-Yáñez, T. (1997). "A Simulation Study of Iterated and Non-iterated Bootstrap Methods for Bias Reduction and Confidence Interval Estimation". *Comm. Statist. -Simula.*, 26(3), 927-946.
- Quenouille, M. (1949). "Approximate test of correlation in time series". *J. Roy. Statist. Soc. Ser. B*, 11, 18-84.
- Rubin, D.B. (1981). "The Bayesian Bootstrap". *Ann. Statist.*, 9, 1, 130-134.
- Schucany, W., Gray, H. and Owen, O. (1971). "On bias reduction in estimation". *J. Amer. Statist. Assoc.*, 66, 524-533.
- Shao, J. (1999). "Mathematical Statistics". Springer.
- Sing, K. (1981). "On the asymptotic accuracy of Efron's bootstrap". *Ann. Statist.* 9, 6, 1.187-1195.
- Tukey, J. (1958). "Bias and confidence in not quite large samples", abstract, *Ann. Math. Statist.*, 29, 614.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* 14, 1261-1350.

INTRODUCCIÓN AL CÁLCULO DE PROBABILIDADES

1. Experimento aleatorio. Espacio muestral asociado. Sucesos. Probabilidad.
2. Variable aleatoria. Función de distribución asociada.
3. Tipos de variables aleatorias. Modelos de probabilidad asociados: funciones de masa de probabilidad y densidad.
4. Transformaciones de variables aleatorias.
5. Características de una variable aleatoria.
6. Variables aleatorias notables: Bernoulli, Binomial, Uniforme y Normal.
7. Vectores aleatorios. Conceptos asociados. Distribución multinomial.
8. Sucesiones de variables aleatorias. Conceptos y resultados asociados.
9. Órdenes de convergencia. Contextos determinista y aleatorio.

5. CARACTERÍSTICAS DE UNA VARIABLE ALEATORIA: SIMETRÍA Y CURTOSIS.

Momentos respecto de la media de orden k

$$\beta_k = E[(X - EX)^k]$$

Casos particulares

K=1 → $\beta_1 = 0$

K=2 → $\beta_2 = \sigma^2$

K=3 → Coeficiente de SIMETRÍA

$$\gamma_1 = \frac{\beta_3}{\sigma^3} = \frac{E[(X - EX)^3]}{\sigma^3}$$

Si X simétrica respecto de μ , $\beta_k = 0$ si k impar

K=4 → Coeficiente de CURTOSIS

$$\gamma_2 = \frac{\beta_4}{\sigma^4} - 3 = \frac{E[(X - EX)^4]}{\sigma^4} - 3$$

$\gamma_2 = 0$ si forma similar a la normal

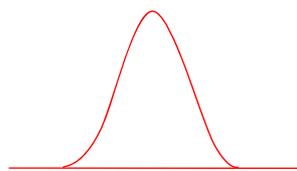
$\gamma_2 > 0$ si más aplastada que la normal

$\gamma_2 < 0$ si menos aplastada que la normal

tpc1

OBSERVACIÓN

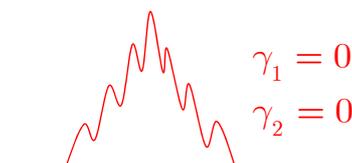
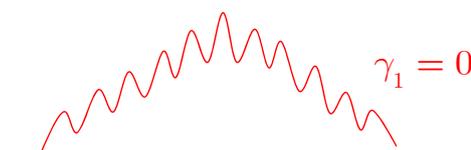
- Notemos que si X es N(0,1) su función de densidad tiene la forma



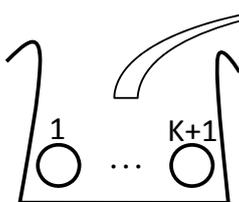
$\gamma_1 = 0$ (simetría)

$\gamma_2 = 0$ (su aplastamiento)

- Para cualquier otra variable aleatoria Y con $\gamma_1 = 0$, su función de densidad (si Y es a.c.) tendrá cierto parecido con la de X (simetría). Si además $\gamma_2 = 0$, entonces su parecido será mayor (además de la simetría tendrá el mismo grado de aplastamiento), salvando las posibles discrepancias que podrían sugerir las figuras que siguen:



7. VECTORES ALEATORIOS: DISTRIBUCIÓN MULTINOMIAL.



n extracciones con reemplazamiento
(n pruebas independientes)

bolas de k+1 tipos
(k+1 posibles resultados)

$$\forall i \in \{1, \dots, k+1\} \begin{cases} P(\textcircled{i}) = p_i \\ X_i = n^\circ \text{ de veces que se obtiene } \textcircled{i} \end{cases}$$

$\vec{X} = (X_1, \dots, X_k)^t \in M_k(n, (p_1, \dots, p_k))$

$$P(X_1 = x_1, \dots, X_k = x_k, X_{k+1} = x_{k+1}) = P(X_1 = x_1, \dots, X_k = x_k) =$$

$$= \frac{n!}{x_1! \cdots x_k! (n - \sum_{i=1}^k x_i)!} p_1^{x_1} \cdots p_k^{x_k} \left(1 - \sum_{i=1}^k p_i\right)^{n - \sum_{i=1}^k x_i}.$$

$$E(\vec{X}) = \begin{pmatrix} np_1 \\ \vdots \\ np_k \end{pmatrix}; \quad V(\vec{X}) = \begin{pmatrix} np_1(1-p_1) & \cdots & -np_1 p_k \\ \vdots & \ddots & \vdots \\ -np_k p_1 & \cdots & np_k(1-p_k) \end{pmatrix}$$

8. SUCESIONES DE VARIABLES ALEATORIAS. CONCEPTOS Y RESULTADOS ASOCIADOS.

- **Sucesión de variables aleatorias**

Cualquier colección numerables de variables aleatorias $\{X_n\}_{n=1}^{\infty}$.

- **Criterios de convergencia para sucesiones de variables aleatorias**

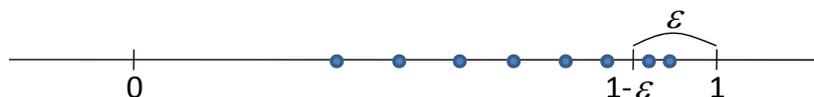
1. **Convergencia en distribución**

$$X_n \xrightarrow[n \rightarrow \infty]{D} X \Leftrightarrow F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x) \quad \forall x \text{ de continuidad de } F$$

(siendo F_n la función de distribución de X_n y F la función de distribución de X)

2. **Convergencia en probabilidad**

$$X_n \xrightarrow[n \rightarrow \infty]{P} X \Leftrightarrow P(|X_n - X| \leq c) \xrightarrow[n \rightarrow \infty]{} 1 \quad \forall c > 0$$



$$\forall c > 0, \forall \varepsilon > 0, \exists n_0 = n_0(c, \varepsilon) / \forall n \geq n_0 \quad P(|X_n - X| \leq c) > 1 - \varepsilon$$

3. Convergencia casi segura

$$X_n \xrightarrow[n \rightarrow \infty]{C.S.} X \Leftrightarrow P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$$

NOTA. La convergencia en probabilidad implica la convergencia de una única sucesión de números reales, mientras que la convergencia casi segura la de muchísimas (tantas como sucesos elementales quepan en ese suceso seguro).

$$C.S. \Rightarrow P. \Rightarrow D.$$

- **Teorema de KHINTCHINE**

Si $\{X_n\}_{n=1}^{\infty} / X_n$ iid X con $EX < \infty$, entonces

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{C.S.} EX$$

NOTA. Dicha convergencia se tendrá también en probabilidad.

- **Teorema de LEVY-LINDEBERG**

Si $\{X_n\}_{n=1}^{\infty} / X_n$ iid X con $EX = \mu < \infty$, $VX = \sigma^2 < \infty$, entonces

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow[n \rightarrow \infty]{D.} N(0, 1)$$

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow[n \rightarrow \infty]{D.} N(0, \sigma^2)$$

$$\bar{X} \xrightarrow[n \rightarrow \infty]{D.} N\left(\mu, \sigma^2/n\right)$$

- **Método DELTA**

Si $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D.} N(0, \sigma^2)$ y $g: \mathbb{R} \rightarrow \mathbb{R}$ diferenciable en μ , entonces

$$\frac{\sqrt{n} g(\bar{X}) - g(\mu)}{\left(g'(\mu)^2 \sigma^2\right)^{1/2}} \xrightarrow[n \rightarrow \infty]{D.} N(0, 1)$$

$$\sqrt{n} \left(g(\bar{X}) - g(\mu)\right) \xrightarrow[n \rightarrow \infty]{D.} N\left(0, g'(\mu)^2 \sigma^2\right)$$

$$g(\bar{X}) \xrightarrow[n \rightarrow \infty]{D.} N\left(g(\mu), g'(\mu)^2 \sigma^2/n\right)$$

OBSERVACIÓN 1

Si $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ es diferenciable en $\mu = (\mu_1, \dots, \mu_k)$, entonces

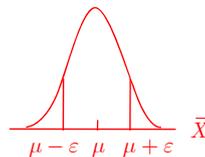
$$g'(\mu)^2 \sigma^2 \rightarrow Dg(\mu) \Sigma Dg(\mu)^t \text{ con } Dg(\mu) = \begin{pmatrix} \frac{\delta g_1}{\delta x_1} & \dots & \frac{\delta g_1}{\delta x_k} \\ \vdots & \ddots & \vdots \\ \frac{\delta g_m}{\delta x_1} & \dots & \frac{\delta g_m}{\delta x_k} \end{pmatrix}_{(x_1, \dots, x_k) = (\mu_1, \dots, \mu_k)}$$

Notemos que si $m=1$ entonces la matriz jacobiana $Dg(\mu)$ se convierte en el vector gradiente de g en μ .

OBSERVACIÓN 2

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} \mu$$
$$P(|\bar{X} - \mu| \leq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 1 \quad \forall \varepsilon > 0$$

¿Cuánto dista de 1 esa probabilidad para un n dado?



- Si $X \in N(\mu, \sigma^2)$ esa distancia se puede medir exactamente:
 $\bar{X} \in N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \in N(0,1) \Rightarrow P\left(\frac{\sqrt{n}}{\sigma} |\bar{X} - \mu| \leq \frac{\sqrt{n}}{\sigma} \varepsilon\right) = P\left(|Z| \leq \frac{\sqrt{n}}{\sigma} \varepsilon\right)$
- Para cualquier X , como $\bar{X} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$, sólo de manera aproximada.
- $\forall X / \gamma_1 = 0$ y $\gamma_2 = 0$, cabe esperar mejor aproximación.

9. ÓRDENES DE CONVERGENCIA

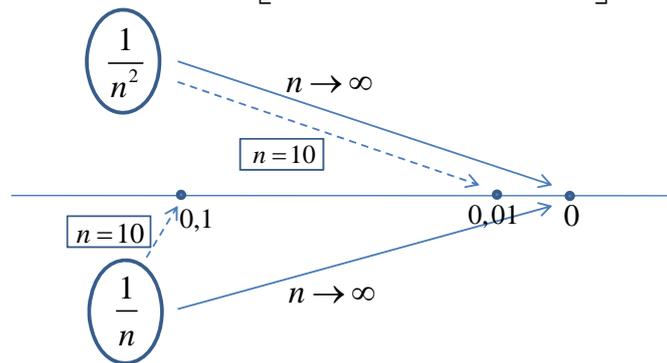
- Contexto DETERMINISTA

$$a_n = o(b_n) \Leftrightarrow \frac{a_n}{b_n} \xrightarrow{n \rightarrow \infty} 0$$

Es decir, si a_n y b_n convergen a 0, a_n lo hace más deprisa que b_n .

Ejemplo:

$$\frac{1}{n^2} = o\left(\frac{1}{n}\right) \left[\frac{\frac{1}{n^2}}{\frac{1}{n}} = \frac{n}{n^2} = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0 \right]$$



$$a_n = O(b_n) \Leftrightarrow \frac{|a_n|}{|b_n|} \leq c \text{ para algún } c, \text{ si } n \geq n_o(c)$$

Es decir, si a_n y b_n convergen a 0, a_n no lo hace más despacio que b_n .

Ejemplo: $\frac{2}{n} + \frac{1}{n^2} = O\left(\frac{1}{n}\right) \left[\frac{\left| \frac{2}{n} + \frac{1}{n^2} \right|}{\left| \frac{1}{n} \right|} = \frac{\frac{2}{n} + \frac{1}{n^2}}{\frac{1}{n}} = 2 + \frac{1}{n} \leq 3, \text{ si } n \rightarrow \infty \right]$

NOTA 1. Claramente: $o \Rightarrow O$ (basta tomar, por ejemplo, $c=1$).

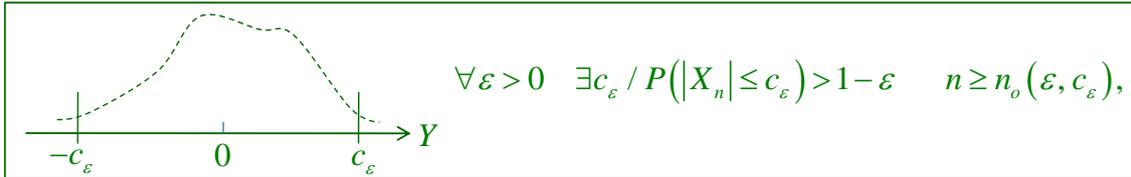
• Contexto ALEATORIO

$$X_n = o_p(b_n) \Leftrightarrow \frac{X_n}{b_n} \xrightarrow[n \rightarrow \infty]{P} 0$$

$$X_n = O_p(b_n) \Leftrightarrow \forall \varepsilon > 0 \exists c_\varepsilon / P\left(\frac{|X_n|}{|b_n|} \leq c_\varepsilon\right) > 1 - \varepsilon \quad \text{si } n \geq n_o(\varepsilon, c_\varepsilon)$$

NOTA 2. Claramente $o_p \Rightarrow O_p$.

NOTA 3. Como se verifica siempre que $X_n \xrightarrow{D} Y \Rightarrow X_n = O_p(1)$, pues



en particular

$$\frac{\bar{X} - \mu_X}{1/\sqrt{n}} \xrightarrow[TCL]{D} N(0, \sigma_X^2) \Rightarrow \frac{\bar{X} - \mu_X}{1/\sqrt{n}} = O_p(1) \Leftrightarrow \boxed{\bar{X} - \mu_X = O_p\left(\frac{1}{\sqrt{n}}\right)}$$

(Análogamente $g(\bar{X}) - g(\mu_X)$)

Tpc4

OBSERVACIÓN

$$\bar{X} - \mu_X = O_p\left(\frac{1}{\sqrt{n}}\right)$$

$\downarrow \quad n \rightarrow \infty \quad \downarrow$
 $0 \quad \quad \quad 0$

No va más despacio hacia 0 que $\frac{1}{\sqrt{n}} = n^{-\frac{1}{2}} \equiv$ cota inferior para la velocidad de convergencia de \bar{X} a μ_X .

$$\frac{1}{n^{\frac{1}{2}}}$$

Velocidad Teoría Clásica (*Indicador emblemático*)



Si $\frac{1}{n}$

¡MEJOR!



Veremos más adelante que, en condiciones muy generales, si tratamos de aproximar, por ejemplo

$$P\left(\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \leq y\right),$$

se tendrá que

$$P\left(\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \leq y\right) - \Phi(y) = O\left(n^{-\frac{1}{2}}\right) \quad \text{TEORÍA CLÁSICA}$$

$$P\left(\sqrt{n}\frac{\bar{X} - \mu}{\sigma} \leq y\right) - P\left(\sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq y\right) = O_p\left(n^{-1}\right). \quad \text{TEORÍA BOOTSTRAP}$$



“versión bootstrap” de
(La distribución de esta “versión bootstrap” se puede conocer en la práctica, de igual forma que se conoce $\Phi(y)$).

Análogamente,

$$X_n = o(b_n) \text{ c.s.} \Leftrightarrow \frac{X_n}{b_n} \xrightarrow[n \rightarrow \infty]{\text{c.s.}} 0$$

$$X_n = O(b_n) \text{ c.s.} \Leftrightarrow P\left(\frac{|X_n|}{|b_n|} \leq c\right) = 1 \quad \text{para algún } c, n \geq n_0(c)$$

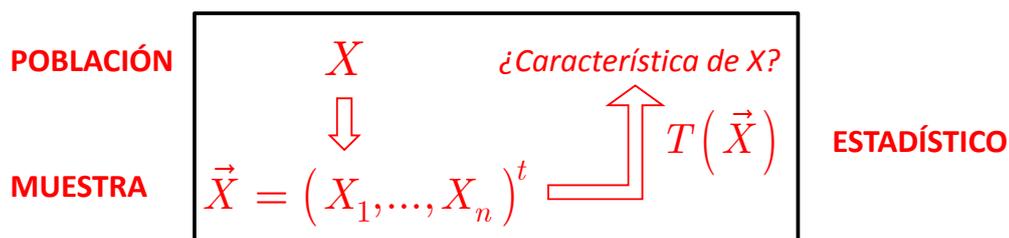
Resumiendo, si a_n y b_n son sucesiones de números reales y X_n de variables aleatorias

	o	O
Determinista	$\frac{a_n}{b_n} \rightarrow 0$	$\frac{ a_n }{ b_n } \leq c \quad n \geq n_0(c)$
P	$\frac{X_n}{b_n} \xrightarrow{P} 0$	$P\left(\frac{ X_n }{ b_n } \leq c_\epsilon\right) > 1 - \epsilon \quad n \geq n_0(\epsilon, c_\epsilon)$
C.S.	$\frac{X_n}{b_n} \xrightarrow{\text{c.s.}} 0$	$P\left(\frac{ X_n }{ b_n } \leq c\right) = 1 \quad n \geq n_0(c)$

INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

1. Conceptos básicos: población, muestra y estadístico. Inferencia paramétrica y no paramétrica.
2. Estimación puntual. Construcción de estimadores. Sesgo y precisión de un estimador: contextos paramétrico y no paramétrico.
3. Estimación mediante intervalos de confianza. Construcción de intervalos de confianza a partir de estadísticos pivotaes: contextos paramétrico y no paramétrico.
4. Estimación mediante contrastes de hipótesis. Construcción de contrastes: contextos paramétrico y no paramétrico.
5. Limitaciones de la inferencia estadística clásica.

1. CONCEPTOS BÁSICOS: POBLACIÓN, MUESTRA, ESTADÍSTICO. INFERENCIA PARAMÉTRICA Y NO PARAMÉTRICA



- **Población.** Cualquier variable aleatoria X con su modelo probabilístico asociado $y = F(x)$ del que deseamos conocer alguna característica.
- **Muestra aleatoria simple de X de tamaño n (m.a.s).** Un vector aleatorio n -dimensional formado por n copias independientes de X ,

$$\vec{X} = (X_1, \dots, X_n) / X_i \text{ iid } X,$$

con su modelo probabilístico asociado $y = F(\vec{x}) = \prod_{i=1}^n F(x_i)$.

- **Estadístico.** Cualquier función de la m.a.s. $T(\vec{X})$ con su modelo probabilístico asociado (función de F).

INFERENCIA

$$\begin{array}{c} X \\ \uparrow \\ \vec{X} = (X_1, \dots, X_n)^t \end{array}$$

- **INFERENCIA PARAMÉTRICA**

Cuando la población pertenece a un modelo paramétrico conocido salvo el valor del parámetro (objetivo de la inferencia).

$$X / F \in \{F_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$$

- **INFERENCIA NO PARAMÉTRICA**

Cuando se desconoce totalmente el modelo poblacional.

$$X / F \text{ desconocida}$$

En este caso, posibles objetivos de la inferencia serán $E(X)$, $F(x)$, x_p (p-cuantil de X), por ejemplo.

2. PRINCIPIOS METODOLÓGICOS DE LA INFERENCIA ESTADÍSTICA. ESTIMACIÓN PUNTUAL

Realiza apreciaciones puntuales (estimaciones) de las características desconocidas mediante estadísticos adecuados para ello: **ESTIMADORES**

A) ESTIMADORES NOTABLES. PROPIEDADES PRINCIPALES

- **Contexto PARAMÉTRICO**

En este contexto, los estimadores usuales son los de máxima verosimilitud.

$$f_\theta(\vec{x}) \begin{cases} \text{Densidad o masa muestral si } \theta \text{ es fijo y } \vec{x} \text{ variable} \\ \text{Función de verosimilitud si } \theta \text{ es variable y } \vec{x} \text{ fijo} \end{cases}$$

$$\vec{x} \rightarrow \hat{\theta}_{m.v.}(\vec{x}) = \arg \max_{\theta} f_\theta(\vec{x})$$

(“Debe ocurrir lo que es más probable”)

$$T(\vec{X}) = \hat{\theta}_{m.v.}(\vec{X}) \text{ es un estimador de máxima verosimilitud de } \theta$$

Conviene recordar que bajo condiciones de regularidad generales siempre existe un estimador de máxima verosimilitud tal que

$$T(\vec{X}) = \hat{\theta}_{m.v.}(\vec{X}) \xrightarrow{D} N\left(\theta, \frac{1}{I(\theta)}\right) = N\left(\theta, \frac{1}{ni(\theta)}\right),$$

siendo

$$I(\theta) = E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log f_{\theta}(\vec{X}) \right)^2 \right]; \quad i(\theta) = E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log f_{\theta}(X_i) \right)^2 \right]$$

la *Información de Fisher* sobre θ proporcionada por una muestra aleatoria simple de la población f_{θ} , o bien la correspondiente a su i -ésima observación, respectivamente. Notemos que, como

$$\sqrt{n}(\hat{\theta}_{m.v.}(\vec{X}) - \theta) \text{ converge en distribución} \Rightarrow \hat{\theta}_{m.v.}(\vec{X}) - \theta = O_p(n^{-1/2})$$

Ejemplo:

$$X \in Ber(p) \rightarrow f_p(x_1, \dots, x_n) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}, \quad x_i = 0, 1$$

$$\log f_p(x_1, \dots, x_n) = \left(\sum_{i=1}^n x_i \right) \log(p) + \left(n - \sum_{i=1}^n x_i \right) \log(1-p)$$

$$\frac{\partial}{\partial p} \log f_p(x_1, \dots, x_n) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = 0 \Leftrightarrow \hat{p}_{m.v.}(\vec{x}) = \frac{1}{n} \sum_{i=1}^n x_i,$$

que corresponde a un máximo de $\log f_p(\vec{x})$.

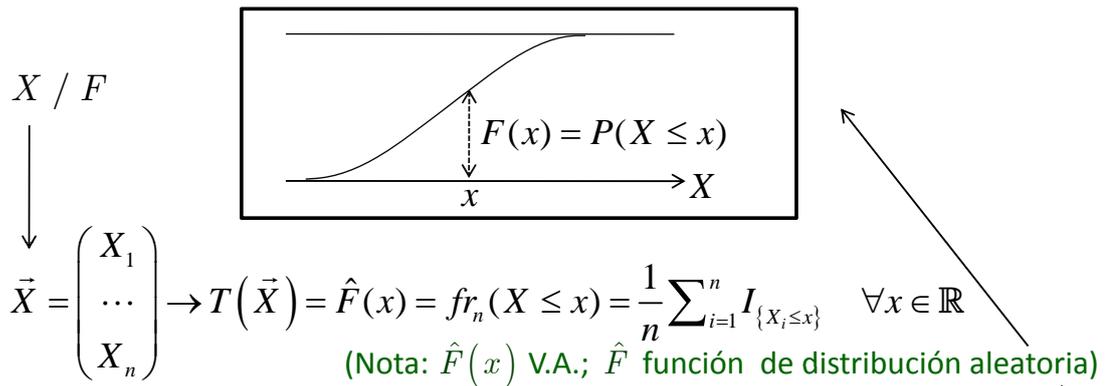
$$\hat{p}_{m.v.}(\vec{X}) = \bar{X}$$

- **Contexto NO PARAMÉTRICO**

En este contexto, los estimadores de mayor interés son la distribución empírica (fundamental en la metodología bootstrap) y los momentos y cuantiles muestrales.

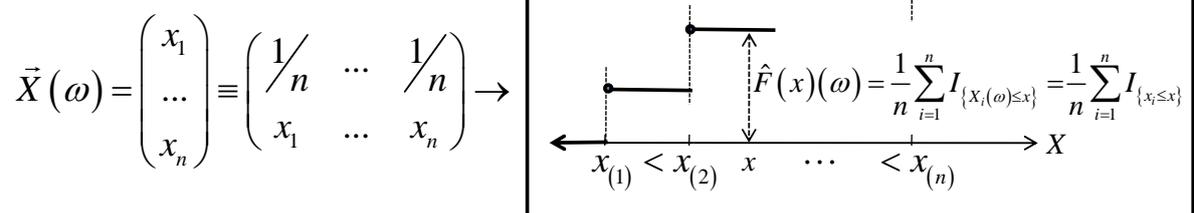
1. Distribución empírica

- **OBJETIVO** (Contexto poblacional): Función de distribución, F



- **ESTIMADOR** (Contexto muestral): Función de distribución empírica, \hat{F}

¿Cómo actúa este estimador?



Aproxima el modelo desconocido de la población por el modelo conocido de cada muestra .

Propiedades notables de la función de distribución empírica

- $\hat{F}(x) \xrightarrow[n \rightarrow \infty]{C.S.} F(x)$ (Ley fuerte grandes números)
 - $\hat{F}(x) \xrightarrow[n \rightarrow \infty]{P} F(x)$ (Ley débil grandes números)
 - $\hat{F}(x) \xrightarrow[n \rightarrow \infty]{D} N\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$ (Teorema central del límite)
- $\sqrt{n}(\hat{F}(x) - F(x))$ converge en distribución $\Rightarrow \hat{F}(x) - F(x) = O_p(n^{-1/2})$

- Converge uniformemente con probabilidad 1 a la función de distribución.

$$\sup_{x \in \mathbb{R}} |\hat{F}(x) - F(x)| \xrightarrow[n \rightarrow \infty]{C.S.} 0$$
 (Teorema de Glivenko - Cantelli)
- Es el estimador no paramétrico de máxima verosimilitud de la función de distribución.
- Es el ECUMV de la función de distribución por ser insesgado y función de la muestra ordenada, que es un estadístico suficiente (siempre) y completo (familia de distribuciones absolutamente continuas).

2. Momentos y cuantiles muestrales

- **OBJETIVO (Contexto poblacional):** Características de F

$$X / F \quad \left\{ \begin{array}{l} \alpha_k = E(X^k) \quad (k = 1 \Rightarrow \alpha_1 = \mu) \\ \beta_k = E((X - \mu)^k) \quad (k = 2 \Rightarrow \beta_2 = \sigma^2) \\ x_p = p - \text{cuantil de } X \end{array} \right.$$

- **ESTIMADOR (Contexto muestral):** Características de \hat{F}

$$\vec{X} = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \text{ m.a.s. de } X / \hat{F} \quad \left\{ \begin{array}{l} \hat{\alpha}_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (k = 1 \Rightarrow \hat{\alpha}_1 = \bar{X}) \\ \hat{\beta}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad (k = 2 \Rightarrow \hat{\beta}_2 = S^2) \\ \hat{x}_p \simeq x_{([np])} \end{array} \right.$$

Propiedades notables de los momentos y cuantiles muestrales

- Estos estimadores de los momentos y cuantiles poblacionales son todos fuertemente consistentes.

$$\hat{\alpha}_k, \hat{\beta}_k, \hat{x}_p \xrightarrow[n \rightarrow \infty]{C.S.} \alpha_k, \beta_k, x_p \quad , \text{ respectivamente.}$$

- También convergen en distribución a variables aleatorias normales que tienen a esas características poblacionales como media:

$$\begin{aligned} \sqrt{n}(\hat{\alpha}_k - \alpha_k) &\xrightarrow[n \rightarrow \infty]{D} N(0, \alpha_{2k} - \alpha_k^2) \\ \sqrt{n}(\hat{\beta}_k - \beta_k) &\xrightarrow[n \rightarrow \infty]{D} N\left(0, \beta_{2k} - \beta_k^2 - 2k\beta_{k-1}\beta_{k+1} + k^2\beta_{k-1}^2\beta_2\right) \\ \sqrt{n}(\hat{x}_p - x_p) &\xrightarrow[n \rightarrow \infty]{D} N\left[0, \left(\frac{\sqrt{p(1-p)}}{f(x_p)}\right)^2\right], \end{aligned}$$

por lo que, en particular, se tendrá para todo k y p :

$$\hat{\alpha}_k - \alpha_k = O_P\left(n^{-1/2}\right); \quad \hat{\beta}_k - \beta_k = O_P\left(n^{-1/2}\right); \quad \hat{x}_p - x_p = O_P\left(n^{-1/2}\right)$$

(Velocidad de convergencia de momentos y cuantiles muestrales hacia los poblacionales)

B) PROPIEDADES GENERALES DE LOS ESTIMADORES (CREDIBILIDAD)

En general (contextos paramétrico y no paramétrico), a cualquier estimador de θ (característica de F) lo denotaremos por $\hat{\theta} = \theta(X_1, \dots, X_n)$.

$$\begin{aligned} \text{Sesgo}_F \text{ de } \hat{\theta} &= E_F \hat{\theta} - \theta \\ \text{Varianza}_F \text{ de } \hat{\theta} &= V_F \hat{\theta} \\ \text{ECM}_F \text{ de } \hat{\theta} &= E_F \left((\hat{\theta} - \theta)^2 \right) = V_F \hat{\theta} + (E_F \hat{\theta} - \theta)^2 \end{aligned}$$

(El error cuadrático medio de un estimador, ECM, es su varianza más su sesgo al cuadrado)

	Var. pequeña	Var. grande
Sesgo pequeño		
Sesgo grande		

Ejemplos:

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow \begin{cases} \text{Sesgo}_F \text{ de } \bar{X} = E_F \bar{X} - \mu = 0 \\ \text{Varianza}_F \text{ de } \bar{X} = \frac{\sigma^2}{n} \end{cases}$
- $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \rightarrow \begin{cases} \text{Sesgo}_F \text{ de } S^2 = E_F S^2 - \sigma^2 = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \\ \text{Varianza}_F \text{ de } S^2 = \frac{\sigma^4}{n^2} 2(n-1) \end{cases}$
- $S_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \rightarrow \begin{cases} \text{Sesgo}_F \text{ de } S_c^2 = E_F S_c^2 - \sigma^2 = \sigma^2 - \sigma^2 = 0 \\ \text{Varianza}_F \text{ de } S_c^2 = \frac{\sigma^4}{(n-1)^2} 2(n-1) \end{cases}$

Observación (de interés práctico).

Sesgo y varianza son indicadores de la credibilidad de un estimador. Por esta razón, en la práctica, debemos acompañar cada estimación puntual que hagamos de sus correspondientes sesgo y varianza estimados (con los mismos datos usados en la estimación puntual). En general, esta segunda estimación puede ser muy complicada.

C) ESTIMACIÓN DEL SESGO Y LA VARIANZA DE UN ESTIMADOR

• Contexto PARAMÉTRICO

Supongamos, por ejemplo, que $X \in N(\mu, \sigma^2)$ y nuestro objetivo es μ . El estimador de máxima verosimilitud de μ es \bar{X} .

$$X \in N(\mu, \sigma^2)$$

↓

$$\vec{X} = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \rightarrow \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \in N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \begin{cases} E_F \bar{X} = \mu \Rightarrow \text{Sesgo}_F \bar{X} = 0 \\ V_F \bar{X} = \frac{\sigma^2}{n} \Rightarrow \widehat{V}_F \bar{X} = \frac{\hat{\sigma}^2}{n} \end{cases}$$

$(\hat{\sigma}^2 = S^2 \text{ ó } S_c^2)$

En cualquier otro contexto paramétrico se procedería análogamente (cuentas más o menos complicadas, pero posibles). Además, en condiciones generales:

$$X / F \in \{F_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$$

↓

$$\vec{X} = (X_1, \dots, X_n)^t \rightarrow \hat{\theta}_{m.v.}(\vec{X}) \xrightarrow[n \rightarrow \infty]{D} N\left(\theta, \frac{1}{I(\theta)}\right) \quad \left(\begin{array}{l} I(\theta) : \text{Información} \\ \text{de Fisher} \end{array} \right)$$

• Contexto NO PARAMÉTRICO

1. Supongamos que queremos estimar $\mu = E_F X$ con \bar{X} .

X / F cualquiera con media μ y varianza σ^2

↓

$$\vec{X} \rightarrow \bar{X} \xrightarrow[n \rightarrow \infty]{D} N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \begin{cases} E_F \bar{X} = \mu \Rightarrow \text{Sesgo}_F \bar{X} = 0 \\ V_F \bar{X} \simeq \frac{\sigma^2}{n} \Rightarrow \widehat{V}_F \bar{X} \simeq \frac{\hat{\sigma}^2}{n} \end{cases}$$

(Lévy-Lindeberg)

2. Supongamos que queremos estimar $g(\mu)$ (g diferenciable en μ) con $g(\bar{X})$

X / F cualquiera con media μ y varianza σ^2

↓

$$\vec{X} \rightarrow g(\bar{X}) \xrightarrow[n \rightarrow \infty]{D} N\left(g(\mu), \frac{g'(\mu)^2 \sigma^2}{n}\right) \Rightarrow$$

(Método delta)

$$\Rightarrow \begin{cases} E_F(g(\bar{X})) \simeq g(\mu) \Rightarrow \text{Sesgo}_F g(\bar{X}) \simeq 0 \\ V_F g(\bar{X}) \simeq \frac{g'(\mu)^2 \sigma^2}{n} \Rightarrow \widehat{V}_F g(\bar{X}) \simeq \frac{g'(\hat{\mu})^2 \hat{\sigma}^2}{n} \end{cases}$$

3. Para un objetivo general θ (cualquier característica de F), y cualquier estimador $\hat{\theta}$ de θ , puede ser muy complicada la estimación de su sesgo y varianza (en base a consideraciones asintóticas análogas a las anteriores).

¡EN TALES SITUACIONES SERÁ DE GRAN UTILIDAD EL BOOTSTRAP!

3. PRINCIPIOS METODOLÓGICOS DE LA INFERENCIA ESTADÍSTICA. ESTIMACIÓN MEDIANTE INTERVALOS DE CONFIANZA (I.C.)

Utiliza un intervalo de números reales para estimar una característica desconocida de la población, θ , con un indicador de riesgo o credibilidad (nivel de confianza) fijado de antemano.

- **Definición de INTERVALO DE CONFIANZA**

Si $T_i(X_1, \dots, X_n)^t$ $i = 1, 2$, son estadísticos tales que

$$P\left(T_1(X_1, \dots, X_n)^t \leq \theta \leq T_2(X_1, \dots, X_n)^t\right) = 1 - \alpha, \quad \alpha \in (0, 1),$$

y observamos los datos $\vec{x} = (x_1, \dots, x_n)^t$, $\vec{X} = \vec{x}$, entonces

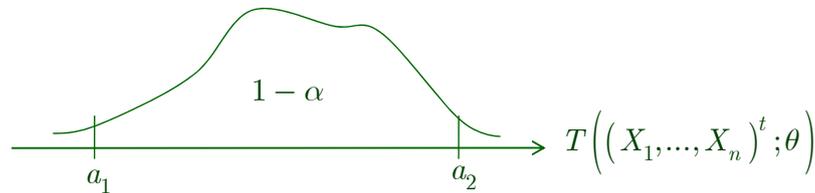
$$\boxed{(T_1(\vec{x}), T_2(\vec{x}))}$$

es un intervalo de confianza para θ de nivel de confianza $1 - \alpha$.

- Método PIVOTAL de construcción de intervalos de confianza

Sea $T\left(\left(X_1, \dots, X_n\right)^t; \theta\right)$ un estadístico “pivotal” (su distribución no depende de θ) y a_i , $i = 1, 2$, tales que

$$P\left(a_1 \leq T\left(\left(X_1, \dots, X_n\right)^t; \theta\right) \leq a_2\right) = 1 - \alpha, \quad \alpha \in (0, 1).$$



Despejando θ :

$$P\left(T_1\left(X_1, \dots, X_n\right)^t \leq \theta \leq T_2\left(X_1, \dots, X_n\right)^t\right) = 1 - \alpha,$$

y si $\vec{X} = \vec{x}$, entonces $\left(T_1(\vec{x}), T_2(\vec{x})\right)$ es un I.C. para θ de nivel $1 - \alpha$.

Por lo tanto, los estadísticos adecuados para construir I.C. son los **PIVOTALES**. Si tenemos dos estadísticos pivotaes para un mismo problema, será mejor el que proporcione I.C. de menor longitud a igualdad de nivel de confianza.

- Obtención de estadísticos pivotaes

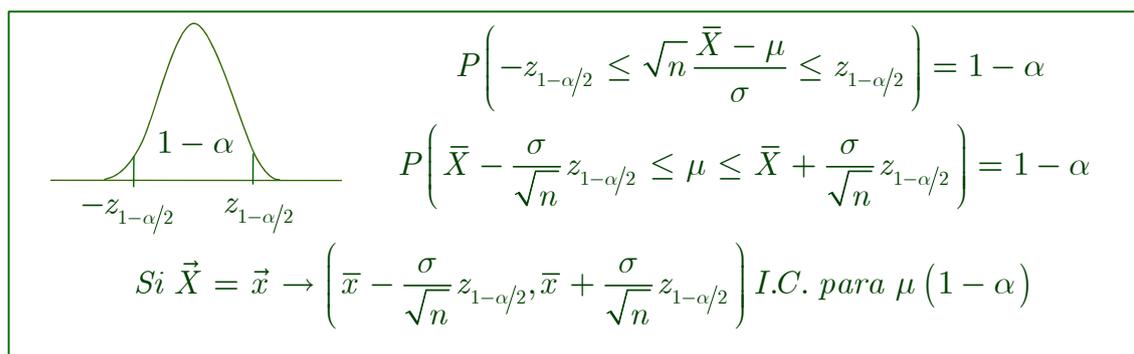
Contexto PARAMÉTRICO

Supongamos, por ejemplo, que $X \in N(\mu, \sigma^2)$ y nuestro objetivo es μ con σ^2 conocida.

$$X \rightarrow \vec{X} = \left(X_1, \dots, X_n\right)^t \rightarrow \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \in N\left(\mu, \frac{\sigma^2}{n}\right)$$

y por lo tanto un estadístico pivotal para μ es

$$T\left(\left(X_1, \dots, X_n\right)^t; \mu\right) = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \in N(0, 1)$$



En cualquier otro contexto paramétrico se procedería análogamente (cuentas más o menos complicadas pero posibles), pues recordemos que en condiciones generales la distribución del siguiente estadístico no depende de θ

$$T\left(\left(X_1, \dots, X_n; \theta\right)^t\right) = -\sum_{i=1}^n \log F_\theta(X_i) \in \Gamma(1, n)$$

Contexto NO PARAMÉTRICO

1. Supongamos que queremos estimar $\mu = E_F X$.

X / F cualquiera con media μ y varianza σ^2

↓

$$\vec{X} \rightarrow \bar{X} \xrightarrow[n \rightarrow \infty]{D} N\left(\mu, \frac{\sigma^2}{n}\right), \text{ (Lévy-Lindeberg)}$$

y por lo tanto un estadístico pivotal asintótico para μ es

$$T\left(\left(X_1, \dots, X_n\right)^t; \mu\right) = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \simeq N(0, 1)$$

(El correspondiente I.C. tendrá un **ERROR DE COBERTURA**)

2. Supongamos ahora que queremos estimar $g(\mu)$ (g diferenciable en μ).

X / F cualquiera con media μ y varianza σ^2

↓

$$\vec{X} \rightarrow g(\bar{X}) \xrightarrow[n \rightarrow \infty]{D} N\left(g(\mu), \frac{g'(\mu)^2 \sigma^2}{n}\right) \text{ (Método delta)}$$

y por lo tanto un estadístico pivotal asintótico para μ es

$$T\left(\left(X_1, \dots, X_n\right)^t; g(\mu)\right) = \sqrt{n} \frac{g(\bar{X}) - g(\mu)}{\left((g'(\mu))^2 \sigma^2\right)^{1/2}} \simeq N(0, 1)$$

(El correspondiente I.C. tendrá un **ERROR DE COBERTURA**)

3. Para un objetivo general θ (cualquier característica de F), puede ser muy complicada la obtención de un estadístico pivotal (en base a consideraciones asintóticas análogas a las anteriores).

¡EN TALES SITUACIONES SERÁ DE GRAN UTILIDAD EL BOOTSTRAP!

4. PRINCIPIOS METODOLÓGICOS DE LA INFERENCIA ESTADÍSTICA. ESTIMACIÓN MEDIANTE CONTRASTES DE HIPÓTESIS

Este principio aborda la estimación de una característica desconocida de la población rechazando o aceptando alguna suposición para la misma con un nivel de riesgo o credibilidad (nivel de significación) fijado de antemano.

CONCEPTOS BÁSICOS

Los introducimos en contexto paramétrico (análogamente se haría en general).

- Sea $\Theta = \Theta_0 \cup \Theta_1$ (Partición del espacio paramétrico)
 $H_0 \equiv \theta \in \Theta_0$, **hipótesis nula**: suposición que se quiere contrastar.
 $H_1 \equiv \theta \in \Theta_1$, **hipótesis alternativa**: suposición complementaria.
- Sea $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$ (Partición del espacio muestral)
 \mathcal{X}_0 , **región de aceptación de H_0**
 \mathcal{X}_1 , **región de rechazo de H_0**

- **TEST DE HIPÓTESIS.** Estadístico que rechaza o acepta H_0 .

$$\varphi(X_1, \dots, X_n)^t = \begin{cases} 1 & \text{si } (X_1, \dots, X_n)^t \in \mathcal{X}_1 \\ 0 & \text{si } (X_1, \dots, X_n)^t \in \mathcal{X}_0 \end{cases}$$

Asociados al test de hipótesis aparecen dos tipos de error:

$$\text{error I} = \frac{\text{rechazar } H_0}{H_0 \text{ cierta}}; \quad \text{error II} = \frac{\text{aceptar } H_0}{H_0 \text{ falsa}}$$

- **FUNCIÓN DE POTENCIA DE UN TEST.** Es la probabilidad de rechazar H_0 en función de θ .

$$\beta_\varphi : \theta \in \Theta \rightarrow \beta_\varphi(\theta) = E_\theta \varphi(X_1, \dots, X_n)^t = P_\theta \left((X_1, \dots, X_n)^t \in \mathcal{X}_1 \right)$$

β_φ permite calcular las probabilidades de cometer los errores I y II:

$$P(\text{error I}) = P \left(\frac{\text{rechazar } H_0}{H_0 \text{ cierta}} \right) = P_{\theta \in \Theta_0} \left((X_1, \dots, X_n)^t \in \mathcal{X}_1 \right) = \beta_\varphi(\theta),$$

$$P(\text{error II}) = P \left(\frac{\text{aceptar } H_0}{H_0 \text{ falsa}} \right) = 1 - P_{\theta \in \Theta_1} \left((X_1, \dots, X_n)^t \in \mathcal{X}_1 \right) = 1 - \beta_\varphi(\theta).$$

- **NIVEL DE SIGNIFICACIÓN DE UN TEST.** Cualquier cota α para la máxima probabilidad de cometer el error I (error de tipo I).

$$\alpha \geq \sup_{\theta \in \Theta_0} \beta_{\varphi}(\theta)$$

Dicho supremo se conoce como **tamaño** del test.

- **OPTIMALIDAD.** Si φ_1 y φ_2 son dos tests de nivel de significación α , φ_1 **será más potente que** φ_2 si

$$\beta_{\varphi_1}(\theta) \geq \beta_{\varphi_2}(\theta) \quad \forall \theta \in \Theta_1$$

Notemos que de esta forma la probabilidad de cometer el error II con φ_1 es menor que con φ_2 . Como ambos tienen el error I acotado de igual forma, claramente es mejor el test φ_1 .

NOTA. Debe tomarse como hipótesis nula la que más coste (social u otros) tenga rechazar. Así, al contrastar si un individuo sospechoso de un delito es CULPABLE o INOCENTE, la hipótesis nula debe ser INOCENTE, por tener más coste social declarar culpable a un inocente que inocente a un culpable.

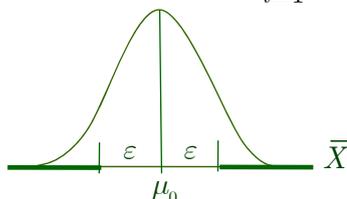
CONSTRUCCIÓN DE CONTRASTES

Contexto PARAMÉTRICO

Supongamos, por ejemplo, que $X \in N(\mu, \sigma^2)$, con σ^2 conocida y deseamos contrastar

$$H_0 \equiv \mu = \mu_0 \quad \text{frente} \quad H_1 \equiv \mu \neq \mu_0$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \in N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \in N(0, 1)$$

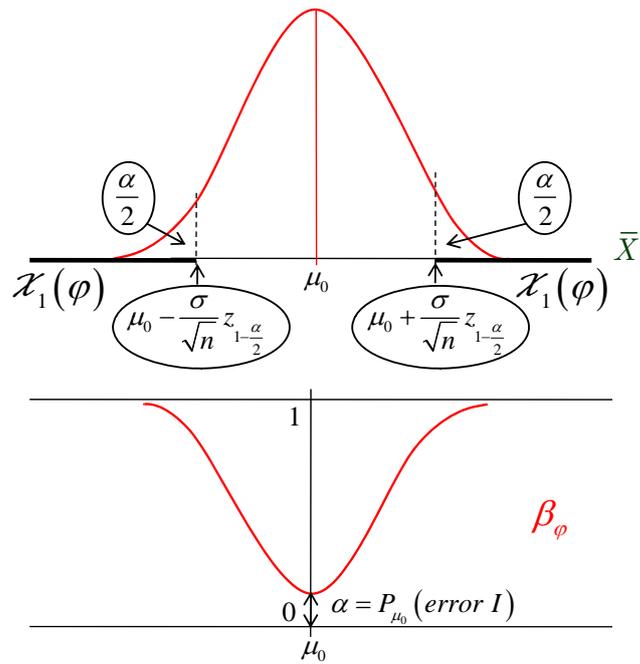


$$\varphi(X_1, \dots, X_n)^t = \begin{cases} 1 & \text{si } |\bar{X} - \mu_0| > \varepsilon \\ 0 & \text{si } |\bar{X} - \mu_0| \leq \varepsilon \end{cases}$$

Para que φ sea de nivel de significación α :

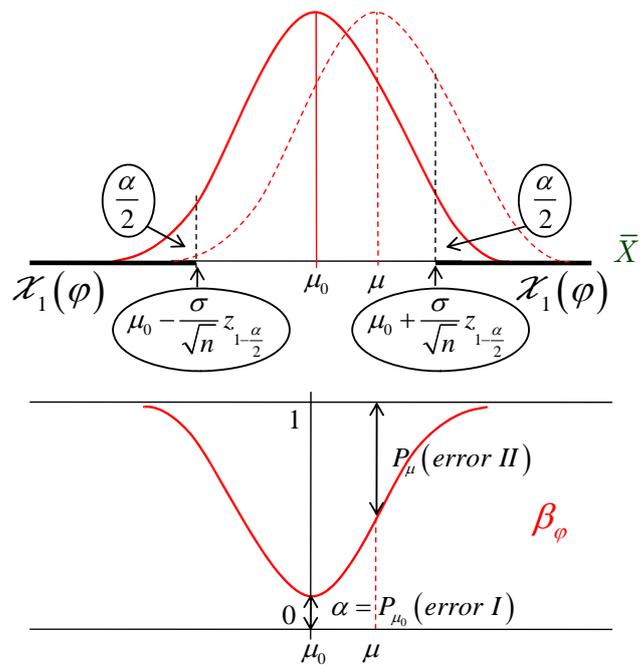
$$\alpha = \beta_{\varphi}(\mu_0) = P_{\mu_0}(|\bar{X} - \mu_0| > \varepsilon) = P_{\mu_0}\left(\frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| > \frac{\sqrt{n}}{\sigma} \varepsilon\right)$$

es decir, $\varepsilon = \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$ $\left(= \frac{S_c}{\sqrt{n}} t_{n-1, 1-\alpha/2} \text{ si } \sigma \text{ desconocida} \right).$



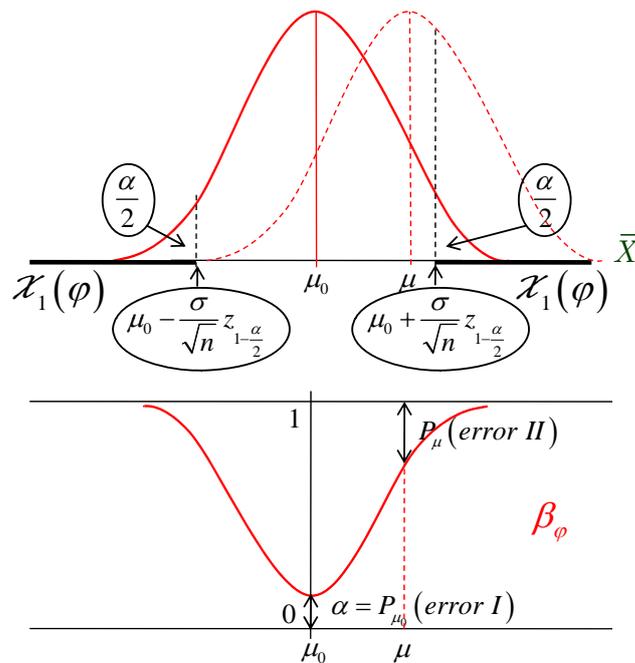
NOTA. Este test es el más potente entre tests de nivel α para este contraste.

En cualquier otro contexto paramétrico se podría proceder análogamente (cuentas más o menos complicadas pero posibles).



NOTA. Este test es el más potente entre tests de nivel α para este contraste.

En cualquier otro contexto paramétrico se podría proceder análogamente (cuentas más o menos complicadas pero posibles).



NOTA. Este test es el más potente entre tests de nivel α para este contraste.

En cualquier otro contexto paramétrico se podría proceder análogamente (cuentas más o menos complicadas pero posibles).

Notemos que con el test anterior, aceptar la hipótesis nula, $\left(\left| \bar{X} - \mu_0 \right| \leq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right)$, es equivalente a que

$$\mu_0 \in \left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right) \text{ I.C. para } \mu (1 - \alpha)$$

Contexto NO PARAMÉTRICO

1. Supongamos, por ejemplo, que nuestro objetivo es μ y sea el contraste

$$H_0 \equiv \mu = \mu_0 \quad \text{frente} \quad H_1 \equiv \mu \neq \mu_0$$

X / F cualquiera con media μ y varianza σ^2

↓

$$\vec{X} \rightarrow \bar{X} \xrightarrow[n \rightarrow \infty]{D} N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \simeq N(0, 1)$$

$$\varphi(X_1, \dots, X_n)^t = \begin{cases} 1 & \text{si } \left| \bar{X} - \mu_0 \right| > \varepsilon \simeq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \\ 0 & \text{en otro caso} \end{cases}$$

⊖ $\hat{\sigma}$

2. Supongamos ahora que nuestro objetivo es $g(\mu)$, con g diferenciable en μ y sea el contraste

$$H_0 \equiv g(\mu) = g(\mu_0) \quad \text{frente} \quad H_1 \equiv g(\mu) \neq g(\mu_0)$$

X / F cualquiera con media μ y varianza σ^2

↓

$$\bar{X} \rightarrow g(\bar{X}) \xrightarrow[n \rightarrow \infty]{D} N\left(g(\mu), \frac{g'(\mu)^2 \sigma^2}{n}\right) \Leftrightarrow \sqrt{n} \frac{g(\bar{X}) - g(\mu)}{g'(\mu) \sigma} \simeq N(0, 1)$$

$$\varphi(X_1, \dots, X_n)^t = \begin{cases} 1 & \text{si } |g(\bar{X}) - g(\mu_0)| > \varepsilon \simeq \frac{g'(\mu) \sigma}{\sqrt{n}} z_{1-\alpha/2} \\ 0 & \text{en otro caso} \end{cases} \quad (\hat{\mu}, \hat{\sigma})$$

NOTA. Análogamente se procedería con otros contrastes unilaterales

3. Para un objetivo general θ (cualquier característica de F), puede ser muy complicada la obtención de un estadístico pivotal.

¡EN TALES SITUACIONES SERÁ DE GRAN UTILIDAD EL BOOTSTRAP!

5. LIMITACIONES DE LA INFERENCIA ESTADÍSTICA CLÁSICA

Para estimar el **Sesgo** y la **Varianza** de un **ESTIMADOR**, hay que conocer características de su distribución (esperanza y varianza).

Para construir **INTERVALOS** de nivel de confianza $1 - \alpha$, y estimar su **error de cobertura**, o **CONTRASTES DE HIPÓTESIS** de nivel de significación α , y estimar su **función de potencia**, a partir de estadísticos pivotaes u otros, hay que conocer la distribución de éstos, o características de variables aleatorias asociadas a estos estadísticos.

- **Contexto PARAMÉTRICO**

La inferencia estadística clásica proporciona respuestas exactas (más o menos complicadas pero posibles)

- **Contexto NO PARAMÉTRICO**

La inferencia estadística clásica proporciona respuestas siempre aproximadas (Teorema Central del Límite). Además, sólo fáciles, para medias o funciones suaves de medias. En otro caso, pueden ser muy complicadas.

Un planteamiento general sugerido por EFRON, muy útil para abordar problemas de inferencia estadística como los anteriores (con indicadores de credibilidad), es considerar variables aleatorias de tipo $R(\vec{X}; F)$ y buscar su *distribución* o *características* de las mismas.

$$\begin{array}{ccc}
 X / F \text{ desconocida} & & \theta(F)? \\
 \downarrow & & \uparrow \\
 \vec{X} = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} & \text{m.a.s. de } X & \rightarrow R(\vec{X}; F)
 \end{array}$$

Ejemplos:

- $R(\vec{X}; F) = \sqrt{n}(\theta(\vec{X}) - \theta(F))$
- $R(\vec{X}; F) = \frac{\sqrt{n}(\theta(\vec{X}) - \theta(F))}{\sigma(\vec{X})}$

La distribución de estas variables permitirá construir intervalos de confianza para $\theta(F)$.

- $R(\vec{X}; F) = \theta(\vec{X}) - \theta(F) \Rightarrow E(R(\vec{X}; F)) = E(\theta(\vec{X})) - \theta(F) \equiv \text{sesgo de } \theta(\vec{X})$
- $R(\vec{X}; F) = \theta(\vec{X}) \Rightarrow V(R(\vec{X}; F)) = V(\theta(\vec{X})) \equiv \text{varianza de } \theta(\vec{X})$
- $R(\vec{X}; F) = I_{\{\theta_1(\vec{X}) \leq \theta(F) \leq \theta_2(\vec{X})\}} - (1 - \alpha) \Rightarrow E(R(\vec{X}; F)) = P(\theta_1(\vec{X}) \leq \theta(F) \leq \theta_2(\vec{X})) - (1 - \alpha) \equiv \text{error de cobertura de ese intervalo aleatorio sobre } \theta(F)$.

Notemos que para estas tres últimas variables basta con conocer una característica de su distribución (en este caso su esperanza) para dar respuesta a importantes problemas de inferencia estadística ligados a la credibilidad de un estimador o un intervalo de confianza.

APROXIMACIÓN CLÁSICA

- Basada en resultados tipo Teorema Central del Límite.
- Fácil sólo si el objetivo son medias o funciones suaves de medias.
- Puede ser muy complicada en otro caso.

APROXIMACIÓN BOOTSTRAP

- Permitirá de modo **fácil** y **siempre** aproximar la distribución de $R(\vec{X}; F)$.
- En algunos casos será mejor que la aproximación clásica.
- La iteración del principio bootstrap permitirá además, correcciones sucesivas, por ejemplo, del sesgo de un estimador y del error de recubrimiento de un intervalo de confianza.

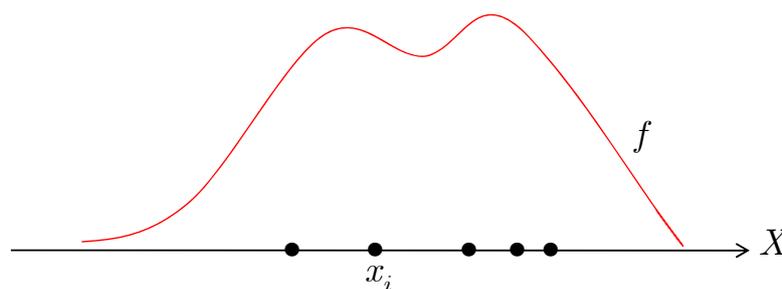
INTRODUCCIÓN A LA SIMULACIÓN Y MÉTODOS DE MONTE CARLO

1. Concepto de muestra artificial.
2. Generadores de dígitos pseudoaleatorios.
3. Métodos universales de simulación de variables continuas.
4. Métodos específicos de simulación de variables continuas.
5. Métodos universales de simulación de variables discretas.
6. Métodos universales de simulación de variables multidimensionales.

Aplicaciones:

- Aproximación por Monte Carlo de la distribución o características de un estadístico (sesgo y precisión: error cuadrático medio). **P1**
- Comparación de dos estimadores puntuales, aproximando por Monte Carlo sus errores cuadráticos medios. **P2**
- Comparación de dos intervalos de confianza aproximando por Monte Carlo sus errores de recubrimiento. **P3**
- Comparación de dos contrastes de hipótesis aproximando por Monte Carlo sus funciones de potencia. **P4**

1. CONCEPTO DE MUESTRA ARTIFICIAL

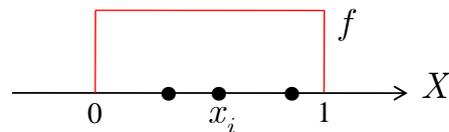


- **MUESTRA NATURAL DE X** : un conjunto de observaciones $\{x_1, \dots, x_n\}$ de dicha población obtenidas mediante labor de campo (muestra de “calle”).
- **MUESTRA ARTIFICIAL DE X** : un conjunto de observaciones $\{x_1, \dots, x_n\}$ de dicha población no obtenidas mediante labor de campo (muestra de “laboratorio”).

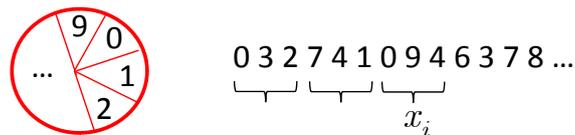
NOTA: La obtención de muestras artificiales, a diferencia de las naturales, requiere el conocimiento del modelo de la población (mecanismo generador de sus observaciones).

2. GENERADORES DE DÍGITOS PSEUDOALEATORIOS

Para obtener muestras artificiales de una población uniforme en $(0,1)$,



la idea es hacer girar un disco con sectores numerados desde 0 hasta 9 y observar su posición de parada respecto de una referencia fija. Agrupando los dígitos observados de 2 en 2, 3 en 3, 4 en 4, ..., se obtienen observaciones artificiales de dicha población con la precisión deseada.



Los principales métodos generadores de dígitos pseudoaleatorios son

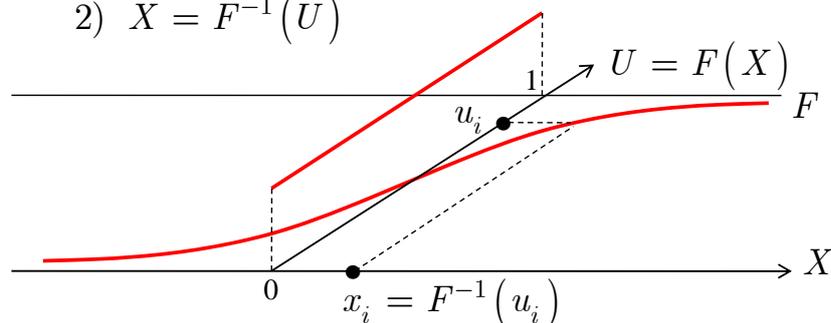
- Método de los cuadrados latinos.
- Método de Lehmer.
- Métodos congruenciales.

3. MÉTODOS UNIVERSALES DE SIMULACIÓN DE V. CONTINUAS

- **MÉTODO DE INVERSIÓN** (Para X con F invertible)

Algoritmo: 1) Generar $U \sim U(0,1)$

2) $X = F^{-1}(U)$



Fundamento:

- $U = F(X) \sim U(0,1)$ si X tiene distribución F

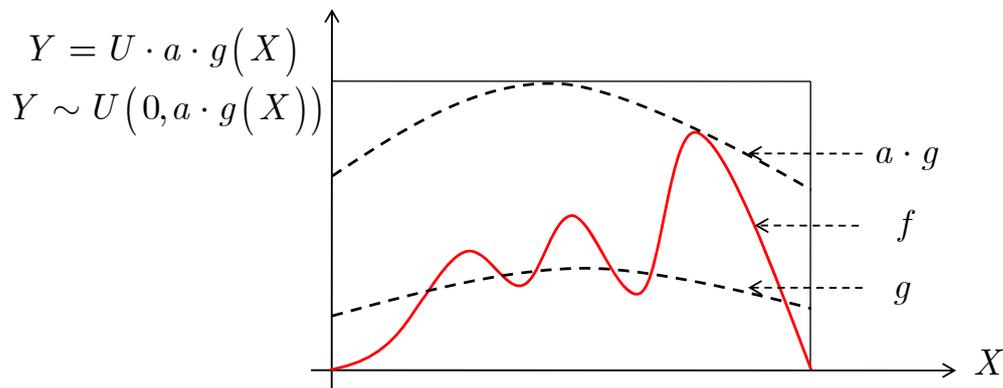
$$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$$
- Consecuencia : $F^{-1}(U)$ tiene distribución F

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

- MÉTODO DE ACEPTACIÓN/RECHAZO (Para X con densidad f)

Sea g densidad fácil de simular tal que $f(x) \leq a \cdot g(x) \quad \forall x \in \mathbb{R}$.

Algoritmo: 1) Generar $X \sim g$; $U \sim U(0,1)$
 2) Si $Y = U \cdot a \cdot g(X) \leq f(X)$ aceptar X
 En otro caso volver a 1).



Fundamento: Veamos que

$$P\left(X \leq x \middle/ X \text{ aceptada}\right) = P(X \leq x) = F(x)$$

$$P(X \text{ aceptada}) = \int_{-\infty}^{\infty} P\left(Y \leq f(X) \middle/ X(z)\right) g(z) dz = \int_{-\infty}^{\infty} \frac{f(z)}{a \cdot g(z)} g(z) dz = \frac{1}{a}$$

$$\begin{aligned}
 P(X \leq x; X \text{ aceptada}) &= \int_{-\infty}^{\infty} P\left(X \leq x; Y \leq f(X) \middle/ X = z\right) g(z) dz = \\
 &= \int_{-\infty}^x \frac{f(z)}{a \cdot g(z)} g(z) dz = \frac{1}{a} F(x)
 \end{aligned}$$

$$P\left(X \leq x \middle/ X \text{ aceptada}\right) = \frac{P(X \leq x; X \text{ aceptada})}{P(X \text{ aceptada})} = \frac{\frac{1}{a} \cdot F(x)}{\frac{1}{a}} = F(x).$$

NOTA: $P(X \text{ aceptada})$ máxima si $a = 1$ ($f \equiv g$).

4. MÉTODOS ESPECÍFICOS DE SIMULACIÓN DE V. CONTINUAS

Para la distribución normal

- **Box-Müller**

$$U_i \sim U(0,1), i = 1,2 \Rightarrow \sqrt{-2 \ln U_1} \cdot \text{sen}^2(2\pi U_2) \sim N(0,1)$$

- **Método basado en el Teorema Central del Límite**

$$U_i \sim U(0,1) \quad i = 1, \dots, 12$$

$$\sqrt{12} \frac{\bar{U} - \mu_U}{\sigma_U} = \frac{\sqrt{12} \left(\frac{1}{12} \sum_{i=1}^{12} U_i - \frac{1}{2} \right)}{1/\sqrt{12}} = \sum_{i=1}^{12} U_i - 6 \simeq N(0,1)$$

- **Polar**

NOTA: Análogamente existen métodos específicos para otras distribuciones.

5. MÉTODOS UNIVERSALES DE SIMULACIÓN DE V. DISCRETAS

- Método de la transformación cuantil.
- Algoritmos basados en árboles binarios. Árboles de Huffman.
- Método de la tabla guía.
- Método de truncamiento.

NOTA: Existen también métodos específicos para estas variables.

6. MÉTODOS UNIVERSALES DE SIMULACIÓN DE VARIABLES MULTIDIMENSIONALES

- Método de las distribuciones condicionadas.
- Métodos de codificación o etiquetado.
- Método de aceptación/rechazo.

SIMULACIÓN Y MÉTODOS DE MONTE CARLO: APLICACIONES

1. Aproximación por Monte Carlo de la distribución o características de un estadístico (sesgo y precisión: error cuadrático medio). P1

Si se conoce el mecanismo que genera los datos, se podrá simular el mecanismo que genera los valores de cualquier variable $R(\vec{X}; F)$.

X / F desconocida



$\vec{X} \rightarrow \text{¿} R = R(\vec{X}; F) \text{?}$
m.a.s. natural

X / F conocida
 ↓
 $\vec{x}^b \rightarrow r^b = R(\vec{x}^b; F)$
 $b = 1, \dots, B$
m.a.s. artificiales

- $\widehat{P}_F(R \leq x)^{MC} = \frac{n^\circ \text{ obs. } r^b \leq x}{B}$
- $\widehat{E}_F R^{MC} = \frac{1}{B} \sum_{b=1}^B r^b$
- $\widehat{V}_F R^{MC} = \frac{1}{B} \sum_{b=1}^B \left(r^b - \frac{1}{B} \sum_{b=1}^B r^b \right)^2$
- $\widehat{c}_p^{MC} = r_{([Bp])}$

NOTA. Las V.A. que aportan esas 4 estimaciones, obtenidas reemplazando r^b por $R^b = R(\vec{X}^b; F)$, $b = 1, \dots, B$, son estadísticos sobre la m.a.s. (R^1, \dots, R^B) que convergen C. S., consistencia fuerte, a los correspondientes 4 objetivos poblacionales:

- $\frac{n^\circ \text{ obs. } R^b \leq x}{B} = \frac{1}{B} \sum_{b=1}^B I_{\{R^b \leq x\}} \xrightarrow[B \rightarrow \infty]{C.S.} E \left\{ I_{\{R \leq x\}} \right\} = P_F(R \leq x)$
- $\frac{1}{B} \sum_{b=1}^B R^b \xrightarrow[B \rightarrow \infty]{C.S.} E_F R$
- $\frac{1}{B} \sum_{b=1}^B \left(R^b - \frac{1}{B} \sum_{b=1}^B R^b \right)^2 \xrightarrow[B \rightarrow \infty]{C.S.} E_F \left\{ (R - E_F R)^2 \right\} = V_F R$
- $R_{([Bp])} \xrightarrow[B \rightarrow \infty]{C.S.} c_p$

NOTA. Si $R = R(\bar{X}; F) = \hat{\theta}_n - \theta \Rightarrow \begin{cases} \widehat{E}_F R^{MC} = \widehat{E}_F \hat{\theta} - \theta = \widehat{\text{sesgo de } \hat{\theta}^{MC}} \\ \widehat{V}_F R^{MC} = \widehat{V}_F \hat{\theta}^{MC} \end{cases}$

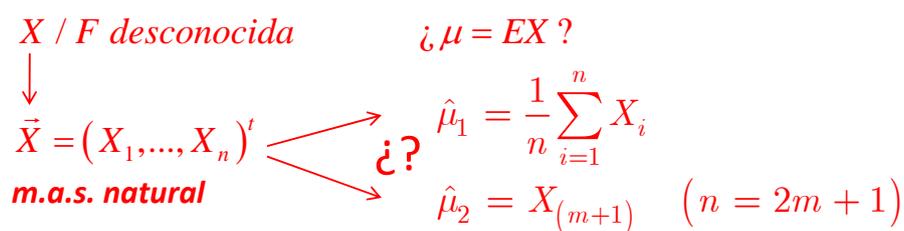
Tabla. Resultado de un estudio de simulación que estima por Monte Carlo (B=10.000) el ECM de \bar{X} como estimador de μ , en diferentes escenarios controlados, y lo compara con el ECM exacto (entre paréntesis).

	Uniforme (0,1)	Normal (0,1)	Exponencial (1)
ECM \bar{X}	MC (exacto: $1/12n$)	MC (exacto: $1/n$)	MC (exacto: $1/n$)
n = 10	0,0084 (0,0083)	0,0973 (0,1000)	0,1077 (0,1000)
n = 20	0,0040 (0,0042)	0,0491 (0,0500)	0,0488 (0,0500)
n = 50	0,0016 (0,0017)	0,0199 (0,0200)	0,0199 (0,0200)

	Uniforme (-1,1)	Normal (5,(0,5 ²))	Exponencial (0,5)
ECM \bar{X}	MC (exacto: $1/3n$)	MC (exacto: $1/4n$)	MC (exacto: $4/n$)
n = 20	0,0169 (0,0167)	0,0124 (0,0125)	0,2012 (0,2000)

SIMULACIÓN Y MÉTODOS DE MONTE CARLO: APLICACIONES

2. Comparación por Monte Carlo de dos estimadores. P2



X / F conocida (escenario controlado) μ conocida

↓

$\vec{x}^b = (x_1, \dots, x_n)^{t^b}, b = 1, \dots, B$ ↗ $\hat{\mu}_1^b$

B m.a.s. artificiales ↘ $\hat{\mu}_2^b$

$$e_i = \widehat{ECM}^{MC} \text{ de } \hat{\mu}_i = \left(\frac{1}{B} \sum_{b=1}^B \hat{\mu}_i^b - \mu \right)^2 + \frac{1}{B} \sum_{b=1}^B \left(\hat{\mu}_i^b - \frac{1}{B} \sum_{b=1}^B \hat{\mu}_i^b \right)^2, \quad i = 1, 2$$

(Notemos que estas estimaciones del ECM son fuertemente consistentes.)

- Elegiríamos el estimador con mínimo ECM estimado, probando con diferentes escenarios controlados.

Tabla. Resultado de un estudio de simulación que compara dos posibles estimadores de μ : media muestral ($\hat{\mu}_1$) y mediana muestral ($\hat{\mu}_2$), estimando por Monte Carlo sus ECM en diferentes escenarios controlados.

ECM	Normal (2,1)		Exponencial (1)		Chicuarado (5)	
	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$
n = 10	0,0998	0,1324	0,1063	0,1607	1,0112	1,5273
n = 50	0,0193	0,0295	0,0190	0,1109	0,2135	0,6823
n = 100	0,0107	0,0161	0,0099	0,1005	0,1006	0,5245

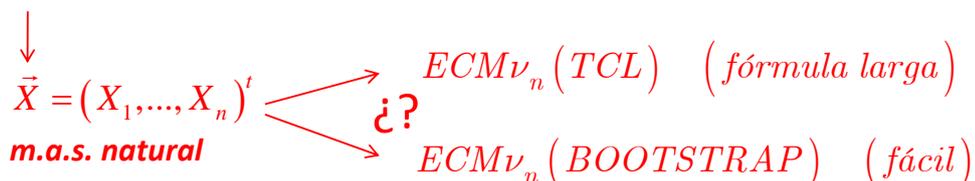
Nota.

- Los contextos exponencial y chicuarado no son adecuados para la mediana muestral (como estimador de la media) por su falta de simetría.
- El bootstrap permitirá estimar el ECM en contexto no paramétrico.

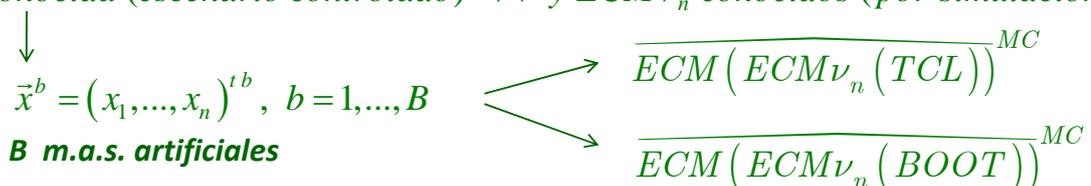
OBSERVACIÓN IMPORTANTE

El método de Monte Carlo no sólo permitirá comparar dos estimadores, sino también dos FILOSOFÍAS para estimar, por ejemplo, el error cuadrático medio de un estadístico no fácil como puede ser una mediana muestral, $ECM\nu_n$.

X / F desconocida

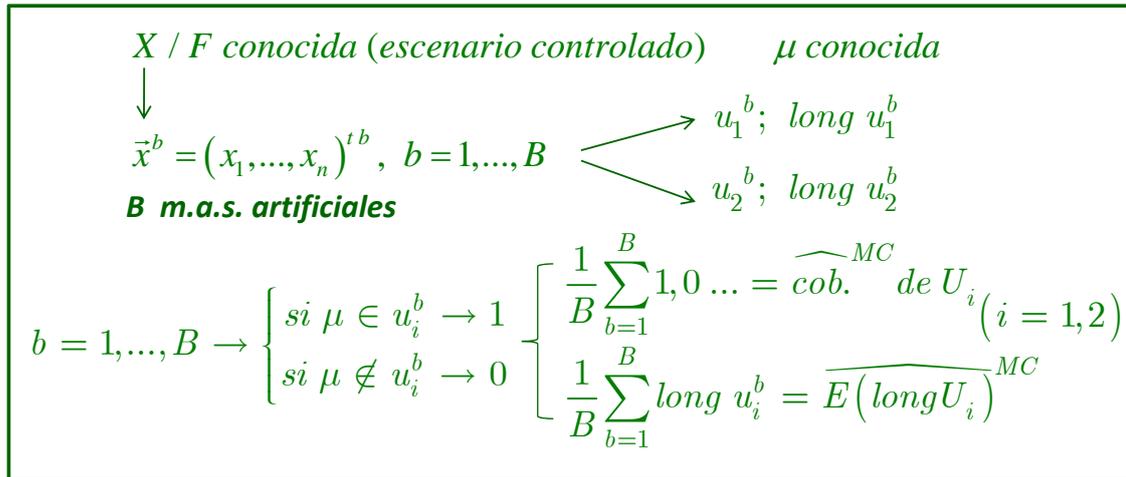
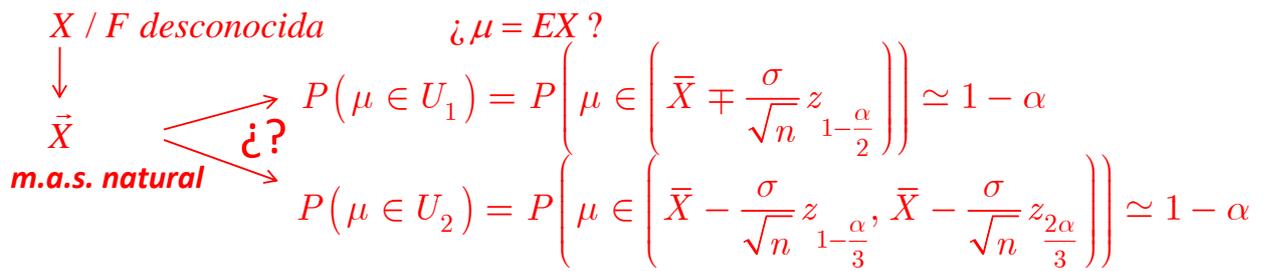


X / F conocida (escenario controlado) $\rightarrow \nu$ y $ECM\nu_n$ conocidos (por simulación)



SIMULACIÓN Y MÉTODOS DE MONTE CARLO: APLICACIONES

3. Comparación por Monte Carlo de dos intervalos de confianza. P3



- Notemos que

$$\frac{1}{B} \sum_{b=1}^B I_{(\mu \in U_i)}^b \xrightarrow[B \rightarrow \infty]{C.S.} E(I_{\mu \in U_i}) = P(\mu \in U_i), i = 1, 2$$

$$\frac{1}{B} \sum_{b=1}^B \text{long } U_i^b \xrightarrow[B \rightarrow \infty]{C.S.} E(\text{long } U_i), i = 1, 2$$

por lo que las estimaciones anteriores son fuertemente consistentes.

- Elegiríamos el intervalo con menor **error de recubrimiento** e_i , probando con diferentes escenarios controlados.

$$e_i = \left| \text{cob. REAL} - \text{cob. NOMINAL} \right| = \left| \widehat{\text{cob.}}^{MC} \text{ de } U_i - (1 - \alpha) \right|$$

A igualdad de cobertura elegiríamos el intervalo de menor longitud.

Tabla. Resultado de un estudio de simulación que compara errores de recubrimiento y longitudes (debajo, entre paréntesis) de dos intervalos de confianza de nivel aproximado 0,95 (obtenidos a partir de U_1 y U_2), estimando por Monte Carlo dichos errores y longitudes con $B=10.000$ en diferentes escenarios controlados.

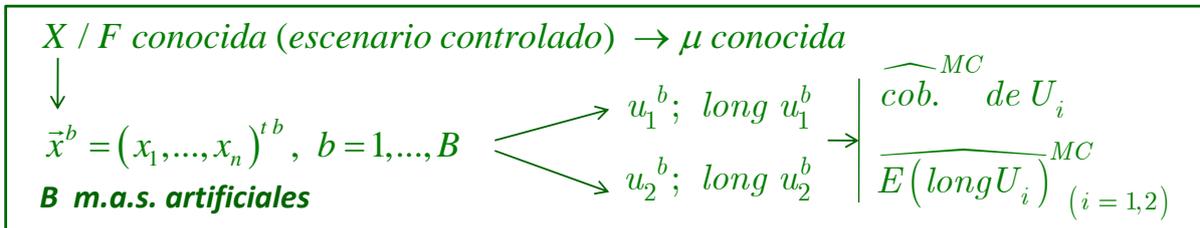
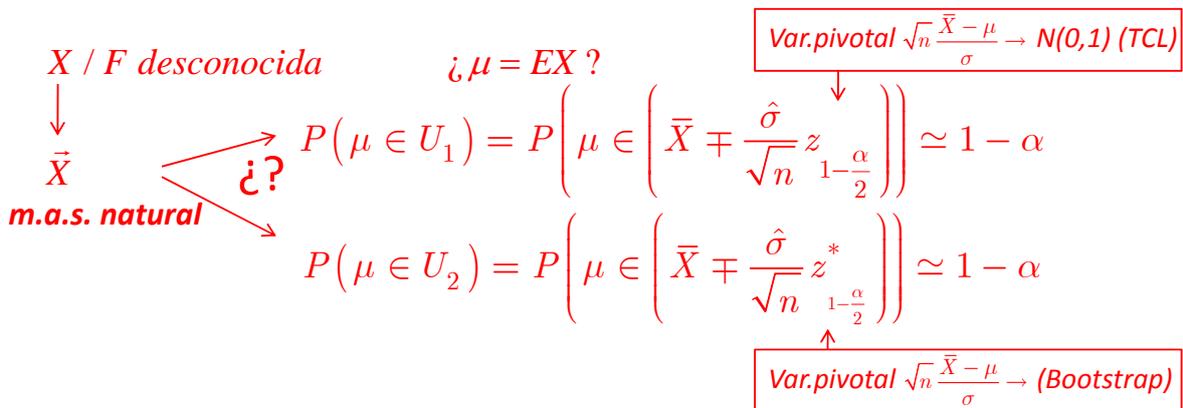
	Normal (5,1)		Exponencial (1/5)		Chic cuadrado (5)	
	U_1	U_2	U_1	U_2	U_1	U_2
n = 10	0,0014 (2,48)	0,0007 (2,51)	0,0048 (6,20)	0,0070 (6,26)	0,0030 (3,92)	0,0047 (3,96)
n = 50	0,0047 (1,11)	0,0020 (1,12)	0,0012 (2,77)	0,0041 (2,80)	0,0021 (1,75)	0,0033 (1,77)
n = 100	0,0017 (0,78)	0,0012 (0,79)	0,0009 (1,96)	0,0022 (1,98)	0,0030 (1,24)	0,0037 (1,25)

Nota.

- En el contexto normal estos intervalos son exactos (nivel 0,95). En los restantes contextos son aproximados (T.C.L.). Se observa que los intervalos obtenidos a partir de U_1 (de longitud mínima en contexto normal) tienen siempre menor longitud que los obtenidos a partir de U_2 .
- Una aproximación alternativa será el bootstrap.

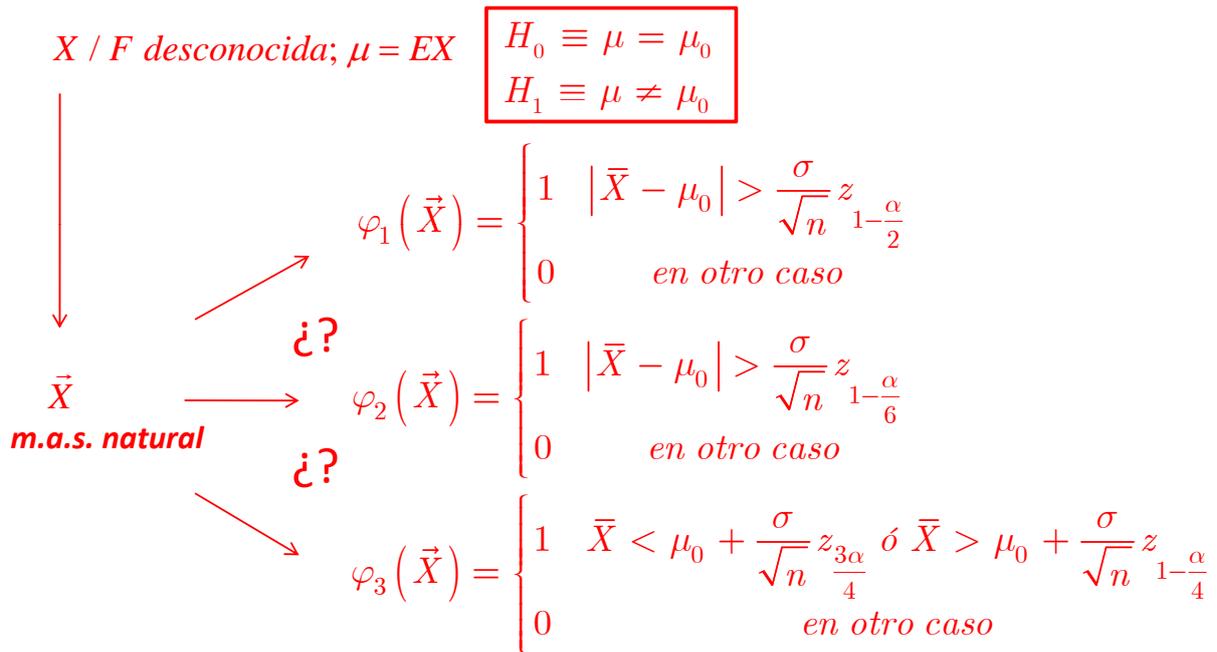
OBSERVACIÓN IMPORTANTE

El método de Monte Carlo no sólo permitirá comparar dos intervalos de confianza, sino también dos FILOSOFÍAS para construir intervalos de confianza, por ejemplo, para una media poblacional, μ :



SIMULACIÓN Y MÉTODOS DE MONTE CARLO: APLICACIONES

4. Comparación por Monte Carlo de dos funciones de potencia. P4



NOTA. En contexto normal, los tres son de nivel α ; además, φ_1 y φ_3 son de tamaño α ; φ_1 y φ_2 son insesgados y φ_3 es sesgado. En contexto no normal todo es aproximado (T.C.L.). Una aproximación alternativa será el bootstrap.

Fijamos μ (muchos valores de μ equiespaciados)

X / F conocida con ese valor de μ (escenario controlado)

↓

$$\vec{x}^b = (x_1, \dots, x_n)^{t^b} \rightarrow \bar{x}^b, \quad b = 1, \dots, B$$

B m.a.s. artificiales

Identificando φ_i con su región de rechazo,

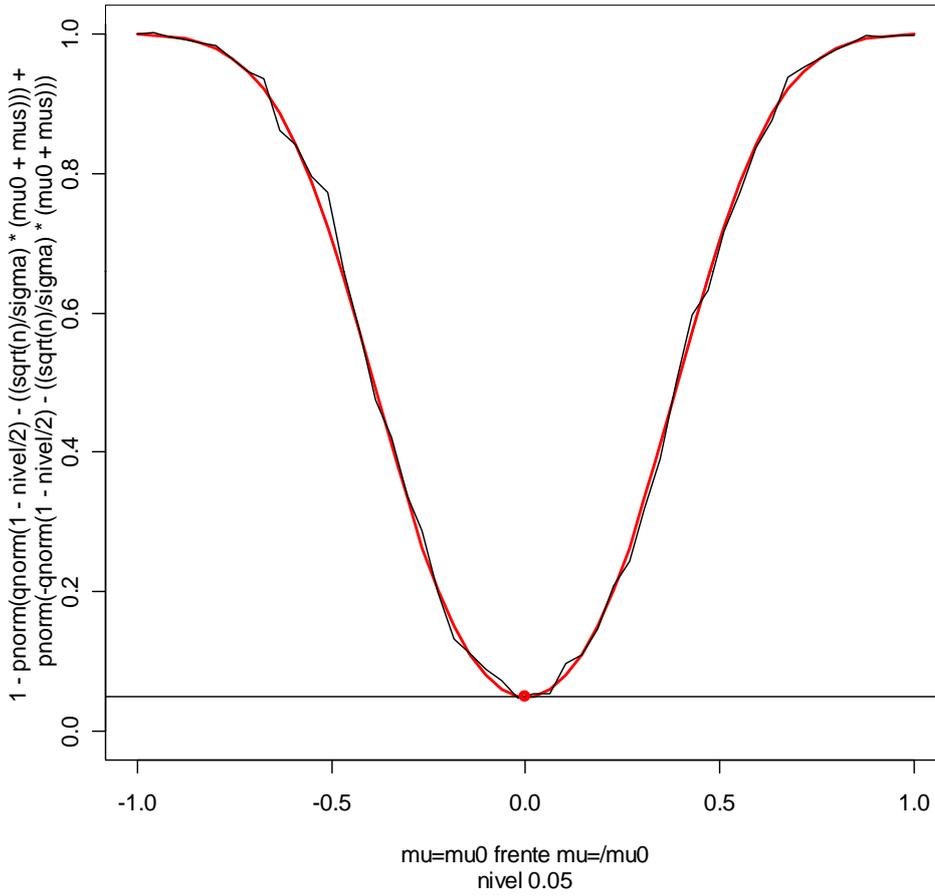
$$b = 1, \dots, B \rightarrow \begin{cases} \text{si } \bar{x}^b \in \varphi_i \rightarrow 1 \\ \text{si } \bar{x}^b \notin \varphi_i \rightarrow 0 \end{cases} \rightarrow \frac{1}{B} \sum_{b=1}^B 1, 0 \dots = \widehat{\beta_{\varphi_i}(\mu)}^{MC}, \quad i = 1, 2, 3$$

Notemos que esa estimación es fuertemente consistente:

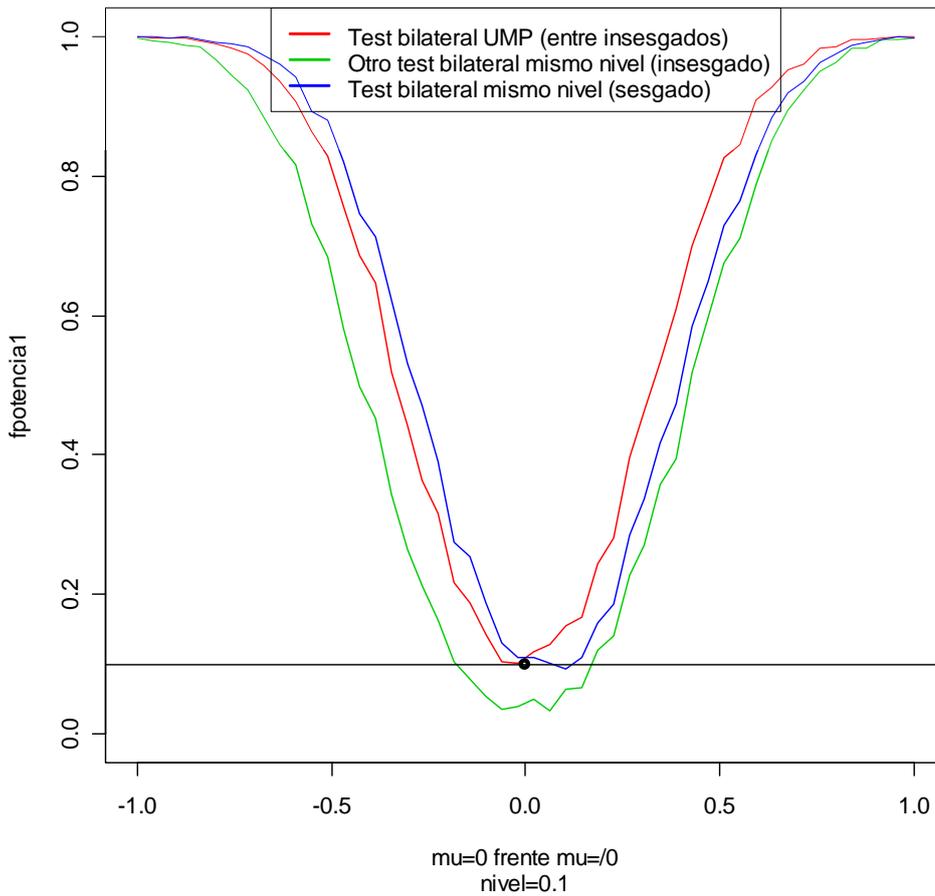
$$\frac{1}{B} \sum_{b=1}^B I_{(\bar{x}^b \in \varphi_i)} \xrightarrow[B \rightarrow \infty]{C.S.} E_{\mu} \left(I_{(\bar{X} \in \varphi_i)} \right) = P_{\mu} \left(\bar{X} \in \varphi_i \right) = \beta_{\varphi_i}(\mu).$$

- Elegiríamos el test cuya función de potencia estimada fuese máxima bajo H_1 (si son comparables), probando con diferentes escenarios.

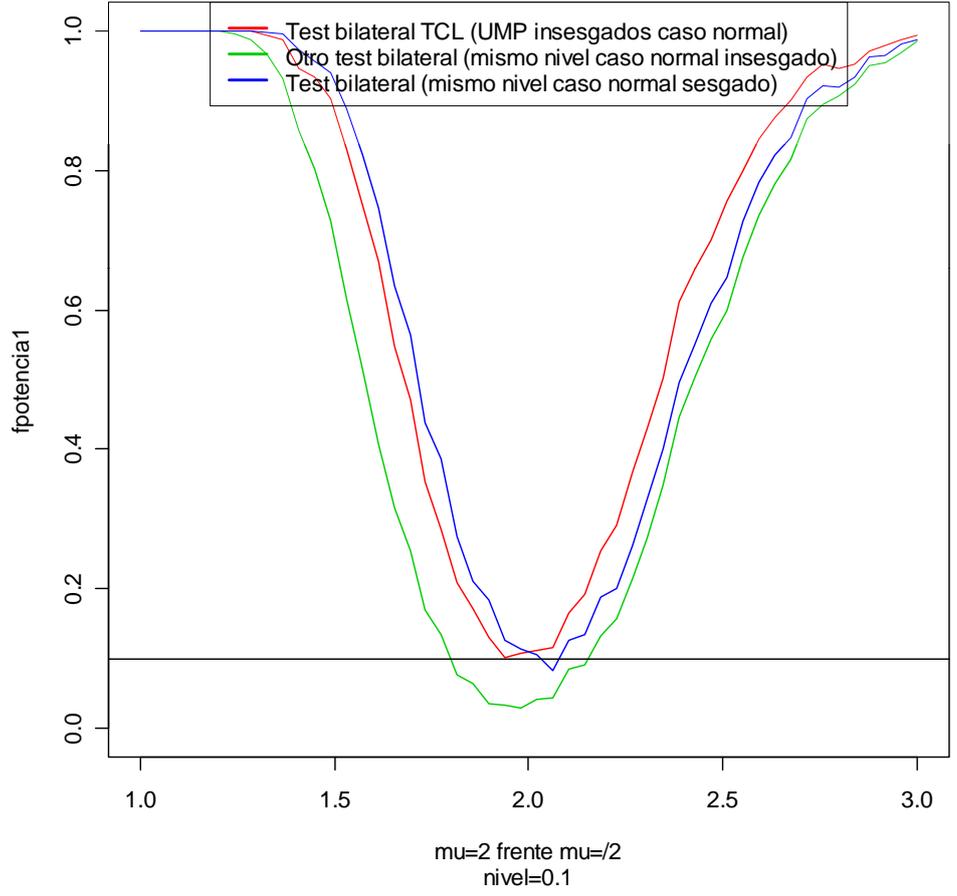
**Función de potencia, test bilateral media (UMP), X normal
(rojo: exacta; negro: MC)**



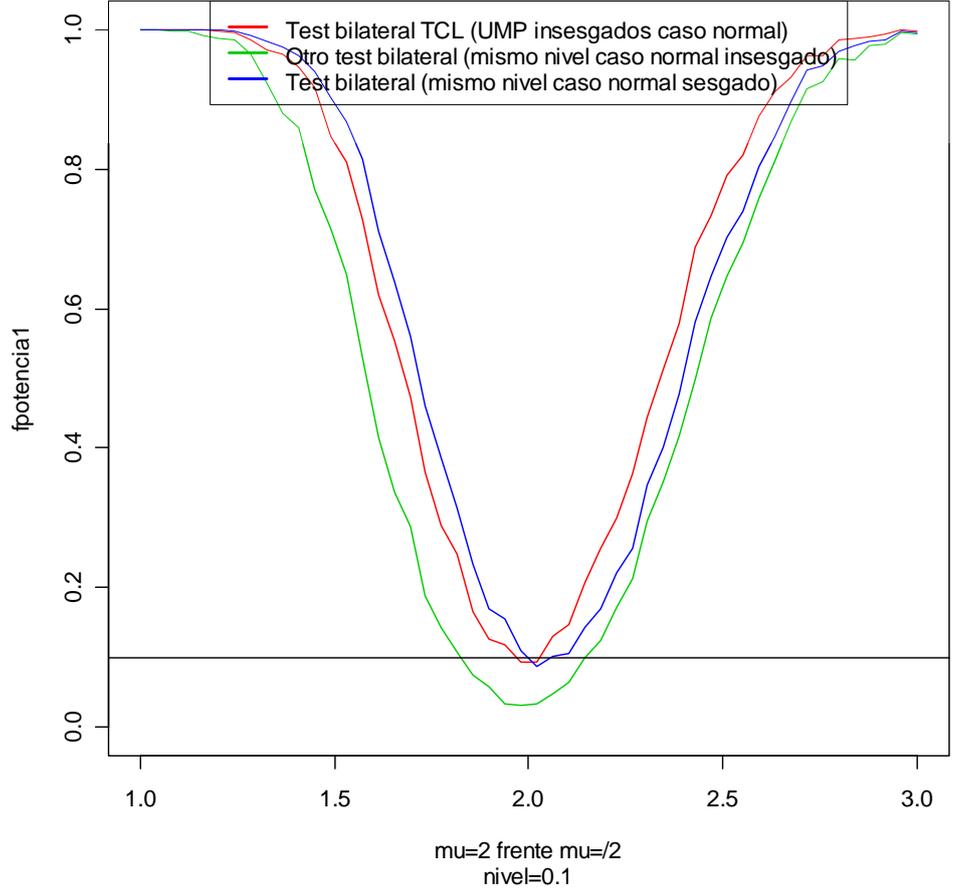
Comparación de funciones de potencia por MC (caso normal)



Comparación de funciones de potencia por MC (caso exponencial)

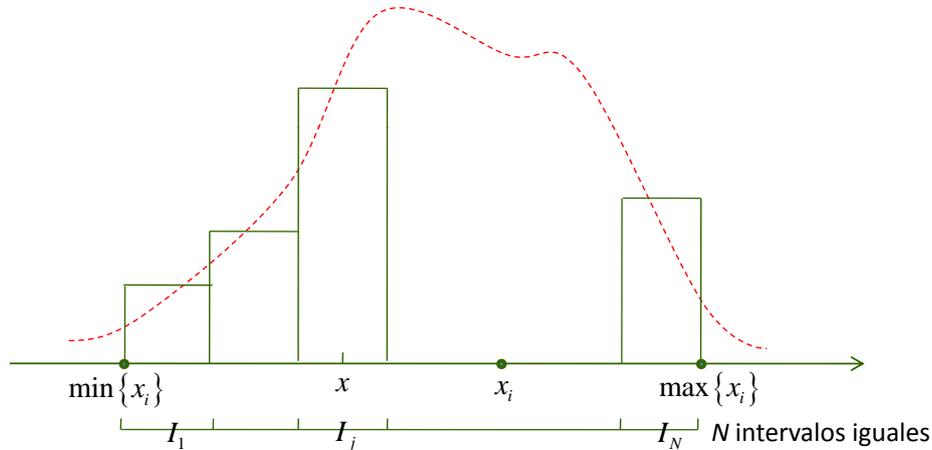


Comparación de funciones de potencia por MC (caso chicuadrado)



$$f(x) = \lim_{h \rightarrow 0} \frac{P(x-h < X \leq x+h)}{2h}$$

1. HISTOGRAMA: aplica la idea de la densidad discretizando el rango muestral.



$$\hat{f}_N(x) = \frac{\frac{\#\{x_i \in I_j\}}{\max\{x_i\} - \min\{x_i\}}}{N} \quad \forall x \in I_j, j \in \{1, \dots, N\}$$

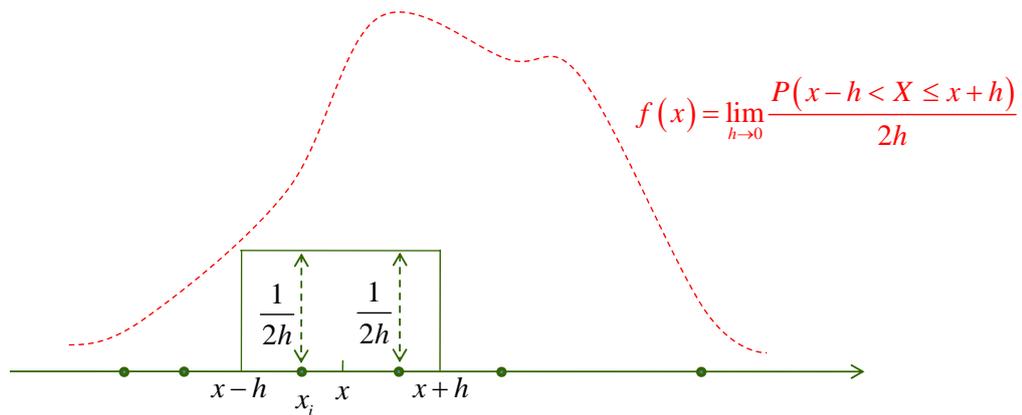
Notemos que el histograma es un estimador de la densidad bien definido:

$$\int_{-\infty}^{\infty} \hat{f}_N(x) dx = \sum_{j=1}^N \frac{\max\{x_i\} - \min\{x_i\}}{N} \times \frac{\frac{\#\{x_i \in I_j\}}{\max\{x_i\} - \min\{x_i\}}}{N} = \sum_{j=1}^N \frac{\#\{x_i \in I_j\}}{n} = \frac{n}{n} = 1$$

NOTA 1. El hecho de discretizar el rango muestral en subintervalos y asignar valor constante sobre ellos provoca que la apariencia del histograma como estimador de la densidad sea “pobre”. Veremos que esta deficiencia la corrige el histograma móvil.

NOTA 2. Al considerar un número fijo de subintervalos en su planteamiento, el histograma será inconsistente como estimador de la densidad.

2. HISTOGRAMA MÓVIL: aplica la idea del histograma fijo, pero de forma continua sobre el rango muestral, en vez de discretizada.



$$\hat{f}_h(x) = \frac{n^{\circ} \text{ obs. } x_i \in (x-h, x+h)}{2h} = \frac{1}{n} \times \frac{n^{\circ} \text{ obs. } x_i \in (x-h, x+h)}{2h} = \frac{1}{n} \times \left(0 + \dots + \frac{1}{2h} + \dots + 0 \right)_{x_i}$$

Esta estimación de la densidad promedia pesos sobre los datos, que valdrán $1/2h$ si los datos están en un entorno de x de radio h , y tomarán el valor 0 en otro caso.

NOTA 1. El histograma móvil admite el siguiente formato:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-x_i}{h}\right)$$

$$\left[K(u) = \begin{cases} \frac{1}{2} & \text{si } |u| \leq 1 \\ 0 & \text{en otro caso} \end{cases} \Rightarrow K\left(\frac{x-x_i}{h}\right) = \begin{cases} \frac{1}{2} & \text{si } x_i \in (x-h, x+h) \\ 0 & \text{en otro caso} \end{cases} \right]$$

NOTA 2. El histograma móvil es un estimador de la densidad bien definido:

$$\int_{-\infty}^{\infty} \hat{f}_h(x) dx = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-x_i}{h}\right) dx = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \frac{1}{h} K(z) h dz = \frac{n}{n} = 1$$

NOTA 3. La elección de h , parámetro ventana, se verá más adelante.

NOTA 4. Observemos que, con pocos datos, el histograma móvil no proporciona una estimación suave de la densidad, pues el entorno de x podría desplazarse ligeramente conservando los mismos datos, y bruscamente perder o ganar alguno (con el mismo peso).

3. MÉTODO NÚCLEO: el estimador de Parzen-Rosemblat generaliza al histograma móvil, al promediar pesos variables sobre los datos que penalizan la distancia de éstos al punto x .

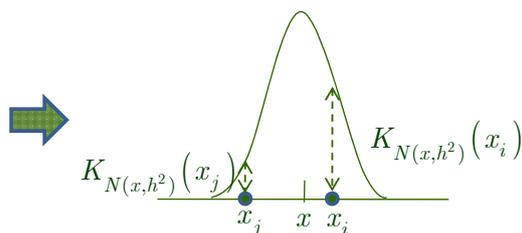
$$\hat{f}_h(x) = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \quad \forall x \in \mathbb{R},$$

donde

- h es el parámetro de suavización o ventana (muy importante).
- K es la función núcleo, cualquier densidad (mucho menos importante).

Si, por ejemplo, K es una densidad $N(0,1)$, $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\}$, entonces

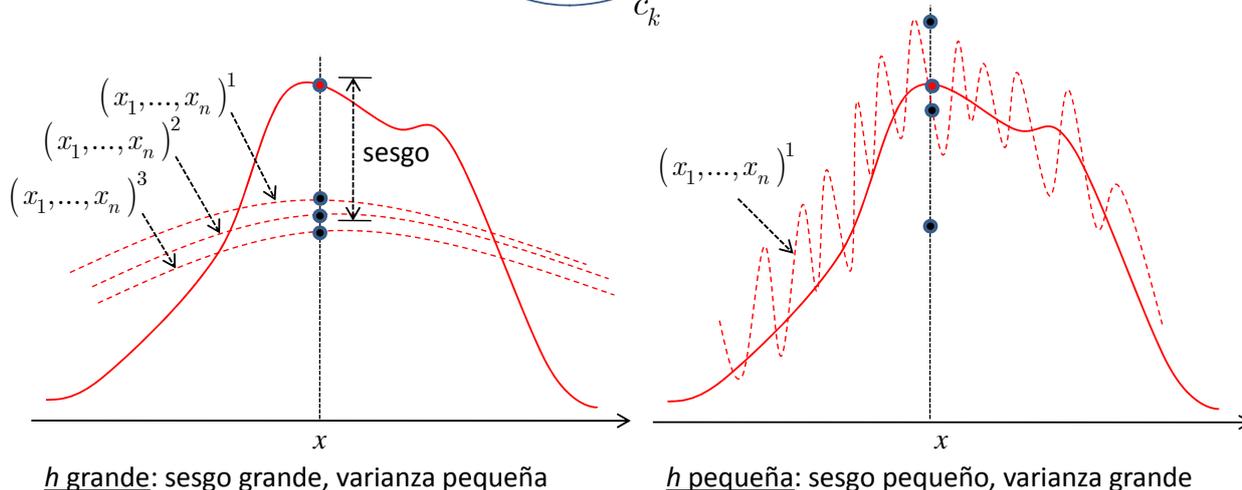
$$\begin{aligned} \frac{1}{h} K\left(\frac{x - x_i}{h}\right) &= \\ &= \frac{1}{h} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2h^2}(x - x_i)^2\right\} = \\ &= K_{N(x, h^2)}(x_i) \end{aligned}$$



Sesgo y varianza del estimador de Parzen-Rosemblat

- $E[\hat{f}_h(x)] = f(x) + \frac{h^2}{2} \left[\int t^2 K(t) dt \right] f''(x) + O(h^4)$
 d_k sesgo de $\hat{f}_h(x)$

- $V[\hat{f}_h(x)] = \frac{1}{nh} \left[\int (K(t))^2 dt \right] f(x) - \frac{1}{n} (f(x))^2 + O(n^{-1}h)$
 c_k



Elección del parámetro de suavización

- Criterio local de error

$$\begin{aligned}MSE_x(h) &= E\left[\left(\hat{f}_h(x) - f(x)\right)^2\right] = \left(\text{sesgo } \hat{f}_h(x)\right)^2 + \text{var } \hat{f}_h(x) = \\&= \frac{h^4}{4} d_k^2 f''(x)^2 + \frac{c_k}{nh} f(x) - \frac{1}{n} f(x)^2 + O(h^6) + O(n^{-1}h)\end{aligned}$$

- Criterio global de error

$$\begin{aligned}MISE(h) &= \int MSE_x(h) dx = \\&= \underbrace{\frac{c_k}{nh} - \frac{1}{n} \int f(x)^2 dx + \frac{d_k^2}{4} h^4 \int f''(x)^2 dx}_{AMISE(h)} + O(n^{-1}h^2) + O(h^6) \\&\quad [h \rightarrow 0 \text{ y } nh \rightarrow \infty]\end{aligned}$$

$$h_{AMISE} = \arg \min_h AMISE(h) = \frac{c_k}{nd_k^2 \int f''(x)^2 dx} = c_0 n^{-1/5}$$

NOTA 1. El estimador de Parzen-Rosemblat generaliza al histograma móvil, en el cual la función núcleo es la densidad uniforme entre -1 y 1.

NOTA 2. El estimador de Parzen-Rosemblat es un estimador de la densidad bien definido (misma demostración que para el histograma móvil).

NOTA 3. Basta con considerar una función núcleo gaussiana para observar que el estimador de Parzen-Rosemblat proporciona una estimación suave de la función de densidad incluso con pocos datos, pues los pesos cambian al cambiar x. Recordemos que con el núcleo uniforme podía no ser así.

NOTA 4. Intuitivamente, la penalización de los datos alejados de x (vecinos no poróximos) viene a compensar el hecho de que la ventana sea distinta de cero.

INTRODUCCIÓN A LA ESTIMACIÓN NO PARAMÉTRICA DE DENSIDAD. APLICACIÓN P5.

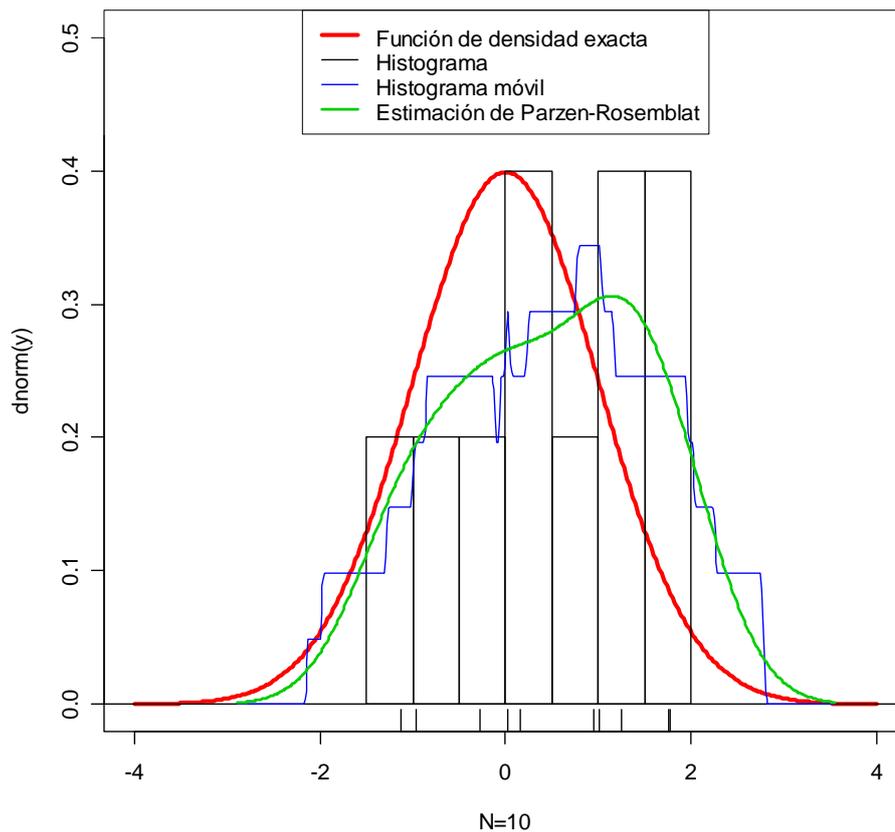
Código fuente en R que compara por Monte Carlo, bajo la suposición de diferentes modelos, el histograma, el histograma móvil y el estimador de Parzen-Rosemblat. Además, estudia el efecto de la elección de la ventana en este último.

```
• x<-rnorm(100,0,1)          CASO N(0,1)
• y<-seq(-4,4,length=400)
• plot(y,dnorm(y),type="l",xlab="N=100",xlim=c(-4,4),ylim=c(0,0.5),col=2,lwd=3,main="Función de densidad (N(0,1))")
• abline(h=0.00)
• rug(x)
• hist(x,freq=FALSE,add=TRUE)
• lines(density(x,kernel="rectangular"),col=4,xlim=c(-4,4),ylim=c(0,0.5))
• lines(density(x),lwd=2,xlim=c(-4,4),ylim=c(0,0.5),col=3)
• legend("top",legend=c("Función de densidad exacta","Histograma","Histograma móvil","Estimación de Parzen-Rosemblat"),lwd=c(3,1,1,2),col=c(2,1,4,3))
```

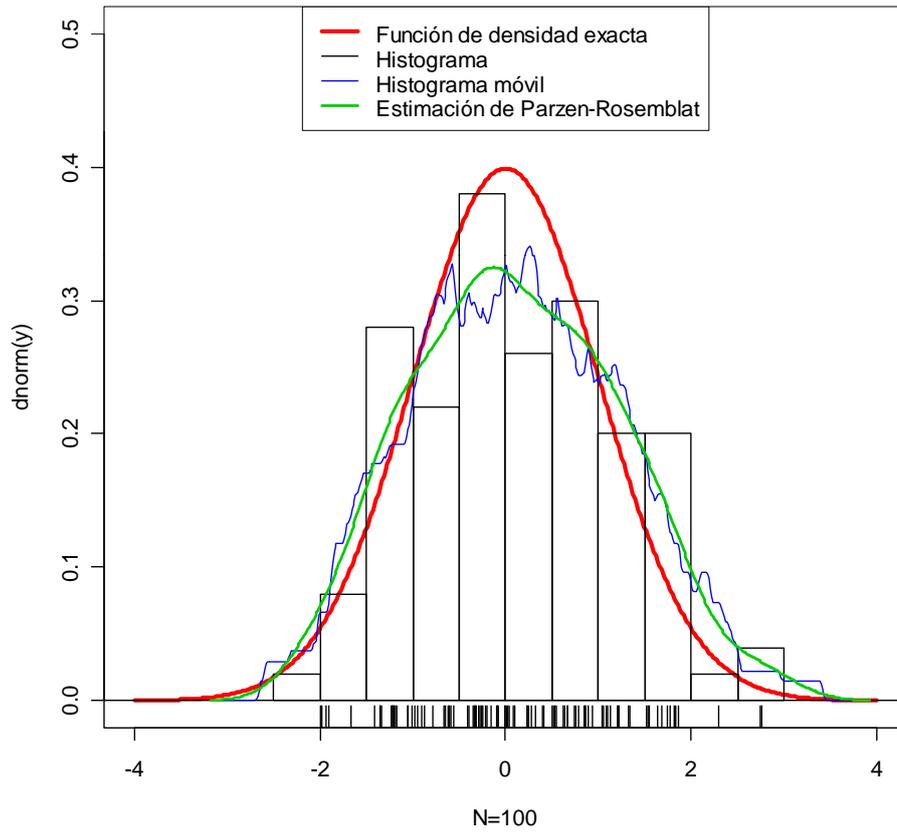
```
• x<-rexp(100,rate=1)       CASO EXPONENCIAL (1)
• y<-seq(-0.1,6,length=400)
• plot(y,dexp(y),type="l",xlab="N=100",xlim=c(0,6),ylim=c(0,0.1),col=2,lwd=3,main="Función de densidad (N(0,1))")
• abline(h=0.00)
• rug(x)
• hist(x,freq=FALSE,add=TRUE)
• lines(density(x,kernel="rectangular"),col=4)
• lines(density(x),lwd=2,col=3)
• legend("top",legend=c("Función de densidad exacta","Histograma","Histograma móvil","Estimación de Parzen-Rosemblat"),lwd=c(3,1,1,2),col=c(2,1,4,3))
```

```
• x<-rchisq(100,6)         CASO CHICUADRADO (6)
• y<-seq(-0.1,25,length=400)
• plot(y,dchisq(y,6),type="l",xlab="N=100",xlim=c(0,25),ylim=c(0,0.16),col=2,lwd=3,main="Función de densidad (N(0,1))")
• abline(h=0.00)
• rug(x)
• hist(x,freq=FALSE,add=TRUE)
• lines(density(x,kernel="rectangular"),col=4)
• lines(density(x),lwd=2,col=3)
• legend("top",legend=c("Función de densidad exacta","Histograma","Histograma móvil","Estimación de Parzen-Rosemblat"),lwd=c(3,1,1,2),col=c(2,1,4,3))
```

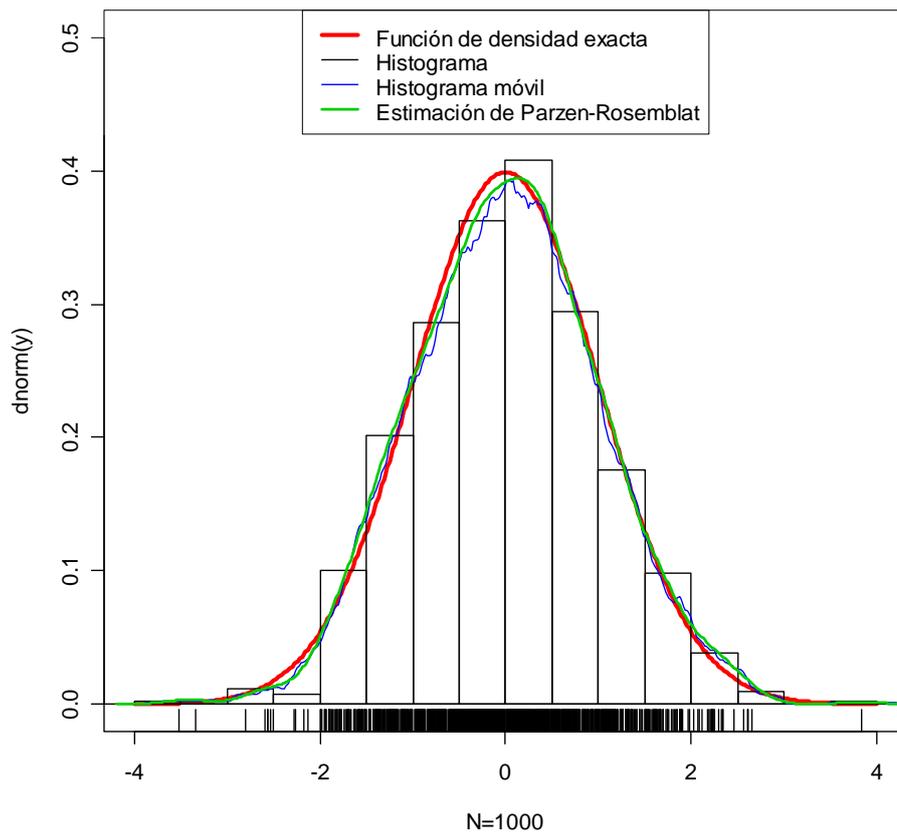
Función de densidad (N(0,1))



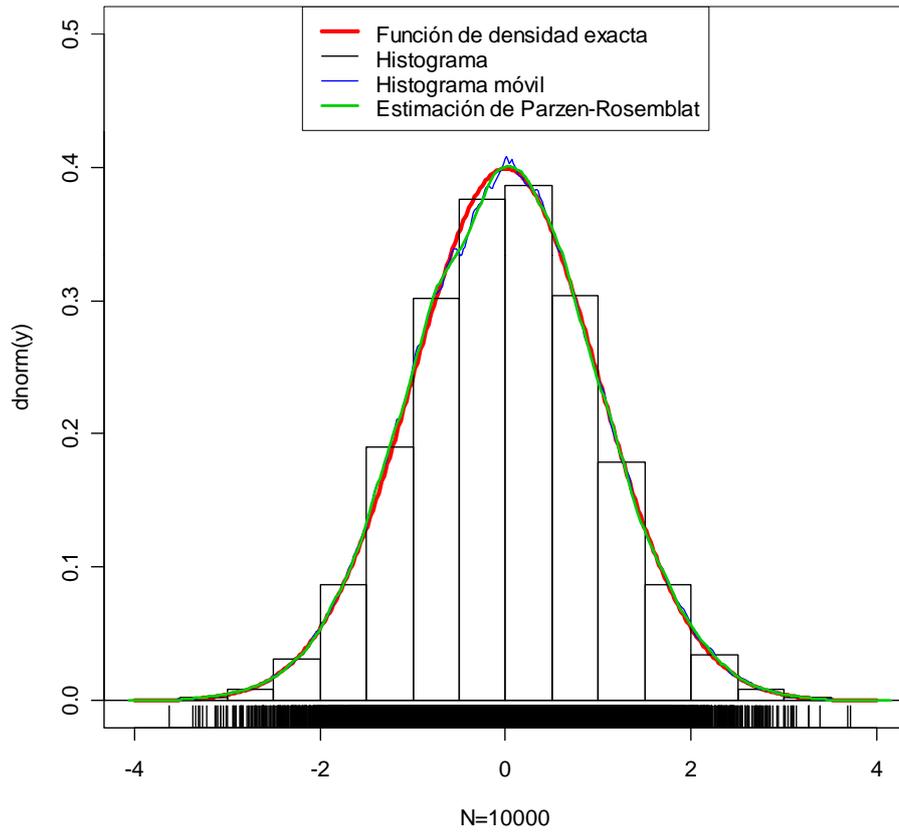
Función de densidad (N(0,1))



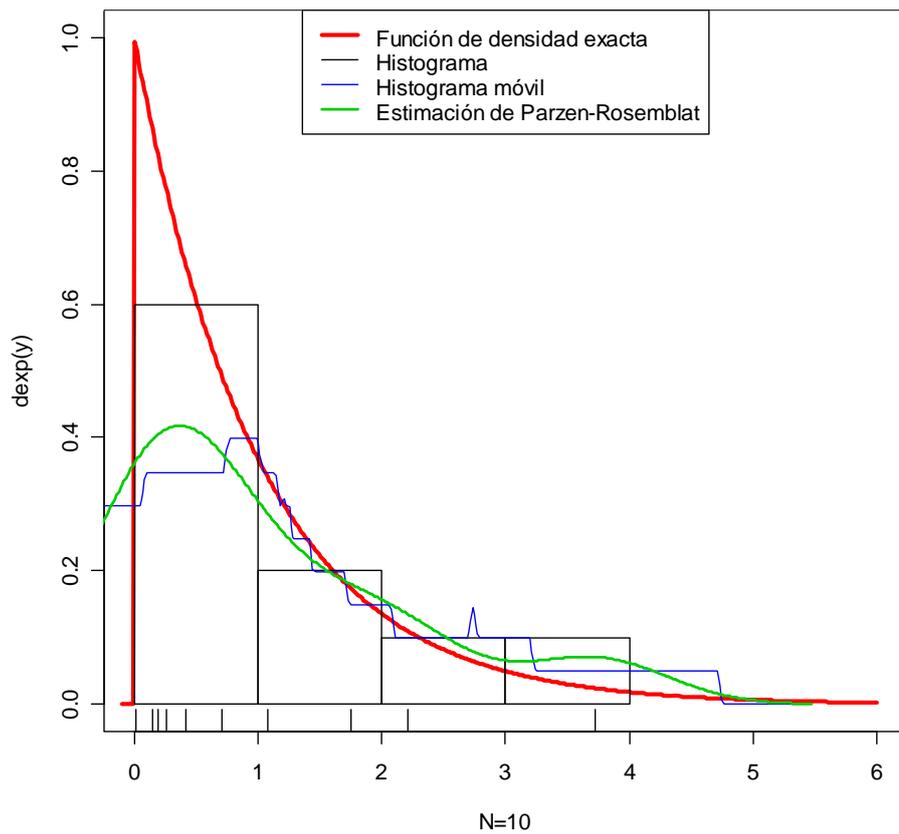
Función de densidad (N(0,1))



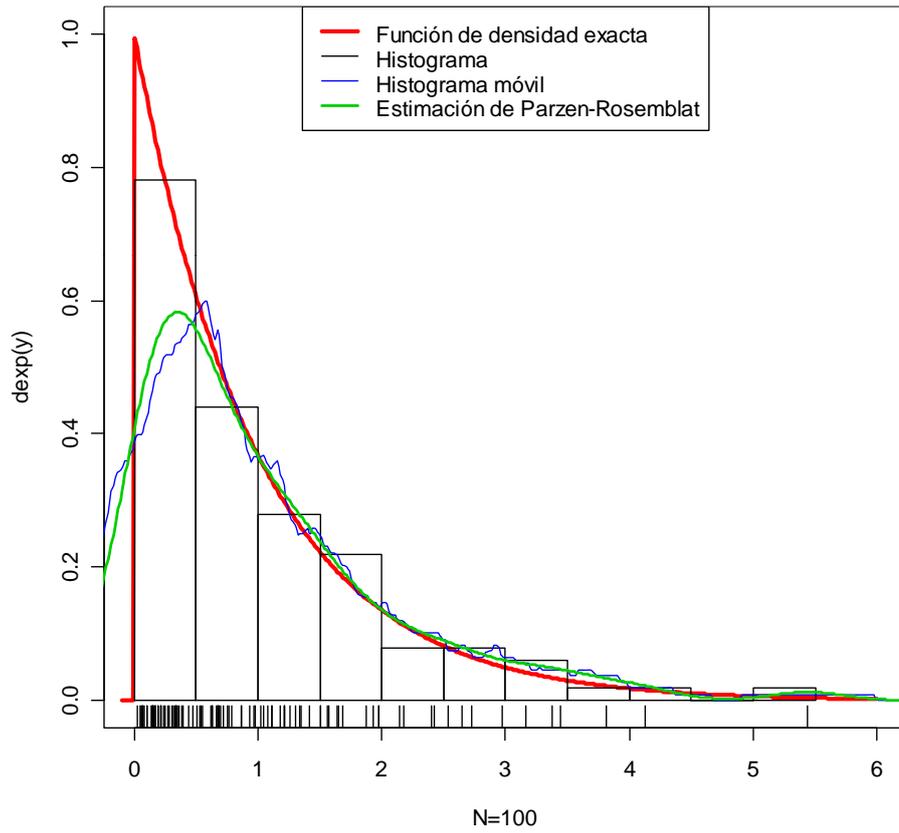
Función de densidad (N(0,1))



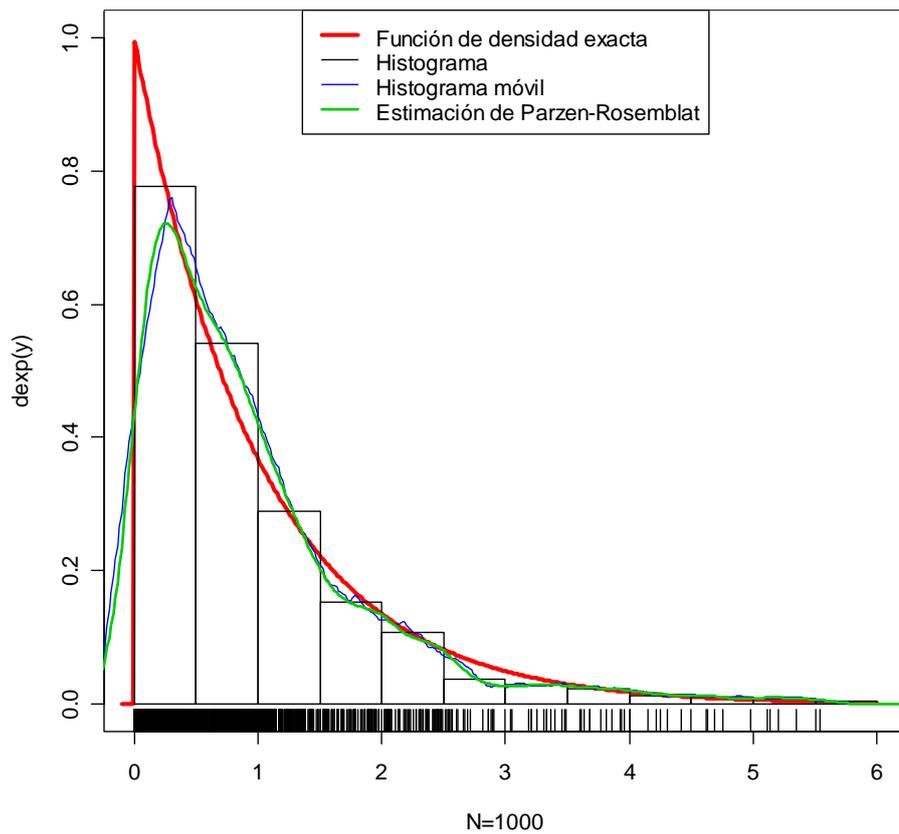
Función de densidad (exponencial (1))



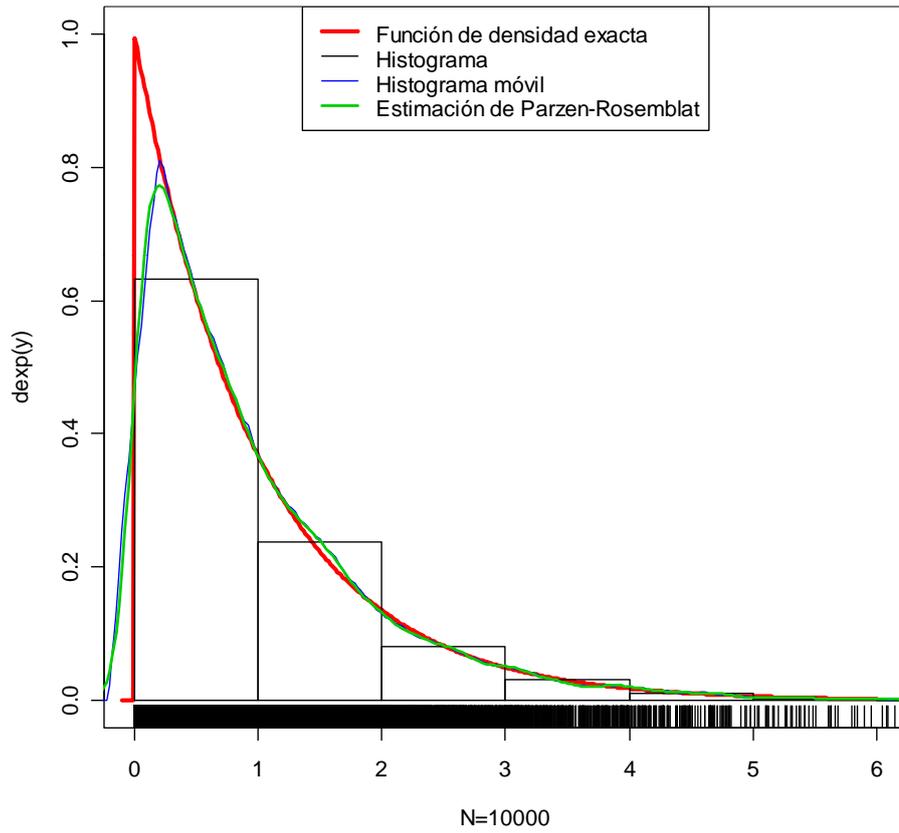
Función de densidad (exponencial (1))



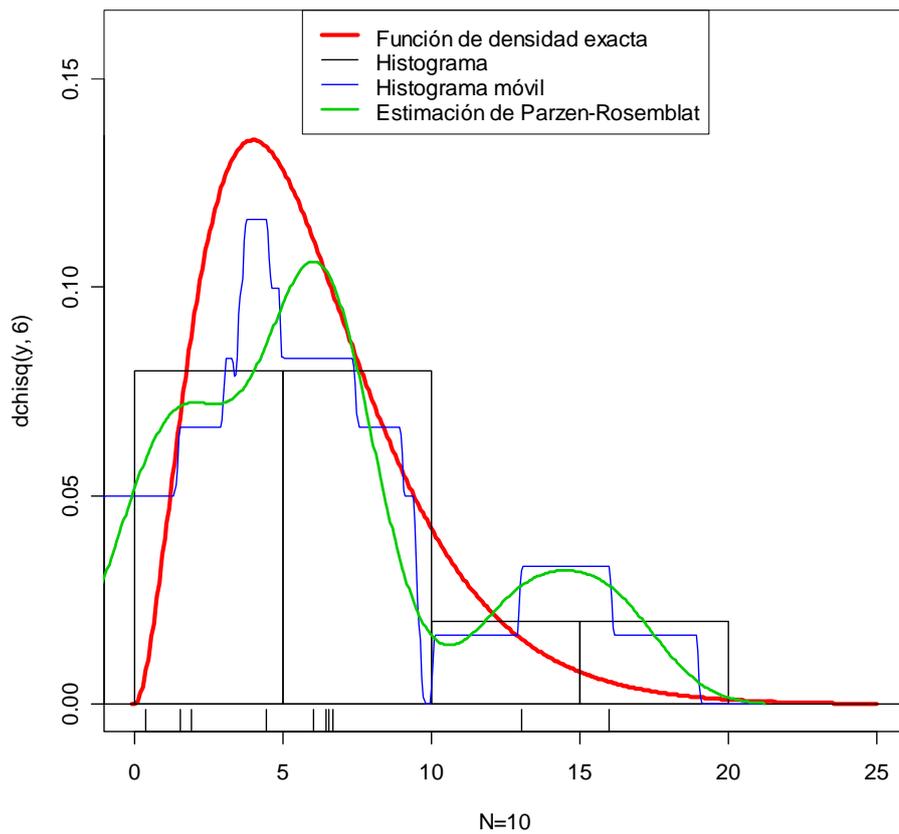
Función de densidad (exponencial (1))



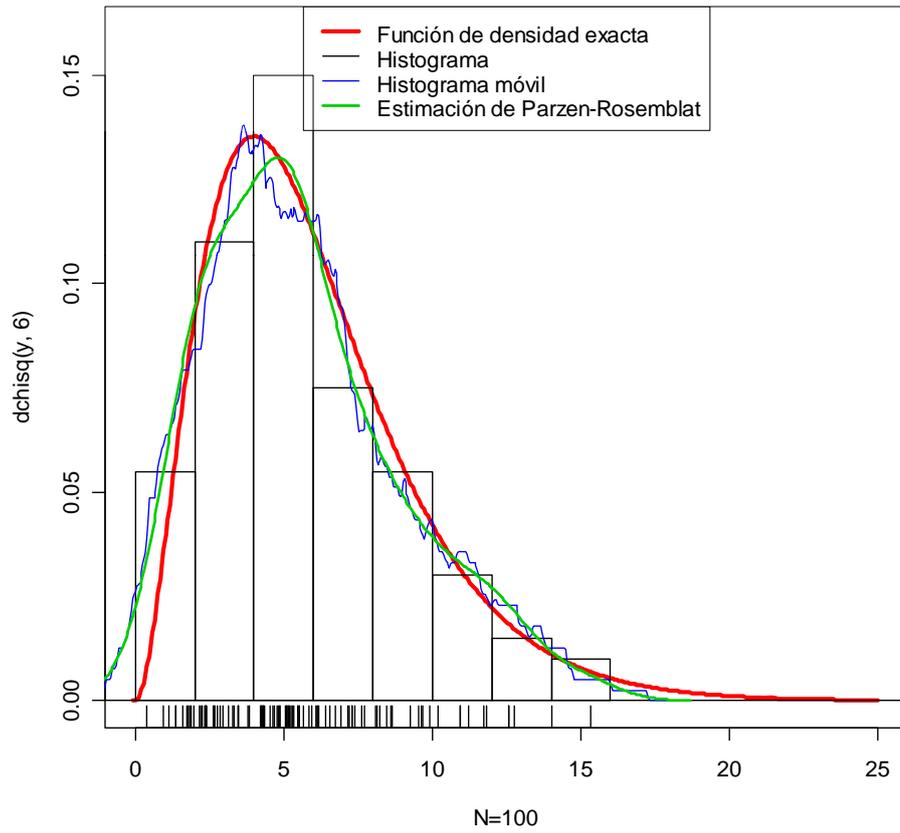
Función de densidad (exponencial (1))



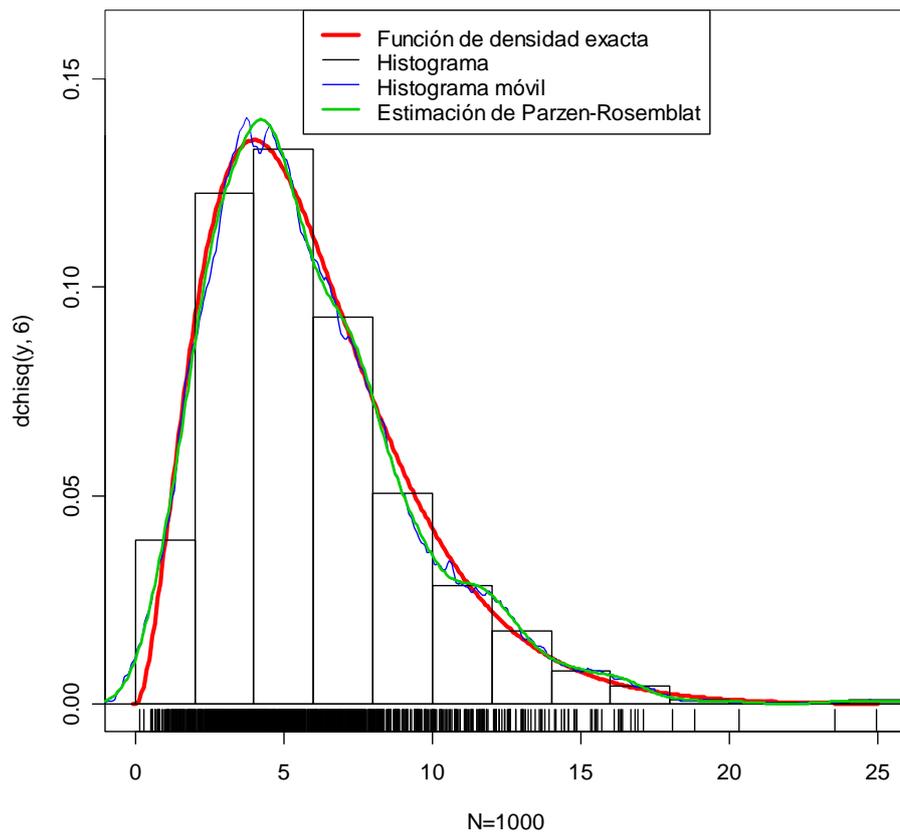
Función de densidad (chi2(6))



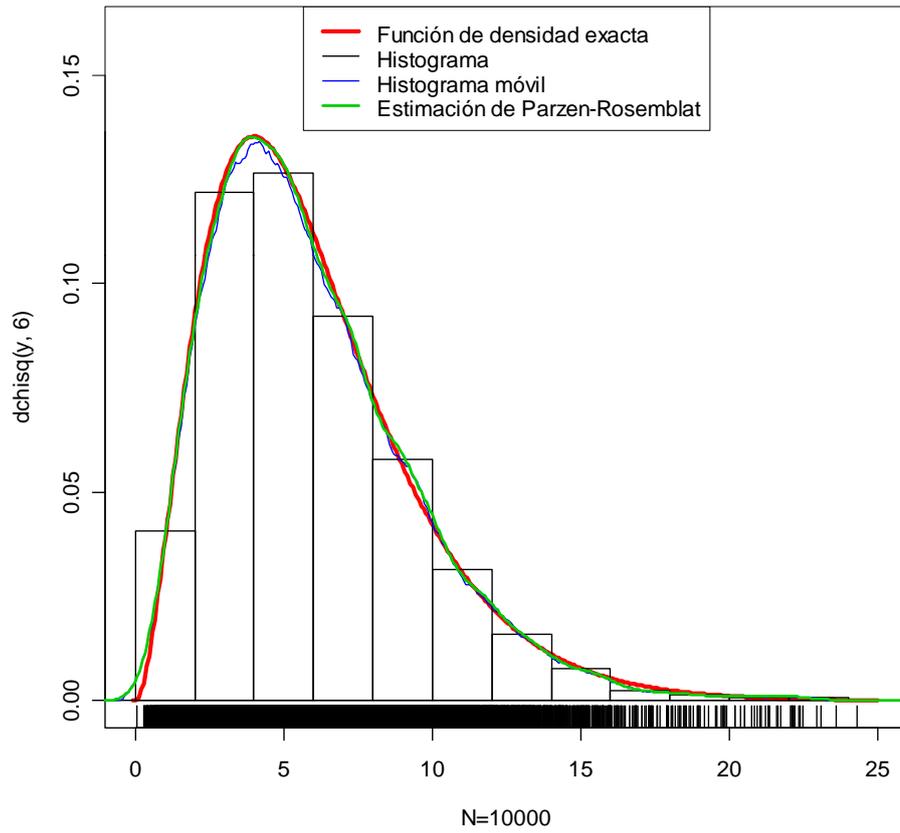
Función de densidad (chi2(6))



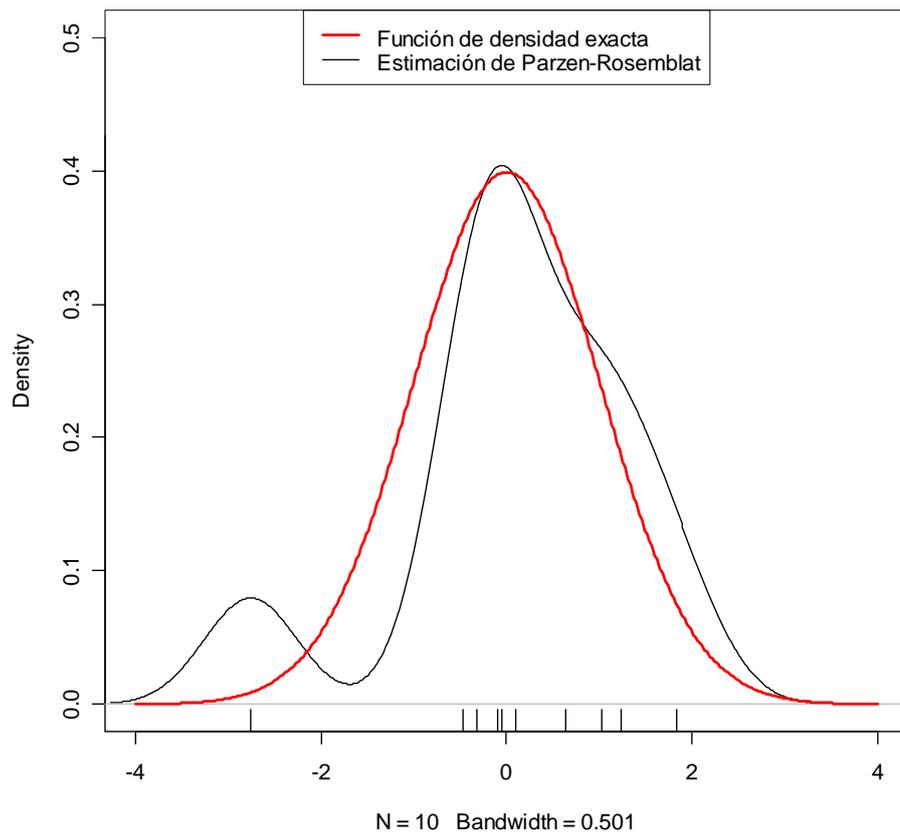
Función de densidad (chi2(6))



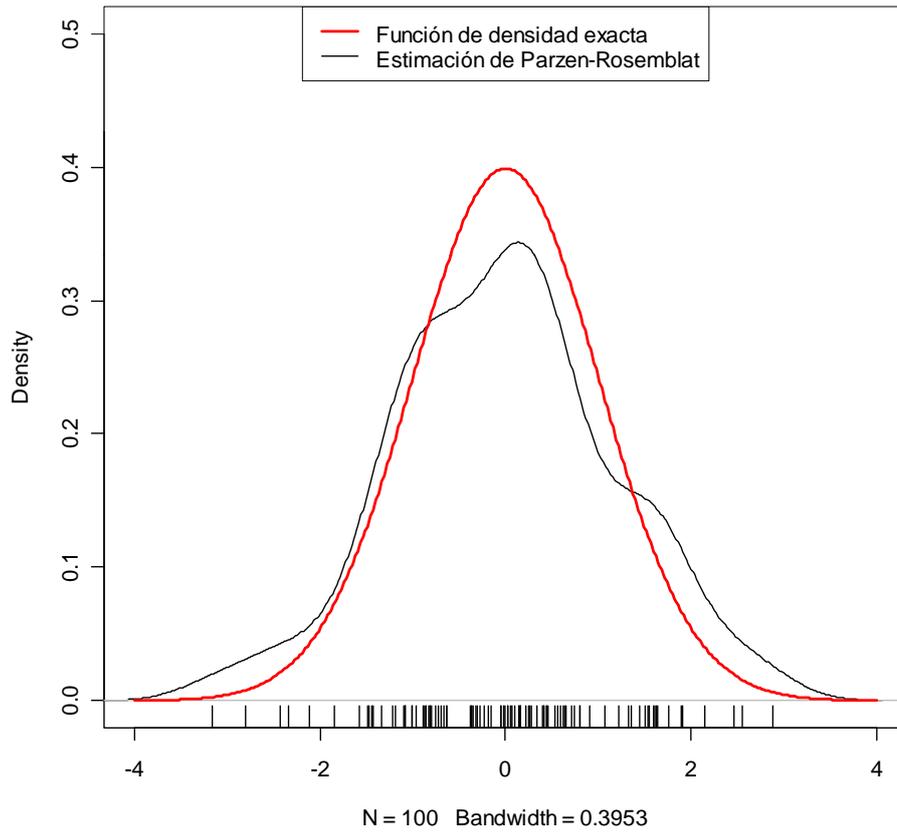
Función de densidad (chi2(6))



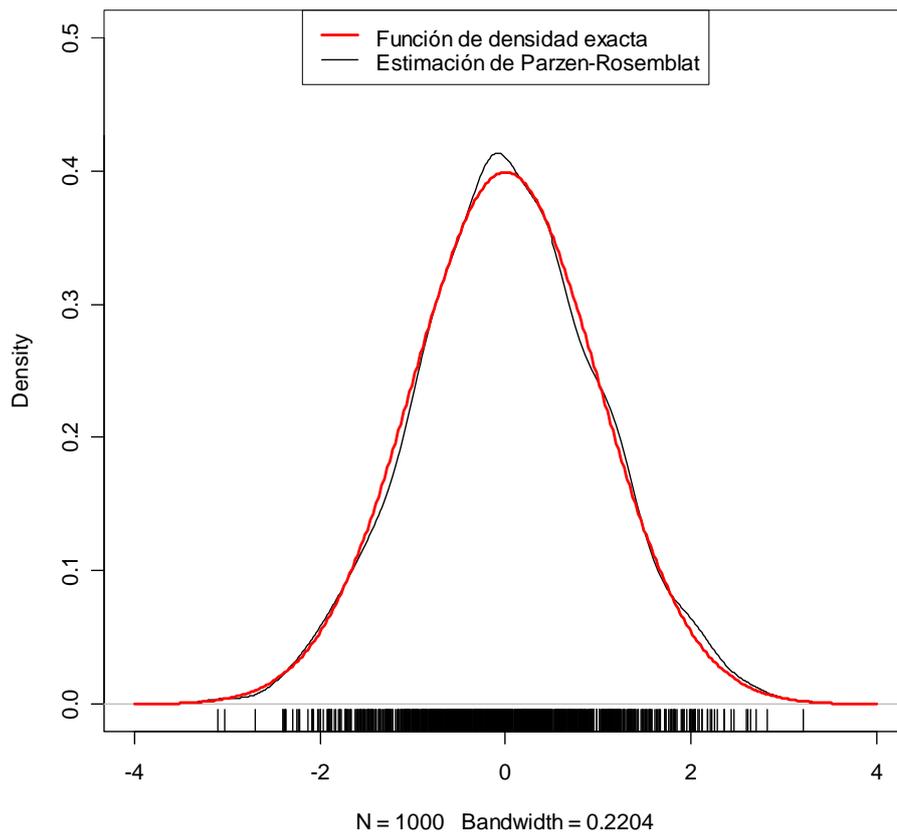
Función de densidad (N(0,1))



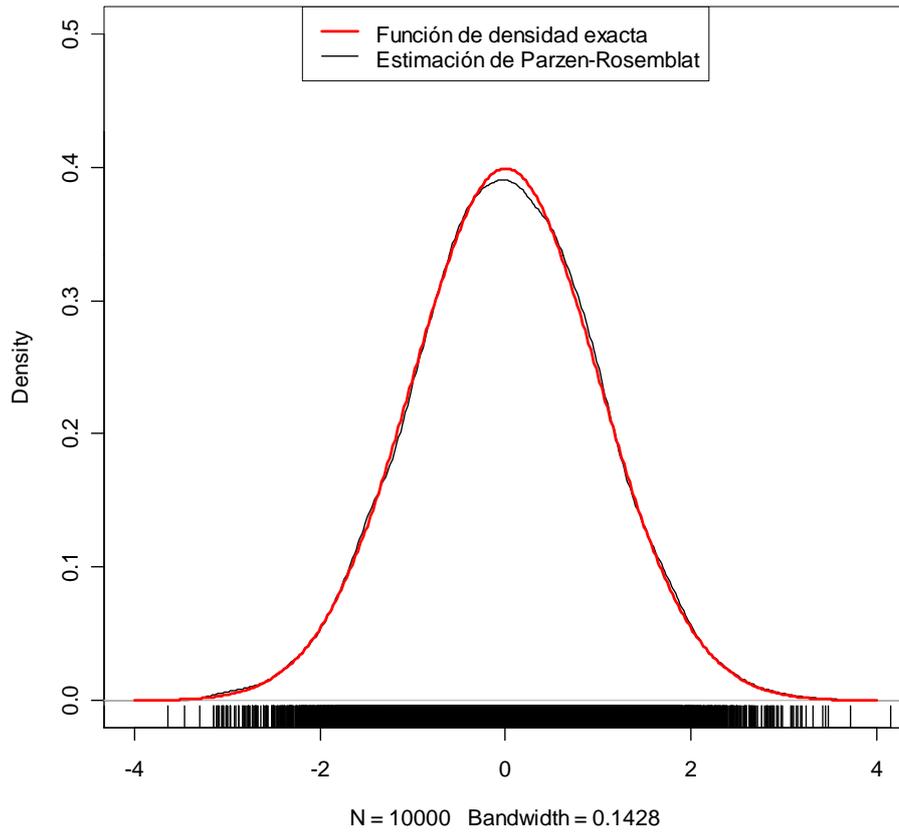
Función de densidad (N(0,1))



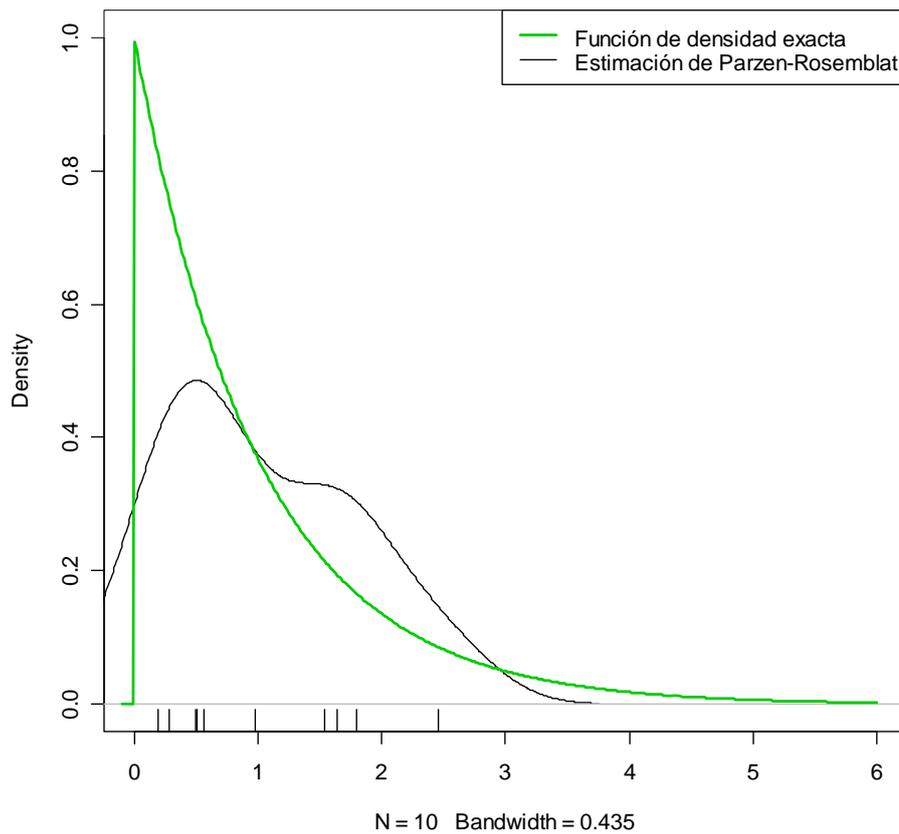
Función de densidad (N(0,1))



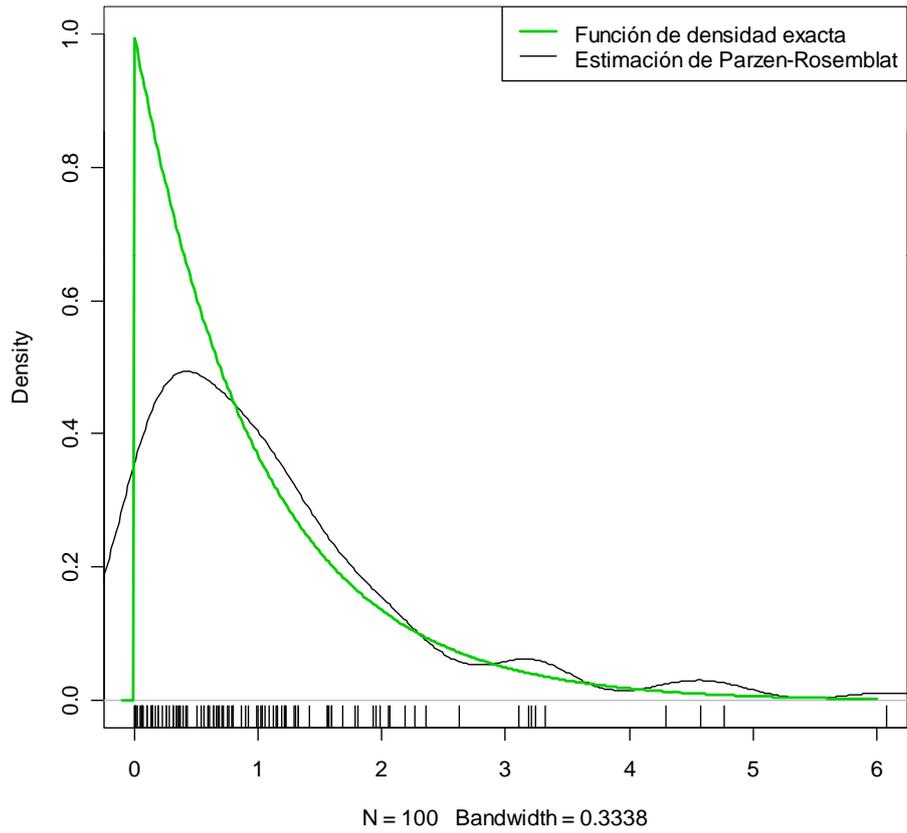
Función de densidad (N(0,1))



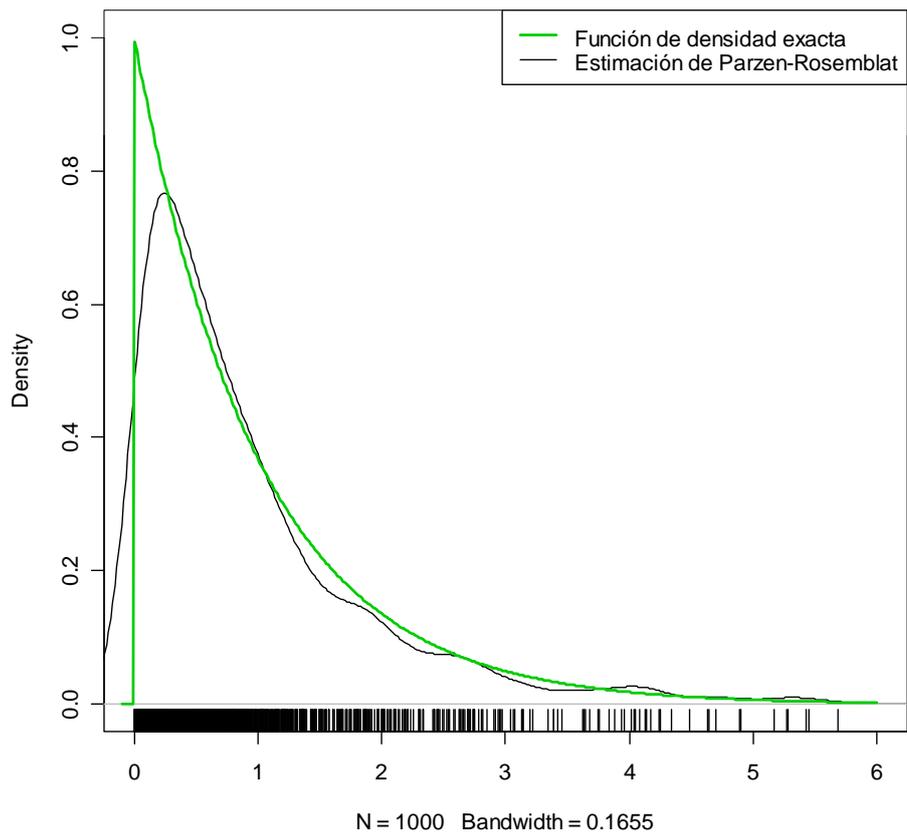
Función de densidad (Exponencial (1))



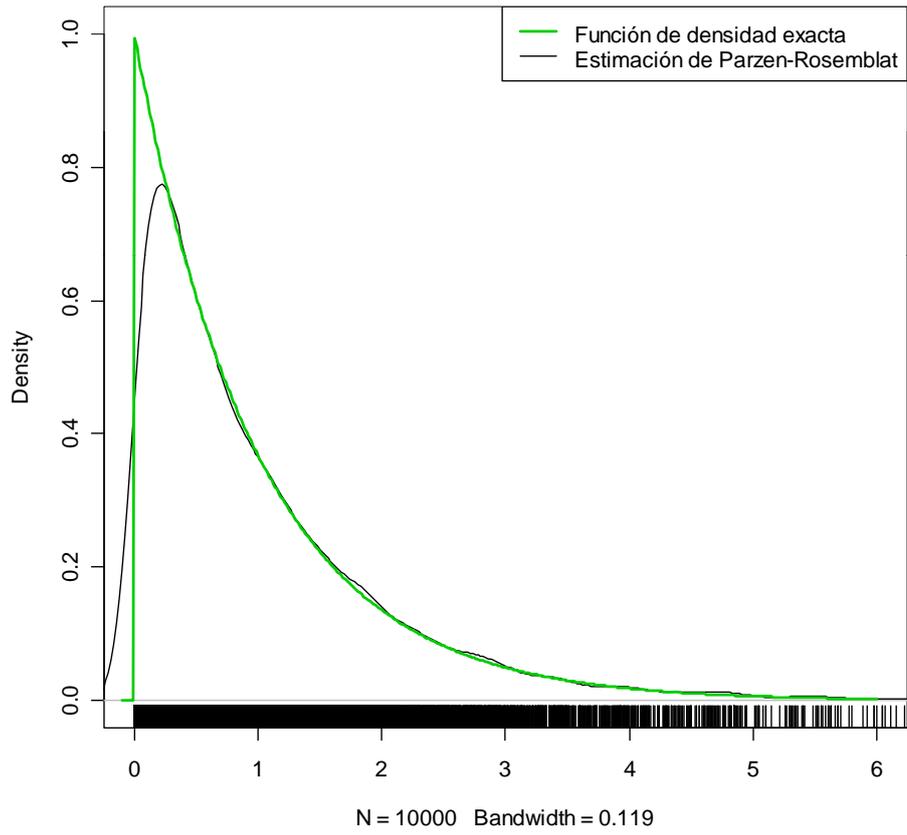
Función de densidad (Exponencial (1))



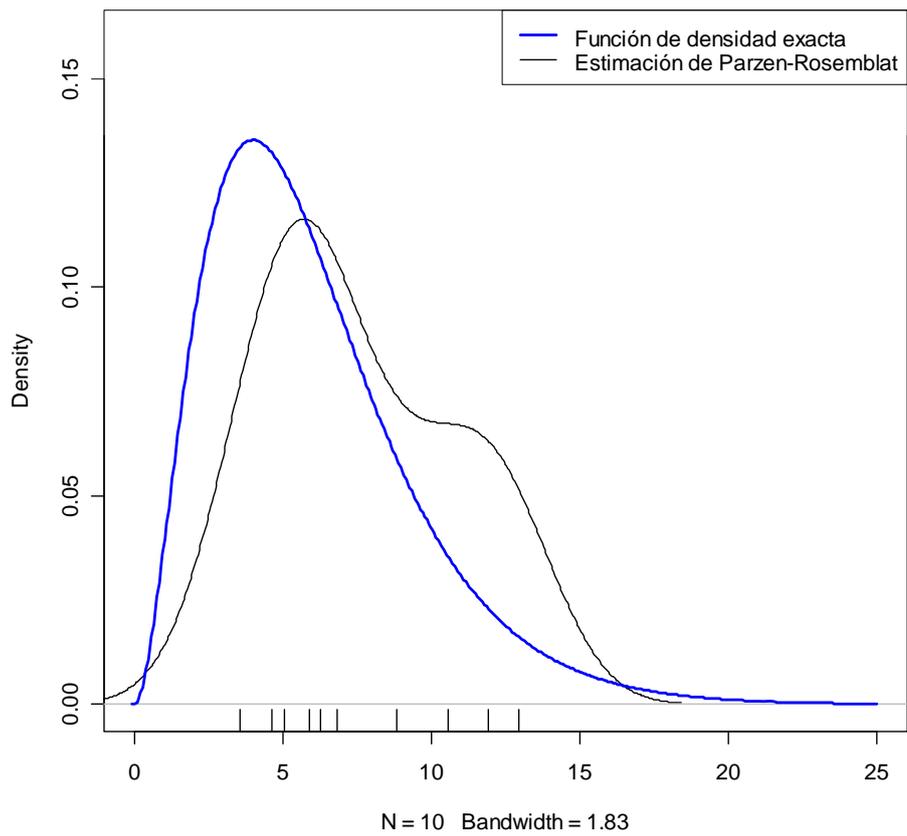
Función de densidad (Exponencial (1))



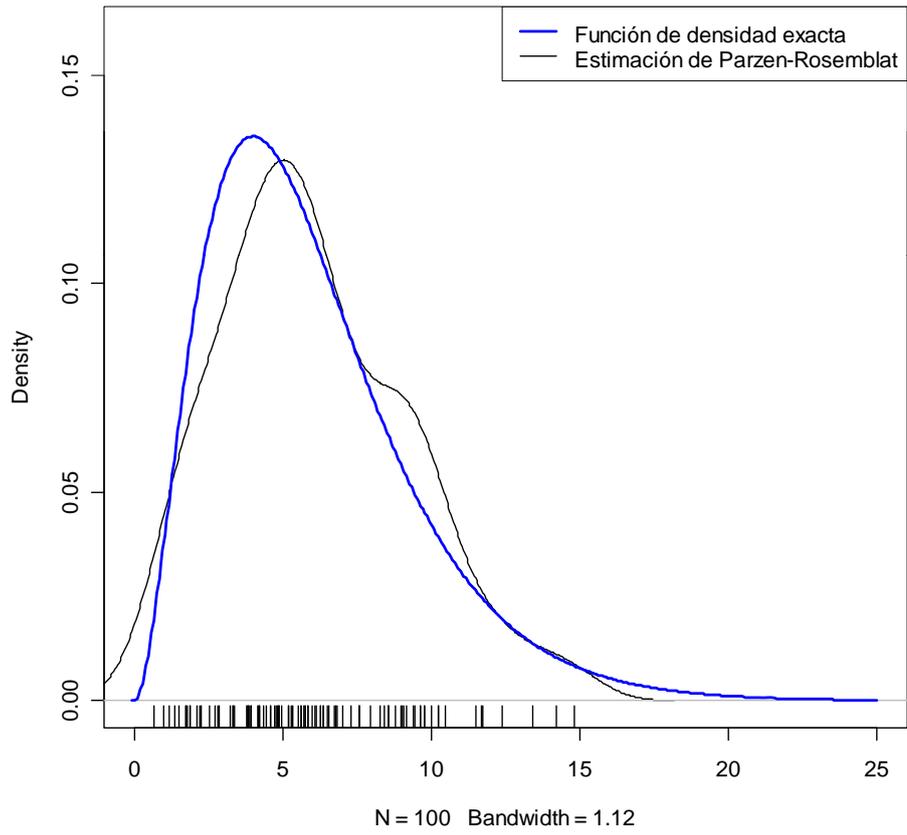
Función de densidad (Exponencial (1))



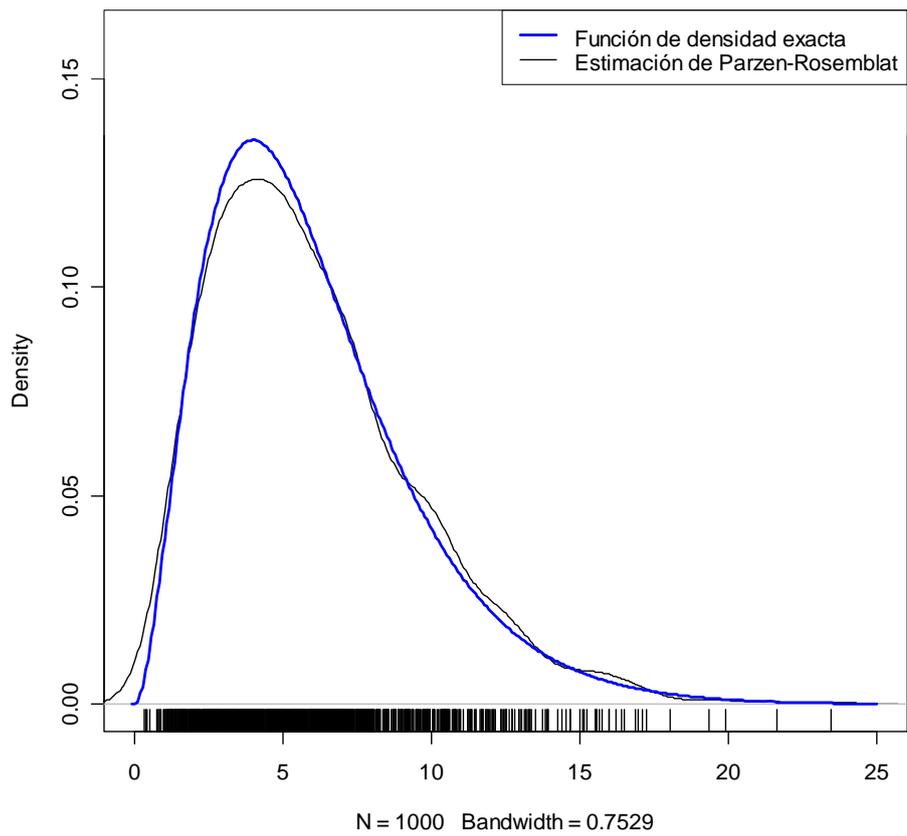
Función de densidad (Chi2 (6))



Función de densidad (Chi2 (6))



Función de densidad (Chi2 (6))



Función de densidad (Chi2 (6))

