

INFERENCIA BASADA EN PROCESOS EMPÍRICOS TRANSFORMADOS.

Enrique M. Cabaña *

Part I

Pruebas clásicas.

1 Introducción. Estadísticos basados en el proceso empírico.

Cada procedimiento de inferencia estadística basado en la información proporcionada por una muestra aleatoria simple no ordenada (X_1, \dots, X_n) de cierta distribución de probabilidades F , puede considerarse también basado en la función de distribución empírica de la muestra

$$F_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \in A\}}$$

o en el proceso empírico relativo a la probabilidad F_0

$$b_n(A) = \sqrt{n}(F_n(A) - F_0(A)),$$

puesto que estas dos últimas medidas aleatorias contienen la misma información que el conjunto desordenado de las n variables aleatorias (X_1, \dots, X_n) que constituyen la muestra.

Consideremos tres ejemplos bien conocidos, que corresponden a pruebas de ajuste, es decir, a pruebas de la hipótesis nula $\mathcal{H}_0: "F = F_0"$. Denotaremos a la función de distribución de una medida, con el mismo símbolo que a la medida, es decir, utilizaremos las notaciones $F_n(x) = F_n((-\infty, x])$, $b_n(x) = b_n((-\infty, x])$ o $F_0(x) = F_0((-\infty, x])$.

Ejemplo 1: Tomemos una partición del recorrido de las variables aleatorias en conjuntos $\{R_1, \dots, R_k\}$, y llamemos $O = (O_1, \dots, O_k)$, $O_h = \sum_{i=1}^n \mathbf{1}_{\{X_i \in R_h\}}$ al vector que cuenta

*Facultad de Ciencias Económicas y de Administración, y Centro de Matemática, Facultad de Ciencias, Universidad de la República, Montevideo, URUGUAY

cuántas observaciones de la muestra pertenecen a cada uno de esos conjuntos. La prueba de ajuste χ^2 rechaza \mathcal{H}_0 cuando el estadístico

$$Q = \sum_{h=1}^k \frac{(O_h - nF_0(R_h))^2}{nF_0(R_h)},$$

cuya distribución asintótica bajo \mathcal{H}_0 es χ_{k-1}^2 , es grande, mayor que una constante que se elige para que el nivel de la prueba sea α .

El estadístico Q puede escribirse de manera equivalente, en términos del proceso empírico: $Q = \sum_{h=1}^k \frac{b_n^2(R_h)}{F_0(R_h)}$.

Ejemplo 2: La prueba de Kolmogorov-Smirnov, consiste en rechazar \mathcal{H}_0 cuando $D_n = \sup_{x \in \mathbf{R}} |F_n(x) - F_0(x)|$ es grande, mayor que una constante $c_n(\alpha)$ que se elige para que el nivel de la prueba sea α , como en el caso anterior. El estadístico D_n puede escribirse en la forma $D_n = \sup_{x \in \mathbf{R}} |b_n(x)|/\sqrt{n}$.

Ejemplo 3: Las pruebas de ajuste del tipo de Cramér-von Mises, rechazan \mathcal{H}_0 cuando estadísticos de la forma $W_n^2 = \int_{-\infty}^{\infty} \psi(F_0(x))(F_n(x) - F_0(x))^2 dF_0(x)$ son más grandes que una constante adecuada al nivel deseado. Cuando la función de pesos ψ se elige constante, igual a 1, W_n es el estadístico de Cramér-von Mises propiamente dicho. Cuando $\psi(x) = (F_0(x)(1 - F_0(x)))^{-1}$, es el estadístico de Anderson y Darling.

Como es obvio, también en este caso puede escribirse el estadístico en función del proceso empírico: $W_n^2 = \int_{-\infty}^{\infty} \psi(F_0(x))(b_n(x))^2 dF_0(x)/\sqrt{n}$.

2 Distribución asintótica del proceso empírico bajo \mathcal{H}_0 .

Las distribuciones asintóticas de los estadísticos de la sección precedente se obtienen inmediatamente a partir de las distribuciones asintóticas del proceso empírico.

En efecto, es bien conocido que bajo la hipótesis nula, el proceso empírico tiene por distribución límite, cuando $n \rightarrow \infty$, la de un puente browniano asociado a la probabilidad F_0 .

Resulta más simple describir el comportamiento asintótico del proceso empírico asociado a una muestra aleatoria simple U_1, \dots, U_n de la distribución uniforme en $[0, 1]$, y generalizar a muestras X_1, \dots, X_n de cualquier distribución F_0 mediante la transformación canónica que lleva X_i en $U_i = F_0(X_i)$.

El puente empírico $b_n^{(U)}(u) = b_n(F_0^{-1}(u))$ ($0 \leq u \leq 1$) es una variable aleatoria en el espacio D de las funciones reales en $[0, 1]$, continuas por la derecha con límite por la izquierda, con la métrica de Prohorov $d(g_1, g_2) = \inf\{\varepsilon : \sup_{u \in [0, 1]} |g_1(u) - g_2(\lambda(u))| < \varepsilon, \sup_{u \in [0, 1]} |\lambda(u) - u| < \varepsilon, \lambda \in \Lambda\}$, $\Lambda = \{\lambda : [0, 1] \rightarrow [0, 1], \text{ creciente, } \lambda(0) = 0, \lambda(1) = 1\}$.

El puente browniano típico es el proceso gaussiano $(b(u))_{0 \leq u \leq 1}$ centrado, con covariancias $\mathbf{E}b(u)b(v) = \min\{u, v\} - uv$, que tiene casi seguramente trayectorias continuas,

y se anula en los extremos del intervalo, de modo que puede considerarse una variable aleatoria con valores en D , concentrada con probabilidad uno en la familia C de las funciones continuas en $[0, 1]$.

La convergencia a la que aludimos más arriba es la *convergencia débil en D* : Para cualquier funcional $\psi : D \rightarrow \mathbf{R}$ continua y acotada, se cumple $\lim_{n \rightarrow \infty} \mathbf{E}\psi(b_n^{(U)}) = \mathbf{E}\psi(b)$.

De la convergencia débil de $b_n^{(U)}$ a b , resulta por ejemplo, que $\sup_{0 \leq u \leq 1} |b_n^{(U)}(u)|$ converge en distribución a $\sup_{0 \leq u \leq 1} |b(u)|$, y que cuando ψ define una funcional continua, $\int \psi(u)(b_n^{(U)}(u))^2 du$ a $\int \psi(u)(b(u))^2 du$.

Deshaciendo el cambio de variables canónico, encontramos que $\sup_t |b_n(t)|$ coincide con $\sup_{0 \leq u \leq 1} |b_n^{(U)}(u)|$ y tiene por lo tanto el mismo límite en ley, que puede escribirse como $\sup_t |\bar{b}^{(F_0)}(t)|$, con $b^{(F_0)}(t) = b(F_0(t))$. A un proceso con la distribución de $b^{(F_0)}$ se lo llama *puente browniano asociado a la probabilidad F_0* .

Encontramos de la misma manera que $\int \psi(F_0(t))(b_n(t))^2 dF_0(t)$ converge en ley a $\int \psi(F_0(t))(b^{(F_0)}(t))^2 dF_0(t)$.

3 Distribución asintótica del proceso empírico bajo alternativas contiguas.

Vamos a introducir ahora una sucesión de alternativas \mathcal{H}_n : " $F = F_n$ ", donde la sucesión de las distribuciones de las muestras aleatorias simples $X_1, \dots, X_n \sim F_n$ es contigua a la sucesión de las distribuciones correspondientes a la hipótesis nula de $X_1, \dots, X_n \sim F_0$.

La contigüidad significa que para toda sucesión de conjuntos $A_n \subset \mathbf{R}^n$ que cumple $\lim_{n \rightarrow \infty} \mathbf{P}\{(X_1, \dots, X_n) \in A_n\} = 0$ cuando X_1, \dots, X_n son independientes e idénticamente distribuidas con distribución F_0 , entonces el límite de la misma probabilidad es también cero cuando $X_1, \dots, X_n \sim F_n$.

Vamos a suponer que se cumplen las siguientes hipótesis, que suelen ser cómodamente verificables, y que implican la contigüidad ([13]):

(i) F_n y F_0 tienen densidades f_n y f_0 respecto de una medida μ .

(ii) $\sqrt{\frac{f_n}{f_0}} = 1 + \frac{\delta k_n}{2\sqrt{n}}$ con funciones k_n uniformemente acotadas por una función K tal que $\int K^2(t) dF_0(t) < \infty$, que convergen en $L^2(\mathbf{R}, dF_0)$ a una función k de norma 1 ($\int (k_n - k)^2 dF_0 \rightarrow 0$, $\int k^2 dF_0 = 1$).

El llamado Tercer Lema de Le Cam ([11]) implica en ese caso que la ley límite de $b_n(A)$ es la de $b^{(F_0)}(A) + \delta \int_A k(t) dF_0(t)$, es decir, que $b_n^{(U)}$ converge débilmente en D a $b + \delta \int \kappa(u) du$, con $\kappa(F_0(t)) = k(t)$.

Como consecuencia, de lo anterior, resulta inmediato describir el comportamiento asintótico bajo alternativas que cumplan las condiciones (i) y (ii), y en particular bajo la hipótesis nula, que corresponde a $\delta = 0$.

Antes de considerar esa aplicación, vamos a presentar algunos ejemplos de alternativas contiguas, y vamos a calcular el *apartamiento asintótico* δk o el *apartamiento asintótico normalizado* o *tipificado* $\delta \kappa$.

3.1 Algunos ejemplos de alternativas contiguas.

3.1.1 Alternativas de cambio de posición.

Si F_0 (con densidad f_0 respecto de cierta medida μ) es la distribución de (X_1, \dots, X_n) bajo \mathcal{H}_0 , tomaremos \mathcal{H}_n : “ $F = F_0(\cdot - \frac{\delta c}{\sqrt{n}}$ ” . Como consecuencia, la densidad asociada a \mathcal{H}_n es $f_n = f_0(\cdot - \frac{\delta c}{\sqrt{n}})$, y entonces, si existe el límite cuadrático de k_n , debe ser

$$k(x) = \lim_{n \rightarrow \infty} k_n(x) = \lim_{n \rightarrow \infty} \frac{2\sqrt{n}}{\delta} \left(\sqrt{\frac{f_0(x - \frac{\delta c}{\sqrt{n}})}{f_0(x)}} - 1 \right) = -c \frac{f'_0(x)}{f_0(x)}.$$

El coeficiente $c = \left(\int \frac{(f'_0(x))^2}{f_0(x)} d\mu(x) \right)^{-1/2}$ se elige para que $\|k\| = 1$.

La Tabla 1 indica los valores de c , $k(x) = \kappa(F_0(x))$ y $\kappa(u)$ correspondientes a algunas distribuciones. Excepto la última (distribución de Cauchy) que no tiene momentos finitos, las distribuciones se han tomado centradas en cero y con variancia uno, para su normalización. La Figura 1 permite comparar entre sí los apartamientos correspondientes a cada una de las distribuciones. En cada caso, puede verificarse que (i) y (ii) se cumplen.

3.1.2 Alternativas de cambio de dispersión.

Nuevamente suponemos que F_0 con densidad f_0 respecto de μ es la distribución asociada a \mathcal{H}_0 , y tomamos \mathcal{H}_n : “ $F = F_0((1 + \frac{\delta c}{\sqrt{n}})\cdot)$. La densidad asociada a \mathcal{H}_n es $f_n(x) = (1 + \frac{\delta c}{\sqrt{n}})f_0((1 + \frac{\delta c}{\sqrt{n}})x)$ y

$$k(x) = \lim_{n \rightarrow \infty} \frac{2\sqrt{n}}{\delta} \left(\sqrt{1 + \frac{\delta c}{\sqrt{n}}} \sqrt{\frac{f_0((1 + \frac{\delta c}{\sqrt{n}})x)}{f_0(x)}} - 1 \right) = c(1 + x \frac{f'_0(x)}{f_0(x)}),$$

con c tal que $c^2 \int (1 + x \frac{f'_0(x)}{f_0(x)})^2 f_0(x) d\mu(x) = 1$. Para las mismas distribuciones de la Tabla 1, los nuevos c , $k(x)$, $\kappa(x)$ se indican en la Tabla 2.

4 Determinación de la potencia de pruebas clásicas frente a alternativas contiguas.

Vamos a plantearnos el cálculo de la potencia asintótica de la prueba de ajuste de Kolmogorov-Smirnov, que corresponde a nuestro Ejemplo 2 de la sección 1. Los casos de los Ejemplos 1 y 3 son análogos.

Tabla 1: Apartamientos asintóticos para alternativas de cambio de posición.

$f_0(x)$	$f'_0(x)$	c	$k(x) = -c \frac{f'_0(x)}{f_0(x)}$	$\kappa(u) = k(F_0^{-1}(u))$
Normal: $\phi(x)$ $= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$-x\phi(x)$	1	x	$\Phi^{-1}(u)$
Doble exponencial: $f_{DE}(x) = \frac{e^{-\sqrt{2} x }}{\sqrt{2}}$	$-\text{sgn}(x)\sqrt{2}f_{DE}(x)$	$\frac{1}{\sqrt{2}}$	$\text{sgn}(x)$	$\text{sgn}(u - \frac{1}{2})$
Logística: $f_{LG}(x)$ $= \frac{\pi}{\sqrt{3}} \frac{e^{\pi x/\sqrt{3}}}{(1+e^{\pi x/\sqrt{3}})^2}$	$\frac{\pi^2 e^{\pi x/\sqrt{3}} (1-e^{\pi x/\sqrt{3}})}{3(1+e^{\pi x/\sqrt{3}})^3}$	$\frac{3}{\pi}$	$\sqrt{3} \frac{e^{\pi x/\sqrt{3}} - 1}{e^{\pi x/\sqrt{3}} + 1}$	$\sqrt{3}(2u - 1)$
Cauchy: $f_{CA}(x)$ $= \frac{1}{\pi(1+x^2)}$	$\frac{-2x}{\pi(1+x^2)^2}$	$\sqrt{2}$	$\frac{2\sqrt{2}x}{1+x^2}$	$\sqrt{2} \sin \pi(2u - 1)$

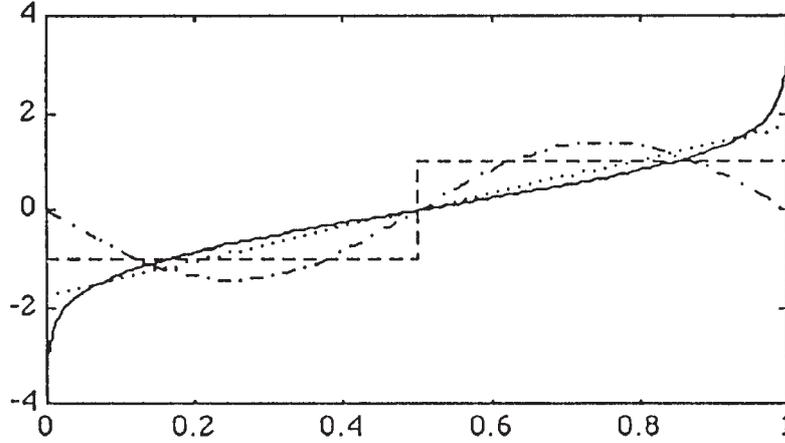


Figura 1: Apartamientos normalizados ($\kappa(u)$, $0 \leq u \leq 1$) para cambios en posición.

La distribución asintótica de $\sqrt{n}D_n = \sup_{x \in \mathbf{R}} |b_n(x)|$ fue obtenida por Kolmogorov ([10]):

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\sqrt{n}D_n \geq c\} = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 c^2}. \quad (1)$$

A partir de los resultados de §2, esa distribución es la de $\sup_{x \in \mathbf{R}} |b^{(F_0)}(x)|$, que coincide con la de $\sup_{0 \leq u \leq 1} |b(u)|$. El resultado de Kolmogorov, puede entonces encontrarse a partir de las propiedades del puente browniano.

Para que el nivel asintótico de la prueba sea α , se elige $c_n(\alpha)$ (ver el Ejemplo 2) igual a $c(\alpha)/\sqrt{n}$, con $c(\alpha)$ tal que $\mathbf{P}\{\sup_{0 \leq u \leq 1} |b(u)| > c(\alpha)\} = \alpha$, esto es, tal que reemplazado en lugar de c en (1) el resultado sea α .

De §3 resulta la potencia asintótica de la prueba cuando la muestra está distribuida con las leyes de la sucesión de alternativas contiguas de apartamento asintótico tipificado $\delta\kappa$:

Tabla 2: Apartamientos asintóticos para alternativas de cambio de dispersión.

Distribución	c	$k(x)$	$\kappa(x)$
Normal	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}(1-x^2)$	$\frac{1}{\sqrt{2}}(1-(\Phi^{-1}(u))^2)$
Doble exponencial	1	$1-\sqrt{2} x $	$1+\log(1- 2u-1)$
Logística	$\frac{3}{\sqrt{3+\pi^2}}$	$\frac{3}{\sqrt{3+\pi^2}}(1-\frac{\pi x}{\sqrt{3}}\frac{e^{\pi x/\sqrt{3}}-1}{e^{\pi x/\sqrt{3}}+1})$	$\frac{3}{\sqrt{3+\pi^2}}(1-(2u-1)\log\frac{u}{1-u})$
Cauchy	$\sqrt{2}$	$\sqrt{2}\frac{1-x^2}{1+x^2}$	$\sqrt{2}\cos\pi(2u-1)$

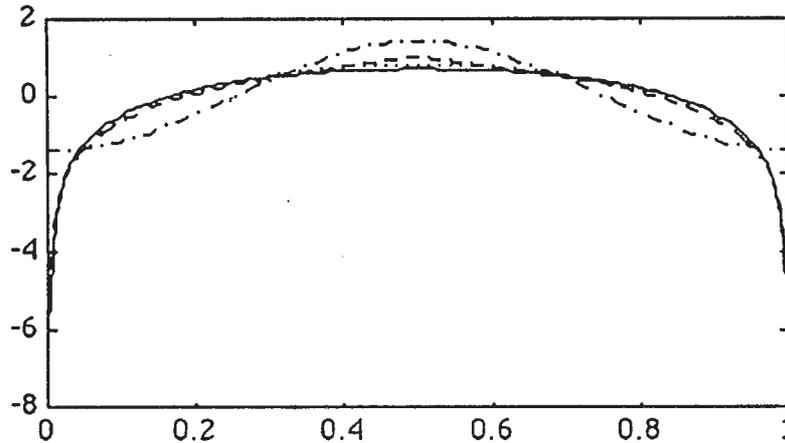


Figura 2: Apartamientos normalizados ($\kappa(u)$, $0 \leq u \leq 1$) para cambios en dispersión.

$$\Pi(\alpha, \delta\kappa) = \mathbf{P}\left\{ \sup_{0 \leq u \leq 1} |b(u) + \delta \int_0^u \kappa(s) ds| > c(\alpha) \right\}. \quad (2)$$

El cálculo exacto de $\Pi(\alpha, \delta\kappa)$ no es posible mediante una fórmula cerrada, pero para cada $\delta\kappa$ puede hacerse aproximadamente por simulación, por ejemplo.

Es obvio que la potencia será tanto mayor, cuanto mayor sea el tamaño δ del apartamiento asintótico y el sesgo $B(u) = \int_0^u \kappa(s) ds$, correspondiente al apartamiento tipificado y normalizado.

Las figuras 3 y 4 indican la forma de $B(u)$ para los apartamientos descritos en las figuras 1 y 2. En particular, puede verse en la figura 3 que el apartamiento integrado máximo se obtiene (en $1/2$) en el caso de los cambios de posición de la distribución doble exponencial.

Vamos a resolver el siguiente problema: De todas las funciones κ que satisfacen $\int \kappa = 0$, $\int \kappa^2 = 1$, ¿cuál es la que produce mayor sesgo $\int_0^u \kappa(s) ds$ en el punto u ?

Lo que buscamos es la función del $L^2([0, 1])$ ortogonal a la constante 1, y de norma 1, que forma el menor ángulo posible con la indicatriz de $[0, u]$. La solución es la proyección de esta indicatriz sobre el hiperplano ortogonal al 1, corregida para que la norma sea 1, esto es: $\frac{(1-u)\mathbf{1}_{(0,u)} - u\mathbf{1}_{(u,1)}}{\sqrt{u(1-u)}}$, y el valor resultante del sesgo en u es $\sqrt{u(1-u)}$, que alcanza

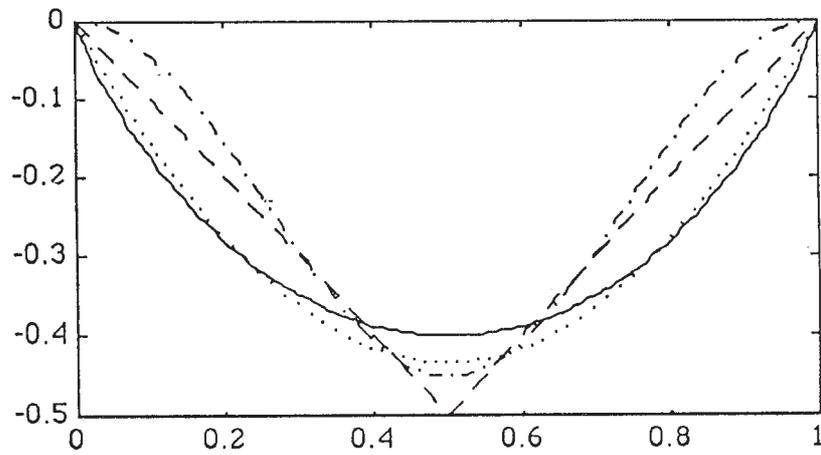


Figura 3: Apartamentos normalizados, tipificados e integrados para cambios en posición.

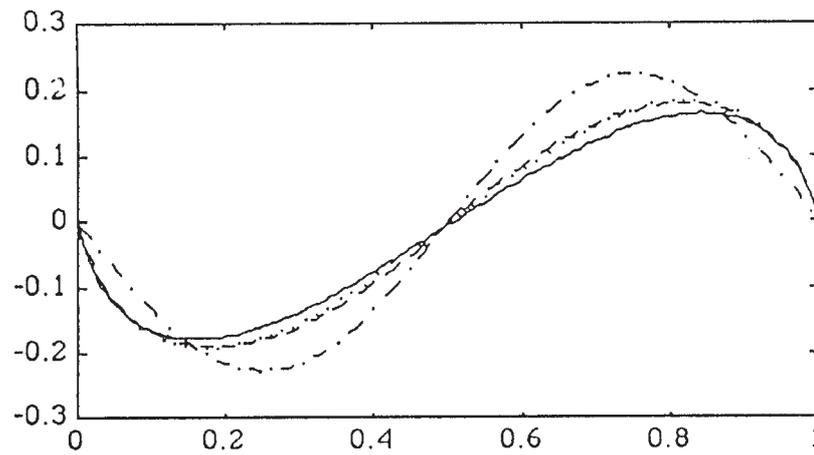


Figura 4: Apartamentos normalizados, tipificados e integrados para cambios en dispersión.

el máximo precisamente en $u = 1/2$, donde la variancia de b es máxima. Para este valor de u , la función κ óptima es $\kappa(s) = \text{sgn}(1/2 - s)$.

Argumentos como los utilizados en [1], [2] muestran, a partir de las observaciones precedentes, que la prueba de Kolmogorov-Smirnov es, para α y $\beta = 1 - \Pi$ pequeños, tan eficiente como la basada en $b_n(1/2)$, que, a su vez, es equivalente a la del cociente de verosimilitudes, cuando $\kappa(s) = \text{sgn}(1/2 - s)$. Este resultado fue obtenido directamente por Capon [5].

Part II

Procesos Empíricos Transformados

5 Motivación

Vamos a proponer pruebas de ajuste similares a las pruebas clásicas en las que el proceso empírico b_n será reemplazado por un nuevo proceso w_n , también definido a partir de la muestra, que diseñaremos para que la prueba tenga mejor eficiencia que la correspondiente prueba clásica en la detección de una familia dada de alternativas contiguas.

Notemos que el proceso empírico clásico asociado a la muestra $\{X\}$ de tamaño 1 con respecto a la distribución de probabilidades F_0 en \mathbf{R} , puede pensarse como la función de distribución de la medida con signo $b_x : A \mapsto b_x(A) = \mathbf{1}_{\{x \in A\}} - F_0(A)$, evaluada en el punto aleatorio $x = X$.

Cuando X tiene distribución F_0 , b_X satisface

$$Eb_X(A) = 0 \text{ para cada } A \text{ medible, y}$$

$$Eb_X(A)b_X(B) = F_0(A \cap B) - F_0(A)F_0(B) \text{ para cualesquiera } A, B \text{ medibles,}$$

y esto significa que b_X tiene los mismos momentos de primero y segundo orden que un puente browniano asociado a F_0 .

De allí resulta, a través de un teorema del límite central, que el proceso empírico $b_n = n^{-1/2} \sum_{i=1}^n b_{X_i}$ tiene bajo \mathcal{H}_0 la distribución asintótica de un puente browniano asociado a F_0 .

El sesgo asintótico de b_n bajo \mathcal{H}_n es

$$\lim \sqrt{n}(F_n(A) - F_0(A)) = \lim \sqrt{n} \int_A \left(\frac{f_n}{f_0} - 1 \right) dF_0 = \delta \int_A \kappa dF_0,$$

y del Tercer Lema de Le Cam resulta la distribución asintótica bajo las alternativas indicada en §3.

6 Construcción formal de los PETs en L^2

Nos planteamos ahora construir medidas w_x tales que

$w_X(A)$ es una variable aleatoria para cada intervalo A ,

$\mathbf{E}w_X(A) = 0$ para cada A , cuando $X \sim F_0$,

$\mathbf{E}w_X(A)w_X(B) = V(A \cap B)$ para cualesquiera A, B y alguna probabilidad V , cuando $X \sim F_0$.

Un teorema del límite central conducirá a concluir que $w_n = n^{-1/2} \sum_{i=1}^n w_{X_i}$ tiene la distribución asintótica de un proceso de Wiener asociado a la probabilidad V .

Como criterio de optimización, procuraremos finalmente que el sesgo asintótico

$$\lim \sqrt{n} \mathbf{E}w_X(A), \quad X \sim F_n$$

sea lo mayor posible en \mathbf{R} .

La familia de medidas aleatorias w_X depende de la elección de la probabilidad V . Supondremos que V tiene densidad a^2 respecto de F_0 , y que $a \neq 0$ F_0 c.s., de modo que V, F_0 son mutuamente absolutamente continuas.

Denotamos $\langle f, g \rangle = \int f g dF_0$ al producto interno en $L^2(\mathbf{R}, dF_0)$, y $\langle f, g \rangle_V = \int f g dV$ al producto interno en $L^2(\mathbf{R}, dV)$.

Si $\Psi = (\psi_0 = 1, \psi_1, \dots)$ es una base ortonormal de $L^2(\mathbf{R}, dF_0)$, $\Psi^V = (\psi_i/a)$ es una base ortonormal de $L^2(\mathbf{R}, dV)$.

Introducimos ahora el desarrollo formal de Fourier de w_x :

$$w_x(A) = \sum_i c_i(x) \int_A \psi_i^V(y) dV(y), \quad c_i(x) = \int \psi_i^V(y) dw_x(y).$$

Las condiciones sobre w_x imponen que se cumpla:

$$\langle c_i, 1 \rangle = \mathbf{E}c_i(X) = \mathbf{E} \int \psi_i^V(y) dw_X(y) = 0$$

y

$$\begin{aligned} \langle c_i, c_j \rangle &= \mathbf{E}c_i(X)c_j(X) = \mathbf{E} \int \psi_i^V(y) dw_X(y) \int \psi_j^V(z) dw_X(z) \\ &= \int \psi_i^V(y) \psi_j^V(y) dV(y) = \langle \psi_i^V, \psi_j^V \rangle_V = \delta_{i,j}, \end{aligned}$$

donde las integrales respecto de w_X son integrales estocásticas respecto de procesos de incrementos ortogonales.

Calculamos ahora

$$w_X(\mathbf{R}) = \sum_i \int_{\mathbf{R}} c_i(X) \psi_i^V(y) dV(y) = \sum_i c_i(X) \langle \psi_i^V, 1 \rangle_V = \sum_i c_i(X) \langle \psi_i, a \rangle.$$

De aquí resulta el sesgo asintótico correspondiente a $X \sim F_n$:

$$\begin{aligned} \lim \mathbf{E} \sqrt{n} w_X(\mathbf{R}) &= \lim \sqrt{n} \sum_i \int \left(\frac{f_n}{f_0} - 1 \right) c_i(x) dF_0(x) \langle \psi_i, a \rangle \\ &\rightarrow \sum_i \langle c_i, k \rangle \langle \psi_i, a \rangle = \langle \mathcal{T}^{-1} k, a \rangle = \langle h, 1 \rangle_V, \end{aligned}$$

donde \mathcal{T} es la isometría que lleva cada ψ_i en c_i y $h = \mathcal{T}^{-1} k/a$.

Dado que $\|h\|_V^2 = \|T^{-1}k\|^2 = \|k\|^2 = 1$, el sesgo resulta máximo cuando $h = 1$, es decir, cuando a se elige igual a

$$\hat{a} = \mathcal{T}^{-1} k.$$

De este desarrollo formal obtenemos

$$\begin{aligned} w_X(A) &= \sum_i c_i(X) \langle \psi_i^V, \mathbf{1}_A \rangle_V = \sum_i \mathcal{T} \psi_i(X) \langle \psi_i, a \mathbf{1}_A \rangle \\ &= \mathcal{T} \sum_i \psi_i(x) \langle \psi_i, a \mathbf{1}_A \rangle|_{x=X} = \mathcal{T}(a \mathbf{1}_A)(X) = \int \mathcal{T}(a \mathbf{1}_A) db_X \end{aligned}$$

y por lo tanto

$$w_n^{(a, \mathcal{T})}(A) = \int \mathcal{T}(a \mathbf{1}_A) db_n. \quad (3)$$

7 Propiedades asintóticas de los PETs.

La expresión (3) puede tomarse como punto de partida. Una vez definidos los PETs de esta manera, se establece directamente sin dificultad que los momentos de w_X son los indicados en la construcción, y es posible aplicar un teorema del límite central y luego el Tercer Lema de Le Cam para concluir que la distribución asintótica de $w_n^{(a, \mathcal{T})}(A)$ es la de

$$\int \mathcal{T}(a \mathbf{1}_A)(db^{(F_0)} + \delta k dF_0) = w^V(A) + \delta \int_A a \mathcal{T}^{-1} k dF_0 = w^V(A) + \delta \int_A h dV.$$

La existencia de un teorema del límite central que asegure esta convergencia, requiere imponer alguna condición sobre a, \mathcal{T} . La convergencia en distribución de $w_n^{(a, \mathcal{T})}(A)$ a $w^V(A)$ sobre la familia \mathcal{J} de las semirrectas de \mathbf{R} , está asegurada (por un Teorema del Límite Central de Ossiander, [14]) cuando la familia de integrandos $\mathcal{G} = \{\mathcal{T}(a \mathbf{1}_A) : A \in \mathcal{J}\}$ está uniformemente acotada por una función $G \in L^2(\mathbf{R}, dF_0)$ y se cumple

$$\int_0^1 \sqrt{\log N_{[]}^{(2)}(\varepsilon, \mathcal{G}, F_0)} d\varepsilon < \infty$$

con $N_{[]}^{(2)}(\varepsilon, \mathcal{G}, F_0)$ igual al mínimo ν para el cual existen conjuntos $\mathcal{L} = \{\ell_1, \dots, \ell_\nu\}$ y $\mathcal{U} = \{u_1, \dots, u_\nu\}$ de funciones de cuadrado integrable tales que, para cada f en \mathcal{G} se encuentran $\ell \in \mathcal{L}$ y $u \in \mathcal{U}$ de modo que $\ell < f < u$ y $\|u - \ell\|^2 \leq \varepsilon^2$.

Part III

Pruebas basadas en PETs.

8 Prueba de Kolmogorov-Smirnov modificada.

Vamos a describir pruebas alternativas a la del Ejemplo 2, siguiendo a [1]. En los trabajos [7] y [4] se analizan las pruebas de ajuste χ^2 y de Cramér-von Mises respectivamente.

Proponemos probar \mathcal{H}_0 mediante la región crítica

$$\sup_{A \in \mathcal{J}} |w_n^{(\hat{a}, T)}(A)| > c(\alpha)$$

con $\hat{a} = T^{-1}k$, $\mathcal{J} = \{(-\infty, x] : x \in \mathbf{R}\}$, y

$$\mathbf{P}\{\sup_{A \in \mathcal{J}} |w^V(A)| > c(\alpha)\} = \alpha. \quad (4)$$

El nivel asintótico es α , de acuerdo a (4), y la potencia asintótica para \mathcal{H}_n es

$$\begin{aligned} & \mathbf{P}\{\sup_x |w^V(x) + \delta V(x)| > c(\alpha)\} \\ &= \mathbf{P}\{\sup_u |w(u) + \delta u| > c(\alpha)\}. \end{aligned}$$

Esto significa en particular que una vez que la función a se ha elegido igual a $T^{-1}k$ para optimizar el sesgo, la potencia asintótica no depende más que de la magnitud δ del apartamiento, pero no de la distribución F_0 de la hipótesis nula, ni de la forma k del apartamiento de las alternativas.

9 Construcción de isometrías.

Dos sencillas observaciones nos permiten construir las isometrías \mathcal{T} de $L^2(\mathbf{R}, dF_0)$ que necesitamos para la construcción de los PETs:

- i₁ Si $\Psi = (\psi_0 = 1, \psi_1, \psi_2, \dots)$ es una base ortonormal de $L^2(\mathbf{R}, dF_0)$, entonces $\mathcal{T} : \psi_i \mapsto \psi_{i+1}$ es una isometría cuyo recorrido es el complemento ortogonal de la constante 1.
- i₂ Si \mathcal{T}_F es una isometría en $L^2(\mathbf{R}, dF)$, entonces

$$(\mathcal{T}_{F_0}g)(x) = (\mathcal{T}_F(g \circ F_0^{-1} \circ F))(F^{-1}(F_0(x)))$$

es una isometría en $L^2(\mathbf{R}, dF_0)$.

La isometría más sencilla que conocemos ha sido usada por Karatzas en la construcción del puente browniano a partir del proceso de Wiener en ([8]), por Efron y Johnstone ([6]) y por Ritov y Wellner ([15]) en conexión con tasas de azar, e introducida en la inferencia estadística por E. V. Khmaladze en [9] con especial énfasis en sus propiedades de martingala que nuestro enfoque no utiliza. Se obtiene a partir del sistema ortonormal de los Polinomios de Laguerre, base de $L^2(\mathbf{R}, e^{-x}dx)$. La transformación

$$\mathcal{T}_L g(x) = g(x) - \int_0^x g(t)dt,$$

que lleva cada polinomio de Laguerre en el siguiente (excepto por el signo) nos conduce, a través de (i₂), a

$$(\mathcal{T}_L g)(x) = g(x) - \int_{-\infty}^x \frac{g(t)dF_0(t)}{1 - F_0(t)} \quad (5)$$

cuya inversa es

$$(\mathcal{T}_L^{-1}h)(x) = h(x) + \frac{1}{1 - F_0(x)} \int_{-\infty}^x h(t)dF_0(t).$$

Para esta isometría, la hipótesis adicional

$$|a|/(1 - F_0)^\rho \in L^2(\mathbf{R}, dF_0) \text{ para algún } \rho > 0$$

basta para asegurar la validez del Teorema del Límite Central que describe la distribución asintótica de los PETs (ver §7).

Otros ejemplos y el detalle de lo que precede pueden verse en [3]. También allí se extienden las isometrías, los PETs y las pruebas de ajuste a un contexto multivariante.

10 Construcción de los estadísticos para la Prueba de Kolmogorov-Smirnov Modificada, con la Isometría de Laguerre.

A partir de la distribución F_0 asociada a \mathcal{H}_0 y del apartamiento normalizado k de las alternativas de interés, se construye $\hat{a} = \mathcal{T}^{-1}k$, para lo que se necesita elegir una isometría.

Con la isometría de Laguerre (5), tenemos

$$\hat{a}(x) = k(x) + \frac{1}{1 - F_0(x)} \int_{-\infty}^x k(t)dF_0(t),$$

y el cambio de variables $a^*(u) = \hat{a}(F_0^{-1}(u))$, nos lleva a

$$a^*(u) = \kappa(u) + \frac{1}{1 - u} \int_0^u \kappa(s)ds.$$

Una vez obtenido \hat{a} , calculamos $w_X((-\infty, x]) = w_X(x)$

$$= \mathcal{T}(\hat{a}\mathbf{1}_{(-\infty, x]})(X) = \hat{a}(X)\mathbf{1}_{\{X \leq x\}} - \int_{-\infty}^x \hat{a}(t)\mathbf{1}_{\{t \leq x\}} \frac{dF_0(t)}{1 - F_0(t)}$$

$$= \hat{a}(X)\mathbf{1}_{\{X \leq x\}} - \hat{A}(X \wedge x), \text{ con } \hat{A}(x) = \int_{-\infty}^x \frac{\hat{a}(t)dF_0(t)}{1 - F_0(t)}.$$

Esta función tiene un salto en X que vale $w_X(X^+) - w_X(X^-) = \hat{a}(X)$, es constante a la derecha de X (vale $\hat{a}(X) - \hat{A}(X)$) y en $(-\infty, X)$ se reduce a $-\hat{A}(x)$.

En el caso particular en que \hat{a} es de signo constante, por ejemplo, negativo, w_n crece en cada intervalo $(X_{(i-1)}, X_{(i)})$ y salta hacia abajo en cada $X_{(i)}$, por lo que el estadístico

$$\sup_x |w_n(x)| = \max_i \{w_n(X_{(i)}^-) \vee (-w_n(X_{(i)}^+))\}$$

se calcula sin dificultad.

Cuando \hat{a} es de signo variable, se requiere mayor cuidado para determinar el estadístico. Por ejemplo, si \hat{a} cambia de signo en los puntos y_1, y_2, \dots, y_ν , entonces

$$\sup_x |w_n(x)| = \max_i \{|w_n(X_{(i)}^-)| \vee |w_n(X_{(i)}^+)|\} \vee \max_j \{|w_n(y_j)|\}.$$

Las tablas siguientes indican las funciones \hat{a} para los ejemplos de las Tablas 1 y 2:

Tabla 3: Parámetro \hat{a} óptimo para alternativas de cambio de posición.

Distribución	$a^*(u)$	$\hat{a}(x)$
Normal	$\Phi^{-1}(u) - \frac{\varphi(\Phi^{-1}(u))}{1-u}$	$x - \frac{\varphi(x)}{1-\Phi(x)}$
Doble exponencial	$\frac{-1}{1-u} \mathbf{1}_{\{u < 1/2\}}$	$\frac{-2}{2 - e^{\sqrt{2}x}} \mathbf{1}_{\{x < 0\}}$
Logística	$\sqrt{3}(u-1)$	$\frac{-\sqrt{3}}{e^{\pi x/\sqrt{3}} + 1}$
Cauchy	$\sqrt{2} \left(\sin(2u-1)\pi - \frac{1 + \cos(2u-1)\pi}{2\pi(1-u)} \right)$	$\frac{2}{1+j^2} \left(x - \frac{1}{\pi - 2\arctan x} \right)$

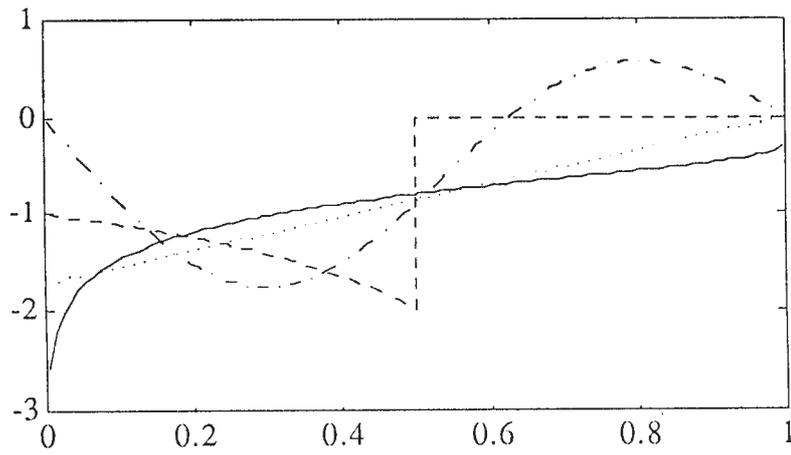


Figura 5: Funciones de "score" tipificadas para cambios en posición.

Tabla 4: Parámetro \hat{a} óptimo para alternativas de cambio de dispersión.

Distribución	$a^*(u)$	$\hat{a}(x)$
Normal	$\frac{1}{\sqrt{2}} \left(1 - (\Phi^{-1}(u))^2 + \frac{\Phi^{-1}(u)\varphi(\Phi^{-1}(u))}{1-u} \right)$	$\frac{1}{\sqrt{2}} \left(1 - x^2 + \frac{x\varphi(x)}{1-\Phi(x)} \right)$
Doble exponencial	$\mathbf{1}_{\{1/2 < u\}} + \left(1 + \frac{\log 2u}{1-u} \right) \mathbf{1}_{\{u \leq 1/2\}}$	$\mathbf{1}_{\{0 < x\}} + \left(1 + \frac{\sqrt{2}x}{1 - e^{\sqrt{2}x/2}} \right) \mathbf{1}_{\{x \leq 0\}}$
Logística	$\frac{3}{\sqrt{3 + \pi^2}} \left(u + (1-u) \log \frac{u}{1-u} \right)$	$\frac{3}{\sqrt{3 + \pi^2}} \left(1 + \frac{\pi x}{\sqrt{3}(1 + e^{\pi x/\sqrt{3}})} \right)$
Cauchy	$\sqrt{2} \cos(2u - 1)\pi + \frac{\sin(2u - 1)\pi}{\sqrt{2}(1-u)\pi}$	$\frac{\sqrt{2}}{1+x^2} \left(1 - x^2 + \frac{x}{\pi/2 - \arctan x} \right)$

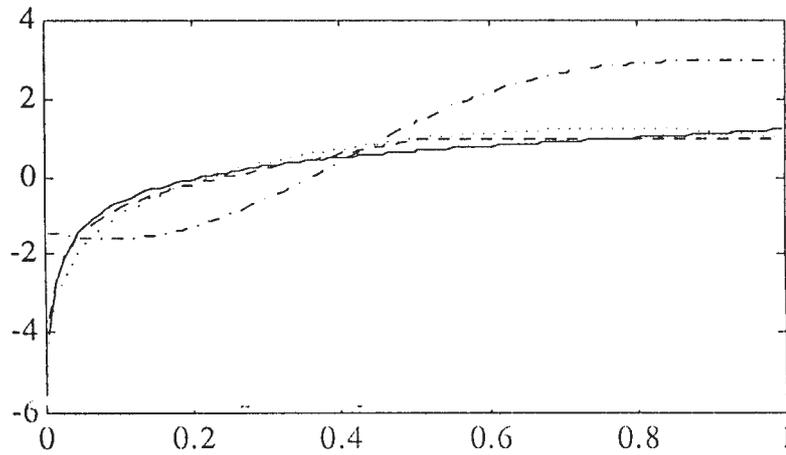


Figura 6: Funciones de “score” tipificadas para cambios en dispersión.

Bibliografía

- [1] Cabaña, A., *Transformations of the empirical process and Kolmogorov-Smirnov tests*. Ann. Statist. Vol. 24 No. 5 (1996), 2020-2035.
- [2] Cabaña, A. and Cabaña, E.M., *Bridge-to-bridge transformations and Kolmogorov-Smirnov Tests*, Comm. in Stat. Theory and Methods **25** (1996) 227-234.
- [3] Cabaña, A. and Cabaña, E.M., *Transformed Empirical Processes and Modified Kolmogorov-Smirnov Tests for multivariate distributions*, Mathematics Preprint Series **208**, June 1996, Universitat de Barcelona, por aparecer en Ann. Statist.
- [4] Cabaña, A. and Cabaña, E.M., *Pruebas de Anderson-Darling modificadas con mejora selectiva de la potencia*, comunicación presentada al IV Congreso "Dr. Antonio A. R. Monteiro", U. Nacional del Sur, Bahía Blanca, Argentina, Abril 1997.
- [5] Capon, J., *On the asymptotic efficiency of the Kolmogorov - Smirnov test*. J. Amer. Statist. Assoc. **60** (1965), pp. 843-853.
- [6] Efron, B. and Johnston, I. *Fisher's information in terms of hazard rate*. Ann. Statist. **18** (1990), 38-62
- [7] Graneri, J.R., *Una prueba de ajuste χ^2 basada en Procesos Empíricos Transformados*, comunicación presentada al IV Congreso "Dr. Antonio A. R. Monteiro", Universidad Nacional del Sur, Bahía Blanca, Argentina, Abril 1997.
- [8] Karatzas, I. and Shreve, S. *Brownian Motion and Stochastic Calculus*. (2nd Edition), Springer-Verlag, New York, 1988-1991.
- [9] Khmaladze, E. V., *Martingale approach in the theory of Goodness-of-fit*. Theory of Probability and its applications, Vol XXVI, No. 2 (1981), 240-257.
- [10] Kolmogorov, A. *Sulla determinazione empirica di una legge di distribuzione*, Giorn. Ist. Ital Attuari, **4** (1933) 83-91.
- [11] Le Cam L., Yang, G.L. *Asymptotics in Statistics. Some basic concepts*. Springer-Verlag, New York, 1990.
- [12] Shorack, G.R.; Wellner, J.A., *Empirical processes with applications to statistics*., John Wiley & Sons , New York, 1986.
- [13] Oosterhoff, J; van Zwet, W.R. *A note on contiguity and Hellinger distance*. Contributions to Statistics, 157-166, Reidel, Dordrecht, 1979.
- [14] Ossiander, M., *A central limit theorem under metric entropy with L_2 bracketing*. Ann. Probab. **15**, (1987) 897-919.
- [15] Ritov, J. and Wellner, J.A. *Censoring, Martingales and the Cox Model*. Contemp. Math. **80** (1988), 191-220.