

An isotonicity test for the regression function

María Jesús López Palomo y José Santos Domínguez Menchero

Resumen This paper deals with the problem of finding a (completely) non-parametric test for the increasing character of the regression function. In this way we propose an statistic based on the L_2 -distance from a non-smooth estimate of the regression to the set of all increasing functions. We show that the test is asymptotically of the prefixed level and that its power under normality is the same than one of the usual chi-square tests.

Un contraste de isotonía para la función de regresión

Resumen. En este trabajo se pretende desarrollar por primera vez un test completamente no paramétrico para contrastar el crecimiento de la función de regresión. Para ello se propone un estadístico basado en la L_2 -distancia de un estimador no suavizado de la regresión al conjunto de todas las funciones crecientes. Se prueba que el test es asintóticamente del nivel prefijado y que su potencia bajo normalidad es la misma que las de los usuales contrastes chi-cuadrado.

1.. Introducción

En muchas situaciones se supone a priori que la función de regresión cumple ciertas propiedades que debería seguir verificando un estimador cualquiera de ella. Algunas de estas propiedades, como la continuidad o la suavidad, se pueden obtener de forma automática con los estimadores habituales (por ejemplo tipo kernel). Otros casos, como el número de extremos de la función o la monotonía de la misma, plantean un problema más complejo. Esta última propiedad ha recibido especial atención: en muchos contextos de tipo económico (ingresos-gastos), médico (evolución de una enfermedad según el suministro de ciertos medicamentos), etc. suele ser evidente dicho carácter creciente. La literatura ha abordado ampliamente la estimación de la regresión bajo la restricción de crecimiento o isotonía (conocida como regresión isotónica); ver Barlow et al. [1], Brunk [3], Hanson et al. [10], Wright [17, 18], Robertson et al. [16], Mukerjee [13], Mammen [12], Cuesta et al. [4, 5], Domínguez y López [7].

El problema de contrastar la isotonía de la función de regresión está menos estudiado en la literatura. En Robertson y Wegman [15] y Robertson et al. [16] se desarrollan procedimientos paramétricos que restringen las observaciones al caso de normalidad: el $\bar{\chi}_{12}^2$ -test y el \bar{E}_{12}^2 -test. También se trata el caso más general de familias de distribución exponenciales. Algunos trabajos recientes consideran este problema de contrastar la isotonía de la función de regresión en un contexto no paramétrico: Bowman et al. [2], Hall y Heckman [9], Ghosal et al. [8], pero presentan diversos inconvenientes: dificultad de cálculo del estadístico o utilización de regresiones lineales que plantean problemas con los “outliers”.

Presentado por Francisco Javier Girón.

Recibido: 7 de Agosto de 2002. Aceptado: 9 de Octubre de 2002.

Palabras clave / Keywords: Regresión isotónica, regresión monótona, Dip.

Mathematics Subject Classifications: 62J99, 62F30

© 2003 Real Academia de Ciencias, España.

En este artículo se presenta un test completamente no paramétrico de la isotonía de la función de regresión, basado en un estadístico de fácil cálculo. Si m es la función de regresión y \hat{m} es un estimador no suavizado de la misma, el contraste está basado en una medida empírica de la isotonía de m (llamada *Dip*) a través de una L_2 -distancia entre \hat{m} y el conjunto de las funciones isotónicas. Bajo ciertas condiciones, el *Dip* y el estadístico $\bar{\chi}_{12}^2$ del contexto paramétrico están fuertemente relacionados. Los resultados conocidos sobre el *Dip* permiten, en un principio, un modelo de diseño de las observaciones completamente aleatorio y el uso de estimadores suaves de la regresión tipo kernel. Sin embargo, por simplicidad, nosotros consideramos en este trabajo un modelo de diseño fijo y cuando se estima la regresión no se utilizarán técnicas de suavizado. Esto permitirá además disponer de un marco común para comparar los resultados obtenidos con los ya existentes en el contexto paramétrico.

Este artículo está organizado de la siguiente forma. En la Sección 2 se introduce el estadístico *Dip* en el cual se basa el contraste completamente no paramétrico desarrollado. Asimismo, se describe el modelo elegido para abordar el contraste de isotonía. En la Proposición 1 se formaliza la idea intuitiva de que las distribuciones más desfavorables, bajo la hipótesis nula de crecimiento de m , son las que tienen regresión constante. En la Sección 3 se estudia el comportamiento asintótico del estimador isotonizado \hat{m}_I bajo regresión constante. Los resultados dados en la Proposición 2 permiten obtener la distribución asintótica bajo regresión constante de cierta modificación del *Dip*, el $s - Dip$. En la Sección 4 se muestra cómo el comportamiento de adecuados $s - Dip$ puede ofrecer información sobre el del *Dip* (Proposición 3). El Teorema 1 obtiene cotas asintóticas para la distribución del *Dip* bajo regresión constante que permiten desarrollar un test para el caso de varianzas conocidas, el *Dip*-test, que asintóticamente es al menos conservador. La Sección 5 es un estudio de simulación. Se obtiene que el *Dip*-test sí es asintóticamente del nivel prefijado, pero en la práctica muy conservador. Esto no parece afectar significativamente a la potencia del test que, en el caso de normalidad y varianzas iguales, es competitiva respecto a la del $\bar{\chi}_{12}^2$ -test aun con pocos datos.

2.. Planteamiento.

Se considera un intervalo $[a,b]$ y una familia de distribuciones de probabilidad en la recta real $\{D(x)\}_{x \in [a,b]}$, con medias $\{m(x)\}_{x \in [a,b]}$ y desviaciones típicas finitas $\{\sigma(x)\}_{x \in [a,b]}$. En este contexto, la función de regresión es entonces la función $m : [a, b] \rightarrow \mathbb{R}$. En el intervalo $[a,b]$ se tomarán n puntos de diseño, $a < x_1 < \dots < x_n = b$ ($n \in \mathbb{N}$) igualmente espaciados $x_{ni} := a + \frac{b-a}{n}i$, $1 \leq i \leq n$. La familia de puntos de diseño $\{x_{ni} : 1 \leq i \leq n\}$ se denotará por \mathcal{X}_n . En cada uno de los n puntos de de diseño, x_{ni} , se dispondrá de r observaciones independientes $Y_{ni}^1, \dots, Y_{ni}^r$ definidas en un mismo espacio probabilístico (Ω, \mathcal{F}, P) de forma que la distribución de Y_{ni}^j es $D(x_{ni})$ $i = 1, \dots, n, j = 1, \dots, r$.

El problema a resolver se puede expresar en los siguientes términos.

Problema P. Dado $n \in \mathbb{N}$, se quiere contrastar la hipótesis nula $H_0 : m$ es creciente, frente a la hipótesis alternativa $H_1 : m$ no es creciente, a partir de n puntos de diseño y r observaciones en cada uno de ellos, $\{Y_{ni}^j\}_{\substack{i=1, \dots, n \\ j=1, \dots, r}}$. ■

Dados n puntos de diseño y r observaciones en cada uno de ellos, un estimador natural de la regresión m es la función $\hat{m} \equiv \hat{m}_{nr} : \mathcal{X}_n \rightarrow \mathbb{R}$ tal que $\hat{m}(x_{ni}) = \sum_{j=1}^r Y_{ni}^j / r$. La isotonización de \hat{m}_{nr} , $\hat{m}_I \equiv \hat{m}_{nrI}$, se define como la mejor $L_2(\mu_n)$ -aproximación creciente a \hat{m}_{nr} : $\|\hat{m}_{nr} - \hat{m}_{nrI}\|_2 = \inf_{g \in \mathcal{H}(n)} \|\hat{m}_{nr} - g\|_2$ siendo $\mathcal{H}(n) := \{g : \mathcal{X}_n \rightarrow \mathbb{R}, \text{ tal que } g \text{ es creciente}\}$ y $L_2(\mu_n)$ el conjunto de todas las funciones $f : \mathcal{X}_n \rightarrow \mathbb{R}$ con la norma dada por la probabilidad empírica, μ_n , que asigna masa o peso n^{-1} a cada punto de diseño de \mathcal{X}_n , $\|f\|_2 := \left[\sum_{i=1}^n \frac{f^2(x_{ni})}{n} \right]^{1/2}$.

En Barlow et al. [1] se demuestra que \hat{m}_{nrI} existe y es única. Además, $\hat{m}_{nrI}(x_{ni})$ es la pendiente por la izquierda en i de la mayor minorante convexa a la función construida por interpolación lineal a partir de los puntos $(0, 0) \cup \left\{ \left(i, \sum_{j=1}^i \hat{m}_{nr}(x_{n,j}) \right) \right\}_{i=1}^n$. Además de este procedimiento de cálculo (que será muy útil

como herramienta técnica en los desarrollos teóricos), se dispone de algunos algoritmos como el método PAVA y las fórmulas Max-min (Barlow et al.[1]).

Como es habitual, extenderemos \hat{m} y \hat{m}_I a todo el intervalo $[a,b]$ como funciones continuas por la izquierda y constantes en $[a,x_{n1}]$ y entre puntos de diseño.

Contexto paramétrico.

Existen diversos resultados relativos a la resolución del Problema P cuando la familia de distribuciones de probabilidad está sujeta a importantes restricciones paramétricas. Así, los resultados fundamentales exigen que las distribuciones sean Normales, $D(x) \equiv N(m(x), \sigma(x))$, exigiendo la mayoría de ellos el conocimiento completo de las varianzas (puede encontrarse un resumen en Robertson et al. [16]). En este caso de normalidad, se sabe que el test de la razón de verosimilitud rechazaría H_0 para valores grandes del estadístico

$$\bar{\chi}_{12}^2 := \sum_{i=1}^n (\hat{m}(x_{ni}) - \hat{m}_p(x_{ni}))^2 \frac{r}{\sigma^2(x_{ni})}$$

donde \hat{m}_p es la isotonización de \hat{m} , pero sustituyendo μ_n por la medida que al punto x_{ni} le asigna peso $1/\sigma^2(x_{ni})$. El problema se resuelve entonces encontrando, para un nivel de significación $\alpha \in [0, 1]$, un valor $H \in \mathbb{R}$ tal que

$$\sup_{\Theta_0} P(\bar{\chi}_{12}^2 > H) = \alpha \tag{1}$$

donde Θ_0 es el conjunto de familias de distribución de probabilidad en la recta real con desviaciones típicas $\{\sigma(x)\}_{x \in [a,b]}$ y función de regresión creciente. Así, se rechazaría el crecimiento de m cuando \hat{m} está suficientemente alejado de su isotonización. Existe también algún desarrollo cuando la varianza es desconocida (Robertson et al. [16]), aunque se exige cierta información parcial como por ejemplo que las varianzas sean iguales (en este caso se trabaja con el estadístico \bar{E}_{12}^2 , de idea similar al $\bar{\chi}_{12}^2$).

Al resolver el test (1) a un nivel α se puede utilizar un resultado (ver Robertson et al. [16]) que indica que las distribuciones más desfavorables bajo H_0 son aquellas con regresión constante. Entonces, para desarrollar el test bastaría encontrar la distribución de $\bar{\chi}_{12}^2$ bajo regresión constante. En Robertson et al. [16] se deriva la distribución exacta de $\bar{\chi}_{12}^2$ bajo esa restricción, resultando ser independiente de r . Para $n > 24$, esta distribución se puede aproximar por otra más sencilla basada en ciertos cumulantes. No obstante, tanto la distribución exacta como la aproximación, son muy complejas.

Enfoque no paramétrico.

En este trabajo nosotros pretendemos avanzar en la obtención de un contraste de regresión isotónica, desarrollando un test más sencillo que el basado en $\bar{\chi}_{12}^2$ y que a la vez pueda ser aplicado en un mayor número de casos, eliminando la normalidad; buscamos un test completamente no paramétrico. Dicho contraste está basado en una medida empírica de m a través de la $L_2(\mu_n)$ - distancia entre el estimador \hat{m} y el conjunto de las funciones crecientes sobre los puntos de diseño. Esta medida recibe el nombre de *Dip* y fue propuesta por primera vez en Domínguez [6], en un modelo de diseño completamente aleatorio. Su consistencia ha sido analizada en Cuesta et al. [4, 5] y Domínguez y López [7] en el modelo de diseño completamente aleatorio. Su definición es la siguiente,

$$Dip \equiv Dip_{nr} := \inf_{g \in \mathcal{H}(n)} \|\hat{m} - g\|_2,$$

de forma que el *Dip* es justamente la $L_2(\mu_n)$ -distancia entre \hat{m} y \hat{m}_I :

$$Dip = \|\hat{m} - \hat{m}_I\|_2 = \left[\sum_{i=1}^n \frac{(\hat{m}(x_{ni}) - \hat{m}_I(x_{ni}))^2}{n} \right]^{1/2}.$$

El Problema P se resolvería entonces, si dado $\alpha \in [0, 1]$, se encuentra una constante $H \in \mathbb{R}$ tal que

$$\sup_{\Theta_0} P(Dip > H) = \alpha. \tag{2}$$

En este contexto no paramétrico el papel que desempeñan las regresiones constantes es el mismo que en el caso paramétrico. El hecho de que, bajo H_0 , las distribuciones más desfavorables sean las que tienen regresión constante es bastante intuitivo: entre todas las posibles situaciones de crecimiento de m , aquellas más difíciles para detectar la isotonía a partir de la nube de puntos son justamente las de regresión constante. En la siguiente proposición se establece el resultado (en Robertson et al. [16] se puede encontrar el resultado correspondiente en el contexto paramétrico). Se puede probar que el papel de las regresiones constantes se mantiene utilizando un modelo de diseño completamente aleatorio, un estimador de la regresión tipo kernel y L_p -distancias en el Dip (ver López-Palomo [11]).

Proposición 1 *Sea m creciente. Entonces, existe una familia $\{D^*(x)\}_{x \in [a,b]}$ de distribuciones de probabilidad y una familia de variables aleatorias $\{Y_{ni}^{*j}\}_{\substack{i=1,\dots,n, \\ j \in \mathbb{N}}}$, verificando que*

- (a) *La función de regresión de D^* , m^* , es constante.*
- (b) *La familia de desviaciones típicas de D^* , $\{\sigma^*(x)\}_{x \in [a,b]}$, coincide con la familia $\{\sigma(x)\}_{x \in [a,b]}$.*
- (c) *Para cada n , $\{Y_{ni}^{*j}\}_{\substack{i=1,\dots,n, \\ j \in \mathbb{N}}}$ son variables aleatorias independientes y la distribución de Y_{ni}^{*j} es $D^*(x_{ni})$ para $i = 1, \dots, n, j \in \mathbb{N}$.*
- (d) *Si \hat{m}^* es la función definida en \mathcal{X}_n ($n \in \mathbb{N}$) por $\hat{m}^*(x_{ni}) := \sum_{j=1}^r \frac{Y_{ni}^{*j}}{r}$ para $i = 1, \dots, n$, \hat{m}_I^* es su isotonización y $Dip^* := \|\hat{m}^* - \hat{m}_I^*\|_2$, entonces se cumple que $Dip \leq Dip^*$.*

DEMOSTRACIÓN. Para cada $x \in [a, b]$, sea Y_x una variable aleatoria (no necesariamente definida en el espacio (Ω, \mathcal{F}, P)) con distribución $D(x)$, y sea $D^*(x)$ la distribución de $Y_x - m(x)$. Igualmente, para cada $n, j \in \mathbb{N}, i = 1, \dots, n$, se define $Y_{ni}^{*j} := Y_{ni}^j - m(x_{ni})$.

Obviamente, se verifican (a) (con $m^*=0$), (b) y (c). Para probar (d), nótese que $\hat{m}^* = \hat{m} - m$ ya que

$$\hat{m}^*(x_{ni}) = \sum_{j=1}^r \frac{Y_{ni}^{*j}}{r} = \sum_{j=1}^r \frac{Y_{ni}^j}{r} - \sum_{j=1}^r \frac{m(x_{ni})}{r} = \hat{m}(x_{ni}) - m(x_{ni})$$

y entonces, $Dip^* = \|\hat{m}^* - \hat{m}_I^*\|_2 = \|\hat{m} - (\hat{m}_I^* + m)\|_2$.

Por otro lado, como $\hat{m}_I^* + m$ es creciente por hipótesis, se tiene que

$$Dip = \inf_{g \in \mathcal{H}(n)} \|\hat{m} - g\|_2 \leq \|\hat{m} - (\hat{m}_I^* + m)\|_2 = Dip^*. \quad \blacksquare$$

En vista del resultado anterior nos planteamos si es razonable buscar la distribución asintótica (no paramétrica) del Dip bajo regresión constante. Para ello, observamos que si la varianza es constante, $\sigma(x) = \sigma^* \forall x \in [a, b]$, los estadísticos Dip y $\bar{\chi}_{12}^2$ están relacionados de la forma: $nrDip^2 = \sigma^{*2}\bar{\chi}_{12}^2$. Esta relación impide resolver el Problema P encontrando la distribución asintótica del Dip bajo regresión constante, puesto que ello implicaría deducir la de $\bar{\chi}_{12}^2$ en el caso de varianzas constantes, que es desconocida incluso añadiendo la restricción de normalidad. Por tanto, se busca un planteamiento distinto a encontrar la distribución asintótica del Dip bajo regresión constante.

3.. Estudio asintótico de \hat{m}_I bajo regresión constante

En primer lugar estudiamos el comportamiento asintótico puntual de \hat{m}_I bajo regresión constante, encontrando que la distribución convenientemente normalizada de $\hat{m}_I - m$ en un punto $x_0 \in (a, b]$, es la pendiente por la izquierda de la mayor minorante convexa en x_0 de un Movimiento Browniano (Proposición 2 (i)). Aunque esta distribución asintótica tiene interés independiente a la resolución del Problema P, desde el punto de vista del Dip sería más interesante disponer de la distribución de $\hat{m}_I - \hat{m}$. En la Proposición

2(ii) se obtiene que el comportamiento asintótico conjunto de $\widehat{m}_I - \widehat{m}$ en s puntos distintos y fijos del intervalo es el de una Normal multidimensional.

En adelante, $\{S^p(x)\}_{x \in [a,b]}$ representará la familia de momentos absolutos centrados de orden $p > 0$ correspondiente a la familia $\{D(x)\}_{x \in [a,b]}$ de distribuciones de probabilidad. Así, para $i = 1, \dots, n, j = 1, \dots, r$, $S^p(x_{ni}) := E|Y_{ni}^j - m(x_{ni})|^p$. Asimismo, denotamos por $\text{slogcom}_{x=x_0} [f(t)]$ (donde “slogcom” proviene de “slope of the greatest convex minorant”) la pendiente por la izquierda en un punto $t_0 \in A$ de la mayor minorante convexa de la función $f : A \rightarrow \mathbb{R}$, siendo A un intervalo real. La convergencia en ley o en distribución, vendrá indicada por el símbolo \mathcal{L} .

Proposición 2 Sea m constante, $\sigma > 0$ continua y $\sup_{x \in [a,b]} S^{2+\delta}(x) < \infty$ para algún $\delta > 0$.

(i) Si r es una función cualquiera de n , se cumple $\forall x_0 \in (a, b)$,

$$\left(\frac{nr}{(b-a)\theta} \right)^{1/2} (\widehat{m}_I(x_0) - m(x_0)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \text{slogcom}_{x=x_0} [W(U(x))],$$

donde W es un Movimiento Browniano en $[0, 1]$ y

$$\theta := \int_a^b \sigma^2(u) du > 0, \quad U(x) := \frac{1}{\theta} \int_a^x \sigma^2(u) du, \quad (x \in [a, b]).$$

(ii) Si r es una función de n tal que $r \rightarrow \infty$ cuando $n \rightarrow \infty$, se cumple para cualesquiera x_1, \dots, x_s , s puntos distintos en (a, b) ,

$$r^{1/2} [\widehat{m}_I(x_j) - \widehat{m}(x_j)]_{j=1}^s \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, [\sigma^2(x_j)]_{j=1}^s \cdot I)$$

donde I es la matriz identidad.

DEMOSTRACIÓN. (i) Considérense $Z_{ni} := r^{1/2}(\widehat{m}(x_{ni}) - m(x_{ni}))$ ($i = 1, \dots, n, n \in \mathbb{N}$). Para cada $y \in [0, 1]$ sean $G(y) := a + (b-a)y$ y W_1, \dots, W_r variables aleatorias independientes definidas en un mismo espacio probabilístico con distribución $D(G(y))$. Por $D_n(y)$ se denotará la distribución de

$$r^{1/2} \left[\sum_{i=1}^r \frac{W_i}{r} - m(G(y)) \right]$$

que tiene por media 0 y por varianza $\sigma^2(G(y))$. La distribución de Z_{ni} ($i = 1, \dots, n, n \in \mathbb{N}$) es $D_n(i/n)$ y se obtiene (Brunk [3]) que

$$\left(\frac{nr}{(b-a)\theta} \right)^{1/2} \text{slogcom}_{x=x_0} [\widehat{\Psi}(x) - \Psi(x)] \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \text{slogcom}_{x=x_0} [W(U(x))]$$

siendo $\Psi(x) := \int_a^x m(u) du$ y su estimador natural $\widehat{\Psi}(x) := \sum_{i=1}^n \widehat{m}(x_{ni}) \chi_{ni}(x)$, $x \in [a, b]$, de manera que

$\chi_{ni}(x) := \int_a^x I_{(x_{ni-1}, x_{ni}]}(u) du$, I_A representa el indicador del conjunto A y $x_{n0} := a$.

El resultado se deriva de la igualdad $\text{slogcom}_{x=x_0} [\widehat{\Psi}(x) - \Psi(x)] = \widehat{m}_I(x_0) - m(x_0)$. Para comprobar dicha igualdad, observése que $\widehat{\Psi}$ es la función construida por interpolación lineal a partir de los puntos $(a, 0) \cup \left\{ \left(a + \frac{(b-a)i}{n}, \frac{b-a}{n} \sum_{j=1}^i \widehat{m}(x_{nj}) \right) \right\}_{i=1}^n$ y que \widehat{m}_I es la pendiente por la izquierda en i de la

mayor minorante convexa de la función $f : [0, n] \rightarrow \mathbb{R}$ construida por interpolación lineal a partir de los puntos $(0, 0) \cup \left\{ \left(i, \sum_{j=1}^i \widehat{m}(x_{nj}) \right) \right\}_{i=1}^n$.

Teniendo en cuenta la relación entre f y $\widehat{\Psi}$:

$$f(t) = \frac{n}{b-a} \widehat{\Psi} \left(a + \frac{(b-a)t}{n} \right), \quad t \in [0, n]$$

se obtiene que $\widehat{m}_I(x_0) = \text{slogcom}_{x=x_0}^{[a,b]} [\widehat{\Psi}(x)]$.

Por último, como m es constante,

$$m(x_0) = \text{slogcom}_{x=x_0}^{[a,b]} [(x-a)m(x_0)] = \text{slogcom}_{x=x_0}^{[a,b]} [\Psi(x)]$$

y por tanto $\widehat{m}_I(x_0) - m(x_0) = \text{slogcom}_{x=x_0}^{[a,b]} [\widehat{\Psi}(x) - \Psi(x)]$, con lo que se demuestra el apartado (i).

(ii) En primer lugar vamos a obtener la distribución asintótica de $\widehat{m} - m$ en $x_0 \in [a, b]$.

Sea, para cada n , $x_0^n := \min\{x_{ni}, x_{ni} \geq x_0, i = 1, \dots, n\}$ y sea Y_{n0}^j la observación j -ésima en el punto x_0^n . Consideramos la función

$$g_n(x) := (x - m(x_0^n))^2 I_{[\varepsilon^2 r \sigma^2(x_0^n), \infty)}([x - m(x_0^n)]^2).$$

Es sencillo ver que $g_n(Y_{n0}^j)$ converge en $L_1(\Omega, \mathcal{F}, P)$ a 0 y se tiene que

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{\sigma^2(x_0)} \int_{\{(Y_{n0}^1 - m(x_0^n))^2 \geq \varepsilon^2 r \sigma^2(x_0^n)\}} (Y_{n0}^1 - m(x_0^n))^2 dP = 0.$$

Entonces, como $\sigma^2(x_0^n)/\sigma^2(x_0) \rightarrow 1$ por ser σ continua y como $Y_{n0}^1, \dots, Y_{n0}^r$ son independientes con distribución $D(x_0^n)$, el Teorema de Lindeberg implica que

$$\left(\frac{r}{\sigma^2(x_0^n)} \right)^{1/2} (\widehat{m}(x_0^n) - m(x_0^n)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

A partir de aquí y teniendo en cuenta que m es constante y \widehat{m} constante a trozos, se obtiene

$$\left(\frac{r}{\sigma^2(x_0)} \right)^{1/2} (\widehat{m}(x_0) - m(x_0)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1). \quad (3)$$

Ahora, sea $(t_1, \dots, t_s) \in \mathbb{R}^s$. Entonces,

$$\begin{aligned} & \sum_{j=1}^s t_j r^{1/2} (\widehat{m}_I(x_j) - \widehat{m}(x_j)) = \\ & = \sum_{j=1}^s t_j \left(\frac{(b-a)\theta}{n} \right)^{1/2} \left(\frac{nr}{(b-a)\theta} \right)^{1/2} (\widehat{m}_I(x_j) - m(x_j)) - \sum_{j=1}^s t_j r^{1/2} (\widehat{m}(x_j) - m(x_j)). \end{aligned}$$

El primer sumando converge en probabilidad a 0 por el apartado (i). A partir de la convergencia dada en (3) se tiene que $t_j r^{1/2} (\widehat{m}(x_j) - m(x_j)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, t_j \sigma(x_j))$, $j = 1, \dots, s$. Además, si n es suficientemente grande, las variables $\widehat{m}(x_1), \dots, \widehat{m}(x_s)$ son independientes y se llega a que el segundo sumando converge en distribución a $\sum_{j=1}^s t_j \sigma(x_j) Z_j$, donde Z_1, \dots, Z_s son variables aleatorias independientes $N(0, 1)$. El apartado (ii) se sigue del Teorema de Cramer-Wold. ■

La Proposición 2 (ii) permite obtener de forma directa (como consecuencia del Teorema de la Aplicación Continua) la distribución asintótica de un estadístico sobre s puntos fijos:

$$r^{1/2} \left[\sum_{j=1}^s \frac{(\widehat{m}_I(x_j) - \widehat{m}(x_j))^2}{s} \right]^{1/2} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \left[\sum_{j=1}^s \frac{\sigma^2(x_j) Z_j^2}{s} \right]^{1/2} \quad (4)$$

donde Z_1, \dots, Z_s son variables aleatorias independientes $N(0, 1)$.

Dicho estadístico se llamará $s - Dip$ y si $a < x_1 < \dots < x_s = b$ ($s \in \mathbb{N}$) son puntos fijos igualmente espaciados, se define

$$s - Dip \equiv s - Dip_{nr} := \left[\sum_{j=1}^s \frac{(\widehat{m}_{nr}(x_j) - \widehat{m}_{nrI}(x_j))^2}{s} \right]^{1/2}.$$

4.. El Dip -test.

Si se dispone de n puntos de diseño y r observaciones en cada uno de ellos, el siguiente resultado muestra que el correspondiente Dip siempre se puede acotar por un adecuado $s - Dip$. Esta acotación será clave para desarrollar el test.

Proposición 3 Sean $n, r, k \in \mathbb{N}$ y sea $j = kn$. Entonces, se tiene que

$$Dip_{nr} \leq n - Dip_{jr}.$$

DEMOSTRACIÓN. Puesto que $j = kn$, se tiene que $x_{ni} = x_{ju}$ para $u = ki, i = 1, \dots, n$. En particular \mathcal{X}_n es un subconjunto de \mathcal{X}_j . Como \widehat{m}_{jrI} es una función creciente sobre el conjunto de puntos \mathcal{X}_j , también lo es sobre el subconjunto \mathcal{X}_n y

$$Dip_{nr} \leq \left[\sum_{i=1}^n \frac{(\widehat{m}_{nr}(x_{ni}) - \widehat{m}_{jrI}(x_{ni}))^2}{n} \right]^{1/2}.$$

Además, $\widehat{m}_{nr}(x_{ni}) = \widehat{m}_{jr}(x_{ni}), i = 1, \dots, n$, y entonces, el último término es igual al estadístico $n - Dip_{jr}$. ■

En el Teorema 1 se obtiene una cota asintótica en n y en r de la distribución del Dip bajo regresión constante.

Teorema 1 Sea m constante, $\sigma > 0$ continua y $\sup_{x \in [a, b]} S^{2+\delta}(x) < \infty$ para algún $\delta > 0$. Entonces,

$$\overline{\lim}_{n \rightarrow \infty} \overline{\lim}_{r \rightarrow \infty} P(Dip \geq H) \leq \alpha$$

donde

$$(i) \quad H^2 \equiv H_{nr}^2 = r^{-1} \left[\sum_{i=1}^n \frac{\sigma^2(x_{ni})}{n} + \left(\frac{2\beta}{n(b-a)} \right)^{1/2} z_{1-\alpha} \right]$$

siendo $z_{1-\alpha}$ el $(1 - \alpha)$ -percentil de la distribución $N(0, 1)$, $\alpha \in (0, 1)$ y $\beta := \int_a^b \sigma^4(x) dx$.

Si además existe la derivada de σ, σ' , en (a, b) y $\sup_{x \in (a, b)} |\sigma'(x)| < \infty$, H puede tomarse de la forma

$$(ii) \quad H^2 \equiv H_{nr}^2 = r^{-1} \left[\frac{\theta}{(b-a)} + \left(\frac{2\beta}{n(b-a)} \right)^{1/2} z_{1-\alpha} \right] \text{ con } \theta := \int_a^b \sigma^2(x) dx.$$

Si $\sigma(x) = \sigma^* \forall x \in [a, b]$,

$$(iii) \quad H^2 \equiv H_{nr}^2 = r^{-1} \sigma^{*2} \left[1 + (2/n)^{1/2} z_{1-\alpha} \right].$$

DEMOSTRACIÓN. Si $n, r \in \mathbb{N}$, para cualquier $A \in \mathbb{R}$ se tiene de la Proposición 3 (con $k = r$) que

$$\overline{\lim}_{r \rightarrow \infty} P(Dip_{nr} \geq Ar^{-1/2}) \leq \overline{\lim}_{r \rightarrow \infty} P(n - Dip_{(rn)r} \geq Ar^{-1/2}). \quad (5)$$

Se considera la función

$$g(u) := \begin{cases} 1, & \text{si } u \leq 2n - 1 \\ i, & \text{si } in \leq u < (i+1)n \quad (i = 2, 3, \dots) \end{cases} \quad (u \in \mathbb{N})$$

y a partir de (4) se tiene que

$$\lim_{u \rightarrow \infty} P(n - Dip_{ug} \geq Ag^{-1/2}) = P \left(\left[\sum_{i=1}^n \frac{\sigma^2(x_{ni}) Z_i^2}{n} \right]^{1/2} \geq A \right).$$

La sucesión $\{P(n - Dip_{(rn)r} \geq Ar^{-1/2})\}_{r=1}^{\infty}$ es una subsucesión de $\{P(n - Dip_{ug} \geq Ag^{-1/2})\}_{u=1}^{\infty}$, puesto que $r = g(rn)$, y ha de tener el mismo límite:

$$\lim_{r \rightarrow \infty} P(n - Dip_{(rn)r} \geq Ar^{-1/2}) = P \left(\left[\sum_{i=1}^n \frac{\sigma^2(x_{ni}) Z_i^2}{n} \right]^{1/2} \geq A \right).$$

Sea ahora $H_n^2 = r H_{nr}^2$ con H_{nr} como en (i). Tomando A como H_n y pasando al límite cuando $n \rightarrow \infty$ se obtiene que

$$\overline{\lim}_{n \rightarrow \infty} \overline{\lim}_{r \rightarrow \infty} P(Dip_{nr} \geq H_{nr}) \leq \overline{\lim}_{n \rightarrow \infty} P \left(\left[\frac{n(b-a)}{2\beta} \right]^{1/2} \left[\sum_{i=1}^n \frac{\sigma^2(x_{ni}) Z_i^2}{n} - \sum_{i=1}^n \frac{\sigma^2(x_{ni})}{n} \right] \geq z_{1-\alpha} \right).$$

Es sencillo ver que el último límite existe y vale $P(N(0, 1) \geq z_{1-\alpha}) = \alpha$, lo que prueba el apartado (i).

De forma análoga y tomando $H_n^2 = r H_{nr}^2$ con H_{nr} como en (ii), se tiene el segundo apartado.

El apartado (iii) se demuestra tomando σ constante en (i). ■

Del resultado anterior se deriva un test no paramétrico (*Dip*-test) que es al menos conservador asintóticamente en n y r : a partir de n puntos de diseño y r observaciones en cada uno de ellos, la isotonía se rechaza si $Dip \geq H$ donde H vendría dado por (i), (ii) ó (iii) en el Teorema 1, dependiendo de las propiedades que se asuman para la varianza. A continuación, se va a ilustrar el comportamiento del *Dip*-test mediante algunas simulaciones.

5.. Simulaciones

Considérense las familias $\{D(x)\}_{x \in [0,1]}$ de distribuciones de probabilidad en el intervalo $[0, 1]$ de forma que $\forall x \in [0, 1]$ se verifique una de las tres condiciones: (i) $D(x) \equiv N(0, 1) + m(x)$, (ii) $D(x) \equiv 2\sqrt{3}U(0, 1) + m(x) - \sqrt{3}$, (iii) $D(x) \equiv E(1) + m(x) - 1$ y además, $m(x) = -Ax$, $A \geq 0$, $\sigma(x) = 1$.

Nótese que el caso $A = 0$ será el caso de regresión constante y servirá para estudiar el nivel de significación. El caso $A > 0$ significa regresión estrictamente decreciente y lo utilizaremos para estudiar la potencia del *Dip*-test.

Fijados el número n de puntos de diseño, el número r de observaciones por punto, el modelo (i), (ii) ó (iii) y la pendiente $-A$ de la regresión, se generarán r observaciones con distribución $D(x_{ni})$ en cada punto de diseño $x_{ni}, i = 1, \dots, n$, y de forma que las nr observaciones resultantes sean independientes. A partir de esta nube de puntos se calculará el estadístico *Dip*, utilizando el algoritmo dado por Puri y Singh [14] para obtener \hat{m}_I .

En virtud de los resultados anteriores, para un nivel de significación $\alpha \in (0, 1)$ se rechazaría la hipótesis de isotonía si $Dip \geq \left[r^{-1} \left(1 + (2/n)^{1/2} z_{1-\alpha} \right) \right]^{1/2}$ donde $z_{1-\alpha}$ es el $(1 - \alpha)$ -percentil de la distribución $N(0, 1)$.

A continuación hemos repetido el proceso generando 10.000 nubes de puntos, para valores fijos de n, r y A y cada uno de los tres modelos anteriores, contabilizando la proporción de rechazos de la isotonía a un nivel prefijado $\alpha = 0,05$. Para valores fijos de n y r , la potencia del test como función de $A > 0$ bajo cada uno de los tres modelos puede ser observada mediante esas proporciones. Salvo en valores muy próximos a $A = 0$, la potencia resultó ser independiente del modelo elegido y reflejaría entonces la potencia real del test en función de A , sin ninguna restricción paramétrica sobre la distribución de probabilidad.

En la Figura 1 se muestran las funciones de potencia cuando $n=20, 50, 100$ y $r=10, 50, 100$. En el eje de abscisas aparecen los valores de A y en el de ordenadas los correspondientes porcentajes de rechazos de la hipótesis nula. Las funciones fueron obtenidas interpolando mediante splines de tercer grado las probabilidades de rechazo para $A \in \{0, 1x : x = 0, 1, \dots, 20\}$. Las diferencias en puntos próximos a $A = 0$ (incluyendo el propio punto) son imperceptibles a efectos de las representaciones gráficas.

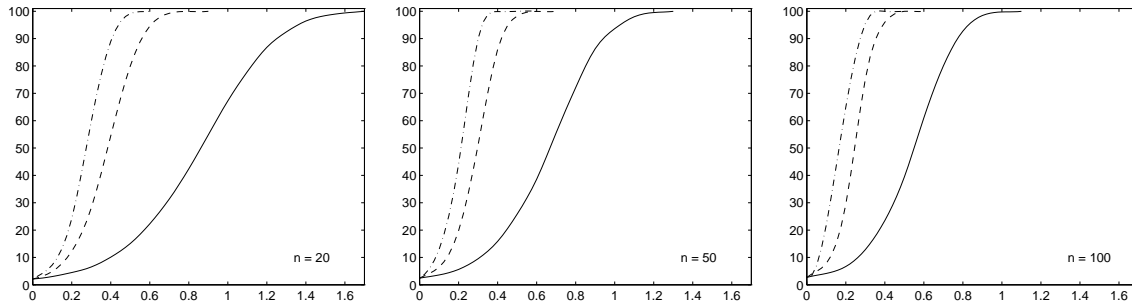


Figura 1. Funciones de potencia del *Dip*-test para distintos valores de n -número de puntos de diseño- y r -número de observaciones en cada punto de diseño-. (- $r=10$; -- $r=50$; -.- $r=100$).

En la práctica, lo importante es qué pendientes puede detectar el test de forma efectiva, es decir con una potencia alta: para un 90% de potencia, $n=20$ y $r=50$ se obtiene $A = 0,57$, e incluso con el menor número de datos manejado, $n=20, r = 10$, A es 1.26, resultados satisfactorios para una desviación típica de 1 y que posteriormente se compararán con los del $\bar{\chi}_{12}^2$ -test bajo normalidad. En la Figura 1 se observa también, como cabe esperar, que para valores fijos de A la potencia crece tanto con n como con r . Sin embargo, el efecto de aumentar r es mucho mayor: nótese que por ejemplo la función de potencia es similar para $n = 20, r = 100$ y para $n = 100, r = 50$, siendo los datos en el primer caso menos de la mitad que en el segundo.

A continuación, se estudia el caso de regresión constante, $A = 0$, aumentando el número de nubes de puntos a 100.000. En el Cuadro 1 se muestran las probabilidades de rechazo para los tres modelos y distintos valores de n y r . Se observa que los porcentajes de rechazo están bastante por debajo del 5%, incluso aunque el número de datos sea muy grande. En el modelo Normal no aparece r porque no depende del número de observaciones por columna (la distribución de $\bar{\chi}_{12}^2$ es independiente de r bajo normalidad y regresión constante, Robertson et al. [16]) y si se observan los otros modelos, a medida que r crece, los

valores se van pareciendo a los de la Normal.

Además se han realizado simulaciones en el caso Normal para valores de n superiores a 1.000 mediante la aproximación de Robertson et al. [16],

$$P\left(Dip \geq \left[r^{-1} \left(1 + (2/n)^{1/2} 1,645\right)\right]^{1/2}\right) = P\left(\bar{\chi}_{12}^2 \geq n \left(1 + (2/n)^{1/2} 1,645\right)\right),$$

y se ha comprobado numéricamente que, cuando $n \rightarrow \infty$, el nivel tiende al 5 %, pero muy lentamente: para valores de $n=10.000$, 100.000 y 500.000 se obtuvieron porcentajes de 4.40, 4.75 y 4.80 respectivamente, alcanzándose efectivamente el nivel prefijado, aunque de forma muy lenta. Así pues, el *Dip*-test no es asintóticamente conservador sino que es del nivel prefijado, aunque éste se alcanza muy despacio.

n	Uniforme				Normal	Exponencial			
	r=1	r=10	r=50	r=100		r=1	r=10	r=50	r=100
20	0.08	1.84	2.03	2.06	2.11	9.45	3.46	2.45	2.26
50	0.10	2.11	2.33	2.45	2.43	12.12	4.02	2.81	2.66
100	0.12	2.37	2.65	2.73	2.70	14.02	4.60	3.16	2.90
500	0.19	3.07	3.41	3.37	3.42	17.02	5.54	3.98	3.60
1.000	0.23	3.32	3.82	3.61	3.72	17.79	5.85	4.18	3.91

Cuadro 1. Porcentaje de rechazos para el *Dip*-test para $A=0$ y un nivel de significación prefijado $\alpha = 0,05$.

Sin embargo, para valores fijos de n y r el test es bastante conservador y se puede pensar que este carácter afecte negativamente a la función de potencia. Para ver que esto no es así, hemos comparado el *Dip*-test con el $\bar{\chi}_{12}^2$ -test bajo normalidad (Figura 2).

Aunque el $\bar{\chi}_{12}^2$ -test es prácticamente del nivel α prefijado para cualquier volumen de datos, en la Figura 2 se ve que la potencia del mismo no mejora significativamente la potencia del *Dip*-test para valores pequeños de n y r , siendo de hecho totalmente similares para n y r grandes.

En conclusión, el *Dip*-test no se ve afectado por su carácter conservador y además en el caso de normalidad su sencillez es una baza importante para usarlo en vez del $\bar{\chi}_{12}^2$ -test específicamente desarrollado para el caso Normal y que presenta dificultades de cálculo en la práctica.

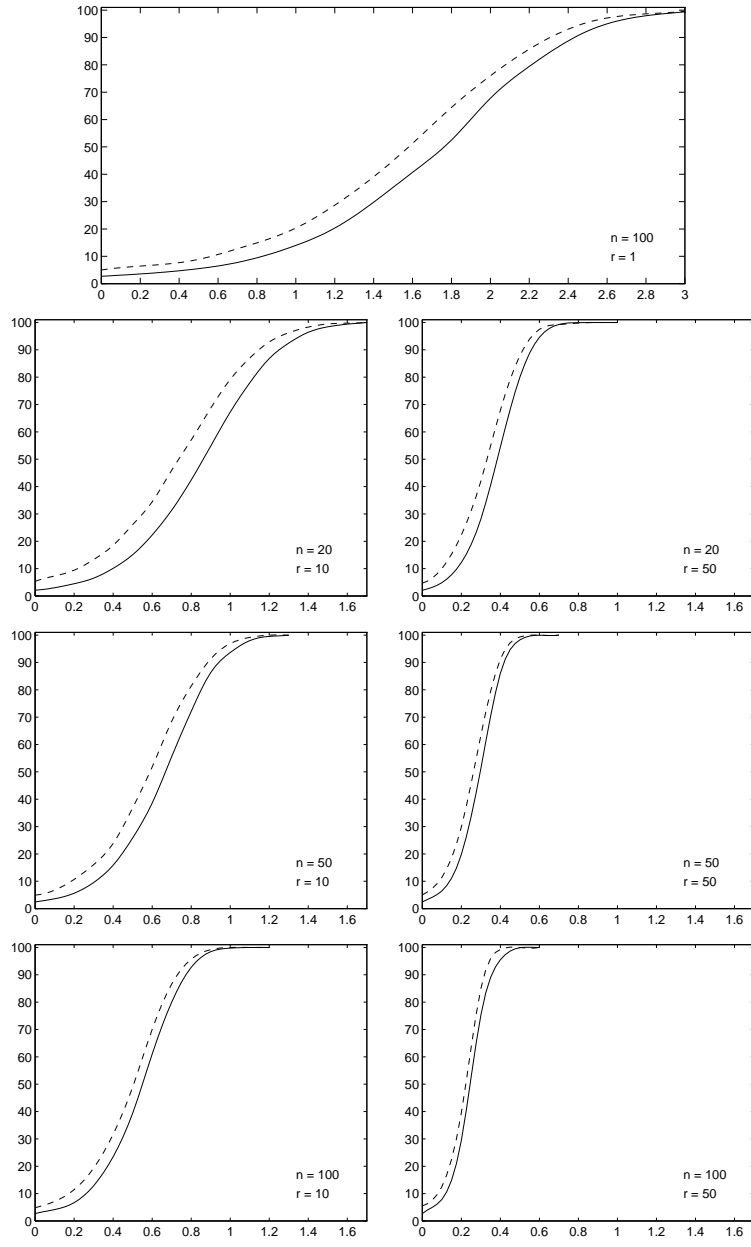


Figura 2. Comparación de las funciones de potencia del $\bar{\chi}_{12}^2$ -test (trazo discontinuo) y del Dip-test (trazo continuo).

Referencias

- [1] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. y Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.
- [2] Bowman, A.W., Jones, M.C. y Gijbels, I. (1988). Testing monotonicity of regression. *J. Comput. Graph. Statist.*, **7**, 489–500.
- [3] Brunk, H.D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference*. Cambridge University Press, 177–195.
- [4] Cuesta, J.A., Domínguez, J.S. y Matrán, C. (1994). Some stochastics on monotone functions. *J. Comput. App. Math.*, **55**, 165–182.
- [5] Cuesta, J.A., Domínguez, J.S. y Matrán, C. (1995). Consistency of L_p -best monotone approximations. *J. Stat. Planning Infer.*, **47**, 295–318.
- [6] Domínguez, J.S. (1991). Medidas de isotonía. Aplicaciones a la regresión no-paramétrica. *Tesis doctoral*, Univ. de Oviedo.
- [7] Domínguez, J.S. y López, M.J. (1997). On the estimation of monotone uniform approximations. *Statist. Probab. Lett.*, **35**, 355–362.
- [8] Ghosal, S., Sen, A. y Van der Vaart, A.W. (2000). Testing monotonicity of regression. *Technical Report WS-514*, Faculteit der Exacte Wetenschappen, Vrije Univ. Amsterdam.
- [9] Hall, P. y Heckman, N.E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.*, Vol. **28**, No. 1, 20–39.
- [10] Hanson, D.L., Pledger, G. y Wright, F.T. (1973). On consistency in monotonic regression. *Ann. Statist.*, **1**, 401–421.
- [11] López-Palomo, M.J. (2000). Desarrollo de procedimientos no paramétricos en el contraste de la regresión isotónica. *Tesis doctoral*, Univ. de Oviedo.
- [12] Mammen, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.*, **19** 2, 724–740.
- [13] Mukerjee, H. (1988). Monotone nonparametric regression. *Ann. Statist.*, **16** 2, 741-750.
- [14] Puri, P.S. y Singh, H. (1990). On recursive formulas for isotonic regression useful for statistical inference under order restrictions. *J. Stat. Planning Infer.*, **24**, 1–11.
- [15] Robertson, T. y Wegman, E.J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.*, **6** 3, 485–505.
- [16] Robertson, T., Wright, F.T. y Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- [17] Wright, F.T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.*, **9** 2, 443–448.
- [18] Wright, F.T. (1982). Monotone regression estimates for grouped observations. *Ann. Statist.*, **10** 1, 278–286.

M. J. López Palomo & J. S. Domínguez Menchero
Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
C/Calvo Sotelo s/n
33071 Oviedo, Spain
mjlp@pinon.ccu.uniovi.es jsdm@pinon.ccu.uniovi.es