

# ESTIMACIÓN NO PARAMÉTRICA DE LA DENSIDAD Y DE LA REGRESIÓN - PREVISIÓN NO PARAMÉTRICA

MICHEL CARBON\*– CHRISTIAN FRANCO†

*Recibido: 28 Agosto 1998*

---

## Resumen

Después de un vistazo sobre la estimación no paramétrica de la densidad y de la regresión, detallamos e interpretamos un método de previsión, llamado de previsión no paramétrica. Mostramos los diferentes aspectos tanto técnicos como prácticos, y comparamos el método propuesto sobre algunos ejemplos respecto a la metodología de Box y Jenkins.

**Palabras-clave:** Estadística no paramétrica, series de tiempo, previsión, regresión, núcleos.

## Abstract

We begin with a short view of non parametric estimation of density and regression; then we give details and interpretation of a forecasting method, called non parametric forecasting. We show different aspects, technical as well as practical, and we compare the method with the Box and Jenkins methodology, over some examples.

**Keywords:** Non-parametric statistics, time series, forecasting, regression, kernels.

**AMS Subject Classification:** 62M10.

## 1 Introduction

Consideremos una serie cronológica  $x_1, x_2, \dots, x_n$ . A partir de estas  $n$  observaciones, se quiere predecir  $x_{n+h}$  ( $h$  es el horizonte de previsión). La primera pregunta que se plantea es ¿cuál es el método que se debe escoger? Aún si la pregunta es tan sencilla, generalmente la respuesta es, desdichadamente, decepcionante. No existe un método mejor que los demás, en todos los casos. Desde hace algún tiempo aparecieron nuevos métodos, llamados no

---

\*Laboratoire de Statistique, Universidad de Rennes II, 6 Avenue Gaston Berger, 35000 Rennes, Francia

†Laboratoire de Probabilités et Statistique, Bâtiment M2, 59650 Villeneuve d'Ascq Cedex, Francia

paramétricos, que parecen aportar una nueva perspectiva sobre este asunto. Estas nuevas técnicas presentan la ventaja de ser fáciles de implementar, de una interpretación más intuitiva y cierta robustez, respecto a los métodos llamados de Box y Jenkins.

En este artículo, haremos un pequeño panorama de ciertos enfoques no paramétricos, tratando de mostrar sobre todo cómo se inserta la previsión no paramétrica. En particular, comenzamos por la estimación no paramétrica de la densidad. El lector interesado puede referirse a los trabajos de Bosq y Lecoutre (1987) o de Praksa Rao (1983). Enseguida, hacemos una presentación bastante rápida de la regresión no paramétrica, de la cual la previsión no paramétrica es un caso particular. Tratamos de mostrar, finalmente, que la interpretación es totalmente natural. Terminamos con ejemplos que muestran una cierta superioridad del método no paramétrico respecto a la metodología de Box y Jenkins.

## 2 Estimación no paramétrica de la densidad

### 2.1 Histograma

El estimador más rudimentario para estimar una densidad es el histograma de las frecuencias. Supongamos que se tiene  $x_1, \dots, x_n$ ,  $n$  observaciones provenientes de una misma ley de probabilidad de densidad  $f$ , donde  $f$  tiene soporte acotado  $[a, b[$ . Para estimar esta densidad  $f$  por el método del histograma, lo que se reduce a aproximar  $f$  por una función escalonada, se corta  $[a, b[$  en  $k$  clases  $[\alpha_i, \alpha_{i+1}[$  donde  $i = 1, \dots, k$ , con  $a = \alpha_1$  y  $b = \alpha_{k+1}$ . El estimador histograma se escribe entonces:  $\forall t \in [a, b[$ ,  $\exists i = 1, \dots, k$  tal que  $t \in [\alpha_i, \alpha_{i+1}[$  y

$$\hat{f}_n(t) = \frac{f_i}{\alpha_{i+1} - \alpha_i},$$

donde  $f_i$  es la frecuencia del número de puntos de la clase correspondiente. Esto se puede escribir también de forma más concisa:  $\forall t \in [a, b[$ ,

$$\hat{f}_n(t) = \sum_{i=1}^k \frac{f_i}{\alpha_{i+1} - \alpha_i} \cdot \mathbf{1}_{[\alpha_i, \alpha_{i+1}[}(t),$$

donde

$$f_i = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{[\alpha_i, \alpha_{i+1}[}(x_j).$$

Es decir:

$$\hat{f}_n(t) = \sum_{i=1}^k \frac{1}{n(\alpha_{i+1} - \alpha_i)} \sum_{j=1}^n \mathbf{1}_{[\alpha_i, \alpha_{i+1}[}(x_j). \quad (1)$$

Para simplificar las notaciones, se supondrá ahora que las clases tienen la misma amplitud, es decir que para todo  $i = 1, \dots, k$ ;  $\alpha_{i+1} - \alpha_i = b(n)$ . Es fácil notar que  $\hat{f}_n$  es una densidad de probabilidad. Si se piensa en la convergencia de este estimador, es claro que  $\hat{f}_n$  será más cercano de la verdadera densidad  $f$  si las amplitudes de las clases son

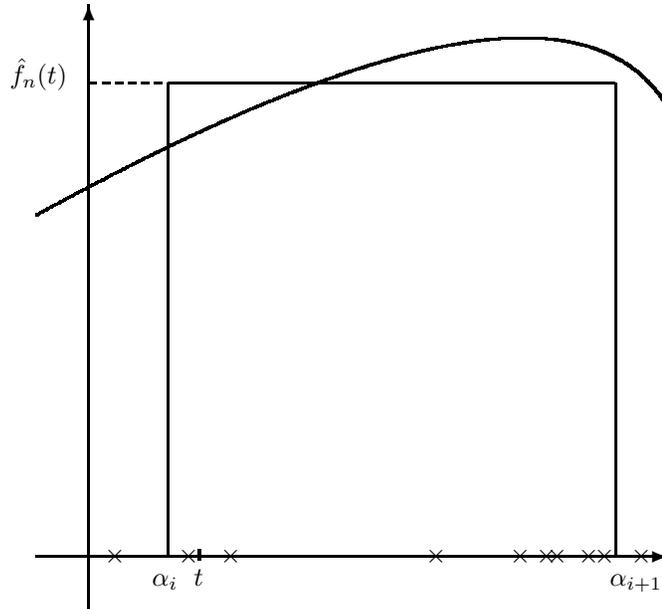


Figure 1: Histograma de las frecuencias

más estrechas, por ello es necesario imponer que  $b(n) \rightarrow 0$  cuando  $n \rightarrow \infty$ . En cambio, es necesario que  $b(n)$  no tienda muy rápido hacia 0, sino se podría tener clases que no contengan ningún punto, y entonces se tendría una función escalonada  $\hat{f}_n$  con gradas de ordenada nula, muy alejada de la realidad. Es necesario que  $b(n)$  tienda hacia 0 con  $n$ , y que, a pesar de ello, haya cada vez más puntos en cada clase, lo que se puede resumir en la condición:

$$nb(n) \rightarrow \infty \quad \text{cuando} \quad n \rightarrow \infty.$$

El error cuadrático medio  $E[\hat{f}_n(x) - f(x)]^2$  es del orden  $n^{-2/3}$ , para una escogencia óptima de  $b(n)$ .

## 2.2 Histograma móvil

Vamos a tratar de mejorar este estimador histograma  $\hat{f}_n$ . Consideremos la clase  $C_i = [\alpha_i, \alpha_{i+1}[$ , e imaginamos que el punto  $t$  de  $C_i$  donde se quiere estimar  $f(t)$  por  $\hat{f}_n(t)$  se sitúe cerca del extremo  $\alpha_i$  (ver figura 1). Entonces todos los puntos de la clase  $C_i$  intervienen en el cálculo de  $\hat{f}_n$ , pero se puede ver que un punto situado cerca de  $\alpha_{i+1}$  será tomado en cuenta, aún si está bastante alejado de  $t$ , y que un punto situado muy cerca de  $t$  en la clase  $C_{i-1}$  no entra en el cálculo de  $\hat{f}_n$ .

Para remediar este inconveniente, se puede usar el histograma móvil, que es una traslación del histograma de forma que el punto  $t$  donde se hace la estimación, se encuentre

en el centro de una clase, más precisamente en el centro de la clase  $[t - h(n), t + h(n)[$  donde  $h(n)$  denota la semiamplitud de una clase. El estimador histograma móvil se escribe entonces:

$$\hat{f}_n(t) = \frac{1}{2nh(n)} \sum_{j=1}^n \mathbf{1}_{[t-h(n), t+h(n)[}(x_j). \quad (2)$$

Observemos que:

$$t - h(n) \leq x_j < t + h(n) \iff -1 \leq \frac{x_j - t}{h(n)} < 1. \quad (3)$$

De donde:

$$\hat{f}_n(t) = \frac{1}{2nh(n)} \sum_{j=1}^n \mathbf{1}_{[-1, 1[}\left(\frac{x_j - t}{h(n)}\right).$$

El estimador se escribe entonces:

$$\hat{f}_n(t) = \frac{1}{nh(n)} \sum_{j=1}^n \mathbf{K}\left(\frac{x_j - t}{h(n)}\right),$$

donde

$$\mathbf{K}(x) = \frac{1}{2} \mathbf{1}_{[-1, 1[}(x).$$

### 2.3 Estimación por el método del núcleo

El estimador construido aún puede ser mejorado. En efecto, ahora que la clase está centrada en  $t$ , se puede notar que todos los puntos de esta clase juegan el mismo rol en cuanto al cálculo de  $\hat{f}_n(t)$ . Sería más sensato pensar que entre más cercano esté un punto de  $t$ , más debe intervenir en el cálculo de  $\hat{f}_n(t)$ . Entonces, la idea más natural es ponderar las observaciones poniendo más peso cuanto más cerca se esté de  $t$ , y menos si se encuentra lejos.

Ya se ha visto un ejemplo de función de peso, denotado  $\mathbf{K}$  en la sección anterior. Era una densidad de probabilidad (la ley uniforme sobre  $[-1/2, 1/2[$ ). Esta función de peso es demasiado brutal y no responde a nuestras preocupaciones. Se escogerán entonces funciones de peso en clases más amplias de densidades, que sobre todo comprenden densidades de soporte no acotado, y que tengan una sola moda con el origen (por ejemplo la ley normal centrada estandarizada).

Debe observarse que, puesto que  $\mathbf{K}$  es una densidad de probabilidad,  $\hat{f}_n(t)$  también es una densidad de probabilidad. El estimador con núcleo se escribe:

$$\hat{f}_n(t) = \frac{1}{nh(n)} \sum_{j=1}^n \mathbf{K}\left(\frac{x_j - t}{h(n)}\right) \quad (4)$$

y  $\mathbf{K}$  se llama un núcleo. En cuanto a las propiedades de convergencia, se muestra que  $\hat{f}_n$  es asintóticamente sin sesgo. También se puede mostrar que, para que  $Var[\hat{f}_n(t)] \rightarrow 0$ , es

necesario que  $h(n) \rightarrow 0$  y  $nh(n) \rightarrow +\infty$  cuando  $n \rightarrow \infty$ . La velocidad de convergencia, esta vez, para la mejor escogencia de  $h(n)$ , y en el sentido del error cuadrático medio es del orden de  $n^{-4/5}$ .

Queda hablar de la escogencia del núcleo y del parámetro  $h(n)$ . Estas escogencias solo se pueden hacer usando ciertos criterios. Sin entrar en todos los detalles, resulta que la escogencia del núcleo no tiene mayor influencia si es escogido en una clase razonable de estimadores. La escogencia de un núcleo gaussiano, por ejemplo, es muy recomendada. La escogencia de  $h(n)$  es, sin embargo, crucial. Se recomienda la siguiente escogencia:

$$h(n) = S_n n^{-1/5},$$

donde  $S_n$  denota la desviación estándar estimada de las observaciones.

Un punto esencial, en favor de esta técnica de estimación, es la fórmula (4) que permanece casi sin modificaciones en el caso multidimensional:

$$\hat{f}_n(t) = \frac{1}{nh^s(n)} \sum_{j=1}^n \mathbf{K} \left( \frac{x_j - t}{h(n)} \right) \quad (5)$$

donde  $t$ , los  $x_j$  están en  $\mathbf{R}^s$ , y donde  $\mathbf{K}$  es una densidad definida sobre  $\mathbf{R}^s$ . En general  $\mathbf{K}$  se escoge como producto de núcleos de  $\mathbb{R}$  en  $\mathbb{R}$ .  $h(n)$  se obtiene por técnicas de validación. Para mayores detalles, ver Devroye y Györfi (1985).

## 2.4 Regresión no paramétrica

Supongamos que el comportamiento de una variable aleatoria  $Y$  esté ligado a otra variable  $X$ . Es clásico tratar de expresar  $Y$  linealmente en función de  $X$ . Es el problema bien conocido de la regresión lineal. Tal vez sería mejor tratar de expresar  $Y$  por una función  $R$  no necesariamente lineal de  $X$ , es decir encontrar una expresión de la forma:

$$Y = R(X) + \varepsilon \quad \text{donde } \varepsilon \text{ es el residuo.}$$

Se puede entonces tratar de determinar  $R$  como solución del problema de minimización:

$$\min E[Y - R(X)]^2.$$

Se obtiene, bajo ciertas condiciones de regularidad (esencialmente de integrabilidad de  $Y$ ):

$$R(X) = E(Y|X),$$

es decir la regresión de  $Y$  sobre  $X$ .

## 2.5 Regresograma

Desde un punto de vista práctico, a partir de  $n$  observaciones  $(x_i, y_i)$ , se va a estimar  $R(t) = E(Y|X = t)$ , fabricando clases  $C_i$  donde se sitúen o no los  $x_j$ . Para la clase  $C_i$  donde se encuentra el punto  $t$ , se efectúa la media de los  $y_j$  correspondientes a los  $x_j$  de esta clase  $C_i$ .

Denotando  $k$  el número de puntos  $x_j$  de la clase  $C_i$ , para todo  $t$  de  $C_i$ , se estima  $R(t)$  por:

$$\hat{R}_n(t) = \frac{1}{k} \sum_{j=1}^k y_j \quad \text{con} \quad k = \sum_{j=1}^n \mathbf{1}_{C_i}(x_j).$$

De donde

$$\hat{R}_n(t) = \frac{\sum_{j=1}^n \mathbf{1}_{C_i}(x_j) y_j}{\sum_{j=1}^n \mathbf{1}_{C_i}(x_j)}.$$

Desde luego, como para el histograma, el regresograma  $\hat{r}_n$  es constante sobre cada clase  $C_i$ . Para mejorarlo, se seguirá entonces el mismo procedimiento que para el histograma.

## 2.6 Regresograma móvil

Como para el histograma móvil, la primera mejoría consiste en centrar la clase en el punto  $t$  donde se estima la regresión. Este regresograma móvil se escribe:

$$\hat{R}_n(t) = \frac{\sum_{j=1}^n \mathbf{1}_{[t-h(n), t+h(n)]}(x_j) y_j}{\sum_{j=1}^n \mathbf{1}_{[t-h(n), t+h(n)]}(x_j)},$$

lo que, gracias a la observación (3), da:

$$\hat{R}_n(t) = \frac{\sum_{j=1}^n \mathbf{1}_{[-1, 1]}[(x_j - t)/h(n)] y_j}{\sum_{j=1}^n \mathbf{1}_{[-1, 1]}[(x_j - t)/h(n)]}.$$

## 2.7 Regresión por el método del núcleo

Por analogía con el caso del histograma, la última etapa consiste en reemplazar la densidad de probabilidad  $\frac{1}{2} \mathbf{1}_{[-1, 1]}$  por un núcleo cualquiera del tipo descrito en el caso del histograma. Entonces, la regresión por el método del núcleo se escribe:

$$\hat{r}_n(t) = \frac{\sum_{j=1}^n \mathbf{K}[(x_j - t)/h(n)] y_j}{\sum_{j=1}^n \mathbf{K}[(x_j - t)/h(n)]}. \quad (6)$$

La velocidad de convergencia, en el sentido del error cuadrático medio, es igualmente del orden de  $n^{-4/5}$ .

También se debe notar que la fórmula (6) no cambia cuando las observaciones están en  $\mathbf{R}^s$  (*i.e.*  $(X, Y)$  esté en  $\mathbf{R}^s \times \mathbf{R}$ ), con la única diferencia que  $t$ , y los  $x_i$  están en  $\mathbf{R}^s$ . Para mayores detalles, ver Härdle (1989).

### 3 Previsión no paramétrica

#### 3.1 Relación entre regresión y previsión

Se considera un proceso estacionario  $(X_t)_{t \in \mathbf{Z}}$ . Se supone haber observado  $X_1, \dots, X_T$ . Se trata de predecir  $X_{T+k}$  (con  $k \in \mathbf{N}^*$ ) a partir de las variables observadas. Un predictor natural de  $X_{T+k}$  basado sobre  $X_1, \dots, X_T$  es dado por  $E(X_{T+k}|X_1, \dots, X_T)$ . Esta esperanza condicional es lamentablemente imposible de estimar correctamente si no se hacen hipótesis suplementarias sobre el proceso  $(X_t)_{t \in \mathbf{Z}}$ . En cambio, si se supone a priori que el proceso es  $r$ -Markoviano, entonces:

$$E(X_{T+k}|X_T, \dots, X_1) = E(X_{T+k}|X_T, \dots, X_{T-r+1}).$$

Se buscará entonces hacer la regresión de  $X_{T+k}$  sobre su propio pasado próximo  $X_T, \dots, X_{T-r+1}$ . Este procedimiento permite a menudo obtener una previsión razonable de  $X_{T+k}$ , aún si el proceso  $(X_t)_{t \in \mathbf{Z}}$  no es  $r$ -Markoviano. Se puede notar que en los métodos de alisamiento, como en la metodología de Box y Jenkins, se realiza usualmente un mismo truncamiento. Por ejemplo, cuando el predictor de  $X_{T+k}$  está basado en una representación autorregresiva infinita del tipo  $\left(X_t = \sum_{i=1}^{\infty} a_i X_{t-i}\right)$ , entonces es calculado efectivamente como una suma finita  $\left(\sum_{i=1}^r \hat{a}_i X_{t-i}\right)$  con los  $\hat{a}_i$  estimados a partir de los datos  $X_1, \dots, X_T$ . Más adelante, indicaremos cómo identificar el mejor  $r$  posible.

#### 3.2 Escritura del predictor con núcleo

Se debe notar que el problema anterior de regresión tiene por caso particular el de la previsión, donde se trata de explicar  $X_{T+k}$  en función de  $X_T, \dots, X_{T-r+1}$ . El predictor con núcleo se escribe entonces:

$$\hat{L}_{T,k}^{(r)} = \frac{\sum_{t=r}^{T-k} \mathbf{K}[(X_T^{(r)} - X_t^{(r)})/h(T)] X_{t+k}}{\sum_{t=r}^{T-k} \mathbf{K}[(X_T^{(r)} - X_t^{(r)})/h(T)]} \quad (7)$$

con  $X_T^{(r)} = (X_T, \dots, X_{T-r+1})$  y  $X_t^{(r)} = (X_t, \dots, X_{t-r+1})$ . Lo que se escribe también:

$$\hat{L}_{T,k}^{(r)} = \sum_{t=r}^{T-k} \hat{\alpha}_{t,T}^{(r)} X_{t+k}, \quad (8)$$

donde

$$\hat{\alpha}_{t,T}^{(r)} = \frac{\mathbf{K}[(X_T^{(r)} - X_t^{(r)})/h(T)]}{\sum_{t=r}^{T-k} \mathbf{K}[(X_T^{(r)} - X_t^{(r)})/h(T)]}. \quad (9)$$

La escritura (8) muestra que el predictor no paramétrico es típicamente un alisador. En efecto, (8) es una media ponderada de los valores pasados. La diferencia fundamental es que aquí  $\hat{\alpha}_{t,T}^{(r)}$  es aleatoria, y no determinista como en el caso de los alisamientos exponenciales.

### 3.3 Interpretación

Si se examina la figura 2, donde el parámetro  $r$  ha sido fijado en 3, se puede notar que las fórmulas (8) y (9), útiles en el cálculo de las previsiones, sirven para calcular en cada punto  $t$  del pasado del proceso un peso (ver (9))  $\hat{\alpha}_{t,T}^{(r)}$ . La previsión en el horizonte 1 se calcula entonces haciendo la media de los  $X_{t+1}$  con las ponderaciones  $\hat{\alpha}_{t,T}^{(r)}$ . La previsión en el horizonte 2 se calcula de forma similar haciendo la media de los  $X_{t+2}$  con las mismas ponderaciones  $\hat{\alpha}_{t,T}^{(r)}$ . Esto se reduce a buscar en el propio pasado del proceso, las secuencias que se parecen más a la última  $X_T^{(r)} = (X_T, \dots, X_{T-r+1})$ . Todas las secuencias  $X_t^{(r)} = (X_t, \dots, X_{t-r+1})$  juegan un papel, por supuesto, pero éste es más importante cuanto mayor sea la similitud entre  $X_t^{(r)} = (X_t, \dots, X_{t-r+1})$  y  $X_T^{(r)} = (X_T, \dots, X_{T-r+1})$ .

### 3.4 Observaciones

Se puede notar que la previsión en el horizonte 2 no hace intervenir la previsión en el horizonte 1. De manera general, el cálculo de las previsiones sucesivas nunca hace intervenir las previsiones en los horizontes anteriores, al contrario de lo que pasa por ejemplo en la metodología de Box y Jenkins. Se deduce por supuesto una robustez más grande del método, por no acumulación de los errores anteriores.

Al principio de esta sección hicimos la hipótesis de estacionalidad. Si el proceso no fuera estacionario, primero habría que eliminar la tendencia mediante una diferenciación ad hoc. Esta estacionalidad es, por supuesto, necesaria, salvo en lo que concierne a una eventual estacionalidad. Una serie que presenta una estacionalidad aporta una información adicional importante y que interesa explotar para afinar nuestras previsiones. Intuitivamente, si se piensa en términos de búsqueda de similitud del pasado al presente, se ve bien que, si hay estacionalidad, nos interesa escoger un  $r$  que sea la longitud de una estacionalidad o de una semi-estación, por ejemplo.

Ahora, imaginemos en la serie un punto aberrante en  $t_0$ . Entonces, toda secuencia del pasado que contenga  $t_0$  tendrá una similitud casi nula con  $X_T^{(r)} = (X_T, \dots, X_{T-r+1})$ , y así no intervendrá en el cálculo de las previsiones. No se presenta ni siquiera el problema de saber si un punto es o no aberrante, o de acuerdo con qué criterio lo es o no, sino que el

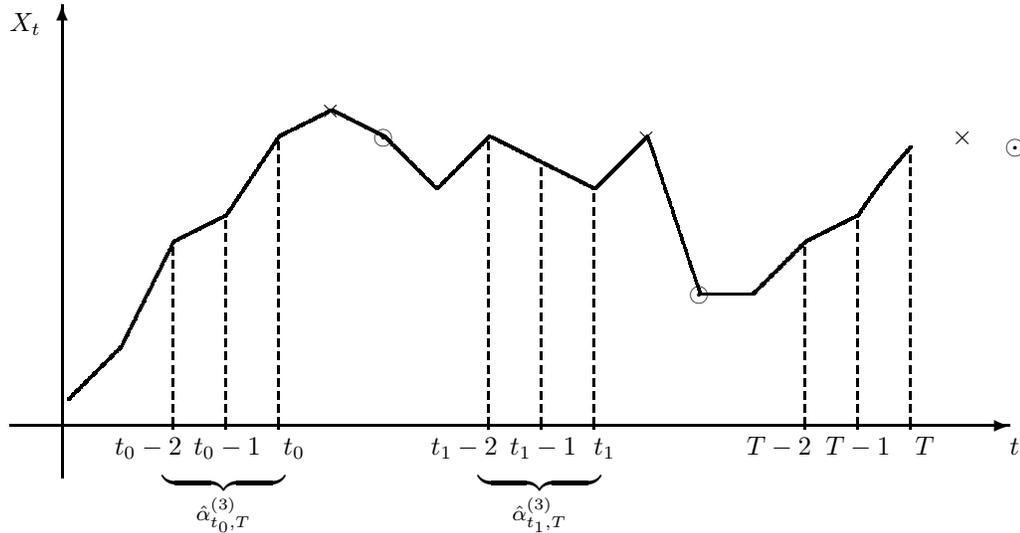


Figure 2: Interpretación del método de previsión no paramétrico: el peso  $\hat{\alpha}_{t_1, T}^{(3)}$  será más importante que el peso  $\hat{\alpha}_{t_0, T}^{(3)}$  porque la secuencia  $(X_{t_0-2}, X_{t_0-1}, X_{t_0})$  se parece más que  $(X_{t_1-2}, X_{t_1-1}, X_{t_1})$  a  $(X_{T-2}, X_{T-1}, X_T)$ . El cálculo de la previsión en el horizonte 1 (respectivamente el horizonte 2), que se materializa por el símbolo  $\times$  (resp.  $\odot$ ), hace intervenir los valores de  $X_{t_0+1}$  y de  $X_{t_1+1}$  (resp.  $X_{t_0+2}$  y  $X_{t_1+2}$ ) (igualmente indicados por el símbolo  $\times$  (resp.  $\odot$ )) ponderados por los pesos  $\hat{\alpha}_{t_0, T}^{(3)}$  y  $\hat{\alpha}_{t_1, T}^{(3)}$ .

mismo es eliminado de manera automática en los cálculos. Esta también es una ventaja notable de esta técnica.

Finalmente, se debe notar la facilidad de implementación informática, lo cual es una ventaja que no se debe despreciar.

### 3.5 Escogencia del parámetro de truncamiento $r$

Para un  $r$  fijo, se puede, para cada  $t$  ( $r+k \leq t \leq T-k$ ), calcular una previsión como se hizo anteriormente por:

$$\hat{X}_{t+k}^{(r)} = \sum_{i=r}^{t-k} \hat{\alpha}_{i,t}^{(r)} X_{i+k} \quad (10)$$

lo que da un error de previsión estimado:

$$e_t^k = |\hat{X}_{t+k}^{(r)} - X_{t+k}| \quad (11)$$

Parece entonces clásico usar un criterio de error cuadrático para estimar el  $r$  adecuado, es decir minimizar:

$$(T - 2k - r + 1)^{-1} \sum_{t=r+k}^{T-k} (e_t^k)^2 \quad (12)$$

Desde luego, se fija un valor máximo  $r_0$  y se utiliza entonces (12) para estimar  $r$ , con  $0 < r \leq r_0$ . Por supuesto, se debe notar que las primeras previsiones calculadas por (10) son pobres en información pues no dependen más que de pocas observaciones. Luego, no se deben comenzar las previsiones más que a partir de una fecha razonable  $t_0$ . (ver Carbon-Delecroix 1993 para de precisiones sobre esta validación).

### 3.6 Intervalo de previsión

Se usarán de nuevo los errores de previsión (ver (11))  $e_t^k$  para  $t = r + k, \dots, T - k$ . Se define  $\hat{q}^k$  el cuantil empírico a 95% (razonamiento idéntico para otros niveles) basado en los  $e_t^k$ . Para ello denotamos  $N_{t_0}^k$  el número de  $e_t^k$  mayores que  $e_{t_0}^k$  y  $t_0$  un índice tal que  $N_{t_0}^k$  sea el mayor entero que verifica

$$\frac{1}{T - 2k - r + 1} N_{t_0}^k \leq 0,05.$$

Entonces, se tiene

$$\hat{q}^k = e_{t_0}^k,$$

lo que da una estimación del intervalo de previsión (à 95 %) para  $X_{T+k}$  :

$$[\hat{X}_{T+k} - \hat{q}^k, \hat{X}_{T+k} + \hat{q}^k]$$

Se puede notar que esta manera de estimar un intervalo de previsión no hace ninguna hipótesis sobre la forma de las leyes subyacentes (Box y Jenkins hacen la hipótesis de normalidad). Además, si hay puntos aberrantes, se encuentran en el 5% restante y no intervienen tampoco aquí en el cálculo del intervalo de previsión.

## 4 Ejemplos y comparaciones

Estudiamos ahora algunas series temporales. Para cada una de ellas, vamos a comparar la eficacia del método no paramétrico respecto al método de Box y Jenkins. Estas series son, o bien simuladas o bien extraídas de la literatura. Para las series simuladas, que son modelos ARMA, por supuesto hemos retenido la identificación óptima para el método de Box y Jenkins. Para las otras series, tomadas de la literatura, hemos retenido las identificaciones indicadas por los autores, para evitar toda ambigüedad. Para cada una de ellas, hemos truncado la serie de manera que se tenga una comparación de las previsiones respecto a las verdaderas observaciones. En las figuras 3 a 14 el gráfico (a) representa los datos utilizados para hacer el conjunto de las previsiones. El gráfico (b) permite comparar las previsiones paramétricas (representadas por los símbolos  $\times$  unidos por guiones), las previsiones no paramétricas (representados por las símbolos  $+$  unidos por un trazo punteado), y los

verdaderos valores (representados como en (a)). Los intervalos de confianza para las previsiones paramétricas y no paramétricas son respectivamente delimitadas por guiones y por puntos.

Hemos definido dos criterios de comparación. Uno se llama *EMO* (error medio observado) definido por:

$$EMO = K^{-1} \sum_{i=T-K+1}^T (|X_i - \hat{X}_i|/|X_i|)$$

donde  $K$  denota el número máximo de previsiones calculadas.

Si se denota por  $\hat{q}_i$  el cuantil estimado del cuantil teórico  $q_i$  definido por:

$$P(|\hat{X}_i - X_i| < q_i) = 0,95,$$

entonces, se define el criterio *EMP* (error medio previsión) por :

$$EMP = K^{-1} \sum_{i=T-K+1}^T (\hat{q}_i/|\hat{X}_i|)$$

Para los dos métodos (no paramétrico y Box-Jenkins), con anterioridad hemos estacionalizado la serie de la misma manera. Desde luego, en los casos en que hay estacionalidad, no se trata preliminarmente por el método no paramétrico.

En las diferentes figuras que siguen, constatamos que :

1. En la mayoría de las simulaciones de modelos ARMA (figures 3, 6, 7, 10 y 14), modelos bien adaptados a la metodología de Box y Jenkins, los resultados obtenidos por el método no paramétrico son muy cercanos de los obtenidos por el método de Box y Jenkins.
2. Para ciertas series se nota a veces un mejor comportamiento de las previsiones no paramétricas (ver por ejemplo la figura 11).
3. El intervalo de previsión en el caso no paramétrico se extiende menos rápidamente que en el caso Box y Jenkins, mostrando la mayor robustez del método (ver por ejemplo la figura 9).

Para otros ejemplos, ver Carbon y Delecroix (1993).

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	0.090%	0.136 %
No paramétrica	0.098%	0.130%

(c)

Figure 3: Comparación de las previsiones paramétricas ( $\times$ ) y no paramétricas (+) para el AR(1)  $X_t = 0.9X_{t-1} + 10000 + \epsilon_t$ , donde  $(\epsilon_t)$  es un ruido blanco fuerte  $\mathcal{N}(0, 5)$  (ver la sección 4 para la descripción de los gráficos)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	3.83%	23.63%
No paramétrica	2.86%	20.29%

(c)

Figure 4: Comparación de las previsiones paramétricas y no paramétricas para la producción de carbón (cf. Pankratz, 1983)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	3.95%	8.85%
No paramétrica	5.77%	8.85%

(c)

Figure 5: Comparación de las previsiones paramétricas y no paramétricas para el tráfico de viajeros (cf. Gouriéroux, 1990)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	1.15	2.04
No paramétrica	0.94	2.93

(c)

Figure 6: Comparación de las previsiones paramétricas y no paramétricas sobre una simulación de la serie  $X_t = \epsilon_t + 0.2X_{t-1} + X_{t-2} - 0.3X_{t-3}$ , donde  $(\epsilon_t)$  es una sucesión i.i.d. de ley uniforme sobre  $[-49, 49]$

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	3.27	2.25
No paramétrica	3.33	2.14

(c)

Figure 7: Comparación de las previsiones paramétricas y no paramétricas sobre una simulación de la serie  $X_t = \epsilon_t + X_{t-1} - 0.1X_{t-2}$ , donde  $(\epsilon_t)$  es una sucesión i.i.d. de ley uniforme sobre  $[-49, 49]$

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	0.33	0.43
No paramétrica	0.33	0.47

(c)

Figure 8: Comparación de las previsiones paramétricas y no paramétricas sobre datos de un proceso químico (cf. Box y Jenkins, 1970)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	2.72%	8.67%
No paramétrica	2.61%	5.11%

(c)

Figure 9: Comparación de las previsiones paramétricas y no paramétricas sobre datos de un proceso químico de concentración (cf. Box y Jenkins, 1970)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	1.55	2.18
No paramétrica	1.71	2.46

(c)

Figure 10: Comparación de las previsiones paramétricas y no paramétricas sobre una simulación de la serie  $X_t = \epsilon_t + 0.8X_{t-1}$ , donde  $(\epsilon_t)$  es una sucesión i.i.d. de ley exponencial de parámetro 2.5.

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	0.25	1.28
No paramétrica	0.15	0.63

(c)

Figure 11: Comparación de las previsiones paramétricas y no paramétricas sobre una simulación de la serie  $X_t = \epsilon_t + 3000 \sin(\frac{\pi}{15}t)$ , donde  $(\epsilon_t)$  es una sucesión i.i.d. de ley exponencial de parámetro  $\frac{1}{300}$ .

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	0.37	1.57
No paramétrica	0.45	0.84

(c)

Figure 12: Comparación de las previsiones paramétricas y no paramétricas sobre la serie *change in business inventories* (cf. Prankratz, 1983)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	7.36%	31.00%
No paramétrica	18.10%	27.58%

(c)

Figure 13: Comparación de las previsiones paramétricas y no paramétricas sobre los permisos de construir (cf. Prankratz, 1983)

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	0.87%	2.45%
No paramétrica	0.95%	2.89%

(c)

Figure 14: Comparación de las previsiones paramétricas y no paramétricas sobre la media móvil  $X_t = 100 + \epsilon_t + 0.8\epsilon_{t-1}$ , donde  $(\epsilon_t)$  es un ruido blanco gaussiano  $\mathcal{N}(0, 1)$

	<i>EMO</i>	<i>EMP</i>
Box y Jenkins	0.54%	22.12%
No paramétrica	0.75%	24.72%

(c)

Figure 15: Comparación de las previsiones paramétricas y no paramétricas sobre el consumo de cigarrillos (cf. Prankratz, 1983)

## References

- [1] Bosq, D. (1973) “Sur l’estimation de la densité d’un processus stationnaire et mélangeant”, *C.R. Acad. Sci. Paris* **277**: 535–538.
- [2] Bosq, D. (1975) “Inégalité de Bernstein pour les processus stationnaires et mélangeants. Applications”, *C.R. Acad. Sci. Paris* **281**: 1095–1098.
- [3] Bosq, D.; Delecroix, M. (1985) “Nonparametric prediction of a Hilbert space valued random variable”, *Stoch. Proc. and their Appl.* **19**: 271–280.
- [4] Bosq, D.; Lecoutre, J.P. (1987) *Théorie de l’Estimation Fonctionnelle*. Economica, Paris.
- [5] Bosq, D.; Lecoutre, J.P. (1992) *Analyse et Prévision des Séries Chronologiques*. Masson, Paris.
- [6] Box, G.; Jenkins, G. (1970) *Time Series Analysis, Forecasting and Control*. Holden Day.
- [7] Carbon, M. (1982) *Sur l’Estimation Asymptotique d’une Classe de Paramètres Fonctionnels pour un Processus Stationnaire*. Thèse de l’Université de Lille.
- [8] Carbon, M. (1983) “Inégalité de Bernstein pour les processus fortement mélangeants non nécessairement stationnaires”, *C. R. Acad. Sci. Paris* **297**: 303–306.
- [9] Carbon, M. (1988) *Inégalités de Grandes Déviations dans les Processus. Applications à l’Estimation Fonctionnelle*. Thèse de Doctorat de l’Université de Paris 6.
- [10] Carbon, M.; Delecroix, M. (1993) “Nonparametric vs. parametric forecasting in time series: a computational point of view”, *Appl. Stoch. Models and Data Anal.* **9**: 215–229.
- [11] Collomb, G. (1981) “Estimation non paramétrique de la régression: revue bibliographique”, *Intern. Stat. Review* **49**: 75–93.
- [12] Collomb, G. (1983) *From Nonparametric Regression to Nonparametric Prediction: Survey on the Mean Square Error and Original Results on the Predictogram*. Lecture Notes in Statistics, **16**, Springer: 182–204.
- [13] Collomb, G. (1984) “Propriétés de convergence presque complète du prédicteur à noyau”, *Z. Wahrsch. Verw. Gebiete* **66**: 441–460.
- [14] Collomb, G. (1985a) “Nonparametric regression: an up to date bibliography”, *Statistics* **2**: 309–324.
- [15] Collomb, G. (1985b) “Nonparametric time series analysis and prediction: uniform almost sure convergence”, *Statistics* **2**: 297–307.

- [16] Deheuvels, P. (1977) “Estimation non paramétrique de la densité par des histogrammes généralisés”, *Rev. Stat. Appl.*: **35**: 5–42.
- [17] Devroye, L.; Györfi, L. (1985) *Nonparametric Density Estimation: the  $L_1$  View*. John Wiley & Sons, New York.
- [18] Doukhan, P. (1991) “Mixing. Properties and examples”. Preprint. Univ. Paris, Orsay.
- [19] Gouriéroux, C.; Monfort, A. (1990) *Séries Temporelles et Modèles Dynamiques*. Economica, Paris.
- [20] Hall, P. (1983) “Large sample optimality of least squared cross-validation in density estimation”, *Ann. of Stat.* **11**: 1156–1174.
- [21] Hall, P. (1984) “Asymptotic properties of integrated squared errors and cross-validation for kernel estimation of a regression function”, *Z. Wahrsch. Verw. Gebiete* **67**: 175–196.
- [22] Härdle, W. (1989) *Applied Nonparametric Regression*. Economic Society Monograph Series, Cambridge University Press.
- [23] Härdle, W.; Marron, J.S. (1985) “Optimal bandwidth selection in nonparametric regression estimation”, *Annals of Statistics* **13**(4): 1465–1481.
- [24] Marron, J.S. (1987) “A comparison of cross-validation techniques in density estimation”, *Ann. of Statist.* **15**: 152–162.
- [25] Pankratz, A. (1983) *Forecasting with Univariate Box-Jenkins Models: Concept and Cases*. John Wiley & Sons, New York.
- [26] Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation*. Academic Press, New York.
- [27] Stone, C.J. (1982) “Optimal global rates of convergence of non parametric regression”, *Ann. of Statist.* **10**: 1040–1053.