Asymptotic normality of the Nadaraya-Watson estimator for non-stationary functional data and applications to telecommunications.

L. ASPIROT∗, K. BERTIN†, G. PERERA∗
† Departamento de Estadística, CIMFAV, Universidad de Valparaíso, Chile
∗ IMERL, Facultad de Ingeniería, Universidad de la República, Uruguay.

September 10, 2008

Abstract

We study a non-parametric regression model where the explanatory variable is a non-stationary dependent functional data and the response variable is scalar. Supposing that the explanatory variable is a non-stationary mixture of stationary processes and general conditions of dependence of the observations (implied in particular by weakly dependence), we obtain the asymptotic normality of the Nadaraya-Watson estimator. Under some additional regularity assumptions on the regression function, we obtain asymptotic confidence intervals for the regression function. We apply this result to estimate the quality of service for an end-to-end connection on a network. Current network traffic leads to non-stationary and dependent data that could be treated using the model described here.

Keywords: Non-parametric regression, functional data, asymptotic normality, non-stationarity, dependence, quality of service.


1 Introduction

Nowadays new services like voice and video are offered over the Internet. For these new applications the need to measure the network performance has increased. Multimedia applications have several requirements in terms of delay, losses and other quality of service parameters. These constraints are stronger than constraints for applications like mail or ftp. Measurements of performance are necessary for different reasons, for example to advance in understanding the behavior of the Internet or to verify the quality of service assured to the new services. Our

∗Instituto de Matemática y Estadística “Rafael Laguardia” (IMERL), Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, Montevideo 11200, Tel-fax: (5982)7110621, Emails: aspirot@fing.edu.uy, perera@fing.edu.uy
†Departamento de Estadística, CIMFAV, Universidad de Valparaíso, Gran Bretaña 1091, 4to piso, Playa Ancha, Valparaíso, Chile, Tel: (56)322508324, Email: karine.bertin@uv.cl
objective in this paper is to propose a model that can be used to measure the quality of service of these new applications in internet.

We consider the following model in which we study the problem of estimating a non-parametric regression function \( \phi \) when the explanatory variable is functional and may be non-stationary and dependent. We observe \( \{X,Y\} = \{(X_i,Y_i) : i = 0,\ldots, n-1\} \) such that

\[
Y_i = \phi(X_i) + \varepsilon_i, \quad i = 0,\ldots, n-1, \tag{1}
\]

where \( \phi \) is a function \( \phi : \mathcal{D} \to \mathbb{R} \), \( X_i \in \mathcal{D} \), \( Y_i \in \mathbb{R} \), the \( \varepsilon_i \)'s are centered and independent of the \( X_i \)'s and \( \mathcal{D} \) is a semi-normed linear space with semi-norm \( \| \cdot \| \). Here \( Y \) will represent the performance of a service such voice or video and we suppose that it depends on the traffic \( X \) in the Internet.

Our aim is to find the asymptotic distribution of the estimator

\[
\hat{\phi}_n(x) = \frac{\sum_{i=1}^{n-1} Y_i K \left( \frac{\|x - X_i\|}{h_n} \right)}{\sum_{i=1}^{n-1} K \left( \frac{\|x - X_i\|}{h_n} \right)} \tag{2}
\]

depending on the traffic \( X \) in the Internet.

The estimator (2) is a generalization of the Nadaraya-Watson estimator (cf. [1]). The asymptotic normality of the Nadaraya-Watson estimator is proved in [2] when the observations \( (X_i,Y_i) \) are independent, and for dependent observations, in [3].

Ferraty, Goia and Vieu ([4], [5]) proposed the estimator (2) to estimate the regression function when the explanatory variable is functional. For weakly dependent and stationary observations, they proved the complete convergence of (2) and obtained rates of convergence when the function \( \phi \) satisfies some regularity conditions. Aspirot et al. (2005) ([6]) have generalized the result of complete convergence for the estimator (2) to a non-stationary case. Masry (2005) ([7]) has proved the asymptotic normality of \( \hat{\phi}_n \) for stationary weakly dependent random variables. Ferraty, Mas and Vieu (2007) ([8]) consider asymptotic normality of the estimator \( \hat{\phi}_n \) in the case of independence from theoretical and practical point of view. Related work on density estimation with functional data can be found in [9] and [10] for mode estimation and in [11] for asymptotic normality of kernel estimators. Surveys on functional data analysis can be found in Ramsay and Silverman (2005) ([12]) and in Ferraty and Vieu (2006) ([13]) for the non-parametric case.

Our goal in this article is to generalize the result of [7] to the case when \( X \) is a non-stationary mixture of stationary processes using some tools presented previously in [14], [15], [16] and [17]. More precisely, we suppose that there exist two random processes \( \xi = \{\xi_n : n \in \mathbb{N}\} \), \( Z = \{Z_n : n \in \mathbb{N}\} \) and a function \( \varphi : \mathcal{D} \times \mathbb{R} \to \mathcal{D} \) such that \( \xi \) is stationary and weakly dependent with values in \( \mathcal{D} \), \( \xi \) is independent of \( Z \), \( Z \) takes values in a finite set \( \{z_1,\ldots, z_m\} \) of \( \mathbb{R} \) and \( X \) satisfies

\[
X_i = \varphi(\xi_i, Z_i), \quad i = 0,\ldots, n-1. \tag{3}
\]

The function \( \phi \) can be very general and serves to express the variable \( X \) as a mixture of stationary processes.
We suppose general conditions of dependence (cf. assumptions in Subsection 3.1), implied in particular by weakly dependence of the observations and of the process ξ, but the process Z may be non-stationary and non-weakly dependent. In Theorem 1, under these conditions and a set of technical assumptions, we obtain the asymptotic normality of the estimator (2). This result is based on central limit theorem for triangular arrays of stationary random fields in \( \mathbb{Z}^d \) [17], where the results depend on the geometry of the index set considered. In Proposition 1 and 2, we present these results of central limit theorems for triangular arrays of a \( \mathbb{R}^m \)-valued centered stationarity random field indexed in \( \mathbb{N} \) or \( \mathbb{Z}^d \). We apply these results to our model, in which we have a field \( X = \varphi(\xi, Z) \) where Z is non-stationary. We consider here our observations indexed in the level sets of Z. Some assumptions about Z should be made in order to ensure properties of the level sets and to have a central limit theorem in this case. In theorem 2, supposing additionally some regularity assumptions on the regression function \( \varphi \) and on the bandwidth \( h_n \), we obtain the asymptotic normality of \( \hat{\varphi}_n(x) - \varphi(x) \) for \( x \in D \), which allows us to construct asymptotic confidence interval for \( \varphi(x) \).

The non-stationary assumption for the data appears naturally when modelling the Internet traffic (cf. [18], [19], [20]). In this paper, we apply our models defined by (1) and (3) to estimate the quality of service in the Internet. The quality of service parameters correspond to some network characteristics like delay, packet loss, etc.. The interest in voice or video applications is based in the fact that these network traffic (the same as many real time applications) have high quality of service requirements (very low delay and losses, etc.). In Section 4, we use the model (3) to estimate quality of service parameters. These quality of service parameters are represented by the variable \( Y \) and we assume that they are a function of the Internet traffic, represented by \( X = \varphi(\xi, Z) \). In the literature, it is usual to assume that the traffic is piecewise stationary, in this model Z is a non-stationary random process that selects between different stationary behaviors or state of the traffic. The process Z indicates for example if the network is very much loaded or not and it could describe seasonal or periodic behaviors. The process ξ is related to the variations of the traffic within a particular state for Z. We estimate quality of service parameters for voice or video traffic when this traffic goes between two nodes over the network. Our procedure measurement is detailed in Subsection 4.2. We give simulations of the estimator (2) and confidence interval, for traffic simulated in Laboratory.

The organization of the paper is as follows. In Section 2, we give some preliminaries on the notions of asymptotically measurable sets and fields and state a central limit theorem for triangular arrays of \( \mathbb{R}^m \)-valued centered stationary random fields. In Section 3, we give our result on the asymptotic normality of the estimator \( \hat{\varphi}_n \) and we make the assumptions under which we obtain this result. Section 4 describes the application of our result to the problem of estimating the quality of service on the Internet. Section 5 is devoted to the proofs.

## 2 Preliminaries

In this section we present some preliminary results. treated in , and [14], [15] and [16]. In subsection 2.1, we define the notion of asymptotically measurable subsets originally introduced in [14], [15] and [16]. This allows, in subsection 2.2, to give a central limit theorem for triangular arrays of \( \mathbb{R}^m \)-valued centered stationary random fields. In subsection 2.3 we present the notion of asymptotically measurable field that allows to state a central limit theorem for a field described
by equation (3). In this section, we consider both the case when the field is indexed in \( \mathbb{N} \) and in \( \mathbb{Z}^d \) and in the following, \( L \) stands for \( \mathbb{N} \) or \( \mathbb{Z}^d \). When \( L = \mathbb{N} \), we denote \( \ell(n) = n \) and when \( L = \mathbb{Z}^d \), we denote \( \ell(n) = (2n + 1)^d \).

2.1 Notion of asymptotically measurable subsets

In what follows \( \text{card}(A) \) indicates the cardinal of subset \( A \).

Definition 1. A subset \( A \subset L \) is said to be an asymptotically measurable subset if, for all \( k \in L \), the following limit exists

\[
F(k, A) = \lim_{n \to \infty} \frac{\text{card}\{A_n \cap (A_n - k)\}}{\ell(n)}.
\]

Definition 2. The subset family \{ \( A^i : i = 1, \ldots, m \) \} in \( L \) is an asymptotically measurable family if, for all \( i \in \{1, \ldots, m\} \), the subsets \( A^i \) are asymptotically measurable and, for all \( k \in L \) and \( i, j \in \{1, \ldots, m\} \), the following limit exists

\[
F(k, A^i, A^j) = \lim_{n \to \infty} \frac{\text{card}\{A_n \cap (A_n^i - k)\}}{\ell(n)}.
\]

2.2 Central limit theorem for triangular arrays of \( \mathbb{R}^m \)-valued centered stationary random field

In the following, the notation \( \Rightarrow \) means convergence in law and \( N_d(0, \Sigma) \) represents a zero mean normal distribution in dimension \( d \) with covariance matrix \( \Sigma \). Moreover, for \( X^n = (X^n_k)_{k \in L} \) a \( \mathbb{R}^m \)-valued centered stationary random field, we denote the components of \( X^n \) by \( X^n = (X^{1,n}, \ldots, X^{m,n}) \) and \( X^n_k = (X^{1,n}_k, \ldots, X^{m,n}_k) \) for \( k \in L \).

Definition 3. We define the class \( B(L) \) as the class of \( \mathbb{R}^m \)-valued centered stationary random fields \( X^n = (X^n_k)_{k \in L} \) that satisfies the following conditions:

\((H_1)\) For all \( i, j \in \{1, \ldots, m\} \) and \( n \in \mathbb{N} \), we have

\[
\sum_{k \in L} \left| E \left\{ X_{0,n}^i X_{k,n}^j \right\} \right| < \infty.
\]

\((H_2)\) Let \( X^{n,J} = (X^{1,n,J}, \ldots, X^{m,n,J}) \) be the truncation by \( J \) of \( X^n \), defined for \( k \in L \) and \( J > 0 \) by

\[
X^{n,J}_k = X^n_k 1_{\{\|X^n_k\| \leq J\}} - E \left[ X^n_k 1_{\{\|X^n_k\| \leq J\}} \right],
\]

where \( \| \cdot \| \) represents the euclidian norm on \( \mathbb{R}^m \).

(i) There exists a sequence \( \gamma(k) \geq 0 \) such that \( \sum_{k \in L} \gamma(k) < \infty \) and such that for all \( k \in L, n \in \mathbb{N}, i, j \in \{1, \ldots, m\} \) and \( J > 0 \) we have

\[
\left| E \left\{ X_{0,n}^{i,J} X_{k,n}^{j,J} \right\} \right| \leq \gamma(k).
\]
Asymptotic normality of the Nadaraya-Watson estimator for non-stationary data

(ii) There exists a sequence \( b(J) \) such that \( \lim_{J \to \infty} b(J) = 0 \) and for all set \( B \subset \mathbb{L}, i \in \{1, \ldots, m\}, n \in \mathbb{N} \) and \( J > 0 \)

\[
E \left[ \left( S_n(B, X_i^{n, J}) - S_n(B, X_i^{n, J}) \right)^2 \right] \leq \frac{b(J)\text{card}(B_n)}{f(n)},
\]

where

\[
S_n(B, X_i^{n, J}) = \frac{1}{f(n)^{1/2}} \sum_{k \in B_n} X_i^{n, J}.
\]

(\( H_3 \)) There exists a sequence of reals numbers \( C(J) \) such that for all \( B \subset \mathbb{L}, n \in \mathbb{N}, J > 0 \) and \( i \in \{1, \ldots, m\} \) we have

\[
E \left[ S_n(B, X_i^{n, J})^4 \right] \leq C(J) \left( \text{card}(B_n) \right)^2.
\]

(\( H_4 \)) There exist a function \( h : \mathbb{R}^+ \to \mathbb{R}^+ \), such that \( \lim_{x \to +\infty} h(x) = 0 \) and a function \( g : \mathbb{R}^+ \times \mathbb{R}^m \to \mathbb{R}^+ \), with \( g(J, t) < \infty \) for all \( J > 0 \) and \( \sup_{t \in \mathbb{R}^m} g(J, t) = g_J < \infty \), such that

\[
\left| E \left[ e^{iS_n(B \cup C, (t, X^{n, J}))} \right] - E \left[ e^{iS_n(B, (t, X^{n, J}))} \right] E \left[ e^{iS_n(C, (t, X^{n, J}))} \right] \right| \leq g(J, t)h(d(B, C)),
\]

for all disjoint sets \( B, C \subset \mathbb{L}, n \in \mathbb{N}, J > 0 \) and \( t \in \mathbb{R}^m \). Here \( \langle \cdot, \cdot \rangle \) represents the scalar product on \( \mathbb{R}^m \).

(\( H_5 \)) There exist sequences \( \gamma^J(i, j, k) \) and \( \gamma(i, j, k), i, j \in \{1, \ldots, m\}, k \in \mathbb{L} \) such that for all \( J > 0 \) we have

\[
\lim_{n \to \infty} E \left\{ X_0^{i, n, J}X_k^{i, n, J} \right\} = \gamma^J(i, j, k),
\]

and

\[
\lim_{J \to \infty} \gamma^J(i, j, k) = \gamma(i, j, k)
\]

for \( i, j \in \{1, \ldots, m\} \) and \( k \in \mathbb{L} \).

We have the two following propositions proved in Section 5. These results generalize results on triangular arrays for \( \mathbb{R} \)-valued random fields obtained in [17] to the case of \( \mathbb{R}^m \)-valued random fields.

**Proposition 1.** If \( X^n = (X_k^n)_{k \in \mathbb{N}} \) belongs to \( B(\mathbb{N}) \), then for any asymptotically measurable family \( \{ A^i : i = 1, \ldots, m \} \) in \( \mathbb{N} \), we have as \( n \) tends to \( \infty \)

\[
(S_n(A^1, X_{1}^{n}), \ldots, S_n(A^m, X_{m}^{n})) \wrightarrow N_m(0, \Sigma),
\]

where for \( i, j \in \{1, \ldots, m\} \)

\[
\Sigma(i, j) = \gamma(i, j, 0)F(i, j, 0) + \sum_{k \geq 1} \{ \gamma(i, j, k)F(i, j, k) + \gamma(j, i, k)F(j, i, k) \}
\]

and

\[
F(i, j, k) = \lim_{n \to \infty} \frac{\text{card}(A_n^i \cap (A_n^j - k))}{n}.
\]
Proposition 2. If \( X^n = (X^n_k)_{k \in \mathbb{Z}^d} \) belongs to \( B(\mathbb{Z}^d) \), then for any asymptotically measurable family \( \{ A^i : i = 1, \ldots, m \} \) in \( \mathbb{Z}^d \), we have as \( n \) tends to \( \infty \)
\[
(S_n(A^1, X^{1,n}), \ldots, S_n(A^m, X^{m,n})) \xrightarrow{w} N_m(0, \Sigma),
\]
where for \( i, j \in \{1, \ldots, m\} \)
\[
\Sigma(i, j) = \sum_{k \in L} \gamma(i, j, k) F(i, j, k)
\]
and
\[
F(i, j, k) = \lim_{n \to \infty} \frac{\text{card}\{A^i_n \cap (A^j_n - k)\}}{2n + 1}.
\]

2.3 Notion of asymptotically measurable field

In what follows \( \frac{a.s.}{n} \) means almost sure convergence as \( n \) tends to \( \infty \) and \( 1_A \) is the indicator function of a set \( A \). In the case \( L = \mathbb{N} \), we set \( D^{(n)} = \{0, \ldots, n - 1\} \) and if \( L = \mathbb{Z}^d \), we set \( D^{(n)} = \{-n, \ldots, n\}^d \).

Definition 4. The \( \mathbb{R}^m \)-valued random field \( Z = (Z_n)_{n \in L} \) is an asymptotically measurable field in \( L \) if there exists a random probability measure \( R_0 \) in \( B \), where \( B \) is the Borel \( \sigma \)-algebra in \( \mathbb{R}^m \), such that
\[
\frac{1}{\ell(n)} \sum_{m \in D^{(n)}} 1_{\{Z_m \in B\}} \xrightarrow{a.s. \over n} R_0(B)
\]
and for all \( k \in L - \{0\} \) there exists a random measure \( R_k \) in \( B_2 \), where \( B_2 \) is the Borel \( \sigma \)-algebra in \( \mathbb{R}^{2m} \), such that for all \( B, C \in B \)
\[
\frac{1}{\ell(n)} \sum_{m \in D^{(n)}} 1_{\{Z_m \in B\}} 1_{\{Z_{m-k} \in C\}} \xrightarrow{a.s. \over n} R_k(B \times C).
\]
If the limit measure is non-random \( Z \) is called a regular field in \( L \).

The following proposition (cf. Proof in [16]) relates asymptotically measurable and regular fields with asymptotically measurable subsets.

Proposition 3. If \( Z = (Z_n)_{n \in L} \) is an asymptotically measurable field, \( B^1, B^2, \ldots, B^m \) are disjoint subsets of \( B \) and \( A^i = \{n \in L : Z_n \in B^i\} \) for \( i = 1, \ldots, m \), then, conditionally to \( Z \) the family \( \{A^1, \ldots, A^m\} \) is an asymptotically measurable family with \( F(k, A^i, A^j) = R_k(B^i, B^j) \) and \( F(0, A^i, A^j) = R_0(B^i) \delta_{ij} \), where \( \delta_{ij} \) is the Kronecker delta.

Remarks: In the next sections we will assume further hypotheses on \( Z \) for the field described in (3) in order to prove the results, but the previous proposition is the tool for conditioning on level sets of the field \( Z \) and applying the results for stationary fields to the non-stationary case (3).
3 Results

In the following, we consider a fixed \( x \in D \) and we will study the asymptotic normality of \( \hat{\phi}_n(x) \) given by (2). Proofs of the results are in Section 5.

The estimator \( \hat{\phi}_n \) can be also written

\[
\hat{\phi}_n(x) = \frac{g_n(x)}{f_n(x)},
\]

where

\[
g_n(x) = \frac{1}{n\psi(h_n)} \sum_{i=0}^{n-1} Y_i K_n(X_i),
\]

\[
f_n(x) = \frac{1}{n\psi(h_n)} \sum_{i=0}^{n-1} K_n(X_i)
\]

with

\[
K_n(u) = K\left(\frac{\|u - x\|}{h_n}\right), \quad u \in D.
\]

and \( \psi(h_n) \) is given in assumption \((A_1)\). We recall that we use the convention \(0/0 = 0\).

In Subsection 3.1, we give the assumptions under which the asymptotic normality of \( \hat{\phi}_n(x) \) is obtained and in Subsection 3.2 we give our results. Subsection 3.3 gives comments on the results.

3.1 Assumptions

We make the following assumptions on the distribution of \((X,Y)\).

\((A_1)\) There exist positive functions \( \psi, \psi_1, \ldots, \psi_m \) defined on \( \mathbb{R}_+ \times D \), functions \( c_1, \ldots, c_k \) defined on \( D \) and a subset \( \Delta \) of \( \{1, \ldots, m\} \) such that for all \( h > 0 \)

\[
P \left[ \|\varphi(\xi_1, z_k) - x\| \leq h \right] = c_k(x)\psi_k(h, x),
\]

where \( \lim_{h \to 0} \psi_k(h, x) = 1 \) if \( k \in \Delta \), and \( \lim_{h \to 0} \psi_k(h, x) = 0 \) if \( k \in \Delta^C \).

In the following, to simplify the notation, we replace \( \psi(h_n, x) \) by \( \psi(h_n) \), but the quantity \( \psi(h_n) \) still depends of \( x \).

\((A_2)\) The functions \( u \mapsto \psi_k(u, x) \) are differentiable on \( \mathbb{R}_+ \), with differential function denoted \( \psi'_k(u, x) \) and satisfy

\[
\lim_{h \to 0} \frac{h}{\psi_k(h, x)} \int_0^1 K(u)\psi'_k(uh, x)du = d_k(x),
\]

where the \( d_k \)'s are functions defined on \( D \).
(A3) The process \( Z \) is regular and satisfies for \( k \in \{1, \ldots, m\} \) as \( n \) tends to \( \infty \) that

\[
\sqrt{n \psi(h_n)} \left( \frac{1}{n} \sum_{i=0}^{n-1} 1\{Z_i = z_k\} - \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k) \right) \xrightarrow{P} 0, \tag{7}
\]

where \( \Rightarrow \) means convergence in probability. For \( k \in \{1, \ldots, m\} \), denote by \( p_k \) the limit

\[
p_k = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k).
\]

Let the \( \mathbb{R}^{2m} \)-value random process \( \tilde{X}^n = (\tilde{X}^{1,n}, \ldots, \tilde{X}^{2m,n}) \) defined for \( i \in \mathbb{N} \) as follows:

- for \( l \in \{1, \ldots, m\} \)
  \[
  \tilde{X}^{l,n}_i = \frac{1}{\sqrt{\psi(h_n)}} K_n (\phi(\xi_i, z_l)) (\phi(\xi_i, z_l) + \varepsilon_i)
  - \frac{1}{\sqrt{\psi(h_n)}} E[K_n (\phi(\xi_i, z_l)) \phi(\xi_i, z_l)] ,
  \]

- for \( l \in \{m + 1, \ldots, 2m\} \)
  \[
  \tilde{X}^{l,n}_i = \frac{1}{\sqrt{\psi(h_n)}} K_n (\phi(\xi_i, z_l)) - \frac{1}{\sqrt{\psi(h_n)}} E[K_n (\phi(\xi_i, z_l))],
  \]

where \( K_n \) is defined by (6).

Moreover we assume the following hypotheses.

(A4) The \( \mathbb{R}^{2m} \)-value random process \( \tilde{X}^n \) belongs to \( B(\mathbb{N}) \).

(A5) The function \( \phi \) is continuous.

(A6) The estimator \( f_n(x) \) defined by (5) converges in probability to \( f(x) > 0 \) where the function \( f \) is defined by

\[
f(u) = \sum_{k \in \Delta} p_k d_k(u) c_k(u), \quad u \in \mathcal{D}.\]

We suppose the following hypothesis on the estimator \( \hat{\phi}_n \).

(A7) The function \( K \) is a positive function with support \([0, 1]\).

(A8) The bandwidth \( h_n \) satisfies

\[
\lim_{n \to \infty} h_n = 0
\]

and

\[
\lim_{n \to \infty} n \psi(h_n) = \infty.
\]
3.2 Asymptotic normality of $\hat{\phi}_n(x)$

The asymptotic normality of $\hat{\phi}_n(x)$ is obtained in three stages: asymptotic normality of the field $\hat{X}^n$ in Proposition 4, of $(g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x)))$ in Proposition 5 and then of $\hat{\phi}_n(x)$ in Theorem 1. Theorem 2 gives the asymptotic normality of $\hat{\phi}_n(x) - \phi(x)$ which allows to construct asymptotic confidence interval for $\phi(x)$.

**Proposition 4.** For $l \in \{1, \ldots, m\}$, denote $A^l = \{k \in \mathbb{N}, Z_k = z_l\}$ and $A^{l+m} = A^l$. If $Z$ is asymptotically measurable, then under Assumption (A4), conditionally on $Z$, we have, as $n$ tends to $\infty$

$$(S_n(A^1, \hat{X}^{1,n}), \ldots, S_n(A^{2m}, \hat{X}^{2m,n})) \overset{w}{\rightarrow} N_{2m}(0, \Sigma),$$

where for $i, j \in \{1, \ldots, 2m\}$

$$\Sigma(i, j) = \gamma(i, j, 0)F(i, j, 0) + \sum_{k \in \mathbb{N}} \{\gamma(i, j, k)F(i, j, k) + \gamma(j, i, k)F(j, i, k)\},$$

with $\gamma(i, j, k), k \in \mathbb{N}$, defined by (4) for the field $\hat{X}^n$ and

$$F(i, j, k) = \lim_{n \rightarrow \infty} \frac{\text{card}\{A^{i}_n \cap (A^{j}_n - k)\}}{n}.$$

**Proposition 5.** Under Assumptions (A3) and (A4), we have, as $n$ tends to $\infty$, that

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x)) \right) \overset{w}{\rightarrow} N(0, A),$$

where

$$A = \begin{pmatrix} a_1(x) & a_2(x) \\ a_2(x) & a_3(x) \end{pmatrix}$$

with

$$a_1(x) = \sum_{i,j=1}^{m} \Sigma(i, j), \quad a_2(x) = \sum_{i=1}^{m} \sum_{j=m+1}^{2m} \Sigma(i, j), \quad a_3(x) = \sum_{i,j=m+1}^{2m} \Sigma(i, j).$$

**Theorem 1.** Under Assumptions (A1)–(A8), we have as $n$ tends to $\infty$

$$\sqrt{n\psi(h_n)} \left( \hat{\phi}_n(x) - E(g_n(x)) \right) \overset{w}{\rightarrow} N(0, \sigma^2(x)),$$

where

$$\sigma^2(x) = \frac{a_1(x) - 2a_2(x)\phi(x) + a_3(x)\phi^2(x)}{f^2(x)}.$$

**Theorem 2.** We suppose that

(B1) the function $\phi$ satisfies

$$|\phi(u) - \phi(v)| \leq L\|u - v\|^\beta,$$

with $\beta > 0$ and $L$ a positive constant,

(B2) the bandwidth $h_n$ satisfies

$$\lim_{n \rightarrow \infty} h_n^\beta \sqrt{n\psi(h_n)} = 0.$$
Then under Assumptions $(A_1)$–$(A_8)$, $(B_1)$ and $(B_2)$, we have as $n$ tends to $\infty$,

$$\sqrt{n\psi(h_n)} \left( \hat{\phi}_n(x) - \phi(x) \right) \xrightarrow{w} N_1(0, \sigma^2(x)),$$

where $\sigma^2(x)$ is defined by (8).

### 3.3 Comments on the results

1- 

2- The conditions $(A_1)$ and $(A_2)$ mean that, conditionally on $Z$, the distance between the observation $X_i$ and the estimation point $x$ has a density. The quantity $\psi_k(h, x)$ plays a role analogous to $h^d$ for multidimensional estimation. The hypothesis $(A_3)$ means that the variable $Z$ does not behave too irregularly. It is assumed that the process has some kind of “stationarity in mean”. However this is not a strong assumption, and for example $(A_3)$ could be satisfied by periodic or semi-periodic random variables. The relation (7) is satisfied for example if the variables $1_{\{Z_i = z_k\}}$ satisfy a central limit theorem for each $k = 1, \ldots, m$. In the hypothesis $(A_6)$, $f(x) > 0$ means roughly that there is a positive probability of finding observations $X_i$ near (in the sense of the semi-norm on $D$) the point $x$ where we want to compute the estimator.” The function $f$ play the role of the density of the whole process $X = \varphi(\xi, Z)$. Moreover the convergence in probability can be obtained under the conditions $(A_1)$ y $(A_2)$, conditions on the moments of $Y_i$ and conditions of mixing (cf. Proof in [6]). The hypotheses $(A_5)$, $(A_7)$ and $(A_8)$ are classical for obtaining asymptotic normality.

3- The hypothesis $(A_4)$ can be replaced by various sets of hypothesis doing a trade-off between conditions on the moments of $Y$ and conditions of mixing. Perera (1997) ([15]) describes several sets of conditions that imply the previous hypotheses.

4- In theorem 1 and 2, we obtain that the variance term is of order smaller than $(n\psi(h_n))^{-1/2}$ and the bias term of order smaller than $h_\beta^n$. Since the quantity $\psi(h_n)$ is unknown, then it is not possible to obtain a plug-in method for $h$. For this reason, we do a sort a cross validation in the application using the bandwidth $h$ that minimizes the sum of the relative error before the time $j$. A possibility to obtain a plug-in bandwidth is to approximate the $X_i$ by its projection on a lineal space of dimension $d$ (for example the lineal space generated the $d$ first elements an orthonormal basis). In this case, a plug-in bandwidth can be obtained by $n^{-\beta/(\beta+d)}$. One has to choose $d$ such it furnish a good approximation for the $X_i$ (cf. Ferraty, Goia and Vieu (2002) for example).

5- With respect to the work of Masry....

### 4 Application to end-to-end quality of service estimation

#### 4.1 Motivation and model

As we say in Introduction, for new services like voice and video, it is important to measure their performance since they have more requirements than mail or ftp for example. There are several measurement techniques, most of them with software implementations, that can be classified
Asymptotic normality of the Nadaraya-Watson estimator for non-stationary data

in active (that send controlled traffic called probe packets) and passive measures (generally measures at the routers). The aim of these techniques is to detect different characteristics of the network, and in particular some performance parameters (delay, losses, available bandwidth, etc.). There are several problems to measure the performance parameters, for example the route of the packets can change, the traffic bit rate is not constant and normally is in bursts, the probe packets can be filtered or altered by one ISP (Internet Service Provider) in the path or there is not clock synchronization between routers and end equipments. Normally the internal routers in the path between two points of interest are not under the control of only one user or one ISP. Therefore, it is not very useful to have measurement procedures that depend on the information of the internal routers. For this reason end-to-end measurements is one of the most developed methodologies during the last years (cf. [21], [22], [23]).

End-to-end measurements are measurements between two nodes of the network. We consider a video application that we want to monitor between two nodes of the network. The network is shared by this multimedia application and many other traffics in the network that are unknown, and we call the complete traffic cross traffic. The quality of service of the multimedia application is represented by the variable \( Y \). We assume that \( Y \) is a function of the cross traffic, represented by \( X \). To take into account non-stationary behaviors of the traffic, we will suppose that \( X \) satisfies the relation (3), that is \( X = \varphi(\xi, Z) \). In this model \( Z \) is a non-stationary random process that selects between different stationary behaviors or state of the traffic. This means that for example we can have two values of \( Z \), one when the network is very loaded and another one when there is few cross traffic over the network. This classification in different types of traffic is non-stationary, has seasonal components and presents dependence between observations. The process \( \xi \) is related to the variations of the traffic within a particular state for \( Z \). Function \( \varphi \) may be very general and describes the variable \( X \) as a non-stationary and dependent process. However, conditioned to \( Z \), this process is stationary and weakly dependent.

We can not measure the cross traffic process but we can obtain data closely related with it. These data are described in the following subsection.

4.2 Measurement procedure

In what follows we describe the procedure to estimate the function \( \phi \) of the model (1) that permits to predict the quality of service. We divide the experiment in two phases. First, we send a burst of small probe packets (pp) of fixed size spaced by a fixed time. Then, immediately after the burst, we send during a short time a video stream. We repeat the previous procedure \( n \) times as it is shown in figure 1.

```
pp            video
```

Figure 1: Measurement procedure scheme

With the probe packets burst we infer the cross traffic between the two nodes of the network. We measure at the output the interarrival time between consecutive probe packets. This time series
is strongly correlated with the cross traffic process that shares the network with the probe traffic. We compute the empirical distribution function of probe packets interarrival times. Then, for each probe packet burst and video sequence \( j \), we have a pair \((X_j, Y_j)\), where \( X_j \) is the empirical distribution function of interarrival times and \( Y_j \) is the performance metric of interest measured from the video stream \( j \).

Therefore, our estimation problem has been transformed into the problem of inferring a function \( \phi : \mathcal{D} \rightarrow \mathbb{R} \) where \( \mathcal{D} \) is the space of the probability distribution functions and \( \mathbb{R} \) is the real line such that the \((X_j, Y_j)\) satisfy (1).

We did the estimations with traffic simulated in Laboratory. For these data, the variable \( Y \) is the one-way delay which is the delay between two points. The simulated traffic is an ON-OFF process. Our ON-OFF process has two variable means belonging respectively to two different intervals (one interval corresponding to loaded traffic: ON and the other interval to unloaded traffic: OFF). During a first exponential time of a given mean, the traffic has a mean choosing randomly in the first interval ON. At the end of the first exponential time, an other mean for the traffic is choosing randomly in the second interval OFF and the traffic has this mean during a second exponential time. Then the same operation is repeated. Here the variable \( Z \) takes two values and the traffic in each of the state of \( Z \) is stationary.

We have 360 observations \((X_j, Y_j)\). For each \( j \in \{1, \ldots, 359\} \), we calculate the estimator \( \hat{\phi}_j \) defined by (2) using the observations \((X_1, Y_1), \ldots, (X_j, Y_j)\) and we represent on figures 2 and 3 the values of \( Y_{j+1} \) and of its prediction \( \hat{\phi}_j(X_{j+1}) \). Moreover for each \( j \in \{1, \ldots, 359\} \), we calculate a confidence interval around \( \hat{\phi}_j(X_{j+1}) \) using a block-bootstrap.

To calculate the confidence bands we use block-bootstrap similar to block-bootstrap used for real data. The size of the block we use is \( \sqrt{i} \) for the i-th prediction. We use bootstrap since the variance that appear in Theorem 1 and 2 are difficult to calculate. The bootstrap seems to function well in our situation. The theoretical justification of this bootstrap is not simple and will be the object of future works. For functional data, wild functional bootstrap is develop in [24] but it is not anymore justifies theoretically.

The estimator \( \hat{\phi}_j \) is calculated using the kernel \( K(x) = (x^2 - 1)^2 1_{[-1,1]}(x) \), the bandwidth that minimizes the sum of the relative error in the estimation before the time \( j \) and the \( L_1 \) norm. The figure 4 represents the relative error of the estimation. The results are relatively accurate for sample of size larger than 100. Our estimator and our confidence intervals are quite accurate for the traffic simulated in Laboratory. The application to traffic in larger network will be object of future works. These first results are interesting since this give a first answer to predict end-to-end quality of service without supposing particular form for the network.
Asymptotic normality of the Nadaraya-Watson estimator for non-stationary data

Figure 2: Estimated delays for all simulated data

Figure 3: Estimated delays for simulated data observations between 230 and 270
5 Proof of the propositions and theorems

5.1 Proof of Propositions 1 and 2

This result is obtained by means of Bernstein’s method (cf. [15], [14]) and it is a generalization of the theorem 4.15 on triangular arrays for \( \mathbb{R} \)-valued random fields obtained by Tablar (2006) ([17], p 86) to \( \mathbb{R}^m \)-valued random fields. To prove that the vectorial field \( (S_n(A^1, X^{1,n}), \ldots, S_n(A^m, X^{m,n})) \xrightarrow{w} N_m(0, \Sigma) \), it is sufficient to prove that for all \( \lambda \in \mathbb{R}^m \)

\[
\langle \lambda, S_n \rangle \xrightarrow{w} N(0, \lambda^t \Sigma \lambda)
\]

where \( S_n = (S_n(A^1, X^{1,n}), \ldots, S_n(A^m, X^{m,n})) \) and \( \lambda^t = (\lambda_1, \ldots, \lambda_m) \).

We have that

\[
\langle S_n, \lambda \rangle = \sum_{i=1}^{m} \lambda_i S_n(A^i, X^{i,n}) = \sum_{i=1}^{m} \frac{1}{(\ell(n))^{1/2}} \lambda_i \sum_{k \in A^i_n} X^{i,n}_k
\]

\[
= \sum_{k \in D(n)} \sum_{i=1}^{m} \lambda_i X^{i,n}_k 1_{\{k \in A^i_n\}}
\]

Let us consider a real valued field \( X^n(\lambda) \) where \( X^{i,n}_k(\lambda) = \lambda_i X^{i,n}_k \) if \( k \in A^i_n \). Then \( \langle S_n, \lambda \rangle = S_n(D(n), X^n(\lambda)) \). As the field \( X^n(\lambda) \) verifies the hypotheses in the work of [17] we can apply the theorem for a real valued random field in order to obtain the result of propositions 1 and 2.

5.2 Proof of Proposition 4

Since \( Z \) is asymptotically measurable, conditionally to \( Z \), the subset family \( (A^1, \ldots, A^{2m}) \) is an asymptotically measurable family in \( \mathbb{N} \). Then applying Proposition 1 to \( \hat{X} \) which belongs to \( B(\mathbb{N}) \), we deduce the result.
Asymptotic normality of the Nadaraya-Watson estimator for non-stationary data

5.3 Proof of Proposition 5

Before proving Proposition 5, here we prove the following lemma.

**Lemma 1.** Let $k \in \{1, \ldots, m\}$. Under the Assumptions (A1), (A2), (A3), (A5), (A7) and (A8), we have

1) the following limit

$$
\lim_{n \to \infty} \frac{1}{\psi(h_n)} E[K_n(\varphi(\xi_1, z_k))] = d_k(x)c_k(x)1_{\{k \in \Delta\}},
$$

2) for $\beta > 0$,

$$
\lim_{n \to \infty} \frac{1}{\psi(h_n)} E \left[ \left\| \frac{\varphi(\xi_1, z_k) - x}{h_n} \right\|^\beta K_n(\varphi(\xi_1, z_k)) \right] \leq d_k(x)c_k(x),
$$

3) that the estimator $f_n(x)$ satisfies

$$
\lim_{n \to \infty} E(f_n(x)) = f(x),
$$

4) and the estimator $g_n(x)$ satisfies

$$
\lim_{n \to \infty} E(g_n(x)) = \phi(x)f(x).
$$

**Proof.** The results 1) and 2) of this lemma come from the fact that under Assumption (A1), the density of $\|\varphi(\xi_1, z_k) - x\|$ is the function $u \to c_k(x)\psi'_k(u, x)$. Then under Assumption (A7)

$$
E[K_n(\varphi(\xi_1, z_k))] = h_n c_k(x) \int_0^1 K(u)\psi'_k(uh_n, x)du
$$

and

$$
E \left[ \left\| \frac{\varphi(\xi_1, z_k) - x}{h_n} \right\|^\beta K_n(\varphi(\xi_1, z_k)) \right] = h_n c_k(x) \int_0^1 K(u)u^\beta\psi'_k(uh_n, x)du
$$

and it is sufficient to use Assumptions (A2) and (A8) to conclude.

Here we prove 3) and 4). We have for $i \in \{1, \ldots, n\}$,

$$
E(K_n(X_i)) = E\{E(K_n(X_i)|Z_i)\} = \sum_{k=1}^m E\{K_n(\varphi(\xi_1, z_k))\}P(Z_i = z_k).
$$

As $(\varphi(\xi_1, z_k))_{i=1,\ldots,n}$ is identically distributed, we obtain that,

$$
E(f_n(x)) = \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E(K_n(\varphi(\xi_1, z_k))) \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k) \right).
$$

(9)

Now applying (A3) and the assertion 1) of this lemma in (9), we obtain 3). Similar calculus give that

$$
E(g_n(x)) = \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E(\phi(\xi_1, z_k) K_n(\varphi(\xi_1, z_k))) \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k) \right).
$$
Then we have
\[
E(g_n(x)) = \phi(x)E(f_n(x)) + R_n,
\]
where
\[
R_n = \sum_{k=1}^{m} \left( \frac{1}{\psi(h_n)} E \left\{ \{ \phi(\varphi(\xi_1, z_k)) - \phi(x) \} K_n(\varphi(\xi_1, z_k)) \right\} \right) \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k)
\]
\[
\leq \sup_{u: ||x - u|| \leq h_n} |\phi(u) - \phi(x)| E(f_n(x))
\]
and we obtain 4) using the continuity of function \( \phi \) (Assumption \((A_5)\)) and using that \( \lim_{n \to \infty} E(f_n(x)) = f(x) \).

Conditionally to \( Z \), we have
\[
\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z) \right) = \left( \sum_{j=1}^{m} S_n(A^j, \tilde{X}_{n,j}), \sum_{j=m+1}^{2m} S_n(A^j, \tilde{X}_{n,j}) \right).
\]

Using Proposition 4, we obtain that, conditionally to \( Z \),
\[
\left( \sum_{j=1}^{m} S_n(A^j, \tilde{X}_{n,j}), \sum_{j=m+1}^{2m} S_n(A^j, \tilde{X}_{n,j}) \right) \overset{w}{\rightarrow} N_2(0, A).
\]

Since \( Z \) is regular, the matrix \( A \) is not random and then we have
\[
\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z) \right) \overset{w}{\rightarrow} N_2(0, A).
\]

Now
\[
(g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x))) = (g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z))
\]
\[
+ (E(g_n(x)|Z) - E(g_n(x)), E(f_n(x)|Z) - E(f_n(x))).
\]

We have
\[
E(f_n(x)|Z) - E(f_n(x)) = \sum_{j=1}^{2m} \frac{E(K_n(\xi_1, z_j))}{\psi(h_n)} \left( \frac{\text{card}(A^j)}{n} - \frac{1}{n} \sum_{i=0}^{n-1} P[Z_i = j] \right).
\]

Under Assumption \((A_3)\) and using assertion 1) of Lemma 1, we have that
\[
\sqrt{n\psi(h_n)} (E(f_n(x)|Z) - E(f_n(x)))
\]
converges in probability to 0 as \( n \) tends to \( \infty \). Using Assumption \((A_3)\), \( \sqrt{n\psi(h_n)} (E(g_n(x)|Z) - E(g_n(x))) \) converges in probability to 0 as \( n \) tends to \( \infty \). Using the lemma of Slutsky, we obtain the result of the proposition.
5.4 Proof of Theorem 1

We have
\[ \hat{\phi}_n(x) - \frac{E(g_n(x))}{E(f_n(x))} = \frac{Q_n(x) - B_n(x)(f_n(x) - E(f_n(x)))}{f_n(x)}, \]
where
\[ B_n(x) = \frac{E(g_n(x))}{E(f_n(x))} - \phi(x) \]
and
\[ Q_n(x) = (g_n(x) - E(g_n(x))) - \phi(x)(f_n(x) - E(f_n(x))). \]

Using 3) and 4) of Lemma 1, we deduce that
\[ \lim_{n \to \infty} B_n(x) = 0. \]

Then using \( A_6 \), we have that \( f_n(x) \) converges in probability to \( f(x) > 0 \) and then it implies that
\[ \frac{B_n(x)(f_n(x) - E(f_n(x)))}{f_n(x)} \]
converges in probability to 0.

Finally, using \( A_6 \), (10) and Proposition 5, applying Slutsky Lemma, we deduce the result of the proposition.

5.5 Proof of Theorem 2

Here we prove that
\[ \lim_{n \to \infty} \sqrt{n \psi(h_n)} \left( \frac{E(g_n(x))}{E(f_n(x))} - \phi(x) \right) = 0. \]

Indeed, we have
\[
|E(g_n(x)) - E(f_n(x))\phi(x)| \leq \frac{1}{\psi(h_n)} \sum_{k=1}^{m} \left( K_n(\varphi(\xi, z_k)) \left( \phi(\varphi(\xi, z_k)) - \phi(x) \right) \right) \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k) \\
\leq L h_\beta \sum_{k=1}^{m} \frac{1}{\psi(h_n)} \left( \frac{\varphi(\xi, z_k)}{h_n} \right) \left\| \varphi(\xi, z_k) - x \right\|_{\beta} \frac{1}{n} \sum_{i=0}^{n-1} P(Z_i = z_k),
\]
where the second line comes from Assumption \( (B_1) \) and the third is a consequence of assertion 2) of Lemma 1. Using (12), Assertion 3) of Lemma 1 and Assumptions \( (B_2) \) and \( (A_6) \), we deduce (11). Theorem 2 is obtained using Theorem 1 and (11).

6 Acknowledgement

This work has been supported by the ARTES group. Karine Bertin has been supported by Project DIPUV 33/05, Project FONDECYT 1061184 and the Laboratory ANESTOC ACT-13. Laura Aspirot has been partially supported by PDT (Programa de Desarrollo Tecnológico) S/C/OP/46/. The authors would like to thank Yoana Donoso for her contributions to the simulations programs.
References


