Set-indexed conditional empirical and quantile processes based on dependent data *

Wolfgang Polonik
Division of Statistics
University of California
Davis, CA95616-8705, USA

Qiwei Yao
Department of Statistics
London School of Economics
London, WC2A 2AE, UK

Abstract

We consider a conditional empirical distribution of the form

$$\widehat{F}_n(C|x) = \sum_{t=1}^{n} \omega_n(X_t - x) I_{\{Y_t \in C\}}$$

indexed by $C \in \mathcal{C}$, where $\{(X_t, Y_t), t = 1, \ldots, n\}$ are observations from a strictly stationary and strong mixing stochastic process, $\{\omega_n(X_t - x)\}$ are kernel weights, and $\mathcal{C}$ is a class of sets. Under the assumption on the richness of the index class $\mathcal{C}$ in terms of metric entropy with bracketing, we have established uniform convergence and asymptotic normality for $\widehat{F}_n(\cdot|x)$. The key result specifies rates of convergences for the modulus of continuity of the conditional empirical process. The results are then applied to derive Bahadur-Kiefer type approximations for a generalized conditional quantile process which, in the case with independent observations, generalizes and improves results of Bhattacharya and Gangopadhyay (1990), and Gangopadhyay and Sen (1993). Potential applications in the areas of estimating level sets and testing for unimodality (or multimodality) of conditional distributions are discussed.

Keywords: Bahadur-Kiefer approximation, conditional distribution, covering number, empirical process theory, generalized conditional quantile, level set, minimum volume predictor, Nadaraya-Watson regression estimator, nonlinear time series, strong mixing.

*Supported partially by the EPSRC Grants L67561 and L16385, and the EU Human Capital and Mobility Program ERB CHRX-CT 940693.

1 Introduction

An empirical process indexed by a class of sets or functions is an interesting mathematical model with various statistical applications (see, for example, Shorack and Wellner 1986). Such a process defined from independent and identically distributed (*i.i.d.*) observations has been extensively studied in literature in past two decades. More recently, empirical processes based on dependent data have been studied under various mixing conditions (*e.g.* Massart 1987, Andrews and Pollard 1994, Doukhan, Massart, and Rio 1995). The extension of the above exploration to *conditional* empirical processes is practically useful and technically more challenging.

Consider observations $\{(X_t, Y_t),\ t = 1, \ldots, n\}$ with $X_t \in \mathbf{R}^d$ and $Y_t \in \mathbf{R}^{d'}$. Let $F(\cdot|x)$ denote the conditional distribution of Y_t given $X_t = x$. Note that $F(C|x) = E\{I_{(Y\in C)}|X_t = x\}$. This regression relationship suggests to consider the following Nadaraya-Watson-type conditional empirical distribution of the form

$$\widehat{F}_n(C|x) = \sum_{t=1}^{n} I_{\{Y_t\in C\}} K\left(\frac{X_t - x}{h}\right) \Big/ \sum_{t=1}^{n} K\left(\frac{X_t - x}{h}\right) \tag{1.1}$$

where $K(\cdot) \geq 0$ is a kernel function on $\mathbf{R}^d$, and $h > 0$ is a bandwidth, C is a measurable set, and $x \in \mathbf{R}^{d'}$. By choosing $C = (-\infty, z], z \in \mathbf{R}^{d'}$, it reduces to the conditional empirical distribution function (cdf) $\widehat{F}_n(z|x) = \widehat{F}_n((-\infty, z]|x)$. The corresponding (classical) conditional empirical process is usually defined as

$$\nu_n(z) = \sqrt{nh^d}(\widehat{F}_n(z|x) - F(z|x)), \quad z \in \mathbf{R}^{d'}$$

where $F(\cdot|x)$ denotes the conditional distribution function of Y_t given $X_t = x$, and x is considered as fixed. To our best knowledge, the study of conditional empirical processes so far has been confined with *i.i.d.* observations $\{(X_t, Y_t),\ t = 1, \ldots, n\}$; see, among others, Stute (1986a,b), Horvath (1988), and Bhattacharya and Gangopadhyay (1990). The conditional empirical quantile process is defined as

$$\xi_n(y|x) = \tilde{g}(y|x)\sqrt{nh^d}\,(\hat{F}_n^{-1}(y|x) - F^{-1}(y|x)),$$

where $\tilde{g}(y|x)$ denotes the conditional density quantile function, i.e. $\tilde{g}(y|x) = g(F^{-1}(y|x))|x)$, $g(\cdot|x)$ denotes the density function of $F(\cdot|x)$, and $\hat{F}_n^{-1}(y|x)$ and $F^{-1}(y|x))$ are the (generalized) inverses of $\widehat{F}_n(z|x)$ and $F(z|x)$ respectively. There are two types frequently used asymptotic approximations for empirical and quantile processes, namely *Bahadur-Kiefer approximations* and *Bahadur representations.* Bahadur-Kiefer approximations are typically of the form

$$\xi_n(y|x) - \nu_n(F^{-1}(y|x)) = \text{“bias”} + \sqrt{nh^d}R_n(y|x) + \text{“higher order terms”},$$

where the rate of convergence of $\sqrt{nh^d}R_n(y|x)$ to zero describes the accuracy of the approximation. A Bahadur representation is an approximation for a quantile process $\xi_n(y|x)$:

$$\xi_n(y|x) = \text{“sum of } i.i.d. \text{ r.v.”} + \text{“bias”} + \sqrt{nh^d}\, R_n(x|z) + \text{“higher order terms”},$$

which is of the similar form to Bahadur-Kiefer approximation but with $\nu_n(F^{-1}(y|x))$ replaced by a sum of *i.i.d.* random variables. Note that we may expect that $\nu_n(F^{-1}(y|x))$ is asymptotically normal, but not necessarily a sum of *i.i.d.* random variables.

Bhattacharya and Gangopadhyay (1990) derived a Bahadur-representation with $R_n(y|x) = O(n^{-3/5}\log n)$ a.s. with both y and x fixed for the case $d = d' = 1$ and the uniform kernel $K(\cdot)$. They also studied nearest neighbor type versions of the empirical and the quantile processes and derived analogous Bahadur representations. Based on a different proof, Gangopadhyay and Sen (1993) re-derived Bhattacharia and Gangopadhyay's result for the nearest neighbor versions with $R_n(x|z) = O_P(n^{-3/5}\log n)$ in their Bahadur representation. Mehra et al. (1991) considered a local version of the conditional quantile estimator proposed by Yang (1984) and studied the asymptotic behavior of the corresponding quantile process. Xiang (1995, 1996) studied a smoothed version of $\hat{F}_n^{-1}(y|x)$ where $\hat{F}_n^{-1}(y|x)$ is smoothed with another kernel locally around y. Xiang derived a Bahadur-representation for this smoothed conditional quantile function. With appropriately selected (two) bandwidths he was able to derive the asymptotic distribution of $\sqrt{nh}\, R_n(y|x)$ for fixed y and x. Bahadur-Kiefer type approximations for unconditional empirical and quantile processes based on mixing processes have also been studied by, for example, Sen (1972), Basu and Singh (1978), or Yoshihara (1995).

In this paper the above studies for conditional processes are generalized in two respects. First, observations are allowed to be dependent, i.e. strong mixing, and secondly we consider set-indexed processes, i.e. the indicators $I_{\{Y_t \le z\}}$ are replaced by $I_{\{Y_t \in C\}}$ (see (1.1)) with sets C lying in some class of (measurable) sets $\mathcal{C}$. This is not only of theoretical interested, but also has some direct statistical applications. Our study is directly motivated by prediction of nonlinear and non-Gaussian time series (Polonik and Yao 2000). We will also discuss potential applications of our results in various other statistical practices such as level set estimation and testing for unimodality.

We always assume that $\{(X_t, Y_t)\}$ is a strictly stationary process, with $X_t \in \mathbf{R}^d$ and $Y_t \in \mathbf{R}^{d'}$. In the usual time series context, Y_t is a scalar and X_t consists of its lagged values. We study the asymptotic behaviour of the conditional empirical process

$$\mathcal{C} \ni C \to \nu_n(C|x) = \sqrt{nh^d}\{\widehat{F}_n(C|x) - F(C|x)\} \tag{1.2}$$

and derive Bahadur-Kiefer approximations for a conditional generalized quantile process (see below). The key result Theorem 2.3 essentially deals with the asymptotic behaviour of the modulus of continuity of $\nu_n(\cdot|x)$. This asymptotic behaviour depends on the richness (or complexity) of the index class $\mathcal{C}$, which is measured in terms of metric entropy with bracketing. In fact if $\mathcal{C}$ is not too rich (see Theorem 2.3 below), the conditional $\mathcal{C}$-indexed empirical process converges weakly, in the sense of Hoffman-Jørgensen (*cf.* van der Vaart and Wellner 1996), to a so-called $F(\cdot|x)$-bridge. Therefore the empirical process behaves like the one based on *i.i.d.* observations in terms of first order asymptotics, as long as the class $\mathcal{C}$ is not too rich. This phenomenon is not a surprise, since only the observations with X_t in a small neighbourhood of x are effectively used in the estimation (1.1). Those observations are not necessarily close with each other in the time space. Indeed, they could be regarded as asymptotically independent under appropriate conditions such as strong mixing; see Hard (1996). On the other hand, it remains at least to us as an open problem to identify the maximum richness of $\mathcal{C}$ (under the strong mixing condition) to retain the above *i.i.d.*-like asymptotic behaviour. The condition specified in this paper restrains $\mathcal{C}$ far from being as rich as in the case of *i.i.d.* observations in order to retain the same asymptotic results. Note that the standard conditional empirical processes indexed by $x \in \mathbf{R}^d$ usually behave asymptotically like those based on *i.i.d.* observations. However, the corresponding class $\mathcal{C} = \{(-\infty, x], x \in \mathbf{R}^d\}$ is very "thin".

We also apply the key result, Theorem 2.3, to derive the Bahadur-Kiefer approximations for a generalized conditional quantile process. Note that for set-indexed processes there is no obvious quantile process or inverse process. Instead we consider processes indexed over an important class of sets, namely the so-called (empirical) minimum volume sets (see below). For $\alpha \in [0,1]$, a set $\mathrm{M}_{\mathcal{C}}(\alpha|x) \in \mathcal{C}$ is called a *conditional MV-set* in $\mathcal{C}$ at level α if

$$\mathrm{M}_{\mathcal{C}}(\alpha|x) \in \mathrm{argmax}\{\mathrm{Leb}(C) : C \in \mathcal{C}, F(C|x) \geq \alpha\}, \tag{1.3}$$

where $\mathrm{Leb}(\cdot)$ denotes Lebesgue measure. Analogously, $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ denotes an *empirical* conditional MV-set if $F(\cdot|x)$ in (1.3) is replaced by the empirical distribution $\widehat{F}_n(\cdot|x)$. We denote their volumes as

$$\mu_{\mathcal{C}}(\alpha|x) = \mathrm{Leb}(\mathrm{M}_{\mathcal{C}}(\alpha|x)) \quad \text{and} \quad \widehat{\mu}_{\mathcal{C}}(\alpha|x) = \mathrm{Leb}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)), \tag{1.4}$$

respectively. The volume process

$$\alpha \to \sqrt{nh^d}\ (\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x)) \tag{1.5}$$

can be considered as a conditional version of a generalized quantile process as defined in Einmahl and Mason (1992). This generalized quantile process turns out to be a generalization of the usual quantile process defined above (see below).

Note that MV-sets are indexed by a one-dimensional parameter. As a consequence we are dealing with one-dimensional processes. Nevertheless, it is essential that set-indexed processes are studied, because empirical minimum volume sets are *random* sets. The Bahadur-Kiefer approximation for the volume process derived in §3 improves and generalizes results of Bhattacharya and Gangopadhyay (1990) and Gangopadhyay and Sen (1993), who both considered the special case of $d' = d = 1$ and the observations being independent. Polonik (1997) established similar Bahadur-Kiefer approximation for an unconditional volume process based on *i.i.d.* observations.

The remainder of the paper is organized as follows. We present the asymptotic results of the process $\{\nu_n(\cdot|x)\}$ in §2. §3 contains the Bahadur-Kiefer approximations for the volume process (1.5). §4 provides a discussion on how the results in this paper can be applied to various statistical applications. §5 contains some key technical arguments.

2 The set-indexed conditional empirical process

In this section, we establish asymptotic properties of the process $\nu_n(\cdot|x)$ defined in (1.2), which include a Glivenko-Cantelli type result, the asymptotic normality of finite dimensional distributions, and the asymptotic behaviour of the modulus of continuity. The last two results imply that $\nu_n(\cdot|x)$ converges to a Gaussian process.

Let $f(\cdot)$ be the density function of X_t. We always assume that $x \in \mathbf{R}^d$ is fixed and $f(x) > 0$. Further, all the non-deterministic quantities are assumed to be measurable, and we write $d_{F(\cdot|x)}(A, B) = F(A\Delta B|x)$. We use c to denote some generic constant, which may be different at different places. We introduce some regularity conditions first.

(A1) The marginal density f is bounded and continuous in a neighbourhood of x.

(A2) The kernel density function K is bounded and symmetric, and $\lim_{u\to\infty} ||u||^d K(u) = 0$.

(A3) $f \in C_{2,d}(b)$, where $C_{2,d}(b)$ denotes the class of bounded real-valued functions with bounded second order partial derivatives.

(A4) $F(\cdot|x)$ has a Lebesgue-density $g(\cdot|x) \in C_{2,d'}(b)$. Moreover, for each $C \in \mathcal{C}$ the function $F(C|\cdot) \in C_{2,d}(b)$ such that $\sup_{C\in\mathcal{C}} \left|\frac{\partial^2}{\partial x_i \partial x_j} F(C|x)\right| < \infty, \;\; \forall\, 1 \le i,j \le d.$

(A5) $|| \int vv^T K(v)\, dv|| < \infty$.

(B1) The joint distribution of (X_t, X_{t+q}) has the density function f_q, and f_q is bounded uniformly over $q \ge 1$.

(B2) The joint density function of (X_s, X_t, X_q, X_r) exists and is bounded from the above by a constant independent of (s,t,q,r).

We call the stationary process $\{(X_t, Y_t)\}$ *strong mixing* if

$$\alpha(j) = \sup_{A\in\mathcal{F}_{-\infty}^0,\, B\in\mathcal{F}_j^\infty} |P(AB) - P(A)P(B)| \to 0, \quad \text{as } j \to \infty, \tag{2.1}$$

where $\mathcal{F}_s^t$ denotes the σ-algebra generated by $\{(X_i, Y_i), s \le i \le t\}$. We use the term *geometrically* strong mixing if $\alpha(j) \le a\, j^{-\beta}$ for some $a > 0$ and $\beta > 1$, and *exponentially* strong mixing if $\alpha(k) \le b\,\gamma^k$ for some $b > 0$ and $0 < \gamma < 1$. Sometimes the condition of strong mixing can be reduced to so-called 2-*strong mixing*, which is defined as in (2.1) with $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_j^\infty$ replaced by $\sigma(X_0, Y_0)$ and $\sigma(X_j, Y_j)$ respectively. We use the terms *geometrically* or *exponentially* 2-strong mixing in a similar manner.

Now we introduce the notion of *metric entropy with bracketing* which provides a measure of richness (or complexity) of a class of sets $\mathcal{C}$. This notion is closely related to *covering numbers.* We adopt L_1-type covering numbers using the bracketing idea. The bracketing reduces to *inclusion* when it is applied to classes of sets rather than classes of functions. For each $\epsilon > 0$, the covering number is defined as

$$\begin{aligned} N_I(\epsilon, \mathcal{C}, F(\cdot|x)) = \inf\{n \in \mathbf{N} : \exists\, C_1, \ldots, C_n \in \mathcal{C} \text{ such that} \\ \forall\, C \in \mathcal{C}\ \exists\, 1 \le i,j \le n \text{ with } C_i \subset C \subset C_j \text{ and } F(C_j \setminus C_i|x) < \epsilon\}. \end{aligned} \tag{2.2}$$

The quantity $\log N_I(\epsilon, \mathcal{C}, F(\cdot|x))$ is called *metric entropy with inclusion* of $\mathcal{C}$ with respect to $F(\cdot|x)$. A pair of sets C_i, C_j is called a *bracket* for C. Estimates for such covering numbers are known for many classes. (See, *e.g.* Dudley 1984.) We will often assume below that either $\log N_I(\epsilon, \mathcal{C}, F(\cdot|x))$ or $N_I(\epsilon, \mathcal{C}, F(\cdot|x))$ behave like powers of ϵ^{-1} : We say that condition (R_γ) holds if

$$\log N_I(\epsilon, \mathcal{C}, F(\cdot|x)) < H_\gamma(\epsilon), \quad \text{for all } \epsilon > 0, \tag{R_γ}$$

where

$$H_\gamma(\epsilon) = \begin{cases} \log(A\,\epsilon^{-r}) & \text{if } \gamma = 0, \\ A\,\epsilon^{-\gamma} & \text{if } \gamma > 0, \end{cases} \tag{2.3}$$

for some constants $A, r > 0$. In fact condition (R_0) holds for intervals, rectangles, balls, ellipsoids, and for classes which are constructed from the above by performing set operations union, intersection and complement finitely many times. The classes of convex sets in $\mathbf{R}^d$ $(d \geq 2)$ fulfill condition (R_γ) with $\gamma = (d-1)/2$. This and other classes of sets satisfying (R_γ) with $\gamma \geq 0$ can be found in Dudley (1987).

Now we are ready to formulate the results on the uniform consistency and the (pointwise) asymptotic normality of $\nu_n(C|x)$.

Theorem 2.1 (Uniform consistency)

Suppose that conditions (A1), (A2) *and* (B1) *hold, and that* $\{(X_t, Y_t)\}$ *is geometrically* 2*-strong mixing with* $\beta > 2(p-1)/(p-2)$. *Let* $\mathcal{C}$ *be a class of measurable sets for which* $N_I(\epsilon, \mathcal{C}, F(\cdot|x)) < \infty$ *for any* $\epsilon > 0$. *Suppose further that* $\forall\ C \in \mathcal{C}$

$$|F(C|y)f(y) - F(C|x)f(x)| \to 0 \quad \text{as } y \to x. \tag{2.4}$$

If $nh^d \to \infty$ *and* $h \to 0$ *as* $n \to \infty$, *then*

$$\sup_{C \in \mathcal{C}} |\widehat{F}_n(C|x) - F(C|x)| \xrightarrow{P} 0.$$

Theorem 2.2 (Asymptotic normality)

Let (A2) – (A5) *and* (B2) *hold, and suppose that* (B1) *holds with* $p = \infty$. *Suppose further that the process* $\{(X_t, Y_t)\}$ *is geometrically strong mixing with* $\beta > 2$. *Let* $h = cn^{-\frac{1}{d+4}}(\log\log n)^{-1}$. *Then as* $n \to \infty$, *for* $m \geq 1$ *and* $C_1, \ldots, C_m \in \mathcal{C}$,

$$\{\nu_n(C_i|x);\ i = 1, \ldots, m\} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where $\Sigma = (\sigma_{i,j})_{i,j=1,\ldots,m}$, *and* $\sigma_{i,j} = \{F(C_i \cap C_j|x) - F(C_i|x)F(C_j|x)\} \int K^2(u)du/f(x)$.

In order to formulate the next theorem which provides the information on the asymptotic behaviour of the modulus of continuity (see remarks below), we need to introduce the following function

$$\Lambda_\gamma(\sigma^2, n) = \begin{cases} \sqrt{\sigma^2 \log \frac{1}{\sigma^2}} & \text{if } \gamma = 0, \\ \max\left((\sigma^2)^{\frac{1-\gamma}{2}}, \left(nh^d\right)^{\frac{3\gamma-1}{2(3\gamma+1)}} \right) & \text{if } \gamma > 0. \end{cases} \tag{2.5}$$

Theorem 2.3 *Suppose that* (A2) – (A5) *and* (B1) *hold, and the process* $\{(X_t, Y_t)\}$ *is exponentially strong mixing. For each* $\sigma^2 > 0$*, let* $\mathcal{C}_\sigma \subset \mathcal{C}$ *be a class of measurable sets with* $\sup_{C \in \mathcal{C}_\sigma} F(C|x) \leq \sigma^2 \leq 1$*, and suppose that* $\mathcal{C}$ *fulfills* (R_γ) *with some* $\gamma \geq 0$*. Further we assume that* $h^d \to 0$ *and* $nh^d \to \infty$ *as* $n \to \infty$ *such that*

$$nh^{d+4} \leq \left(\Lambda_\gamma(\sigma^2, n)\right)^2, \text{ and}$$
$$\frac{nh^d \left(\sigma^2 \log \frac{1}{\sigma^2}\right)^{1+\gamma}}{\log n} \to \infty \quad \text{as } n \to \infty.$$

Further we assume that $\sigma^2 \geq h^2$*. For* $\gamma > 0$ *and* $d = 1, 2$ *the latter has to be replaced by* $\sigma^2 \geq h^d \log(1/h^d)$*. Then for every* $\epsilon > 0$ *there exists a constant* $M > 0$ *such that*

$$P\left(\sup_{C \in \mathcal{C}_\sigma} |\nu_n(C|x)| \geq M\, \Lambda_\gamma(\sigma^2, n)\right) \leq \epsilon$$

for all sufficiently large n*.*

Remark 2.4 (a) Note that Λ_n tends to zero as $n \to \infty$ provided $\gamma < 1/3$. In this case, Theorem 2.3 entails the tightness of the conditional set-indexed empirical process. To see this, note that trivially $\sup_{C,D \in \mathcal{C}} |\nu_n(C|x) - \nu_n(D|x)| \leq 2 \sup_{B \in \mathcal{C} \setminus \mathcal{C}} |\nu_n(B|x)|$ where $\mathcal{C} \setminus \mathcal{C} = \{C \setminus D, C, D \in \mathcal{C}\}$. Without the loss of generality, we may assume that $\emptyset \in \mathcal{C}$ such that $\mathcal{C} \subset \mathcal{C} \setminus \mathcal{C}$. Now, it is easy to see that $N_I(\epsilon, \mathcal{C}, F(\cdot|x)) \leq N_I(\epsilon, \mathcal{C} \setminus \mathcal{C}, F(\cdot|x)) \leq (N_I(\epsilon/2, \mathcal{C}, F(\cdot|x)))^2$. This implies that (R_γ) holds for $\mathcal{C}$ if and only if it holds for $\mathcal{C} \setminus \mathcal{C}$. Hence, an application of Theorem 2.3 to the class $\mathcal{C} \setminus \mathcal{C}$ together with Theorem 2.2 entails, by standard arguments, that the set-indexed process converges in distribution to a so-called $F(\cdot|x)$-bridge, provided $\gamma < 1/3$. An $F(\cdot|x)$-bridge is a Gaussian process with almost surely continuous sample paths and covariance structure as given in Theorem 2.2 (*e.g.* Pollard 1984). Taking into account possible non-measurability the convergence in distribution should be understood in the sense of Hoffman-Jørgensen (see van der Vaart and Wellner 1996).

(b) It is well-known in the empirical process theory that an unconditional empirical process based on *i.i.d.* observations is tight if (R_γ) holds with the sharp bound $\gamma < 1$ (see Alexander 1984). The same conclusion holds for a conditional empirical process as long as the process is formed from a set of *i.i.d.* observations. However, for the empirical processes based on dependent data under the strong mixing condition, we assume in this paper $\gamma < 1/3$ to achieve the tightness. It was indicated on page 128 of Andrews and Pollard (1994) that the tightness of an (unconditional) empirical process can be established by using the method of Massart (1987) under the condition that $\gamma < 1/4$. (Note that the parameter β in Andrews and Pollard (1994) is equal to 2γ in our notation.) Hence, we have enlarged the upper bound from 1/4 to 1/3. However it remains as an open problem if a further improvement is possible, and if further we can reach the upper bound 1 for strong mixing processes.

(c) To demonstrate that our general results lead to well-known (optimal) rates of convergence in special cases, we briefly discuss the case $\gamma = 0$. With $h = c_n \left(\frac{\sigma^2}{n}\right)^{\frac{1}{d+4}}$, where $c_n \to c > 0$ as $n \to \infty$, the results below follow from Theorem 2.3 immediately.

(c1) Let $\sigma^2 = 1$, we have that

$$n^{\frac{2}{d+4}} \sup_{C \in \mathcal{C}} |\widehat{F}_n(C|x) - F(C|x)| = O_P(1).$$

(c2) Let $\{\mathcal{C}_n\}$ be a sequence of classes of sets with $\mathcal{C}_n \subset \mathcal{C}$ and $\sup_{C \in \mathcal{C}_n} F(C|x) \leq \sigma_n^2 \leq 1$. Let $\sigma^2 = \sigma_n^2 \to 0$ such that the conditions of Theorem 2.3 hold. Then

$$\left(\frac{n}{\sigma_n^2}\right)^{\frac{2}{d+4}} \sup_{C \in \mathcal{C}_n} |\widehat{F}_n(C|x) - F(C|x)| = O_P(\sqrt{\log n}).$$

3 Bahadur-Kiefer-type approximations

In this section we study the behaviour of the volume process defined in (1.5), which can be regarded as a generalized quantile process. Note that $\widehat{\mu}_{\mathcal{C}}(\alpha|x) = \text{Leb}(\widehat{\text{M}}_{\mathcal{C}}(\alpha|x))$, and

$$\widehat{\text{M}}_{\mathcal{C}}(\alpha|x) \in \text{argmax}\{\text{Leb}(C) : \widehat{F}_n(C|x) \geq \alpha\}.$$

We assume throughout this section, that empirical MV-sets with finite ν-measure exist for every $\alpha \in [0,1]$. This assumption is satisfied for all standard choices of the class $\mathcal{C}$. Replacing the Lebesgue measure by a general function $\lambda : \mathcal{C} \to \mathbf{R}$, the process defined in (1.5) becomes a conditional version of the generalized quantile function as defined in Einmahl and Mason (1992). It reduces to the conditional quantile if we let $\mathcal{C} = \{(-\infty, x], x \in \mathbf{R}\}$ and $\lambda((-\infty, x]) = x$. In fact we have that the MV-set $\widehat{\text{M}}_{\mathcal{C}}(\alpha|x) = (-\infty, \widehat{F}_n^{-1}(\alpha|x)]$ on the one hand, and the "volume" $\widehat{\mu}_{\mathcal{C}}(\alpha|x) = \lambda((-\infty, \widehat{F}_n^{-1}(\alpha|x)]) = \widehat{F}_n^{-1}(\alpha|x)$ on the other hand. Hence, a conditional quantile may be regarded as an MV-sets itself, and as well as its "volume".

A classical (unconditional) empirical MV-sets is the so-called *shorth* which is the MV-interval at the level 1/2. The term 'shorth' was first introduced by Andrews et al. (1972) referring to the *mean* of the data lying inside the MV-interval at the level 1/2, which is different from current practice. Rousseeuw (1986) introduced the MV-ellipsoid in the context of robust estimation for multivariate location and scatter.

A very important type of MV-sets are the so-called level sets defined in terms of probability density functions. Suppose that $F(\cdot|x)$ has Lebesgue density $g(\cdot|x)$. Denote

$$\Gamma_{g(\cdot|x)}(\lambda) = \{x \in \mathbf{R}^d : g(\cdot|x) \geq \lambda\}, \quad \lambda > 0, \tag{3.1}$$

the *level sets* of $g(\cdot|x)$. It is easy to see that if $\Gamma_{g(\cdot|x)}(\lambda) \in \mathcal{C}$, it is an MV-set at the level $\alpha_\lambda = F(\Gamma_{g(\cdot|x)}(\lambda)|x)$.

Theorem 3.1 below presents Bahadur-Kiefer type rates of approximation for the set-indexed conditional empirical process. Note that $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ depends on the bandwidth h through $\widehat{F}_n(\cdot|x)$, which is not reflected explicitly in the notation.

Theorem 3.1 (Generalized Bahadur-Kiefer approximation)

Suppose that the conditions of Theorem 2.3 hold. Assume that $\mu_{\mathcal{C}}(\cdot|x)$ is differentiable with Lipschitz-continuous derivative $\mu'_{\mathcal{C}}(\cdot|x)$, and the condition (R_γ) holds for $\mathcal{C}$. Let further $\alpha \in (0,1)$ be fixed and suppose that $\mathrm{M}_{\mathcal{C}}(\alpha|x)$ is unique up to Leb-nullsets, that $F(\mathrm{M}_{\mathcal{C}}(\beta|x)|x) = \beta$ for all β in a neighborhood of α, and that $\mu'_{\mathcal{C}}(\alpha|x) > 0$. If for h and σ^2 satisfying the conditions of Theorem 2.3 *we have that as $n \to \infty$,*

$$d_{F(\cdot|x)}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x), \mathrm{M}_{\mathcal{C}}(\alpha|x)) = O_P(\sigma^2),$$

then as $n \to \infty$,

$$|(\widehat{F}_n - F)(\mathrm{M}_{\mathcal{C}}(\alpha|x)) + \frac{1}{\mu'_{\mathcal{C}}(\alpha|x)}(\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x))| = O_P\left(\frac{\Lambda_\gamma(\sigma^2, n)}{(nh^d)^{1/2}}\right).$$

In order to evaluate explicit rates from this theorem, we need to know the rates of convergence σ^2 for the empirical MV-sets. To this end, we assume that the level sets of the conditional density are (essentially) unique MV-sets. More precisely, it is assumed that for $\alpha \in [0,1]$ there exists a level λ_α such that for any $\mathrm{M}_{\mathcal{C}}(\alpha)$ we have

$$d_{\mathrm{Leb}}(\Gamma_{g(\cdot|x)}(\lambda_\alpha), \mathrm{M}_{\mathcal{C}}(\alpha|x)) = 0. \tag{3.2}$$

This assumption is fulfilled for all α if $\Gamma_{g(\cdot|x)}(\lambda) \in \mathcal{C}$ *for all* $\lambda \geq 0$, and $g(\cdot|x)$ has no flat parts (*i.e.* $\mathrm{Leb}\{y : g(y|x) = \lambda\} = 0 \;\; \forall\, \lambda > 0$). In addition we assume that $g(\cdot|x)$ is regular at the level λ_α, in the sense that

$$\mathrm{Leb}\{y : |g(y|x) - \lambda_\alpha| \leq \epsilon\} = O(\epsilon). \tag{3.3}$$

Under (3.2) and (3.3) rates of convergence for MV-sets are derived in Polonik and Yao (2000). Using these rates we obtain the following corollary.

Corollary 3.2 *Let conditions* (A2) – (A5), (B1) *and* (B2) *hold, and suppose that the process* $\{(X_t, Y_t)\}$ *is exponentially strong mixing. Let* $\alpha \in [0,1]$ *such that (3.2) and (3.3) hold. Further let*

$$h = c \max\left(n^{-\frac{1}{d+(3+\gamma)}}, n^{-\frac{1}{d+2(3\gamma+1)}}\right).$$

For $\gamma > 0$ *and* $d = 1$ *the term* $n^{-\frac{1}{d+(3+\gamma)}}$ *in the definition of* h *has to be replaced by* $n^{-\frac{1}{d+(3+\gamma)}} \log(n)$. *Then for every* $\eta > 0$ *we have as* $n \to \infty$ *that*

$$|(\widehat{F}_n - F)(\Gamma_{g(\cdot|x)}(\lambda_\alpha)|x) + \lambda_\alpha(\widehat{\mu}_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x))| = \begin{cases} O_P(n^{-\frac{2}{d+(3+\gamma)}+\eta}) & \text{if } \gamma < 1/5, \\ O_P(n^{-\frac{2}{d+2(3\gamma+1)}}) & \text{if } \gamma \geq 1/5. \end{cases}$$

Finally, we state a theorem giving Bahadur-Kiefer approximations for the standard conditional one-dimensional empirical process indexed by $y \in \mathbf{R}$ (see Introduction). Let

$$q(\alpha) = q(\alpha|x) = F^{-1}(\alpha|x)$$

denote the conditional quantile, and let

$$q_n(\alpha|x) = \widehat{F}_n^{-1}(\alpha|x),$$

where F^{-1} and $\widehat{F}_n^{-1}$ denote the generalized inverses of $F(\cdot|x)$ and $\widehat{F}_n(\cdot|x)$, respectively. Since we now use the optimal bandwidth, the bias comes into play (see also Lemma 5.2 in the Appendix). We define

$$\Psi_2(C|x) = \frac{1}{f(x)}\langle \nabla F(C|x), \int vK(v)\langle v, (\nabla f)(x)\rangle d\,v\rangle + \frac{1}{2}\int v^T \nabla^2 F(C|x) v K(v)\, d\,v \tag{3.4}$$

where ∇ and ∇^2 denote gradient and Hessian operator respectively. To simplify notation we write $\Psi_2(y|x)$ instead of $\Psi_2((-\infty, y]|x)$.

Corollary 3.3 (Bahadur-Kiefer approximation for the usual conditional empirical process) *Let conditions* (A2) – (A5), (B1) *and* (B2) *hold with* $\mathcal{C} = \{(-\infty, y], y \in \mathbf{R}\}$, *and suppose that the process* $\{(X_t, Y_t)\}$ *is exponentially strong mixing. Suppose further, that for a fixed* $\alpha \in (0,1)$ *the function* $g(\cdot|x)$ *is continuous at* $q(\alpha|x)$ *and that* $g(q(\alpha|x)|x) > 0$. *Let* $h = cn^{-1/d+4}$. *Then as* $n \to \infty$, *it holds that*

$$|(\widehat{F}_n - F)(q(\alpha|x)|x) + \Psi_2(q(\alpha|x)|x) + g(q(\alpha|x)|x)(q_n(\alpha|x) - q(\alpha|x))| = O_P\left(n^{-\frac{3}{d+4}}\sqrt{\log n}\right).$$

The approximation holds also almost surely with rate $O\left(n^{-\frac{3}{d+4}}(\log n)^{\frac{3}{4}}\right)$

Remark 3.4 (a) Although the class $\mathcal{C}$ in Corollary 3.3 satisfies (R_γ) with $\gamma = 0$, the rates in Corollary 3.3 are faster than that derived from Corollary 3.2 with $\gamma = 0$. In fact, the quantiles converge at the rate of $1/\sqrt{nh^d}$, whereas the estimators of level sets converge at a slower rate, although both of them are MV-sets. Note that quantiles are MV-sets in the class of intervals of the form $(-\infty, y], y \in \mathbf{R}$, which have one fixed end-point at $-\infty$. Hence, the estimation of a quantile reduces to the estimation of its "length", which can be fulfilled at the rate of $1/\sqrt{nh^d}$. However, estimation of a general MV-set is much more involved, and hence the convergence is slower. (It is well-known that the classical shorth can be estimated at the rate of $n^{-1/3}$ only, whereas the length of the shorth can be estimated at the rate of $n^{-1/2}$.)

(b) Corollary 3.3 improves results from Bhattacharya and Gangopadhyay (1990) and from Gangopadhyay and Sen (1993). Bhattacharya and Gangopadhyay dealt with the *i.i.d.* case using a uniform kernel with one-dimensional X_i, i.e. $d = 1$. Note that in the *i.i.d.* case, the Bahadur-Kiefer approximation given in Corollary 3.3 actually gives a Bahadur representation. The convergence rate obtained by Bhattacharya and Gangopadhyay (1990)for their Bahadur representation is $O\left(n^{-\frac{3}{5}} \log n\right)$, which is slower than ours by a factor $(\log n)^{\frac{1}{4}}$. Gangopadhyay and Sen (1993) used a different method of derive a Bahadur representation of order $O_P\left(n^{-\frac{3}{5}} \log n\right)$ which is slower that ours by a factor of $\sqrt{\log n}$.

(c) The above a.s. approximation rate is of the form $(nh^d)^{-3/4}(\log(nh^d))^{3/4}$. Hence, up to log-factors it is in alignment with the rates for unconditional quantile process. For example, the almost sure rate for one-dimensional unconditional processes (*i.e.* $d = 1$) with *i.i.d.* observations is $O(n^{-3/4}(\log \log n)^{3/4})$ (Kiefer 1967).

(d) The density quantile function $g(q(\alpha)|x)$ in Theorem 3.3 corresponds to λ_α in Theorem 3.1. Note that both of them have the same geometric interpretation as the values of the (conditional) density at the boundary of the corresponding MV-set which are $\Gamma_{g(\cdot|x)}(\lambda_\alpha)$ and $(-\infty, q(\alpha|x)]$, respectively.

4 Discussion

Apart from its direct application in prediction of nonlinear and non-Gaussian time series (Polonik and Yao 2000, and references therein), a conditional empirical MV-set is also interesting (i) as an estimator for a level set of a conditional density, and (ii) to be used in tests for unimodality of conditional distributions. In this section, we discuss how the above theoretical results can be applied to these two applications.

First, we briefly illustrate how to derive the L_1-rate of convergence for a conditional empirical MV-set by applying Theorem 3.1 iteratively. It can be shown that the L_1-distance between the empirical and the true MV-set can be estimated from above by a sum of several terms including the difference of the empirical process and the generalized quantile process. (See Polonik and Yao 2000 for details.) Hence, Bahadur-Kiefer rates derived in Theorem 3.1 are useful. Note, however, an explicit rate σ^2 is needed in applying Theorem 3.1, and further, it is not necessary in Theorem 3.1 to let σ^2 converge to 0. Now we start with $\sigma^2 = 1$. Then Theorem 3.1 yields the first Bahadur-Kiefer approximation rate which in turn can be used to derive the first rate σ^2 for $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$. Further, this rate for $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ can be plugged into Theorem 3.1 to yield a faster Bahadur-Kiefer type approximation. This faster Bahadur-Kiefer rate leads to a faster rate of convergence for $\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x)$ and so on. The iteration will be continued until the rate of convergence is saturated.

The testing for unimodality of conditional distribution is an interesting and challenging problem in statistics. It has been observed that the conditional distribution of (nonlinear) time series given its lagged values could be multimodal. Further, the number of modes may vary over different places in the state space. Polonik and Yao (2000) proposed a heuristic device to detect the possible multimodality based on coverage probabilities of unconnected regions. A more rigorous statistical test can be constructed as follows based on conditional MV-sets. We only consider a special case when Y is univariate (*i.e.* $d' = 1$).

To predict Y from X, the best predictive region among a candidate class $\mathcal{C}$ is the MV-set of $\mathcal{C}$ in the sense that the MV-set has the minimum Lebesgue measure. Obviously this best predictor depends on the choice of the class $\mathcal{C}$. In view of simple prediction, there is strong temptation to let $\mathcal{C}$ be the class of all intervals $\mathcal{I}_1$. However, such a $\mathcal{C}$ is only pertinent when the conditional density $g(\cdot|x)$ is unimodal. Indeed, if $g(\cdot|x)$ is, for example, bimodal, we should let $\mathcal{C} = \mathcal{I}_2$ which is the class of unions of at most two intervals. In this case, the MV-set of $\mathcal{I}_1$ may have much larger Lebesgue measure than that of $\mathcal{I}_2$. Hence, the comparison of the volumes (Lebesgue measures) for the MV-sets in different set classes gives us the information on the number of modes of the underlying conditional distribution. This idea has been explored by Polonik (1997) in testing the unimodality for unconditional distribution.

To test the null hypothesis that $g(\cdot|x)$ is unimodal, we define the statistic

$$T_{n,A}(x) = \sup_{\alpha \in A}\{\hat{\mu}_{\mathcal{I}_2}(\alpha|x) - \hat{\mu}_{\mathcal{I}_1}(\alpha|x)\} \tag{4.1}$$

where $A \subset [0,1]$. Obviously, we may replace $\mathcal{I}_1$ and $\mathcal{I}_2$ in the above expression by appropriate $\mathcal{C}$ and $\mathcal{D}$ (with $\mathcal{C} \subset \mathcal{D}$) respectively for testing different hypotheses. Now, it follows from Theorem 3.1 and its proof that under the null hypothesis

$$\hat{\mu}_{\mathcal{I}_2}(\alpha|x) = \mu_{\mathcal{I}_2}(\alpha|x) + \mu'_{\mathcal{I}_2}(\alpha|x)\Big((\hat{F}_n - F)(\Gamma_{g(\cdot|x)}(\lambda_\alpha)|x) + (nh^d)^{-1/2}\omega_{\nu_{n,\mathcal{I}_2}}(\sigma_n^2)\Big) + R_n,$$

where $\omega_{\nu_{n,\mathcal{I}_2}}$ denotes the modulus of continuity of $\nu_{n,\mathcal{I}_2}$ which is the conditional empirical process indexed by $\mathcal{I}_2$, and σ_n^2 denotes the L_1-rate of convergence of $\widehat{\mathrm{M}}_{\mathcal{I}_2}(\alpha|x)$ to $\Gamma_{g(\cdot|x)}(\lambda_\alpha)$. The remainder term R_n is of smaller order. The analogous expansion also holds for $\hat{\mu}_{\mathcal{I}_1}(\alpha|x)$. Since under the null hypothesis $\mu_{\mathcal{I}_2}(\alpha|x) = \mu_{\mathcal{I}_1}(\alpha|x)$, the statistic $T_{n,\{\alpha\}}(x)$ converges to 0 under the null hypothesis and the rate of convergence is $(nh^d)^{-1/2}\omega_{\nu_{n,\mathcal{I}_2}}(\sigma_n^2)$. The rates given in Corollary 3.2 for $\gamma = 0$ are explicit rates for this quantity for some particular h. Since the statistic $T_{n,A}$ is defined as a supremum, we need to show that the results in Theorem 3.1 and Corollary 3.2 hold uniformly for $\alpha \in A$, which can be validated at least for $A \subset [\epsilon, 1-\epsilon](\epsilon > 0)$ under appropriate conditions on the smoothness of $g(\cdot|x)$ (see Polonik 1997 for the global *i.i.d.* case).

The idea of excess mass provides an alternative approach to test the uni- or multi-modality. The excess mass approach was introduced independently by Müller and Sawitzki (1987) and Hartigan (1987). (For further work see Nolan 1991, and Polonik 1995a,b). Adapted to the conditional empirical processes, the basic statistic is of the form $E_{n,\mathcal{C}}(\lambda|x) = \sup_{C\in\mathcal{C}}(\widehat{F}_n(C|x) - \lambda\,\mathrm{Leb}(C))$, which might be called a conditional empirical excess mass functional. As a function of λ it contains information about mass concentration of the underlying distribution. Similar to the above, we compare the functionals under different classes $\mathcal{C}$. Namely, we define the test statistic

$$T_n(x) = \sup_{\lambda>0}(E_{n,\mathcal{I}_2}(\lambda|x) - E_{n,\mathcal{I}_1}(\lambda|x)),$$

which is a conditional version of the test statistic proposed by Müller and Sawitzki (1987, 1991). The rates of convergence of $T_n(x)$ under the hypothesis of unimodality can be derived from Theorem 2.3. It can be shown that $\sqrt{nh}\,T_n(x)$ can be estimated from above by $\sup_{\lambda>0}(\nu_n(\Gamma_{n,\mathcal{I}_2}(\lambda)|x) - \nu_n(\Gamma_{g(\cdot|x)}(\lambda)|x))$. Here $\Gamma_{n,\mathcal{I}_2}(\lambda)|x) \in \mathcal{I}_2$ denotes the conditional empirical λ-cluster which is the maximizer of the excess mass statistic $E_{n,\mathcal{I}_2}(\lambda|x)$ defined above. See Polonik (1995a) for unconditional cases with *i.i.d.* observations. If L_1-rates of convergence σ_n^2 for the sets $\Gamma_{n,\mathcal{I}_2}(\lambda)$ to

$\Gamma_{g(\cdot|x)}(\lambda)$ can be derived, then we have $\sqrt{nh}\, T_n(x) \le O_P(\omega_{\nu_n,\mathcal{I}_2}(\sigma_n^2))$, and rates of convergence of the quantity on the right-hand side of the last inequality immediately follow from Theorem 2.3. The rates σ_n^2 can be derived by using ideas from Polonik (1995) and the results of the present paper.

Finally, we point out that the above test can be generalized to tests for other null hypotheses if we replace $\mathcal{I}_1$ and $\mathcal{I}_2$ by $\mathcal{C}$ and $\mathcal{D}$ (with $\mathcal{C} \subset \mathcal{D}$) in the definition of the test statistic. This generalization can be treated analogously, provided information about the metric entropy with bracketing of $\mathcal{D}$ (and hence also about $\mathcal{C}$) is known. This shows the actual strength of Theorem 2.3.

5 Appendix: Proofs

Some proofs of the above results are more or less standard nowadays. We outlines these proofs only. The proof of Theorem 2.3 is most technically involved, where we adapted the so-called chaining idea in empirical process theory. The key of the adaptation is to establish an exponential inequality (see Lemma 5.3 below) for which the proof is provided. Using this exponential inequality, Theorem 2.3 can be proved following the main ideas of the proof of Theorem 2.3 of Alexander (1984). Therefore the detailed derivation is omitted.

Throughout the proofs we use the notation:

$$\varphi_n(C|x) = \frac{1}{nh^d} \sum_{t=1}^{n} I\{Y_t \in C\} K\left(\frac{X_t - x}{h}\right), \tag{5.1}$$

and define

$$f_n(x) = \frac{1}{nh^d} \sum_{t=1}^{n} K(\frac{X_t - x}{h}). \tag{5.2}$$

The corresponding theoretical functions are $\varphi(C|x) = F(C|x)f(x)$ and $f(x)$ itself. We write $K_h(y) = \frac{1}{h^d} K\left(\frac{y}{h}\right)$. Moreover, unless stated otherwise x is assumed to fixed such that $f(x) > 0$. Let us first introduce two technical lemmas without proofs:

Lemma 5.1 *Suppose that f is continuous at x. Suppose further that f is bounded, and that* (2.4) *holds. Then we have $\forall\, C \in \mathcal{C}$ that as $n \to \infty$*

$$|E(\varphi_n(C|x)) - \varphi(C|x)| = o_P(1), \quad \textit{and} \quad |E(\widehat{F}_n(C|x)) - F(C|x)| = o_P(1).$$

If (2.4) *does hold uniformly over $C \in \mathcal{C}$ so do the assertions.*

In the following lemma we give the exact asymptotic behaviour of the bias terms. The proofs consist of tedious, but straightforward calculations using Taylor expansions. Details are omitted.

Lemma 5.2 *Suppose that (A2) – (A5) hold. Let* $\Psi_1(C|x) = \langle \nabla F(C|x), \int vK(v)\langle v, \nabla f(x)\rangle dv\rangle + \frac{1}{2} f(x) \int v^T \nabla^2 F(C|x) v K(v)\, dv \; + \; \frac{1}{2} F(C|x) \int v^T \nabla^2 f(x) v K(v)\, dv$ *and let* Ψ_2 *as defined in* (3.4). *Then we have for each* x *as* $n \to \infty$ *that uniformly in* $C \in \mathcal{C}$

$$(i) \quad h^{-2}(E\varphi_n(C|x) - \varphi(C|x)) \to \Psi_1(C|x)$$

$$(ii) \quad h^{-2}(E\widehat{F}_n(C|x) - F(C|x)) \to \Psi_2(C|x).$$

Proof of Theorem 2.1: We use the following decomposition:

$$\widehat{F}_n(C|x) - F(C|x) = \frac{1}{f(x)}(\varphi_n(C|x) - \varphi(C|x)) - \frac{\widehat{F}_n(C|x)}{f(x)}(f_n(x) - f(x)). \tag{5.3}$$

¿From this it is easy to see that we only need to show that as $n \to \infty$

$$\sup_{C \in \mathcal{C}} |\varphi_n(C|x) - \varphi(C|x)| = o_P(1) \quad \text{and} \tag{5.4}$$

$$|f_n(x) - f(x)| = o_P(1). \tag{5.5}$$

(5.5) is well known to hold under the present conditions (cf. Bosq 1996). That for every fixed $C \in \mathcal{C}$ we have $|\varphi_n(C|x) - \varphi(C|x)| = o_P(1)$ can be shown by similar arguments, and is omitted here. It remains to conclude *uniform* consistency from this by using finite metric entropy with inclusion. By observing that both ϕ_n and ϕ are monotonic functions in the sense that $\phi_{(n)}(C) \le \phi_{(n)}(D)$ for $C \subset D$, the bracketing idea can be used in exactly the same way as is done to proof uniform convergence of the empirical process (e.g. Pollard 1984, proof of Theorem 2). Details are omitted.

Proof of Theorem 2.2: Using the fact that under the given conditions $f_n(x)$ is consistent we have

$$\nu_n(C|x) = \sqrt{nh^d}\,\left(\frac{1}{f(x)}(\varphi_n(C|x) - \varphi(C|x)) - \frac{F(C|x)}{f(x)}(f_n(x) - f(x))\right)(1 + o_P(1)). \tag{5.6}$$

Using the notation $\tilde{b}_t(C) = I\{Y_t \in \mathcal{C}\} K(\frac{X_t - x}{h})$ we obtain

$$\nu_n(C|x) \quad = \quad \frac{1}{\sqrt{nh^d}} \sum_{t=1}^{n} W_{tn}(C|x) \tag{5.7}$$

$$+ \; \sqrt{nh^d}\, \frac{1}{f(x)}\Big(E\varphi_n(C) - \varphi(C)\Big) - \sqrt{nh^d}\, \frac{F(C|x)}{f(x)}\left(EK_h(X_t - x) - f(x)\right), \tag{5.8}$$

where $W_{tn}(C|x) = \frac{1}{f(x)}(\tilde{b}_t(C) - E\tilde{b}_t(C)) - \frac{F(C|x)}{f(x)}\left(K(X_t - x) - EK(X_t - x)\right)$. It is well known that under the assumptions of the theorem the bias of f_n converges to zero at a rate h^2. The same holds for the bias of $\varphi_n(C)$ (Lemma 5.2). Hence, the assumptions on h assure that the terms in (5.8) are asymptotically negligible. It remains to show that the right-hand side in (5.7) is asymptotically normal with the given variance. To see this, the proof of Theorem 2.3 of Bosq (1996) can easily by adapted. [1] As for adapting the estimates given there one can use the fact that $\tilde{b}_t(C) \leq K(\frac{X_t - x}{h})$. Calculation of the asymptotic variance-covariance matrix is lengthy but straightforward.

Proof of Theorem 2.3: We start with the decomposition of $\nu_n(C|x)$ given in (5.6) above. Since under the present assumptions $\sqrt{nh^d}(f_n(x) - f(x)) = O_P(1)$ (*e.g.* Bosq 1996), the second summand in the main term (inside the brackets) in (5.6) is of the order $O_P(\sigma^2)$. It remains to show, that $\sqrt{nh^d}\sup_{C \in \mathcal{C}_\sigma}(\varphi_n(C|x) - \varphi(C|x))$ is of the desired order.
To see this first note that $E\varphi_n(C|x) - \varphi_n(C|x)$ is of the (uniform) order $O_P(h^2)$. Hence, the assumption $\sqrt{nh^d}\, h^2 \leq \Lambda_\gamma(\sigma^2, n)$ ensures that the bias-terms is of the required order. Therefore, with $\tilde{\nu}_n(C|x) = \sqrt{nh^d}\,\{\varphi_n(C|x) - E\varphi_n(C|x)\}$ it remains to show that the assertion of the theorem holds with ν_n replaced by $\tilde{\nu}_n$. This is done by adapting the chaining idea (known from empirical process theory) to the present situation. The crucial ingredient is the following exponential inequality (5.9). Using this, the proof of Theorem 2.3 of Alexander (1984) (and the corresponding Correction (1987)) can be adapted to the present situation. This adaptation is somehow involved, because the free parameter r in (5.9) has to be chosen appropriately at various places throughout the proof, and one has to make sure that every choice satisfies the constraint $r \in [1, n/2]$. Details are omitted, however. Only the proof for (5.9) is presented.

Lemma 5.3 *Under the assumptions of Theorem 2.3, for each $\epsilon > 0$ and each integer $r \in [1, n/2]$ there exist positive constants c, c_1, c_2 such that for $C \in \mathcal{C}_\sigma$ and large enough n*

$$P(|\tilde{\nu}_n(C|x)| > \epsilon) \leq \; 4\exp\left(-\frac{\epsilon^2}{c\left(\sigma^2 + \sqrt{\frac{n}{h^d}}\frac{\epsilon}{r}\right)}\right) + \exp\left[-c_1\frac{n}{r} + c_2\left(\log r + (0 \vee \log\frac{n}{h^d\epsilon^2})\right)\right]. \tag{5.9}$$

Proof of Lemma 5.3: The following result is crucial. If $F(C|x) \geq \frac{h^d}{log(1/h^d)}$ and $\sigma^2 \geq h^2$, then

$$\sum_{k=1}^{\infty} \mathrm{Cov}\,(b_1(C), b_{k+1}(C)) = O\left(h^{-d}(F(C|x))\right). \tag{5.10}$$

[1] In the proof of Theorem 2.3 of Bosq (1996) $r^{3/4}$ has to be replaced by $r^{5/4}$ in the formula preceding (2.40).

In order to derive this result we proof the following two estimates:

$$\text{Cov}\,(b_1(C), b_{k+1}(C)) \;= O\left((F(C|x)h^{-d})^{1/p}\right) \quad \text{for } p > 1, \tag{5.11}$$

$$\text{Cov}\,(b_1(C), b_{k+1}(C)) \;= O\left(p'\alpha(k)^{1/p'} F(C|x)^{1-1/p'} \left(h^{-d}\right)^{1+1/p'}\right) \quad \text{for } p' > 1. \tag{5.12}$$

Estimate (5.11) can be seen as follows. With $1/p + 1/q = 1$ we have

$$\begin{aligned}
&\text{Cov}\,(b_1(C), b_{k+1}(C)) \le \int I\{Y_1 \in C\} K_h(X_1 - x) K_h(X_{k+1} - x) dP \\
\le\ & \left(\int \left(1\{Y_1 \in C\} K_h^{1/p}(X_1 - x)\right)^p dP\right)^{1/p} \cdot \left(\int \left(K_h^{1/q}(X_1 - x) K_h(X_{k+1} - x)\right)^q dP\right)^{1/q} \\
\le\ & c_1 \left(F(C|x) f(x) + O(h^2)\right)^{1/p} \cdot \left(\int K_h(u - x) du\right)^{1/q} \left(\int K_h^q(v - x) dv\right)^{1/q} \\
=\ & \left(F(C|x) f(x) + O(h^2)\right)^{1/p} \cdot O\left(h^{-d + d/q}\right) = O\left(\left((F(C|x) + h^2) h^{-d}\right)^{1/p}\right).
\end{aligned}$$

By assumption $F(C|x) \ge h^2$, and hence (5.11) follows. For the proof of (5.12) we use the following inequality. Let $p', q', r' > 1$ be integers with $1/p' + 1/q' + 1/r' = 1$, then

$$\text{Cov}(b_{k+1}(C), b_1(C)) \le 2\, p' (2\alpha(k))^{1/p'} ||b_{k+1}(C)||_{q'} ||b_1(C)||_{r'}. \tag{5.13}$$

Inequality (5.13) is known as Davydov's inequality, and is a corollary to the so-called Rio's inequality (see Bosq 1996, Corollary 1.1). Using similar arguments as above, one can see that

$$||b_{k+1}(C)||_{q'} = O\left((F(C|x) + h^2)^{1/q'} h^{-d(1-1/q')}\right)$$

and

$$||b_1(C)||_{r'} = O\left((F(C|x) + h^2)^{1/r'} h^{-d(1-1/r')}\right).$$

Plugging these results into (5.13) gives (5.12). Combining the estimates (5.11) and (5.12) leads to (5.10) by using similar arguments as in the proof of Theorem 2.1 in Bosq (1996).

We now indicate how to obtain the exponential inequality (5.9). We apply (1.26) from Bosq (1996) to $\tilde{\nu}_n(C|x) = \sqrt{nh^d}\, \frac{1}{n} \sum_{t=1}^{n} (b_t(C) - Eb_t(C))$. Since $|b_t(C)| \le M_1 h^{-d}$, for some constant $M_1 > 0$ this gives for $\epsilon > 0$

$$\begin{aligned}
P(|\tilde{\nu}_n(C|x)| > \epsilon) &= P\left(|\varphi_n(C|x) - E\varphi_n(C|x)| > \frac{\epsilon}{\sqrt{nh^d}}\right) \\
&\le 4 \exp\!\Big(-\frac{\epsilon^2 r}{8nh^d v^2(r)}\Big) + 22 \left(1 + \frac{4\, M_1\, n^{1/2}}{h^{d/2} \epsilon}\right)^{1/2} r\, \alpha\left(\left[\frac{n}{2r}\right]\right) \qquad (5.14)
\end{aligned}$$

where here $[x]$ denotes integer part of x, $\alpha(\cdot)$ denotes the strong mixing coefficient defined in (2.1), and r is an integer with $1 \le r \le n/2$. Further,

$$v^2(r) = \frac{8r^2}{n^2}\sigma^2(r) + \frac{c_1}{2n^{1/2}h^{3d/2}}\,\epsilon$$

and
$$\sigma^2(r) = O\left(s\,\mathrm{Var}(b_1(C)) + s\sum_{k=1}^{s-1} \mathrm{Cov}\,(b_1(C), b_{k+1}(C))\right),$$

where $s = n/2r$. It is well known that $h^d\mathrm{Var}(b_1(C)) = F(C|x)f(x)\int K^2 + O(h^2)$. Together with (5.10) this gives

$$\sigma^2(r) = O(sh^{-d}F(C|x)) = O(sh^{-d}\sigma^2), \tag{5.15}$$

which leads to the estimate

$$v^2(r) = O\left(s^{-1}h^{-d}\sigma^2 + \frac{\epsilon}{\sqrt{nh^{3d}}}\right). \tag{5.16}$$

Plugging this estimate into (5.14) we obtain for some constant $c > 0$

$$P(|\tilde\nu_n(C|x)| > \epsilon) \le 4\exp\left(-\frac{\epsilon^2}{c\left(\sigma^2 + \sqrt{\frac{n}{h^d}}\,\frac{\epsilon}{r}\right)}\right) + 22\left(1 + 4\epsilon^{-1}\sqrt{\frac{n}{h^d}}\right)^{1/2} r\,\gamma^s. \tag{5.17}$$

Using $r \le n/2$ we obtain for the last term the following estimate which completes the proof:

$$\begin{aligned} 22\left(1+4\epsilon^{-1}\sqrt{\tfrac{n}{h^d}}\right)^{1/2} r\,\gamma^s \;&\le \exp\left[-\frac{n}{2\,r}\log\frac{1}{\gamma} + \log\frac{n}{2} + \log 22 + \frac{1}{2}\log\left(1+4\epsilon^{-1}\sqrt{\frac{n}{h^d}}\right)\right] \\ &\le \exp\left[-c_1\frac{n}{r} + c_2\left(\log q + (0 \vee \log\frac{n}{h^d\epsilon^2})\right)\right]. \end{aligned}$$

Proof of Theorem 3.1: Let $q_n(\alpha) = \sqrt{nh^d}\,\frac{f(x)}{\mu_k'(\alpha|x)}\left(\widehat\mu_{\mathcal{C}}(\alpha|x) - \mu_{\mathcal{C}}(\alpha|x)\right)$. Using similar arguments as in Polonik (1997), proof of Lemma 7.1, one can see that on the set $B_n = \{d_{F(\cdot|x)}(\widehat{\mathrm{M}}_{\mathcal{C}}(\alpha|x), \mathrm{M}_{\mathcal{C}}(\alpha|x)) < \sigma^2\} \cup \{|\alpha_n^\pm - \alpha| < \sigma^2\} \cup \{\alpha_n^\pm \in (0,1)\}$ we have

$$\sqrt{nh^d}\,\frac{1}{\mu_k'(\alpha|x)}\left(\mu_{\mathcal{C}}(\alpha_n^-|x) - \mu_{\mathcal{C}}(\alpha|x)\right) \le q_n(\alpha) \le \sqrt{nh^d}\,\frac{1}{\mu_k'(\alpha|x)}\left(\mu_{\mathcal{C}}(\alpha_n^+|x) - \mu_{\mathcal{C}}(\alpha|x)\right) \tag{5.18}$$

where

$$\alpha_n^\pm = \alpha \pm ((\widehat F_n - F)(\mathrm{M}_{\mathcal{C}}(\alpha|x)|x) + \omega_{\nu_n(\cdot|x)}(\sigma^2). \tag{5.19}$$

and $\omega_{\nu_n(\cdot|x)}(\epsilon) = \sup_{\{C,D\in\mathcal{C}:\, d_{F(\cdot|x)}(C,D)<\epsilon\}}|\nu_n(C|x) - \nu_n(D|x)|$ denotes the modulus of continuity of the conditional empirical process $\nu_n(C|x) = \sqrt{nh^d}\;(\widehat F_n(C|x) - F(C|x))$. ¿From (5.18) and (5.19) together with the fact that $P(B_n) \to 0$ as $n \to \infty$ the assertion follows by applying a one-term Taylor expansion.

Proof of Theorem 3.3: First note that all the results proven above hold analogously if we choose $\mathcal{C} = \{(-\infty, y] : y \in \mathbf{R}\}$ and replace *Leb* by the function ν defined through $\nu((-\infty, y]) = y$. A first application of Theorem 2.3 with $\sigma^2 = 1$ shows that $\sup_{\{y \in \mathbf{R}\}} \sqrt{nh^d}\, |\widehat{F}_n(y|x) - F(y|x)| = O_P(1)$ as $n \to \infty$ (cf. Remark 2.4, (c1)). From this it follows that as $n \to \infty$

$$|\widehat{F}_n(q_n(\alpha|x)|x) - F(q_n(\alpha|x)|x)| = O_P\left((nh^d)^{-1/2}\right),$$

and since $\widehat{F}_n(q_n(\alpha|x)|x) = F(q(\alpha|x)|x) + o_P(1/nh^d)$ we obtain

$$|F(q_n(\alpha|x)|x) - F(q(\alpha|x)|x)| = O_P\left((nh^d)^{-1/2}\right). \tag{5.20}$$

Observing $|F(q_n(\alpha|x)|x) - F(q(\alpha|x)|x)| = d_F((-\infty, q_n(\alpha|x)], (-\infty, q(\alpha|x)])$, and applying Theorem 3.1 a second time, but now with $\sigma^2 = (nh^d)^{-1/2}$, gives the asserted rate of convergence in probability by choosing $h = n^{-\frac{1}{d+4}}$. Note that here we also have to take into account the bias which for this choice of h does not vanish (cf. Lemma 5.2).

As for the almost sure rate, note that the first rate (see (5.20) above) needs to hold almost surely. By using the exponential inequalities derived in the proof of Theorem 2.3 and applying Borel-Cantelli-Lemma one can see that $\sup_{\{y \in \mathbf{R}\}} \sqrt{nh^d}\, |\widehat{F}_n(y|x) - F(y|x)| = O(\sqrt{\log n})a.s.$ Then, as above, plugging in the resulting rate for $d_F((-\infty, q_n(\alpha|x)], (-\infty, q(\alpha|x)])$ into Theorem 2.3 and applying Borel-Cantelli-Lemma a second time leads to the assertion.

References

Alexander K.S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041-1067, Correction *Ann. Probab.* **15** 428-430.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rodgers, W.H., and Tukey, J.W. (1972). Robust estimation of location: survey and advances. Princeton Univ. Press, Princeton, N.J.

Andrews W.K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Inst. Stat. Review* **62** 119-132.

Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577-580.

Basu, G.J. and Singh, K. (1978): On deviations between empirical and quantile processes for mixing random variables. *J. Multivariate Anal.* **8** 532-549.

Bhattacharya, P.K. and Gangopadhyay, A.K. (1990): Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400-1415.

Bosq, D. (1996). Nonparametric statistics for stochastic processes. Lecture Notes in Statistics No. 110, Springer, New York.

Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré* **31** 393-427.

Dudley, R.M. (1974). Metric entropy of classes of sets with differentiable boundaries. *J. Approx. Theorie* **10** 227-236.

Dudley, R.M. (1984). A course in empirical processes. *Ecole d'Ete de Probabilites de Saint Flour XII-1982, Lecture Notes in Math.* **1097** 1-142, Springer, New York.

Einmahl, J.H.J. and Mason, D.M. (1992). Generalized quantile processes. *Ann. Statist.* **20**, 1062-1078.

Gangopadhyay A.K. and Sen, P.K. (1993). Contiguity in Bahadur type representation of a conditional quantile and application in conditional quantile process. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift,* J.K. Gosh, S.K. Mitra, K.R. Parathasarathy and B.L.S. Prakasa Rao *(eds.)* 219 - 232, Wiley Eastern Limited, Publishers

Hard, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer.

Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267-270.

Horvath, L. (1988). Asymptotics of conditional empirical processes. *J. Multivariate Analysis* **26** 184 - 206

Hyndman, R.J. (1995). Forecast regions for non-linear and non-normal time series models. *Int. J. Forecasting,* **14**, 431-441.

Hyndman, R.J. (1996). Computing and graphing highest density regions. *Ameri. Statist.* **50**, 361-365.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323-1342.

Kiefer, J. (1970). Deviations between the sample quantile process and the sample d.f. In *Non-Parametric techniques in statistical inference.* M.L. Puri, (Ed.). Cambridge University Press, London/New York

Lientz, B.P. (1970). Results on nonparametric modal intervals. *SIAM J. Appl. Math.* **19** 356-366

Massart, P. (1987). Invariance principles for empirical processes: the weakly dependent case. PhD-Thesis, Paris-Sud

Mehra, K.L., Sudhakara Rao, M., and Upadrasta, S.P. (1991). A smooth conditional quantile estimator and related applications of conditional empirical processes. *J. Multivariate Anal.* **37** 151-179

Müller, D.W. and Sawitzki, G. (1987). Using excess mass estimates to investigate the modality of a distribution. Preprint No. 398, SFB 123, Universität Heidelberg

Müller, D.W. and Sawitzki, G. (1991). Excess mass estimates and tests of multimodality. *J. Amer. Statist. Assoc.* **86** 738-746

Nolan, D. (1991). The excess mass ellipsoid. *J. Multivariate Anal.* **39** 348-371

Pollard, D. (1984). Convergence of stochastic processes. Springer, New York

Polonik, W. (1995a). Measuring mass concentration and estimating density contour clusters - an excess mass approach. *Ann. Statist.* **23** 855-881

Polonik, W. (1995). Density estimation under qualitative assumptions in higher dimensions. *J. Multivariate Anal.* **55** 61- 81

Polonik, W. (1997). Minimum volume sets and generalized quantile processes. *Stoch. Processes and Appl.* **69** 1-24

Polonik, W. and Yao, Q. (2000). Conditional minimum volume predictive regions for stochastic processes. *J. Ameri. Statist. Assoc.* **95**, 509-515

Sen, P.K. (1972): On the Bahadur's representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multivariate Anal.* **2** 77-95

Shorack G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

Stute, W. (1986a). Conditional empirical processes. *Ann. Statist.* **14** 638-647.

Stute, W. (1986b). On almost sure behaviour of conditional empirical distribution functions. *Ann. Probab.* **14** 891-901.

van der Vaart A. and Wellner J.A. (1996) *Weak convergence and empirical processes.* Springer, New York.

Xiang, X. (1995). On Bahadur-Kiefer representation of a kernel conditional quantile estimator. *Nonparametric Statistics* **5** 275-287

Xiang, X. (1996). A kernel estimator of a conditional quantile. *J. Multivariate Anal.* **59** 206-216

Yang, S.-S. (1985). A smooth nonparametric estimator of a quantile function . *J. Amer. Statist. Assoc.* **80**, 1004 - 1011.

Yoshihara, K. (1995): The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Stat. & Prob. Letters* **24** 299-304.