The Annals of Statistics
1992, Vol. 20, No. 2, 779–796

STATISTICAL INFERENCE FOR CONDITIONAL CURVES: POISSON PROCESS APPROACH

BY M. FALK AND R.-D. REISS

Katholische Universität Eichstätt and Universität GH Siegen

A Poisson approximation of a truncated, empirical point process enables us to reduce conditional statistical problems to unconditional ones. Let $(\mathbf{X}, \mathbf{Y})$ be a $(d+m)$-dimensional random vector and denote by $F(\cdot|\mathbf{x})$ the conditional d.f. of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. Applying our approach, one may study the fairly general problem of evaluating a functional parameter $T(F(\cdot|\mathbf{x}_1), \dots, F(\cdot|\mathbf{x}_p))$ based on independent replicas $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ of $(\mathbf{X}, \mathbf{Y})$. This will be exemplified in the particular cases of nonparametric estimation of regression means and regression quantiles besides other functionals.

1. Introduction. Influenced by the articles of Nadaraya (1964) and Watson (1964), the nonparametric estimation of the regression function $E(Y|\mathbf{X} = \cdot)$ aroused an increasing interest. Since $E(Y|\mathbf{X} = \mathbf{x})$ is the conditional mean, that is, the mean value of the conditional d.f.

$$F(\cdot|\mathbf{x}) = P(Y \le \cdot|\mathbf{X} = \mathbf{x})$$

of Y given $\mathbf{X} = \mathbf{x}$, one may write

$$E(Y|\mathbf{X} = \mathbf{x}) = \int y F(dy|\mathbf{x}).$$

For estimating the mean and the conditional mean one may utilize the sample mean and, respectively, certain conditional sample means as, for example, those defined in (4.3). Other functionals of the conditional d.f. are of interest as well, even if one is only interested in the question of evaluating the conditional mean. Recall that the median is another measure of the center of a distribution. Moreover, in robust statistics one studies trimmed means or solutions of certain equations to obtain estimators of the mean that are robust against errors and the deviation of the actual model from certain parametric models. The sample median is a robust statistic of that type in a limiting sense. Notice that an unconditional statistical procedure always has its conditional counterpart.

For a detailed discussion of such statistical topics we refer to Huber (1981) in the unconditional setup and, in the conditional case, to Stone (1977), as an early important reference, and Stute (1986), Härdle, Janssen and Serfling (1988), Samanta (1989), Truong (1989) and Manteiga (1990) to cite the most

Received September 1989; revised October, 1990.

AMS 1980 *subject classifications*. Primary 62J99; secondary 60G55.

Key words and phrases. Regression functionals, Poisson process, empirical process, Hellinger distance.

recent contributions. Our approach enables a unified treatment of such conditional questions to some extent. The primary aim of this paper is not to discuss various specific applications, but to present the new method.

The basic tool of our approach is a Poisson approximation of truncated, empirical processes. The line of research taken up by Falk and Reiss (1992) will be crucial for the present paper. Apart from replacing the original point process by a Poisson process with the same intensity measure [see also Deheuvels and Pfeifer (1988) and Reiss (1989)], we adopt the machinery available for Poisson processes to achieve a further simplification of the statistical model. The final Poisson process only depends on our target "parameter," namely, the conditional d.f. $F(\cdot|x)$.

In Section 2 the basic method and basic theorems are developed in the case of conditioning at a single point. The extension to several points is the topic of Section 3. Our examples in Section 4 will primarily concern the conditional (≡ regression) mean and the conditional median where the second parameter will be studied within the more general framework of conditional quantiles. Other functional parameters that are dealt with are the conditional d.f. itself and the functional parameter related to U-statistics. Finally, the relevance and importance of the projection pursuit technique is indicated in another example.

2. Approximate Poisson model. Let $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ be independent replicas of the random vector $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ is $\mathbb{R}^d$-valued and $\mathbf{Y}$ is $\mathbb{R}^m$-valued; that is, $(\mathbf{X}, \mathbf{Y})$ is $(d+m)$-dimensional. Let $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ be fixed. Our results will concern the conditional d.f. $F(\cdot|\mathbf{x})$ of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$.

Suppose that $(\mathbf{X}, \mathbf{Y})$ has a Lebesgue density, say $f(\mathbf{z}, \mathbf{y})$, for $\mathbf{z}$ near $\mathbf{x}$ and that the marginal density of $\mathbf{X}$, say g, is continuous at $\mathbf{x}$ with $g(\mathbf{x}) > 0$.

We consider only those observations from the sample $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ where the first coordinate lies in a small cube in $\mathbb{R}^d$ with center $\mathbf{x}$; that is, the statistical inference is based on $(\mathbf{X}_i, \mathbf{Y}_i)$ such that

$$\mathbf{X}_i \in \prod_{j=1}^{d} \left[x_j - a_{n,j}^{1/d}/2, x_j + a_{n,j}^{1/d}/2\right] =: \left[\mathbf{x} - \mathbf{a}_n^{1/d}/2, \mathbf{x} + \mathbf{a}_n^{1/d}/2\right]$$

with $\mathbf{a}_n = (a_{n,1}, \dots, a_{n,d}) \in (0, \infty)^d$ converging to zero as $n \to \infty$.

Speaking in terms of empirical point processes, we investigate the performance of the process N_n defined by

$$N_n(B) := \sum_{i=1}^{n} \varepsilon_{((\mathbf{X}_i - \mathbf{x})/\mathbf{a}_n^{1/d}, \mathbf{Y}_i)}(B \cap S), \qquad B \in \mathbb{B}^{d+m},$$

where $S := [-1/2, 1/2]^d \times \mathbb{R}^m$, $\varepsilon_z(\cdot)$ denotes the Dirac measure with mass at z and the operation $(\mathbf{X}_i - \mathbf{x})/\mathbf{a}_n^{1/d}$ is meant componentwise; that is, with $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,d})$ we have

$$(\mathbf{X}_i - \mathbf{x})/\mathbf{a}_n^{1/d} = \left((X_{i,j} - x_j)/a_{n,j}^{1/d}\right)_{j=1}^{d}.$$

Put

$$\gamma(n) := \prod_{j=1}^{d} a_{n,j}^{1/d}.$$

Notice that $\gamma(n) = c(n)$ if $a_{n,1} = \cdots = a_{n,d} = c(n)$.

We will prove in the following that, under certain regularity conditions on the density f, the truncated, empirical point process N_n may be approximated (in Hellinger distance H) by the Poisson process

$$N_n^* := \sum_{i=1}^{\tau(n)} \varepsilon_{(\mathbf{U}_i, \mathbf{W}_i)},$$

where $\tau(n)$ is a Poisson r.v. with parameter $ng(\mathbf{x})\gamma(n)$, $\mathbf{U}_i$ is uniformly distributed on $[-1/2, 1/2]^d$, $\mathbf{W}_i$ is distributed on $\mathbb{R}^m$ according to the conditional d.f. $F(\cdot|\mathbf{x})$ of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ and $\tau(n), \mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{W}_1, \mathbf{W}_2, \dots$ are independent.

Given random elements X and Y we will write

$$H(\mathscr{L}(X), \mathscr{L}(Y)) = \left(\int \left(p_X^{1/2} - p_Y^{1/2}\right)^2 d\mu\right)^{1/2}$$

for the Hellinger distance between the distributions $\mathscr{L}(X)$ and $\mathscr{L}(Y)$ of X and Y where p_X and p_Y are μ-densities of the distributions of X and Y.

THEOREM 1. *Suppose that for* $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)$ *in a small neighborhood of* $\mathbf{0} \in \mathbb{R}^d$ *the following condition holds*:

$$f(\mathbf{x} + \boldsymbol{\delta}, \mathbf{y})^{1/2} = f(\mathbf{x}, \mathbf{y})^{1/2}\{1 + O(|\boldsymbol{\delta}| r(\mathbf{y}))\}, \qquad \mathbf{y} \in \mathbb{R}^m, \tag{2.1}$$

for some real-valued function r *satisfying* $\int r^2(y) f(\mathbf{x}, \mathbf{y})\, d\mathbf{y} < \infty$.

Then if $|\mathbf{a}_n|$ *is sufficiently small*,

$$H(\mathscr{L}(N_n), \mathscr{L}(N_n^*)) = O\left(\gamma(n) + (n\gamma(n))^{1/2} |\mathbf{a}_n^{1/d}|\right).$$

The crucial point of the preceding approximation is that the random vectors $\mathbf{Y}_i$ belonging to the $\mathbf{X}_i$ in a neighborhood of $\mathbf{x}$ may approximately be replaced by some i.i.d. random vectors $\mathbf{W}_i$ that are distributed according to the conditional d.f. $F(\cdot|\mathbf{x})$ that is central to our investigations. This opens the way to tackle the estimation of functionals of $F(\cdot|\mathbf{x})$ by means of standard arguments.

REMARKS. (i) If $a_{n,1} = \cdots = a_{n,d} = c(n)$, then the bound in the preceding theorem reduces to $O(c(n) + (nc(n)^{(d+2)/d})^{1/2})$.

(ii) Condition (2.1), which is suggested by the mean value theorem, is closely related to the Hellinger differentiability of the path of conditional distributions

$$P_{\boldsymbol{\delta}} := P\{\mathbf{Y} \in \cdot | \mathbf{X} = \mathbf{x} + \boldsymbol{\delta}\} \to P_0 = P\{\mathbf{Y} \in \cdot | \mathbf{X} = \mathbf{x}\}, \qquad |\boldsymbol{\delta}| \to 0,$$

namely,

$$\left(\frac{f(\mathbf{x}+\boldsymbol{\delta},\mathbf{y})}{g(\mathbf{x}+\boldsymbol{\delta})}\right)^{1/2}=\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x})}\right)^{1/2}\left\{1+\frac{\langle\boldsymbol{\delta},\mathbf{h}(\mathbf{y})\rangle}{2}+O(|\boldsymbol{\delta}|r_{\boldsymbol{\delta}}(\mathbf{y}))\right\}$$

for some real-valued remainder function $r_{\boldsymbol{\delta}}$ satisfying

$$\int r_{\boldsymbol{\delta}}^2(\mathbf{y})f(\mathbf{x},\mathbf{y})\,d\mathbf{y}\to 0 \quad \text{as } |\boldsymbol{\delta}|\to 0,$$

and $\langle\mathbf{z},\mathbf{y}\rangle=\sum_{i=1}^d z_i y_i$ denoting the inner product on $\mathbb{R}^d$. Moreover, the $\mathbb{R}^d$-valued function $\mathbf{h}$ is called the Hellinger derivative of the path $P_{\boldsymbol{\delta}}$.

Hellinger differentiability, which was introduced by Le Cam (1966), is one of those differentiability concepts useful in connection with general asymptotic, statistical theory [see also Pfanzagl (1985)]. We particularly refer to Theorem 2 where a refinement of condition (2.1) will be utilized.

Finally, notice that $|\mathbf{a}_n^{1/d}|$ may be bounded away from zero, but $n\gamma(n)\to 0$ as $n\to\infty$.

The following two auxiliary results are crucial for the proof of Theorem 1. The first one is taken from Falk and Reiss [(1992), Theorem 2].

LEMMA 1. *Let $\tilde{N}_n$ be a Poisson point process on $\mathbb{R}^{d+m}$ with the same intensity measure μ_n as the empirical point process N_n; that is,*

$$\begin{aligned}\mu_n(B)&=E(\tilde{N}_n(B))=E(N_n(B))\\&=nP\{((\mathbf{X}-\mathbf{x})/\mathbf{a}_n^{1/d},\mathbf{Y})\in B\cap S\},\qquad B\in\mathbb{B}^{d+m}.\end{aligned}$$

Then,

$$\begin{aligned}H(\mathscr{L}(N_n),\mathscr{L}(\tilde{N}_n))&\le 3^{1/2}P\{((\mathbf{X}-\mathbf{x})/\mathbf{a}_n^{1/d},\mathbf{Y})\in S\}\\&=3^{1/2}P\{\mathbf{x}-\mathbf{a}_n^{1/d}/2\le\mathbf{X}\le\mathbf{x}+\mathbf{a}_n^{1/d}/2\}.\end{aligned}$$

The next result is immediate from an inequality given by Liese and Vajda [(1987), proof of Corollary 3.31].

LEMMA 2. *Let N_1, N_2 be Poisson point processes on $\mathbb{R}^k$ with intensity measures having densities p_1, p_2 w.r.t. some dominating measure ν. Then*

$$\begin{aligned}H^2(\mathscr{L}(N_1),\mathscr{L}(N_2))&=2\left(1-\exp\left(-\tfrac{1}{2}\int\left(p_1^{1/2}-p_2^{1/2}\right)^2d\nu\right)\right)\\&\le\int\left(p_1^{1/2}-p_2^{1/2}\right)^2d\nu\le\int|p_1-p_2|\,d\nu.\end{aligned}$$

Now we are ready to establish the proof of Theorem 1.

PROOF OF THEOREM 1. The intensity measure μ_n of N_n is determined by (if $|\mathbf{a}_n|$ is small)

$$\mu_n\Big([-1/2, t_i]_{i=1}^d \times (-\infty, s_j]_{j=1}^m\Big) = EN_n\Big([-1/2, t_i]_{i=1}^d \times (-\infty, s_j]_{j=1}^m\Big)$$
$$= nP\{\mathbf{Y} \le \mathbf{s}, \mathbf{x} - \mathbf{a}_n^{1/d}/2 \le \mathbf{X} \le \mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{t}\},$$

where $\mathbf{s} = (s_1, \dots, s_m) \in \mathbb{R}^m$ and $\mathbf{t} = (t_1, \dots, t_d) \in [-1/2, 1/2]^d$. The last term equals

$$n\int_{(-\infty,\mathbf{s}]}\int_{[\mathbf{x}-\mathbf{a}_n^{1/d}/2,\mathbf{x}+\mathbf{a}_n^{1/d}\mathbf{t}]} f(\mathbf{z},\mathbf{y})\,d\mathbf{z}\,d\mathbf{y}$$
$$= n\gamma(n)\int_{(-\infty,\mathbf{s}]}\int_{[-1/2,\mathbf{t}]} f(\mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{z},\mathbf{y})\,d\mathbf{z}\,d\mathbf{y};$$

that is, μ_n has the Lebesgue density

$$p_n(\mathbf{z},\mathbf{y}) = n\gamma(n) f(\mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{z},\mathbf{y}) 1_{[-1/2,1/2]^d}(\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^d, \mathbf{y} \in \mathbb{R}^m. \tag{2.2}$$

Let now N_n^{**} be a Poisson process with intensity measure μ_n. Lemma 1 implies

$$\begin{aligned} H(\mathscr{L}(N_n), \mathscr{L}(N_n^{**})) &\le CP\{\mathbf{x} - \mathbf{a}_n^{1/d}/2 \le \mathbf{X} \le \mathbf{x} + \mathbf{a}_n^{1/d}/2\} \\ &= C\int_{[\mathbf{x}-\mathbf{a}_n^{1/d}/2,\mathbf{x}+\mathbf{a}_n^{1/d}/2]} g(\mathbf{z})\,d\mathbf{z} \\ &= C\gamma(n)\int_{[-1/2,1/2]^d} g(\mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{z})\,d\mathbf{z} \\ &= O(\gamma(n)). \end{aligned} \tag{2.3}$$

Denote again by $F(\cdot|\mathbf{x})$ the conditional d.f. of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. The intensity measure μ_n^* of N_n^* is determined by

$$\mu_n^*([-1/2,\mathbf{t}] \times (-\infty,\mathbf{s}]) = ng(\mathbf{x})\gamma(n)F(\mathbf{s}|\mathbf{x})\prod_{i=1}^d (t_i + 1/2)$$
$$= ng(\mathbf{x})\gamma(n)\prod_{i=1}^d (t_i + 1/2)\int_{(-\infty,\mathbf{s}]} f(\mathbf{x},\mathbf{y})/g(\mathbf{x})\,d\mathbf{y}$$
$$= n\gamma(n)\int_{(-\infty,\mathbf{s}]}\int_{[-1/2,\mathbf{t}]} f(\mathbf{x},\mathbf{y})\,d\mathbf{z}\,d\mathbf{y},$$

where $\mathbf{t} \in [-1/2, 1/2]^d$, $\mathbf{s} \in \mathbb{R}^m$. Consequently, μ_n^* has the Lebesgue density

$$p_n^*(\mathbf{z},\mathbf{y}) = n\gamma(n) f(\mathbf{x},\mathbf{y}) 1_{[-1/2,1/2]^d}(\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^d, \mathbf{y} \in \mathbb{R}^m. \tag{2.4}$$

Lemma 2 yields

$$
\begin{aligned}
H^2(\mathscr{L}(N_n^{**}), \mathscr{L}(N_n^*)) &\le \int_{\mathbb{R}^{d+m}} \left(p_n(\mathbf{z},\mathbf{y})^{1/2} - p_n^*(\mathbf{z},\mathbf{y})^{1/2}\right)^2 d\mathbf{z}\, d\mathbf{y} \\
&= n\gamma(n) \int_S \left(f(\mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{z}, \mathbf{y})^{1/2} - f(\mathbf{x},\mathbf{y})^{1/2}\right)^2 d\mathbf{z}\, d\mathbf{y} \\
&= n\gamma(n) \int_{\mathbb{R}^m} \int_{[-1/2,1/2]^d} f(\mathbf{x},\mathbf{y}) \\
&\quad \times O\left(|\mathbf{a}_n^{1/d}\mathbf{z}|^2 r(\mathbf{y})^2\right) d\mathbf{z}\, d\mathbf{y} \\
&= O\left(n\gamma(n) |\mathbf{a}_n^{1/d}|^2\right)
\end{aligned} \tag{2.5}
$$

by condition (2.1). The assertion of Theorem 2 now follows from (2.3) and (2.5). □

According to Theorem 1 we know that those $\mathbf{X}_i$ falling into the cube $[\mathbf{x} - \mathbf{a}_n^{1/d}/2, \mathbf{x} + \mathbf{a}_n^{1/d}/2]$ can asymptotically be replaced by random vectors that are uniformly distributed on $[-1/2, 1/2]^d$. As a consequence there is asymptotically no loss of information if we consider, in place of the empirical point process N_n, the marginal process $N_n(\cdot|\mathbf{x})$ in the second component; compare this with the technical remark made in Section 4, Example 3. We have

$$
\begin{aligned}
N_n(A|\mathbf{x}) := N_n(\mathbb{R}^d \times A) &= \sum_{i=1}^{n} \varepsilon_{\mathbf{Y}_i}(A) \varepsilon_{\mathbf{X}_i}\left([\mathbf{x} - \mathbf{a}_n^{1/d}/2, \mathbf{x} + \mathbf{a}_n^{1/d}/2]\right) \\
&= \sum_{i=1}^{K(n)} \varepsilon_{\mathbf{Z}_i}(A), \qquad A \in \mathbb{B}^m,
\end{aligned}
$$

where

$$
K(n) = \sum_{i=1}^{n} \varepsilon_{\mathbf{X}_i}\left([\mathbf{x} - \mathbf{a}_n^{1/d}/2, \mathbf{x} + \mathbf{a}_n^{1/d}/2]\right)
$$

and the $\mathbf{Z}_i$ are defined by the $\mathbf{Y}_i$ values pertaining to those $\mathbf{X}_i$ lying in $[\mathbf{x} - \mathbf{a}_n^{1/d}/2, \mathbf{x} + \mathbf{a}_n^{1/d}/2]$.

Clearly, $N_n(\cdot|\mathbf{x})$ may be approximated by the corresponding marginal process $N_n^*(\cdot|\mathbf{x})$ of N_n^*. We have

$$
N_n^*(\cdot|\mathbf{x}) = N_n^*(\mathbb{R}^d \times \cdot) = \sum_{i=1}^{\tau(n)} \varepsilon_{\mathbf{W}_i}.
$$

By passing from the original process to the marginal one, the dimension of the involved space is reduced. This will enable an improvement of the bound given in Theorem 1 by imposing a stronger condition on f. This condition comes more closely to the Hellinger differentiability of the path

$$P_{\boldsymbol{\delta}} = P\{\mathbf{Y} \in \cdot | \mathbf{X} = \mathbf{x} + \boldsymbol{\delta}\}, \qquad |\boldsymbol{\delta}| \text{ small},$$

mentioned above.

THEOREM 2. *Suppose that for* $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)$ *in a small neighborhood of* $\mathbf{0} \in \mathbb{R}^d$ *the following condition holds*:

$$f(\mathbf{x} + \boldsymbol{\delta}, \mathbf{y})^{1/2} = f(\mathbf{x}, \mathbf{y})^{1/2}\{1 + \langle \boldsymbol{\delta}, \mathbf{h}(\mathbf{y})\rangle + O(|\boldsymbol{\delta}|^{1+\beta} r(y))\} \tag{2.6}$$

for some $\beta > 0$, *where* $\mathbf{h}$ *and* r *satisfy the conditions*

$$\int (|\mathbf{h}(\mathbf{y})|^4 + r(\mathbf{y})^4) f(\mathbf{x}, \mathbf{y})\, d\mathbf{y} < \infty.$$

Then, if $|\mathbf{a}_n|$ *is sufficiently small*,

$$H(\mathscr{L}(N_n(\cdot|\mathbf{x})), \mathscr{L}(N_n^*(\cdot|\mathbf{x}))) = O\big(\gamma(n) + (n\gamma(n))^{1/2} |\mathbf{a}_n^{1/d}|^{1+\min(1,\beta)}\big).$$

REMARK. If we choose $a_{n,1} = a_{n,2} = \cdots = a_{n,d} = c(n)$ and if $\beta = 1$, which is a typical value, then the preceding bound equals $O(c(n) + (nc(n)^{(d+4)/d})^{1/2})$ compared to $O(c(n) + (nc(n)^{(d+2)/d})^{1/2})$ in Theorem 1.

PROOF OF THEOREM 2. With the notation of the proof of Theorem 1 we have

$$\begin{aligned}
&H(\mathscr{L}(N_n(\cdot|\mathbf{x})), \mathscr{L}(N_n^*(\cdot|\mathbf{x}))) \\
&\quad \le H(\mathscr{L}(N_n(\cdot|\mathbf{x})), \mathscr{L}(N_n^{**}(\cdot|\mathbf{x}))) \\
&\qquad + H(\mathscr{L}(N_n^{**}(\cdot|\mathbf{x})), \mathscr{L}(N_n^*(\cdot|\mathbf{x}))) \\
&\quad \le H(\mathscr{L}(N_n), \mathscr{L}(N_n^{**})) + H(\mathscr{L}(N_n^{**}(\cdot|\mathbf{x})), \mathscr{L}(N_n^*(\cdot|\mathbf{x})))
\end{aligned} \tag{2.7}$$

by the monotonicity theorem due to Csiszár (1963) [see also Liese and Vajda (1987)] where $N_n^{**}(\cdot|\mathbf{x})$ denotes the marginal process in the second component pertaining to N_n^{**}.

From the proof of Theorem 1 we know that $H(\mathscr{L}(N_n), \mathscr{L}(N_n^{**})) = O(\gamma(n))$. Moreover, the densities of the intensity measures pertaining to the Poisson processes $N_n^{**}(\cdot|\mathbf{x})$ and $N_n^*(\cdot|\mathbf{x})$ on $\mathbb{R}^m$ are given by

$$p_n(\mathbf{y}|\mathbf{x}) = n\gamma(n) \int_{[-1/2,1/2]^d} f(\mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{z}, \mathbf{y})\, d\mathbf{z}, \qquad \mathbf{y} \in \mathbb{R}^m,$$

and

$$p_n^*(\mathbf{y}|\mathbf{x}) = n\gamma(n) f(\mathbf{x}, \mathbf{y}), \qquad \mathbf{y} \in \mathbb{R}^m.$$

Consequently, by Lemma 2 and by condition (2.6) we obtain

$$
\begin{aligned}
&H^2(\mathscr{L}(N_n^{**}(\cdot|\mathbf{x})), \mathscr{L}(N_n^{*}(\cdot|\mathbf{x}))) \\
&\quad \le \int_{\mathbb{R}^m} \left(p_n(\mathbf{y}|\mathbf{x})^{1/2} - p_n^*(\mathbf{y}|\mathbf{x})^{1/2}\right)^2 d\mathbf{y} \\
&\quad = n\gamma(n) \int_{\mathbb{R}^m} \left\{ \left(\int_{[-1/2,1/2]^d} f(\mathbf{x} + \mathbf{a}_n^{1/d}\mathbf{z}, \mathbf{y})\, d\mathbf{z} \right)^{1/2} - f(\mathbf{x},\mathbf{y})^{1/2} \right\}^2 d\mathbf{y} \\
&\quad = n\gamma(n) \int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y}) \left\{ \left[\int_{[-1/2,1/2]^d} \left(1 + \langle \mathbf{a}_n^{1/d}\mathbf{z}, \mathbf{h}(\mathbf{y})\rangle \right.\right.\right. \qquad (2.8) \\
&\qquad\qquad \left.\left.\left. + O\left(|\mathbf{a}_n^{1/d}\mathbf{z}|^{1+\beta} r(\mathbf{y})\right)\right)^2 d\mathbf{z} \right]^{1/2} - 1 \right\}^2 d\mathbf{y} \\
&\quad \le n\gamma(n) \int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y}) \left\{ \int_{[-1/2,1/2]^d} \left[\left(1 + \langle \mathbf{a}_n^{1/d}\mathbf{z}, \mathbf{h}(\mathbf{y})\rangle \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. + O\left(|\mathbf{a}_n^{1/d}\mathbf{z}|^{1+\beta} r(\mathbf{y})\right)\right)^2 - 1 \right] d\mathbf{z} \right\}^2 d\mathbf{y},
\end{aligned}
$$

where the last inequality follows from $|a^{1/2} - 1| \le |a - 1|$, $a \ge 0$.

The inner integral is of order

$$
O\Big(|\mathbf{a}_n^{1/d}|^2 |\mathbf{h}(\mathbf{y})|^2 + |\mathbf{a}_n^{1/d}|^{2(1+\beta)} r(y)^2 + |\mathbf{a}_n^{1/d}|^{1+\beta} |r(\mathbf{y})| + |\mathbf{a}_n^{1/d}|^{2+\beta} |\mathbf{h}(\mathbf{y})| |r(\mathbf{y})| \Big)
$$

by the fact that

$$
\int_{[-1/2,1/2]^d} \langle \mathbf{a}_n^{1/d}\mathbf{z}, \mathbf{h}(\mathbf{y})\rangle\, d\mathbf{z} = 0.
$$

Consequently,

$$
H(\mathscr{L}(N_n^{**}(\cdot|\mathbf{x})), \mathscr{L}(N_n^{*}(\cdot|\mathbf{x}))) = O\left((n\gamma(n))^{1/2} |\mathbf{a}_n^{1/d}|^{1+\min(1,\beta)} \right). \qquad (2.9)
$$

Combining (2.7) and (2.9) the proof is complete. □

3. Extension to several points. Next we will generalize the preceding results for a single point $\mathbf{x}$ to a set $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ of several points where $p = p(n)$ may increase as n increases.

Consider now only those observations from the sample $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ where the first coordinate lies in one of the cubes with center $\mathbf{x}_v$, $v = 1, \dots, p$;

that is, we consider only those $(\mathbf{X}_i, \mathbf{Y}_i)$ such that

$$\mathbf{X}_i \in \bigcup_{v=1}^{p} I_v,$$

where

$$I_v := I_{v,n} := \left[\mathbf{x}_v - \mathbf{a}_{n,v}^{1/d}/2, \mathbf{x}_v + \mathbf{a}_{n,v}^{1/d}/2\right], \qquad v = 1, \dots, p,$$

and

$$\mathbf{x}_v = (x_{v,1}, \dots, x_{v,d}) \in \mathbb{R}^d,$$

$$\mathbf{a}_{v,n} = (a_{v,n,1}, \dots, a_{v,n,d}) \in (0, \infty)^d$$

for $v = 1, \dots, p$.

In this sequel, we suppose that I_v, $1 \le v \le p$, are pairwise disjoint and that the marginal density of $\mathbf{X}$, say g, is continuous at $\mathbf{x}_v$ with $g(\mathbf{x}_v) > 0$ for $v = 1, \dots, p$.

We therefore consider the vector $(N_{n,1}, \dots, N_{n,p})$ of truncated, empirical point processes on $S^p = ([-1/2, 1/2]^d \times \mathbb{R}^m)^p$ where

$$N_{n,v}(B) := \sum_{i=1}^{n} \varepsilon_{((\mathbf{X}_i - \mathbf{x}_v)/\mathbf{a}_{v,n}^{1/d}, Y_i)}(B \cap S), \qquad B \in \mathbb{B}^{d+m}.$$

The random vector $(N_{n,1}, \dots, N_{n,p})$ will be approximated in Hellinger distance by the vector $(N_{n,1}^*, \dots, N_{n,p}^*)$ of independent Poisson processes where

$$N_{n,v}^* := \sum_{i=1}^{\tau_v(n)} \varepsilon_{(\mathbf{U}_{v,i}, \mathbf{W}_{v,i})}, \qquad v = 1, \dots, p,$$

$\tau_v(n)$ is a Poisson r.v. with parameter

$$ng(\mathbf{x}_v) \prod_{i=1}^{d} a_{v,n,i}^{1/d} =: ng(\mathbf{x}_v)\gamma_v(n),$$

$\mathbf{U}_{v,i}$ is uniformly distributed on $[-1/2, 1/2]^d$, $\mathbf{W}_{v,i}$ is distributed according to the conditional distribution $F(\cdot|\mathbf{x}_v)$ of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}_v$ and $\tau_v(n), \mathbf{U}_{v,1}, \mathbf{U}_{v,2}, \dots, \mathbf{W}_{v,1}, \mathbf{W}_{v,2}, \dots$, $v = 1, \dots, p$, are mutually independent. This result is the content of the next theorem.

THEOREM 3. *Suppose that for* $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)$ *in a small neighborhood of* $\mathbf{0} \in \mathbb{R}^d$ *and* $v = 1, \dots, p$,

$$(3.1) \qquad f(\mathbf{x}_v + \boldsymbol{\delta}, y)^{1/2} = f(\mathbf{x}_v, \mathbf{y})^{1/2}\{1 + O(|\boldsymbol{\delta}| r_v(\mathbf{y}))\}, \qquad \mathbf{y} \in \mathbb{R}^m,$$

for some real-valued function r_v *satisfying* $\int r_v^2(\mathbf{y}) f(\mathbf{x}, \mathbf{y})\, d\mathbf{y} < \infty$. *Then, if the* *bandwidth* $|\mathbf{a}_{v,n}|$ *is sufficiently small,*

$$H\big(\mathscr{L}(N_{n,1},\dots,N_{n,p}),\mathscr{L}(N^*_{n,1},\dots,N^*_{n,p})\big) = O\left[\sum_{v=1}^{p}\gamma_v(n) + \left(\sum_{v=1}^{p} n\gamma_v(n)\big|\mathbf{a}_{v,n}^{1/d}\big|^2\right)^{1/2}\right].$$

According to Theorem 3 the original processes may approximately be replaced by independent Poisson processes, a fact, that considerably simplifies further statistical analysis. The independence of $N^*_{n,1},\dots,N^*_{n,p}$ originates from the property of a Poisson process that the process evaluated at pairwise disjoint sets yields a sequence of independent r.v.'s.

REMARK. If we choose $a_{v,n,1} = \cdots = a_{v,n,d} = c(n)$, then the bound in the preceding theorem reduces to $O(pc(n) + (pnc(n)^{(d+2)/d})^{1/2})$.

PROOF OF THEOREM 3. Denote by $\tilde{N}_{n,v}$ the nonstandardized empirical point process pertaining to $\mathbf{x}_v$; that is,

$$\tilde{N}_{n,v}(B) := \sum_{i=1}^{n}\varepsilon_{(\mathbf{X}_i,\mathbf{Y}_i)}(B\cap(I_v\times\mathbb{R}^m)), \qquad B\in\mathbb{B}^{d+m}, v=1,\dots,p,$$

and put for $B\in\mathbb{B}^{d+m}$,

$$\tilde{N}_n(B) := \sum_{v=1}^{p}\tilde{N}_{n,v}(B) = \sum_{i=1}^{n}\varepsilon_{(\mathbf{X}_i,\mathbf{Y}_i)}\left(B\cap\left(\bigcup_{v=1}^{p} I_v\times\mathbb{R}^m\right)\right).$$

Let now $\tilde{N}^{**}_n$ be a Poisson process with the same intensity measure as $\tilde{N}_n$. In analogy to the proof of Theorem 1 we obtain from Lemma 1,

$$H\big(\mathscr{L}(\tilde{N}_n),\mathscr{L}(\tilde{N}^{**}_n)\big) = O\left(\sum_{v=1}^{p}\gamma_v(n)\right). \tag{3.2}$$

Moreover, for $v=1,\dots,p$ put

$$\tilde{N}^{**}_{n,v}(B) := \tilde{N}^{**}_n(B\cap(I_v\times\mathbb{R}^m)), \qquad B\in\mathbb{B}^{d+m}.$$

Then $\tilde{N}^{**}_{n,v}$, $v=1,\dots,p$, are independent Poisson processes since I_v, $v=1,\dots,p$, are pairwise disjoint and, moreover,

$$\tilde{N}^{**}_n = \sum_{v=1}^{p}\tilde{N}^{**}_{n,v}.$$

Next observe that

$$\begin{aligned} &H\big(\mathscr{L}(\tilde{N}_{n,1},\dots,\tilde{N}_{n,p}),\mathscr{L}(\tilde{N}^{**}_{n,1},\dots,\tilde{N}^{**}_{n,p})\big)\\ &\quad= H\big(\mathscr{L}(\tilde{N}_n),\mathscr{L}(\tilde{N}^{**}_n)\big) = O\left(\sum_{i=1}^{p}\gamma_v(n)\right). \end{aligned} \tag{3.3}$$

Finally, denote by $N^{**}_{n,v}$ the standardized version of $\tilde{N}^{**}_{n,v}$. We have

$$\begin{aligned}
&H\big(\mathscr{L}(N_{n,1},\dots,N_{n,p}),\mathscr{L}(N^*_{n,1},\dots,N^*_{n,p})\big)\\
&\quad\le H\big(\mathscr{L}(N_{n,1},\dots,N_{n,p}),\mathscr{L}(N^{**}_{n,1},\dots,N^{**}_{n,p})\big)\\
&\qquad+H\big(\mathscr{L}(N^{**}_{n,1},\dots,N^{**}_{n,p}),\mathscr{L}(N^*_{n,1},\dots,N^*_{n,p})\big)\\
&\quad\le H\Big(\mathscr{L}\big(\tilde{N}_{n,1},\dots,\tilde{N}_{n,p}\big),\mathscr{L}\big(\tilde{N}^{**}_{n,1},\dots,\tilde{N}^{**}_{n}\big)\Big)\\
&\qquad+H\big(\mathscr{L}(N^{**}_{n,1},\dots,N^{**}_{n,p}),\mathscr{L}(N^*_{n,1},\dots,N^*_{n,p})\big)\\
&\quad\le O\left(\sum_{v=1}^{p}\gamma_v(n)\right)+\left(\sum_{v=1}^{p}H^2(\mathscr{L}(N^{**}_{n,v}),\mathscr{L}(N^*_{n,v}))\right)^{1/2}\\
&\quad=O\left[\sum_{v=1}^{p}\gamma_v(n)+\left(\sum_{v=1}^{p}\left(n\gamma_v(n)|\mathbf{a}_n^{1/d}|^2\right)\right)^{1/2}\right]
\end{aligned}$$

by the arguments in the proof of Theorem 1. □

Now let $N_n(\cdot|\mathbf{x}_v)$ and $N^*_n(\cdot|\mathbf{x}_v)$ be defined as in Section 2. Notice that $N^*_n(\cdot|\mathbf{x}_1),\dots,N^*_n(\cdot|\mathbf{x}_p)$ are independent. The following extension of Theorem 2 is now straightforward by using the arguments of the preceding proof.

THEOREM 4. *Suppose that for* $\boldsymbol{\delta}=(\delta_1,\dots,\delta_d)$ *in a small neighborhood of* $\mathbf{0}\in\mathbb{R}^d$ *and* $v=1,\dots,p$,

$$f(\mathbf{x}_v+\boldsymbol{\delta},\mathbf{y})^{1/2}=f(\mathbf{x}_v,\mathbf{y})^{1/2}\left\{1+\langle\boldsymbol{\delta},\mathbf{h}_v(\mathbf{y})\rangle+O\big(|\boldsymbol{\delta}|^{1+\beta_v}r_v(\mathbf{y})\big)\right\} \tag{3.4}$$

for some $\beta_v>0$, *where* $\mathbf{h}_v$ *and* r_v *satisfy* $\int(|\mathbf{h}_v(\mathbf{y})|^4+r_v(\mathbf{y})^4)f(\mathbf{x}_v,\mathbf{y})\,d\mathbf{y}<\infty$.
Then, if $|\mathbf{a}_{v,n}|$ *is small,*

$$\begin{aligned}
&H\Big(\mathscr{L}\big(N_n(\cdot|\mathbf{x}_1),\dots,N_n(\cdot|\mathbf{x}_p)\big),\mathscr{L}\big(N^*_n(\cdot|\mathbf{x}_1),\dots,N^*_n(\cdot|\mathbf{x}_p)\big)\Big)\\
&\quad=O\left[\sum_{v=1}^{p}\gamma_v(n)+\left(\sum_{v=1}^{p}\left(n\gamma_v(n)|\mathbf{a}_{v,n}^{1/d}|^{2(1+\min(1,\beta_v))}\right)\right)^{1/2}\right].
\end{aligned}$$

REMARK. If we choose $a_{v,n,1}=\cdots=a_{v,n,d}=c(n)$, $v=1,\dots,p$, and if $\beta=1$, then the preceding bounds equals $O(pc(n)+(pnc(n)^{(d+4)/d})^{1/2})$.

4. Applications to conditional curve estimation. We are interested in the performance of functionals of the empirical processes N_n or $N_n(\cdot|x)$ as estimators of functional parameters of the conditional d.f. $F(\cdot|x)$. In Sections 2 and 3 we calculated the error that is made when the empirical processes are replaced by the Poisson processes N^*_n and $N^*_n(\cdot|x)$. It remains to deal with i.i.d. random variables W_i with common d.f. $F(\cdot|x)$ and a random sample size given by a Poisson r.v. Making use of a further reduction (cf. Theorem 5 and Lemma 3), classical results for nonrandom sample sizes can be made applicable.

Let us recall some basic concepts of standard statistical inference. Denote by F_n the empirical d.f. based on n i.i.d. random variables with common d.f. F. To estimate a real-valued functional parameter $T(F)$ we may utilize a statistical functional $T(F_n)$ or $T_n(F_n)$. We give a list of examples:

$T(F)$	$T(F_n)$ or $T_n(F_n)$	
$F(t)$	empirical d.f.	$F_n(t)$
$F^{-1}(q)$	empirical q-quantile	$F_n^{-1}(q)$
$\int z\,dF(z)$	empirical mean	$\int z\,dF_n(z)$
$\int\int h(z,y)\,dF(z)\,dF(y)$	U-statistic	$\int\int h(z,y)\,dF_n(z)\,dF_n(y)$
$F(t)$	kernel estimator	$F_{n,\beta(n)}(t)$
$F^{-1}(q)$	kernel estimator	$F^{-1}_{n,\beta(n)}(q)$,

where given a kernel u with $\int u(y)\,dy = 1$,

$$F_{n,\beta(n)}(t) = \int \frac{1}{\beta(n)} u\left(\frac{t-y}{\beta(n)}\right) F_n(y)\,dy = \int U\left(\frac{t-y}{\beta(n)}\right) dF_n(y)$$

and

$$F^{-1}_{n,\beta(n)}(q) = \int_0^1 \frac{1}{\beta(n)} u\left(\frac{q-y}{\beta(n)}\right) F_n^{-1}(y)\,dy$$

with $U(z) = \int_{-\infty}^z u(y)\,dy$. In the conditional setup we replace:

(a) the d.f. F by the conditional d.f. $F(\cdot|x)$;

(b) the functional T_n, if depending on n, by $T_{K(n)}$ where again

$$K(n) = \sum_{i=1}^n 1_{[x-a_n/2,\,x+a_n/2]}(X_i);$$

(c) the empirical d.f. F_n by the conditional empirical d.f. $F_n(\cdot|x)$ defined by

$$\begin{aligned} F_n(t|x) &= \frac{1}{K(n)} \sum_{i=1}^n 1_{(-\infty,t]}(Y_i) 1_{[x-a_n/2,\,x+a_n/2]}(X_i) \\ &= \frac{1}{K(n)} \sum_{i=1}^{K(n)} 1_{(-\infty,t]}(Z_i) \end{aligned} \tag{4.1}$$

where the Z_i are defined by the y-values pertaining to those X_i lying in the interval $[x - a_n/2, x + a_n/2]$. Put $F_n(\cdot|x) = 0$ if $K(n) = 0$.

Notice that there is a one-to-one correspondence between the empirical point process $N_n(\cdot|x)$ in Section 2 and the conditional, empirical d.f. $F_n(\cdot|x)$ (or, alternatively, the pertaining empirical counting process). The use of empirical d.f.'s instead of empirical measures merely acknowledges traditions in statistics.

In the following considerations the standard normal distribution is taken as the limiting distribution because this is perhaps the most important case. Other examples may be obtained by treating sample extremes and conditional sample extremes where the limiting distributions are different from the normal one.

THEOREM 5. *Suppose that, with* $\sigma > 0$, $\delta \in (0,1)$ *and* $C > 0$,

$$\sup_t \left| P\left\{ \frac{k^{1/2}}{\sigma}\big(T_k(F_k) - T(F(\cdot|x))\big) \le t \right\} - \Phi(t) \right| \le Ck^{-\delta}, \qquad k \in \mathbb{N}, \tag{4.2}$$

where the empirical d.f. F_k *is based on* k *i.i.d. random variables with common d.f.* $F(\cdot|x)$; *it is implicitly assumed that the r.v.'s* $T_k(F_k)$ *are measurable.*

Then, with $F_n(\cdot|x)$ *denoting the conditional empirical d.f. in* (4.1), *we have*

$$\sup_t \left| P\left\{ \frac{(na_n g(x))^{1/2}}{\sigma}\big(T_{K(n)}(F_n(\cdot|x)) - T(F(\cdot|x))\big) \le t \right\} - \Phi(t) \right|$$
$$\le D(na_n g(x))^{-\min(1/2,\delta)} + H(\mathscr{L}(N_n(\cdot|x)), \mathscr{L}(N_n^*(\cdot|x))),$$

where $D > 0$ *only depends on* C [*with the convention that* $T_{K(n)}$ *is replaced by* T *if* (4.2) *is formulated for* T *in place of* T_k].

PROOF. Recall that $F_n(\cdot|x)$ is the empirical d.f. pertaining to $N_n(\cdot|x)$. In analogy to this notation one may write

$$F_n^*(\cdot|x) = \frac{1}{\tau(n)} \sum_{i=1}^{\tau(n)} 1_{(-\infty,t]}(W_i)$$

for the empirical d.f. pertaining to $N_n^*(\cdot|x) = \sum_{i=1}^{\tau(n)} \varepsilon_{W_i}$. Recall that $W_1, W_2, W_3, \ldots$ are i.i.d. r.v.'s with common d.f. $F(\cdot|x)$. If F_k denotes the empirical d.f. based on the sample $W_1, \ldots, W_k$ of size k, then, obviously,

$$F_{\tau(n)} = F_n^*(\cdot|x).$$

Check that $T_{K(n)}(F_n(\cdot|x))$ may be replaced by $T_{\tau(n)}(F_n^*(\cdot|x)) = T_{\tau(n)}(F_{\tau(n)})$ within the error bound $H(\mathscr{L}(N_n(\cdot|x)), \mathscr{L}(N_n^*(\cdot|x)))$. Now the asserted inequality is immediate from Lemma 3 applied to $V_k = T_k(F_k)$, $\tau = \tau(n)$ and $\mu = T(F(\cdot|x))$. □

LEMMA 3. *Let* $\sigma > 0$, $\delta \in (0,1)$, $C > 0$ *and* $\mu \in \mathbb{R}$ *be such that for the sequence of r.v.'s* $V_1, V_2, V_3, \ldots$,

$$\sup_t \left| P\left\{ \frac{k^{1/2}}{\sigma}(V_k - \mu) \le t \right\} - \Phi(t) \right| \le Ck^{-\delta}, \qquad k \in \mathbb{N}.$$

Let τ *be a Poisson r.v. with parameter* $\lambda \ge 1$ *that is independent of* $V_1, V_2, V_3, \ldots$. *Then,*

$$\sup_t \left| P\left\{ \frac{\lambda^{1/2}}{\sigma}(V_\tau - \mu) \le t \right\} - \Phi(t) \right| \le D\lambda^{-\min(1/2,\delta)},$$

where D *only depends on* C. (*By convention we have* $V_\tau = 0$ *if* $\tau = 0$.)

PROOF. We have

$$\begin{aligned}
&\sup_t \left| P\left\{ \frac{\lambda^{1/2}}{\sigma}(V_\tau - \mu) \le t \right\} - \Phi(t) \right| \\
&\quad = \sup_t \left| \int \left(P\left\{ \frac{\lambda^{1/2}}{\sigma}(V_k - \mu) \le t \right\} - \Phi(t) \right) d\mathscr{L}(\tau)(k) \right| \\
&\quad \le \int \left(\sup_t \left| P\left\{ \frac{k^{1/2}}{\sigma}(V_k - \mu) \le t \right\} - \Phi\left(t \left(\frac{\lambda}{k} \right)^{1/2} \right) \right| \right) d\mathscr{L}(\tau)(k) \\
&\quad \le \int \left(\sup_t \left| P\left\{ \frac{k^{1/2}}{\sigma}(V_k - \mu) \le t \right\} - \Phi(t) \right| \right) d\mathscr{L}(\tau)(k) \\
&\qquad + \int_{\mathbb{N}} \left(\sup_t \left| \Phi(tk^{1/2}) - \Phi(t\lambda^{1/2}) \right| \right) d\mathscr{L}(\tau)(k) + e^{-\lambda}.
\end{aligned}$$

Applying Lemma 4 below we conclude that the first term is bounded by $DE(\max(\tau, 1))^{-\delta}) \le D\lambda^{-\delta}$, where D denotes here and in the following a generic constant.

For the second term we have, by the mean value theorem,

$$\begin{aligned}
&\int_{\mathbb{N}} \left(\sup_t \left| \varphi(tk^{1/2}) - \varphi(t\lambda^{1/2}) \right| \right) d\mathscr{L}(\tau)(k) \\
&\qquad = \int_{\mathbb{N}} \left(\sup_t \{ \varphi(\xi)|t|\, |k^{1/2} - \lambda^{1/2}| \} \right) d\mathscr{L}(\tau)(k),
\end{aligned}$$

where ξ is between $tk^{1/2}$ and $t\lambda^{1/2}$. From the monotonicity properties of φ we deduce that the preceding term is bounded by

$$\begin{aligned}
&\int_{\mathbb{N}} \sup_t \{ (\varphi(tk^{1/2}) + \varphi(t\lambda^{1/2}))|t| \} |k^{1/2} - \lambda^{1/2}|\, d\mathscr{L}(\tau)(k) \\
&\quad \le \int_{\mathbb{N}} \sup_t \{ \varphi(t)|t| \} \frac{k^{1/2} - \lambda^{1/2}}{k^{1/2}}\, d\mathscr{L}(\tau)(k) \\
&\qquad + \int \sup_t \{ \varphi(t)|t| \} \frac{k^{1/2} - \lambda^{1/2}}{\lambda^{1/2}}\, d\mathscr{L}(\tau)(k) \\
&\quad \le DE\left(\frac{|\tau^{1/2} - \lambda^{1/2}|}{(\max(\tau, 1))^{1/2}} \right) + \frac{D}{\lambda^{1/2}} E(|\tau^{1/2} - \lambda^{1/2}|) \\
&\quad \le DE\left(\frac{1}{\max(\tau, 1)} \right)^{1/2} E\left(|\tau^{1/2} - \lambda^{1/2}|^2 \right)^{1/2} + \frac{D}{\lambda^{1/2}} E(|\tau^{1/2} - \lambda^{1/2}|) \\
&\quad \le \frac{D}{\lambda^{1/2}} E\left(\left(\frac{|\tau - \lambda|}{\tau^{1/2} + \lambda^{1/2}} \right)^2 \right)^{1/2} + \frac{D}{\lambda^{1/2}} E\left(\frac{|\tau - \lambda|}{\tau^{1/2} + \lambda^{1/2}} \right) \\
&\quad \le \frac{D}{\lambda} E\left((\tau - \lambda)^2 \right)^{1/2} = \frac{D}{\lambda^{1/2}},
\end{aligned}$$

which completes the proof. □

The following lemma has been crucial for the derivation of the preceding auxiliary result.

LEMMA 4. *Let τ be Poisson distributed with parameter $\lambda \geq 1$. Then*

$$E\big((\max(\tau, 1))^{-\delta}\big) \leq 2\lambda^{-\delta}$$

for any $\delta \in (0, 1)$.

PROOF. We will prove that

$$E\big((\max(\tau, 1))^{-\delta}\big) \leq \frac{2}{\lambda} E(\tau^{1-\delta})$$

and, hence, the Cauchy–Schwarz inequality yields

$$E(\tau \wedge 1)^{-\delta} \leq \frac{2}{\lambda} E(\tau)^{1-\delta} = \frac{2}{\lambda} \lambda^{1-\delta} = 2\lambda^{-\delta}.$$

Check that

$$\begin{aligned} E\big((\max(\tau, 1))^{-\delta}\big) &= e^{-\lambda}\left\{ \sum_{k=1}^{\infty} \frac{1}{k^{\delta}} \frac{\lambda^k}{k!} + 1 \right\} \\ &= e^{-\lambda}\left\{ \sum_{k=1}^{\infty} \frac{1}{kk!} \lambda^k k^{1-\delta} + 1 \right\} \\ &= e^{-\lambda}\left\{ \sum_{k=1}^{\infty} \lambda^k \frac{1}{(k+1)k!} \frac{k+1}{k} k^{1-\delta} + 1 \right\} \\ &\leq 2\frac{e^{-\lambda}}{\lambda}\left\{ \sum_{k=1}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} k^{1-\delta} + \lambda \right\} \\ &\leq 2\frac{e^{-\lambda}}{\lambda}\left\{ \sum_{k=1}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} (k+1)^{1-\delta} + \lambda \right\} \\ &\leq \frac{2}{\lambda} E(\tau^{1-\delta}). \end{aligned}$$

The proof is complete. □

In order not to overload the present paper, we will merely treat some simple examples to some extent where the statistical functional is given by $T(F_n)$. A more extended account of applications will be given somewhere else.

EXAMPLE 1 (Conditional distribution function). Put $T(F) = F(y)$ where y is fixed. We have

$$\begin{aligned} \sup_t \left| P\left\{ \left\{ \frac{na_n g(x)}{F(y|x)(1 - F(y|x))} \right\}^{1/2} (F_n(y|x) - F(y|x)) \leq t \right\} - \Phi(t) \right| \\ = O\big((na_n)^{-1/2}\big) + H(\mathscr{L}(N_n(\cdot|x)), \mathscr{L}(N_n^*(\cdot|x))). \end{aligned}$$

EXAMPLE 2 (Conditional quantiles). Put $T(F) = F^{-1}(q)$, $q \in (0,1)$ fixed. Assume that $F(\cdot|x)$ is continuously differentiable near $F(\cdot|x)^{-1}(q)$ with $f_x(F(\cdot|x)^{-1}(q)) > 0$, where $f_x = F(\cdot|x)'$. Then (4.2) is satisfied with $\mu = F(\cdot|x)^{-1}(q)$, $\sigma^2 = q(1-q)/f_x(F(\cdot|x)^{-1}(q))^2$ and $\delta = 1/2$.

Consequently,

$$\sup_t \left| P\left\{ \frac{(na_n g(x))^{1/2}}{\sigma} \left(F_n(\cdot|x)^{-1}(q) - F(\cdot|x)^{-1}(q) \right) \leq t \right\} - \Phi(t) \right|$$
$$= O\left((na_n)^{-1/2}\right) + H(\mathscr{L}(N_n(\cdot|x)), \mathscr{L}(N_n^*(\cdot|x))).$$

EXAMPLE 3 (Regression function). Assume that (4.2) holds for $T(F) := \int zF(dz)$, $\mu = \int zF(dz|x)$ and $\sigma^2 := \int (z-\mu)^2 F(dz|x)$. Then

$$\sup_t \left| P\left\{ \frac{(na_n g(x))^{1/2}}{\sigma} \left(\int zF_n(dz|x) - \int zF(dz|x) \right) \leq t \right\} - \Phi(t) \right|$$
$$= O\left((na_n)^{-1/2}\right) + H(\mathscr{L}(N_n(\cdot|x)), \mathscr{L}(N_n^*(\cdot|x))).$$

Notice that

$$\int zF_n(dz|x) = \sum_{i=1}^n Y_i u_0\left(\frac{X_i - x}{a_n}\right) \Big/ \sum_{i=1}^n u_0\left(\frac{X_i - x}{a_n}\right) \tag{4.3}$$

is the Nadaraya-Watson estimator with particular kernel $u_0 = 1_{[-1/2,1/2]}$. If u_0 is replaced by some arbitrary kernel u with $\int u(z)\,dz = 1$ vanishing outside of $[-1/2, 1/2]$, then it is well known that the asymptotic normality holds with σ^2 replaced by $\sigma^2 \int u^2(z)\,dz$. Applying the Cauchy–Schwarz inequality we get

$$\int u^2(z)\,dz \geq \left(\int u(z)\,dz \right)^2 = 1 = \int u_0^2(z)\,dz$$

showing that the asymptotic variance is minimized by taking the uniform kernel u_0.

EXAMPLE 4 (Conditional V-statistics). Let $h: \mathbb{R}^2 \to \mathbb{R}$ be a symmetric kernel such that $\int\int |h(z,y)|^3 F(dz|x)F(dy|x) < \infty$, $\int |h(z,z)|^{3/2} F(dz|x) < \infty$ and $\sigma_0^2 := \int (\int h(z,y)F(dy|x) - \mu_0)^2 F(dz|x) > 0$, where

$$\mu_0 = \int\int h(z,y) F(dz|x) F(dy|x) =: T(F(\cdot|x)).$$

Then the Berry–Esseen theorem for V-statistics [cf. Serfling (1980), Sections 5 and 6] implies that (4.2) is satisfied and, therefore,

$$\sup_t \left| P\left\{ \frac{(na_n g(x))^{1/2}}{\sigma_0} \left(\int\int h(z,y) F_n(dz|x) F_n(dy|x) - \mu_0 \right) \leq t \right\} - \Phi(t) \right|$$
$$= O\left((na_n)^{-1/2}\right) + H(\mathscr{L}(N_n(\cdot|x)), \mathscr{L}(N_n^*(\cdot|x))).$$

EXAMPLE 5 (Projection pursuit technique). The trouble with high-dimensional data $\mathbf{X}_i$ is that local areas of the sample space are almost empty and, consequently, we will not be able to pick up local features such as $P(\mathbf{Y} \le \cdot|\mathbf{X} = \mathbf{x})$ unless the sample size is gigantic. To bypass this curse of dimensionality, it is therefore reasonable to reduce the dimension of the observations $\mathbf{X}_i$ by applying projection pursuit techniques [see, e.g., Huber (1985) and Hall (1988)]. One may estimate the conditional d.f.,

$$F_{\boldsymbol{\vartheta}}(\cdot|x) := P(\mathbf{Y} \le \cdot|\langle \boldsymbol{\vartheta}, \mathbf{X}\rangle = x),$$

of $\mathbf{Y}$ given $\mathbf{X}$ in direction $\boldsymbol{\vartheta}$ where $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_d)$ is a d-dimensional unit vector.

Fix a unit vector $\boldsymbol{\vartheta} \in \mathbb{R}^d$ and $x \in \mathbb{R}$. We consider now only those observations from the sample $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ where the projection of the first coordinate in direction $\boldsymbol{\vartheta}$ lies in a small interval in $\mathbb{R}$ with center x; that is, we consider the empirical process

$$N_{n,\boldsymbol{\vartheta}}(B) := \sum_{i=1}^{n} \varepsilon_{((\langle \boldsymbol{\vartheta}, \mathbf{X}_i\rangle - x)/a_n, \mathbf{Y}_i)}(B \cap S), \qquad B \in \mathbb{B}^{1+m},$$

where

$$S := [-1/2, 1/2] \times \mathbb{R}^m.$$

It is clear that the results of Sections 2 and 3 are applicable with $(\mathbf{X}_i, \mathbf{Y}_i)$ replaced by $(\langle \boldsymbol{\vartheta}, \mathbf{X}_i\rangle, Y_i)$.

For further applications of the preceding result to M, L and R estimators one may apply Berry–Esseen theorems for the respective functionals as, for example, given in the monograph by Serfling (1980). Since our results are formulated with bounds on the remainder term of the approximation, we are also able to investigate second order efficiency (deficiency) of kernel estimators in the conditional framework [see Reiss (1989) for results in the unconditional case]. One of the advantages of our approach is that highly technical investigations have merely to be carried out for the unconditional problem of i.i.d. random variables [e.g., Falk and Reiss (1990)]. Several modifications of the present approach are possible and have still to be explored in detail.

Acknowledgment. This paper was written while the first author was on leave at the Philipps-University Marburg. He is indebted to Volker Mammitzsch for providing a stimulating atmosphere.

REFERENCES

CSISZÁR, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci. Ser. A* **8** 85–108.

DEHEUVELS, P. and PFEIFER, D. (1988). Poisson approximation of multinomial distributions and point processes. *J. Multivariate Anal.* **25** 65–89.

FALK, M. and REISS, R.-D. (1990). Bootstrapping conditional curves. Report 242, Univ. Siegen.

FALK, M. and REISS, R.-D. (1992). Poisson approximation of empirical processes. *Statist. Probab. Lett.* To appear.

HALL, P. (1988). Estimating the direction in which a data set is most interesting. *Probab. Theory Related Fields* **80** 51–77.

HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

HUBER, P. J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–475.

LE CAM, L. (1966). Likelihood functions for large numbers of independent observations. In *Research Papers in Statistics: Festschrift for J. Neyman* (F. N. David, ed.) 167–187. Wiley, London.

LIESE, F. and VAJDA, I. (1987). *Convex Statistical Distances. Teubner-Texte zur Mathematik* **95**. Teubner, Leipzig.

MANTEIGA, W. G. (1990). Asymptotic normality of generalized functional estimators dependent on covariables. *J. Statist. Plann. Inference* **24** 377–390.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

PFANZAGL, J. (1985). *Asymptotic Expansion for General Statistical Models. Lecture Notes in Statist.* **31**. Springer, New York.

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer, New York.

SAMANTA, M. (1989). Nonparametric estimation of conditional quantiles. *Statist. Probab. Lett.* **7** 407–412.

STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5** 595–645.

STUTE, W. (1986). Conditional empirical processes. *Ann. Statist.* **14** 638–647.

TRUONG, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** 606–613.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

MATH.-GEOGRAPHISCHE FAKULTÄT
KATH. UNIVERSITÄT EICHSTÄTT
OSTENSTR. 26-28
D-8078 EICHSTÄTT
GERMANY

FB 6, UNIVERSITÄT GH SIEGEN
HÖLDERLINSTR. 3
D-5900 SIEGEN
GERMANY