LAWS OF SMALL NUMBERS: SOME APPLICATIONS TO CONDITIONAL CURVE ESTIMATION

M. Falk and F. Marohn (Eichstätt, Germany)

To the memory of Professor József Mogyoródi

Introduction

In recent years there has been increasing interest in a general theory concerning *rare events*, for which a handy and traditional name is *laws of small numbers*. Whenever one is concerned with rare events i.e., events with a small probability of occurence, the *Poisson*-distribution shows up in a natural way which is statistical folklore.

So the basic idea is simple, but its applications are nevertheless far-reaching and require therefore a complex mathematical machinery. The closely related book by David Aldous [1] "Probability Approximations via the Poisson Clumping Heuristic" demonstrates this need in an impressive way. But this book focuses narrowly on examples, though ranging over many fields of probability theory, and does not try to constitute a complete account of any field.

We will describe in the following in a quite informal way a general theory first and then apply this theory to a specific subfield of regression analysis. In prose: If we are interested only in those random elements among independent replicates of a random element Z, which fall into a given subset A of the sample space, the best way to describe this *random* sample (with Binomial sample size) is via the concept of *truncated empirical point processes.* If the probability for Z falling into A is small, then the Poisson approximation entails that we can approximate the truncated empirical point process by a Poisson point process with the sample size now being a Poisson random variable. This is what we will call *first step Poisson process approximation.*

Often, those random elements falling into A follow closely an ideal or limiting distribution; replacing their actual distribution by this ideal one, we generate a *second step Poisson process approximation* to the initial truncated empirical process.

Within certain error bounds, we can therefore handle those observations among the original sample, which fall into A, like ideal observations, whose stochastic behavior depends solely upon a few (unknown) parameters. This approach permits the application of standard methods to statistical questions concerning the original and typically nonparametric sample.

If the subset A is located in the center of the distribution of Z, then *regression analysis* turns out to be within the scope of laws of small numbers. If the subset A is however located at the border, then *extreme value theory* is typically covered by our theory. These specifications lead to characteristic results in each case.

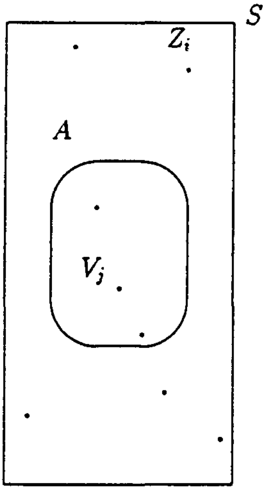
As the Hellinger distance provides a more accurate bound for the approximation of product measures in terms of their margins, as does the Kolmogorov-Smirnov or the variational distance, we will focus on the formulation of laws of small numbers within the Hellinger distance.

1. Foundations: First and second order Poisson process approximations

Let Z be a random element (re) in a sample space S bearing a σ-algebra $\mathcal{B}$ and let $Z_1, \ldots, Z_n$ be independent replicates of Z. Fix a subset $A \in \mathcal{B}$ and consider only those observations among $Z_1, \ldots, Z_n$ falling into A. Arranged in the order of their outcome, we can denote these $Z_i \in A$ by $V_1, \ldots, V_{K_A(n)}$, where the random number

$$K_A(n) := \sum_{i=1}^{n} 1_A(Z_i)$$

is Binomial distributed $B(n,p)$ with probability $p = P\{Z \in A\}$:



It is intuitively clear and is for example verified in Falk and Reiss [10] that $V_1, V_2, \ldots$ are independent replicates of a re V, whose range is the set A and whose distribution is the conditional distribution of Z given $Z \in A$

$$P\{V \in B\} = P\{Z \in B \mid Z \in A\} = \frac{P\{Z \in B \cap A\}}{P\{Z \in A\}}, \quad B \in \mathcal{B}.$$

Moreover, $K_A(n)$ and $V_1, V_2, \ldots$ are independent.

If $p = P\{Z \in A\}$ is small and n is large, it is well known that $B(n,p)$ can be approximated within a reasonable error bound by the Poisson distribution

$$P_\lambda(\{k\}) := e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

with $\lambda = np$.

The practical significance of the Poisson approximation of Binomial distributions was presumably first discovered by L.von Bortkiewicz [4]. He also seems to be the first to term this approximation a *law of small numbers*. For recent references we refer to the article by Arratia et al. [2] and the literature cited therein. A bound for the Hellinger distance between $B(n,p)$ and P_{np} was established by Falk and Reiss [10].

The random number $K_A(n) = \sum_{i=1}^n 1_A(Z_i)$ will consequently for $p = P\{Z \in A\}$ small and n large behave like a Poisson random variable (rv) $\tau_A(n)$ with parameter np

$$K_A(n) \underset{\mathcal{D}}{\sim} \tau_A(n),$$

where $\underset{\mathcal{D}}{\sim}$ indicates approximation in distribution. As $K_A(n)$ is independent of those Z_i falling into A, $\tau(n)$ will share this property and so we arrive at our first law of small numbers or

First order Poisson process approximation
$V_1, \ldots, V_{K_A(n)} \underset{\mathcal{D}}{\sim} V_1, \ldots, V_{\tau_A(n)}.$

The error of the preceding approximation is determined only by the error of the approximation of $K_A(n)$ by $\tau_A(n)$ or of $B(n,p)$ by P_{np}, respectively.

Different to the *global Poissonization* technique, where the fixed sample size n is replaced by a Poisson rv $\tau(n)$ with parameter n

$$Z_1, \ldots, Z_n \underset{\mathcal{D}}{\sim} Z_1, \ldots, Z_{\tau(n)},$$

$\tau(n)$ independent of $Z_1, Z_2, \ldots$, our approach is a *local* Poissonization in the set A.

Let now the set A depend on the sample size n i.e. $A = A_n$, such that the sequence $A_1, A_2, \ldots$ of sets is decreasing

$$A_1 \supset A_2 \supset \ldots$$

In this case, the conditional distribution of Z, given $Z \in A_n$, can often be approximated by some ideal limiting distribution i.e.,

$$P\{V \in \cdot\} = P\{Z \in \cdot \mid Z \in A_n\} \underset{\text{nlarge}}{\sim} P\{W \in \cdot\},$$

where W is a re with this ideal distribution and the error of this approximation decreases with increasing sample size n.

This observation suggests the second law of small numbers or

Second order Poisson process approximation

$$V_1, \ldots, V_{\tau_A(n)} \underset{\mathcal{D}}{\sim} W_1, \ldots, W_{\tau_A(n)},$$

where $W_1, W_2, \ldots$ are independent replicates of W; $\tau_A(n)$ and the sequence $W_1, W_2, \ldots$ are independent.

The error of this approximation is obviously determined by the distance of the distributions of V and W. Combining the first and second order Poisson approximation we arrive at the approximation

$$V_1, \ldots, V_{K_A(n)} \underset{\mathcal{D}}{\sim} W_1, \ldots, W_{\tau_A(n)},$$

with the total error being the sum of two errors, which are completely different in nature.

If the subsets A_n are locataed in the center of the distribution of Z, then *regression analysis* turns out to be within the scope of the laws of small numbers as we will see in the next section. If the subsets A_n are located at the border, then *extreme value theory* is typically covered by our preceding approach. This can easily be motivated if one is interested only in those observations among an iid sample which exceed a certain threshold. As these are the largest observations, extreme value theory shows up in a natural way (cf. Leadbetter et al. [18], Galambos [13], Resnick [25], Reiss [23], Davison and Smith [7] and the literature cited therein). For details we refer to Falk et all. [12], where the preceding quite informal introduction is made rigorous via the concept of point processes.

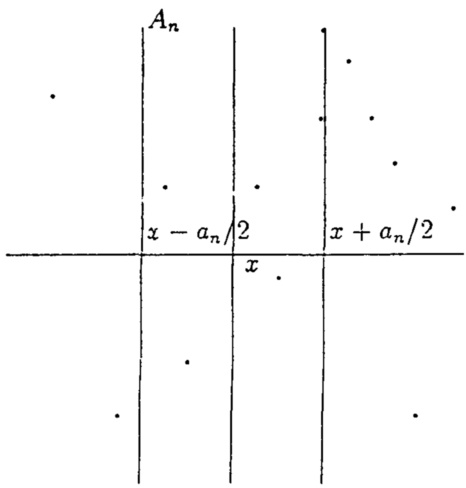
In the present article we demonstrate in the next section how the preceding approach can be made rigorous in regression analysis and we will utilize it to nonparametric and semiparametric conditional curve estimation. In the semiparametric setup we will derive in particular asymptotically optimal estimators from local asymptotic normality (LAN) of certain Poisson processes which approximately describe our initial sample. In the following example we first describe in an informal way, why regression analysis turns out to be within the scope of laws of small numbers.

Example. Let $Z = (X, Y)$ be a random vector in $\mathbb{R}^2$ and fix $x \in \mathbb{R}$. We are interested in the conditional distribution function (df) $F(\cdot \mid x) := P\{Y \leq \cdot \mid X = x\}$ of Y given $X = x$.

In this case we choose the data window

$$A_n := [x - a_n/2, x + a_n/2] \times \mathbb{R}$$

with windowwidth $a_n > 0$ for the data $Z_i = (X_i, Y_i)$, $i = 1, \ldots, n$:



Then,

$$\begin{aligned} K_{A_n}(n) :&= \sum_{i=1}^{n} 1_{A_n}(Z_i) \\ &= \sum_{i=1}^{n} 1_{[x-a_n/2, x+a_n/2]}(X_i) \end{aligned}$$

is $B(n,p)$-distributed with

$$p_n = P\{Z \in A_n\} = P\{X \in [x - a_n/2, x + a_n/2]\} \sim g(x)a_n,$$

where we assume that X has a density g, say, near x being continuous and positive at x.

If $Z = (X,Y)$ has a joint density f, say, on A_n, then we obtain for $t \in [0,1]$ and $s \in \mathbb{R}$

$$\begin{aligned} &P\{V \le (x - a_n/2 + t_{a_n}, s)\} \\ &= P\{Z \le (x - a_n/2 + ta_n, s), z \in A_n\}/P\{Z \in A_n\} \\ &= P\{x - a_n/2 \le X \le x - a_n/2 + ta_n, Y \le s\}/p_n \\ &= \int_{x-a_n/2}^{x-a_n/2+ta_n} \int_{-\infty}^{s} f(u,w)du\ dw \\ &= \int_{-\infty}^{s} a_n \int_0^t f(x + a_n u - a_n/2, w)du\ dw/p_n \\ &\xrightarrow[a_n \to 0]{} \int_{-\infty}^{s} s \int_0^t f(x,w)du\ dw/g(x) \\ &= t\int_{-\infty}^{s} f(x,w)dw/g(x) = tF(s \mid x) \end{aligned}$$

under suitable regularity conditions on f (near x).

Consequently, we obtain the approximation

$$V_1, \dots, V_{K_{A_n}(n)} \underset{\mathcal{D}}{\sim} (U_1, W_1), \dots, (U_{\tau_{A_n}(n)}, W_{\tau_{A_n}(n)}),$$

where U is on $[x - a_n/2, x + a_n/2]$ uniformly distributed, W_i follows the conditional df $F(\cdot \mid x)$, $\tau_{A_n}(n)$ is Poisson $P_{na_n g(x)}$ distributed and $\tau_{A_n}(n), W_1, W_2, \dots, U_1, U_2, \dots$ are all independent!

In this example our approach entails that the information we are interested in is essentially contained in the second component of V_i.

We close this section with some quite informal remarks why we prefer the Hellinger distance in our considerations.

Let X, Y be re's with values in some measurable space $(S, \mathcal{B})$. The Hellinger distance (between the distributions) of X and Y is defined by

$$H(X,Y) := \left(\int (f^{1/2} - g^{1/2})^2 d\mu\right)^{1/2},$$

where μ is a dominating measure and f, g are μ-densities of the distributions of X and Y.

For vectors $\mathbf{X} = (X_1, \dots, X_k)$, $\mathbf{Y} = (Y_1, \dots, Y_k)$ of independent copies of X and Y we have

$$H(\mathbf{X}, \mathbf{Y}) \le k^{1/2} H(X, Y),$$

whereas for the variational distance $d(X,Y) := \sup_{B \in \mathcal{B}} \mid P\{X \in B\} - P\{Y \in B\} \mid$ we only get the bound

$$d(\mathbf{X}, \mathbf{Y}) \le k d(X, Y).$$

Since in general $d(\cdot,\cdot) \le H(\cdot,\cdot)$ we deduce

$$d(\mathbf{X}, \mathbf{Y}) \le k^{1/2} H(X, Y)$$

(cf. section 3.3 in Reiss [23]).

If $d(X,Y)$ and $H(X,Y)$ are therefore of the same order, the Hellinger distance provides a more accurate bound than the variational distance i.e., $k^{1/2}$ compared with k.

Within our framework we obtain consequently for the second order Poisson process approximation the bound

$$\begin{aligned} &H((V_1, \dots, V_{\tau_A(n)}),\ (W_1, \dots, W_{\tau_A(n)})) \\ &\le \left(\int H^2((V_1, \dots, V_k),\ (W_1, \dots, W_k))\ \mathcal{L}(\tau_A(n))(dk) \right)^{1/2} \\ &\le \left(\int k H^2(V, W) \mathcal{L}(\tau_A(n))(dk) \right)^{1/2} \\ &= H(V, W) E(\tau_A(n))^{1/2} = H(V, W)(np)^{1/2}, \end{aligned}$$

where the first inequality is suggested by the convexity theorem (see Lemma 3.1.3 in Reiss [24]). By $\mathcal{L}(Z)$ we denote the distribution of a random element Z. On the other hand,

$$\begin{aligned} &d((V_1, \dots, V_{\tau_A(n)}), (W_1, \dots, W_{\tau_A(n)})) \\ &= \int k d(V, W) \mathcal{L}(\tau_A(n))(dk) = d(V, W) np. \end{aligned}$$

Consequently, we obtain the estimate

$$\begin{aligned} &d(V_1, \dots, V_{\tau_A(n)}), (W_1, \dots, W_{\tau_A(n)})) \\ &\le H((V_1, \dots, V_{\tau_A(n)}), (W_1, \dots, W_{\tau_A(n)})) \\ &\le H(V, W)(np)^{1/2}, \end{aligned}$$

which is more accurate than $d(V,W)np$ if $np \geq 1$ and $d(V,W)$ and $H(V,W)$ are of the same order; but this is typically the case.

We note that the preceding quite informal considerations can be made rigorous via the concept of point processes (cf. Falk and Reiss [10,11], Falk et al. [12], Reiss [24]).

2. Estimation of conditional curves

This section is devided into three parts. In the first part we make the Poisson process approach to regression analysis rigorous. Part two and three are concerned with nonparametric and semiparametric applications, respectively.

2.1. The Poisson process approach

Let $Z = (X, Y)$ be a $(d+m)$-dimensional random vector and denote again by $F(\cdot \mid x) := P\{Y \leq \cdot \mid X = x\}$ the conditional df of Y given $X = x$, $x \in \mathbb{R}^d$. Applying our approach one may study the fairly general problem of evaluating a functional parameter $T(F(\cdot \mid x))$ based on independent replicates $Z_i = (X_i, Y_i)$, $i = 1, \ldots, n$, of Z. This can be done in a nonparametric setup, where typical functionals are the regression mean $T_1(F) = \int t \; F(dt)$ on the regression quantile $T_2(F) = F^{-1}(q) = \inf\{t \in \mathbb{R} : F(t) \geq q\}$, $q \in (0,1)$, as well as in a parametric setup, where $F(\cdot \mid x)$ is a member of a parametric family $\{F_\vartheta(\cdot \mid x) : \vartheta \in \Theta\}$, $\Theta \in \mathbb{R}^k$, and $T(F_\vartheta(\cdot \mid x)) := \vartheta$.

While classical nonparametric regression analysis focuses on the problem of estimating the conditional mean $T_1(F(\cdot \mid x)) = \int t \; F(dt \mid x)$ (a recent reference is, for example, Eubank [8]), the estimation of general regression functionals $T(F(\cdot \mid x))$ has been receiving increasing interest only in recent years (see, for example, Stute [30], Härdle et al. [15], Truong [31], Samanta [26], Manteiga [21], Jones and Hall [17], Goldstein and Messer [14], Bhattacharya and Gangopadhyay [3], Chaudhuri [5].

Statistical inference based on $(X_1, Y_1), \ldots, (X_n, Y_n)$ for a functional $T(F(\cdot \mid x))$ has obviously to be based on those Y_i among $Y_1, \ldots, Y_n$, whose corresponding X_i-values are *close* to x. Choose therefore as in the example in the preceding section a windowwidth $a_n = (a_{n1}, \ldots, a_{nd}) \in (0, \infty)^d$ and define as the data-window for X_i

$$\begin{aligned} S_n &:= \times_{j=1}^{d} [x_j - a_{nj}^{1/d}/2,\ x_j + a_{nj}^{1/d}/2] \\ &=: [x - a_n^{1/d}/2,\ x + a_n^{1/d}/2]. \end{aligned}$$

The data set Y_i with $X_i \in S_n$ is described in a mathematically precise way by the truncated empirical point process

$$N_n(B) := \sum_{i=1}^{n} \varepsilon_{Y_i}(B)\varepsilon_{X_i}(S_n) = \sum_{i=1}^{K(n)} \varepsilon_{V_i}(B), \ B \in \mathbb{B}^m,$$

where

$$K(n) := \sum_{i=1}^{n} \varepsilon_{X_i}(S_n)$$

is the number of those Y_i with $X_i \in S_n$ which we denote by $V_1, V_2, \ldots$ By $\varepsilon_x(B) = 1_B(x) = 1$ if $x \in B$ and 0 otherwise we denote the Dirac-measure with mass one at x. From Lemma 1 in Falk and Reiss [10] we know that $K(n)$ and $V_1, V_2, \ldots$ are independent, where

$$P\{V \in \cdot\} = P\{Y \in \cdot \mid X \in S_n\},$$

and $K(n)$ is $B(n, p_n)$-distributed with $p_n = P\{X \in S_n\} \sim$ volume of S_n if $\| a_n \|$ is small (under suitable regularity conditions). By $\| \ \|$ we denote the Euclidean norm.

If we replace in N_n the sample size $K(n)$ by a Poisson rv $\tau(n)$ with parameter $E(K(n)) = np_n$, which is also independent of $V_1, V_2, \ldots$, then we obtain the Poisson process approximation N_n^* of N_n, defined by

$$N_n^*(B) := \sum_{i=1}^{\tau(n)} \varepsilon_{V_i}(B), \ B \in \mathbb{B}^m.$$

The error of this approximation is determined only by the error of the approximation of $K(n)$ by $\tau(n)$ (see Theorem 2 in Falk and Reiss [10]).

Theorem 2.1.1. (**First order Poisson process approximation**). *We have for the Hellinger distance*

$$H(N_n, N_n^*) \leq CP\{X \in S_n\}, \tag{1}$$

where C is a universal constant with $C \leq \sqrt{3}$.

It is intuitively clear and was already shown in the example of the previous section (with $d = m = 1$) that for $\| a_n \| \to 0$

$$P\{V \in \cdot\} = P\{Y \in \cdot \mid X \in S_n\} \underset{\|a_n\| \to 0}{\longrightarrow} P\{Y \in \cdot \mid X = x\}.$$

This implies the approximation of N_n^* by the Poisson process

$$N_n^{**}(B) := \sum_{i=1}^{\tau^*(n)} \varepsilon_{W_i}(B), \ B \in \mathbb{B}^m,$$

where $W_1, W_2, \ldots$ are independent replicates of a random vector W with target df $F(\cdot \mid x)$, and $\tau^*(n)$ is a Poisson rv with parameter $n \ vol(S_n) g(x)$; g denoting the marginal density of X and $vol(S_n) := \Pi_{j=1}^d a_{nj}^{1/d}$ the volume of S_n. The rv $\tau^*(n)$ and $W_1, W_2, \ldots$ are again independent.

Theorem 2.1.2. (Second order Poisson process approximation). *Suppose that the random vector (X,Y) has a joint density f on the strip $[x-\varepsilon_0, x+\varepsilon_0] \times \mathbb{R}^m$ for some $\varepsilon_0 \in (0,\infty)^d$, which satisfies uniformly for $\varepsilon \in (-\varepsilon_0, \varepsilon_0) \subset \mathbb{R}^d$ and $y \in \mathbb{R}^m$ the expansion*

$$(2) \qquad f(x+\varepsilon, y)^{1/2} = f(x,y)^{1/2}\Big\{1 + \langle \varepsilon, r_1(y)\rangle + O(\| \varepsilon \|^2 \, r_2(y))\Big\},$$

where $\int (\| r_1(y) \|^4 + (r_2(y))^4) f(x,y)dy < \infty$ and $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^d$. Then we have for $\| a_n \|$ small

$$(3) \qquad H(N_n^*, N_n^{**}) = O\Big((n \ vol(S_n))^{1/2} \, \| a_n^{1/d} \|^2\Big).$$

Corollary 2.1.3. *Combining (1) and (3) we obtain under the conditions of Theorem 2.1.2 the bound*

$$H(N_n, N_n^{**}) = O\Big((vol(S_n) + (n \ vol(S_n))^{1/2} \, \| a_n^{1/d} \|^2\Big).$$

With equal binwidths $a_{n1} = \ldots = a_{nd} = c_n$, the preceding bound reduces to $O(c_n + (nc_n^{(d+4)/d})^{1/2})$. While the function r_2 in expansion (2) collects the remainders left over, the function r_1 reflects the dependence of the conditional distribution of Y given X near x from the conditional distribution of Y given $X = x$.

Example. Suppose that (X,Y) is bivariate normally distributed i.e.,

$$f(z,y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)} \exp\Big\{ -\frac{1}{2(1-\rho^2)}\Big(\Big(\frac{z-\mu_1}{\sigma_1}\Big)^2 - 2\rho\Big(\frac{z-\mu_1}{\sigma_1}\Big)\Big(\frac{y-\mu_2}{\sigma_2}\Big) + \Big(\frac{y-\mu_2}{\sigma_2}\Big)^2\Big)\Big\}, \ z,y \in \mathbb{R},$$

where $\mu_1, \mu_2, \in \mathbb{R}, \sigma_1, \sigma_2 > 0$ and $\rho \in (-1,1)$. Taylor expansion of the exponential function at 0 entails the expansion

$$\begin{aligned}\left(\frac{f(x+\varepsilon,y)}{f(x,y)}\right)^{1/2} &= \exp\left\{-\frac{1}{4(1-\rho^2)}\left(\frac{2(x-\mu_1)\varepsilon+\varepsilon^2}{\sigma_1^2}+\frac{2\rho\varepsilon}{\sigma_1\sigma_2}(y-\mu_2)\right)\right\}\\ &= 1+\varepsilon\frac{1}{2(1-\rho^2)}\left(\frac{\rho}{\sigma_1\sigma_2}(y-\mu_2)-\frac{x-\mu_1}{\sigma_1^2}\right)\\ &+O\left(\varepsilon^2 d_1\exp(d_2\mid y\mid)(1+y^2)\right)\\ &=: 1+\varepsilon r_1(y)+O\left(\varepsilon^2 r_2(y)\right)\end{aligned}$$

with some appropriate positive constants d_1, d_2.

Proof of Theorem 2.1.2. The densities of the intensity measures on $\mathbb{R}^m$ pertaining to N_n^* and N_n^{**} are given by

$$f_n^*(y) = n\ vol(S_n)\int_{[-\frac{1}{2},\frac{1}{2}]^d} f(x+a_n^{1/d}z,y)dz,\ y\in\mathbb{R}^m,$$

and

$$f_n^{**}(y) = n\ vol(S_n)f(x,y),\ y\in\mathbb{R}^m.$$

By the monotonicity theorem due to Csiszár [6] (see also Liese and Vajda [20] or Theorem 3.2.1 in Reiss [24])and expansion (2) we obtain

$$\begin{aligned}H^2(N_n^{**},N_n^*) &\le \int_{\mathbb{R}^m}\left(f_n^{**}(y)^{1/2}-f_n^*(y)^{1/2}\right)^2 dy\\ &= n\ vol(S_n)\int_{\mathbb{R}^m}\left\{\left(\int_{[-\frac{1}{2},\frac{1}{2}]^d} f(x+a_n^{1/d}z,y)dz\right)^{1/2}-f(x,y)^{1/2}\right\}^2 dy\\ &= O\left(n\ vol(S_n)\parallel a_n^{1/d}\parallel^4\right).\end{aligned}$$

The preceding approach can be extended to several points $x_1,\ldots,x_r$ with the corresponding bounds summing up (see Falk and Reiss [11] for details).

2.2. Applications: The nonparametric case

The usual nonparametric estimate of a functional $T(F)$ based on an iid sample $Y_1, \ldots, Y_n$ with common df F is $T(F_n)$, where $F_n(t) := n^{-1}\sum_{i=1}^n 1_{(-\infty,t]}(Y_i)$ denotes the pertaining empirical df. Within our framework, the empirical df

$$\begin{aligned}\hat{F}_n(t \mid S_n) :&= K(n)^{-1}\sum_{i=1}^n 1_{(-\infty,t]}(Y_i)1_{S_n}(X_i)\\ &= N_n(\mathbb{R}^m)^{-1}N_n((-\infty,t]),\ t \in \mathbb{R}^m,\end{aligned}$$

pertaining to those Y_i among $Y_1, \ldots, Y_n$ with $X_i \in S_n$, suggests itself as a nonparametric estimate of $F(\cdot \mid x)$. The resulting estimate of $T(F(\cdot \mid x))$ is $T(\hat{F}_n(\cdot \mid S_n))$. Observe that $\hat{F}_n$ is the standardized df pertaining to the random measure N_n.

For the mean value functional T_1 we obtain for example

$$T_1(\hat{F}_n(\cdot \mid S_n)) = \int t\hat{F}_n(dt \mid S_n) = \frac{\sum_{i=1}^n Y_i 1_{S_n}(X_i)}{\sum_{i=1}^n 1_{S_n}(X_i)}$$

which is the Nadaraya-Watson estimator. Following Stone [28,29] and Truong [31] we call $T(\hat{F}_n(\cdot \mid S_n))$ *kernel estimator* of a *general regression functional* $T(F(\cdot \mid x))$.

In the following we suppose for the sake of a clear presentation that the dimension m of Y is 1.

Theorem 2.2.1. *Suppose that for some $\sigma > 0$, $\delta \in (0, 1/2]$ and $C > 0$*

$$\sup_{t\in\mathbb{R}}\left|P\left\{\frac{k^{1/2}}{\sigma}\Big(T(F_k(\cdot \mid x)) - T(F(\cdot \mid x))\Big) \le t\right\} - \Phi(t)\right| \le Ck^{-\delta},\ k \in \mathbb{N}, \tag{4}$$

where $F_k(\cdot \mid x)$ denotes the empirical df pertaining to k independent rvs with common df $F(\cdot \mid x)$. If the vector (X, Y) satisfies condition (2), then we obtain for the kernel estimator $T(\hat{F}_n(\cdot \mid S_n))$ with equal binwidths $a_{n1} = \ldots = a_{nd} = c_n$

$$\begin{aligned}&\sup_{t\in\mathbb{R}} P\left|\left\{\frac{(nc_n g(x))^{1/2}}{\sigma}\Big(T(\hat{F}_n(\cdot \mid S_n)) - T(F(\cdot \mid x))\Big) \le t\right\} - \Phi(t)\right|\\ &= O\Big((nc_n)^{-\delta} + c_n + (nc_n^{(d+4)/d})^{1/2}\Big).\end{aligned}$$

With the particular choice $c_n = O(n^{-d/(d+4)})$, we obtain roughly the rate $O_P(n^{-2/(d+4)})$ for $T(\hat{F}_n(\cdot \mid S_n)) - T(F(\cdot \mid x))$ which is known to be the optimal attainable accuracy under suitable regularity conditions in case of the mean value functional (Stone [28,29]), and quantile functional (Chaudhuri [5]) (for a related result we refer to Truong [31]).

The proof of Theorem 2.2.1 is based on the following elementary result (see Lemma 1 in Falk and Reiss [11]).

Lemma 2.2.2. *Let $V_1, V_2, \ldots$ be a sequence of rvs such that for some $\sigma > 0, \mu \in \mathbb{R}$ and $\delta \in (0, 1/2]$*

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

Then we have with τ being a Poisson rv with parameter $\lambda > 0$ and independent of each V_i, $i = 1, 2, \ldots$

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

where D depends only on C (with the convention $V_\tau = 0$ if $\tau = 0$).

Proof of Theorem 2.2.1. Put $V_k := T(F_k(\cdot \mid x))$, $k = 1, 2, \ldots$, and $\mu := T(F(\cdot \mid x))$. Since $T(\hat{F}_n(\cdot \mid S_n))$ is a functional of the empirical point process, we obtain from Corollary 2.1.3

$$\sup_{t\in\mathbb{R}} \left| P\left\{ \frac{(nc_n g(x))^{1/2}}{\sigma} \Big(T(\hat{F}_n(\cdot \mid S_n)) - T(F(\cdot \mid x)) \Big) \le t \right\} - \Phi(t) \right|$$

$$= \sup_{t\in\mathbb{R}} \left| P\left\{ \frac{(nc_n g(x))^{1/2}}{\sigma}(V_{\tau^*(n)} - \mu) \le t \right\} - \Phi(t) \right| + O\Big(c_n + (nc_n^{(d+4)/d})^{1/2} \Big),$$

where $\tau^*(n) = N_n^{**}(\mathbb{R})$ is Poisson distributed with parameter $\lambda = nc_n g(x)$ and independent of each $V_1, V_2, \ldots$ The assertion is now immediate from Lemma 2.2.2.

Condition (4) is satisfied for a large class of functionals T for which a Berry-Esseen result is available i.e., U- and V-statistics, M, L and R estimators. See, for example, the monograph by Serfling [27].

2.3. Applications: The semiparametric case

Assume now that the conditional distribution $P_\vartheta\{Y \in \cdot \mid X = x\} = Q_\vartheta(\cdot)$ of $Y(\in \mathbb{R}^m)$ given $X = x \in \mathbb{R}^d$ is a member of a parametric family, where the parameter space Θ is an open subset of $\mathbb{R}^k$. Under suitable regularity conditions we establish asymptotically optimal estimates based on N_n of the true underlying parameter ϑ_0. Since the estimation problem involves the joint density of (X, Y) as an infinite dimensional nuisance parameter, we actually have to deal with a special semiparametric problem: Since we observe data Y_i whose X_i-values are only *close* to x, our set of data $V_1, \ldots, V_{K(n)}$, on which we will base statistical inference, is usually *not* generated according to our target conditional distribution $Q_\vartheta(\cdot)$ but to some distribution being close to $Q_\vartheta(\cdot)$. This error is determined by the joint density f of (X, Y) which is therefore an infinite dimensional nuisance parameter. As a main tool we utilize local asymptotic normality (LAN) of the Poisson process N_n^{**}. (For a general approach to semiparametric problems we refer to the book by Pfanzagl [22].)

Suppose that for $\vartheta \in \Theta$ the probability measure $Q_\vartheta(\cdot)$ has Lebesgue-density q_ϑ. We suppose that the density f of the random vector (X, Y) exists on a strip $[x - \varepsilon_0, x + \varepsilon_0] \times \mathbb{R}^m$ and is a member of the following class of functions

$$\mathcal{F}(C_1, C_2, C_3)$$

$$=: \Bigg\{ f : [x - \varepsilon_0, x + \varepsilon_0] \times \mathbb{R}^m \to \mathbb{R}_+ \text{ such that } 0 < g_f(x) := \int f(x, y) dy \le C_1$$

$$\text{and for any } \varepsilon \in (0, \varepsilon_0]$$

$$\left| \frac{f^{1/2}(x + \varepsilon, y)}{f^{1/2}(x, y)} - (1 + \langle \varepsilon, h_f(y) \rangle) \right| \le C_2 \parallel \varepsilon \parallel^2 r_f(y)$$

$$\text{for some functions } h_f : \mathbb{R}^m \to \mathbb{R}^d,\ r_f : \mathbb{R}^m \to \mathbb{R} \text{ satisfying}$$

$$\int (\parallel h_f(y) \parallel^4 + r_f^4(y)) f(x, y) dy \le C_3 \Bigg\},$$

where C_1, C_2, C_3 are fixed positive constants. The leading term h_f of the above expansion reflects the dependence between the conditional distribution of Y given X near x and $X = x$ (compare with condition (2)).

The class of possible distributions Q of (X,Y), which we consider, is then characterized by

$$\mathcal{P} := \mathcal{P}(\mathcal{F}(C_1,C_2,C_3),\Theta)$$
$$:= \Big\{ P \mid \mathbb{R}^{d+m} : P \text{ has density } f \in \mathcal{F}(C_1,C_2,C_3) \text{ on } [x-\varepsilon_0, x+\varepsilon_0]$$
$$\text{such that the conditional density } f(\cdot \mid x) := f(x,\cdot)\Big/ \int f(x,y)dy$$
$$\text{is an element of } \{q_\vartheta : \vartheta \in \Theta\}\Big\}.$$

Note that $\mathcal{P}(\mathcal{F}(C_1,C_2,C_3),\Theta)$ forms a semiparametric family of distributions, where the densities $f \in \mathcal{F}(C_1,C_2,C_3)$ form the nonparametric part (in which we are primarily not interested), and where the k-dimensional parametric part (we are primarily interested in) is given by Θ. As a consequence, we index expectations, distributions etc. by $E_{f,\vartheta}, \mathcal{L}_{f,\vartheta}$ etc.

A main tool for the solution of our estimation problem is the following extension of Corollary 2.1.3 which follows by a careful study of the proof of Theorem 2 in Falk and Reiss [11]. By this result we can handle our data $V_1,\dots,V_{K(n)}$ within a certain error bound as being independently generated according to Q_ϑ, where the independent sample size is a Poisson rv $\tau^*(n)$ with parameter $n\ vol(S_n)g_f(x)$; in other words, we can handle the empirical point process N_n (which we observe) within this error bound as the ideal Poisson process N_n^{**}.

Lemma 2.3.1. *We have*

$$\sup_{\mathcal{P}(\mathcal{F}(C_1,C_2,C_3),\Theta)} H(\mathcal{L}_{f,\vartheta}(N_n), \mathcal{L}_{g_f(x),\vartheta}(N_n^{**}))$$
$$= O\Big(vol(S_n) + (n\ vol(S_n))^{1/2} \parallel a_n^{1/d} \parallel^2 \Big).$$

Notice that in the preceding result the distribution of the Poisson process $N_n^{**}(\cdot) = \sum_{i=1}^{\tau^*(n)} \varepsilon_{W_i}(\cdot)$ depends only on ϑ and the real parameter $g_f(x) = \int f(x,y)dy$, with $n\ vol(S_n)g_f(x)$ being the expectation of the Poisson rv $\tau^*(n)$.

By the preceding model approximation we can reduce the semiparametric problem $\mathcal{L}_{f,\vartheta}(N_n)$ with unknown $f \in \mathcal{F}(C_1,C_2,C_3)$ and $\vartheta \in \Theta$ to the $(k+1)$-dimensional paremetric problem

$$\mathcal{L}_{b,\vartheta}(N_n^{**}) = \mathcal{L}_{b,\vartheta}\Big(\sum_{i=1}^{\tau^*(n)} \varepsilon_{W_i} \Big),$$

where $\tau^*(n)$ is a Poisson rv with expectation $n\ vol(S_n)b$, $b \in (0, C_1], W_1, W_2, \ldots$ are iid random vectors with distribution Q_ϑ and $\tau^*(n)$ and $W_1, W_2, \ldots$ are independent.

We require Hellinger differentiability of the family $\{q_\vartheta : \vartheta \in \Theta\}$ of densities at any $\vartheta_0 \in \Theta$ i.e., we require the following expansion

$$(5) \qquad q_\vartheta^{1/2}(\cdot) = q_{\vartheta_0}^{1/2}(\cdot)\Big(1 + \langle \vartheta - \vartheta_0, v_{\vartheta_0}(\cdot)\rangle/2 + \parallel \vartheta - \vartheta_0 \parallel r_{\vartheta,\vartheta_0}(\cdot)\Big),$$

for some measurable function $v_{\vartheta_0} = (v_{01}, \ldots, v_{0k})^t, v_{0i} \in L_2(Q_{\vartheta_0}), i = 1, \ldots, k$ and some remainder term $r_{\vartheta,\vartheta_0}$ satisfying

$$\parallel r_{\vartheta,\vartheta_0} \parallel_{L_2(Q_{\vartheta_0})} = \left(\int r^2_{\vartheta,\vartheta_0}(y) Q_{\vartheta_0}(dy)\right)^{1/2} \underset{\|\vartheta-\vartheta_0\|\to 0}{\longrightarrow} 0.$$

Denote by $M(\mathbb{R})$ the space of all finite point measures on $\mathbb{R}$, endowed with the smallest σ-algebra $\mathcal{M}(\mathbb{R})$ such that all projections $M(\mathbb{R}) \ni \mu \mapsto \mu(B)$, $B \in \mathbb{B}$, are measurable, and define the statistical experiment $E_n = (M(\mathbb{R}), \mathcal{M}(\mathbb{R}), \{\mathcal{L}_{\vartheta_0+t\delta_n}(N_n^{**}) : t \in \Theta_n\})$, where $\delta_n = (n\ vol(S_n))^{-1/2}$ and $\Theta_n = \{t \in \mathbb{R}^k : \vartheta_0 + t\delta_n \in \Theta\}$. Throughout the rest we suppose that $n\ vol(S_n) \to \infty$ as $n \to \infty$.

It is well known that condition (5) implies the local asymptotic normality (LAN) of the statistical experiments $(\mathbb{R}^m, \mathbb{B}^m, \{Q_{\vartheta_0+tn^{-1/2}} : t \in \Theta_n\})$. Without further assumptions this result remains true for E_n if the marginal density $g_f(x)$ of X at point x does not depend on ϑ, which is intuitively clear and which can immediately be seen from the likelihood process of E_n. But this would be a rather restrictive condition. The following result, which is adopted from Falk and Marohn [9], shows that in order to get LAN of E_n it suffices to require the function $g(x)$ to be smooth at ϑ_0.

Theorem 2.3.2. *(LAN of E_n). Fix $b > 0$. Under condition (5) we have with $b_n = b + o(\delta_n)$ and $\vartheta_n = \vartheta_0 + t\delta_n$*

$$\frac{d\mathcal{L}_{b_n,\vartheta_n}(N_n^{**})}{d\mathcal{L}_{b,\vartheta_0}(N_n^{**})}(\cdot) = \exp\Big(\langle t, Z_{n,\vartheta_0}(\cdot)\rangle_{b,\vartheta_0} - \frac{1}{2} \parallel t \parallel^2_{b,\vartheta_0} + R_{n,\vartheta_0,t}(\cdot)\Big)$$

with central sequence $Z_{n,\vartheta_0} : M(\mathbb{R}) \to \mathbb{R}^k$ given by

$$Z_{n,\vartheta_0}(\mu) = (\delta_n \mu(\mathbb{R}^m))^{-1}\Gamma^{-1}(\vartheta_0)\int v_{\vartheta_0} d\mu$$

and $R_{n,\vartheta_0,t} \to 0$ in $\mathcal{L}_{b,\vartheta_0}(N_n^{**})$-probability, where $\langle s,t\rangle_{b,\vartheta_0} := s'b\Gamma(\vartheta_0)t$, $s,t, \in \mathbb{R}^k$, and the $k \times k$-matrix $\Gamma(\vartheta_0) := (\int v_{0i}v_{0j}dQ_{\vartheta_0})_{i,j\in\{1,\ldots,k\}}$ is assumed to be positive definite.

Note that under alternatives of the form $b_n = b + o(\delta_n)$, $\vartheta_n = \vartheta_0 + t\delta_n$, the central sequence Z_{n,ϑ_0} *does not depend on* the nuisance parameter b, which will become lateron the value of the marginal density of X at x. If we allow $b_n = b + O(\delta_n)$ instead, then LAN of $(E_n)_n$ still holds, but the central sequence depends on the nuisance parameter b, which cannot be estimated without affecting the asymptotics (see Falk and Marohn [9] for details).

We recall the famous convolution theorem of Hajek (see, for example, Ibragimov and Has'minskii [16], Theorem 9.1, p. 154). Suppose that condition (5) holds for $\vartheta_0 \in \Theta$ and that $T_n(N_n^{**})$ is an *asymptotically δ_n-regular sequence of estimators in* ϑ_0 based on N_n^{**} i.e.,

$$\delta_n^{-1}(T_n(N_n^{**}) - \vartheta_0 - t\delta_n) \to_{\mathcal{D}} G \text{ for all } t \in \mathbb{R}^k$$

under $\vartheta_0 + t\delta_n$ for some probability measure G on $\mathbb{R}^k$ where $\to_{\mathcal{D}}$ denotes convergence in distribution. Then there exists a probability measure H on $\mathbb{R}^k$ such that

$$G = H * \mathcal{N}\Big(0, b^{-1}\Gamma^{-1}(\vartheta_0)\Big),$$

where $\mathcal{N}(0, b^{-1}\Gamma^{-1}(\vartheta_0))$ is the standard normal distribution on $(\mathbb{R}^k, \langle\cdot,\cdot\rangle_{b,\vartheta_0})$ with mean vector 0 and covariance matrix $b^{-1}\Gamma^{-1}(\vartheta_0)$ and $*$ denotes convolution.

In view of this convolution theorem, a δ_n-regular sequence of estimators $T_n(N_n^{**})$ is called *asymptotially efficient* in ϑ_0 (in the sense of Fisher) if

$$\delta_n^{-1}(T_n(N_n^{**}) - \vartheta_0) \to_{\mathcal{D}} \mathcal{N}\Big(0, b^{-1}\Gamma^{-1}(\vartheta_0)\Big)$$

under ϑ_0.

By Theorem 2.3.2 we know that Z_{n,ϑ_0} is central and hence,

$$\delta_n Z_{n,\vartheta_0}(N_n^{**}) + \vartheta_0 = \tau^*(n)^{-1}\Gamma^{-1}(\vartheta_0) \sum_{i=1}^{\tau^*(n)} v_{\vartheta_0}(W_i) + \vartheta_0$$

is asymptotically efficient in ϑ_0 for each $b > 0$. Note that this is true however only under the condition $b_n = b + o(\delta_n)$ in which case Z_{n,ϑ_0} is central. If we replace the unknown underlying parameter ϑ_0 by any δ_n^{-1}-consistent estimator $\hat{\vartheta}_{\tau^*(n)} = \hat{\vartheta}_n(N_n^{**})$ of ϑ_0 i.e., $\delta_n^{-1}(\hat{\vartheta}_{\tau^*(n)} - \vartheta_0)$ is stochastically bounded under ϑ_0, we obtain that

$$\hat{\kappa}_n(N_n^{**}) := \delta_n Z_{n,\hat{\vartheta}_{\tau^*(n)}}(N_n^{**}) + \hat{\vartheta}_{\tau^*(n)}$$

is asymptotically efficient in ϑ_0, whenever the function $\vartheta \to \Gamma(\vartheta)$ is continuous at ϑ_0 and

$$(6) \qquad \sup_{\|\vartheta_0-\vartheta\|\le K\delta_n} \left|\delta_n Z_{n,\vartheta_0}(N_n^{**}) + \vartheta_0 - \delta_n Z_{n,\vartheta}(N_n^{**}) - \vartheta\right| = o_P(\delta_n)$$

under ϑ_0 (and b) for any $K > 0$.

Denote by $F = F_{\vartheta_0}$ the distribution function of Q_{ϑ_0} and by $F_l(t) := l^{-1}\sum_{i=1}^{l} 1_{(-\infty,t]}(W_i)$, $t \in \mathbb{R}^m$, the empirical df pertaining to an iid sample $W_i, \dots, W_l$ with common distribution Q_{ϑ_0}. Using conditioning techniques, elementary calculations show that condition (6) is satisfied if the function $\vartheta \to \Gamma(\vartheta)$ is continuous at ϑ_0 and the following two conditions hold

$$(7) \qquad \sup_{\|\vartheta_0-\vartheta\|\le Kl^{-1/2}} \left|l^{1/2}\int (v_{\vartheta_0}(s) - v_\vartheta(s))(F_l - F)(ds)\right| = o_P(1)$$

as $l \to \infty$ for any $K > 0$ and

$$(8) \qquad \left(\int v_\vartheta(s)F(ds) + \Gamma(\vartheta_0)(\vartheta - \vartheta_0)\right) \Big/ \parallel \vartheta - \vartheta_0 \parallel \underset{\|\vartheta-\vartheta_0\|\to 0}{\longrightarrow} 0.$$

Note that $\sqrt{n}$-consistency of $\hat\vartheta_n(W_1, \dots, W_n)$ implies δ_n^{-1}-consistency of $\hat\vartheta_{\tau^*(n)} = \hat\vartheta_{\tau^*(n)}(W_1, \dots, W_{\tau^*(n)})$. We remark that under the present assumptions $\sqrt{n}$-consistent estimators actually exist (cf. LeCam [19], Proposition 1, p. 608).

In the following we discuss one standard family $\{Q_\vartheta : \vartheta \in \Theta\}$ (of possible conditional distributions) which satisfies conditions (5) and (6). Further examples can easily be constructed as well.

Example (Exponential families). Let $\{Q_\vartheta : \vartheta \in \Theta\}$, $\Theta \subset \Theta^*$ open, be a k-parametric exponential family of probability measures on $\mathbb{R}$ with natural parameter space $\Theta^* \subset \mathbb{R}^k$ i.e.,

$$q_\vartheta(x) = \frac{dQ_\vartheta}{d\nu}(x) = \exp(\langle\vartheta, T(x)\rangle - K(\vartheta)),\ x \in \mathbb{R},$$

for some σ-finite measure ν on $\mathbb{R}$ and some measurable map $T = (T_1, \dots, T_k) : \mathbb{R} \to \mathbb{R}^k$, where the functions $\{1, T_1, \dots, T_k\}$ are linear independent on the complement of each ν-null set and $K(\vartheta) := \log \int \exp(\langle\vartheta, T(x)\rangle)d\nu(x)$. It is well known that the function $\vartheta \to E_\vartheta T$ is analytic in the interior of Θ^*. From Theorem 1.194 in Witting [32] we conclude that for $\vartheta_0 \in \Theta^*$ the family $\{Q_\vartheta\}$ is Hellinger-differentiable at ϑ_0 with derivative

$$v_{\vartheta_0}(x) = \nabla \log q_{\vartheta_0}(x) = T(x) - E_{\vartheta_0}T$$

where $\nabla = (\frac{\partial}{\partial \vartheta_i})_{i=1,\ldots,k}$ denotes the nabla-operator. In this case, we get $\Gamma(\vartheta_0) = Cov_{\vartheta_0} T$ and condition (8) is implied by

$$\frac{E_\vartheta T - E_{\vartheta_0} T - \nabla E_{\vartheta_0} T(\vartheta - \vartheta_0)}{\| \vartheta - \vartheta_0 \|} \to 0$$

for $\vartheta \to \vartheta_0$ and $\nabla E_{\vartheta_0} T = Cov_{\vartheta_0} T$. Notice that $Cov_{\vartheta_0} T$ is positive definite by the linear independence of $\{1, T_1, \ldots, T_k\}$ (Witting [32], Theorem 1.153). Condition (7) trivially holds since the integrand is independent of s.

We can rewrite $\hat{\kappa}_n(N_n^{**})$ in the form

$$\hat{\kappa}(N_n^{**}) = (N_n^{**}(\mathbb{R}))^{-1}\Gamma^{-1}(\hat{T}(N_n^{**})) \int v_{\hat{T}(N_n^{**})} dN_n^{**} + \hat{T}(N_n^{**})$$

with $\hat{T} : M(\mathbb{R}) \to \mathbb{R}^k$ given by

$$\hat{T}(\mu) = \hat{\vartheta}_{\mu(\mathbb{R})}(W_1, \ldots, W_{\mu(\mathbb{R})})$$

if $\mu = \sum_{i=1}^{\mu(\mathbb{R})} \varepsilon_{w_i}$ is an atomization of μ.

The preceding considerations are summarized in the following result with Poisson process $N_n^{**} = \sum_{i=1}^{\tau^*(n)} \varepsilon_{W_i}$.

Theorem 2.3.3. *Fix $b > 0$ and suppose that the family $\{Q_\vartheta : \vartheta \in \Theta\}$ satisfies conditions (5) and (6) for any $\vartheta_0 \in \Theta (\subset \mathbb{R}^k)$. Let $\hat{\vartheta}_n = \hat{\vartheta}_n(W_1, \ldots, W_n)$ be any $\sqrt{n}$-consistent estimator of each ϑ_0 and put $\hat{T}(N_n^{**}) := \hat{\vartheta}_{\tau^*(n)}(W_1, \ldots, W_{\tau^*(n)})$. If $b_n = b + o(\delta_n)$ then*

$$\begin{aligned}\hat{\kappa}(N_n^{**}) &= (N_n^{**}(\mathbb{R}))^{-1}\Gamma^{-1}(\hat{T}(N_n^{**})) \int v_{\hat{T}(N_n^{**})} + \hat{T}(N_n^{**}) \\ &= \tau^*(n)^{-1}\Gamma^{-1}(\hat{T}(N_n^{**})) \sum_{i=1}^{\tau^*(n)} v_{\hat{T}(N_n^{**})}(W_i) + \hat{T}(N_n^{**})\end{aligned}$$

is an asymptotically efficient estimator i.e., asymptotically efficient in ϑ_0 for all $\vartheta_0 \in \Theta$.

By means of Lemma 2.3.1 and Theorem 2.3.2, we can now establish asymptotic efficiency of an estimator $\hat{\kappa}(N_n)$ of ϑ_0 along regular paths in $\mathcal{P}(\mathcal{F}(C_1, C_2, C_3), \Theta)$.

Definition 2.3.4. *A path $\lambda \to P_{\vartheta_0 + \lambda t} \in \mathcal{P}(\mathcal{F}(C_1, C_2, C_3), \Theta)$, $t \in \mathbb{R}^k$, $\lambda \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$, is regular in ϑ_0, if the corresponding marginal densities of X satisfy $| g_{\vartheta_0 + \lambda t}(x) - g_{\vartheta_0}(x) | = o(\lambda)$ for $\lambda \to 0$.*

Now, we are in the proper position to state our main result.

Theorem 2.3.5. *Suppose that the family $\{Q_\vartheta : \vartheta \in \Theta\}$ satisfies condition (5) and (6) for any $\vartheta_0 \in \Theta$. Let $vol(S_n) \to 0, \| a_n \| \to 0$, $n\ vol(S_n) \| a_n \|^{4/d} \to 0$, and $n\ vol(S_n) \to \infty$. Then*

$$\hat{\kappa}(N_n) := (N_n(\mathbb{R}))^{-1}\Gamma^{-1}(\hat{T}(N_n)) \int v_{\hat{T}(N_n)} dN_n + \hat{T}(N_n)$$

is asymptotically efficient in the sense

$$\delta_n^{-1}(\hat{\kappa}(N_n) - \vartheta_0 - t\delta_n) \to_{\mathcal{D}} \mathcal{N}\Big(0, \Gamma^{-1}(\vartheta_0)/g_{\vartheta_0}(x)\Big)$$

under regular paths $P_{\vartheta_0 + t\delta_n}$ in $\mathcal{P}$, whereas for any other estimator sequence $T_n(N_n)$ of ϑ_0 based on N_n, which is asymptotically δ_n-regular along regular paths $P_{\vartheta_0+t\delta_n}$, we have

$$\delta_n^{-1}(T_n(N_n) - \vartheta_0 - t\delta_n) \to_{\mathcal{D}} H * \mathcal{N}\Big(0, \Gamma^{-1}(\vartheta_0)/g_{\vartheta_0}(x)\Big)$$

for some probability measure H on $\mathbb{R}^k$.

Proof. By Lemma 2.3.1 we can replace N_n by N_n^{**} and hence, the assertion follows from the asymptotic efficiency of $\hat{\kappa}(N_n^{**})$ established in Theorem 2.3.3 together with elementary computations.

Remark . If we choose $a_{n,1} = \ldots = a_{n,d} = c_n$, then we obtain $vol(S_n) = c_n$, $n \cdot vol(S_n) \| a_n \|^{4/d} = O(nc_n^{(d+4)/d})$ and $\delta_n = (nc_n)^{-1/2}$. The choice $c_n = l^2(n)n^{-d/(d+4)}$ with $l(n) \to 0, n \to \infty$, results in δ_n of minimum order $O(l(n)^{-1}n^{-2/(d+4)})$. The factor $l(n)^{-1}$ which may converge to infinity at an arbitrarily slow rate actually ensures that the approximation of N_n by N_n^{**} is close enough so that asymptotically the nonparametric part of the problem of the estimation of ϑ_0 i.e., the joint density of (X, Y), is suppressed. In particural, it ensures the asymptotically unbiasedness of the optimal estimator sequence $\hat{\kappa}(N_n)$.

References

[1] **Aldous D.**, *Probability Approximations via the Poisson Clumping Heuristic*, Springer, New York, 1989.

[2] **Arratia R., Goldstsein L. and Gordon L.**, Poisson approximation and the Chen-Stein method (with discussion), *Statistical Science*, 5(1990),403-434.

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

[4] **von Bortkiewicz L.**, *Das Gesetz der kleinen Zahlen,* Teubner, Leipzig, 1898.

[5] ***Chaudhuri P.**, Nonparametric estimates of regression quantiles and their* local Bahadur representation, *Ann. Statist.*, **19***(1991),760-777.*

[6] **Csiszár I.**, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, *MTA Mat. Kut. Int. Kõzl.*, **8**(1963),85-108.

[7] **Davison A.C. and Smith R.L.**, Models for exceedances over high thresholds (with discussion), *J.R. Statist.Soc. B*, **52**(1990),393-442.

[8] **Eubank R.L.**, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1988.

[9] **Falk M. and Marohn F.**, Asymptotic optimal tests for conditional distributions, Preprint Katholische Universität, Eichstätt, 1991.

[10] **Falk M. and Reiss R.-D.**, Poisson approximation of empirical processes, *Statist.Probab.Letters,* (to appear in 1992).

[11] **Falk M. and Reiss R.-D.**, Statistical inference of conditional curves: Poisson process approach, *Ann.Statist.*, (to appear in 1992).

[12] **Falk M., Hüsler J. and Reiss R.-D.**, *Laws of small numbers: Extremes and rare events*, DMV Seminar, Birkhäuser, Basel, forthcoming 1992.

[13] **Galambos J.**, *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Krieger, Malabar, Florida, 1987.

[14] **Goldstein L. and Messer K.**, Optimal plug-in estimators for nonparametric functional estimation, Technical report # 277, Stanford University, 1990.

[15] **Härdle W., Janssen P. and Serfling R.J.**, Strong uniform consistency rates for estimators of conditional functionals, *Ann. Statist.*, **16**(1988),1428-1449.

[16] **Ibragimov I.A. and Has'minskii R.Z.**, *Statistical Estimation*, Springer, Berlin, 1981.

[17] **Jones M.C. and Hall P.**, Mean squared error properties of kernel estimates of regression quantiles, *Statist. Probab. Letters*, **10**(1990),283-289.

[18] **Leadbetter M.R., Lindgren G. and Rootzén H.**, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer, New York, 1983.

[19] **LeCam L.**, *Asymptotic Methods in Statistical Decision Theory*, Springer Series in Statistics, Springer, New York, 1986.

[20] **Liese F. and Vajda I.**, *Convex Statistical Distances*, Teubner-Texte zur Mathematik 95, Teubner, Leipzig, 1987.

[21] **Manteiga W.G.**, Asymptotic normality of generalized functional estimators dependent on covariables, *J. Statist. Planning Inference*, **24**(1990),377-390.

[22] **Pfanzagl J.** *Estimation in Semiparametric Models (Some Recent Developments)*, *Lecture Notes in Statistics*, **63**, Springer, Berlin-Heidelberg, 1990.

[23] **Reiss R.-D.**, *Approximate Distributions of Order Statistics (With Applications to Nonparametric Statistics)*, Springer Series in Statistics, Springer, New York, 1989.

[24] **Reiss R.-D.** *A Course on Point Processes*, forthcoming monograph, to appear in Springer Series in Statistics, Springer, New York, 1992.

[25] **Resnick S.I.**, *Extreme Values, Regular Variation, and Point Processes*, Applied Probability 4, Springer, New York, 1987.

[26] **Samanta M.**, Nonparametric estimation of conditional quantiles, *Statist. Probab.Letters*, **7**(1989),407-412.

[27] **Serfling R.J.**, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

[28] **Stone C.J.**, Optimal rates of convergence for nonparametric estimators, *Ann.Statist.*, **8**(1980),1348-1360.

[29] **Stone C.J.**, Optimal global rates of convergence for nonparametric regression, *Ann.Statist*, **10**(1982),1040-1053.

[30] **Stute W.**, On almost sure convergence of conditional empirical distribution functions, *Ann.Probab.*, **14**(1986),891-901.

[31] **Truong Y.K.** Asymptotic properties of kernel estimators based on local medians, *Ann.Statist.*, **17**(1989),606-617.

[32] **Witting H.** *Mathematische Statistik I.*, Teubner, Stuttgart, 1985.

M. Falk and F. Marohn
Mathematisch–Geographische Fakultät
Katholische Universität Eichstätt
Ostenstr. 26-28
W-8078 Eichstätt, Germany