A General Kernel Functional Estimator with Generalized Bandwidth
– Strong Consistency and Applications –

Rafael Weißbach
Fakultät Statistik, Universität Dortmund
e-mail: Rafael.Weissbach@uni-dortmund.de *

March 26, 2004

Abstract

We consider the problem of uniform asymptotics in kernel functional estimation where the bandwidth can depend on the data. In a unified approach we investigate kernel estimates of the density and the hazard rate for uncensored and right-censored observations. The model allows for the fixed bandwidth as well as for various variable bandwidths, e.g. the nearest neighbor bandwidth. An elementary proof for the strong consistency of the generalized estimator is given that builds on the local convergence of the empirical process against the cumulative distribution function and the Nelson-Aalen estimator against the cumulative hazard rate, respectively.

*Supported by the DFG, SFB 475 "Reduction of Complexity in Multivariate Structures", B1.
AMS subject classifications. 62N01
Keywords. Functional estimation, Density, Hazard rate, Kernel smoothing, Uniform consistency, Empirical process, Nearest neighbor bandwidth, Random censorship.

1 Introduction

For the description of continuous univariate observations various statistical methods are available. Without parametric assumptions estimation of the cumulative distribution function is theoretically well established and common practice by means of the empirical process. If smoothness of the distribution is assumed further insight can be gained from estimation of the density. Rosenblatt (1956) introduced the method of kernel estimation, i.e. the convolution of the empirical process with a density centered at the origin, named kernel. The idea was resumed by Parzen (1962) allowing for various kernels but still depending on a fixed bandwidth. Motivated by practical drawbacks criticism was raised with respect to the inflexibility of the bandwidth definition. Wagner (1975) proposed a variable bandwidth for the density estimation that keeps the number of observations in the convolution window fixed rather than the window width itself. This nearest neighbor bandwidth is still parameterized by a one-dimensional parameter, namely the number of nearest neighbors. Consistency proofs were given for the kernel density estimation with fixed bandwidth e.g. by Parzen (1962), Silverman (1978) and Stute (1982a). Consistency for the case of the nearest neighbor bandwidth were considered by Wagner (1975) and Ralescu (1995). A large area of application for distribution estimation is found in the context of survival analysis where the survival distribution is estimated by the method of Kaplan and Meier (1958) which is heavily referred to in applied work. In this context right-censoring is of major concern known to cause a bias for the distribution estimation when ignored. Adaption for right-censoring in density estimation is scarce for the variable bandwidth with the exception of Schäfer (1986). However, in the context of survival studies the hazard rate was given preference over the density because of its notion as "instantaneous risk of failure". Similar arguments for the kernel estimation of the hazard rate with respect to the bandwidth hold as for the density estimation. Consistency for censored data and data-adaptive bandwidth definitions were considered e.g. by Tanner and Wong (1984), Gefeller and Dette (1992) or Müller and Wang (1994). We strive to prove the consistency for all cases mentioned in a unified approach and generalize to help avoid further consistency considerations for yet uninspected cases of a cross-product of censoring, bandwidth definition and functional estimation. We given a practically relevant example for such a combination in the context of biometrical survival analysis.

2 Model and notation

Let $X_1, \ldots, X_n$ be i.i.d.univariate random variables with density $f(\cdot)$, then

$$f_n(x) := \int_{\mathbb{R}} K_b(x-y)dF_n(y) = 1/(nb)\sum_{i=1}^{n} K((x-X_i)/b) \tag{1}$$

is the classical Parzen-estimator of the density, derived from the convolution idea, with empirical distribution function

$$F_n(x) := 1/n \sum_{i=1}^{n} I_{\{X_i \le x\}}(x),$$

kernel function $K(\cdot)$ and and $K_b(\cdot) := 1/bK(\cdot/b)$ (see Parzen (1962)).

The fixed bandwidth incorporates a *varying numbers of data points* in the estimation for different values of x. The latter leads to the well-known trade-off between a strong bias for large density areas and a variability for small density areas as well as it is not data-adaptive smoothing.

The nearest neighbor (NN) bandwidth was introduced by Wagner (1975) for density estimation to reduce the deficits of the fixed bandwidth. The data-adaptive bandwidth $R_n^{NN}(t)$ keeps the number k of data points involved in the bandwidth choice at each point t equal. The first formalization of the bandwidth was with respect to the order statistics of the differences from the data points to the reference point but to enable a generalization to censored data we use the following

$$R_n^{NN}(t) := \inf\left\{r > 0 : |F_n(t-r/2) - F_n(t+r/2)| \ge k/n\right\}. \tag{2}$$

The definition realizes the collection of empirical mass in a window around a "time" t of magnitude k/n, i.e. a summation of k jumps of the empirical process indicating k nearest neighbors.

To incorporate the fixed bandwidth and the censoring information in the following analysis and to introduce a generalized bandwidth we replace the empirical process $F_n(\cdot)$ in (2) by a general stochastic process for which we require monotony. We denote the process $\tilde{\Psi}_n(\cdot)$ as "smoothing process" and define

$$R_n(t) := \inf\left\{r > 0 : \left|\tilde{\Psi}_n\left(t - r/2\right) - \tilde{\Psi}_n\left(t + r/2\right)\right| \geq p_n\right\} \tag{3}$$

with univariate bandwidth parameter $p_n \in \mathbb{R}^+$. Let us summarize three important examples:

- $\tilde{\Psi}_n(\cdot) = F_n(\cdot)$ with $p_n = k/n$ yields the k-nearest neighbor bandwidth $R_n^{NN}(t)$,
- $\tilde{\Psi}_n(\cdot) = S_n(\cdot)$ (Kaplan-Meier estimate of the survival function for censored data) with $p_n = k/n$ yields the k-nearest neighbor bandwidth in the survival analytic setup (cf. Gefeller and Dette (1992)) and
- $\tilde{\Psi}_n(\cdot) = c \cdot id(\cdot) + d$ with $p_n = |c| \cdot b$ yields the fixed bandwidth b for any c and d.

Consider now that not only the empirical process, as in (1), can be smoothed to get an estimate of the density. You might as well, in a censored scenario of a survival analysis, smooth the Nelson-Aalen-estimate of the cumulative hazard rate to achieve an estimate of the hazard rate. In general you may estimate the derivative $\psi(\cdot)$ of any functional $\Psi(\cdot)$ you have an estimate $\Psi_n(\cdot)$ of, i.e.

$$\psi_n(x) := \int_{\mathbb{R}} \frac{1}{R_n(t)} K\left(\frac{x-t}{R_n(t)}\right) d\Psi_n(t), \tag{4}$$

where we have now already made use of the generalized bandwidth $R_n(\cdot)$.

The most used scenarios in practice and according estimators of the cumulative function are given in Table I.

Table I: Estimates of functions used to estimate derivatives: Right-censored and uncensored design

	Uncensored scenario	Censored scenario
Cumulative distribution function	Empirical distribution function	Kaplan-Meier survival estimator
Cumulative hazard rate	Simplified Nelson-Aalen estimator	Nelson-Aalen estimator

Remark 1. The suggested estimator is a variable kernel estimator and does not share the disadvantage of the genuine nearest-neighbor estimator of not integrating the density estimate to 1 (cf. Breiman et al. (1977)).

3 Strong consistency

To state the strong consistency of the estimate $\psi_n(\cdot)$ (4) smoothness assumptions on the function to estimate, assumptions on the rate of convergence of the bandwidth, and assumptions on the kernel are crucial for the analysis. For the functions $\Psi : \mathbb{R} \to \mathbb{R}_0^+$ and $\tilde{\Psi} : \mathbb{R} \to \mathbb{R}_0^+$ we assume

$$|\Psi(x) - \Psi(y)| \leq M|x - y|, \quad |\Psi(x) - \Psi(y)| \geq m|x - y| \tag{5}$$

$\forall x, y \in [A, B] \subset \mathbb{R}$ $(A < B)$ where $0 < m \leq M < \infty$ and $0 < \tilde{m} \leq \tilde{M} < \infty$. The same has to hold for $\tilde{\Psi}(\cdot)$ with constants $\tilde{M}$ and $\tilde{m}$ This implies the existence of the derivatives $\psi(\cdot)$ and $\tilde{\psi}(\cdot)$ with $m \leq |\psi(x)| \leq M$ and $\tilde{m} \leq |\tilde{\psi}(x)| \leq \tilde{M}$ $\forall x \in [A, B]$. Furthermore we assume $\psi(\cdot)$ and $\tilde{\psi}(\cdot)$ to be Lipschitz-continuous on $[A, B]$ with constants L_ψ and $L_{\tilde{\psi}}$.

The due to their monotony the stochastic processes $\Psi_n(\cdot)$ and $\tilde{\Psi}_n(\cdot)$ imply measures on $\mathbb{R}$. E.g. for $\Psi_n(\cdot)$ and interval $I = [I_l; I_u]$ the mass is defined by $\Psi_n(I) := |\Psi_n(I_l) - \Psi_n(I_u)|$. As asymptotic behavior we assume that exists a sequence of a right-continuous and monotonous stochastic processes $\Psi_n(x)$ with $0 < D < \infty$ s.t.

$$P\left\{\lim_{n\longrightarrow\infty}\sup \sup_{I\subset[A,B],\Psi(I)\leq p_n} \frac{|\Psi_n(I)-\Psi(I)|}{\sqrt{\log(n)p_n/n}} = D\right\} = 1. \tag{6}$$

The assumed convergence properties $0 < p_n \in \mathbb{R} < 1$, $p_n \longrightarrow 0$ and $(np_n)/\log(n) \longrightarrow \infty$ resemble those of the fixed bandwidth in MISE analysis. Keep in mind here that p_n is linked closely to the fixed bandwidth as seen in the examples of the generalized bandwidth (3). The same has to hold for $\tilde{\Psi}_n(x)$ with constant $0 < \tilde{D} < \infty$.

Such properties are proven for the empirical distribution, the product limit (Kaplan-Meier) survival function estimate and the Nelson-Aalen estimator, i.e. all functionals listed in Table I, by Stute (1982b) and Schäfer (1986) and are trivial for $\tilde{\Psi}_n(\cdot) = c \cdot id(\cdot) + d$.

Lipschitz-continuity and strict positiveness on a closed interval $[A, B]$ are assumed for the functions $\psi(\cdot)$ and $\tilde{\psi}(\cdot)$ to simplify the calculations. The the positiveness of $\psi(\cdot)$ is not necessary for the density estimation with fixed bandwidth considered in Silverman (1978) and b) the hazard rate estimation with fixed bandwidth under random censoring considered in Diehl and Stute (1988). Continuity, piecewise Lipschitz continuity L_K and bounded total variation $V(K)$ are mild restrictions on the kernel because valid for all kernels used in practice despite the gaussian density kernel.

The stated properties assumed allow us to state a convergence of $\psi_n(\cdot)$.

Theorem 3.1 *Under the above conditions exists a constant* $D_0 \leq \max\{D_1 + D_2, D_3\}$ *such that*

$$P\left\{\lim_{n\longrightarrow\infty}\sup \frac{\sup_{x\in[a,b]}|\psi_n(x)-\psi(x)|}{\sqrt{\log(n)/(np_n)}+p_n} = D_0\right\} = 1 \qquad \forall\ [a,b] \in (A,B),$$

where $D_1 := 2\tilde{D}M\tilde{M}^2\tilde{m}^{-2}(\sup(K)+L_K\tilde{M}\tilde{m}^{-1})$, $D_2 := D\tilde{M}M^{1/2}V(K)\tilde{m}^{-1/2}$ *and* $D_3 := 2\tilde{M}^3ML_KL_{\tilde{\psi}}\tilde{m}^{-5} + 2\sup(K)L_{\tilde{\psi}}\tilde{M}^2M\tilde{m}^{-4} + L_{\psi}\tilde{m}^{-1}$.

The core of the proof is the common technic of integration by parts (see Parzen (1962)) which decomposes the error into a factor of total variation of the kernel and a factor for the local proximity of the stochastic processes $\Psi_n(\cdot)$ to its limit $\Psi(\cdot)$. The total variation is calculated in an elementary fashion. The local convergence of the stochastic process $\Psi_n(\cdot)$ to $\Psi(\cdot)$ with the assumed rate of $((\log np_n)/n)^{1/2}$ is achieved by exponential inequalities, i.e. exponential bounds on the sum of bounded random variables stated by Bennett (1962) and Hoeffding (1962) who refer to the modification of the Tchebycheff inequality by Bernstein (1924). The additional contribution of the variability of the bandwidth to the error is taken into account adapting the method of proof in Schäfer (1986).

The complete proof is given in the Appendix.

Remark 2. For the kernel density estimation with fixed bandwidth (1) Theorem 3.1 states the strong consistency. To see this, choose $\Psi_n(\cdot)$ as empirical process $F_n(\cdot)$ and if we consider the third example of the generalized bandwidth 3 , i.e. choose $\tilde{\Psi}_n(\cdot)$ as linear function $c \cdot id(\cdot) + d$ with smoothing parameter $p_n = |c| \cdot b$.

With the further simplification of a triangular kernel, i.e. $K(\cdot) = 2|\cdot|I_{[-1/2,1/2]}$ the example allows us to illustrate the proof here conceptionally.

Let us consider the difference of the estimate and its expectation, i.e. the smoothed density $\bar{f}(x)$.

$$|f_n(x) - \bar{f}(x)| = \left| \int_{x-b/2}^{x+b/2} K_b(x-t)dF_n(t) - \int_{x-b/2}^{x+b/2} K_b(x-t)dF(t) \right| \tag{7}$$

From a heuristic point of view it is clear, that the difference (7) is dependent on the proximity of $F_n(t)$ to $F(t)$. The proximity is only relevant within the interval $[x - b/2, x + b/2]$ because outside the kernel vanishes. The kernel however must have an impact because it quantifies the distribution of empirical mass within the interval. To understand the inequality (11) consider the discrete analog. Let be $\{x - b/2 = t_1 < t_2 < \ldots < t_p < t_{p+1} = x + b/2\}$ be a partition of the interval $[x - b/2, x + b/2]$. With the notation of $A_i := [t_i, t_{i+1}],\ i = 1, \ldots, p$, the estimator is $f_n(x) \approx \sum_{i=1}^{p} K_b(x - t_i)F_n(A_i)$. For the smoothed density holds

$\bar{f}(x) \approx \sum_{i=1}^{p} K_b(x-t_i)F(A_i)$, such that the absolute difference is

$$|f_n(x) - \bar{f}(x)| \approx |\sum_{i=1}^{p} K_b(x-t_i)(F_n(A_i) - F(A_i)| \quad (8)$$

$$\leq \sup_{I \in [x-b/2, x+b/2]} |F_n(I) - F(I)|| \sum_{i=1}^{p} K_b(x-t_i)| \quad (9)$$

$$= \sup_{I \in [x-b/2, x+b/2]} |F_n(I) - F(I)|1/bO(1). \quad (10)$$

The first term in (8) quantifies the local proximity of the empirical process to the cumulative distribution function. The rate of convergence for a bounded density $f(\cdot) < M$ is $O(\sqrt{\log(n)b/n})$ (see Schäfer (1986)) so that the total rate is $O(\sqrt{\log(n)/(bn)})$.

The second term in (8) is asymptotically the total variation - $V(K) := \int |dK|$ - formally arising when applying partial integration to (7). As a consequence we have in the continuous version

$$|f_n(x) - \bar{f}(x)| \leq \sup_{I \subset [x-b/2, x+b/2]} |F_n(I) - F(I)| V_{[x-b/2, x+b/2]}(K_b(x-t)) \quad (11)$$

The total variation for the triangular kernel is clearly $\int_{-b/2}^{b/2} 4/b^2 dt = 4/b$. Again with Schäfer (1986) we have for sufficiently large n almost sure

$$|f_n(x) - \bar{f}(x)| \leq 36\sqrt{M \log n/(nb)}$$

For sufficiently large n holds for the bias

$$|\bar{f}(x) - f(x)| \leq \int_{[x-b/2, x+b/2]} K_b(x-t)|f(t) - f(x)|dt$$

$$\leq \sup_{t \in [x-b/2, x+b/2]} |f(t) - f(x)| \leq \sup_{t \in [x-b/2, x+b/2]} L_f|x-t|$$

$$\leq 1/2 L_f b$$

(with Lipschitz-constant of the density $f(\cdot)$ L_f). Note that the order of convergence of the bias b does not resemble that of b^2 typically derived in MISE context (cf. Wand and Jones (1995)). The reason is the lack of assumption of symmetry for the kernel function which allows for boundary kernels.

Hence almost sure,

$$|f_n(x) - f(x)| \leq 36\sqrt{M \log n/(nb)} + 1/2L_f b$$

independently of x on $[a, b]$ and for sufficiently large n. Such that the uniform consistency holds for sequences b_n fulfilling $b_n \to 0$ and $(nb_n)/\log n \to \infty$. Keep in mind that b_n and p_n differ only by a constant factor.

Remark 3. As a meaningful example, Theorem 3.1 warrants the uniform asymptotics for the hazard rate estimator with nearest-neighbor bandwidth for right-censored data. As common we say that $X_i = \max\{T_i, C_i\}$ are the observed either survival times T_i or censoring times C_i and $\delta_i = I_{\{X_i = T_i\}}$ indicate the censoring for $i = 1, \ldots, n$ independent observations with interesting hazard function $h(\cdot)$ of the T_i's. Keep in mind the second example of a generalized bandwidth 3 and let $H_n(\cdot)$ represent the Nelson-Aalen estimator of the cumulative hazard rate for censored data

$$H_n(x) = \sum_{i:X_{(i)} \leq x} \frac{\delta_{(i)}}{n - i + 1}.$$

Hence the suggested hazard rate estimate has the form

$$\begin{aligned} \hat{h}_{R_n^{NN}}(x) &:= \sum_{i=1}^{n} \frac{\delta_{(i)}}{n-i+1} \frac{1}{R_n^{NN}(X_i)} K\left(\frac{X_i - x}{R_n^{NN}(X_i)}\right) \\ &= \int_{\mathbb{R}} \frac{1}{R_n^{NN}(t)} K\left(\frac{X_i - x}{R_n^{NN}(t)}\right) dH_n(t). \end{aligned}$$

For that approach strong consistency was not considered before although being of relevance in survival analysis of biometrical data.

Remark 4. The asymptotic rate of the boundary of the uniform error

$$\sqrt{\log(n)(np_n)} + p_n$$

implies a maximal rate of convergence of of $(\log n/(p_n n))^{1/3}$ if the bandwidth parameter p_n satisfies the same rate. It has to be mentioned that the optimal rate of uniform absolute convergence as well as the consequent rate of the bandwidth of $(\log n/(p_n n))^{1/3}$ has only been reported before for the fixed bandwidth density estimate by Schäfer (1986).

A Appendix

We give an elementary proof for Theorem 3.1 since the counting process methodology as in Andersen et al. (1993) is not applicable. Keep the assumptions from Section 3 in mind. Note first that it is equivalent to state the almost sure asymptotics

$$\textit{Exists constant } D \leq 1 \textit{ s.t. } P\left\{\limsup_{n\to\infty} S_n(Y_1(\omega),\ldots,Y_n(\omega))/a_n = D\right\} = 1.$$

as

$$\textit{For all } \alpha > 1 \textit{ holds } P\left\{\omega | \exists N \in \mathbb{N}\ \forall n > N:\ S_n(Y_1(\omega),\ldots,Y_n(\omega)) \leq \alpha a_n\right\} = 1.$$

with positive zero-sequence a_n. The results follows from Hewitt and Savage (1955).

The proof of Theorem 3.1 is conducted in three steps. First the random bandwidth $R_n(t)$ is replaced by its deterministic analogue $r_n(t) := \inf\{r > 0 || \tilde{\Psi}(t + r/2) - \tilde{\Psi}(t - r/2)| \geq p_n\}$. In the second - and crucial - step the convergence of the kernel estimate with variable - but deterministic - bandwidth $r_n(t)$ to the function $\psi(\cdot)$ convoluted with the kernel is investigated. In the last step the convergence of the bias, i.e. the difference of the latter convoluted $\psi(\cdot)$ to $\psi(\cdot)$, is quantified. The differences add up to the overall difference:

$$\begin{aligned}\sup_{x\in[a,b]} |\psi_n(x) - \psi(x)| \leq & \sup_{x\in[a,b]} |\psi_n(x) - \breve{\psi}_n(x)| \\ & + \sup_{x\in[a,b]} |\breve{\psi}_n(x) - \bar{\psi}_n(x)| + \sup_{x\in[a,b]} |\bar{\psi}_n(x) - \psi(x)|,\end{aligned} \tag{12}$$

with

$$\breve{\psi}_n(x) := \int_{\mathbb{R}} \frac{1}{r_n(t)} K\left(\frac{x-t}{r_n(t)}\right) d\Psi_n(t) \quad \text{and the expected estimate}$$
$$\bar{\psi}_n(x) := \int_{\mathbb{R}} \frac{1}{r_n(t)} K\left(\frac{x-t}{r_n(t)}\right) d\Psi(t).$$

The assumption of the compact support $[-1/2; 1/2]$ of the kernel function $K(\cdot)$ simplifies the proof since we can restrict the integration support for sufficiently large n using $r_n(\cdot)$ and $R_n(t)$ respectively. The the positivity $\tilde{\psi}(\cdot) > \tilde{m}$ (see assumption (5)) enables to bound the support for $r_n(\cdot)$ by $I_n(x) := [x - p_n/(2\tilde{m}), x + p_n/(2\tilde{m})]$. For sufficiently large n the interval $I_n(x)$ is lying in any closed interval $[a, b]$ contained in (A, B). For the random bandwidth the support is almost surely bounded by the interval $I_n^\alpha(x) := [x - \alpha p_n/\tilde{m}, x + \alpha p_n/\tilde{m}]$ for $R_n(\cdot)$, because

$$P\left\{\exists N \in \mathbb{N}\ \forall n > N : \sup_{x \in [a,b]} \sup_{t \in \mathbb{R} \setminus I_n^\alpha(x)} K\left(\frac{t-x}{R_n(t)}\right) = 0\right\} = 1. \tag{13}$$

for any $\alpha > 1$. The result is clear noting the local convergence rate of $\tilde{\Psi}_n(\cdot)$ (see 6).

A.1 Convergence of the stochastic bandwidth

We can now bound the first term in the inequality (12). Almost sure for sufficiently large n (see (13)) is

$$|\psi_n(x) - \breve{\psi}_n(x)| = \left| \int_{\mathbb{R}} \frac{1}{R_n(t)} K\left(\frac{x-t}{R_n(t)}\right) - \frac{1}{r_n(t)} K\left(\frac{x-t}{r_n(t)}\right) d\Psi_n(t) \right| \tag{14}$$

Prior to inspection of the latter bound we note that for $[a, b] \subset (A, B)$ there exists a constant $E \leq \tilde{D}/\tilde{m}$ s.t.

$$P\left\{\lim \sup_{n \longrightarrow \infty} \frac{\sup_{t \in [a,b]} |R_n(t) - r_n(t)|}{\sqrt{(\log(n) p_n)/n}} = E\right\} = 1.$$

This follows from the positive bounds of $\tilde{\psi}(\cdot)$ (see (5)) and the local convergence of $\tilde{\Psi}_n(\cdot)$ (see again (6)).

Now follows for the integrant in (14)

$$\left|\frac{1}{R_n(t)}K\left(\frac{x-t}{R_n(t)}\right)-\frac{1}{r_n(t)}K\left(\frac{x-t}{r_n(t)}\right)\right|$$

$$\leq \left|\frac{1}{R_n(t)}K\left(\frac{x-t}{R_n(t)}\right)-\frac{1}{r_n(t)}K\left(\frac{x-t}{R_n(t)}\right)\right|+\left|\frac{1}{r_n(t)}K\left(\frac{x-t}{R_n(t)}\right)-\frac{1}{r_n(t)}K\left(\frac{x-t}{r_n(t)}\right)\right|$$

$$\leq \sup(K)\left|\frac{1}{R_n(t)}-\frac{1}{r_n(t)}\right|+\frac{\tilde{M}}{p_n}\left|K\left(\frac{x-t}{R_n(t)}\right)-K\left(\frac{x-t}{r_n(t)}\right)\right| \tag{15}$$

because of the Lipschitz-continuity of $\tilde{\Psi}(\cdot)$.

For the first absolute follows

$$\begin{aligned}\sup_{t\in[a,b]}\left|\frac{1}{R_n(t)}-\frac{1}{r_n(t)}\right| &= \sup_{t\in[a,b]}\left|\frac{r_n(t)-R_n(t)}{R_n(t)r_n(t)}\right| \\ &\leq \tilde{D}\alpha^2\tilde{M}^2/\tilde{m}\sqrt{\log(n)/(np_n^3)}\end{aligned}$$

because of $\inf_{t\in[a,b]} r_n(t) \geq p_n/\tilde{M}$, $\sup_{t\in[a,b]}|R_n(t)-r_n(t)| \leq C/\tilde{m}\sqrt{(\log(n)p_n)/n}$ $\forall\ C > \tilde{D}$ and hence for $C = \tilde{D}\alpha$ and $\inf_{t\in[a,b]} R_n(t) \geq p_n/(\alpha\tilde{M})$. The last statement for $\inf_{t\in[a,b]} R_n(t)$ follows from the first two together with the convergence rate for p_n determined in Section 3.

For the second absolute in (15) follows

$$\left|K\left(\frac{x-t}{R_n(t)}\right)-K\left(\frac{x-t}{r_n(t)}\right)\right| \leq k\left(\sup_{x\in[a,b],t\in I_n^\alpha(x)}\left|\frac{x-t}{R_n(t)}-\frac{x-t}{r_n(t)}\right|\right)$$

with modul of continuity $k(\cdot)$ of K defined as $k(\delta) := \sup\{|K(x)-K(y)| :\ |x-y| \leq \delta\}$. Furthermore,

$$\begin{aligned}\left|\frac{x-t}{R_n(t)}-\frac{x-t}{r_n(t)}\right| &= |x-t|\left|\frac{1}{R_n(t)}-\frac{1}{r_n(t)}\right| \\ &\leq \alpha p_n/\tilde{m}\tilde{D}\alpha^2\tilde{M}^2/\tilde{m}\sqrt{\log(n)/(np_n^3)}\end{aligned}$$

because $t \in I_n^\alpha(x)$. Hence the modul of continuity $k(\cdot)$ of K is bounded by

$$\left| K\left(\frac{x-t}{R_n(t)}\right) - K\left(\frac{x-t}{r_n(t)}\right) \right| \leq \alpha^3 L_K \tilde{D} \tilde{M}^2 / \tilde{m}^2 \sqrt{\log(n)/(np_n)},$$

because the kernel is assumed to be piecewise Lipschitz-continuous and for sufficiently large n we can assume that only finite discontinuities exist in the support of K.

The Lipschitz-continuity of $\Psi_n(\cdot)$ (5) and the local convergence (6) ensure that $\Psi_n(I_n^\alpha(x)) \leq 2\alpha^2 p_n M/\tilde{m}$ which simplifies (14) to

$$\begin{aligned} |\psi_n(x) - \breve{\psi}_n(x)| &\leq 2\alpha^2 p_n M/\tilde{m} \Big[\sup(K) \tilde{D} \alpha^2 \tilde{M}^2 / \tilde{m} \sqrt{\log(n)/(np_n^3)} \\ &+ \tilde{M}/p_n L_K \alpha^3 \tilde{D} \tilde{M}^2 / \tilde{m}^2 \sqrt{\log(n)/(np_n)} \Big] \\ &= \alpha^5 2 M \tilde{M}^2 / \tilde{m}^2 \tilde{D} \Big[\sup(K)/\alpha \sqrt{\log(n)/(np_n)} + L_K \tilde{M}/\tilde{m} \sqrt{\log(n)/(np_n)} \Big] \\ &\leq \alpha^5 2 \tilde{D} M \tilde{M}^2 / \tilde{m}^2 \left(\sup(K) + L_K \tilde{M}/\tilde{m} \right) \sqrt{\log(n)/(np_n)}. \end{aligned}$$

Since $x \in [a, b]$ and $\alpha > 1$ have been arbitrary we have

$$P\left\{ \lim \sup_{n \longrightarrow \infty} \frac{\sup_{x \in [a,b]} |\psi_n(x) - \breve{\psi}_n(x)|}{\sqrt{\log(n)/(np_n)}} = D_1' \right\} = 1$$

for

$$D_1' \leq 2\tilde{D} M \tilde{M}^2 / \tilde{m}^2 \left(\sup(K) + L_K \tilde{M}/\tilde{m} \right) =: D_1$$

with $0 < \tilde{D} < \infty$ defined by the local convergence of the stochastic process $\tilde{\Psi}_n(\cdot)$.

A.2 Convergence of the convoluted process

For the second term in (12) let be $\alpha > 1$. For sufficiently large n it holds $|\breve{\psi}_n(x) - \bar{\psi}_n(x)| = |\int_{I_n(x)} \frac{1}{r_n(t)} K(\frac{t-x}{r_n(t)}) d\Psi_n(t) - \int_{I_n(x)} \frac{1}{r_n(t)} K(\frac{t-x}{r_n(t)}) d\Psi(t)|$ with $I_n(x) = [x - p_n/(2\tilde{m}), x + p_n/(2\tilde{m})]$ uniformly for all $x \in [a, b]$. Integration by parts for signed measures yields

$$\left|\breve{\psi}_n(x) - \bar{\psi}_n(x)\right| \leq V_{I_n(x)}\left(\frac{1}{r_n(t)}K\left(\frac{t-x}{r_n(t)}\right)\right) \sup_{I Intervall \subset I_n(x)} |\Psi(I) - \Psi_n(I)|,$$

since $\frac{1}{r_n(t)}K(\frac{t-x}{r_n(t)})$ is continuous with respect to t on $I_n(x)$ for sufficiently large n due the Lipschitz condition

$$|r_n(t) - r_n(s)| \leq 2|s-t|$$

for $s, t \in [A, B]$. The latter follows directly from the definition of $r_n(\cdot)$.

Using elementary results for the total variation V as in Schäfer (1986) proofs

$$V_{I_n(x)}\left(\frac{1}{r_n(\cdot)}K\left(\frac{\cdot - x}{r_n(\cdot)}\right)\right) \leq \alpha\tilde{M}V_{(-1,1)}(K)/p_n.$$

Together with

$$\sup_{I Intervall \subset I_n(x)} |\Psi(I) - \Psi_n(I)| \leq C\sqrt{\log(n)Mp_n/(n\tilde{m})}$$

for all $C > D$ and sufficiently large n (see (6)) reveals for $C := D\alpha$: $\exists\ N$ such that $\forall\ n > N$

$$\begin{aligned}\left|\breve{\psi}_n(x) - \bar{\psi}_n(x)\right| &\leq \alpha\tilde{M}V(K)/p_n D\alpha\sqrt{\log(n)Mp_n/(n\tilde{m})} \\ &= \alpha^2 D\tilde{M}\sqrt{M}V(K)/\sqrt{\tilde{m}}\sqrt{\log(n)/(np_n)}.\end{aligned}$$

Since again $\alpha > 1$ was arbitrary now exists $D_2' \leq D\tilde{M}\sqrt{M}V(K)/\sqrt{\tilde{m}} := D_2$ such that

$$P\left\{\lim\sup_{n\longrightarrow\infty} \frac{\sup_{x\in[a,b]}|\breve{\psi}_n(x) - \bar{\psi}_n(x)|}{\sqrt{\log(n)/(np_n)}} = D_2'\right\} = 1.$$

A.3 Convergence of the bias

The last term in (12) is bounded by $|\bar{\psi}_n(x) - \psi(x)| \leq \int_{I_n^1(x)} |\frac{1}{r_n(t)}K(\frac{t-x}{r_n(t)}) - \frac{1}{r_n(x)}K(\frac{t-x}{r_n(x)})||d\Psi(t)| + \int_{I_n^1(x)} \frac{1}{r_n(x)}K(\frac{t-x}{r_n(x)})|\psi(t) - \psi(x)|dt$ for $I_n(x) \subset I_n^1(x) = [x - p_n/\tilde{m}, x + p_n/\tilde{m}]$ and n sufficiently large.

Now, the second addend of the latter boundary is smaller than $\int_{I_n^1(x)} \frac{1}{r_n(x)} K(\frac{t-x}{r_n(x)}) dt \sup_{t \in I_n^1(x)} |\psi(t) - \psi(x)| \leq 1 \cdot L_\psi p_n / \tilde{m}$ because of the Lipschitz-continuity of $\psi(\cdot)$.

For the integrand to the first addend holds for $t \in I_n^1(x)$ and uniformly in $x \in [a,b] \in (A,B)$ $|\frac{1}{r_n(t)} K(\frac{t-x}{r_n(t)}) - \frac{1}{r_n(x)} K(\frac{t-x}{r_n(x)})| \leq = \frac{1}{r_n(t)} |K(\frac{t-x}{r_n(t)}) - K(\frac{t-x}{r_n(x)})| + K(\frac{t-x}{r_n(x)}) |\frac{1}{r_n(t)} - \frac{1}{r_n(x)}|$. The first factor of the first addend is clearly bounded by $\tilde{M}/p_n$ because of the definition of $r_n(t)$ and $\tilde{\psi}(\cdot) < \tilde{M}$. The second factor is bounded by the continuity of K $k(\sup_{x \in [a,b], t \in I_n^1(x)} |\frac{t-x}{r_n(t)} - \frac{t-x}{r_n(x)}|)$. The first factor of the second addend is bounded by $\sup(K)$. For the second factor note that $p_n = \int_{t-r_n(t)/2}^{t+r_n(t)/2} |\tilde{psi}(\xi)| d\xi$ together with the Lipschitz-continuity of $\tilde{psi}(\cdot)$ and $|\tilde{psi}(\cdot)| > \tilde{m}$ imply that for $[a,b] \subset (A,B)$ $|r_n(s) - r_n(t)| \leq L_{\tilde{\psi}} \tilde{m}^{-2} p_n |x - t|$ $\forall\, s,t \in [a,b]$. So that the factor is bounded by $\tilde{M}^2/p_n^2 L_{\tilde{\psi}} p_n \tilde{m}^{-2} |x-t|$. Rearranging the deeds show the integrand to be $\leq \tilde{M}/p_n k(p_n L_{\tilde{\psi}} \tilde{M}^2/\tilde{m}^4) + \sup(K) L_{\tilde{\psi}} \tilde{M}^2/\tilde{m}^3$.

Hence,

$$\lim \sup_{n \longrightarrow \infty} \frac{\sup_{x \in [a,b]} |\bar{\psi}_n(x) - \psi(x)|}{p_n} \leq D_3,$$

with

$$D_3 = 2\tilde{M}^3 M L_K L_{\tilde{\psi}} / \tilde{m}^5 + 2 \sup(K) L_{\tilde{\psi}} \tilde{M}^2 M / \tilde{m}^4 + L_\psi / \tilde{m}.$$

Summation of the three parts completes the proof.

References

Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer, New York.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association* **57** 33–45.

Bernstein, S. (1924). Sur une modification de l'inéqualité de tchebichef. *Annals Science Institute Sav. Ukraine, Sect. Math.1.*

Breiman, L., Meisel, W., and Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics* **19** 135–144.

Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *Journal of Multivariate Analysis* **25** 299–310.

Gefeller, O. and Dette, H. (1992). Nearest neighbour kernel estimation of the hazard function from censored data. *Journal of Statistical Computation and Simulation* **43** 93–101.

Hewitt, E. and Savage, L. (1955). Symmetric measures on cartesian products. *Transactions of the American Mathematical Society* **80** 470–501.

Hoeffding, W. (1962). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58** 13ff.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53** 457–481.

Müller, H.-G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50** 61–76.

Parzen, E. (1962). On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics* **33** 1065–1076.

Ralescu, S. (1995). The law of the iterated logarithm for the multivariate nearest neighbor density estimators. *Journal of Multivariate Analysis* **53** 159–179.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* **27** 832–837.

Schäfer, H. (1986). Local convergence of empirical measures in the random censorship situation with application to density and rate estimators. *Annals of Statistics* **14** 1240–1245.

Silverman, B. (1978). Weak and strong uniform consistency of the kernel estimate of the density and its derivates. *Annals of Statistics* **6** 177–184.

Stute, W. (1982a). The law of the logarithm for kernel density estimators. *Annals of Probability* **10** 414–422.

Stute, W. (1982b). The oscillation behaviour of empirical processes. *Annals of Probability* **10** 86–107.

Tanner, M. and Wong, W. (1984). Data-based nonparametric estimation of the hazard function with application to model diagnostics and exploratory analysis. *Journal of the American Statistical Association* **79** 174–182.

Wagner, T. (1975). Nonparametric estimates of probability densities. *IEEE Transactions on Infromation Theory* **21** 438–440.

Wand, M. and Jones, M. (1995). *Kernel Smoothing.* Chapman & Hall, London.