Noname manuscript No.
(will be inserted by the editor)

Strong Consistency for Delta Sequence Ratios

Wladyslaw Poniatowski · **Rafael Weißbach**

Received: date / Accepted: date

Abstract Almost sure convergence for ratios of delta functions establishes global and local strong consistency for a variety of estimates and data generations. For instance, the empirical probability function from independent identically distributed random vectors, the empirical distribution for univariate independent identically distributed observations, and the kernel hazard rate estimate for right-censored and left-truncated data are covered. The convergence rates derive from the Bennett-Hoeffding inequality.

Keywords kernel smoothing · hazard rate · left-truncation · right-censoring · empirical process

PACS 02.50.Ey

Mathematics Subject Classification (2000) 62G05 · 62G20 · 62N02 · 60G57 · 60G50

1 Introduction

The analysis of continuous univariate observations is frequent in statistical work. Without parametric assumptions, the estimation of the cumulative distribution function with the empirical distribution function is common practice. Consistency may be established by normalizing the empirical process (see e.g. [Shorack and Wellner(1986)]).

If smoothness of the distribution is assumed, further insight can be gained from estimating the density. [Parzen(1962)] introduced the method of kernel estimation, i.e. the convolution of the empirical distribution function with a density centered at the origin, named kernel. Consistency proofs were given for the kernel density estimation with fixed bandwidth

Wladyslaw Poniatowski
Institut für Wirtschafts- und Sozialstatistik, Fakultät Statistik, Technische Universität Dortmund, 44221 Germany.
Tel.: +49/231/7555939
Fax: +49/231/7555284
E-mail: vladyslav.ponyatovskyy@uni-dortmund.de

Rafael Weißbach
Institut für Wirtschafts- und Sozialstatistik, Fakultät Statistik, Technische Universität Dortmund, 44221 Germany.

e.g. by [Parzen(1962)], [Silverman(1978)] and [Stute(1982a)]. A large field of application for distribution estimation was found in connection with survival analysis where the cumulative hazard function distribution is estimated by the method of [Aalen(1978)]. In this context right-censoring is a major concern. The design is typically reflected by analyzing the bivariate, although partly unobservable vector of survival times and censoring times. In connection with survival studies the hazard rate is superior to the density because of its interpretation as 'instantaneous risk of failure'. Consistency proofs for the hazard rate are given in [Schäfer(1986)]. Combining the estimates, [Weißbach(2006)] follows a generalized formulation allowing for density, the hazard function and various data designs, including right-censoring. Most of the above work considers the univariate sequence of observations, censored or not, and accounts for the censoring indicator. The aim of this paper is to explore the idea of vector-valued (partly unobserved) random sequences in order to incorporate (i) further data defects such as left-truncation and (ii) further functions, such as the probability function. Whereas the empirical distribution function, the kernel density estimate and the kernel hazard rate estimate are delta sequences [Walter and Bulum(1979)], we generalize to a *ratio* of delta sequences to accommodate for left-truncation. Using the exponential Bennett-Hoeffding inequality [Hoeffding(1963)], we establish global consistency and, for the smoothed functionals, local convergence rates.

2 Model and Notation

Let $(X_t)_{t\in[0,\infty)}, (Y_t)_{t\in[0,\infty)}, \ldots$ be finite state space stochastic processes with combined random vector of observations $\mathbf{S} : \Omega \to \mathbb{R}^d$. For example, in the survival analysis with right-censored observations one may observe only $X = \min(T,C)$ and $\delta = \mathbf{1}_{\{T<C\}}$, where T is a lifetime censored by an independent random variable C.

Suppose that the process $(X_t)_{t\in[0,\infty)}$ has a positive and differentiable functional characteristic $\Psi(\cdot)$ and its first derivative $\psi(\cdot)$ has to be estimated. We assume that the estimate $\Psi_n(\cdot)$ for $\Psi(\cdot)$ is present, where $\Psi_n(\cdot)$ bases on the series $(\mathbf{S}_i)_{i=1,\ldots,n}$ of i.i.d. vectors of observations. A general approach to the estimation of the derivative $\psi(\cdot)$ from $\Psi_n(\cdot)$ is the well-known kernel estimator first proposed for the density function. In further investigations, the kernel density estimator has been extended, and the consistency of a general kernel estimator with variable bandwidth, for instance, was established in [Weißbach(2006)].

We assume the function $\Psi : \mathbb{R} \to \mathbb{R}_0^+$ and the "smoother" $\tilde{\Psi} : \mathbb{R} \to \mathbb{R}_0^+$ with first derivatives $\psi(\cdot)$, $\tilde{\psi}(\cdot)$ and estimates $\Psi_n(\cdot)$, $\tilde{\Psi}_n(\cdot)$ respectively to be Lipschitz continuous, strictly monotonic increasing and bounded on $[A,B]$.

Moreover, we assume the kernel function $K(\cdot)$ to be bounded and $\Psi_n(x)$ to be right continuous and monotonic increasing on $[A,B]$ with $0 < D < \infty$, so that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi(J) - \Psi_n(J)|}{\sqrt{\log(n)p_n/n}} = D\right\} = 1$$

where $p_n > 0$, $p_n \to 0$ and $(np_n)/\log(n) \to \infty$, i.e. the local rate of $\mathscr{O}(\sqrt{\log(n)p_n/n})$ for the measure $\Psi_n(J)$ is required. The theoretical and empirical measures of the interval $J := [a,b] \subseteq [A,B]$ are defined as $\Psi(J) := \int_J \psi(x)dx$ and $\Psi_n(J) := \Psi_n(b) - \Psi_n(a)$, respectively. The same assumptions and definitions are valid - with a $\tilde{D}$ - for $\tilde{\Psi}(\cdot)$.

The kernel estimator for $\psi(x)$ with variable bandwidth

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

is uniformly consistent on support $[A,B]$ with rate of $\mathscr{O}([\log(n)/(np_n)]^{1/2}+p_n)$. Here, the general bandwidth bases on the "smoother"

$$R_n(x) := \inf\left\{r>0 : \left|\tilde{\Psi}_n\left(x-\frac{r}{2}\right)-\tilde{\Psi}_n\left(x+\frac{r}{2}\right)\right| \geq p_n\right\}.$$

and incorporates e.g. fixed bandwidth or nearest neighbor bandwidth [Weißbach(2006)].

The kernel estimator (1) depends only on the estimate $\Psi_n(\cdot)$ and not directly upon the observations $(\mathbf{S}_i)_{i=1,\dots,n}$. Hence, of interest is to construct a general estimate of $\Psi(\cdot)$ and $\tilde{\Psi}(\cdot)$ with local rate of $\mathscr{O}(\sqrt{\log(n)p_n/n})$ from the present observations. The next goal is to obtain an estimator for $\psi(\cdot)$ by kernel smoothing.

3 General locally consistent estimator

Some estimates with a local rate of $\mathscr{O}(\sqrt{\log(n)p_n/n})$ have already been described in the literature. The empirical distribution function, the Kaplan-Meier estimator and the Nelson-Aalen estimator attain this rate of convergence (see [Stute(1982b), Schäfer(1986)]. Generalizing these cases allows new local consistent estimates for other stochastic processes or other characteristics $\Psi(\cdot)$ and $\psi(\cdot)$.

Let $(\mathbf{S}_i)_{i=1,\dots,n}$ be a series of independent identically distributed random vectors $\mathbf{S}_i : \Omega \to \mathbb{R}^d$. We assume the function $G:\mathbb{R}\to\mathbb{R}_0^+$ to be continuous and the function $G_n:\mathbb{R}\times(\mathbb{R}^d)^n\to\mathbb{R}_0^+$, $(x,s_1,\dots,s_n)\mapsto G_n(x)(s_1,\dots,s_n)$ to be symmetric for each fixed $x\in\mathbb{R}$ and $s_1,\dots,s_n\in\mathbb{R}^d$. Further, we use the simplified notation $G_n(x,\omega)$ for $G_n(x)(\mathbf{S}_1(\omega),\dots,\mathbf{S}_n(\omega))$.

Additionally, we define the mapping $\Delta:\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}_0^+, (x,s)\mapsto\Delta^x(s)$ with simplified notation $\Delta_i^x(\omega)$ for $\Delta^x(\mathbf{S}_i(\omega))$. The estimate for $\Psi(\cdot)$ can be constructed as follows:

$$\Psi_n(x) := \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{S_i^1\leq x\}}\cdot\Delta_i^x}{G_n(S_i^1)}, \tag{2}$$

where S_i^1 is the first element of the vector $\mathbf{S}_i$.

The local consistency of the estimate (2) needs some assumptions on the function to estimate, on the observed random variables and on the rate of $G_n(\cdot)$ to $G(\cdot)$.

(L1) We assume the interval $[A,B]$ with $A\in\mathbb{R}$, $B\in\mathbb{R}$ and $A<B$.
(L2) We assume the function $\Psi:\mathbb{R}\to\mathbb{R}^+$ to be continuous, positive and strictly monotonic increasing on $[A,B]$.
(L3) We assume $0\leq\Delta_i^x\leq\Delta_{max}<\infty$ for each $x\in[A,B]$.
(L4) Let a constant $M := \sup_{x\in[A,B]}[G(x)]^{-1} < \infty$ exist, we assume then $[\mathbf{1}_{\{x\leq a\}}\Delta_i^a-\mathbf{1}_{\{x\leq b\}}\Delta_i^b][G(x)]^{-1}<2\Delta_{max}M$ for $i=1,\dots,n$, $a,b\in[A,B]$, $x\in\mathbb{R}$.
(L5) We assume also $[\mathbf{1}_{\{x\leq a\}}\Delta_i^a-\mathbf{1}_{\{x\leq b\}}\Delta_i^b][G(x)-G_n(x)]=0$ for all $x\notin[a,b]\subseteq[A,B]$.
(L6) Furthermore, we assume $E\left(\frac{\mathbf{1}_{\{S_i^1\leq x\}}\cdot\Delta_i^x}{G(S_i^1)}\right)=\Psi(x)$ for each fixed $x\in[A,B]$
(L7) We assume for $G(x)$ and $G_n(x)$ a constant $0\leq D\leq D_G<\infty$ to exist, so that

$$P\left\{\limsup_{n\to\infty}\frac{\sup_{x\in[A,B]}|G(x)-G_n(x)|}{\sqrt{\log(n)/n}}=D\right\}=1$$

(L8) J is an interval with $J:=[a,b]\subseteq[A,B]$ and $a\leq b$.
(L9) For the series (p_n) we assume $p_n>0$, $p_n\to 0$ and $np_n/\log(n)\to\infty$.

Theorem 1 *Under conditions (L1)-(L9), there exists a constant*

$$0 \le D \le 2(\sqrt{2 \cdot (2\Delta_{max} M + \Psi(B))} + D_G M)$$

so that

$$P\left\{ \limsup_{n\to\infty} \frac{\sup_{J \subseteq [A,B], \Psi(J) \le p_n} |\Psi(J) - \Psi_n(J)|}{\sqrt{\log(n) p_n / n}} = D \right\} = 1.$$

In order to prove Theorem 1 we specify the help function

$$\Psi_n^*(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{1}_{\{S_i^1 \le x\}} \cdot \Delta_i^x}{G(S_i^1)}$$

and the help measure

$$\Psi_n^*(J) := \Psi_n^*(b) - \Psi_n^*(a).$$

The aim is to split the difference $|\Psi_n(J) - \Psi(J)|$ in two parts using the help measure and to show the bounds almost sure for each term separately. Section A.3 is the detailed proof of Theorem 1.

4 General globally consistent estimator

To complete the construction of the general estimator (2) we need a $\mathscr{O}(\sqrt{\log(n)/n})$ consistent estimator for the function $G(x)$. In fact, estimates with this rate of convergence are also known. The rate of $\mathscr{O}(\sqrt{\log(n)/n})$ has be established for the empirical distribution function by [Földes and Rejtö(1981)].

Let $(\mathbf{S}_i)_{i=1,\dots,n}$ be the series of i.i.d. random vectors from Section 3. We define the mapping $H : \mathbb{R}^d \to \mathscr{P}(\mathbb{R})$ with simplified notation $H_i(\omega)$ for $H(\mathbf{S}_i(\omega))$. A possible construction of $G_n(x)$ can be derived from the generalization of the empirical distribution function as follows:

$$G_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x, \tag{3}$$

where the definition of Γ_i^x is identical to Δ_i^x from Section 3.

Similar to the local consistent estimate (2), the global consistency of (3) needs some assumptions on the function to estimate and on the observed random variables.

(G1) We assume the interval $[A,B]$ with $A \in \mathbb{R}$, $B \in \mathbb{R}$ and $A < B$.
(G2) We assume $0 \le \Gamma_i^x \le \Gamma_{max} < \infty$ for each $x \in [A,B]$.
(G3) We assume the function $G : \mathbb{R} \to \mathbb{R}_0^+$ to be nonnegative with $G(x) \le G_{max} < \infty$ for each $x \in [A,B]$.
(G4) Furthermore, we assume $E\left(\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x\right) = G(x)$ for each fixed $x \in [A,B]$.

Theorem 2 *Under the conditions (G1)-(G4), there exists a constant*

$$0 \le D \le \sqrt{2G_{max} \cdot (\Gamma_{max} + G_{max})}$$

so that

$$P\left\{ \limsup_{n\to\infty} \frac{\sup_{x \in [A,B]} |G(x) - G_n(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1$$

The proof of Theorem 2 is similar to that of Theorem 1 and is presented in Section A.4 in more detail.

5 Applications

Despite the simplicity of the general estimators (2) and (3), these can be applied on some known stochastic processes. The advantage here is the known rate of convergence. In this paper we show the rate of convergence for the well-known empirical distribution function, the kernel density estimator and the hazard rate estimator for left-truncated and right-censored data.

5.1 Convergence of empirical time-dependent probability functions

In this section we present the general approach to estimate time-dependent probability functions. The results of this section will be used in the further sections to simplify the proofs.

Let $(\mathbf{S}_i)_{i=1,\dots,n}$ be the series of i.i.d. vectors $\mathbf{S} : \Omega \to \mathbb{R}^d$ from Section 3, where $S_i^d \in \{s_1,\dots,s_k \in \mathbb{R}\}$ for $i = 1,\dots,n$ is assumed. We define also a random set H_i like in Section 4.

We assume $G(x) := P(x \in H_i, S_i^d = S_C) \leq 1 =: G_{max}$ to be a nonnegative function on $[A,B]$ and $G_n(x) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{x\in H_i, S_i^d = S_C\}}$ to be an estimate for $G(x)$, where $S_C \in \{s_1,\dots,s_k \in \mathbb{R}\}$ is a constant. Then a constant $D \leq 2$ exists, so that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{x\in[A,B]} |G(x) - G_n(x)|}{\sqrt{\log(n)/n}} = D\right\} = 1. \tag{4}$$

To prove this convergence we define the random variables $\Gamma_i^x = \mathbf{1}_{\{S_i^d = S_C\}} \leq 1 =: \Gamma_{max}$, $i = 1,\dots,n$ for each fixed $x \in \mathbb{R}$. We get then the following expectation for each fixed $x \in [A,B]$:

$$\begin{aligned} &E(\mathbf{1}_{\{x\in H_i\}} \cdot \Gamma_i^x) = E(\mathbf{1}_{\{x\in H_i\}} \cdot \mathbf{1}_{\{S_i^d = S_C\}}) \\ &= E(\mathbf{1}_{\{x\in H_i, S_i^d = S_C\}}) = P(x \in H_i, S_i^d = S_C) = G(x) \end{aligned}$$

The assumptions (G1)-(G4) are fulfilled and the convergence follows from Theorem 2.

5.2 Convergence of empirical probability distribution and kernel density estimator

Let $S_1,\dots,S_n$ be univariate i.i.d. random variables with a strictly positive and Lipschitz continuous distribution $F(x)$ and density $f(x)$ functions on $[A,B]$. First, we examine the rate of convergence of the empirical distribution function $F_n(x) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{S_i \leq x\}}$.

Theorem 3 *Let $(S_i \in \mathbb{R})_{i=1,\dots,n}$ be a series of i.i.d. random variables with a distribution function $F(x) := P(S_i \leq x)$ on $[A,B]$.*

Then a constant $D \leq 2$ exists, so that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{x\in[A,B]} |F(x) - F_n(x)|}{\sqrt{\log(n)/n}} = D\right\} = 1$$

Proof We define i.i.d. random vectors $(S_i \in \mathbb{R}, S_C)_{i=1,\dots,n}$, where $S_C < \infty$ is a constant. We define also the mapping $H(x_1,x_2) := \{y : x_1 \le y\}$ and random sets $H_i = \{y \in \mathbb{R} : S_i \le y\}$, $i = 1,\dots,n$. Similar to Section 5.1 we define a time-dependent probability function $G(x) := P(x \in H_i, S_C = S_C)$ and its estimate $G_n(x) := \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{\{x \in H_j, S_C = S_C\}}$.

The convergence of $F_n(x)$ results from equalities $F(x) = G(x)$ and $F_n(x) = G_n(x)$ and the convergence formulation (4). □

From the general kernel estimator (1) we obtain now the following density kernel estimator:

$$f_n(x) := \sum_{i=1}^{n} \frac{1}{n \cdot R_n(S_i)} K\left(\frac{x - S_i}{R_n(S_i)}\right) \mathbf{1}_{\{S_i \le x\}}, \tag{5}$$

where the "smoother" is the same distribution function $F(x)$.

The convergence of the estimate (5) needs additionally to assume the Lipschitz continuity of $F(\cdot)$ and $f(\cdot)$ on $[A,B]$. We define now the measures $F(J) := \int_J f(s)ds = F(b) - F(a)$ and $F_n(J) := F_n(b) - F_n(a)$ for an interval $J := [a,b] \subseteq [A,B]$. As we know, the local convergence of $F_n(x)$, i.e. the convergence of the empirical measure $F_n(J)$, must be proven to show the convergence of $f_n(x)$.

Theorem 4 *Let $(S_i \in \mathbb{R})_{i=1,\dots,n}$ be a series of i.i.d. random variables with Lipschitz continuous and strictly monotonic increasing on $[A,B]$ distribution function $F(x) := P(S_i \le x)$ and the series (p_n) fulfills the assumption (L9).*

Then a constant $D \le 2\sqrt{6}$ exists, so that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{J \subseteq [A,B], F(J) \le p_n} |F(J) - F_n(J)|}{\sqrt{\log(n) p_n / n}} = D\right\} = 1 \tag{6}$$

Proof We use the following assumptions to prove Theorem 4:

- We assume Δ_i^x, $i = 1,\dots,n$ to be a constant $\Delta_i^x = \Delta_{max} := 1$ for each fixed $x \in [A,B]$.
- We define $\Psi(x) := P(S_i \le x) = F(x)$.
- The functions $G(x)$ and $G_n(x)$ will be also assumed to be constant: $G(x) = M := 1$ and $G_n(x) := 1$.
- Moreover, we define a constant $D_G := 0$.

Our goal now is to show the assumptions (L1)-(L9).

From Section A.1 follows $|G(x) - G_n(x)| = 0 \le C_G\sqrt{\log(n)/n}$ for all $n > 0$, $x \in [A,B]$ and $C_G > D_G$, so that the assumptions (L5) and (L7) are fulfilled.

With constant Δ_i^x we get the assumption (L4) for $a,b \in [A,B]$, $x \in \mathbb{R}$ as follows:

$$[\mathbf{1}_{\{x \le a\}}\Delta_i^a - \mathbf{1}_{\{x \le b\}}\Delta_i^b][G(x)]^{-1} = \mathbf{1}_{\{a \le x \le b\}}\Delta_{max}M < 2\Delta_{max}M,$$

where $M = [G(x)]^{-1} = 1$.

The assumption (L6) can be obtained for each fixed $x \in [A,B]$ trivially as follows:

$$E\left(\frac{\mathbf{1}_{\{S_i \le x\}} \cdot \Delta_i^x}{G(S_i)}\right) = E\left(\mathbf{1}_{\{S_i \le x\}}\right) = \Psi(x).$$

The assumptions (L1)-(L9) are fulfilled and the convergence (6) follows from Theorem 1. □

As we know from Section 2, the rate of $\mathcal{O}([\log(n)/(np_n)]^{1/2} + p_n)$ on $[A,B]$ results for the density kernel estimator (5) from Theorem 4.

5.3 Convergence of hazard rate estimate from left-truncated and right-censored data

As a more complicated case we examine now a survival process with left-truncated and right-censored observations. We assume $(T_i \in \mathbb{R}_0^+)_{i\in\mathbb{N}}$, $(C_i \in \mathbb{R}_0^+)_{i\in\mathbb{N}}$ and $(L_i \in \mathbb{R})_{i\in\mathbb{N}}$ to be independent series of i.i.d. random variables with Lipschitz continuous distributions $F^T(x)$, $F^C(x)$ and $F^L(x)$ respectively, where T_i are the lifetimes and pairs $[L_i, C_i]$ represent bounds of observations under condition $L_i < C_i$ for $i \in \mathbb{N}$. Because of the variable left bound of observation L_i, only objects with $L_i < T_i$ can be observed.

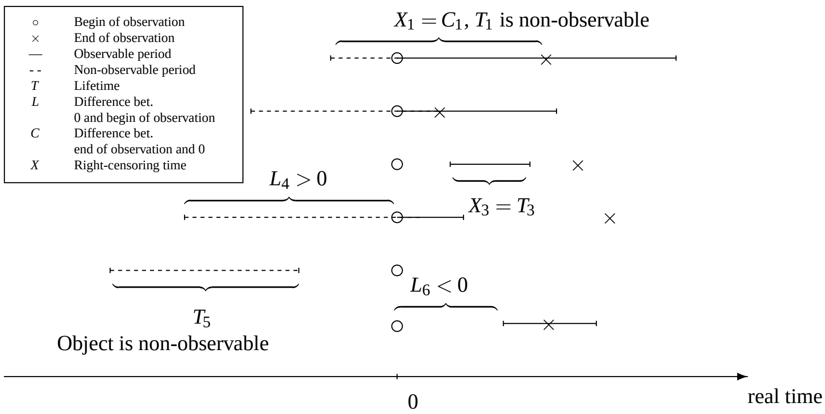

Fig. 1 Survival process and bounds of observation in the real time

As we can see in Fig. 1, the right bound of observation C_i can cut the lifetime on the right side, so that T_i is not always to observe. Hence, we define the observable censoring random variable $X_i = \min(T_i, C_i)$ and censoring index $\delta_i = \mathbf{1}_{\{T_i<C_i\}}$. Thus, one can observe only vectors (L_i, X_i, δ_i) with $L_i < X_i$ in a survival process with left-truncated and right-censored data.

A survival process can be described with a bent for a change of state (death or insolvency) or in the hazard rate [Andersen et al(1993)Andersen, Borgan, Gill, and Keiding] as follows:

$$\lambda(t)dt = P(t \le T < t + dt \mid t \le T) = \frac{dF^T(t)}{1 - F^T(t)}.$$

Another way to describe a survival process is to use the cumulative hazard rate:

$$\Lambda(t) := \int_0^t \lambda(s)ds = \int_0^t \frac{dF^T(s)}{1 - F^T(s)} = -\log[1 - F^T(t)]$$

The advantage of the cumulative hazard rate is expressed in the possibility of its estimation by the left-truncated and right-censored observations [Cao et al(2005)Cao, Janssen, and Veraverbeke] as follows:

$$\Lambda_n(t) := \sum_{i=1}^{n} \frac{\mathbf{1}_{\{X_i \le t, \delta_i = 1 | L_i \le X_i\}}}{nG_n(X_i)} = \sum_{i:X_{(i)} \le t} \frac{\delta_i}{\sharp\{j : L_j \le X_{(i)} \le X_j\}}, \tag{7}$$

where $G_n(t) = n^{-1}\sum_{i=1}^{n} \mathbf{1}_{\{L_i \le t \le X_i | L_i \le X_i\}}$ is the consistent estimate of the probability function $G(x) = P(L_i \le t \le X_i | L_i \le X_i)$.

Theorem 5 *Let $(L_i, X_i)_{i=1,\dots,n}$ with $L_i \le X_i$ be i.i.d. random vectors of observations.*

Then a constant $D \le 2$ exists, so that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{t\in[A,B]} |G(t) - \hat{G}_n(t)|}{\sqrt{\log(n)/n}} = D\right\} = 1. \tag{8}$$

Proof The trick of the proof is to define the random sets $H_i := \{y \in \mathbb{R} : L_i \le y \le X_i \mid L_i \le X_i\}$, constants Z_i and a constant Z_C, so that $Z_i = Z_C$ for $i = 1, \dots, n$. We can express then the functions $G(t)$ and $G_n(t)$ as follows:

$$\begin{aligned} G(t) &= P(L_i \le t \le X_i | L_i \le X_i) \\ &= P(L_i \le t \le X_i, Z_i = Z_C | L_i \le X_i) = P(t \in H_i, Z_i = Z_C) \end{aligned}$$

and

$$\begin{aligned} G_n(t) &= n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{L_i \le t \le X_i | L_i \le X_i\}} \\ &= n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{L_i \le t \le X_i, Z_i = Z_C | L_i \le X_i\}} = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{t \in H_i, Z_i = Z_C\}}. \end{aligned}$$

The convergence (8) follows from Section 5.1. □

Unfortunately, some calculations need also an estimate of the hazard rate. Hence, the goal is here to estimate the hazard function from the equation (7). As a possible solution we apply the kernel estimator (1) with "smoother" $\Lambda(t)$ for the hazard rate as follows :

$$\lambda_n(t) := \sum_{i: X_{(i)} \le t} \frac{1}{R_n(X_{(i)})} K\left(\frac{t - X_{(i)}}{R_n(X_{(i)})}\right) \frac{\delta_i}{\sharp\{j : L_j \le X_{(i)} \le X_j\}} \tag{9}$$

We define now an interval $[A, B]$ so that $F^L(A) > 0$ and $F^X(B) < 1$ and assume $\Lambda(t)$ and $\lambda(t)$ to be strictly positive and Lipschitz continuous on $[A, B]$. The convergence of $\lambda_n(t)$ can be easily shown, if the theoretical measure $\Lambda_n(J) := \Lambda_n(b) - \Lambda_n(a)$ converges to $\Lambda(J) := \int_J \lambda(t)dt = \Lambda(b) - \Lambda(a)$ with local rate of $\mathcal{O}(\sqrt{\log(n)p_n/n})$, where $J := [a, b] \in [A, B]$.

Theorem 6 *Let $(L_i, X_i, \delta_i)_{i=1,\dots,n}$ with $L_i \le X_i$ be i.i.d. random vectors of observations and the series (p_n) fulfills the assumption (L9).*

Then a constant $D \le 2(\sqrt{2 \cdot (M + \Lambda(B))} + 2M)$ exists, so that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{J\subseteq[A,B], \Lambda(J)\le p_n} |\Lambda(J) - \Lambda_n(J)|}{\sqrt{\log(n)p_n/n}} = D\right\} = 1 \tag{10}$$

where $M := \sup_{t\in[A,B]}[G(t)]^{-1}$.

Proof First, we check the assumptions (L1)-(L9) for the local convergence. The random variable Δ_i^x can be defined as a constant $\Delta_i^x = \mathbf{1}_{\{\delta_i=1|L_i\le X_i\}} \le 1 =: \Delta_{max}$. Additionally, we define $\Psi(x) := \Lambda(x)$ and $D_G := 2$.

From Theorem 5 and Section A.1 we know that $|G(x) - \hat{G}_n(x)| \le C\sqrt{\log(n)/n}$ a.s. for large n, $x \in \mathbb{R}$ and $C > D_G$.

We examine now the function $G(x)$ and write it as

$$\begin{aligned} G(x) &= \alpha^{-1}F^L(x)(1-F^T(x))(1-F^C(x)) \\ &= \int_{-\infty}^{x} \alpha^{-1}(1-F^T(x))(1-F^C(x))f^L(s)ds, \end{aligned}$$

where $\alpha = P(L_i \le X_i)$.

Since the distribution functions $F^T(x)$, $F^C(x)$ and $F^L(x)$ are Lipschitz continuous the functions $G(x)$ and $\Lambda(x)$ are also Lipschitz continuous for all $x \in [A,B]$. The assumptions (L4) and (L5) are fulfilled because of the constant Δ_i^x.

Now we prove the assumption (L6) as follows:

$$\begin{aligned} &E\left(\frac{\mathbf{1}_{\{X_i\le x|L_i\le X_i\}}\cdot\Delta_i^x}{G(X_i)}\right) = E\left(\frac{\mathbf{1}_{\{X_i\le x,\delta_i=1|L_i\le X_i\}}}{G(X_i)}\right) \\ =&\sum_{\delta_1=0}^{1}\int_{-\infty}^{\infty}\frac{\mathbf{1}_{\{x_1\le x,\delta_1=1\}}}{G(x_1)}dF^{X,\delta}(x_1,\delta_1) = \int_{-\infty}^{x}\frac{dF^{X,\delta}(x_1,1)}{G(x_1)}, \end{aligned} \tag{11}$$

where $F^{X,\delta}(x,y)$ is the distribution function of random vectors (X_i,δ_i) under condition $L_i \le X_i$, $i = 1,\ldots,n$.

For the intervals $J \subseteq [A,B]$ we can get the derivative $dF^{X,\delta}(x_1,1)$ from the probability $P(X_i \in J, \delta_i = 1 \mid L_i \le X_i)$ as follows:

$$\begin{aligned} &P(X_i\in J,\delta_i=1\mid L_i\le X_i) = \alpha^{-1}P(X_i\in J,\delta_i=1,L_i\le X_i) \\ =&\alpha^{-1}[P(T_i\in J,T_i\le C_i,L_i\le T_i,T_i\le C_i) \\ &+P(C_i\in J,T_i\le C_i,L_i\le C_i,C_i<T_i)] = \alpha^{-1}P(T_i\in J,L_i\le T_i\le C_i). \end{aligned} \tag{12}$$

The probabilities $P(X_i \in J, \delta_i = 1 \mid L_i \le X_i)$ and $P(T_i \in J, L_i \le T_i \le C_i)$ can be written as

$$\begin{aligned} &P(X_i\in J,\delta_i=1\mid L_i\le X_i) = E(\mathbf{1}_{\{X_i\in J,\delta_i=1|L_i\le X_i\}}) \\ =&\sum_{\delta_1=0}^{1}\int_{-\infty}^{\infty}\mathbf{1}_{\{x_1\in J,\delta_1=1\}}dF^{X,\delta}(x_1,\delta_1) = \int_{x_1\in J}dF^{X,\delta}(x_1,1) \end{aligned} \tag{13}$$

and

$$\begin{aligned} &\alpha^{-1}P(T_i\in J,L_i\le T_i\le C_i) = \alpha^{-1}E(\mathbf{1}_{\{T_i\in J,L_i\le T_i\le C_i\}}) \\ =&\int_{t\in\mathbb{R}}\int_{c\in\mathbb{R}}\int_{l\in\mathbb{R}}\alpha^{-1}\mathbf{1}_{\{t\in J\}}\mathbf{1}_{\{l\le t\}}\mathbf{1}_{\{t\le c\}}dF^T(t)dF^C(c)dF^L(l) \\ =&\int_{t\in J}\alpha^{-1}F^L(t)(1-F^C(t))dF^T(t). \end{aligned} \tag{14}$$

The derivative $dF^{X,\delta}(x_1,1)$ follows then for the one-element intervals $J=[j,j]$, $j\in[A,B]$ from the equations (12), (13) and (14):

$$\begin{aligned}&\int_{x_1\in J} dF^{X,\delta}(x_1,1)=\int_{t\in J}\alpha^{-1}F^L(t)(1-F^C(t))dF^T(t)\\ =&dF^{X,\delta}(j,1)=\alpha^{-1}F^L(j)(1-F^C(j))dF^T(j).\end{aligned} \tag{15}$$

We substitute (15) in (11) and get the assumption (L6):

$$\begin{aligned}&E\left(\frac{\mathbf{1}_{\{X_i\le x|L_i\le X_i\}}\cdot\Delta_i^x}{G(X_i)}\right)=\int_{-\infty}^{x}\frac{dF^{X,\delta}(x_1,1)}{G(x_1)}\\ =&\int_{-\infty}^{x}\frac{\alpha^{-1}F^L(x_1)(1-F^C(x_1))dF^T(x_1)}{G(x_1)}\\ =&\int_{-\infty}^{x}\frac{\alpha^{-1}F^L(x_1)(1-F^C(x_1))dF^T(x_1)}{\alpha^{-1}F^L(x_1)(1-F^C(x_1))(1-F^T(x_1))}=\int_{-\infty}^{x}\frac{dF^T(x_1)}{1-F^T(x_1)}=\Lambda(x)=\Psi(x)\end{aligned}$$

The assumptions (L1)-(L9) are fulfilled and the convergence (10) follows from Theorem 1. □

One can see that the observations $L_i\le 0$, $i=1,\dots,n$ are present in the right-censored case. The Nelson-Aalen estimator is then the special case of the presented here estimate (7).

As known from Section 2, the rate of convergence of the estimate (9) follows from Theorem 6 and is $\mathscr{O}([\log(n)/(np_n)]^{1/2}+p_n)$ on $[A,B]$.

Acknowledgements The work of the authors was supported by the Deutsche Forschungsgemeinschaft (SFB 475, 'Reduction of complexity in multivariate data structures').

References

[Aalen(1978)] Aalen O (1978) Nonparametric estimation of partial transition probabilities in multiple decrement models. Annals of Statistics 6:534–545

[Andersen et al(1993)Andersen, Borgan, Gill, and Keiding] Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) Statistical Models Based on Counting Processes. Springer

[Cao et al(2005)Cao, Janssen, and Veraverbeke] Cao R, Janssen P, Veraverbeke N (2005) Relative hazard rate estimation for right censored and left truncated data. Sociedad de Estadística e Investigación Operativa Test 14(1):257–280

[Földes and Rejtö(1981)] Földes A, Rejtö L (1981) Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. Annals of Statistics 9:122–129

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

[Hoeffding(1963)] Hoeffding W (1963) Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association 58:13ff

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

[Schäfer(1986)] 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

[Shorack and Wellner(1986)] Shorack G, Wellner J (1986) Empirical processes with applications to statistics. John Wiley & Sons, New York

[Silverman(1978)] 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(1982a)] Stute W (1982a) The law of the logarithm for kernel density estimators. Annals of Probability 10:414–422

[Stute(1982b)] Stute W (1982b) The oscillation behaviour of empirical processes. Annals of Probability 10:86–107
[Walter and Bulum(1979)] Walter G, Bulum J (1979) Probability density estimation using delta sequences. Annals of Statistics 3:30–48
[Weißbach(2006)] Weißbach R (2006) A general kernel functional estimator with general bandwidth - strong consistensy and applications. Nonparametric Statistics 18(1):1–12

A Proofs

A.1 Limes superior formulation of the convergence

As a result of Hewitt and Savage [Hewitt and Savage(1955)], it is equivalent to state there exists a constant $D \leq 1$ s.t.

$$P\left\{\limsup_{n\to\infty} \frac{S_n(\mathbf{Y}_1(\omega),\ldots,\mathbf{Y}_n(\omega))}{a_n}\right\} = 1.$$

As for all $a > 1$, it holds that

$$P\{\omega \mid \exists N \in \mathbb{N} \forall n > N : S_n(\mathbf{Y}_1(\omega),\ldots,\mathbf{Y}_n(\omega)) \leq \alpha a_n\} = 1.$$

With positive zero-sequence a_n, a sequence of i.i.d. random d-vectors $\mathbf{Y}_i : \Omega \to (\mathbb{R})^d$ and a sequence of measurable symmetric mappings $S_n : \mathbb{R}^n \to \mathbb{R}_0^+$.

A.2 Distribution of the mean value

We assume $T_1,\ldots,T_n$ to be independent bounded random variables with expectation 0 and dispersion σ^2, so that

$$E(T_i) = 0$$

$$|T_i| \leq b$$

$$\sigma^2 := Var(T_i)$$

for $i = 1,...,n$.

From the two sided version of Theorem 3 from [Hoeffding(1963)] and the inequality $\log(1+x) \geq 2x/(2+x)$, $x \geq 0$ results the following inequality for each $\varepsilon > 0$:

$$P(|\overline{T}| \geq \varepsilon) \leq 2\exp\left(-\frac{n\varepsilon^2}{2\sigma^2 + 2b\varepsilon/3}\right).$$

A.3 Local convergence

The proof of Theorem 1 follows in four steps.

First, we show the exponential bounds for the following distribution of the difference $|\Psi_n^*(J) - \Psi(J)|$:

$$P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon) < 2\exp\left(\frac{-n\varepsilon^2}{2(2\Delta_{max}M + \Psi(B))(p+\varepsilon)}\right) \tag{16}$$

for all $p > 0$, $\varepsilon > 0$, $n \in \mathbb{N}_{>0}$ and each fixed $J \subseteq [A,B]$ with $\Psi(J) \leq p$.

Because of the definition (2) and the assumption (L3) is

$$\Psi_n^*(J) - \Psi(J) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{1}_{\{S_i^1 \le b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \le a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J)\right) \tag{17}$$

the arithmetical mean of the n independent, bounded and like R_J distributed random variables for each fixed $J \subseteq [A,B]$, where

$$R_J := \frac{\mathbf{1}_{\{S_i^1 \le b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \le a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J).$$

The expectation, the dispersion and the bounds of R_J can be then calculated for fixed $J \subseteq [A,B]$ with $\Psi(J) \le p$.

The expectation of R_J follows from the assumption (L6):

$$E(R_J) = E\left(\frac{\mathbf{1}_{\{S_i^1 \le b\}} \cdot \Delta_i^b}{G(S_i^1)}\right) - E\left(\frac{\mathbf{1}_{\{S_i^1 \le a\}} \cdot \Delta_i^a}{G(S_i^1)}\right) - \Psi(b) + \Psi(a) = 0. \tag{18}$$

From assumption (L4), we get the following bounds of $|R_J|$ on $[A,B]$:

$$\begin{aligned} |R_J| &= \left|\frac{\mathbf{1}_{\{S_i^1 \le b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \le a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J)\right| \\ &< 2\Delta_{max}M + \Psi(B) - \Psi(A) < 2\Delta_{max}M + \Psi(B) =: g. \end{aligned} \tag{19}$$

The dispersion of R_J can be obtained from the expectation (18) and bounds (19) as follows:

$$\begin{aligned} \sigma_J^2 :=& Var(R_J) = E\left[\left(\frac{\mathbf{1}_{\{S_i^1 \le b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \le a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J)\right)^2\right] \\ &< 2\Delta_{max}M \cdot E\left(\frac{\mathbf{1}_{\{S_i^1 \le b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \le a\}} \cdot \Delta_i^a}{G(S_i^1)}\right) \\ &= 2\Delta_{max}M \cdot \Psi(J) < g \cdot p. \end{aligned} \tag{20}$$

From the equations (17), (18), (19), (20) and the inequality from Section A.2 result the following right bounds:

$$P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon) < 2\exp\left(\frac{-n\varepsilon^2}{2(\sigma_J^2 + g\varepsilon/3)}\right) < 2\exp\left(\frac{-n\varepsilon^2}{2g(p+\varepsilon)}\right)$$

for each fixed interval $J \subseteq [A,B]$ with $\Psi(J) \le p$.

On the second step we show the inequality

$$\sup_{J \subseteq [A,B], \Psi(J) \le p_n} |\Psi_n^*(J) - \Psi(J)| \le C\sqrt{\log(n)p_n/n} \tag{21}$$

almost sure for a constant $C > \sqrt{2(2\Delta_{max}M + \Psi(B))}$ and large n.

On the right side of the inequality (16) p and ε could be substituted with p_n and $\varepsilon_n := C\sqrt{\log(n)p_n/n}$, $C > 0$, $n > 1$ respectively as follows:

$$< 2 \cdot \exp\left(-\log(n)\frac{C^2}{2g}\frac{p_n}{(p_n + \varepsilon_n)}\right) = 2n^{-\frac{C^2}{2g}\frac{p_n}{(p_n+\varepsilon_n)}} =: A_n.$$

The series (A_n) is then summable starting from some large $n < \infty$ only for

$$\beta_n := \frac{C^2}{2g}\frac{p_n}{(p_n + \varepsilon_n)} > 1.$$

From $\varepsilon_n/p_n = C\sqrt{\log(n)/(np_n)}$ and the assumption (L9) follow $\varepsilon_n/p_n \to 0$ and $p_n/(p_n+\varepsilon_n) \to 1$ for large n. The condition $\beta_n > 1$ can be then achieved with $C^2/2g > 1$ or $C > \sqrt{2g}$.

Consequently, the series (A_n) is summable from some large $n < \infty$ and only for $C > \sqrt{2g}$. For each $J \subseteq [A,B]$ with $\Psi(J) \leq p_n$ we get then

$$\exists C > \sqrt{2g}\ \exists m < \infty, m \in \mathbb{N} : \sum_{n=m}^{\infty} P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon_n) < \sum_{n=m}^{\infty} A_n < \infty$$

and

$$\forall m < \infty, m \in \mathbb{N} : \sum_{n=1}^{m} P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon_n) \leq m < \infty$$

Because of the summability of $P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon_n)$, the following probability results from Borel-Cantelli lemma for $C > \sqrt{2g}$:

$$P\left(\limsup_{n\to\infty} |\Psi_n^*(J) - \Psi(J)| > \varepsilon_n\right) = 0,$$

i.e $|\Psi_n^*(J) - \Psi(J))|$ doesn't exceed ε_n for the most n. For large n and for all $J \subseteq [A,B]$ with $\Psi(J) \leq p_n$, we get almost surely

$$|\Psi_n^*(J) - \Psi(J))| \leq C\sqrt{\log(n) p_n/n}$$

The same inequality is valid for the supremum of $|\Psi_n^*(J) - \Psi(J)|$ on $[A,B]$:

$$\sup_{J\subseteq[A,B], \Psi(J)\leq p_n} |\Psi_n^*(J) - \Psi(J)| \leq C\sqrt{\log(n) p_n/n}$$

for $C > \sqrt{2g}$ and large n almost surely.

On the third step we prove the following inequality on the basis of the previous results:

$$\sup_{J\subseteq[A,B], \Psi(J)\leq p_n} |\Psi_n(J) - \Psi_n^*(J)| \leq C \cdot p_n \sqrt{\log(n)/n}$$

almost sure for some $C > D_G \cdot M$ and large n.

From the assumption (L7) and the limes superior formulation from Section A.1 we get the following right bounds for $G(x) - G_n(x)$:

$$G(x) - G_n(x) \leq |G(x) - G_n(x)| \leq \sup_{x\in[A,B]} |G(x) - G_n(x)| \leq C_1' \cdot \sqrt{\log(n)/n}$$

almost sure for $C_1' > D_G$, large n and all $x \in [A,B]$. These bounds can be rewritten for $G_n(x)$ as follows:

$$G_n(x) \geq G(x) - C_1'\sqrt{\log(n)/n} \geq \inf_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n}.$$

From the assumption (L7) follows $\inf_{t\in[A,B]} G(t) > 0$. Because of $\sqrt{\log(n)/n} \to 0$, the following inequations are fulfilled for $x \in [A,B]$ and large n:

$$\inf_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n} > 0,$$

$$\frac{1}{G_n(x)} \leq \frac{1}{\inf_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n}}$$

and

$$\frac{|G(x) - G_n(x)|}{G_n(x)} \leq \frac{C_1' \cdot \sqrt{\log(n)/n}}{\inf\limits_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n}}.$$

The following bounds for $\Psi_n^*(J) - \Psi(J)$ and $\Psi_n^*(J)$ result from the equation (21) almost sure for $J \subseteq [A,B]$ with $\Psi(J) \leq p_n$, large n and $C_2' > \sqrt{2\cdot(2\Delta_{max}M + \Psi(B))}$:

$$\begin{aligned}\Psi_n^*(J) - \Psi(J) &\leq |\Psi_n^*(J) - \Psi(J)| \\ &\leq \sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n^*(J) - \Psi(J)| \leq C_2'\sqrt{\log(n)p_n/n}\end{aligned}$$

and consequently

$$\Psi_n^*(J) \leq \Psi(J) + C_2'\sqrt{\log(n)p_n/n} \leq p_n + C_2'\sqrt{\log(n)p_n/n}.$$

We obtain then the following equation from the assumption (L5) almost sure for each $J \subseteq [A,B]$ with $\Psi(J) \leq p_n$ and large n:

$$\begin{aligned}|\Psi_n(J) - \Psi_n^*(J)| &= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{G_n(S_i^1)} - \frac{1}{G(S_i^1)}\right|(\mathbf{1}_{\{S_i^1\leq b\}}\cdot\Delta_i^b - \mathbf{1}_{\{S_i^1\leq a\}}\cdot\Delta_i^a) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{G(S_i^1) - G_n(S_i^1)}{G_n(S_i^1)}\right|\frac{\mathbf{1}_{\{S_i^1\leq b\}}\cdot\Delta_i^b - \mathbf{1}_{\{S_i^1\leq a\}}\cdot\Delta_i^a}{G(S_i^1)} \\ &\leq \frac{C_1'\sqrt{\log(n)/n}\cdot\Psi_n^*(J)}{\inf\limits_{t\in[A,B]} G(t) - C_1'\sqrt{\log(n)/n}} \leq \frac{C_1'\sqrt{\log(n)/n}\cdot\left(p_n + C_2'\sqrt{\log(n)p_n/n}\right)}{\inf\limits_{t\in[A,B]} G(t) - C_1'\sqrt{\log(n)/n}}.\end{aligned}$$

The transformation $p_n + C_2'\sqrt{\log(n)p_n/n} = p_n[1 + C_2'\sqrt{\log(n)/(p_n n)}]$ is evident, where the term $C_2'\sqrt{\log(n)/(p_n n)}$ can be neglected for large n because of the assumption (L9). For large n, we can also neglect the term $\sqrt{\log(n)/n}$ in numerator. For all $J \subseteq [A,B]$ with $\Psi(J) \leq p_n$ and for large n, we get then the following inequality almost sure:

$$|\Psi_n(J) - \Psi_n^*(J)| \leq \frac{C_1'}{\inf\limits_{t\in[A,B]} G(t)} p_n\sqrt{\log(n)/n} = C_1' \cdot M \cdot p_n\sqrt{\log(n)/n}$$

The following right bound results for some $C > D_G \cdot M$ and large n almost sure:

$$\sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n(J) - \Psi_n^*(J)| \leq C \cdot p_n\sqrt{\log(n)/n}.$$

On the last step we examine the expression $\sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n(J) - \Psi(J)|$. This overall difference can be represented by the sum of the deviations of the theoretical and empirical measures $\Psi(J)$ and $\Psi_n(J)$ from the help measure $\Psi_n^*(J)$ as follows:

$$\begin{aligned}&\sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n(J) - \Psi(J)| \\ &\leq \sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n(J) - \Psi_n^*(J)| + \sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n^*(J) - \Psi(J)|\end{aligned}$$

Because of the assumption (L9), $\frac{p_n\sqrt{\log(n)/n}}{\sqrt{\log(n)p_n/n}} = \sqrt{p_n}$ approaches zero, i.e. $p_n\sqrt{\log(n)/n} \leq \sqrt{\log(n)p_n/n}$ is valid for large n.

From the previously mentioned right bounds of $|\Psi_n(J) - \Psi_n^*(J)|$ and $|\Psi_n^*(J) - \Psi(J)|$ follows the existence of the constant

$$C > \sqrt{2\cdot(2\Delta_{max}M + \Psi(B))} + D_G \cdot M,$$

so that

$$\sup_{J\subseteq[A,B],\Psi(J)\leq p_n} |\Psi_n(J)-\Psi(J)| \leq C(\sqrt{\log(n)p_n/n}+p_n\sqrt{\log(n)/n}) \leq 2C\sqrt{\log(n)p_n/n}$$

is valid almost surely for large n. Because of the symmetry of $\Psi_n(J)$ the limes superior formulation of the convergence follows from Section A.1. □

A.4 Global convergence

The proof of Theorem 3 follows in two steps.

First, we take a look at the difference $|G_n(x)-G(x)|$ and show exponential bounds of its distribution:

$$P(|G_n(x)-G(x)|>\varepsilon) < 2\exp\left(\frac{-n\varepsilon^2}{2(G_{max}+\Gamma_{max})(G_{max}+\varepsilon)}\right) \tag{22}$$

for $\varepsilon \in \mathbb{R}_{>0}$, $n \in \mathbb{N}_{>0}$ and each fixed $x \in [A,B]$.

From the definition of the estimate (3) and the assumption (G2) follows the construction of the aforesaid deviation:

$$G_n(x)-G(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x - G(x)\right) \tag{23}$$

as the arithmetical mean of n independent, bounded and like R_x distributed random variables for each fixed $x \in [A,B]$, where $R_x := \mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x - G(x)$.

The expectation of R_x follows for fixed $x \in [A,B]$ from the assumption (G4):

$$\begin{aligned} E(R_x) &= E\left(\mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x - G(x)\right) \\ &= E\left(\mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x\right) - G(x) = 0 \end{aligned} \tag{24}$$

Further, the bounds of $|R_x|$ on $[A,B]$ can be calculated as follows:

$$|R_x| = \left|\mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x - G(x)\right| < \Gamma_{max}+G_{max} =: g. \tag{25}$$

We get then the following bounds for the disperse of R_x on $[A,B]$:

$$\begin{aligned} \sigma_x^2 := Var(R_x) &= E\left[\left(\mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x - G(x)\right)^2\right] \\ &= E\left(\mathbf{1}_{\{x\in H_i\}}\cdot(\Gamma_i^x)^2\right) - (G(x))^2 \leq E\left(\mathbf{1}_{\{x\in H_i\}}\cdot(\Gamma_i^x)^2\right) \\ &\leq \Gamma_{max}\cdot E\left(\mathbf{1}_{\{x\in H_i\}}\cdot\Gamma_i^x\right) \\ &= \Gamma_{max}\cdot G(x) < g\cdot G_{max} \end{aligned} \tag{26}$$

From the equations (23), (24), (25), (26) and the inequality from Section A.2 result the following right bounds:

$$P(|G_n(x)-G(x)|>\varepsilon) < 2\exp\left(\frac{-n\varepsilon^2}{2(\sigma_x^2+g\varepsilon/3)}\right) < 2\exp\left(\frac{-n\varepsilon^2}{2g(G_{max}+\varepsilon)}\right)$$

for each fixed $x \in [A,B]$.

On the second step we transform the right side of the inequality (22) by substituting ε with $\varepsilon_n := C'\sqrt{G_{max}}\cdot\sqrt{\log(n)/n}$, $C'>0$, $n>1$ respectively as follows:

$$< 2\cdot\exp\left(-\log(n)\frac{(C')^2}{2g}\frac{G_{max}}{(G_{max}+\varepsilon_n)}\right) = 2n^{-\frac{(C')^2}{2g}\frac{G_{max}}{(G_{max}+\varepsilon_n)}} =: A_n,$$

where $g = \Gamma_{max} + G_{max}$.

The summability of the series (A_n) is evident starting from some large n only for

$$\beta_n := \frac{(C')^2}{2g} \frac{G_{max}}{(G_{max} + \varepsilon_n)} > 1$$

The condition $\beta_n > 1$ can be valid only for $(C')^2/2g > 1$ because of $G_{max}/(G_{max}+\varepsilon_n) < 1$ and $G_{max}/(G_{max} + \varepsilon_n) \to 1$.

Consequently, the series (A_n) is summable starting from some large $n < \infty$ and only for $C' > \sqrt{2g}$. Similar to Section A.3 the following bounds for $|G_n(x) - G(x)|$ and its supremum can be shown by the Borel-Cantelli lemma:

$$\sup_{x \in [A,B]} |G_n(x) - G(x)| \leq C\sqrt{\log(n)/n}$$

for $C > \sqrt{2gG_{max}}$ and large n almost surely.

Because of symmetry of $G_n(x)$, we obtain from Section A.1 the limes superior formulation of the convergence:

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{x\in[A,B]} |G_n(x) - G(x)|}{\sqrt{\log(n)/n}} = D\right\} = 1$$

where a constant $0 \leq D \leq \sqrt{2G_{max} \cdot (\Gamma_{max} + G_{max})}$ exists. □