Dept. of Math. University of Oslo
Statistical Research Report No. 6
ISSN 0806–3842 December 2006

On the errors committed by functionals of empirical processes

Steffen Grønneberg
Department of Mathematics, University of Oslo
P.O. Box 1053 Blindern, N-0316 Oslo, Norway
e-mail: steffeng@math.uio.no

Abstract: Consider a uniformly strongly consistent statistical functional. We present a new method for extracting limit distributions of the type "the last time" and "the number of times" an error larger than ϵ is committed for Hadamard differentiable estimators. We then apply our techniques to several estimators, including the Nelson-Aalen estimator and the quantiles of the Kaplan-Meier estimator.

Key words: The last n, the number of ϵ-misses, functional delta method, empirical process

1. Introduction

It is of great interest to study the differential calculus of statistical functionals based on empirical processes, as it can painlessly yield useful approximations of complex estimators using functional delta methods.

The only randomness of statistical functionals originates from the empirical process. This means we can figure out by how much and where our functional errs if we know this for the empirical process. We present a framework to extend known characteristics of the errors committed by empirical processes to Hadamard-differentiable statistical functionals and give several examples. Specifically, we investigate the asymptotic distribution of N_ε, the last time an ε-error is committed and Q_ε, the number of errors larger than ε.

These sizes are of interest in statistical estimation theory. Hjort & Fenstad (1992) showed that in most real valued estimators, the limit of $\varepsilon^2 N_\varepsilon$ depends only on the asymptotic variance, which thus gives new motivation to its use as a benchmark for the efficiency of an estimator. We will show that the Hadamard-differential plays an analogous rôle for a large class of statistical functionals. Further, once the limit distribution of N_ε is extracted, it can readily be applied to create sequential fixed-volume confidence regions in the style of Section 8F of Hjort & Fenstad (1992).

Hadamard-differentiability respects a chain-rule, so our results can often be extended in a tree structure. Our main example regards the hierarchy of

$$\text{Nelson-Aalen} \mapsto \text{Kaplan-Meier} \mapsto \text{Quantile estimator under censoring}$$

which are Hadamard-differential maps of each other.

We now turn to an explanation of the general heuristics used in both our and previously known results, then briefly summarize past developments. Further, section 2 presents our new results. Most proofs are in Appendix A. We continue with applications in Section 3 and end with Appendix B which states and proves a needed result for the multivariate empirical process.

Consider a variable D_n, typically some distance or precision measure associated with n observed data points. Assume $Pr\{D_n \to 0\} = 1$ and $Z_m(u) = \sqrt{m} D_{\lfloor mu \rfloor} \xrightarrow[m\to\infty]{\mathscr{L}} Z(u)$ in each Skorokhod space $\mathbb{D}[a, b]$ for each interval $0 < a < b$. Let $N_\varepsilon = \max\{n : D_n > \varepsilon\}$, which is $a.s.$ finite from the assumed $a.s.$ convergence. In many cases we have

$$Pr\left\{\varepsilon^2 N_\varepsilon \geq y\right\} \doteq Pr\left\{\max_{n\geq m} \sqrt{m} D_n \geq \sqrt{y}\right\} = Pr\left\{\max_{u\geq 1} Z_m(u) \geq \sqrt{y}\right\},$$

where $m = \lfloor y/\varepsilon^2 \rfloor$. This is close to showing $\varepsilon^2 N_\varepsilon \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \{\max_{u\geq 1} Z(u)\}^2$ - as in Hjort & Fenstad (1992). The same argument can be extended to other functionals than N_ε. Hjort & Fenstad (1992) also considered $Q_\varepsilon = \#\{n : D_n > \varepsilon\}$, using the same heuristics.

Asymptotics of this type was initialized by Stute (1983), which was limited to a certain class of real valued M-estimators, and greatly generalized in Hjort & Fenstad (1992). Barbe et al. (1998) looked at a certain class of mixing processes, and Atlagh et al. (2005) regarded real valued asymptotically linear estimators. However, the only previously known results for stochastic processes are found in Hjort & Fenstad (1992). They were also the only ones which so far has considered other functionals, such as their Q_ε. As for statistical applications, Barbe et al. (1998) applied the N_ε convergence to stopping times for the Gibbs-sampler and Hjort & Fenstad (1992) found confidence regions with shrinking boundaries and tests with power one.

We also mention that Hjort & Fenstad (1992) showed that their treatment of the empirical process can be extended to related estimators and gave the Crámer-von Mises statistic as an example. Although our results is a development of this, their method would not work without a detailed analysis of each individual statistical functional. As we only use the differential structure in our approximation results, our perspective allows us to give very general conditions for our results and is further simple to apply. Hjort & Fenstad (1992) also gave results for functional differentiable real-valued estimators, but did not regard any functional differentiable estimators that are stochastic processes.

2. Main results

We will be concerned with stochastic processes in the Skorokhod-space $\mathbb{D}$ of all cádlág functions (that is, right continuous functions with left hand limits) and will utilize two different metrizations. We will primarily use the uniform norm as the metric of choice for weak convergence, and secondarily use the Skorokhod topology as an auxiliary tool. For the uniform case, we will use the ball σ-field generated by all ε-balls, as used in Gill (1989), due to Dudley. We will also be using the Hadamard-differentiability and delta method of Gill (1989). The needed definitions and results follows the discussion given after Equation (2.1) As for the Skorokhod-space, we will utilize a multivariate generalization given in Bickel & Wichura (1971) which allows multidimensional time needed by estimators based on a multivariate empirical process. The multivariate Skorokhod-space will only be used in Appendix B, as weak convergence in the uniform metric and the multivariate Skorokhod metric is equivalent if the target variable is continuous, see Bickel & Wichura (1971, section 3).

Let $F_1, F_2, \dots$ be a sequence of stochastic processes (with respect to the ball σ-field) on $\mathbb{D}(T)$ which converges strongly to a deterministic F. Here $\mathbb{D}(T)$ is the space of all cádlág functions indexed by a given hypercube $T \subseteq \mathbb{R}^q$. We will allow vector valued functions and set $|F| := \max_{1\leq i\leq p} |F^i|$ when $F = (F^1, F^2, \dots, F^d) \in \mathbb{R}^d$. Suppose further that

$$K_m(s,t) = \sqrt{m}\left\{F_{\lfloor sm\rfloor}(t) - F(t)\right\}$$

has a continuous weak limit $K(s,t)$ (with respect to Dudley's weak convergence definition). For a given linear and continuous functional $\phi'_F(\alpha)$ on $\mathbb{D}$, write

$$\sqrt{n}[\phi(F_n) - \phi(F)] = \phi'_F(\sqrt{n}[F_n - F]) + R_n \tag{2.1}$$

where R_n is a remainder. If ϕ is Hadamard-differentiable and $\phi'_F(\alpha)$ is its differential, Gill (1989) gives $R_n \xrightarrow[n\to\infty]{\mathscr{P}} 0$ in his Theorem 3 - the functional delta method. For examples on how to compute Hadamard-differentials, see Gill (1989) and section 5.2 of Shao (2003).

The following is our basic approximation result, and will also present the setting for which our results are valid. We defer the proof to Appendix A.

Lemma 2.1. *Let $\phi : B_1 \mapsto B_2$ where B_1, B_2 are two linear subspaces of $\mathbb{D}(T)^d$ and $T \subseteq \mathbb{R}^q$ is a hypercube. Assume that $F_1, F_2, \dots$ are stochastic processes indexed by $\mathbb{D}(T)^d$ which converge uniformly to F a.s. and that $\|\phi(F_n) - \phi(F)\|_\infty$ converge almost surely to zero. Suppose ϕ is Hadamard-differentiable at each K_m and at the weak limit K, tangentially to a subspace of $\mathbb{D}(T)^d$ which includes F. Then $\sqrt{m}\sup_{n\geq m}\|\phi(F_n) - \phi(F)\|_\infty$ and $\sup_{n\geq m}\|\phi'_F(\sqrt{m}[F_n - F])\|_\infty$ have the same asymptotic distribution.*

Note that we do not impose the *i.i.d.* setting on our variables, yet we will not investigate any non-*i.i.d.* variables in our examples. The assumption $\|\phi(F_n) - \phi(F)\|_\infty \xrightarrow[n\to\infty]{\text{a.s.}} 0$ is motivated by the fact that we are interested in e.g. the last time $\|\phi(F_n) - \phi(F)\|_\infty > 0$, as described in the introduction. Instead of this assumption, we could also have required ϕ to be continuous (w.r.t. the uniform norm). This would transfer the *a.s.* convergence of the empirical process to our functional, but is stronger than our assumptions.

Theorem 2.2. *Assume the setting of Lemma 2.1 and that*

$$\sqrt{m}[F_{\lfloor ms \rfloor}(t) - F(t)] \xrightarrow[m\to\infty]{\mathscr{L}} K(s,t) \tag{2.2}$$

for some continuous K*, where* $s \in (0,1]$ *and* $t \in T$*. Assume that* ϕ'_F *is continuous at* $K(s,t)$*, where* ϕ'_F *treats* s *as a given constant (so it is still a functional of* $\mathbb{D}(T)^d$*). Assume further that for any* $\delta > 0$ *we have*

$$\lim_{c\to\infty} \limsup_{m\to\infty} Pr\left\{ \sup_{n\geq cm} \|\phi'_F(\sqrt{m}[F_n - F])\|_\infty \geq \delta \right\} = 0. \tag{2.3}$$

We then have that

$$\varepsilon^2 N_\varepsilon \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \left(\sup_{0\leq s\leq 1} \sup_t |\phi'_F\{K_F(s,\cdot)\}(t)| \right)^2,$$

where the notation is intended as a reminder as to that ϕ'_F *regards* s *as an index.*

Proof. Notice that $Pr\left\{\varepsilon^2 N_\varepsilon > y\right\} = Pr\left\{\sup_{n\geq m} \|\phi(F_n) - \phi(F)\|_\infty > \varepsilon\right\} = Pr\left\{\sqrt{m}\sup_{n\geq m} \|\phi(F_n) - \phi(F)\|_\infty > \sqrt{y_0}\right\}$ where $m := \lfloor y/\varepsilon^2 \rfloor$ and $y_0 = \varepsilon^2 \lfloor y/\varepsilon^2 \rfloor$ (which goes uniformly to y as $\varepsilon \to 0$). By Lemma 2.1 we only have to be concerned with the weak convergence of $\phi'_F(\sqrt{m}[F_n - F])$. Now notice that by invoking the continuous mapping theorem on both of the supremum and ϕ'_F in conjunction with the assumed tail-inequality given in Equation (2.3), we get

$$\begin{aligned} \sup_{s\geq 1}\sup_t |\phi'_F(\sqrt{m}[F_{\lfloor ms\rfloor}(\cdot) - F(\cdot)])(t)| &\xrightarrow[m\to\infty]{\mathscr{L}} \sup_{s\geq 1}\sup_t |\phi'_F\{K_F(s^{-1},\cdot)\}(t)| \\ &= \sup_{0\leq s\leq 1} \sup_t |\phi'_F\{K_F(s,\cdot)\}(t)|, \end{aligned} \tag{2.4}$$

by Theorem 4.2 of Billingsley (1968), which completes the proof. □

We will apply this when F_n is the empirical process. Thus K will be a Kiefer-process, as given in Appendix B.

Our method is simple to apply. This simplicity resides in the following result. The proof is given in Appendix A.

Lemma 2.3. *Assuming the setting of Lemma 2.1. If for any* $\delta > 0$*,*

$$\lim_{c\to\infty} \limsup_{m\to\infty} Pr\left\{ \sqrt{m} \sup_{n\geq cm} \|F_n - F\|_\infty \geq \delta \right\} = 0 \tag{2.5}$$

we also have that for any $\delta > 0$*,*

$$\lim_{c\to\infty} \limsup_{m\to\infty} Pr\left\{ \sup_{n\geq cm} \|\phi'_F(\sqrt{m}[F_n - F])\|_\infty \geq \delta \right\} = 0. \tag{2.6}$$

Our results and techniques are strong enough to tackle other properties of the errors as well. Define $Q_\varepsilon(a)$ by the number of times $\phi(F_n)$ is further than ε

away from $\phi(F)$ (w.r.t. the uniform norm) among $n \geq a/\varepsilon^2$. Hjort & Fenstad (1992) shows that we can write

$$Q_\varepsilon(a) = \int_{\lfloor a/\varepsilon^2 \rfloor}^{\infty} I\left\{\|\phi(F_{\lfloor s \rfloor}) - \phi(F)\|_\infty > \varepsilon\right\} ds.$$

The following Theorem shows that our machinery is still valid in this case, and is proved in Appendix A.

Theorem 2.4. *Assume the setting and assumptions of Theorem 2.2. We then get*

$$\varepsilon^2 Q_\varepsilon(a) \xrightarrow[\varepsilon \to 0]{\mathscr{L}} Q(a) = \int_a^\infty I\left\{\|\phi'_F\{K(s,\cdot)\}(t)\|_\infty \geq 1\right\} dt.$$

3. Applications

To illustrate the potential of our method, we will consider several examples. We will only regard the limit distributions of N_ε and not Q_ε,, as this is entirely analogous.

Example 3.1. First of all, consider the identity map $\phi = \iota$, which is linear and thus obviously Hadamard-differentiable. The differential is given by $\phi'_F(\alpha) = \alpha$, which is obviously continuous. This returns us to the empirical process result of Hjort & Fenstad (1992).

As Hadamard-differentiability follows a chain-rule (see Gill, 1989)) we can extend results in a hierarchical manner, as the following example shows.

Example 3.2. We now look at the Nelson-Aalen and Kaplan-Meier estimators of survival analysis. Let $X_1, \ldots X_n$ be *i.i.d.* life times with c.d.f. G and $C_1, \ldots, C_n$ *i.i.d.* censoring times with c.d.f. H and bivariate c.d.f. F. Note that we are under the rather restrictive assumption of simple random censoring. We will also use the following more natural assumption on the time. We will let time end at $\tau = K^{-1}(1)$, where $1-K = (1-G)(1-H)$. Uniform strong consistency is given in e.g. Theorem 1 of Shorack & Wellner (1986, Chapter 7.3).

This example reveal both the power and the bottleneck of our method. On the one hand, the extraction of a rather complex limit result is reduced to the computation of Hadamard-differentials. On the other hand is the limitations imposed by the empirical process theory of Appendix B; it only handles the *i.i.d.* case. If we were to use a more general empirical process, our results would be more powerful and could potentially be used with more general censoring mechanisms.

We observe $(Z_1, \Delta_1), \ldots, (Z_n, \Delta_n)$ where $Z_i = X_i \wedge C_i$ and $\Delta_i = I\{X_i \leq C_i\}$. The cumulative hazard function is

$$\Lambda(t) := \int_{[0,t]} \frac{1}{Pr\{Z \geq x\}} dPr\{Z \leq x, \Delta = 1\}.$$

This can be written as a functional of the bivariate cumulative distribution function $F = F_{X,\Delta}(x,\delta) = Pr\{X \leq x, \Delta \leq \delta\}$ as

$$\phi(F_{X,\Delta})(t) = \int_{[0,t]} \frac{1}{1 - F_{X,\Delta}(x,\delta)|_{\delta=1}} d(F_{X,\Delta}(x,\delta)|_{\delta=1} - F_{X,\Delta}(x,\delta)|_{\delta=0}).$$

Let F_n be the bivariate empirical *c.d.f.*. The Nelson-Aalen estimator can then be written as $\phi(F_n)$. Notice that ϕ can be decomposed into simpler maps, so we can use the chain-rule. Let $A \in D[\mathbb{R}^2]$ and regard the sequence of maps

$$\begin{aligned} A &\mapsto (A, A) \mapsto (1 - A(x,\delta)|_{\delta=1}, A(x,\delta)_{\delta=1} - A(x,\delta)|_{\delta=0}) \\ &\mapsto (\frac{1}{1 - A(x,\delta)|_{\delta=1}}, A(x,\delta)_{\delta=1} - A(x,\delta)|_{\delta=0}) \\ &\mapsto \int_{[0,t]} \frac{1}{1 - A(x,\delta)|_{\delta=1}} d(A(x,\delta)|_{\delta=1} - A(x,\delta)|_{\delta=0}), \end{aligned}$$

in which the first two maps are linear and continuous (and thus Hadamard-differentiable), and the two last ones are Hadamard-differentiable by Lemma 3 of Gill (1989) on the domain $\{(A, B) : \int |dA| < \infty, B > 0\}$ under the assumption that $1/A$ is of bounded variation. We will thus assume that this is true for $(A(x), B(x)) = (Pr\{Z \geq x\}, Pr\{Z \leq x, \Delta = 1\})$. van der Vaart & Wellner (1996, example 3.9.19) state that the derivative map of the composition of the last two maps is $\psi_{A,B}(\alpha, \beta) = \int_{[0,t]}(1/B)d\alpha - \int_{[0,t]}(\beta/B^2)dA$, where the first integral is interpreted in a partial integration fashion if α is of unbounded variation. The composition of the first two maps has differential $\varphi'_A(\alpha) = (\alpha(x,1), \alpha(x,1) - \alpha(x,0))$. From the chain-rule (Gill, 1989), we can conclude that

$$\begin{aligned} \phi'_{F_{X,\Delta}}(\alpha)(t) &= \psi'_{\varphi(F_{X,\Delta})} \circ \varphi'_{F_{X,\Delta}}(\alpha)(t) \\ &= \int_{[0,t]} \frac{1}{F(x,1) - F(x,0)} d\alpha(x,1) - \int_{[0,t]} \frac{\alpha(x,1) - \alpha(x,0)}{F(x,1) - F(x,0)} dF(x,1). \end{aligned}$$

This is a continuous map, and we invoke Theorem 2.2 w.r.t. the bivariate empirical process convergence and tail bounds of Appendix B to conclude that

$$\varepsilon^2 N_\varepsilon(\Lambda) \xrightarrow[\varepsilon \to 0]{\mathscr{L}} \left(\sup_{0 \leq s \leq 1} \sup_{0 \leq t \leq \tau} \left| \int_{[0,t]} \frac{1}{F(x,1) - F(x,0)} dK(s^{-1}, x, 1) - \int_{[0,t]} \frac{K(s^{-1}, x, 1) - K(s^{-1}, x, 0)}{F(x,1) - F(x,0)} dF(x,1) \right| \right)^2.$$

We continue with the Kaplan-Meier estimator and notice that it can be written as $\mathbb{K}_n = 1 - \Psi(-\hat{\Lambda}_n)$, where $\hat{\Lambda}_n$ is the Nelson-Aalen estimator and

$$\Psi(A)(s,t] = \prod_{s<u\leq t} (1 + dA(u))$$

where $\prod$ is product integration (see Andersen et al. (1992) for details). Its Hadamard-differential is given by Lemma 3.9.30 of van der Vaart & Wellner (1996) as

$$\Psi'_A(\alpha)(t) = \alpha(t) + \int_{(0,t]} \prod_{(0,s)} (1 + dA) dA(s)[\alpha(t) - \alpha(s)] + \int_{(0,t]} \alpha(r-) dA(r) \times \prod_{(r,t]} (1+dA) + \int\int_{0<s<r\leq t} \prod_{(0,s)} (1+dB) dB(s)[h(r-) - h(s)] \prod_{(r,t]} (1+dA) dA(r),$$

where $\alpha \in \mathbb{D}$ may be of unbounded variation. Using the chain rule of Hadamard-differentiation, we can regard the Kaplan-Meier estimator as $\Psi \circ -\phi(F_n)$. This yields

$$\varepsilon^2 N_\varepsilon(\mathbb{K}) \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \left(\sup_{0\le s\le 1} \sup_{0\le t\le\tau} \left| \Psi'_{-\phi(F)} \circ -\phi'_F\{K(s^{-1},\cdot)\}(t) \right| \right)^2,$$

where the Kiefer-process still has three-dimensional time.

Continuing even further, let us regard the quantile estimator under censoring, using the inverse Kaplan-Meier functional. It is strongly consistent. We will treat the quantile $p \in (0,1)$ as given. The quantile estimator under censoring is even uniformly strongly consistent for varying quantiles by Theorem 18.4.1 of Shorack & Wellner (1986) if the observations have a density f and $f(F^{-1})$ is continuous. The Hadamard-differentiability (tangentially to the set of functions that are continuous at $F^{-1}(p)$) of $\Phi(T)(p) = T^{-1}(p) = \inf\{x : T(x) \ge p\}$ is secured if $f(F^{-1})$ is also strictly positive, see Gill (1989, Corollary 2). The differential is given as $\Phi'_F(\alpha) = -\alpha(F^{-1}(p))/f(F^{-1}(p)$. Assuming $\mathbb{K} = 1-\Psi(-\phi(F_{X,\Delta}))$ is continuous at $F^{-1}(p)$, we can conclude that the quantile estimator can be written as $\Phi \circ (\Psi \circ -\phi(F_n))$ and

$$\varepsilon^2 N_\varepsilon(\mathbb{K}^{-1}(p)) \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \left(\sup_{0\le s\le 1} |\Phi'_{\Psi\circ-\phi(F)} \circ \Psi'_{-\phi(F)} \circ -\phi'_F\{K(s^{-1},\cdot)\}| \right)^2$$

using the chain rule. This limit is rather complex once written out, and it seems implausible that it would be found using other techniques.

The following example shows consistency between our and known results in a non-trivial case.

Example 3.3. Let us consider the quantile estimator $\hat{\xi}_{pn} = F_n^{-1}(p)$, which is $a.s.$ converging to $\xi_p = F^{-1}(p)$. By the Bahadur-representation (Serfling, 1980, Theorem 2.5.1) or the Lipschitz differentiability of the quantile function, it can be seen using section 4 or section 3.C of Hjort & Fenstad (1992) that

$$\varepsilon^2 N_\varepsilon \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \frac{p(1-p)}{f(\xi_p)^2} \left(\sup_{0\le s\le 1} |W_s| \right)^2. \tag{3.1}$$

Using the Hadamard-differentiability (discussed in Example 3.2) of $\hat{\xi}_{pn} = \Phi(F_n) = F_n^{-1}(p)$ we get that

$$\varepsilon^2 N_\varepsilon \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \left(\sup_{0\le s\le 1} \left| \frac{K_F(s,\xi_p)}{f(\xi_p)} \right| \right)^2,$$

which is seen to be equal in distribution to the limit of eq. (3.1) by the covariance structure of the Kiefer process given by eq. (B.1). Notice that

$$\mathrm{Cov}\,(K_F(s_1,\xi_p), K_F(s_2,\xi_p)) = (s_1 \wedge s_2)(F(\xi_p) - F(\xi_p)^2) = (s_1 \wedge s_2)p(1-p),$$

which means $K_F(s,\xi_p) =_D \sqrt{p(1-p)}W_s$ as they are both Gaussian.

Example 3.4. As a final example, we note that the functionals

$$\phi_1(T)(x) = \frac{T(x+\delta) - T(x-\delta)}{2\delta}$$

and

$$\phi_2(T, G) = T - G$$

are both linear and continuous, thus Hadamard differentiable with differential equal to them selves. When ϕ_1 is applied to the empirical process or the Kaplan-Meier functional, we get a density estimator. And when ϕ_2 is applied to two empirical processes or Kaplan-Meier estimators (assuming independence of the observations) we get a test statistic for the null hypothesis of equality in distribution. Both of the statistical functionals conform to the assumptions of Theorem 2.2.

4. Discussion

This paper extends the empirical process results of Hjort & Fenstad (1992) to a much larger class of functionals. As the class of Hadamard-differentiable functionals is vast, we could have given several even more complex examples. The main restrictions of our results are the empirical process theory used and the strict measurability conditions imposed by the weak convergence theory we use. As indicated by the survival analysis example, it would be very interesting to use a more general empirical process theory. We further conjecture that our measurability conditions could be loosened considerably by using the weak convergence theory of Hoffmann-Jørgensen as described in e.g. van der Vaart & Wellner (1996). Further, it would be interesting to see if the first-order results of this paper could be extended to second-order results in the style of Hjort & Fenstad (1995).

During the work with this paper, two new functionals were thought of. Let $R_\varepsilon(a, b, c)$ (for ratio) be given by

$$R_\varepsilon(a,b,c) = \frac{\int_{\lfloor a/\varepsilon^2 \rfloor}^{\infty} I\left\{b\varepsilon < \|\phi(F_{\lfloor s \rfloor}) - \phi(F)\|_\infty < c\varepsilon\right\} ds}{\int_{\lfloor a/\varepsilon^2 \rfloor}^{\infty} I\left\{\|\phi(F_{\lfloor s \rfloor}) - \phi(F)\|_\infty\} > \varepsilon\right\} ds}, \tag{4.1}$$

which was kindly suggested by Professor Nils Lid Hjort. Further, let

$$M_\varepsilon(a) = \frac{\int_{\lfloor a/\varepsilon^2 \rfloor}^{\infty} \|\phi(F_{\lfloor s \rfloor}) - \phi(F)\|_\infty I\left\{\|\phi(F_{\lfloor s \rfloor}) - \phi(F)\|_\infty > \varepsilon\right\} ds}{\int_{\lfloor a/\varepsilon^2 \rfloor}^{\infty} I\left\{\|\phi(F_{\lfloor s \rfloor}) - \phi(F)\|_\infty\} > \varepsilon\right\} ds}, \tag{4.2}$$

be the average size of the errors larger than ε among $n \geq a/\varepsilon^2$. We conjecture that

$$\varepsilon^{-1} M_\varepsilon(a) \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \frac{1}{Q(a)} \int_a^\infty \|\phi'_F\{K(s,\cdot)\}(t)\|_\infty I\left\{\|\phi'_F\{K(s,\cdot)\}(t)\|_\infty \geq 1\right\} dt,$$

and

$$R_\varepsilon(a,b,c) \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \frac{1}{Q(a)} \int_a^\infty I\left\{b \leq \|\phi'_F\{K(s,\cdot)\}(t)\|_\infty \leq c\right\} dt.$$

under the assumptions of Theorem 2.2. To prove this, we could follow the procedure used for N_ε and Q_ε.

5. Acknowledgments

This paper was written as a part of the author's master thesis. I would like to express my thanks to my advisor Professor Nils Lid Hjort. Both for introducing me to the techniques and problems concerning N_ε-related results and for keeping me interested by continuously presenting new and interesting problems. In specific, this paper would never have come to be unless he suggested that I investigated the N_ε properties of the Nelson-Aalen estimator. I would further like to thank Marion Haugen for pointing out inconsequent notation used in the paper and Trygve Karper for many interesting discussions on analysis.

Appendix A. Proofs

The following probabilistic observation will be needed for the proof of Lemma 2.1.

Lemma A.1. *Let $(X_n)_{n=1}^{\infty}$ be a sequence of non-negative random variables that converges a.s. to zero with the additional property that $f(n)X_n \xrightarrow[n\to\infty]{\mathscr{P}} 0$, where $f : \mathbb{N} \mapsto \mathbb{R}$ is an arbitrary positive and strictly increasing function. We then also have*

$$f(n) \sup_{m\geq n} X_m \xrightarrow[n\to\infty]{\mathscr{P}} 0. \tag{A.1}$$

Proof. Use $\bar{S}$ to note the complement of a given set S. Let $\varepsilon > 0$ and

$$A_n^\varepsilon := \{\omega \in \Omega : f(n)X_n(\omega) \geq \varepsilon\}.$$

The assumed convergence in probability is equivalent to $Pr\{A_n^\varepsilon\} \to_{n\to\infty} 0$ for any given $\varepsilon > 0$. Now set $B_n^\varepsilon := \{\omega \in \Omega : f(n)\sup_{m\geq n} X_m \geq \varepsilon\}$. Let Ω^* be the set of unit measure where $X_n \to 0$. We show that $Pr\left\{(\bar{A}_n^\varepsilon \setminus \bar{B}_n^\varepsilon)\bigcap \Omega^*\right\} \to 0$, which implies $Pr\{\bar{B}_n^\varepsilon\} \to 1$ since $Pr\{\bar{A}_n^\varepsilon\} \to 1$ and $\bar{B}_n^\varepsilon \subseteq \bar{A}_n^\varepsilon$. Observe that

$$(\bar{A}_n^\varepsilon \setminus \bar{B}_n^\varepsilon)\bigcap \Omega^* = \left\{\omega \in \Omega^* : X_n(\omega) < \varepsilon/f(n), \sup_{m\geq n} X_n(\omega) \geq \varepsilon/f(n)\right\}.$$

As $\omega \in \Omega^*$, where there is convergence, there exists a finite index $m_{\text{sup}}^{(n)}$ where $\sup_{m\geq n} X_n(\omega)$ is reached by the non-negativeness of X_n. Notice that $X_{m_{\text{sup}}^{(n)}} \geq \varepsilon/f(n)$ implies $X_{m_{\text{sup}}^{(n)}} \geq \varepsilon/f(m_{\text{sup}}^{(n)})$ by the monotonicity of $f(\cdot)$. This gives

$$\begin{aligned}&\left\{\omega \in \Omega^* : X_n(\omega) < \frac{\varepsilon}{f(n)}, X_{m_{\text{sup}}^{(n)}} \geq \frac{\varepsilon}{f(n)}\right\}\\ &\quad\subseteq \left\{\omega \in \Omega^* : X_n(\omega) < \frac{\varepsilon}{f(n)}, X_{m_{\text{sup}}^{(n)}} \geq \frac{\varepsilon}{f(m_{\text{sup}}^{(n)})}\right\} = (\bar{A}_n^\varepsilon \setminus \bar{A}_{m_{\text{sup}}^{(n)}}^\varepsilon)\bigcap \Omega^*.\end{aligned}$$

As $m_{\text{sup}}^{(n)}$ is just an increasing subsequence of $\mathbb{N}$, the sequence $\{A_{m_{\text{sup}}^{(n)}}^\varepsilon\} \subseteq \{A_n^\varepsilon\}$ will also converge in probability to zero. Since ε was arbitrary, the convergence of Equation (A.1) follows. $\square$

Note that the needed measurability conditions are included in the assumptions of the delta-method as given in Gill (1989), so we need not stress this in the proofs.

Proof of Lemma 2.1. Define $Q_n = R_n/\sqrt{n}$. Notice that $\phi(F_n) - \phi(F) = \phi'_F(F_n - F) + Q_n$ from the assumed linearity of ϕ'_F. This implies

$$Pr\left\{\sqrt{m}\sup_{n\geq m}\|\phi(F_n) - \phi(F)\|_\infty < y\right\} = Pr\left\{\sqrt{m}\sup_{n\geq m}\|\phi'_F(F_n - F) + Q_n\|_\infty < y\right\}.$$

From the triangle inequality of both the supremum over n and T, the difference between $\sqrt{m}\sup_{n\geq m}\|\phi'_F(F_n - F) + Q_n\|_\infty$ and $\sqrt{m}\sup_{n\geq m}\|\phi'_F(F_n - F)\|_\infty$ is dominated by $\sqrt{m}\sup_{n\geq m}\|Q_n\|_\infty$. They will thus have the same asymptotic distribution if we can show that $\sqrt{m}\sup_{n\geq m}\|Q_n\|_\infty \xrightarrow[n\to\infty]{\mathscr{P}} 0$. Note that from assumption of Hadamard-differentiability, we have $\sqrt{n}\|Q_n\|_\infty \xrightarrow[n\to\infty]{\mathscr{P}} 0$ per definition, so that Lemma A.1 reduces the problem to show that $\|Q_n\|_\infty \xrightarrow[n\to\infty]{\text{a.s.}} 0$ is true. We assumed $\|\phi(F_n) - \phi(F)\|_\infty \xrightarrow[n\to\infty]{\text{a.s.}} 0$ and since ϕ'_F is continuous with respect to the uniform norm (this is an assumption of Hadamard-differentiability in Gill (1989)), we also get that $\|\phi'_F(F_n - F)\|_\infty \xrightarrow[n\to\infty]{\text{a.s.}} \|\phi'_F(0)\|_\infty = 0$ from the assumption $\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{a.s.}} 0$. Now notice that

$$\begin{aligned}0 \leq \|Q_n\|_\infty &= \|Q_n + \phi'_F(F_n - F) - \phi'_F(F_n - F)\|_\infty \\ &\leq \|Q_n + \phi'_F(F_n - F)\|_\infty + \|\phi'_F(F_n - F)\|_\infty \\ &= \|\phi(F_n) - \phi(F)\|_\infty + \|\phi'_F(F_n - F)\|_\infty \xrightarrow[n\to\infty]{\text{a.s.}} 0,\end{aligned}$$

which completes the proof. □

Proof of Lemma 2.3. We will gain the limit in Equation (2.6) by extending Equation (2.5)) from the continuity of ϕ'_F.

To shorten notation, let $H^i_m(s,t) := \sqrt{m}[F^i_{\lfloor ms\rfloor}(t) - F^i(t)]$, where F^i is defined by the i'th row-element of the vector F. Let further $H_m(s,t) := \sqrt{m}[F_{\lfloor ms\rfloor}(t) - F(t)]$. Now, let $\delta > 0$ and define

$$\begin{aligned}B^\delta_{c,m} &:= \left\{\omega\in\Omega : \sqrt{m}\max_{1\leq i\leq d}\sup_{n\geq cm}\sup_{t\in T}|F^i_n(t) - F^i(t)| < \delta\right\} \\ &= \left\{\omega\in\Omega : \max_{1\leq i\leq d}\sup_{s\geq c}\sup_{t\in T}|H^i_m(s,t)| < \delta\right\} \\ &= \left\{\omega\in\Omega : \max_{1\leq i\leq d} H^i_m([c,\infty),T) \subseteq (-\delta,\delta)\right\}\end{aligned}$$

and

$$\begin{aligned} A^{\delta}_{c,m} &:= \left\{\omega\in\Omega : \sup_{n\geq cm}\|\phi'_F(\sqrt{m}[F_n-F])\|_\infty<\delta\right\} \\ &= \left\{\omega\in\Omega : \sup_{s\geq c}\sup_{t\in T}|\phi'_F\{H_m(s,t)\}|<\delta\right\} \\ &= \left\{\omega\in\Omega : \phi'_F\{H_m([c,\infty),T)\}\subseteq(-\delta,\delta)\right\}. \end{aligned}$$

As ϕ'_F is continuous with respect to the uniform norm, we can choose a l that is so large that $\max_{1\leq i\leq d} H^i_m([c,\infty),T)\subseteq(-\frac{\delta}{l},\frac{\delta}{l})$ implies $\phi'_F\{H_m([c,\infty),T)\}\subseteq(-\delta,\delta)$. Thus, $A^{\delta/l}_{c,m}\subseteq B^{\delta}_{c,m}$ which gives $P\{A^{\delta/l}_{c,m}\}\leq Pr\{B^{\delta}_{c,m}\}$. We now use the assumption given by Equation (2.5). This applies for all elements of the vector $F=(F^1,F^2,\ldots,F^q)$. This gives, for some $\Delta>0$, that

$$\begin{aligned} 0 &\leq \lim_{c\to\infty}\limsup_{m\to\infty} Pr\left\{\max_{1\leq i\leq d}\sup_{n\geq cm}\|\sqrt{m}[F^i_n-F^i]\|_\infty\geq\Delta\right\} \\ &\leq \lim_{c\to\infty}\limsup_{m\to\infty} Pr\left\{\bigcap_{1\leq i\leq d}\{\sup_{n\geq cm}\|\sqrt{m}[F^i_n-F^i]\|_\infty\geq\Delta\}\right\}=0. \end{aligned}$$

By using $\Delta=\delta/l$ we get

$$\lim_{c\to\infty}\limsup_{m\to\infty} Pr\left\{\sqrt{m}\sup_{n\geq cm}\|\phi'_F(F_n-F)\|_\infty\geq\delta\right\}=0,$$

which completes the proof. □

We need the following Lemma to prove Theorem 2.4.

Lemma A.2. *We have that*

$$\int_{\lfloor am\rfloor/m}^{\infty} I\left\{\sqrt{m}\|\phi(F_{\lfloor mt\rfloor})-\phi(F)\|_\infty>1\right\}\,dt \tag{A.2}$$

and

$$\int_{\lfloor am\rfloor/m}^{\infty} I\left\{\|\phi'_F(\sqrt{m}[F_{\lfloor mt\rfloor}-F])\|_\infty>1\right\}\,dt \tag{A.3}$$

have the same asymptotic distribution.

Proof. Let $A=\lfloor am\rfloor/m$. Observe that for $\gamma\in(0,1)$ and $a,b\geq0$ we have that

$$I\{a+b\geq1\}\leq I\{a\geq1-\gamma\}+I\{b\geq\gamma\}.$$

By the definition of Hadamard-differentiability, we have that $\sqrt{m}[\phi(F_{\lfloor mt\rfloor})-\phi(F)]=\sqrt{m}\phi'_F(F_{\lfloor mt\rfloor}-F)+\sqrt{m}Q_{\lfloor mt\rfloor}$. This implies that the difference between

Equation (A.2) and Equation (A.3) can be written as

$$\int_A^\infty I\left\{\sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)+Q_{\lfloor mt\rfloor}\|_\infty \geq 1\right\} - I\left\{\sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\|_\infty > 1\right\}\,dt$$
$$\leq \int_A^\infty I\left\{\sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\| \geq 1-\gamma\right\} - I\left\{\sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\|_\infty > 1\right\}$$
$$+ I\left\{\sqrt{m}\|Q_{\lfloor mt\rfloor}\|_\infty \geq \gamma\right\}\,dt$$
$$= \int_A^\infty I\left\{1-\gamma \leq \sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\| \leq 1\right\} + I\left\{\sqrt{m}\|Q_{\lfloor mt\rfloor}\|_\infty \geq \gamma\right\}\,dt,$$

from the triangle inequality of the uniform norm. We will first bound the integral of the first addend, then the integral of the second. We know that $\sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\|_\infty \xrightarrow[t\to\infty]{\text{a.s.}} 0$ from the continuity of both ϕ_F' and the supremum-mapping. Let Ω_m^* be the unit-measure set where we do have this convergence. Let

$$L_m(1/2) := \sup_{n\geq m}\sup\left\{t \geq 0 : \sqrt{n}\|\phi_F'(F_{\lfloor nt\rfloor}-F)\| \geq 1/2\right\},$$

which is finite on $\bigcap_{n\geq m}\Omega_n^*$ and decreasing. Assume $\gamma < 1/2$ and observe that this gives

$$\int_A^\infty I\left\{1-\gamma \leq \sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\| \leq 1\right\}\,dt$$
$$= \int_A^{L_m(1/2)} I\left\{1-\gamma \leq \sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\| \leq 1\right\}\,dt,$$

which reduces our bounding procedure to a finite area. Note that the $L_m(1/2)$ depends on m. As it is a decreasing sequence, this will not cause any problems: we can rather use $L_{m_0}(1/2)$ for some large m_0. As $\phi_F'(\sqrt{m}[F_{\lfloor mt\rfloor}-F])$ converges weakly to $\phi_F'(K_t(s))$, which is continuous, the modulus of continuity in the multivariate Skorokhod-space of Bickel & Wichura (1971) converge to zero in probability. In specific, the same is true for the Skorokhod modulus of continuity in t alone. Call this $w''(\delta)$. Let $\delta > 0$ and choose m so large and γ so small, that $B_{\gamma,m}(\delta) := \{\omega : w''(\gamma) < \delta\}$ measures up close to one. Assume we are in $(\bigcap_{n\geq m}\Omega_n^*) \cap B_{\gamma,m}(\delta)$. This gives

$$\int_A^{L_{m_0}(1/2)} I\left\{1-\gamma \leq \sqrt{m}\|\phi_F'(F_{\lfloor mt\rfloor}-F)\| \leq 1\right\}\,dt \leq (L_{m_0}(1/2)-A)\delta,$$

where δ can be made arbitrarily small by choosing γ small enough. This completes the bounding procedure of the first addend. For the second, define $D_m(\gamma) = \{\omega \in \Omega : \sqrt{m}\sup_{t\geq 1}\|Q_{\lfloor mt\rfloor}\|_\infty < \gamma\}$. By the proof of Lemma 2.1, the measure of $D_m(\gamma)$ converges to unity for any $\gamma > 0$. As $\gamma < 1/2$, we then get that $\int_A^\infty I\{\sqrt{m}\|Q_{\lfloor mt\rfloor}\|_\infty > 1\}\,dt = 0$ in $D_m(\gamma)$. Thus, we can bound the difference between Equation (A.2) and Equation (A.3) in $D_m(\gamma)\cap(\bigcap_{n\geq m}\Omega_n)\cap B_{\gamma,m}(\delta)$ by choosing m large enough and δ and γ small enough. As all these sets are asymptotically of unit measure, we have convergence in probability. □

Proof of Theorem 2.4. As in Hjort & Fenstad (1992), we use

$$\varepsilon^2 Q_\varepsilon = \int_{\lfloor am\rfloor/m}^{\infty} I\left\{\sqrt{m}\|\phi(F_{\lfloor mt\rfloor}) - \phi(F)\|_\infty > 1\right\} dt.$$

By the previous Lemma, we only have to be concerned with the structurally simpler

$$\int_{\lfloor am\rfloor/m}^{\infty} I\left\{\|\phi'_F(\sqrt{m}[F_{\lfloor mt\rfloor} - F])\|_\infty > 1\right\} dt.$$

By the Continuous Mapping Theorem, the result is obvious if we limit the integral to the finite region $[\lfloor am\rfloor/m, c]$ where c is large. To show that this suffices, we show a tail inequality parallel to eq. (2.3). By Lemma 2.3, this is simple. Choose m and c so large that $\sup_{n\geq cm}\|\phi'_F(\sqrt{m}[F_n - F])\|_\infty \geq 1$ with probability less than η. This means that the tail above c is zero with probability larger than $1-\eta$.

Appendix B. Multivariate empirical processes

We now provide the needed asymptotics for empirical processes with multidimensional time. That is, processes in some subspace of $\mathbb{D}(T)$, typically $T = \mathbb{R}^q$. We will use the multidimensional generalization of Skorokhod-space given in Bickel & Wichura (1971). All of their results use $T = [0,1]^q$. This will be sufficient, as our results are also valid when T is more general by considerations of the inverse transformation of the cumulative distribution function F (which may be entirely arbitrary), see Remark 1 of Csörgo (1981).

Let $t = (t_1, \ldots, t_q) \in \mathbb{R}^q$. Define the multivariate empirical process by $F_n(t) = \frac{1}{n}\sum_{i=1}^n Z_i(t)$, where $Z_i(t) = I_{C(t)}(X_i) - Q(C(t))$ and $C(t) = \Pi_p[0, t_p]$. Theorem 1 in Chapter 1.4 of Borovkov (1998) gives

$$\|F_n(t) - F(t)\|_\infty = \sup_t |F_n(t) - F(t)| \xrightarrow[n\to\infty]{\text{a.s.}} 0,$$

needed to initialize a discussion on N_ε-related functionals. To use Theorem 2.2 in conjunction with Lemma 2.3, we need both convergence of the partial sum process of Equation (2.2) and the asymptotically negligible tail described by Equation (2.5). Define $G_n(t) = \sqrt{n}F_n(t) = \frac{1}{n^{1/2}}\sum_{i=1}^n Z_i(t)$ and $X_n(s,t) = (\frac{\lfloor ns\rfloor}{n})^{1/2} G_{\lfloor ns\rfloor}(t) = \frac{1}{n^{1/2}}\sum_{i=1}^{\lfloor ns\rfloor} Z_j(t)$, where $s \in [0,1]$ and $t \in T$.

Theorem B.1. *In the current setting, assume each element in X_i is independent and that $T = [0,1]^q$. We then have $X_n(s,t) \xrightarrow[n\to\infty]{\mathscr{L}} X$, where X is a D_{q+1} dimensional continuous Gaussian zero mean process with*

$$Cov(X(s_1,t_1), X(s_2,t_2)) = (s_1 \wedge s_2)\Gamma(t_1,t_2), \tag{B.1}$$

where $\Gamma(t_1,t_2) = Cov(Z_1(t_1), Z_1(t_2))$ (it is thus a Kiefer process). Further we have that

$$Pr\left\{\sup_{0\leq s\leq 1, t\in[0,1]^q} |X_m(s,t)| \geq b\right\} \leq A/b^4 \tag{B.2}$$

for some universal constant A.

Proof. The convergence is Theorem 6 in Bickel & Wichura (1971) combined with the comment on empirical processes following its proof. As commented in the proof of the Theorem in section 4 of Hjort & Fenstad (1992), the tail-bound (B.2) is found by Bickel & Wichura's Theorem 1 in conjunction with their inequality (1). □

We mention that the Kiefer process is continuous as it is a continuous transformation (see Csörgo & Révés, 1975) of a Brownian Sheet. Specifically,

$$K(t_1, t_2, \ldots, t_{p-1}, s) =_D W(t_1, t_2, \ldots, t_p, s) - (\Pi_{i=1}^p t_i) W(1, \ldots, 1, s).$$

The continuity of a Brownian sheet is given in e.g. Khoshnevisan (2002, Theorem 3.2.1.).

We will be concerned with vectors of multivariate empirical processes and we will use $|F(t)| = \max_{1 \le i \le m} |F^i(t)|$ in this context. The previous Theorem gives us the weak convergence required by Equation (2.2) if we use the product topology and assume that all observations are independent. Now we only need the tail inequality of Equation (2.5).

Lemma B.2. *In the current multidimensional and multivariate setting, we have that*

$$\lim_{c\to\infty} \limsup_{m\to\infty} Pr \left\{ \sqrt{m} \sup_{n \ge cm} \sup_{t\in[0,1]^q} |F_n(t) - F(t)| \ge \delta \right\} = 0.$$

Proof. Notice that

$$\begin{aligned} & Pr \left\{ \sup_{n\ge cm} \sup_{t\in[0,1]^q} \max_{1\le i\le d} |F_n^i(t) - F^i(t)| \ge \delta \right\} \\ \le \quad & Pr \left\{ \max_{1\le i\le d} \sup_{n\ge cm} \sup_{t\in[0,1]^q} |F_n^i(t) - F^i(t)| \ge \delta \right\} \\ \le \quad & Pr \left\{ \bigcap_{1\le i\le d} [\omega \in \Omega : \sup_{n\ge cm} \sup_{t\in[0,1]^q} |F_n^i(t,\omega) - F^i(t,\omega)| \ge \delta] \right\} \end{aligned}$$

which can be made arbitrarily small by the one dimensional case, which in turn is implied by Equation (B.2). □

References

Andersen, P., Borgan, Ø., Gill, R. & Keiding, N. (1992). *Statistical Models Based on Counting Processes.* Springer Series in Statistics. Springer Verlag.

Atlagh, M., Broniatowski, M. & Celant, G. (2005). Last exit times for a class of asymptotically linear estimators. *Communications in statistics* .

Barbe, P., Doisy, M. & Garel, B. (1998). Last passage time for the empirical mean of some mixing processes. *Statistics & Probability Letters* , 237–245.

Bickel, P. & Wichura, M. (1971). Multiparameter Stochastic processes. *The Annals of Mathematical Statistics* **42**, 1656–1670.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

Borovkov, A. (1998). *Mathematical Statistics.* Gordon and Breach Science Publishers.

Csörgo, M. & Révés, P. (1975). A new method to prove strassen type laws of invariance principle. II. *Z. Wahrscheinlichkeitstheorie verw. Gebeite* **31**, 261–269.

Csörgo, S. (1981). Limit behaviour of the empirical characteristic function. *The Annals of Probability* **9**, 130–144.

Gill, R. D. (1989). Non- and Semi-parametric Maximum Likelihodd Estimators and the Von Mises Method (Part 1). *Scandinavian Journal of Statistics* **16**, 97–128.

Hjort, N. L. & Fenstad, G. (1992). On the last time and the number of times an estimator is more than ϵ from its target value. *The Annals of Statistics* **20**, 469–489.

Hjort, N. L. & Fenstad, G. (1995). Second-order asymptotics for the number of times an estimator is more than ϵ from its target value. *Journal of Statistical Planning and Inference* **48**, 261–275.

Khoshnevisan, D. (2002). *Multiparameter Processes - An introduction to Random Fields.* Springer Monographs in Mathematics. Springer.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

Shao, J. (2003). *Mathematical Statistics.* Springer Texts in Statistics. Springer.

Shorack, G. & Wellner, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley.

Stute, W. (1983). Last Passage Times of M-Estimators. *Scandinavian Journal of Statistics* **10**, 301–305.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer Series in Statistics. Springer.