Last exit times for a class of asymptotically linear estimators*

M. Atlagh[1], M. Broniatowski[2], and G. Celant[3]

[1] Département de Mathématiques,
Université Louis Pasteur, 4 Rue René Descartes, 67000 Strasbourg, France
[2] LSTA,
Université Paris 6, 4 Place Jussieu, 75005 Paris, France
[3] Dipartimento di Scienze Statistiche,
Via C.Battisti 241, Padova, Italy

Abstract. We study the last exit time for the Glivenko-Cantelli statistics indexed by some class of functions. Also we provide upper bounds for its tail distribution. Our first example is the Glivenko-Cantelli statistics indexed by a subclass of a Sobolev space; we next consider last exit times for adaptive semiparametric estimates in the spirit of Klaassen, for which we provide the distribution and tail bounds uniformly upon the nuisance parameter.

AMS 2000 Classification:Primary 62F35
Keywords: Last exit time, Functional Glivenko-Cantelli statistics, Adaptive estimation, Semiparametric estimation.

1 Introduction

Rates of convergence for statistical estimators usually focus of asymptotic equivalents for the distance between the estimate and the parameter it intends to approximate. When the estimate is strongly consistent, which is to say that it converges almost surely, then the time which is necessary in order that it stays in some neighborhood of the true value of the parameter from then on is a well defined random variable, which bears a very intuitive sense and which, sometimes, can be evaluated, at least for small neighborhoods of the parameter. In this context, the situation is quite similar to the case when we consider a deterministic sequence x_n converging to x in a metric space: given some (small) ε, which is the order of magnitude of the integer $N(\varepsilon)$ such that, for all n larger than $N(\varepsilon)$, the distance between x_n and x remains forever smaller than ε? This class of problems is usually referred to as "last exit times" problems, considering that the terms of the sequence of estimates may stay outside the ε-neighborhood of its limit only when when n is smaller than $N(\varepsilon)$. This notion has been presented for sequences of M-estimates by [Stute, 1983], and has been extended to the last exit time for the Glivenko-Cantelli statistics by [Hjort and Fenstad, 1992]; extensions for the case when the sample is drawn from a Markov chain or from a strongly mixing

* Work partly supported by Ateneo di Padova, n. CPDAO23415, Italy

sequence have been studied by [Barbe *et al.*, 1999], and some extensions for U-statistics have recently been proposed by [Bose and Chatterjee, 2001]. The present paper follows this chain of works and provides some insight in the range of adaptive semi parametric estimates; it also provides some information on the tail of the distribution of last exit times for those types of estimates. The main tool to be imported for the obtention of the law of $N(\varepsilon)$ for such estimates is uniformity with respect to the nuisance parameter. This is achieved through Gaussian approximations for the so-called sequential empirical process, a device which has been proposed in the form which is to be used here by [Sheehy and Wellner, 1992].

The structure of the paper is as follows: The first section is devoted to the obtention of a general result on last exit times for the Glivenko-Cantelli statistics indexed by a class of functions, following the line defined by [Stute, 1983]. The second section specializes this result for various situations, with an emphasis towards semi parametric adaptive estimators in the spirit of [Klaassen, 1987].

2 Last exit time for the functional Glivenko-Cantelli Statistics

A sample $(X_1, ..., X_n)$ is given , with i.i.d. components following a common distribution P on some space $\mathcal{X}$. For f a real valued measurable function on $\mathcal{X}$ we denote Pf the expectation of f with respect to P, i.e. $Pf := \int f dP$. Denote P_n the empirical measure pertaining to the sample, $P_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ where δ_x is the Dirac mass at point x. In the sequel $\mathcal{F}$ denotes a subclass of functions in $\mathcal{L}^2(P)$ for which we assume that for all x in $\mathcal{X}$, the condition

$$\sup_{f\in\mathcal{F}} |f(x) - Pf| \quad \text{is finite} \tag{1}$$

holds. Define

$$N_\varepsilon := \sup\left\{ n \geq 1 : \sup_{f\in\mathcal{F}} |(P_n - P)(f)| \geq \varepsilon \right\},$$

which denotes the last exit time for the Glivenko–Cantelli statistics indexed by $\mathcal{F}$. Since $\mathcal{F}$ is a Donsker class, it satisfies the Glivenko–Cantelli property, namely

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} |(P_n - P)f| = 0 \quad \text{a.s.},$$

which implies that N_ε is a.s. finite. We consider the limiting distribution of $\varepsilon^2 N_\varepsilon$ when ε tends to 0.

Following [Stute, 1983], with $N_\varepsilon(f) := \sup\{n \geq 1 \, : \, |P_n f - Pf| > \varepsilon\}$, we have, for fixed f in $\mathcal{F}$,

Proposition 1 *Let f belongs to $\mathcal{L}^2(P)$. Then*

(i) $\lim_{\varepsilon\to 0} \varepsilon^2 N_\varepsilon(f) \stackrel{\mathrm{d}}{=} W^2_{\max}(f) := \sigma^2(f) \sup_{0\le s\le 1} W^2(s)$,
where $\sigma^2(f) = Pf^2 - (PF)^2$, *and* $W(s)$ *is the standard Wiener process.*

(ii) $\lim_{\lambda\to\infty} \lim_{\varepsilon\to 0} \dfrac{P\{\varepsilon^2 N_\varepsilon(f) > \lambda\}}{\psi\left(\frac{\sqrt{\lambda}}{\sigma(f)}\right)} = 1$, *where* ψ *denotes the upper tail of the standard normal distribution* $\psi(\lambda) := P[N(0,1) > \lambda]$.

We obtain an information pertaining to the moments of the r.v. $\varepsilon^2 N_\varepsilon(f)$ for small ε.

A sequence of r.v.'s Y_n is r-quick convergent to 0 whenever, for all $\varepsilon > 0$, $E(N_\varepsilon^r) := E(\sup\{n \ge 1 : |X_n| \ge \varepsilon\})^r$ is finite.

This property has been used by [Lai, 1981] in order to assess optimality properties of probability ratio tests in sequential analysis.

As a consequence of Proposition 1 we have

Corollary *Let* f *belong to* $\mathcal{L}^2(P)$. *The sequence* $(P_n - P)(f)$ *is* r*-quick convergent to 0 for all* $r > 0$.

Proof of Proposition 1

(i) for all $f \in \mathcal{L}^2(P)$, it holds $N_\varepsilon(f) = \sup\left\{n \ge 1 : \frac{1}{n}\left|\sum_{i=1}^n (f(X_i) - Pf)\right| \ge \varepsilon\right\}$
$= \sup\left\{n \ge 1 : \frac{\nu_n(f)}{\sqrt{n}} \ge \varepsilon\right\}$.

Let y be a positive number. Then $P\left\{\varepsilon^2 N_\varepsilon(f) \ge y\right\} = P\left\{N_\varepsilon(f) \ge \frac{y}{\varepsilon^2}\right\}$.
Define $m := < y/\varepsilon^2 >$, the smallest integer larger or equal y/ε^2. Then

$$\begin{aligned}
P\{\varepsilon^2 N_\varepsilon(f) \ge y\} &= P\{N_\varepsilon(f) \ge m\} \\
&= P\left\{\sup\left\{n \ge 1 : \frac{1}{n}\left|\sum_{i=1}^n (f(X_i) - Pf)\right| \ge \varepsilon\right\} \ge m\right\} \\
&= P\left\{\sup_{n\ge m} \frac{1}{n}\left|\sum_{i=1}^n (f(X_i) - Pf)\right| \ge \sqrt{\frac{y_0}{m}}\right\},
\end{aligned}$$

where $y_0 := m\varepsilon^2$. We thus have $0 \le y_0 - y \le \varepsilon^2$, which entails that y_0 tends to y as ε tends to 0.

For all f in $L^2(P)$, we therefore have

$$\begin{aligned}
P\{\varepsilon^2 N_\varepsilon(f) \ge y\} &\le P\left\{\sqrt{m}\sup_{n\ge m} \frac{1}{n}\left|\sum_{i=1}^n (f(X_i) - Pf)\right| \ge \sqrt{y_0}\right\} \\
&= P\left\{\sqrt{m}\sup_{t\ge 1} \frac{1}{tm}\left|\sum_{i=1}^{mt} (f(X_i) - Pf)\right| \ge \sqrt{y_0}\right\} \\
&= P\left\{\sigma(f)\sup_{t\ge 1} \frac{1}{t}\left|\frac{1}{\sigma(f)\sqrt{m}}\sum_{i=1}^{mt} (f(X_i) - Pf)\right| \ge \sqrt{y_0}\right\} \\
&=: P(E_m).
\end{aligned}$$

For all $r > 1$, set

$$A_m(r) := \left\{ \sigma(f) \sup_{1 \le t \le r} \frac{1}{t} \left| \frac{1}{\sigma(f)\sqrt{m}} \sum_{i=1}^{mt} f(X_i) - Pf) \right| \ge \sqrt{y_0} \right\}$$

and

$$B_m(r) := \left\{ \sigma(f) \sup_{t > r} \frac{1}{t} \left| \frac{1}{\sigma(f)\sqrt{m}} \sum_{i=1}^{mt} (f(X_i) - Pf) \right| \ge \sqrt{y_0} \right\}.$$

For all $r > 1$,

$$P(E_m) = P(A_m(r) \cup B_m(r)),$$

and therefore

$$\begin{aligned} \lim_{\varepsilon \to 0} P\left(\varepsilon^2 N_\varepsilon(f) \ge y\right) &= \lim_{m \to \infty} P(E_m) = \lim_{r \to \infty} \lim_{m \to \infty} P(A_m(r) \cup B_m(r)) \\ &= \lim_{r \to \infty} \lim_{m \to \infty} P(A_m(r)) + P(B_m(r)) - P(A_m(r) \cap B_m(r)). \end{aligned}$$

If

$$\lim_{r \to \infty} \lim_{m \to \infty} P(B_m(r)) = 0, \tag{2}$$

then

$$\lim_{m \to \infty} P(E_m) = \lim_{r \to \infty} \lim_{m \to \infty} P(A_m(r)).$$

The proof of (2) is easy.
Following Donsker Invariance Principle, the processes

$$\left(\frac{1}{\sigma(f)\sqrt{m}} \sum_{i=1}^{mt} (f(X_i) - Pf); \quad t \in [1, r] \right)_{m \ge 1}$$

converge in distribution in $D[1, r]$ to the Brownian process $W(t)$. By continuity of the supremum it therefore holds

$$\lim_{m \to \infty} \sup_{1 \le t \le r} \frac{1}{t\sigma(f)\sqrt{m}} \left| \sum_{i=1}^{mt} (f(X_i) - Pf) \right| \overset{\mathrm{d}}{=} \sup_{1 \le t \le r} \left| \frac{W(t)}{t} \right|.$$

For all $t \in [1, r]$, the process $W^*(\frac{1}{t}) = \frac{W(t)}{t}$ is also a Brownian process. Therefore

$$\sup_{1 \le t \le r} \left| \frac{W(t)}{t} \right| \overset{\mathrm{d}}{=} \sup_{\frac{1}{r} \le s \le 1} |W^*(s)|.$$

Hence whenever (2) holds, we have

$$\begin{aligned} \lim_{\varepsilon \to 0} P(\varepsilon^2 N_\varepsilon(f) \ge y) &= P\left(\sigma(f) \lim_{r \to \infty} \sup_{\frac{1}{r} \le s \le 1} |W(s)| \ge \sqrt{y} \right) \\ &= P\left\{ \sigma(f) \sup_{0 \le s \le 1} |W(s)| \ge \sqrt{y} \right\} \\ &= P\left\{ W^2_{\max}(f) \ge y \right\}, \end{aligned}$$

where $W^2_{\max}(f) = \sigma^2(f) \sup_{0 \le s \le 1} W^2(s)$.

(ii) The maximal variance of $\sigma(f)W(s)$ equals $\sigma^2(f)$ and is obtained when $s = 1$.
Let

$$I_h := \{s \in [0,1] : s\sigma^2(f) \geq \sigma^2(f) - h^2\} = \left[1 - \frac{h^2}{\sigma^2(f)}, 1\right],$$

and note that $E(\sigma^2(f)W(1)) = \sigma^2(f)$.
The uniform a.s. continuity of $\sigma(f)W(s)$ on $[0,1]$ entails that

$$\lim_{h\to 0} \frac{1}{h} E\left\{\sup_{s\in I_h} \sigma(f)|W(s) - W(1)|\right\} = 0,$$

which proves that the conditions in [Adler, 1990], Theorem 5.5 are fulfilled, proving the claim. ■

Let us turn to the uniform case, that is, consider the limiting distribution of $\varepsilon^2 N_\varepsilon$, $\varepsilon \to 0$. We will assume

H1: $\mathcal{F}$ is a Donsker class
H2: For all $x \in \mathcal{X}$, $\sup_{f\in\mathcal{F}} |f(x) - P(f)|$ is finite
H3: $\mathcal{F}$ is a permissible class of function, implying that $\sup_{f\in\mathcal{F}} |P_n f - Pf|$ is measurable.

Define a Gaussian process Z_P defined on $[0,1]\times\mathcal{F}$, a version of which has uniformly bounded sample paths which are uniformly continuous on $[0,1] \times \mathcal{F}$ when equipped with the $\tilde{\rho}_P$ pseudo-metric defined on $[0,1] \times \mathcal{F}$ defined through $\tilde{\rho}_P\left((s,f),(t,g)\right) := |s-t| + P(f-g)^2$. The existence of such process is a consequence of hypothesis (H1) above (see [Sheehy and Wellner, 1992]). The *Kiefer-Müller* process Z_P is centered for any s and f, and its covariance operator is given by

$$cov\left[Z_P(s,f), Z_P(t,g)\right] = (s \wedge t)(Pfg - PfPg).$$

We then have

Proposition 2 *When $\mathcal{F}$ satisfies (H1), (H2) and (H3),*

$$\lim_{\varepsilon\to 0} \varepsilon^2 N_\varepsilon \overset{\mathrm{d}}{=} \sup_{(s,f)\in\mathcal{F}'} |Z_P(s,f)|^2,$$

where $\mathcal{F}' := [0,1] \times \mathcal{F}$.

Proof

Let y be some positive number, and $m :=< y/\varepsilon^2 >$. It holds, setting $y_0 = m\varepsilon^2$,

$$\begin{aligned}
&P(\varepsilon^2 N_\varepsilon \geq y) = P(N_\varepsilon \geq m) \\
&= P\left(\sqrt{m} \sup_{n \geq m} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^{n} (f(X_i) - P(f) \right| \geq \sqrt{y_0}\right) \\
&= P\left(\sup_{f \in \mathcal{F}} \sup_{t \geq 1} \frac{1}{t} \left| \frac{1}{\sqrt{m}} \sum_{i=1}^{mt} (f(X_i) - Pf) \right| \geq \sqrt{y_0}\right) \\
&= P\left(\sup_{f \in \mathcal{F}} \sup_{t \geq 1} \frac{1}{t} \left|\mathbb{Z}_m(t,f)\right| \geq \sqrt{y_0}\right),
\end{aligned}$$

where $\mathbb{Z}_m$ is the sequential empirical process, an element of $\ell^\infty([0,\infty) \times \mathcal{F})$ defined through $\mathbb{Z}_m(t,f) := \frac{1}{\sqrt{m}} \sum_{i=1}^{mt} (f(X_i) - Pf)$. Note that for all fixed t, by (1), $\mathbb{Z}_m(t,.)$ belongs to $l^\infty(\mathcal{F})$, the set of all bounded sequences defined from $\mathcal{F}$ onto $\mathbb{R}$.

Following [Sheehy and Wellner, 1992], Theorem 11, for any $r > 0$, $\mathbb{Z}_m$ converges in distribution in $\ell^\infty([0,r] \times \mathcal{F})$ to Z_P. For all $r \geq 1$, it thus holds

$$\lim_{m \to \infty} \sup_{f \in \mathcal{F}} \sup_{1 \leq t \leq r} \left| \frac{\mathbb{Z}_m(t,f)}{t} \right| \stackrel{\text{d}}{=} \sup_{f \in \mathcal{F}} \sup_{1 \leq t \leq r} \left| \frac{Z_P(t,f)}{t} \right|.$$

Define $Z^*(\frac{1}{t}, f) := \frac{Z_P(t,f)}{t}$, for $t \in [1,r]$ and $f \in \mathcal{F}$. The centered Gaussian process Z^* is a Kiefer-Müller process indexed by $[1,r] \times \mathcal{L}^2(\mathcal{X})$; its covariance operator is defined, for f, g in $\mathcal{L}^2(\mathcal{X})$ and s, t in $[1,r]$, by

$$E\left(Z^*\left(\frac{1}{s}, f\right) Z^*\left(\frac{1}{t}, g\right)\right) = \left(\frac{1}{s} \wedge \frac{1}{t}\right) (Pfg - (Pf)(Pg)),$$

whence

$$\sup_{1 \leq t \leq r} \left| \frac{Z_P(t,f)}{t} \right| \stackrel{\text{d}}{=} \sup_{1 \leq t \leq r} \left| Z^*\left(\frac{1}{t}, f\right) \right| \stackrel{\text{d}}{=} \sup_{\frac{1}{r} \leq s \leq 1} |Z^*(s,f)|.$$

It follows by continuity that for any $r \geq 1$,

$$\lim_{m \to \infty} P\left(\sup_{f \in F} \sup_{1 \leq t \leq r} \frac{1}{t} \mathbb{Z}_m(t,f) \geq \sqrt{y_0}\right) = P\left(\sup_{\frac{1}{r} \leq s \leq 1} |Z^*(s,f)| \geq \sqrt{y_0}\right),$$

which, since y tends to 0 as $m \to \infty$ equals $P\left(\sup_{\frac{1}{r} \leq s \leq 1} |Z^*(s,f)|^2 \geq y\right)$.

In order to prove Proposition 2, it remains to prove that

$$\lim_{r \to \infty} \lim_{m \to \infty} P\left(\sup_{f \in \mathcal{F}} \sup_{t \geq r} \frac{1}{tm} \left| \sum_{i=1}^{mt} (f(X_i) - Pf) \right| \geq \sqrt{y_0}\right) = 0.$$

This follows , as (2), selecting some f in $\mathcal{F}$ and noting that
$\sup_{f \in \mathcal{F}} \sup_{t \geq r} \frac{1}{tm} \left| \sum_{i=1}^{mt} (f(X_i) - Pf) \right| \geq \frac{1}{[rm]} \left| \sum_{i=1}^{[rm]} (f(X_i) - Pf) \right|$. ■

3 Last exit times for adaptive estimates

Let $X_1^n := (X_1, \ldots, X_n)$ be an i.i.d. sample with X_1 distributed by $P_{\theta,g}$ on $\mathbb{R}^k$. The parameter of interest is θ, with $\theta \in \Theta$, an open set, an $g \in \mathcal{G}$, the set of nuisance parameters. A locally asymptotically linear estimator T_n of θ satisfies

$$\sqrt{n}\left(T_n - \theta_n - \frac{1}{n}\sum_{i=1}^{n} J(X_i, \theta_n, g)\right) = o_{\theta_n,g}(1). \tag{3}$$

In (3) (θ_n) is any sequence such that

$$\sqrt{n}(\theta_n - \theta) = O(1) \tag{4}$$

and J satisfies

$$\int J(x, \theta, g) dP_{\theta,g}(x) = 0 \tag{5}$$

together with

$$\int |J(x, \theta, g)|^2 dP_{\theta,g}(x) < \infty, \tag{6}$$

for all $\theta \in \Theta$ and $g \in \mathcal{G}$.

All the o and O notation are meant "in probability" where random variables are involved.

The function J in (3) is the influence function for T_n. An estimate S_n of θ is adaptive and efficient whenever there exists a function $J(x, \theta, g)$ such that, for all sequence (θ_n) satisfying (4), it holds

$$\lim_{n\to\infty} \sqrt{n}(S_n - \theta_n) \overset{\mathrm{d}}{=} N\left(0, \Sigma_{\theta,g}^{-1}\right) \tag{7}$$

where $\Sigma_{\theta,g}$ is the covariance matrix of $J(X, \theta, g)$ for fixed regular θ and g. When the J function coincides with the usual score function $\frac{\dot{f}(x,\theta,g)}{f(x,\theta,g)}$ with f the density of $P_{\theta,g}$, (7) is the classical normal limit behavior for ML estimates under contiguity of the sequence of measures $P_{\theta_n,g}$ to $P_{\theta,g}$ for all (θ_n) satisfying (7) and $g \in \mathcal{G}$, which we will assume from now on. Regularity of θ and g is defined in [Bickel, 1982].

Adaptive estimates are efficient for all $g \in \mathcal{G}$, even though the knowledge of g may not be used in the construction of the estimates. Under the above setting, constructions of efficient adaptive estimates have been proposed in [Beran, 1978], [Schick, 1986] and [Klaassen, 1987] among others. We will follow Klaassen's approach based on influence functions and provide some insight on last exit times for his estimates, strenghtening his assumptions when needed. We just need some kind of uniformity with respect to g as stated now.

Assume that there exists a sequence of functions $J_n(x, \theta_n, X_1^n)$ defined on $(\mathbb{R}^k, \Theta, \mathbb{R}^{kn})$ and a function $J(x, \theta, g)$ defined on $(\mathbb{R}^k, \Theta, \mathcal{G})$, such that

$$\sup_{g\in\mathcal{G}} \int J_n(x, \theta_n, X_1^n) dP_{\theta_n,g}(x) = o_{\theta_n}(1) \tag{8}$$

and

$$\sup_{g\in\mathcal{G}} \sqrt{n} \int [J_n(x,\theta_n,X_1^n) - J(x,\theta_n,g)]^2 df_{\theta_n g}(x) = o_{\theta_n}(1). \tag{9}$$

Assume further that we can construct a sequence of preliminary estimates S_n of θ such that

$$\sqrt{n}(S_n - \theta) = O_\theta(1). \tag{10}$$

Then we can construct a uniformly locally asymptotically linear adaptive estimate T_n of θ, satisfying therefore

$$\sup_{g\in\mathcal{G}} \sqrt{n} \left(T_n - \theta_n - \frac{1}{n}\sum_{i=1}^{n} J(X_i,\theta_n,g) \right) = o_{\theta_n}(1) \tag{11}$$

for all sequence (θ_n) satisfying (4). Display (11) proves that the estimate J_n of the Influence function J, together with an initial estimate of θ, provides explicit estimates of θ enjoying asymptotic normality and second order efficiency in the sense of Rao.

We now state additional conditions which entail some knowledge n the last exit time for the above adaptive estimate T_n. We assume that $J(x,\theta,g)$ is regular on a bounded open subset S of the image of X_1, say $I_m X_1$, uniformly upon θ and g. Also J is assumed to be constant outside S. Such conditions entail robustness for T_n. Precisely, assume

(H1) There exists $q > k/2$ such that $\sup_{\theta,g} \|J(\cdot,\theta,g)\|_{W_2^q} < \infty$, where $\|\cdot\|_{W_2^q}$ is the L_2–Sobolev norm of order q on S.
(H2) There exists $K > 0$ such that for all a in $I_m\ X_1 \setminus S$, for all (θ,g), $J(a,\theta,g) = K$.
(H3) $I_m\ X_1$ is convex or is a countable union of convex sets with non intersecting closures.

Under (H1), (H2) and (H3) the class $\mathcal{J}$ of functions $J(\cdot,\theta,g)$ is Donsker. When $I_m X_1 = [0,1]^k$, Theorem 7.7.1 in [Dudley, 1982], entails that for some $K_1 > 0$, for any $\varepsilon > 0$, the entropy number of $\mathcal{J}$ satisfies

$$\log N(\varepsilon, W_2^q, \mathcal{J}) \le K_1 e^{-k/q}.$$

Denote $N_\varepsilon := \sup\{n \ge 1 : |T_n - \theta| > \varepsilon\}$ where $|\cdot|$ is the Euclidian norm.

Applying Proposition 3 yields

Corollary *Under all the above assumptions, plus (H1), (H2) and (H3),*

(i) $\lim_{\varepsilon\to 0} \varepsilon^2 N_\varepsilon \overset{\mathrm{d}}{=} \sup_{\theta,g} \sup_{0\le s\le 1} |Z(s,J(\cdot,\theta,g))|^2$ *where* $Z(s,f)$ *is the functional Kiefer-Müller Process as defined in Section 1.*
(ii) *When* $I_m X_1 = [0,1]^k$*, then there exists some positive constant* K *such that for all* $\lambda > 0$*,*

$$\lim_{\varepsilon\to 0} P(\varepsilon^2 N_\varepsilon > \lambda) \le 2K(\sqrt{\lambda})^{1+k/q} \psi(\sqrt{\lambda}/\sigma_y)$$

where $\sigma_y^2 := \sup_{\theta,g} \int J^2(x,\theta,g) dP_{\theta,g}(x)$*.*

References

[Adler, 1990]R.J. Adler. *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*. Lecture Notes Monograph Series IMS, 1990.

[Barbe *et al.*, 1999]P. Barbe, M. Doisy, and B. Garel. Last passage time for the empirical mean of some mixing processes. *Stat. Prob. Letters*, 40(3):237–245, 1999.

[Beran, 1978]R. Beran. An efficient and robust adaptive estimator of location. *Annals of Statistics*, 6(2):293–313, 1978.

[Bickel, 1982]P.J. Bickel. On adaptive estimation. *Annals of Statistics*, 10(3):647–671, 1982.

[Bose and Chatterjee, 2001]A. Bose and S. Chatterjee. Last passage times of minimum contrast estimators. *Jour. Austral. Math. Soc.*, 71:1–10, 2001.

[Dudley, 1982]R.M. Dudley. *A Course on Empirical Processes*. Lecture Notes in Mathematics, Springer, 1982.

[Hjort and Fenstad, 1992]N.L. Hjort and G. Fenstad. On the last time and the number of times an estimator is more than ϵ from its large value. *Annals of Statistics*, 820(1):469–489, 1992.

[Klaassen, 1987]A.J. Klaassen. Consistent estimation of the influence function of locally asymtotically linear estimators. *Annals of Statistics*, 15(4):1548–1562, 1987.

[Lai, 1981]T.L. Lai. Asymptotic optimality of invariant sequential probability ratio tests. *Annals of Statistics*, 9(1):318–333, 1981.

[Schick, 1986]A. Schick. On asymptotically efficient estimation in semiparametric models. *Annals of Statistics*, 14(3):1139–1151, 1986.

[Sheehy and Wellner, 1992]A. Sheehy and J.A. Wellner. Uniform Donsker classes of functions. *Annals of Statistics*, 20(4):1983–2030, 1992.

[Stute, 1983]W. Stute. Last passage times of M-estimators. *Scand. Jour. Statistics*, 10:301–305, 1983.