A functional central limit theorem for the partial sums of sorted i.i.d. random variables

Jean-François Marckert  David Renault

CNRS, LaBRI, Université de Bordeaux
351 cours de la Libération
33405 Talence cedex, France
email: name@labri.fr

Abstract

Let $(X_i, i \geq 1)$ be a sequence of i.i.d. random variables with values in $[0,1]$, and $f$ be a function such that $\mathbb{E}(f(X_1)^2) < +\infty$. We show a functional central limit theorem for the process $t \mapsto \sum_{i=1}^{n} f(X_i)1_{X_i \leq t}$.

Keywords: Empirical process, Donsker class

AMS classification: 62G30

1 Introduction

Let $(X_1, X_2, \ldots)$ be a sequence of i.i.d. random variables (r.v.) with values in $[0,1]$, having distribution $\mu$, distribution function $F$, and defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which the expectation operator is denoted $\mathbb{E}$. In this paper we are interested in proving a functional limit theorem for the sequence of processes $(Z_n, n \geq 1)$ defined by

$$Z_n(t) := \frac{1}{n} \sum_{i=1}^{n} f(X_i)1_{X_i \leq t}, \quad t \in [0,1]$$

where $f : [0,1] \to \mathbb{R}$ is a measurable function. Let $(\hat{X}_i, 1 \leq i \leq n)$ be the sequence $(X_i, 1 \leq i \leq n)$ sorted in increasing order, and for any $t \in [0,1]$, denote by

$$N_n(t) = \# \{i : 1 \leq i \leq n, X_i \leq t \}$$

the number of $X_i$’s smaller than $t$. Clearly, for $t \in [0,1]$,

$$Z_n(t) = \frac{1}{n} \sum_{k=1}^{N_n(t)} f(\hat{X}_i).$$

Hence, $Z_n$ encodes the partial sums of functions of sorted i.i.d. r.v., as mentioned in the title of this paper. In order to state a central limit theorem for $Z_n$ the existence of $\text{Var}(f(X_1)) < +\infty$ is clearly needed, but it is not sufficient to control the fluctuations of $Z_n$ on all intervals. Standard
considerations about the binomial distribution implies that $N_n(t_2) - N_n(t_1)$ is quite concentrated around $n(F(t_2) - F(t_1))$ (for $t_1 < t_2$). Conditionally on $(N_n(t_1), N_n(t_2)) = (n_1, n_2)$,

$$Z_n(t_2) - Z_n(t_1) \overset{(d)}{=} \frac{1}{n} \sum_{k=1}^{n_2-n_1} f(X_{(t_1,t_2)}(k))$$

(2)

where $\overset{(d)}{=}$ means “equals in distribution”, and where $(X_{(t_1,t_2)}(k), 1 \leq k \leq n_2 - n_1)$ is a family of i.i.d. r.v., whose common distribution is that of $X$ conditional on $X \in (t_1, t_2]$. Hence, to get a functional central limit theorem for $Z_n$, the variances of these distributions need to be controlled.

The following hypothesis Hyp is designed for that purpose:

Hyp: there exists an increasing function $T : [0, 1) \to \mathbb{R}^+$ such that:

\[
\begin{align*}
T(x) \ln(x) &\xrightarrow{x \to 0} 0, \\
\forall I \text{ interval } &\subset [0; 1], \quad \text{Var} (f(X) | X \in I) \leq \frac{T(\mu(I))}{\mu(I)}
\end{align*}
\]

where $\text{Var}(g(X) | X \in I)$ denotes the variance of $g(X)$ conditional on $X \in I$ (by convention, we set $\mathbb{E}(g(X) | X \in I) = 0$ when $\mathbb{P}(X \in I) = 0$).

When $f$ is bounded by $\gamma$ on $[0; 1]$, the function $T(x) = \gamma^2 x$ satisfies Hyp (see also the discussion below Theorem 1).

Consider the mean of $Z_n$

$$Z(t) := \mathbb{E}(Z_n(t)) = \mathbb{E} (f(X)1_{X \leq t}), \quad (3)$$

(this can be shown to be a càdlàg process when $\mathbb{E}(|f(X)|) < +\infty$) and

$$Y_n(t) = \sqrt{n} [Z_n(t) - Z(t)]. \quad (4)$$

The aim of this paper is to show the following result:

**Theorem 1.** Let $(X_i, i \geq 0)$ be a sequence of i.i.d. r.v. taking their values in $[0, 1]$ and $f : [0, 1] \to \mathbb{R}$ a measurable function satisfying Hyp, then

$$Y_n \xrightarrow{\mathbb{D}} Y$$

in $D[0, 1]$, the space of càdlàg functions on $[0, 1]$ equipped with the Skorokhod topology, where $(Y_t, t \in [0, 1])$ is a centered Gaussian process with variance function

$$\text{Var}(Y_s) = F(s) \text{Var}(f(X) | X \leq s) + F(s)(1 - F(s)) \mathbb{E}(f(X) | X \leq s)^2$$

(5)

and with covariance function, for $0 \leq s < t \leq 1$

$$\text{Cov}(Y_s, Y_t - Y_s) = -F(s)(F(t) - F(s)) \mathbb{E}(f(X) | X \leq s) \mathbb{E}(f(X) | s < X \leq t).$$

(6)
We discuss a bit the conditions in the theorem. Assume that the \( X_i \)'s are i.i.d. uniform on \([0,1]\), and that \( f(x) = 1/x^\alpha \) for some \( \alpha > 0 \). The r.v. \( f(X) = 1/X^\alpha \) possesses a variance iff \( \alpha < 1/2 \), and then it is in the domain of attraction of the normal distribution only in this case (Theorem 1 needs this hypothesis for the convergence of \( Y_n(1) \)). The largest \( \text{Var}(f(X) | X \in (a, a + \varepsilon)) \) is obtained for \( a = 0 \), in which case we get

\[
\text{Var}(f(X) | X \in (0, \varepsilon)) = \frac{\varepsilon^{-2\alpha} \alpha^2}{(1 - 2\alpha)(1 - \alpha)^2},
\]

and one can check that \( \alpha < 1/2 \) is also the condition for the existence of a function \( T \) satisfying Hyp. Hyp appears to be a minimal assumption in that sense.

The first result concerning the convergence of empirical processes is due to Donsker’s Theorem \([2]\). It says that when \( f \) is constant equal to 1, then \( Y_n \) converges in \( D[0,1] \) to the standard Brownian bridge \( b \) up to a time change. A kind of miracle arises then, since the same analysis works for all distributions \( \mu \) by a simple time change. This is not the case here.

Apart from strong convergence theorems à la Komlós-Major-Tusnády \([4]\), modern results about the convergence of empirical processes – see Shorack & Wellner \([7]\) and van der Vaart & Wellner \([8]\) – much rely on the concept of Donsker classes, which we discuss below.

Denote by \( P_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k} \), the empirical measure associated with the sample \((X_i, 1 \leq i \leq n)\). As a measure, \( P_n \) operates on any set \( F \) of measurable functions \( \phi : [0,1] \to \mathbb{R} \),

\[
P_n \phi = \int_x \phi(x) dP_n(x) = \sum_{k=1}^{n} \phi(X_k)/n.
\]

The empirical process is the signed measure \( G_n := \sqrt{n}(P_n - \mu) \). By the standard central limit theorem, for a given function \( \phi \) (such that \( \mu \phi^2 < +\infty \)), \( G_n \phi \xrightarrow{d} \mathcal{N}(0, \mu(\phi - \mu \phi)^2) \), where \( \mathcal{N}(m, \sigma^2) \) designates the normal distribution with mean \( m \) and variance \( \sigma^2 \).

A \( P \)-Donsker class is a set of measurable functions \( F \) such that \( (G_n \phi, \phi \in F) \) converges in distribution to \( (G \phi, \phi \in F) \), in the \( L_\infty \) topology (it is a central limit theorem for a process index by a set of functions). This means that:

- the convergence of the finite dimensional distributions holds: (meaning that for any \( k \), any \( \phi_1, \ldots, \phi_k \in F \), \( G_n \phi_1, \ldots, G_n \phi_k \xrightarrow{d} \mathcal{N} \) \( (N_1, \ldots, N_k) \) and \( N \) is a centered Gaussian vector with covariance matrix \( \text{Cov}(N_i, N_j) = \mu [\langle \phi_i - \mu \phi_i, \phi_j - \mu \phi_j \rangle] \).
- the sequence \( (G_n \phi, \phi \in F) \) is tight in \( L_\infty \).

The proof that a set forms a Donsker class is usually not that simple, and numerous criteria can be found in the literature. In our case, the set of functions \( F \) is the following one:

\[
F_f = \{(x \mapsto \phi_t(x) = f(x)1_{x \leq t}, t \in [0,1])\}.
\]
We were unable to find such a criterion for this class, but notice that if such a result existed, it would imply Theorem 1 only for the topology $L_\infty$, a topology which is weaker than ours. Of course, Theorem 1 implies that $\mathcal{F}_f$ forms a Donsker class.

**Note.** In fact classes $\mathcal{F}_f$ for non decreasing $f$, or for functions $f$ whose level sets are given by two intervals at most (such that $x \mapsto x^2$, $x \mapsto \cos(2\pi x)$, $x \mapsto \sin(2\pi x)$) are Donsker, since they are VC subgraph class (see Vapnik & Chervonenkis [9]).

If we consider the variables $X_i$'s in the formula (1), as random times, then $Z_n(t)$ corresponds (up to the normalisation) to the sum of $f(X_i)$ for all events $X_i$ appearing before time $t$, where $f$ is some cost function. The process $Y_n$ appears to be the suitable tool to measure the fluctuations of $Z_n$.

We would like to mention [5], a work at the origin of the present paper, written by the same authors. In [5], the convergence of rescaled trajectories made with sorted increments (in $\mathbb{C}$) to a deterministic convex is shown. For this purpose a weaker version of Theorem 1 is established.

We provide a proof of our theorem in an old fashioned style. We prove the convergence of the finite dimensional distributions, and then establish the tightness in $D[0,1]$; even if the proof is a bit technical, we think that several tricks make it interesting in its own right.

## 2 Proof of Theorem 1

The proof starts with that of the convergence of the finite dimensional distributions (FDD) convergence of $Y_n$: this is classical as we will see. Let $\theta_0 := 0 < \theta_1 < \theta_2 < \cdots < \theta_K = 1$ for some $K \geq 1$ be fixed. In the sequel, for any function (random or not) $L$ indexed by $\theta$, $\Delta L(\theta_j)$ will stand for $L(\theta_j) - L(\theta_{j-1})$. For any $\ell \leq K$

$$
\Delta Y_n(\theta_\ell) = \sqrt{n} \left[ \Delta Z_n(N_n(\theta_j)) - \Delta Z(\theta_\ell) \right],
$$

where by convention $Z_n(N_n(\theta_{-1})) = Z(\theta_{-1}) = 0$. The convergence of the FDD of $Y_n$ follows the convergence in distribution of the increments $(\Delta Y_n(\theta_\ell), 0 \leq \ell \leq K)$. Notice that

$$
\Delta Z(\theta_j) = E \left( f(X)1_{\theta_{j-1} < X \leq \theta_j} \right).
$$

If for some $j$, $\theta_{j-1}$ and $\theta_j$ are chosen in such a way that $\Delta F(\theta_j) = 0$ then the $j$th increment in (7) is 0 almost surely (this is the case for the 0th increment if $\mu(\{0\}) = 0$). We now discuss the asymptotic behaviour of the other increments: let $J = \{ j \in \{0,\ldots,K\} : \Delta F(\theta_j) \neq 0 \}$.

Let $(n_j, j \in J)$ be some fixed integers summing to $n$. Denote by $\mu_{\theta_{j-1}, \theta_j}$ the law of $X$ conditioned by $\{ \theta_{j-1} < X \leq \theta_j \}$. Conditional on $(N_n(\theta_j) = n_j, j \in J)$, the variables $\Delta Z_n(N_n(\theta_j)), j \in J$ are independent, and $\Delta Z_n(N_n(\theta_j))$ is a sum of $n_j - n_{j-1}$ i.i.d. copies of variables under $\mu_{\theta_{j-1}, \theta_j}$, denoted from now on $(X_{\theta_{j-1}, \theta_j}(k), k \geq 1)$.

\footnote{We thank Emmanuel Rio for this information}
Since $(\Delta N_n(\theta_j), j \in J) \sim \text{Multinomial}(n, (\Delta F(\theta_j), j \in J))$,

$$\left(\frac{\Delta N_n(\theta_j) - n\Delta F(\theta_j)}{\sqrt{n}}, j \in J\right) \xrightarrow{d} (G_j, j \in J)$$

(9)

where $(G_j, j \in J)$ is a centered Gaussian vector with covariance function,

$$\text{cov}(G_k, G_\ell) = -\Delta F(\theta_k) \cdot \Delta F(\theta_\ell) + 1_{k=\ell} \Delta F(\theta_k),$$

formula valid for any $0 \leq k, \ell \leq K$. Putting together the previous considerations, we have

$$\Delta Y_n(\theta_j) = \sum_{m=1}^{\Delta N_n(\theta_j)} \frac{f(X_{\theta_j-\delta}(m)) - \mathbb{E}(f(X_{\theta_j-\delta}))}{\sqrt{n}}$$

$$+ \left(\frac{\Delta N_n(\theta_j) - n\Delta F(\theta_j)}{\sqrt{n}}\right) \mathbb{E}(f(X_{\theta_j-\delta})))$$

(10)

(11)

Using (9) and the central limit theorem, we then get that

$$\left(\Delta Y_n(\theta_j), 0 \leq j \leq K\right) \xrightarrow{d} \left(\sqrt{\Delta F(\theta_j) G_j + G_j \mathbb{E}(f(X_{\theta_j-\delta}))), 0 \leq j \leq K\right)$$

(12)

where the family of r.v. $(G_j, j \leq K)$ and $(\tilde{G}_j, j \leq K)$ are independent, and the r.v. $\tilde{G}_j$ are independent centered Gaussian r.v. with variance $\text{Var}(f(X_{\theta_j-\delta}))$ (this allows one to determine the variance and covariance (5) and (6)). Notice that here only the finiteness of $\text{Var}(f(X_{\theta_j-\delta}))$ and $\mathbb{E}(f(X_{\theta_j-\delta}))$ are used.

It remains to show the tightness of the sequence $(Y_n, n \geq 0)$ in $D[0,1]$. A criterion for the tightness in $D[0,1]$ can be found in Billingsley [1, Thm. 13.2]: a sequence of processes $(Y_n, n \geq 1)$ with values in $D[0,1]$ is tight if, for any $\varepsilon \in (0,1),$

$$\lim_{\delta \to 0} \limsup_n \mathbb{P}(\omega'(Y_n, \delta) \geq \varepsilon) = 0$$

where $\omega'(f, \delta) = \inf(t_i) \max_{s, t \in [t_i, t_{i+1})} |f(s) - f(t)|$, and the partitions $(t_i)$ range over all partitions of the form $0 = t_0 < t_1 < \cdots < t_n \leq 1$ with $\min(t_i - t_{i-1}, 1 \leq i \leq n) \geq \delta$.

We now compare our current model formed by a set $\{X_1, \ldots, X_n\}$ of $n$ i.i.d. copies of $X$ denoted from now on by $\mathbb{P}_n$, with a Poisson point process $\mathbb{P}_n$ on $[0,1]$ with intensity $n\mu$, denoted by $\mathbb{P}_{P_n}$. Conditionally on $\#P_n = k$, the $k$ points $P_n := \{X'_1, \ldots, X'_k\}$ are i.i.d. and have distribution $\mu$, and then $\mathbb{P}_{P_n}(\cdot | \#P = n) = \mathbb{P}_n$. The Poisson point process is naturally equipped with a filtration $\sigma := \{\sigma_t = \sigma(\{P \cap [0,t]\}, t \in [0,1])\}$.

We are here working under $\mathbb{P}_{P_n}$, and we let $N(\theta) = \#(P_n \cap [0,\theta])$; notice that under $\mathbb{P}_n$, $N$ and $N_n$ coincide.

Before starting, recall that if $N \sim \text{Poisson}(b)$, for any positive $\lambda$,

$$\mathbb{P}(N \geq x) = \mathbb{P}(e^{\lambda N} \geq e^{\lambda x}) \leq \mathbb{E}(e^{\lambda N - \lambda x}) = e^{-b + be^{-\lambda} - \lambda x}$$

(13)

$$\mathbb{P}(N \leq x) = \mathbb{P}(e^{-\lambda N} \geq e^{-\lambda x}) \leq \mathbb{E}(e^{-\lambda N + \lambda x}) = e^{-b - be^{-\lambda} + \lambda x}.$$
We explain now why the tightness of \((Y_n, n \geq 1)\) under \(\mathbb{P}_{P_n}\) implies the same result under \(\mathbb{P}_n\). Let \(m = \inf \{x \in [0,1], F(x) \geq 1/2\}\) be the median of \(\mu\).

**Lemma 2.1.** There exists a constant \(\gamma\) (which depends on \(\mu\)), such that for any \(\sigma_m\)-measurable event \(A\),

\[
\mathbb{P}_n(A) = \mathbb{P}_{P_n}(A | \#P = n) \leq \gamma \mathbb{P}_{P_n}(A). \tag{15}
\]

**Proof of the Lemma**

We have

\[
\mathbb{P}_{P_n}(A | \#P = n) = \sum_k \frac{\mathbb{P}_{P_n}(A, \#(P \cap [0,m]) = k) \mathbb{P}(\#P \cap [m,1] = n - k)}{\mathbb{P}(\#P = n)}
\leq \sum_k \mathbb{P}_{P_n}(A, \#(P \cap [0,m]) = k) \sup_{k'} \frac{\mathbb{P}(\#P \cap [m,1] = n - k')}{\mathbb{P}(\#P = n)}
\leq \gamma \mathbb{P}_{P_n}(A)
\]

where \(\gamma = \sup_{n \geq 1} \sup_{k'} \frac{\mathbb{P}(\#P \cap [m,1] = n - k')}{\mathbb{P}(\#P = n)}\), which is indeed finite since \(\mathbb{P}(\#P = n) \sim (2\pi n)^{-1/2}\), and since \(\#P \cap [m,1] \sim \text{Poisson}(n/2)\), and then the probability that its value is \(k\) is bounded above by some \(d/\sqrt{n}\) according to Petrov [6, Thm. 7 p. 48]. □

Thanks to Lemma 2.1, if the sequence of restrictions \((Y_n|[0,m], n \geq 1)\) of \(Y_n\) on \([0,m]\) is tight on \(D[0,m]\) under \(\mathbb{P}_{P_n}\) then so it is under \(\mathbb{P}_n\) (the same proof works on \(D[m,1]\) by a time reversal argument). To end the proof, we show that \((Y_n|[0,m], n \geq 1)\) is indeed tight under \(\mathbb{P}_{P_n}\).

Take then some (small) \(\eta \in (0,1), \varepsilon > 0\); we will show that one can find a finite partition \((t_i, i \in I)\) of \([0,m]\) and a \(\delta \in (0,m)\) such that

\[
\limsup_n \mathbb{P}_n(\omega'(Y_n, \delta) \geq \varepsilon) \leq \eta,
\tag{16}
\]

which is sufficient for our purpose.

We decompose the process \(Y_n\) as suggested by (10) and (11),

\[
Y_n(\theta) = Y'_n(\theta) + Y''_n(\theta) \tag{17}
\]

where

\[
Y'_n(\theta) = \sum_{m=1}^{N_n(\theta)} \frac{f(X_{[0,\theta]}(m)) - \mathbb{E}(f(X_{[0,\theta]}))}{\sqrt{n}}
\tag{18}
\]

\[
Y''_n(\theta) = \left(\frac{N_n(\theta) - nF(\theta)}{\sqrt{n}}\right) \frac{Z_\theta}{F(\theta)}. \tag{19}
\]

(If \(F(\theta)\) then set \(Y''_n(\theta) = 0\) instead of (19)).

The tightness of each of the sequences \((Y'_n, n \geq 1)\) and \((Y''_n, n \geq 1)\) in \(D[0,1]\) suffices to deduce that of \((Y_n, n \geq 1)\). We then proceed separately.
**Tightness of \((Y'_n, n \geq 1)\)**

To control the jumps of \(Y'_n\), we will need to localise the large atoms of \(\mu\). Let \(A = \{x \in [0, m], \mu(\{x\}) > 0\}\) be the set of positions of the atoms of \(\mu\) in \([0, m]\), and let \(\mathcal{A}^a := \{x \in A : \mu(\{x\}) \geq a\}\). Clearly \#\(\mathcal{A}^a \leq 1/a\) and \([0, m] \setminus \mathcal{A}^a\) forms a finite union of open connected intervals \((O_x, x \in G)\), with extremities \((t'_i, i \in I)\). The intervals \((O_x, x \in G)\) can be further cut as follows:

- do nothing to those such that \(\mu(O_x) < 2a\),
- those such that \(\mu(O_x) > 2a\) are further split. Since they contain no atom with mass \(\geq a\), they can be split into smaller intervals having all their weights in \([a, 2a]\) except for at most one (in each interval \(O_x\) which may have a weight smaller than \(a\)).

Once all these splittings have been done, a list of at most \(3/a\) intervals are obtained (in fact less than that), all of them having a weight smaller than \(2a\). Name \(G_a = (O_x, x \in I_a)\) the collection of obtained open intervals, indexed by some set \(I_a\), and by \((t'_i, i \geq 0)\) the partitions obtained. Take \(O\) one of these intervals. One has \#\((P_n \cap O)\) is Poisson with parameter \(n \mu(O) \leq na\). Consider again (10), (11) and Hyp. Set, for any \(L \geq 1\),

\[
S^{(n)}_L := \sum_{\ell=1}^L f(X_O(\ell)) - \mathbb{E}(f(X_O))
\]

Let

\[
\omega(Y'_n, O) = \sup\{|Y'_n(s) - Y'_n(t)|, s, t \in O\}
\]

be the modulus of continuity of \(Y'_n\) on \(O\). We have, for any \(\alpha \in (0, 1/2)\),

\[
\mathbb{P}(\omega(Y'_n, O) \geq x) \leq \mathbb{P}\left(\#(P_n \cap O) - n \mu(O) \geq n^{1/2+\alpha}\right)
+ \sup_{L \in \Gamma_n} \mathbb{P}\left(\sup \left\{|S^{(n)}_i - S^{(n)}_j|, i, j \leq L\right\} \geq x\right)
\]

(20)

where

\[
\Gamma_n = \left[\frac{n \mu(O) - n^{1/2+\alpha}, n \mu(O) + n^{1/2+\alpha}}{2}\right]
\]

Using (13) and (14), one sees that

\[
\mathbb{P}\left(\|P(n \mu(O)) - n \mu(O)\| \geq n^{\alpha+1/2}\right) \leq ce^{-c'n^\alpha}
\]

for some \(c > 0, c' > 0\) and \(n\) large enough (for this take \(x = n \mu(O) + n^{1/2+\alpha}, \lambda = 1/\sqrt{n}\) in (13) and, \(x = n \mu(O) - n^{1/2+\alpha}, \lambda = 1/\sqrt{n}\) in (14)).

Let us take care of the second term in (20). Clearly,

\[
\sup \left\{|S^{(n)}_i - S^{(n)}_j|, i, j \leq L\right\} = \max_{i \leq L} S^{(n)}_i - \min_{j \leq L} S^{(n)}_j
\]

According to Petrov [6, Thm.12 p50],

\[
\mathbb{P} \left(\max_{i \leq L} S^{(n)}_i \geq x\right) \leq 2\mathbb{P} \left(S^{(n)}_L \geq x - \sqrt{\frac{2L \text{Var}(f(X_O))}{n}}\right)
\]
and then
\[ P\left( \max_{i \leq L} S_i^{(n)} \geq x \right) \leq 2P\left( S_L^{(n)} \geq x - C_n(O) \right), \]
for \( C_n(O) = \sqrt{\frac{2LT(\mu(O))}{n\mu(O)}} \), and a similar inequality holds for \( \min_{i \leq L} S_i^{(n)} \). Since
\[ P\left( \max_{i \leq L} S_i^{(n)} - \min_{j \leq L} S_j^{(n)} \geq x \right) \leq P\left( \max_{i \leq L} S_i^{(n)} \geq x/2 \right) + P\left( -\min_{j \leq L} S_j^{(n)} \geq x/2 \right) \]
\[ \leq 2P\left( S_L^{(n)} \geq \frac{x}{2} - C_n(O) \right) + 2P\left( S_L^{(n)} \leq - \frac{x}{2} + C_n(O) \right). \]
To get some bounds, we use the central limit theorem for \( S_L^{(n)} \), and take \( x = \varepsilon, a > 0 \) such that 
\( T(a) = \varepsilon^2 \delta^2 \) for some small \( \delta > 0 \) (recall that \( T \) is increasing and therefore invertible), and any sequence \( L_n \) such that \( L_n/n \rightarrow \mu(O) \) (any sequence \( L = L_n \) such that \( L_n \in \Gamma_n \) satisfies this, and then we can control the supremum with this method). We have
\[ P\left( S_L^{(n)} \geq \varepsilon/2 - C_n(O) \right) = P\left( \frac{S_L^{(n)}}{\sqrt{\mu(O)\text{Var}(f(X_O))}} \geq \frac{\varepsilon/2 - C_n(O)}{\sqrt{\mu(O)\text{Var}(f(X_O))}} \right). \]
For \( n \) large enough,
\[ C_n(O) \leq \sqrt{4T(\mu(O))} \leq 2\varepsilon \delta \]
and therefore
\[ \limsup_n P\left( S_L^{(n)} \geq \frac{\varepsilon}{2} - C_n(O) \right) \leq \Phi \left( \frac{\varepsilon/2 - 2\varepsilon \delta}{\sqrt{\mu(O)\text{Var}(f(X_O))}} \right) \]
where \( \Phi \) is the tail function of the standard Gaussian distribution.

Finally, if \( \delta \) is chosen sufficiently small \( (2\delta < 1/2) \), since \( \mu(O)\text{Var}(f(X_O)) \leq T(\mu(O)) \leq T(a) = \varepsilon^2 \delta^2 \), then on each interval \( O \in G_a \),
\[ P\left( \sup \left\{ |S_i^{(n)} - S_j^{(n)}|, i, j \leq L \right\} \geq \varepsilon \right) \leq 4\Phi \left( \frac{1/2 - 2\delta}{\delta} \right) \]
and this independently of the choice of the interval \( O \) in \( G_a \), for \( n \) large enough.

The control of the intervals all together can be achieved using the union bound; since they are at most \( 3/T^{-1}(\varepsilon^2 \delta^2) \) such intervals, by the union bound
\[ P_{\omega_n} \left( \sup_{O \in G_a} \omega(Y_n^r, O) \geq \varepsilon \right) \leq \frac{3}{T^{-1}(\varepsilon^2 \delta^2)} \left( 4\Phi \left( \frac{1/2 - 2\delta}{\delta} \right) + ce^{-c\gamma n^\gamma} \right). \]
Since \( \Phi(x) \sim \exp(-x^2/2)/(\sqrt{2\pi}x) \), and \( T(x) \ln(x) \rightarrow 0 \), which implies that for any \( \varepsilon > 0 \), and \( \gamma > 0 \) there exists a \( \delta \) sufficiently small such that
\[ T(\varepsilon^{-\gamma/\delta^2}) < \varepsilon^2 \delta^2 \text{ or equivalently } \frac{1}{T^{-1}(\varepsilon^2 \delta^2)} < e^{\gamma/\delta^2} \]
and as a result the probability can be taken as small as wanted. \( \square \)
Tightness of \( (Y''_n, n \geq 1) \)

Recall (19). We work here under \( \mathbb{P}_n \) and we only consider the interval \( I = \{\theta : F(\theta) > 0\} \) since \( Y''_n(\theta) \) equals 0 on its complement. Since on \( I, \theta \mapsto \frac{Z_\theta}{F(\theta)} \) is càdlàg (and does not depend on \( n \)), it suffices to see why \( \left( \frac{N_n(\theta) - nF(\theta)}{\sqrt{n}}, n \geq 0 \right) \) is tight in \( D[0,1] \), but this is clear since this is a consequence of the convergence of the standard empirical process (Donsker [2]). \( \square \)

References

[1] P. Billingsley, (1999) *Convergence of probability measures*, Wiley Series in Probab. and Stat, John Wiley & Sons Inc.

[2] M.D. Donsker, (1952) *Justification and extension of Doob’s heuristic approach to the Kolmogorov-Smirnov theorems*, Annals of Mathematical Statistics., 23:277–281, 1952

[3] O. Kallenberg (1997), *Foundations of Modern Probability*. Probability and Its Applications. Springer, New York, NY.

[4] J. Komlós, P. Major, G. Tusnády, (1975) *An approximation of partial sums of independent RV’s and the sample DF*, I. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32, 111–131.

[5] J.F. Marckert, D. Renault (2012) *Compact convexes of the plane and probability theory*, preprint, arXiv:1208.5408.

[6] V.V. Petrov (1975) *Sums of independent random variables*, Springer-Verlag, New York.

[7] G.R. Shorack, J.A. Wellner, (1986) *Empirical processes with applications to statistics*, Wiley Series in Probab. and Math. Stat.: Probab. and Math. Stat. John Wiley & Sons, Inc., New York.

[8] A.W. van der Vaart, J.A. Wellner, Jon A. (1996) *Weak convergence and empirical processes. With applications to statistics*. Springer Series in Stat. Springer-Verlag, New York.

[9] V.N. Vapnik, A.YA. Chervonenkis (1971) *On the uniform convergence of relative frequencies of events to their probabilities*, Theory of probab. and appl., Vol. XVI, 2, 264–280.