Some asymptotic properties of the hybrids of empirical and partial-sum processes

Sergio Alvarez-Andrade

Abstract

The motivation of this paper is to study some properties of the local times (when it exists) of the hybrids of empirical and partial-sum processes defined by

\[ \bar{A}_n(t) = \sum_{1 \leq i \leq n} H(X_i)1_{\{X_i \leq t\}}\epsilon_i, \quad -\infty < t < \infty, \]

namely by using knowing results on empirical process and Brownian local times.

1. Introduction

We consider the process defined by (1.1) under some conditions. In the first section, we recall some properties of this process and we give some references about it. In the second section, we establish some results about the local times of this process. Let

\[ (1.1) \quad \bar{A}_n(t) = \sum_{1 \leq i \leq n} H(X_i)1_{\{X_i \leq t\}}\epsilon_i, \quad -\infty < t < \infty, \]

denote the hybrids of empirical and partial sums processes, under the following regularity conditions:

(A) the sequences of real random variables \( \{X_i\}_{i \geq 1} \) and \( \{\epsilon_i\}_{i \geq 1} \) are independent.

2000 Mathematics Subject Classification: Primary: 60J55, 60F15; Second.: 60F17, 62G30.
Keywords: Local times, compensated Poisson process, hybrids of empirical and partial-sum processes, Brownian motion.
(B) \{X_i, 1 \leq i < \infty\} are independent and identically distributed (i.i.d.)
random variables with common distribution function \(F\).

(C) \{\epsilon_i, 0 \leq i < \infty\} are i.i.d. random variables with
\(P(\epsilon_i = -1) = P(\epsilon_i = 1) = 1/2\).

(D) the function \(H\) is positive and has bounded variations on the real line.

We assume without loss of generality that all the random variables and
processes introduced so far and later on this paper can be defined on the
same probability space (cf. Appendix 2 in [5]). Throughout this paper we
denote by \(\log x = \log(x \vee \epsilon)\) and by \(\log n\) the \(n\)-times-iterated logarithm.

Our aim here is to establish a result of type
\[
\sup_{x,y \in \mathbb{R}: |x-y| \leq b_n} \left| \xi^*_t(A^*_n) - \tilde{\xi}^*_t(W_n(G_n)) \right| = O(n^{-\delta'(4+\delta')/4}) \log n \log 2 n, \text{ a.s.,}
\]
where \(\xi^*_t(A^*_n)\) is the local time of the normalized version of \(A_n\) (see Re-
mrk 1.1) and \(\tilde{\xi}^*_t(W_n(G_n))\) is the local time of \(W_n(G_n)\), where \(W_n\) is a
standard Brownian motion and \(G_n\) is a time transformation.

Diebolt et al. in [6], Diebolt in [7] and Diebolt and Laib in [8] studied the
process \(\{\bar{A}_n(t)\}\) under some more general conditions on the sequence \(\epsilon_i\), they
showed that \(A^*_n(t) = n^{-1/2} \bar{A}_n(t)\) converges weakly to a time-transformed
Wiener process (Brownian motion) and obtained upper bounds for the rate
of convergence.

The time transformation for the limiting Wiener process given in [7] is
\[
\bar{G}_n(t) = \int_{-\infty}^{t} H^2(s) dF_n(s),
\]
where
\[
F_n(t) = \frac{1}{n} \sum_{1 \leq i \leq n} 1_{\{X_i \leq t\}}, -\infty < t < \infty,
\]
is the empirical distribution function.

Later, Horváth in [10], showed that the random time change “\(\bar{G}_n(t)\)” can
be replaced with a non-random time change says
\[
(1.2) \quad \bar{G}(t) = \int_{-\infty}^{t} H^2(s) dF(s),
\]
where \(F\) is the common distribution function of the \(X_i\)'s, without reducing
the rate of the approximation given in [7]. He also gave the almost sure ap-
proximation of the two-parameter process \(\{A^*_n(t), -\infty < t < \infty, 1 \leq n < \infty\}\),
by a two-parameter Wiener process.
The case $H \equiv 1$ has been studied for instance (see also references therein) in Heusler and Mason [9] where they defined the “randomly weighted empirical process” associated with $\bar{A}_n(t)$. Horváth, in [10], showed how to detect a possible change in the distribution of independent observations based on approximations for the particular bootstrapped empirical process based on independent weights (see also [11]). Some statistical applications were developed in [16, p. 398] (see also [6]).

Remark 1.1 Without loss of generality, there are independent, identically distributed random variables $\{Y_i, 1 \leq i < \infty \}$ uniform on $[0, 1]$ such that

$$X_i = Q(Y_i),$$

with $Q(y) = \inf \{x : F(x) \geq y \}$ i.e., the quantile function of $F$ (see [10, p. 5]). Then, we can consider

$$A_n(t) = \sum_{1 \leq i \leq n} V(Y_i)1_{\{Y_i \leq t\}} \epsilon_i, \quad 0 \leq t \leq 1,$$

in the place of $\bar{A}_n(t)$, because

$$\bar{A}_n(t) = \sum_{1 \leq i \leq n} H(Q(Y_i))1_{\{Q(Y_i) \leq t\}} \epsilon_i = \sum_{1 \leq i \leq n} V(Y_i)1_{\{Y_i \leq F(t)\}} \epsilon_i,$$

holds with $V = H \circ Q$. Therefore, by (D), we can assume without loss of generality that

$$\sup_{t \in [0,1]} |V(t)| \leq 1.$$

As a consequences of Remark 1.1, we have that $A^*_n(t)$ is of the form $A^*_n(t) = A_n(t)/\sqrt{n}$, where we replace $\bar{A}_n(t)$ by $A_n(t)$ without loss of generality and we can replace the times changes (corresponding to $\bar{G}_n(t)$ and $G(t)$) by

$$G_n(t) = \int_0^t V^2(s) dE_n(s) \quad \text{and} \quad G(t) = \int_0^t V^2(s) ds, \quad t \in [0, 1].$$

where

$$E_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i \leq t\}}, \quad t \in [0, 1]$$

denotes the uniform empirical distribution function related to the i.i.d. sequence $\{Y_i, 1 \leq i < \infty \}$ i.i.d. of uniform distribution on $[0, 1]$. 
2. Local times results

2.1. Preliminaries and result

We begin by recalling the following definition given in [14].

**Definition 1** For any stochastic process $X = \{X(t); t \geq 0\}$, define (when it exists) the crossing process of $X$ as

$$C(X) = \{C^r(X); (x,t) \in \mathbb{R} \times [0,1]\},$$

with $C^r_t(x) = \#\{s \leq t, X(s) = x\}$, where $\#A$ denotes the cardinality of the set $A$. Define (when it exists) the local time of $X$ as

$$L(X) = \{L^r(x,t); (x,t) \in \mathbb{R} \times [0,1]\},$$

where $L^r(x,t) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^1 \mathbf{1}_{\{x-\epsilon \leq X(s) \leq x+\epsilon\}} ds$ a.s.

Define the empirical uniform process associated to the sequence $\{Y_i, 1 \leq i < \infty\}$ of i.i.d. random variables uniformly distributed on the interval $[0,1]$ by

$$\alpha_n(t) = \sqrt{n}(E_n(t) - t), \quad 0 \leq t \leq 1,$$

where $E_n(t)$ is given by (1.5). The crossing process related to $\alpha_n(t)$ is given by

$$C^r_t(\alpha_n) = \#\{s \leq t; \alpha_n(s) = x\}$$

and by Lemma 1 of [15], we obtain the local time at a level $x$ up to time $t$ by

$$L^r_t(\alpha_n) = \frac{C^r_t(\alpha_n)}{\sqrt{n}}.$$

Let us define the local time of $A^*_n(t) = \sum_{1 \leq i \leq n} V(Y_i) \mathbf{1}_{\{Y_i \leq t\}} \epsilon_i / \sqrt{n}$ by

$$\xi^r_t(A^*_n) = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} \mathbf{1}_{\{A^*_n(s) = x\}}, \quad t \in [0,1], x \in \mathbb{R}.$$

and in the light of the strong approximation given by Diebolt in [7], we define the following local time

$$\hat{\xi}^r_t(W_n(G_n)) = \int_0^1 \eta^r_s(G_n) d_s L^r_s(W_n),$$

where $W_n(t) = W(nt) / \sqrt{n}$ with $W$ is a standard Wiener process and $\eta^r_s(G_n)$ denoting the local time at a level $s$ up to $t$ of the process $G_n(t)$ (see (2.7) below). Our aim is now to establish the following result.
Proposition 1 Assume that conditions (A), (B), (C) and (D) are fulfilled, and that $A_n^*$ admits a local time $\xi_t^*(A_n^*)$ and define $\xi_t^*(A_n^*) = \sup_{t \in [0,1]} \xi_t^*(A_n^*)$.

Then on a suitable probability space on which one can define a sequence of Brownian motions $W_n$ and a sequence $A_n^*$, we have

$$\sup_{x,y \in \mathbb{R}} |\xi_t^*(A_n^*) - \tilde{\xi}_t^*(W_n(G_n))| = O(n^{-\delta/((4+\delta')}}) \log n \log_2 n), \text{ a.s.}$$

where $b_n = n^{-1/(4+\delta)}$, for some constant $\delta' > 0$.

The proof of this result will be given in the next subsection, and it will be obtained as a consequence of three lemmas. In the following lines we give some remarks about the local times $\xi_t^*(A_n^*)$ and we recall some important results about the local time of the empirical process and the Brownian process.

Putting $U_n(t) = \sum_{i=1}^{[nt]} \epsilon_i/\sqrt{n}$ and $M_n(t) = \sum_{1 \leq i \leq n} V(Y_i)1_{\{Y_i \leq t\}}/n$, $t \in [0,1]$, then we have $A_n^* \overset{d}{=} U_n(M_n(t))$. By this last equality (in distribution) and from the definition of the local time $\xi_t^*(A_n^*)$, we have $U_n(M_n(t)) = x$ if $M_n(t)$ is the return time to $x$ of $U_n$, for $x = 0, \pm 1, \pm 2, \ldots$. Consequently, $\xi_t^*(A_n^*)$ can be written as

$$\xi_t^*(A_n^*) = \int_0^1 \delta_t^*(M_n) d\nu_t^*(U_n),$$

for $0 \leq t \leq 1$ and $x \in \mathbb{R}$, where $\delta_t^*(M_n)$ is the local time of $M_n(t)$ (see (2.8) below) and $\nu_t^*(U_n)$ is the local time of $U_n(t)$ (see (2.6) below).

We recall an other important result about local time for $\alpha_n(t)$, given by Theorem 2 of [15].

Theorem 1 On a probability space, $(\Omega, \mathcal{F}, \mathcal{P})$, there exists a sequence of standard Brownian bridges, $\{B_n(t); 0 \leq t \leq 1\}_{n \geq 1}$, and a sequence of uniform empirical processes, $\{\alpha_n(t); 0 \leq t \leq 1\}_{n \geq 1}$, such that for all $\epsilon > 0$,

$$\sup_{x \in \mathbb{R}} \sup_{0 \leq t \leq 1} \left| L_t^x(\alpha_n) - L_t^x(B_n) \right| = o\left(n^{-1/4} \log n^{3/4+\epsilon}\right) \text{ a.s.}$$

Finally, we recall some important facts related to the local time of Brownian motion.

Let $\{W(t), t \geq 0\}$ be a one-dimensional Brownian motion with $W(0) = 0$ and let $\{L_t^x(W), t \geq 0, x \in \mathbb{R}\}$ denote its local time process. That is, for any Borel function $f \geq 0$,

$$\int_0^t f(W(s)) \, ds = \int_{-\infty}^\infty f(x)L_t^x(W) \, dx, \quad (t \geq 0),$$
finally we can define \( \{L_t^x(W), \ t \geq 0, x \in \mathbb{R}\} \) by

\[
L_t^x(W) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^t 1_{\{x \leq W(s) < x+\epsilon\}} ds,
\]

see for instance [13, p.1295].

Remember that by the scaling property of Brownian motion and putting
\( W_n(t) = W(nt)/\sqrt{n} \), we have

\[
\frac{1}{\sqrt{n}} L_{nt}^{\sqrt{n}x}(W_n) = L_t^x(W),
\]

where \( L_{nt}^{\sqrt{n}x}(W_n) \) is the local time of \( W(nt) \) (see for instance [18]).

In the next subsection we give the proof of Proposition 2.1 and define and study other local times related to our processes.

### 2.2. Proof of Proposition 1

We have to deal with the following processes and their corresponding local times.

For the process \( U_n(t) \) related to the random walk given by the random variables \( \epsilon_i \)'s, we define the corresponding local time by

\[
\nu(x', \tau)(U_n) = \#\left\{ j : 0 \leq j \leq \tau, U_n(j) = x' \right\}
= \#\left\{ j : 0 \leq j \leq \lfloor nt \rfloor, \sum_{i=1}^j \epsilon_i = \lfloor \sqrt{n}x \rfloor \right\}, \ x \in \mathbb{R}.
\]

We refer for instance to [3] or to [17] among others.

Let \( \eta_t^x \) be the local time at a level \( x \) up to \( t \) of

\[
G_n(t) = \int_0^t V^2(s) dE_n(s),
\]

i.e.

\[
\eta_t^x(G_n) = \frac{1}{\sqrt{n}} \sum_{s \leq t} 1_{\{\int_s^t V^2(s) dE_n(s) = x\}},
\]

where \( E_n(t) \), is the uniform empirical distribution function (see (1.5)).

Finally, the local time for \( M_n(t) \) is defined by

\[
\delta_t^x(M_n) = \frac{1}{\sqrt{n}} \sum_{s \leq t} 1_{\{M_n(s) = x\}}, \ t \in [0, 1], x \in \mathbb{R}.
\]
Remark 2.1 Our definition of the local times in (2.7) and in (2.8) follows the definition given for the local time of the compensated compound Poisson process \( \{Z(t)\} \) given in [14, (1.1)]. More precisely by [15, (3.1)], we have that

\[
\{\alpha_n(t); \ t \geq 0\} \equiv \{Z_n(t); \ t \geq 0 \mid Z_n(1) = 0\},
\]

where \( \alpha_n(t) \) is the uniform empirical process and \( Z_n(t) = Z(nt)/\sqrt{n} \) with \( \{Z_n(t)\}_{n \geq 1} \) is a sequence of compensated Poisson process with expected arrival rate of \( 1/n \). Remark that \( \{Z_n(1) = 0\} \equiv \{N(n) = n\} \) for \( N(n) \) a Poisson random variable with mean \( n \). Roughly speaking for study (2.7), in the light of crossing comparison (see [15, p. 339]) we can use

\[
1 \{\int_0^t V^2(s)d\alpha_n(s) = \int_0^t V^2(s)dZ_n(s)|Z_n(1) = 0\} \Leftrightarrow 1 \{\int_0^t V^2(s)dE_n(s) = \int_0^t V^2(s)dN_n(s)|N_n(n) = n\}.
\]

The proof of proposition 1 will be based on the following lemmas.

**Lemma 2.1** We assume that (A), (B), (C) and (D) hold. Then we can define a sequence of Wiener process \( \{W_n(x) = W(nx)/\sqrt{n}, \ 1 \leq n < \infty\} \), where \( W \) is a standard Wiener process such that \( \{W_n(x), \ 0 \leq x < \infty, \ 1 \leq n < \infty\} \) and \( \{Y_i, \ 1 \leq i < \infty\} \) are independent and

\[
\sup_{t \in [0,1]} \left| A^*_n(t) - W_n(G_n(t)) \right| = O \left( \log n / \sqrt{n} \right) \text{ a.s.}
\]

**Proof.** This lemma is a direct consequence of Theorem 1 in [7] (see also Theorem A in [10]).

**Lemma 2.2** Under the same conditions of Proposition 1 and for \( \varepsilon' > 0 \), we have almost surely

\[
\sup_{x, y \in \mathbb{R}, |x - y| \leq \varepsilon} \left| \xi^*_1(A_n^*) - \xi^*_1(A_n^*) \right| = O \left( \log n (\log(1/\varepsilon))^{1/2} \right)
\]

\[
+ o \left( n^{-(1/4) + \varepsilon'} \sqrt{\log_2 n} \right).
\]

**Proof.** From (2.4), we have

\[
\sup_{x, y \in \mathbb{R}, |x - y| \leq \varepsilon} \left| \xi^*_1(A_n^*) - \xi^*_1(A_n^*) \right| \leq \sup_{x, y \in \mathbb{R}, |x - y| \leq \varepsilon} \left| \int_0^1 \delta^*_1(M_n) d_s (\nu^*_x(U_n) - \nu^*_y(U_n)) \right|.
\]
This and a result of [15, p. 336] and by integration by parts in the Stieltjes sense and by using the fact that \( \nu_0'(U_n) = 0 \), implies that for all \( \varepsilon' > 0 \), the following occurs with probability one:

\[
\sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} |\xi_t^x(A_n^*) - \xi_t^y(A_n^*)| \\
\leq \sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| \int_0^1 \delta_t^x(M_n)ds (L_s^x(W_n) - L_s^y(W_n)) \right| \\
+ o\left(n^{-1/4} + \varepsilon' \int_0^1 \delta_t^x(M_n)ds \right),
\]

(2.13)

where the \( o() \) term on the inequality is obtained by using the estimation of the distance between the local time of a random walk and the local time of a Brownian motion, see for instance (10.1) of [17]) and by a scale change. We will see that in fact in the place of (2.13), we have

\[
\sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| \int_0^1 \delta_t^x(M_n)ds (L_s^x(W_n) - L_s^y(W_n)) \right| \\
+ o\left(n^{-1/4} + \varepsilon' \sqrt{\log_2 n} \right),
\]

(2.14)

where the \( \sqrt{\log_2 n} \) term in the right hand side is given by the law of iterated logarithm (LIL), see (2.15) below. Remark that in the last equality the little-o terms goes to zero (a.s.) as \( n \to \infty \) uniformly in \( \varepsilon \). In order to supplement the proof we establish the following Facts:

**Fact 1.** Put \( \bar{M}_n(t) = \int_0^t V(\alpha)d\alpha Z_n(\alpha) | Z_n(1) = 0 \), then in the light of crossing comparison and (2.9), we have \( \delta_t^s(M_n) \equiv \delta_t^{s'}(M_n) \), where \( s' = s - \sqrt{n} \int_0^s V(\alpha)d\alpha \), moreover by [4, p. 1051] (see also [1]), we obtains almost surely that

\[
\sup_{s' \in \mathbb{R}} \left| \delta_t^{s'}(M_n) - L_1^{s'}(W_{1n}) \right| = o\left(n^{-\nu}\right), \quad (0 < \nu < 1/4),
\]

where \( W_{1n}(t) = W_n \left( \int_0^t V^2(\alpha)d\alpha \right) \) where we have used (3.9) of [10], moreover by the LIL for \( L_1^{s'}(W_{1n}) \) i.e.

\[
\limsup_{n \to \infty} \frac{L_1^{s'}(W_{1n})}{\sqrt{\log_2 n}} = \limsup_{n \to \infty} \frac{L_1^{0}(W_{1n})}{\sqrt{\log_2 n}} = \sqrt{2} \text{ a.s.,}
\]

see [12], we have

\[
\delta_t^s(M_n) = O\left((\log_2 n)^{1/2}\right) + o\left(n^{-\nu}\right) = O\left((\log_2 n)^{1/2}\right) \text{ a.s.}
\]

(2.15)
Fact 2. By Fact 1, we have that (2.13) can be replaced by (2.14), because \( \int_0^1 \delta^1_s(M_n)ds = O(\sqrt{\log_2 n}) \) a.s. Moreover, we have that (2.14) is with probability 1 a
\[
O\left( (\log_2 n)^{1/2} \sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| \int_0^1 d_s(L^x_s(W_n) - L^y_s(W_n)) \right| \right) + o\left( n^{-(1/4) + \varepsilon'} \sqrt{\log_2 n} \right),
\]
by using \( L^x_0(W_n) = 0 \) and by (1.3) of [13] (for instance), we have that (2.14) is with probability 1 a
\[
O\left( \log_2 n(\varepsilon \log(1/\varepsilon)^{1/2}) \right) + o\left( n^{-(1/4) + \varepsilon'} \sqrt{\log_2 n} \right).
\]
Fact 1 and Fact 2 give the announced result.

The following lemma does not give the optimal rate but gives sufficient condition in our case (in the light of the precedent two Lemmas), see remark at the end of the proof if we want a best rate of convergence.

Lemma 2.3 For the sequence \( \{W_n\}_{n \geq 1} \) of Brownian motion and a nonrandom \( \varepsilon > 0 \), we have
\[
\sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| \int_0^1 \eta^x_s(G_n)ds \left( L^x_s(W_n) - L^y_s(W_n) \right) \right| \leq c' \varepsilon^{1/2 - \delta} \log n \left( \log_2 n \right)^{1/2} a.s.
\]
where \( c' \) is a constant independent of \( n \), and for some \( 0 < \delta < 1/2 \).

Proof.
\[
\sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| \int_0^1 \eta^x_s(G_n)ds \left( L^x_s(W_n) - L^y_s(W_n) \right) \right| = O \left( \sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| L^x_s(W_n) - L^y_s(W_n) \right| \left( 1 + \int_0^1 \eta^x_s(G_n)ds \right) \right),
\]
the equality is obtained by integration by parts. As a consequence of the crossing comparison and with the similar arguments as in the proof of Lemma 2.2, we have that the only term that we must study is
\[
\sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \sup_{0 \leq s \leq 1} \left| L^x_s(W_n) - L^y_s(W_n) \right|.
\]
But, this term is almost surely upper bounded by \( c \varepsilon^{1/2 - \delta} \log n \), where \( c \) is a constant independent of \( n \), this estimates is obtained by using [2, (3.7)].
Remark 2.2  We claim that in order to have a best rate of convergence in Lemma 2.3, we could utilize in the same way as in [1, p. 393], knowing that

\[ P \left( \sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} |L_t^x(W_n) - L_t^y(W_n)| \geq 2 \sqrt{\varepsilon \log \left( \frac{1}{\varepsilon} \right) + \sqrt{\varepsilon \lambda}} \right) \leq c_1 \exp\left( -\lambda^2 / c_1 \right), \]

where \( c_1 \) is a constant and \( \lambda > 0 \).

Proof of Proposition.

We want to evaluate

\[ \Delta_n = \sup_{x \in \mathbb{R}} \left| \xi_t^x(A_n^*) - \tilde{\xi}_t^x(W_n(G_n)) \right| \]

in the light of Lemma 3.1 in [1] (see also Fact 7.2 in [4, p. 1052]), we have

\[ \Delta_n \leq \sup_{0 \leq t \leq 1} \frac{|A_n^*(t) - W_n(G_n(t))|}{\varepsilon^2} + \sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} |\xi_t^x(A_n^*) - \xi_t^y(A_n^*)| + \sup_{x,y \in \mathbb{R}, |x-y| \leq \varepsilon} \left| L_t^x(W_n(G_n)) - L_t^y(W_n(G_n)) \right|. \]

By the previous Lemma’s, and replacing \( \varepsilon \) by \( b_n \), we have the announced result.

Acknowledgment. I am indebted to an anonymous referee for his helpful comments and corrections.

References


Recibido: 20 de septiembre de 2005
Revisado: 7 de febrero de 2007

Sergio Alvarez-Andrade
Laboratoire de Mathématiques Appliquées de Compiègne
Université de Technologie de Compiègne
B.P. 529, F-60205 Compiègne Cedex, France
Sergio.Alvarez@utc.fr