Article

Convergence in Total Variation to a Mixture of Gaussian Laws

Luca Pratelli [1] and Pietro Rigo [2,*

[1]   Accademia Navale, Viale Italia 72, 57100 Livorno, Italy; pratel@mail.dm.unipi.it
[2]   Dipartimento di Matematica "F. Casorati", Universita' di Pavia, via Ferrata 1, 27100 Pavia, Italy
*   Correspondence: pietro.rigo@unipv.it

Abstract: It is not unusual that $X_n \xrightarrow{dist} VZ$ where X_n, V, Z are real random variables, V is independent of Z and $Z \sim \mathcal{N}(0,1)$. An intriguing feature is that $P(VZ \in A) = E\left\{\mathcal{N}(0, V^2)(A)\right\}$ for each Borel set $A \subset \mathbb{R}$, namely, the probability distribution of the limit VZ is a mixture of centered Gaussian laws with (random) variance V^2. In this paper, conditions for $d_{TV}(X_n, VZ) \to 0$ are given, where $d_{TV}(X_n, VZ)$ is the total variation distance between the probability distributions of X_n and VZ. To estimate the rate of convergence, a few upper bounds for $d_{TV}(X_n, VZ)$ are given as well. Special attention is paid to the following two cases: (i) X_n is a linear combination of the squares of Gaussian random variables; and (ii) X_n is related to the weighted quadratic variations of two independent Brownian motions.

1. Introduction

All random elements involved in the sequel are defined on a common probability space $(\Omega, \mathcal{F}, P)$. We let $\mathcal{B}$ denote the Borel σ-field on $\mathbb{R}$ and $\mathcal{N}(a, b)$ the Gaussian law on $\mathcal{B}$ with mean a and variance b, where $a \in \mathbb{R}$, $b \geq 0$, and $\mathcal{N}(a, 0) = \delta_a$. Moreover, Z always denotes a real random variable such that:

$$Z \sim \mathcal{N}(0, 1).$$

In plenty of frameworks, it happens that:

$$X_n \xrightarrow{dist} VZ, \tag{1}$$

where X_n and V are real random variables and V is independent of Z. Condition (1) actually occurs in the CLT, both in its classical form (with $V = 1$) and in its exchangeable and martingale versions (Examples 3 and 4). In addition, condition (1) arises in several recent papers with various distributions for V. See, e.g., [1–8].

An intriguing feature of condition (1) is that the probability distribution of the limit:

$$P(VZ \in A) = \int \mathcal{N}(0, V^2)(A) \, dP, \quad A \in \mathcal{B},$$

is a mixture of centered Gaussian laws with (random) variance V^2. Moreover, condition (1) can be often strengthened into:

$$d_W(X_n, VZ) \to 0, \tag{2}$$

where $d_W(X_n, VZ)$ is the Wasserstein distance between the probability distributions of X_n and VZ. In fact, condition (2) amounts to (1) provided the sequence (X_n) is uniformly integrable; see Section 2.1.

A few (engaging) problems are suggested by conditions (1) and (2). One is:

(*) Give conditions for $d_{TV}(X_n, VZ) \to 0$, where

$$d_{TV}(X_n, VZ) = \sup_{A \in \mathcal{B}} |P(X_n \in A) - P(VZ \in A)|.$$

Under such (or stronger) conditions, estimate the rate of convergence, i.e., find quantitative bounds for $d_{TV}(X_n, VZ)$.

Problem (*) is addressed in this paper. Before turning to results, however, we mention an example.

Example 1. *Let B be a fractional Brownian motion with Hurst parameter H and*

$$X_n = \frac{n^{1+H}}{2} \int_0^1 t^{n-1}(B_1^2 - B_t^2)\, dt.$$

The asymptotics of X_n and other analogous functionals of the B-paths (such as weighted power variations) is investigated in various papers. See, e.g., [5,7–10] and references therein. We note also that:

$$\int_0^1 t^n B_t\, dB_t = \frac{X_n}{n^H} - \frac{H}{2H + n} \quad \text{for each } H \geq 1/4,$$

where the stochastic integral is meant in Skorohod's sense (it reduces to an Ito integral if $H = 1/2$).

Let $a(H) = 1/2 - |1/2 - H|$ and $V = \sqrt{H\Gamma(2H)}\, B_1 \sim \mathcal{N}(0, H\Gamma(2H))$. In [8], it is shown that, for every $\beta \in (0,1)$, there is a constant k (depending on H and β only) such that:

$$d_{TV}(X_n, VZ) \leq k\, n^{-\beta\, a(H)} \quad \text{for all } n \geq 1,$$

where Z is a standard normal random variable independent of V. Furthermore, the rate $n^{-\beta\, a(H)}$ is quite close to be optimal; see condition (2) of [8].

In Example 1, problem (*) admits a reasonable solution. In fact, in a sense, Example 1 is our motivating example.

This paper includes two main results.

The first (Theorem 1) is of the general type. Suppose $l_n := \int |t\, \phi_n(t)|\, dt < \infty$, where ϕ_n is the characteristic function of X_n. (In particular, X_n has an absolutely continuous distribution). Then, an upper bound for $d_{TV}(X_n, VZ)$ is provided in terms of l_n and $d_W(X_n, VZ)$. In some cases, this bound allows to prove $d_{TV}(X_n, VZ) \to 0$ and to estimate the convergence rate. In Example 5, for instance, such a bound improves on the existing ones; see Theorem 3.1 of [6] and Remark 3.5 of [7]. However, for the upper bound to work, one needs information on l_n and $d_W(X_n, VZ)$, which is not always available. Thus, it is convenient to have some further tools.

In the second result (Theorem 2), the ideas underlying Example 1 are adapted to weighted quadratic variations; see [5,8,9]. Let B and B' be independent standard Brownian motions and

$$X_n = n^{1/2} \sum_{k=0}^{n-1} f(B_{k/n} - B'_{k/n}) \{(\Delta B_{k/n})^2 - (\Delta B'_{k/n})^2\},$$

where $f : \mathbb{R} \to \mathbb{R}$ is a suitable function, $\Delta B_{k/n} = B_{(k+1)/n} - B_{k/n}$ and $\Delta B'_{k/n} = B'_{(k+1)/n} - B'_{k/n}$. Under some assumptions on f (weaker than those usually requested in similar problems), it is shown that

$d_{TV}(X_n, VZ) = O(n^{-1/4})$, where $V = 2\sqrt{\int_0^1 f^2(\sqrt{2}B_t)\,dt}$. Furthermore, $d_{TV}(X_n, VZ) = O(n^{-1/2})$ if one also assumes $\inf|f| > 0$. (We recall that, if a_n and b_n are non-negative numbers, the notation $a_n = O(b_n)$ means that there is a constant c such that $a_n \le c\,b_n$ for all n).

2. Preliminaries

2.1. Distances between Probability Measures

In this subsection, we recall a few known facts on distances between probability measures. We denote by $(S, \mathcal{E})$ a measurable space and by μ and ν two probability measures on $\mathcal{E}$.

The total variation distance between μ and ν is:

$$\|\mu - \nu\| = \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

If X and Y are $(S, \mathcal{E})$-valued random variables, we also write:

$$d_{TV}(X, Y) = \|P(X \in \cdot) - P(Y \in \cdot)\| = \sup_{A \in \mathcal{E}} |P(X \in A) - P(Y \in A)|$$

to denote the total variation distance between the probability distributions of X and Y.

Next, suppose S is a separable metric space, $\mathcal{E}$ the Borel σ-field and

$$\int d(x, x_0)\,\mu(dx) + \int d(x, x_0)\,\nu(dx) < \infty \quad \text{for some } x_0 \in S,$$

where d is the distance on S. The Wasserstein distance between μ and ν is:

$$W(\mu, \nu) = \inf_{X \sim \mu, Y \sim \nu} E[d(X, Y)],$$

where inf is over the pairs (X, Y) of $(S, \mathcal{E})$-valued random variables such that $X \sim \mu$ and $Y \sim \nu$. By a duality theorem, $W(\mu, \nu)$ admits the representation:

$$W(\mu, \nu) = \sup_f \left| \int f\,d\mu - \int f\,d\nu \right|,$$

where sup is over those functions $f : S \to \mathbb{R}$ such that $|f(x) - f(y)| \le d(x, y)$ for all $x, y \in S$; see, e.g., Section 11.8 of [11]. Again, if X and Y are $(S, \mathcal{E})$-valued random variables, we write:

$$d_W(X, Y) = W\big[P(X \in \cdot),\, P(Y \in \cdot)\big]$$

to mean the Wasserstein distance between the probability distributions of X and Y.

Finally, we make precise the connections between convergence in distribution and convergence according to Wasserstein distance in the case $S = \mathbb{R}$. Let X_n and X be real random variables such that $E|X_n| + E|X| < \infty$ for each n. Then, the following statements are equivalent:

- $\lim_n d_W(X_n, X) = 0$;
- $X_n \xrightarrow{dist} X$ and $E|X_n| \to E|X|$;
- $X_n \xrightarrow{dist} X$ and the sequence (X_n) is uniformly integrable.

2.2. Two Technical Lemmas

The following simple lemma is fundamental for our purposes.

Lemma 1. *If a_1, $a_2 \in \mathbb{R}$, $0 \leq b_1 \leq b_2$ and $b_2 > 0$, then:*

$$\|\mathcal{N}(a_1, b_1) - \mathcal{N}(a_2, b_2)\| \leq 1 - \sqrt{\frac{b_1}{b_2}} + \frac{|a_1 - a_2|}{\sqrt{2\pi b_2}}.$$

Lemma 1 is well known; see e.g. Proposition 3.6.1 of [12] and Lemma 3 of [8]. Note also that, if $a_1 = a_2 = a$, Lemma 1 yields:

$$\|\mathcal{N}(a, b_1) - \mathcal{N}(a, b_2)\| \leq \frac{|b_1 - b_2|}{b_i} \quad \text{for each } i \text{ such that } b_i > 0.$$

The next result, needed in Section 4, is just a consequence of Lemma 1. In such a result, $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces, $g_n : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ Borel functions, and X and Y random variables with values in $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Lemma 2. *Let ν be the probability distribution of Y. If X is independent of Y and*

$$g_n(X, y) \sim \mathcal{N}(0, \sigma_n^2(y)), \quad g(X, y) \sim \mathcal{N}(0, \sigma^2(y)), \quad \sigma^2(y) > 0$$

for ν-almost all $y \in \mathcal{Y}$, then:

$$d_{TV}\Big(g_n(X, Y), g(X, Y)\Big) \leq \min\Big\{ E\Big(\frac{|\sigma_n(Y) - \sigma(Y)|}{\sigma(Y)}\Big), \ E\Big(\frac{|\sigma_n^2(Y) - \sigma^2(Y)|}{\sigma^2(Y)}\Big) \Big\}.$$

Proof. Since X is independent of Y,

$$d_{TV}\Big(g_n(X, Y), g(X, Y)\Big) = \sup_{A \in \mathcal{B}} \Big| \int \Big(P\big(g_n(X, y) \in A\big) - P\big(g(X, y) \in A\big)\Big) \nu(dy) \Big|$$

$$\leq \int \big\| P\big(g_n(X, y) \in \cdot\big) - P\big(g(X, y) \in \cdot\big) \big\| \nu(dy).$$

Thus, since $g_n(X, y)$ and $g(X, y)$ have centered Gaussian laws and $g(X, y)$ has strictly positive variance, for ν-almost all $y \in \mathcal{Y}$, Lemma 1 yields:

$$d_{TV}\Big(g_n(X, Y), g(X, Y)\Big) \leq \int \frac{|\sigma_n(y) - \sigma(y)|}{\sigma(y)} \nu(dy) = E\Big(\frac{|\sigma_n(Y) - \sigma(Y)|}{\sigma(Y)}\Big)$$

$$\text{and} \quad d_{TV}\Big(g_n(X, Y), g(X, Y)\Big) \leq \int \frac{|\sigma_n^2(y) - \sigma^2(y)|}{\sigma^2(y)} \nu(dy) = E\Big(\frac{|\sigma_n^2(Y) - \sigma^2(Y)|}{\sigma^2(Y)}\Big).$$

□

3. A General Result

As in Section 1, let X_n, V and Z be real random variables, with $Z \sim \mathcal{N}(0, 1)$ and V independent of Z. Since $|V| Z \sim VZ$, it can be assumed $V \geq 0$. We also assume $E|X_n| + E|VZ| < \infty$, so that we can define:

$$d_n = d_W(X_n, VZ).$$

In addition, we let:

$$X'_n = X_n + d_n^{1/2} U,$$

where U is a standard normal random variable independent of $(X_n, V, Z : n \geq 1)$.

We aim to estimate $d_{TV}(X_n, VZ)$. Under some conditions, however, the latter quantity can be replaced by $d_{TV}(X_n, X'_n)$.

Lemma 3. *For each $\alpha < 1/2$,*

$$|d_{TV}(X_n, VZ) - d_{TV}(X_n, X'_n)| \leq d_n^{1/2} + d_n^{1/2-\alpha} + P(V < d_n^\alpha).$$

In addition, if $E(1/V) < \infty$, then:

$$|d_{TV}(X_n, VZ) - d_{TV}(X_n, X'_n)| \leq d_n^{1/2} \{1 + E(1/V)\}.$$

Proof. The Lemma is trivially true if $d_n = 0$. Hence, it can be assumed $d_n > 0$. Define $X''_n = VZ + d_n^{1/2} U$ and note that:

$$|d_{TV}(X_n, VZ) - d_{TV}(X_n, X'_n)| \leq d_{TV}(X'_n, X''_n) + d_{TV}(X''_n, VZ).$$

For each $A \in \mathcal{B}$,

$$P(X'_n \in A) = \int \mathcal{N}(X_n, d_n)(A) \, dP \quad \text{and} \quad P(X''_n \in A) = \int \mathcal{N}(VZ, d_n)(A) \, dP.$$

Hence, Lemma 1 yields:

$$d_{TV}(X'_n, X''_n) = \sup_{A \in \mathcal{B}} \left| \int \left(\mathcal{N}(X_n, d_n)(A) - \mathcal{N}(VZ, d_n)(A) \right) dP \right|$$
$$\leq \int \|\mathcal{N}(X_n, d_n) - \mathcal{N}(VZ, d_n)\| \, dP \leq \frac{E|X_n - VZ|}{d_n^{1/2}}.$$

On the other hand, the probability distribution of X''_n can also be written as:

$$P(X''_n \in A) = \int \mathcal{N}(0, V^2 + d_n)(A) \, dP.$$

Arguing as above, Lemma 1 implies again:

$$d_{TV}(X''_n, VZ) \leq \int \|\mathcal{N}(0, V^2 + d_n) - \mathcal{N}(0, V^2)\| \, dP$$
$$\leq E\left(1 - \frac{V}{\sqrt{V^2 + d_n}}\right) \leq E\left(\frac{d_n^{1/2}}{\sqrt{V^2 + d_n}}\right) \leq \frac{d_n^{1/2}}{\epsilon} + P(V < \epsilon)$$

for each $\epsilon > 0$. Letting $\epsilon = d_n^\alpha$ with $\alpha < 1/2$, it follows that:

$$|d_{TV}(X_n, VZ) - d_{TV}(X_n, X'_n)| \leq \frac{E|X_n - VZ|}{d_n^{1/2}} + d_n^{1/2-\alpha} + P(V < d_n^\alpha). \tag{3}$$

Inequality (3) holds true for *every* joint distribution for the pair (X_n, VZ). In particular, inequality (3) holds if such a joint distribution is taken to be one that realizes the Wasserstein distance, namely, one such that $E|X_n - VZ| = d_n$. In this case, one obtains:

$$|d_{TV}(X_n, VZ) - d_{TV}(X_n, X'_n)| \leq d_n^{1/2} + d_n^{1/2-\alpha} + P(V < d_n^\alpha).$$

Finally, if $E(1/V) < \infty$, it suffices to note that:

$$d_{TV}(X''_n, VZ) \leq E\left(\frac{d_n^{1/2}}{\sqrt{V^2 + d_n}}\right) \leq d_n^{1/2} E(1/V).$$

$\square$

For Lemma 3 to be useful, $d_{TV}(X_n, X'_n)$ should be kept under control. This can be achieved under various assumptions. One is to ask X_n to admit a Lipschitz density with respect to Lebesgue measure.

Theorem 1. *Let ϕ_n be the characteristic function of X_n and*

$$l_n = \int |t\,\phi_n(t)|\,dt = 2\int_0^\infty t\,|\phi_n(t)|\,dt.$$

Given $\beta \geq 1$, suppose $\sup_n E|X_n|^\beta < \infty$ and $d_n \to 0$. Then, there is a constant k, independent of n, such that:

$$d_{TV}(X_n, X'_n) \leq k\left(l_n\,d_n^{1/2}\right)^{\beta/(\beta+1)}.$$

In particular,

$$d_{TV}(X_n, VZ) \leq d_n^{1/2} + d_n^{1/2-\alpha} + P(V < d_n^\alpha) + k\left(l_n\,d_n^{1/2}\right)^{\beta/(\beta+1)}$$

for each $\alpha < 1/2$, and

$$d_{TV}(X_n, VZ) \leq d_n^{1/2}\left\{1 + E(1/V)\right\} + k\left(l_n\,d_n^{1/2}\right)^{\beta/(\beta+1)} \qquad \text{if } E(1/V) < \infty.$$

It is worth noting that, if $\beta = 1$, the condition $\sup_n E|X_n| < \infty$ follows from $d_n \to 0$. On the other hand, $d_n \to 0$ can be weakened into $X_n \xrightarrow{dist} VZ$ whenever $\sup_n E|X_n|^\beta < \infty$ for some $\beta > 1$; see Section 2.1.

Proof of Theorem 1. If $l_n = \infty$, the Theorem is trivially true. Thus, it can be assumed $l_n < \infty$.

Since ϕ_n is integrable, the probability distribution of X_n admits a density f_n with respect to Lebesgue measure. In addition,

$$|f_n(x) - f_n(y)| = (1/2\pi)\left|\int (e^{-itx} - e^{-ity})\,\phi_n(t)\,dt\right|$$

$$\leq \frac{|x-y|}{2\pi}\int |t\,\phi_n(t)|\,dt = \frac{l_n\,|x-y|}{2\pi}.$$

Given $t > 0$, it follows that:

$$2\,d_{TV}(X_n, X'_n) \leq 2\int \|P(X_n \in \cdot) - P(X_n + d_n^{1/2}u \in \cdot)\|\,\mathcal{N}(0,1)(du)$$

$$= \int\int |f_n(x) - f_n(x - d_n^{1/2}u)|\,dx\,\mathcal{N}(0,1)(du)$$

$$\leq P(|X_n| > t) + P(|X'_n| > t) + \int_{-t}^t \int |f_n(x) - f_n(x - d_n^{1/2}u)|\,\mathcal{N}(0,1)(du)\,dx.$$

Since $\sup_n E|X_n|^\beta < \infty$ and $d_n \to 0$, one obtains:

$$P(|X_n| > t) + P(|X'_n| > t) \leq P(|X_n| > t) + P(|X_n| > t/2) + P(d_n^{1/2}|U| > t/2)$$

$$\leq 2\,P(|X_n| > t/2) + \frac{d_n^{\beta/2}E|U|^\beta}{(t/2)^\beta}$$

$$\leq \frac{2\,E|X_n|^\beta + d_n^{\beta/2}E|U|^\beta}{(t/2)^\beta} \leq \frac{k^*}{t^\beta}$$

for some constant k^*. Hence,

$$2\, d_{TV}(X_n, X'_n) \leq \frac{k^*}{t^\beta} + \frac{l_n\, d_n^{1/2}}{2\pi} \int_{-t}^{t} \int |u|\, \mathcal{N}(0,1)(du)\, dx$$

$$\leq \frac{k^*}{t^\beta} + \frac{l_n\, d_n^{1/2}}{\pi}\, t \quad \text{for each } t > 0.$$

Minimizing over t, one finally obtains:

$$2\, d_{TV}(X_n, X'_n) \leq c(\beta)\, (k^*)^{1/(\beta+1)} \left(\frac{l_n\, d_n^{1/2}}{\pi}\right)^{\beta/(\beta+1)},$$

where $c(\beta)$ is a constant that depends on β only. This concludes the proof. □

Theorem 1 provides upper bounds for $d_{TV}(X_n, VZ)$ in terms of l_n and d_n. It is connected to Proposition 4.1 of [4], where d_{TV} is replaced by the Kolmogorov distance.

In particular, Theorem 1 implies that $d_{TV}(X_n, VZ) \to 0$ provided $V > 0$ a.s. and

$$\lim_n d_n^{1/2} l_n = \lim_n d_W(X_n, VZ)^{1/2} \int |t\, \phi_n(t)|\, dt = 0.$$

In addition, Theorem 1 allows to estimate the convergence rate. As an extreme example, if $d_n \to 0$, $E(1/V) < \infty$ and $\sup_n\{l_n + E|X_n|^\beta\} < \infty$ for all $\beta \geq 1$, then:

$$d_{TV}(X_n, VZ) = O(d_n^\alpha) \quad \text{for every } \alpha < 1/2.$$

We next turn to examples. In each such examples, Z is a standard normal random variable independent of all other random elements.

Example 2. (Classical CLT). *Let $V = 1$ and $X_n = (1/\sqrt{n}) \sum_{i=1}^{n} \xi_i$, where $\xi_1, \xi_2, \ldots$ is an i.i.d. sequence of real random variables such that $E(\xi_1) = 0$ and $E(\xi_1^2) = 1$. In this case, $d_n = O(n^{-1/2})$; see Theorem 2.1 of [13]. Suppose now that $E|\xi_1|^\beta < \infty$ for all $\beta \geq 1$ and ξ_1 has a density f (with respect to Lebesgue measure) such that $\int |f'(x)|\, dx < \infty$. Then, $\sup_n\{l_n + E|X_n|^\beta\} < \infty$ for all $\beta \geq 1$, and Theorem 1 yields:*

$$d_{TV}(X_n, Z) = O(n^{-\alpha}) \quad \text{for each } \alpha < 1/4.$$

This rate, however, is quite far from optimal. Under the present assumptions on ξ_1, in fact, $d_{TV}(X_n, Z) = O(n^{-1/2})$; see Theorem 1 of [14].

We finally prove $\sup_n\{l_n + E|X_n|^\beta\} < \infty$. It is well known that $E|\xi_1|^\beta < \infty$ for all β implies $\sup_n E|X_n|^\beta < \infty$ for all β. Hence, it suffices to prove $\sup_n l_n < \infty$. Let ϕ be the characteristic function of ξ_1 and $q = \int |f'(x)|\, dx$. An integration by parts yields $|\phi(t)| \leq q/|t|$ for each $t \neq 0$. By Lemma 1.4 of [15], one also obtains $|\phi(t)| \leq 1 - (1/43)(t/q)^2$ for $|t| < 2q$ (just let $b = 2q$ and $c = 1/2$ in Lemma 1.4 of [15]). Since $\phi_n(t) = \phi(t/\sqrt{n})^n$ for each $t \in \mathbb{R}$,

$$|\phi_n(t)| \leq \left(\frac{q\,\sqrt{n}}{|t|}\right)^n \quad \text{for } |t| \geq q\,\sqrt{n} \quad \text{and}$$

$$|\phi_n(t)| \leq \left(1 - \frac{t^2}{43\, q^2 n}\right)^n \quad \text{for } |t| < q\,\sqrt{n}.$$

Using these inequalities, $\sup_n l_n < \infty$ follows from a direct calculation.

As noted above, the rate provided by Theorem 1 in the classical CLT is not optimal. While not exciting, this fact could be expected. Indeed, Theorem 1 is a general result, applying to *arbitrary* X_n, and should not be requested to give optimal bounds in a very special case (such as the classical CLT).

Example 3. (Exchangeable CLT). *Suppose now that (ξ_n) is an exchangeable sequence of real random variables with $E(\xi_1^2) < \infty$. Define*

$$V = \sqrt{E(\xi_1^2 \mid \mathcal{T}) - E(\xi_1 \mid \mathcal{T})^2} \quad and \quad X_n = \frac{\sum_{i=1}^{n}\{\xi_i - E(\xi_1 \mid \mathcal{T})\}}{\sqrt{n}},$$

where $\mathcal{T}$ is the tail σ-field of (ξ_n). By de Finetti's theorem,

$$d_{TV}(X_n, VZ) \leq E\Big(\|P(X_n \in \cdot \mid \mathcal{T}) - \mathcal{N}(0, V^2)\|\Big).$$

Hence, $d_{TV}(X_n, VZ) \to 0$ provided $\|P(X_n \in \cdot \mid \mathcal{T}) - \mathcal{N}(0, V^2)\| \xrightarrow{P} 0$. As to Theorem 1, note that $X_n \xrightarrow{dist} VZ$ (see e.g. Theorem 3.1 of [16]) and

$$E(X_n^2) = E\{E(X_n^2 \mid \mathcal{T})\} = n\, E\Big\{E\Big(\big(\frac{\sum_{i=1}^{n}(\xi_i - E(\xi_1 \mid \mathcal{T}))}{n}\big)^2 \mid \mathcal{T}\Big)\Big\}$$
$$= n\, E(V^2/n) = E(V^2) < \infty.$$

Furthermore, $l_n \leq E\Big\{\int |t|\, |E(e^{itX_n} \mid \mathcal{T})|\, dt\Big\}$. Thus, by Theorem 1, $d_{TV}(X_n, VZ) \to 0$ whenever

$$E(\xi_1^2 \mid \mathcal{T}) > E(\xi_1 \mid \mathcal{T})^2 \text{ a.s.} \quad and \quad \lim_n d_n^{1/2}\, E\Big\{\int |t|\, |E(e^{itX_n} \mid \mathcal{T})|\, dt\Big\} = 0.$$

Example 4. (Martingale CLT). *Let*

$$X_n = \sum_{j=1}^{k_n} \xi_{n,j},$$

where $(\xi_{n,j} : n \geq 1, j = 1, \ldots, k_n)$ is an array of real square integrable random variables and $k_n \uparrow \infty$. For each $n \geq 1$, let:

$$\mathcal{F}_{n,0} \subset \mathcal{F}_{n,1} \subset \ldots \subset \mathcal{F}_{n,k_n}$$

be sub-σ-fields of $\mathcal{F}$ with $\mathcal{F}_{n,0} = \{\varnothing, \Omega\}$. A well known version of the CLT (see e.g. Theorem 3.2 of [17]) states that $X_n \xrightarrow{dist} VZ$ provided:

(i)   *$\xi_{n,j}$ is $\mathcal{F}_{n,j}$-measurable and $E(\xi_{n,j} \mid \mathcal{F}_{n,j-1}) = 0$ a.s.;*
(ii)   *$\sum_j \xi_{n,j}^2 \xrightarrow{P} V^2, \max_j|\xi_{n,j}| \xrightarrow{P} 0, \sup_n E(\max_j \xi_{n,j}^2) < \infty$;*
(iii)   *$\mathcal{F}_{n,j} \subset \mathcal{F}_{n+1,j}$.*

Condition (iii) can be replaced by:

(iv)   *V is measurable with respect to the σ-field generated by $\mathcal{N} \cup (\cap_{n,j}\mathcal{F}_{n,j})$ where $\mathcal{N} = \{A \in \mathcal{F} : P(A) = 0\}$.*

Note also that, under (i), one obtains $E(X_n^2) = \sum_{j=1}^{k_n} E(\xi_{n,j}^2)$.

Now, in addition to (i)–(ii)–(iii) or (i)–(ii)–(iv), suppose $\sup_n \sum_{j=1}^{k_n} E(\xi_{n,j}^2) < \infty$. Then, Theorem 1 (applied with $\beta = 2$) implies $d_{TV}(X_n, VZ) \to 0$ whenever $V > 0$ a.s. and $\lim_n d_n^{1/2} l_n = 0$. Moreover,

$$d_{TV}(X_n, VZ) = O\Big(\big(l_n\, d_n^{1/2}\big)^{2/3}\Big) \qquad if\ E(1/V) + l_n < \infty\ for\ each\ n.$$

Our last example is connected to the second order Wiener chaos. We first note a simple fact as a lemma.

Lemma 4. *Let $\xi = (\xi_1, \ldots, \xi_k)$ be a centered Gaussian random vector. Define:*

$$Y = \sum_{j=1}^{k} a_j \left\{\xi_j^2 - \gamma_j^2\right\},$$

where $a_j \in \mathbb{R}$ and $\gamma = (\gamma_1, \ldots, \gamma_k)$ is an independent copy of ξ. Then, the characteristic function ψ of Y can be written as:

$$\psi(t) = E\left\{e^{-t^2 S}\right\}, \quad t \in \mathbb{R}, \text{ where } S = \sum_{i,j} a_i \, a_j \, E(\xi_i \xi_j) \, (\xi_i + \gamma_i) \, (\xi_j + \gamma_j).$$

Proof. Let $\sigma_{i,j} = E(\xi_i \xi_j)$, $\xi^* = (\xi + \gamma)/\sqrt{2}$ and $\gamma^* = (\xi - \gamma)/\sqrt{2}$. Then,

$$(\xi^*, \gamma^*) \sim (\xi, \gamma), \quad Y = 2 \sum_j a_j \, \xi_j^* \gamma_j^*, \quad S = 2 \sum_{i,j} a_i a_j \sigma_{i,j} \xi_i^* \xi_j^*.$$

Therefore,

$$\psi(t) = E\left\{E\left(e^{itY} \mid \xi^*\right)\right\} = E\left\{e^{-2t^2 \sum_{i,j} a_i a_j \sigma_{i,j} \xi_i^* \xi_j^*}\right\} = E\left\{e^{-t^2 S}\right\}.$$

$\square$

Example 5. (Squares of Gaussian random variables). *For each $n \geq 1$, let $(\xi_{n,1}, \ldots, \xi_{n,k_n})$ be a centered Gaussian random vector and*

$$X_n = \sum_{j=1}^{k_n} a_{n,j} \left\{\xi_{n,j}^2 - E(\xi_{n,j}^2)\right\} \qquad \text{where } a_{n,j} \in \mathbb{R}.$$

Take an independent copy $(\gamma_{n,1}, \ldots, \gamma_{n,k_n})$ of $(\xi_{n,1}, \ldots, \xi_{n,k_n})$ and define:

$$Y_n = \sum_{j=1}^{k_n} a_{n,j} \left\{\xi_{n,j}^2 - \gamma_{n,j}^2\right\},$$

$$S_n = \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} a_{n,i} \, a_{n,j} \, E(\xi_{n,i} \xi_{n,j}) \, (\xi_{n,i} + \gamma_{n,i}) \, (\xi_{n,j} + \gamma_{n,j}).$$

Note that S_n is a (random) quadratic form of the covariance matrix $(E(\xi_{n,i}\xi_{n,j}) : 1 \leq i, j \leq k_n)$. Therefore, $S_n \geq 0$.

Since $|\phi_n|^2$ agrees with the characteristic function of Y_n, Lemma 4 yields:

$$|\phi_n(t)|^2 = E\left\{e^{-t^2 S_n}\right\}.$$

Being $S_n \geq 0$, it follows that:

$$|\phi_n(t)|^2 = E\left\{e^{-t^2 S_n} 1_{\{S_n \geq t^{\frac{\epsilon-4}{2}}\}}\right\} + E\left\{e^{-t^2 S_n} 1_{\{S_n < t^{\frac{\epsilon-4}{2}}\}}\right\}$$

$$\leq e^{-t^{\epsilon/2}} + P(S_n < t^{\frac{\epsilon-4}{2}}) = e^{-t^{\epsilon/2}} + P\left(S_n^{-2-\epsilon} > t^{4+\frac{\epsilon(2-\epsilon)}{2}}\right)$$

$$\leq e^{-t^{\epsilon/2}} + E\left(S_n^{-2-\epsilon}\right) t^{-4-\frac{\epsilon(2-\epsilon)}{2}} \quad \text{for all } \epsilon > 0 \text{ and } t > 0.$$

Hence,

$$\frac{l_n}{2} = \int_0^\infty t\, |\phi_n(t)|\, dt \le 1 + \int_1^\infty t\, |\phi_n(t)|\, dt$$

$$\le 1 + \int_1^\infty t\, e^{\frac{-t^{\epsilon/2}}{2}}\, dt + \sqrt{E\!\left(S_n^{-2-\epsilon}\right)} \int_1^\infty t^{-1-\frac{\epsilon(2-\epsilon)}{4}}\, dt,$$

so that $\sup_n l_n < \infty$ *whenever* $\sup_n E\!\left(S_n^{-2-\epsilon}\right) < \infty$ *for some* $\epsilon \in (0,2)$.

To summarize, applying Theorem 1 with $\beta = 2$, *one obtains:*

$$d_{TV}(X_n,\, VZ) = O(d_n^{1/3}) \tag{4}$$

provided $X_n \xrightarrow{dist} VZ$, *for some V independent of Z, and*

$$E(1/V) + \sup_n \left\{ E\!\left(S_n^{-2-\epsilon}\right) + E(X_n^2) \right\} < \infty \quad \text{for some } \epsilon \in (0,2).$$

The bound (4) requires strong conditions, which may be not easily verifiable in real problems. However, the above result is sometimes helpful, possibly in connection with the martingale CLT of Example 4. As an example, the conditions for (4) are not hard to be checked when $\xi_{n,1}, \ldots, \xi_{n,k_n}$ *are independent for fixed n. We also note that, to our knowledge, the bound (4) improves on the existing ones. In fact, letting* $p = 2$ *in Theorem 3.1 of [6] (see also Remark 3.5 of [7]) one only obtains* $d_{TV}(X_n,\, VZ) = O(d_n^{1/5})$.

4. Weighted Quadratic Variations

Theorem 1 works nicely if one is able to estimate d_n and l_n, which is usually quite hard. Thus, it is convenient to have some further tools. In this section, $d_{TV}(X_n,\, VZ)$ is upper bounded via Lemma 2. We focus on a special case, but the underlying ideas are easily adapted to more general situations. The results in [8], for instance, arise from a version of such ideas.

For any function $x : [0,1] \to \mathbb{R}$, denote:

$$\Delta x(k/n) = x((k+1)/n) - x(k/n) \quad \text{where } n \ge 1 \text{ and } k = 0,1,\ldots,n-1.$$

Let $q \ge 2$ be an integer, $f : \mathbb{R} \to \mathbb{R}$ a Borel function, and $J = \{J_t : 0 \le t \le 1\}$ a real process. The weighted q-variation of J on $\{0, 1/n, 2/n, \ldots, 1\}$ is:

$$J_n^* = \sum_{k=0}^{n-1} f(J_{k/n}) \left(\Delta J_{k/n}\right)^q.$$

As noted in [5], to fix the asymptotic behavior of J_n^* is useful to determine the rate of convergence of some approximation schemes of stochastic differential equations driven by J. Moreover, the study of J_n^* is also motivated by parameter estimation and by the analysis of single-path behaviour of J. See [5,9,18–21] and references therein.

More generally, given an $\mathbb{R}^2$-valued process:

$$(I, J) = \{(I_t, J_t) : 0 \le t \le 1\},$$

one could define:

$$(I, J)_n^* = \sum_{k=0}^{n-1} f(I_{k/n}) \left(\Delta J_{k/n}\right)^q.$$

The weight $f(I_{k/n})$ of $\left(\Delta J_{k/n}\right)^q$ depends now on I. Thus, in a sense, $(I, J)_n^*$ can be regarded as the weighted q-variation of J relative to I.

Here, we focus on:

$$X_n = n^{1/2} \sum_{k=0}^{n-1} f\big(B_{k/n} - B'_{k/n}\big) \big\{ (\Delta B_{k/n})^2 - (\Delta B'_{k/n})^2 \big\}, \tag{5}$$

where B and B' are independent standard Brownian motions. Note that, letting $q = 2$ and $I = B - B'$, one obtains:

$$X_n = n^{1/2} \big\{ (I, B)^*_n - (I, B')^*_n \big\}.$$

Thus, $n^{-1/2} X_n$ can be seen as the difference between the quadratic variations of B and B' relative to $I = B - B'$.

We aim to show that, under mild assumptions on f, the probability distributions of X_n converge in total variation to a certain mixture of Gaussian laws. We also estimate the rate of convergence. The smoothness assumptions on f are weaker than those usually requested in similar problems; see, e.g., [5].

Theorem 2. *Let B and B' be independent standard Brownian motions and Z a standard normal random variable independent of (B, B'). Define X_n by Equation (5) and*

$$V = 2 \sqrt{\int_0^1 f^2\big(\sqrt{2} B_t\big)\, dt}\,.$$

Suppose $E(1/V^2) < \infty$ and

$$|f(x) - f(y)| \le c\, |x - y|\, e^{|x| + |y|}$$

for some constant c and all $x, y \in \mathbb{R}$. Then, there is a constant k independent of n satisfying:

$$d_{TV}(X_n, VZ) \le k\, n^{-1/4}.$$

Moreover, if $\inf|f| > 0$, one also obtains $d_{TV}(X_n, VZ) \le k\, n^{-1/2}$.

To understand better the spirit of Theorem 2, think of the trivial case $f = 1$. Then, the asymptotic behavior of $X_n = n^{1/2} \sum_{k=0}^{n-1} \big\{ (\Delta B_{k/n})^2 - (\Delta B'_{k/n})^2 \big\}$ can be deduced by classical results. In fact, $d_{TV}(X_n, 2Z) = O(n^{-1/2})$ and this rate is optimal; see Theorem 1 of [14]. On the other hand, since $V = 2$, the same conclusion can be drawn from Theorem 2.

We finally prove Theorem 2.

Proof of Theorem 2. First note that $T = (B + B')/\sqrt{2}$ and $Y = (B - B')/\sqrt{2}$ are independent standard Brownian motions and

$$X_n = 2 n^{1/2} \sum_{k=0}^{n-1} f\big(\sqrt{2}\, Y_{k/n}\big)\, \Delta T_{k/n}\, \Delta Y_{k/n}.$$

Note also that:

$$VZ \sim 2 T_1 \sqrt{\int_0^1 f^2\big(\sqrt{2} Y_t\big)\, dt}.$$

Thus, in order to apply Lemma 2, it suffices to let $\mathcal{X} = \mathcal{Y} = C[0,1]$, $X = T$, and

$$g_n(x,y) = 2\,n^{1/2} \sum_{k=0}^{n-1} f\left(\sqrt{2}\,y(k/n)\right) \Delta x(k/n)\, \Delta y(k/n),$$

$$g(x,y) = 2\,x(1) \sqrt{\int_0^1 f^2\left(\sqrt{2}y(t)\right) dt}.$$

For fixed $y \in \mathcal{Y}$, $g_n(T,y)$ and $g(T,y)$ are centered Gaussian random variables. Since $E\left\{\left(\Delta T_{k/n}\right)^2\right\} = 1/n$,

$$\sigma_n^2(y) = E\left\{g_n(T,y)^2\right\} = 4 \sum_{k=0}^{n-1} f^2\left(\sqrt{2}\,y(k/n)\right) \left(\Delta y(k/n)\right)^2$$

$$\text{and} \quad \sigma^2(y) = E\left\{g(T,y)^2\right\} = 4 \int_0^1 f^2\left(\sqrt{2}y(t)\right) dt.$$

On noting that $\sigma^2(Y) \sim V^2$, one also obtains:

$$\sigma^2(Y) > 0 \text{ a.s.} \quad \text{and} \quad E\left\{1/\sigma^2(Y)\right\} = E(1/V^2) < \infty.$$

Next, define:

$$a_n = (1/4)\,E\left\{|\sigma_n^2(Y) - \sigma^2(Y)|\right\}$$

$$= E\left\{\left|\sum_{k=0}^{n-1} f^2\left(\sqrt{2}\,Y_{k/n}\right) \left(\Delta Y_{k/n}\right)^2 - \int_0^1 f^2\left(\sqrt{2}Y_t\right) dt\right|\right\}.$$

By Lemma 2 and the Cauchy–Schwarz inequality,

$$d_{TV}(X_n,\,VZ)^2 = d_{TV}\left(g_n(T,Y),\,g(T,Y)\right)^2 \le E\left(\frac{|\sigma_n(Y) - \sigma(Y)|}{\sigma(Y)}\right)^2$$

$$\le E\left\{1/\sigma^2(Y)\right\} E\left\{\left(\sigma_n(Y) - \sigma(Y)\right)^2\right\}$$

$$\le E(1/V^2)\,E\left\{|\sigma_n^2(Y) - \sigma^2(Y)|\right\} = 4\,E(1/V^2)\,a_n.$$

If $\inf|f| > 0$, since $\sigma^2(Y) \ge 4 \inf f^2$, Lemma 2 implies again:

$$d_{TV}(X_n,\,VZ) \le E\left(\frac{|\sigma_n^2(Y) - \sigma^2(Y)|}{\sigma^2(Y)}\right) \le \frac{E\left(|\sigma_n^2(Y) - \sigma^2(Y)|\right)}{4 \inf f^2} = \frac{a_n}{\inf f^2}.$$

Thus, to conclude the proof, it suffices to show that $a_n = O(n^{-1/2})$.
Define $c^* = \max(c,\,|f(0)|)$ and note that:

$$|f(s)| \le c^*\, e^{2|s|} \quad \text{and} \quad |f(s)^2 - f(t)^2| \le 2\,c\,c^*|s - t|\, e^{3(|s|+|t|)} \quad \text{for all } s,\,t \in \mathbb{R}.$$

Define also:

$$a_n^{(1)} = E\left\{\left|\sum_{k=0}^{n-1} f^2\left(\sqrt{2}\,Y_{k/n}\right) \left(\left(\Delta Y_{k/n}\right)^2 - 1/n\right)\right|\right\} \quad \text{and}$$

$$a_n^{(2)} = E\left\{\left|(1/n) \sum_{k=0}^{n-1} f^2\left(\sqrt{2}\,Y_{k/n}\right) - \int_0^1 f^2\left(\sqrt{2}Y_t\right) dt\right|\right\}.$$

Since $a_n \le a_n^{(1)} + a_n^{(2)}$, it suffices to see that $a_n^{(i)} = \mathrm{O}(n^{-1/2})$ for each i. Since Y has independent increments and $E\big\{ \big((\Delta Y_{k/n})^2 - 1/n \big)^2 \big\} = 2/n^2$,

$$
\begin{aligned}
(a_n^{(1)})^2 &= E\Big\{ \Big| \sum_{k=0}^{n-1} f^2\big(\sqrt{2}\,Y_{k/n}\big) \big((\Delta Y_{k/n})^2 - 1/n \big) \Big| \Big\}^2 \\
&\le \sum_{k=0}^{n-1} E\Big\{ f^4\big(\sqrt{2}\,Y_{k/n}\big) \big((\Delta Y_{k/n})^2 - 1/n \big)^2 \Big\} \\
&= \sum_{k=0}^{n-1} E\big\{ f^4\big(\sqrt{2}\,Y_{k/n}\big) \big\} E\Big\{ \big((\Delta Y_{k/n})^2 - 1/n \big)^2 \Big\} \\
&= (2/n^2) \sum_{k=0}^{n-1} E\big\{ f^4\big(\sqrt{2}\,Y_{k/n}\big) \big\} \\
&\le \frac{2\,(c^*)^4\, E\big\{ e^{8\sqrt{2}M} \big\}}{n} \qquad \text{where } M = \sup_{0 \le t \le 1} |Y_t|.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
a_n^{(2)} &= E\Big\{ \Big| (1/n) \sum_{k=0}^{n-1} f^2\big(\sqrt{2}\,Y_{k/n}\big) - \int_0^1 f^2\big(\sqrt{2}\,Y_t\big)\,dt \Big| \Big\} \\
&\le \sum_{k=0}^{n-1} \int_{k/n}^{(k+1)/n} E\big\{ \big| f^2\big(\sqrt{2}\,Y_{k/n}\big) - f^2\big(\sqrt{2}\,Y_t\big) \big| \big\}\,dt \\
&\le 2\sqrt{2}\,c\,c^* \sum_{k=0}^{n-1} \int_{k/n}^{(k+1)/n} E\big\{ |Y_{k/n} - Y_t|\, e^{6\sqrt{2}M} \big\}\,dt \\
&\le 2\sqrt{2}\,c\,c^* \sqrt{E\big\{ e^{12\sqrt{2}M} \big\}} \sum_{k=0}^{n-1} \int_{k/n}^{(k+1)/n} \sqrt{E\big\{ (Y_{k/n} - Y_t)^2 \big\}}\,dt \\
&\le 2\sqrt{2}\,c\,c^* \sqrt{E\big\{ e^{12\sqrt{2}M} \big\}}\, \frac{1}{\sqrt{n}}.
\end{aligned}
$$

Therefore, $a_n^{(i)} = \mathrm{O}(n^{-1/2})$ for each i, and this concludes the proof. □

Author Contributions: Each author contributed in exactly the same way to each part of this paper.

References

1. Azmoodeh, E.; Gasbarra, D. New moments criteria for convergence towards normal product/tetilla laws. *arXiv* **2017**, arXiv:1708.07681.
2. Eichelsbacher, P.; Thäle, C. Malliavin-Stein method for Variance-Gamma approximation on Wiener space. *Electron. J. Probab.* **2015**, *20*, 1–28. [CrossRef]
3. Gaunt, R.E. On Stein's method for products of normal random variables and zero bias couplings. *Bernoulli* **2017**, *23*, 3311–3345. [CrossRef]
4. Gaunt, R.E. Wasserstein and Kolmogorov error bounds for variance-gamma approximation via Stein's method I. *arXiv* **2017**, arXiv:1711.07379.
5. Nourdin, I.; Nualart, D.; Tudor, C.A. Central and non-central limit theorems for weighted power variations of fractional Brownian motion. *Ann. I.H.P.* **2010**, *46*, 1055–1079. [CrossRef]
6. Nourdin, I.; Poly, G. Convergence in total variation on Wiener chaos. *Stoch. Proc. Appl.* **2013**, *123*, 651–674. [CrossRef]

7.   Nourdin, I.; Nualart, D.; Peccati, G. Quantitative stable limit theorems on the Wiener space. *Ann. Probab.* **2016**, *44*, 1–41. [CrossRef]

8.   Pratelli, L.; Rigo, P. Total Variation Bounds for Gaussian Functionals. 2018, Submitted. Available online: http://www-dimat.unipv.it/rigo/frac.pdf (accessed on 10 April 2018)

9.   Nourdin, I.; Peccati, G. Weighted power variations of iterated Brownian motion. *Electron. J. Probab.* **2008**, *13*, 1229–1256. [CrossRef]

10. Peccati, G.; Yor, M. Four limit theorems for quadratic functionals of Brownian motion and Brownian bridge. In *Asymptotic Methods in Stochastics, AMS, Fields Institute Communication Series*; Amer. Math. Soc.: Providence, RI, USA, 2004; pp. 75–87.

11. Dudley R.M. *Real Analysis and Probability*; Cambridge University Press: Cambridge, UK, 2004.

12. Nourdin, I.; Peccati, G. *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*; Cambridge University Press: Cambridge, UK, 2012.

13. Goldstein, L. L^1 bounds in normal approximation. *Ann. Probab.* **2007**, *35*, 1888–1930. [CrossRef]

14. Sirazhdinov, S.K.H.; Mamatov, M. On convergence in the mean for densities. *Theory Probab. Appl.* **1962**, *7*, 424–428. [CrossRef]

15. Petrov, V.V. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*; Clarendon Press: Oxford, UK, 1995.

16. Berti, P.; Pratelli, L.; Rigo, P. Limit theorems for a class of identically distributed random variables. *Ann. Probab.* **2004**, *32*, 2029–2052.

17. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Applications*; Academic Press: New York, NY, USA, 1980.

18. Barndorff-Nielsen, O.E.; Graversen, S.E.; Shepard, N. Power variation and stochastic volatility: A review and some new results. *J. Appl. Probab.* **2004**, *44*, 133–143. [CrossRef]

19. Gradinaru, M.; Nourdin, I. Milstein's type schemes for fractional SDEs. *Ann. I.H.P.* **2009**, *45*, 1058–1098. [CrossRef]

20. Neuenkirch, A.; Nourdin, I. Exact rate of convergence of some approximation schemes associated to SDEs driven by a fractional Brownian motion. *J. Theor. Probab.* **2007**, *20*, 871–899. [CrossRef]

21. Nourdin, I. A simple theory for the study of SDEs driven by a fractional Brownian motion in dimension one. In *Seminaire de Probabilites*; Springer: Berlin, Germany, 2008; Volume XLI, pp. 181–197.