

# A note on recovering the Brownian motion component from a Lévy process

Konstantin Borovkov\*

## Abstract

González Cázares and Ivanovs (2021) suggested a new method for “recovering” the Brownian motion component from the trajectory of a Lévy process that required sampling from an independent Brownian motion process. We show that such a procedure works equally well without any additional source of randomness if one uses normal quantiles instead of the ordered increments of the auxiliary Brownian motion process.

**Keywords:** Brownian motion; Lévy process; quantile function; high frequency sampling.

**MSC2020 subject classifications:** 60F99; 60G51; 60J65.

Submitted to ECP on March 15, 2022, final version accepted on July 10, 2022.

## 1 Introduction and main results

The present note is complementing a recent interesting paper [4] (see also related paper [3]) presenting an original idea on how to “recover” the Brownian motion component from the observed Lévy process

$$X_t = Y_t + \sigma W_t, \quad t \in [0, 1],$$

where the standard Brownian motion  $W$  is independent of the pure jump process  $Y$  and  $\sigma > 0$ . More precisely, the authors of [4] recovered the path of the Brownian bridge  $\{W_t - tW_1\}_{t \in [0,1]}$  (noting that it is not possible to consistently “extract” the linear drift from the Brownian motion process trajectory, due to the equivalence of the distributions of Brownian motion processes with different linear drifts).

Apart from being an interesting mathematical result by itself, such a separation can be useful in some applications as well. For instance, if the Brownian component is interpreted as noise, it enables one to recover (up to a linear drift) the signal  $Y$  from the observed process  $X$ . Further comments (and some relevant references) on how one can benefit from such a separation in statistical problems can be found in [4].

The first method used in [4] for recovering the Brownian motion component was based on a construction that curiously required randomization. To describe that method, we need to introduce some notations. For an  $n$ -tuple  $\mathcal{X} = (x_1, \dots, x_n)$  of real numbers that are all different from each other (this will a.s. be the case for all the random samples considered in this note, so without loss of generality we will assume in what follows that they all do have this property, omitting “a.s.” in the respective relations),

---

\*The University of Melbourne, Australia. E-mail: borovkov@unimelb.edu.au

denote by  $R_k(\mathcal{X}) := \sum_{j=1}^n \mathbf{1}(x_j \leq x_k)$  the rank of  $x_k$  in  $\mathcal{X}$ ,  $k = 1, \dots, n$ , and by  $(\mathcal{X})_j$  and  $[\mathcal{X}]_j$  the  $j$ th component of  $\mathcal{X}$  and  $j$ th order statistic for  $\mathcal{X}$ , respectively, so that  $(\mathcal{X})_k = x_k$  and  $[\mathcal{X}]_{R_k(\mathcal{X})} = x_k$ . For a random process  $\{V_t\}_{t \in [0,1]}$  and  $n \geq 1$ , we set  $\Delta_n V := (V_{1/n} - V_0, V_{2/n} - V_{1/n}, \dots, V_1 - V_{(n-1)/n}) \in \mathbb{R}^n$ .

Now assume that  $\{W'_t\}_{t \in [0,1]}$  is a standard Brownian motion process that is independent of  $X$  and put

$$W_t^{(n)} := \sum_{i=1}^{\lfloor nt \rfloor} [\Delta_n W']_{R_i(\Delta_n X)}, \quad t \in [0, 1].$$

In words, we first re-order one-step increments of  $W'$  on the grid  $k/n$ ,  $k = 0, 1, \dots, n$ , such that the sequence of their ranks is the same as the sequence of the ranks of the increments of  $X$  on the same grid, and then form  $W^{(n)}$  as the process of the partial sums of that re-ordered sequence of the increments of  $W'$ .

The following theorem re-states the first part of the main result in [4]. Set  $\beta^* := \inf \{p > 0 : \int_{(-1,1)} |x|^p \Pi(dx) < \infty\}$ , where  $\Pi$  is the jump measure of  $Y$ .

**Theorem 1.1.** For any  $p \in (\beta^*, 2] \cup \{2\}$ , as  $n \rightarrow \infty$ ,

$$\sup_{t \in [0,1]} |W_t - W_t^{(n)} - (W_1 - W_1^{(n)})t| = o_P(n^{-1/2+p/4}). \tag{1.1}$$

The natural question that arises here is about the role of the independently sampled process  $W'$ : why do we need this auxiliary random object? Does it necessarily need to be a standard Brownian motion? It is possible to modify the suggested method to avoid using any auxiliary independent random processes?

Simulations showed that the described scheme still works when  $W'$  is an independent fractional Brownian motion process with an arbitrary Hurst parameter  $H \in (0, 1)$  (one just needs to scale the process so that the marginal distributions of the components of  $\Delta_n W'$  would be the same as for  $\Delta_n W$ ). This observation suggested that the true role of the auxiliary process  $W'$  is just to provide approximations to normal quantiles and that one can “recover”  $W$  using this kind of approach without sampling any independent random process. We show in the present note that this is the case indeed.

Denote by  $\Phi$  the standard normal distribution function, by  $\bar{\Phi} := 1 - \Phi$  the distribution tail of  $\Phi$ , by  $\varphi$  the density of  $\Phi$ , and by  $Q := \Phi^{-1}$  the standard normal quantile function. Finally, set

$$\widetilde{W}_t^{(n)} := n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} Q(R_i(\Delta_n X)/(n+1)), \quad t \in [0, 1]. \tag{1.2}$$

Our main result is the following theorem.

**Theorem 1.2.** The assertion of Theorem 1.1 remains true if the process  $W^{(n)}$  is replaced in it with  $\widetilde{W}^{(n)}$ .

## 2 Proofs

One of the key elements of our proof of Theorem 1.2 is the bound on the rate of approximation of the quantile function  $Q$  by the step function

$$Q_n(u) := \sum_{k=1}^n Q\left(\frac{k}{n+1}\right) \mathbf{1}\left(u \in \left[\frac{k-1}{n}, \frac{k}{n}\right)\right), \quad u \in (0, 1), \tag{2.1}$$

that we state in the following lemma.

**Lemma 2.1.** For any  $n \geq 1$ , one has

$$\int_0^1 (Q(u) - Q_n(u))^2 du \leq \frac{3.73}{n}.$$

The proof of this bound uses three elementary auxiliary results established below in Lemmata 2.2–2.4.

**Lemma 2.2.** Let  $f \in C^1(a_0, b_0)$  be convex and non-decreasing on  $[a, b] \subset (a_0, b_0)$ . Then, for any  $v_0 \in [a, b]$ ,

$$\int_a^b (f(v) - f(v_0))^2 dv \leq \frac{1}{3} (f'(b))^2 (b - a)^3.$$

*Proof of Lemma 2.2.* For  $v, v_0 \in [a, b]$ , one has

$$|f(v) - f(v_0)| = \left| \int_{v_0}^v f'(s) ds \right| \leq f'(b) |v - v_0|$$

as  $f'(s) \geq 0$  is non-decreasing by assumption. Hence

$$\int_a^b (f(v) - f(v_0))^2 dv \leq (f'(b))^2 \int_a^b (v - v_0)^2 dv \leq \frac{1}{3} (f'(b))^2 ((b - v_0)^3 - (a - v_0)^3),$$

completing the proof since clearly  $y^3 - x^3 \leq (y - x)^3$  for  $x \leq 0 \leq y$ .  $\square$

**Lemma 2.3.** For any  $u \in (0.5, 1)$ , one has  $1 - u \leq \sqrt{\frac{\pi}{2}} \varphi(Q(u))$ .

*Proof of Lemma 2.3.* The desired inequality follows from the observation that it turns into equality at the endpoints  $u = 0.5$  and  $u = 1$  and that its RHS is a concave function as  $[\varphi(Q(u))]' = -\sqrt{2\pi} e^{Q^2(u)/2} < 0$ .  $\square$

**Lemma 2.4.** Assume that  $f(v)$ ,  $v \in [a, b]$ , is a non-decreasing function,  $v_0 \in [a, \frac{1}{2}(a + b)]$ . Then

$$I_1 := \int_a^b (f(v) - f(v_0))^2 dv \leq \int_a^b (f(v) - f(a))^2 dv =: I_2.$$

*Proof of Lemma 2.4.* Expanding the squares, one has

$$\begin{aligned} I_2 - I_1 &= 2(f(v_0) - f(a)) \int_a^b f(v) dv + (f^2(a) - f^2(v_0))(b - a) \\ &= (f(v_0) - f(a)) \left[ 2 \int_a^b f(v) dv - (f(v_0) + f(a))(b - a) \right] \geq 0 \end{aligned}$$

since, due to the monotonicity of  $f$ ,

$$\begin{aligned} \int_a^b f(v) dv &= \int_a^{v_0} \dots + \int_{v_0}^b \dots \geq f(a)(v_0 - a) + f(v_0)(b - v_0) \\ &= \frac{1}{2} f(a)(b - a) + \frac{1}{2} f(v_0)(b - a) + (f(v_0) - f(a)) \left( \frac{1}{2}(a + b) - v_0 \right), \end{aligned}$$

where the last term is non-negative by the assumptions.  $\square$

*Proof of Lemma 2.1.* One has

$$\int_0^1 (Q(u) - Q_n(u))^2 du = \sum_{k=1}^n \int_{(k-1)/n}^{k/n} \left( Q(u) - Q\left(\frac{k}{n+1}\right) \right)^2 du. \quad (2.2)$$

By symmetry, it is enough to bound the terms with  $k \geq n/2$  only, assuming for simplicity that  $n$  is even.

As  $Q$  is clearly convex and increasing on  $(\frac{1}{2}, 1)$ , for  $n/2 < k < n$  we get by Lemmata 2.2 and 2.3 that

$$\int_{(k-1)/n}^{k/n} \left(Q(u) - Q\left(\frac{k}{n+1}\right)\right)^2 du \leq \frac{1}{3} \left(Q'\left(\frac{k}{n}\right)\right)^2 n^{-3} = \frac{n^{-3}}{3\varphi^2(Q(\frac{k}{n}))} \leq \frac{\pi n^{-1}}{6(n-k)^2}.$$

Therefore

$$\sum_{k=n/2+1}^{n-1} \int_{(k-1)/n}^{k/n} \left(Q(u) - Q\left(\frac{k}{n+1}\right)\right)^2 du \leq \frac{\pi}{6n} \sum_{k=n/2+1}^{n-1} \frac{1}{(n-k)^2} \leq \frac{\pi}{6n} \sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^3}{36n}. \tag{2.3}$$

For the last term in the sum on the RHS of (2.2), setting  $q := Q(1 - 1/n)$ , from Lemma 2.4 we obtain that

$$\begin{aligned} J_n &:= \int_{(n-1)/n}^1 \left(Q(u) - Q\left(\frac{n}{n+1}\right)\right)^2 du \leq \int_{(n-1)/n}^1 (Q(u) - q)^2 du \\ &= \int_{(n-1)/n}^1 Q(u)^2 du - 2q \int_{(n-1)/n}^1 Q(u) du + q^2 n^{-1}. \end{aligned}$$

Integrating by parts, we get

$$\begin{aligned} \int_{(n-1)/n}^1 Q(u)^2 du &= \mathbf{E}(Z_1^2; Z_1 > q) = \int_q^\infty z^2 \varphi(z) dz \\ &= [-z\varphi(z)]_q^\infty + \int_q^\infty \varphi(z) dz = q\varphi(q) + \bar{\Phi}(q), \end{aligned}$$

whereas

$$\int_{(n-1)/n}^1 Q(u) du = \mathbf{E}(Z_1; Z_1 > q) = \int_q^\infty z\varphi(z) dz = - \int_q^\infty d\varphi(z) = \varphi(q).$$

Since  $\bar{\Phi}(q) = n^{-1}$  and  $q^2 n^{-1} - q\varphi(q) = q^2(\bar{\Phi}(q) - \varphi(q)/q) < 0$  by the well-known inequality for the normal Mills' ratio (see e.g. Ch. VII.1 in [2]), we conclude that

$$J_n \leq q\varphi(q) + n^{-1} - 2q\varphi(q) + q^2 n^{-1} \leq n^{-1}.$$

Together with (2.3) and an elementary bound for the constant  $\frac{\pi^3}{18} + 2 < 3.73$  this completes the proof of Lemma 2.1.  $\square$

One more key element of our proof of Theorem 1.2 is the following convergence result for the processes of partial sums of exchangeable random variables.

**Lemma 2.5.** *Let  $\gamma_{n,1}, \gamma_{n,2}, \dots, \gamma_{n,n}$ ,  $n \geq 1$ , be a triangular array of random variables that are exchangeable in each row and such that  $\sum_{k=1}^n \gamma_{n,k} = 0$  a.s. and, as  $n \rightarrow \infty$ ,*

$$\sum_{k=1}^n \gamma_{n,k}^2 = o_P(1). \tag{2.4}$$

*Then  $\max_{1 \leq i \leq n} \left| \sum_{k=1}^i \gamma_{n,k} \right| = o_P(1)$ .*

The assertion of Lemma 2.5 is an immediate consequence of Theorem 3.13 in [5] on convergence of partial sums processes. In our case, the limiting process in that theorem is identically equal to zero, the convergence of characteristic triples following from the assumptions and the obvious observation that  $\max_{1 \leq k \leq n} |\gamma_{n,k}| = o_P(1)$  from (2.4).

*Proof of Theorem 1.2.* As in the proof of Theorem 1.1 in [4], we will start with the observation that

$$\sup_{t \in [0,1]} |W_t - W_{\lfloor nt \rfloor/n}| = O_P(n^{-1/2}(\ln n)^{1/2}), \tag{2.5}$$

which is an immediate consequence of the Lévy’s modulus of continuity theorem (see e.g. p. 114 in [6]). Therefore, in the problem of bounding the error in the version of (1.1) with  $\widetilde{W}^{(n)}$  instead of  $W^{(n)}$ , we only need to consider the maximum of the absolute deviations on the grid  $t = \frac{i}{n}$ ,  $1 \leq i \leq n$ .

To this end, we observe that  $\widetilde{W}_1^{(n)} = 0$  since  $Q(\frac{1}{2} + h) + Q(\frac{1}{2} - h) = 0$ ,  $h \in [0, \frac{1}{2})$ , and hence, letting

$$\eta_{n,k} := (\Delta_n W)_k - n^{-1/2}Q(R_k(\Delta_n X)/(n + 1)) - n^{-1}W_1,$$

one has

$$W_{i/n} - \widetilde{W}_{i/n}^{(n)} - (W_1 - \widetilde{W}_1^{(n)})i/n = \sum_{k=1}^i \eta_{n,k}. \tag{2.6}$$

Next, similarly to the decomposition of  $\xi_{ni}$  on p. 2422 in [4], we write  $\eta_{n,k} = \widetilde{\eta}_{n,k} + \widehat{\eta}_{n,k}$ , where

$$\begin{aligned} \widetilde{\eta}_{n,k} &:= [\Delta_n W]_{R_k(\Delta_n X)} - n^{-1/2}Q(R_k(\Delta_n X)/(n + 1)) - n^{-1}W_1, \\ \widehat{\eta}_{n,k} &:= (\Delta_n W)_k - [\Delta_n W]_{R_k(\Delta_n X)}. \end{aligned}$$

That

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^i \widehat{\eta}_{n,k} \right| = o_P(n^{-1/2+p/4}) \quad \text{as } n \rightarrow \infty \tag{2.7}$$

was proved on p. 2425 in [4]. To complete the proof of our theorem, we will now show that

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^i \widetilde{\eta}_{n,k} \right| = O_P(n^{-1/2}(\ln \ln n)^{1/2}) \quad \text{as } n \rightarrow \infty. \tag{2.8}$$

First note that it is not hard to verify that (2.8) is equivalent to the assertion that, for any positive sequence  $\varepsilon_n \rightarrow 0$ ,

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^i \widetilde{\eta}_{n,k} \right| = o_P(\varepsilon_n^{-1}n^{-1/2}(\ln \ln n)^{1/2}) \quad \text{as } n \rightarrow \infty. \tag{2.9}$$

Further, it is easy to see that the random variables  $\widetilde{\eta}_{n,1}, \dots, \widetilde{\eta}_{n,n}$  are exchangeable and  $\sum_{k=1}^n \widetilde{\eta}_{n,k} = 0$ . Moreover, one has

$$\sum_{k=1}^n \widetilde{\eta}_{n,k}^2 = O_P(n^{-1} \ln \ln n) \quad \text{as } n \rightarrow \infty. \tag{2.10}$$

To prove (2.10), note that the components of the vector

$$\mathcal{Z}_n = (Z_1, \dots, Z_n) := n^{1/2}((\Delta_n W)_1, \dots, (\Delta_n W)_n)$$

are independent standard normal random variables, and let  $\overline{Z}_n := n^{-1} \sum_{k=1}^n Z_k$ .

Setting  $\zeta_{n,k} := [\mathcal{Z}_n]_k - Q\left(\frac{k}{n+1}\right)$ ,  $k = 1, \dots, n$ , we observe that  $\bar{\zeta}_n := n^{-1} \sum_{k=1}^n \zeta_{n,k} = \bar{Z}_n$  and hence

$$\tilde{\eta}_{n,k} = n^{-1/2}([\mathcal{Z}_n]_{R_k(\Delta_n X)} - Q(R_k(\Delta_n X)/(n+1)) - \bar{Z}_n) = n^{-1/2}(\zeta_{n,R_k(\Delta_n X)} - \bar{\zeta}_n).$$

Therefore,

$$\sum_{k=1}^n \tilde{\eta}_{n,k}^2 = n^{-1} \sum_{k=1}^n (\zeta_{n,k} - \bar{\zeta}_n)^2 = n^{-1} \sum_{k=1}^n \zeta_{n,k}^2 - \bar{\zeta}_n^2.$$

Here  $\bar{\zeta}_n^2 = \bar{Z}_n^2 \stackrel{d}{=} n^{-1} Z_1^2 = O_P(n^{-1})$ . Further, denoting by

$$Q_n^*(u) := \sum_{k=1}^n [\mathcal{Z}_n]_k \mathbf{1}\left(u \in \left[\frac{k-1}{n}, \frac{k}{n}\right)\right), \quad u \in (0, 1),$$

the empirical quantile function for  $\mathcal{Z}_n$  and recalling notation  $Q_n$  from (2.1), one has

$$\begin{aligned} n^{-1} \sum_{k=1}^n \zeta_{n,k}^2 &= n^{-1} \sum_{k=1}^n \left([\mathcal{Z}_n]_k - Q\left(\frac{k}{n+1}\right)\right)^2 = \int_0^1 (Q_n^*(u) - Q_n(u))^2 du \\ &\leq 2 \int_0^1 (Q_n^*(u) - Q(u))^2 du + 2 \int_0^1 (Q(u) - Q_n(u))^2 du. \end{aligned}$$

It follows from Theorem 1 in [1] that the first term in the second line is  $O_P(n^{-1} \ln \ln n)$ , whereas the second term in that line is  $O(n^{-1})$  by Lemma 2.1. This completes the proof of (2.10)

Now, for an arbitrary positive sequence  $\varepsilon_n \rightarrow 0$ , set  $\gamma_{n,k} := \varepsilon_n n^{1/2} (\ln \ln n)^{-1/2} \tilde{\eta}_{n,k}$ ,  $k = 1, \dots, n$ . It follows from (2.10) that  $\sum_{k=1}^n \gamma_{n,k}^2 = O_P(\varepsilon_n^2) = o_P(1)$ , and therefore Lemma 2.5 implies the desired relation (2.9) (and hence (2.8)). The assertion of Theorem 1.2 follows from representation (2.6) and relations (2.5), (2.7) and (2.8).  $\square$

## References

- [1] Berthet, P. and Fort, J. C.: Exact rate of convergence of the expected  $W_2$  distance between the empirical and true Gaussian distribution. *Electron. J. Probab.* **25**, (2020), paper no. 12, 16 pp. MR4059190
- [2] Feller, W.: An Introduction to Probability Theory and Its Applications. V. 1. 3rd edn. Wiley, New York–London–Sydney, 1968. xviii+509 pp. MR0228020
- [3] Fomichov, V., González Cázares, J. and Ivanovs, J.: Implementable coupling of Lévy process and Brownian motion. *Stoch. Proc. Appl.* **142**, (2021), 407–431. MR4321329
- [4] González Cázares, J. and Ivanovs, J.: Recovering Brownian and jump parts from high-frequency observations of a Lévy process. *Bernoulli*, **27**, (2021), 2413–2436. MR4303889
- [5] Kallenberg, O.: Probabilistic Symmetries and Invariance Principles. Springer, New York, 2005. xii+510 pp. MR2161313
- [6] Karatzas, I. and Shreve, S. E.: Brownian motion and stochastic calculus. 2nd edn. Springer, New York, 1991. xxiv+470 pp. MR1121940

**Acknowledgments.** The author is grateful to the anonymous referee for spotting a few typos and providing useful comments that helped to improve the exposition of this paper.