Enseign. Math. (2) 72 (2026), 147–159
DOI 10.4171/LEM/1094

A proof of the central limit theorem using the 2-Wasserstein metric

Calvin Wooyoung Chin

Abstract. We prove the Lindeberg–Feller central limit theorem without using characteristic functions or Taylor expansions, but instead by measuring how far a distribution is from the standard normal distribution according to the 2-Wasserstein metric. This falls under the category of renormalization group methods. The facts we need about the metric are explained and proved in detail. We illustrate the idea on a classical version of the central limit theorem before going into the main proof.

Mathematics Subject Classification 2020: 60F05.

Keywords: central limit theorem, 2-Wasserstein metric, renormalization group method, optimal transport.

1. Introduction

Let X and Y be independent random variables with mean 0 and variance 1. The central question of this note is whether the distribution of

$$\frac{X+Y}{\sqrt{2}}$$

is "closer" to the standard normal than those of X and Y in some appropriate sense. Assuming this is true, the idea is that if $X_1, X_2, \ldots$ are independent with mean 0 and variance 1, then

$$\frac{X_1+X_2+X_3+X_4}{\sqrt{4}} = \frac{\frac{X_1+X_2}{\sqrt{2}} + \frac{X_3+X_4}{\sqrt{2}}}{\sqrt{2}}$$

would be closer to the standard normal than $(X_1+X_2)/\sqrt{2}$ and $(X_3+X_4)/\sqrt{2}$, which are in turn closer than X_1, X_2, X_3, and X_4. We might repeat this to show that weighted averages of 8 terms, 16 terms, ... are increasingly closer to the standard normal, hopefully leading to a version of the central limit theorem (CLT).

This idea, referred to as the renormalization group approach, is well known and capable of proving the CLT; see [5] for a nice summary of the literature. When the CLT is proven in this way, the CLT itself is often a test bed, and the actual goal is to apply the same idea to harder problems. In this note, however, we would like to put a little more emphasis on making the proof of the CLT more accessible. Our approach falls under the category that defines an actual topological metric between distributions. Other such approaches include [4, 5], where the Zolotarev metric [9] or a metric based on the characteristic function is used.

Before introducing the Zolotarev metric in [4], the authors briefly discuss the 2-Wasserstein metric but choose not to use it because it is not suited for the application of the Banach contraction method. The 2-Wasserstein distance between (the distributions of) X and Y is given by

$$W_2(X,Y) := \inf_{X',Y'} \sqrt{\mathbf{E}\big[(X'-Y')^2\big]},$$

where X' and Y' range over random variables with the same distributions as X and Y. The point is that the dependence or correlation between X' and Y', i.e., the coupling, is not specified.

The Wasserstein metric is a well-known metric in probability theory that appears in the context of optimal transport and the Wasserstein generative adversarial networks (WGANs) in machine learning, for example. There is a p-Wasserstein metric corresponding to the L^p metric for every $p \in [1,\infty]$. In this note, we use the 2-Wasserstein metric to prove the central limit theorem.

An important property of the 2-Wasserstein metric is the following. Throughout this note, $\Rightarrow$ denotes convergence in distribution.

Proposition 1. *Let $X, X_1, X_2, \ldots$ be random variables with mean 0 and variance 1. If $W_2(X_n, X) \Rightarrow 0$, then $X_n \Rightarrow X$.*

Proposition 2. *Let $X, Y, X_1, X_2, \ldots$ be random variables with mean 0 and variance 1. If $X_n \Rightarrow X$, then $W_2(X_n, Y) \to W_2(X, Y)$.*

Thanks to these facts, we can prove the central limit theorem by examining the 2-Wasserstein distances. Our key result is the following. Throughout the note, let Z be a standard normal random variable.

Theorem 3. *If X and Y are independent random variables with mean 0 and variance 1, then*

$$W_2\left(\frac{X+Y}{\sqrt{2}}, Z\right)^2 \le \frac{W_2(X,Z)^2 + W_2(Y,Z)^2}{2}.$$

The equality holds if and only if both X and Y are standard normal.

The condition for equality is the nontrivial part. Theorem 3 is powerful enough to imply the following theorem that says the sum of small independent random variables has a distribution that is close to the normal, explaining why the normal distribution is so ubiquitous.

Theorem 4 (Lindeberg–Feller). *For each $\varepsilon > 0$, let M_ε be the supremum of*

$$W_2\left(\sum_{j=1}^{n} X_j, Z\right),$$

where $X_1, \dots, X_n$ ranges over any finite sequence of independent mean-zero random variables with $|X_j| \le \varepsilon$ for $j = 1, \dots, n$ and $\sum_{j=1}^{n} \mathbf{E}[X_j^2] = 1$. Then, $M_\varepsilon \to 0$ as $\varepsilon \to 0$.

This is one way to state the Lindeberg–Feller theorem [2, Theorem 3.4.10], which is the most general central limit theorem one typically sees. After the theorem in [2], there is a brief remark on why this implies the usual Lindeberg–Levy CLT.

Our approach to the central limit theorem is more quantitative or “soft analytic” than many classical proofs. Unlike the standard proof [2, Theorem 3.4.1] using characteristic functions or the proof [7, Section 2.2.3] using the moment method, this proof avoids the use of Fourier analysis. Other such approaches include the one by Trotter [8] that replaces characteristic functions with linear operations on some function space, and the proof by the so-called Lindeberg swapping [3]. Unlike those proofs, the one we present here does not even use Taylor expansions.

In Section 2, we will use Theorem 3 to complete the proof idea we sketched in the beginning of this introduction. This part is not new, and similar proofs can be found in the literature cited above. After introducing the notion of inverses of cumulative distribution functions in Section 3, we will prove Theorem 3 in Section 4. In Section 5 that follows, we will prove the Lindeberg–Feller theorem (Theorem 4).

Any proof of the central limit theorem involves some measure theoretic probability. In our case, Propositions 1 and 2 and a couple other facts fall under this category. Since the proofs of these are rather standard and do not contain new idea, we collected them at the end in Section 6 for those who are interested.

2. Proof of a version of the Lindeberg–Lévy CLT

To understand how one can use Theorem 3 to prove the central limit theorem, let us consider the following easier version. Let

$$S_n := X_1 + \cdots + X_n.$$

Theorem 5 (Lindeberg–Lévy, bounded, lacunary). *Let $X_1, X_2, \ldots$ be i.i.d. random variables with $|X_1| \leq B < \infty$, $\mathbf{E}X_1 = 0$, and $\mathbf{E}X_1^2 = 1$. Then,*

$$\frac{S_{2^n}}{\sqrt{2^n}} \Rightarrow Z,$$

where Z is a standard normal random variable.

From this, it is not difficult to derive the usual Lindeberg–Lévy theorem where X_1 can be unbounded and 2^n is replaced with n. However, we omit the details as we will prove the stronger Theorem 4.

We need the following two lemmas, which are proved in Section 6.

Lemma 6. *If there is a $B < \infty$ with $\mathbf{E}X_n^2 \leq B$ for all $n \in \mathbf{N}$, then some subsequence of $(X_n)_{n \in \mathbf{N}}$ converges in distribution.*

Lemma 7. *If $X_n \Rightarrow X$ and there is a $B < \infty$ with $\mathbf{E}X_n^4 \leq B$ for all $n \in \mathbf{N}$, then $\mathbf{E}X_n^2 \Rightarrow \mathbf{E}X^2$ and $\mathbf{E}X_n \to \mathbf{E}X$.*

Proof of Theorem 5. By Theorem 3, the sequence

$$\left(W_2\left(\frac{S_{2^n}}{\sqrt{2^n}}, Z\right)\right)_{n \in \mathbf{N}}$$

is nonincreasing. We want to show that this converges to 0.

Since

$$\mathbf{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^4\right] = \frac{n\mathbf{E}X_1^4 + 3n(n-1)(\mathbf{E}X_1^2)^2}{n^2} \leq B^4 + 3$$

for all $n \in \mathbf{N}$, Lemmas 6 and 7 imply that, for some subsequence $n_1 < n_2 < \cdots$, we have

$$\frac{S_{2^{n_k}}}{\sqrt{2^{n_k}}} \Rightarrow L \quad \text{as } k \to \infty,$$

where L is some random variable with mean 0 and variance 1.

By Proposition 2, we have

$$W_2(L, Z) = \lim_{k \to \infty} W_2\left(\frac{S_{2^{n_k}}}{\sqrt{2^{n_k}}}, Z\right) = \inf_{n \in \mathbf{N}} W_2\left(\frac{S_{2^n}}{\sqrt{2^n}}, Z\right). \tag{1}$$

If L' is an independent copy of L, then

$$\frac{S_{2^{n_k+1}}}{\sqrt{2^{n_k+1}}} \Rightarrow \frac{L + L'}{\sqrt{2}},$$

and thus,

$$W_2\left(\frac{L+L'}{\sqrt{2}}, Z\right) = \lim_{k\to\infty} W_2\left(\frac{S_{2^{n_k+1}}}{\sqrt{2^{n_k+1}}}, Z\right) = \inf_{n\in\mathbf{N}} W_2\left(\frac{S_{2^n}}{\sqrt{2^n}}, Z\right). \tag{2}$$

By (1) and (2), we have

$$W_2\left(\frac{L+L'}{\sqrt{2}}, Z\right)^2 = \frac{W_2(L,Z)^2 + W_2(L',Z)^2}{2}.$$

Theorem 3 then implies that L is standard normal. Since

$$\lim_{n\to\infty} W_2\left(\frac{S_{2^n}}{\sqrt{2^n}}, Z\right) = \inf_{n\in\mathbf{N}} W_2\left(\frac{S_{2^n}}{\sqrt{2^n}}, Z\right) = W_2(L,Z) = 0,$$

we have $S_{2^n}/\sqrt{2^n} \Rightarrow Z$. ■

3. Inverses of CDFs and the 2-Wasserstein metric

For a random variable X, let us denote its cumulative distribution function by $F_X\colon \mathbf{R} \to [0,1]$. We say that $(X_n)_{n\in\mathbf{N}}$ converges in distribution to X and write $X_n \Rightarrow X$ if

$$F_{X_n}(t) \to F_X(t) \quad \text{as } n \to \infty$$

for all $t \in \mathbf{R}$, where F_X is continuous.

However, it is not so easy to use this definition directly to prove convergence in distribution. As a result, one often uses some equivalent condition that is easier to show. Typically, the Portmanteau theorem [6, Theorem 8.4.1] provides such a condition, but in this note, we will instead use *inverses* of cumulative distribution functions. These are well-known tools in probability, and the relevant proofs are rather elementary. We will mention where one can find proofs whenever we omit some details.

If F_X is a strictly increasing function, then its inverse

$$F_X^{-1}\colon (0,1) \to \mathbf{R}$$

has an interesting property: if we view this as a random variable defined on the sample space $(0,1)$, then it has the same distribution as X. To see this, notice that

$$\mathbf{P}(F_X^{-1} \le x) = F_X(x) = \mathbf{P}(X \le x) \quad \text{for all } x \in \mathbf{R}.$$

A similar thing can be done even if F_X is not strictly increasing. In general, we consider the “inverse” $F_X^{\leftarrow}\colon (0,1) \to \mathbf{R}$ given by

$$F_X^{\leftarrow}(t) := \inf\{x \in \mathbf{R} : F_X(x) \ge t\}.$$

It can be shown that $F_X^{\leftarrow}$ is nondecreasing, left-continuous and has the same distribution as X if we view it as a random variable defined on $(0, 1)$. For a proof, see [6, Section 2.5.2].

An important property of inverses for the purposes of this note is the following. It is a known result that appears to date back to [1].

Proposition 8. *For any random variables X and Y with finite variances, we have*

$$\mathbf{E}[XY] \le \int_0^1 F_X^{\leftarrow} F_Y^{\leftarrow}. \tag{3}$$

The full proof of Proposition 8 can be found in Section 6, but the basic idea is simple: it is nothing more than a continuous version of the rearrangement inequality. The inequality says that if $x_1 \le \cdots \le x_n$ and $y_1 \le \cdots \le y_n$, then

$$x_1 y_{\sigma(1)} + \cdots + x_n y_{\sigma(n)} \le x_1 y_1 + \cdots + x_n y_n$$

for any permutation $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$. The pair $(F_X^{\leftarrow}, F_Y^{\leftarrow})$ is a way to couple X and Y so that X increases as Y increases.

Thanks to Proposition 8, we can express $W_2(X, Y)$ in terms of $F_X^{\leftarrow}$ and $F_Y^{\leftarrow}$.

Corollary 9. *For any random variables X and Y with mean 0 and variance 1, we have*

$$W_2(X, Y)^2 = \int_0^1 (F_X^{\leftarrow} - F_Y^{\leftarrow})^2.$$

Proof. For any X' and Y' with the same distribution as X and Y, we have

$$\mathbf{E}\big[(X' - Y')^2\big] = 2 - 2\mathbf{E}[X'Y'] \ge 2 - 2\int_0^1 F_X^{\leftarrow} F_Y^{\leftarrow} = \int_0^1 \big(F_X^{\leftarrow} - F_Y^{\leftarrow}\big)^2.$$

Thus, the infimum of the possible $\mathbf{E}[(X' - Y')^2]$ is $\int_0^1 (F_X^{\leftarrow} - F_Y^{\leftarrow})^2$. ∎

4. Proof of Theorem 3

Theorem 3. *If X and Y are independent random variables with mean 0 and variance 1, then*

$$W_2\left(\frac{X + Y}{\sqrt{2}}, Z\right)^2 \le \frac{W_2(X, Z)^2 + W_2(Y, Z)^2}{2}.$$

The equality holds if and only if both X and Y are standard normal.

Proof. Let $f_1 = F^{\leftarrow}_{X/\sqrt{2}}$, $f_2 = F^{\leftarrow}_{Y/\sqrt{2}}$, and $g = F^{\leftarrow}_{Z/\sqrt{2}}$. We have

$$\int_0^1 (f_1(x)-g(x))^2\,dx = \frac{W_2(X,Z)^2}{2} \quad\text{and}\quad \int_0^1 (f_2(x)-g(x))^2\,dx = \frac{W_2(Y,Z)^2}{2}$$

by Corollary 9. Let $F, G\colon (0,1)^2 \to \mathbf{R}$ be given by

$$F(x,y) = f_1(x) + f_2(y) \quad\text{and}\quad G(x,y) = g(x) + g(y).$$

If we view $(0,1)^2$ as the sample space where area is interpreted as probability, then F and G are random variables having the same distributions as $(X+Y)/\sqrt{2}$ and Z. Notice that

$$\begin{aligned}\text{(4)}\qquad W_2\left(\frac{X+Y}{\sqrt{2}}, Z\right)^2 &\le \int_0^1\int_0^1 (F(x,y)-G(x,y))^2\,dx\,dy \\ &= \int_0^1 (f_1(x)-g(x))^2\,dx + \int_0^1 (f_2(y)-g(y))^2\,dy \\ &= \frac{W_2(X,Z)^2 + W_2(Y,Z)^2}{2}.\end{aligned}$$

Assume that the equality holds, and let us show that X and Y are standard normal. We claim that $G(x_1,y_1) = G(x_2,y_2)$ implies $F(x_1,y_1) = F(x_2,y_2)$. To show this, suppose that $F(x_1,y_1) < F(x_2,y_2)$. Since f_1 and f_2 are left-continuous, for some small $\varepsilon > 0$, the squares $R_1 := (x_1-\varepsilon, x_1] \times (y_1-\varepsilon, y_1]$ and $R_2 := (x_2-\varepsilon, x_2] \times (y_2-\varepsilon, y_2]$ satisfy $\sup F(R_1) < \inf F(R_2)$. Since g is an increasing bijection that is continuous, we can take small squares $S_1 = (a, a+\delta) \times (b, b+\delta) \subset R_1$ and $S_2 = (c, c+\delta) \times (d, d+\delta) \subset R_2$ such that $\inf G(S_1) > \sup G(S_2)$. Let $\phi\colon S_1 \to S_2$ be given by $\phi(a+x, b+y) := (c+x, d+y)$ and define

$$H(x,y) = \begin{cases} F(\phi(x,y)) & \text{if } (x,y) \in S_1, \\ F(\phi^{-1}(x,y)) & \text{if } (x,y) \in S_2, \\ F(x,y) & \text{otherwise.} \end{cases}$$

Since

$$\begin{aligned}&\int_0^1\int_0^1 F(x,y)G(x,y)\,dx\,dy - \int_0^1\int_0^1 H(x,y)G(x,y)\,dx\,dy \\ &\quad= \int_0^\delta\int_0^\delta (F(a+x,b+y) - F(c+x,d+y))G(a+x,b+y)\,dx\,dy \\ &\qquad+ \int_0^\delta\int_0^\delta (F(c+x,d+y) - F(a+x,b+y))G(c+x,d+y)\,dx\,dy \\ &\quad= \int_0^\delta\int_0^\delta (F(a+x,b+y) - F(c+x,d+y)) \\ &\qquad\cdot (G(a+x,b+y) - G(c+x,d+y))\,dx\,dy < 0,\end{aligned}$$

we have

$$\int_0^1 \int_0^1 (H(x,y) - G(x,y))^2 \, dx \, dy < \int_0^1 \int_0^1 (F(x,y) - G(x,y))^2 \, dx \, dy.$$

Since H has the same distribution as F, the definition of W_2 tells us that the equality cannot hold in (4), contrary to our assumption. Thus, we have $F(x_1, y_1) = F(x_2, y_2)$, as claimed.

We now know that F is constant on $G^{-1}(c)$ for all $c \in \mathbf{R}$. Fix an $n \in \mathbf{N}$, and let $x_k \in \mathbf{R}$ be such that $g(x_k) = k/n$ for each $k \in \mathbf{Z}$. Since $(x_k, x_{-k}) \in G^{-1}(0)$ and $(x_{k+1}, x_{-k}) \in G^{-1}(1/n)$, the value of

$$f_1(x_{k+1}) - f_1(x_k) = F(x_{k+1}, x_{-k}) - F(x_k, x_{-k})$$

is the same for all $k \in \mathbf{Z}$. Notice that $g(x_{k+1}) - g(x_k)$ is the same for all $k \in \mathbf{Z}$, too. Since this argument applies for all $n \in \mathbf{N}$ and f_1 is nondecreasing, we must have $f_1 = \alpha g + \beta$ for some $\alpha > 0$ and $\beta \in \mathbf{R}$. As $\int f_1 = 0$ and $\int f_1^2 = 1$, the only possibility is $f_1 = g$. The same argument gives $f_2 = g$. ■

5. Proof of the Lindeberg–Feller CLT

Theorem 4 (Lindeberg–Feller). *For each $\varepsilon > 0$, let M_ε be the supremum of*

$$W_2\left(\sum_{j=1}^{n} X_j, Z\right),$$

where $X_1, \dots, X_n$ ranges over any finite sequence of independent mean-zero random variables with $|X_j| \le \varepsilon$ for $j = 1, \dots, n$ and $\sum_{j=1}^{n} \mathbf{E}[X_j^2] = 1$. Then, $M_\varepsilon \to 0$ as $\varepsilon \to 0$.

Proof. Let $M := \lim_{\varepsilon \to 0+} M_\varepsilon$. For each $n \in \mathbf{N}$, take independent mean-zero random variables $X_{n1}, \dots, X_{nm_n}$ ($m_n \in \mathbf{N}$) with $|X_{nj}| \le 1/n$ for $j = 1, \dots, m_n$ and $\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2] = 1$ such that

$$W_2\left(\sum_{j=1}^{m_n} X_{nj}, Z\right) \ge M - 1/n. \tag{5}$$

Since $\max_{j=1}^{m_n} \mathbf{E}[X_{nj}^2] \to 0$ as $n \to \infty$, we can choose $1 \le k_n \le m_n$ for each $n \in \mathbf{N}$ so that $v_n := \sum_{j=1}^{k_n} \mathbf{E}X_{nj}^2 \to 1/2$ as $n \to \infty$. By passing to a subsequence if necessary, we may assume that

$$\frac{1}{\sqrt{v_n}} \sum_{j=1}^{k_n} X_{nj} \Rightarrow L_1 \quad \text{and} \quad \frac{1}{\sqrt{1 - v_n}} \sum_{j=k_n+1}^{m_n} X_{nj} \Rightarrow L_2$$

for some L_1 and L_2, by using Lemma 6.

Since $|X_{nj}| \leq 1$, we have $\mathbf{E}[X_{nj}^4] \leq \mathbf{E}[X_{nj}^2]$, and thus,

$$\mathbf{E}\left[\left(\sum_{j=1}^{k_n} X_{nj}\right)^4\right] \leq \sum_{j=1}^{k_n} \mathbf{E}[X_{nj}^4] + 3\left(\sum_{j=1}^{k_n} \mathbf{E}X_{nj}^2\right)^2 \leq v_n + 3v_n^2$$

for all $n \in \mathbf{N}$. By Lemma 7, L_1 has mean 0 and variance 1. The same holds for L_2. By Proposition 2, we have

$$W_2(L_1, Z) = \lim_{n\to\infty} W_2\left(\frac{1}{\sqrt{v_n}} \sum_{j=1}^{k_n} X_{nj}, Z\right) \leq M.$$

Similarly, we have $W_2(L_2, Z) \leq M$.

Notice that we have

$$\sum_{j=1}^{m_n} X_{nj} \Rightarrow \frac{L_1 + L_2}{\sqrt{2}}.$$

By Proposition 2 and (5), we have

$$W_2\left(\frac{L_1 + L_2}{\sqrt{2}}, Z\right) = \lim_{n\to\infty} W_2\left(\sum_{j=1}^{m_n} X_{nj}, Z\right) = M.$$

Since

$$M^2 = W_2\left(\frac{L_1 + L_2}{\sqrt{2}}, Z\right)^2 \leq \frac{W_2(L_1, Z)^2 + W_2(L_2, Z)^2}{2} \leq M^2,$$

Theorem 3 implies that L_1 and L_2 are standard normal. Therefore, we have

$$M = W_2(Z, Z) = 0.$$

■

6. Measure theoretic probabilistic proofs

We can prove the following by directly using the definition of convergence in distribution.

Lemma 6. *If there is a $B < \infty$ with $\mathbf{E}X_n^2 \leq B$ for all $n \in \mathbf{N}$, then some subsequence of $(X_n)_{n\in\mathbf{N}}$ converges in distribution.*

Proof. Let $q \in \mathbf{Q}$. By the Bolzano–Weierstrass theorem, there is a subsequence of $(F_{X_n}(q))_{n\in\mathbf{N}}$ that converges to a finite number. Using diagonalization, we can find $n_1 < n_2 < \cdots$ such that

$$\lim_{k\to\infty} F_{X_{n_k}}(q) = F_q \quad \text{for some } F_q \in \mathbf{R}$$

for all rational $q \in (0, 1)$.

Now, let $F\colon \mathbf{R} \to [0,1]$ be given by

$$F(x) := \inf\{F_q : q \in (x, \infty) \cap \mathbf{Q}\}.$$

Then, it can be shown that F is a nondecreasing right continuous function such that $F_{n_k}(x) \to F(x)$ as $k \to \infty$ for all $x \in \mathbf{R}$ at which F is continuous. For more details, see [6, Lemma 9.6.2].

By Chebyshev's inequality, for each $M > 0$, we have

$$F_{X_n}(-M) = \mathbf{P}(X_n \le -M) \le B/M^2 \quad \text{and} \quad 1 - F_{X_n}(M) = \mathbf{P}(X_n > M) \le B/M^2.$$

This implies $F(-M) \le B/M^2$ and $F(M) \ge 1 - B/M^2$, and thus, we have $F(x) \to 1$ as $x \to \infty$ and $F(x) \to 0$ as $x \to -\infty$. This shows that F is a cumulative distribution function. ∎

An important property of inverses is that $X_n \Rightarrow X$ if and only if

$$F_{X_n}^{\leftarrow}(t) \to F_X^{\leftarrow}(t) \quad \text{as } n \to \infty$$

for all $t \in (0,1)$, where $F_X^{\leftarrow}$ is continuous. See [6, Proposition 8.3.1 and Theorem 8.3.2] for a proof. Replacing convergence in distribution with convergence with probability 1 is useful when we prove facts like the following.

Lemma 7. *If $X_n \Rightarrow X$ and there is a $B < \infty$ with $\mathbf{E}X_n^4 \le B$ for all $n \in \mathbf{N}$, then $\mathbf{E}X_n^2 \Rightarrow \mathbf{E}X^2$ and $\mathbf{E}X_n \to \mathbf{E}X$.*

Proof. Since we can replace X_n and X with $F_{X_n}^{\leftarrow}$ and $F_X^{\leftarrow}$, it is enough to assume $X_n \to X$ with probability 1 instead of $X_n \Rightarrow X$. Let $M > 0$ be such that

$$\mathbf{P}(X^2 = M) = 0.$$

Since $X_n 1_{\{|X_n| \le M\}} \to X 1_{\{|X| \le M\}}$ on $\{|X| \ne M\}$, which has probability 1, the bounded convergence theorem implies

$$\mathbf{E}[X_n^2; X_n^2 \le M] \to \mathbf{E}[X^2; X^2 \le M].$$

Here, $\mathbf{E}[X; \varphi(X)]$ denotes $\mathbf{E}[X 1_{\{\varphi(X)\}}]$. Since

$$\mathbf{E}[X_n^2; X_n^2 > M] \le \mathbf{E}[X_n^4/M^2] \le \frac{B}{M^2},$$

letting $M \to \infty$ gives $\mathbf{E}[X_n^2] \to \mathbf{E}[X^2]$. The proof for $\mathbf{E}[X_n] \to \mathbf{E}[X]$ is similar. ∎

Out interest in inverses in this note came from the following "continuous version" of the rearrangement inequality. The proof also resembles its discrete analogue.

Proposition 8. *For any random variables X and Y with finite variances, we have*

$$\mathbf{E}[XY] \le \int_0^1 F_X^{\leftarrow} F_Y^{\leftarrow}. \tag{3}$$

Proof. First, assume that X and Y are simple; i.e., there are $a_1, \dots, a_n, b_1, \dots, b_m \in \mathbf{R}$ such that $X \in \{a_1, \dots, a_n\}$ and $Y \in \{b_1, \dots, b_m\}$. Let X' and Y' be random variables with the same distribution as X and Y. Since the set of possible

$$(\mathbf{P}(X' = a_i, Y' = b_j))_{i \le n, j \le m}$$

is compact, the supremum of possible $\mathbf{E}[X'Y']$ is attained.

If (X', Y') has a different joint distribution from $(F_X^{\leftarrow}, F_Y^{\leftarrow})$, then there are $a, b, c, d \in \mathbf{R}$ with

$$a < c, \quad b > d, \quad \mathbf{P}(X' = a, Y' = b) > 0, \quad \text{and} \quad \mathbf{P}(X' = c, Y' = d) > 0;$$

i.e., "X' does not increase as Y' increases." Now, take events $E \subset \{X' = a, Y' = b\}$ and $F \subset \{X' = c, Y' = d\}$ with the same nonzero probability and swap the values of Y' on E and F to form a new random variable Y''. (We can assume that the underlying sample space is good enough to do this.) Since

$$\mathbf{E}[XY''] - \mathbf{E}[XY'] = (ad + bc - ab - cd)\mathbf{P}(E) = (c - a)(b - d)\mathbf{P}(E) > 0,$$

the supremum mentioned above is not attained by (X', Y'). Since this is true whenever (X', Y') has a different joint distribution from $(F_X^{\leftarrow}, F_Y^{\leftarrow})$, the supremum must be attained by $(F_X^{\leftarrow}, F_Y^{\leftarrow})$. This proves (3).

Now, assume that X and Y are not necessarily simple. We can take simple $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ such that $|X_n| \le |X|$, $|Y_n| \le |Y|$, $X_n \to X$, and $Y_n \to Y$, $|F_{X_n}^{\leftarrow}| \le |F_X^{\leftarrow}|$, $|F_{Y_n}^{\leftarrow}| \le |F_Y^{\leftarrow}|$, $F_{X_n}^{\leftarrow} \to F_X^{\leftarrow}$, and $F_{Y_n}^{\leftarrow} \to F_Y^{\leftarrow}$. Since $|X_n Y_n| \le |XY|$, the dominated convergence theorem [6, Theorem 5.3.3] gives $\mathbf{E}[X_n Y_n] \to \mathbf{E}[XY]$. Similarly, we have $\int_0^1 F_{X_n}^{\leftarrow} F_{Y_n}^{\leftarrow} \to \int_0^1 F_X^{\leftarrow} F_Y^{\leftarrow}$. As $\mathbf{E}[X_n Y_n] \le \int_0^1 F_{X_n}^{\leftarrow} F_{Y_n}^{\leftarrow}$, letting $n \to \infty$ gives (3). ■

Recall the following corollary.

Corollary 9. *For any random variables X and Y with mean 0 and variance 1, we have*

$$W_2(X, Y)^2 = \int_0^1 (F_X^{\leftarrow} - F_Y^{\leftarrow})^2.$$

Using this characterization of 2-Wasserstein metric, we can prove the relation between W_2 and the convergence in distribution.

Proposition 1. *Let $X, X_1, X_2, \dots$ be random variables with mean 0 and variance 1. If $W_2(X_n, X) \Rightarrow 0$, then $X_n \Rightarrow X$.*

Proof. Assume that $X_n \not\Rightarrow X$. Then, there is a $t \in (0,1)$, where $F_X^{\leftarrow}$ is continuous such that $F_{X_n}^{\leftarrow}(t) \not\to F_X^{\leftarrow}(t)$. Let $\varepsilon > 0$ and $n_1 < n_2 < \cdots$ be such that

$$\left|F_{X_{n_k}}^{\leftarrow}(t) - F_X^{\leftarrow}(t)\right| \geq 2\varepsilon \quad \text{for all } k \in \mathbf{N}.$$

Since $F_X^{\leftarrow}$ is continuous at t, there is a $\delta > 0$ such that $|F_X^{\leftarrow}(s) - F_X^{\leftarrow}(t)| \leq \varepsilon$ for all $s \in (t-\delta, t+\delta)$.

Let $k \in \mathbf{N}$. If $F_{X_{n_k}}^{\leftarrow}(t) \leq F_X^{\leftarrow}(t) - 2\varepsilon$, then on $(t-\delta, t]$ we have $F_X^{\leftarrow} - F_{X_{n_k}}^{\leftarrow} \geq \varepsilon$. If $F_{X_{n_k}}^{\leftarrow}(t) \geq F_X^{\leftarrow}(t) + 2\varepsilon$, then on $[t, t+\delta)$ we have $\inf X_{n_k} - F_X^{\leftarrow} \geq \varepsilon$. In either case, we have

$$\int_0^1 \left(F_{X_{n_k}}^{\leftarrow} - F_X^{\leftarrow}\right)^2 \geq \varepsilon^2\delta.$$

Since this is true for all $k \in \mathbf{N}$, we have $W_2(X_n, X) \not\to 0$. ∎

Proposition 2. *Let* $X, Y, X_1, X_2, \ldots$ *be random variables with mean* 0 *and variance* 1. *If* $X_n \Rightarrow X$, *then* $W_2(X_n, Y) \to W_2(X, Y)$.

Proof. Since we can replace X_n, X, and Y with $F_{X_n}^{\leftarrow}$, $F_X^{\leftarrow}$, and $F_Y^{\leftarrow}$, it is enough to show that if $X_n \to X$ with probability 1, then $\mathbf{E}[(X_n - Y)^2] \to \mathbf{E}[(X - Y)^2]$. As $\mathbf{E}X_n^2 = \mathbf{E}X^2 = \mathbf{E}Y^2 = 1$, it suffices to show $\mathbf{E}[(X_n - X)Y] \to 0$. Since

$$\left|\mathbf{E}[(X_n - X)Y]\right|^2 \leq \mathbf{E}\left[(X_n - X)^2\right]\mathbf{E}[Y^2]$$

by the Cauchy–Schwarz inequality, showing $\mathbf{E}[(X_n - X)^2] \to 0$ is enough.

Let $\varepsilon > 0$ be given. Choose $M > 1$ with $\mathbf{E}[X^2; |X| > M - 1] \leq \varepsilon$. Let $\phi\colon \mathbf{R} \to \mathbf{R}$ be the continuous function that satisfies $\phi(x) = x^2$ if $|x| \leq M - 1$, $\phi(x) = 0$ if $|x| \geq M$, and is linear on the remaining intervals. By the bounded convergence theorem, we have

$$\lim_{n\to\infty} \mathbf{E}[\phi(X_n)] = \mathbf{E}[\phi(X)] \geq 1 - \varepsilon,$$

and thus, $\limsup_{n\to\infty} \mathbf{E}[X_n^2; |X_n| > M] \leq \varepsilon$.

Since $(X_n - X)^2 \leq 2(X_n^2 + X^2)$, we have

$$\begin{aligned}\mathbf{E}\left[(X_n - X)^2; |X_n - X| > 2M\right] &\leq 2\mathbf{E}[X_n^2 + X^2; |X_n| \geq |X|, |X_n| > M] \\ &\quad + 2\mathbf{E}[X_n^2 + X^2; |X_n| < |X|, |X| > M] \\ &\leq 4\mathbf{E}[X_n^2; |X_n| > M] + 4\mathbf{E}[X^2; |X| > M].\end{aligned}$$

This implies

$$\limsup_{n\to\infty} \mathbf{E}\left[(X_n - X)^2; |X_n - X| > 2M\right] \leq 8\varepsilon.$$

On the other hand, the bounded convergence theorem implies

$$\mathbf{E}\left[(X_n - X)^2; |X_n - X| \leq 2M\right] \to 0.$$

Combining the two pieces together and letting $\varepsilon \to 0$ finish the proof. ∎

References

[1] P. Bártfai, Über die Entfernung der Irrfahrtswege. *Studia Sci. Math. Hungar.* **5** (1970), 41–49. Zbl 0274.60048 MR 0275499

[2] R. Durrett, *Probability—theory and examples*. 5th edn., Camb. Ser. Stat. Probab. Math. 49, Cambridge University Press, Cambridge, 2019. Zbl 1440.60001 MR 3930614

[3] J. W. Lindeberg, Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** (1922), no. 1, 211–225. Zbl 48.0602.04 MR 1544569

[4] R. Neininger and L. Rüschendorf, A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14** (2004), no. 1, 378–418. Zbl 1041.60024 MR 2023025

[5] S. Ott, A note on the renormalization group approach to the central limit theorem. 2023, arXiv:2303.13905v3.

[6] S. I. Resnick, *A probability path*. Birkhäuser, Boston, MA, 1999. Zbl 0944.60002 MR 1664717

[7] T. Tao, *Topics in random matrix theory*. Grad. Stud. Math. 132, American Mathematical Society, Providence, RI, 2012. Zbl 1256.15020 MR 2906465

[8] H. F. Trotter, An elementary proof of the central limit theorem. *Arch. Math.* **10** (1959), 226–234. Zbl 0086.34002 MR 0108847

[9] V. M. Zolotarev, Metric distances in spaces of random variables and their distributions. *Math. USSR, Sb.* **30** (1976), no. 3, 373–401. Zbl 0383.60022

(Reçu le 23 mars 2024)

Calvin Wooyoung Chin, Apple Inc., One Apple Park Way, Cupertino, CA 95014, USA;
e-mail: wooyoung_chin@apple.com; *e-mail:* cwychin@icloud.com