# A BERRY–ESSÉEN TYPE ESTIMATE FOR LÉVY'S METRIC

Sebastian Maaß

ETH Zürich, Departement Mathematik, Rämistrasse 101, 8092 Zürich, Switzerland
E-Mail: sebastian@maass.name, URL: http://www.sebastian.maass.name

**Abstract**

We present some upper estimates for Lévy's metric and, as an application, a Berry–Esséen type estimate for the rate of convergence in the Central Limit Theorem in Lyapunov's version in terms of Lévy's metric which improves the original one.

## 1 Introduction

All results providing an estimate of the speed of convergence to the normal distribution can be classified into four groups, on the one hand by using Kolmogorov's metric or any other (e.g. Lévy's metric), and on the other hand by using characteristic functions in the proof or working on the original space of distribution functions. Except for the case of calculating the distance w.r.t. Lévy's metric without using characteristic functions, all variants can be found in the literature. In this paper, we therefore present some upper estimates for Lévy's metric and, as an application, a Berry–Esséen type estimate for the Central Limit Theorem in Lyapunov's version in terms of Lévy's metric which improves the original one.

Our results are motivated as follows. In lectures on probability theory, characteristic functions are typically introduced as a purely technical tool mainly to prove the Central Limit Theorem. By proving the latter directly on the space of distribution functions, we avoid technical steps which could distract from the underlying mathematical ideas. Furthermore, the use of Lévy's metric instead of Kolmogorov's metric has the advantage that the former always metricizes convergence in distribution, whereas the latter only metricizes convergence in distribution to continuous limit distributions (which admittedly is sufficient for the proof of the Central Limit Theorem). Finally, our estimate of the rate of convergence to the normal distribution w.r.t. Lévy's metric is better than the original estimate from Berry by far and even topical estimates w.r.t. Kolmogorov's metric are not better than ours.

*AMS 2000 subject classifications.* 60F05, 41A25
*Key words and phrases.* Lévy's metric, Central Limit Theorem, Berry–Esséen Theorem.

Since the use of Lévy's metric is not very common, we provide some historical notes. In 1925, Lévy introduced a metric on the space of distribution functions in a very informal way (cf. Lévy (1925) [11, p. 194 – 195, 199 – 200]). In fact, he defined the distance between two distribution functions to be the Hausdorff distance (without referring to it) between the graphs of these distribution functions in $\mathsf{R}^2$ after filling the jumps with line segments. In 1937, Lévy put his definition in concrete terms in a footnote (cf. Lévy (1937) [12, p. 241]) and, in two different versions, in a note (cf. Fréchet (1937) [7, p. 333 – 334]). Unfortunately, the most intuitive and least technical parts of Lévy's work regarding his metric (cf. Lévy (1925) [11, p. 194 – 195] and Fréchet (1937) [7, p. 333 – 334]) are almost completely disregarded in the literature. Hence it is not surprising that some facts concerning Lévy's metric – like the relation to the Hausdorff distance (cf. Zolotarev (1997) [15, p. 64]) – have been reinvented. By mostly presenting a definition without its simple geometrical interpretation in the literature, Lévy's metric mainly got the status of being a curiosity, at best good enough for exercises in probability books.

This paper is organized as follows. In Section 2, Lévy's metric is defined and elementary properties are proved. We also state some geometrical interpretations of Lévy's metric. Relations between Lévy's metric, Fan's metric, and Kolmogorov's metric are shown in Section 3. Some of these relations can be used to improve estimates of Lévy's metric and to show easily the well-known fact that stochastic convergence implies convergence in distribution. In Section 4, we provide some new estimates for Lévy's metric only using the absolute moments of the random variables involved. One of these is used to estimate the rate of convergence to the normal distribution w.r.t. Lévy's metric and we compare this result with existing ones.

## 2 Definition and basic properties

Let $(\Omega, \mathcal{A}, P)$ denote a probability space and for any random variable $X$ let $F_X$ denote the distribution function of $X$, i.e. $F_X(x) = P(X \le x)$. $E(X)$ will denote the expected value of $X$, $V(X)$ will denote the variance of $X$.

**Proposition 2.1** *For two random variables $X, Y : \Omega \to \mathsf{R}$, let $d_L(X,Y)$ be defined by*

$$d_L(X,Y) := \inf\left\{ h \ge 0 \,\middle|\, F_X(x) \le F_Y(x+h) + h, \; F_Y(x) \le F_X(x+h) + h \;\forall x \in \mathsf{R} \right\}.$$

*Then $d_L$ is a pseudo-metric on the space of random variables.*

**Definition 2.2** *The pseudo-metric $d_L$ is called **Lévy's metric**.*

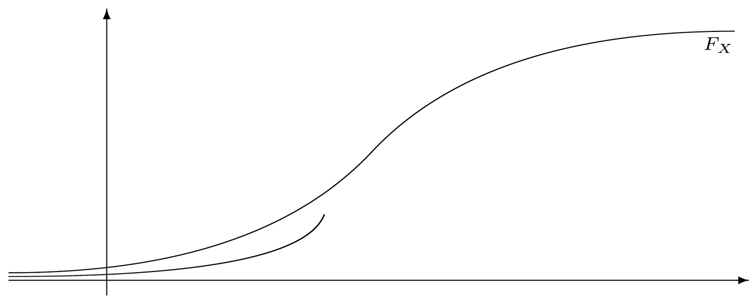
Usually, Lévy's metric is defined on the set of distribution functions (where it actually is a metric) instead on their corresponding random variables. Since all estimates of Lévy's metric in this article only use absolute moments of the random variables involved, we have slightly changed the domain.

The pseudo-metric $d_L$ has an intuitive geometrical interpretation. Both conditions in the definition of Lévy's metric can be rewritten as $F_Y(x+h) \geq F_X(x)-h$ and $F_Y(x-h) \leq F_X(x)+h$ for all $x \in \mathsf{R}$. If $F_Y$ is continuous then, by the Mean Value Theorem, these conditions guarantee that $d_L \leq h$ if and only if $F_Y$ meets every square $S_{x,h}$ with corners at $(x-h, F_X(x)+h)$ and $(x+h, F_X(x)-h)$ for all $x \in \mathsf{R}$. Generally, $d_L \leq h$ if and only if the completed graph $\overline{F_Y}$ of the distribution function $F_Y$, being defined as a subset of $\mathsf{R}^2$ by $\overline{F_Y} := \{(x,y) \in \mathsf{R}^2 \mid y \in [\lim_{x' \nearrow x} F_Y(x'), \lim_{x' \searrow x} F_Y(x')]\}$, meets every square $S_{x,h}$, $x \in \mathsf{R}$. This condition is equivalent to the claim that the Hausdorff distance (w.r.t. the metric $d$ on $\mathsf{R}^2$, defined by $d((x_1,x_2),(y_1,y_2)) := \max\{|x_1-y_1|, |x_2-y_2|\}$), between the sets $\overline{F_X}$ and $\overline{F_Y}$,

$$d_{\text{Hausdorff}}(\overline{F_X}, \overline{F_Y}) = \max\left\{ \max_{\bar{x} \in \overline{F_X}} \min_{\bar{y} \in \overline{F_Y}} d(\bar{x}, \bar{y}), \max_{\bar{y} \in \overline{F_Y}} \min_{\bar{x} \in \overline{F_X}} d(\bar{x}, \bar{y}) \right\},$$

is smaller than or equal to $h$. Therefore, Lévy's metric can be expressed as the Hausdorff distance on the set of completed graphs of distribution functions, i.e.

$$d_L(X,Y) = d_{\text{Hausdorff}}(\overline{F_X}, \overline{F_Y}).$$



**Proof.** It is obvious that $M \supset ]d_L(X,Y), \infty[$. By continuity from the right of $F_X$, resp. $F_Y$, we obtain $d_L(X,Y) \in M$. □

We now collect some rather elementary results on Lévy's metric.

**Proposition 2.4** *Let $X, Y$ be two random variables. Then*

*(a)* $d_L(X+c, Y+c) = d_L(X,Y)$ *for all* $c \in \mathsf{R}$,

*(b)* $d_L(X,Y) = d_L(-X,-Y)$,

*(c)* $d_L(X_1+X_2, Y_1+Y_2) \le d_L(X_1,Y_1) + d_L(X_2,Y_2)$,

*(d)* $d_L(X,Y) \le d_L(|X-Y|, 0)$.

Part (c) is already known (cf. Zolotarev (1967) [14, Lemma 1]) but our proof is shorter and easier. The parts (a) and (b) can intuitively be seen by interpreting Lévy's metric geometrically as the Hausdorff distance between the corresponding completed distribution functions since the geometrical relation of the distribution functions w.r.t. each other does not change by the given operations, translation and reflection w.r.t. the point $(0, \frac{1}{2})$. Nevertheless, a formal proof will be given below.

**Proof.**

(a) This follows directly from the definition of Lévy's metric and from $F_{x+c}(x) = F_X(x-c)$.

(b) We prove $h > d_L(X,Y)$ implies $h \ge d_L(-X,-Y)$ and $h > d_L(-X,-Y)$ implies $h \ge d_L(X,Y)$. Suppose $h > d_L(X,Y)$. Then there exists an $\varepsilon > 0$ such that $F_X(x) \le F_Y(x+h-\varepsilon)+h-\varepsilon$ and $F_Y(x) \le F_X(x+h-\varepsilon)+h-\varepsilon$ for all $x \in \mathsf{R}$. Furthermore,

$$\begin{aligned}
& F_X(x) \le F_Y(x+h-\varepsilon) + h - \varepsilon \\
\Leftrightarrow\ & P(-X \ge -x) \le P(-Y \ge -x-h+\varepsilon) + h - \varepsilon \\
\Leftrightarrow\ & P(-X < -x) \ge P(-Y < -x-h+\varepsilon) - h + \varepsilon \\
\Leftrightarrow\ & P(-X < -x+h) + h \ge P(-Y < -x+\varepsilon) + \varepsilon \\
\Rightarrow\ & F_{-Y}(-x) \le F_{-X}(-x+h) + h.
\end{aligned}$$

Analogously, we conclude $F_{-X}(-x) \le F_{-Y}(-x+h)+h$ from $h > d_L(X,Y)$, i.e. $h > d_L(X,Y)$ implies $h \ge d_L(-X,-Y)$. The reverse direction, i.e. $h > d_L(-X,-Y)$ implies $h \ge d_L(X,Y)$, can be obtained in the same way.

(c) Let $h_i := d_L(X_i, Y_i)$, $i = 1, 2$. Then

$$\begin{aligned}
F_{X_1+X_2}(x) &= \int F_{X_1}(x-y)\, dF_{X_2}(y) \\
&\le \int F_{Y_1}(x-y+h_1) + h_1\, dF_{X_2}(y) \\
&= \int F_{X_2}(x-y+h_1)\, dF_{Y_1}(y) + h_1 \\
&\le \int F_{Y_2}(x-y+h_1+h_2) + h_2\, dF_{Y_1}(y) + h_1 \\
&= F_{Y_1+Y_2}(x+h_1+h_2) + h_1 + h_2
\end{aligned}$$

Analogously, we obtain $F_{Y_1+Y_2}(x) \leq F_{X_1+X_2}(x + h_1 + h_2) + h_1 + h_2$ and thus the desired inequality.

(d) From the triangular inequality follows

$$d_L(X, Y) \leq d_L(X - Y, 0) + d_L(Y, Y) = d_L(X - Y, 0).$$

Since $F_{|X-Y|} \leq F_{X-Y}$ we find for every $h \geq d_L(|X - Y|, 0)$

(i) $F_0(x) \leq F_{|X-Y|}(x + h) + h \leq F_{X-Y}(x + h) + h \quad$ for all $x \in \mathsf{R}$.

The inequality

(ii) $F_{X-Y}(x) \leq F_0(x + h) + h \quad$ for all $x \geq -h$

is trivially true.
To prove

(iii) $F_{X-Y}(x) \leq F_0(x + h) + h \quad$ for all $x < -h$,

we calculate, using $F_{|X-Y|}(h) \geq F_0(0) - h = 1 - h$,

$$\begin{aligned}\sup_{x<-h} F_{X-Y}(x) &= \sup_{x>h} P(Y - X \geq x) \\ &\leq \sup_{x>h} P(|X - Y| \geq x) \\ &= 1 - \inf_{x>h} P(|X - Y| < x) \\ &= 1 - F_{|X-Y|}(h) \\ &\leq h \\ &= F_0(x + h) + h.\end{aligned}$$

□

# 3 Relations between Lévy's metric and other probability metrics

Probability metrics are commonly introduced to metricize different types of convergence. In this section, we mention Fan's metric and Kolmogorov's metric and enlighten their relations to Lévy's metric.

It is well-known that Lévy's metric metricizes convergence in distribution (cf. e.g. Galambos (1988) [8, Section 4.3]), i.e. for a sequence of random variables $X_n$ holds

$$F_{X_n}(x) \to F_X(x) \;\; \forall \text{ continuity points } x \text{ of } F_X \iff d_L(X_n, X) \to 0.$$

Another probability metric which will turn out to be nicely related to Lévy's metric is ***Fan's metric*** $d_F$ which is defined on the set of random variables by (cf. Fan (1944) [6] or Dudley (1989) [4, p. 226])

$$d_F(X, Y) := \inf\big\{h \in \mathsf{R} \mid P(|X - Y| > h) \leq h\big\}.$$

It is well-known (cf. e.g. Dudley (1989) [4, Theorem 9.2.2]) that this metric metricizes stochastic convergence, i.e.

$$\forall\, \varepsilon > 0 : \; P(|X_n - X| > \varepsilon) \to 0 \iff d_F(X_n, X) \to 0.$$

A third metric of interest in our context is

$$d_K(X,Y) := \|F_X - F_Y\|_\infty,$$

sometimes referred to as ***Kolmogorov's metric***. As it can easily be shown from the definition, Kolmogorov's metric metricizes convergence in distribution if and only if the limit distribution function is continuous.

In the following proposition, we collect some relations between Lévy's metric, Fan's metric and Kolmogorov's metric. Part (c) is well-known and Part (d) has already been stated (without proof) by Zolotarev (cf. Zolotarev (1997) [15, p. 65]).

**Proposition 3.1** *Let $X, Y$ be random variables. Then*

*(a)* $d_F(X,Y) = d_L(|X-Y|, 0)$.

*(b)* $d_L \le d_F$.

*(c)* $d_L \le d_K$.

*(d)* $d_K(X,Y) \le (1 + \|F'_X\|_\infty) \cdot d_L(X,Y)$ *if $F_X$ is di erentiable.*

**Proof.**

(a) One easily verifies that the inequalities in the definition of $d_L(|X-Y|, 0)$ hold for all $h > d_F(X,Y)$ and do not hold for all $h < d_F(X,Y)$.

(b) This follows directly from (a) and Proposition 2.4 (d).

(c) Elementary calculations show that the inequalities in the definition of $d_L$ hold for all $h > d_K(X,Y)$.

(d) Let $x \in \mathsf{R}$ and suppose $F_Y(x) > F_X(x)$. Then, by differentiability of $F_X$, $F_X(x + d_L(X,Y)) \le F_X(x) + d_L(X,Y) \cdot \|F'_X\|_\infty$ and, by definition of Lévy's metric and Lemma 2.3, $F_Y(x) \le F_X(x + d_L(X,Y)) + d_L(X,Y)$. Thus $F_Y(x) \le F_X(x) + (1 + \|F'_X\|_\infty) \cdot d_L(X,Y)$.
Analogously, we obtain $F_Y(x) \ge F_X(x) - (1 + \|F'_X\|_\infty) \cdot d_L(X,Y)$ in the case $F_Y(x) < F_X(x)$ from $F_X(x - d_L(X,Y)) \ge F_X(x) - d_L(X,Y) \cdot \|F'_X\|_\infty$ and $F_Y(x) \ge F_X(x - d_L(X,Y)) - d_L(X,Y)$. □

These results have important implications. First, from Proposition 3.1 (b) directly follows the well-known fact that stochastic convergence implies convergence in distribution. Second, Proposition 3.1 (b) can and will in Corollary 4.8 be used to adopt upper estimates for Fan's metric between two random variables as some w.r.t. Lévy's metric. Third, Proposition 3.1 (c) and (d) will help to compare the rate of convergence of a sequence of random variables to the standard normal distribution when one is given w.r.t. Kolmogorov's metric and the other in terms of Lévy's metric.

We conclude this section with stating an upper estimate for Fan's metric between two random variables which in some cases will improve estimates w.r.t. Lévy's metric in the way remarked in the preceding paragraph. This estimate has the advantage that it only uses the variances of the random variables.

**Proposition 3.2** *Let $X, Y$ be two random variables with $E(X) = E(Y)$. Then*

$$d_F(X,Y) \le \sqrt[m+1]{E(|X-Y|^m)} = \left(\|X-Y\|_m\right)^{\frac{m}{m+1}}$$

*for all $m \in \mathbb{N}$. Additionally, if $X, Y$ are independent then*

$$d_F(X,Y) \le \sqrt[3]{V(X)+V(Y)}.$$

**Proof.** First, suppose $d_F(X,Y) > 0$. From the definition of Fan's metric follows $d_F(X,Y) - \varepsilon < 1 - F_{|X-Y|}(d_F(X,Y) - \varepsilon)$ for every $\varepsilon \in ]0, d_F(X,Y)[$. By the Generalized Chebyshev Inequality and by $E(X) = E(Y)$, we get $1 - F_{|X-Y|}(d_F(X,Y) - \varepsilon) \le E(|X-Y|^m)(d_F(X,Y) - \varepsilon)^{-m}$. This implies $d_F(X,Y) - \varepsilon < \sqrt[m+1]{E(|X-Y|^m)}$ and since $\varepsilon \in ]0, d_F(X,Y)[$ was chosen arbitrarily, we obtain $d_F(X,Y) \le \sqrt[m+1]{E(|X-Y|^m)}$. This inequality obviously also holds for $d_F(X,Y) = 0$. Now suppose $X$ and $Y$ are independent. Then

$$d_F(X,Y) \le \sqrt[3]{V(X-Y)} = \sqrt[3]{V(X)+V(Y)}.$$ □

# 4 Main Results

The Central Limit Theorem in Lyapunov's version states that whenever

$$\frac{1}{\sigma^3(S_n)} \sum_{i=1}^{n} E\Big(|X_i|^3\Big) \xrightarrow[n\to\infty]{} 0 \tag{4.1}$$

holds for a sequence $(X_n)_{n\in\mathbb{N}}$ of independent identically distributed random variables with $E(X_n) = 0$ for all $n \in \mathbb{N}$ and having finite second and third absolute moments then the normalized partial sum $\frac{S_n}{\sigma(S_n)}$, $S_n := \sum_{i=1}^{n} X_i$, is asymptotically standard normal distributed. The Berry–Esséen Theorem then provides an estimate of the rate of convergence of the distributions of the sequence of partial sums to the standard normal distribution in terms of Kolmogorov's metric depending on the converging term in formula (4.1) (cf. e.g. Berry (1941) [1, Theorem 5]),

$$d_K\left(\frac{S_n}{\sigma(S_n)}, Y\right) \le 3.6 \cdot \sqrt[4]{\frac{1}{\sigma^3(S_n)} \sum_{i=1}^{n} E\Big(|X_i|^3\Big)}, \tag{4.2}$$

with $Y$ being standard normal distributed. Obviously, the Berry–Esséen Theorem implies the Central Limit Theorem.

In this section, we show that there is a Berry–Esséen type theorem for Lévy's metric which, compared to the standard theorem, yields better estimates. To achieve this objective, we provide a class of upper estimates for Lévy's metric all referring to some absolute moments of the random variables.

We adopt a method of proving the Central Limit Theorem directly on the set of distribution functions from Huber (cf. Huber (1975) [10, p. 49 – 53]) which evidently originates from Lindeberg (cf. Lindeberg (1922) [13]). Huber estimated the difference between $F_{S_n}$ and the distribution function of the standard normal distribution without using a probability metric. We incorporate this method in Lemma 4.2, Proposition 4.3 and Theorem 4.9 but by adding the use of Lévy's metric we attach importance to the measurement of the speed of convergence to the normal distribution. Furthermore, the class of estimates for Lévy's metric presented here depends on the use of sufficiently often differentiable approximations of indicator functions of the type $x \mapsto 1_{]-\infty,x_0]}(x)$, $x_0 \in \mathsf{R}$, and we show optimal choices of such functions in Proposition 4.4 and Proposition 4.5.

As a preliminary result, we start with a special version of Taylor's Theorem and the Fundamental Theorem of Calculus. For a natural number $m$, denote by $\mathcal{C}_b^m(\mathsf{R})$ the linear space of bounded, real-valued functions on $\mathsf{R}$ having $m$ bounded continuous derivatives. Furthermore, denote by $\mathcal{F}^m(\mathsf{R})$ the linear subspace of $\mathcal{C}_b^{m-1}(\mathsf{R})$ consisting of all functions $f : \mathsf{R} \to \mathsf{R}$ with $f^{(m)}(x)$ existing for all $x \in \mathsf{R}$ except for a finite subset $A_f$ of $\mathsf{R}$ and with $\|f^{(m)}\|_\infty := \|f^{(m)}|_{\mathbb{R}\setminus A_f}\|_\infty < \infty$.

**Lemma 4.1**

*(a)* Fundamental Theorem of Calculus
*Let $f \in \mathcal{F}^1(\mathsf{R})$ and $x_0 \in \mathsf{R}$. Then*

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(t)\, dt$$

*holds for every $x \in \mathsf{R}$.*

*(b)* Taylor's Theorem
*Let $f \in \mathcal{F}^m(\mathsf{R})$ and $x_0 \in \mathsf{R}$. Then*

$$f(x) = \sum_{k=0}^{m-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_m(x_0, x)$$

*holds for all $x \in \mathsf{R}$ with*

$$\|R_m(x_0, x)\|_\infty \leq \frac{\|f^{(m)}\|_\infty}{m!}|x - x_0|^m.$$

**Proof.**

(a) If there isn't any discontinuity point of $f'$ between $x_0$ and $x$ the result holds by the classical Fundamental Theorem of Calculus. Now suppose there is exactly one discontinuity point $a$ of $f'$ between $x_0$ and $x$, w.l.o.g. suppose $x_0 < a < x$. Then, applying the classical Fundamental Theorem of Calculus and continuity of $f$, for sufficiently small $\varepsilon > 0$,

$$\begin{aligned} f(x) &= \big[f(x) - f(a+\varepsilon)\big] + \big[f(a+\varepsilon) - f(a-\varepsilon)\big] + \big[f(a-\varepsilon)f(x_0)\big] \\ &= \int_{a+\varepsilon}^{x} f'(t)\, dt + \big[f(a+\varepsilon) - f(a-\varepsilon)\big] + \int_{x_0}^{a-\varepsilon} f'(t)\, dt \\ &\xrightarrow[\varepsilon\to 0]{} \int_{x_0}^{x} f'(t)\, dt. \end{aligned}$$

If $x_0$ or $x$ is a discontinuity point of $f'$ itself this result remains true by boundedness of $f'$, resp. continuity of $f$. By induction, we obtain the general result.

(b) For a given $f \in \mathcal{F}^m(\mathsf{R})$, denote by $A_f \subset \mathsf{R}$ the set of real numbers $x$ for which $f^{(m)}(x)$ does not exists. Furthermore, denote by $\alpha$ the minimal distance between two elements in $A_f$, $\alpha := \min\{|y-z| \mid y, z \in A_f\}$. Define the sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{C}_b^m(\mathsf{R})$ in the following way. Set

$$f_n^{(m)}(x) := f^{(m)}(x)$$

if $\min_{y\in A_f} |x-y| \geq \frac{\alpha}{2n}$ and

$$f_n^{(m)}(x) := \lambda f^{(m)}(y - \tfrac{\alpha}{2n}) + (1-\lambda) f^{(m)}(y + \tfrac{\alpha}{2n})$$

if $x = \lambda(y - \frac{\alpha}{2n}) + (1-\lambda)(y + \frac{\alpha}{2n})$ for some $y \in A_f$ with $|x-y| < \frac{\alpha}{2n}$ and $\lambda \in [0,1]$. Furthermore, set recursively $f_n^{(k)}(x) := f^{(k)}(0) + \int_0^x f_n^{(k-1)}(t)\,dt$ for each $k = m-1, \ldots, 0$. By construction, $\|f_n^{(m)}\|_\infty \leq \|f^{(m)}\|_\infty$. From (a) follows

$$\|f_n^{(m-1)} - f^{(m-1)}\|_\infty \leq |A_f| \cdot \tfrac{1}{2} \cdot \big(2 \cdot \|f^{(m-1)}\|\big) \cdot \big(2 \tfrac{\alpha}{2n}\big) = 2 \cdot |A_f| \cdot \|f^{(m-1)}\| \tfrac{\alpha}{2n}.$$

Hence, for every $x \in \mathsf{R}$, the term $|f_n^{(k)}(x) - f^{(k)}(x)|$, $k \in \{0, \ldots, m-1\}$, can be estimated only using $2 \cdot |A_f| \cdot \|f^{(m-1)}\| \frac{\alpha}{2n}$, $k$, and $|x|$. Thus, $f_n^{(k)}$ converges pointwise to $f_n^{(k)}$. Applying Taylor's Theorem for every $f_n$ and using pointwise convergence of $f_n^{(k)}$ to $f^{(k)}$, $k = 0, .., m-1$, yields the desired results. □

Now we come to a fundamental lemma in this section.

**Lemma 4.2**

*(a) Let $f \in \mathcal{F}^m(\mathsf{R})$, $m \in \mathsf{N}$ and let $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ be pairwise independent random variables with $E(X_i^k) = E(Y_i^k)$, $i = 1, \ldots, n$, $k = 1, \ldots, m-1$. Then with $S_n := \sum_{i=1}^n X_i$ and $T_n := \sum_{i=1}^n Y_i$*

$$\big|E(f(S_n)) - E(f(T_n))\big| \leq \frac{\|f^{(m)}\|_\infty}{m!} \left( \sum_{i=1}^n E\Big(|X_i|^m\Big) + \sum_{i=1}^n E\Big(|Y_i|^m\Big) \right).$$

*(b) If especially $n = 1$ then (a) also holds if $X_1$ and $Y_1$ are dependent.*

**Proof.** By Lemma 4.1,

$$f(X_1 + \cdots + X_n) = \sum_{k=1}^{m-1} \frac{f^{(k)}(S_{n-1})}{k!} X_n^k + R_m(S_{n-1}, S_n) \tag{4.3}$$

with

$$R_m(S_{n-1}, S_n) \leq \frac{\|f^{(m)}\|_\infty}{m!} |X_n|^m.$$

Integrating both sides of Equation (4.3) yields (using independence in the case $n > 1$)

$$E\big(f(S_n)\big) = \sum_{k=1}^{m-1} \frac{1}{k!} E\big(f^{(k)}(S_{n-1})\big) E\big(X_n^k\big) + E\big(R_m(S_{n-1}, S_n)\big)$$

and with $E(X_i^k) = E(Y_i^k)$ follows

$$\Big|E\big(f(S_{n-1}+X_n)\big) - E\big(f(S_{n-1}+Y_n)\big)\Big| \le \frac{\|f^{(m)}\|_\infty}{m!}\Big(E\big(|X_n|^m\big) + E\big(|Y_n|^m\big)\Big). \tag{4.4}$$

Using

$$\begin{aligned}&\Big|E\big(f(S_n)\big) - E\big(f(T_n)\big)\Big| \\ &\le \sum_{i=0}^{n-1}\left|E\left(f\Big(\sum_{j=1}^{n-i} X_j + \sum_{j=n-i+1}^{n} Y_j\Big)\right) - E\left(f\Big(\sum_{j=1}^{n-i-1} X_j + \sum_{j=n-i}^{n} Y_j\Big)\right)\right|\end{aligned}$$

and the corresponding analogous version of inequality (4.4), we get the desired results of (a) and (b). □

The subsequent proposition provides the announced class of upper estimates for Lévy's metric.

**Proposition 4.3**

*(a) Let $f \in \mathcal{F}^m(\mathbb{R})$, $m \in \mathbb{N}$, with*

$$\begin{aligned} f(x) &= 1 && \text{if } x \le 0\,, && (4.5a)\\ f(x) &\in [0,1] && \text{if } 0 < x < 1\,, && (4.5b)\\ f(x) &= 0 && \text{if } x \ge 1\,. && (4.5c)\end{aligned}$$

*Let $X_1,\dots,X_n$, $Y_1,\dots,Y_n$ be pairwise independent random variables with $E(X_i^k) = E(Y_i^k)$, $i = 1,\dots,n$, $k = 1,\dots,m-1$. Define $S_n := \sum_{i=1}^n X_i$ and $T_n := \sum_{i=1}^n Y_i$. Then*

$$d_L(S_n,T_n) \le \sqrt[m+1]{\frac{\|f^{(m)}\|_\infty}{m!}\left(\sum_{i=1}^n E\Big(|X_i|^m\Big) + \sum_{i=1}^n E\Big(|Y_i|^m\Big)\right)}\,. \tag{4.6}$$

*(b) If especially $n = 1$ then (a) also holds if $X_1$ and $Y_1$ are dependent.*

**Proof.** Define $f_{x_0,h} : \mathbb{R} \to \mathbb{R}$ by $f_{x_0,h}(x) := f(\frac{x-x_0}{h})$. To prove (a), resp. (b), we use Lemma 4.2 (a), resp. (b), and obtain

$$\begin{aligned} F_{S_n}(x_0) &= E\big(1_{]-\infty,x_0]}\circ S_n\big)\\ &\le E\big(f_{x_0,h}\circ S_n\big)\\ &\le E\big(f_{x_0,h}\circ T_n\big) + \frac{\|f^{(m)}\|_\infty}{h^m m!}\left(\sum_{i=1}^n E\Big(|X_i|^m\Big) + \sum_{i=1}^n E\Big(|Y_i|^m\Big)\right)\\ &\le E\big(1_{]-\infty,x_0+h]}\circ T_n\big) + \frac{\|f^{(m)}\|_\infty}{h^m m!}\left(\sum_{i=1}^n E\Big(|X_i|^m\Big) + \sum_{i=1}^n E\Big(|Y_i|^m\Big)\right)\\ &= F_{T_n}(x_0+h) + h^{-m}\cdot\frac{\|f^{(m)}\|_\infty}{m!}\left(\sum_{i=1}^n E\Big(|X_i|^m\Big) + \sum_{i=1}^n E\Big(|Y_i|^m\Big)\right).\end{aligned}$$

With

$$\tilde{h} := \sqrt[m+1]{\frac{\|f^{(m)}\|_\infty}{m!}\left(\sum_{i=1}^{n} E\Big(|X_i|^m\Big) + \sum_{i=1}^{n} E\Big(|Y_i|^m\Big)\right)}$$

follows $F_{S_n}(x_0) \le F_{T_n}(x_0 + \tilde{h}) + \tilde{h}$ and due to symmetry we also obtain $F_{T_n}(x_0) \le F_{S_n}(x_0 + \tilde{h}) + \tilde{h}$. By the definition of $d_L$, we get (a), resp. (b). □

The main feature of Proposition 4.3 is that Lévy's metric can be estimated only using absolute moments of the random variables involved. By choosing $f$ in this proposition in an optimal way, i.e. minimizing $\|f^{(m)}\|_\infty$ for a given $m$, we will be able to provide a rate of convergence to the normal distribution, i.e. providing the Berry–Esséen type theorem for Lévy's metric.

We now put Proposition 4.3 (a) in concrete terms for $m \in \{1,2,3\}$. The occurring regularities of the form of the functions used in the proof gives rise for conjecturing that this result also holds for every natural number $m$ (cf. Conjecture 4.6).

**Proposition 4.4** *Let* $X_1,\dots,X_n$, $Y_1,\dots,Y_n$ *be pairwise independent random variables satisfying* $E(X_i^k) = E(Y_i^k)$ *with* $i = 1,\dots,n$, $k = 1,\dots,m-1$ *and* $m \in \{1,2,3\}$. *Then*

$$d_L(S_n, T_n) \le \sqrt[m+1]{\frac{4^{m-1}}{m}\left(\sum_{i=1}^{n} E\Big(|X_i|^m\Big) + \sum_{i=1}^{n} E\Big(|Y_i|^m\Big)\right)}\,. \tag{4.7}$$

**Proof.** Define $f_1, f_2, f_3 : \mathsf{R} \to [0,1]$ by

$$f_1(x) := 1 - \int_0^x 4^0 \cdot 0! \cdot 1_{[0,1[}(t_1)\,dt_1\,, \tag{4.8a}$$

$$f_2(x) := 1 - \int_0^x \int_0^{t_1} 4^1 \cdot 1! \cdot \Big(1_{[0,\frac{1}{2}[} - 1_{[\frac{1}{2},1[}\Big)(t_2)\,dt_2\,dt_1\,, \tag{4.8b}$$

$$f_3(x) := 1 - \int_0^x \int_0^{t_1} \int_0^{t_2} 4^2 \cdot 2! \cdot \Big(1_{[0,\frac{1}{4}[\cup[\frac{3}{4},1[} - 1_{[\frac{1}{4},\frac{3}{4}[}\Big)(t_3)\,dt_3\,dt_2\,dt_1\,. \tag{4.8c}$$

By Lemma 4.1 (b), $f_i \in \mathcal{F}^i(\mathsf{R})$ and $\|f_i^{(i)}\|_\infty = 4^{i-1} \cdot (i-1)!$. Applying Proposition 4.3 (a) finishes the proof. □

The subsequent proposition answers the question of optimality of the result given in the preceding proposition.

**Proposition 4.5** *The functions* $f_i$, $i = 1,2,3$, *defined in* (4.8) *satisfy* $\|f_i^{(i)}\|_\infty = \inf\{\|f^{(i)}\|_\infty \mid f \in \mathcal{F}^i(\mathsf{R})$ *with* $(4.5)\}$. *Therefore, Estimate* (4.7) *is the best possible concretion of Estimate* (4.6).

**Proof.** Suppose $g_1 \in \mathcal{F}^1(\mathsf{R})$ satisfies the conditions (4.5). Then

$$0 = g_1(1) = g_1(0) + \int_0^1 g_1'(t_1)\,dt_1 \ge g_1(0) + \int_0^1 -\|g_1'\|_\infty\,dt_1 = 1 - \|g_1'\|_\infty,$$

i.e. $\|g_1'\|_\infty \geq 1$, hence $f_1$ is an optimal function.
Now suppose $g_2 \in \mathcal{F}^2(\mathsf{R})$ satisfies the conditions (4.5). W.l.o.g., we can assume that $g_2$ satisfies the symmetric property $g_2(t) = 1 - g_2(1-t)$ since otherwise we use the function $\overline{g_2} \in \mathcal{F}^2(\mathsf{R})$, defined by

$$\overline{g_2}(t) := \tfrac{1}{2} \cdot g_2(t) + \tfrac{1}{2} \cdot (1 - g_2(1-t)).$$

It satisfies the conditions (4.5), $\overline{g_2}(t) = 1 - \overline{g_2}(1-t)$ and $\|\overline{g_2}''\|_\infty \leq \|g_2''\|_\infty$. Then, using symmetry of $g_2$, $g_2(0) = 1$ and $g_2'(0) = 0$,

$$\begin{aligned}
\tfrac{1}{2} = g_2(\tfrac{1}{2}) = g_2(0) &+ \int_0^{\frac{1}{2}} \left( g_2'(0) + \int_0^{t_1} g_2''(t_2)\, dt_2 \right) dt_1 \\
&\geq 1 + \int_0^{\frac{1}{2}} \int_0^{t_1} -\|g_2''\|_\infty \, dt_2 \, dt_1 \\
&= 1 + \int_0^{\frac{1}{2}} -\|g_2''\|_\infty \cdot t_1 \, dt_1 \\
&= 1 - \tfrac{1}{8} \cdot \|g_2''\|_\infty,
\end{aligned}$$

i.e. $\|g_2''\|_\infty \geq 4$, hence $f_2$ is an optimal function.
Finally, suppose $g_3 \in \mathcal{F}^3(\mathsf{R})$ satisfies the conditions (4.5). From the symmetric property of $g_3$, $g_3(t) = 1 - g_3(1-t)$ follows $g_3'(t) = g_3'(1-t)$ and $g_3''(t) = -g_3''(1-t)$, hence $g_3''(\frac{1}{2}) = 0$. Furthermore, $g_3''(t) \geq -\|g_3'''\|_\infty \cdot t$, $t \in [0, \frac{1}{4}]$, and $g_3''(t) \geq \|g_3'''\|_\infty \cdot (t - \frac{1}{2})$, $t \in [\frac{1}{4}, \frac{1}{2}]$. Therefore,

$$\begin{aligned}
\tfrac{1}{2} = g_3(\tfrac{1}{2}) = g_3(0) &+ \int_0^{\frac{1}{2}} \left( g_3'(0) + \int_0^{t_1} g_3''(t_2)\, dt_2 \right) dt_1 \\
&= 1 + \int_0^{\frac{1}{4}} \int_0^{t_1} g_3''(t_2)\, dt_2 \, dt_1 + \int_{\frac{1}{4}}^{\frac{1}{2}} \int_0^{t_1} g_3''(t_2)\, dt_2 \, dt_1 \\
&\geq 1 + \int_0^{\frac{1}{4}} \int_0^{t_1} -\|g_3'''\|_\infty \cdot t_2 \, dt_2 \, dt_1 \\
&\quad + \int_{\frac{1}{4}}^{\frac{1}{2}} \left( \int_0^{\frac{1}{4}} -\|g_3'''\|_\infty \cdot t_2 \, dt_2 + \int_{\frac{1}{4}}^{t_1} -\|g_3'''\|_\infty \cdot (t_2 - \tfrac{1}{2})\, dt_2 \right) dt_1 \\
&= 1 - \tfrac{1}{64} \cdot \|g_3'''\|_\infty,
\end{aligned}$$

i.e. $\|g_3'''\|_\infty \geq 32$, hence $f_3$ is an optimal function. □

The Estimate (4.7) has also been proved by the author to be valid for some more natural numbers $m$. Optimality of Estimate (4.7) has also been proved for $m = 4$. This gives rise to formulate the subsequent conjecture.

**Conjecture 4.6** *Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be pairwise independent random variables satisfying $E(X_i^k) = E(Y_i^k)$, $i = 1, \ldots, n$, $k = 1, \ldots, m-1$ and $m \in \mathsf{N}$. Then*

$$d_L(S_n, T_n) \leq \sqrt[m+1]{\frac{4^{m-1}}{m} \left( \sum_{i=1}^n E\Big(|X_i|^m\Big) + \sum_{i=1}^n E\Big(|Y_i|^m\Big) \right)}\,.$$

Before turning to our main theorem, we give two rather simple applications of Proposition 4.4.

For a random variable $X$, denote by $M(X)$ a median of $X$, i.e.

$$M(X) \in \left[\sup\left\{x \in \mathsf{R} \mid F_X(x) \leq \tfrac{1}{2}\right\}, \inf\left\{x \in \mathsf{R} \mid F_X(x) \geq \tfrac{1}{2}\right\}\right],$$

and by $\tau(X) := E(|X - M(X)|)$ the average absolute deviation from the median. In insurance mathematics, a multiple $\alpha\tau(X)$ of $\tau(X)$, $\alpha > 0$, has been suggested as a risk loading since the premium principle $E(X) + \alpha\tau(X)$ can be represented as a (non-additive) Choquet integral (cf. Denneberg (1994) [3, Exercise 5.4]) having favorable properties for applications. In situations like this one, where the volatility parameter $\tau(X)$ is used, the following corollary of Proposition 4.4 may be of interest.

**Corollary 4.7** *Let $X, Y$ be two random variables with $M(X) = M(Y)$. Then*

$$d_L(X, Y) \leq \sqrt{\tau(X) + \tau(Y)}.$$

**Proof.** Since, by Proposition 2.4 (a), $d_L(X, Y) = d_L(X - MX, Y - MX) = d_L(X - MX, Y - MY)$, the statement directly follows from Proposition 4.4 for $m = 1$. □

For $m = 2$, Proposition 4.4 gets the subsequent form.

**Corollary 4.8** *Let $X, Y$ be two random variables with $E(X) = E(Y)$. Then*

$$d_L(X, Y) \leq \sqrt[3]{2\Big(V(X) + V(Y)\Big)}. \tag{4.9}$$

*Additionally, if $X, Y$ are independent then*

$$d_L(X, Y) \leq \sqrt[3]{V(X) + V(Y)}.$$

The last assertion directly follows from Proposition 3.2 using Proposition 3.1 (b).

In his 1967 paper, Zolotarev proved a weaker upper estimate of Lévy's metric than given in (4.9) (cf. Zolotarev (1967) [14, Lemma 2]),

$$d_L(X, Y) \leq \sqrt[3]{4 \max\Big\{V(X), V(Y)\Big\}}.$$

As the main application of Proposition 4.4 we now state a Berry–Esséen type estimate of the rate of convergence to the normal distribution in terms of Lévy's metric.

**Theorem 4.9** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables with $E(X_n) = 0$, $V(X_n) = \sigma_n^2 > 0$ and $E(|X_n|^3)$ finite. Furthermore, let $Y$ be a standard normal distributed random variable which is independent of $X_n$ for all $n \in \mathsf{N}$. Then*

$$d_L\left(\frac{S_n}{\sigma(S_n)}, Y\right) \leq 1.93 \cdot \sqrt[4]{\frac{1}{\sigma^3(S_n)} \sum_{i=1}^{n} E\left(|X_i|^3\right)}.$$

---

This proof contains significant gaps or some errors since the use of Chebychev's Inequality seems to be applied by mistake for negative values.

**Proof.** Let $(Y_n)$ be a sequence of independent standard normal distributed random variables. Successively using that $\sum_{i=1}^{n} \frac{\sigma(X_i)}{\sigma(S_n)} Y_i$ is standard normal distributed, Proposition 4.4 with $m = 3$, $\sigma^3(X_i) \leq E(|X_i|^3)$ (Jensen's Inequality) and $E(|Y_i|^3) = \sqrt{\frac{8}{\pi}}$ for all $i \leq n$, we obtain

$$\begin{aligned} d_L\left(\frac{S_n}{\sigma(S_n)}, Y\right) &= d_L\left(\frac{S_n}{\sigma(S_n)}, \sum_{i=1}^{n} \frac{\sigma(X_i)}{\sigma(S_n)} Y_i\right) \\ &\leq \sqrt[4]{\frac{16}{3\sigma^3(S_n)} \left(\sum_{i=1}^{n} E\left(|X_i|^3\right) + \sum_{i=1}^{n} \sigma^3(X_i) E\left(|Y_i|^3\right)\right)} \\ &\leq \sqrt[4]{\frac{16}{3\sigma^3(S_n)} \left(\sum_{i=1}^{n} E\left(|X_i|^3\right)\left(1 + E\left(|Y_i|^3\right)\right)\right)} \\ &\leq \sqrt[4]{\frac{16}{3\sigma^3(S_n)} \left(1 + \sqrt{\frac{8}{\pi}}\right)\left(\sum_{i=1}^{n} E\left(|X_i|^3\right)\right)} \\ &\leq 1.93 \cdot \sqrt[4]{\frac{1}{\sigma^3(S_n)} \sum_{i=1}^{n} E\left(|X_i|^3\right)}. \end{aligned}$$

□

A natural question arising now is how to compare the standard Berry–Esséen estimate of the rate of convergence w.r.t. Kolmogorov's metric to the one obtained above. Using $\|F'_Y\|_\infty = (\sqrt{2\pi})^{-1}$ for a standard normal distributed random variable $Y$, Proposition 3.1 (c), (d) and Theorem 4.9 together yield that an estimate in terms of Kolmogorov's metric,

$$d_K\left(\frac{S_n}{\sigma(S_n)}, Y\right) \leq C \cdot \sqrt[4]{\frac{1}{\sigma^3(S_n)} \sum_{i=1}^{n} E\left(|X_i|^3\right)},$$

is better than the one in Theorem 4.9 if $C < 1.93$, incomparable if $C \in [1.93, 2.70]$, and worse if $C > 2.70$. Since $C = 3.6$ in Berry's original estimate (cf. Inequality (4.2), our estimate is an improvement. Breiman has mentioned, that there exist unpublished calculations giving bounds as low as $C = 2.05$ (cf. Breiman (1992) [2, p. 184]). This bound is incomparable to our result, but this also means that it is not better than ours.

# 5 Conclusions

It remains as an open problem to prove Conjecture 4.6. Although all relevant cases of this conjecture, i.e. those cases actually used in this article, have been proved in Proposition 4.4, it would be a nice result. Another task remaining to do is to provide a rate of convergence for sequences of independent distributed random variables (having certain additional properties) converging in distribution to a Poisson distributed random variable. Such a result cannot be expressed in terms of Kolmogorov's metric since the limit distribution is not continuous and would therefore expose the advantages of Lévy's metric.

## References


[1] Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. **49**, 122-136.

[2] Breiman, L. (1968, 1992 republished). Probability. SIAM, Philadelphia.

[3] Denneberg, D. (1994, 1997 2nd ed.). Non-additive measure and integral. Kluwer, Dordrecht.

[4] Dudley, R. M. (1989). Real analysis and probability. Chapman & Hall/CRC, Boca Raton.

[5] Esséen, C. G. (1945). Fourier analysis of distribution functions. Acta Math. **77**, 1-125.

[6] Fan, K. (1944). Entfernung zweier zufälliger Größen und die Konvergenz nach Wahrscheinlichkeit. Math. Z. **49**, 681-683.

[7] Fréchet, M. (1937, 1950 2nd ed.). Recherches théoriques modernes sur la calcul des probabilités. Gauthier-Villars, Paris.

[8] Galambos, J. (1988). Advanced probability theory. Dekker, New York.

[9] Gnedenko, B. V. and Kolmogorov, A. N. (1954). Limit distributions for sums of independent random variables. Addison-Wesley, Cambridge.

[10] Huber, P. J. (1975). Einführung in die Wahrscheinlichkeit und Statistik. Verlag der Fachvereine an der ETH Zürich, Zürich.

[11] Lévy, P. (1925). Calcul des probabilités. Gauthier-Villars, Paris.

[12] Lévy, P. (1937, 1954 2nd ed.). Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris.

[13] Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Math. Zeitschr. **15**, 211-225.

[14] Zolotarev, V. M. (1967). A generalization of the Lindeberg-Feller theorem. Theor. Prob. Appl. **12**, 608-618.

[15] Zolotarev, V. M. (1997). Modern theory of summation of random variables. VSP, Utrecht.