A FRIENDLY PROOF OF THE BERRY-ESSEEN THEOREM

ROMAN VERSHYNIN

Abstract. A gem of classical probability, the Berry-Esseen theorem provides a non-asymptotic form of the central limit theorem. This note gives a friendly and intuitive exposition of the classical Fourier-analytic proof of Esseen's smoothing inequality and, as a consequence, a general Berry-Esseen theorem for non-i.i.d random variables. The exposition is suitable for use in a basic graduate course in probability.

The Berry–Esseen theorem is a beautiful result in classical probability that gives a non-asymptotic form of the central limit theorem.

Theorem 0.1 (Berry–Esseen theorem [1, 2])**.** *Let $X_1, \ldots, X_n$ be independent mean-zero random variables whose sum $S_n \coloneqq X_1 + \cdots + X_n$ satisfies $\mathrm{Var}(S_n) = 1$. Let G be a standard normal random variable. Then*

$$\sup_{a \in \mathbb{R}} \Big| \mathbb{P}\{S_n \le a\} - \mathbb{P}\{G \le a\} \Big| \le C \sum_{k=1}^{n} \mathbb{E}|X_k|^3,$$

provided the right-hand side is finite, where C is an absolute constant.

To see how this implies the classical central limit theorem, let $Y_1, Y_2, \ldots$ be independent mean-zero random variables with unit variances, and with uniformly bounded third moments: $\sup_k \mathbb{E}|Y_k|^3 = O(1)$. Applying Theorem 0.1 for $X_k = Y_k/\sqrt{n}$, we obtain

$$\sup_{a \in \mathbb{R}} \Big| \mathbb{P}\Big\{ \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \le a \Big\} - \mathbb{P}\{G \le a\} \Big| = O\Big(\frac{1}{\sqrt{n}}\Big). \tag{0.1}$$

Thus $(Y_1 + \cdots + Y_n)/\sqrt{n}$ converges in distribution to G, recovering the classical central limit theorem (under a third moment assumption, which can be weakened). Importantly, (0.1) establishes *uniform* convergence of the distribution functions, provides an explicit and optimal *rate* of convergence $O(n^{-1/2})$, and holds *non-asymptotically*, that is, for any finite n.

In this note, we present a gentle and intuitive exposition of the classical Fourier-analytic proof of the Berry-Esseen theorem. No prior background in Fourier analysis is assumed; we recall all we need in Section 2.

There are several existing textbook-level expositions of proofs of the Berry-Esseen theorem, including the classical Fourier-analytic approach (see, for example, [4]) as well as proofs based on Stein's method (see, for example, [5]).

1. The proof strategy

Here is a bird's-eye, non-rigorous sketch of the proof of Theorem 0.1.

Partially supported by the NSF Grant DMS 2451011 and the U.S. Air Force Grant FA9550-25-1-0294.

1.1. **Smoothing.** We begin by writing

$$\mathbb{P}\{S_n \le a\} = \mathbb{E}\,\mathbf{1}_{\{S_n \le a\}}. \tag{1.1}$$

The difficulty is that the function $x \mapsto \mathbf{1}_{\{x \le a\}}$ is too rough: it has a discontinuous drop from 1 to 0 at the point a. To mitigate this issue, we smooth the indicator function: approximate it with a function $f(x)$ that decreases from 1 to 0 gradually. With such approximation, (1.1) becomes

$$\mathbb{P}\{S_n \le a\} \approx \mathbb{E}\, f(S_n).$$

A rigorous version of this smoothing step will be given in Lemma 3.1.

1.2. **Fourier transform.** Next, we replace $f(x)$ by an even simpler function – a complex exponential $e^{2\pi i t x}$. This can be done using the Fourier transform inversion formula (recalled in (2.2) below):

$$\mathbb{E}\, f(S_n) = \mathbb{E} \int_{\mathbb{R}} e^{2\pi i t S_n}\, \widehat{f}(t)\, dt = \int_{\mathbb{R}} \mathbb{E}\left[e^{2\pi i t S_n}\right]\, \widehat{f}(t)\, dt.$$

Applying the same method to a standard normal random variable G, subtracting the two expressions, and using Jensen's inequality, we obtain

$$\left|\mathbb{P}\{S_n \le a\} - \mathbb{P}\{G \le a\}\right| \lesssim \int_{\mathbb{R}} \left|\mathbb{E}\, e^{2\pi i t S_n} - \mathbb{E}\, e^{2\pi i t G}\right| |\widehat{f}(t)|\, dt. \tag{1.2}$$

This reduces the problem to comparing the *characteristic functions* $\mathbb{E}\, e^{itS_n}$ and $\mathbb{E}\, e^{itG}$. For the standard normal random variable G, a simple computation yields

$$\mathbb{E}\, e^{itG} = \exp(-t^2/2), \quad t \in \mathbb{R}. \tag{1.3}$$

1.3. **Taylor approximation.** For S_n, independence of the random variables X_k gives

$$\mathbb{E}\, e^{itS_n} = \mathbb{E}\left[e^{itX_1}\right] \cdots \mathbb{E}\left[e^{itX_n}\right],$$

which reduces the problem to computing the characteristic function $\mathbb{E}\, e^{itX_k}$ for each random variable X_k separately. By assumption,

$$\mathbb{E}\, X_k = 0 \quad \text{and} \quad \mathbb{E}\, X_k^2 =: \sigma_k^2 < \infty.$$

Using the Taylor approximation $e^z \approx 1 + z + z^2/2$ and taking expectations, we get

$$\begin{aligned} \mathbb{E}\, e^{itX_k} &\approx \mathbb{E}\big(1 + itX_k - t^2 X_k^2/2\big) \\ &= 1 - \sigma_k^2 t^2/2 \approx \exp\big(-\sigma_k^2 t^2/2\big). \end{aligned}$$

Multiplying over k gives

$$\mathbb{E}\, e^{itS_n} \approx \exp\big(-\sigma_1^2 t^2/2\big) \cdots \exp\big(-\sigma_n^2 t^2/2\big) = \exp\big(-t^2/2\big), \tag{1.4}$$

since $\sigma_1^2 + \cdots + \sigma_n^2 = 1$ by assumption. Comparing to (1.3), we see that

$$\mathbb{E}\, e^{itS_n} \approx \mathbb{E}\, e^{itG}.$$

Substituting this approximation into (1.2) completes the heuristic "proof".

1.4. **A challenge and a fix.** The main issue with this heuristic argument lies in the quality of the Taylor approximation. We cannot expect

$$\mathbb{E}\, e^{itX_k} \approx \exp(-\sigma_k^2 t^2/2) \tag{1.5}$$

uniformly over all $t \in \mathbb{R}$. For example, X_k is a Rademacher distribution, the characteristic function equals $\cos(t)$, which does not even decay to zero as $|t| \to \infty$.

There is an elegant fix for this issue: choose the smoothing function f whose Fourier transform $\widehat{f}$ is supported on a *compact interval* centered at the origin. The integrand in (1.2) would vanish outside that interval, and so it suffices to establish the approximation (1.5) only for t close to the origin. This idea leads to *Esseen's smoothing inequality* (Theorem 3.3).

In Lemma 4.1, we will prove (1.5) in a neighborhood of the origin. Multiplying over k, we will deduce a rigorous form of (1.4) in Lemma 4.3. Plugging it into the smoothing inequality will complete the proof of Theorem 0.1.

Now let's do this step by step.

2. Background on Fourier transform

First, let's recall some basic facts about Fourier transform that we will use in the proof.

The *Fourier transform* of an integrable function $f : \mathbb{R} \to \mathbb{C}$ is the bounded and continuous function $\widehat{f} : \mathbb{R} \to \mathbb{C}$ defined as

$$\widehat{f}(t) = \int_{\mathbb{R}} e^{-2\pi itx} f(x)\, dx. \tag{2.1}$$

If f is a *Schwartz function* (that is, f and all its derivatives decay faster than any polynomial at infinity), then $\widehat{f}$ is also a Schwartz function.

Any Schwartz function f can be reconstructed from its Fourier transform using the *Fourier inversion formula*:

$$f(x) = \int_{\mathbb{R}} e^{2\pi itx} \widehat{f}(x)\, dx. \tag{2.2}$$

Comparing with (2.1), we see that applying the Fourier transform twice yields $f(-x)$. In particular, the Fourier transform is an involution on even functions.

The Fourier transform preserves the L^2 norm: for a Schwartz function f, the *Plancherel identity* says that

$$\int_{\mathbb{R}} |f(x)|^2\, dx = \int_{\mathbb{R}} |\widehat{f}(t)|^2\, dt. \tag{2.3}$$

The *convolution* of two integrable functions f and g is the integrable function $f * g$ defined as

$$(f * g)(x) = \int_{\mathbb{R}} f(y) g(x-y)\, dy.$$

The *convolution theorem* states that

$$\widehat{f * g} = \widehat{f} \cdot \widehat{g} \quad \text{pointwise.} \tag{2.4}$$

3. Smoothing

We can write the cumulative distribution function of a random variable X as

$$\mathbb{P}\{X \le a\} = \mathbb{E}\, \mathbf{1}_{(-\infty,0]}(X - a),$$

where $\mathbf{1}_{(-\infty,0]}(x)$ is the indicator of $(-\infty, 0]$. We now smooth this indicator, replacing it by a function that transitions gradually from 1 to 0, with most of the transition occurring in an ε-neighborhood of the origin.

Lemma 3.1 (Smoothing the discrepancy)**.** *Let φ be a probability density function, which is also a Schwartz function. Consider a smoothed version of the indicator function of $(-\infty, 0]$:*

$$f = \mathbf{1}_{(-\infty,0]} * \varphi.$$

Let X and Y be random variables. Assume that the probability density function of Y is bounded by M. Then we have for any $\varepsilon > 0$:

$$\sup_{a\in\mathbb{R}}\Big|\mathbb{P}\{X \le a\} - \mathbb{P}\{Y \le a\}\Big| \le 2\sup_{a\in\mathbb{R}}\Big|\mathbb{E}\, f\Big(\frac{X-a}{\varepsilon}\Big) - \mathbb{E}\, f\Big(\frac{Y-a}{\varepsilon}\Big)\Big| + C_\varphi M\varepsilon. \tag{3.1}$$

Here C_φ depends only on the choice of the smoothing function φ.

Proof. Step 1. Regularity of the discrepancy function. By rescaling, it is enough to consider $\varepsilon = 1$. Our task is to bound the discrepancy function

$$\Delta(a) \coloneqq F_X(a) - F_Y(a),$$

where $F_X(a) \coloneqq \mathbb{P}\{X \le a\}$ and $F_Y(a) \coloneqq \mathbb{P}\{Y \le a\}$ are the cumulative distribution functions.

The function F_X increases, while F_Y cannot increase too fast because its derivative is bounded by M. Putting this together, we see that Δ cannot decrease too fast:

$$\Delta(\bar{a} + t) \ge \Delta(\bar{a}) - Mt \quad \text{for any } \bar{a} \in \mathbb{R},\ t \ge 0. \tag{3.2}$$

In particular, once Δ achieves its maximum, Δ will have to remain close to its maximum for a while. To make this precise, assume without loss of generality that

$$\bar{\Delta} \coloneqq \sup_{a\in\mathbb{R}}|\Delta(a)| = \sup_{a\in\mathbb{R}}\Delta(a). \tag{3.3}$$

(Otherwise replace X, Y by $-X, -Y$.) Choose a point $\bar{a}$ where $\Delta(\bar{a}) \ge 0.9\bar{\Delta}$. Then (3.2) gives

$$\Delta(\bar{a} + t) \ge 0.9\bar{\Delta} - Mt \ge \bar{\Delta}/2$$

as long as $0 \le t \le \bar{\Delta}/3M$. In other words, we found an interval of length $\bar{\Delta}/3M$ on which Δ is bounded below by $\bar{\Delta}/2$.

Step 2. The smoothed discrepancy function. Our task is to compare the discrepancy function $\Delta(a)$ to its smoothed version

$$\Delta_f(a) \coloneqq \mathbb{E}\, f(X-a) - \mathbb{E}\, f(Y-a).$$

To express Δ_f in terms of Δ, note that

$$\mathbb{E}\, f(X-a) = \mathbb{E}\int_{\mathbb{R}} \mathbf{1}_{(-\infty,0]}(X-a-y)\varphi(y)\,dy = \int_{\mathbb{R}} F_X(a+y)\varphi(y)\,dy$$

by the Fubini theorem. Express $\mathbb{E}\, f(Y-a)$ similarly, subtract, and obtain

$$\Delta_f(a) = \int_{-\infty}^{\infty} \Delta(a+y)\varphi(y)\,dy. \tag{3.4}$$

Step 3. Integrating. In Step 1, we found an interval $[a_0 - T, a_0 + T]$ with $T = \bar{\Delta}/6M$ and on which Δ is bounded below by $\bar{\Delta}/2$. Let's decompose (3.4):

$$\Delta_f(b) = \underbrace{\int_{|y|\le T} \Delta(a_0+y)\varphi(y)\,dy}_{I_1} + \underbrace{\int_{|y|>T} \Delta(a_0+y)\varphi(y)\,dy}_{I_2}.$$

To bound I_2, recall that (3.3) implies that Δ is bounded below by $-\bar{\Delta}$ everywhere. Moreover, φ is a Schwartz function, so $\int_{|y|>T} \varphi(y)\, dy \le C_\varphi / T$. Thus

$$I_2 \ge -\bar{\Delta} \cdot \frac{C_\varphi}{T}.$$

To bound I_1, recall that Δ is bounded below by $\bar{\Delta}/2$ in the range of integration. Moreover, φ is a probability density function, so its total integral equals 1. Thus

$$I_1 \ge \frac{\bar{\Delta}}{2} \cdot \Big(1 - \frac{C_\varphi}{T}\Big).$$

Adding the two bounds, we conclude that

$$\Delta_f(a_0) \ge \frac{\bar{\Delta}}{2} - \frac{3C_\varphi \bar{\Delta}}{2T} \ge \frac{\bar{\Delta}}{2} - 9C_\varphi M.$$

Rearranging the terms yields $\bar{\Delta} \le 2\Delta_f(a_0) + 18C_\varphi M$, which completes the proof. □

In Lemma 3.1, we replaced the indicator function $\mathbf{1}_{(-\infty,0]}$ by a smooth function f. Now, we further replace f with a very particular choice: the complex exponential.

Lemma 3.2 (Smoothed discrepancy via characteristic functions)**.** *There exists a probability density function φ, which is also a Schwartz function, with the following property. Consider a smoothed version of the indicator function of $(-\infty, 0]$:*

$$f = \mathbf{1}_{(-\infty,0]} * \varphi.$$

Let X and Y be random variables. Then we have for any $\varepsilon > 0$:

$$\sup_{a\in\mathbb{R}} \Big| \mathbb{E}\, f\Big(\frac{X-a}{\varepsilon}\Big) - \mathbb{E}\, f\Big(\frac{Y-a}{\varepsilon}\Big) \Big| \le \int_{-1/\varepsilon}^{1/\varepsilon} \Big| \frac{\mathbb{E}\, e^{itX} - \mathbb{E}\, e^{itY}}{t} \Big|\, dt.$$

Proof. Step 1. Choosing a smoothing function. By translation and dilation, we can assume that $a = 0$ and $\varepsilon = 1$, so it suffices to prove the following version of the conclusion:

$$|\mathbb{E}\, f(X) - \mathbb{E}\, f(Y)| \le \int_{-1}^{1} \Big| \frac{\mathbb{E}\, e^{itX} - \mathbb{E}\, e^{itY}}{t} \Big|\, dt.$$

Choose any probability density function φ, which is also a Schwartz function, and whose Fourier transform is supported in the interval $[-1/2\pi, 1/2\pi]$.

(Why does such φ exist? Take any even Schwartz function ψ supported in $[-1/4\pi, 1/4\pi]$ and satisfying $\int_{\mathbb{R}} \psi(x)^2\, dx = 1$, and set

$$\varphi(t) := \widehat{\psi}(t)^2.$$

Since ψ is even and Schwartz, $\widehat{\psi}$ is real-valued and Schwartz, so φ is nonnegative and Schwartz. Moreover, Plancherel identity (2.3) gives $\int_{\mathbb{R}} \widehat{\psi}(t)^2\, dt = \int_{\mathbb{R}} \psi(x)^2\, dx = 1$, showing that φ is a probability density function. Finally, we have $\widehat{\varphi} = \psi * \psi$: to see this, first apply the convolution theorem (2.4) and then the Fourier inversion formula (2.2), noting that ψ is even. Thus, $\widehat{\varphi}$ is supported on $[-1/2\pi, 1/2\pi]$.)

Step 2. Bounding the Fourier transform. Since the function f approaches 1 at $-\infty$, it is not integrable. To fix this, let us consider the integrable function

$$f_M = \mathbf{1}_{(-M,0]} * \varphi.$$

Since the indicators $\mathbf{1}_{(-M,0]}$ increase to $\mathbf{1}_{(-\infty,0]}$ pointwise as $M \to \infty$, the monotone convergence theorem implies that the functions f_M increase to f pointwise. Another application of the monotone convergence theorem yields

$$\mathbb{E}\, f_M(X) \to \mathbb{E}\, f(X), \quad \mathbb{E}\, f_M(Y) \to \mathbb{E}\, f(Y) \quad \text{as } M \to \infty. \tag{3.5}$$

We claim that

$$\widehat{f_M} \text{ is supported on } \Big[-\frac{1}{2\pi}, \frac{1}{2\pi}\Big] \quad \text{and} \quad \big|\widehat{f_M}(t)\big| \le \frac{1}{\pi|t|}, \quad t \in \mathbb{R}. \tag{3.6}$$

Indeed, the convolution theorem (2.4) shows that $\widehat{f_M} = \widehat{\mathbf{1}_{(-M,0]}} \cdot \widehat{\varphi}$. Now,

$$\widehat{\mathbf{1}_{(-M,0]}}(t) = \int_{-M}^{0} e^{-2\pi i t x}\, dx = \frac{e^{2\pi i t M} - 1}{2\pi i t}, \quad \text{so} \quad \Big|\widehat{\mathbf{1}_{(-M,0]}}(t)\Big| \le \frac{1}{\pi|t|}.$$

Moreover, by construction, $|\widehat{\varphi}(t)|$ vanishes outside $[-1/2\pi, 1/2\pi]$ and is bounded by $\int_{\mathbb{R}} |\varphi(x)|\, dx = 1$ since φ is a probability density function. Combining these bounds proves the claim.

Step 3. Integrating. Using the inverse Fourier transform formula, we can write

$$\mathbb{E}\, f_M(X) = \mathbb{E} \int_{\mathbb{R}} e^{2\pi i t X} \widehat{f_M}(t)\, dt = \int_{\mathbb{R}} \mathbb{E}\left[e^{2\pi i t X}\right] \widehat{f_M}(t)\, dt$$

by Fubini theorem (note that f_M is Schwartz by construction, so $\widehat{f_M}$ is Schwartz, too). Write the same for $\mathbb{E}\, f_M(Y)$, subtract and use Jensen's inequality to get

$$\begin{aligned} |\mathbb{E}\, f_M(X) - \mathbb{E}\, f_M(Y)| &\le \int_{\mathbb{R}} \big|\mathbb{E}\, e^{2\pi i t X} - \mathbb{E}\, e^{2\pi i t Y}\big| \big|\widehat{f_M}(t)\big|\, dt \\ &\le \int_{-1/2\pi}^{1/2\pi} \big|\mathbb{E}\, e^{2\pi i t X} - \mathbb{E}\, e^{2\pi i t Y}\big| \cdot \frac{1}{\pi|t|}\, dt \quad \text{(using (3.6))}. \end{aligned}$$

Make the change of variable $s = 2\pi t$, take limit as $M \to \infty$ using (3.5), and the proof is complete. □

Combining Lemmas 3.1 and 3.2, we immediately obtain

Theorem 3.3 (Esseen's smoothing inequality [3])**.** *Let X and Y be random variables. Assume that the probability density function of Y is bounded by M. Then we have for any $\varepsilon > 0$:*

$$\sup_{a \in \mathbb{R}} \Big| \mathbb{P}\{X \le a\} - \mathbb{P}\{Y \le a\} \Big| \le 2 \int_{-1/\varepsilon}^{1/\varepsilon} \left| \frac{\mathbb{E}\, e^{itX} - \mathbb{E}\, e^{itY}}{t} \right| dt + CM\varepsilon.$$

Here C is an absolute constant.

4. Characteristic functions

Esseen's smoothing inequality (Theorem 3.3) reduces the problem of approximating a cumulative distribution function $\mathbb{P}\{X \le a\}$ to approximating the characteristic function $\mathbb{E}\, e^{itX}$. So, what can we say about the characteristic function?

Lemma 4.1 (The characteristic function of a random variable)**.** *Let X be a random variable with*

$$\mathbb{E}\, X = 0, \quad \mathbb{E}\, X^2 = \sigma^2, \quad \mathbb{E}|X|^3 = \rho^3 < \infty. \tag{4.1}$$

Then

$$\mathbb{E}\, e^{itX} = \exp\Big(-\frac{\sigma^2 t^2}{2} + O(\rho^3 t^3)\Big) \quad \textit{whenever} \quad |t| \le \frac{1}{\rho} \tag{4.2}$$

and

$$\left|\mathbb{E}\, e^{itX}\right| \le \exp\Big(-\frac{\sigma^2 t^2}{2} + O(\rho^3 t^3)\Big) \quad \textit{for any} \quad t \in \mathbb{R}. \tag{4.3}$$

In the statement and proof of this lemma, we use the $O(\cdot)$ notation to hide factors that are bounded by absolute constants. Precisely, $O(a)$ stands for θa where θ is some quantity that satisfies $|\theta| \le C$, where C is an absolute constant. The quantity θ and the constant C may change from line to line.

Proof. Step 1. Approximating the exponential function. To prove (4.2), write a Taylor approximation of the exponential function:

$$e^{ix} = 1 + ix - \frac{x^2}{2} + \theta_0 x^3 \quad \text{for some } \theta_0 = \theta_0(x) \text{ satisfying } |\theta_0| \le \frac{1}{6}.$$

(This holds since the third derivative of e^{ix} is bounded by 1 in modulus.) Substitute $x = tX$ and take expectation to get

$$\mathbb{E}\, e^{itX} = 1 + it\,\mathbb{E}\, X - \frac{t^2\,\mathbb{E}\, X^2}{2} + t^3\,\mathbb{E}[\theta_0 X^3]$$

for some random variable θ_0 satisfying $|\theta_0| \le \frac{1}{6}$ pointwise. Using the assumptions (4.1), we get

$$\mathbb{E}\, e^{itX} = 1 - \frac{\sigma^2 t^2}{2} + \theta \rho^3 t^3 \quad \text{for some } \theta = \theta(x) \text{ satisfying } |\theta| \le \frac{1}{6}. \tag{4.4}$$

For convenience, let us rewrite this as

$$\mathbb{E}\, e^{itX} = 1 - \frac{a}{2} + \theta b, \quad \text{where} \quad a = \sigma^2 t^2 \text{ and } b = \rho^3 t^3, \tag{4.5}$$

and note that

$$a^2 \le |b| \le 1. \tag{4.6}$$

(The second inequality follows from the assumption $|t| \le \frac{1}{\rho}$. To check the first inequality, note that $\sigma \le \rho$ by Jensen's inequality, so $a^2 = \sigma^4 t^4 \le \rho^4 t^4 = |b|^{4/3} \le |b|$.)

Step 2. Linearizing the logarithmic function. Now write a Taylor approximation of the logarithmic function:

$$\ln(1+x) = x + O(x^2) \quad \text{whenever} \quad |x| \le \frac{2}{3}.$$

We can use this for $x := -\frac{a}{2} + \theta b$, since (4.6) and (4.4) guarantee that $|x| \le \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$. We get

$$\ln\Big(1 - \frac{a}{2} + \theta b\Big) = -\frac{a}{2} + \theta b + O\Big(\big(-\frac{a}{2} + \theta b\big)^2\Big).$$

Expanding the square and letting the $O(\cdot)$ notation absorb the factors bounded by absolute constants, we conclude that

$$\ln\Big(1 - \frac{a}{2} + \theta b\Big) = -\frac{a}{2} + O(b) + O(a^2) + O(ab) + O(b^2) = -\frac{a}{2} + O(b),$$

where the last step follows from (4.6). Recalling (4.5), we see that we proved that

$$\ln\big(\mathbb{E}\, e^{itX}\big) = -\frac{\sigma^2 t^2}{2} + O(\rho^3 t^3),$$

as claimed.

Step 3. Deducing the bound (4.3). In the range $|t| \le \frac{1}{\rho}$, the bound (4.3) follows from (4.2). And if $|t| > \frac{1}{\rho}$, the bound is nearly trivial. Indeed, since $|e^{ix}| = 1$ holds pointwise, we have

$$|\mathbb{E}\, e^{itX}| \le 1.$$

On the other hand, $\sigma \le \rho$ and $\rho|t| > 1$ yield $\sigma^2 t^2/2 \le \rho^2 t^2/2 \le \rho^3 |t|^3$, so

$$\exp\Big(-\frac{\sigma^2 t^2}{2} + \rho^3 |t|^3 \Big) \ge \exp(0) = 1,$$

and (4.3) follows. □

Remark 4.2 (No approximation everywhere). One might ask whether the bound (4.2) could hold for all $t \in \mathbb{R}$, in which case we won't need the separate bound (4.3). This is false in general. For instance, a characteristic function $\mathbb{E}\, e^{itX}$ may have compact support; then the left-hand side of (4.2) vanishes for large $|t|$, while the right-hand side remains positive.

Our next goal is to approximate the characteristic function of a sum of independent random variables. To do this, we will multiply the bounds from Lemma 4.1 to obtain:

Lemma 4.3 (The characteristic function of a sum)**.** *There exist absolute constants $C, c > 0$ so that the following holds. Let $X_1, \ldots, X_n$ be independent random variables that satisfy*

$$\mathbb{E}\, X_k = 0, \quad \mathbb{E}\, X_k^2 = \sigma_k^2, \quad \mathbb{E} |X_k|^3 = \rho_k^3 < \infty.$$

Assume that

$$\sum_{k=1}^n \sigma_k^2 = 1 \quad \textit{and let} \quad \rho^3 := \sum_{k=1}^n \rho_k^3.$$

Then the sum $S_n := X_1 + \cdots + X_n$ satisfies

$$\Big| \mathbb{E}\, e^{itS_n} - e^{-t^2/2} \Big| \le C \rho^3 |t|^3 e^{-t^2/4} \quad \textit{whenever} \quad |t| \le \frac{c}{\rho^3}.$$

Proof. Step 1. Assume that $|t| \le \frac{1}{\rho}$. Then $|t| \le \frac{1}{\rho_k}$ for each k, which allows us to apply (4.2) and get

$$\mathbb{E}\, e^{itX_k} = \exp\Big(-\frac{\sigma_k^2 t^2}{2} + \theta_k \rho_k^3 t^3 \Big) \quad \text{for some } \theta_k = \theta_k(t) \text{ satisfying } |\theta_k| \le C.$$

By independence, this yields

$$\mathbb{E}\, e^{itS_n} = \prod_{k=1}^n \mathbb{E}\, e^{itX_k} = \exp\Big(-\frac{t^2}{2} + \theta \rho^3 t^3 \Big) \quad \text{for some } \theta = \theta(t) \text{ satisfying } |\theta| \le C.$$

Therefore

$$\Big| \mathbb{E}\, e^{itS_n} - e^{-t^2/2} \Big| = e^{-t^2/2} \Big| e^{\theta \rho^3 t^3} - 1 \Big| \le C_1 \rho^3 |t|^3 e^{-t^2/2},$$

and we are done. (The last bound follows once we apply the Taylor approximation of the exponential function $|e^x - 1| \le |x| e^{|x|}$ for $x := \theta \rho^3 t^3$ and note that $|x| \le C$ by assumption on t.)

Step 2. Assume that $\frac{1}{\rho} < |t| \le \frac{c}{\rho^3}$. Arguing similarly to Step 1, but applying (4.3) instead, we obtain

$$\big| \mathbb{E}\, e^{itS_n} \big| \le \exp\Big(-\frac{t^2}{2} + \theta \rho^3 t^3 \Big) \quad \text{for some } \theta = \theta(t) \text{ satisfying } |\theta| \le C.$$

Choosing the absolute constant $c > 0$ in the assumption on t small enough, we can make sure that $\theta \rho^3 t^3 \le t^2/4$. This gives

$$\left| \mathbb{E}\, e^{itS_n} \right| \le e^{-t^2/4}.$$

Hence, by triangle inequality, we conclude that

$$\left| \mathbb{E}\, e^{itS_n} - e^{-t^2/2} \right| \le e^{-t^2/4} + e^{-t^2/2} \le 2\rho^3 |t|^3 e^{-t^2/4},$$

since $\rho^3 |t|^3 \ge 1$ by assumption. The lemma is proved. □

5. Proof of Theorem 0.1

Now we are ready to prove the Berry-Esseen Theorem 0.1. Set

$$\rho^3 := \sum_{k=1}^{n} \mathbb{E}|X_k|^3.$$

Apply the Esseen's smoothing inequality (Theorem 3.3) for the standard normal random variable $Y = G$ and for $1/\varepsilon = c/\rho^3$. Since the density of G is bounded by an absolute constant and its characteristic function equals $e^{-t^2/2}$, we obtain

$$\sup_{a \in \mathbb{R}} \left| \mathbb{P}\{X \le a\} - \mathbb{P}\{G \le a\} \right| \le 2 \int_{-c/\rho^3}^{c/\rho^3} \left| \frac{\mathbb{E}\, e^{itX} - e^{-t^2/2}}{t} \right| dt + C_1 \rho^3.$$

Now substitute the bound on the characteristic function of S_n given by Lemma 4.3. We get

$$\sup_{a \in \mathbb{R}} \left| \mathbb{P}\{X \le a\} - \mathbb{P}\{Y \le a\} \right| \le C_2 \rho^3 \int_{-\infty}^{\infty} t^2 e^{-t^2/4}\, dt + C_1 \rho^3 \le C_3 \rho^3.$$

The Berry-Esseen Theorem 0.1 is proved. □

Acknowledgements

I am grateful to Pedro Abdalla and Guangyi Zou, who kindly read an early version of this note and helped eliminate several typos and inaccuracies.

References

[1] A. C. Berry, *The accuracy of the Gaussian approximation to the sum of independent variates*, Transactions of the AMS 49 (1941), 122–136.

[2] C.-G. Esseen, *On the Liapunoff limit of error in the theory of probability*, Arkiv för Matematik, Astronomi och Fysik A28 (1942), 1–19.

[3] C.-G. Esseen, *Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law*, Acta Math. 77 (1945), 1–125.

[4] W. Feller, *An introduction to probability theory and its applications.* 2nd ed., Vol. II. John Wiley & Sons, 1971.

[5] C. Stein, L. Goldstein, Q.-M. Shao, *Normal approximation by Stein's method,* Probability and its Applications, Springer, 2011.

Department of Mathematics, University of California, Irvine, U.S.A.
Email address: rvershyn@uci.edu