The Berry-Esseen theorem

Jordan Bell

June 3, 2015

1 Cumulative distribution functions

For a random variable $X : (\Omega, \mathscr{F}, P) \to \mathbb{R}$, we define its **cumulative distribution function** $F_X : \mathbb{R} \to [0,1]$ by

$$F_X(x) = P(X \leq x) = \int_{\{X \leq x\}} dP = \int_{t \leq x} d(X_*P)(t) = (X_*P)((-\infty, x]).$$

A **distribution function** is a function $F : \mathbb{R} \to [0,1]$ such that (i) $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$, (ii) $F(\infty) = \lim_{x\to\infty} F(x) = 1$, (iii) F is nondecreasing, (iv) F is **right-continuous**: for each $x \in \mathbb{R}$,

$$F(x+) = \lim_{t \downarrow x} F(t) = F(x).$$

It is a fact that the cumulative distribution function of a random variable is a distribution function and that for any distribution function F there is a random variable X for which $F = F_X$.

Let γ_1 be the **standard Gaussian measure on** $\mathbb{R}$: γ_1 has density

$$p(t, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

with respect to Lebesgue measure on $\mathbb{R}$. Let Φ be the cumulative distribution function of γ_1:

$$\Phi(x) = \gamma_1((-\infty, x]) = \int_{-\infty}^x d\gamma_1(t) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

We first prove the following lemma about distribution functions.[1]

Lemma 1. Suppose that F is a distribution function, that $G : \mathbb{R} \to \mathbb{R}$ satisfies

$$G(-\infty) = \lim_{x\to-\infty} G(x) = 0, \qquad G(\infty) = \lim_{x\to\infty} G(x) = 1,$$

[1] Kai Lai Chung, *A Course in Probability Theory*, third ed., p. 236, Lemma 1; cf. Allan Gut, *Probability: A Graduate Course*, second ed., p. 358, Lemma 6.1.

and that G is differentiable and its derivative satisfies

$$M = \sup_{x \in \mathbb{R}} |G'(x)| < \infty. \tag{1}$$

Writing

$$\Delta = \frac{1}{2M} \sup_{x \in \mathbb{R}} |F(x) - G(x)|,$$

there is some $a \in \mathbb{R}$ such that for all $T > 0$,

$$\begin{aligned} & 2MT\Delta \left(3 \int_0^{T\Delta} \frac{1 - \cos x}{x^2} dx - \pi \right) \\ & \leq \left| \int_{\mathbb{R}} \frac{1 - \cos Tx}{x^2} (F(x+a) - G(x+a)) dx \right|. \end{aligned}$$

Proof. Because $G(-\infty) = 0$ and $G(\infty) = 1$, there is some compact interval K such that $-1 < G(x) < 2$ for $x \in \mathbb{R} \setminus K$. Then, because G is continuous it is bounded on K, showing that G is bounded on $\mathbb{R}$, and because $M > 0$ we get $\Delta < \infty$.

Write $H = F - G$. Because $H(\infty) = 0$ and $H(-\infty) = 0$, there is a compact interval K for which

$$2M\Delta = \sup_{x \in \mathbb{R}} |H(x)| = \sup_{x \in K} |H(x)|.$$

By the Bolzano-Weierstrass theorem, either there is a sequence $u_n \in K$ increasing to some $u \in K$ such that $|H(u_n)| \uparrow 2M\Delta$ or there is a sequence $u_n \in K$ decreasing to some $u \in K$ such that $|H(u_n)| \uparrow 2M\Delta$.[2] In the first case, either there is a subsequence v_n of u_n such that $|H(v_n)| = H(v_n)$ or there is a subsequence v_n of u_n such that $|H(v_n)| = -H(v_n)$. In the first subcase we get $H(u-) = 2M\Delta$, thus

$$F(u-) - G(u) = 2M\Delta. \tag{2}$$

In the second subcase we get $H(u-) = -2M\Delta$, thus

$$F(u-) - G(u) = -2M\Delta. \tag{3}$$

In the second case, either there is a subsequence v_n of u_n such that $|H(v_n)| = H(v_n)$ or there is a subsequence v_n of u_n such that $|H(v_n) = -H(v_n)$. In the first subcase we get $H(u+) = 2M\Delta$, thus

$$F(u) - G(u) = 2M\Delta. \tag{4}$$

In the second subcase we get $H(u+) = 2M\Delta$, thus

$$F(u) - G(u) = -2M\Delta. \tag{5}$$

[2] In the proof in Chung there are merely two cases but it is not explained why those are exhaustive.

We now deal with the subcase (3). Let $a = u - \Delta$. For $|x| < \Delta$, by (1) we have

$$|G(x+a) - G(u)| = \left| \int_u^{u+x-\Delta} G'(y)dy \right| \leq |x - \Delta| M = (\Delta - x)M,$$

whence

$$G(x+a) \geq G(u) + (x - \Delta)M.$$

Because $x + a = x + u - \Delta < u$ and as F is nondecreasing and using (3),

$$\begin{aligned} F(x+a) - G(x+a) &\leq F(u-) - G(x+a) \\ &\leq F(u-) - (G(u) + (x - \Delta)M) \\ &= -M(x + \Delta). \end{aligned}$$

Then, because $x \mapsto \frac{1-\cos Tx}{x^2} x$ is an odd function,

$$\begin{aligned} \int_{-\Delta}^{\Delta} \frac{1 - \cos Tx}{x^2}(F(x+a) - G(x+a))dx &\leq -M \int_{-\Delta}^{\Delta} \frac{1 - \cos Tx}{x^2}(x + \Delta)dx \\ &= -2M\Delta \int_0^{\Delta} \frac{1 - \cos Tx}{x^2} dx. \end{aligned}$$

On the other hand,

$$\begin{aligned} &\left| \int_{(-\infty,-\Delta)\cup(\Delta,\infty)} \frac{1 - \cos Tx}{x^2}(F(x+a) - G(x+a))dx \right| \\ \leq &2M\Delta \int_{(-\infty,-\Delta)\cup(\Delta,\infty)} \frac{1 - \cos Tx}{x^2} dx \\ = &4M\Delta \int_{\Delta}^{\infty} \frac{1 - \cos Tx}{x^2} dx. \end{aligned}$$

Thus

$$\begin{aligned} &\int_{\mathbb{R}} \frac{1 - \cos Tx}{x^2}(F(x+a) - G(x+a))dx \\ \leq &-2M\Delta \int_0^{\Delta} \frac{1 - \cos Tx}{x^2} dx + 4M\Delta \int_{\Delta}^{\infty} \frac{1 - \cos Tx}{x^2} dx \\ = &2M\Delta \left(-3 \int_0^{\Delta} \frac{1 - \cos Tx}{x^2} dx + 2 \int_0^{\infty} \frac{1 - \cos Tx}{x^2} dx \right) \\ = &2M\Delta \left(-3 \int_0^{\Delta} \frac{1 - \cos Tx}{x^2} dx + 2 \cdot \frac{\pi T}{2} \right) \\ = &2MT\Delta \left(-3 \int_0^{T\Delta} \frac{1 - \cos x}{x^2} dx + \pi \right), \end{aligned}$$

which yields the claim of the lemma, for the subcase (3). □

We now prove a lemma that gives an inequality for characteristic functions.[3] We remark that because F is a distribution function, it makes sense to speak about the measure induced by F, and because G is of bounded variation and is continuous, its variation function V_G is continuous and the functions $V_G - G$ and V_G are nondecreasing, and it thus makes sense to speak about the signed measure induced by $G = V_G - (V_G - G)$, which is equal to the difference of the measures induced by V_G and $V_G - G$.

Lemma 2. Suppose that F is a distribution function, that $G : \mathbb{R} \to \mathbb{R}$ satisfies

$$G(-\infty) = \lim_{x \to -\infty} G(x) = 0, \qquad G(\infty) = \lim_{x \to \infty} G(x) = 1,$$

that G is differentiable and of bounded variation and that its derivative satisfies

$$M = \sup_{x \in \mathbb{R}} |G'(x)| < \infty,$$

and that

$$\int_{\mathbb{R}} |F - G| dx < \infty.$$

Write

$$\Delta = \frac{1}{2M} \sup_{x \in \mathbb{R}} |F(x) - G(x)|$$

and

$$f(t) = \int_{\mathbb{R}} e^{itx} dF(x), \qquad g(t) = \int_{\mathbb{R}} e^{itx} dG(x).$$

Then for all $T > 0$,

$$\Delta \leq \frac{1}{\pi M} \int_0^T \frac{|f(t) - g(t)|}{t} dt + \frac{12}{\pi T}.$$

Proof. For any $t \in \mathbb{R}$, because $(F - G)(-\infty) = 0$ and $(F - G)(\infty) = 0$ and because $\int_{\mathbb{R}} |F - G| dx < \infty$, integrating by parts gives

$$\begin{aligned} f(t) - g(t) &= \int_{\mathbb{R}} e^{itx} dF(x) - \int_{\mathbb{R}} e^{itx} dG(x) \\ &= \int_{\mathbb{R}} e^{itx} d(F - G)(x) \\ &= -it \int_{\mathbb{R}} (F - G)(x) e^{itx} dx. \end{aligned}$$

Take a to be the real number that Lemma 1 yields. As

$$\frac{f(t) - g(t)}{-it} e^{-ita} (T - |t|) = (T - |t|) \int_{\mathbb{R}} (F(x + a) - G(x + a)) e^{itx} dx,$$

[3] Kai Lai Chung, *A Course in Probability Theory*, third ed., p. 237, Lemma 2; Zhengyan Lin and Zhidong Bai, *Probability Inequalities*, p. 29, Theorem 4.1.a.

we obtain, using Fubini's theorem,

$$
\begin{aligned}
&\int_{-T}^{T} \frac{f(t)-g(t)}{-it} e^{-ita}(T-|t|)dt \\
=&\int_{-T}^{T} \left((T-|t|) \int_{\mathbb{R}} (F(x+a)-G(x+a))e^{itx}dx \right) dt \\
=&\int_{\mathbb{R}} (F(x+a)-G(x+a)) \left(\int_{-T}^{T} (T-|t|)e^{itx}dt \right) dx \\
=&2\int_{\mathbb{R}} (F(x+a)-G(x+a)) \frac{1-\cos Tx}{x^2} dx.
\end{aligned}
$$

Therefore, because F and G are real valued and thus $|f(-t)-g(-t)| = |\overline{f(t)-g(t)}| = |f(t)-g(t)|$,

$$
\begin{aligned}
\left| \int_{\mathbb{R}} (F(x+a)-G(x+a)) \frac{1-\cos Tx}{x^2} dx \right| &\leq \frac{1}{2} \int_{-T}^{T} \frac{|f(t)-g(t)|}{|t|} (T-|t|)dt \\
&= \int_0^T \frac{|f(t)-g(t)|}{t} (T-t)dt \\
&\leq T \int_0^T \frac{|f(t)-g(t)|}{t} dt.
\end{aligned}
$$

Using this with Lemma 1,

$$
2MT\Delta \left(3\int_0^{T\Delta} \frac{1-\cos x}{x^2} dx - \pi \right) \leq T \int_0^T \frac{|f(t)-g(t)|}{t} dt.
$$

But

$$
\begin{aligned}
3\int_0^{T\Delta} \frac{1-\cos x}{x^2} dx - \pi &= 3\int_0^{\infty} \frac{1-\cos x}{x^2} dx - 3\int_{T\Delta}^{\infty} \frac{1-\cos x}{x^2} dx - \pi \\
&\geq 3\int_0^{\infty} \frac{1-\cos x}{x^2} dx - 6\int_{T\Delta}^{\infty} \frac{1}{x^2} dx - \pi \\
&= 3 \cdot \frac{\pi}{2} - \frac{6}{T\Delta} - \pi \\
&= \frac{\pi}{2} - \frac{6}{T\Delta},
\end{aligned}
$$

with which we have

$$
2MT\Delta \left(3\int_0^{T\Delta} \frac{1-\cos x}{x^2} dx - \pi \right) \geq 2MT\Delta \cdot \left(\frac{\pi}{2} - \frac{6}{T\Delta} \right) = MT\Delta\pi - 12M,
$$

and hence

$$
MT\Delta\pi - 12M \leq T \int_0^T \frac{|f(t)-g(t)|}{t} dt,
$$

i.e.

$$\Delta \leq \frac{12}{\pi T} + \frac{1}{\pi M} \int_0^T \frac{|f(t) - g(t)|}{t} dt,$$

proving the claim. □

2 Berry-Esseen theorem

Let $X_{n,j}$, $n \geq 1$, $1 \leq j \leq k_n$, be L^3 random variables, with $k_n \to \infty$, such that for each n, the random variables $X_{n,j}$, $1 \leq j \leq k_n$, are independent, and such that for all n and j,

$$E(X_{n,j}) = 0.$$

Let $F_{n,j}$ be the cumulative distribution function of $X_{n,j}$:

$$F_{n,j}(x) = P(X_{n,j} \leq x).$$

Let $f_{n,j}$ be the characteristic function of $X_{n,j}$ (equivalently, the characteristic function of $F_{n,j}$):

$$f_{n,j}(t) = \int_{\mathbb{R}} e^{itx} d(X_{n,j*}P)(x) = \int_{\mathbb{R}} e^{itx} dF_{n,j}(x).$$

Write, for $n \geq 1$,

$$S_n = \sum_{j=1}^{k_n} X_{n,j},$$

and let F_n be the cumulative distribution function of S_n:

$$F_n(x) = P(S_n \leq x)$$

Also, let f_n be the characteristic function of S_n (equivalently, the characteristic function of F_n). Because $X_{n,j}$, $1 \leq j \leq k_n$, are independent, we have $S_{n*}P = (X_{n,1*}P) * \cdots * (X_{n,k_n*}P)$ and hence

$$f_n(t) = \int_{\mathbb{R}} e^{itx} d(S_{n*}P)(x) = \prod_{j=1}^{k_n} f_{n,j}(t).$$

For $n \geq 1$ and $1 \leq j \leq k_n$, write

$$\sigma_{n,j}^2 = E(X_{n,j}^2), \qquad s_n^2 = \sum_{j=1}^{k_n} \sigma_{n,j}^2$$

and

$$\gamma_{n,j} = E(|X_{n,j}|^3), \qquad \Gamma_n = \sum_{j=1}^{k_n} \gamma_{n,j}.$$

We further assume that for each n,

$$s_n^2 = \sum_{j=1}^{k_n} \sigma_{n,j}^2 = 1. \tag{6}$$

We will use the following inequality which we state separately because it is of general use.

Lemma 3. For $n \geq 1$ and $|z| < 1$,

$$\left| \log(1+z) - \sum_{m=1}^{n-1} \frac{(-1)^{m-1} z^m}{m} \right| \leq \frac{|z|^n}{n(1-|z|)}.$$

We now prove an inequality for f_n, the characteristic function of S_n.[4]

Lemma 4. For $n \geq 1$, if $|t| < \frac{1}{2\Gamma_n^{1/3}}$ then

$$|f_n(t) - e^{-t^2/2}| \leq \Gamma_n |t|^3 e^{-t^2/2}.$$

Proof. For $1 \leq j \leq k_n$ and $l \geq 0$ and $v \in \mathbb{R}$,

$$f_{n,j}^{(l)}(v) = (i)^l E(X_{n,j}^l e^{ivX_{n,j}}).$$

Thus

$$f_{n,j}(0) = 1, \quad f'_{n,j}(0) = iE(X_{n,j}) = 0, \quad f''_{n,j}(0) = -E(X_{n,j}^2) = -\sigma_{n,j}^2,$$

and

$$f'''_{n,j}(v) = -iE(X_{n,j}^3 e^{ivX_{n,j}}).$$

Then by Taylor's theorem, there is some s between 0 and t such that

$$f_{n,j}(t) = 1 - \frac{\sigma_{n,j}^2}{2} t^2 - \frac{iE(X_{n,j}^3 e^{isX_{n,j}})}{6} t^3.$$

Put

$$-iE(X_{n,j}^3 e^{isX_{n,j}}) = \theta \gamma_{n,j},$$

for which

$$|\theta| = \frac{|E(X_{n,j}^3 e^{isX_{n,j}})|}{E(|X_{n,j}|^3)} \leq 1.$$

Because the L^2 norm is upper bounded by the L^3 norm and because $|t| < \frac{1}{2\Gamma_n^{1/3}}$,

$$|\sigma_{n,j} t| \leq |\gamma_{n,j}^{1/3} t| \leq |\Gamma_n^{1/3} t| < \frac{1}{2},$$

[4]Kai Lai Chung, *A Course in Probability Theory*, third ed., p. 239, Lemma 3.

and hence

$$
\begin{aligned}
|f_{n,j}(t) - 1| &= \left| -\frac{\sigma_{n,j}^2}{2} t^2 + \frac{\theta \gamma_{n,j} t^3}{6} \right| \\
&\le \frac{1}{2} |\sigma_{n,j} t|^2 + \frac{\gamma_{n,j}}{48 \Gamma_n} \\
&< \frac{1}{8} + \frac{1}{48} \\
&< \frac{1}{4}.
\end{aligned}
$$

Lemma 3 and the inequality $|a+b|^2 \le 2(|a|^2 + |b|^2)$ then tell us that

$$
\begin{aligned}
|\log f_{n,j}(t) - (f_{n,j}(t) - 1)| &\le \frac{|f_{n,j}(t) - 1|^2}{2(1 - |f_{n,j}(t) - 1|)} \\
&< \frac{2}{3} |f_{n,j}(t) - 1|^2 \\
&= \frac{2}{3} \left| -\frac{\sigma_{n,j}^2}{2} t^2 + \frac{\theta \gamma_{n,j}}{6} t^3 \right|^2 \\
&\le \frac{4}{3} \left(\frac{\sigma_{n,j}^4}{4} t^4 + \frac{|\theta|^2 \gamma_{n,j}^2}{36} t^6 \right).
\end{aligned}
$$

Because $\sigma_{n,j} \le \gamma_{n,j}^{1/3}$ and $|\theta| \le 1$,

$$
\begin{aligned}
|\log f_{n,j}(t) - (f_{n,j}(t) - 1)| &\le \frac{4}{3} \left(\frac{\sigma_{n,j} \gamma_{n,j}}{4} t^4 + \frac{\gamma_{n,j}^2}{36} t^6 \right) \\
&= \frac{4}{3} \left(\frac{|\sigma_{n,j} t|}{4} + \frac{|\gamma_{n,j}^{1/3} t|^3}{36} \right) \gamma_{n,j} |t|^3 \\
&\le \frac{4}{3} \left(\frac{1}{2 \cdot 4} + \frac{1}{8 \cdot 36} \right) \gamma_{n,j} |t|^3 \\
&= \frac{37}{216} \gamma_{n,j} |t|^3 \\
&< \frac{1}{5} \gamma_{n,j} |t|^3.
\end{aligned}
$$

Combining this with $f_{n,j}(t) = 1 - \frac{\sigma_{n,j}^2}{2} t^2 + \frac{\theta \gamma_{n,j}}{6} t^3$,

$$
\left| \log f_{n,j}(t) + \frac{\sigma_{n,j}^2}{2} t^2 \right| \le \left| \frac{\theta \gamma_{n,j}}{6} t^3 \right| + \frac{1}{5} \gamma_{n,j} |t|^3 \le \frac{1}{6} \gamma_{n,j} |t|^3 + \frac{1}{5} \gamma_{n,j} |t|^3 \le \frac{1}{2} \gamma_{n,j} |t|^3.
$$

Because this is true for each $1 \le j \le k_n$ and because, according to (6),

$\sum_{j=1}^{k_n} \sigma_{n,j}^2 = 1$,

$$\left|\log f_n(t) + \frac{t^2}{2}\right| \leq \frac{|t|3^2}{2} \sum_{j=1}^{k_n} \gamma_{n,j} = \frac{|t|^3}{2}\Gamma_n.$$

For any $z \in \mathbb{C}$ it is true that $|e^z - 1| \leq |z|e^{|z|}$, so the above yields

$$\begin{aligned}
|f_n(t)e^{t^2/2} - 1| &= |\exp(\log(f_n(t)e^{t^2/2})) - 1| \\
&\leq |\log(f_n(t)e^{t^2/2})| \exp\left(\left|\log(f_n(t)e^{t^2/2})\right|\right) \\
&= \left|\log f_n(t) + \frac{t^2}{2}\right| \exp\left(\left|\log(f_n(t)e^{t^2/2})\right|\right) \\
&\leq \frac{|t|^3}{2}\Gamma_n \exp\left(\frac{|t|^3}{2}\Gamma_n\right).
\end{aligned}$$

But $|t|^3 < \frac{1}{8\Gamma_n}$, so

$$|f_n(t)e^{t^2/2} - 1| \leq \frac{|t|^3}{2}\Gamma_n e^{1/16} \leq |t|^3\Gamma_n,$$

which completes the proof. □

The next lemma gives a different bound on the characteristic function of S_n.[5]

Lemma 5. For $n \geq 1$, if $|t| < \frac{1}{4\Gamma_n}$ then

$$|f_n(t)| \leq e^{-t^2/3}.$$

Proof. First, for a distribution function F with characteristic function f,

$$\begin{aligned}
|f(t)|^2 &= f(t)\overline{f(t)} \\
&= \int_{\mathbb{R}} e^{itx} dF(x) \cdot \int_{\mathbb{R}} e^{-itx} dF(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{it(x-y)} dF(x)\right) dF(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \cos t(x-y) + i \sin t(x-y) dF(x)\right) dF(y).
\end{aligned}$$

Because $|f(t)|^2$ is real it follows that

$$|f(t)|^2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \cos t(x-y) dF(x)\right) dF(y).$$

[5]Kai Lai Chung, *A Course in Probability Theory*, third ed., p. 240, Lemma 4.

Using

$$\left|\cos u - \left(1 - \frac{u^2}{2}\right)\right| \le \frac{|u|^3}{6}, \qquad |a+b|^p \le 2^{p-1}(|a|^p + |b|^p),$$

we have

$$\left|\cos t(x-y) - \left(1 - \frac{(t(x-y))^2}{2}\right)\right| \le \frac{2}{3}\left(|tx|^3 + |ty|^3\right) = \frac{2|t|^3}{3}(|x|^3 + |y|^3)$$

and then

$$|f(t)|^2 \le \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1 - \frac{(t(x-y))^2}{2} + \frac{2|t|^3}{3}(|x|^3 + |y|^3) dF(x) \right) dF(y).$$

Using this for $f_{n,j}$, and using that $E(X_{n,j}) = 0$,

$$\begin{aligned}
|f_{n,j}(t)|^2 &\le \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1 - \frac{(t(x-y))^2}{2} + \frac{2|t|^3}{3}(|x|^3 + |y|^3) dF_{n,j}(x) \right) dF_{n,j}(y) \\
&= \int_{\mathbb{R}} 1 - \frac{t^2\sigma_{n,j}^2}{2} - \frac{t^2 y^2}{2} + \frac{2|t|^3\gamma_{n,j}}{3} + \frac{2|t|^3|y|^3}{3} dF(y) \\
&= 1 - \frac{t^2\sigma_{n,j}^2}{2} - \frac{t^2\sigma_{n,j}^2}{2} + \frac{2|t|^3\gamma_{n,j}}{3} + \frac{2|t|^3\gamma_{n,j}}{3} \\
&= 1 - t^2\sigma_{n,j}^2 + \frac{4|t|^3\gamma_{n,j}}{3}.
\end{aligned}$$

Because $1 + u \le e^u$ for all $u \in \mathbb{R}$,

$$|f_{n,j}(t)|^2 \le \exp\left(-t^2\sigma_{n,j}^2 + \frac{4|t|^3\gamma_{n,j}}{3}\right).$$

Then, by (6),

$$\begin{aligned}
|f_n(t)|^2 &= \prod_{j=1}^{k_n} |f_{n,j}(t)|^2 \\
&\le \prod_{j=1}^{k_n} \exp\left(-t^2\sigma_{n,j}^2 + \frac{4|t|^3\gamma_{n,j}}{3}\right) \\
&= \exp\left(-t^2 \sum_{j=1}^{k_n} \sigma_{n,j}^2 + \frac{4|t|^3}{3} \sum_{j=1}^{k_n} \gamma_{n,j}\right) \\
&= \exp\left(-t^2 + \frac{4|t|^3}{3}\Gamma_n\right).
\end{aligned}$$

As $|t| < \frac{1}{4\Gamma_n}$,

$$|f_n(t)| \le \exp\left(-\frac{t^2}{2} + \frac{2|t|^3}{3}\Gamma_n\right) \le \exp\left(-\frac{t^2}{2} + \frac{2|t|^2}{12}\right) = e^{-t^2/3},$$

proving the claim. □

We now combine Lemma 4 and Lemma 5.[6]

Lemma 6. For $n \geq 1$, if $|t| < \frac{1}{4\Gamma_n}$ then

$$|f_n(t) - e^{-t^2/2}| \leq 16\Gamma_n |t|^3 e^{-t^2/3}.$$

Proof. Either $|t| < \frac{1}{2\Gamma_n^{1/3}}$ or $\frac{1}{2\Gamma_n^{1/3}} \leq |t| < \frac{1}{4\Gamma_n}$. In the first case, Lemma 4 tells us

$$|f_n(t) - e^{-t^2/2}| \leq \Gamma_n |t|^3 e^{-t^2/2} \leq \Gamma_n |t|^3 e^{-t^2/3} \leq 16\Gamma_n |t|^3 e^{-t^2/3}.$$

In the second case, Lemma 5 tells us

$$|f_n(t)| \leq e^{-t^2/3},$$

and so, as in this case we have $1 \leq 8\Gamma_n |t|^3$,

$$|f_n(t) - e^{-t^2/2}| \leq |f_n(t)| + e^{-t^2/2} \leq e^{-t^2/3} + e^{-t^2/2} \leq 2e^{-t^2/3} \leq 16\Gamma_n |t|^3 e^{-t^2/3},$$

showing that the claim is true in both cases. □

We finally prove the **Berry-Esseen theorem**.[7]

Theorem 7 (Berry-Esseen theorem)**.** There is some $A_0 < 36$ such that for each $n \geq 1$,

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq A_0 \Gamma_n.$$

Proof. Let Z be a random variable with $Z_* P = \gamma_1$, i.e. whose cumulative distribution function is Φ. By (6) and because $X_{n,j}$, $1 \leq j \leq k_n$, are independent and satisfy $E(X_{n,j}) = 0$,

$$E(S_n^2) = \sum_{j=1}^{k_n} E(X_{n,j}^2) = \sum_{j=1}^{k_n} \sigma_{n,j}^2 = 1.$$

If $x < 0$ then by Chebyshev's inequality

$$F_n(x) = P(S_n \leq x) = P(-S_n \geq -x) \leq \frac{1}{x^2} E(|S_n|^2) = \frac{1}{x^2}$$

and

$$\Phi(x) = P(Z \leq x) = P(-Z \geq -x) \leq \frac{1}{x^2} E(|Z|^2) = \frac{1}{x^2}.$$

[6] Kai Lai Chung, *A Course in Probability Theory*, third ed., p. 240, Lemma 5.

[7] Kai Lai Chung, *A Course in Probability Theory*, third ed., p. 235, Theorem 7.4.1; cf. Allan Gut, *Probability: A Graduate Course*, second ed., p. 356, Theorem 6.2; John E. Kolassa, *Series Approximation Methods in Statistics*, p. 25, Theorem 2.6.1; Alexandr A. Borovkov, *Probability Theory*, p. 659, Theorem A5.1; Ivan Nourdin and Giovanni Peccati, *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*, p. 71, Theorem 3.7.1.

If $x > 0$ then also by Chebyshev's inequality

$$1 - F_n(x) = 1 - P(S_n \le x) = P(S_n > x) \le \frac{1}{x^2}$$

and

$$1 - \Phi(x) = 1 - P(Z \le x) = P(Z > x) \le \frac{1}{x^2}.$$

Therefore, because F_n and Φ are nonnegative and $1 - F_n$ and $1 - \Phi$ are nonnegative, for all $x \in \mathbb{R}$ we have

$$|F_n(x) - \Phi(x)| \le \frac{1}{x^2}.$$

Then, because $|F_n| \le 1$ and $|\Phi| \le 1$,

$$\int_{\mathbb{R}} |F_n(x) - \Phi(x)| dx \le \int_{|x|\le 1} 2dx + \int_{|x|>1} \frac{1}{x^2} dx = 6 < \infty.$$

$\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \le \frac{1}{\sqrt{2\pi}}$. We apply Lemma 2 with $F = F_n$, $G = \Phi$, and $M = \frac{1}{\sqrt{2\pi}}$, and because the characteristic function of Φ is $\phi(t) = e^{-t^2/2}$, we obtain for $T = \frac{1}{4\Gamma_n}$,

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| &\le \frac{2}{\pi} \int_0^{\frac{1}{4\Gamma_n}} \frac{|f_n(t) - \phi(t)|}{t} dt + \frac{96 M \Gamma_n}{\pi} \\
&= \frac{2}{\pi} \int_0^{\frac{1}{4\Gamma_n}} \frac{|f_n(t) - e^{-t^2/2}|}{t} + \frac{96\Gamma_n}{\pi\sqrt{2\pi}}.
\end{aligned}$$

Then applying Lemma 6,

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| &\le \frac{2}{\pi} \int_0^{\frac{1}{4\Gamma_n}} \frac{16\Gamma_n t^3 e^{-t^2/3}}{t} dt + \frac{96\Gamma_n}{\pi\sqrt{2\pi}} \\
&= \Gamma_n \left(\frac{32}{\pi} \int_0^{\frac{1}{4\Gamma_n}} t^2 e^{-t^2/3} dt + \frac{96}{\pi\sqrt{2\pi}} \right).
\end{aligned}$$

This proves the claim with

$$A_0 = \frac{32}{\pi} \int_0^{\infty} t^2 e^{-t^2/3} dt + \frac{96}{\pi\sqrt{2\pi}} = \frac{32}{\pi} \cdot \frac{3\sqrt{3\pi}}{4} + \frac{96}{\pi\sqrt{2\pi}} = 35.64\ldots.$$

□