Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2017

Lecture 11: The Central Limit Theorem

Relevant textbook passages:

Pitman [16]: Section 3.3, especially p. 196

Larsen–Marx [12]: Section 4.3, (Appendix 4.A.2 is optional)

11.1 Fun with CDFs

Recall a random variable X on the probability space $(S, \mathcal{E}, P)$ has a **cumulative distribution function** (cdf) F defined by

$$F(x) = P(X \leqslant x) = P\big(\{s \in S : X(s) \leqslant p\}\big).$$

11.1.1 Quantiles

Pitman [16]: pp.319–323

Assume X is a random variable with a **continuous, strictly increasing** cumulative distribution function F.

- Then the equation

$$P(X \leqslant x) = F(x) = p$$

has a unique solution $x_p = F^{-1}(p)$ for every p with $0 < p < 1$, namely $x_p = F^{-1}(p)$.

- (If we allow x to take on the values $\pm\infty$, then we can be sure that a solution exists for the cases $p = 0$ and $p = 1$.)

11.1.1 Definition *When X has a continuous, strictly increasing cumulative distribution function F, the value $x_p = F^{-1}(p)$ is called the p^{th}* ***quantile*** *of the distribution F. The mapping $F^{-1} : p \mapsto x_p$ is called* ***quantile function*** *of the distribution F.*

Some of the quantiles have special names.

- When p has the form of a percent, x_p is called a **percentile**.
- The $p = 1/2$ quantile is called the **median**.
- The $p = 1/4$ quantile is the **first quartile**, and the $p = 3/4$ is called the **third quartile**.
- The interval between the first and third quartiles is called the **interquartile range**.
- There are also **quintiles**, and **deciles**, and, I'm sure, others as well.

11.1.2 Proposition *Let F be the cdf of the random variable X. Assume that F is continuous. Then $F \circ X$ is a random variable that is uniformly distributed on $[0, 1]$. In other words,*

$$P(F(X) \leqslant p) = p, \qquad (0 \leqslant p \leqslant 1).$$

Note that we are not assuming that F is strictly increasing.

Proof: Since F is continuous, its range has no gaps, that is, the range includes $(0,1)$. Now fix $p \in [0,1]$. There are three cases, $p = 1$, $0 < p < 1$, and $p = 1$.

- The case $p = 1$ is trivial, since F is bounded above by 1.
- In case $0 < p < 1$, define

$$\begin{aligned} x_p &= \sup\{x \in \boldsymbol{R} : F(x) \leqslant p\} \\ &= \sup\{x \in \boldsymbol{R} : F(x) = p\} \end{aligned}$$

and note that x_p is finite. In fact, if F is strictly increasing, then x_p is just the p^{th} quantile defined above.

By continuity,

$$F(x_p) = p.$$

By the construction of x_p, for all $x \in \boldsymbol{R}$,

$$F(x) \leqslant p \iff x \leqslant x_p. \tag{1}$$

So, replacing x by X above,

$$F(X) \leqslant p \iff X \leqslant x_p$$

so

$$P\left(F(X) \leqslant p\right) = P\left(X \leqslant x_p\right) = F(x_p) = p.$$

- The above argument works for $p = 0$ if 0 is in the range of F. But if 0 is not in the range of F, then $F(X) > 0$ a.s., so $P\left(F(X) = 0\right) = 0$.

∎

11.1.2 The fake inverse of a CDF

Even if F is not continuous, so that the range of F has gaps, there is still a "fake inverse" of F that is very useful. Define

$$F^\star \colon (0,1) \to \boldsymbol{R}$$

by

$$F^\star(p) = \inf\{x \in \boldsymbol{R} : F(x) \geqslant p\}.$$

(This notation is my own.)

> When F is strictly increasing and continuous, then $F^\star$ is just F^{-1} on $(0,1)$.

More generally, flat spots in F correspond to jumps in $F^\star$ and vice-versa. The key property is that for any $p \in (0,1)$ and $x \in \boldsymbol{R}$,

$$F^\star(p) \leqslant x \iff p \leqslant F(x), \tag{2}$$

or equivalently

$$F(x) \geqslant p \iff x \geqslant F^\star(p).$$

Compare this to equation (1).

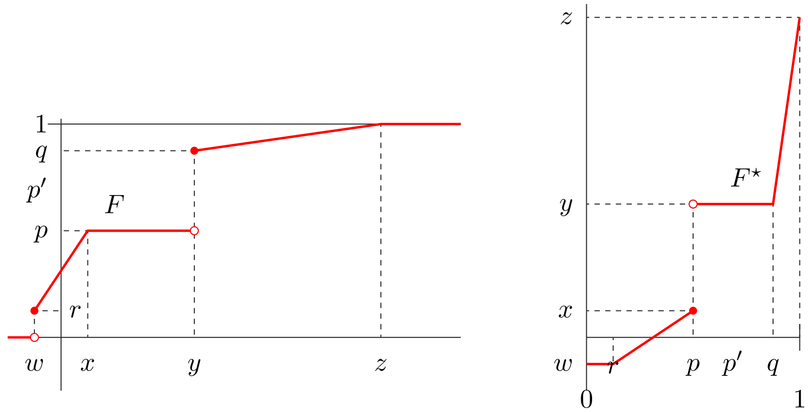

Figure 11.1. Construction of $F^\star$ from F.

11.1.3 Example Refer to Figure 11.1 to get a feel for the relationship between flats and jumps. The cumulative distribution function F has a flat spot from x to y, which means $P(x < X < y) = 0$. Now $\{v : F(v) \geqslant p\} = [x, \infty)$, so $F^\star(p) = x$.

It also has jump from p to q at y, which means $P(y) = q - p$. For p' in the gap $p < p' < q$, we have $\{v : F(v) \geqslant p'\} = [y, \infty)$, so $F^\star(p') = y$. This creates a flat spot in $F^\star$. □

11.1.4 Proposition *Let U be a Uniform[0,1] random variable. That is, for every $p \in [0, 1]$, $P(U \leqslant p) = p$. Note that $P(U \in (0, 1)) = 1$ so with probability one $F^\star(U)$ is defined. Then $F^\star(U)$ is a random variable and*

the cumulative distribution function of $F^\star(U)$ is F.

Proof: From (2),

$$P(F^\star(U) \leqslant x) = P(U \leqslant F(x)) = F(x),$$

where the last equality comes from the fact that U is uniformly distributed. ∎

The usefulness of this result is this:

If you can have a uniform random variable U, then you can create a random variable X with any distribution F via the transformation $X = F^\star(U)$.

11.2 ⋆ Stochastic dominance and increasing functions

Recall that a random variable X **stochastically dominates** the random variable Y if for every x,

$$P(X > x) \geqslant P(Y > x).$$

This is equivalent to $F_X(x) \leqslant F_Y(x)$ for all x.

11.2.1 Proposition *Let X stochastically dominate Y, and let g be a nondecreasing real function. Then* Not in Pitman

$$\boldsymbol{E}\, g(X) \geqslant \boldsymbol{E}\, g(Y).$$

If g is strictly increasing, then the inequality is strict unless $F_X = F_Y$.

As a special case, when $g(x) = x$, we have $\boldsymbol{E}\, X \geqslant \boldsymbol{E}\, Y$.

I will prove this in two ways. The first proof is for the special where X and Y are strictly bounded in absolute value by b, and have densities f_X and f_Y, and the function g is continuous continuously differentiable. Then the expected value of $g(X)$ is obtained via the integral

$$\int_{-b}^{b} g(x) f_X(x)\,dx,$$

so integrating by parts we see this is equal to

$$g(t)F_X(t)\Big|_{-b}^{b} - \int_{-b}^{b} F_X(x) g'(x)\,dx.$$

Likewise $\boldsymbol{E}\,g(Y)$ is equal to

$$g(t)F_Y(t)\Big|_{-b}^{b} - \int_{-b}^{b} F_Y(x) g'(x)\,dx.$$

Now by assumption $F_X(-b) = F_Y(-b) = 0$ and $F_X(b) = F_Y(b) = 1$, so the first term in each expectation is identical. Since g is nondecreasing, $g' \geqslant 0$ everywhere. Since $F_X \leqslant F_Y$ everywhere, we conclude $\boldsymbol{E}\,g(X) \geqslant \boldsymbol{E}\,g(Y)$.

The second proof is more general, and relies on the inverse of the CDF (aka the **quantile function**. Last time we showed that even if a cdf F is not invertible because it has jumps, we can use the function

$$F^\star(p) = \inf\{x \in \boldsymbol{R} : F(x) \geqslant p\}.$$

We showed that if U is a Uniform[0,1] random variable, then $F^\star(U)$ is a random variable that has distribution F. Now observe that if $F_X \leqslant F_Y$, then $F_X^\star \geqslant F_Y^\star$. Therefore, since g is nondecreasing,

$$g\big(F_X^\star(U)\big) \geqslant g\big(F_Y^\star(U)\big)$$

Since expectation is a positive operator, we have

$$\boldsymbol{E}\,g\big(F_X^\star(U)\big) \geqslant \boldsymbol{E}\,g\big(F_Y^\star(U)\big).$$

But $g\big(F_X^\star(U)\big)$ has the same distribution as $g(X)$ and $g\big(F_Y^\star(U)\big)$ has the same distribution as $g(Y)$, so their expectations are the same. q.e.d.

For yet a different sort of proof, based on convex analysis, see Border [3].

11.3 ⋆ When are distributions close?

I have already argued that the DeMoivre–Laplace Limit Theorem says that the Binomial(n,p) distributions can be approximated by the Normal distributions when n is large. The argument consisted mostly of showing how the Normal density approximates the Binomial pmf. In this lecture I will formalize the notion of closeness for distributions.

Not in Pitman.

The appropriate notion is what is called **convergence in distribution**. Informally, two distributions are close if their cumulative distribution functions are close. This is fine if both cdfs are continuous. Then we can define the distance between F and G to be $\sup_x |F(x) - G(x)|$. This definition is reasonable in a few other cases as well. For instance, on the last homework assignment, you had to plot the empirical cdf of data from the coin-tossing experiment, and it was pretty close everywhere to a straight line.

But there is another kind of closeness we wish to capture. Suppose X is a random variable that takes on the values 1 and 0, each with probability 1/2; and Y is a random variable that takes on the values 1.00001 and 0, each with probability 1/2. Should their distributions be considered close? I'd like to think so. There is a simple way to incorporate both notions of closeness, which is captured by the following somewhat opaque definition.

11.3.1 Definition *The sequence F_n of cumulative distribution functions* ***converges in distribution*** *to the cumulative distribution function F, written*

$$F_n \xrightarrow{\mathcal{D}} F$$

if for all t at which F is continuous,

$$F_n(t) \to F(t),$$

We also say the random variables X_n converge in distribution to X, written $X_n \xrightarrow{\mathcal{D}} X$ if their cdfs converge in distribution.

Add pictures.

11.3.2 Example Let X_n be a random variable that takes on the values $1 + (1/n)$ and 0, each with probability 1/2, and let X take on the values 1 and 0, each with probability 1/2. Then the cdf F_n of X_n has jumps of size 1/2 at 0 and $1 + (1/n)$, while the cdf F of X has jumps of size 1/2 at 0 and 1. You can check that at every t except at $t = 1$, which is a point of discontinuity of F, $F_n(t) \to F(t)$. Thus $F_n \xrightarrow{\mathcal{D}} F$. □

11.3.3 Remark Note that convergence in distribution depends only on the cdf, so random variables can be defined on different probability spaces and still converge in distribution. This is not true of convergence in probability or almost-sure convergence.

Aside: It is possible to quantify how close two distribution functions are using the **Lévy metric**, which is defined by

$$d(F, G) = \inf \Big\{ \varepsilon > 0 : \big(\forall x\big) \ \Big[F(x - \varepsilon) - \varepsilon \leqslant G(x) \leqslant F(x + \varepsilon) + \varepsilon \Big] \Big\}.$$

It can be shown that $F_n \xrightarrow{\mathcal{D}} F$ if and only if $d(F_n, F) \to 0$. See, for instance, Billingsley [2, Problem 4, p. 21–22].

11.3.4 Fact *It can be shown that if X_n has cdf F_n and X has cdf F, and if g is a bounded continuous function, then*

$$F_n \xrightarrow{\mathcal{D}} F \implies \boldsymbol{E}\, g(X_n) \to \boldsymbol{E}\, g(X).$$

See, e.g., Breiman[4, § 8.3, pp. 163–164]. In fact, this can be taken as a definition for convergence in distribution.

Aside: The fact above does not imply that if $X_n \xrightarrow{\mathcal{D}} X$, then $\boldsymbol{E}\, X_n \to \boldsymbol{E}\, X$, since the function $f(x) = x$ is not bounded. Ditto for the variance. There is a more complicated result for unbounded functions that applies often enough to be useful. Here it is.

Let $X_n \xrightarrow{\mathcal{D}} X$, and let g and h be continuous functions such that $|h(x)| \to \infty$ as $x \to \pm\infty$, and $|g(x)/h(x)| \to 0$ as $x \to \pm\infty$. If $\limsup_{n\to\infty} \boldsymbol{E}(|h(X_n)|) < \infty$, then $\boldsymbol{E}\big(g(X_n)\big) \to \boldsymbol{E}\big(g(X)\big)$. See Breiman [4, § 8.3, pp. 163–164, exercise 14]. So as a consequence,

If $X_n \xrightarrow{\mathcal{D}} X$, and $\boldsymbol{E}(|X_n^3|)$ is bounded, then $\boldsymbol{E}(X_n) \to \boldsymbol{E}(X)$ and $\boldsymbol{Var}(X_n) \to \boldsymbol{Var}(X)$.

11.4 Central Limit Theorem

We saw in the Law of Large Numbers that if S_n is the sum of n independent and identically distributed random variable with finite mean μ and standard deviation σ, then the sample mean

$A_n = S_n/n$ converges to μ as $n \to \infty$. This is because the mean of $A_n = S_n/n$ is μ and the standard deviation is equal to $\sigma/\sqrt{n}$, so the distribution is collapsing around μ.

What if we don't want the distribution to collapse? Let's standardize the average A_n:

$$A_n^* = \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{\frac{S_n}{n} - \mu}{\sigma/\sqrt{n}} = \frac{n}{n}\frac{\frac{S_n}{n} - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}.$$

Or equivalently, by Proposition 6.12.2, let's look at $S_n/\sqrt{n}$ rather than S_n/n. The mean of $S_n/\sqrt{n}$ is $\sqrt{n}\mu$ and the standard deviation is σ. The standardization is

$$\frac{\frac{S_n}{\sqrt{n}} - \sqrt{n}\mu}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma},$$

which is the same as the standardization of A_n.

One version of the Central Limit Theorem tells what happens to the distribution of the standardization of A_n. The DeMoivre–Laplace Limit Theorem is a special case of the Central Limit Theorem that applies to the Binomial Distribution. Cf. Pitman [16, p. 196] and Larsen–Marx [12, Theorem 4.3.2, pp. 246–247].

11.4.1 Central Limit Theorem, v. 1 *Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables. Let $\mu = \boldsymbol{E}\, X_i$ and $\sigma^2 = \boldsymbol{Var}\, X_i$. Define S_n by $S_n = \sum_{i=1}^n X_i$.*

Then

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} N(0,1).$$

The proof of this result is beyond the scope of this course, but I have included a completely optional appendix to these notes sketching it.

In fact, the CLT (as it is known to its friends) is even more general.

11.5 ⋆ A CLT for non-identical distributions

What if the X_is are not identically distributed? The CLT can still hold. The following result due to J. W. Lindeberg [13] is taken from Feller [8, Theorem VIII.4.3, p. 262]. See also Loève [14, 21.B, p. 292]; Hall and Heyde [10, 3.2]. Feller [6, 7] provides a stronger theorem that does not require finite variances.

11.5.1 Lindeberg's CLT *Let $X_1, X_2, \dots$ be independent random variables (not necessarily identically distributed) with $\boldsymbol{E}\, X_i = 0$ and $\boldsymbol{Var}\, X_i = \sigma_i^2 > 0$. Define*

$$S_n = \sum_{i=1}^n X_i.$$

Let

$$s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2 = \boldsymbol{Var}\, S_n,$$

so that s_n is the standard deviation of S_n. Assume that for every $t > 0$, the following ***Lindeberg condition*** *is satisfied:*

$$\frac{1}{s_n^2}\sum_{i=1}^n \boldsymbol{E}\, X_i^2 \mathbf{1}_{(|X_i| \geqslant t s_n)} \to 0 \text{ as } n \to \infty$$

then

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{D}} N(0,1).$$

11.5.2 Remark Lindeberg's Condition seems a bit opaque. One of its implications is that for every $\varepsilon > 0$, for all n large enough, we have

$$\sigma_i < \varepsilon s_n, \qquad i = 1, \dots, n.$$

If all the X_i are bounded in absolute value by the same bound, then Lindeberg's Condition holds if and only if $s_n \to \infty$. [8, p. 264]

> The importance of this result is that for a standardized sum of a large number of *independent* random variables, each of which has a negligible part of the total variance, the distribution is approximately normal.
>
> This is the explanation of the observed fact that many characteristics of a population have a normal distribution. In applied statistics, it is used to justify the assumption that data are normally distributed.

11.6 ⋆ The Berry–Esseen Theorem

One of the nice things about the Weak Law of Large Numbers is that it gave us information on how close we probably were to the mean. The Berry–Esseen Theorem gives information on how close the cdf of the standardized sum is to the standard normal cdf. The statement and a proof may be found in Feller [8, Theorem XVI.5.1., pp. 542–544]. See also Bhattacharya and Ranga Rao [1, Theorem 12.4, p. 104].

11.6.1 Berry–Essen Theorem *Let $X_1, \dots, X_n$ be independent and identically distributed with expectation 0, variance $\sigma^2 > 0$, $\boldsymbol{E}(|X_i^3|) = \rho < \infty$. Let F_n be the cumulative distribution function of $S_n/\sqrt{n}\sigma$. Then for all $x \in \boldsymbol{R}$,*

$$|F_n(x) - \Phi(x)| \leqslant \frac{3\rho}{\sigma^3\sqrt{n}}.$$

11.7 ⋆ The CLT and densities

Under the hypothesis that each X_i has a density and finite third absolute moment, it can be shown that the density of $S_n/\sqrt{n}$ converges uniformly to the standard normal density at a rate of $1/\sqrt{n}$. See Feller [8, Section XVI.2, pp. 533ff].

11.8 ⋆ The "Second Limit Theorem"

Fréchet and Shohat [9] give an elementary proof of a generalization of another useful theorem on convergence in distribution, originally due to Andrey Andreyevich Markov (Андре́й Андре́евич Ма́рков),[1] which he called the Second Limit Theorem. More importantly their proof is in English. The statement here is taken from van der Waerden [19, § 24.F, p. 103]. You can also find this result in Breiman [4, Theorem 8.48, pp. 181–182].

11.8.1 Markov–Fréchet–Shohat Second Limit Theorem *Let F_n be a sequence of cdfs where each F_n has finite moments $\mu_{k,n}$ of all orders $k = 1, 2, \dots$, and assume that for each k, $\lim_{n\to\infty} \mu_{k,n} = \mu_k$, where each μ_k is finite. Then there is a cdf F such that the k^{th} moment of F is μ_k. Moreover, if F is uniquely determined by its moments, then*

$$F_n \xrightarrow{\mathcal{D}} F.$$

[1] Not to be confused with his son, Andrey Andreyevich Markov, Jr. (Андре́й Андре́евич Ма́рков).

An important case is the standard Normal distribution, which is determined by its moments

$$\mu_k = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{(2m)!}{2^m m!} & \text{if } k = 2m. \end{cases}$$

Mann and Whitney [15] used this to derive the asymptotic distribution of their eponymous test statistic [19, p. 277], which we shall discuss in Section 27.6.

11.9 ⋆ Slutsky's Theorem

My colleague Bob Sherman assures me that the next result, which he refers to as Slutsky's Theorem [18], is incredibly useful in his work. This version is taken from Cramér [5, § 20.6, pp. 254–255].

11.9.1 Theorem (Slutsky's Theorem) *Let $X_1, X_2, \ldots,$ be a sequence of random variables with cdfs $F_1, F_2, \ldots,$. Assume that $F_n \xrightarrow[n\to\infty]{\mathcal{D}} F$. Let $Y_1, Y_2, \ldots,$ satisfy $\operatorname{plim} Y_n = c$, where c is a real constant. Then, (slightly abusing notation),*

$$(X_n + Y_n) \xrightarrow{\mathcal{D}} F(x - c), \qquad (X_n Y_n) \xrightarrow{\mathcal{D}} F(x/c)\ (c > 0), \qquad (X_n/Y_n) \xrightarrow{\mathcal{D}} F(cx)\ (c > 0).$$

One of the virtues of this theorem is that you do not need to assume anything about the independence or dependence of the random variables X_n and Y_n. The proof is elementary, and may be found in Cramér [5, § 20.6, pp. 254–255], Jacod and Protter [11, Theorem 18.8, p. 161], or van der Waerden [19, § 24.G, pp.103–104].

Appendix

Stop here.

You are not responsible for what follows.

11.10 ⋆ The characteristic function of a distribution

Some of you may have taken **ACM 95**, in which case you have come across Fourier transforms. If you haven't, you may still find this comment weakly illuminating. You may also want to look at Appendix 4.A.2 in Larsen–Marx [12].

You may know that the exponential function extends nicely to the complex numbers, and that for a real number u,

$$e^{iu} = \cos u + i \sin u.$$

where of course $i = \sqrt{-1}$.

The **characteristic function** $\varphi_X : \boldsymbol{R} \to \mathbb{C}$ of the random variable X is a complex-valued function defined on the real line by

$$\varphi_X(u) = \boldsymbol{E}\, e^{iuX}.$$

This expectation always exists as a complex number (integrate real and imaginary parts separately) since $|e^{iux}| = 1$ for all u, x.

11.10.1 Fact *If X and Y have the same characteristic function, they have the same distribution.*

11.10.2 Fact (Fourier Inversion Formula) *If F has a density f and characteristic function φ, then*

$$\varphi(u) = \int_{-\infty}^{\infty} e^{iux} f(x)\,dx$$

and we have the ***inversion formula***

$$f(x) = \int_{-\infty}^{\infty} e^{-iux} \varphi(u)\,du.$$

11.10.1 Characteristic Function of an Independent Sum

11.10.3 Fact *Let X, Y be independent random variables on $(S, \mathcal{E}, P)$. Then*

$$\begin{aligned}
\varphi_{X+Y}(u) &= \boldsymbol{E}(e^{iu(X+Y)}) \\
&= \boldsymbol{E}(e^{iuX} e^{iuY}) \\
&= \boldsymbol{E}\, e^{iuX}\, \boldsymbol{E}\, e^{iuY} \quad \text{by independence} \\
&= \varphi_X(u)\varphi_Y(u).
\end{aligned}$$

11.10.2 Characteristic Function of a Normal Random Variable

Let X be a standard normal random variable. Then

$$\varphi_X(u) = \boldsymbol{E}\, e^{iuX} = \frac{1}{\sqrt{2\pi}} \int_{\boldsymbol{R}} e^{iux} e^{\frac{1}{2}x^2}\,dx = e^{-\frac{u^2}{2}}.$$

(This is *not* obvious.)

11.11 ⋆ The characteristic function and convergence in distribution

11.11.1 Theorem

$$F_n \xrightarrow{\mathcal{D}} F \iff \varphi_n(u) \to \varphi(u) \quad \text{for all } u.$$

Proof: Breiman [4, 8.30 and 8.31]. ∎

11.12 ⋆ The characteristic function and the CLT

Sketch of proof of the CLT: We consider the case where $\boldsymbol{E}\, X_i = 0$. (Otherwise subtract the mean everywhere.)

Compute the characteristic function φ_n of $\dfrac{X_1 + \cdots + X_n}{\sqrt{n}\sigma}$.

$$\begin{aligned}
\varphi_n(u) &= \boldsymbol{E}\; e^{iu\left\{\frac{X_1+\cdots+X_n}{\sqrt{n}\sigma}\right\}} \\
&= \boldsymbol{E}\left(\Pi_{k=1}^{n} e^{iu\left\{\frac{X_k}{\sqrt{n}\sigma}\right\}}\right) \\
&= \prod_{k=1}^{n} \boldsymbol{E}\, e^{iu\left\{\frac{X_k}{\sqrt{n}\sigma}\right\}} \quad \text{(independence)} \\
&= \left[\boldsymbol{E}\, e^{iu\left\{\frac{X_1}{\sqrt{n}\sigma}\right\}}\right]^n \quad \text{(identical distribution)} \\
&= \left[\boldsymbol{E}\left(1 + iu\frac{X_1}{\sqrt{n}\sigma} + \frac{1}{2}\left(iu\frac{X_1}{\sqrt{n}\sigma}\right)^2 + o\left[\left(\frac{iuX_1}{\sqrt{n}\sigma}\right)^2\right]\right)\right]^n \\
&\quad \text{(Taylor Series as } \boldsymbol{E}\, X_1 = 0\ \boldsymbol{E}\, X_1^2 = \sigma^2) \\
&= \left[1 + 0 - \frac{u^2}{2n}\frac{\sigma^2}{\sigma^2} + o\left(\frac{-u^2}{2n}\right)\right]^n
\end{aligned}$$

Now use the following well-known fact: $\lim_{x\to 0}(1 + ax + o(bx))^{\frac{1}{x}} \to e^a$, so as $n \to \infty$, $\varphi_n(u) \to e^{-\frac{u^2}{2}}$, which is the characteristic function of $N(0, 1)$. ∎

Proof of the well-known fact:

$$\begin{aligned}
\ln\left((1 + ax + o(bx))^{\frac{1}{x}}\right) &= \frac{1}{x}\ln\big(1 + ax + o(bx)\big) \\
&= \tfrac{1}{x}[\ln 1 + ax\ln'(1) + o(x)] \\
&= \tfrac{1}{x}[ax + o(x)] \\
&= a + \tfrac{o(x)}{x} \to a \text{ as } x \to 0.
\end{aligned}$$

Therefore $e^{(1+ax+\gamma(bx))^{\frac{1}{x}}} \to e^a$ as $x \to 0$. ∎

11.13 ⋆ Other Proofs

The Fourier transform approach to the CLT is algebraic and does not give a lot of insight. There are other approaches that have a more probabilistic flavor. Chapter 2 of Ross and Peköz [17] provides a very nice elementary, but long, proof based on constructing a new set random variables (the Chen–Stein construction) with the same distribution as the sequence $\dfrac{X_1 + \cdots + X_n}{\sqrt{n}\sigma}$ in such a way that we can easily compute their distance from a Standard Normal.

Bibliography

[1] R. N. Bhattacharya and R. Ranga Rao. 1986. *Normal approximation and asymptotic expansions*, corrected and enlarged edition ed. Malabar, Florida: Robert E. Krieger Publishing Company. Reprint of the 1976 edition published by John Wiley & Sons. The new edition has an additional chapter and mistakes have been corrected.

[2] P. Billingsley. 1968. *Convergence of probability measures*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

[3] K. C. Border. 1991. Functional analytic tools for expected utility theory. In C. D. Aliprantis, K. C. Border, and W. A. J. Luxemburg, eds., *Positive Operators, Riesz Spaces, and Economics*, number 2 in Studies in Economic Theory, pages 69–88. Berlin: Springer–Verlag. http://www.hss.caltech.edu/~kcb/Courses/Ec181/pdf/Border1991-Dominance.pdf

[4] L. Breiman. 1968. *Probability*. Reading, Massachusetts: Addison Wesley.

[5] H. Cramér. 1946. *Mathematical methods of statistics*. Number 34 in Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. Reprinted 1974.

[6] W. Feller. 1935. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* 40(1):521–559. DOI: 10.1007/BF01218878

[7] ——— . 1937. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. ii. *Mathematische Zeitschrift* 42(1):301–312. DOI: 10.1007/BF01160080

[8] W. Feller. 1971. *An introduction to probability theory and its applications*, 2d. ed., volume 2. New York: Wiley.

[9] M. Fréchet and J. Shohat. 1931. Errata: A proof of the generalized second limit-theorem in the theory of probability. *Transactions of the American Mathematical Society* 33(4):999. DOI: 10.1090/S0002-9947-1931-1500512-4

[10] P. Hall and C. C. Heyde. 1980. *Martingale limit theory and its application*. Probability and Mathematical Statistics. New York: Academic Press.

[11] J. Jacod and P. Protter. 2004. *Probability essentials*, 2d. ed. Berlin and Heidelberg: Springer.

[12] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[13] J. W. Lindeberg. 1922. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* 15(1):211–225. DOI: 10.1007/BF01494395

[14] M. Loève. 1977. *Probability theory*, 4th. ed. Number 1 in Graduate Texts in Mathematics. Berlin: Springer–Verlag.

[15] H. B. Mann and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18(1):50–60. http://www.jstor.org/stable/2236101.pdf

[16] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[17] S. M. Ross and E. A. Peköz. 2007. *A second course in probability*. Boston: ProbabilityBookstore.com.

[18] E. Slutsky. 1925. Ueber stochastische Asymptotem und Grenzwerte. *Metron* 5(3):3.

[19] B. L. van der Waerden. 1969. *Mathematical statistics*. Number 156 in Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. New York, Berlin, and Heidelberg: Springer–Verlag. Translated by Virginia Thompson and Ellen Sherman from *Mathematische Statistik*, published by Springer-Verlag in 1965, as volume 87 in the series Grundlerhen der mathematischen Wissenschaften.