36710-36752 Advanced Probability Overview — Fall 2020

8. Characteristic Functions and CLT

Instructor: Alessandro Rinaldo

Associated reading: Sec 7.1, 7.3 of Ash and Doléans-Dade; Sec 3.3, 3.4 of Durrett.

Overview

In this set of lecture notes we present the Central Limit Theorem. There are many different ways to prove the CLT. We will follow the common approach using characteristic functions. Characteristic functions are essentially Fourier transformations of distribution functions, which provide a general and powerful tool to analyze probability distributions.

1 Characteristic Functions

Recall that in order to check convergence in distribution for a sequence of random quantities X_n, we need to show convergence of $\mathrm{E} f(X_n)$ for all bounded continuous function f. We have shown that when $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^1, \mathcal{B}^1)$, it suffices to check convergence of $F_n(x)$ for all continuity points x of F. For the case in which $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^p, \mathcal{B}^p)$, there is a useful technique for determining if a sequence of random vectors converges in distribution. It is based on a characterization of distributions by something simpler than the means of all bounded continuous functions. The means of a special collection of bounded continuous functions, namely $\{\exp(it^\top x) : t \in \mathbb{R}^p\}$, are enough to characterize a distribution. From here on in the notes, i is one of the complex square-roots of -1.

Definition 1 (Characteristic Function). *The function $\phi_X(t) = \mathrm{E}\exp(it^\top X)$ is called the* characteristic function *(cf) of X.*

(Mathematicians will recognize the cf as the Fourier transform of f_X, the density function of X.) Every distribution on $\mathbb{R}^p$ has a cf regardless of whether moments exist. Recall from complex analysis that $\exp(iu) = \cos(u) + i\sin(u)$. So, we see that $\exp(it^\top x)$ is indeed bounded as a function of x for each t.

Example 2 (Normal distribution). *Let $f_X(x) = \exp(-x^2/2)/\sqrt{2\pi}$ be the density of X.*

Then

$$\begin{aligned}\phi_X(t) &= \frac{1}{\sqrt{2\pi}}\int \exp(itx - x^2/2)dx \\ &= \frac{1}{\sqrt{2\pi}}\int \exp\left(-\frac{1}{2}[x-it]^2 - \frac{t^2}{2}\right)dx \\ &= \exp(-t^2/2).\end{aligned}$$

Example 3 (Uniform distribution). *Let $f(x) = 1/2$ for $-1 < x < 1$. Then*

$$\phi(t) = \frac{1}{2}\int_{-1}^{1}\exp(itx)dx = \frac{\exp(it)-\exp(-it)}{2it} = \frac{\sin(t)}{t}.$$

Example 4 (Cauchy distribution). *Let $f_X(x) = [\pi(1+x^2)]^{-1}$. Then $\phi_X(t) = \exp(-|t|)$. To prove this requires contour integration.*

Remark 5 (Continuity). *Of course all cf's are continuous by the dominated convergence theorem. Since $|\exp(it^\top x) - \exp(iu^\top x)| \le 2$ for all t, u, x, we can pass the limit as $u \to t$ under the integral in $\int[\exp(it^\top x) - \exp(iu^\top x)]d\mu_X(x)$ to get 0 for the limit.*

Remark 6 (Smoothness). *The smoothness of the cf is related to the existence of moments. Now, suppose that X is a random variable with finite mean. We can write*

$$|\exp(ix)-1|^2 = |\cos(x) + i\sin(x) - 1|^2 = 2 - 2\cos(x) = 2\int_0^x \sin(t)dt \le 2\int_0^x tdt = x^2.$$

This implies that $|\exp(ix)-1| \le |x|$ for all x. Clearly, $|\exp(ix)-1| \le 2$ for all x also. So

$$|\exp(ix)-1| \le \min\{2, |x|\}. \tag{1}$$

This implies that $[\exp(ixt)-1]/t$ is bounded by a μ_X-integrable function $|x|$. By the dominated convergence theorem, we can pass the limit as $t \to 0$ under the integral to get that $\phi'(0)$ exists and equals $i\mathrm{E}(X)$. With a bit more effort similar results hold if higher moments exist: $\phi^{(k)}(0) = i^k\mathrm{E}X^k$.

Some basic properties of cf's are summarized below.

Proposition 7 (Basic properties of cf). *All cf's have the following properties:*

1. $\phi(0) = 1$, $|\phi(t)| \le 1$,
2. $\phi(-t) = \overline{\phi(t)}$ *(complex conjugate),*
3. $|\phi(t+h) - \phi(t)| \le \mathrm{E}|e^{ihX} - 1|$ *(uniform continuity),*

4. $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$.

The next result gives a sufficient condition for $\phi(t)$ to be a cf.

Theorem 8 (Polya's Criterion). *Let ϕ be continuous, real, nonnegative, symmetric, decreasing and convex on $[0,\infty)$, such that $\phi(0) = 1$, $\lim_{t\to\infty}\phi(t) = 0$, then ϕ is a characteristic function.*

Proposition 9 (Cf of sum of independent r.v.'s). *If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.*

The remaining theorems about convergence in distribution are

- the inversion/uniqueness theorem that says that each cf corresponds to a unique distribution,
- the continuity theorem that says that $X_n \xrightarrow{\mathcal{D}} X$ if and only if $\phi_{X_n}(t) \to \phi_X(t)$ for all t (the "only if" direction being trivial), and
- the central limit theorem that says that certain normalized sums of independent (not necessarily identically distributed) random variables with finite variance converge in distribution to a standard normal distribution.

1.1 Inversion formula and uniqueness

Theorem 10 (Inversion and uniqueness). *Let ϕ be the cf for the probability P on $(\mathbb{R}^p, \mathcal{B}^p)$. Let A be a rectangular region of the form*

$$A = \{(x_1,\ldots,x_p) : a_j \le x_j \le b_j \text{ for all } j\},$$

where $a_j < b_j$ for all j and $P(\partial A) = 0$. For each $T > 0$, let

$$B_T = \{(t_1,\ldots,t_p) : -T \le t_j \le T \text{ for all } j\}.$$

Then

$$P(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^p} \int_{B_T} \prod_{j=1}^{p} \left[\frac{\exp(-it_ja_j) - \exp(-it_jb_j)}{it_j}\right] \phi(t) dt_1 \cdots dt_p.$$

Distinct probability measures have distinct cf's.

The proof relies on the following interesting result which we state without proof. The proof is outlined in Exercise 1.7.5 of Durrett.

Lemma 11.

$$\lim_{T\to\infty}\int_{-T}^{T}\frac{\sin(ct)}{t}dt = \begin{cases} \pi & \textit{if } c>0, \\ 0 & \textit{if } c=0, \\ -\pi & \textit{if } c<0. \end{cases} \tag{2}$$

Because dt/t is invariant measure with respect to scale changes on $(0,\infty)$, the integral doesn't depend on $|c|$ for $c \neq 0$.

Sketch of Proof, Theorem 10. Basically, replace $\phi(t)$ by $\int \prod_{j=1}^{p} \exp(it_j x_j) dP(x)$, change the order of integration, pass the limit inside the integral over x, combine the two products into one, rewrite $\exp(-it_j c_j)$ in terms of sines and cosines (for $c_j \in \{x_j - a_j, x_j - b_j\}$), notice that the cosine terms integrate to 0 over t_j, and apply the above formula to the sine terms. When x_j is between a_j and b_j, the limit of the integral over t_j yields $\pi - (-\pi) = 2\pi$. When x_j is outside of $[a_j, b_j]$, the limit yields either $\pi - \pi$ or $-\pi - (-\pi)$, both 0.

Proof: [Proof of Theorem 10] Apply Fubini's theorem to write

$$\begin{aligned} &\int_{B_T}\prod_{j=1}^{p}\left[\frac{\exp(-it_ja_j)-\exp(-it_jb_j)}{it_j}\right]\phi(t)dt_1\cdots dt_p \qquad (3)\\ &= \int_{\mathbb{R}^p}\int_{B_T}\prod_{j=1}^{p}\left[\frac{\exp(it_j[x_j-a_j])-\exp(it_j[x_j-b_j])}{it_j}\right]dt_1\cdots dt_j d\mu(x). \end{aligned}$$

We can do this beacuse the integrand is bounded by $\prod_{j=1}^{p}|b_j - a_j|$ according to Equation (1) and the set over which we are integrating has finite product measure. Rewrite the jth factor in the integrand on the right-side of (3) as

$$\frac{\cos(t_j[x_j-a_j])-\cos(t_j[x_j-b_j])+i\sin(t_j[x_j-a_j])-i\sin(t_j[x_j-b_j])}{it_j}.$$

Since the integration over t_j is from $-T$ to T and $\{\cos(t_j[x_j - a_j]) - \cos(t_j[x_j - b_j])\}/t_j$ is bounded and an odd function, its integral is 0. We rewrite the right side of (3) as

$$\int_{\mathbb{R}^p}\int_{B_T}\prod_{j=1}^{p}\left[\frac{\sin(t_j[x_j-a_j])}{t_j}-\frac{\sin(t_j[x_j-b_j])}{t_j}\right]dt_1\cdots dt_j d\mu(x). \tag{4}$$

Define

$$\begin{aligned} g_T(x) &= \int_{B_T}\prod_{j=1}^{p}\left[\frac{\sin(t_j[x_j-a_j])}{t_j}-\frac{\sin(t_j[x_j-b_j])}{t_j}\right]dt_1\cdots dt_p \\ &= \prod_{j=1}^{p}\int_{-T}^{T}\frac{\sin(t_j[x_j-a_j])}{t_j}-\frac{\sin(t_j[x_j-b_j])}{t_j}dt_j. \end{aligned}$$

This function is uniformly bounded for all T and x because Lemma 11 implies that

$$\sup_{T,c}\left|\int_{-T}^{T}\frac{\sin(ct)}{t}dt\right| < \infty\,.$$

Hence by DCT the limit as $T \to \infty$ of the integral in Equation (4) equals $\int \lim_{T\to\infty} g_T(x)d\mu(x)$. If we define

$$\psi_{a,b}(x) = \begin{cases} 0 & \text{if } x < a, \\ \pi & \text{if } x = a, \\ 2\pi & \text{if } a < x < b, \\ \pi & \text{if } x = b, \\ 0 & \text{if } x > b, \end{cases}$$

then Lemma 11 says that $\lim_{T\to\infty} g_T(x) = \prod_{j=1}^{p} \psi_{a_j,b_j}(x_j)$, which equals $(2\pi)^p$ for $x \in \text{int}(A)$ and equals 0 for $x \in \overline{A}^C$. Since $\mu(\partial A) = 0$, we have

$$\frac{1}{(2\pi)^p}\int_{\mathbb{R}^p}\lim_{T\to\infty} g_T(x)d\mu(x) = \mu(A)\,.$$

At most countably many hyperplanes perpendicular to the coordinate axes can have positive μ probability. So, the rectangular regions A with $\mu(\partial A) = 0$ form a π-system that generate $\mathcal{B}^p$. It follows from the inversion formula that $\phi_1 = \phi_2$ implies $\mu_1 = \mu_2$. That is, the characteristic function determines the distribution. ■

The following theorem allows us to simplify some future proofs by doing only the $p = 1$ case.

Lemma 12 (Cramér-Wold). *Let X and Y be p-dimensional random vectors. Then X and Y have the same distribution if and only if $\alpha^\top X$ and $\alpha^\top Y$ have the same distribution for every $\alpha \in \mathbb{R}^p$.*

Proof: We know that X and Y have the same distribution if and only if $\phi_X(t) = \phi_Y(t)$ for every $t \in \mathbb{R}^p$. This is true if and only if $\phi_X(s\alpha) = \phi_Y(s\alpha)$ for all $\alpha \in \mathbb{R}^p$ and all $s \in \mathbb{R}$. But $\phi_X(s\alpha)$ is the cf of $\alpha^\top X$ (as a function of s) and $\phi_Y(s\alpha)$ is the cf of $\alpha^\top Y$. So, $\phi_X(s\alpha) = \phi_Y(s\alpha)$ for all $\alpha \in \mathbb{R}^p$ and all $s \in \mathbb{R}$ if and only if $\alpha^\top X$ and $\alpha^\top Y$ have the same distribution for every $\alpha \in \mathbb{R}^p$. ■

If the characteristic function is integrable, a continuous density exists. We will not prove this result.

Proposition 13. *If ϕ is the cf of the cdf F on $(\mathbb{R}, \mathcal{B}^1)$ and if ϕ is integrable, then F has a density*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx)\phi(t)dt, \tag{5}$$

which is continuous.

1.2 The continuity theorem

The connection between characteristic functions and convergence in distribution is the following.

Theorem 14 (Continuity theorem). *Let $\{P_n\}_{n=1}^{\infty}$ be a sequence of probabilities on $(\mathbb{R}^p, \mathcal{B}^p)$, and let P be another probability. Let ϕ_n be the cf for P_n, and let ϕ be the cf for P. Then $P_n \stackrel{\mathcal{D}}{\to} P$ if and only if $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ for all $t \in \mathbb{R}^p$.*

Proof: The "only if" direction follows directly from the definition of convergence in distribution since $\exp(itx)$ is a bounded continuous function of x for all t. For the "if" direction, start with $p = 1$, and construct the following bound

$$\begin{aligned}
\frac{1}{u}\int_{-u}^{u}[1-\phi_n(t)]dt &= \int_{-\infty}^{\infty}\frac{1}{u}\int_{-u}^{u}[1-\exp(itx)]dtdP_n(x) \\
&= 2\int_{-\infty}^{\infty}\left(1-\frac{\sin(ux)}{ux}\right)dP_n(x) \\
&\geq 2\int_{\{x:|x|\geq 2/u\}}\left(1-\frac{1}{|ux|}\right)dP_n(x) \\
&\geq P_n(\{x:|x|\geq 2/u\}),
\end{aligned}$$

where the first equality is justfied by Fubini's theorem. Let $\epsilon > 0$. Since $\phi(0) = 1$ and ϕ is continuous, there exists u such that $\int_{-u}^{u}[1-\phi(t)]dt/u < \epsilon$. Since ϕ_n converges to ϕ, the dominated convergence theorem implies that $\int_{-u}^{u}[1-\phi_n(t)]dt/u < 2\epsilon$ for sufficiently large n, say for $n > N$. Let $a \geq 2/u$ be such that $P_n(\{x : |x| > a\}) < 2\epsilon$ for $n = 1, \ldots, N$. Then $P_n(\{x : |x| > a\}) < 2\epsilon$ for all n and the sequence $\{P_n\}_{n=1}^{\infty}$ is tight. For $p > 1$, apply this same reasoning to each coordinate distribution to piece together the necessary compact set to show that the sequence of distributions is tight. By Helly-Bray theorem, there exists a subsequence $\{P_{n_k}\}_{k=1}^{\infty}$ that converges in distribution. By the "only if" part of the theorem, the cf's for this subsequence converge to ϕ. The only distribution with cf ϕ is P (Theorem 10), hence $P_{n_k} \stackrel{\mathcal{D}}{\to} P$. Since every convergent subsequence converges to P, the last claim in Helly-Bray says that, $P_n \stackrel{\mathcal{D}}{\to} P$. ∎

Example 15. *For each j, let Y_j have a uniform distribution on the interval $[-1, 1]$ and let $X_n = \sqrt{\frac{3}{n}}\sum_{j=1}^{n} Y_j$. Then the cf of X_n is*

$$\phi_n(t) = \left(\frac{\sin\left(t\sqrt{3/n}\right)}{t\sqrt{3/n}}\right)^n.$$

We can write $\sin(t) = t - t^3/6 + o(t^3)$ so that, for each t,

$$\frac{\sin\left(t\sqrt{3/n}\right)}{t\sqrt{3/n}} = 1 - \frac{t^2}{2n} + o(1/n),$$

as $n \to \infty$. It follows easily that $\lim_{n\to\infty} \phi_n(t) = \exp(-t^2/2)$. This is the cf of the standard normal distribution.

The following two results are useful in proving convergence in distribution.

Corollary 16. *If $\lim_{n\to\infty} \phi_n(t)$ exists for all t and is continuous at 0, then the limit is a cf, and the distributions converge to the distribution with that cf.*

The continuity at 0 was all that was needed to establish that the sequence of distributions was tight. Another corollary (thanks to Cramér and Wold) is the following

Corollary 17. *If $\{X_n\}_{n=1}^{\infty}$ is a sequence of p-dimensional random vectors and X is a random vector, then $X_n \stackrel{\mathcal{D}}{\to} X$ if and only if $\alpha^\top X_n \stackrel{\mathcal{D}}{\to} \alpha^\top X$ for all $\alpha \in I\!R^p$.*

2 Central Limit Theorem

Theorem 18 (Lindeberg-Feller central limit theorem). *Let $\{r_n\}_{n=1}^{\infty}$ be a sequence of integers. For each $n = 1, 2, \ldots$, let $X_{n,1}, \ldots, X_{n,r_n}$ be independent random variables with $X_{n,k}$ having mean 0 and finite nonzero variance $\sigma_{n,k}^2$. Define $\sigma_n^2 = \sum_{k=1}^{r_n} \sigma_{n,k}^2$ and $S_n = \sum_{k=1}^{r_n} X_{n,k}$. Assume that, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{k=1}^{r_n} \mathrm{E}\left[X_{n,k}^2 \mathbf{1}(|X_{n,k}| \geq \epsilon\sigma_n)\right] = 0. \tag{6}$$

Then S_n/σ_n converges in distribution to the standard normal distribution.

The proof of Theorem 18 works by applying the continuity theorem 14. We must show that the cf of S_n/σ_n converges to $\exp(-t^2/2)$ for all t. The proof has two (lengthy) steps. One is to approximate the cf $\phi_{n,k}$ of each $X_{n,k}/\sigma_n$ by $1 - t^2\sigma_{n,k}^2/(2\sigma_n^2)$. The other is to approximate $\exp(-t^2/2)$ by $\prod_{k=1}^{r_n}[1 - t^2\sigma_{n,k}^2/(2\sigma_n^2)]$.
Also, notice that $X_{n,k}$ is divided by σ_n in all formulas in the statement of the theorem. Hence, without loss of generality, we can assume that $\sigma_n = 1$ for all n. We do this in the proof, given at the end of this set of lecture notes.

Example 19 (iid CLT). *If $X_1, X_2, \ldots,$ are iid with mean 0 and variance σ^2, then let $r_n = n$ and $X_{n,k} = X_k$ for all n and all $k \leq n$. Then $\sigma_n^2 = n\sigma^2$ and*

$$\frac{1}{\sigma_n^2} \sum_{k=1}^{n} \mathrm{E}\left[X_{n,k}^2 \mathbf{1}(|X_{n,k}| \geq \epsilon\sigma_n)\right] = \frac{1}{\sigma^2}\mathrm{E}\left[X_1^2 \mathbf{1}(|X_1| \geq \epsilon\sqrt{n}\sigma)\right] \to 0,$$

by DCT.

Example 20 (Lyapounov CLT). *Instead of assuming Equation* (6)*, we assume that there exists* $\delta > 0$ *such that* $\mathrm{E}[|X_{n,k}|^{2+\delta}] < \infty$ *and that*

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{\sigma_n^{2+\delta}} \mathrm{E}\left[|X_{n,k}|^{2+\delta}\right] = 0. \tag{7}$$

Since $|X_{n,k}|^2 \le |X_{n,k}|^{2+\delta}/[\epsilon^\delta \sigma_n^\delta]$ *when* $|X_{n,k}| > \epsilon\sigma_n$*, we have that the sum in Equation* (6) *is bounded by*

$$\frac{1}{\sigma_n^2} \sum_{k=1}^{r_n} \mathrm{E}\left[X_{n,k}^2 \mathbf{1}(|X_{n,k}| > \epsilon\sigma_n)\right] \le \frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{\sigma_n^{2+\delta}} \mathrm{E}\left[|X_{n,k}|^{2+\delta}\right].$$

Hence, if Equation (7) *holds, so does Equation* (6)*.*

Example 21. *Let* $Y_1, Y_2, \ldots$ *be independent Poisson random variables with the parameter of* Y_k *being* $1/k$*. Then let* $X_{n,k} = Y_k - 1/k$ *for all* n *and all* $k \le n$*. Now,* $\sigma_n^2 = L_n = \sum_{k=1}^n 1/k$*. For* $\delta = 1$*,* $\mathrm{E}(X_{n,k}^3) = 1/k$ *also. Hence*

$$\mathrm{E}|X_{n,k}|^3 \le \mathrm{E}\left(\left[X_{n,k} + \frac{1}{k}\right]^3\right) = \frac{1}{k} + \frac{3}{k^2} + \frac{1}{k^3} \le \frac{5}{k}.$$

The sum on the left of Equation (7) *is bounded by* $5/\sqrt{L_n}$*, which goes to 0. So,* $[\sum_{k=1}^n Y_k - L_n]/\sqrt{L_n}$ *converges in distribution to standard normal. Notice that* $L_n = \log(n) + c_n$ *where* c_n *is bounded. By Theorem 18,* $[\sum_{k=1}^n Y_k - \log(n)]/\sqrt{\log(n)}$ *converges in distribution to standard normal also.*

Proposition 22. *If the* $X_{n,k}$ *are uniformly bounded and if* $\lim_{n\to\infty} \sigma_n^2 = \infty$*, then Equation* (6) *will hold.*

Example 23 (Bernoulli distribution). *If* $X_{n,k}$ *has a Bernoulli distribution with parameter* $1/k$ *and* $r_n = n$*, the condition holds. The theorem does not apply, however, if the Bernoulli parameter is* $1/k^2$*. Indeed, if the Bernoulli parameter is* $1/k^2$*,* $\sum_{k=1}^n X_{n,k}$ *converges almost surely according to the basic* L^2 *convergence theorem. As another example, if* $r_n = n$ *and the Bernoulli parameter is* $k/(n+1)$ *for* $k = 1, \ldots, n$*, then* $\sigma_n^2 = n(n+2)/[6(n+1)]$*. In fact,* r_n *could be as small as* $n^{1/2+\epsilon}$ *for* $0 < \epsilon \le 1/2$*, and the theorem would still apply. This example cannot be described as a single sequence as all of the distributions of* $X_{n,k}$ *change as* n *changes.*

Example 24 (Delta method). *Suppose that* $Y_1, Y_2, \ldots$ *are iid with common mean* η *and common variance* σ^2*. Let* $X_n = \frac{1}{n}\sum_{j=1}^n Y_j$*. Then* $\sqrt{n}(X_n - \eta) \xrightarrow{\mathcal{D}} Z$*, where* Z *has a normal distribution with mean 0 and variance* σ^2*. If* g *is a function with derivative* g' *at* η*, then* $\sqrt{n}[g(X_n) - g(\eta)]$ *converges in distribution to a normal distribution with mean 0 and variance* $[g'(\eta)]^2\sigma^2$*.*

A multivariate central limit theorem exists for iid sequences, and the proof combines the univariate central limit theorem together with the method of the Cramér-Wold lemma 12 and the Continuity theorem 14.

Theorem 25 (Multivariate Central Limit Theorem). *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of iid random vectors with common mean vector η and common covariance matrix Σ. Let $\overline{X}_n$ be the average of the first n of these vectors. Then $Z_n = \sqrt{n}(\overline{X}_n - \eta)$ converges in distribution to multivariate normal with zero mean vector and covariance matrix Σ.*

Proof: By Corollary 17, all we need to show is that, for all α, $\alpha^\top Z_n \xrightarrow{\mathcal{D}} N(0, \alpha^\top \Sigma\alpha)$. For every vector α, let $Y_k = \alpha^\top X_k$ which are iid with common mean $\alpha^\top\eta$ and common variance $\alpha^\top\Sigma\alpha$. Let $\sigma_n^2 = n\alpha^\top\Sigma\alpha$. If $\alpha^\top\Sigma\alpha = 0$, then $\Pr(Y_k = \alpha^\top\eta) = 1$ and $\Pr(\alpha^\top Z_n = 0) = 1$ for all n, which means that $\alpha^\top Z_n \xrightarrow{\mathcal{D}} N(0, \alpha^\top\Sigma\alpha)$. For the rest of the proof, assume that $\alpha^\top\Sigma\alpha > 0$. Theorem 18 says that

$$\frac{n\alpha^\top\overline{X}_n - n\alpha^\top\eta}{\sigma_n} = \frac{\alpha^\top Z_n}{\sqrt{\alpha^\top\Sigma\alpha}} \xrightarrow{\mathcal{D}} N(0,1).$$

Multiply by $\sqrt{\alpha^\top\Sigma\alpha}$ to get that $\alpha^\top Z_n \xrightarrow{\mathcal{D}} N(0, \alpha^\top\Sigma\alpha)$. ■

A multivariate central limit theorem also exists for general independent sequences, but it is very cumbersome to state. (Imagine replacing all of the σ^2's and σ_n^2's in Theorem 18 by matrices.)

Proof of Lindeberg-Feller CLT

To prove the central limit theorem, we will need to be able to approximate arbitrary characteristic functions. First, by various integrations by parts and reasoning similar to that which achieved Equation (1), we can obtain the following bound.

Lemma 26.

$$\left|\exp(ix) - \left[1 + ix - \frac{x^2}{2}\right]\right| \le \min\left\{|x|^3, x^2\right\}.$$

In terms of the cf of a random variable X with mean 0 and variance σ^2, this equation says that

$$\left|\phi_X(t) - \left[1 - \frac{t^2\sigma^2}{2}\right]\right| \le \mathrm{E}\left[\min\{|Xt|^3, (Xt)^2\}\right]. \tag{8}$$

Notice that only a second moment is required in order for the mean on the far right to exist. In order to apply a bound like this to a sum like S_n, we need to approximate a product of cf's by a product of approximations. The following simple results are useful. Their proofs are contained in another course document.

Proposition 27. *Let $z_1, \ldots, z_m$ and $w_1, \ldots, w_m$ be complex numbers with modulus at most 1. Then*

$$\left|\prod_{k=1}^{m} z_k - \prod_{k=1}^{m} w_k\right| \le \sum_{k=1}^{m} |z_k - w_k|$$

Proof: We shall use induction. The result is trivially true when $m = 1$. Assume that it is true for $m = m_0$. For $m = m_0 + 1$, we have

$$\begin{aligned}
\left|\prod_{k=1}^{m_0+1} z_k - \prod_{k=1}^{m_0+1} w_k\right| &= \left|\prod_{k=1}^{m_0+1} z_k - w_{m_0+1}\prod_{k=1}^{m} z_k + w_{m_0+1}\prod_{k=1}^{m} z_k - \prod_{k=1}^{m_0+1} w_k\right| \\
&\le \left|\prod_{k=1}^{m} z_k\right| |z_{m_0+1} - w_{m_0+1}| + \left|\prod_{k=1}^{m} z_k - \prod_{k=1}^{m} w_k\right| |w_{m_0+1}| \\
&\le \sum_{k=1}^{m} |z_k - w_k| + |z_{m_0+1} - w_{m_0+1}|.
\end{aligned}$$

■

Proposition 28. *For complex z, $|\exp(z) - 1 - z| \le |z|^2 \exp(|z|)$.*

Proof: Write $\exp(z) - 1 - z = \sum_{k=2}^{\infty} z^k/k!$. Since $k! < (k+2)!$ for $k \ge 0$, we have

$$\left|\sum_{k=2}^{\infty} \frac{z^k}{k!}\right| \le |z|^2 \sum_{k=0}^{\infty} \frac{|z|^k}{(k+2)!} \le |z|^2 \exp(|z|).$$

■

Proof: [Proof of Theorem 18] Without loss of generality, we assume that $\sigma_n = 1$ for all n. The cf of S_n is

$$\phi_n(t) = \prod_{k=1}^{r_n} \phi_{n,k}(t).$$

According to Equation (8), for each n, k, and t,

$$\begin{aligned}
\left|\phi_{n,k}(t) - \left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| &\le \mathrm{E}\left[\min\{|X_{n,k}t|^3, (X_{n,k}t)^2\}\right] \\
&\le \mathrm{E}\left[|tX_{n,k}|^3 \mathbf{1}(|X_{n,k}| < \epsilon)\right] + \mathrm{E}\left[|tX_{n,k}|^2 \mathbf{1}(|X_{n,k}| \ge \epsilon)\right] \\
&\le \epsilon|t|^3\sigma_{n,k}^2 + t^2\mathrm{E}\left[|X_{n,k}|^2 \mathbf{1}(|X_{n,k}| \ge \epsilon)\right].
\end{aligned}$$

It follows that

$$\sum_{k=1}^{r_n} \left|\phi_{n,k}(t) - \left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| \le \epsilon|t|^3 + t^2 \sum_{k=1}^{r_n} \mathrm{E}\left[|X_{n,k}|^2 \mathbf{1}(|X_{n,k}| \ge \epsilon)\right].$$

The last sum goes to 0 as $n \to \infty$ according to Equation (6). Since ϵ is arbitrary, we have

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \left| \phi_{n,k}(t) - \left[1 - \frac{t^2 \sigma_{n,k}^2}{2} \right] \right| = 0. \tag{9}$$

In order to apply Proposition 27, we need $\sigma_{n,k}^2$ to all be small. For each $\epsilon > 0$, we have

$$\begin{aligned} \sigma_{n,k}^2 &= \mathrm{E}\left[|X_{n,k}|^2 \mathbf{1}(|X_{n,k}| \le \epsilon) \right] + \mathrm{E}\left[|X_{n,k}|^2 \mathbf{1}(|X_{n,k}| > \epsilon) \right] \\ &\le \epsilon^2 + \mathrm{E}\left[|X_{n,k}|^2 \mathbf{1}(|X_{n,k}| > \epsilon) \right]. \end{aligned}$$

It follows from Equation (6) that

$$\lim_{n\to\infty} \max_k \sigma_{n,k}^2 = 0. \tag{10}$$

Next, fix $t \neq 0$ and notice that for n sufficiently large $0 < t^2 \sigma_{n,k}^2 / 2 < 1$ for all k simultaneously. It follows from Proposition 27 and Equation (9) that

$$\lim_{n\to\infty} \left| \phi_n(t) - \prod_{k=1}^{r_n} \left[1 - \frac{t^2 \sigma_{n,k}^2}{2} \right] \right| = 0. \tag{11}$$

Since $\sigma_n^2 = 1$, we have that $\exp(-t^2/2) = \prod_{k=1}^{r_n} \exp(-t^2 \sigma_{n,k}^2 / 2)$. For n large enough so that $t^2 \sigma_{n,k}^2 / 2 < 1$ for all k write

$$\begin{aligned} \left| \exp\left(-\frac{t^2}{2} \right) - \prod_{k=1}^{r_n} \left[1 - \frac{t^2 \sigma_{n,k}^2}{2} \right] \right| &\le \sum_{k=1}^{r_n} \left| \exp\left(-\frac{t^2 \sigma_{n,k}^2}{2} \right) - 1 + \frac{t^2 \sigma_{n,k}^2}{2} \right| \\ &\le \frac{t^4}{4} \sum_{k=1}^{r_n} \sigma_{n,k}^4 \exp\left(\frac{t^2}{2} \right) \\ &\le \frac{t^4}{4} \max_k \sigma_{n,k}^2 \exp\left(\frac{t^2}{2} \right), \end{aligned} \tag{12}$$

where the first inequality follows from Proposition 27, the second follows from Proposition 28, and the third follows from the fact that $\sigma_n^2 = 1$. Finally, the last term in Section 2 goes to 0 according to Equation (10). Combining this with Equation (11) says that $\lim_{n\to\infty} \phi_n(t) = \exp(-t^2/2)$. ■