1 Lecture 12 — Central Limit Theorem; Characteristic Functions

The previous lecture closed the strong law of large numbers. We now turn to fluctuations: properly normalised, sums $S_n = X_1 + \cdots + X_n$ of iid mean-zero random vectors converge in distribution to a Gaussian. The proof rests on three pillars assembled here — Prohorov's tightness theorem, the characteristic function and its uniqueness, and Lévy's continuity lemma — from which the central limit theorem (CLT) drops out by a Taylor expansion.

1.1 Tightness and Prohorov's theorem

Definition 1.1: Uniform tightness

A collection $\{\mu_i\}_{i\in I}$ of probability measures on a metric space is *uniformly tight* if, for every $\varepsilon > 0$, there exists a compact set K_ε such that

$$\mu_i(K_\varepsilon) \; > \; 1-\varepsilon \qquad \text{for all } i \in I.$$

Prohorov's theorem is the probabilistic analogue of Bolzano–Weierstrass: it upgrades tightness (a uniform mass-control condition) to sequential weak compactness.

Theorem 1.2: Prohorov

Let $\{\mu_i\}_{i=1}^{\infty}$ be a sequence of probability measures on a metric space. If the sequence is uniformly tight, then it is relatively (sequentially) compact for weak convergence: every subsequence μ_{i_k} has a further subsubsequence $\mu_{i_{k_r}} \Rightarrow \mu$ for some probability measure μ (possibly depending on the subsequence).

The next proposition is the standard subsubsequence trick: if every subsequence has a further subsubsequence with the same weak limit, then the whole sequence converges to that limit.

Proposition 1.3: Subsubsequence criterion for weak convergence

Let $\{\mu_i\}_{i=1}^{\infty}$ and μ be probability measures. Suppose that for every subsequence μ_{i_k} there exists a further subsubsequence $\mu_{i_{k_r}} \Rightarrow \mu$. Then $\mu_i \Rightarrow \mu$.

1.2 Gaussian measures

Definition 1.4: Gaussian measure on $\mathbb{R}$

A Borel measure γ on $(\mathbb{R}, \mathcal{B})$ is *Gaussian* with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$ if

$$\gamma((a,b]) \; = \; \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\Big[-\frac{1}{2\sigma^2}(x-m)^2\Big] \, d\lambda(x).$$

For $\sigma = 0$ we set $\gamma = \delta_m$ (Dirac mass at m) and call γ a *degenerate* Gaussian measure.

Definition 1.5: Gaussian measure on $\mathbb{R}^d$

A Borel measure γ on $(\mathbb{R}^d, \mathcal{B})$ is *Gaussian* if for every linear functional $f\colon \mathbb{R}^d \to \mathbb{R}$ the induced measure $\gamma \circ f^{-1}$ on $(\mathbb{R}, \mathcal{B})$ is Gaussian. Equivalently, every linear combination of the coordinates is one-dimensional Gaussian.

Definition 1.6: Gaussian random variable

A random variable Z from a probability space $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}^d, \mathcal{B})$ is *Gaussian* if its law $\gamma := \mu \circ Z^{-1}$ is a Gaussian measure on $(\mathbb{R}^d, \mathcal{B})$.

Remark 1.1. For vectors $u, v \in \mathbb{R}^d$ we use the Euclidean inner product $\langle u, v\rangle = \sum_{i=1}^d u_i v_i$ and write $|u|^2 = \langle u, u\rangle$. A collection $\{X_i\}_{i=1}^\infty$ is *iid* if the X_i are pairwise independent and share a common law ("random variables induce measures").

1.3 Characteristic functions

The characteristic function is the Fourier transform of a probability measure; it linearises convolution and, by uniqueness below, encodes the measure completely.

Definition 1.7: Characteristic function

For a probability measure μ on $(\mathbb{R}^d, \mathcal{B})$, the *characteristic function* $\tilde{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ is

$$\tilde{\mu}(t) \;:=\; \int \exp\{\, i\langle x, t\rangle \,\}\, d\mu(x).$$

When $\tilde{\mu}$ is integrable against Lebesgue measure on $\mathbb{R}^d$, the inverse transform recovers a density:

$$p(x) \;=\; (2\pi)^{-d} \int \tilde{\mu}(t)\, \exp\{\, -i\langle x, t\rangle \,\}\, d\lambda(t), \qquad \lambda\text{-a.e.},$$

with p the probability density function of μ.

Definition 1.8: Convolution of measures

For two measures μ, ν on $(\mathbb{R}^d, \mathcal{B})$, the *convolution* $\mu * \nu$ is the measure

$$(\mu * \nu)(B) \;:=\; \int \nu(B - x)\, d\mu(x), \qquad B \in \mathcal{B},$$

where $B - x = \{y \in \mathbb{R}^d : y + x \in B\}$. The operation $*$ is associative and commutative; the characteristic function of $\mu * \nu$ is $\tilde{\mu}\,\tilde{\nu}$; and if X, Y are independent with laws μ, ν, then $X + Y$ has law $\mu * \nu$.

Theorem 1.9: Uniqueness of characteristic functions

Let μ and ν be probability measures on $(\mathbb{R}^d, \mathcal{B})$. If $\tilde{\mu} = \tilde{\nu}$, then $\mu = \nu$.

Remark 1.2. The proof goes via Gaussian smoothing. Let γ_σ be the mean-zero Gaussian on $\mathbb{R}^d$ with covariance $\sigma^2 I$ and put $\mu^{(\sigma)} := \mu * \gamma_\sigma$, $\nu^{(\sigma)} := \nu * \gamma_\sigma$. The smoothed measures admit explicit densities

$$q^{(\sigma)}(x) = (2\pi)^{-d} \int \tilde{\nu}(t) \exp\left[-i\langle x, t\rangle - \tfrac{1}{2}\sigma^2 |t|^2\right] d\lambda(t),$$

and similarly for $p^{(\sigma)}$ with $\tilde{\mu}$. Hence $\tilde{\mu} = \tilde{\nu}$ forces $\mu^{(\sigma)} = \nu^{(\sigma)}$ for every $\sigma > 0$. Realising $\mu^{(\sigma)}$ as the law of $X + \sigma Z$ (with $X \sim \mu$, $Z \sim \gamma_1$ independent) and letting $\sigma \downarrow 0$ gives $X + \sigma Z \to X$ almost surely, hence in probability and so in distribution: $\mu^{(\sigma)} \Rightarrow \mu$, and likewise $\nu^{(\sigma)} \Rightarrow \nu$. Uniqueness of weak limits gives $\mu = \nu$.

1.4 Lévy's continuity lemma

Convergence of characteristic functions, plus tightness, controls weak convergence of the underlying measures.

Lemma 1.10: Lévy continuity

Let $\{\mu_i\}_{i=1}^\infty$ be a uniformly tight sequence of probability measures on $\mathbb{R}^d$. If the characteristic functions satisfy $\tilde{\mu}_i(v) \to \tilde{\mu}(v)$ for every $v \in \mathbb{R}^d$, then $\mu_i \Rightarrow \mu$, where μ is the (unique) probability measure with characteristic function $\tilde{\mu}$.

Remark 1.3. By Prohorov (**??**), every subsequence μ_{i_k} has a further weakly convergent subsubsequence $\mu_{i_{k_r}} \Rightarrow \mu^*$. Continuity of the integrand forces $\widetilde{\mu^*} = \tilde{\mu}$ on all of $\mathbb{R}^d$, and uniqueness of characteristic functions (**??**) identifies $\mu^* = \mu$. The subsubsequence criterion (**??**) then promotes this to convergence of the full sequence.

1.5 The central limit theorem

We can now prove the headline result. The hypothesis is just iid plus a finite second moment.

Theorem 1.11: Central limit theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let $\{X_n\}_{n=1}^\infty$ be iid random vectors on $(\mathbb{R}^d, \mathcal{B})$ with

$$\mathbb{E}X_n = 0 \quad \text{and} \quad \mathbb{E}\,|X_n|^2 < \infty.$$

Set $S_n = \sum_{j=1}^n X_j$. Then

$$n^{-\frac{1}{2}} S_n \xrightarrow{d} Z,$$

where Z is a Gaussian random vector on $\mathbb{R}^d$ with mean zero and covariance Σ given by $\Sigma_{jk} = \mathbb{E}[X_{nj} X_{nk}]$.

The strategy of the proof is a two-step: *tightness* of the normalised sums via a second-moment Chebyshev bound, and *characteristic-function convergence* via Taylor expansion. The two ingredients meet in Lévy's lemma.

Remark 1.4 (tightness via Chebyshev). Since the X_j are mean zero and independent, $\mathbb{E}\langle X_j, X_k\rangle =$

0 for $j \neq k$, so

$$\mathbb{E}\,|n^{-\frac{1}{2}}S_n|^2 \;=\; \frac{1}{n}\,\mathbb{E}\left[\,\sum_{j,k=1}^{n}\langle X_j, X_k\rangle\right] \;=\; \mathbb{E}\,|X_j|^2.$$

For any $\varepsilon > 0$, choose $M_\varepsilon > 0$ with $\mathbb{E}|X_j|^2/M_\varepsilon^2 < \varepsilon$; Chebyshev's inequality gives $\mathbb{P}(|n^{-\frac{1}{2}}S_n| > M_\varepsilon) < \varepsilon$, uniformly in n. The sequence $\{n^{-\frac{1}{2}}S_n\}$ is therefore uniformly tight.

Remark 1.5 (characteristic-function expansion). Fix $v \in \mathbb{R}^d$. The scalars $\langle v, X_j\rangle$ are iid real-valued with $\mathbb{E}\langle v, X_j\rangle = 0$ and $\mathbb{E}\langle v, X_j\rangle^2 < \infty$. Define

$$h(v) \;:=\; \mathbb{E}\exp(\,i\langle v, X_j\rangle\,).$$

Then $h(0) = 1$, $\nabla h(0) = 0$ and $\nabla^2 h(0) = -\Sigma$ where $\Sigma = \mathbb{E}[X_j X_j^\top]$. Taylor's theorem gives

$$h(v) \;=\; 1 \;-\; \tfrac{1}{2}\,v^\top\Sigma\,v \;+\; o(|v|^2).$$

Independence then yields, for any fixed v,

$$\mathbb{E}\exp\{\,i\langle n^{-\frac{1}{2}}S_n, v\rangle\,\} \;=\; h(n^{-\frac{1}{2}}v)^n \;=\; \left(1 - \frac{v^\top\Sigma v}{2n} + o\left(\frac{|v|^2}{n}\right)\right)^n \;\longrightarrow\; \exp\{-\tfrac{1}{2}v^\top\Sigma v\}$$

as $n \to \infty$. The right-hand side is the characteristic function of the mean-zero Gaussian Z on $\mathbb{R}^d$ with covariance Σ. Combining with tightness and Lévy's continuity (Result 1.7) gives $n^{-\frac{1}{2}}S_n \Rightarrow Z$ — convergence in distribution.

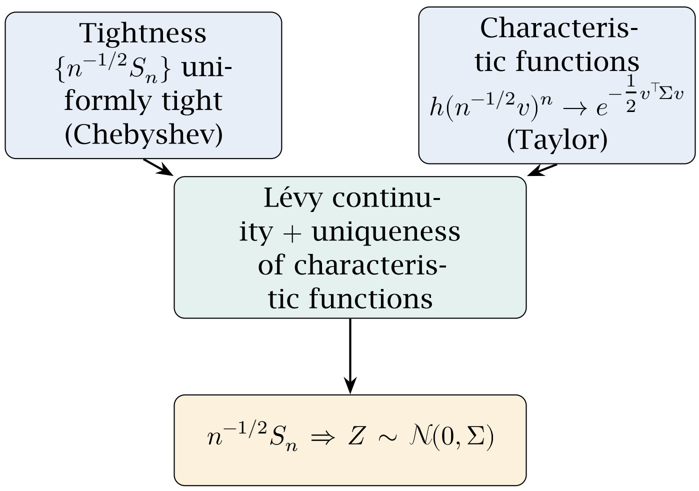

Figure 1. Architecture of the CLT proof: tightness and pointwise convergence of characteristic functions feed into Lévy's lemma; the limiting characteristic function identifies the Gaussian Z.

Remark 1.6. The covariance entry $\Sigma_{jk} = \mathbb{E}[X_{nj}X_{nk}]$ is independent of n by the iid hypothesis; the limiting Gaussian is the same regardless of which copy of X_n one uses to compute it. In the scalar case $d = 1$ the conclusion reduces to the familiar $n^{-1/2}S_n \Rightarrow \mathcal{N}(0,\sigma^2)$ with $\sigma^2 = \mathbb{E}X_1^2$.