Contents

2 Central limit theorem
  2a Introduction .................................................. 21
  2b Convolution ..................................................... 22
  2c The initial distribution does not matter .................... 24
  2d From smooth functions to indicators ....................... 26

2 Central limit theorem

2a Introduction

Discrete probability spaces are enough here as long as all random variables
are discrete (otherwise $\Omega = \mathbb{R}^n$ fits); to this end use triangle arrays.

Let $X_1, X_2, \ldots$ be independent identically distributed random variables,
and $S_n = X_1 + \cdots + X_n$.

2a1 Theorem. ¹ Let $\mathbb{E} X_1 = 0$ and $\mathbb{E} X_1^2 = 1$. Then

$$
\mathbb{P} \left( a \sqrt{n} < S_n < b \sqrt{n} \right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, dx \quad \text{as } n \to \infty
$$

whenever $-\infty \leq a \leq b \leq \infty$.

Clearly, the De Moivre-Laplace Theorem 1a20 is a special case.

More than 10 proofs are well-known. Some use Stirling formula. Some
use Brownian motion. Some prove convergence to the normal distribution.
Some prove first convergence to some distribution, and then identify it.

Moment method: first, find $\lim_n \mathbb{E} \left( \frac{S_n}{\sqrt{n}} \right)^k$ assuming all moments finite
(otherwise, truncate); then approximate the indicator of an interval by polynomi-
als.

Fourier transform ("characteristic function"): first, $\lim_n \mathbb{E} \exp \left( i \lambda \frac{S_n}{\sqrt{n}} \right) =
\exp \left( -\frac{\lambda^2}{2} \right)$; then approximate the indicator of an interval by trigonometric
sums.

Smooth test functions (Lindeberg): first, $\mathbb{E} f \left( \frac{S_n}{\sqrt{n}} \right) - \mathbb{E} f \left( \frac{\tilde{S}_n}{\sqrt{n}} \right) \to 0$ as
$n \to \infty$ for $f \in C^3$; then approximate the indicator of an interval by such
smooth functions. This will be done here.

¹[KS, Sect. 10.1, Th. 10.5]; [D, Sect. 2.4, Theorem (4.1)].
2b Convolution

The convolution $\nu * f$ of a probability distribution $\nu$ on $\mathbb{R}$ and a bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ is a function $\mathbb{R} \to \mathbb{R}$ defined by

$$ (\nu * f)(x) = \int f(x + y) \nu(dy). $$

For a discrete $\nu$ the convolution is a linear combination of shifts. In general it may be thought of as an integral combination of shifts. Probabilistically, $$(P_X * f)(a) = \mathbb{E} f(a + X).$$

2b1 Lemma. If $f$ is bounded and continuous then also $\mu * f$ is, and $\|\mu * f\| \leq \|f\|$.

Here and below the norm is supremal (rather than $L_2$):

$$ \|f\| = \sup_{x \in \mathbb{R}} |f(x)|. $$

Proof. Boundedness: $|\mathbb{E} f(a + X)| \leq \sup |f(\cdot)|$. Continuity: if $a_n \to a$ then $f(a_n + x) \to f(a + x)$ pointwise, thus $\mathbb{E} f(a_n + X) \to \mathbb{E} f(a + X)$ by the bounded convergence theorem.

For independent $X, Y$ we have $P_{X+Y} * f = P_Y * P_X * f$ (it means, $(P_Y * (P_X * f))$), since

$$ (P_{X+Y} * f)(a) = \mathbb{E} f(a + X + Y) = \iint f(a + x + y) P_X(dx)P_Y(dy) = \int \left( \int f(a+x+y)P_X(dx) \right) P_Y(dy) = \int (P_X * f)(a+y) P_Y(dy) = (P_Y * (P_X * f))(a). $$

We define the convolution of two probability distributions $\mu, \nu$ by $(\mu * \nu)(B) = (\mu \times \nu)(\{(x, y) : x + y \in B\})$, then $P_{X+Y} = P_X * P_Y$ for independent $X, Y$, and we may interpret $P_Y * P_X * f$ as $(P_Y * P_X) * f$ equally well.

Convolution for discrete:

$$(P_X * f)(a) = \sum_x p_X(x) f(a + x);$$

$$P_{X+Y}(a) = \sum_{(x,y):x+y=a} p_X(x)p_Y(y) = \sum_x p_X(x)p_Y(a-x).$$

---

1The definition generalizes easily to finite signed measures and bounded Borel functions, but we do not need it.

2Well, it is required by the definition above...
Convolution for absolutely continuous:

\[(P_X * f)(a) = \int p_X(x)f(a + x) \, dx;\]

\[p_{X+Y}(a) = \int p_X(x)p_Y(a - x) \, dx.\]

Some examples:

\[\text{Binom}(m, p) * \text{Binom}(n, p) = \text{Binom}(m + n, p), \quad \text{— binomial}\]

\[N(a_1, \sigma_1^2) * N(a_2, \sigma_2^2) = N(a_1 + a_2, \sigma_1^2 + \sigma_2^2). \quad \text{— normal}\]

The latter equality can be checked by integration, or obtained from the former by a limiting procedure, but better note that the standard two-dimensional normal distribution \(N(0, 1) \times N(0, 1)\) has the density\(^1\)

\[\frac{1}{2\sqrt{\pi}e^{-x^2/2}} \cdot \frac{1}{2\sqrt{\pi}e^{-y^2/2}} = \frac{1}{2\pi}e^{-(x^2+y^2)/2}\]

invariant under rotations; thus, \(X \cos \alpha + Y \sin \alpha \sim N(0, 1)\) for all \(\alpha\).

**2b2 Lemma.** If \(f\) has a bounded and continuous derivative, then also \(\mu * f\) has, and \((\mu * f)' = \mu * f'\).

**Proof.** We have a bounded continuous \(g\) satisfying \(f(b) = f(a) + \int_a^b g(x) \, dx\). Thus,

\[(\mu * f)(b) = \int f(b + y) \mu(dy) = \int \left( f(a + y) + \int_{a+y}^{b+y} g(x) \, dx \right) \mu(dy) =
\]

\[= \int f(a + y) \mu(dy) + \int \left( \int_a^b g(x + y) \, dx \right) \mu(dy) =
\]

\[= (\mu * f)(a) + \int_a^b \left( \int g(x + y) \mu(dy) \right) dx = (\mu * f)(a) + \int_a^b (\mu * g)(x) \, dx.\]

\[\square\]

The same holds for \(f''\) and \(f'''\).\(^2\)

---

\(^1\)In addition, integrating it in polar coordinates we get \(\frac{1}{2\pi} \int_0^{2\pi} e^{-r^2/2} r \, dr \) \(\int_0^{2\pi} \, d\varphi = 1\), which shows that \(1/\sqrt{2\pi}\) is the right coefficient for the density of \(N(0, 1)\). (See also Proof of 1a20.)

\(^2\)And so on, of course, but we need only three derivatives.
2c The initial distribution does not matter

Let $\mu, \nu$ be two probability distributions on $\mathbb{R}$ satisfying

$$
\int x \mu(dx) = \int x \nu(dx) = 0, \quad \int x^2 \mu(dx) = \int x^2 \nu(dx) = 1.
$$

We consider independent random variables $X_1, \ldots, X_n$ distributed $\mu$, and independent random variables $Y_1, \ldots, Y_n$ distributed $\nu$. Note that $\mathbb{E} X_1 = \mathbb{E} Y_1 = 0$ and $\mathbb{E} X_1^2 = \mathbb{E} Y_1^2 = 1$.

2c1 Proposition. If $f, f', f'', f'''$ are continuous and bounded on $\mathbb{R}$ then

$$
\mathbb{E} f(\frac{X_1 + \cdots + X_n}{\sqrt{n}}) - \mathbb{E} f(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}) \to 0 \text{ as } n \to \infty.
$$

The proof will be given after a corollary.

2c2 Corollary. $\mathbb{E} f(\frac{X_1 + \cdots + X_n}{\sqrt{n}}) \to \sqrt{\frac{2}{\pi}} \int_{-\infty}^{+\infty} f(x) e^{-x^2/2} dx$ for all $n$.

Proof of the corollary. Let $Y_1$ be normal $N(0, 1)$, then $Y_1 + \cdots + Y_n$ is also normal, thus

$$
\mathbb{E} f(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}) = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{+\infty} f(x) e^{-x^2/2} dx \text{ for all } n.
$$

We start proving the proposition.

We have $\int (a + bx + cx^2) \mu(dx) = \int (a + bx + cx^2) \nu(dx)$ for all $a, b, c \in \mathbb{R}$. Similarly,

$$
\int (a + bx + cx^2) \mu_n(dx) = \int (a + bx + cx^2) \nu_n(dx);
$$

here and below $\mu_n$ is the distribution of $X_1/\sqrt{n}$, and $\nu_n$ — of $Y_1/\sqrt{n}$; that is, $\int f(\frac{x}{\sqrt{n}}) \mu(dx) = \int f \, d\mu_n$ (and the same for $\nu$). These $\mu_n, \nu_n$ are useful, since

$$
(2c3) \quad \mathbb{E} f(\frac{X_1 + \cdots + X_n}{\sqrt{n}}) = (\mu_n * \cdots * \mu_n * f)(0) = (\mu_n * \cdots * \mu_n * f)(0),
$$

and the same for $Y$ and $\nu$. 

2c4 Lemma. There exist \( \varepsilon_n \to 0 \) such that for every \( f \) (as in 2c1) and every \( n \),
\[
\left| \int f \, d\mu_n - \int f \, d\nu_n \right| \leq \frac{\varepsilon_n}{n} (\|f''\| + \|f'''\|).
\]

2c5 Remark. These \( \varepsilon_n \) depend on \( \mu, \nu \) (but not \( f \)). If \( \mu, \nu \) have third moments then moreover
\[
\left| \int f \, d\mu_n - \int f \, d\nu_n \right| \leq \frac{1}{6n^{1.5}} \|f'''\| (E|X_1|^3 + E|Y_1|^3).
\]

Proof of the lemma. We define \( g \) by
\[
f(x) = f(0) + f'(0)x + \frac{1}{2} f''(0)x^2 + g(x);
\]
g is continuous but not bounded;
\[
|g(x)| \leq \|f'''\| \cdot \frac{1}{6}|x|^3.
\]
We have \( \int (f - g) \, d\mu_n = \int (f - g) \, d\nu_n \), therefore
\[
\left| \int f \, d\mu_n - \int f \, d\nu_n \right| \leq \int |g| \, d\mu_n + \int |g| \, d\nu_n,
\]
which leads immediately to 2c5, but we need an argument that does not require the third moments. We note that \( |\frac{1}{2} f''(0)x^2 + g(x)| \leq \frac{1}{2} \|f''\||x|^2 \), therefore
\[
|g(x)| \leq \|f''\| \cdot |x|^2,
\]
and split the integral:
\[
\int |g| \, d\mu_n = \int \left| g \left( \frac{x}{\sqrt{n}} \right) \right| \mu(dx) \leq \int_{|x| \leq n^{1/12}} \|f''\| \cdot \frac{1}{6} \left( \frac{x}{\sqrt{n}} \right)^3 \mu(dx) + \int_{|x| > n^{1/12}} \|f''\| \cdot \left( \frac{x}{\sqrt{n}} \right)^2 \mu(dx) \leq \frac{\varepsilon_n}{n} (\|f''\| + \|f'''\|) \quad \text{where}
\]
\[
\varepsilon_n = \max \left( \frac{1}{24n^{1/6}} \int_{|x| > n^{1/12}} x^2 \mu(dx), \int_{|x| \leq n^{1/12}} \left( \frac{x}{\sqrt{n}} \right)^3 \mu(dx) \right);
\]
the same holds for \( \int |g| \, d\nu_n \).

\footnote{The exponent 1/12 may be replaced with any other number between 0 and 1/6.}
Proof of Proposition 2c1. By (2c3) it is sufficient to prove that $|(\mu_n * f)(0) - (\nu_n * f)(0)| \to 0$. Applying Lemma 2c4 to a shifted function $x \mapsto f(a + x)$ we get

$$
\|\mu_n * f - \nu_n * f\| \leq \frac{\varepsilon_n}{n} (\|f''\| + \|f'''\|).
$$

We turn $\mu_n * f$ into $\nu_n * f$ gradually:

$$
\mu_n * f - \nu_n * f = \sum_{k=0}^{n-1} (\mu_n^{*(n-k)} * \nu_n^{*k} * f - \mu_n^{*(n-k-1)} * \nu_n^{*(k+1)} * f) = \sum_{k=0}^{n-1} \mu_n^{*(n-k-1)} * (\mu_n * f_k - \nu_n * f_k),
$$

where $f_k = \nu_n^{*k} * f$. Now, $\|f''_k\| \leq \|f''\|$, $\|f'''_k\| \leq \|f'''\|$, and $\|\mu_n^{*(n-k-1)} * (...)\| \leq \|(...)\|$; thus,

$$
\|\mu_n * f - \nu_n * f\| \leq \sum_{k=0}^{n-1} \frac{\varepsilon_n}{n} (\|f''\| + \|f'''\|) = \varepsilon_n (\|f''\| + \|f'''\|) \to 0
$$
as $n \to \infty$.

\[ \square \]

2d From smooth functions to indicators

2d1 Lemma. There exists a function $\varphi : \mathbb{R} \to \mathbb{R}$ having three bounded derivatives and such that $\varphi(x) = 0$ for all $x \leq -1$, $\varphi(x) = 1$ for all $x \geq 0$.

Proof. The function $\psi(x) = (1 - x^2)^4$ for $|x| \leq 1$, otherwise 0, has two (in fact, three) continuous derivatives. We take $\varphi(x) = \frac{1}{c} \int_{-\infty}^{\infty} \psi(t) \, dt$ where $c = \int_{-\infty}^{\infty} \psi(t) \, dt$.

Let $X_1, \ldots, X_n$ be as in 2c1. By 2c2 for every $a \in \mathbb{R}$ and $\varepsilon > 0$,

$$
\mathbb{E} \varphi \left( \frac{1}{\varepsilon} \left( \frac{X_1 + \cdots + X_n}{\sqrt{n}} - a \right) \right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi \left( \frac{1}{\varepsilon} (x - a) \right) e^{-x^2/2} \, dx
$$
as $n \to \infty$. Taking into account that

$$
\mathbb{P} \left( \frac{X_1 + \cdots + X_n}{\sqrt{n}} \geq a \right) \leq \mathbb{E} \varphi \left( \frac{1}{\varepsilon} \left( \frac{X_1 + \cdots + X_n}{\sqrt{n}} - a \right) \right)
$$
we get

$$
\limsup_{n \to \infty} \mathbb{P} \left( \frac{X_1 + \cdots + X_n}{\sqrt{n}} \geq a \right) \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi \left( \frac{1}{\varepsilon} (x - a) \right) e^{-x^2/2} \, dx.
$$
The right-hand side converges to \( \frac{1}{\sqrt{2\pi}} \int_0^a e^{-x^2/2} \, dx \) as \( \varepsilon \to 0 \). Thus, \( \limsup \mathbb{P}(X_1 + \cdots + X_n \geq a\sqrt{n}) \leq \mathbb{P}(\xi \geq a) \) where \( \xi \sim N(0, 1) \); or equivalently, \( \liminf \mathbb{P}(X_1 + \cdots + X_n < a\sqrt{n}) \geq \mathbb{P}(\xi < a) \). Similarly, \( \limsup \mathbb{P}(-X_1 - \cdots - X_n \geq a\sqrt{n}) \leq \mathbb{P}(\xi \geq a) \), that is, \( \limsup \mathbb{P}(X_1 + \cdots + X_n \leq -a\sqrt{n}) \leq \mathbb{P}(\xi \leq -a) \), or equivalently, \( \limsup \mathbb{P}(X_1 + \cdots + X_n \leq a\sqrt{n}) \leq \mathbb{P}(\xi \leq a) \). We have

\[
\mathbb{P}(\xi < a) \leq \liminf \mathbb{P}(X_1 + \cdots + X_n < a\sqrt{n}) \leq \\
\leq \limsup \mathbb{P}(X_1 + \cdots + X_n \leq a\sqrt{n}) \leq \mathbb{P}(\xi \leq a) = \mathbb{P}(\xi < a),
\]

therefore

\[
\lim_{n \to \infty} \mathbb{P}(X_1 + \cdots + X_n < a\sqrt{n}) = \mathbb{P}(\xi < a).
\]