Contents

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

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

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

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

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 $\lim_n \mathbb{E}\, \big(\frac{S_n}{\sqrt{n}}\big)^k$ assuming all moments finite (otherwise, truncate); then approximate the indicator of an interval by polynomials.

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

Smooth test functions (Lindeberg): first, $\mathbb{E}\, f\big(\frac{S_n}{\sqrt{n}}\big) - \mathbb{E}\, f\big(\frac{\tilde S_n}{\sqrt{n}}\big) \to 0$ 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 ν on $\mathbb{R}$ and a bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ is a function $\mathbb{R} \to \mathbb{R}$ defined by[1]

$$(\nu * f)(x) = \int f(x+y)\,\nu(\mathrm{d}y)\,.$$

For a discrete ν 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[2] then also $\mu * f$ is, and $\|\mu * f\| \le \|f\|$.

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

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

Proof. Boundedness: $|\mathbb{E}\, f(a+X)| \le \sup |f(\cdot)|$. 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. $\square$

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

$$(P_{X+Y} * f)(a) = \mathbb{E}\, f(a+X+Y) = \iint f(a+x+y)\,P_X(\mathrm{d}x)P_Y(\mathrm{d}y) =$$
$$\int \Big(\int f(a+x+y) P_X(\mathrm{d}x) \Big) P_Y(\mathrm{d}y) = \int (P_X * f)(a+y)\,P_Y(\mathrm{d}y) = \big(P_Y * (P_X * f)\big)(a)\,.$$

We define the convolution of two probability distributions μ, ν by $(\mu*\nu)(B) = (\mu \times \nu)\big(\{(x,y) : x+y \in B\}\big)$, 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:

$$(P_X * f)(a) = \sum_x p_X(x) f(a+x)\,;$$
$$p_{X+Y}(a) = \sum_{(x,y):x+y=a} p_X(x) p_Y(y) = \sum_x p_X(x) p_Y(a-x)\,.$$

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

[2] Well, it is required by the definition above…

Convolution for absolutely continuous:

$$(P_X * f)(a) = \int p_X(x) f(a+x)\,\mathrm{d}x\,;$$
$$p_{X+Y}(a) = \int p_X(x) p_Y(a-x)\,\mathrm{d}x\,.$$

Some examples:

$$\mathrm{Binom}(m,p) * \mathrm{Binom}(n,p) = \mathrm{Binom}(m+n,p)\,, \qquad \text{— binomial}$$
$$N(a_1,\sigma_1^2) * N(a_2,\sigma_2^2) = N(a_1+a_2,\sigma_1^2+\sigma_2^2)\,. \qquad \text{— normal}$$

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]

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

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

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 $f(b) = f(a) + \int_a^b g(x)\,\mathrm{d}x$. Thus,

$$(\mu * f)(b) = \int f(b+y)\,\mu(\mathrm{d}y) = \int \Big(f(a+y) + \int_{a+y}^{b+y} g(x)\,\mathrm{d}x \Big) \mu(\mathrm{d}y) =$$
$$= \int f(a+y)\mu(\mathrm{d}y) + \int \Big(\int_a^b g(x+y)\,\mathrm{d}x \Big) \mu(\mathrm{d}y) =$$
$$= (\mu * f)(a) + \int_a^b \Big(\int g(x+y)\,\mu(\mathrm{d}y) \Big) \mathrm{d}x = (\mu * f)(a) + \int_a^b (\mu * g)(x)\,\mathrm{d}x\,.$$

□

The same holds for f'' and f'''.[2]

[1] In addition, integrating it in polar coordinates we get $\frac{1}{2\pi} \big(\int_0^\infty \mathrm{e}^{-r^2/2} r\,\mathrm{d}r \big) \big(\int_0^{2\pi} \mathrm{d}\varphi \big) = 1$, 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 μ, ν be two probability distributions on $\mathbb{R}$ satisfying

$$\int x\,\mu(\mathrm{d}x) = \int x\,\nu(\mathrm{d}x) = 0\,, \quad \int x^2\,\mu(\mathrm{d}x) = \int x^2\,\nu(\mathrm{d}x) = 1\,.$$

We consider independent random variables $X_1, \dots, X_n$ distributed μ, and independent random variables $Y_1, \dots, Y_n$ distributed ν. 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

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

The proof will be given after a corollary.

2c2 Corollary.

$$\mathbb{E}\,f\Big(\frac{X_1 + \dots + X_n}{\sqrt{n}}\Big) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-x^2/2}\,\mathrm{d}x\,.$$

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

$$\mathbb{E}\,f\Big(\frac{Y_1 + \dots + Y_n}{\sqrt{n}}\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) \mathrm{e}^{-x^2/2}\,\mathrm{d}x \quad \text{for all } n\,.$$

□

We start proving the proposition.

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

$$\int (a + bx + cx^2)\,\mu_n(\mathrm{d}x) = \int (a + bx + cx^2)\,\nu_n(\mathrm{d}x)\,;$$

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

$$\mathbb{E}\,f\Big(\frac{X_1 + \dots + X_n}{\sqrt{n}}\Big) = (\mu_n * \dots * \mu_n * f)(0) = (\mu_n^{*n} * f)(0)\,, \tag{2c3}$$

and the same for Y and ν.

2c4 Lemma. There exist $\varepsilon_n \to 0$ such that for every f (as in 2c1) and every n,

$$\Big| \int f \, \mathrm{d}\mu_n - \int f \, \mathrm{d}\nu_n \Big| \le \frac{\varepsilon_n}{n} \big(\|f''\| + \|f'''\| \big) \, .$$

2c5 Remark. These ε_n depend on μ, ν (but not f). If μ, ν have third moments then moreover

$$\Big| \int f \, \mathrm{d}\mu_n - \int f \, \mathrm{d}\nu_n \Big| \le \frac{1}{6n^{1.5}} \|f'''\| \big(\mathbb{E} \, |X_1|^3 + \mathbb{E} \, |Y_1|^3 \big) \, .$$

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;

$$|g(x)| \le \|f'''\| \cdot \frac{1}{6} |x|^3 \, .$$

We have $\int (f-g) \, \mathrm{d}\mu_n = \int (f-g) \, \mathrm{d}\nu_n$, therefore

$$\Big| \int f \, \mathrm{d}\mu_n - \int f \, \mathrm{d}\nu_n \Big| \le \int |g| \, \mathrm{d}\mu_n + \int |g| \, \mathrm{d}\nu_n \, ,$$

which leads immediately to 2c5, but we need an argument that does not require the third moments. We note that $|\frac{1}{2} f''(0) x^2 + g(x)| \le \frac{1}{2} \|f''\| x^2$, therefore

$$|g(x)| \le \|f''\| \cdot |x|^2 \, ,$$

and split the integral:[1]

$$\begin{aligned} \int |g| \, \mathrm{d}\mu_n &= \int \Big| g\Big(\frac{x}{\sqrt{n}} \Big) \Big| \, \mu(\mathrm{d}x) \le \\ &\le \underbrace{\int_{|x| \le n^{1/12}} \|f'''\| \cdot \frac{1}{6} \Big| \frac{x}{\sqrt{n}} \Big|^3 \, \mu(\mathrm{d}x)}_{O(n^{-7/6})} + \underbrace{\int_{|x| > n^{1/12}} \|f''\| \cdot \Big| \frac{x}{\sqrt{n}} \Big|^2 \, \mu(\mathrm{d}x)}_{o(1/n)} \le \\ &\le \frac{\varepsilon_n}{n} \big(\|f''\| + \|f'''\| \big) \quad \text{where} \\ &\qquad \varepsilon_n = \max \Big(\frac{1}{24 n^{1/6}}, \int_{|x| > n^{1/12}} x^2 \, \mu(\mathrm{d}x) \Big) \, ; \end{aligned}$$

the same holds for $\int |g| \, \mathrm{d}\nu_n$. □

[1] 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 $|(\mu_n^{*n} * f)(0) - (\nu_n^{*n} * f)(0)| \to 0$. Applying Lemma 2c4 to a shifted function $x \mapsto f(a+x)$ we get

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

We turn μ_n^{*n} into ν_n^{*n} gradually:

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

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

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

as $n \to \infty$. □

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 $x \le -1$, $\varphi(x) = 1$ for all $x \ge 0$.

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

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

$$\mathbb{E} \, \varphi \bigg(\frac{1}{\varepsilon} \Big(\frac{X_1 + \dots + X_n}{\sqrt{n}} - a \Big) \bigg) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi \Big(\frac{1}{\varepsilon} (x-a) \Big) \mathrm{e}^{-x^2/2} \, \mathrm{d}x$$

as $n \to \infty$. Taking into account that

$$\mathbb{P} \Big(\frac{X_1 + \dots + X_n}{\sqrt{n}} \ge a \Big) \le \mathbb{E} \, \varphi \bigg(\frac{1}{\varepsilon} \Big(\frac{X_1 + \dots + X_n}{\sqrt{n}} - a \Big) \bigg)$$

we get

$$\limsup_{n\to\infty} \mathbb{P} \Big(\frac{X_1 + \dots + X_n}{\sqrt{n}} \ge a \Big) \le \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi \Big(\frac{1}{\varepsilon} (x-a) \Big) \mathrm{e}^{-x^2/2} \, \mathrm{d}x .$$

The right-hand side converges to $\frac{1}{\sqrt{2\pi}}\int_a^\infty \mathrm{e}^{-x^2/2}\,\mathrm{d}x$ as $\varepsilon \to 0$. Thus, $\limsup \mathbb{P}\big(X_1+\dots+X_n \ge a\sqrt{n}\big) \le \mathbb{P}\big(\xi \ge a\big)$ where $\xi \sim N(0,1)$; or equivalently, $\liminf \mathbb{P}\big(X_1+\dots+X_n < a\sqrt{n}\big) \ge \mathbb{P}\big(\xi < a\big)$. Similarly, $\limsup \mathbb{P}\big(-X_1-\dots-X_n \ge a\sqrt{n}\big) \le \mathbb{P}\big(\xi \ge a\big)$, that is, $\limsup \mathbb{P}\big(X_1+\dots+X_n \le -a\sqrt{n}\big) \le \mathbb{P}\big(\xi \le -a\big)$, or equivalently, $\limsup \mathbb{P}\big(X_1+\dots+X_n \le a\sqrt{n}\big) \le \mathbb{P}\big(\xi \le a\big)$. We have

$$\begin{aligned}\mathbb{P}\big(\xi < a\big) \le \liminf \mathbb{P}\big(X_1+\dots+X_n < a\sqrt{n}\big) \le \\ \le \limsup \mathbb{P}\big(X_1+\dots+X_n \le a\sqrt{n}\big) \le \mathbb{P}\big(\xi \le a\big) = \mathbb{P}\big(\xi < a\big)\,,\end{aligned}$$

therefore

$$\lim_{n\to\infty} \mathbb{P}\big(X_1+\dots+X_n < a\sqrt{n}\big) = \mathbb{P}\big(\xi < a\big)\,.$$