7 The central limit theorem

The central limit theorem states that if we take n samples, not necessarily normally distributed, with mean μ and standard deviation σ, then the sample mean is approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$ for sufficiently large n.

Definition 7.1. Let $X, X_1, X_2, \ldots$ be real random variables. We say that (X_n) converges to X in distribution, symbolically $X_n \xrightarrow{\mathcal{D}} X$, if

$$\lim_{n\to\infty} \mathbf{P}(X_n \le x) = \mathbf{P}(X \le x)$$

at all points where the distribution function of X is continuous.

Remark 7.2. Almost sure convergence implies convergence in distribution.

Theorem 7.3 (Central limit theorem (CLT))**.** Let $X_1, X_2, \ldots$ be square integrable real random variables i.i.d. with mean m and standard deviation σ. Put $S_n = X_1 + \cdots + X_n$. Then

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \mathbf{P}(X \le x) \quad \text{for any } x \in \mathbb{R},$$

where X is a standard normal random variable. In other words,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} X.$$

Exercise 7.1. A fair coin is tossed 10,000 times; let X be the number of times it comes up heads. Use the central limit theorem and a table of values (printed or electronic) of

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

to estimate

(i) the probability that $4950 \le X \le 5050$;

(ii) the number k such that $|X - 5000| \le k$ with probability 0.98.

Exercise 7.2. Let $X_1, X_2, \ldots$ be i.i.d real random variables with $X_1 \sim \mathrm{Rad}_{1/2}$, i.e., $\mathbf{P}(X_1 = -1) = \mathbf{P}(X_1 = 1) = 1/2$, and let $S_n = X_1 + \ldots + X_n$. Show that $\mathbf{P}(S^* = \infty) = 1$.

Exercise 7.3. Let $X_1, X_2, \ldots$ be i.i.d real random variables with $X_1 \sim \mathrm{Rad}_{1/2}$; let $S_n = X_1 + \ldots + X_n$, and let $Z_n = S_n - n$. (Z_n is the excess of heads over tails in n tosses, if $X_n = 1$ when heads and $X_n = 0$ when tails on the nth toss.) Show that

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{Z_n}{\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Exercise 7.4. Let $X_1, X_2, \ldots$ be i.i.d Poisson random variables with parameter $\lambda = 1$. Let $S_n = X_1 + \cdots + X_n$. Show that

$$\lim_{n\to\infty} \frac{S_n - n}{\sqrt{n}} = X,$$

where X is a standard normal random variable.

Exercise 7.5. Show that

$$\lim_{n\to\infty} e^{-n} \left(\sum_{k=0}^{n} \frac{n^k}{k!} \right) = \frac{1}{2}.$$

(Hint: use exercise 7.4.)