Central Limit Theorem [nln9]

The central limit theorem is a major extension of the law of large numbers. It explains the unique role of the Gaussian distribution in statistical physics.

Given are a large number of statistically independent random variables $X_i, i = 1, \ldots, N$ with equal probability distributions $P_X(x_i)$. The only restriction on the shape of $P_X(x_i)$ is that the moments $\langle X_i^n \rangle = \langle X^n \rangle$ are finite for all n.

Goal: Find the probability distribution $P_Y(y)$ for the random variable $Y = (X_1 - \langle X \rangle + \cdots + X_N - \langle X \rangle)/N$.

$$P_Y(y) = \int dx_1\, P_X(x_1) \cdots \int dx_N\, P_X(x_N) \delta\left(y - \frac{1}{N} \sum_{i=1}^{N} [x_i - \langle X \rangle] \right).$$

Characteristic function:

$$\Phi_Y(k) \equiv \int dy\, e^{iky} P_Y(y), \quad P_Y(y) = \frac{1}{2\pi} \int dk\, e^{-iky} \Phi_Y(k).$$

$$\begin{aligned} \Rightarrow \ \Phi_Y(k) &= \int dx_1\, P_X(x_1) \cdots \int dx_N\, P_X(x_N) \exp\left(i \frac{k}{N} \sum_{i=1}^{N} [x_i - \langle X \rangle] \right) \\ &= \left[\bar{\Phi}\left(k/N\right) \right]^N, \end{aligned}$$

$$\begin{aligned} \bar{\Phi}\left(\frac{k}{N}\right) &= \int dx\, e^{i(k/N)(x - \langle X \rangle)} P_X(x) = \exp\left(-\frac{1}{2} \left(\frac{k}{N}\right)^2 \langle\langle X^2 \rangle\rangle + \cdots \right) \\ &= 1 - \frac{1}{2} \left(\frac{k}{N}\right)^2 \langle\langle X^2 \rangle\rangle + \mathrm{O}\left(\frac{k^3}{N^3}\right), \end{aligned}$$

where we have performed a cumulant expansion to leading order.

$$\Rightarrow \ \Phi_Y(y) = \left[1 - \frac{k^2 \langle\langle X^2 \rangle\rangle}{2N^2} + \mathrm{O}\left(\frac{k^3}{N^3}\right) \right]^N \xrightarrow{N \to \infty} \exp\left(-\frac{k^2 \langle\langle X^2 \rangle\rangle}{2N} \right).$$

where we have used $\lim_{N\to\infty}(1 + z/N)^N = e^z$.

$$\Rightarrow \ P_Y(y) = \sqrt{\frac{N}{2\pi \langle\langle X^2 \rangle\rangle}} \exp\left(-\frac{Ny^2}{2\langle\langle X^2 \rangle\rangle} \right) = \frac{1}{\sqrt{2\pi \langle\langle Y^2 \rangle\rangle}}\, e^{-y^2/2\langle\langle Y^2 \rangle\rangle}$$

with variance $\langle\langle Y^2 \rangle\rangle = \langle\langle X^2 \rangle\rangle / N$

Note that regardless of the form of $P_X(x)$, the average of a large number of (independent) measurements of X will be a Gaussian with standard deviation $\sigma_Y = \sigma_X / \sqrt{N}$.