

PROBABILITY THEORY

MATH 154

Unit 19: Central limit theorem

19.1. A non-constant random variable $X \in \mathcal{L}^2$ can be **normalized** to $X^* = \frac{(X - \mathbb{E}[X])}{\sigma(X)}$. It has now zero **mean** $\mathbb{E}[X^*] = 0$ and **variance** $\sigma(X^*) = \sqrt{\text{Var}[X^*]} = 1$. Not every random variable can be normalized as such. A Cauchy distributed random variable for example has no finite variance and so can not be scaled to have variance 1. We need $X \in \mathcal{L}^2$.

Theorem 1 (Central limit theorem). *Given $X_n \in \mathcal{L}^2$ which are IID with mean 0 and finite variance $\sigma^2 > 0$. Then $S_n/(\sigma\sqrt{n}) \rightarrow N(0, 1)$ in distribution.*

19.2. The CLT can be encoded more briefly as $S_n^* \rightarrow^d N(0, 1)$. Lets look first at some properties of the density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

called **normal distribution** $N(0, \sigma^2)$. We have $\mathbb{E}[|X|^p] = 2 \int_0^\infty x^p f(x) dx$, which is after a substitution $u = x^2/(2\sigma^2)$ equal to

$$\frac{1}{\sqrt{\pi}} 2^{p/2} \sigma^p \int_0^\infty u^{\frac{1}{2}(p+1)-1} e^{-u} du .$$

The integral to the right is by definition equal to $\frac{2^{p/2}}{\sqrt{\pi}} \Gamma(\frac{1}{2}(p+1))$. The characteristic function is $\phi_X(t) = e^{-t^2\sigma^2/2}$.

19.3. Here is the proof of the CLT:

Proof. By the Lévy criterion for weak convergence and noting that $e^{-t^2/2}$ is continuous, we only to show that for every $t \in \mathbb{R}$

$$\mathbb{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] \rightarrow e^{-t^2/2} .$$

Denote by ϕ_X the characteristic function of X_k . Because $\mathbb{E}[X_k] = 0$ and $\mathbb{E}[X_k^2] = \sigma^2$, and $e^{itX_k} = 1 + itX_k - t^2X_k^2/2 + o(t^2X_k^2)$. Taking expectations gives

$$\phi_{X_k}(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2) .$$

Now use that the sum of independent random variables $aX_k = X_k/(\sigma\sqrt{n})$ produces the product of characteristic functions and also that $\phi_{aX}(t) = \phi_X(at)$.

$$\begin{aligned} \mathbb{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] &= \phi_X\left(\frac{t}{\sigma\sqrt{n}}\right)^n \\ &= \left(1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{1}{n}\right)\right)^n = e^{-t^2/2} + o(1). \end{aligned}$$

□

19.4. Remark: There is a different proof that only needs independence and would allow different distributions for each X_i . It needs some assumptions however, like $M = \sup_i \|X_i\|_3 < \infty$ and $\delta = \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \text{Var}[X_i] > 0$. One can then show that

$$\lim_{n \rightarrow \infty} \mathbb{P}[S_n^* \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy, \quad \forall x \in \mathbb{R}.$$

A $N(0, \sigma^2)$ distributed random variable X satisfies $\mathbb{E}[|X|^p] = \frac{1}{\sqrt{\pi}} 2^{p/2} \sigma^p \Gamma(\frac{1}{2}(p+1))$ and so $\mathbb{E}[|X|^3] = \sqrt{\frac{8}{\pi}} \sigma^3$. But the proof is more technical.

19.5. Let $\mathcal{P}$ denote the space of probability measure μ on $(\mathbb{R}, \mathcal{B})$ which have the properties that $\int_{\mathbb{R}} x^2 d\mu(x) = 1, \int_{\mathbb{R}} x d\mu(x) = 0$. Define a map $T : \mathcal{P} \rightarrow \mathcal{P}$ as

$$T\mu(A) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A\left(\frac{x+y}{\sqrt{2}}\right) d\mu(x) d\mu(y).$$

We could realize this map by taking for a given μ a random variable X with that law, then build a new random variable Y with the same distribution that is independent, then look at the law of $(X+Y)/\sqrt{2}$.

Theorem 2 (Renormalisation fixed point). *The only fixed point of T on $\mathcal{P}$ is the law $N(0, 1)$ of the standard normal distribution. It is an attractive fixed point in the sense that $T^n \mu \rightarrow N(0, 1)$ starting with any initial condition μ .*

Proof. If μ is the law of a random variables X, Y with $\text{Var}[X] = \text{Var}[Y] = 1$ and $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Then $T(\mu)$ is the law of the normalized random variable $(X+Y)/\sqrt{2}$ because the independent random variables X, Y can be realized on the probability space $(\mathbb{R}^2, \mathcal{B}, \mu \times \mu)$ as coordinate functions $X((x, y)) = x, Y((x, y)) = y$. Then $T(\mu)$ is obviously the law of $(X+Y)/\sqrt{2}$. Now use that $T^n(X) = (S_{2^n})^*$ converges in distribution to $N(0, 1)$. □

19.6. An other cool fact is that we can see that for normalized random variables with continuous PDF f that has finite **differential entropy** $S(X) = -\int_{\mathbb{R}} f(x) \log(f(x)) dx$, the entropy increases when applying T . The Gibbs inequality $D[p, q] \geq 0$ for the Kullback-Leibler divergence can show this. It uses that the entropy of $N(0, 1)$ is $\log(\sqrt{2\pi}e) = 1.41894\dots$ ¹

Theorem 3. *The normal distribution is the distribution of maximal entropy among all distributions of finite differential entropy in $\mathcal{P}$.*

¹Pretty cool: it is the log of the geometric mean of 2π and e .