PROBABILITY THEORY

MATH 154

Unit 19: Central limit theorem

19.1. Any non-constant random variable $X \in \mathcal{L}^2$ can be **normalized** to $X^* = \frac{(X-\mathrm{E}[X])}{\sigma(X)}$. This normalized variable has zero **mean** $\mathrm{E}[X^*] = 0$ and **variance** $\sigma(X^*) = \sqrt{\mathrm{Var}[X^*]} = 1$. We can not normalize every random variable. A Cauchy distributed random variable for example has no finite variance and so can not be scaled to have variance 1. But we can normalize any non-constant random variable 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 $S_n/(\sigma\sqrt{n}) \to N(0,1)$ in distribution.*

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

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

which belongs to the **normal distribution** $N(0,\sigma^2)$. We we have $\mathrm{E}[|X|^p] = 2\int_0^\infty x^p f(x)\, dx$ 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 $\frac{2^{p/2}}{\sqrt{\pi}} \Gamma(\frac{1}{2}(p+1))$. We can also compute the characteristic function. $\phi_X(t) = e^{-t^2\sigma^2/2}$. To the proof:

Proof. By the Levy criterion for weak convergence and noting that $e^{-t^2/2}$ has no atoms, we have to show that for all $t \in \mathbb{R}$

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

Denote by ϕ_{X_n} the characteristic function of X_n. As this is independent of n, we just write ϕ. Since by assumption $\mathrm{E}[X_n] = 0$ and $\mathrm{E}[X_n^2] = \sigma^2$, we can use, using the Taylor formula with remainder term:

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

This works despite that $\phi(t)$ does not necessarily have a full Taylor expansion. Lets use the **Landau notation** $o(f)$ for a term which could be replaced by a function g

satisfying $g(t)/f(t) \to 0$.

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

□

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

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

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

19.4. 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 the map on $\mathcal{P}$ as follows:

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

Technically, we can 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 $T^n\mu \to N(0,1)$ starting with any initial condition μ.*

Proof. If μ is the law of a random variables X, Y with $\mathrm{Var}[X] = \mathrm{Var}[Y] = 1$ and $\mathrm{E}[X] = \mathrm{E}[Y] = 0$. 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 $N(0,1)$. □

19.5. An other cool fact is that we can see that for normalized random variables with continuous PDF f that has finite **differentiable entropy** $S(X) = -\int_{\mathbb{R}} f(x)\log(f(x)) \, dx$, the entropy increases when applying T. The Gibbs inequality from lecture 12 $D[p,q] \geq 0$ gives after a short computation. It uses that the entropy of $N(0,1)$ is $\log(\sqrt{2\pi e}) = 1.41894....$ [1]

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

Oliver Knill, knill@math.harvard.edu, Math 154, Spring, 2025

[1]Pretty cool: the log of the geometric mean of 2π and e.