PROOF OF CLT USING LINDEBERG PRINCIPLE

Section 1

Introduction

Our look at the concept of universality starts from the Central Limit Theorem: if we add i.i.d. random variables X_i of finite variance and zero mean, then the law of this sums, when rescaled properly convergences to the Gaussian law.

Theorem 1.1 (Central limit theorem)**.** *Let $X_1, X_2, \dots$ be i.i.d. random variables of finite variance σ^2. Then $n^{-1/2}\sum_{i=1}^n (X_i - \mathbb{E}X_i)$ converges in law to $N(0, \sigma^2)$.*

CLT is an example of the phenomenon of universality because the resulting law, the Gaussian law, depends very little on the exact details of the initial law of X_i - indeed only the variance counts.

One of the proofs that explains this universality rather well is is called the *Lindeberg exchange principle.*

The basic idea comes in two observations:

(1) The Gaussian law itself is a stable law: if we take $Y_1, Y_2, \dots$ to be i.i.d Gaussians of zero mean, then for every $n \geq 1$, the law of $n^{-1/2}\sum_{i=1} Y_i$ has the same Gaussian law.
(2) Starting now from general i.i.d. random variables X_i, one observes that if we swap the variables X_i one by one for Gaussians Y_i of the same mean and variance, then the error by doing so at every step is so small that in fact it is negligible in the limit!

This key step is encapsulated in the following proposition, which we state for convenience for variables with unit variance.

Proposition 1.2 (Lindeberg Exchange Principle)**.** *Let $X_1, X_2, \dots$ be i.i.d. zero mean unit variance random variables, let $Y \sim N(0,1)$. Define $S_n := n^{-1/2}\sum_{i=1}^n X_i$. Then for any $\delta > 0$, there is a $n_\delta \in \mathbb{N}$ so that for all $n > n_\delta$ and for all $f : \mathbb{R} \to \mathbb{R}$ smooth with uniformly bounded derivatives up to third order, we have that*

$$|\mathbb{E}f(S_n) - \mathbb{E}f(Y)| < 2\delta(\sup_{x\in\mathbb{R}} |f'(x)| + \sup_{x\in\mathbb{R}} |f''(x)| + \sup_{x\in\mathbb{R}} |f'''(x)|).$$

Before proving the proposition, let us see how CLT follows from this proposition. In fact there are many ways to argument this. For example one could just argue that smooth functions of compact support are dense in the space of continuous functions. Let us here give a very direct argument:

Proof of the CLT:. By subtracting the mean and rescaling we can assume that X_i are zero mean and unit varaince.

Let us denote by $S_n := n^{-1/2}\sum_{i=1}^n X_i$. Then we have that $\mathbb{P}(S_n \leq x) = \mathbb{E}(\mathbf{1}_{\{S_n \leq x\}})$. Now we can bound $\mathbf{1}_{\{S_n \leq x\}}$ from above by smooth functions f_ϵ that are equal to 1 in $(-\infty, x]$ are equal to 0 in $[x + \epsilon, \infty)$ such that the third derivative is bounded by $C\epsilon^{-3}$.

Thus using the proposition we have that for n large enough

$$\mathbb{P}(S_n \leq x) \leq \mathbb{E}f_\epsilon(Y) + 6C\delta\epsilon^{-3},$$

where $Y \sim N(0,1)$. In particular Y has bounded density and thus $\mathbb{E}f_\epsilon(Y) \leq \mathbb{P}(Y \leq x) + c\epsilon$ for some $c > 0$. Hence

$$\mathbb{P}(S_n \leq x) \leq \mathbb{P}(Y \leq x) + 6C\epsilon^{-3}\delta + c\epsilon.$$

Now, given δ we can choose $\epsilon = \delta^{1/4}$ to get

$$\mathbb{P}(S_n \leq x) \leq \mathbb{P}(Y \leq x) + \widetilde{C}\delta^{1/4}$$

for all n large enough. By taking δ arbitrarily small, we conclude that $\limsup_{n\in\mathbb{N}} \mathbb{P}(S_n \leq x) \leq \mathbb{P}(Y \leq x)$. We can similarly prove a lower bound to conclude that $\lim_{n\to\infty} \mathbb{P}(S_n \leq x) = \mathbb{P}(Y \leq x)$ and the theorem follows. $\square$

We now prove the proposition. The exchange principle itself would come out a bit more cleanly if we assumed bounded third moments, but it is also nice how it mixes together with the so called truncation method.

Proof of Lindeberg Exchange Principle: Let $Y_1, Y_2 \ldots$ be i.i.d. standard Gaussians. For $k \geq 1$, write

$$S_{n,k} := \frac{\sum_{i=1}^{k-1} X_i + \sum_{i=k}^{n} Y_i}{n^{1/2}},$$

Notice that $S_{n,n+1} = S_n$ and $S_{n,1} \sim N(0,1)$. Thus we can write

$$f(S_n) - f(Y) = \sum_{k=1}^{n} f(S_{n,k+1}) - f(S_{n,k}). \tag{1.1}$$

Our aim will be to control each individual summand. To do this write further

$$S^0_{n,k} := \frac{\sum_{i=1}^{k-1} X_i + \sum_{i=k+1}^{n} Y_i}{n^{1/2}},$$

where we have omitted the k-th term altogether.

By third-order Taylor's approximation we can write a.s.

$$f(S_{n,k+1}) = f(S^0_{n,k}) + \frac{X_k}{n^{1/2}} f'(S^0_{n,k}) + \frac{X_k^2}{2n} f''(S^0_{n,k}) + \frac{X_k^3}{6n^{3/2}} f'''(x_L),$$

with x_L between $S_{n,k+1}$ and $S^0_{n,k}$ and similarly

$$f(S_{n,k}) = f(S^0_{n,k}) + \frac{Y_k}{n^{1/2}} f'(S^0_{n,k}) + \frac{Y_k^2}{2n} f''(S^0_{n,k}) + \frac{X_k^3}{6n^{3/2}} f'''(x_L),$$

We would want to take expectations now and conclude by summing up. Indeed, the two first moments would then just cancel, because of matching moments of X_k and Y_k and independence of $S^0_{n,k}$ from both Y_k and X_k (check that you know how to conclude in this case!). However, a priori we don't know that the third moment exists and thus the third order term could cause us problems.

To circumvent this problem we use a truncation method: for $\delta > 0$ small, write

$$X_k = X_k \mathbf{1}_{\{|X_k| \leq \delta n^{1/2}\}} + X_k \mathbf{1}_{\{|X_k| > \delta n^{1/2}\}}.$$

Denote the first term by $\widetilde{X}_k$ and the second term by $X_{k,>}$. We will use the exchange principle above for $\widetilde{X}_k$ instead of X_k and control the contribution of the second terms separately.

So set $\widetilde{S}_n = n^{-1/2} \sum_{k=1}^{n} \widetilde{X}_k$. Notice that $\mathbb{E}\widetilde{X}_k \neq 0$ and $\mathbb{E}\widetilde{X}_k^2 \neq 1$, so by the end of telescoping in (1.1) with $\widetilde{X}_k$ instead of X_k, we are left with

$$|\mathbb{E} f(\widetilde{S}_n) - f(Y)| \leq n^{1/2} |\sup_{x\in\mathbb{R}} f'(x)| |\mathbb{E}\widetilde{X}_k| + |\sup_{x\in\mathbb{R}} f''(x)| \mathbb{E}\widetilde{X}_k^2 + n^{-1/2} |\sup_{x\in\mathbb{R}} f'''(x)| \mathbb{E}|\widetilde{X}_k|^3. \tag{1.2}$$

Now observe that $\mathbb{E}|\widetilde{X}_k|^3 \leq \delta n^{1/2} \mathbb{E}|\widetilde{X}_k|^2 = \delta n^{1/2}$, as $\mathbb{E}\widetilde{X}_k^2 \leq \mathbb{E}X_k^2$. Moreover we claim that:

Claim 1.3. *For n large enough $|\mathbb{E}\widetilde{X}_k^2 - 1| < \delta^2$ and $|\mathbb{E}\widetilde{X}_k| = |\mathbb{E}X_{k,>}| < \delta n^{-1/2}$.*

These two things together imply that the RHS of Equation (1.2) can be bounded by

$$\delta|\sup_{x\in\mathbb{R}} f'(x)| + \delta^2|\sup_{x\in\mathbb{R}} f''(x)| + \delta|\sup_{x\in\mathbb{R}} f'''(x)|.$$

Moreover, this claim also helps us deal with the part coming from the tails $X_{k,>}$. Denote $E_n = n^{-1/2}\sum_{k=1}^n X_{k,>}$. Then again by Taylor expansion and the claim

$$|\mathbb{E}f(S_n) - \mathbb{E}f(\widetilde{S}_n)| \leq |\sup_{x\in\mathbb{R}} f'(x)||\mathbb{E}E_n| \leq |\sup_{x\in\mathbb{R}} f'(x)|n^{-1/2}\sum_{k=1}^n |\mathbb{E}X_{k,>}| \leq |\sup_{x\in\mathbb{R}} f'(x)|\delta.$$

Thus we conclude the proposition by proving the claim:

Proof of claim: The first part just comes from the fact that $\mathbb{E}X_k^2 = 1$, the fact that $|\tilde{X}_k| \to |X_k|$ as $n \to \infty$ and the monotone convergence theorem. Notice that this also implies that $\mathbb{E}|X_{k,>}|^2 < \delta^2$.

For the second part, the equality comes from the fact that $\mathbb{E}X_k = 0$. Thus by Jensen it suffices to show that $\mathbb{E}|X_{k,>}| < \delta n^{-1/2}$. But this follows from the previous part as

$$n^{1/2}\delta\mathbb{E}|X_{k,>}| \leq \mathbb{E}|X_{k,>}|^2 < \delta^2.$$

□

□

In fact Proposition 1.2 also gives a rate of convergence that gets better when we control more derivatives.

Exercise 1.1. *Suppose $X_1, X_2, \ldots$ are i.i.d. such that for $k \geq 3$ the first $k-1$ moments of X_1 match with those of a standard Gaussian Y and the k-th moment is finite. Let $S_n := n^{-1/2}\sum_{i=1}^n X_i$ as before and f be smooth with uniformly bounded derivatives.*

Prove that $|\mathbb{E}f(S_n) - f(Y)| = O(n^{1-\frac{k}{2}})$.