TOPICS IN PROBABILITY. PART II: UNIVERSALITY

Exercise sheet 7: Universality and CLT

Exercise 1. *Does the CLT remain true if we do not assume that variances of the variables are finite? More precisely, let $(X_i)_i$ be i.i.d. random variables with mean zero and infinite variance. Does $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ converge in law (to normal distribution)?*

$\star$: *Suppose X_i's are symmetric with $\mathbb{P}[|X_i| > x] = x^{-\alpha}$ for $\alpha \in (0,2)$ for all $x \geq 1$. For which α's is X_i integrable? Is it possible to find a different normalization of $\sum_{i=1}^n X_i$ such that the resulting variable converges in law (to some probability law)? What is the intuitive explanation?*

Proof. The canonical statement of CLT does not remain true for general variables of infinite variance. For instance, let X_i's be as in the star question. We will show on the next exercise sheet (Exercise 3 Sheet 8) that their characteristic function satisfies $\phi(t) \sim 1 - C|t|^\alpha(1+o(1))$ as $|t|$ tends to 0 for an appropriate constant $C > 0$. Thus, as $n \to \infty$,

$$\phi_{\frac{\sum^n X_i}{n^{1/2}}}(u) = \phi(t/n^{1/2})^n \sim \left(1 - C\frac{|u|^\alpha(1+o(1))}{n^{\alpha/2}}\right)^n \sim e^{-C|u|^\alpha n^{1-\alpha/2}(1+o(1))} \sim 0.$$

For $\alpha \in (1,2)$, the variables X_i's are integrable, hence, the counterexample to the first part is found. $\square$

Exercise 2 (Lindeberg-Feller CLT)**.** *Let $(X_i)_i$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = \sigma_i^2 < \infty$ (not necessarily equal to one another). Let $s_n^2 := \sum_{i=1}^n \sigma_n^2$. Show by adjusting the proof of Lindeberg exchange principle to this more general setting that if for any $\varepsilon > 0$,*

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i| > \varepsilon s_n\}}] \xrightarrow{n\to\infty} 0,$$

then $\frac{1}{s_n}\sum_{i=1}^n X_i$ converges in law to a standard normal variable.

Hint: Take Y_i's (in the proof of the exchange principle) to be independent central normal with variance σ_i^2, note that now $\tilde{X}_k$'s are not identically distributed — how does it affect (1.2)?

Find a sequence of independent random variables $(X_i)_i$ with mean zero and variance one such that $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ does not converge in law to a standard normal variable.

Proof. Let Y_i's be independent centered normal with variance σ_i^2. We set

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

so that $S_{n,n+1} = S_n$ and $S_{n,1} \sim N(0,1)$, and

$$S_{n,k}^0 := \frac{\sum_{i=1}^{k-1} X_i + \sum_{i=k+1}^n Y_i}{s_n}.$$

Analogously to the proof in the lecture, using Taylor's theorem up to second order with a remainder, we get that a.s.,

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

where $|R_n(f,u)| \le \max(\|f'''\|_\infty, \|f''\|_\infty)\min(u^3/6, u^2) =: C\min(u^3/6, u^2)$. More precisely, we used that $f(y) - f(x) = f'(x)(y-x) + f''(x)(y-x)^2 + R(x,y,f)$ with the remainder bounded by $|R(x,y,f)| \le \min(\frac{\sup_{z\in[x,y]}|f'''(z)|}{6}(y-x)^3, \sup_{z\in[x,y]}|f''(z)|(y-x)^2)$[1]. Note that R_n is integrable as X_k's and Y_k's are square-integrable. Since furthermore $\mathbb{E}[X_k] = \mathbb{E}[Y_k] = 0$, $\mathbb{E}[X_k^2] = \sigma_k^2 = \mathbb{E}[Y_k^2]$ (and X_k, Y_k are independent of $S^0_{n,k}$), we obtain that

$$|\mathbb{E}[f(S_n) - f(Y)]| \le C\sum_{k=1}^n \left(\mathbb{E}[\min(|X_k|^3/(6s_n^3), X_k^2/s_n^2)] + \mathbb{E}[\min(|Y_k|^3/(6s_n^3), Y_k^2/s_n^2)]\right),$$

where Y is a standard normal random variable.

For $\varepsilon > 0$ small, write

$$X_k = X_k\mathbf{1}_{\{|X_k|\le\varepsilon s_n\}} + X_k\mathbf{1}_{\{|X_k|>\varepsilon s_n\}} =: \widetilde{X}_k + X_{k,>};$$

and analogously $Y_k = \widetilde{Y}_k + Y_{k,>}$. We then have

$$\mathbb{E}[\min(|X_k|^3/(6s_n^3), X_k^2/s_n^2)] \le \frac{1}{6s_n^3}\mathbb{E}[|\widetilde{X}_k|^3] + \frac{1}{s_n^2}\mathbb{E}[X_{k,>}^2] \le \frac{\varepsilon\sigma_k^2}{6s_n^2} + \frac{1}{s_n^2}\mathbb{E}[X_{k,>}^2]. \tag{0.1}$$

Note that by assumption, $\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}[X_{k,>}^2]$ converges to zero as n tends to infinity. Therefore, for all n sufficiently large, $\sum_{k=1}^n \mathbb{E}[\min(|X_k|^3/(6s_n^3), X_k^2/s_n^2)] \le \varepsilon/4$. Furthermore, Lindeberg's condition implies that $\max_k \frac{\sigma_k^2}{s_n^2} \to 0$ as n tends to infinity. Indeed, for any $\delta > 0$,

$$\max_k \sigma_k^2/s_n^2 = \max_k \mathbb{E}[X_k^2/s_n^2] \le \delta^2 + \sum_j \mathbb{E}[X_k^2/s_n^2\mathbf{1}_{\{|X_k|\delta s_n\}}] \xrightarrow{n\to\infty} \delta^2 \xrightarrow{\delta\downarrow 0} 0.$$

Thus, Y_k's also satisfy Lindeberg's condition: for any $\varepsilon > 0$,

$$\frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}[Y_k^2\mathbf{1}_{\{|Y_k|>\varepsilon s_n\}}] \le \sum_{k=1}^n \frac{\sigma_k^2}{s_n^2}\mathbb{E}[Y^4]\max_k \mathbb{P}[|Y| > \varepsilon s_n/\sigma_k] \le 3\mathbb{P}[|Y| > \varepsilon s_n/\max_k \sigma_k] \xrightarrow{n\to\infty} 0.$$

Hence, (0.1) also holds for Y_k's, and we conclude that for any $\varepsilon > 0$, there exists n_ε such that for all $n \ge n_\varepsilon$,

$$|\mathbb{E}[f(S_n) - \mathbb{E}[f(Y)]| \le \varepsilon(\|f'\|_\infty + \|f''\|_\infty + \|f'''\|_\infty).$$

With the same proof as in the lecture notes the desired CLT follows from the exchange principle.

Consider a sequence of independent random variables $(X_n)_n$, for each n supported in $\{-2^n, 0, 2^n\}$ with $\mathbb{P}[X_n = 2^n] = \mathbb{P}[X_n = -2^n] = \frac{1}{2}4^{-n}$. Note that $\mathbb{E}[X_n] = 0, \mathrm{Var}[X_n] = 1$

[1]this, in turn, follows from iterative application of fundamental theorem of calculus: $f(y) - f(x) = \int_x^y f'(z)\mathrm{d}z = \cdots = f'(x)(y-x) + f''(x)(y-x)^2/2 + \int_x^y\int_x^z\int_x^u f'''(s)\mathrm{d}s\mathrm{d}u\mathrm{d}z$ and the latter summand is bounded by both $\sup|f'''||y-x|^3/6$ and $2\sup|f''|(y-x)^2/2$.

and its characteristic function is given by $\phi_n(t) = 1 - 4^{-n}(1 - \cos(2^n t))$ for each $n \in \mathbb{N}, t \in \mathbb{R}$. Therefore,

$$\phi_{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \left(1 - 4^{-i}(1 - \cos(2^i t/\sqrt{n}))\right) = \exp\left(\sum_{i=1}^n \log(1 - 4^{-i}(1 - \cos(2^i t/\sqrt{n})))\right)$$
$$\geq \exp\left(\sum_{i=1}^n \log(1 - 2 * 4^{-i})\right) \xrightarrow[\text{unif in } t]{n\to\infty} \exp\left(\sum_{i=1}^\infty \log(1 - 2 * 4^{-i})\right) =: \delta > 0.$$

Since characteristic function of any symmetric non-degenerate normal distribution is given by e^{-Ct^2} for some $C > 0$ and converges to zero as t tends to infinity, we conclude that $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ cannot converge in law to normal distribution. ☐

Exercise 3 (Sherrington-Kirkpatrick ground state)**.**

Let us consider statistical mechanics model of n spins, i.e. particles that can be in one of two states $\{\pm 1\}$, due to Sherrington and Kirkpatrick, which models a rough energy landscape by introducing random interactions between the spins. More precisely, for a configuration of spins $\sigma \in \{-1, 1\}^n$, let the energy $H(\sigma)$ be defined as

$$H(\sigma) = \frac{1}{n^{3/2}} \sum_{1 \leq i < j \leq n} X_{ij}\sigma_i\sigma_j,$$

where X_{ij} are independent with zero mean and unit variance. The ground-state energy, that is, the energy the system attains at zero temperature (when in thermal equilibrium), is given by

$$Z = \min_{\sigma \in \{-1,1\}^n} H(\sigma).$$

One of the basic questions to ask is whether Z is universal, more precisely, whether it depend significantly on the distribution of X_{ij}'s or not? Universality is important from the physical perspective: it states that macroscopic observations are insensitive to the microscopic details in the description of physical systems. We want to apply Universality theorem proven in the lecture, but Z is not three times differentiable w.r.t. X_{ij}. The solution is to introduce a suitable smooth approximation of the minimum.

- *Show that for any $\beta > 0$,*

$$|Z - Z_\beta| \leq \frac{n \log 2}{\beta}, \quad Z_\beta = -\frac{1}{\beta}\log\left(\sum_{\sigma \in \{-1,1\}^n} e^{-\beta H(\sigma)}\right).$$

- *Combine part one with the universality theorem to show that the expected ground-state energy $\mathbb{E}[Z]$ is insensitive to the distribution of the variables X_{ij}. Assume that the third moments of X_{ij}'s exist and are uniformly bounded.*
- $\star$*: extend to the case of unbounded/non-existent third moments.*

Proof. Let us prove the upper bound for $Z - Z_\beta$,

$$Z - Z_\beta = Z + \frac{1}{\beta}\log\left(\sum_{\sigma \in \{-1,1\}^n} e^{-\beta H(\sigma)}\right) \leq Z + \frac{1}{\beta}\log\left(\sum_{\sigma \in \{-1,1\}^n} e^{-\beta Z}\right) = \frac{n}{\beta}\log 2.$$

On the other hand, if $\alpha \in \{-1,1\}^n$ is the (random) minimizing configuration, then

$$Z - Z_\beta = \frac{1}{\beta}\log\left(\sum_{\sigma\in\{-1,1\}^n} e^{-\beta(H(\sigma)-Z)}\right) = \frac{1}{\beta}\log\left(1 + \sum_{\sigma\neq\alpha}\underbrace{e^{-\beta(H(\sigma)-Z)}}_{\geq 0}\right) \geq 0 \geq -\frac{n}{\beta}\log 2.$$

As for the second part of the exercise, let us first assume that X_{ij}'s have third moments which are uniformly bounded by $C > 0$ and that Z^0 corresponds to Z with all X_{ij} being independent standard Gaussians, then by the universality theorem

$$\begin{aligned}|\mathbb{E}[Z] - \mathbb{E}[Z^0]| &\leq |\mathbb{E}[Z_\beta] - \mathbb{E}[Z^0_\beta]| + 2\frac{n}{\beta}\log 2 \leq 2\frac{n}{\beta}\log 2 + \frac{2C}{6}\sum_{1\leq i<j\leq n}\left\|\partial^3 Z_\beta/\partial {X_{ij}}^3\right\|_\infty \\ &\leq 2\frac{n}{\beta}\log 2 + \frac{2C}{6}\sum_{1\leq i<j\leq n} 3\frac{\beta^2}{n^{9/2}} \leq 2\frac{n}{\beta}\log 2 + \frac{C\beta^2}{n^{5/2}} \leq \tilde{C}\left(\frac{n}{\beta} + \left(\frac{\beta}{n^{5/4}}\right)^2\right).\end{aligned}$$

So, choosing $\beta = n^u$ for any $u \in (1, 5/4)$ yields a good bound (which converges to 0 as n tends to infinity) on the difference of $\mathbb{E}[Z]$ and $\mathbb{E}[Z^0]$.

Extension to the general case (no assumption on third moments) can be done fully analogously to the proof of Exercise 1 Sheet 7.

Discussion of Z^0: Note that since we are summing only over X_{ij} with $i < j$, we can consider matrix X, which has entries X_{ij} for $i \leq j$ and X_{ji} for $i > j$, so that X is symmetric. Then there exist orthogonal matrix U and a diagonal matrix D so that $X = U^T D U$ and so

$$\begin{aligned}H(\sigma) &= \frac{1}{2n}(U\sigma)^T\frac{D}{\sqrt{n}}(U\sigma) - \frac{\sum_i X_{ii}}{2n^{3/2}} = \frac{\|\sigma\|^2}{2n}\frac{(U\sigma)^T\frac{D}{\sqrt{n}}(U\sigma)}{\|U\sigma\|^2} - \frac{\sum_i X_{ii}}{2n^{3/2}} \\ &= \frac{\left\|\sqrt{\frac{D}{\sqrt{n}}}(U\sigma)\right\|^2}{2\|U\sigma\|^2} - \frac{\sum_i X_{ii}}{2n^{3/2}} = \frac{\left\|\sqrt{\frac{X}{\sqrt{n}}}\sigma\right\|^2}{2\|\sigma\|^2} - \frac{\sum_i X_{ii}}{2n^{3/2}},\end{aligned}$$

where $\sqrt{D}$ is diagonal with potentially complex entries so that $\sqrt{D}^2 = D$ and $\sqrt{X} = U^T\sqrt{D}U$. The second term converges almost surely and in L^2 to zero as n tends to infinity. So, at the end of the day, exploring limiting behaviour of Z^0 is equivalent to investigating operator norm of $\sqrt{X/\sqrt{n}}$ restricted to the space of spin configurations, which is in turn equivalent to investigate the operator norm of $X/\sqrt{n}$ restricted to the space of spin configurations. The latter can be bounded by absolute values of maximal and minimal eigenvalues of $X/\sqrt{n}$. $\square$