Homework 2
Math 270B, Winter 2020, Prof. Roman Vershynin

Problem 1 (Scheffé's theorem)

(a) (For continuous distributions) Prove that if probability density functions of X_n converge to probability density function of X pointwise, then X_n converges to X weakly.

(b) (For discrete distributions) Prove that if probability mass functions of X_n converge to probability mass function of X pointwise, then X_n converges to X weakly.

(c) (No converse) In general, weak convergence does not imply pointwise convergence of probability density functions. Show this by example.
sol:
(a). Let f_n be the density function of X_n and f be the density function of X.
We recall the generalized dominated convergence theorem: Let $\{h_n\}$ and $\{g_n\}$ be sequences of measurable functions, g_n non-negative and integrable. If (1)$|h_n| \le g_n$ on $\mathbb{R}$, (2) $f_n(x) \to f$ and $g_n \to g$ pointwisely, and (3) $\lim_{n\to\infty} \int g_n = \int g < \infty$. Then, $\lim_{n\to\infty} \int h_n = \int h < \infty$

Now $\forall x$, let $E = (-\infty, x]$, let $h_n = (f - f_n) \cdot \mathbf{1}_E$ and $g_n = f_n + f$. We see that $|h_n| \le g_n$ on $\mathbb{R}$, and $h_n \to 0 = h$, $g_n \to 2f = g$ pointwisely. Also, $\forall n$, $\int_{\mathbb{R}} g_n = \int_{\mathbb{R}} f_n + \int_{\mathbb{R}} f_n = \int_{\mathbb{R}} 2f = 2$. Now, we apply the generalized dominated convergence theorem, we have that $\lim_{n\to\infty} \int_{\mathbb{R}} h_n \to 0$, then we have $lim_{n\to\infty} \int_E f_n \to \int_E f$. Hence, $\mathbb{P}\{X_n \le x\} \to \mathbb{P}\{X \le x\}, \forall x$.

(b). Let the collection of ranges (which is countable) of X_n and X be $\{a_k\}_{k=1}^{\infty} \subset \mathbb{R}$. $\forall x \in \mathbb{R}$, let $N(x)$ (if exists) be the largest integer s.t. $a_{N(x)} \le x$. Then $\forall \epsilon$, based on pointwise convergence of pmf, we have:

$$\forall k, 1 \le k \le N(x), \quad \exists M_k \quad s.t. \quad |\mathbb{P}\{X_n = a_k\} - \mathbb{P}\{X = a_k\}| < \frac{\epsilon}{N(x)}$$

. Take $M = \max_{1 \le k \le N(x)} M_k$. Then we have

$$|\mathbb{P}\{X_n \le x\} - \mathbb{P}\{X \le x\}| \le \sum_{k=1}^{N(x)} |\mathbb{P}\{X_n = a_k\} - \mathbb{P}\{X = a_k\}| < \epsilon \ \ \forall n > M$$

If no such $N(x)$ exsits, then either $(x > a_k, \forall k)$ $\mathbb{P}\{X_n \le x\} = \mathbb{P}\{X \le x\} = 1, \forall n$, or $(x < a_k, \forall k)$ $\mathbb{P}\{X_n \le x\} = \mathbb{P}\{X \le x\} = 0, \forall n$.
Hence, $\forall x$, $\mathbb{P}\{X_n \le x\} \to \mathbb{P}\{X \le x\}$, so $X_n \xrightarrow{w} X$.

(c). I recycle the example from the previous homework: $f_n(x) = \begin{cases} 2 & x \in [\frac{2k}{2^n}, \frac{2k+1}{2^n}) \\ 0 & else \end{cases}$

for $k = 0, 1, \cdots, 2^{n-1} - 1$. Then, $X_n \xrightarrow{D}$ Uniform([0,1]), but $f_n(x) \not\to 1$ anywhere.

Problem 2 (Weak limit of normal random variables)

Consider normal random variables $X_n \sim N(\mu_n, \sigma_n^2)$. Assume X_n converge weakly to some random variable X. Prove that $X \sim N(\mu, \sigma^2)$ where $\mu = \lim \mu_n$ and $\sigma^2 = \lim s_n^2$ (and both limits exist).

sol: Since $X_n \xrightarrow{D} X$, by continuity theorem, we know that the sequence of probability measures $\{P_n\}$ is tight. Then, $\forall \epsilon > 0\ \exists M > 0$, s.t. $\mathbb{P}_n(\{|x| > M\}) = \mathbb{P}\{|X_n| > M\} < \epsilon, \forall n$. Now, suppose $\sup_n |\mu_n| = \infty$, then we have $|\mu_N| > M$ for some N, so $\mathbb{P}\{|X_N| > M\} \geq \frac{1}{2}$. This is a contradiction. Hence, $\sup_n \mu_n < \infty$.

Meanwhile, we note that $\mathbb{P}\{|X_n| > M\} = 1 - \int_{-M}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu_n)^2}{2\sigma^2}} dx \geq 1 - \int_{-M}^{M} \frac{1}{\sigma\sqrt{2\pi}} dx = 1 - \frac{2M}{\sigma_n\sqrt{2\pi}}$. Then, if $\sup_n \sigma_n = \infty$ we have that $\forall M, \forall \epsilon > 0$, $\mathbb{P}\{|X_m| > M\} > \epsilon$ for some m. This contradicts to tightness.
Hence, both μ_n and σ_n are bounded. Then, we first find a convergence subsequence n_k s.t μ_k converges. Then, we find a convergence subsequence of σ_{n_k}. We call this further subsequence $\{n_l\}$. Now, we have that both $\{\mu_{n_l}\}$ and $\{\sigma_{n_l}\}$ converge. Denote the limits μ and σ.
By Continuity theorem, we have $\psi_X(t) = \lim_{n\to\infty} \psi_{X_n}(t), \forall t$. But

$$\lim_{l\to\infty} \psi_{X_{n_l}}(t) = \lim_{l\to\infty} e^{i\mu_{n_l} - \sigma_{n_l}^2 t^2/2} = e^{\lim_{l\to\infty}(i\mu_{n_l}) - \sigma_{n_l}^2 t/2} = e^{i\mu - \sigma^2 t/2}$$

Therefore, we have that $\psi_X(t) = e^{i\mu - \sigma^2 t/2}$, which means X has distribution $N(\mu, \sigma^2)$. Also, by $\psi_{X_n}(t) \to \psi_X(t)$, we see that $\mu = \lim_{n\to\infty} \mu_n$ and $\sigma = \lim_{n\to\infty} \sigma_n$

Problem 3 (No convergence in probability in CLT)

Let $X_1, X_2, \ldots$ be independent Rademacher random variables[1] Let $S_n = X_1 + \cdots + X_n$.

(a) Prove that the sequence $(S_n/\sqrt{n})$ is unbounded almost surely.

(b) Prove that $(S_n/\sqrt{n})$ does not converge in probability.

sol: Let $\mathcal{F}_k = \sigma(X_k, X_{k+1}, \cdots)$. Denote the tail algebra $\tau = \cap_{k=1}^{\infty} \mathcal{F}_k$. For any fixed $N \in \mathbb{N}$, let event $A = \{\limsup_n \frac{S_n}{\sqrt{n}}$ is unbounded$\}$ and $A_m = \{\limsup_n \frac{S_n}{\sqrt{n} > m}\}$. We see that $\forall m, A_m \in \tau$, as the $\limsup_n \frac{S_n}{\sqrt{n}}$ does not depend on finitely many X_k (If given the result, change finitely many of X_k will not change the result). Then, by Kolmogorove 0-1 law. We have that $\mathbb{P}\{A_m\} = 1$ or 0. We have that

$$\mathbb{P}\{A_m\} = \lim_{n\to\infty} \mathbb{P}\left\{\cup_{N\geq n} \frac{S_N}{\sqrt{N}} > m\right\} \geq \lim_{n\to\infty} \mathbb{P}\left\{\frac{S_n}{\sqrt{n}} > m\right\} \to \mathbb{P}\{N(0,1) > m\} > 0$$

. Hence, $\mathbb{P}\{A_m\} = 1, \forall m$. Then $\mathbb{P}\{A\} = \mathbb{P}\{\cap_{m=1}^{\infty} A_m\} = 1$
(b). Now, suppose that $\frac{S_n}{\sqrt{n}} \to Z \sim N(0,1)$ in probability. We also have that $\frac{S_{2n}}{\sqrt{2n}} \xrightarrow{P} Z$.

[1]A Rademacher random variable takes values $-1, 1$ with probability $1/2$ each.

Then, we have that $\frac{S_n}{\sqrt{n}} - \frac{S_{2n}}{\sqrt{2n}} \to 0$ in probability. However, we have that $\sqrt{2}\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}} = \frac{S_{2n}-S_n}{\sqrt{n}} \to N(0,1)$, and it is independent to $\frac{S_n}{\sqrt{n}}$. Therefore, we have that

$$\mathbb{P}\left\{-1 \le \frac{S_n}{\sqrt{n}} \le 0, 2 \le \frac{S_{2n}-S_n}{\sqrt{n}} \le 4\right\} = \mathbb{P}\left\{-1 \le \frac{S_n}{\sqrt{n}} \le 0\right\}\mathbb{P}\left\{2 \le \frac{S_{2n}-S_n}{\sqrt{n}} \le 4\right\}$$
$$\to \mathbb{P}\left\{-1 \le Z \le 0\right\}\mathbb{P}\left\{2 \le Z \le 4\right\} > 0$$

However, this means that $\mathbb{P}\left\{-1 \le \frac{S_n}{\sqrt{n}} \le 0, \frac{1}{\sqrt{2}} \le \frac{S_{2n}}{\sqrt{2n}} \le \frac{4}{\sqrt{2}}\right\} \to \mathbb{P}\left\{-1 \le Z \le 0\right\}\mathbb{P}\left\{2 \le Z \le 4\right\} > 0$, which contridicts to $\frac{S_n}{\sqrt{n}} - \frac{S_{2n}}{\sqrt{2n}} \to 0$ in probability.

Problem 4 (Non-summable variances yield CLT)

Let $X_1, X_2, \ldots$ be independent random variables such that there exists $M > 0$ so that $|X_i| \le M$ almost surely for all i. Show that if $\sum_i \mathrm{Var}(X_i) = \infty$ then the sum $S_n = X_1 + \cdots + X_n$ satisfies

$$\frac{S_n - \mathbb{E}\, S_n}{\sqrt{\mathrm{Var}(S_n)}} \to N(0,1) \quad \text{weakly.}$$

It suffices to show that a.s. uniformly bounded of X_is implies the Lindbery-Feller condition:

$$\forall \epsilon > 0, \lim_{n\to\infty} \frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}[(X_i-\mu_i)^2 \cdot \mathbf{1}_{\{|X_i-\mu_i|>\epsilon s_n\}}] = 0, \quad \text{where } s_n^2 = \sum_{i=1}^n \mathrm{Var}\, X_i$$

As we have that $\mu_i = \mathbb{E}[X_i] \le M, \forall i$, we have that $(X_i - \mu_i)^2 \le 4M^2$ a.s. Then, by Chebyshev's inequality (the version of $\mathbb{P}\left\{|X_i - \mu_i| \ge k\right\} \le \frac{\mathrm{Var}\, X_i}{k^2}$), we have $\forall \epsilon > 0$

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}[(X_i-\mu_i)^2 \cdot \mathbf{1}_{\{|X_i-\mu_i|>\epsilon s_n\}}] \le \frac{1}{s_n^2}\sum_{i=1}^n 4M^2 \cdot \mathbb{P}\left\{|X_i - \mu_i| \ge \epsilon s_n\right\} = \frac{1}{s_n^2}\frac{\sum_{i=1}^n 4M^2\,\mathrm{Var}\, X_i}{\epsilon^2 s_n^2} = \frac{4M^2}{s_n^2} \to 0$$

Therefore, by Lindberg-Feller CLT, $\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathrm{Var}\, S_n}} \xrightarrow{D} N(0,1)$

Problem 5 (Lyapunov's CLT)

Let $X_1, X_2, \ldots$ be independent random variables with zero means and unit variances. (Do not assume that X_i have the same distribution though.) Assume that

$$\sup_i \mathbb{E}\,|X_i|^{2+\delta} < \infty$$

for some $\delta > 0$. Prove that

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \to N(0,1) \quad \text{weakly.}$$

sol: $s_n = \sqrt{\sum_{i=1}^n \mathrm{Var}(X_i)} = \sqrt{n}$. Let $M = \sup_i \mathbb{E}|X_i|^{2+\delta}$, then $\forall \epsilon > 0$, when $|X_i| > \epsilon s_n$, we have $(\frac{|X_i|}{\epsilon\sqrt{n}})^\delta > 1$. Hence,

$$\frac{1}{n}\sum_{i=0}^n \mathbb{E}[|X_i|^2 \cdot \mathbf{1}_{\{|X_i|>\epsilon\sqrt{n}\}}] < \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(\frac{|X_i|}{\epsilon\sqrt{n}})^\delta |X_i|^2 \cdot \mathbf{1}_{\{|X_i|>\epsilon\sqrt{n}\}}]$$
$$= \frac{1}{\epsilon^\delta n^{1+\delta/2}}\sum_{i=1}^n \mathbb{E}[|X_i|^{2+\delta/2} \cdot \mathbf{1}_{\{|X_i|>\epsilon\sqrt{n}\}}] \le \frac{nM}{\epsilon^\delta n^{1+\delta/2}} \to 0, \text{ as } \delta > 0$$

Therefore, we obtain the condition for Lindbery-Feller CLT, so by the theorem, $\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$

Problem 6 (Bayes formula)

Let (Ω, Σ, P) be a probability space and $\mathcal{F} \subset \Sigma$ be a sub-σ-algebra. Consider two events $E \in \Sigma$ and $F \in \mathcal{F}$.

(a) Check that

$$P(F|E) = \frac{\mathbb{E}\left[P(E|\mathcal{F})\mathbf{1}_F\right]}{\mathbb{E}\, P(E|\mathcal{F})}.$$

(b) Specialize this equation to the case where $\mathcal{F}$ is generated by a partition $\Omega = F_1 \sqcup \cdots \sqcup F_n$, i.e. $\mathcal{F} = \sigma(F_1, \ldots, F_n)$. Deduce Bayes formula in its familiar form:

$$P(F_i|E) = \frac{P(E|F_i)\, P(F_i)}{\sum_i P(E|F_i)\, P(F_i)}.$$

sol: (a). $\frac{\mathbb{E}[P(E|\mathcal{F})\cdot\mathbf{1}_F)]}{\mathbb{E}[P(E|\mathcal{F})]} = \frac{\mathbb{E}[\mathbb{E}[\mathbf{1}_E|\mathcal{F}]\cdot\mathbf{1}_F]}{\mathbb{E}[\mathbb{E}[\mathbf{1}_E|\mathcal{F}])]} = \frac{\mathbb{E}[\mathbb{E}[\mathbf{1}_E\mathbf{1}_F|\mathcal{F}]]}{P(E)} = \frac{\mathbb{E}[\mathbb{E}[\mathbf{1}_{E\cap F}|\mathcal{F}]]}{\mathbb{P}\left\{E\right\}} = \frac{\mathbb{P}\left\{E\cap F\right\}}{\mathbb{P}\left\{E\right\}} = P(F|E)$

(b). Similar to (a), we can also get that $\mathbb{P}\{E|F_i\} = \frac{\mathbb{P}\left\{E\cap F_i\right\}}{\mathbb{P}\left\{F_i\right\}}$. Then, $\mathbb{P}\{F_i|E\} = \frac{\mathbb{E}[\mathbb{P}\left\{E|\mathcal{F}\right\}\cdot\mathbf{1}_{F_i}]}{\mathbb{E}P(E|\mathcal{F})}$

$$= \frac{\mathbb{P}\{E\cap F_i\}}{\mathbb{E}[\sum_i \mathbf{1}_{F_i}\cdot\mathbb{E}[\mathbf{1}_E|\mathcal{F}]]} = \frac{\mathbb{P}\{E|F_i\}\mathbb{P}\{F_i\}}{\sum_i \mathbb{E}[\mathbb{E}[\mathbf{1}_{F_i}\mathbf{1}_E|\mathcal{F}]} = \frac{\mathbb{P}\{E|F_i\}\mathbb{P}\{F_i\}}{\sum_i \mathbb{P}\{E\cap F_i\}} = \frac{\mathbb{P}\{E|F_i\}\mathbb{P}\{F_i\}}{\sum_i \mathbb{P}\{E|F_i\}\mathbb{P}\{F_i\}}$$

Problem 7 (Conditional Cauchy-Schwarz inequality)

Show that

$$(\mathbb{E}[XY|\mathcal{F}])^2 \le \mathbb{E}[X^2|\mathcal{F}]\cdot\mathbb{E}[Y^2|\mathcal{F}]$$

almost surely.

sol: For $\forall\theta$, we have $0 \le \mathbb{E}[(X+\theta Y)^2|\mathcal{F}] = \mathbb{E}[X^2|\mathcal{F}] + 2\theta\mathbb{E}[XY|\mathcal{F}] + \theta^2\mathbb{E}[Y^2|\mathcal{F}]$. Which means this quadratic equation w.r.t. θ has at most on root. Hence, $\Delta = (2\mathbb{E}[XY|\mathcal{F}])^2 - 4(\mathbb{E}[X^2|\mathcal{F}])(\mathbb{E}[Y^2|\mathcal{F}]) \le 0$. Therefore, we have that, $(\mathbb{E}[XY|\mathcal{F}])^2 \le \mathbb{E}[X^2|\mathcal{F}]\mathbb{E}[Y^2|\mathcal{F}]$

Problem 8 (Law of total variance)

Define conditional variance of X by

$$\mathrm{Var}(X|\mathcal{F}) := \mathbb{E}[X^2|\mathcal{F}] - (\mathbb{E}[X|\mathcal{F}])^2.$$

Show that

$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X|\mathcal{F})) + \mathrm{Var}(\mathbb{E}[X|\mathcal{F}]).$$

sol: LHS=

$$\mathbb{E}\Big[\mathbb{E}[X^2|\mathcal{F}] - (\mathbb{E}[X|\mathcal{F})^2]\Big] + \mathbb{E}\Big[(E[X|\mathcal{F}])^2\Big] - (\mathbb{E}[\mathbb{E}[\mathbb{X}|\mathcal{F}]])^2$$

since $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[X]$

$$= E[X^2] - E[(\mathbb{E}[X|\mathcal{F}])^2] + (E[(\mathbb{E}[X|\mathcal{F}])^2]) - (E[X])^2$$

$$= \mathbb{E}[X^2] - E[X^2] = \mathrm{Var}\, X$$

Problem 9 (Conditioning always reduces second moment)

Let $Y := \mathbb{E}[X|\mathcal{F}]$. Show that if $\mathbb{E}(Y^2) = \mathbb{E}(X^2)$ then $X = Y$ a.s.

sol: Consider that $(X - \mathbb{E}[X|\mathcal{F}]) = (X - Y)$. We have that $\mathrm{Var}(X - Y)$

$$= \mathbb{E}\Big[\mathbb{E}[(X-Y)^2|\mathcal{F}] - (\mathbb{E}[X-Y|\mathcal{F}])^2\Big] - \mathbb{E}((\mathbb{E}[X-Y|\mathcal{F}])^2$$

$$= \mathbb{E}[X^2] + \mathbb{E}[Y^2] - \mathbb{E}[2Y\mathbb{E}[X|\mathcal{F}]]$$

since $Y = \mathbb{E}[X|\mathcal{F}]$ and $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$, we arrive

$$\mathrm{Var}\,(X - Y) = 2\mathbb{E}[Y^2] - 2\mathbb{E}[Y^2] = 0$$

. Then, we have that $Y - X$ is a constant a.s. Since $\mathbb{E}[X - Y] = 0$, we have $X = Y$ a.s.