SDS 387, Fall 2024
Homework 2

Due October 3, by midnight on Canvas.

1. Show that it X_n and Y_n are independent for all n and $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, then

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ Y \end{bmatrix},$$

where X and Y are independent.
Use characteristic functions. Assume for simplicity that the X_n's and the Y_n's are univariate random variables, though the same argument is directly applicable to the vector case. Let $t = (t_1, t_2)^\top \in \mathbb{R}^2$ and let $\phi_{(X_n,Y_n)}(t)$ be the (joint) characteristic function of (X_n, Y_n) at t. Then, by independence,

$$\phi_{(X_n,Y_n)}(t) = \phi_{X_n}(t_1)\phi_{Y_n}(t_2) \to \phi_X(t_1)\phi_Y(t_2) = \phi_{(X,Y)}(t).$$

Since this is true for any choice of $t \in \mathbb{R}^2$ the result follows from the Continuity Theorem for characteristic functions.

2. In class we showed that, if $X_n \xrightarrow{d} X$ and $Y_n - X_n \xrightarrow{d} 0$ then $Y_n \xrightarrow{d} X$. Use this result to prove that, if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ for some constant, then

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ c \end{bmatrix}.$$

Note that X_n and Y_n are not necessarily independent.
By problem 1., $\begin{bmatrix} X_n \\ c \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ c \end{bmatrix}$. Next,

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} - \begin{bmatrix} X_n \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ Y_n - c \end{bmatrix} \xrightarrow{d} 0.$$

The conclusion follows from the result we showed in class.

3. Consider the settings of the above problem. Prove the following results, known together as Slutsky's theorem:

$$X_n Y_n \xrightarrow{d} Xc \quad \text{and} \quad X_n + Y_n \xrightarrow{d} X + c.$$

This follows directly from the continuous mapping theorem since addition $(x, y) \mapsto x+y$ and multiplication $(x, y) \mapsto xy$ are continuous functions.

4. **Polya's Theorem**. Let $\{X_n\}$ be a sequence of random variables in $\mathbb{R}$ converging to X, a random variable with a continuos c.d.f. F_X. Show that

$$\lim_n \sup_{x \in \mathbb{R}} |F_{X_n}(x) - F_X(x)| = 0,$$

where F_{X_n} is the c.d.f of X_n. The above result says that if X is continuous, then the convergence of the c.d.f.'s is uniform over $\mathbb{R}$, not just point-wise. You may (though you do not need to) proceed as follows.

(a) Let $\epsilon \in (0,1)$ be arbitrary (small). Next, let $-\infty = x_0 < x_1 < \ldots, x_k < x_{k+1} = \infty$ be such that $F(x_i) - F(x_{i-1}) \leq \epsilon$ for all $i = 1, \ldots, k+1$. This is possible. Why? Let $k+1 = \lceil \frac{1}{\epsilon} \rceil$ and, for $i = 1, \ldots, k$, let $x_i = F_X^{-}(i/(k+1)) = \inf\{c \in \mathbb{R} \colon F_X(c) \geq i/(k+1)\}$. Then $F_X(x_i) - F_X(x_{i-1}) = \frac{1}{k+1} \leq \epsilon$. This construction is possible since F is a continuous c.d.f..

(b) For any $x \in \mathbb{R}$ there exists one $i \in \{1, \ldots, k\}$ such that $x \in (x_{i-1}, x_i]$. Show that $F_{X_n}(x) - F_X(x) \leq F_{X_n}(x_i) - F_X(x_i) + \epsilon$ and that $F_{X_n}(x) - F_X(x) \geq F_{X_n}(x_{i-1}) - F_X(x_{i-1}) - \epsilon$. Conclude that

$$\sup_{x \in \mathbb{R}} |F_{X_n}(x) - F_X(x)| \leq \max_{i=0,\ldots,k} |F_{X_n}(x_i) - F_X(x_i)| + \epsilon.$$

For any x, let i be the unique index such that $x \in (x_{i-1}, x_i]$. Then

$$F_{X_n}(x) - F_X(x) \leq F_{X_n}(x_i) - F_X(x_{i-1}) \leq +F_{X_n}(x_i) - F_X(x_i) + \epsilon$$

and

$$F_{X_n}(x) - F_X(x) \geq F_{X_n}(x_{i-1}) - F_X(x_i) \geq +F_{X_n}(x_{i-1}) - F_X(x_{i-1}) - \epsilon.$$

This proves that, for any $x \in \mathbb{R}$,

$$\sup_{x \in \mathbb{R}} |F_{X_n}(x) - F_X(x)| \leq \max_{i=0,\ldots,k} |F_{X_n}(x_i) - F_X(x_i)| + \epsilon.$$

(c) Deduce the result from the inequality above. From the last expression, taking limits and using the fact that $X_n \xrightarrow{d} X$ and $\max_{i=0,\ldots,k} |F_{X_n}(x_i) - F_X(x_i)| \to 0$ because k is finite and independent of n,

$$\lim_n \sup_{x \in \mathbb{R}} |F_{X_n}(x) - F_X(x)| \leq \epsilon.$$

Since ϵ is arbitrary positive number, $\lim_n \sup_{x \in \mathbb{R}} |F_{X_n}(x) - F_X(x)| = 0$.

5. **Some O_P and o_P calculus**.

(a) Show that $O_p(1) + O_p(1) = O_P(1)$.

If $X_n = O_P(1)$ and $Y_n = O_P(1)$ then, for every $\epsilon \in (0,1)$ there exists $M > 0$ such that $\mathbb{P}(|X_n| > M) < \epsilon$ and $\mathbb{P}(|Y_n| > \epsilon) < \epsilon$, eventualy. Thus, eventually,

$$\mathbb{P}(|X_n + Y_n| > 2M) \leq \mathbb{P}(|X_n| > M) + \mathbb{P}(|Y_n| > M) < 2\epsilon.$$

(b) Show that $o_p(1) + o_p(1) = o_P(1)$.
If $X_n = o_P(1)$ and $Y_n = o_P(1)$ then, for every $\epsilon \in (0,1)$, $\mathbb{P}(|X_n| > \epsilon) < \epsilon$ and $\mathbb{P}(|Y_n| > M) < \epsilon$, eventualy. Thus, eventually,

$$\mathbb{P}(|X_n + Y_n| > 2\epsilon) \leq \mathbb{P}(|X_n| > \epsilon) + \mathbb{P}(|Y_n| > \epsilon) < 2\epsilon.$$

(c) Show that $O_P(1)o_p(1) = o_p(1)$. (Note that we can equivalently write this as $O_P(o_p(1)) = o_P(O_P(1)))$.
If $X_n = O_P(1)$ and $Y_n = o_P(1)$ the, for any every $\epsilon \in (0,1)$ there exists $M > 0$ such that $\mathbb{P}(|X_n| > M) < \epsilon$ and $\mathbb{P}(|Y_n| > \epsilon/M) < \epsilon$, eventually. Thus, eventually,

$$\mathbb{P}(|X_n Y_n| > \epsilon) \leq \mathbb{P}(|X_n| > M) + \mathbb{P}(|Y_n| > \epsilon/M)n < 2\epsilon.$$

(d) If $X_n = o_p(1)$, can we conclude that $X_n = O_P(1)$? Explain. Yes!

(e) What can you say about the asymptotic behavior of the stochastic quantity $\frac{1}{O_P(1)}$?
If $X_n = O_P(1)$, we cannot conclude that $1/X_n = O_P(1)$ because it may very well be the case that $X_N = o_P(1)$ (see last point); i.e., $1/X_n$ could diverge. But we can say that $1/X_n$ does not converge in probability to 0, otherwise

6. Give an example of a sequence of independent, centered random variables $X_1, X_2, \ldots$, all with unit variances, such that $\sqrt{n}\overline{X}_n$ does not converge in distribution to $N(0,1)$. *Hint: Construct a sequence of independent centered random variables such that the probability that $X_n = 0$ converges to 1 exponentially.*
Following the hint, let $X_1, X_2, \ldots$ be independent random varibles such that, for each i, X_i takes values 0 and $\pm e^{i/2}$ with probability $1 - 1/e^i$ and $1/(2e^i)$, respectively. Then, for all i, $\mathbb{E}[X_i] = 0$ and $\mathbb{V}[X_i] = 1$. However, $\sqrt{n}\overline{X}_n$ does not converge to any distribution with a continuous c.d.f. at 0 - in particular, it will not converge to a Gaussian. To see this, it is sufficient to show that

$$\lim_n \mathbb{P}(\sqrt{n}\ \overline{X}_n = 0) = \lim_n \mathbb{P}(S_n = 0) > 0,$$

where $S_n = \sum_{i=1}^n X_i$. This can be established as follows:

$$\begin{aligned}
\lim_n \mathbb{P}(S_n = 0) &\geq \lim_n \mathbb{P}(X_i = 0, \text{ all } i) \\
&= 1 - \lim_n \mathbb{P}(\cup_i \{X_i \neq 0\}) \\
&\geq 1 - \sum_{i=1}^{\infty} \mathbb{P}(X_i \neq o) \\
&= 1 - \sum_{i=1}^{\infty} e^{-1} \\
&\geq 1 - \sum_{i=0}^{\infty} e^{-1} \\
&= \frac{1-e}{1+e} > 0,
\end{aligned}$$

where the first identity is De Morgan's law and the second inequality follows from the union bound.

7. Let $Y_1, Y_2, \ldots$ be i.i.d. with mean zero and unit variance and let $X_k = \sigma_k Y_k$. Show that the LF condition in this case reduces to

$$\lim_n \frac{\max_{k=1,\ldots,n} \sigma_k^2}{\sum_{i=1}^n \sigma_k^2} = 0$$

Let $B_n^2 = \sum_{k=1}^n \sigma_k^2$. The (LF) condition is, for any $\epsilon > 0$,

$$\begin{aligned}
\frac{1}{B_n^2} \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbb{1}_{\{|X_k| \geq \epsilon B_n\}}] &= \frac{1}{B_n^2} \sum_{k=1}^n \sigma_k^2 \mathbb{E}[Y_1^2 \mathbb{1}_{\{|Y_1| \geq \epsilon B_n / \sigma_k\}}] \\
&\leq \frac{1}{B_n^2} \sum_{k=1}^n \sigma_k^2 \sqrt{\mathbb{P}\left(|Y_1| \geq \epsilon B_n / (\max_k \sigma_k)\right)} \\
&= \sqrt{\mathbb{P}\left(|Y_1| \geq \epsilon B_n / (\max_k \sigma_k)\right)} \\
&\leq \frac{1}{\epsilon} \sqrt{\frac{\max_{k=1,\ldots,n} \sigma_k^2}{\sum_{i=1}^n \sigma_k^2}} \\
&\to 0,
\end{aligned}$$

where we have used Cauchy-Schwarts in the second

8. Read the proofs of Theorem 1 and 2 in the paper *Variable selection via nonconcave penalized likelihood and its oracle properties,* by J. Fan and R. Li, Journal of American Statistical Association, 2001, 96, 1348-1360. This will show you how O_P and o_P notation is useful. Available here.

9. **Optional reading assignment.** In class, we saw an example of why the triangular array setup is desirable for proving CLTs when the data-generating distribution is not fixed and may change with n. Here is an example from the literature: Lemma 6 of the paper Hypothesis Testing For Densities and High-Dimensional Multinomials: Sharp Local Minimax Rates by S. Balakrishnan and L. Wasserman.