Each problem is worth 10 points and only five randomly chosen problems will be graded if there are more than 5 problems. Please indicate whom you worked with, it will not affect your grade in any way.

1. (Characterization of Normal rvs) Let $X_1, X_2, \ldots, X_k$ be i.i.d. r.v. with finite second moment, such that $(X_1 + X_2 + \cdots + X_k)/\sqrt{k} \overset{d}{=} X_1$ for some $k > 1$, i.e., the distribution of $(X_1 + X_2 + \cdots + X_k)/\sqrt{k}$ is the same as the distribution of $X_1$. Prove that $X_1$ must be a Normal rv with mean 0.

2. (Total variation distance.) Given two probability measure $\mu, \nu$ on $\mathbb{R}$ we define the total variation distance as

$$d_{TV}(\mu, \nu) = \sup_A |\mu(A) - \nu(A)| \in [0, 1].$$

(i) Show that this is indeed a distance and strictly stronger than convergence in distribution.

(ii) If there exists a countable subset $D$ such that $\mu(D) = \nu(D) = 1$, show that

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{d \in D} |\mu(\{d\}) - \nu(\{d\})|.$$

(iii) For two probability measures $\mu, \nu$ with densities $f, g$, respectively show that

$$d_{TV}(\mu, \nu) = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| \, dx.$$

3. (Couplings.) A coupling between two measures $\mu, \nu$ on $\mathbb{R}$ is defined as a joint distribution of a random vector $(X, Y)$ such that marginally $X \sim \mu$ and $Y \sim \nu$. Consider two probability measures $\mu, \nu$ with densities $f, g$, respectively.

(i) Show that

$$d_{TV}(\mu, \nu) \leq \inf_{(X,Y)} \mathbb{P}(X \neq Y)$$

(ii) Let $X, Y$ be r.v.s with CDFs $F, G$, respectively. Let $I$ be a Bernoulli($p$) r.v. independent of $X, Y$. Show that the random variable $IX + (1 - I)Y$ has CDF $pF + (1 - p)G$.

(iii) Prove that the inequality in (i) is an equality, so that there exists a coupling for which $\mathbb{P}(X \neq Y) = d_{TV}(\mu, \nu)$.

Hint: For the function $h = \min(f, g)$, by 2.(iii) the distance is $(1 - \int h(x) \, dx)$.

4. (The Lévy Metric.) Show that

$$\rho(F, G) = \inf \{\varepsilon : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \text{ for all } x\}$$

defines a metric on the space of distributions and $\rho(F_n, F) \to 0$ if and only if $F_n \xrightarrow{(d)} F$.

5. (Kolmogorov-Smirnov distance.) i) Prove that, if $F_n \xrightarrow{(d)} F$ and $F$ is continuous everywhere, then $d_{KS}(F_n, F) \to 0$, where the distance

$$d_{KS}(F, G) = \sup_x |F(x) - G(x)|$$

is called the Kolmogorov-Smirnov distance.

ii) Prove that, convergence in Kolmogorov-Smirnov distance is strictly weaker than convergence in total-variation distance.
6. **Converging together lemma.** (i) If \( X_n \Rightarrow X, Y_n \Rightarrow c \) where \( c > 0 \) is a constant and \( X_n, Y_n \) are defined on the same probability space, then \( X_n + Y_n \Rightarrow X + c \) and \( X_n Y_n \Rightarrow cX \). (We assumed \( c > 0 \) only to make the proof simpler.)

(ii) If \( X_1, X_2, \ldots \) are i.i.d. with \( \mathbb{E} X_1 = 0, \text{Var}(X_1) < \infty \), using SLLN, CLT and part (i) prove that

\[
\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{X_1^2 + X_2^2 + \cdots + X_n^2}} \Rightarrow N(0,1).
\]

This is called **Self-normalized Central Limit theorem.**

7. **CLT using Characteristic functions.** (a) Let \( \phi_n(t) \) be the characteristic function of \( X_n \) and \( \phi(t) \) be the characteristic function of \( X \). Prove that \( \phi_n(t) \to \phi(t) \) as \( n \to \infty \) for all \( t \in \mathbb{R} \), implies that \( X_n \Rightarrow X \).

(b) For a r.v. \( X \) with \( \mathbb{E} X = 0, \mathbb{E} X^2 = \sigma^2 < \infty \) show that

\[
\mathbb{E}(e^{itX}) = 1 - \frac{1}{2} t^2 \sigma^2 + o(t^2) \text{ as } t \to 0.
\]

(c) Use the above results to give a proof of the basic Central Limit Theorem: If \( X_1, X_2, \ldots \) are i.i.d. with mean 0 and Variance 1, then,

\[
\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \xrightarrow{(d)} N(0,1).
\]