Homework 2

Due Monday, February 6, 2016

Exercise 1. *Let P and Q be probability measures on $\mathbb{R}^d$ with finite second moments (i.e. if $X \sim P$ and $Y \sim Q$ then $E|X|^2 < \infty$ and $E|Y|^2 < \infty$). The L_2 Wasserstein distance d_W between P and Q is defined as*

$$d_W^2(P,Q) := \inf_{\mathcal{L}(X,Y)\in\Pi(P,Q)} E|X-Y|^2$$

where $\Pi(P,Q)$ is the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals given by P and Q. In particular, if X and Y are two d-dimensional random vectors, defined on the same probability space, then the distribution of the 2d-dimensional random vector (X,Y) is in $\Pi(P,Q)$ if and only if $X \sim P$ and $Y \sim Q$.

Show that there exists d-dimensional random vectors X^* and Y^*, definied on the same probability space, such that $\mathcal{L}(X^*,Y^*) \in \Pi(P,Q)$ and

$$d_W^2(P,Q) := E|X^* - Y^*|^2.$$

The Central Limit Theorem we showed in class can be generalized in many ways. The following generalization considers a triangular array of random variables which allow the variances of the random variables in the sum to vary with n.

Claim 1 (Lindeberg-Feller)**.** *Suppose $X_{i,n}$, for $n \in \mathbb{N}$ and $i \le n$, forms a triangular array of independent random variables which satisfy the following conditions*

1. $EX_{i,n} = 0$ *for all* n *and* $i \le n$;
2. $\sum_{i=1}^n EX_{i,n}^2 = 1$ *for all* n;
3. $\sum_{i=1}^n E(X_{i,n}^2 I_{\{|X_{i,n}|\ge\delta\}}) \to 0$ *as* $n \to \infty$ *for all* $\delta > 0$.

Then

$$\sum_{i=1}^n X_{i,n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

In fact, Lindeberg's conditions are also necessary in the following sense.

Claim 2 (Lindeberg-Feller)**.** *Suppose $X_{i,n}$, for $n \in \mathbb{N}$ and $i \le n$, forms a triangular array of independent random variables which satisfy the following conditions*

1. $EX_{i,n} = 0$ *for all* n *and* $i \le n$;
2. $\sum_{i=1}^n EX_{i,n}^2 = 1$ *for all* n;
3. $\max_{1\le i\le n} EX_{i,n}^2 \to 0$;
4. $\sum_{i=1}^n X_{i,n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Then

$$\sum_{i=1}^n E(X_{i,n}^2 I_{\{|X_{i,n}|\ge\delta\}}) \xrightarrow{n\to\infty} 0$$

for all $\delta > 0$.

We will not cover Claim 2 in class since the standard proof requires Fourier analysis which we want to avoid.

Exercise 2. *Show Claim 1 by extending the Lindeberg method we used in class to prove the Central Limit Theorem.*

Exercise 3. *Give a counterexample to the conjecture: If $X_1, X_2, \ldots$ are independent random variables with $EX_n = 0$ and $EX_n^2 = 1$ for all n, then $(X_1 + \cdots + X_n)/\sqrt{n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.*

Hint: *Consider X_n to take values $-v_n, 0, v_n$ with probabilities $p_n/2, 1-p_n, p_n/2$. You are free to use Claim 2 (without proof) in your solution.*

Exercise 4. *Let $Z_1, Z_2, \ldots$ be iid random variables with density $\delta(z)$ with respect to Lebesque measure. Think of each Z_n recording the time it takes a random person to run a mile. Now let*

$$X_n := \begin{cases} 1 & \text{if } Z_n > \max(Z_1,\ldots,Z_{n-1}) \\ 0 & \text{otherwise.} \end{cases}$$

In other words, X_n is a 1 if the n^{th} random person breaks a speed record. Show

$$\frac{S_n - E(S_n)}{\sqrt{var(S_n)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where $S_n := X_1 + \cdots + X_n$. Hint: *Start by arguing that the X_i's are independent Bernoulli random variables. You can use the following fact to find $P(X_n = 1)$:*

$$P(Z_n > Z_1, \ldots, Z_n > Z_{n-1}) = \int_{\mathbb{R}} P(z > Z_1, \ldots, z > Z_{n-1})\delta(z)dz.$$