Homework 2

STAT W4413: Nonparametric Statistics

DUE: Tuesday, February 9, 12:00 noon

(1) Please sign your home work with your name and UNI number.

(2) Homework must be submitted into the Statistics Homework Boxes room 904 on the 9th floor of SSW building.

(3) Homework is due Tuesday, February 9, 12:00 noon.

(4) No late homework, under any circumstances, will be accepted.

(5) At the end of semester, one of your lowest homework scores will be dropped before the final grade is calculated.

(6) Your submitted solutions should consist of (i) the hand written (or printout) of the results with all the details.

1. Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. drawn from $F(x)$. Let $a \in \mathbb{R}$. In the discussion of "goodness of fit" measures we deal with random variables of the form

$$Y_n = \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq a) \right).$$

Characterize the limiting distribution of Y_n (Characterize the distribution $\{Y_n\}_{n=1}^{\infty}$ converges to). Furthermore, characterize the limiting distribution of W_n defined as

$$W_n = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq a) - F(a) \right).$$

Hint: First use the Weak Law of Large Numbers to find the limit in probability of $\{Y_n\}_{n=1}^{\infty}$, then conclude using the fact that convergence in probability implies convergence in distribution.

Hint 2: Find the mean and variance of $\mathbb{I}(X_i \leq a)$, then use the Central Limit Theorem to derive the limiting distribution of $\{W_n\}_{n=1}^{\infty}$

2. Study the proof of Slutsky's theorem for $X_n + Y_n \xrightarrow{d} X + c$ from the slides. Then prove the following:

   (a) Why do Equalities or Inequalities (a), (b), (c), and (d) hold?

   (b) Use conditioning argument similar to that in the slides to show that

$$\mathbb{P}(X_n + Y_n \leq t) \geq \mathbb{P}(X_n \leq -c - \epsilon + t) - \mathbb{P}(|Y_n - c| \geq \epsilon).$$

3. Prove that if $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $X_n + Y_n \xrightarrow{p} X + Y$.
   Hint: Use an argument similar to the argument you used for the last problem, a triangle inequality: $|X_n + Y_n - X - Y| \leq |X_n - X| + |Y_n - Y|$, and the fact that $\mathbb{P}(|X_n - X| + |Y_n - Y| \geq \varepsilon) \leq \mathbb{P}(|X_n - X| \geq \varepsilon/2) + \mathbb{P}(|Y_n - Y| \geq \varepsilon/2)$

4. Here you prove a special form of the continuous mapping theorem. Suppose that $\{X_i\}_{i=1}^{\infty}$ converges to X in distribution. Let f be an increasing continuous function. Prove that $\{f(X_i)\}_{i=1}^{\infty}$ converges in distribution to $f(X)$.
   Hint: Use the fact that for a strictly increasing function the inverse function is well defined.

5. A population of 10^8 voters choose between two candidates A and B. A fraction 0.5005 of them plan to vote for candidate A and the rest for candidate B. A fair poll with sample size n is performed, i.e., the n samples are i.i.d. and done with replacement (same person may be polled more than once). Find a good estimate of n such that the probability that candidate A wins is greater than 0.99.
   Hint: Define a random variable $X_i = 1$ if i^{th} voter votes for candidate A, $X_i = 0$ if i^{th} voter votes for candidate B; and use Central Limit Theorem.

6. Let $\{X_i\}_{i=1}^{\infty}$ be iid with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = 1$. Define the random variables

$$Y_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\sum_{i=1}^{n} \frac{1}{n}(X_i - \bar{X})^2}}$$

   Characterize the limiting distribution of Y_n.
   Hint: Study the hints provided in the slides about the t-dsitribution.