Homework 4
STAT W4413: Nonparametric Statistics

DUE: Friday, April 1, 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 Friday, April 1, 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, (ii) printout of the relevant figures (if applicable), and (iii) the printout of the source code (if applicable).

1. (20p) In this problem we would like to use the χ^2 test to check the independence of two random variables. Suppose that we have two sets of random variables, $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, where $X_i, Y_i \in \{-1, 0, 1\}$. Our null hypothesis is that these two random variables are independent.

   (a) Write down the independence hypothesis as a composite null hypothesis on the joint distribution of (X, Y).
   Hint: X_i and Y_i are independent if and only if their joint probability mass function is equal to the product of the univariate probability mass functions, i.e., for all $n, m \in \{-1, 0, 1\}$

$$P(X_i = m, Y_i = n) = P(X_i = m)P(Y_i = n)$$

Hint: Use the fact that $P(X_i = -1) + P(X_i = 0) + P(X_i = 1) = 1$ and $P(Y_i = -1) + P(Y_i = 0) + P(Y_i = 1) = 1$ to reduce the number of parameters in your composite null hypothesis to four parameters.

(b) What is the χ^2 statistic for this test? What are the limiting distribution's degrees of freedom?

Hint: Calculate the MLE under the null hypothesis.

(c) In the table below we have summarized the rating people provided for two different products. The first location (top-left) in the table specifies the number of people that did not like any of the two products for instance. Run a χ^2 test to check if the ratings are independent. Calculate the p-value.

	$Y = -1$	$Y = 0$	$Y = 1$
$X = -1$	1	4	10
$X = 0$	2	3	16
$X = 1$	4	10	50

2. (20p) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_X$, where F_X is the cumulative distribution function. In Homework 1, you proved that $F_X(X_i)$ is uniformly distributed between $[0, 1]$. Here we consider a different problem. Suppose that we order these n random variables to obtain $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. We also define $Y_i = F_X(X_{(i)})$. What is the joint probability density function of $Y_1, Y_2, \ldots, Y_n$?

Hint: Explain why Y_i are ordered too and use this to find the joint pdf of $Y_1, Y_2, \ldots, Y_n$.

3. (20p) Consider the KS test for $n = 5$. Our goal is to characterize the distribution of K through Monte Carlo simulation. For this purpose, we need 10000 iid samples for K. Explain how you can generate such samples. Use these samples to obtain the empirical distribution of K and plot your distribution. What is the connection between what you are doing and MC simulation.

4. (20p) Check the calculation of Cramer-Von Mises test in Lecture 11 Slides 15 & 16. Try to derive the equations yourself. Describe:

(a) How you derive Equality (a).

(b) How you derive Equality (b).

(c) Describe why under the null hypothesis the distribution of the Cramer-Von Mises statistic, C, is independent of (invariant to) the null distribution F_0.

Hint: Argument follows from the same result for the KS test.

(d) Consider $X_1 = 0.99, X_2 = 0.18, X_3 = 0.84, X_4 = 0.76, X_5 = 0.3$. Calculate the Cramer-Von Mises statistic for $F_0 = Unif[0, 1]$. Do NOT use any software for this part of the problem.

Hint: Use the formula we derived above and the fact that $F_0(X_{(i)}) = X_{(i)}$ for uniform distribution.

5. (20p) We denote the rank of $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$ with $r(X_1), r(X_2), \ldots, r(X_n)$, respectively. Also, let $X_{(1)}, X(2), \ldots, X_{(n)}$ be the ordered version of $X_1, X_2, \ldots, X_n$. Assume that F is a continuous function and calculate the following:

   (a) What is the joint distribution of $r(X_1), r(X_2), \ldots, r(X_n)$? Prove it.

   (b) $\mathbb{E}(r(X_i))$.

   (c) $\mathbb{E}(r^2(X_i))$.

   (d) The probability mass function of $r(X_2), \ldots, r(X_n)$ given that $r(X_1) = 1$.

   (e) Probability density function of $X_{(r(X_i))}$.