INDIVIDUAL TEST
S.-T YAU COLLEGE MATH CONTESTS 2012

Probability and Statistics

Please solve 5 out of the following 6 problems,
or highest scores of 5 problems will be counted.

1. Solve the following two problems:

1) An urn contains b black balls and r red balls. One of the balls was drawn at random, and putted back in the urn with a additional balls of the same color. Now suppose that the second ball drawn at random is red. What is the probability that the first ball drawn was black?

2) Let (X_n) be a sequence of random variables satisfying

$$\lim_{a\to\infty} \sup_{n\geq 1} P(|X_n| > a) = 0.$$

Assume that sequence of random variables (Y_n) converges to 0 in probability. Prove that (X_nY_n) converges to 0 in probability.

2. Solve the following two problems:

1) Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G}$ be a sub-algebra of $\mathcal{F}$. Assume that X is a non-negative integrable random variable. Set $Y = E[X|\mathcal{G}]$. Prove that

(a)$[X > 0] \subset [Y > 0]$,a.s.;

(b)$[Y > 0] = \text{ess.inf}\{A : A \in \mathcal{G},\ [X > 0] \subset A\}$.

2) Let X and Y have a bivariate normal distribution with zero means, variances σ^2 and τ^2, respectively, and correlation ρ. Find the conditional expectation $E(X|X+Y)$.

3. Suppose that $\{p(i,j) : i = 1, 2, \cdots, m; j = 1, 2, \cdots, n\}$ is a finite bivariate joint probability distribution, that is,

$$p(i,j) > 0,\ \sum_{i=1}^{m}\sum_{j=1}^{n} p(i,j) = 1.$$

(i) Can $\{p(i,j)\}$ be always expressed as

$$p(i,j) = \sum_k \lambda_k a_k(i) b_k(j)$$

for some finite $\lambda_k \geq 0,\ \sum_k \lambda_k = 1,\ a_k(i) \geq 0,\ \sum_{i=1}^m a_k(i) = 1,\ b_k(j) \geq 0,\ \sum_{j=1}^n b_k(j) = 1$?

(ii) Prove or disprove the above relation by use of conditional probability.

4. Let $X_1, \cdots, X_m$ be an independent and identically distributed (i.i.d.) random sample from a cumulative distribution function (CDF) F, and $Y_1, \cdots, Y_n$ an i.i.d. random sample from a CDF G. We want to test $H_0 : F = G$ versus $H_1 : F \neq G$. The total sample size is $N = m + n$. Consider the following two nonparametric tests.

- The Wilcoxon rank sum tests. The test proceeds by first ranking the pooled X and Y samples and then taking the sum of the ranks associated with the Y sample. Let $R_{y_1}, \cdots, R_{y_n}$ be the rankings of the sample $y_1 < \cdots < y_n$ from the pooled sample in increasing order. The Wilcoxon rank sum statistic is defined as $W = \sum_{j=1}^{n} R_{y_j}$.
- The Mann-Whitney U-test. Let $U_{ij} = 1$ if $X_i < Y_j$, and $U_{ij} = 0$ otherwise. The Mann-Whitney U-statistic is defined as $U = \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij}$. The probability $\gamma = P(X < Y)$ can be estimated as $U/(mn)$. The decision rule is based on assessing if $\gamma = 0.5$.

Assume that there are no tied data values.

(a) Show that $W = U + \frac{1}{2}n(n+1)$, which shows that the two test statistics differ only by a constant and yield exactly the same p-values.

(b) Using the central limit theorem, the Wilcoxon rank sum statistic W can be converted to a Z-variable, which provides an easy-to-use approximation. The transformation is

$$Z_W = \frac{W - \mu_W}{\sigma_W},$$

where μ_W and σ_W^2 are the mean and variance of W under H_0. Show that $\mu_W = \frac{1}{2}n(N+1)$ and $\sigma_W^2 = \frac{1}{12}mn(N+1)$.

5. Let X be a random variable with $EX^2 < \infty$, and $Y = |X|$. Assume that X has a Lebesgue density symmetric about 0. Show that random variables X and Y are uncorrelated, but they are not independent.

6. Let $E_1, \cdots, E_n$ be i.i.d. random variables with $E_i \sim \text{Exponential}(1)$. Let $U_1, \cdots, U_n$ be i.i.d. uniformly (on [0,1]) distributed random variables. Further, assume that $E_1, \cdots, E_n$ and $U_1, \cdots, U_n$ are independent.

(a) Find the density of $X = (E_1 + \cdots + E_m)/(E_1 + \cdots + E_n)$, where $m < n$.

(b) Show that $Y = \frac{(n-m)X}{m(1-X)}$ is distributed as the F-distribution with degrees of freedom $(2m, 2(n-m))$

(c) Find the density of $(U_1 \cdots U_n)^{-X}$.