Master's Exam - Fall 2006

October 5, 2006
1:00 pm – 5:00 pm

NAME:____________________________________

A. The number of points for each problem is given.

B. There are 12 problems with varying numbers of parts.

1 – 5	Probability (115 points)
6 – 12	Statistics (140 points)

C. Write your answers on the exam paper itself. If you need more room you may use the extra sheets provided. Answer as many questions as you can on each part. Tables are provided.

D. Professors Sakhanenko and Stapleton will be in their offices (441 and 429) during most of the exam, with at least one being available at all times. One of them will come to the exam room every 30 minutes to see if there are any questions.

PROBABILITY PART.

1. For each of the following experiments and events determine the probability of the event. A deck of cards has 9 cards with numbers 1, 1, 2, 2, 2, 2, 2, 3, 3.

(a) (4 pts.) *Experiment:* Three cards are drawn randomly without replacement. *Event:* At least two of the cards chosen have numbers larger than one.

(b) (5 pts.) *Experiment:* Cards are drawn randomly without replacement until a card with the number 3 is chosen. *Event:* The first 3 is chosen on the 6th draw.

(c) (6 pts.) *Experiment:* 600 cards are drawn randomly with replacement. *Event:* The sum of the numbers on the 600 cards is larger than 1250. (Use an approximation).

(d) (7 pts.) *Experiment:* Same as in (c). *Event:* The number of 3's chosen is less than 160. (Use an approximation)

(e) (8 pts.) *Experiment:* Two cards are drawn randomly without replacement. This is repeated independently 72 times. *Event:* Two or more of the 72 trials result in 33's. (Both cards have 3's.) (Use an approximation)

2. Let $f(x) = 0.5x$ for $0 \leq x \leq 2$. Let X_1 and X_2 be a random sample from this density.

(a) (5 pts.) Given a random variable U, uniformly distributed on $[0, 1]$, how could such a random variable X_1 be generated from U?

(b) (5 pts.) Let $M = \max(X_1, X_2)$. Find the cdf for M.

(c) (8 pts.) Find the cdf and the density for $Y = (X_1 - 1)^2$.

(d) (6 pts.) Find the correlation coefficient $\rho(X_1, 2X_1 - X_2)$.

(e) (6 pts.) Let $W = X_1 X_2$. Find $E(W)$.

3. (a) (5 pts.) State the Weak Law of Large Numbers.

(b) (10 pts.) Use the Chebyshev's Inequality to prove the Weak Law of Large Numbers.

4. Let X be a random variable with the density $3x^2, 0 < x < 1$. Conditionally on $X = x$, let Y be uniformly distributed on $[0, x^2]$.

(a) (7 pts.) Find the marginal density of Y.

(b) (10 pts.) Find $Z = \mathbf{E}(X|Y)$.

(c) (10 pts.) Find the joint cdf of X and Y.

5. Let X be a Bernoulli random variable with $0 < p < 1$. Conditionally on $X = x$, let Y be Poisson random variable with the parameter $\lambda(x + 1)$, $\lambda > 0$.

(a) (8 pts.) Find the probability distribution function of Y.

(b) (5 pts.) Find $P(X = 0|Y = y), y = 0, 1, 2, ...$

STATISTICS PART.

6. Let $a \in R$ be a unknown parameter. Let $X_1, .., X_n$ be a random sample from the distribution with density $f(x) = e^{a-x}$ for $x \geq a$, 0 for $x < a$.

(a) (10 pts.) Find the maximum likelihood estimator of a.

(b) (10 pts.) Find the method of moments estimator of a.

(c) (6 pts.) Find the bias of the MLE of a.

7. The presence of excess protein in urine is a symptom of kidney distress among diabetics. Urinary protein was measured for 8 patients before and after nine weeks of captopril therapy. The following data correspond to the amounts of urinary protein in g/24 hours before and after therapy.

Before	17.0	16.2	10.4	9.2	7.9	5.8	5.4	5.7
After	5.7	5.6	3.4	6.5	0.7	6.1	6.7	2.9

(a) (12 pts.) State a parametric model, define hypotheses, and carry out a test at level $\alpha = 0.05$ which will enable you to decide whether the captopril therapy is effective in lowering the amount of urinary protein.

(b) (8 pts.) Perform a nonparametric test of the same hypotheses. Find the p-value for this test.

8. In order to estimate the change $\Delta = p_1 - p_2$ in the proportion of a population of 45 thousand MSU students who wanted a decrease in tuition between time t_1 and time t_2, random samples of sizes $n_1 = 500$ and $n_2 = 400$ were taken independently at the two times. Let X_1 and X_2 be the numbers favoring a decrease in tuition for the two samples.

(a) (8 pts.) Define notation and give a formula for a 95% confidence interval on Δ.

(b) (5 pts.) Apply the formula for the case $X_1 = 400, X_2 = 240$.

(c) (7 pts.) Explain the meaning of your interval in such a way that someone who had not ever studied statistics would understand.

(d) (8 pts.) If equal sample sizes, $n = n_1 = n_2$ were to be used how large must n be in order that the estimator $\hat{\Delta}$ have probability at least 0.99 of being within 0.05 of Δ?

9. Let $(X_1, X_2, X_3, X_4, X_5)$ be a random sample from the Bernoulli distribution with a unknown parameter $0 < p < 1$. Suppose that we wish to test $H_0 : p = 0.2$ vs $H_a : p > 0.2$.

(a) (5 pts.) Consider the test which rejects for $T = X_1 + ... + X_5 > 3$. What is the level of significance α?

(b) (8 pts.) Use a theorem to prove that this test is uniformly most powerful for this α level.

(c) (7 pts.) Find the power of this test for $p = 0.5$.

10. Let (x_i, Y_i) be observed for $i = 1, ..., n$. Suppose that the x_i are constants, and that $Y_i = \frac{\beta}{x_i} + \varepsilon_i$, where β is a unknown parameter, and the ε_i are independent, each with the $N(0, \sigma^2)$ distribution.

(a) (8 pts.) Show that the least squares estimator of β is $\hat{\beta} = (\sum_{i=1}^n Y_i x_i^{-1})/(\sum_{i=1}^n x_i^{-2})$. Do **NOT** use matrix or vector space methods.

(b) (5 pts.) Show that $\hat{\beta}$ is an unbiased estimator of β.

(c) (5 pts.) Find Var$(\hat{\beta})$.

(d) (8 pts.) For the following (x_i, Y_i) pairs find a 95% confidence interval on β: (1, 3), (2, 5), (1, 5), (2, 9).

11. (10 pts.) A jury panel included 12 men and 8 women. By a procedure described as "random" a jury of 6 was selected. The jury included just one woman. Find the p-value of a test of the null hypothesis that the selection of jurors was random against the alternative that there is discrimination against women in the selection of jurors.

12. (10 pts.) Let X_1, X_2, X_3 be a random sample from a continuous cdf $F(x)$. Let Y_1, Y_2, Y_3, Y_4 be a random sample from the cdf $G(y) = F(y - D)$. Suppose that you wish to test $H_0 : D = 0$, vs $H_a : D \neq 0$.

For X's: 23, 18, 15, and Y's: 27, 24, 21, 20 find the exact p-value for the Wilcoxon Rank Sum test of H_0 vs H_a.