Master's Exam - Spring 2006

March 16, 2006
1:00 pm – 5:00 pm

NAME:________________________________

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

B. There are 11 problems with varying numbers of parts.

1 – 5	Probability (115 points)
6 – 11	Statistics (120 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. Answers can be written in terms of standard normal cdf Φ, normal α-quantiles z_α ($P(Z \le z_\alpha) = \alpha$), t-quantiles t_α.

D. Professors Zuo, Sakhanenko and Stapleton will be in their offices (440, 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 10 cards with numbers 1, 1, 2, 2, 3, 3, 3, 4, 4, 4.

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

(b) (5 pts.) *Experiment:* Cards are drawn randomly without replacement until a card with the number 4 is chosen. *Event:* The first 4 is chosen on the 5th draw.

(c) (6 pts.) *Experiment:* 100 cards are drawn randomly with replacement. *Event:* The sum of the numbers on the 100 cards is larger than 300. (Use an approximation).

(d) (7 pts.) *Experiment:* Same as in (c). *Event:* The number of 4's chosen is less than 25. (Use an approximation)

(e) (8 pts.) *Experiment:* Two cards are drawn randomly without replacement. This is repeated independently 120 times. *Event:* Two or more of the 120 trials result in 44's. (Both cards have 4's.) (Use an approximation)

2. Let $f(x, y) = y + 4x$ for $0 < x < y < 1$. Let (X, Y) be a random vector from this density.

(a) (7 pts.) Find the cdf and the density for Y.

(b) (5 pts.) Let Y_1 and Y_2 be a random sample from the density in (a). Let $M = \max(Y_1, Y_2)$. Find the cdf for M.

(c) (6 pts.) Find the correlation coefficient $\rho(Y_1 - Y_2, Y_1 + Y_2)$.

(d) (7 pts.) Let $W = X/Y$. 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 $f(x) = 4x^3$ for $0 \le x \le 1$. Conditionally on $X = x$, let Y be uniformly distributed on $[0, x]$.

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

(b) (6 pts.) Given a random variable U, uniformly distributed on $[0, 1]$, how could a random variable X be generated from U?

(c) (7 pts.) Find $Z = \mathbf{E}(Y|X)$.

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

5. Consider the reliability diagram below. The system has 5 components numbered 1, 2, 3, 4, 5. Let A_k be the event that component k works as it should. Suppose that $A_1, ..., A_5$ are independent. Denote $p_k = P(A_k)$ for each k.

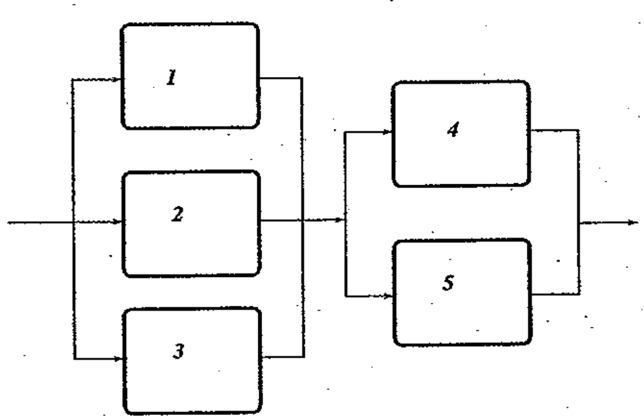

Figure 1: For problem 5.

(a) (5 pts.) Express the event B that the entire system works successfully in terms of A_k.

(b) (5 pts.) Express $P(B)$ in terms of p_k.

(c) (5 pts.) Express $P(A_1|B)$ in terms of p_k.

STATISTICS PART.

6. Let $\lambda > 0$ be a unknown parameter. Let $X_1, ..., X_n$ be a random sample from the distribution with density $f(x) = \lambda x e^{-\lambda x^2}$ for $x \geq 0$.

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

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

7. The following data correspond to a weight loss for 7 people. Each person was put on a low-carb diet and on a high-carb diet. The order of diets was determined randomly for each person.

Low-carb diet	2.8	2.3	3.1	3.0	2.4	3.6	3.7
High-carb diet	2.0	2.0	2.1	1.8	2.5	1.7	3.3

(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 low-carb diet is more effective than the high-carb diet for the weight loss.

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 $n_2 = 600$ 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 90% confidence interval on Δ.

(b) (5 pts.) Apply the formula for the case $X_1 = 250, X_2 = 350$.

(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 0.95 of being within 0.01 of Δ?

9. Let (X_1, X_2) be a random sample from the geometric distribution with unknown parameter $0 < p < 1$. Suppose that we wish to test $H_0 : p = 0.25$ vs $H_a : p > 0.25$.

(a) (5 pts.) Consider the test which rejects for $T = X_1 + X_2 < 3$. What is the level of significance α?

(b) (10 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 $Y_i = \beta x_i + \varepsilon_i$, where β is an 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 $\hat{\beta} = (\sum_{i=1}^{n} x_i Y_i)/(\sum_{i=1}^{n} x_i^2)$. 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 $\text{Var}(\hat{\beta})$.

(d) (8 pts.) For the following (x_i, Y_i) pairs find a 90% confidence interval on β: (1, 3), (2, 5), (3, 7), (4, 9).

11. (12 pts.) Let $X_1, ..., X_n$ be independent random variables from the distribution with density $f(x) = \frac{1}{\lambda} e^{-x/\lambda}, x > 0$. Consider the two estimators $\lambda^* = nX_{(1)}$ and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

(a) Show that λ^* is an unbiased estimator of λ.

(b) Determine the relative efficiency (relative size of variances) of λ^* to $\bar{X}$ and find its limiting value as $n \to \infty$.