Fall 2007

**PROBABILITY PART.**

1. For each of the following experiments and events determine the probability of the event. A deck of cards has 11 cards with letters A, B, R, A, C, A, D, A, B, R, A.

(a) (4 pts.) *Experiment:* Four cards are drawn randomly without replacement. *Event:* The cards spell C, A, R, D. (Order does not matter).

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

(c) (6 pts.) *Experiment:* 605 cards are drawn randomly with replacement. *Event:* The number of cards with the letter R is larger than 131. (Use an approximation).

(d) (7 pts.) *Experiment:* The experiment in (a) is repeated independently 99 times. *Event:* Only once the word "CARD" is chosen. (In this order. Use an approximation)

(e) (7 pts.) *Experiment:* The whole deck is randomly rearranged. *Event:* The rearrangement reads "ABRACADABRA". (In this order).

2. Let $f(x) = 2x$ for $0 \leq x \leq 1$. Let $X_1, X_2$, and $X_3$ 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) (10 pts.) Let $S = X_1 - X_2$. Find the cdf for $S$.

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

(d) (6 pts.) Find the correlation coefficient $\rho(X_2 + 2X_3, 5X_1 + 4X_2 - 3X_3)$.

(e) (6 pts.) Let $W = \frac{X_1^2}{X_2 X_3}$. Find $E(W)$.

3. (a) (3 pts.) Let $X_n$ have the cdf $F_n$ for $n = 1, 2, ...$ Let $X$ have the cdf $F$. Define: $X_n$ converges in distribution to $X$ as $n \to \infty$.

(b) (10 pts.) Let $Y_n$ have the cdf $F_n(x) = x^n$ for $0 \leq x \leq 1$, 0 for $x < 0$, 1 for $x > 1$, $n = 1, 2, ...$ Show that $Y_n$ converge in distribution as $n \to \infty$. Find the limiting distribution.

(c) (5 pts.) State the Central Limit Theorem.

4. Let $(X, Y)$ be a random vector with the joint density $1/\pi$, for $x^2 + y^2 < 1$.

(a) (7 pts.) Find the marginal density of $X$.

(b) (7 pts.) Find the conditional density of $Y$ given $X = x, -1 < x < 1$.

(c) (7 pts.) Find $Z = \mathbf{E}(Y^2|X)$.

(d) (4 pts.) Find $V = \mathbf{E}(X + Y^2|X)$.

5. Let $X$ have the density $3x^2$ on $[0, 1]$. Conditionally on $X = x$, let $Y = 2$ with probability $x$ and let $Y = -2$ with probability $1 - x$.

(a) (7 pts.) Find the probability mass distribution function of $Y$.

(b) (6 pts.) Find the conditional density of $X$ given $Y = -2$.

## STATISTICS PART.

6. Let $a$ be a unknown parameter. Let $X_1, .., X_n$ be a random sample from the density $2e^{-2(x-a)}$ on $(a, \infty)$.

(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 10 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 | 14.4 | 8.4 | 9.1 | 12.8 | 8.9 | 7.7 | 6.1 | 8.3 | 10.4 | 15.1 |
| After | 11.7 | 6.3 | 6.9 | 10.2 | 6.5 | 8.2 | 7.1 | 7.0 | 12.2 | 13.1 |

(a) (12 pts.) State a parametric model, define hypotheses, and carry out a test at level $\alpha = 0.01$ 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 with $\alpha = 0.01$. 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 $n_1 = 1000$ and $n_2 = 800$ 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 99% confidence interval on $\Delta$.

(b) (5 pts.) Apply the formula for the case $X_1 = 870, X_2 = 640$.

(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 0.02 of $\Delta$?

9. Let $(X_1, ..., X_9)$ be a random sample from the Normal distribution with a unknown mean $\mu$ and known standard deviation $\sigma = 1$. Suppose that we wish to test $H_0 : \mu = 2$ vs $H_a : \mu > 2$.

(a) (5 pts.) Consider the test which rejects for $T = X_1 + ... + X_9 > 25.5$. What is the level of significance $\alpha$?

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

(c) (7 pts.) Find the power of this test for $\mu = 3$.

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 $\beta$ is a unknown parameter, and the $\varepsilon_i$ are independent, each with the $N(0, \sigma^2)$ distribution.

(a) (8 pts.) Show that the least squares estimator of $\beta$ 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 $\beta$.

(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 $\beta$:

$$(1,5), (2,-3), (2,-4), (2/3, 1).$$

11. (10 pts.) A jury panel included 12 men and 6 women. By a procedure described as "random" a jury of 7 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 $X_1, X_2, X_3, X_4, X_5$ 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: 17, 18, 13, 15, 16, and $Y$'s: 10, 11, 14, 12, find the exact $p$-value for the Wilcoxon Rank Sum test of $H_0$ vs $H_a$.