Spring 2008

PROBABILITY PART.

1. (a) (7 pts.) A group of 32 boys and 32 girls is randomly divided into two equal groups. Find the probability that each group will be equally divided into boys and girls.

(b) (7 pts.) If 10 children are randomly selected from a group of 32 boys and 32 girls, what is the probability that there will be 3 girls selected?

2. Suppose that 6.5% of men and 2.0% of women are color-blind. Assume that 51.4% of a population are women and 48.6% are men.

(a) (3 pts.) Find the probability that a randomly selected person is color-blind.

(b) (3 pts.) Given that a randomly selected person is not color-blind, what is the conditional probability that the selected person was a man?

(c) (6 pts.) Suppose that a random sample of 50 persons is selected. Find (approximate) probability that there were 1 color-blind persons in the sample.

(d) (6 pts.) Suppose that a random sample of 500 persons is selected. Find (approximate) probability that there were 20 color-blind persons in the sample.

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 $U_1^{(n)}, ..., U_n^{(n)}$ be independent uniformly distributed on $[0, n]$ random variables. Let

$$M_n = \min(U_1^{(n)}, ..., U_n^{(n)}), n = 1, 2, ...$$

Let X be exponentially distributed random variable with parameter 1, i.e its density is $e^{-x}, x > 0$. Show that M_n converges in distribution to X as $n \to \infty$.

(c) (3 pts.) Define: X_n converges in probability to 0 as $n \to \infty$.

/

4. Let Y have the density $f(y) = 2(1-y)$ for $0 \le y \le 1$. Given $Y = y$, let X be uniformly distributed random variable on the interval $[0, \sqrt{1-y}]$.

(a) (5 pts.) Find the joint density of (X, Y).

(b) (7 pts.) Find the marginal density of X.

(c) (10 pts.) Find the conditional density of Y given $X = x, 0 < x < 1$.

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

(e) (5 pts.) Find $V = \mathbf{E}(X^3 + XY|X)$.

5. Let $f(x) = 2.5x^4$ for $-1 \le x \le 1$. Let X_1, X_2, X_3 and X_4 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 $M = X_1 X_2$. Find the cdf for M. (Hint: graph the area of integration).

(c) (8 pts.) Find the cdf and the density for $Y = e^{X_1^2}$.

(d) (6 pts.) Find the correlation coefficient $\rho(5X_3 + X_1 - 2X_2 + 10X_4, X_4 - 3X_1 - 5X_2 + 8X_3)$.

(e) (6 pts.) Find $E(W)$ for

$$W = \frac{X_1^2 X_2^{1/3}}{X_3^4}$$

STATISTICS PART.

6. Let $\alpha > 0$ be a unknown parameter. Let $(X_1, .., X_n)$ be a random sample from the density

$$f(x) = \frac{2x}{\alpha^2} \text{ for } 0 \le x \le \alpha$$

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

(b) (6 pts.) Find the method of moments estimator of α.

(c) (7 pts.) Find the bias of the MLE of α.

7. Let $f_0(x) = \frac{x}{2}, 0 \le x \le 2$ and $f_1(x) = \frac{x^3}{4}, 0 \le x \le 2$. Suppose that we wish to test $H_0 : X$ has density f_0 vs $H_a : X$ has density f_1.

(a) (8 pts.) Construct the most powerful test of $\alpha \in (0,1)$ significance level based on X.

(b) (3 pts.) Which theorem did you use in part (a)?

(c) (7 pts.) Find the power of this test.

8. How does energy intake compare to energy expenditure? To study this issue, a random sample of professional soccer players was taken. Below the results are summarized in a table (MJ/day):

Player	1	2	3	4	5	6	7
Expenditure	14.4	12.1	14.3	14.2	15.2	15.0	17.8
Intake	14.6	9.2	11.8	11.6	12.9	15.5	16.3

(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 there is a significant difference between intake and expenditure.

(b) (8 pts.) Perform a nonparametric test of the same hypotheses with $\alpha = 0.05$. Find the p-value for this test.

9. (12 pts.) The following table provides data on body weight gain (g) for two independent samples: a sample of animals given a 1 mg/pellet dose of a certain soft steroid and a sample of control animals:

Steroid	30.1	45.3	42.2	40.5	39.8
Control	39.1	31.4	38.8	34.2	33.7

State a nonparametric model, define hypotheses, and carry out a test at level $\alpha = 0.02$ which will enable you to decide whether there is a significant difference in weight gain between steroid and control groups.

10. Let (x_i, Y_i) be observed for $i = 1, ..., n$. Suppose that the x_i are constants, and that $Y_i = \beta x_i^2 + \varepsilon_i$, 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 $\hat{\beta} = (\sum_{i=1}^n Y_i x_i^2)/(\sum_{i=1}^n x_i^4)$. Do **NOT** use matrix or vector space methods.

(b) (5 pts.) Show that $\hat{\beta}$ is an unbiased estimator of β. Do **NOT** use matrix or vector space methods.

(c) (6 pts.) Find $\mathrm{Var}(\hat{\beta})$. Do **NOT** use matrix or vector space methods.

(d) (8 pts.) For the following (x_i, Y_i) find a 99% confidence interval on β: (1, 4), (2, 10), (1, 2), (2, 14).

11. (12 pts.) A random sample of individuals who drive to work in a Detroit area is obtained, and each individual is categorized with respect to both size of vehicle and commuting distance (in miles). Does the accompanying data suggest that there is an association between type of vehicle and commuting distance?

Distance	< 10	$[10, 20]$	> 20
Subcompact	6	27	17
Compact	8	36	16
Midsize	21	46	33
Full-size	15	18	7

State a model, define hypotheses, and carry out a test at level $\alpha = 0.1$.

12. Let p_1 and p_2 be the proportions of a population of 45 thousand MSU students who wanted a decrease in tuition at time t_1 and time t_2, respectively. In order to estimate $\Delta = p_1 - 2p_2$ random samples of sizes $n_1 = 500$ and $n_2 = 1000$ 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 $X_1 = 400, X_2 = 850$.

(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?