Stat/Math 3460 - Winter 2006
Intermediate Statistical Theory
Course Outline

Instructor:	Dr. E. Susko (Chase 115) susko@mathstat.dal.ca `http://www.mathstat.dal.ca/~tsusko/`
Time:	TR 1:05-2:25
Place:	LSC 214
Office hours:	TR 9:30-11:30 (tentative)
Text:	Probability and Statistical Inference Volume 2: Statistical Inference J.G. Kalbfleisch

Marking Scheme

Assignments	35%
Midterm	25% Thursday, March 2 in class
Final exam	40%

Intended Material

1. Introduction to likelihood estimation
2. Likelihood estimation in multiple dimensions
3. Confidence intervals
4. Optimality properties of likelihood estimation
5. Significance tests
6. Sufficiency and conditional tests
7. Analysis of normal measurements
8. Additional topics

There will be no supplemental examination in this course.

Stat/Math 3460 - Intermediate Statistical Theory - Winter 2006
Final Examination

Name: _____
Student ID #: _____

Each question is equally weighted.

1. Let $x_1, \ldots, x_n$ be a random sample from the distribution with probability density function

$$f(x;\theta) = \theta x^{\theta-1}, \qquad 0 < x < 1$$

   (a) Derive the maximum likelihood estimator of θ.

   (b) Obtain an expression for the observed information, $J(\theta)$.

   (c) Suppose $n = 3$, $x_1 = 0.1$, $x_2 = 0.3$ and $x_3 = 0.5$. Construct an approximate 95% confidence interval for θ using the large sample normality of $\hat{\theta}$.

2. If $x_1, \ldots, x_n$ are a random sample from the $N(\mu, \sigma^2)$ distribution, one can show that $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \sum(x_i - \bar{x})^2/n$ and

$$J(\mu,\sigma) = \begin{bmatrix} n/\sigma^2 & 2n(\hat{\mu}-\mu)/\sigma^3 \\ 2n(\hat{\mu}-\mu)/\sigma^3 & -n/\sigma^2 + 3\{n\hat{\sigma}^2 + n(\hat{\mu}-\mu)^2\}/\sigma^4 \end{bmatrix}$$

   You do not need to show this.

   (a) What is the maximum likelihood estimator of $\gamma = \mu/\sigma$?

   (b) Give an expression involving $\hat{\gamma}$ *alone* for a 95% confidence interval for γ.

3. *Not in final version* The following table gives the number of beetles which dies within 6 days and the number which survived at each of six concentrations of an insecticide.

Concentration	0.711	0.852	0.959	1.066	1.202	1.309
Number Dead	15	24	26	24	29	29
Number Alive	35	25	24	26	21	20

   It is quite common for data like this to model the probability of a death at dose d through a logistic regression relationship

$$\log[p/(1-p)] = \alpha + \beta d$$

   implying that the porbability of death at dose d is $p = \exp(\alpha + \beta d)/[1 + \exp((\alpha + \beta d)]$. For the data in the table it can be shown that $\hat{\alpha} = 3.356$, $\hat{\beta} = -0.165$ and

$$J(\hat{\alpha}, \hat{\beta}) = \begin{bmatrix} 5.858 & 6.187 \\ 6.187 & 6.756 \end{bmatrix}$$

   (a) Obtain an approximate 99% confidence interval for $\alpha + 2\beta$.

   (b) Use the result of (3a) to obtain a 99% confidence interval for the probability of death at dose $d = 2$.

4. A seed dealer claims that his plans have a germination probability of 80%. A customer purchased three packages of seeds, one package for each type of plant and planted 100 seeds from each package. The number of seeds germinating in one month were as follows:

	Type 1	Type 2	Type 3
Germination	75	66	81
No Germination	25	34	19

You need not show the derivation of maximum likelihood estimators for the questions below; of course, you may want to do so if you are unsure what they are.

(a) Test the hypothesis that the germination probability is 80% for all types of plants. *You should be able to use a single test, not three separate ones*

(b) Test the hypothesis that the germination probability is the same for all types of plants (although not necessarily 80%)

(c) Assuming the germination probability is the same for all plants, test the hypothesis that it is 80%.

5. Suppose that patients arrive for treatment according to a Poisson process in time, with 20 arrivals per year on average. The treatment is successful for a fraction θ of patients. Let X be the number of successful treatments and Y the number of unsuccessful treatments in a one year period. Then X and Y are independent Poisson variates with means 20θ and $20(1-\theta)$.

(a) Determine the maximum likelihood estimator $\hat{\theta}$.

(b) Find a sufficient statistic that includes $\hat{\theta}$ as a component. Give reasons for your answer.

(c) Show that $T = X + Y$ is ancillary for θ.

(d) Obtain the distribution of $\hat{\theta}$ that would be appropriate for constructing confidence intervals. You need not carry out the confidence interval construction.

6. For each of the following questions, place a check mark next to the most appropriate answer.

(a) A random sample $x_1, \ldots, x_n$ from a Poisson distribution with mean μ and a sample from a Poisson distribution with mean $\lambda\mu$ are obtained. A 14.65% likelihood interval for λ is found to be $[0.9, 1.2]$. The p-value for a likelihood ratio statistic test of $H_0 : \lambda = 1$ would

i. _____ be less than or equal to 0.05

ii. _____ would be greater than or equal to 0.05

iii. _____ more information is needed to know whether the p-value would be less than or greater than 0.05.

(b) A roulette wheel has 18 red slots, 18 black slots and 2 green slots. A likelihood ratio statistic test is to be carried out on the hypothesis $H_0 : p_R = p_B = 18/38$ where p_R and p_B are the probabilities that the ball lands in the red or black slots. Assuming this hypothesis is correct, the probability that a p-value less than 0.75 is obtained is approximately

i. _____ 0.75

ii. _____ 0.25

iii. _____ 0.05

(c) For a random sample $x_1, \ldots, x_n$ from an exponential distribution: $f(x;\theta) = \theta^{-1}\exp[-x/\theta]$, it is known that $E[\bar{X}] = \theta$.

$$\begin{aligned} l(\theta) &= -n\log[\theta] - n\bar{x}/\theta \\ S(\theta) &= -n/\theta + n\bar{x}/\theta^2 \\ J(\theta) &= -n/\theta^2 + 2n\bar{x}/\theta^3 \end{aligned}$$

An estimator T is has expectation θ. A lower bound for the variance of T is

i. _____ n/θ^2
ii. _____ θ^3/n
iii. _____ $\theta^2/(2n)$
iv. _____ $2\theta^2 + \theta/n$
v. _____ none of the above

(d) Suppose that the hypothesis $H_0 : p = 1/2$ is tested using $D = |X - 1/2|$ as a test statistic where X has a binomial$(10, p)$ distribution. The hypothesis is rejected when the p-value is less than 0.05. The Type II error for the experiment

i. _____ is 0.05
ii. _____ is 0.95
iii. _____ is 0.50
iv. _____ cannot be determined without additional information.

(e) The exponential probability density function is $f(x;\theta) = \theta^{-1}\exp[-x/\theta]$ and $P\{X \geq x;\theta\} = \exp[-x/\theta]$. A single observation X is drawn from an exponential distribution with unknown θ. The coverage probability for the random interval $[0, 2X]$ is

i. _____ $e^{-1/2}$
ii. _____ e^{-2}
iii. _____ $1 - e^{-1/2}$
iv. _____ $1 - e^{-2}$
v. _____ Cannot be calculated without knowing the value of θ.