Statistics Masters Exam: Part II (Afternoon)

May 13, 2004
1:30 – 4:30 P.M.

Instructions:

1. Write your student number here: ________________________________

2. Write your name here: _________________________________________

3. Write your Student Number ONLY (not your name) at the top of each page.

4. Start each problem on a new page. Use one side of the paper only — that is, do not work on the back of the page.

5. The exam is closed-book. But you may use up to three pages of notes (8.5" x 11", both sides).

6. You must hand in your work and this exam packet at the end of this session.

7. There are four problems. Each problem will be given equal weight in the overall grade.

8. GOOD LUCK!!!!
1. Let $\theta = (\theta_1, \theta_2)$ be a vector of parameters. Suppose that $S_1 = S_1(x, \theta_2)$ is a sufficient statistic for $\theta_1$ whenever $\theta_2$ is fixed and known, where $x = (x_1, \ldots, x_n)$ a sample of $n$ observations. Further, let $S_2 = S_2(x, \theta_1)$ be sufficient for $\theta_2$ whenever $\theta_1$ is fixed and known. Assume that $\theta_1$ and $\theta_2$ vary independently, $\theta_1 \in \Omega_1$ and $\theta_2 \in \Omega_2$, and the set $\{x : p(x, \theta) > 0\}$ does not depend on $\theta$ where $p(x, \theta)$ is the density of $x$.

   a. Show that if $S_1$ and $S_2$ do not depend on $\theta_2$ and $\theta_1$, respectively, then $(S_1, S_2)$ is sufficient for $\theta$.

   b. Show that part a fails for the $n(\theta_1, \theta_2)$ distribution.
2. Suppose $X_1, ..., X_n$ are iid $n(\mu, \sigma^2)$ with $\mu$ and $\sigma^2$ both unknown.

   a. Show that the size $\alpha$ likelihood ratio test of $H_0 : \mu = 0$, $\sigma^2$ unspecified versus $H_A : \mu \neq 0$, $\sigma^2$ unspecified can be based on the t-statistic. You should provide statements to support the arguments you use in the derivation.

   b. Construct the size $\alpha$ likelihood ratio test of $H_0 : \mu = 0$, $\sigma^2 = 1$ versus $H_A : \mu \neq 0$ or $\sigma^2 \neq 1$, or both. Explain how you would determine the critical value or the p-value of the test.
3. Suppose $X_1, \ldots, X_n$ are iid from a Poisson distribution with parameter $\theta$.

a. Find the limiting distribution for $\sqrt{n}(\bar{X} - \theta)$.

b. State the limiting distribution for $\sqrt{n}(g(\bar{X}) - g(\theta))$, assuming that $g(u)$ is sufficiently smooth.

c. Find $g(u)$ so that the asymptotic variance in part b does not depend on $\theta$. (Then $g(u)$ is called a variance stabilizing transformation.)

d. Use the result in part c to construct an approximate 95% confidence interval for $\theta$. 
4. Suppose you have one observation $x$ from a $n(\theta, 1)$ distribution.

a. Suppose $\theta \in \Omega = (-\infty, \infty)$ and we put a conjugate prior on $\Omega$. Find the Bayes estimate assuming square error loss.

b. Suppose we believe that $\theta$ can take only two values, $\theta_1$ and $\theta_2$, with equal probabilities. Show that the Bayes estimate (with square error loss) is

$$\frac{\theta_1 e^{-\frac{1}{2}(x-\theta_1)^2} + \theta_2 e^{-\frac{1}{2}(x-\theta_2)^2}}{e^{-\frac{1}{2}(x-\theta_1)^2} + e^{-\frac{1}{2}(x-\theta_2)^2}}$$
The following three questions cover stochastic processes.

Students who have just finished the current Stat 515 must work on Problem 5 and choose one from Problems 6 and 7.

Other students may choose any two from Problems 5, 6, and 7.
5. Simulation. Consider the following two density functions:

\[ g(x) = \begin{cases} 
1 & \text{if } x \in [0, 1] \\
0 & \text{if } x \notin [0, 1] 
\end{cases} \]

\[ f(y) = \begin{cases} 
2y & \text{if } y \in [0, 1] \\
0 & \text{if } y \notin [0, 1] 
\end{cases} \]

Suppose we wish to estimate the cumulative distribution function of \( f(\cdot) \) using simulation. That is, we are interested in the integral

\[ F(t) = \int_0^t f(y) \, dy = t^2 \quad (\text{for } t \in [0, 1]) \]

and we pretend that we do not have an explicit expression for it.

(a) Given a simulated IID sample from \( f \), say \( Y_1, Y_2, ..., Y_B \), how would you estimate \( F(t) \), for a fixed value of \( t \)? (Hint: \( F(t) \) is the expectation of an indicator function.)

(b) What would the simulation variance be, when expressed as a function of \( t \)?

(c) Suppose we had instead a simulated sample \( X_1, ..., X_B \) from the density \( g \). How would one use importance sampling to generate an estimator of \( F(t) \)?

(d) What would the simulation variance for importance sampling be, expressed as a function of \( t \)?

(e) For what values of \( t \) would the importance sampling estimator have higher efficiency in terms of its asymptotic standard error?
6. Poisson process

(a) Suppose that the number $N$ of typographical errors in a new text is Poisson distributed with mean $\lambda$. Suppose that each of the errors in the text are found by proofreader (call him $A$) with probability $p_A$, with each occurrence being independent. If we let $X_A$ be the number of errors found by this proofreader and $Y_A$ be the number of errors not found, show that $X_A, Y_A$ are independent Poisson variables, and give their parameters. (Hint: condition on $N$.)

(b) Now suppose that there is a second proofreader (call her $B$), who finds the errors with probability $p_B$, independently of $A$. We now categorize the errors made into four types:

- **Type 1**: both A and B find the error
- **Type 2**: A finds the error but B does not
- **Type 3**: B finds the error, but A does not
- **Type 4**: neither A nor B finds the error

Argue that, for given $N = n$, the counts $Y_1, Y_2, Y_3, Y_4$ of the number of errors of each type have a multinomial distribution. Give the parameters (in terms of $n, p_A,$ and $p_B$) and the probability mass function.

(c) Show that $Y_1, Y_2, Y_3, Y_4$ are (unconditionally) independent Poisson, and give the parameters.
7 §. (Markov Chains). Trials are performed in sequence. If the preceding two trials are successes, then the next trial is a success with probability 0.8. Otherwise, the next trial is a success with probability 0.5.

(a) Write this model as a Markov Chain, and give the probability transition matrix.

(b) Is this Chain irreducible? positive recurrent? Aperiodic?

(c) In the long run, what proportion of trials are successes?