McGill University

Part A Examination in Statistics

Theory Paper

Department of Mathematics & Statistics

Date: 9 May 2005

Time: 13:00

Instructions

• Answer two questions out of Section P. Only two questions will be marked.

• Answer four questions out of Section S. Only four questions will be marked.

• If you do not indicate which questions you wish to have marked, the questions will be marked in the order in which they appear in the answer book until the quota has been reached.

• All questions are weighted equally (20 marks each).

• A passing grade must be achieved in both sections (P and S) to pass the Theory Paper.

• Each question will be assessed independently by at least two members of the statistics group, and the final result determined after discussion within the Part A Exam Subcommittee.

• Good luck!

This exam comprises the cover and 9 questions on 6 pages.
Section P:  Answer two questions out of questions P1 to P3.

P1. Let $\Omega = [0,1]$. For each $n \geq 1$ define the $n$th dyadic partition of $[0,1]$ by $t_0 = 0 < t_1 = 1/2^n < \cdots < t_k = k/2^n < \cdots < t_{2^n} = 1$. An interval determined by this partition is $\{0\}$ or $(t_k, t_{k+1})$, $0 \leq k \leq 2^n - 1$.

Show that the $\sigma$-algebra $B[0,1]$ of Borel subsets of $[0,1]$ is the smallest $\sigma$-algebra that contains all the $\sigma$-algebras $B_n$, where $B_n$ is the $\sigma$-algebra of finite unions of the intervals determined by the $n$th partition.

Hint: Use the fact that for $\omega \in [0,1]$ either $\omega = 0$ or there exists a unique sequence $d_1(\omega), d_2(\omega), \ldots, \in \{0,1\}$ which does not end in an unbroken string of zeros such that

$$\sum_{i=1}^{n} \frac{d_i(\omega)}{2^i} < \omega \leq \sum_{i=1}^{n} \frac{d_i(\omega)}{2^i} + \frac{1}{2^n}$$

for all $n \in \mathbb{N}$.

P2. Let $X$ and $Y$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{G}$ be a $\sigma$-subalgebra of $\mathcal{F}$.

(a) [7 marks]
Using the Radon-Nikodym theorem, carefully define the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ in the case where $X \geq 0$.

(b) [4 marks]
Suppose $X$ and $Y$ are independent, and that $\phi : \mathbb{R}^2 \to [0, \infty)$ is measurable. For each $x$, let $h(x) = \mathbb{E}[\phi(x,Y)]$. Show that $h(X)$ is a version of $\mathbb{E}[\phi(X,Y)|X]$.

(c) [5 marks]
Suppose $X$ and $Y$ have joint density function given by

$$f(x,y) = \begin{cases} 2 & \text{if } 0 \leq x \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}[X|\sigma(Y)]$, where $\sigma(Y)$ is the $\sigma$-subalgebra of $\mathcal{F}$ generated by $Y$.

(d) [4 marks]
Let $A, B \in \mathcal{F}$, and let $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$. Find a version of $\mathbb{P}[A|\mathcal{G}]$.

[Continued ]
**P3. (a) [10 marks]**

Let the random variable $X$ on $(\Omega, \mathcal{F}, P)$ induce the probability space $(\mathbb{R}^1, \mathcal{B}^1, \mu)$, that is, $\mu(A) = P[X \in A]$ for all $A \in \mathcal{B}^1$, $\mathcal{B}^1$ the $\sigma$-field of Borel sets in $\mathbb{R}^1$. Let $f \geq 0$ be Borel measurable. Prove that

$$\int_{\Omega} f(X(\omega)) P(d\omega) = \int_{\mathbb{R}^1} f(x) \mu(dx).$$

provided either side exists.

**P3. (b) [10 marks]**

Let $X_1, \ldots, X_n, \ldots$ be Borel measurable random variables satisfying $\mathbb{E} [\sum_{i=1}^{\infty} |X_i|] < \infty$. Show that

$$\mathbb{E} \left[ \sum_{i=1}^{\infty} X_i \right] = \sum_{i=1}^{\infty} \mathbb{E} [X_i],$$

assuming that $\sum_{i=1}^{n} X_i$ converges with probability 1 as $n \to \infty$. 

[End of Part P. Exam continues overleaf.]
Section S: Answer four questions out of questions S1 to S6.

S1. Note: The following three subquestions (a), (b) and (c) are independent of each other.

(a) [4 marks]
Let \( X_1 \) and \( X_2 \) be independent, continuous and identically distributed random variables with cdf \( F \) and pdf \( f \). Show that \( \Pr[X_1 \leq X_2] = 1/2 \).

(b) [7 marks]
Let \( X \sim U(0, 1) \) and let \( Z = \max(X, 1/2) \). Find the Lebesgue decomposition of the cumulative distribution function of \( Z \).

(c) [9 marks total]
Let \( [X, Y]' \) be uniformly distributed on the half-disk bounded by \( x = 0 \) and by the arc of radius 1 from \(-\pi/2\) to \(\pi/2\), as per the Figure below.

i. [3 marks]
Are \( X \) and \( Y \) independent?
Prove your answer.

ii. [6 marks]
Let \( \Theta \) be the angle of the vector \([X, Y]'\) from the horizontal (as displayed), and let \( R = \sqrt{X^2 + Y^2} \) be the length of this vector. Show that \( R \) and \( \Theta \) are independent and find their respective pdf’s.

[Continued]
S2. **Note:** The following two subquestions (a) and (b) are independent of each other.

**(a)** [10 marks]
Let \(X_1, X_2, \cdots\) be a sequence of independent and identically distributed random variables with common distribution \(F\). Write
\[
M_n = \max_{1 \leq i \leq n} X_i, \quad n = 1, 2, \cdots
\]
Suppose further that
\[
\lim_{x \to \infty} e^x [1 - F(x)] = b > 0.
\]
Show that \(M_n - \log(bn)\) converges in distribution to some random variable \(Y\) and find its moment-generating function.

*Hint:* If \(\lim_{n \to \infty} a_n = a\), then \(\lim_{n \to \infty} \left[1 + a_n/n + o(1/n)\right]^n = \exp(a)\).

**(b)** [10 marks]
Let \(X \sim \text{Poisson}(\lambda), \lambda > 0\). Show that for \(c > 0\),
\[
P[X > c] \leq \exp(c - \lambda) \left(\frac{\lambda}{c}\right)^c.
\]

*Hint:* Use Markov’s inequality to bound \(P[\exp(tX) > k]\) above for \(t \in \mathbb{R}\).

S3. **(a)** [6 marks]
Prove from the definition of completeness that the Poisson family of distributions indexed by the parameter \(\lambda \in [0, +\infty)\) is complete.

Let \(X_1, \ldots, X_n\) be a sample of independent and identically distributed \(\text{Poisson}(\lambda)\) random variables, \(\lambda > 0\).

**(b)** [6 marks]
Find a minimal sufficient statistic for \(\lambda\) based on this sample.

**(c)** [2 mark]
Find a simple unbiased estimator for \(P_{\lambda} [X_1 = 0]\).

**(d)** [6 marks]
Hence find the unique minimum variance unbiased estimator for \(P_{\lambda} [X_1 = 0]\).
S4. Let \( X = [X_1, X_2, \ldots, X_n]' \) be a random vector of jointly continuous random variables (r.v.s), whose joint probability density function is given by \( f_X(x; \theta) \). The parameter \( \theta \) is assumed to be a real valued scalar. Let \( g(\theta) \) be an everywhere differentiable real valued function of \( \theta \) and suppose that the statistic \( T = T(X) \) satisfies \( \mathbb{E}_\theta[T(X)] = g(\theta) \).

(a) [4 marks]

State the Cramér-Rao inequality for \( \text{Var}_\theta[T] \), under appropriate regularity conditions. State the conditions.

(b) [4 marks]

Under appropriate regularity conditions, prove that

\[
\mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta} \log f_X(X; \theta) \right] = 0.
\]

(c) [5 marks]

Show that

\[
\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_X(X; \theta) \right)^2 \right] = -\mathbb{E}_\theta \left[ \frac{\partial^2}{\partial \theta^2} \log f_X(X; \theta) \right]
\]

under an appropriate regularity condition. State this condition.

(d) [7 marks]

Suppose now that the \( X_i \)'s are independent and identically distributed with common marginal probability density function \( f_X(x; \theta) \). Prove that

\[
\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_X(X; \theta) \right)^2 \right] = n \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_X(X_1; \theta) \right)^2 \right].
\]

Present the details of your reasoning carefully.

[Continued ]
S5. Let \( X \sim N(\mu, 1) \) and \( Y \sim N(\nu, 1) \) be independent, where we assume \( \nu \neq 0 \). Let \( \theta = \mu/\nu \).

(a) [4 marks]
Show that \( Z(\theta) = \frac{X - \theta Y}{\sqrt{1 + \theta^2}} \) is a pivot for \( \theta \).

(b) [8 marks]
Use the pivot \( Z(\theta) \) to find the value \( k(\alpha) > 0 \) such that if
\[
A(\theta) = \{ (x, y) \in \mathbb{R}^2 : (y^2 - k(\alpha))\theta^2 - 2\theta xy + x^2 \leq k(\alpha) \},
\]
then \( A(\theta_0) \) is an acceptance region for an \( \alpha \) test of \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \neq \theta_0 \), \( 0 < \alpha < 1/2 \).

(c) [8 marks]
Show that inverting the acceptance region found in (b) yields a \( 1 - \alpha \) confidence set that is equal to \( \mathbb{R} \) with positive probability. Do not attempt to find this probability. Express in simple words the condition on observations \( x \) and \( y \) which leads to a \( 1 - \alpha \) confidence set equal to the real line. (You can answer this question even if you did not answer (b)).

S6. Let \( X = [X_1, \ldots, X_n]' \) be a random sample from the Pareto(\( \beta, \gamma \)) distribution, with probability density function
\[
f_{\gamma,\beta}(x) = \frac{\beta \gamma^\beta}{\beta x^{\beta+1}} \mathbb{1}[x \geq \gamma], \quad \beta, \gamma > 0.
\]
Assume that \( \gamma \) is fixed and known and \( \beta \) is unknown.

(a) [10 marks]
Show that the \( \Gamma(r, \theta) \), \( r, \theta > 0 \) family is conjugate for the Pareto(\( \beta, \gamma \)), \( (\beta > 0, \gamma \) fixed) family given \( X_{(1)} = \min \{ X_1, \ldots, X_n \} \) (that is, show that the posterior distribution of \( \beta \) is also in the Gamma family). (Hint: \( X_{(1)} \) is also a Pareto random variable.)

(b) [3 marks]
Let \( \beta \sim \Gamma(r, \theta) \). Find the Bayes rule (estimator) of \( \beta \) under squared-error loss.

(c) [3 marks]
In no more than two sentences, criticize the use of a posterior for \( \beta \) conditioned on \( X_{(1)} \).

(d) [4 marks]
Indicate how you would produce a shortest \( 1 - \alpha \) credible interval for \( \beta \). Be specific but do not carry out the procedure.

End of Theory Paper.