Student Name: ______________________ ID: ______________________

McGill University

Faculty of Science

Department of Mathematics and Statistics

Statistics Part A Comprehensive Exam

Theory Paper

Date: Tuesday, May 14, 2013 Time: 13:00 – 17:00

Instructions

- Answer only **two** questions out of Section P. If you answer more than two questions, then only the **FIRST TWO questions will be marked.**
- Answer only **four** questions out of Section S. If you answer more than four questions, then only the **FIRST FOUR questions will be marked.**

Questions	Marks
P1	
P2	
P3	
S1	
S2	
S3	
S4	
S5	
S6	

This exam comprises the cover page and four pages of questions.

Section P
Answer only two questions out of P1–P3

P1.

Consider Lebesgue measure λ on $((0,1],\mathcal{B})$, where $\mathcal{B}$ are the Borel sets of $(0,1]$. Fix a bijection $\sigma : \{1,2,\ldots\} \to \{1,2,\ldots\}$, and define a map $T : (0,1] \to (0,1]$ as follows. If $x \in (0,1]$ has dyadic expansion $0.x_1x_2\ldots$ set $T(x) = 0.x_{\sigma(1)}x_{\sigma(2)}\ldots$.

(Note: if this expansion is not unique then $x = m/2^n$ for some $m,n \in \mathbb{N}$ with $m \leqslant 2^n$; in this case we use the expansion that ends with an infinite sequence of zeros rather than an infinite sequence of ones.)

(a) Prove that $T((0,m/2^n])$ is measurable for all $m,n \in \mathbb{N}$ with $m \leqslant 2^n$. **(4 Marks)**

(b) Show that $\{(0,x] : x = m/2^n, m,n \in \mathbb{N}, m \leqslant 2^n\}$ generates $\mathcal{B}((0,1])$; conclude that T and T^{-1} are both measurable maps. **(8 Marks)**

(**Hint:** *one of these follows from the other by replacing σ by σ^{-1}.*)

(c) Show that that $\lambda(T(B)) = \lambda(B)$ for all $B \in \mathcal{B}$. **(8 Marks)**

(**Hint:** *First show this for some simpler collection of sets, then use the Carathéodory Extension theorem.*)

P2.

Let A,B,C be strictly positive, independent random variables with distribution function F. Show that the quadratic equation $Az^2 + Bz + C = 0$ has real zeros with probability

$$\int_0^\infty \int_0^\infty F\left(\frac{x^2}{4y}\right) \mathrm{d}F(x)\mathrm{d}F(y)\,.$$

(20 Marks)

P3.

Fix a non-negative random variable X on a probability space $(\Omega,\mathcal{F},\mathbb{P})$, and a σ-algebra $\mathcal{G} \subset \mathcal{F}$.

(a) Show that almost surely $\mathbb{E}[X \mid \mathcal{G}] = \int_0^\infty \mathbb{P}[X > t \mid \mathcal{G}]\mathrm{d}t$. **(10 Marks)**

(b) Show that almost surely $\mathbb{P}[X \geqslant a \mid \mathcal{G}] \leqslant a^{-t}\mathbb{E}[X^t \mid \mathcal{G}]$. **(10 Marks)**

Section S
Answer only four questions out of S1–S6

S1.

Let (S_1, S_2) be a random pair uniformly distributed on the triangle

$$\mathcal{S} = \left\{(x, y) \in [0, \infty)^2 \;:\; x + y \leqslant 1\right\}$$

and set $S_3 = 1 - S_1 - S_2$.

(a) For arbitrary $s_1, s_2, s_3 \in \mathbb{R}$, compute the probability

$$\Pr(S_1 > s_1, S_2 > s_2, S_3 > s_3).$$

(7 marks)

(b) Determine the distribution of S_3. **(6 marks)**

(c) Show that the variables

$$\frac{S_1}{S_1 + S_2} \quad \text{and} \quad \left(\frac{S_1 + S_2}{S_1 + S_2 + S_3}\right)^2$$

are independent and uniformly distributed on the interval $(0, 1)$. **(7 marks)**

S2.

Let $X_i, i = 1, 2, \ldots$ be independent Bernoulli(p) random variables and let $Y_n = (X_1 + \cdots + X_n)/n$. Define $\theta_n = Y_n(1 - Y_n)$.

(a) Show that θ_n is a consistent estimator of $\mathrm{var}(X_i) = p(1 - p)$. **(4 marks)**

(b) Determine the limit in distribution, as $n \to \infty$, of the sequence

$$\sqrt{n}\{\theta_n - p(1 - p)\}.$$

(8 marks)

(c) Determine the limit in distribution of the sequence

$$n\{\theta_n - p(1 - p)\}$$

as $n \to \infty$ when $p = 1/2$. **(8 marks)**

S3.

Let N be a Poisson(λ) random variable, independent of the sequence of i.i.d. random variables $X_i, i = 1, 2, \ldots$ satisfying $\Pr(X_i > 0) = 1$. Define

$$Z = \sum_{i=1}^{N} X_i.$$

(a) Suppose that $Y_i, i = 1, 2, \ldots$ is a sequence of i.i.d. random variables, independent of N and such that $\Pr(Y_i > 0) = 1$. Suppose further that X_i is independent of Y_j when $j \neq i$, but possibly dependent when $i = j$. Compute the linear correlation coefficient $\varrho(Z, W)$, where

$$W = \sum_{i=1}^{N} Y_i.$$

Under which conditions does $\varrho(Z, W)$ exist? Can Z and W ever be independent? Justify your answers. **(8 marks)**

(b) Suppose that the variables X_i follow the logarithmic series distribution with parameter $p \in (0, 1)$, viz.

$$\Pr(X_i = k) = \left\{-\frac{1}{\ln(p)}\right\} \frac{(1-p)^k}{k}, \quad k = 1, 2, \ldots$$

Prove that the distribution of Z is Negative binomial$(-\lambda/\ln(p), p)$. **(8 marks)**

(**Hint:** *You may use, without proof, that the moment generating functions of the Poisson*(λ) *and Negative binomial*(r, p) *distributions are, respectively, given by* $e^{\lambda(e^t-1)}$, $t \in \mathbb{R}$ *and* $[p/\{1-(1-p)e^t\}]^r$, $t < -\ln(1-p)$.)

(c) Is the family of distributions of Z from part **(b)** an exponential family? Justify your answer. **(4 marks)**

S4.

Let $X_1, \ldots, X_n$ be mutually independent copies of a discrete random variable X. Assume that there exists $\theta \in (0, 1)$ such that

$$\Pr(X = x) = P_\theta(x) \equiv \begin{cases} \theta & \text{if } x = -1, \\ (1-\theta)^2\theta^x & \text{if } x = 0, 1, \ldots \\ 0 & \text{elsewhere.} \end{cases}$$

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \xi_n = \sum_{i=1}^{n} \mathbf{1}(X_i = -1),$$

where $\mathbf{1}(A)$ denotes the indicator of the event A.

(a) Show that $(\overline{X}_n, \xi_n)$ is jointly sufficient for θ. **(4 marks)**

(b) Find an algebraic expression for $\mathrm{E}(\bar{X}_n)$. You may use the fact that $\sum_{x=0}^{\infty} x\theta^x = \theta(1-\theta)^{-2}$.

(2 marks)

(c) Using the result from part (b) or otherwise, check whether the family of distributions $\{P_\theta : \theta \in (0, 1)\}$ is complete for the statistic $(\overline{X}_n, \xi_n)$. **(2 marks)**

(d) Derive the maximum likelihood estimator θ_n of θ. **(2 marks)**

(e) Show that θ_n is a consistent estimator of θ. **(2 marks)**

(f) Compute the Fisher information for θ, viz.

$$I(\theta) = \mathrm{E}\left[\left\{\frac{\partial}{\partial\theta} \ln P_\theta(X)\right\}^2\right].$$

You may use the fact that $\sum_{x=0}^{\infty} x^2\theta^x = \theta(1+\theta)(1-\theta)^{-3}$.

(4 marks)

(g) Construct an approximate 95% confidence interval for θ based on θ_n. Justify your answer. **(4 marks)**

S5.

Let $\{f_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ be a class of densities with common support $\mathcal{X} \subseteq \mathbb{R}$. Assume that for all $x \in \mathcal{X}$, the map $\theta \mapsto f_\theta(x)$ is differentiable in θ. Given a random variable X with density f_θ and a (measurable) function $g : \Theta \to \mathbb{R}$, consider two classes of estimators defined by

$$\begin{aligned}\mathcal{U}_g &= \{\delta : \forall_{\theta\in\Theta} \quad \mathrm{E}_\theta[\delta(X)] = g(\theta)\},\\ \mathcal{U}_0 &= \{\xi : \forall_{\theta\in\Theta} \quad \mathrm{E}_\theta[\xi(X)] = 0\}.\end{aligned}$$

In other words, $\mathcal{U}_g$ is the set of all unbiased estimators $\delta(X)$ of $g(\theta)$ while $\mathcal{U}_0$ is the set of all unbiased estimators $\xi(X)$ of 0. Let also $\psi(X, \theta)$ be a (measurable) function of X and θ.

(a) Show that the following statements are equivalent: **(6 marks)**

(i) $\mathrm{cov}_\theta[\delta(X) - \delta'(X), \psi(X, \theta)] = 0$ for all $\theta \in \Theta$ and all $\delta, \delta' \in \mathcal{U}_g$;

(ii) $\mathrm{E}_\theta[\xi(X)\psi(X, \theta)] = 0$ for all $\theta \in \Theta$ and $\xi \in \mathcal{U}_0$.

When (ii) holds, the "estimating function" $\psi(X, \theta)$ is said to be perpendicular to $\mathcal{U}_0$.

(b) Fix $\delta_0 \in \mathcal{U}_g$ and suppose that $\psi(X, \theta)$ is not constant and perpendicular to $\mathcal{U}_0$. Show that

$$(\mathrm{cov}_\theta[\delta_0(X), \psi(X, \theta)])^2 \leqslant \mathrm{var}_\theta[\delta(X)]\mathrm{var}_\theta[\psi(X, \theta)]$$

holds for all $\delta \in \mathcal{U}_g$. **(4 marks)**

(c) Now suppose that the identity

$$\frac{\partial}{\partial\theta}\int_{\mathcal{X}} T(x)f_\theta(x)\mathrm{d}x = \int_{\mathcal{X}} T(x)\frac{\partial}{\partial\theta}f_\theta(x)\mathrm{d}x$$

is valid for any (measurable) function $T : \mathcal{X} \to \mathbb{R}$.

(i) Show that the score function, viz.

$$s(X, \theta) = \frac{\partial}{\partial\theta}\ln f_\theta(X),$$

is perpendicular to $\mathcal{U}_0$. **(5 marks)**

(ii) Assume that g is differentiable on Θ. Show that, for all $\delta \in \mathcal{U}_g$,

$$\mathrm{cov}_\theta[\delta(X), s(X, \theta)] = g'(\theta).$$

Using part **(b)** or otherwise, derive the Fréchet–Cramér–Rao lower bound for the variance of $\delta(X)$ for any $\delta \in \mathcal{U}_g$. **(5 marks)**

S6.

Let X be a single observation from a density f_θ whose parameter $\theta \in \{\theta_0, \theta_1\}$ is unknown. Consider the problem of testing $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$. Let $k > 0$ and $\gamma \in [0, 1]$ be fixed constants.

(a) Show that the decision rule

$$\delta(x) = \begin{cases} 1 & \text{if } f_1(x) > k f_0(x), \\ \gamma & \text{if } f_1(x) = k f_0(x), \\ 0 & \text{if } f_1(x) < k f_0(x), \end{cases}$$

is a most powerful test of its size. **(8 marks)**

(b) Show that given any $\alpha \in (0, 1)$, there exists a test δ of the form given in part **(a)** which is most powerful of size α. **(8 marks)**

(c) Given a decision rule δ of the form **(a)** which is most powerful of size α, show that $\alpha \leqslant 1 - \beta$, where β stands for the probability of a Type II error. **(4 marks)**