Student Name: ______________________ ID: ______________________

McGill University

Faculty of Science

Department of Mathematics and Statistics

Statistics Part A Comprehensive Exam

Theory Paper

Date: Tuesday, May 13, 2014 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.

(a) State Fubini's theorem. **(5 marks)**

(b) Show that if X and Y are random variables with joint probability density function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$, then the function g defined by

$$g(x) = \int_{\mathbb{R}} f(x, y)dy$$

is a probability density function for X. **Hint.** Recall the change of measure formula: if X has law μ then for any bounded measurable function $f : \mathbb{R} \to \mathbb{R}$, $\mathrm{E}(f(X)) = \int f(x)\mathrm{d}\mu$. **(8 marks)**

(c) Show that if f and g are two densities for X then the set $\{x : f(x) \neq g(x)\}$ has Lebesgue measure zero. **(7 marks)**

P2. In this question $(X_i, i \geqslant 1)$ is an arbitrary sequence of real random variables.

(a) What does it mean for X_i to converge in distribution to a random variable X as $i \to \infty$? **(5 marks)**

(b) Show that there exist positive constants $a_1, a_2, \ldots$ such that $a_n X_n$ converges in distribution to 0. **(5 marks)**

(c) Let $X_1, X_2, \ldots$ be identically distributed random variables with finite second moment. Show that for all $\epsilon > 0$, $n\mathbb{P}[|X_1| \geqslant \epsilon\sqrt{n}] \to 0$. **(5 marks)**

(d) Let $X_1, X_2, \ldots$ be identically distributed random variables with finite second moment. Show that $n^{-1/2} \max_{1 \leqslant k \leqslant n} X_k$ converges in probability to 0. **(5 marks)**

P3.

(a) Suppose that X is a random variable taking only positive integer values. Show that for all integers $n, m \geqslant 1$, we have $\mathrm{P}[X \geqslant n + m | X > n] = \mathrm{P}[X \geqslant m]$ if and only if there is $p \in (0, 1)$ such that for all integers $m \geqslant 1$ we have $\mathrm{P}[X = m] = p^{m-1}(1 - p)$. **(10 marks)**

(b) Suppose that X is a continuous random variable taking only non-negative values. Show that for all $s, t \geqslant 0$, we have $\mathrm{P}[X \geqslant s + t | X \geqslant s] = \mathrm{P}[X \geqslant t]$ if and only if there is $c > 0$ such that for all $t \geqslant 0$ we have $\mathrm{P}[X \geqslant t] = e^{-ct}$. **(10 marks)**

Section S
Answer only four questions out of S1–S6

S1. Consider a Dirichlet distributed random vector (X_1, X_2, X_3) with parameters $\alpha_1, \alpha_2, \alpha_3 > 0$, that is, $X_3 = 1 - X_1 - X_2$ and the density of (X_1, X_2) is

$$f(x_1, x_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} (1 - x_1 - x_2)^{\alpha_3 - 1}$$

for all $x_1, x_2 > 0$ such that $x_1 + x_2 < 1$.

(a) What can you say about the density of (X_1, X_2, X_3)? **(3 marks)**

(b) Determine the marginal distributions of X_i, $i = 1, \ldots, 3$. **(6 marks)**

(c) Compute the correlation between X_1 and $X_2 + X_3$. *Justify every step you make.*

(5 marks)

(d) Suppose that $Y_1 \sim \text{Beta}(\alpha_1, \alpha_2 + \alpha_3)$ and $Y_2 \sim \text{Beta}(\alpha_2, \alpha_3)$ are independent. Prove that

$$(X_1, X_2, X_3) \stackrel{d}{=} (Y_1, Y_2(1 - Y_1), (1 - Y_1)(1 - Y_2))$$

where $\stackrel{d}{=}$ denotes equality in distribution. *Hint: show first that* $(X_1, X_2) \stackrel{d}{=} (Y_1, Y_2(1 - Y_1))$. **(6 marks)**

S2. Consider the inverse Gaussian distribution with parameters $\lambda > 0$ and $\mu > 0$. Its density is given by

$$f(x; \lambda, \mu) = \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left\{-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$$

(a) Show that the inverse Gaussian family of distributions is an exponential family. Identify the canonical parameters and determine the canonical parameter space.

(7 marks)

(b) Suppose that X is an inverse Gaussian random variable. Compute the correlation between X and $1/X$. **(7 marks)**

(c) Show that the correlation of any two random variables Y and Z with finite second moments satisfies $|\text{corr}(Y, Z)| \leqslant 1$. Can the bound $|\text{corr}(Y, Z)| = 1$ be attained for any pair (Y, Z) of random variables with given marginal distributions? Justify your answer. **(6 marks)**

S3. Suppose that $\alpha, \beta > 0$ and $(X_1, P_1), \ldots, (X_k, P_k)$ are independent random vectors such that

$$\begin{aligned} X_i | P_i &\sim \text{Binomial}(n_i, P_i), \quad i = 1, \ldots, k, \\ P_i &\sim \text{Beta}(\alpha, \beta). \end{aligned}$$

Denote the total number of successes by $Y = \sum_{i=1}^{k} X_i$.

(a) Compute the expectation and variance of Y. **(6 marks)**

(b) Determine the distribution of Y when $n_1 = \cdots = n_k = 1$. **(7 marks)**

(c) Suppose that W and Z are random variables with finite expectations. Determine a function h such that $W - h(Z)$ is orthogonal to $g(Z)$, viz.

$$\mathrm{E}\Big[\big\{W - h(Z)\big\}g(Z)\Big] = 0,$$

for any measurable function g such that $\mathrm{E}\{g(Z)\}$ is finite. *Show your work and justify every step you make.* **(7 marks)**

S4. Find a nontrivial set of sufficient statistics in each of the following cases:

(a) Random variables $X_{jk} (j = 1, \cdots, m\ ;\ k = 1, \cdots r)$ have the form $X_{jk} = \mu + \eta_j + \varepsilon_{jk}$, where the η_j's and the ε_{jk}'s are independently normally distributed with zero means and variances respectively σ_b^2 and σ_w^2. The unknown parameters are thus $(\mu, \sigma_b^2, \sigma_w^2)$. **(10 marks)**

(b) Independent binary random variables $Y_1, \cdots, Y_n$ are such that the probability of the value one depends on an explanatory variable x, which takes corresponding values $x_1, \cdots, x_n$, through the model

$$\log\left[\frac{P(Y_j = 1)}{P(Y_j = 0)}\right] = \gamma + \beta x,$$

where γ and β are scalar-valued constants. **(10 marks)**

S5. If we wish to study the distribution of X, the number of albino children (or children with a rare anomaly) in families with proneness to produce such children, a convenient sampling method is first to discover an albino child and through it obtain the albino count X^w of the family to which it belongs. If the probability of detecting an albino is β, then the probability that a family with k albinos is recorded is $w(k) = 1 - (1-\beta)^k$, assuming the usual independence of Bernoulli trials. In such a case

$$p_{X^w}(k) = P(X^w = k) = \frac{w(k)P(X = k)}{\mathbb{E}[w(X)]}, \quad k = 0, 1, 2, \cdots$$

(a) Suppose X has the *Pascal Distribution*, that is

$$P(X = k) = \frac{\alpha^k}{(1+\alpha)^{k+1}}, \quad k = 0, 1, 2, \cdots$$

Find $\mathbb{E}(X)$ and show that

$$\lim_{\beta \to 0} \frac{w(k)}{\mathbb{E}[w(X)]} = \frac{k}{\mathbb{E}(X)}.$$

State clearly the assumptions you need to establish this result. **(7 marks)**

(b) Suppose β is small enough, such that the result of Part (b) is applicable. Is this probability distribution a member of Exponential family? Let $X_1^w, \cdots, X_n^w$ be a sample of size from p_{X^w}. Find a complete sufficient statistic for α. **(7 marks)**

(c) Using the asymptotic distribution of α find a 95% confidence interval for α.

(6 marks)

S6. Let $X_i \overset{iid}{\sim} N(\theta, 1)$, $i = 1, 2, \cdots, n$. Consider the sequence

$$\delta_n = \begin{cases} \overline{X}_n, & \text{if } |\overline{X}_n| \geqslant 1/n^{1/4}, \\ a\overline{X}_n, & \text{if } |\overline{X}_n| \leqslant 1/n^{1/4}. \end{cases}$$

Show that $\sqrt{n}(\delta_n - \theta) \xrightarrow{\mathcal{L}} N(0, \nu(\theta))$, where $\nu(\theta) = 1$ if $\theta \neq 0$ and $\nu(\theta) = a^2$ if $\theta = 0$. Is $\nu(\theta)$ greater than or equal to the information bound? (**Hint:** condition on $|\overline{X}_n|$).

(20 marks)