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 $f_{X;Y}$ : $R^2$ Ñ r 0; 8q, then the function g de ned by

$$g p x q \quad \text{»}_{\mathbb{R}} f p x; y q d y$$

is a probability density function for X. Hint. Recall the change of measure formula: if $_3$X has law then for any bounded measurable function f : R Ñ R, Epf pX qq f pxqd . (8 marks)

(c) Show that if f and g are two densities for X then the set tx : f pxq gpxqu has Lebesgue measure zero. (7 marks)

**P2.** In this question $pX_i; i$ ¥ 1q 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 Ñ 8 ? (5 marks)

(b) Show that there exist positive constants $a_1; a_2; :::$ such that $a_n X_n$ converges in distribution to 0. (5 marks)

(c) Let $X_1; X_2; :::$ be identically distributed random variables with nite second moment. Show that for all ¡ 0, nPr|$X_1$| ¥ ? ns Ñ 0. (5 marks)

(d) Let $X_1; X_2; :::$ be identically distributed random variables with nite second moment. Show th[illegible] {2 $\max_{1 ¤ k ¤ n}$ (5 marks)

1 Ñ 0 7.9701 Tf 6.585 0 Td [(1)]TJ/F38 7.9

## 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 $Y_1 \sim \mathrm{Beta}(\alpha_1, \alpha_2 + \alpha_3)$ and $Y_2 \sim \mathrm{Beta}(\alpha_2, \alpha_3)$ 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 $\mu > 0$ and $\lambda > 0$. Its density is given by

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

(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 71.9552 Tf 12.091 0 Td [(wh)-5(at)]TJ/F42 9 Tf -2.37 -1.841 Td ... (c)]TJ/F42 9.9626 Tf 11.956 0 Td [(that)]TJ

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

$$X_i \mid P_i \sim \text{Binomial}(n_i, P_i), \quad i = 1, \ldots, k;$$
$$P_i \sim \text{Beta}(\alpha, \beta).$$

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.

$$E\left[\{W - h(Z)\}\, g(Z)\right] = 0,$$

for any measurable function $g$ such that $E\{g(Z)\}$ 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 $X_{jk}$ $(j = 1, \ldots, m;\ k = 1, \ldots, r)$ have the form $X_{jk} = \mu + \eta_j + \varepsilon_{jk}$, where the $\eta_j$'s and the $\varepsilon_{jk}$'s are independently normally distributed with zero means and variances respectively $\sigma_b^2$ and $\sigma_w^2$. The unknown parameters are thus $(\mu, \sigma_b^2, \sigma_w^2)$. (10 marks)

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

$$\log \frac{P(Y_j = 1)}{P(Y_j = 0)} = \alpha + \beta x,$$

where $\alpha$ and $\beta$ 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 $\alpha$, then the probability that a family with $k$ albinos is recorded is $w(k) = 1-(1-\alpha)^k$, assuming the usual independence of Bernoulli trials. In such a case

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

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

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

Find $E(X)$ and show that

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

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

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

(c) Using the asymptotic distribution of $\hat{\theta}$ find a 95% confidence interval for $\theta$.

(6 marks)

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

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

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

(20 marks)