McGill University

Faculty of Science

Department of Mathematics and Statistics

Part A Examination

Statistics: Theory Paper

Date: 5th May 2015 Time: 1pm-5pm

Instructions

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

- ***You may use any result that is known to you, but you must state the name of the result (law/theorem/lemma/formula/inequality) that you are using, and show the work of verifying the condition(s) for that result to apply.***
- ***For the problems with multiple parts, you are allowed to assume the conclusion from the previous part in order to solve the next part, whether or not you have completed the previous part.***

P1. Assume that $\{A_n : n \geq 1\}$ is a sequence of **independent** events on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that, if $\alpha_n := \min\{\mathbb{P}(A_n), 1 - \mathbb{P}(A_n)\}$, then $\sum_{n\geq 1} \alpha_n = \infty$. Show that all the singletons of $(\Omega, \mathcal{F}, \mathbb{P})$ are $\mathbb{P}-$null sets, i.e., for every $\omega \in \Omega$ such that $\{\omega\} \in \mathcal{F}$, $\mathbb{P}(\{\omega\}) = 0$.

20 MARKS

P2. Let $\{X_n : n \geq 1\}$ be a sequence of **identically distributed** $\mathbb{R}$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\mathbb{E}[X_1^2] < \infty$.

Show that

(a) for every $\epsilon > 0$, $\lim\limits_{n\to\infty} n \cdot \mathbb{P}(|X_1| \geq \epsilon\sqrt{n}) = 0$;

(b) $\dfrac{1}{\sqrt{n}} \max\limits_{1\leq k\leq n} |X_k| \to 0$ in probability as $n \to \infty$.

20 MARKS

P3. Let $\{X_n : n \geq 1\}$ be a sequence of **independent** and **identically distributed** $\mathbb{R}$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\mathbb{E}[|X_1|] < \infty$. Set $S_n := \sum_{j=1}^{n} X_j$ for every $n \geq 1$.

Show that

(a) $\left\{\dfrac{S_n}{n} : n \geq 1\right\}$ is uniformly integrable;

(b) $\dfrac{S_n}{n} \to \mathbb{E}[X_1]$ in L^1, as $n \to \infty$.

20 MARKS

S1. Consider the probability distribution for random vector $\mathbf{X} = (X_1, ..., X_k)^\top$ with joint mass function given by

$$f_{X_1,...,X_k}(x_1, ..., x_k; \theta_1, \ldots, \theta_{k+1}) = \prod_{i=1}^{k+1} \theta_i^{x_i}$$

with support

$$\left\{ (x_1, ..., x_k) : x_i \in \{0, 1\}, \sum_{i=1}^{k} x_i \in \{0, 1\} \right\},$$

where x_{k+1} is defined by

$$x_{k+1} = 1 - (x_1 + \ldots + x_k).$$

The parameters of the distribution satisfy $0 \leq \theta_i \leq 1$ for all i, and

$$\sum_{i=1}^{k+1} \theta_i = 1.$$

(a) Express the joint probability mass function in Exponential Family form and identify the natural (or canonical) parameters.

4 MARKS

(b) Derive the expectation and variance-covariance matrix of $\mathbf{X}$.

6 MARKS

(c) Derive the (joint) moment generating function of $\mathbf{X}$, $M_{\mathbf{X}}(\mathbf{t})$.

4 MARKS

(d) Derive the marginal distribution of

$$T = \sum_{i=1}^{k} X_i.$$

2 MARKS

(e) Derive the joint distribution of

$$T_1 = X_1 + X_2 \qquad T_2 = X_3 + \cdots + X_k.$$

4 MARKS

NB. It is not sufficient to state the solutions without proof.

S2. (a) Suppose that Z_1 and Z_2 are independent and identically distributed Normal$(0,1)$ random variables. Let Y be the random variable defined by

$$Y = \mathbb{1}_{(0,\infty)}(Z_1 - Z_2) = \begin{cases} 1 & Z_1 > Z_2 \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbb{1}_A(.)$ is the indicator function for the set A. Let $W = \Phi(Z_1)$, where $\Phi(.)$ is the standard normal distribution function.

(i) Find the marginal distributions of Y and W.

3 MARKS

(ii) Find

$$\mathbb{E}_{W|Y}[W|Y = 1].$$

8 MARKS

(b) If Z is a Normal$(0,1)$ random variable, prove that

$$\mathrm{P}_Z[-3 < Z < 3] \geq \frac{8}{9}.$$

3 MARKS

(c) Consider the following three level hierarchical model

LEVEL 3 :	$r \in \{1, 2, \ldots\}$	Fixed parameter
LEVEL 2 :	$V \sim \text{Gamma}(r/2, r/2)$	
LEVEL 1 :	$X \mid V = v \sim \text{Normal}(0, v^{-1})$	

Find the kurtosis of X, κ, defined – whenever the relevant expectations are finite – by

$$\kappa = \frac{\mathbb{E}_X[(X-\mu)^4]}{\sigma^4}$$

where μ and σ^2 are the expectation and variance of X respectively.

6 MARKS

S3. (a) Suppose that continuous random variables $X_1, \ldots, X_n$ are independent and identically distributed with cdf F_X specified by

$$F_X(x) = \frac{x^2}{1+x^2} \qquad x > 0$$

and zero otherwise. Let Y_n be the maximum order statistic derived from $X_1, \ldots, X_n$.

Show that the probability density function of Y_n can be approximated, for large n, by the function

$$f(y) = \left(\frac{2n}{y^3}\right) \exp\{-n/y^2\} \qquad y > 0$$

and zero otherwise. 5 MARKS

(b) Suppose that $V_1, \ldots, V_n$ are independent and identically distributed random variables having a Poisson(λ) distribution. Let $W_i = \mathbb{1}_0(V_i)$ be the indicator random variable that takes the value one if $V_i = 0$, and zero otherwise, for $i = 1, \ldots, n$.

(i) Determine the asymptotic behaviour, as $n \longrightarrow \infty$, of the random variable

$$\overline{W}_n = \frac{1}{n} \sum_{i=1}^{n} W_i$$

and find a large sample approximation to the distribution of $\overline{W}_n$.

5 MARKS

(ii) Determine the asymptotic behaviour, as $n \longrightarrow \infty$, of the random variable

$$M_n = \frac{1}{n} \sum_{i=1}^{n} V_i(V_i - 1)$$

and find a large sample approximation to the distribution of M_n.

5 MARKS

(c) Suppose that $X_{11}, \ldots, X_{1n}$ are identically distributed random variables having an Exponential distribution with expectation μ_1, and $X_{21}, \ldots, X_{2n}$ are identically distributed random variables having an Exponential distribution with expectation μ_2, with all random variables mutually independent. Let

$$\overline{X}_{n1} = \frac{1}{n} \sum_{i=1}^{n} X_{1i} \qquad \qquad \overline{X}_{n2} = \frac{1}{n} \sum_{i=1}^{n} X_{2i}.$$

For large n, find a normal approximation to the distribution of

$$R_n = \frac{\overline{X}_{n1}}{\overline{X}_{n2}}.$$

5 MARKS

S4. Let $X_1, \ldots, X_n$ be i.i.d. random variables from a normal distribution Normal(μ, σ^2) with unknown parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Let $\theta = (\mu, \sigma^2)$, and let $\mathbf{X} = (X_1, \ldots, X_n)$.

(a) Show that

$$T(\mathbf{X}) = \frac{X_1 - \overline{X}_n}{S_n}$$

is an ancillary statistic, where $\overline{X}_n$ and S_n^2 are the sample mean and variance, respectively.

3 MARKS

(b) Prove that the statistic $T(\mathbf{X})$ in (a) is *independent* of the statistic $U(\mathbf{X}) = (\overline{X}_n, S_n^2)$. If using the results of any well-known theorem, clearly state the theorem.

4 MARKS

(c) Consider the probability $\eta(\theta) = \mathrm{P}(X_1 \leq c)$, for a known constant c.

(i) Using the results in (a) – (b), find the UMVUE of $\eta(\theta)$. Also, find the Cramér-Rao lower bound for the variance of any unbiased estimator of $\eta(\theta)$.

8 MARKS

(ii) Find the maximum likelihood estimator (MLE) of $\eta(\theta)$. Comment on the relationship between the MLE and the UMVUE when n becomes large.

5 MARKS

A fact to be used in part (c): Let $f_{T(\mathbf{X})}(t)$ be the pdf of the statistic $T(\mathbf{X})$. Then,

$$f_{T(\mathbf{X})}(t) > 0 \ , \ \ \textit{if} \ \ 0 < |t| < \frac{n-1}{\sqrt{n}}$$

and $f_{T(\mathbf{X})}(t) = 0$, otherwise.

S5. (a) Let X be a random variable with a pdf/pmf belonging to a parametric family $\mathcal{F}$ indexed by parameter θ, $\mathcal{F} = \{f(x;\theta) : \theta \in \Theta\}$.

Describe in detail the *monotone likelihood ratio property* for this family.

2 MARKS

(b) Consider a random variable X with the pdf

$$f(x;\theta) = \frac{\exp\{-(x-\theta)\}}{(1+\exp\{-(x-\theta)\})^2} \qquad ; \qquad x \in \mathbb{R} \ , \ \theta \in \mathbb{R}.$$

(i) Show that this family has a monotone likelihood ratio property.

2 MARKS

(ii) If X is a single observation from this pdf, find, giving precise details of the rejection region, the uniformly most powerful (UMP) test of size α based on X of

$$\begin{aligned} H_0 &: \theta \leq 0 \\ H_1 &: \theta > 0 \end{aligned}$$

5 MARKS

(c) Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from the Uniform$(\theta, \theta+1)$ distribution.

(i) Find a maximum likelihood estimator of θ.

3 MARKS

(ii) Consider a test of the hypotheses

$$\begin{aligned} H_0 &: \theta = 0 \\ H_1 &: \theta > 0. \end{aligned}$$

defined by the rejection region $\mathcal{R} = \{\mathbf{X} : X_{(n)} \geq 1 \text{ or } X_{(1)} \geq k\}$ for some k, where $(X_{(1)}, X_{(n)})$ are the minimum and maximum order statistics.

Find k such that this test has size α, and find its power function, $\beta(\theta)$.

Recall that for a random sample of size n from distribution with pdf f_X with support $\mathbb{X}$ and cdf F_X, the order statistics $(X_{(1)}, X_{(n)})$ have joint pdf

$$f_{X_{(1)},X_{(n)}}(t,u) = n(n-1)f_X(t)f_X(u)\{F_X(u) - F_X(t)\}^{n-2} \qquad (t,u) \in \mathbb{X}, t < u.$$

8 MARKS

S6. (a) Let $X_1, \ldots, X_n$ be i.i.d. random variables with a common pdf/pmf $f(x;\theta)$, and some unknown real-valued parameter θ.

(i) Describe in detail two methods for constructing *exact* confidence intervals with confidence coefficient $1-\alpha$ for the parameter θ based on $X_1, \ldots, X_n$. 2 MARKS

(ii) Suppose that

$$f(x;\theta) = \frac{1}{\Gamma(r)\theta^r} x^{r-1} e^{-x/\theta} \quad , \quad 0 < x < \infty$$

where $\theta > 0$ is unknown, and r is a known positive integer. Based on $X_1, \ldots, X_n$, construct the confidence interval for θ with confidence coefficient $1-\alpha$ that has the shortest (expected) length.

3 MARKS

(b) Consider using the likelihood ratio statistic $\lambda_n(\mathbf{X})$ for testing

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned}$$

based on a random sample of size n from a one parameter density $f(x;\theta)$.

Show that, under regularity conditions to be stated, the maximum likelihood estimator of θ, $\widehat{\theta}_n$, is asymptotically normally distributed. Also, show that under H_0, as $n \longrightarrow \infty$,

$$-2\log\lambda_n(\mathbf{X}) \xrightarrow{d} \chi^2_1$$

6 MARKS

You may assume that $\widehat{\theta}_n$ is a consistent estimator of θ under the stated regularity assumptions.

(c) Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with a common Poisson(λ) distribution, where $\lambda > 0$ is unknown. Let $\mathbf{X} = (X_1, \ldots, X_n)$, and $\tau(\lambda) = \Pr(X_1 = 1)$.

The UMVUE of $\tau(\lambda)$ is, for any finite n,

$$T(\mathbf{X}) = \widehat{\lambda}_n \left(\frac{n-1}{n}\right)^{n\widehat{\lambda}_n - 1}$$

where $\widehat{\lambda}_n$ is the maximum likelihood estimator (MLE) of λ.

(i) Find the asymptotic variance of $T(\mathbf{X})$. 2 MARKS

(ii) Find the asymptotic variance of the MLE of $\tau(\lambda)$, denoted $U(\mathbf{X})$. 2 MARKS

(iii) Using $U(\mathbf{X})$, construct an approximate confidence interval with confidence coefficient $1-\alpha$ for $\tau(\lambda)$. 5 MARKS