McGill University

Faculty of Science

Department of Mathematics and Statistics

Part A Examination

Statistics: Theory Paper

Date: Tuesday August 14, 2018 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. Show that the Borel $\sigma-$algebra on $\mathbb{R}$ is **not** generated by singletons on $\mathbb{R}$, i.e.,

$$\mathcal{B}(\mathbb{R}) \neq \sigma(\{\{x\} : x \in \mathbb{R}\}).$$

20 MARKS

P2. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, suppose that $X_n, Y_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ with $|X_n| \leq Y_n$ a.s. for every $n \geq 1$. Further assume that there exist random variables X and Y such that $X_n \to X$ a.s. and $Y_n \to Y$ a.s.. Prove that, if $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[Y]$, then $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \to X$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. (*Hint:* Consider $Y_n - |X_n|$).

20 MARKS

P3. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, suppose that $\{Y_n : n \geq 1\}$ is a sequence of **independent** and **identically distributed** random variables such that $\mathbb{E}[Y_1] = 1$ and Y_1 is not a.s. constant, i.e., $\mathbb{P}(Y_1 = 1) < 1$. Set $T_0 \equiv 1$ and $T_n := \Pi_{j=1}^{n} Y_j$ for every $n \geq 1$.

a) Show that $E[T_{n+1}|F_n] = T_n$ for every $n \geq 0$.

10 MARKS

b) Show that $\lim_{n\to\infty} T_n = 0$ a.s.. (*Hint:* Consider $\ln T_n$.)

10 MARKS

S1. Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind each of the other 2, a goat. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens one of the remaining doors behind which there is a goat. (Suppose this is door No. 3). He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? Explain. (Note: A "Yes" or "No" answer alone will not earn any marks.)

20 MARKS

S2. Suppose the number of defects per yard in a certain fabric, Y, has a Poisson distribution with parameter λ. The parameter λ is assumed to be a random variable with a density function given by

$$f(\lambda) = \begin{cases} e^{-\lambda}, & \text{if} \quad \lambda \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

a) Find the mean and the variance of Y. 10 MARKS

b) How likely is it that Y exceeds 9?

10 MARKS

S3. The variance stabilizing transformations are transformations for which the resulting statistic has an asymptotic variance that is independent of the parameters of interest. For each of the following cases, find the asymptotic distribution of the transformed statistic $\nu(\cdot)$ and show that it is variance stabilizing.

a) $T_n = \overline{X}_n$, where $X_i \overset{iid}{\sim}$ Poisson(λ) and $\nu(T_n) = \sqrt{T_n}$.

10 MARKS

b) $T_n = \overline{X}_n$, where $X_i \overset{iid}{\sim}$ Ber(p) and $\nu(T_n) = \arcsin\sqrt{T_n}$.

10 MARKS

S4. An experimenter wishes to study the number of a certain type of bacterium in samples of water. Assume that the number of this type of bacterium, say X, in a sample of water is a Poisson random variable with unknown parameter λ. Suppose that $X_1, X_2, \ldots, X_n$ are a random sample of bacteria counts in n water samples.

In what follows we are interested in the probability of observing no bacteria of this type in a randomly selected water sample, i.e. $\eta(\lambda) = P(X = 0) = e^{-\lambda}$.

a) Find the UMVUE of $\eta(\lambda)$. Call it $\delta(\boldsymbol{X})$, where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.

5 MARKS

b) Find the variance of $\delta(\boldsymbol{X})$, as well as the Cramer-Rao lower bound (CRLB) for the variance of any unbiased estimator of η.

5 MARKS

c) Using the result of part (b), find the relative efficiency of $\delta(\boldsymbol{X})$. Find the limit of the relative efficiency of $\delta(\boldsymbol{X})$, as $n \to \infty$, and explain your finding.

5 MARKS

d) Find the MLE of $\eta(\lambda)$. Using the MLE, construct an approximate $100(1-\alpha)\%$ confidence interval for $\eta(\lambda)$, for large n.

5 MARKS

S5. Let $X_1, X_2, \ldots, X_m$ be i.i.d. from $N(\mu_1, \sigma_1^2)$, and $Y_1, Y_2, \ldots, Y_n$ be i.i.d. from $N(\mu_2, \sigma_2^2)$. The two samples are independent. The unknown parameters are $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

(a) Find the MLE and UMVUE of $\eta = (\mu_1 - \mu_2)^2$. Show that both the MLE and UMVUE are consistent estimators of η, as $m, n \to \infty$.

5 MARKS

(b) Assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown. Suppose that we are interested in testing

$$H_0 : \mu_1 = \mu_2 \ , \ \ H_1 : \mu_1 \neq \mu_2.$$

Use the likelihood ratio statistic to design a test of size exactly equal to α for testing H_0 versus H_1.

5 MARKS

(c) Assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown. Find the shortest $100(1 - \alpha)\%$ confidence interval for the difference $\mu_1 - \mu_2$.

5 MARKS

(d) Using the likelihood ratio statistic, design a test for testing

$$H_0 : \sigma_1 = \sigma_2 = \sigma \ \text{(unknown)} \ , \ \ H_1 : \sigma_1 \neq \sigma_2$$

at a significance level α.

5 MARKS

S6. Let $Y_1, Y_2, \ldots, Y_n$ be independent Bernoulli random variables each with probability of success given as

$$P(Y_i = 1; t_i) = \frac{\exp(\alpha + \beta t_i)}{1 + \exp(\alpha + \beta t_i)} \;\; ; \;\; i = 1, 2, \ldots, n.$$

The unknown parameters are α and β, and $t_1, t_2, \ldots, t_n$ are some known constants.

a) Find a minimal sufficient statistic for $\boldsymbol{\theta} = (\alpha, \beta)$. 6 MARKS

b) It is known that there is no closed form for the MLE of $\boldsymbol{\theta}$. Explain, in full details, how the Newton-Raphson algorithm can be use to find an approximation to the MLE of $\boldsymbol{\theta}$.

7 MARKS

c) Assume the regularity conditions for this parametric family. Using Wald statistic, for large n: **(i)** construct an approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$; **(ii)** explain how to test $H_0 : \boldsymbol{\theta} = (\alpha_0, \beta_0)$ versus $H_1 : \boldsymbol{\theta} \neq (\alpha_0, \beta_0)$, at an approximate significance level α.

7 MARKS