McGill University

Faculty of Science

Department of Mathematics and Statistics

Part A Examination

Statistics: Theory Paper

Date: 10th 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. Let $\{X_n : n \geq 1\}$ be a sequence of **independent** and **identically distributed** $\mathbb{R}-$valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Further assume that X_n's are **NOT** a.e. constant, i.e., for every $c \in \mathbb{R}$, $\mathbb{P}(X_n = c) < 1$. Show that $\mathbb{P}(X_n > X_{n+1} \text{ i.o. }) = 1$. 20 MARKS

P2. Assume that $\{Y_n : n \geq 1\}$ is a sequence of **independent** and **identically distributed** $\mathbb{R}-$valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the common distribution

$$\mathbb{P}\left(Y_n = \frac{1}{2}\right) = \mathbb{P}\left(Y_n = \frac{3}{2}\right) = \frac{1}{2}.$$

Set $T_n := \Pi_{j=1}^n Y_j$ for every $n \geq 1$. Show that $\lim_{n\to\infty} T_n = 0$ a.s..
(*Hint:* Consider $\log T_n$.) 20 MARKS

P3. Let $\{Y_n : n \geq 1\}$ be a sequence of $\mathbb{R}$-valued random variables on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n : n \geq 0\}, \mathbb{P})$ such that, for every $n \geq 1$, $Y_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $Y_n \in m\mathcal{F}_n$ and $\mathbb{E}[Y_n | \mathcal{F}_{n-1}] = 0$. Further assume that $\sum_{n\geq 1} \frac{\mathbb{E}[Y_n^2]}{n^2} < \infty$. Set $X_0 \equiv 0$, and $X_n := \sum_{j=1}^n \left(\frac{Y_j}{j}\right)$ for every $n \geq 1$.

(i) Show that $\{X_n : n \geq 0\}$ is a martingale with respect to $\{\mathcal{F}_n : n \geq 0\}$. 8 MARKS

(ii) Prove that the Strong Law of Large Number holds for the sequence $\{Y_n : n \geq 1\}$, that is,

$$\frac{\sum_{j=1}^n Y_j}{n} \to 0 \text{ as } n \to \infty \text{ a.s..}$$

12 MARKS

S1. Suppose 5% of all people filing the long income tax form seek deductions that they know are illegal, while 2% incorrectly list deductions because they are unfamiliar with income tax regulations. Of the 5% who are guilty of cheating, 80% will deny knowledge of the error if confronted by an investigator. If the filer of the long form is confronted with an unwarranted deduction and he or she denies the knowledge of the error, what is the probability that he or she is guilty? 20 MARKS

S2. Let $X_1, \cdots, X_n$ be iid $N(\mu, \sigma^2)$.

(a) Using the method of moment generating functions, show that $\overline{X} \sim N(\mu, \sigma^2/n)$. *Hint:* If $X \sim N(\mu, \sigma^2)$, then $M_X(t) = \exp(\mu t + \sigma^2 t^2/2)$.) 5 MARKS

(b) Show that $(X_i - \mu)^2/\sigma^2 \sim \chi_1^2$. Deduce, using the method of moment generating functions that

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)},$$

where χ_k^2 is the Chi-squared distribution with k degrees of freedom. (*Hint*: If $W \sim \chi_k^2$, then $M_W(t) = (1-2t)^{-k/2}$ for $t < 1/2$). 5 MARKS

(c) Show that

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = n\left(\frac{\overline{X} - \mu}{\sigma}\right)^2 + \frac{(n-1)S^2}{\sigma^2},$$

where $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n - 1$ is the sample variance. 3 MARKS

(c) It is possible to show, when X_is are normally distributed, that $\overline{X}$ and $(X_1 - \overline{X}, \cdots, X_n - \overline{X})$ are independent. Using this fact and parts (a, b, c) show that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

7 MARKS

S3. Let X_i be independent $Bernoulli(p)$ random variables and let $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) Consider $Y_n\,(1 - Y_n)$, the estimate of variance. Show that

$$\sqrt{n}\,[Y_n\,(1-Y_n) - p(1-p)] \xrightarrow{D} \begin{cases} N\left[0, (1-2p)^2 p(1-p)\right], & \text{if} \quad p \neq 1/2, \\ 0, & \text{if} \quad p = 1/2. \end{cases}$$

10 MARKS

(b) Show that for $p = 1/2$, we have

$$n\left[\frac{1}{4} - Y_n\,(1-Y_n)\right] \xrightarrow{D} \frac{1}{4}\chi_1^2.$$

.

10 MARKS

S4. (a) Let $X_1, \ldots, X_n$ be i.i.d. random variables from an exponential distribution with pdf

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \;\;, \;\; x > 0$$

and some unknown parameter $\theta > 0$. Let $\mathbb{X} = (X_1, \ldots, X_n)$.

Show that $T(\mathbb{X}) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a complete sufficient statistic for θ. If using the results of any well-known theorem, clearly state the theorem.

2 MARKS

(b) Design a UMP test of size of $\alpha \in (0,1)$ for testing the hypotheses

$$H_0 : \theta \geq \theta_0 \;\;, \;\; H_1 : \theta < \theta_0$$

for some known value θ_0.

3 MARKS

(c) In details show that the statistic $T(\mathbb{X})$ in (a) is *independent* of the statistic $U(\mathbb{X}) = X_1/\overline{X}_n$. If using the results of any well-known theorem, clearly state the theorem.

3 MARKS

(d) Consider the probability $\eta(\theta) = P(X_1 > t)$, for some known constant $t > 0$. Using the results in (b)-(c), find the UMVUE of $\eta(\theta)$. Also, find the Cramér-Rao lower bound (CRLB) for the variance of any unbiased estimator of $\eta(\theta)$.

8 MARKS

(d) Find the maximum likelihood estimator (MLE) of $\eta(\theta) = P(X_1 > t)$, for some known constant $t > 0$. Show that both methods (the MLE and UMVUE) provide approximately the same estimate of $\eta(\theta)$ when the sample size is large.

4 MARKS

S5. (a) Suppose that the probability of death $p(x)$ is related to the dose x of a certain drug in the following manner

$$p(x) = \frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}$$

where $\beta_0 > 0$ and $\beta_1 \in \mathbb{R}$ are unknown. Denote the parameter vector $\theta = (\beta_0, \beta_1)$.

In an experiment, k different doses $x_1, x_2, \ldots, x_k$ of the drug are considered. Each dose x_i is applied to a number n_i of certain animals and the number Y_i of deaths among them is recorded. Note that $(x_, x_2, \ldots, x_k)$ and $(n_1, n_2, \ldots, n_k)$ are known constants and $Y_1, Y_2, \ldots, Y_k$ are independent binomial random variables such that

$$Y_i \sim Bin(n_i, p(x_i)) \ , \ \ i = 1, 2, \ldots, k.$$

(i) Find a minimal sufficient statistic for θ. Is your statistic complete?

4 MARKS

(ii) Using the maximum likelihood theory, describe (in details) an approximate procedure for testing the hypothesis $H_0 : \theta = \theta^*$, for some pre-specified value θ^*, at a significance level $\alpha \in (0, 1)$.

6 MARKS

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

(i) Find a maximum likelihood estimator of θ. Also find the moment estimator of θ.

4 MARKS

(ii) Are any of the estimators in Part (i) sufficient statistics for θ? Justify your answers.

2 MARKS

(iii) Consider the data

$$1.07, 1.11, 1.31, 1.51, 1.69, 1.72, 1.92, 2.24, 2.62, 2.98$$

which are randomly generated from the above uniform distribution with a known θ.

Are the sample mean and median of this data sensible point estimates of θ? Justify your answer. If your answer to the previous question is "no", provide a sensible estimate of θ.

4 MARKS

S6. Consider an experiment in which X is a random variable taking values either 1 or 2, according to the toss of a fair coin, and let Y be a random variable with a conditional distribution

$$(Y|X = x) \sim N(\theta, x)$$

for some unknown $\theta \in \mathbb{R}$, and $x = 1$ or 2.

Let $(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$ be an i.i.d random sample of (X, Y) which has a joint probability density function say $f(x, y; \theta)$. The following statistical inference is performed using this random sample.

(a) Find a minimal sufficient statistic for θ.

2 MARKS

(b) Find the MLE of θ. Assuming the standard regularity conditions for the above parametric family, write down the asymptotic distribution of the MLE, as $n \to \infty$.

5 MARKS

(c) Construct a Wald-type confidence interval with confidence coefficient $1 - \alpha$ for θ. Specify the mathematical forms of both the lower and upper bounds of the interval.

5 MARKS

(d) Using the likelihood ratio statistic construct a confidence interval with confidence coefficient $1 - \alpha$ for θ. Specify the mathematical forms of both the lower and upper bounds of the interval.

5 MARKS

(e) Comment on the (expected) length of the two intervals in Parts (c) and (d).

3 MARKS