Imperial College London

MATH95013

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2020

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Statistical Methods

Date: 29th April 2020

Time: 11.00am – 13.00pm (BST)

Time Allowed: 2 Hours

Upload Time Allowed: 30 Minutes

This paper has 4 Questions.

Candidates should start their solutions to each question on a new sheet of paper.

Each sheet of paper should have your CID, Question Number and Page Number on the top.

Only use 1 side of the paper.

Allow margins for marking.

Any required additional material(s) will be provided. Calculators may be used.

Credit will be given for all questions attempted.

Each question carries equal weight.

SUBMIT YOUR ANSWERS WITH COMPLETED COVERSHEETS WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

1. Choose one answer for each part. Partial credit may be awarded for working if an incorrect answer is selected. There is no negative marking.

(a) Question 1 on this exam consists of 5 multiple-choice subquestions, each with 5 choices. If you were to guess for each subquestion, we can model your answering of Question 1 as a sequence of 5 independent events. What is the probability that you get at least 4 subquestions correct if you guess at random? (4 marks)

(i) 0.00128, (ii) 0.00160, (iii) 0.00512, (iv) 0.00544, (v) 0.00672.

(b) For the probability density function

$$f(x) = \begin{cases} c\sqrt{x} & 0 \le x < 4, \\ 0 & \text{otherwise,} \end{cases}$$

what is the value of c? (4 marks)

(i) $3/16$, (ii) $16/3$, (iii) $2\sqrt{3}$, (iv) $\frac{1}{2\sqrt{3}}$, (v) $1/2$.

(c) Suppose $X \sim \text{Gamma}(\alpha_1, 1)$ and $Y \sim \text{Gamma}(\alpha_2, 1)$, with X and Y independent. Let $Z_1 = X + Y$ and $Z_2 = \frac{X}{X+Y}$. What is the support of (Z_1, Z_2)? (4 marks)

(i) $\mathbb{R} \times \mathbb{R}$, (ii) $\mathbb{R}^+ \times \mathbb{R}^+$, (iii) $\mathbb{R} \times (0, 1)$, (iv) $\mathbb{R}^+ \times (0, 1)$, (v) $(0, 1) \times (0, 1)$.

(d) Suppose $X \sim \text{N}(\mu = 0, \sigma^2 = 1)$ and let $X_n = -2X$ for $n = 1, 2, \dots$. Then, which **one** of the following statements is true? (4 marks)

(i) $X_n \xrightarrow{\mathcal{D}} 2X$ with $2X \sim \text{N}(\mu = 0, \sigma^2 = 2)$, (ii) $X_n \xrightarrow{\mathcal{P}} 2X$ with $2X \sim \text{N}(\mu = 0, \sigma^2 = 2)$,

(iii) $X_n \xrightarrow{\mathcal{D}} 2X$ with $2X \sim \text{N}(\mu = 0, \sigma^2 = 4)$, (iv) $X_n \xrightarrow{\mathcal{P}} 2X$ with $2X \sim \text{N}(\mu = 0, \sigma^2 = 4)$,

(v) $X_n \xrightarrow{\mathcal{D}} 2X$ with $2X \sim \text{N}(\mu = 0, \sigma^2 = 1)$.

(e) Suppose X is a continuous random variable with probability density function

$$f(x) = \begin{cases} 2x & 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Which of the following is the moment generating function, $M_X(t)$, of X? (4 marks)

(i) $2x\text{e}^t$, (ii) $\frac{2t^3}{3}\text{e}^t$, (iii) $\frac{2}{t^2}\left(t\text{e}^t - \text{e}^t + 1\right)$, (iv) $\frac{2}{t^2}\left(t\text{e}^t - \text{e}^t\right)$, (v) $\frac{1}{t^2}\left(t\text{e}^t - \text{e}^t\right)$.

(Total: 20 marks)

2. (a) Imperial College wishes to conduct a survey on undergraduate student satisfaction. 100 currently enrolled undergraduate students, randomly selected, are asked whether they are overall satisfied with their education. There are only two possible responses, yes and no, and 71 students answer that, yes, they are satisfied.

(i) What is the population to which inferences from the sample of 100 students may be generalized? Also, if p represents the proportion of the population who are overall satisfied, suggest a model for these data. (2 marks)

(ii) Derive the *maximum likelihood estimator* for p. Is this estimator unbiased? (4 marks)

(iii) Now, find the *maximum likelihood estimate* of p and give an approximate 95% confidence interval (CI) for p. Interpret the CI. (Please give the CI to 3 decimal places.) (4 marks)

You may find the following statistical table for the standard normal distribution helpful:

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0	0.5	0.9	0.816	1.8	0.964	2.8	0.997
0.1	0.540	1.0	0.841	1.9	0.971	3.0	0.999
0.2	0.579	1.1	0.864	2.0	0.977	3.5	0.9998
0.3	0.618	1.2	0.885	2.1	0.982	1.282	0.9
0.4	0.655	1.3	0.903	2.2	0.986	1.645	0.95
0.5	0.691	1.4	0.919	2.3	0.989	1.96	0.975
0.6	0.726	1.5	0.933	2.4	0.992	2.326	0.99
0.7	0.758	1.6	0.945	2.5	0.994	2.576	0.995
0.8	0.788	1.7	0.955	2.6	0.995	3.09	0.999

(b) Suppose $X|M = m \sim \text{Binomial}(m, \alpha)$ and $M \sim \text{Poisson}(\lambda)$.

(i) What is the joint probability mass function of X and M? (4 marks)

(ii) What is the marginal distribution of X? What is $\text{E}(X)$? (Please show *all* your work.) (6 marks)

(Total: 20 marks)

3. Astrophysicists often study distant objects by counting the number of photons at specific energies recorded by a detector. The data produced consist of photon counts from a number of energy bins. We refer to the (mean) energy in bin i as E_i (in units of kiloelectron volts, keV), and the count in bin i as Y_i, for $i = 1, \ldots, n$. Thus, both E_i and Y_i are known, for $i = 1, \ldots, n$ (the E_i are known constants). We can then model the photon counts as independent Poisson random variables:

$$Y_i | \alpha, \beta \sim \mathrm{Poisson}(\Lambda_i),$$

where $\Lambda_i = \alpha E_i^{-\beta}$ and (α, β) are model parameters of primary scientific interest. Suppose we further know $\beta = 1.6$ such that there is only one unknown model parameter, α.

(a) Write down the likelihood function (as a function of α). (4 marks)

(b) Derive the *maximum likelihood estimator* for α, denoted $\hat{\alpha}_{\mathrm{MLE}}$. Is $\hat{\alpha}_{\mathrm{MLE}}$ unbiased? (4 marks)

(c) For the actual observed data, there are $n = 883$ energy bins $E_1, \ldots, E_{883}$. Now that we have proposed and fitted a model, we may want to check our model. If you had access to the observed data, how could you formally check whether the fitted model could have reasonably generated the data? (Please give a description of the steps involved and relevant equations. You obviously cannot perform the necessary calculations without access to the observed data.) (4 marks)

(d) Suppose that we now treat α as a random variable and specify a prior distribution $f(\alpha) \propto 1$, i.e.,

$$\alpha \sim \mathrm{Uniform}(0, \infty).$$

(Recall that this is an *improper* distribution as the probability density function, $f(\alpha)$, does not integrate to one.) What is the distribution of $\alpha | \underline{\mathrm{Y}}$, where $\underline{\mathrm{Y}} = (Y_1, \ldots, Y_n)$? This is the *posterior distribution*. (4 marks)

(e) Consider a new estimator for α, $\hat{\alpha}_{\mathrm{Bayes}} = \mathrm{E}(\alpha | \underline{\mathrm{Y}})$. Give an equation for $\hat{\alpha}_{\mathrm{Bayes}}$, using the prior distribution defined in (d). Is $\hat{\alpha}_{\mathrm{Bayes}}$ an unbiased estimator for α? How do $\hat{\alpha}_{\mathrm{Bayes}}$ and $\hat{\alpha}_{\mathrm{MLE}}$ compare as the amount of data increases? Finally, suggest a method for quantifying the uncertainty on the estimate yielded from $\hat{\alpha}_{\mathrm{Bayes}}$. (4 marks)

(Total: 20 marks)

4. (a) Suppose X and Y are continuous random variables with joint probability density function

$$f_{XY}(x,y) = \begin{cases} c(x^2+y^2) & 0<y<1,\, 0<x<1, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a normalizing constant that does not depend on x or y.

(i) Find the value of c and then compute $P(X < \frac{1}{2}, Y > \frac{1}{2})$. (2 marks)

(ii) Derive the marginal density of X, $f_X(x)$, and then derive $\mathrm{E}(X)$. (4 marks)

(iii) Derive $f_{Y|X}(y|x)$, compute $P(Y < \frac{3}{4} | X = \frac{1}{3})$, and derive $\mathrm{E}(Y|X = \frac{1}{3})$. (4 marks)

(b) Let W and R be independent random variables with $W \sim \mathrm{Beta}(\alpha, \beta)$ and $R \sim \mathrm{Beta}(\alpha+\beta, \theta)$.

(i) What is the joint probability density function of (W, R)? (2 marks)

(ii) Let $U = WR$ and $V = W$. What is the joint probability density function of (U, V)? (4 marks)

(iii) Show that $U \sim \mathrm{Beta}(\alpha, \beta+\theta)$. (4 marks)

(Total: 20 marks)

DISCRETE DISTRIBUTIONS

	range $\mathbb{X}$	*parameters*	*pmf* f_X	*cdf* F_X	$\mathsf{E}[X]$	$\mathsf{Var}[X]$	*mgf* M_X
$Bernoulli(\theta)$	$\{0,1\}$	$\theta \in (0,1)$	$\theta^x(1-\theta)^{1-x}$		θ	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$

CONTINUOUS DISTRIBUTIONS

		parameters	*pdf*	*cdf*	E$[X]$	Var$[X]$	*mgf*
$Uniform(\alpha,\beta)$	(α,β)	$\alpha<\beta\in\mathbb{R}$	$\dfrac{1}{\beta-\alpha}$	$\dfrac{x-\alpha}{\beta-\alpha}$	$\dfrac{(\alpha+\beta)}{2}$	$\dfrac{(\beta-\alpha)^2}{12}$	$\dfrac{e^{\beta t}-e^{\alpha t}}{t\,(\beta-\alpha)}$
$Exponential(\lambda)$	$\mathbb{R}^+$	$\lambda\in\mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda-t}\right)$
$Gamma(\alpha,\beta)$	$\mathbb{R}^+$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$		$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	$\left(\dfrac{\beta}{\beta-t}\right)^\alpha$
$Normal(\mu,\sigma^2)$	$\mathbb{R}$	$\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t+\sigma^2t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\dfrac{\nu+1}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\left\{1+\dfrac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		0 (if $\nu>1$)	$\dfrac{\nu}{\nu-2}$ (if $\nu>2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$		$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx$ is the *gamma function*, and $\Gamma(\frac{1}{2})=\sqrt{\pi}$.