OCR
Oxford Cambridge and RSA

Friday 23 June 2017 – Morning

A2 GCE MATHEMATICS (MEI)

4769/01 Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4769/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found inside the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer any **three** questions.
- Do **not** write in the barcodes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER/INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

Option 1: Estimation

1 The random variable X has the probability density function

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0,$$

where λ is a positive parameter. $X_1, X_2, \ldots, X_n$ are independent random variables distributed as X.

(i) Show that $\tilde{\lambda}$, the maximum likelihood estimator for λ, is $(\bar{X})^{-1}$ where $\bar{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

(You should justify that this gives the **maximum** likelihood.) **[10]**

You are given that

$$E(\tilde{\lambda}) = \frac{n}{n-1}\lambda$$

for $n > 1$ and

$$\mathrm{Var}(\tilde{\lambda}) \approx -\left(\frac{d^2(\ln L)}{d\lambda^2}\right)^{-1}$$

when n is large and where L is the likelihood used in part **(i)**.

(ii) Show that the mean square error in $\tilde{\lambda}$ is approximately

$$\lambda^2\left(\frac{1}{n} + \frac{1}{(n-1)^2}\right),$$

when n is large. **[5]**

(iii) Find the value of k for which $\hat{\lambda} = k(\bar{X})^{-1}$ is an unbiased estimator of λ. **[3]**

(iv) Assuming n is large, show that $\hat{\lambda}$ is a more efficient estimator than $\tilde{\lambda}$. **[6]**

Option 2: Generating Functions

2 The random variable X has a Poisson distribution with parameter λ.

(i) Show that the probability generating function of X is

$$G_X(t) = e^{\lambda(t-1)}.$$ **[4]**

The random variable Y is the sum of n independent random variables each distributed as X.

(ii) Find $G_Y(t)$, the probability generating function of Y, and hence state the distribution of Y, including its parameter(s). **[3]**

(iii) Explain why

$$G_Y(-1) = P(Y \text{ is even}) - P(Y \text{ is odd})$$

and hence find the probability that Y takes an even value. **[6]**

(iv) A discrete random variable has probability generating function $G(t)$ and moment generating function $M(t)$. Show that $M(t) = G(e^t)$. **[2]**

(v) By considering the moment generating function of Y, show that $Z = \frac{Y - \lambda n}{\sqrt{\lambda n}}$ has approximately the Normal distribution $N(0, 1)$ when λn is large. **[9]**

Option 3: Inference

3 Two randomly selected groups of people were asked by a psychologist to complete a memory test. Group A took the test using distinctive smells which the psychologist thinks will improve recall. Group B took the test without any distinctive smells being present. The results for the two groups (using arbitrary units) are given below. Higher numbers indicate better recall.

Group A scores:	25.8	27.4	26.2	23.5	28.3	26.4	27.2
Group B scores:	25.6	24.9	23.7	25.9	26.9		

(i) Use a t test at the 5% significance level to investigate whether the presence of distinctive smells improves performance in memory tests. **[10]**

(ii) State two distributional assumptions required in using the t test for these data. **[2]**

(iii) If the assumptions in part **(ii)** are not justified, a non-parametric test could be used. Carry out this alternative test at the 5% significance level to investigate whether the presence of distinctive smells improves performance in memory tests. **[11]**

(iv) If the assumptions in part **(ii)** are justified, why is it preferable to use a t test rather than a non-parametric test? **[1]**

Option 4: Design and Analysis of Experiments

4 **(i)** What is meant by randomised block design? Describe an experimental situation in which it would be appropriate to use randomised block design. Justify why randomised block design is appropriate in this situation. **[5]**

(ii) The usual one-way analysis of variance model can be written as

$$x_{ij} = \mu + \alpha_i + e_{ij} .$$

Interpret the terms in the model. State any distributional assumptions required in the e_{ij} term. **[5]**

Two drugs and a placebo are used in a trial on 15 patients, allocated randomly to three different groups. The level of a hormone in the blood, x (in arbitrary units), is measured two hours after the treatment is administered. The results are shown in the table below.

	Placebo	Drug A	Drug B
Hormone level	101	102	94
	105	94	90
	98	86	108
	88	108	100
	87	110	106
$\sum x$	479	500	498
$\sum x^2$	46143	50400	49836

(iii) Conduct the usual one-way analysis of variance test at the 5% significance level to investigate whether there is a difference between hormone levels in the three situations. State your hypotheses carefully and report briefly your conclusions. **[12]**

(iv) Jamie thinks that the analysis of variance test used in part **(iii)** may not be valid. He uses a goodness of fit test at the 5% significance level to investigate whether a Normal distribution can be fitted to the 15 data items. He concludes that a Normal distribution does not fit and he claims that the analysis of variance is therefore not appropriate.

Explain why, although Jamie has carried out a goodness of fit test correctly, his claim about analysis of variance being inappropriate is not correct. **[2]**

END OF QUESTION PAPER

OCR
Oxford Cambridge and RSA

