OCR

RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2644

Probability & Statistics 4

Friday **17 JANUARY 2003** Afternoon 1 hour 20 minutes

Additional materials:
Answer booklet
Graph paper
List of Formulae (MF8)

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

- Write your Name, Centre Number and Candidate Number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- Give non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question or is clearly appropriate.
- You are permitted to use a graphic calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is 60.
- Questions carrying smaller numbers of marks are printed earlier in the paper, and questions carrying larger numbers of marks later in the paper.
- **You are reminded of the need for clear presentation in your answers.**

This question paper consists of 4 printed pages.

[Turn over

OCR PROB & STATS 4

1 The probability generating function of the random variable X is given by $G(t) = e^{2t-a}$, where a is a constant. Find

(i) the value of a, [2]

(ii) $E(X)$. [2]

2 At a particular restaurant 55% of customers order an aperitif and 80% order wine. 30% of customers order wine without ordering an aperitif.

(i) Find the percentage of customers who order both an aperitif and wine. [2]

(ii) Given that a customer ordered wine, find the probability that the customer ordered an aperitif. [2]

(iii) 40% of customers order a starter with their meal, independently of whether they order an aperitif and independently of whether they order wine. Show that at most 29% of customers order an aperitif, a starter and wine. [3]

3 (i) Find the number of different arrangements of these letters.

X X X Y Y Y Y Y [2]

The continuous random variables X and Y have identical distributions. Three observations of X and five observations of Y are taken and placed in rank order, smallest first. The sum of the ranks of the values of X is denoted by R.

(ii) Assuming that all arrangements are equally likely, find $P(R \leqslant 8)$. [2]

A chemical is produced in two factories, A and B. A customer who obtains the chemical from both factories suspects that there is a greater amount of impurity in the chemical produced by factory A than in the chemical produced by factory B. An analysis of five random samples from factory A and three random samples from factory B gave the following percentages of impurity.

Factory A	3.5	4.7	3.8	5.2	4.6
Factory B	2.6	1.8	4.5		

(iii) Use a suitable non-parametric test, at the 5% significance level, to decide whether there is evidence to support the customer's suspicion. [4]

— JANUARY 2003

4 The discrete random variables X and Y have the joint probability distribution given in the following table.

		X		
		0	1	2
Y	0	0	0	0.1
	1	0	0.2	0.3
	2	0.1	0.3	0

(i) Show that $\mathrm{Var}(X) = 0.41$ and write down $\mathrm{Var}(Y)$. [4]

(ii) Tabulate the probability distribution of T, where $T = X + Y$, and use it to find $\mathrm{Var}(T)$. [3]

(iii) Hence or otherwise find $\mathrm{Cov}(X, Y)$ and state, giving a reason, whether X and Y are independent. [3]

5 The time, X months, to the first breakdown of a particular make of washing machine may be modelled by a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \dfrac{4a^4}{x^5} & x \geqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where a is an unknown positive constant.

(i) Find $\mathrm{E}(X)$. [2]

Two randomly chosen washing machines break down independently after X_1 and X_2 months.

(ii) Show that T, where $T = \lambda X_1 + (\frac{3}{4} - \lambda)X_2$, is an unbiased estimator of a for any value of the constant λ. [2]

(iii) Find the value of λ for which the variance of T is a minimum. [3]

(iv) Given that $X_1 = 3.45$ and $X_2 = 1.75$, estimate a using this value of λ. Explain why, in this case, this is not a useful estimate of a. [2]

6 The random variable X has a probability density function which involves a parameter α. The moment generating function of X is $(1 - 2t)^{-\alpha}$. Find

(i) $\mathrm{E}(X)$, [2]

(ii) $\mathrm{Var}(X)$. [2]

This distribution, with parameter α, is denoted by $\mathrm{D}(\alpha)$. The sum of three observations of a random variable having a $\mathrm{D}(2)$ distribution and one observation of a random variable having a $\mathrm{D}(1)$ distribution is denoted by S. All observations are independent.

(iii) Write down, in simplified form, the moment generating function of S, and hence identify the distribution of S in terms of D. [3]

(iv) Find $\mathrm{E}(S^3)$. [3]

7 Portfolios of students' work, used for assessment on a university course, are each marked independently by two lecturers, Ms Abel and Mr Barnes. The marks given to a random sample of 10 students in the year 2002 are as follows.

Student	1	2	3	4	5	6	7	8	9	10
Ms Abel's mark	37	48	62	24	53	72	41	29	53	24
Mr Barnes' mark	41	54	74	26	66	62	34	32	62	29

(i) Use an appropriate Wilcoxon test, at the 5% significance level, to decide whether the average marks given by Ms Abel and Mr Barnes are different. [5]

(ii) Why, in this case, is a Wilcoxon test preferred to a sign test? [1]

The final mark awarded is the sum of the two marks, as shown in the table below.

Student	1	2	3	4	5	6	7	8	9	10
Final mark	78	102	136	50	119	134	75	61	115	53

The median mark of all students for this portfolio in 2001 was 120.

(iii) Use a sign test, at the 10% significance level, to decide whether there is evidence that the median mark has decreased. [5]

(iv) Give a reason why, in this case, a sign test may be more appropriate than a Wilcoxon test. [1]

JAN 2003 (S4)

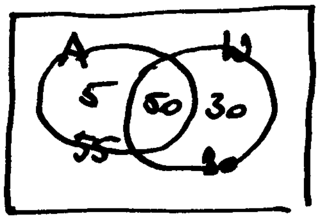
① (i) $G(1)=1$ so $e^{2-a}=1$ and $\underline{a=2}$ [2]

(ii) $G'(t)=2e^{2t-a}$ so $E(X)=G'(1)=\underline{\underline{2}}$ [2]

[4]

---

②



(i) 50% [2]

(ii) 5/8 [2]

(iii) $P(S\cap A\cap W) = -P(S)-P(A)-P(W)$
$+P(S\cap A)+P(S\cap W)+P(A\cap W)+P(S\cup A\cup W)$
$=P(S\cup A\cup W)-(.55+.8+.4)+(.32+.22+.5)$
$=P(S\cup A\cup W)-0.71$ but $P(S\cup A\cup W)\le 1$
$\therefore P(S\cap A\cap W)\le 1-0.71=0.29$ [3]

[7]

---

③ (i) $\dfrac{8!}{3!\,5!}=\underline{56}$ [2]

(ii) 123, 124, 125, 134 so $4/56=\underline{1/14}$ [2]

(iii) BBAABAAA so $R_B=8$, $R_A=19$

H_0: Distributions of impurities identical between A and B

H_1: Factory A has higher average.

$1/14>1/20$ [or $8\ge 7$] so do not reject H_0.

Insufficient evidence that more impurity in chemical from factory A. [4]

[8]

---

④ (i)

x	0	1	2
$P(X=x)$	0.1	0.5	0.4

$E(X)=1.3$

$V(X)=2.1-1.3^2=0.41$ [4]

(ii)

t	2	3
$P(T=t)$	0.4	0.6

$E(T)=2.6$

$V(T)=7.0-2.6^2=0.24$ [3]

(iii) $V(T)=V(X)+V(Y)+2\text{Cov}(X,Y)$

$0.24=0.41+0.41+2\text{Cov}(X,Y)$

$\text{Cov}(X,Y)=-0.29$

(iv) $\text{Cov}(X,Y)\ne 0$ so not independent [3]

[10]

---

⑤ (i) $\displaystyle\int_a^\infty 4a^4x^{-4}\,dx=\left[\frac{-4a^4}{3x^3}\right]_a^\infty=\frac{4a}{3}$ [2]

(ii) $\lambda\cdot\dfrac{4a}{3}+\left(\dfrac{3}{4}-\lambda\right)\dfrac{4a}{3}=a$ so unbiased [2]

(iii) $Var=\left\{\lambda^2+\left(\frac{3}{4}-\lambda\right)^2\right\}\sigma^2=\left(2\lambda^2-\frac{3\lambda}{2}+\frac{9}{16}\right)\sigma^2$

Differentiate: min. for $4\lambda=\frac{3}{2}$, $\lambda=\frac{3}{8}$. [3]

(iv) $\frac{3}{8}\times 3.45+\frac{3}{8}\times 1.75=1.95$

but this is bigger than 1.75 and clearly $a\le 1.75$. [2]

[9]

---

⑥ (i) $M'(0)=2\alpha$ [2]

(ii) $M''(0)=4\alpha(\alpha+1)$

$\therefore V(X)=4\alpha^2+4\alpha-4\alpha^2=4\alpha$ [2]

(iii) $[(1-2t)^{-2}]^3[(1-2t)^{-1}]=(1-2t)^{-7}=D(7)$ [3]

(iv) $\dfrac{d^3}{dt^3}(1-2t)^{-7}=-7\cdot-8\cdot-9\cdot(-2)^3(1-2t)^{-10}$ [3]

$\therefore M^{(3)}(0)=4032$

[10]

---

⑦

Diff	4	6	12	2	13	−10	−7	3	9	5
Rank	3	5	9	1	10	−8	−6	2	7	4

$Q=14$, $P=41$ so $T=14$

H_0: median difference is 0 (population median)

H_1: median difference $\ne 0$.

Critical value is 8. $14>8$ so do not reject H_0.

Insufficient evidence that average marks differ. [5]

(ii) Wilcoxon test uses more information. [1]

(iii) − − + − − + − − − −

$B(10, 0.5)$: $P(\le 2)=0.0547$
<0.1

Reject H_0. Significant evidence that median mark has decreased.

[H_0: $m=120$. H_1: $m<120$, where m is the population median] [5]

(iv) To use ~~Wilcoxon~~ we would need to know that the ~~distribution~~ distribution of marks was symmetric. [1]

[12]