OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 2644

Probability & Statistics 4

Wednesday **23 JUNE 2004** 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 3 printed pages and 1 blank page.

1 For the events A and B, $\mathrm{P}(A) = \frac{1}{3}$, $\mathrm{P}(B) = \alpha$ and $\mathrm{P}(A \cup B) = \frac{8}{9}$.

(i) For what value of α are A and B mutually exclusive? [2]

(ii) For what value of α are A and B independent? [3]

2 In a certain large population of car owners, the proportion that have fully comprehensive insurance is p. In a random sample of n owners from this population, the number that have fully comprehensive insurance is denoted by X.

(i) Show that $Y = \dfrac{X}{n}$ is an unbiased estimator of p. [3]

(ii) A particular random sample of 150 car owners contains 94 who have fully comprehensive insurance. Use this information to calculate an estimate of the variance of Y. [2]

3 The effect of a certain drug, intended to increase the rate of heart-beat, was measured on 16 volunteers. These volunteers were divided into two groups of 8. The first group was treated with the drug and the second group was treated with a placebo (i.e. a treatment not containing a drug). The heart beats were measured one hour after treatment. After a week the experiment was repeated with the same volunteers, this time with the treatments reversed. The results, in beats per minute, for the 16 patients were as follows.

Volunteer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Drug	102	86	75	81	61	98	69	79	58	71	80	69	84	92	70	94
Placebo	98	84	73	89	66	84	88	71	56	63	98	63	70	81	61	89

A test is to be carried out, at the 5% significance level, of whether a majority of people treated with the drug would have an increased heart-beat rate.

(i) Carry out the test using the sign test. [6]

(ii) State an advantage of using the Wilcoxon signed-rank test rather than the sign test. [1]

4 The continuous random variable X has probability density function given by

$$\mathrm{f}(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda(x-a)} & x \geqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where a and λ are constants.

(i) Show that the moment generating function of X is $\dfrac{\lambda \mathrm{e}^{at}}{\lambda - t}$. [3]

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

(iii) Show that $\mathrm{Var}(X)$ is independent of a. [4]

5 Three unbiased dice, A, B and C, are thrown together. The total number of sixes occurring on dice A and B is denoted by X and the total number of sixes occurring on dice B and C is denoted by Y. The joint probability distribution of X and Y is given in the following table.

		X		
		0	1	2
Y	0	$\frac{125}{216}$	$\frac{25}{216}$	0
	1	$\frac{25}{216}$	$\frac{30}{216}$	a
	2	0	a	b

(i) Show that $a = \frac{5}{216}$ and find the value of b. [3]

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

(iii) Find $\mathrm{Cov}(X, Y)$. [4]

(iv) Give a reason why $\mathrm{Var}(X - Y)$ is not equal to $\mathrm{Var}(X) + \mathrm{Var}(Y)$. [1]

6 A bus company operates one service each day between Leeds and Leicester, and records show that during 2003 the median time for the journey was 207.5 minutes. For 2004, the times, T minutes, of the first 80 journeys were used to test whether the median time had changed. It was decided to use the Wilcoxon signed-rank test with a significance level of 2%. The values of $T - 207.5$ were found (none of which was zero) and the values of $|T - 207.5|$ were ranked, smallest first. The sum of the ranks corresponding to the positive differences was 1114.

(i) Assuming that the sample is representative of the times for the whole year, carry out the test. [8]

(ii) What distributional assumption is required for the validity of the test? [1]

(iii) Comment on the validity of the assumptions made in parts (i) and (ii). [2]

7 The probability generating function of a discrete random variable R is given by

$$\mathrm{G}_R(t) = t^3(a + bt)^3,$$

where a and b are constants. It is given that $\mathrm{E}(R) = 4$.

(i) Show that $b = \frac{1}{3}$ and find the value of a. [6]

(ii) Find $\mathrm{Var}(R)$. [3]

The sum of n independent observations of R is denoted by S.

(iii) Show that

$$\mathrm{P}(S = 3n + 3) = \frac{2^{3n-4}n(3n-1)(3n-2)}{3^{3n}}.$$ [4]

1	(i)	Use $P(A \cap B)=0$ and			
		$P(A \cup B)=P(A)+P(B)-P(A \cap B)$	M1		
		$\alpha = {}^5/_9$	A1	**2**	
	(ii)	Use $P(A \cap B)=P(A)P(B)$			
		or conditional probability formula	M1		
		$\frac{1}{3} + \alpha - {}^8/_9 = \frac{1}{3}\alpha$	A1		aef
		$\alpha = {}^5/_6$	A1	**3**	
2	(i)	$E(Y)=E(X/n)=E(X)/n$	M1		
		$=np/n$	A1		
		$=p$ so Y is an unbiased estimate of p	A1	**3**	Must have statement
	(ii)	$Var(Y)=p(1-p)/n$	M1		Correct formula with est(p)
		0.00156 (3SF)	A1	**2**	
3	(i)	H_0: $p = \frac{1}{2}$, or $p \le \frac{1}{2}$, H_1: $p > \frac{1}{2}$ (where p is the population proportion whose heart-beat rate increases after drug)	B1		
		Form differences ,D-P (or P-D) and count the number of + (or -) values, X	M1		
		Obtain 12 (or 4)	A1		
		Attempt to find $P(X \ge 12)$ (or $X \le 4$)			
		EITHER: from table ; giving 0.0384	M1A1√		ft 12 or 4
		Compare correctly with 0.05 and Reject H_0, and accept that heart-beat rate increased in a majority of people taking drug	B1		ft 0.0384 Only from tail prob
		OR: z-value with cc	M1		
		1.75	A1		
		Compare correctly with 1.645 and state conclusion as above.	B1√		ft 1.75 Only from tail prob
		OR: CR from table ; $X \ge 12$ (or $X \le 4$)	M1A1		
		4 (or 12) is in CR, so reject H_0 and conclusion as above	B1√	**6**	
	(ii)	More information is used	B1	**1**	Magnitude of differences

4	(i)	$\int_a^\infty \lambda e^{-\lambda(x-a)} e^{xt}\,dx$			
		Integral with correct limits	B1		
		Correctly integrated, $[\lambda/(\lambda-t)]e^{-x\lambda+xt+\lambda a}$	B1		
		Given answer obtained convincingly	B1	**3**	
	(ii)	EITHER [α]			
		Reasonable attempt at $M'(t)$ with intention find $M'(0)$, $\lambda a e^{at}/(\lambda-t)+\lambda e^{at}/(\lambda-t)^2$	M1		Using quotient or product rule.
		OR [β] (See (iii)[β])			
		Attempt to expand mgf to find coefficient of t	M1		
		OR[γ]			
		Attempt to use pdf, correct form and limits	M1		
		$a + 1/\lambda$	A1	**2**	
	(iii)	EITHER [α]:			
		$M''(t)$			
		$[\lambda a^2/(\lambda-t) + 2\lambda a/(\lambda-t)^2 + 2\lambda/(\lambda-t)^3]e^{at}$	M1A1		Reasonable attempt M1
		$\text{Var}(X)=M''(0) -[M'(0)]^2$	M1		With attempt to evaluate
		$1/\lambda^2$, (independent of a)	A1	**4**	
		OR [β] which includes (ii)			
		$M(t)=(1+t/\lambda+t^2/\lambda^2)(1+at+\frac{1}{2}a^2t^2)$	M1A1		M1 for at least 3 correct terms
		$E(X)=a+1/\lambda$	B1		
		$E(X^2)=a^2+2a/\lambda + 2/\lambda^2$	B1		
		$\text{Var}(X)$ as in [α]	M1A1	**6**	
		OR [γ]			
		Change variable to y, where $y=x-a$	M1		
		$f(y) = \lambda e^{-\lambda y}$,	A1		
		$\text{Var}(Y)=\text{Var}(X)$ and $f(y)$ is independent of a	M1		
		So $\text{Var}(X)$ is independent of a	A1	**4**	Or by integration to find $E(X^2)$
5	(i)	$P(X=2,Y=1) =P(A6,B6,C\ \text{not}6)$	M1		With attempt to evaluate
		$=(1/6)(1/6)(5/6)=5/216$ AG	A1		
		$b=1/216$	B1	**3**	Could be found first
	(ii)	$p(0)=150/216, p(1)=60/216, p(2)=6/216$	B1		aef
		$E(X)=\frac{1}{3}$	B1√	**2**	ft marginal distribution
	(iii)	$E(XY)=30/216+20/216+2/216\ (=1/4)$	M1A1		M1 with one error
		$E(XY)-E(X)E(Y)$	M1		With attempt to evaluate
		5/36	A1	**4**	Accept 0.139
	(iv)	Covariance is not zero	B1	**1**	

6	(i)	H_0: m=207.5 H_1: $m \neq 207.5$	B1		Or in words
		P~N(μ, σ^2);with			
		μ=80××81/4, σ^2=80×81×161/24	M1		At least one correct parameter
		N(1620,43470)	A1		May be implied later
		P(P≤1114); z=(1114.5-μ)/σ	M1		With or without or with wrong cc
		Correct expression	A1		
		z =-2.425 OR –2.427	A1		
		EITHER[α]			
		Compare correctly with –2.326 or –2.054	M1		
		Use –2.326 and Reject H_0 and accept a change in the median time	A1√		ft z
		OR [β}			
		Compare Φ(z) correctly with 0.01 or 0.02	M1		
		Use 0.1 and conclusion as above	A1√	**8**	**ft z**
	(ii)	T has a symmetric distribution	B1	**1**	
	(iv)	e.g Road conditions during first 80 days unlikely to be typical	B1		Reasonable comment not given in question.
		Distribution of times unlikely to be symmetrical	B1	**2**	Accept normality of P as an assumption in (i)-CLT etc.
7	(i)	$(a+b)^3$=1 or $a+b$=1	B1		Seen
		$G'(t)=3t^2(a+bt)^3 + 3bt^3(a+bt)^2$	B1		aef
		Use G′(1)=4	M1		
		Method of solving equations	M1		
		b=⅓ ; a=⅔	A1A1	**6**	
	(ii)	$G''(t)$ correct, any form	B1		
		Use $G''(1)+G'(1)-\{G'(1)\}^2$	M1		
		= ⅔	A1√	**3**	ft G″(t). Or from marginal distn
	(iii)	$G_S(t)= [t^3(⅔+⅓t)^3]^n$	M1		
		Attempt to find coefficient of t^{3n+3}	M1		
		${}^{3n}C_3(⅔)^{3n-3}(⅓)^3$	A1√		ft slightly incorrect $G_S(t)$
		Given answer obtained correctly	A1	**4**	