OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS 4735

Probability & Statistics 4

Wednesday **21 JUNE 2006** Afternoon 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
List of Formulae (MF1)

TIME 1 hour 30 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 graphical 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 72.
- 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 **(i)** State whether the following are true or false for the random variables X and Y.

(a) X and Y are independent $\Longrightarrow \text{Cov}(X, Y) = 0$.

(b) $\text{Cov}(X, Y) = 0 \Longrightarrow X$ and Y are independent.

(c) $\text{Cov}(X, Y) \neq 0 \Longrightarrow X$ and Y are not independent.

[2]

(ii) Given that $\text{Var}(X) = 2$, $\text{Var}(Y) = 3$ and $\text{Var}(2X - Y) = 6$, find $\text{Cov}(X, Y)$. [4]

2 Out of a sample of 60 pairs of twin boys it was found that the first-born in 37 of the pairs was taller than the second-born. In the remaining 23 pairs the second-born was taller.

Stating clearly a necessary assumption, carry out a test at the 5% significance level of whether, in a majority of pairs of twin boys, the first-born is taller than the second-born. [9]

3 5% of valves manufactured in a certain factory are faulty. Each of the valves is tested by a machine which classifies 98% of faulty valves as faulty. It classifies 96% of non-faulty valves as non-faulty. F denotes the event 'a valve is faulty' and C denotes the event 'a valve is classified as faulty'.

(i) Show that $\text{P}(C) = 0.087$. [3]

(ii) Find $\text{P}(F \mid C)$. [3]

Each month 5000 valves are sold, all of which the machine classified as non-faulty.

(iii) Find the expected number of faulty valves sold each month. [4]

4 The continuous random variable X has a uniform distribution with probability density function given by

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b, \\ 0 & \text{otherwise,} \end{cases}$$

where a and b are constants. This distribution is denoted by $\text{U}(a, b)$.

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

For the independent random variables X_1 and X_2, $X_1 \sim \text{U}(-1, 0)$ and $X_2 \sim \text{U}(0, 1)$.

(ii) Find the moment generating function of T, where $T = X_1 + X_2$. [2]

S denotes the sum of two independent observations of Y, where $Y \sim \text{U}(-\frac{1}{2}, \frac{1}{2})$.

(iii) Show that S has the same moment generating function as T, and state what this indicates about the distributions of S and T. [4]

5 A random sample of 13 observations of a continuous random variable is taken, and the values ranked from 1 to 13. Four of these rankings are selected at random. The order in which the rankings are selected is irrelevant.

(i) Calculate how many possible different selections there are of the 4 rankings. [2]

The sum of the 4 rankings is denoted by R.

(ii) List all the selections of 4 rankings for which $R \leqslant 13$, and hence obtain the exact value of $\mathrm{P}(R \leqslant 13)$. [3]

The distributions of the continuous random variables X and Y have the same shape. A random sample of 4 observations of X and a random sample of 9 observations of Y are taken and the 13 observations are ranked. The sum of the ranks of the 4 observations of X is 13.

(iii) Naming the test used, and stating the null and alternative hypotheses, show that the samples give evidence of a difference in the medians of X and Y at a significance level smaller than 2%. [5]

6 The discrete random variable Y has probability generating function given by

$$\mathrm{G}(t) = \frac{0.8t}{1 - 0.2t}.$$

(i) Show that $\mathrm{E}(Y) = \frac{5}{4}$. [3]

(ii) Express $\mathrm{P}(Y = r)$ in terms of r, giving the possible values of r. [4]

(iii) By identifying the probability distribution of Y, or otherwise, find $\mathrm{Var}(Y)$. [3]

(iv) Find $\mathrm{P}(T \geqslant 8)$, where T is the sum of 6 independent observations of Y. [3]

7 During the Great War, *Invictor* tanks were used, but records of how many were manufactured have been lost. It may be assumed that the tanks had integer serial numbers ranging consecutively from 1 to n, where n is unknown. Suppose that a randomly selected tank has serial number denoted by X.

(i) Write down $\mathrm{E}(X)$ and show that $\mathrm{Var}(X) = \frac{1}{12}(n^2 - 1)$. [4]

The remains of two *Invictor* tanks have serial numbers denoted by X_1 and X_2.

(ii) Show that $N_1 = X_1 + X_2 - 1$ is an unbiased estimator of n. [2]

The larger of X_1 and X_2 is denoted by M.

(iii) Show that $\mathrm{P}(M = r) = \dfrac{2(r-1)}{n(n-1)}$, for $r = 2, 3, \ldots, n$. [3]

(iv) Using the result $\sum_{r=2}^{n} r(r-1) = \frac{1}{3}(n^3 - n)$, find $\mathrm{E}(M)$ and hence construct another unbiased estimator, N_2, of n. [3]

(v) Given that the variance of N_1 is $\frac{1}{6}(n^2 - n - 2)$ and that N_1 is a more efficient estimator than N_2, obtain an inequality for $\mathrm{Var}(M)$. [3]

Statistics 4 4735 June 2006
MARK SCHEME

1	**(i)**	(a) True (b) False **(c)** True	B2	**2**	B0 for 0,1 correct, B1 for 2 correct, B2 for 3 correct.
	(ii)	Var(2X-Y)=	M1		Using formula
		4Var(X)+Var(Y)-4Cov(X,Y)	A1		
		6=11-4Cov(X,Y)	M1		Obtain cov
		Cov(X,Y)=5/4	A1	**4**	cao
2		EITHER: sample is random OR twin pairs chosen independently	B1		
		H_0:m_F=m_S , H_1:$m_F > m_S$	B1		For both using medians
		Use of B(60,0.5)	M1		
		Normal approx with μ=30, σ^2=15	A1		Both
		EITHER: z=(36.5-30)/$\sqrt{15}$	M1		Standardising
		=1.678	A2		A1 if correct apart from missing or wrong cc
		OR:CR is (X-30-0.5)/$\sqrt{15}$ >1.645	M1		Setting up inequality
		X≥37	A2		A1 if correct apart from missing or wrong c.c.
		EITHER: 1.678> 1.645 OR: Sample value 37 in CR	M1		Correct comparison
		There is evidence that the first-born male twins are taller than the second -born twin in a majority of cases.	A1		Conclusion in context
		OR: p-value: 0.0467 > 1.645	M1		
		Completion	A1	**9**	
		NB: Exact Bin (60,0.5) p-value is 0.04623 from graphical calculator: full credit			
3	(i)	P(C)=P(C│F)P(F)+P(C│F′)P(F′)	M1		Use of formula
		=0.98×0.05 + 0.04×0.95	A1		
		0.087 **AG**	A1	**3**	
	(ii)	P(F│C)= $\dfrac{0.05\times0.98}{0.05\times0.98+0.95\times0.04}$	M1A1		
		=0.5632	A1	**3**	art 0.563 or 49/87
	(iii)	P(F│C′)=P(C′│F)P(F) / P(C′)	M1		Conditional prob.
		0.02×0.05/0.913 [0.001095]	A1		
		5000×above = 5.476., 5.48.	M1A1√	**4**	ft a conditional prob.

4	**(i)**	$M_X(t)=\int_a^b \frac{1}{b-a}e^{xt}dt$	M1		Correct integral with limits
		$= \left[\frac{e^{xt}}{(b-a)t}\right]_a^b$	B1		Correct integral
		$=\frac{e^{bt}-e^{at}}{(b-a)t}$ AG	A1	**3**	
	(ii)	Product of mgfs	M1		
		$\left(\frac{1-e^{-t}}{t}\right)\left(\frac{e^{t}-1}{t}\right)$	A1	**2**	
	(iii)	$M_S(t)=\left(\frac{e^{\frac{1}{2}t}-e^{-\frac{1}{2}t}}{t}\right)^2$	M1		Square of $M_Y(t)$
		$= (e^t-2+e^{-t})/t^2$	A1dep		Correctly shown
		mgfs of S and T are same	depA1		Correctly shown
		S and T have identical distributions	B1	**4**	
5	(i)	$^{13}C_4$	M1		Use of formula
		715	A1	**2**	
	(ii)	1234,1235,1236,1237,1245,1246, 1345	B2		B1 for 5 or 6
		7/715	B1√	**3**	ft (i)
	(iii)	Wilcoxon Rank Sum Test	B1		
		H_0:m_X=m_Y, H_1:m_X≠m_Y	B1		Both, involving medians
		Use P(R ≤13)	M1		
		2×7/715×100 =1.958% < 2%	M1		Comparing correctly
		Reject H_0, evidence of difference in medians at a significance level of (smaller than) 2%	A1	**5**	
		SR: If tables used, B1B1 M1 for CV with correct comparison for rejection M1 for rejection at 2% (not <) Max 4/5			

Q	Part	Answer	Mark	Total	Notes
6	(i)	$G'(t)=[0.8(1-0.2t)+0.16t]/(1-0.2t)^2$	M1		Quotient or product rule
			A1		
		$G'(t)=0.8/0.8^2=5/4$ **AG**	A1	**3**	
	(ii)	$G(t)=0.8t\,(1-0.2t)^{-1}$	M1		Use binomial expansion
		$=0.8t(1+0.2t+0.04t^2+\ \ )$	A1		At least 2 correct terms
		$P(Y=r)=0.8(0.2)^{r-1}$	A1		
		$r=1,2,3,\ldots.$	A1	**4**	OR from G(0.8)
	(iii)	EITHER: $Y\sim G(0.8)$	B1		Parameter not required
		$\mathrm{Var}(Y)=(1-0.8)/0.8^2$	M1		
		$=0.3125$	A1		
		OR: $G''(t)=0.32/(1-0.2t)^3$	B1		
		Use $G''(1)+G'(1)-(G'(1))^2$	M1		
		0.3125	A1	**3**	
	(iv)	$G_T(t)=0.8^6t^6(1-0.2t)^{-6}$	B1		$(G_Y(t))^6$
		$P(T\geq 8)=1-0.8^6(1+6\times 0.2)$	M1		Two terms in bracket
		$=0.42328$	A1	**3**	art 0.423
7	(i)	$E(X)=\frac{1}{2}(n+1)$	B1		
		$\mathrm{Var}(X)={}^1/_n\sum r^2-\frac{1}{4}(n+1)^2$	M1		Use of variance formula
		$={}^1/_6(n+1)(2n+1)-\frac{1}{4}(n+1)^2$	A1		
		$={}^1/_{12}(n^2-1)$ AG	A1	**4**	Correctly obtained
	(ii)	$E(N_1)=E(X_1)+E(X_2)-1$	M1		
		$=\frac{1}{2}(n+1)+\frac{1}{2}(n+1)-1$			
		$=n$,**(** so N_1 is an unbiased estimator of n**)**	A1	**2**	
	(iii)	$P(M=r)=$			
		EITHER: $P(X_1<r,X_2=r)+P(X_1=r,X_2<r)$	M1		
		$=((r-1)/n)(1/(n-1))+(1/n)(r-1)/(n-1)$	A1		
		$=2(r-1)/[n(n-1)]$ AG, $r=2,3,4,\ldots$	A1		
		OR: Choose 1 from r-1 and 1 from 1	M1		
		${}^{r-1}C_1\times{}^1C_1/{}^nC_2$	A1		
		$=(r-1)/[\frac{1}{2}n(n-1)]$=AG	A1	**3**	
	(iv)	$E(M)=\dfrac{2}{n(n-1)}\sum_{r=2}^{n}r(r-1)$	M1		
		$=\frac{2}{3}(n+1)$	A1		
		$N_2={}^3/_2M-1$	A1√	**3**	ft E(M)
	(v)	$\mathrm{Var}(N_1)<\mathrm{Var}(N_2)$ or equivalent	M1		Stated or implied
		${}^1/_6(n^2-n-2)<{}^9/_4\mathrm{Var}(M)$	A1		
		$\mathrm{Var}(M)>{}^2/_{27}(n^2-n-2)$	A1√	**3**	ft N_2