

Paper 3B/4B: Further Statistics 1 Mark Schemes

Question	Scheme	Marks	AOs	
1	$H_0 : \lambda = 5 (\lambda = 2.5) \quad H_1 : \lambda > 5 (\lambda > 2.5)$	B1	2.5	
	$X \sim \text{Po} (2.5)$	B1	3.3	
	Method 1:	Method 2:		
	$P(X \geq 7) = 1 - P(X \leq 6)$ $= 1 - 0.9858$	$P(X \geq 5) = 0.1088$ $P(X \geq 6) = 0.042$	M1	1.1b
	$= 0.0142$	$\text{CR } X \geq 6$	A1	1.1b
	$0.0142 < 0.05 \quad 7 \geq 6$ or 7 is in critical region or 7 is significant Reject H_0 . There is evidence at the 5% significance level that the level of pollution has increased. or There is evidence to support the scientists claim is justified	A1cso	2.2b	
(5 marks)				
Notes:				
B1: Both hypotheses correct using λ or μ and 5 or 2.5				
B1: Realising that the model Po(2.5) is to be used. This may be stated or used				
M1: Using or writing $1 - P(X \leq 6)$ or $1 - P(X < 7)$ a correct CR or $P(X \geq 5) = \text{awrt } 0.109$ and $P(X \geq 6) = \text{awrt } 0.042$				
A1: awrt 0.0142 or CR $X \geq 6$ or $X > 5$				
M1: A fully correct solution and drawing a correct inference in context				

Question	Scheme	Marks	AOs
2(a)	$P(X \geq 1) = 1 - P(X = 0)$ $1 - P(X = 0) = 0.049$	B1	3.1b
	$P(X = 0) = 0.951$	B1	1.1b
	$x^5 = 0.951$ $x = 0.99$	M1	3.1b
	$p = 0.01$	A1	1.1b
	$X \sim B(1000, 0.01)$	M1	3.3
	Mean = $np = 10$	A1ft	1.1b
	Variance = $np(1 - p) = 9.9$	A1ft	1.1b
		(7)	
(b)	$X \sim \text{Po}(\text{"10"})$ then require: $P(X > 6) = 1 - P(X \leq 6)$	M1	3.4
	$= 1 - 0.1301$		
	$= 0.870$	A1	1.1b
		(2)	
(c)	The approximation is valid as : the number of calls is large	B1	2.4
	The probability of connecting to the wrong agent is small	B1	2.4
		(2)	
(d)	The answer is accurate to 2 decimal place	B1	3.2b
		(1)	
(12 marks)			
Notes:			
(a) B1: Realising that the $P(\text{at least 1 call}) = 1 - P(X = 0)$ B1: Calculating $P(X = 0) = 0.951$ M1: Forming the equation $x^5 = \text{"their 0.951"}$ may be implied by $p = 0.01$ A1: 0.01 only M1: Realising the need to use the model $B(1000, 0.01)$ This may be stated or used A1: Mean = 10 or ft their p but only if $0 < p < 1$ A1: Var = 9.9 or ft their p but only if $0 < p < 1$			
(b) M1: Using the model $\text{Po}(\text{"their 10"})$ (this may be written or used) and $1 - P(X \leq 6)$ A1: awrt 0.870 Award M1 A1 for awrt 0.870 with no incorrect working			
(c) B1: Explaining why approximation is valid - need the context of number and calls B1: Need the context connecting, wrong agent			
(d) B1: Evaluating the accuracy of their answer in (b). Allow 2 significant figures			

Question	Scheme	Marks	AOs
3(a)	Expected value for 2 = $150 \times P(X = 2)$	M1	3.4
	= 28.3015...	A1	1.1b
	Expected value for 4 or more = $150 - (53.8 + 56.6 + 28.3 + 8.9)$ = 2.4	A1ft	1.1b
	H ₀ : Bin(20, 0.05) is a suitable model H ₁ : Bin(20, 0.05) is not a suitable model	B1	2.5
	Combining last two groups	M1	2.1
	≥ 3		
	Observed frequency 19		
	Expected frequency 11.3		
	$\nu = 4 - 1 = 3$	B1	1.1b
	Critical value, $\chi^2(0.05) = 7.815$	B1	1.1a
	Test statistic = $\frac{(43 - 53.8)^2}{53.8} + \frac{(62 - 56.6)^2}{56.6} + \dots$	M1	1.1b
	= 8.117	A1	1.1b
	In critical region, sufficient evidence to reject H ₀ , accept H ₁ Significant evidence at 5% level to reject the manager's model	A1	3.5a
		(10)	
(b)	$\nu = 4 - 2 = 2$		
	4 classes due to pooling	B1	2.4
	2 restrictions (equal total and mean/proportion)	B1	2.4
		(2)	
(c)	H ₀ : Binomial distribution is a good model H ₁ : Binomial distribution is not a good model	B1	3.4
	Critical value, $\chi^2(0.05) = 5.991$ Test statistic is not in critical region, insufficient evidence to reject H ₀ There is evidence that the Binomial distribution is a good model	B1	3.5a
		(2)	
(14 marks)			

Question 3 notes:**(a)****M1:** Using the binomial model $150 \times p^2 \times (1-p)^{18}$ may be implied by 28.3**A1:** awrt 28.3**A1:** awrt 2.4 or ft their “28.3”**B1:** Both hypotheses correct using the correct notation or written out in full**M1:** For recognising the need to combine groups**B1:** Number of degrees of freedom = 3 may be implied by a correct CV**B1:** awrt 7.82**M1:** Attempting to find $\sum \frac{(O_i - E_i)^2}{E_i}$ or $\sum \frac{O_i^2}{E_i} - N$ may be implied by awrt 8.12**A1:** awrt 8.12**A1:** Evaluating the outcome of a model by drawing a correct inference in context**(b)****B1:** Explaining why there are 4 classes**B1:** Explanation of why 2 is subtracted**(c)****B1:** Correct hypotheses for the refined model**B1:** The CV awrt 5.99 and drawing the correct inference for the refined model

Question	Scheme	Marks	AOs
4	Po(2.3) $n = 100$ $\mu = 2.3$ $\sigma^2 = 2.3$		
	$CLT \Rightarrow \bar{X} \approx N\left(2.3, \frac{2.3}{100}\right)$	M1 A1	3.1a 1.1b
	$P(\bar{X} > 2.5) = P\left(Z > \frac{2.5 - 2.3}{\sqrt{0.023}}\right)$	M1	3.4
	$= P(Z > 1.318..)$		
	$= 0.09632...$	A1	1.1b
		(4)	
(4 marks)			
Notes:			
M1: For realising the need to use the CLT to set $\bar{X} \approx$ normal with correct mean May be implied by using the correct normal distribution A1: For fully correct normal stated or used M1: Use of the normal model to find $P(\bar{X} > 2.5)$. Can be awarded for $\frac{2.5 - 2.3}{\sqrt{0.023}}$ or awrt 1.32 A1: awrt 0.0963			

Question	Scheme	Marks	AOs
5(a)	$\binom{7}{1} \times 0.15^2 \times (0.85)^6$	M1	3.3
	$= 0.05940\dots = \text{awrt } \underline{\underline{0.0594}}$	A1	1.1b
		(2)	
(b)	The model is only valid if:		
	the games (trials) are independent	B1	3.5b
	the probability of winning a prize, 0.15, is constant for each game	B1	3.5b
		(2)	
(c)	$18 = \frac{r}{p} \quad \text{and} \quad 6^2 = \frac{r(1-p)}{p^2}$	M1 A1	3.1b 1.1b
	Solving: $2p = 1 - p$	M1	1.1b
	$p = \frac{1}{3} (> 0.15)$ so Mary has the greater chance of winning a prize	A1	3.2a
		(4)	
	(8 marks)		
Notes:			
5(a)			
M1: For selecting an appropriate model negative binomial or B(7, 0.15) with an extra success in 8 th trial e.g.			
$\binom{7}{1} 0.15 \times (0.85)^6 \times 0.15$ Allow $\binom{7}{1} 0.85 \times (0.15)^6 \times 0.85$ may be implied by awrt 0.0594			
A1: awrt 0.0594			
(b)			
B1: Stating the first assumption that games are independent			
B1: Stating the second assumption that the probability remains constant			
(c)			
M1: Forming an equation for the mean or for the standard deviation			
A1: Both equations correct			
M1: Solving the 2 equations leading to $2p = 1 - p$			
A1: For $p = \frac{1}{3}$ followed by a correct deduction			

Question	Scheme	Marks	AOs
6(a)	$G_X(1) = 1$ gives	M1	2.1
	$k \times 6^2 = 1$ so $k = \frac{1}{36}$ *	A1*cs0	1.1b
		(2)	
(b)	$P(X=3) = \text{coefficient of } t^3$ so $G_X(t) = k(\dots + 4t^3 \dots)$	M1	1.1b
	$[P(X=3)] = \frac{1}{9}$	A1	1.1b
		(2)	
(c)	$G'_X(t) = 2k(3+t+2t^2) \times (1+4t)$	M1	2.1
	$E(X) = G'_X(1) = 2k(3+1+2) \times (1+4)$	M1	1.1b
	$= \frac{5}{3}$	A1	1.1b
	$G''_X(t) = 2k[(3+t+2t^2) \times 4 + (1+4t)^2]$	M1 A1	2.1 1.1b
	$G''_X(1) = 2k[6 \times 4 + 5^2] \quad \left\{ = \frac{49}{18} \right\}$	M1	1.1b
	$\text{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2 = \frac{49}{18} + \frac{5}{3} - \frac{25}{9}$	M1	2.1
	$= \frac{29}{18}$ *	A1*cs0	1.1b
		(8)	
(d)	$G_{2X+1}(t) = \frac{t}{36}(3+t^2+2(t^2)^2)^2$ [$\times t$ or sub t^2 for t]	M1	3.1a
	$= G_{2X+1}(t) = \frac{t}{36}(3+t^2+2t^4)^2$	A1	1.1b
		(2)	
(14 marks)			
Notes:			
(a)			
M1: Stating $G_X(1) = 1$			
A1*: Fully correct proof with no errors cs0			
(b)			
M1: Attempting to find the coefficient of t^3 . May be implied by obtaining $\frac{1}{9}$ or awrt 0.11			
A1: $\frac{1}{9}$, allow awrt 0.111			

Question 6 notes continued:**(c)****M1:** Attempting to find $G'_x(t)$. Allow Chain rule or multiplying out the brackets and differentiating**M1:** Substituting $t = 1$ into $G'_x(t)$ **A1:** $\frac{5}{3}$, allow awrt 1.67**M1:** Attempting to find $G''_x(t)$ **A1:** $2k \left[(3+t+2t^2) \times 4 + (1+4t)^2 \right]$ or $k(48t^2 + 24t + 26)$ o.e.**A1:** $2k[6 \times 4 + 5^2]$ o.e.**M1:** Using $G''_x(1) + G'_x(1) - [G'_x(1)]^2$ to find the Variance**A1*:** $\frac{29}{18}$ cso**(d)****M1:** Realising the need to $\times t$ or sub t^2 for t **A1:** $\frac{t}{36} (3 + t^2 + 2t^4)^2$, or $\frac{t}{36} (9 + 6t^2 + 13t^4 + 4t^6 + 4t^8)$ o.e.

Question	Scheme	Marks	AOs
7(a)	$X \sim B(20, 0.2)$ and seek c such that $P(X \leq c) < 0.10$	M1	3.3
	$[P(X \leq 1) = 0.0692]$ CR is $X \leq 1$	A1	1.1b
		(2)	
(b)	Size = <u>0.0692</u>	B1ft	1.2
		(1)	
(c)	$Y = \text{no. of spins until red obtained so } Y \sim \text{Geo}(0.2)$	M1	3.3
	$\mu = \frac{1}{p}$ so if $p < 0.2$ then mean is <u>larger</u> so seek d so that $P(Y \geq d) < 0.10$	M1	2.4
	$P(Y \geq d) = (0.8)^{d-1}$	M1	3.4
	$(0.8)^{d-1} < 0.10 \Rightarrow d-1 > \frac{\log(0.1)}{\log(0.8)}$	M1	1.1b
	$d > 11.3..$	A1	1.1b
	CR is $Y \geq 12$	A1	2.2b
		(6)	
(d)	Size = $[0.8^{11} = 0.085899...] = \mathbf{0.0859}$	B1	1.1b
		(1)	
(e)(i)	Power = $P(\text{reject } H_0 \text{ when it is false}) = P(X \leq 1 \mid X \sim B(20, p))$	M1	2.1
	$= (1-p)^{20} + 20(1-p)^{19} p$	M1	1.1b
	$= (1-p)^{19} (1+19p) *$	A1*cs0	1.1b
(ii)	Power = $(1-p)^{11}$	B1	1.1b
		(4)	
(f)	Sam's test has smaller $P(\text{Type I error})$ (or size) so is better	B1	2.2a
	Power of Sam's test = 0.1755...	B1	1.1b
	Power of Tessa's test = $0.85^{11} = 0.1673...$	B1	1.1b
	So for $p = 0.15$ Sam's test is recommended	B1	2.2b
		(4)	
(18 marks)			

Question 7 notes:	
(a)	
M1:	Realising the need to use the model Using $B(20,0.2)$ with method for finding the CR or implied by a correct CR
A1:	$X \leq 1$ or $X < 2$
(b)	
B1:	awrt 0.0692
(c)	
M1:	Realising that the model $\text{Geo}(0.2)$ is needed. This may be written or used
M1:	Realising the key step that they need to find $P(Y \geq d) < 0.10$
M1:	Using the model $(0.8)^{d-1}$
M1:	Using the model $(0.8)^{d-1} < 0.10$ and finding a method to solve leading to a value/range of values for d
A1:	For $d > 11.3..$
A1:	For $Y \geq 12$ or $Y > 11$ (a correct inference)
(d)	
B1ft:	awrt 0.0692. ft their answer to part (c)
(e)(i)	
M1:	Using $B(20, p)$ and realizing they need to find $P(X \leq 1)$ o.e. This may be used or written
M1:	Using $P(X=0) + P(X=1)$
A1*:	Fully correct proof (no errors) cso
(ii)	
B1:	For $(1-p)^{11}$
(f)	
B1:	Making a deduction about the tests using the answers to part(b) and (d)
B1:	awrt 0.0176
B1:	awrt 0.167
B1:	A correct inference about which test is recommended