2.2 Further Statistics: Poisson and Geometric Distributions: Test Pack

AQA 7367 · Calculator allowed · about 70 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 Typographical errors on a page of a Revision Library worksheet occur singly and at random, independently of each other, at a constant average rate of 1.8 errors per page.

(a) State two conditions, in the context of this scenario, that must hold for the number of errors per page to be well modelled by a Poisson distribution. **(2)**

(b) Let X be the number of errors on a randomly chosen page. State the distribution of X, including its parameter. **(1)**

(c) Find P(X = 2). **(2)**

(d) Find the probability that a randomly chosen page contains at least one error. **(2)**

(e) Two pages are chosen independently at random. Using a suitable distribution for the total number of errors on the two pages, find the probability that there is at most one error in total. **(3)**

(Total for Question 1 is 10 marks)

---

2 For each of parts (a) to (c), state which of the Binomial, Poisson or Geometric distribution would be most appropriate to model the random variable described, giving a reason. No calculations are required for parts (a) to (c).

(a) The number of trains, out of the next 20 trains departing a particular platform at a UK railway station, that are more than 5 minutes late, given that each train is independently more than 5 minutes late with the same fixed probability. **(2)**

(b) The number of power cuts recorded at a substation in Cardiff during a randomly chosen week, given that power cuts happen singly, independently and at a constant average rate. **(2)**

(c) The number of times a fair six-sided die is rolled, up to and including the first time a six is obtained. **(2)**

(d) A call centre receives calls singly and at random, independently of each other, at a constant average rate of 3.4 calls per 5-minute period. Find the probability that the call centre receives more than 6 calls in a 5-minute period. **(3)**

(Total for Question 2 is 9 marks)

© Revision Library | revisionlibrary.co.uk

3 An orchard in Kent has 3000 apple trees. Each tree is infected with a particular leaf disease independently of the others, with probability 0.0025. Let X be the number of infected trees.

(a) State the exact distribution of X, and explain why a Poisson approximation is appropriate here, giving the parameter of the approximating distribution. **(3)**

(b) Using the Poisson approximation, find P(X = 5). **(2)**

(c) Using the Poisson approximation, find P(X > 10). **(3)**

(d) The exact binomial probability P(X = 5) is 0.1094, to 4 significant figures. Comment on the suitability of the Poisson approximation used in part (b). **(2)**

(Total for Question 3 is 10 marks)

---

4 The random variable X has distribution B(50, 0.02).

(a) State two conditions under which a Poisson distribution is a suitable approximation to a binomial distribution, and give the parameter of the approximating distribution. **(2)**

(b) Use the binomial distribution to find P(X = 1), and use the Poisson approximation to find the same probability, each to 4 significant figures. **(4)**

(c) Find the percentage error in using the Poisson approximation here, to 2 significant figures. **(2)**

(Total for Question 4 is 8 marks)

© Revision Library | revisionlibrary.co.uk

5 A player attempts a task repeatedly until the first success. Each attempt succeeds with probability 0.5, independently of the others. Let X be the number of the attempt on which the first success occurs, so X ~ Geo(0.5).

(a) Find P(X = 7) and P(X > 7), each to 3 significant figures. **(3)**

(b) Find E(X) and Var(X). **(4)**

(c) Find the least number of attempts n such that the probability of at least one success within n attempts exceeds 0.95. **(4)**

(Total for Question 5 is 11 marks)

---

6 Let X ~ Geo(p) model the number of independent trials, each with constant probability of success p, up to and including the first success. It is a standard result that $P(X > n) = (1-p)^n$ for a positive integer n.

(a) For positive integers s and t, show that P(X > s + t | X > s) = P(X > t). State clearly what this result demonstrates about the geometric distribution. **(6)**

(Total for Question 6 is 6 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.2 Further Statistics: Poisson and Geometric Distributions: Test Pack

Question 1

(a) any one valid condition in context, e.g. errors occur independently of each other **B1**

(a) a second valid condition in context, e.g. errors occur singly (not two at the same point) and/or at a constant average rate over a page of fixed length, oe **B1**

(a) Answer: Any two of: errors occur independently of each other; errors occur singly, not simultaneously; errors occur at a constant average rate; pages are of comparable, fixed length.

(b) X ~ Po(1.8), cao **B1**

(b) Answer: X ~ Po(1.8)

(c) correct use of $P(X=2) = e^{-1.8} * 1.8^2 / 2!$ **M1**

(c) awrt 0.268 **A1**

(c) Answer: P(X = 2) = awrt 0.268

(d) uses $P(X \geq 1) = 1 - P(X=0)$ **M1**

(d) awrt 0.835 **A1**

(d) Answer: P(X ≥ 1) = awrt 0.835

(e) identifies total number of errors on two pages ~ Po(3.6), using the additive property (oe) **M1**

(e) attempts $P(X=0) + P(X=1)$ with $\lambda = 3.6$ **M1**

(e) awrt 0.126 **A1**

(e) Answer: P(at most 1 error in total) = awrt 0.126

Question 2

(a) Binomial **B1**

(a) reason: a fixed number of independent trials (n = 20), each with the same probability of the event 'more than 5 minutes late' (oe) **B1**

(a) Answer: Binomial, since there is a fixed number of independent trials (20 trains), each with the same probability of being more than 5 minutes late.

(b) Poisson **B1**

(b) reason: events occur singly, at random, independently and at a constant average rate over a fixed interval, with no natural upper bound on the count (oe) **B1**

(b) Answer: Poisson, since power cuts occur singly, independently and at a constant average rate over the fixed interval of one week.

(c) Geometric **B1**

(c) reason: independent repeated trials with a constant probability of success (rolling a six), continued until the first success (oe) **B1**

(c) Answer: Geometric, since the die is rolled repeatedly and independently, with constant probability 1/6 of success, until the first six is obtained.

(d) identifies X ~ Po(3.4) and uses $P(X>6) = 1 - P(X \leq 6)$ **M1**

(d) attempts to sum $P(X=0)$ to $P(X=6)$ with $\lambda = 3.4$ **M1**

(d) awrt 0.0579 **A1**

(d) Answer: P(X > 6) = awrt 0.0579

Question 3

(a) X ~ B(3000, 0.0025) **B1**

(a) n is very large and p is very small (so np is of moderate size), justifying the Poisson approximation (oe) **B1**

(a) parameter $\lambda = np = 3000 * 0.0025 = 7.5$ **B1**

(a) Answer: X ~ B(3000, 0.0025) exactly; approximated by Po(7.5) since n is very large and p is very small.

(b) correct use of $P(X=5) = e^{-7.5} * 7.5^5 / 5!$ with $\lambda = 7.5$ **M1**

(b) awrt 0.109 **A1**

(b) Answer: P(X = 5) = awrt 0.109 (Poisson approximation)

(c) uses $P(X>10) = 1 - P(X \leq 10)$ **M1**

(c) sums Poisson probabilities from 0 to 10 with $\lambda = 7.5$ **M1**

(c) awrt 0.138 **A1**

(c) Answer: P(X > 10) = awrt 0.138

(d) the Poisson approximation (0.1094 to 4 s.f.) agrees with the exact binomial value to 4 significant figures (oe) **B1**

(d) concludes the approximation is very accurate here, because n = 3000 is very large and p = 0.0025 is very small, giving a moderate np **B1**

(d) Answer: The Poisson approximation agrees with the exact binomial value to 4 significant figures (both 0.1094), so it is very accurate here since n is very large and p is very small.

Question 4

(a) n is large and p is small **B1**

(a) X is approximately Po(50 x 0.02) = Po(1) **B1**

(a) Answer: Po(1)

(b) $50C1 \times 0.02 \times 0.98^{49}$ **M1**

(b) 0.3716 **A1**

(b) e^{-1} 1! **M1**

(b) 0.3679 **A1**

(b) Answer: binomial 0.3716, Poisson 0.3679

(c) |approximate - exact|/exact x 100 **M1**

(c) 1.0% **A1**

(c) Answer: 1.0%

Question 5

(a) $P(X = 7) = (1 - 0.5)^6 \times 0.5$ **M1**

(a) $P(X > 7) = (1 - 0.5)^7$ **M1**

(a) 0.00781 **A1**

(a) Answer: 0.00781; 0.00781

(b) E(X) = 1/p **M1**

(b) 2 **A1**

(b) $Var(X) = (1 - p)/p^2$ **M1**

(b) 2.000 **A1**

(b) Answer: E(X) = 2.000, Var(X) = 2.000

(c) $P(X \leq n) = 1 - (1 - 0.5)^n > 0.95$ **M1**

(c) $(1 - 0.5)^n < 0.05$, and takes logs (changing the inequality) **M1**

(c) n > 4.322 **A1**

(c) n = 5 **A1**

(c) Answer: n = 5

© Revision Library | revisionlibrary.co.uk

Question 6

(a) quotes $P(X > n) = (1-p)^n$ for a positive integer n **M1**

(a) recognises that the event $\{X > s+t\}$ is a subset of $\{X > s\}$ (since $s+t > s$), so $\{X > s+t\}$ intersect $\{X > s\} = \{X > s+t\}$ **B1**

(a) applies the conditional probability formula $P(X > s+t \mid X > s) = P(X > s+t \text{ and } X > s) / P(X > s)$ **M1**

(a) substitutes to get $(1-p)^{s+t} / (1-p)^s$ **A1**

(a) simplifies to $(1-p)^t = P(X > t)$, cso **A1**

(a) states that this demonstrates the memoryless property: given that the first s trials have failed, the distribution of the number of further trials needed is the same as it was at the start **B1**

(a) Answer: $P(X > s+t \mid X > s) = (1-p)^t = P(X > t)$ (shown); this is the memoryless property of the geometric distribution.

© Revision Library | revisionlibrary.co.uk