2. Statistical Distributions

AS – Discrete Distributions

The Binomial Distribution

For a fixed number of trials, *n*, each with a probability *p* of occurring, the probability of a number *x* of successes is given by the formula

$$P(X = x) = C_x^n p^x (1 - p)^{n-x}$$

Binomial distribution tables (and calculators) give you cumulative probability $P(X \leq x)$

Eg1 The random variable $X \sim B(10, 0.35)$, find:
(a) $P(X \leq 6)$
(b) $P(X \geq 5)$
(c) $P(X = 6)$
(d) $P(4 \leq X \leq 7)$

(a) $P(X \leq 6)$ using a calc $= 0.974$

(b) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.7515 = 0.2485$

(c) $P(X=6) = {}^{10}C_6 (0.35^6)(0.65^4) = 0.0689$

(d) $P(4 \leq X \leq 7) = P(X \leq 7) - P(X \leq 3) = 0.99518 - 0.51383$
$= 0.4814$

The binomial distribution can be appropriately applied under the following conditions:
- the trials are independent
- the trials have a constant probability of success
- there are a fixed number of trials
- there is only success or failure.

The Poisson Distribution

The Poisson distribution is a discrete probability distribution which is used to model the number of events occurring randomly within a given interval of time and space.

In a particular interval, the probability of an event *X* occurring *x* number of times is given by:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

where $\lambda = \mu = E(X)$ and x = 0, 1, 2, 3, ...

If the probabilities are distributed in this way, it is written $X \sim Po(\lambda)$

As with the binomial distribution, tables give you the cumulative probability $P(X \leq x)$

Eg2 The number of telephone calls received at an exchange during a weekday morning follows a Poisson distribution with a mean of 6 calls per 5 minute period. Find the probability that
(a) there are no calls received in the next five minutes
(b) 3 calls are received in the next five minutes
(c) fewer than 2 calls are received between 11:00 and 11:05
(d) more than 2 calls are received between 11:30 and 11:35

$X \sim Po(6)$

(a) $P(X=0) = e^{-6}\left(\frac{6^0}{0!}\right) = 0.0025$

(b) $P(X=3) = e^{-6}\left(\frac{6^3}{3!}\right) = 0.0892$

(c) $P(X<2) = P(X=0) + P(X=1) = e^{-6}\left[1 + \frac{6^1}{1!}\right] = 0.174$

(d) $P(X>2) = 1 - P(X \leq 2) = 1 - 0.0620 = 0.938$

Calc PoissonCD

The Poisson distribution can be appropriately used when
- *n* is large (usually > 50) **and**
- *p* is small (usually < 0.1)

The Poisson distribution can be used as an approximation to the binomial distribution

If X ~ B(n, p) with large n and small p, then X ~ Po(np)

Eg3 The probability that a wrapped chocolate biscuit is double wrapped is 0.01. Use a suitable approximation to find the probability that of the next 60 biscuits you unwrap:
(a) none are double wrapped
(b) at least 2 are double wrapped

let X represent the n° of biscuits that are double wrapped.

$n > 50,\ p < 0.1$

~~X~B(50×0.1) X~Po(~~

$X \sim Po(60 \times 0.01)$

$X \sim Po(0.6)$

(a) $P(X=0) = e^{-0.6}\left(\frac{0.6^0}{0!}\right) = 0.549$

(b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.878$

$= 0.122$

The Discrete Uniform Distribution

A discrete uniform distribution is a distribution where all the outcomes are equally likely, for example the outcome when a fair dice is thrown.

If X is a discrete variable and is uniformly distributed on the set {1, 2, 3, 4, ..., N} then

$$P(x) = \frac{1}{N}$$

Eg4 A fair octagonal spinner numbered from 1 to 8 is spun and the number obtained X is recorded. This process is repeated a set number of times.

Find $P(2 \leq X < 5)$ $= P(2) + P(3) + P(4)$

$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$

$= \frac{3}{8}.$

At both AS and A2 you may be given questions where the appropriate distribution to apply is not provided. In each of the following situations decide which are best modelled by binomial, Poisson or uniform distributions. Give a reason for your decision.

(a) A customer service department receives an average of 15 complaints per hour. What is the probability that there are at most 10 complaints in a certain period of one hour? Poisson $Po(15)$

(b) The demand for rental cars from a car hire firm has a mean of 8 per day. Find the probability that on a randomly chosen day, the demand is for 8 or more cars. Posision $Po(8)$

(c) A bag contains 6 red and 4 black balls. One ball is taken out, its colour noted and then returned to the bag. The process is repeated until 5 balls' colours have been recorded. What is the probability that all 5 balls are red? $B(5, 0\cdot6)$

(d) 5% of a batch of components are faulty. Find the probability that in a randomly selected batch of 20 components, 4 components are faulty. $B(20, 0\cdot05)$

(e) A test containing multiple choice questions with a choice of answers A, B, C or D consists of 20 questions. The question setter needs to be sure that the chance of passing the test simply by guessing the answers is less than 0.05. How many marks should the pass mark be set at for the probability of passing by guessing to be less than 0.05? $B(20, 0\cdot05)$

(f) Daisies are growing randomly in a meadow. There is an average of six daisy plants per square metre of meadow. Find the probability that a randomly picked area of $0.5m^2$ of the meadow contains no daisies. $Po(6)$

(g) A salesman in a car showroom sells on average 12 cars per week. What is the probability that he sells less than 5 cars in a randomly selected week? $Po(12)$

(h) A one-mile stretch of road contains an average of 3 potholes. Find the probability that in a half-mile stretch of the same road there will be no potholes. $Po(3)$

(i) It is known that a packet of seeds has a probability of 0.95 of all the seeds germinating. Find the probability that in a packet of 50 seeds at most 5 seeds will not germinate. $B(50, 0\cdot95)$

(j) In an accident and emergency department it is known that the average wait to be seen by a doctor is 2 hours. Find the probability of a patient turning up on a random day and waiting less than 1 hour. $Po(2)$

Exercise 2.1

1. It is known that 25% of the bulbs in a box produce yellow flowers. A customer buys 20 of these bulbs. Find the probability that:
   (a) exactly 4 bulbs produce yellow flowers
   (b) fewer than 8 bulbs produce yellow flowers.

2. The number of items of junk mail arriving by post each day at a house can be modelled by a Poisson distribution with mean 3.4 .
   (a) Without using tables, calculate:
      (i) $P(X = 4)$
      (ii) $P(X \le 2)$.
   (b) Using tables, determine $P(4 \le X \le 7)$.

3. Each time a darts player throws a dart at the bulls-eye they hit the bulls-eye with a probability 0.08. The darts player throws 100 darts at the bulls-eye. Use a Poisson approximation to find the probability that she hits the bulls-eye fewer than 5 times.

4. On a turtle farm, turtles are bred and hatched from eggs under controlled conditions.
   (a) The probability of producing a female turtle from an egg is 0.4 under the controlled conditions. The probability of producing a female from an egg is independent of other eggs hatching to produce female turtles. When 20 eggs are kept under the controlled conditions, find the probability that:
      (i) exactly 10 female turtles are produced
      (ii) more than 7 female turtles are produced. [5]
   (b) During the hatching process, the probability that an egg fails to hatch is 0.05. When 300 eggs are kept under the controlled conditions, use the Poisson approximation to find the probability that the number of eggs failing to hatch is less than 10. [3]

5. (a) A factory manufactures cups. The manager knows from past experience that 5% of the cups produced are defective. Given a random sample of 50 of these cups, determine the probability that the number of defective cups in this sample is:
      (i) exactly 2
      (ii) between 3 and 8 (both inclusive). [6]
   (b) The factory also manufactures plates. The manager knows that 1·5% of the plates produced are defective. A random sample of 250 plates is taken.
      (i) Explain why the Poisson distribution can be used as an approximation to the binomial distribution, to model the number of plates that are defective.
      (ii) Use an appropriate Poisson distribution to find, approximately, the probability that the sample of plates contains exactly 4 defective plates. [5]

6 (a) The random variable X has the binomial distribution $B(20, 0{\cdot}2)$.

(i) Without the use of tables, calculate $P(X = 6)$.

(ii) Determine $P(2 \leq X \leq 8)$. [5]

(b) The random variable Y has the binomial distribution $B(200, 0{\cdot}0123)$. Use the Poisson distribution to determine the approximate value of $P(Y = 3)$. [3]

7 (a) When a certain type of seed is planted, there is a probability of 0.7 that it produces red flowers. A gardener plants 20 of these seeds. Calculate the probability that:

(i) exactly 15 seeds produce red flowers

(ii) more than 12 seeds produce red flowers. [6]

(b) When a different type of seed is planted, there is a probability of 0.09 that it produces white flowers. The gardener plants 150 of these seeds. Use an appropriate Poisson distribution to determine, approximately, the probability that exactly 10 seeds produce white flowers. [3]

Numerical Answers

(1a) 0.1896 (b) 0.8982

(2ai) 0.1858 (ii) 0.3397 (b) 0.4185

(3) 0.0996

(4ai) 0.1171 (ii) 0.5841 (b) 0.0699

(5ai) 0.2611 (ii) 0.4587 (b) 0.194

(6ai) 0.109 (ii) 0.9208 (b) 0.212

(7ai) 0.179 (ii) 0.772 (b) 0.076

AS Distribution

① $N = 20 \quad p = 0.25$

(a) $P(X = 4) = P(X \le 4) - P(X \le 3)$

$= 0.4148 - 0.2252$

$= 0.1896$

Or $^{20}C_4 (0.25^4)(0.75)^{16}$

$= 0.1896$

(b) $P(X < 8) = P(X \le 7) = 0.8982$

② $X \sim Po(3.4)$

(a)(i) $P(X = 4) = e^{-3.4}\left(\frac{3.4^4}{4!}\right) = 0.1858$

(ii)(~~b~~) $P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$= e^{-3.4}\left(1 + \frac{3.4^1}{1!} + \frac{3.4^2}{2!}\right)$

$= 0.3397$

(b) $P(4 \le X \le 7) = P(X \le 7) - P(X \le 3)$

$= 0.9764 - 0.5584$

$= 0.4185$

③ $\lambda = np = 100 \times 0.08 = 8$

$X \sim Po(8)$

$P(X \le 5) = P(X \le 4) = 0.0996$

(4) (a) $p = 0.4 \quad n = 20$

(i) $P(X=10) = {}^{20}C_{10}\, 0.4^{10}\, 0.6^{10} = 0.1171$

(ii) $P(X>7) = 1 - P(X \leq 7) = 1 - 0.41589 = 0.5841$

(b) $\lambda = np = 300 \times 0.05 = 15$

$X \sim Po(15)$

$P(X<10) = P(X \leq 9) = 0.0699$

(5) (a) $n = 50 \quad p = 0.05$

(i) $P(X=2) = {}^{50}C_2\,(0.05^2)(0.95^{48}) = 0.2611$

(ii) $P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2)$

$= 0.9998 - 0.5405 = 0.4587$

(b) Use Poisson if N is large & p is small

(i) here ~~NP~~ $n = 250 \quad p = 0.015$, $\lambda = np = 3.75$

(ii) $X \sim Po(3.75)$

$P(X=4) = e^{-3.75}\left(\frac{3.75^4}{4!}\right) = 0.194$

(6) (a) $B(20, 0.2)$

(i) $P(X=6) = {}^{20}C_6\,(0.2^6)(0.8^{14}) = 0.109$

(ii) $P(2 \leq X \leq 8) = P(X \leq 8) - P(X \leq 1)$

$= 0.9900 - 0.0692 = 0.9208$

(iii) $B(200, 0.0123)$

$\lambda = np = 2.46$, $X \approx Po(2.46)$ $\quad P(Y=3) = \cancel{e^{-2.46}\left(\frac{3}{3!}\right)} = e^{-2.46}\left(\frac{2.46^3}{3!}\right)$

$= 0.212$

7. (a) $n = 20 \quad p = 0.7$

(i) $P(X = 15) = {}^{20}C_{15}(0.7^{15})(0.3^{5}) = 0.179$

(ii) $P(X > 12) = \cancel{P(} \; 1 - P(X \leq 12) = 1 - 0.2277 = 0.772$

(b) $\lambda = np = 150 \times 0.09 = 13.5$

$X \sim Po(13.5)$

$P(X = 10) = e^{-13.5}\left(\frac{13.5^{10}}{10!}\right) = 0.076$