1 Univariate discrete probability distributions

In this chapter we shall study some particular types of discrete distribution that arise in the physical world and derive their means and variances. We shall also determine their m.g.fs. Some of the major types of discrete distribution include Bernoulli, binomial, negative binomial, geometric ,hyper-geometric and Poisson.

1.1 The Bernoulli Distribution:

A Bernoulli trial is a random experiment which has only two mutually exclusive outcomes i.e. the occurrence (success) or non occurrence (failure) of an event. Example of such experiment include tossing a coin resulting in head or tails, sitting an exam result in passing or failing ,testing positive or negative of a disease, being dead or alive. Here the occurrence of one event excludes the occurrence of the other. The probability of success is denoted as p while the probability of failure is denoted as q so that $p + q = 1 \; or \; p = 1 - q \; or \; q = 1 - p$. Let X be a random variable defined such that:

$$X = \begin{cases} \; if \; the \; outcome \; of \; a \; trial \; is \; a \; success \\ 0 \; if \; the \; outcome \; of \; a \; trial \; is \; a \; failure \end{cases}$$

Definition:

A random variable X is defined to have Bernoulli distribution. If the density function of X is given by

$$f(x) = \{^{p^x(1-p)^{1-x}, x=0.1}_{0, \quad elsewhere}$$

The moments of X

$$E(x^k) = E(x) = p, \; k = 1, 2, 3, ---$$

The variance of x is given by

$$Var(x) = E(x^2) - [E(x)]^2$$

$$= p - p^2$$

$$= p(1 - p) = pq$$

The m.g.f of x is given by

$$M_X(t) = E(e^{tx}) = q + pe^t$$

As shown previously.

1.2 The Binomial Distribution

If a Bernoulli trial is repeated more than once, we have a binomial distribution. i.e if we toss a coin once, it is a Bernoulli trial, if we toss it twice, thrice, etc we have a Binomial distribution. If we test one individual for a disease where the outcome is either positive or negative we have a Bernoulli trial, when we test 2,3,..individuals we have a Binomial distribution. The number of successes in n trials is a random variable having a binomial distribution with parameters n and p

In other words, a random variable X is said to have a Binomial distribution if it counts the number of successes when n (fixed number) identical, independent Bernoulli trials are performed.

So, to identify a Binomial random variable 5 questions must be asked.

1. Are there a fixed number of trials (n)?
2. Is each trial a Bernoulli trial i.e. does each trial have 2 complementary outcomes?
3. Are the Bernoulli trials identical i.e. is the probability of success p the same for all the trials?
4. Are the trials independent i.e. does the outcome of one trial not affect the outcome of another trial?
5. Does X count the number of successes? If the answer is "yes" to all 5 questions then X is a Binomial random variable. Notation:

Definition :A random variable x is defined to have a binomial distribution if the discrete probability function of x is given by:

$$f(x; n, p) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, - - - n, 0 < p < 1$$

In short hand we write $X \sim Bin(n,\ p)$where X is a positive integer, n is the number of trials and p is the probability of success (depending on what you are interested in)

Example 1

A fair coin is tossed three times. Find the probability of throwing exactly 2 heads

Let X be number of heads. The probability of getting a head and a tail in a fair coin is equal so $p = q = 0.5$

$$X \sim Bin\,(3,\ 0.5)$$

$$p(X = 2) = \binom{3}{2}(0.5)^2(1 - 0.5)^{3-2} = 0.375$$

Figure 1: Example 2

A student answers 10 questions in a multiple-choice test by guessing each answer. For each question, there are 5 possible answers, only one of which is correct. If we consider a "success" as getting a question right and consider the 10 questions as 10 independent Bernoulli trials, then X ~ Bin(10 , 0.2) where X is the random variable representing the number of correct answers. What is the probability that the student chooses

a) 3 answers correctly?
b) 7 answers correctly?
c) fewer than 3 answers correctly?
d) at least 5 answers correctly?

Solution:
a. $p\,(X = 3) = f(3) = 10C3\,(0.2)^3\,(0.8)^7$
b. $p\,(X = 7) = f\,(7) = 10C7\,(0.2)^7\,(0.8)^3 = 0.000786$
Attempt the others

Figure 2: Mean and Variance

Mean and standard deviation of a binomial random variable

If X is a binomial random variable with n trials, probability of success p and probability of failure q, then the mean, variance and standard deviation of X can be calculated by using the following formulae.

$$\text{mean} = E(X) = \mu = np$$
$$\text{var}(X) = \sigma^2 = npq$$
$$\text{standard deviation } (X) = \sqrt{npq}.$$

Example

For T = the number of tails when a coin is flipped 3 times, n = 3, p = q = ½ .

$$E(T) = \mu = 3 \times \frac{1}{2} = \frac{3}{2}$$
$$\sigma = \sqrt{3 \times 0.5 \times 0.5} = \sqrt{0.75} = 0.866$$

Note: A Binomial random variable with n=1 is simply a Bernoulli trial and is sometimes referred to as a Bernoulli distribution.

1.3 Poisson distribution

1.3.1 Poisson random variable:

If random variable X counts the number of events that occur at random in an interval of time or space, then X is a Poisson random variable. The average number of events that occur in the time/space interval is denoted by µ or λetc. A short hand way of referring to a Poisson distributed random variable X with average (mean) rate of occurrence µ or λ etc, is X ~ Po(µ) or $X \sim Po(\lambda)$etc.

Examples

1. The number of bad cheques presented for daily payment at a bank. (interval=time)
2. The number of road deaths per month. (interval=time)
3. The number of bacteria in a given culture. (interval here cold be volume e.g. mm^3 of culture
4. The number of defects per square meter on metal sheets being manufactured. (interval here is space)
5. The number of mistakes per typewritten page. (interval here is space)

The probability that X events occur in time/space is given by:

$$p\left(X=x\right)=p\left(X\right)=\frac{\mu^{x}e^{-\mu}}{x!},\;\; for\ x=0,1,2,3,4,...\ and\ \mu>0$$

Example 1

A secretary claims an average mistake rate of 1 per page. A sample page is selected at random and 5 mistakes found. What is the probability of her making 5 or more mistakes if her claim of 1 mistake per page on average is correct?

Figure 3: Solution

In this case $\mu=1$ is claimed and X the number of mistakes ≥ 5. If the claim is true,

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

$$= 1 - \left(e^{-1} + e^{-1} + \frac{e^{-1}}{2!} + \frac{e^{-1}}{3!} + \frac{e^{-1}}{4!}\right)$$

$= 1 - 0.9963$

$= 0.0037.$

The above calculation shows that if the claim of 1 mistake per page on average is true, there is only a 37 in 10 000 chance of getting 5 or more mistakes per page. This remote chance of 5 or more mistakes when an average of 1 mistake per page is true casts doubt on whether the claim of 1 mistake per page on average is in fact true.

Example 2

At a particular restaurant 4 plates are broken, on average, each week. What is the probability that

a) 2 plates are broken next week?

b) at most 4 plates are broken next week?

c) more than 3 plates are broken next week?

Figure 4: Solution

a) Let X = number of plates broken in a week.

Then X ~ Po(4)

$$P(X=2) = \frac{e^{-4}4^2}{2\,!} = 0.1465$$

b) $P(X \leq 4) = f(0) + f(1) + f(2) + f(3) + f(4)$

$$= \frac{e^{-4}4^0}{0\,!} + \frac{e^{-4}4^1}{1\,!} + \frac{e^{-4}4^2}{2\,!} + \frac{e^{-4}4^3}{3\,!} + \frac{e^{-4}4^4}{4\,!}$$

$$= 0.6288$$

Figure 5: Solution

c) $P(X > 3) = 1 - P(X \leq 3)$

$$= 1 - f(0) - f(1) - f(2) - f(3)$$

$$= 1 - \frac{e^{-4}4^0}{0\,!} - \frac{e^{-4}4^1}{1\,!} - \frac{e^{-4}4^2}{2\,!} - \frac{e^{-4}4^3}{3\,!}$$

$$= 0.5665$$

Figure 6: Mean and variance

Mean and standard deviation of a Poisson random variable

- The mean and variance of the Poisson distribution are given by E(X) = μ and var(X) = μ.
- In the case of the Poisson approximation to the binomial distribution

$$E(X) = var(X) = np$$
$$\text{standard deviation} = \sqrt{np}\,.$$

Note

If the average rate of occurrence of μ is given for a particular time/space interval length/size, probability calculations can also be carried out for an interval length/size which is different to the one given.

Example

Calls arrive at switchboard at an average rate of 1 every 15 seconds. What is the probability of not more than 5 calls arriving during a particular minute?

Solution

A mean rate of 1 every 15 seconds is equivalent to a mean rate of 4 every minute. Since the question concerns an interval of 1 minute, $\mu = 4$ (not $\mu = 1$)

Figure 7:

$$P(X \le 5) = e^{-4} + \frac{4^1e^{-4}}{1!} + \frac{4^2e^{-4}}{2!} + \frac{4^3e^{-4}}{3!} + \frac{4^4e^{-4}}{4!} + \frac{4^5e^{-4}}{5!} = 0.7851$$