PROBABILIY AND STATISTICS I

LECTURE ELEVEN

Discrete random variables

Lecturer: Dr. Emily Roche

INTRODUCTION

This lecture will focus on discrete probability distributions.

Intended learning outcomes

At the end of this lecture, you will be able to derive fundamental discrete probability distribution functions, derive the expected values and interpret them and show understanding of the different applications of various discrete functions

References

These lecture notes should be supplemented with relevant topics from the book listed in the Bibliography at the end of the lecture.

Probability distributions

In practical problems, the statistician is often confronted with the necessity of discussing a population which he cannot examine every member. There are two types of populations i.e

1. A finite population which contains a finite number of members e.g population of books in the library.
2. An infinite population which contains an infinite number of members e.g population of pressure at every point in the atmosphere.

Two concepts in connection with the population have to be distinguished namely:

(i) An existent population which is a population of concrete objects

(ii) A hypothetical population as one having no existence in reality but only in imagination.

Example:

Consider the throws of a dice. Each throw can be regarded as an individual. There's an infinite number of throws which can be made with the die provided it does not ware out.

All the possible throws of the die constitute a population. Though we can concretely observe some members of this population by throwing a die, we can never produce them all. Furthermore, even if the die were never thrown at all there would still be a population of possible throws.

Definition:

Hypothetical population consists of all the possible ways in which an observed event can occur. All we need is to picture all the circumstances before the event actually happens.

The hypothetical population is characterized by a number of constants called parameters. Such constants/parameters include:

- Measures of position
- Measures of dispersion together with the moments and measures of skewness
- In multi-variate populations the various total and partial correlations.

Discrete probability distributions

1. **Bernoulli distribution**

A Bernoulli trial is a random experiment which has only two mutually exclusive outcomes i.e

Success [occurrence of an event of interest]

Failure [non-occurrence of an event of interest]

Examples:

- Tossing a coin results in heads or tails
- Testing a manufactured item. (The item can either be effective or non-effective)
- Sitting an examination can either result in a pass or fail.

Let X be a random variable such that:

$$X = \begin{cases} 1 & \text{if the outcome is a success} \\ 0 & \text{if the outcome is a fail} \end{cases}$$

The simple distribution is completely defined and characterized by a single parameter, p, where

$$p = Prob[Success] \text{ i.e}$$

$$P[X = 1] = p \quad \text{and} \quad P[X = 0] = 1 - p = q$$

The density function can be expressed compactly as

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0,1 \\ 0 & elsewhere \end{cases}$$

This is a Bernoulli distribution.

Bernoulli mathematical properties

1. The mean or expected value

$$E(X) = \sum_{x=0}^{1} xf(x) \quad = \sum_{x=0}^{1} xp^x(1-p)^{(1-x)} = 0 + p = p$$

2. The variance

$$Var\,(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_{x=0}^{1} x^2 f(x)$$

$$= \sum_{x=0}^{1} x^2 p^x (1-p)^{(1-x)}$$

$$= 0 + p = p$$

$$Var(X) = p - p^2$$

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

2. **Binomial probability distribution**

Binomial probability distribution is a set of probabilities for discrete events. Discrete events are those whose results or outcomes can be counted. Binomial probabilities are commonly encountered in business situations e.g. in quality control activities the binomial probabilities are frequently used especially when determining the probability of having a certain number of defective items in a given consignment.

The binomial probability distribution is usually characterized by the fact that the binomial events have to fulfill the following properties

i. Each event has two possible outcomes only known as success or failure
ii. The probability of each outcome is independent of the previous outcomes
iii. The sample size is generally fixed
iv. The probabilities of success and failure tend to approach 0.5 if the sample size increases (in the event when an unbiased coin is thrown a number of times)
v. The probabilities are given by the following equation

$$P(r) = {}^{n}C_{r} p^{r}(1-p)^{n-r}$$

$$= \frac{n!}{r!\,(n-r)!} p^{r}(1-p)^{n-r}$$

Where p = Probability of success

r = number of successes

n = sample size

$q = 1 - P$ = Probability of failure

Example

A medical survey was conducted in order to establish the proportion of the population which was infected with cancer. The results indicated that 40% of the population were suffering from the disease.

A sample of 6 people was later taken and examined for the disease. Find the probability that the following outcomes were observed

Only one person had the disease

Exactly two people had the disease

At most two people had the disease

At least two people had the disease

Three or four people had the disease

Solution

$$P(\text{a person having cancer}) = 40\% = 0.4 = P$$

$$P(\text{a person not having cancer}) = 60\% = 0.6 = 1 - p = q$$

P (only one person having cancer)

$$= \ {}^{6}C_{1}(0.4)(0.6)^{5} = \frac{6!}{1!\,5!}(0.4)^{1}(0.6)^{5} = 0.1866$$

P (exactly 2 people had the disease)

$$= \ {}^{6}C_{2}(0.4)^{2}(0.6)^{4} = \frac{6!}{2!\,4!}(0.4)^{2}(0.6)^{4} = 0.311$$

P (at most 2 people had the disease) = $P(0) + P(1) + P(2) = P(0)\ or\ P(1)\ or\ P(2)$

Each of the probabilities are computed and then added. $P(1)$ and $P(2)$ have been computed above, so

$$P(0) = \ {}^{6}C_{0}(0.4)^{0}(0.6)^{6} = \frac{6!}{0!\,6!}(0.4)^{0}(0.6)^{6} = 0.0467$$

Therefore P (at most 2 people had the disease) = $0.0467 + 0.1866 + 0.311 = 0.5443$

P (at least 2 people had the disease) = $P\ (2) + P\ (3) + P\ (4) + P\ (5) + P\ (6)$

This is easier calculated using the complementation rule as = $1 - [P\ (0) + P\ (1)]$

because of the fact that probability of a sample space $P(S) = 1$

$$= 1 - (0.0467 + 0.1866) = 0.7667$$

P (3 or 4 people had the disease) = $P\ (3) + P\ (4)$

$$= \ {}^{6}C_{3}(0.4)^{3}(0.6)^{3} + \ {}^{6}C_{4}(0.4)^{4}(0.6)^{2} = \frac{6!}{3!\,3!}(0.4)^{3}(0.6)^{3} + \frac{6!}{2!\,4!}(0.4)^{4}(0.6)^{2}$$

$$= 0.27648 + 0.13824 = 0.41472$$

Binomial mathematical properties

The mean or expected value = $n \times p = np$

Where; n = Sample Size

p = Probability of success

The variance = npq

Where; q = probability of failure = $1 - p$

The standard deviation = $\sqrt{npq}$

Example

A firm is manufacturing 45,000 units of nuts. The probability of having a defective nut is 0.15

Calculate the following

i. The expected number of defective nuts
ii. The variance and standard deviation of the defective nuts in a daily consignment of 45,000

Solution

$$n = 45{,}000; \quad p = 0.15; \quad q = 0.85$$

i. The expected number of defective nuts = $np = 45000 \times 0.15 = 6750.$
ii. The variance = $npq = 45000 \times 0.15 \times 0.85 = 5737.5$

The standard deviation = $\sqrt{npq} = \sqrt{5737.5} = 75.74$

3. Poisson probability distribution

This is a set of probabilities which is obtained for discrete events which are described as being rare. Occasions similar to binominal distribution but have very low probabilities and large sample size.

Examples of such events in business are as follows:

i. Telephone congestion at midnight
ii. Traffic jams at certain roads at 9 o'clock at night
iii. Sales boom
iv. Attaining an age of 100 years (Centurion)

Poisson probabilities are frequently applied in business situations in order to determine the numerical probabilities of such events occurring.

The formula used to determine such probabilities is as follows

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

Where x = Number of successes

λ = mean number of the successes in the sample ($\lambda = np$)

Example

A manufacturer assures his customers that the probability of having defective item is 0.005. A sample of 1000 items was inspected. Find the probabilities of having the following possible outcomes

i. Only one is defective

ii. At most two defective

Solution

$$\lambda = np = 1000 \times 0.005 = 5$$

$P(\text{Only one is defective}) = P(x = 1) = \frac{2.718^{-5}5^{1}}{1!} = 0.0337$

$P(\text{at most two defective}) = P(x \leq 2) = P(0) + P(1) + P(2) = 0.00674 + 0.0337 + 0.08427 = 0.012471$

Poisson mathematical properties

The mean or expected value = $np = \lambda$

Where; n = Sample Size; p = Probability of success

The variance $= np = \lambda$

Standard deviation = $\sqrt{np} = \sqrt{\lambda}$

Example

The probability of a rare disease striking a given population is 0.003. A sample of 10000 was examined. Find the expected number of people suffering from the disease and hence determine the variance and the standard deviation.

Solution

Sample size $n = 10000$

$$P \text{ (a person suffering from the disease) } = 0.003 = p$$

∴ expected number of people suffering from the disease

$$Mean = np = \lambda = 10000 \times 0.003 = 30$$

$$Variance = np = 30$$

$$Standard\ deviation = \sqrt{30} = 5.477$$

Bibliography

Gupta, SP (Dr.), (2014). *Statistical methods* (43rd Ed.). Sultan Chand & Sons.

S. C. Gupta and V. K. Kapoor, (2020). *Fundamentals of mathematical Statistics* (12th Ed). Sultan Chand & Sons.