Discrete Probability Distributions

Discrete Probability Distributions

Introduction

It is often possible to model real systems by using the same or similar random experiments and their associated random variables. Numerical random variables may be classified in two broad but distinct categories called discrete random variables and continuous random variables. Often, discrete random variables are associated with counting while continuous random variables are associated with measuring. In HELM 42. you will meet contingency tables and deal with non-numerical random variables. Generally speaking, discrete random variables can take values which are separate and can be listed. Strictly speaking, the real situation is a little more complex but it is sufficient for our purposes to equate the word discrete with a finite list. In contrast, continuous random variables can take values anywhere within a specified range. This Section will familiarize you with the idea of a discrete random variable and the associated probability distributions. The Workbook makes no attempt to cover the whole of this large and important branch of statistics but concentrates on the discrete distributions most commonly met in engineering. These are the binomial, Poisson and hypergeometric distributions.

Prerequisites

Before starting this Section you should ...

- understand the concepts of probability

Learning Outcomes

On completion you should be able to ...

- explain what is meant by the term discrete random variable
- explain what is meant by the term discrete probability distribution
- use some of the discrete probability distributions which are important to engineers

1. Discrete probability distributions

We shall look at discrete distributions in this Workbook and continuous distributions in HELM 38. In order to get a good understanding of discrete distributions it is advisable to familiarise yourself with two related topics: permutations and combinations. Essentially we shall be using this area of mathematics as a calculating device which will enable us to deal sensibly with situations where *choice* leads to the use of very large numbers of possibilities. We shall use combinations to express and manipulate these numbers in a compact and efficient way.

Permutations and Combinations

You may recall from HELM 35.2 concerned with probability that if we define the probability that an event A occurs by using the definition:

$$\mathsf{P}(A) = \frac{\text{The number of equally likely experimental outcomes favourable to } A}{\text{The total number of equally likely outcomes forming the sample space}} = \frac{a}{n}$$

then we can only find $\mathsf{P}(A)$ provided that we can find both a and n. In practice, these numbers can be very large and difficult if not impossible to find by a simple counting process. Permutations and combinations help us to calculate probabilities in cases where counting is simply not a realistic possibility.

Before discussing permutations, we will look briefly at the idea and notation of a factorial.

Factorials

The **factorial** of an integer n commonly called 'factorial n' and written $n!$ is defined as follows:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1 \qquad n \geq 1$$

Simple examples are:

$$3! = 3\times2\times1 = 24 \qquad 5! = 5\times4\times3\times2\times1 = 120 \qquad 8! = 8\times7\times6\times5\times4\times3\times2\times1 = 40320$$

As you can see, factorial notation enables us to express large numbers in a very compact format. You will see that this characteristic is very useful when we discuss the topic of permutations. A further point is that the definition above falls down when $n = 0$ and we define

$$0! = 1$$

Permutations

A **permutation** of a set of distinct objects places the objects **in order**. For example the set of three numbers $\{1, 2, 3\}$ can be placed in the following orders:

1,2,3   1,3,2   2,1,3   2,3,1   3,2,1   3,1,2

Note that we can choose the first item in 3 ways, the second in 2 ways and the third in 1 way. This gives us $3\times2\times1 = 3! = 6$ distinct orders. We say that the set $\{1, 2, 3\}$ has the distinct permutations

1,2,3   1,3,2   2,1,3   2,3,1   3,2,1   3,1,2

Example 1
Write out the possible permutations of the letters A, B, C and D.

4*3*2*1 = 24

In general we can order n distinct objects in $n!$ ways.
Suppose we have r different types of object. It follows that if we have n_1 objects of one kind, n_2 of another kind and so on then the n_1 objects can be ordered in $n_1!$ ways, the n_2 objects in $n_2!$ ways and so on. If $n_1 + n_2 + \cdots + n_r = n$ and if p is the number of permutations possible from n objects we may write

$$p \times (n_1! \times n_2! \times \cdots \times n_r!) = n!$$

and so p is given by the formula

$$p = \frac{n!}{n_1! \times n_2! \times \cdots \times n_r!}$$

Very often we will find it useful to be able to calculate the number of permutations of n objects taken r at a time. Assuming that we do not allow repetitions, we may choose the first object in n ways, the second in $n-1$ ways, the third in $n-2$ ways and so on so that the r^{th} object may be chosen in $n-r+1$ ways.

Example 2
Find the number of permutations of the four letters A, B, C and D taken three at a time.

4P3 = 24

In general the numbers of permutations of n objects taken r at a time is

$$n(n-1)(n-2)\dots(n-r+1) \qquad \text{which is the same as} \qquad \frac{n!}{(n-r)!}$$

This is usually denoted by ${}^nP_r$ so that

$${}^nP_r = \frac{n!}{(n-r)!}$$

If we allow repetitions the number of permutations becomes n^r (can you see why?).

Example 3

Find the number of permutations of the four letters A, B, C and D taken two at a time.

4P2 = 12

Combinations

A **combination** of objects takes **no account of order** whereas a permutation does. The formula ${}^nP_r = \dfrac{n!}{(n-r)!}$ gives us the number of ordered sets of r objects chosen from n. Suppose the number of sets of r objects (taken from n objects) in which order is not taken into account is C. It follows that

$$C \times r! = \frac{n!}{(n-r)!} \quad \text{and so } C \text{ is given by the formula} \qquad C = \frac{n!}{r!(n-r)!}$$

We normally denote the right-hand side of this expression by ${}^nC_r$ so that

$${}^nC_r = \frac{n!}{r!(n-r)!} \qquad \text{A common alternative notation for } {}^nC_r \text{ is } \binom{n}{r}.$$

Example 4

How many car registrations are there beginning with $NP05$ followed by three letters? Note that, conventionally, I, O and Q may not be chosen.

(a) How many different signals consisting of five symbols can be sent using the dot and dash of Morse code?

(b) How many can be sent if five symbols *or less* can be sent?

Your solution

A box contains 50 resistors of which 20 are deemed to be 'very high quality' , 20 'high quality' and 10 'standard'. In how many ways can a batch of 5 resistors be chosen if it is to contain 2 'very high quality', 2 'high quality' and 1 'standard' resistor?

2. Random variables

A random variable X is a quantity whose value cannot be predicted with certainty. We assume that for every real number a the probability $\mathsf{P}(X = a)$ in a trial is well-defined. In practice, engineers are often concerned with two broad types of variables and their probability distributions: discrete random variables and their distributions, and continuous random variables and their distributions. Discrete distributions arise from experiments involving counting, for example, road deaths, car production and aircraft sales, while continuous distributions arise from experiments involving measurement, for example, voltage, corrosion and oil pressure.

Discrete random variables and probability distributions

A random variable X and its distribution are said to be discrete if the values of X can be presented as an ordered list say $x_1, x_2, x_3, \ldots$ with probability values $p_1, p_2, p_3, \ldots$. That is $\mathsf{P}(X = x_i) = p_i$. For example, the number of times a particular machine fails during the course of one calendar year is a discrete random variable.

More generally a discrete distribution $f(x)$ may be defined by:

$$f(x) = \begin{cases} p_i & \text{if } x = x_i \quad i = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

The distribution function $F(x)$ (sometimes called the cumulative distribution function) is obtained by taking sums as defined by

$$F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} p_i$$

We sum the probabilities p_i for which x_i is less than or equal to x. This gives a step function with jumps of size p_i at each value x_i of X. The step function is defined for all values, not just the values x_i of X.

Key Point 1

Probability Distribution of a Discrete Random Variable

Let X be a random variable associated with an experiment. Let the values of X be denoted by $x_1, x_2, \ldots, x_n$ and let $\mathsf{P}(X = x_i)$ be the probability that x_i occurs. We have two necessary conditions for a valid probability distribution:

- $\mathsf{P}(X = x_i) \geq 0 \quad$ for all $\; x_i$
- $\displaystyle\sum_{i=1}^{n} \mathsf{P}(X = x_i) = 1$

Note that n may be uncountably large (infinite).

(These two statements are sufficient to guarantee that $\mathsf{P}(X = x_i) \leq 1$ for all x_i.)

Example 5

Turbo Generators plc manufacture seven large turbines for a customer. Three of these turbines do not meet the customer's specification. Quality control inspectors choose two turbines at random. Let the discrete random variable X be defined to be the number of turbines inspected which meet the customer's specification.

(a) Find the probabilities that X takes the values $0, 1$ or 2.

(b) Find and graph the cumulative distribution function.

3. Mean and variance of a discrete probability distribution

If an experiment is performed N times in which the n possible outcomes $X = x_1, x_2, x_3, \ldots, x_n$ are observed with frequencies $f_1, f_2, f_3, \ldots, f_n$ respectively, we know that the mean of the distribution of outcomes is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\displaystyle\sum_{i=1}^{n} f_ix_i}{\displaystyle\sum_{i=1}^{n} f_i} = \frac{1}{N}\sum_{i=1}^{n} f_ix_i = \sum_{i=1}^{n}\left(\frac{f_i}{N}\right)x_i$$

(Note that $\displaystyle\sum_{i=1}^{n} f_i = f_1 + f_2 + \cdots + f_n = N$.)

The quantity $\dfrac{f_i}{N}$ is called the **relative frequency** of the observation x_i. Relative frequencies may be thought of as akin to probabilities; informally we would say that the chance of observing the outcome x_i is $\dfrac{f_i}{N}$. Formally, we consider what happens as the number of experiments becomes very large. In order to give meaning to the quantity $\dfrac{f_i}{N}$ we consider the limit (if it exists) of the quantity $\dfrac{f_i}{N}$ as $N \to \infty$. Essentially, we define the probability p_i as

$$p_i = \lim_{N\to\infty} \frac{f_i}{N}$$

Replacing $\dfrac{f_i}{N}$ with the probability p_i leads to the following definition of the mean or **expectation** of the discrete random variable X.

Key Point 2

The Expectation of a Discrete Random Variable

Let X be a random variable with values $x_1, x_2, \ldots, x_n$. Let the probability that X takes the value x_i (i.e. $\mathsf{P}(X = x_i)$) be denoted by p_i. The mean or **expected value** or **expectation** of X, which is written $\mathsf{E}(X)$ is defined as:

$$\mathsf{E}(X) = \sum_{i=1}^{n} x_i\ \mathsf{P}(X = x_i) = p_1x_1 + p_2x_2 + \cdots + p_nx_n$$

The symbol μ is sometimes used to denote $\mathsf{E}(X)$.

The expectation $\mathsf{E}(X)$ of X is the value of X which we expect on average. In a similar way we can write down the expected value of the function $g(X)$ as $\mathsf{E}[g(X)]$, the value of $g(X)$ we expect on average. We have

$$\mathsf{E}[g(X)] = \sum_{i}^{n} g(x_i)f(x_i)$$

In particular if $g(X) = X^2$, we obtain $\qquad \mathsf{E}[X^2] = \sum_{i}^{n} x_i^2 f(x_i)$

The variance is usually written as σ^2. For a frequency distribution it is:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \mu)^2 \qquad \text{where } \mu \text{ is the mean value}$$

and can be expanded and 'simplified' to appear as:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \mu^2$$

This is often quoted in words:

The variance is equal to the mean of the squares minus the square of the mean.

We now extend the concept of variance to a random variable.

Key Point 3

The Variance of a Discrete Random Variable

Let X be a random variable with values $x_1, x_2, \ldots, x_n$. The variance of X, which is written $\mathsf{V}(X)$ is defined by

$$\mathsf{V}(X) = \sum_{i=1}^{n} p_i(x_i - \mu)^2$$

where $\mu \equiv \mathsf{E}(X)$. We note that $\mathsf{V}(X)$ can be written in the alternative form

$$\mathsf{V}(X) = \mathsf{E}(X^2) - [\mathsf{E}(X)]^2$$

The standard deviation σ of a random variable is $\sqrt{\mathsf{V}(X)}$.

Example 6

A traffic engineer is interested in the number of vehicles reaching a particular crossroads during periods of relatively low traffic flow. The engineer finds that the number of vehicles X reaching the crossroads per minute is governed by the probability distribution:

x	0	1	2	3	4
$\mathsf{P}(X = x)$	0.37	0.39	0.19	0.04	0.01

(a) Calculate the expected value, the variance and the standard deviation of the random variable X.

(b) Graph the probability distribution $\mathsf{P}(X = x)$ and the corresponding cumulative probability distribution $F(x) = \sum_{x_i \le x} \mathsf{P}(X = x_i)$.

Find the expectation, variance and standard deviation of the number of Heads in the three-coin toss experiment.

Your solution

Exercises

1. A machine is operated by two workers. There are sixteen workers available. How many possible teams of two workers are there?

2. A factory has 52 machines. Two of these have been given an experimental modification. In the first week after this modification, problems are reported with thirteen of the machines. What is the probability that both of the modified machines are among the thirteen with problems assuming that all machines are equally likely to give problems,?

3. A factory has 52 machines. Four of these have been given an experimental modification. In the first week after this modification, problems are reported with thirteen of the machines. What is the probability that exactly two of the modified machines are among the thirteen with problems assuming that all machines are equally likely to give problems?

4. A random number generator produces sequences of independent digits, each of which is as likely to be any digit from 0 to 9 as any other. If X denotes any single digit, find $\mathsf{E}(X)$.

5. A hand-held calculator has a clock cycle time of 100 nanoseconds; these are positions numbered $0, 1, \ldots, 99$. Assume a flag is set during a particular cycle at a random position. Thus, if X is the position number at which the flag is set.

   $$\mathsf{P}(X = k) = \frac{1}{100} \qquad k = 0, 1, 2, \ldots, 99.$$

   Evaluate the average position number $\mathsf{E}(X)$, and σ, the standard deviation.

   (Hint: The sum of the first k integers is $k(k+1)/2$ and the sum of their squares is: $k(k+1)(2k+1)/6$.)

6. Concentric circles of radii 1 cm and 3 cm are drawn on a circular target radius 5 cm. A darts player receives 10, 5 or 3 points for hitting the target inside the smaller circle, middle annular region and outer annular region respectively. The player has only a 50-50 chance of hitting the target at all but if he does hit it he is just as likely to hit any one point on it as any other. If $X =$ 'number of points scored on a single throw of a dart' calculate the expected value of X.

Answers