

Random Variables

Previously we defined an experiment as any process that generates will defined outcomes.

We can describe the outcomes of an experiment using numbers and this numerical description is called a random variable. A random variable is a numerical description of the outcome of the experiment.

Random Variables

Remark: We will call random variables R.V.s for short. We associate a number with each outcome of the experiment and the particular numerical value of the random variable depends on the outcome.

For example, consider the experiment of picking an employee at random from an office. We are interested in the number of years work experience the employee has.

Random Variables

The outcome of the experiment could be one year, two years and so on. The random variable is the number of experience the employee has and the value of the random variable can change every time we repeat the experiment.

Random Variables

We can repeat the experiment of picking an employee at random many times and count the number of times that the outcome was one year, two years and so on. Recall that the probability $p = r/n$ where n is the number of time we carry out an experiment (i.e. number of trials) r is the number of times we get the result we are interested in (i.e. number of successes)

Random Variables

We can therefore associate probabilities with the various possible values that the random variable might take.

The values of the random variable and the associated probabilities form a probability distribution.

Years (R.V.) 1 2 3 4 5 Probability 0.2 0.3 0.3 0.1 0.1

Random Variables

The value of the random variable is unknown before we carry out the experiment but using the probability distribution in the table above, we can say that the probability that an employee selected at random will have one years experience is 0.2 or 20% Although experiments have outcomes that are naturally described using numbers, other do not. For example, an employee might be asked if they enjoy their work. The random variable is enjoyment of work, and the outcomes are yes and no.

Random Variables

We can arbitrarily assign the value zero to an employee who says no and one to an employee who says yes.

Random Variables

In this way, the random variable still provides a numerical description of the outcome of the experiment. Many methods of statistical analysis assume that data collected follows a known probability distribution. We will later examine the most commonly used probability distributions.

Random Variables

- A random variable is defined as a numerical event whose value is determined by a chance process.
- When probability values are assigned to all possible numerical values of a random variable X , either by a listing or by a mathematical function, the result is a *probability distribution*.

Random Variables

- The sum of the probabilities for all the possible numerical outcomes must equal one.
- Individual probability values may be denoted by the symbol $f(x)$, which indicates that a mathematical function is involved, by $P(X = x)$, which recognizes that the random variable can have various specific values, or simply by $P(x)$.

Random Variables

- The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be ‘heads’ or ‘tails’.
- However, we often want to represent outcomes as numbers.
- A *random variable* is a function that associates a unique numerical value with every outcome of an experiment.
- The value of the random variable will vary from trial to trial as the experiment is repeated.

Random Variables

- Numeric values can be assigned to outcomes that are not usually considered numeric.
- For example, we could assign a ‘head’ a value of 0, and a ‘tail’ a value of 1, or vice versa.

Random Variables

There are two types of random variable - discrete and continuous. The distinction between both types will be important later on in the course.

Examples

- A coin is tossed ten times. The random variable X is the number of tails that are noted. X can only take the values $\{0, 1, \dots, 10\}$, so X is a discrete random variable.
- A light bulb is burned until it burns out. The random variable Y is its lifetime in hours. Y can take any positive real value, so Y is a continuous random variable.

Discrete Random Variable

- A discrete random variable is one which may take on only a countable number of distinct values such as $\{0, 1, 2, 3, 4, \dots\}$.
- Discrete random variables are usually (but not necessarily) counts.
- If a random variable can take only a finite number of distinct values, then it must be discrete.

Discrete Random Variable

- Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten.

Continuous Random Variable

- A continuous random variable is one which takes an infinite number of possible values.
- Continuous random variables are usually measurements.
- Examples include height, weight, the amount of sugar in an orange, the time required to run a computer simulation.

Discrete Random Variables

- For a discrete random variable observed values can occur only at isolated points along a scale of values. In other words, observed values must be integers.
- Consider a six sided die: the only possible observed values are 1, 2, 3, 4, 5 and 6.
- It is not possible to observe values that are real numbers, such as 2.091.
- (*Remark: it is possible for the average of a discrete random variable to be a real number.*)

Discrete Random Variables

- Therefore, it is possible that all numerical values for the variable can be listed in a table with accompanying probabilities.
- There are several standard probability distributions that can serve as models for a wide variety of discrete random variables involved in business applications.

Discrete probability distributions

The discrete probability distributions that described in this course are

- the binomial distribution,
- the geometric distribution,
- the hypergeometric distribution,
- the Poisson distributions.