1. Introduction & Properties of Random Variables

Welcome to the series of E-learning modules on illustration on random variables and their properties. In this module we discuss about illustrations on continuous random variable, properties of random variable like location and mode, their measures and comparison of different measures of location and dispersion.

By the end of this session, you will be able to:

- Understand illustration on continuous random variable
- Understand the properties location and dispersion of Random Variable
- Understand different measures of locations, namely, mean, median and mode
- Compare location measures
- Understand different measures of dispersions namely, inter-quartile range, semi-inter-quartile range, mean absolute deviation, variance and standard deviation
- Compare dispersion measures

Let us now consider an illustration on random variable.
We have discussed about discrete random variables and their distributions in the last semester. Let us consider some of the illustration on continuous random variables.
Consider modeling the distribution of the age that a person dies at. Age of death, measured perfectly with all the decimals and no rounding, is a continuous random variable.
Another example of a continuous variable is Body temperature, assuming any value between thirty and forty five degrees Celsius.

Some of the other examples where the random variable follows continuous distribution are:

- The speed of a car
- The concentration of a chemical in a water sample
- Tensile strengths
- Heights of people in a population
- Lengths or areas of manufactured components
- Life time of a battery
- The annual rain fall in a region etc.

Let us now discuss the properties of random variables.
We have discussed the idea of probability distributions, in particular the distributions of discrete variables. Now we proceed to continuous variables.
Here we will discuss two important properties by which we may characterize random variables, regardless of the probability distribution are location and dispersion.

For any random variable, the domain gives all possible values of the variable and the probability distribution gives the likelihood of those values. Together these two pieces of information provide a complete description of the properties of the random variable.

Two important properties of a variable that are contained in this description are:

- Location: the central tendency of the variable, which describes a value around which the variable tends to cluster, and
- Dispersion: the typical range of values that might be expected to be observed in experiments. This gives some idea of the spread in values that might result from our experiment or experiments

The probability distribution contains all the information needed to determine these characteristics, as well as still more esoteric descriptors of the properties of random variables. Since we have discussed the distributions of discrete variables, we will tend to use these in our discussions and examples; the same concepts apply, with very little modification, to continuous variables.

Now let us discuss the Property of location that is central tendency of a random variable.
The central tendency, or location, of a variable can be indicated by any (or all) of the following: the mean, the median, or the mode. Although most people are familiar with means, the other two properties are actually easier to understand.

2. Mode, Median, Mean

Mode

Let us first consider the mode.

The mode is the most probable value of a discrete variable. More generally, it is the maximum of the probability distribution function. In continuous distribution, it is the value of x for which the probability density function (pdf) attains its maximum.

The value of ‘x’ mode such that

f of x mode is equal to P maximum, the maximum value of the probability density function

Multi-modal probability distributions have more than one mode. The distributions with two modes are bimodal, and so on. Although multi-modal distributions may have several local maxima, there is usually a single global maximum that is the single most probable value of the random variable.

Median

Now let us consider median, which is denoted by Q two.

The median is only a little more complicated than the mode: it is the value Q two such that,

Probability of (x less than Q TWO) is equal to Probability of (x is greater than Q TWO)

In other words, there is an equal probability of observing a value greater than or less than the median.

The median is also the second quartile— hence *the origin of the symbol Q TWO.*

Any distribution can be divided into four equal “pieces,” such that:

Probability of (x is less than Q one) is equal to Probability of (Q one less than x less than Q TWO) is equal to Probability of (Q TWO is less than x less than Q three) is equal to Probability of (x greater than Q three)

The boundaries Q one, Q TWO (that is, the median), and Q three are the quartiles of the probability distribution.

Mean

Now finally let us consider mean.

Before defining the mean, it is helpful to discuss a mathematical operation called the weighted sum. Almost everybody performs weighted sums — especially students calculating test averages or grade point averages!

For example, let’s imagine that a student has taken two tests and a final test, scoring eighty five and eighty points on the tests, and seventy five points on the final. An “un-weighted average” of these three numbers is eighty points; however, the final is worth (that is, weighted) more than the tests.

Suppose that the instructor feels that the final is worth sixty per cent of the test grade, while the other two tests are worth twenty per cent each. The weighted sum would be calculated as follows. Let W one is equal to w two is equal to zero point two, and w threeis equal to zero point six,

then weighted score is:

Summation w I into (score) is equal to =w one into zero point eight , five) plus w two into zero point eight zero plus w three into zero point seven, five is equal to seventy eight.

The final score, seventy eight, is a weighted sum. Since the final test is given more weight than the two other tests, the weighted sum is closer to the final score seventy five than is the un-weighted average eighty. Grade point averages are calculated on a similar principle, where the weights for each grade are determined by the course credit hours.

To choose an example from chemistry, the atomic weights listed in the periodic tables are weighted averages of isotope masses; the weights are determined by the relative abundance of the isotopes.

In general, a weighted sum is represented by the expression weighted sum is equal to summation w_i into x_i
Where x_i is the individual values, and w_i are the corresponding weights.
When the sum of all the individual weights is one (that is w_i is equal to one) then the weighted sum is often referred to as a weighted average.

The mean or expected value, of the continuous variable is defined as follows.
Expectation of X is equal to mu x is equal to integration over minus infinity to infinity, x into f of x dx.
Where f of (x) is a mathematical function that defines the probability distribution of the random variable x.

3. Comparison Between the Different Measures of Location

Now let us compare the different measures of location, which is mode, median and mean.
We have defined three different indicators of the location of a random variable: the mean, median and mode. Each of these has a slightly different meaning.
Imagine that you are betting on the outcome of a particular experiment:

- If you choose the mode, you are essentially betting on the single most likely outcome of the experiment.
- If you choose the median, you are equally likely to be larger or smaller than the outcome.
- If you choose the mean, you have the best chance of being closest to the outcome

A random variable cannot be predicted exactly, but each of the three indicators gives a sense of the value the random variable is likely to be near. In most applications, the mean gives the best single description of the location of the variable.

Here we have given two graphs, named as ‘a’ and ‘b’

Figure (a)

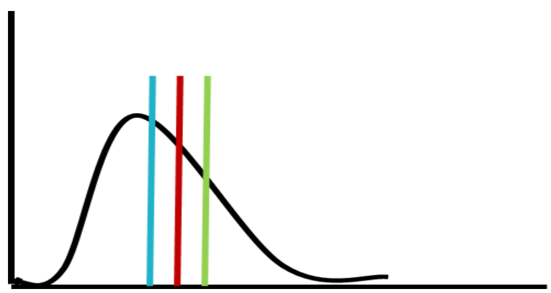
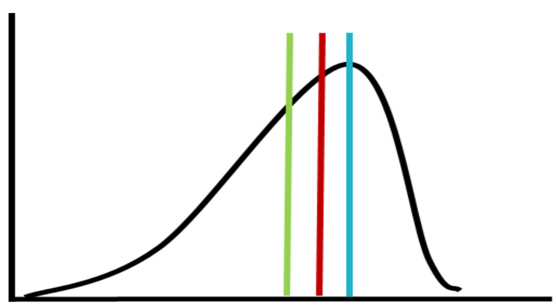


Figure (b)



In the above figure, among the three vertical lines, the blue line denotes the mode, the red line median and the green line mode.
Figure (a) denotes the positively skewed distribution and figure (b) denotes the negatively skewed distribution.

Now from above graph let us observe how different the values of the mean, median and

mode are.
It turns out that the three values are different only for asymmetric distributions, such as the two shown in figure in the previous slide. If a distribution is skewed to the right (or positively skewed as in fig (a)) then
mu X is greater than Q TWO which is greater than x mode

For distributions skewed to the left (negatively skewed as in fig.(b))
mu X is less than Q TWO which is less than x mode
For symmetrical ("bell-shaped") distributions, the mean, median and mode all have exactly the same value.

Illustration
Obtain mean median and mode for the following continuous frequency distribution having probability density function f of (x) is equal to three into x square, where zero less than or equal to x less than or equal to one and comment on the type of the distribution.

Solution :
Mean of the distribution is given by,
Expectation of X is equal to mu X is equal to minus infinity to infinity x into f of x d x
Is equal to three into integral over zero to one, x into x square d x
Is equal to three into x power four divided by four, ranges from zero to one
By substituting the limits and simplifying we get,
Mu x is equal to three divided by four which is equal to zero point seven five.

Median divides the whole distribution into two equal parts. Suppose we consider median Q TWO as some unknown number m, then
Integral from minus infinity to m, f of x d x is equal to integral over m to infinity; f of x d x is equal to half.
Consider any one integral, that is the first one, and let us solve for m

Integral over minus infinity to m f of x d x is equal to three into integral over zero to m x square d x is equal to half.
On integration, we get left hand side as three into x cube divided by three, ranges from zero to m which is equal to half
Implies, m cube is equal to half
Or m is equal to one divided by cube root of two which is equal to zero point seven, nine, three, seven.

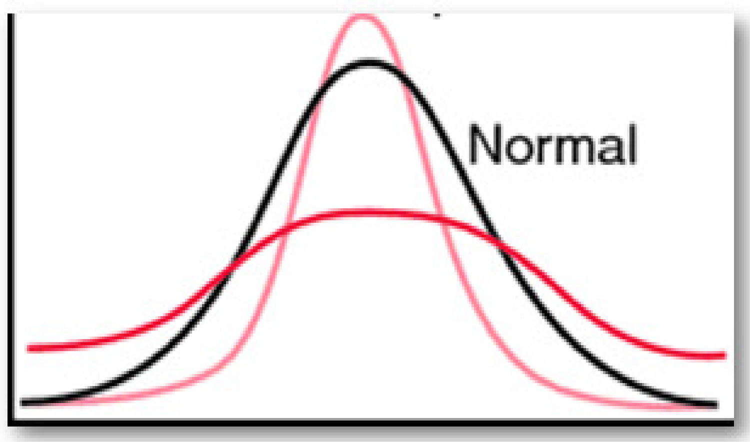
Mode is the value where the distribution attains its maximum. By looking at the density function it is obvious that the function attains its maximum when x is equal to 1, the extreme point.
Hence x mode is equal to one
Observe that mew x (which is zero point seven, five) is less than Q TWO (that is zero point seven, nine, three, seven) is less than x mode (which is one)
Hence the distribution is negatively skewed.

4. Dispersion of Continuous Random Variables

Now let us consider the second property that is dispersion of continuous random variables. Some variables are more “variable,” more uncertain, than others. Of course, theoretically speaking, a variable may assume any one of the range of values in the domain. However, when speaking of the variability of a random variable, we generally mean the range of values that would commonly (i.e., most probably) be observed in an experiment. This property is called the dispersion of the random variable. Dispersion refers to the range of values that are commonly assumed by the variable.

Experiments that produce outcomes that are highly variable will be more likely to give values that are farther from the mean than similar experiments that are not as variable. In other words, probability distributions tend to be broader as the variability increases. Following figure compares the probability functions or “probability density functions” of two continuous variables.

Figure 1



In this graph there are three curves. The black curve which is at the center is kept as base to see the variation of other two distributions.
The variable described by the broader probability distribution that is the dark red line, is more likely to be farther from the mean than the other variable shown as the light red line. The black curve shows the ideal curve which almost gives the correct value.

As with the mean, it is convenient to describe variability with a single value. Three common ways to do so are:

i. Inter-quartile range and semi-inter-quartile range
ii. Mean absolute deviation
iii. Variance standard deviation

Now let us them consider one by one.

The first one is the inter-quartile range and the semi-inter-quartile range
The inter-quartile range, IQR, is the difference between the first and third quartiles
IQR is equal to Q three minus Q one
This is a measure of dispersion because the "wider" a distribution gets, the greater the difference between the quartiles.

The semi-inter-quartile range, QR is probably the more commonly used measure of dispersion; it is simply half the inter-quartile range.
QR is equal to (Q three minus Q one) divided by two

Now consider the second measure, The mean absolute deviation
Since the dispersion describes the spread of a random variable about its mean, it makes sense to have a quantitative descriptor of this quantity. The mean absolute deviation, MD, is exactly what it sounds like: the expected value (that is, the mean) of the absolute deviation of a variable from its mean value, mew X.
MD is equal to Expectation of (modulus of x minus mew x)

The concept behind the mean absolute deviation is quite simple: it indicates the mean or 'typical' distance of a variable from its own mean, μ x
For a continuous random variable,
MD is equal to integration from minus infinity to infinity, modulus of X minus mew X into f of x d x.

5. Variance and Standard Deviation

Now consider the last measure, variance and standard deviation.
Like the mean absolute deviation, the variance and standard deviation measure the dispersion of a random variable about its mean mu x. The variance of a random variable x, sigma X square, is the expected value of (x minus mu X) the whole square, which is the squared deviation of x from its mean value.
That is sigma x square is equal to expectation of X minus mu X the whole square.

As you can see, the concept of the variance is very similar to that of the mean absolute deviation. In fact, the variance is sometimes called the mean squared deviation. The variance for continuous variables is given by
Sigma x square is equal to integration over minus infinity to infinity, x minus mew x the whole square into f of x into d x.
Higher variance signifies greater variability of a random variable.

One problem with using the variance to describe the dispersion of a random variable is that the units of variance are the squared units of the original variable. For example, if x is a length measured in m, then sigma x square has units of m square.
The standard deviation, sigma x, has the same units as x, and so is a little more convenient at times. The standard deviation is simply the positive square root of the variance.

Sometimes the variability of a random variable is specified by the relative standard deviation or RSD.
Therefore, RSD is equal to sigma x divided by mu x.

Now let us compare the different measures of dispersion.
We have described three common ways to measure a random variable's dispersion: semi-inter-quartile range, mean absolute deviation, and the standard deviation. These measures are all related to each other, so, in a sense, it makes no difference which measure we use.

In fact, for distributions that are only moderately skewed, Mean absolute Deviation is nearly equal to zero point eight into sigma and semi inter quartile range is nearly equal to zero point six seven into sigma. For a variety of reasons, the variance and standard deviation are the best measures of dispersion of a random variable.

Let us consider the same illustration which was explained earlier and we find the variance of the distribution and hence standard deviation.
Obtain variance and hence standard deviation for the following continuous frequency distribution having probability density function, f of x is equal to three x square , where x less than or equal to one and more than or equal to zero.

Solution:
We know that sigma x square is equal to integral from minus infinity to infinity, x minus mu x

whole square into f of x dx
Where,
Mu X is equal to minus infinity to infinity x into f of x dx.
This is equal to three into integral over zero to one, x into x square dx.
This is equal to three into x to the power four divided by four, ranges from zero to one.
By substituting the limits and simplifying we get,
Mu x is equal to three divided by four which is equal to zero point seven five.

Now let us find variance sigma x square.
Sigma square is equal to three into integral over zero to one, x minus zero point seven five whole square into x square d x
Squaring and taking x square inside the bracket, we get
three into integral over zero to one, x raised to the power four minus one point five into x cube plus zero point five six two five x square d x.
this is equal to three into x raised to the power five divided by five minus one point five into x to the power four divided by four plus zero point five six two five into x cube divided by three, the range for x is from zero to one.
Which is equal to three into one divided by five minus one point five divided by four plus zero point five, six, two, five divided by three is equal to zero point zero three seven three.
Standard deviation is given by positive square root of zero point zero three seven three is equal to zero point one nine three one.

Here's a summary of our learning in this session where we have:

- Understood illustration on continuous random variable
- Understood the properties of random variable namely, location and dispersion
- Understood different measures of locations, namely, mean, median and mode
- Understood the comparison of location measures
- Explained different measures of dispersions namely, inter-quartile range, semi-inter-quartile range, mean absolute deviation, variance and standard deviation and
- Understood comparison of dispersion measures