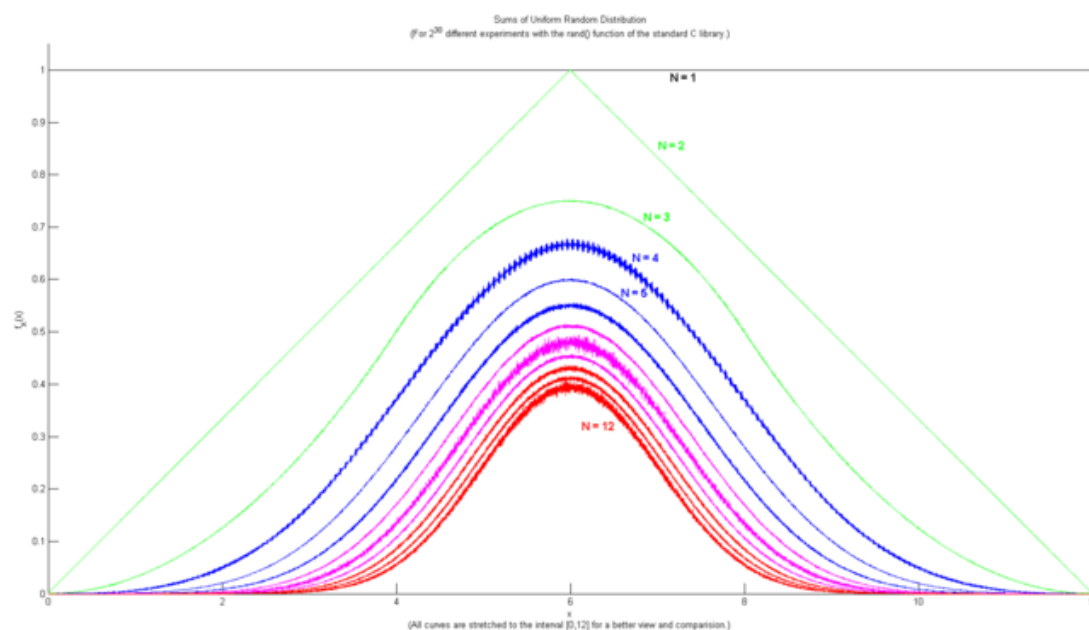


Statistics: Random Variables

[See online here](#)

The probability models included in this article explain sample space, types of random variables and expected values of each sample. The properties of the expected value of each sample in a population, properties of variance and continuous random variable are also explained in this article. The issues in probability models and random variables are another part of this article.



Probability Models

A **mathematical representation of a random phenomenon** with sample spaces, probabilities associated with each event and events within the sample spaces is known as a **probability model**. Here, random variables are numeric values which are generated based on the results of a random event. Random variables are a central point of a probability model.

A random variable X is a real-valued function defined on a sample space, such that B_x as defined above is an event for all x .

Example: The random variable is any number we get from a random selection. Tossing a coin, we can get head (0) or tail (1). Either of the outcomes is a random variable. In this case, a random variable is not numeric.

Sample space

The set of values of a random variable is known as its sample space. For example, if we throw a dice, we have a possibility of outcome 1,2,3,4,5,6. This set of values is expected

as a result of the event i.e. throwing dice is the sample space.

Types of Random Variables

There are **two main types** of random variables i.e. **discrete** and **continuous random variable**.

Continuous random variable

This type of random variable has an **uncountable list of outcomes** associated with the event. These are random variables which are associated with measurements like the score of a soccer team, the height of a group of people and student test result. It is the collection of all the possible values of probability and random variables which are expected to occur as a result of the related event.

Example: An advertisement is on air 10 - 15 times a day on different channels. It is not possible for the audience to make an estimate or count after how much time it will be on air for 24 hours. It can be 10 minutes, 1 hour or 10 hours.

Another example of continuous random variables is the time a passenger has to wait on a bus stop waiting for it. It can be 2 minutes, 10 minutes or a few seconds. It is not possible to list all the possible times associated with the arrival of the bus at the stop.

Discrete random variables

A discrete random variable is a variable which has **measurable/countable outcomes**. The probability of a discrete random variable is known as a **discrete probability**.

If X can take any value in a set D that is countable, then X is said to be discrete. Usually, D is some subset of the integers, so we assume in future that any discrete random variable is integer-valued unless it is stated otherwise.

Example: A jar is filled with 10 red, 12 purple and 5 blue marble pieces. Here the number of red marbles is countable and the probability associated with it can be measurable; hence, it is a discrete random variable.

Another example of a discrete random variable is the number of students present in a class. It is countable and the probability associated with the presence of students and can be measured easily by using related formulas.

Furthermore, even within these two classes of a random variable, there are further subcategories, which it is often useful to distinguish. Here is a short list of some of them:

Constant random variable. If $X(\omega) = C$ for all ω , where C is a constant, then X is a constant.

Indicator random variable. If X can take only the values 0 or 1, then X is said to be an indicator. If we define the event on which $X=1$, $A = \{\omega: X(\omega)=1\}$, then X is said to be the indicator of A .

Expected values

The measure of the center of a random variable is known as the expected value. The evaluation of characteristics of a random variable is considered expected value associated with an event.

Once a probability model is established, the related expected values can be predicted. Finding the expected value for continuous variables is a difficult process, hence it should be restricted to discrete random variables. The random variable is denoted by X hence the expected value of a random variable is shown as $E[X]$.

The following mathematical formula can be used to measure expected values associated with a discrete random variable:

$$E[X] = \sum xP(X = x)$$

Example

The roll of dice has a sample space of 1, 2, 3, 4, 5, and 6. The probability model for a dice roll is:

$$P(X=1) = P(X=6) = 1/12, P(X=2) = P(X=4) = P(X=6) = 1/6, \text{ and } P(X=3) = 1/3.$$

The expected value of these discrete random variables will be as follows:

$$E[X] = 1(1/12) + 2(1/6) + 3(1/3) + 4(1/6) + 5(1/12) + 6(1/6) = 3 \frac{1}{6}.$$

Properties of Expected Value

The properties of expected value of discrete random variables include:

Let C be **the constant**

X be **a random variable**

With a constant, the expected value of a random variable will be as follows:

$$E[X+c] = E[X] + c$$

$$E[cX] = cE[X]$$

Suppose X and Y are **two random variables**, then the expected value will be as follows:

$$E[X+Y] = E[X] + E[Y]$$

Now, having three constants (a, b, c) with two random variables (X, Y), the expected value can be calculated as follows:

$$E[aX + c] = aE[X] + c$$

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

Medians of Random Variables

Median is information about the "middle" value of the random variable $F(X) = 0.5$

Symmetric Random Variable

If a continuous random variable is symmetric about a point μ , then both the median and the expectation of the random variable are equal to μ

Evaluating Spread

Spread is the **measure of similarity or a varied set of observed values** for a particular set of data. The measures of spread include **range, quartiles, and interquartile range**.

Quantiles of Random variables

- The P th quantile of a random variable X
- A probability of P that the random variable takes a value less than the P th quantile
- **Upper quartile**
 - The 75th percentile of the distribution
- **Lower quartile**
 - The 25th percentile of the distribution
 - Interquartile range is the distance between the two quartiles

In order to find out how spread out (varied) a distribution is, it has to be measured by using specific formula for standard deviation. For finding out **standard deviation value**, firstly, we have to measure the variance by using this formula:

$$\text{Var}(X) = \sum (x - E[X])^2 P(X=x)$$

In the above formula:

X = Random variable

$\text{Var}(X)$ = Variance of X

Let the Variance of a random variable of a dice roll by using the above formula will be:

$$\text{Var}(X) = (1-3 \frac{1}{6})^2 (\frac{1}{12}) + (2-3 \frac{1}{6})^2 (\frac{1}{6}) + (3-3 \frac{1}{6})^2 (\frac{1}{3}) + (4-3 \frac{1}{6})^2 (\frac{1}{6}) + (5-3 \frac{1}{6})^2 (\frac{1}{12}) + (6-3 \frac{1}{6})^2 (\frac{1}{6})$$

$$\text{Var}(X) = 2.361$$

Standard deviation is the square root of the variance. When putting the value of variance in the formula of standard deviation, we get:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

$$\text{SD}(X) = \sqrt{2.361}$$

$$\text{SD}(X) = \sqrt{1.5366}$$

Properties of Variance

If there are three constants (a, b, c) with two random variables (X, Y), the properties of variance will be as follows:

$$\text{Var}(X + c) = \text{Var}(X)$$

$$\text{Var}(aX) = a^2 \text{Var}(X).$$

$$\text{Var}(X+Y) = \text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$$

$$\text{Var}(aX+ bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$$

Example: Taking the previous example of the roll of the dice from expected value

section, suppose the variance of the second roll of the dice is calculated as:

$$\text{Var (Y)} = 1.98$$

$$\text{Var (X)} = 2.361$$

Then the **variance** and **standard deviation** of dice game with score $(3X + 2Y + 1)$ will be as follows:

Variance:

$$\text{Var (3X - 2Y + 1)} = 9\text{Var (X)} + 4\text{Var (Y)}$$

$$= 9(2.361) + 4(1.98)$$

$$= 29.169$$

Standard deviation:

$$\text{SD (3X - 2y + 1)} = \sqrt{\text{Var (3X - 2y + 1)}}$$

$$= \sqrt{29.169}$$

$$= 5.401$$

Note on Continuous Random Variable

Continuous random variables model random phenomena. Medical students will not be tested or required to deal with the calculation of expected values and variances associated with random variables. It is part of the calculus, which is not relevant to the course of medical students.

Issues in Probability Models and Random Variable

Some of the issues to be encountered by the students while using probability models or dealing with random variables include:

The probability models are not always correct. The probability of data collection should be questioned in order to ensure accuracy.

If by mistake, a wrong or unsuitable probability model will be chosen during the research process, it will nullify the effect of the whole of the data collected. If a probability model is wrong, the outcomes related to it are wrong as well.

Dependent variables should always be taken into account with great consideration. Expected values of a random variable are required to be added. For the variance of expected values, the two random variables should always be independent; otherwise, the results will be wrong.

We only add variances of independent variables. The standard deviation of the same data should not be added.

Variances of independent variables will always be added, no matter if you are locating the difference between the two variables.

References

[Random Variables](#). via mathsisfun.com

[Mean and Variance of Linear Combinations](#). via onlinecourses.science.psu.edu

[Continuous Random Variables](#). via revisionmaths.com

[Probability Models](#). via stat.yale.edu

[Random Variables: Definition, Types & Examples](#). via study.com

Legal Note: Unless otherwise stated, all rights reserved by Lecturio GmbH. For further legal regulations see our [legal information page](#).

Notes