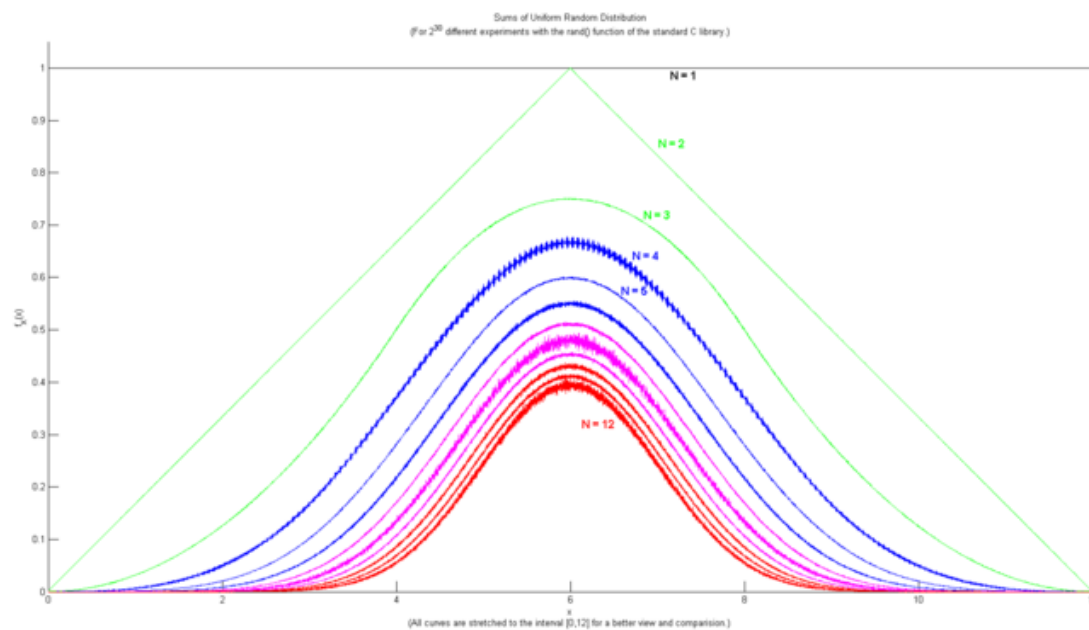


Statistics: Random Variables

[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 probability model is a **mathematical representation of a random phenomenon**. The model is comprised of a sample space, events that occur within that sample space, and the probability that an event will occur.

Within the model, random variables are numeric values that are generated based on the results of a random event. Random variables are the central point of the probability model.

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

Example: A random variable is any number we get from a random selection. Tossing a coin, we can get head (0) or tail (1). Either outcome 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, the possible outcomes are 1,2,3,4,5, or 6. One of these values is expected as a result of the event.

Types of Random Variables

There are **two main types** of random variables, which are **discrete** and **continuous random variables**.

Continuous random variable

This type of random variable has an **uncountable list of outcomes** associated with the event. These are random variables associated with measurements, like a soccer team's score, the height of a group of people, and student test results. It is the collection of all the possible values of probability and random variables that 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. The audience can't estimate or count the total time the advertisement will be on the air over 24 hours. It could be 10 minutes, 1 hour, or 10 hours.

Another example of a continuous random variable is the time a passenger has to wait at a bus stop for the bus to arrive. It could be 2 minutes, 10 minutes, or a few seconds. It is impossible to list all the possible times associated with how long it will take for the bus to arrive.

Discrete random variable

A discrete random variable 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 that any discrete random variable is an integer unless it is stated otherwise.

Example: A jar is filled with 10 red, 12 purple, and 5 blue marble pieces. The number of red marbles is countable, and the probability associated with it can be measured; 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 can be measured easily with related formulas.

Furthermore, even within these two classes of a random variable, there are further subcategories. Here is a shortlist 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 considered an indicator. If we define the event on which $X=1$, $A = \{\omega: X(\omega)=1\}$, then X is said to be an indicator of A .

Expected values

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

Once a probability model is established, the related expected values can be predicted. Finding the expected value for a continuous variable is difficult, so it should be restricted to discrete random variables. The random variable is denoted by X , so its expected value 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:

$$P(X=1) = P(X=6) = 1/12, P(X=2) = P(X=4) = P(X=5) = 1/6, \text{ and } P(X=3) = 1/3.$$

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

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

Properties of Expected Value

The properties of the expected value of a discrete random variable 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**. The expected value will be as follows:

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

With three constants (a , b , c) and 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

The 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
 - The interquartile range is the distance between the two quartiles

A distribution's spread (varied) has to be measured using a specific formula for standard deviation. To find out the **standard deviation value**, we have to measure the variance by using this formula:

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

X = Random variable

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

Using the above formula, the variance of a random variable of a dice roll will be:

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

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

The standard deviation is the square root of the variance. When putting the value of the variance into the standard deviation formula, 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 variance properties 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 the 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 dice game's **variance** and **standard deviation** 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 Variables

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

Issues in Probability Models and Random Variables

Probability models are not always correct. The probability of data collection should be questioned to ensure accuracy. If, by mistake, a wrong or unsuitable probability model is selected during the research process, then the outcomes based on the model are wrong as well.

Dependent variables should always be considered carefully. The expected values of a random variable must be added. For the variance of expected values, the two random variables should always be independent; otherwise, the results will be invalid.

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, regardless of whether 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