Unit 6: Common Probability Distributions

Contents

1.0 Introduction

In this section we look at the branch of statistics that deals with analysis of random events. Probability is the numerical assessment of likelihood on a scale from 0 (impossibility) to 1 (absolute certainty). Probability is usually expressed as the ratio between the number of ways an event can happen and the total number of things that can happen (e.g., there are 13 ways of picking a diamond from a deck of 52 cards, so the probability of picking a diamond is 13/52, or ¼). Probability theory grew out of attempts to understand card games and gambling. As science became more rigorous, analogies between certain biological, physical, and social phenomena and games of chance became more evident (e.g., the sexes of newborn infants follow sequences similar to those of coin tosses). As a result, probability became a fundamental tool of modern genetics and many other disciplines.

2.0 Intended Learning Outcomes (ILOs)

By the end of this unit, the reader should be able to:

- Explain the role of probability distribution functions in simulations
- Describe Probability theory
- Explain the fundamental concepts of Probability theory
- Explain Random Variable
- Explain Limiting theorems
- Describe Probability distributions in simulations

- List common Probability distributions.

3.0 Main Content

3.1 Distribution Functions and Simulation

Many simulation tools and approaches are ***deterministic***. In a deterministic simulation, the input parameters for a model are represented using single values (which typically are described either as "the best guess" or "worst case" values). Unfortunately, this kind of simulation, while it may provide some insight into the underlying mechanisms, is not well-suited to making predictions to support decision-making, as it cannot quantitatively address the risks and uncertainties that are inherently present.

However, it is possible to quantitatively represent uncertainties in simulations. ***Probabilistic simulation*** is the process of explicitly representing these uncertainties by specifying inputs as probability distributions. If the inputs describing a system are uncertain, the prediction of future performance is necessarily uncertain. That is, the result of any analysis based on inputs represented by probability distributions is itself a probability distribution. Hence, whereas the result of a deterministic simulation of an uncertain system is a *qualified statement* ("if we build the dam, the salmon population could go extinct"), the result of a probabilistic simulation of such a system is a *quantified probability* ("if we build the dam, there is a 20% chance that the salmon population will go extinct"). Such a result (in this case, quantifying the risk of extinction) is typically much more useful to decision-makers who might utilize the simulation results.

3.2 Probability Definitions

The word *probability* does not have a consistent direct definition. In fact, there are two broad categories of **probability interpretations**, whose adherents possess different (and sometimes conflicting) views about the fundamental nature of probability:

1. Frequentists talk about probabilities only when dealing with experiments that are random and well-defined. The probability of a random event denotes the *relative frequency of occurrence* of an experiment's outcome, when repeating the experiment. Frequentists consider probability to be the relative frequency "in the long run" of outcomes.
2. Bayesians, however, assign probabilities to any statement whatsoever, even when no random process is involved. Probability, for a Bayesian, is a way to represent an individual's *degree of belief* in a statement, or an objective degree of rational belief, given the evidence

   The scientific study of probability is a modern development. Gambling shows that there has been an interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions of use in those problems only arose much later.

Probability Distribution

A probability distribution gathers together all possible outcomes of a random variable (i.e. any quantity for which more than one value is possible), and summarizes these outcomes by indicating the probability of each of them. While a probability distribution is often associated with the bell-shaped curve, recognize that such a curve is only indicative of one specific type of probability, the so-called normal probability distribution. However, in real life, a probability distribution can take any shape, size and form.

Example: Probability Distribution

For example, if we wanted to choose a day at random in the future to schedule an event, and we wanted to know the probability that this day would fall on a Sunday, as we will need to avoid scheduling it on a Sunday. With seven days in a week, the probability that a random day would happen to be a Sunday would be given by one-seventh or about 14.29%. Of course, the same 14.29% probability would be true for any of the other six days.

In this case, we would have a uniform probability distribution: the chances that our random day would fall on any particular day are the same, and the graph of our probability distribution would be a straight line.

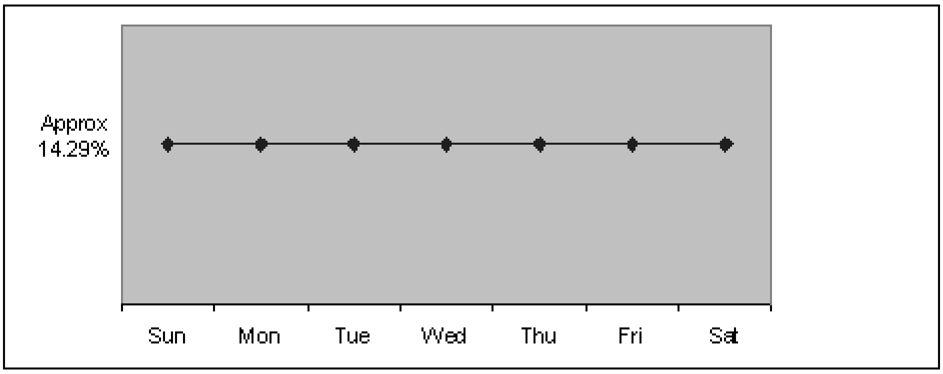

fig.1: Uniform Probability Distribution

Probability distributions can be simple to understand as in this example, or they can be very complex and require sophisticated techniques (e.g., option pricing models, Monte Carlo simulations) to help describe all possible outcomes.

3.3 Random Variables

Random variable is **discrete** random variables if it can take on a finite or countable number of possible outcomes. The previous example asking for a day of the week is an example of a discrete variable, since it can only take seven possible values. Monetary variables expressed in dollars and cents are always discrete, since money is rounded to the nearest $0.01.

A random variable is **continuous** random variable if it has has infinite possible outcomes. A rate of return (e.g. growth rate) is continuous:

- a stock can grow by 9% next year or by 10%, and in between this range it could grow by 9.3%, 9.4%, 9.5%
- Clearly there is no end to how precise the outcomes could be broken down; thus it's described as a continuous variable.

Outcomes in Discrete vs. Continuous Variables

The rule of thumb is that a discrete variable can have all possibilities listed out, while a continuous variable must be expressed in terms of its upper and lower limits, and greater-than or less-than indicators. Of course, listing out a large set of possible outcomes (which is usually the case for money variables) is usually impractical – thus money variables will usually have outcomes expressed as if they were continuous.

Examples:

- Rates of return can theoretically range from –100% to positive infinity.
- Time is bound on the lower side by 0.
- Market price of a security will also have a lower limit of $0, while its upper limit will depend on the security – stocks have no upper limit (thus a stock price's outcome $\geq$ $0),
- Bond prices are more complicated, bound by factors such as time-to-maturity and embedded call options. If a face value of a bond is $1,000, there's an upper limit (somewhere above $1,000) above which the price of the bond will not go, but pinpointing the upper value of that set is imprecise.

3.4 Probability Function

A probability function gives the probabilities that a random variable will take on a given list of specific values. For a discrete variable, if (x_1, x_2, x_3, x_4 …) are the complete set of possible outcomes, p(x) indicates the chances that X will be equal to x. Each x in the list for a discrete variable will have a p(x). For a continuous variable, a probability function is expressed as f(x).

The two key properties of a probability function, p(x) (or f(x) for continuous), are the following:

1. $0 \leq p(x) \leq 1$, since probability must always be between 0 and 1.
2. Add up all probabilities of all distinct possible outcomes of a random variable, and the sum must equal 1.

Determining whether a function satisfies the first property should be easy to spot since we know that probabilities always lie between 0 and 1. In other words, p(x) could never be 1.4 or –0.2. To illustrate the second property, say we are given a set of three possibilities for X: (1, 2, 3) and a set of three for Y: (6, 7, 8), and given the probability functions f(x) and g(y).

x	f(x)	y	g(y)
1	0.31	6	0.32
2	0.43	7	0.40
3	0.26	8	0.23

For all possibilities of f(x), the sum is 0.31+0.43+0.26=1, so we know it is a valid probability function. For all possibilities of g(y), the sum is 0.32+0.40+0.23 = 0.95, which violates our second principle. Either the given probabilities for g(y) are wrong, or there is a fourth possibility for y where g(y) = 0.05. Either way it needs to sum to 1.

Probability Density Function

A probability density function (or pdf) describes a probability function in the case of a continuous random variable. Also known as simply the "density", a probability density function is denoted by "f(x)". Since a pdf refers to a continuous random variable, its probabilities would be expressed as ranges of variables rather than probabilities assigned to individual values as is done for a discrete variable. For example, if a stock has a 20% chance of a negative return, the pdf in its simplest terms could be expressed as:

x	f(x)
< 0	0.2
≥ 0	0.8

3.5 Mathematical Treatment of Probability

In mathematics, a probability of an event *A* is represented by a real number in the range from 0 to 1 and written as P(*A*), p(*A*) or Pr(*A*). An impossible event has a probability of 0, and a certain event has a probability of 1. However, the converses are not always true: probability 0 events are not always impossible, nor probability 1 events certain. The rather subtle distinction between "certain" and "probability 1" is treated at greater length in the article on "almost surely".

The *opposite* or *complement* of an event *A* is the event [not *A*] (that is, the event of *A* not occurring); its probability is given by P(not *A*) = 1 - P(*A*). As an example, the chance of not rolling a six on a six-sided die is 1 – (chance of rolling a six) $1 - \frac{1}{6} = \frac{5}{6}$. .

Joint Probabilty

If both the events *A* and *B* occur on a single performance of an experiment this is called the ***intersection or joint probability*** of *A* and *B*, denoted as and $P(A \cap B)$.

If two events, *A* and *B* are ***independent*** then the joint probability is:

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B),$$

for example, if two coins are flipped the chance of both being heads is: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Mutually Exclusive Events

If either event *A* or event *B* or both events occur on a single performance of an experiment this is called the union of the events *A* and *B* denoted as $P(A \cup B)$. If two events are ***mutually exclusive*** then the probability of either occurring is:

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - 0 = P(A) + P(B)$$

For example, the chance of rolling a 1 or 2 on a six-sided die is

$$P(1 \text{ or } 2) = P(1) + P(2) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{1}{3}.$$

If the events are not mutually exclusive then

$$P\left(A \text{ or } B\right) = P(A \cup B) = P\left(A\right) + P\left(B\right) - P\left(A \text{ and } B\right).$$

For example, when drawing a single card at random from a regular deck of cards, the chance of getting a heart or a face card (J,Q,K) (or one that is both) is $\frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{11}{26}$, because of the 52 cards of a deck 13 are hearts, 12 are face cards, and 3 are both: here the possibilities included in the "3 that are both" are included in each of the "13 hearts" and the "12 face cards" but should only be counted once.

Conditional Probability

This is the probability of some event *A*, given the occurrence of some other event *B*. Conditional probability is written $P(A|B)$, and is read "the probability of *A*, given *B*". It is defined by:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

If $P(B) = 0$ then $P(A \mid B)$ is undefined.

Summary of probabilities

Event	Probability
A	P(A) □ [0,1]
Not A	P(A) = 1-P(A)
A or B	P(A U B) = P(A) + P(B)-P(A∩B) = P(A)+P(B) If A and B are mutually exclusive

A and B	$P(A \cap B) = P(A\|B)P(B)$ $= P(A)P(B)$ If A and B are independent
A given B	$P(A\|B) = P(A \cap B)/P(B)$

Two or more events are mutually exclusive if the occurrence of any one of them excludes the occurrence of the others.

3.6 Probability theory

Like other theories, the theory of probability is a representation of probabilistic concepts in formal terms—that is, in terms that can be considered separately from their meaning. These formal terms are manipulated by the rules of mathematics and logic, and any results are then interpreted or translated back into the problem domain.

There have been at least two successful attempts to formalize probability, namely the Kolmogorov formulation and the Cox formulation. In Kolmogorov's formulation sets are interpreted as events and probability itself as a measure on a class of sets. In Cox's theorem, probability is taken as a primitive (that is, not further analyzed) and the emphasis is on constructing a consistent assignment of probability values to propositions. In both cases, the laws of probability are the same, except for technical details.

Probability theory is a mathematical science that permits one to find, using the probabilities of some random events, the probabilities of other random events connected in some way with the first.

The assertion that a certain event occurs with a probability equal, for example, to 1/2, is still not, in itself, of ultimate value, because we are striving for definite knowledge. Of definitive, cognitive value are those results of probability theory that allow us to state that the probability of occurrence of some event A is very close to 1 or (which is the same thing) that the probability of the non-occurrence of event A is very small. According to the principle of “disregarding sufficiently small probabilities,” such an event is considered practically reliable. Such conclusions, which are of scientific and practical interest, are usually based on the assumption that the occurrence or non-occurrence of event A depends on a large number of factors that are slightly connected with each other.

Consequently, it can also be said that **probability theory** is a mathematical science that clarifies the regularities that arise in the interaction of a large number of random factors.

To describe the regular connection between certain conditions S and event A, whose occurrence or non-occurrence under given conditions can be accurately established, natural science usually uses one of the following schemes:

(a) For each realization of conditions S, event A occurs. All the laws of classical mechanics have such a form, stating that for specified initial conditions and forces acting on an object or system of objects, the motion will proceed in an unambiguously definite manner.

(b) Under conditions S, event A has a definite probability $P(A/S)$ equal to p.

> Thus, for example, the laws of radioactive emission assert that for each radioactive substance there exists the specific probability that, for a given amount of a substance, a certain number of atoms N will decompose within a given time interval.

Let us call the frequency of event A in a given set of n trials (that is, of n repeated realizations of conditions S) the ratio $h = m/n$ of the number m of those trials in which A occurs to the total number of trials n. The existence of a specific probability equal to p for an event A under conditions S is manifested in the fact that in almost every sufficiently long series of trials, the frequency of event A is approximately equal to p.

Statistical laws, that is, laws described by a scheme of type (b), were first discovered in games of chance similar to dice. The statistical rules of birth and death (for example, the probability of the birth of a boy is 0.515) have also been known for a long time. A great number of statistical laws in physics, chemistry, biology, and other sciences were discovered at the end of the 19th and in the first half of the 20th century.

The possibility of applying the methods of probability theory to the investigation of statistical laws, which pertain to a very wide range of scientific fields, is based on the fact that the probabilities of events always satisfy certain simple relationships, which will be discussed in the next section. The investigation of the properties of probabilities of events on the basis of these simple relationships is also a topic of probability theory.

3.6.1 Fundamental concepts of Probability theory.

The fundamental concepts of probability theory as a mathematical discipline are most simply defined in the framework of so-called elementary probability theory. Each trial T considered in elementary probability theory is such that it is ended by one and only one of the events E_1, E_2, … , E_s (by one or another, depending on the case). These events are called outcomes of the trial. Each outcome E_k is connected with a positive number p_k, the probability of this outcome. The numbers p_k must add up to 1. Events A, which consist of the fact that "either E_i, or E_j … , or E_k occurs," are then considered. The outcomes E_i, … , E_k are said to be favorable to A, and according to the definition, it is assumed that the probability $P(A)$ of event A is equal to the sum of the probabilities of the outcomes favorable to it:

(1) $$P(A) = p_i + p_j + \dots + p_k$$

The particular case $p_1 = p_2 = p_s = 1/s$ leads to the formula:

(2) $$P(A) = r/s$$

Formula (2) expresses the so-called classical definition of probability according to which the probability of some event A is equal to the ratio of the number r of outcomes favorable to A to the number s of all "equally likely" outcomes. The classical definition of probability only reduces the concept of probability to the concept of equal possibility, which remains without a clear definition.

EXAMPLE. In the tossing of two dice, each of the 36 possible outcomes can be designated by *(i, j)*, where *i* is the number of pips that comes up on the first dice and *j*, the number on the second. The outcomes are assumed to be equally likely. To the event *A*, “the sum of the pips is 4,” three outcomes are favorable: (1,3); (2,2); (3,1). Consequently, $P(A) = 3/36 = 1/12$.

Starting from certain given events, it is possible to define two new events: their union (sum) and intersection (product). Event *B* is called the **union** of events A_1, A_2, …, A_r if it has the form “A_1 or A_2, …, or A_r occurs.”

Event *C* is called the **intersection** of events A_1, A_2 …, A_r if it has the form “A_1, and A_2, … , and A_r occurs.”
The union of events is designated by the symbol ∪, and the intersection, by ∩. Thus, we write:

$B = A_1 \cup A_2 \cup \ldots \cup A_r$ $C = A_1 \cap A_2 \cap \ldots \cap A_r$
Events *A* and *B* are called disjoint if their simultaneous occurrence is impossible—that is, if among the outcomes of a trial not one is favourable to *A* and *B* simultaneously.
Two of the basic theorems of probability theory are connected with the operations of union and intersection of events; these are the theorems of addition and multiplication of probabilities.

3.6.2 Theorem of Addition of Probabilities.

If events A_1, A_2, …, A_r are such that each two of them are disjoint, then the probability of their union is equal to the sum of their probabilities.

Thus, in the example presented above of tossing two dice, event *B*, “the sum of the pips does not exceed 4,” is the union of three disjoint events A_2, A_3, A_4, consisting of the fact the sum of the pips is equal to 2, 3, and 4, respectively. The probabilities of these events are 1/36, 2/36, and 3/36, respectively. According to the theorem of addition of probabilities, probability *P(B)* is:

$1/36 + 2/36 + 3/36 = 6/36 = 1/6$

The conditional probability of event *B* under condition *A* is determined by the formula:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

which, as can be proved, corresponds completely with the properties of frequencies.
Events A_1, A_2, …, A_r are said to be independent if the conditional probability of each of them, under the condition that some of the remaining events have occurred, is equal to its

"absolute" probability.

3.6.3 Theorem of Multiplication of Probabilities.

The probability of the intersection of events A_1, A_2, …, A_r is equal to the probability of event A_1 multiplied by the probability of event A_2 under the condition that A_1 has occurred, …, multiplied by the probability of A_r under the condition that A_1, A_2, …, A_{r-1} have occurred. For independent events, the multiplication theorem reduces to the formula:

$$P(A_1 \cap A_2 \cap \ldots \cap A_r) = P(A_1) \times P(A_2) \times \ldots \times P(A_r) \ldots\ldots\ldots\ldots (3)$$

that is, the probability of the intersection of independent events is equal to the product of the probabilities of these events. Formula (3) remains correct, if on both sides some of the events are replaced by their inverses.

EXAMPLE:

Four shots are fired at a target, and the hit probability is 0.2 for each shot. The target hits by different shots are assumed to be independent events. What is the probability of hitting the target three times?

Each outcome of the trial can be designated by a sequence of four letters [for example, (s, f, f, s) denotes that the first and fourth shots hit the target (success), and the second and third miss (failure)]. There are 2 · 2 · 2 · 2 = 16 outcomes in all. In accordance with the assumption of independence of the results of individual shots, one should use formula (3) and the remarks about it to determine the probabilities of these outcomes. Thus, the probability of the outcome (s, f, f, f) is set equal to 0.2 x 0.8 x 0.8 x 0.8 = 0.1024; here,
0.8 = 1 - 0.2 is the probability of a miss for a single shot. For the event "three shots hit the target," the possible outcomes are: (s, s, s, f), (s, s, f, s), (s, f, s, s), and (f, s, s, s) are favorable and the probability of each is the same:
0.2 · 0.2 · 0.2 · 0.8 = · · · = 0.8 · 0.2 · 0.2 · 0.2 = 0.0064
Consequently, the desired probability is 4 x 0.0064 = 0.0256.
Generalizing the discussion of the given example, it is possible to derive one of the fundamental formulas of probability theory: if events A_1, A_2, …, A_n are independent and each has a probability p, then the probability of exactly m such events occurring is:

$$P_n(m) = C_n^{\ m}(1- p)^{n- m} \ldots\ldots\ldots\ldots (4)$$

Here, $C^m{}_n$ denotes the number of combinations of n elements taken m at a time. For large n, the calculation using formula (4) becomes difficult. In the preceding example, let the number of shots equal 100; the problem then becomes one of finding the probability x that the number of hits lies in the range from 8 to 32. The use of formula (4) and the addition theorem gives an accurate, but not a practically useful, expression for the desired probability

$$x = \sum_{m=8}^{32} C_{100}^{m}(0.2)^{m}(0.8)^{100-m}$$

The approximate value of the probability x can be found by the **Laplace theorem**

$$x \approx \frac{1}{\sqrt{2\pi}} \int_{-3}^{+3} e^{-z^2/2}\, dz = 0.9973$$

with the error not exceeding 0.0009. The obtained result demonstrates that the event $8 \leq m \leq 32$ is practically certain. This is the simplest, but a typical, example of the use of the limit theorems of probability theory.

Another fundamental formula of elementary probability theory is the so-called total probability formula: if events $A_1, A_2, \ldots, A_r$ are disjoint in pairs and their union is a certain event, then the probability of any event B is the sum

$$P(B) = \sum_{k=1}^{r} P(B/A_k)P(A_k)$$

The theorem of multiplication of probabilities turns out to be particularly useful in the consideration of compound trials. Let us say that trial T consists of trials $T_1, T_2, \ldots, T_{n-1}, T_n$, if each outcome of trial T is the intersection of certain outcomes $A_i, B_i, \ldots, x_k, Y_l$ of the corresponding trials $T_1, T_2, \ldots, T_{n-1}, T_n$. From one or another consideration, the following probabilities are often known:

$\boldsymbol{P(A_1), P(B_j/A_i), \ldots, P(Y_l/A_i \cap B_j \cap \ldots \cap X_k)}$**..(5)**

According to the probabilities of (5), probabilities $P(E)$ for all the outcomes of E of the compound trial and, in addition, the probabilities of all events connected with this trial can be determined using the multiplication theorem (just as was done in the example above).

Two types of compound trials are the most significant from a practical point of view:

(a) the component trials are independent, that is, the probabilities (5) are equal to the unconditional probabilities $P(A_i), P(B_j), \ldots, P(Y_l)$; and
(b) the results of only the directly preceding trial have any effect on the probabilities of the outcomes of any trial—that is, the probabilities (5) are equal, respectively, to $P(A_i), P(B_j/A_i), \ldots, P(Y_l/X_k)$.

In this case, it is said that the trials are connected in a Markov chain. The probabilities of all the events connected with the compound trial are completely determined here by the initial probabilities $P(A_i)$ and the transition probabilities $P(B_j/A_i), \ldots, P(Y_l/X_k)$.
Often, instead of the complete specification of a probability distribution of a random variable, it is preferable to use a small number of numerical characteristics. The most frequently used are the mathematical expectation and the dispersion.

In addition to mathematical expectations and dispersions of these variables, a joint distribution of several random variables is characterized by correlation coefficients and so forth. The meaning of the listed characteristics is to a large extent explained by the **limit theorems**

3.7 The Limit theorems.

In the formal presentation of probability theory, limit theorems appear in the form of a superstructure over its elementary sections, in which all problems have a finite, purely arithmetic character. However, the cognitive value of probability theory is revealed only by the limit theorems. Thus, the **Bernoulli theorem** proves that in independent trials, the frequency of appearance of any event, as a rule, deviates little from its probability, and the **Laplace theorem** indicates the probabilities of one or another deviation. Similarly, the meaning of such characteristics of a random variable as its mathematical expectation and dispersion is explained by the law of large numbers and the **central limit theorem**.

Let $\boldsymbol{X_1, X_2, \ldots, X_n}$, be independent random variables that have one and the same probability distribution with $\mathbf{E}X_K = a$, $\mathbf{D}X_K = \sigma^2$ and Y_n be the arithmetic mean of the first n variables of sequence such that:

$Y_n = (X_1 + X_2 + X_2 + \cdot\cdot\cdot + X_n)/n$

In accordance with the law of large numbers, for any $\varepsilon > 0$, the probability of the inequality $| Y_n - a | \leq \varepsilon$ has the limit 1 as $n \to \infty$, and thus Y_n, as a rule, differs little from a.

The **central limit theorem** makes this result specific by demonstrating that the deviations of Y_n from a are approximately subordinate to a normal distribution with mean zero and dispersion $8^2/n$. Thus, to determine the probabilities of one or another deviation of Y_n from a for large n, there is no need to know all the details about the distribution of the variables X_n; it is sufficient to know only their dispersion.

In the 1920's it was discovered that even in the scheme of a sequence of identically distributed and independent random variables, limiting distributions that differ from the normal can arise in a completely natural manner. Thus, for example, if X_1 is the time until the first reversion of some randomly varying system to the original state, and X_2 is the time between the first and second reversions, and so on, then under very general conditions the distribution of the sum $X_1 + \cdot\cdot\cdot + X_n$ (that is, of the time until the Aith reversion), after multiplication by $n^{-1/\alpha}$(α is a constant less than 1), converges to some limiting distribution. Thus, the time until the *nth* reversion increases, roughly speaking, as $n^{1/\alpha}$, that is, more rapidly than n (in the case of applicability of the law of large numbers, it is of the order of n).

The mechanism of the emergence of the majority of limiting regularities can be understood ultimately only in connection with the theory of random processes.

Random processes.

In a number of physical and chemical investigations of recent decades, the need has arisen to consider, in addition to one-dimensional and multidimensional random variables, random processes—that is, processes for which the probability of one or another of their courses is defined. In probability theory, a random process is usually considered as a one-parameter family of random variables *X(t)*. In an overwhelming number of applications, the parameter *t* represents time, but this parameter can be, for example, a point in space, and then we usually speak of a random function. In the case when the parameter t runs through the integer-valued numbers, the random function is called a **random sequence**. Just as a random variable is characterized by a distribution law, a random process can be characterized by a set of joint distribution laws for $X(t_1)$, $X(t_2)$, …, $X(t_n)$ for all possible moments of $t_1, t_2, \ldots, t_n$ for any $n > 0$.

3.8 Probability Distribution Functions

In probability theory and statistics, a **probability distribution** identifies either the probability of each value of a random variable (when the variable is discrete), or the probability of the value falling within a particular interval (when the variable is continuous). The probability distribution describes the range of possible values that a random variable can attain and the probability that the value of the random variable is within any (measurable) subset of that range.

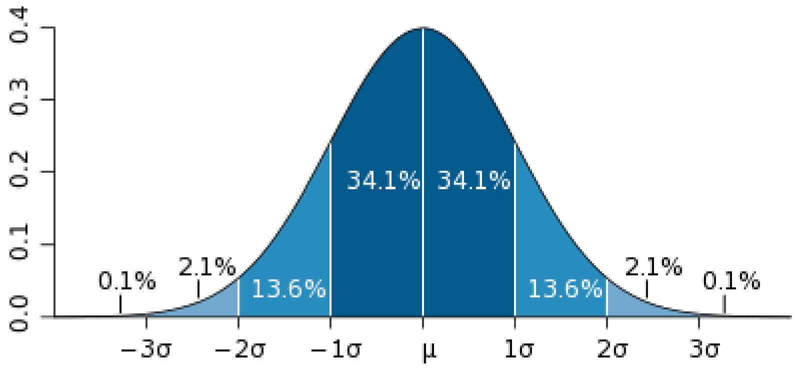

Figure 2: The Normal distribution, often called the "bell curve".

When the random variable takes values in the set of real numbers, the probability distribution is completely described by the *cumulative distribution function*, whose value at each real x is the probability that the random variable is smaller than or equal to x.

The concept of the probability distribution and the random variables which they describe underlies the mathematical discipline of probability theory, and the science of statistics. There is spread or variability in almost any value that can be measured in a population (e.g. height of people, durability of a metal, sales growth, traffic flow, etc.); almost all measurements are made with some intrinsic error; also in physics many processes are described probabilistically, from the kinetic properties of gases to the quantum mechanical description of fundamental particles. For these and many other reasons, simple numbers are often inadequate for describing a quantity, while probability distributions are often more appropriate.

3.8.1 Probability distributions of real-valued random variables

Because a probability distribution Pr on the real line is determined by the probability of a real-valued random variable X being in a half-open interval $(-\infty, x]$, the probability distribution is completely characterized by its cumulative distribution function given as:

$$F(x) = \Pr[X \le x] \qquad \forall x \in \mathbb{R}.$$

a. Discrete probability distribution

A probability distribution is called ***discrete*** if its cumulative distribution function only increases in jumps. More precisely, a probability distribution is discrete if there is a finite or countable set whose probability is 1.

For many familiar discrete distributions, the set of possible values is discrete in the sense that all its points are isolated points. But, there are discrete distributions for which this countable set is dense on the real line.

Discrete distributions are characterized by a probability mass function, *p* such that:

$$\Pr[X = x] = p(x).$$

b. Continuous probability distribution

By one convention, a probability distribution μ is called ***continuous*** if its cumulative distribution function $F(x) = \mu(-\infty, x]$ is continuous and, therefore, the probability measure of singletons $\mu\{x\} = 0$.

Another convention reserves the term ***continuous probability distribution*** for absolutely continuous distributions. These distributions can be characterized by a probability density function: a non-negative Lebesgue integrable function *f* defined on the real numbers such that

$$F(x) = \mu(-\infty, x] = \int_{-\infty}^{x} f(t)\,dt.$$

Discrete distributions and some continuous distributions do not admit such a density.

Terminologies

The **support** of a distribution is the smallest closed interval/set whose complement has probability zero. It may be understood as the points or elements that are actual members of the distribution.

A **discrete random variable** is a random variable whose probability distribution is discrete. Similarly, a **continuous random variable** is a random variable whose probability distribution is continuous.

Some properties

- The probability density function of the sum of two independent random variables is the **convolution** of each of their density functions.
- The probability density function of the difference of two independent random variables is the **cross-correlation** of their density functions.
- Probability distributions are not a vector space – they are not closed under linear

combinations, as these do not preserve non-negativity or total integral 1 – but they are closed under convex combination, thus forming a convex subset of the space of functions (or measures).

In mathematics and, in particular, functional analysis, **convolution** is a mathematical operation on two functions *f* and *g*, producing a third function that is typically viewed as a modified version of one of the original functions. Convolution is similar to cross-correlation. It has applications that include statistics, computer vision, image and signal processing, electrical engineering, and differential equations

3.9 Summary of Common Probability Distributions

The following is a list of some of the most common probability distributions, grouped by the type of process that they are related to.

Note that all of the univariate distributions below are singly-peaked; that is, it is assumed that the values cluster around a single point. In practice, actually-observed quantities may cluster around multiple values. Such quantities can be modeled using a mixture distribution.

3.9.1 Related to real-valued quantities that grow linearly (e.g. errors, offsets)

- Normal distribution (aka Gaussian distribution), for a single such quantity; the most common continuous distribution;
- Multivariate normal distribution (aka multivariate Gaussian distribution), for vectors of correlated outcomes that are individually Gaussian-distributed;

3.9.2 Related to positive real-valued quantities that grow exponentially (e.g. prices, incomes, populations)

- Log-normal distribution, for a single such quantity whose log is normally distributed
- Pareto distribution, for a single such quantity whose log is exponentially distributed; the prototypical power law distribution.

3.9.3 Related to real-valued quantities that are assumed to be uniformly distributed over a (possibly unknown) region

- Discrete uniform distribution, for a finite set of values (e.g. the outcome of a fair die)
- Continuous uniform distribution, for continuously-distributed values.

3.9.4 Related to Bernoulli trials (yes/no events, with a given probability)

- *Bernoulli* distribution, for the outcome of a single Bernoulli trial (e.g. success/failure, yes/no);
- *Binomial* distribution, for the number of "positive occurrences" (e.g. successes, yes votes, etc.) given a fixed total number of independent occurrences;
- *Negative binomial* distribution, for binomial-type observations but where the quantity

of interest is the number of failures before a given number of successes occurs’

- *Geometric* distribution, for binomial-type observations but where the quantity of interest is the number of failures before the first success; a special case of the negative binomial distribution.

3.9.5 Related to sampling schemes over a finite population

- *Binomial* distribution, for the number of "positive occurrences" (e.g. successes, yes votes, etc.) given a fixed number of total occurrences, using sampling with replacement
- *Hypergeometric* distribution, for the number of "positive occurrences" (e.g. successes, yes votes, etc.) given a fixed number of total occurrences, using sampling without replacement
- *Beta-binomial* distribution, for the number of "positive occurrences" (e.g. successes, yes votes, etc.) given a fixed number of total occurrences, sampling using a Polya urn scheme (in some sense, the "opposite" of sampling without replacement)

3.9.6 Related to categorical outcomes (events with *K* possible outcomes, with a given probability for each outcome)

- *Categorical* distribution, for a single categorical outcome (e.g. yes/no/maybe in a survey); a generalization of the Bernoulli distribution;
- *Multinomial* distribution, for the number of each type of categorical outcome, given a fixed number of total outcomes; a generalization of the binomial distribution;
- *Multivariate* hyper geometric distribution, similar to the multinomial distribution, but using sampling without replacement; a generalization of the hyper geometric distribution;

3.9.7 Related to events in a Poisson process (events that occur independently with a given rate)

- *Poisson* distribution, for the number of occurrences of a Poisson-type event in a given period of time
- *Exponential* distribution, for the time before the next Poisson-type event occurs

3.9.8 Useful for hypothesis testing related to normally-distributed outcomes

- *Chi-square* distribution, the distribution of a sum of squared standard normal variables; useful e.g. for inference regarding the sample variance of normally-distributed samples
- *Student's* t distribution, the distribution of the ratio of a standard normal variable and the square root of a scaled chi squared variable; useful for inference regarding the mean of normally-distributed samples with unknown variance
- *F-distribution*, the distribution of the ratio of two scaled chi squared variables; useful e.g. for inferences that involve comparing variances or involving R-squared (the squared correlation coefficient)

3.9.9 Useful as conjugate prior distributions in Bayesian inference

- *Beta distribution*, for a single probability (real number between 0 and 1); conjugate to the Bernoulli distribution and binomial distribution
- *Gamma distribution*, for a non-negative scaling parameter; conjugate to the rate parameter of a Poisson distribution or exponential distribution, the precision (inverse variance) of a normal distribution, etc.
- *Dirichlet distribution*, for a vector of probabilities that must sum to 1; conjugate to the categorical distribution and multinomial distribution; generalization of the beta distribution
- *Wishart distribution*, for a symmetric non-negative definite matrix; conjugate to the inverse of the covariance matrix of a multivariate normal distribution; generalzation of the gamma distribution

4.0 Self-Assessment Exercise(s)

Answer the following questions:

1. Define probability distribution
2. What is the relationship between a random variable and probability distribution
3. List the distributions related to:
    a. Bernoulli trials
    b. Categorical outcomes
    c. Hypothesis testing
4. The student is expected to familiarize him/herself with these probability distributions and their applications.

5.0 Conclusion

The basis of simulation is randomness. Here we have discussed this fundamental basis which offers us the possibility to quantitatively represent uncertainties in simulations. With **Probabilities** in simulation we can explicitly represent uncertainties by specifying inputs as probability distributions.

6.0 Summary

In this unit we discussed the following:

- Defined Probability as
- Discussed the fundamental concepts of probability theory
- The limit theorem
- Random variables and Random processes
- Probability distributions

- Provided a listing of common probability distributions grouped by their related processes

7.0 Further Readings

- Devore, J. L. (2018). *Probability and statistics for engineering and the sciences*. Toronto, Ontario: Nelson.
- Georgii, H. (2013). *Stochastics: Introduction to probability and statistics*. Berlin: De Gruyter.
- Giri, N. C. (2019). *Introduction to probability and statistics*. London: Routledge.
- Johnson, R. A., Miller, I., & Freund, J. E. (2019). *Miller & Freunds probability and statistics for engineers*. Boston: Pearson Education.
- Laha, R. G., & Rohatgi, V. K. (2020). *Probability theory*. Mineola, NY: Dover Publications.
- Mathai, A. M., & Haubold, H. J. (2018). *Probability and statistics: A course for physicists and engineers*. Boston: De Gruyter.
- Pishro-Nik, H. (2014). *Introduction to probability, statistics, and random processes*. Blue Bell, PA: Kappa Research, LLC.
- Spiegel, M. R., Schiller, J. J., & Srinivasan, R. A. (2013). *Schaums outline of probability and statistics*. New York: McGraw-Hill.