1. Introduction

Welcome to the series of E-learning modules on the concept of convergence in distribution and in probability. In this module, we are going to cover the concept of convergence in distribution and in probability, definition, properties, applications and relationship between convergence in distribution and convergence in probability.

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

- Explain the convergence in distribution
- Explain the examples and properties of convergence in distribution
- Explain convergence in probability
- Explain the examples and properties of convergence in probability
- Explain the relationship between convergence in distribution and convergence in probability

In probability theory, there exist several different notions of convergence of random variables. An important concept in probability theory is the convergence of sequences of random variables to some limit random variable. It has a vital role to play in statistics and in stochastic processes.

The same concepts are known in more general mathematics as stochastic convergence. They formalize the idea that a sequence of essentially random or unpredictable events can sometimes be expected to settle into a behaviour that is essentially unchanging when items far enough into the sequence are studied.

The different possible notions of convergence relate to how such behaviour can be characterised. Two readily understood behaviours are that the sequence eventually takes a constant value and that value in the sequence continues to change. However, it can be described by an unchanging probability distribution.

2. Convergence in Distribution

Definition
A sequence {X1, X2, etc} of random variables is said to converge in distribution, or converge weakly, or converge in law to a random variable X if , limit of Fn of x as n tends to infinity is equal to F of x
For every number x in R at which F is continuous. Here Fn and F are the cumulative distribution functions of random variables Xn and X correspondingly.

The requirement that only the continuity points of *F* should be considered is essential. An example given below explains why we require the distribution functions that converge only at continuity points for the limiting distribution function.

For example:
If Xn are distributed uniformly in intervals [0, 1 by n], then this sequence converges in distribution to a degenerate random variable X is equal to 0. Indeed, Fn of x is equal to 0 for all n when x less than or equal to 0, and Fn(x) is equal to 1 for all x greater than or equal to 1 by n when n is greater than zero .

However, for this limiting random variable
F of zero is equal to 1, even though Fn of zero is equal to zero for all n. Thus, the convergence of cumulative distribution functions fails at the point x equal to 0, where F is discontinuous.
Convergence in distribution may be denoted as
Xn tends to X with small letter d or Xn tends to X with capital letter D or Xn tends to X with capital letter L
For example, if X is standard normal we can write, Xn tends to Normal with mean zero and variance 1 in distribution
It should be clear what we mean by Xn tends to X in distribution

The random variables Xn converge in distribution to a random variable X having distribution function F. Similarly, we have Fn tends to F in distribution if there is a sequence of random variables {Xn}, where Xn has distribution function Fn, and a random variable X having distribution function F, so that
Xn tends to X in distribution

The first obvious fact to notice is that convergence in distribution only involves the distributions of the random variables. Thus, the random variables need not even be defined on the same probability space (that is, they need not be defined for the same random experiment). This is in sharp contrast to the other modes of convergence:

- Convergence with probability 1
- Convergence in probability
- Convergence in kth mean

However, we cannot deny the fact that convergence in distribution is the weakest of all of these modes of convergence. It is nonetheless very important. The central limit theorem is one of the two fundamental theorems of probability, which is a theorem about convergence in distribution.

The definition of convergence in distribution may be extended from random vectors to more complex random elements in arbitrary metric spaces, and even to the “random variables”, which are not measurable - a situation which occurs for example in the study of empirical processes. This is the “weak convergence of laws without laws being defined” - except asymptotically.

In this case, the term weak convergence is preferable and we say that a sequence of random elements {Xn} converges weakly to X if
Outer Expected value of h of Xn tends to Expected value of h of x for all continuous bounded functions h dot. Here E star denotes the outer expectation, which is the expectation of a “smallest measurable function g that dominates h of (Xn)”.

Properties

- Since F of a is equal to Probability of X less than or equal to a, the convergence in distribution means that the probability for Xn to be in a given range is approximately equal to the probability that the value of X is in that range, provided n is sufficiently large
- In general, convergence in distribution does not imply that the sequence of corresponding probability density functions will also converge
- As an example one may consider random variables with densities, ƒn of x is equal to (1 minus cos of (2 into phi into n into x). These random variables converge in distribution to a uniform with limits 0 to 1, whereas their densities do not converge at all.
- Portmanteau lemma provides several equivalent definitions of convergence in distribution. Although these definitions are less intuitive, they are used to prove a number of statistical theorems
- Continuous mapping theorem states that for a continuous function g of dot , if the sequence {Xn} converges in distribution to X, then so does {g of (Xn)} converge in distribution to g of X
- Levy’s continuity theorem: the sequence {Xn} converges in distribution to X if and only if the sequence of corresponding characteristic functions Phi n converges point wise to the characteristic function Phi of X
- Convergence in distribution is metrizable by the Levy–Prokhorov metric
- A natural link to convergence in distribution is the Skorokhod's representation theorem

Example:
1) Dice Factory
Suppose a new dice factory has just been built. The first few dice come out quite biased due to imperfections in the production process. The outcome from tossing any of them will follow a distribution marked different from the desired *uniform distribution*.
As the factory is improved, the dice become less and less loaded, and the outcomes from tossing a newly produced dice will follow the uniform distribution more and more closely.

---

2) Tossing coins
Let *Xn* be the fraction of heads after tossing up an unbiased coin *n* times. Then, *X*1 has the Bernoulli distribution with expected value mu is equal to point five and variance sigma square is equal to zero point two five. The subsequent random variables *X*2, *X*3, etc will all be

distributed Binomially.

As n grows larger, this distribution will gradually start to take shape more and more similar to the *bell curve* of the normal distribution. If we shift and rescale Xn appropriately, then Zn is equal to root n into (Xn bar minus mu by sigma) will be converging in distribution to the standard normal, the result that follows from the celebrated *central limit theorem*.

Convergence in distribution is the weakest form of convergence, since it is implied by all other types of convergence. However, convergence in distribution is very frequently used in practice. Most often, it arises from application of the central limit theorem. With this mode of convergence, we increasingly expect to see the next outcome in a sequence of random experiments becoming better and better modelled by a given probability distribution.

3. Convergence in Probability

Among several deferent modes of convergence, convergence in probability is one of them. Here, we consider sequences X1, X2, etc of random variables instead of real numbers. As with real numbers, we would like to have an idea of what it means to converge.

In general, convergence will be to some limiting random variable. However, this random variable might be a constant, so it also makes sense to talk about convergence to a real number.
The concept of convergence in probability is based on the following intuition: Two random variables are "close to each other" if there is a high probability that their difference is very small.

The basic idea behind this type of convergence is that the probability of an "unusual" outcome becomes smaller and smaller as the sequence progresses. The concept of convergence in probability is used very often in statistics. For example, an estimator is called consistent if it converges in probability to the quantity being estimated. Convergence in probability is also the type of convergence established by the weak law of large numbers.

Let Xn be a sequence of random variables defined on a sample space. Let X be a random variable and epsilon a strictly positive number. In other words, the probability of Xn being far from X should go to zero when n increases. Formally, we should have limit of n tends to infinity Probability of modulus of Xn minus X greater than epsilon is equal to zero
(Note that Xn is a sequence of real numbers, therefore the above limit is the usual limit of a sequence of real numbers).

Furthermore, the condition limit of n tends to infinity Probability of modulus of Xn minus X greater than epsilon is equal to zero should be satisfied for any variable (also for very small, which means that we are very restrictive on our criterion for deciding whether Xn is far from X). This leads us to the following definition of convergence:

Definition
A sequence {*Xn*} of random variables converges in probability towards X if for all epsilon greater than zero
limit of n tends to infinity Probability of modulus of Xn minus X greater than epsilon is equal to zero
Formally, pick any epsilon greater than zero and any delta greater than zero. Let *P* be the probability that *Xn* is outside the ball of radius epsilon centred at *X*. Then, for *Xn* to converge in probability to *X* there should exist a number *N delta* such that for all *n greater than or equal to N delta* the probability *P* is less than delta.

Convergence in probability is denoted by adding the letter p over an arrow indicating convergence, or using the "plim" probability limit operator as follows:
Xn tends to X with a small letter p or capital P over an arrow or plim as n tends to infinity Xn is equal to X.

Properties

- Convergence in probability implies convergence in distribution
- Convergence in probability does not imply almost sure convergence
- In the opposite direction, convergence in distribution implies convergence in probability only when the limiting random variable *X* is a constant
- The continuous mapping theorem states that for every continuous function *g of dot,* if Xn converges to X on probability, then g of Xn converges to g of X in probability
- Convergence in probability defines a topology on the space of random variables over a fixed probability space

Example:
Height of a person
Consider the following experiment. First, pick a random person in the street. Let X be his/her height, which is Xn, a random variable. Then, you start asking other people to estimate this height by eye. Let Xn be the average of the first n responses. Then (provided there is no systematic error), by the law of large numbers, the sequence Xn will converge in probability to the random variable X.

Archer:
A person takes a bow and starts shooting arrows at a target. Let Xn be his score in nth shot. Initially, he will be very likely to score zeros, but as the time goes and his archery skill increases, he will become more and more likely to hit the bullseye and scores a maximum of 10 points.

After the years of practice, the probability that he scores anything but 10 will be getting increasingly smaller and smaller. Thus, the sequence Xn converges in probability to X is equal to 10. Note that Xn roughly does not converge. However, no matter how professional the archer becomes, there will always be a small probability of making an error. Thus, the sequence {Xn} will never turn stationary. There will always be non-perfect scores at a lesser frequency.

4. Relationship between Convergence in Probability and Convergence in Distribution

Result 1:
Suppose that Xn, n is equal to 1, 2, etc and X are random variables (defined on the same probability space) with distribution functions Fn, n is equal to 1, 2, etc. and F, respectively. If Xn converges to X as n tends to infinity in probability then the distribution of Xn converges to the distribution of X as as n tends to infinity. That is, if Xn converges to X in probability then, Xn converges to X in distribution

Proof
Probability of Xn less than or equal to t is equal to Probability of Xn less than or equal to t intersection Probability of modulus of Xn minus X less than or equal to e plus Probability of Xn less than or equal to t intersection Probability of modulus of Xn minus X greater than e
Less than or equal to Probability of X less than or equal to t plus e plus Probability of modulus of Xn minus X greater than e
First, pick e small enough so that
Probability of X less than or equal to t plus e less than or equal to Probability of X less than or equal to t plus eta by 2.
(Since F is right continuous.)
Then, pick n large enough so that Probability of modulus of Xn minus X greater than e is less than eta by 2

Then, for n large enough, we have
Probability of (Xn less than or equal to t) is less than or equal to Probability of (X less than or equal to t plus e)plus Probability of (modulus of Xn minus X greater than e) implies Fn of t less than or equal to Probability of (X less than or equal to t) plus eta by 2 plus eta by 2
Which is equal to F of t plus eta
Therefore, Fn of t is less than or equal to F of t plus eta
Similarly, for n large enough, if F is continuous at t, we have
Fn of t is greater than or equal to F of t minus eta. This shows that
Limit of n tends to infinity Fn of t is equal to F of t at continuity points of F. However, the converse is not true.

Result 2:
Suppose that Xn converges to C in distribution then, Xn converges to C in probability where, C is a constant.
Thus, when the limit is a constant, convergence in probability and convergence in distribution are equivalent.

5. Applications of Limits

There are several important cases where a special distribution converges to another special distribution as a parameter approaches a limiting value. Indeed, such convergence results are part of the reason why such distributions are *special* in the first place. For example:

- Convergence of the Hypergeometric Distribution to the Binomial
- Convergence of the Binomial Distribution to the Poisson

In case of Sequences of Random Variables:

- Limit Theorems can be used to obtain properties of estimators as the sample sizes tend to infinity
  - Convergence in Probability – Limit of an estimator
  - Convergence in Distribution – Limit of a Cumulative distribution function
- Central Limit Theorem - Large Sample Distribution of the Sample Mean of a Random Sample
- Weak Law of Large Numbers (WLLN): Let X1, up to Xn be iid random variables with Expected value of Xi is equal to mu and variance of Xi is equal to sigma square is less than infinity. Then, the sample mean converges in probability to mu

To summarize, we have the following implications for the various modes of convergence and no other implications hold in general.

- Convergence with probability 1 implies convergence in probability
- Convergence in mean implies convergence in probability
- Convergence in probability implies convergence in distribution

It follows that convergence with probability 1, convergence in probability, and convergence in mean all imply convergence in distribution, so the latter mode of convergence is indeed the weakest. However, convergence in probability and convergence in distribution are equivalent when the limiting variable is a constant. Of course, a constant can be viewed as a random variable defined on any probability space.

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

- The concept of convergence in distribution
- The properties and examples for convergence in distribution
- The concept of convergence in probability
- The properties and examples for convergence in probability
- The applications of these modes of convergence
- The relationship between convergence in distribution and convergence in probability