Convergence of Random Variables

Different modes of convergence explained in simple terms

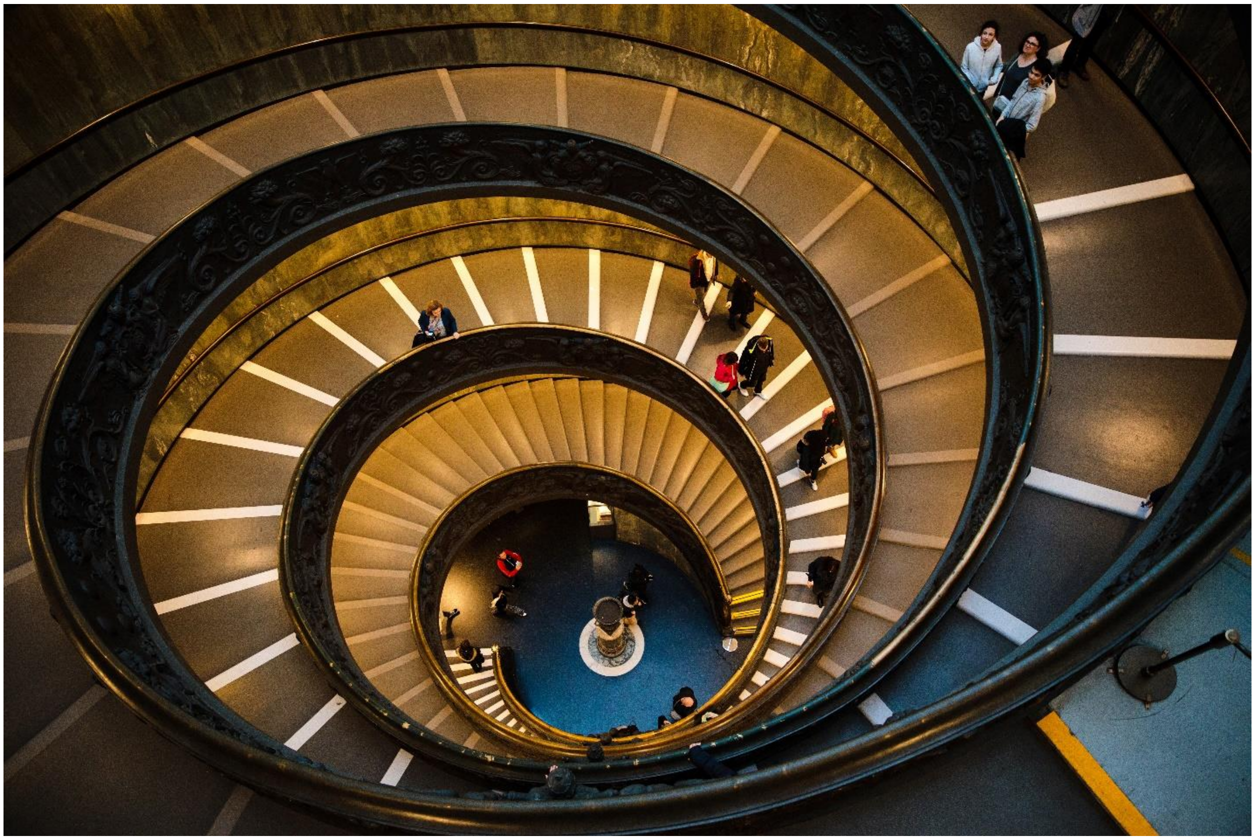


Photo by YIFEI CHEN on [Unsplash]

Convergence of random variables (RVs) implies that a sequence of random variables follows a fixed behavior when repeated for large number of times

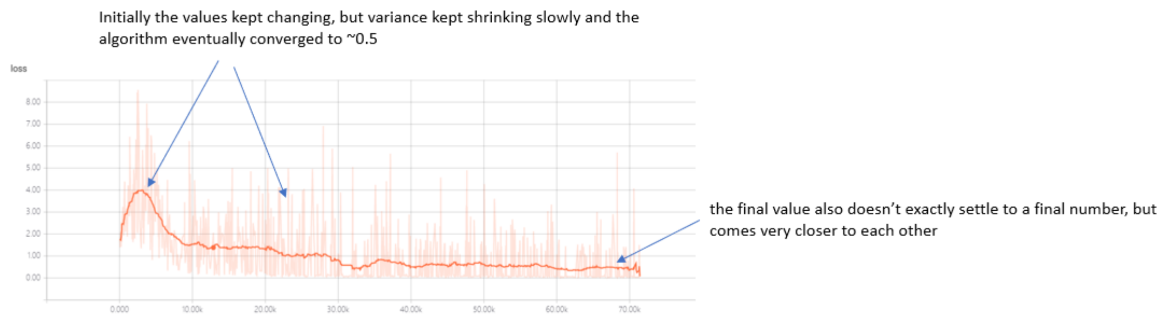
The sequence of RVs (X_n) keeps changing values initially and settles to a number closer to X eventually. But, what does 'convergence to a number close to X' mean? Often RVs might not exactly settle to one final number, but variance keeps getting smaller with them leading the series to converge to become very close to X

So, lets learn a notation to explain the above phenomenon:

$$X_n \xrightarrow[n \to \infty]{} X$$

, which implies that the series converge to final value X as n grows larger.

As Data Scientists, we often talk about whether an algorithm is converging or not? Now, let's observe above convergence properties with an example.



convergence plot of an algorithm

Now that we are thorough with the concept of convergence, lets pay more attention to how "close" should the "close" should be in the context shared above?

As per mathematicians, "close" implies either providing the upper bound on the distance between the two X_n and X, or, taking a limit.

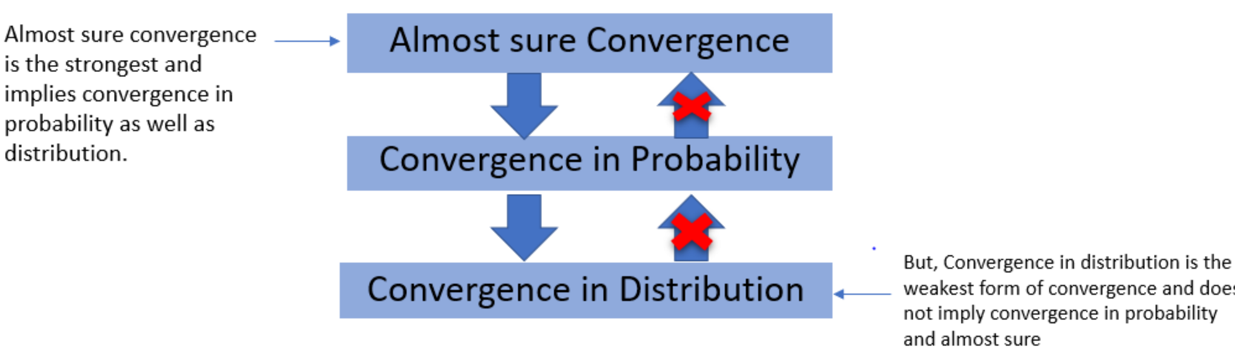
Below, we understand 3 key types of convergence based on taking limits:

1) Almost surely
2) Convergence in probability
3) Convergence in distribution

But why do we have different types of convergence when all it does is settle to a number? Well, that's because, there is no one way to define the convergence of RVs.

Relationship among different modes of convergence:

If a series converges almost surely which is a stronger convergence, then that series converges in probability and distribution as well. But, vice versa is not true



Convergence in distribution:

Intuition: It implies that as n grows larger, we become better in modelling the distribution and in turn the next output.

Definition: A series of real number RVs converges in distribution if the cdf of X_n converges to cdf of X as n grows to ∞

$$\lim_{n \to \infty} F_n(x) = F(x),$$

where F represents the cdf and should be continuous for all x € R

As it only depends on the cdf of the *sequence of random variables and the limiting random variable*, it does not require any dependence between the two. So, convergence in distribution doesn't tell anything about either the joint distribution or the probability space unlike convergence in probability and almost sure

Notation: $X_n \overset{d}{\to} X$

Example: Central limit theorem (CLT).

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$$

$\overline{X_n}$ is the sample mean of a series of 'n' i.i.d. RVs; and μ and σ are the mean and standard deviation of the population

Question: Let X_n be a sequence of random variables X_1, X_2,...such that its cdf is defined as

F(x) = (1-(1-1/nx)^nx). Prove if it converges in distribution, if X~ exp(1)

Solution: Lets first calculate the limit of cdf of X_n:

$$\begin{aligned}
\lim_{n \to \infty} F_{X_n}(x) &= \lim_{n \to \infty} \left(1 - \left(1 - \frac{1}{n}\right)^{nx}\right) \\
&= 1 - \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^{nx} \\
&= 1 - e^{-x} \\
&= F_X(x), \qquad \text{for all } x.
\end{aligned}$$

As the cdf of X_n is equal to the cdf of X, it proves that the series converges in distribution.

Conceptual Analogy: During initial ramp up curve of learning a new skill, the output is different as compared to when the skill is mastered. Over a period since the new skill is learnt so as to not make big errors, it is safe to say that output converges in distribution

Convergence in probability:

Intuition: Let's say we fix the distance as ε, where we want to measure the two RVs. Now, the probability that X_n differs from the X by more than ε is 0. Put differently, the probability of unusual outcome keeps shrinking as the series progresses.

Definition: A series X_n is said to converge in probability to X if and only if:

$$\lim_{n \to \infty} \Pr\left(|X_n - X| > \varepsilon\right) = 0.$$

Unlike convergence in distribution, convergence in probability depends on the joint cdfs i.e. X_n and X are dependent

Notation: $X_n \xrightarrow{p} X$

Example: Weak law of large numbers

As 'weak' and 'strong' law of large numbers are different versions of Law of Large numbers (LLN) and are primarily distinguished based on the modes of convergence, we will discuss them later.

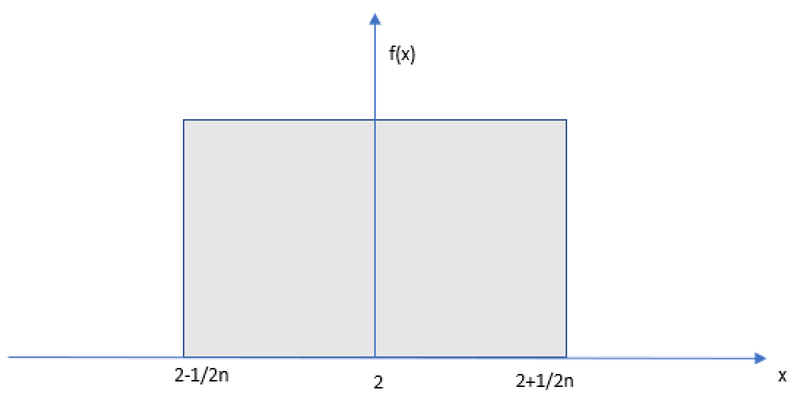
Question: Let X_n be a sequence of random variables X_1, X_2,...such that

X_n ~ Unif (2-1/2n, 2+1/2n)

For a given fixed number 0< ε<1, check if it converges in probability and what is the limiting value

Solution: For X_n to converge in probability to a constant number, we need to find whether P(|X_n-2| > ε) goes to 0 for a certain ε

Let's see how the distribution looks like and what is the region beyond which the probability that the RV deviates from the converging constant beyond a certain distance becomes 0.



For a fixed distance ε < 1/2n, X_n converges in probability to 2 i.e. $P(|X_n - 2| > \varepsilon) = 0 \; \forall \; \varepsilon > 1/2n$

Conceptual Analogy: If the performance of a school is measured based on 10 randomly selected students from each class, we may not get the true ranking for school. However, as we keep considering more and more students from each class, we arrive at true ranking of the school.

Almost sure convergence:

Intuition: the probability that X_n converges to X for a very high value of n is almost sure i.e. prob is 1.

Definition: the infinite sequence of RVs X1(ω), X2(ω)… Xn(w) has a limit with probability 1, which is X(ω).

$$\Pr\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

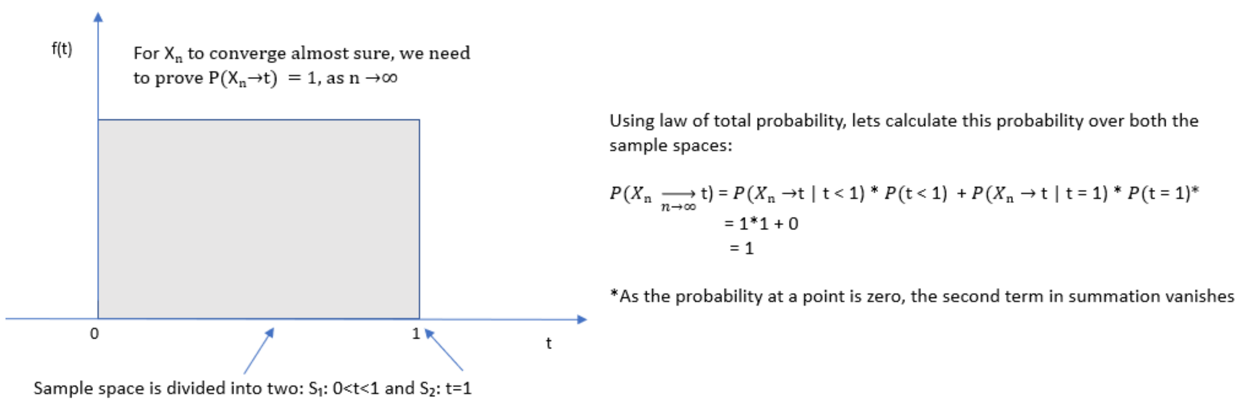
where Ω: the sample space of the underlying probability space over which the random variables are defined

Notation: $X_{\mathrm{n}} \xrightarrow{a.s.} X$

Example: Strong Law of convergence

Question: Let X_n be a sequence of random variables X_1, X_2,...such that

$X_n = t + t^n$, where T ~ Unif(0, 1)



Conceptual Analogy: If a person donates a certain amount to charity from his corpus based on the outcome of coin toss, then X1, X2 implies the amount donated on day1, day2. The corpus will keep decreasing with time, such that the amount donated in charity will reduce to 0 almost surely i.e. with a probability of 1.

Distinction between the convergence in probability and almost surely:

- Limit is outside the probability in convergence in probability, while limit is inside the probability in almost sure.
- 'Weak' law of large numbers is a result of the convergence in probability which is weak convergence. The intuition behind calling it as weak convergence is that it states that sample mean $\overline{X_{\mathrm{n}}}$ will be closer to population mean μ with increasing n but leaving the scope that ε error can occur infinite number of times. However, almost sure convergence is a more constraining one and says that the probability of the difference between the two means being lesser than ε is 1. The difference is very subtle and mostly theoretical.