Convergence

Convergence is an important concept in Probability and Analysis. So we begin by recalling some ideas of convergence from an earlier course.

The fundamental notion is that of convergence of a sequence (of numbers) x_n. We say that x_n converges to 0 as n goes to infinity if, we can ensure that it is as small as we like once we choose n sufficiently large. In other words, for any $c > 0$, we can find k so that $|x_n| < c$ for $n > k$.

A fundamental fact (principle of Archimides) is that the fractions $1/n$ go to zero as n goes to infinity. So we can replace the above definition with the following: for any positive integer $N > 0$, there is a $k > 0$ so that $|x_n| < 1/N$ for $n > k$. This is sometimes a bit nicer as it *suggests* that we can write k (a positive integer) as (some kind of) a function of M (also a positive integer). In many cases of interest, this can actually be done. However, it is far from "automatic" in general.

The above notion of convergence leads us (as shown by Cauchy) from rational numbers to their limits, the real numbers.

In this course, we are primarily dealing with random variables, which do not (except for constant random variables!) have a predetermined value. Hence, it does not make sense to talk about $|Z_n| < c$ as a True/False statment, but only as a probability! Moreover, it is too much to demand that $P(|Z_n| < c) = 1$ for all large n. So we need to ask for this probability to go to 1 with n.

Convergence in Probability

While stating the weak Law of Large numbers, we came across such a notion of convergence of a sequence Z_n of random variables. The sequence is said to converge to 0 *in probability* if, for any $c > 0$, $P(|Z_n| \geq c)$ goes to zero as n goes to infinity.

Thus, we can re-state the Weak Law of Large Numbers as follows.

Given a sequence of independent random variables X_i such that $\sigma^2(X_i) \leq M$ for some fixed M, let $Y_n = (X_1 + \cdots + X_n)/n$ denote the "running" averages, then $Y_n - E(Y_n)$ converges to 0 in probability.

Why do we call this *weak* and how do we make it stronger?

Let us choose c small, for example $c = 1/N$ for a large integer N. Then the above condition means that we can make the probability of the event $|Y_n - Y| \geq 1/N$ very small, for example smaller than $1/M$ for some large integer M, by choosing n large enough.

Ideally, we *would* like to have a condition which says that the *union* of $|Y_n - Y| \geq 1/N$ is very small. This means that we have high probability for the event that

$|Y_n - Y| < 1/N$ for *all* (sufficiently large) n. The above condition is *weaker* than this new condition.

Almost Sure Convergence

We now look at the stronger condition that for all $c > 0$ (small), and for all M (large integer), there is an k so that

$$P(\cup_{n>k}(|Y_n - Y| > c)) < 1/M$$

This is our stronger notion of convergence, which is called convergence *almost everywhere* or *almost sure* convergence. The latter term is used to underline that the event that Y_n converges to Y has probability 1.

We note that this condition *implies* the previous condition, but is not, in any obvious way, implied by it. Another way of describing this condition is to say that the probability that $|Y_n - Y| > c$ for infinitely many n (or infinitely often) is 0.

This stronger notion of convergence is what the (Strong) Law of Large Numbers provides. It states:

Given a sequence of independent random variables X_i such that $\sigma^2(X_i) \leq M$ for some fixed M, let $Y_n = (X_1 + \cdots + X_n)/n$ denote the "running" averages, then $Y_n - E(Y_n)$ converges to 0 almost surely.

We will not provide a direct proof of this during this course, but it can be derived as a consequence of the Central Limit Theorem which is stated below.

LLN Wars

Clearly, the strong law as stated above is stronger than the weak law! Once the strong law was proved people tried to prove the weak law with *less* assumptions about the distributions of X_i. For example, we can get rid of the hypothesis that $\sigma^2(X_i)$ are uniformly bounded; after all the statement only involves the expectations $E(X_i)$.

Thus we can ask if the weak law holds for independent *identically distributed* random variables under the assumption that $E(|X_i|)$ is finite. This version of the weak law was proved. However, the strong law stuck back as Kolmogorov was able to prove the strong law under the same conditions!

However, the weak law fought back! On the one hand, the assumption the $E(|X_i|)$ is bounded was replaced with some even weaker assumptions on the behaviour of X_i in the event it takes large values, and yet the weak law was proved. On the other hand Kolmogorov showed that *any* version of the strong law must involve a bound on $E(|X_i|)$!

Central Limit Theorem

Let us move away from these violent and messy battles over laws and ask a different question. What can be said about the *distribution* of Y_n.

In the special case of the "biased" coin, we have already seen the de Moivre-Laplace Theorem as follows.

Let X_i be the random variable that returns 1 (success) with probability p and 0 (failure) with probability $1-p$. Moreover, assume that the X_i are independent variables for the different i's. Then $S_n = \sum_i X_i$ *counts* the number of successes and is given by the Binomial distribution:

$$P(S_n = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

The de Moivre-Laplace Theorem implies that we have an asymptotic identity

$$P(a < |S_n - np|/\sqrt{np(1-p} \leq b) \simeq \frac{1}{\sqrt{2\pi}} \int_a^b dte^{-t^2/2}$$

We note that the running average Y_n is just S_n/n and $E(Y_n) = p = E(X_i)$. Moreover, $\sigma^2(X_i) = p(1-p)$.

Given a sequence of indenpendent identically distributed random variables X_i with mean m and variance s^2. Let $Y_n = (X_1 + \cdots + X_n)/n$ be the running averages as before. The Central Limit Theorem states that we have an asymptotic identity

$$P(a < \sqrt{n}(Y_n - m)/s \leq b) \simeq \frac{1}{\sqrt{2\pi}} \int_a^b dte^{-t^2/2}$$

Note that a and b remain fixed as n goes to infinity.

A different way of saying this is to say that the distribution functions of $\sqrt{n}(Y_n - m)/s$ converge to a normal distribution function. This kind of convergence is called convergence in distribution. **Warning**: It is important to note that this does not say *anything* about the size of $|Y_n - Y|$ for some "limiting" random variable Y.

In order to understand the reasons behind the Central Limit Theorem (and other limit theorems), we need to understand Characteristic Functions.