Limits of Random Variables

One can show that a random variable X is a limit (in some suitable sense) of a sequence of discrete random variables. The idea is similar the way in which real numbers are limits of rational numbers or continuous functions are limits of polynomial functions.

Even if we may never encounter real numbers in practice (for example, the actual numerical result of a physical experiment or a computer calculation is always a rational number!), the mathematical idealisation of a real number is useful in many ways. for example, calculus can give us different ways to calculate the same number by using a different limiting process which may be convenient. This is then a good approxiation to the rational number we want.

Similarly, the limiting distributions of a (convergent) sequence of discrete distribtions may be computable in a number of ways which have no *obvious* relation with the original sequence.

Prime Numbers

Limiting behaviours are of interest. We know that there are infinitely many prime numbers (theorem proved by Euclid). How are these numbers distributed. Put it differently, if we choose a number at random between 2 and N what is the probability p_N that this number is prime?

Here is a *heuristic* argument. The probability that the number is not divisible by 2 is $(1 - 1/2)$, that it is not divisible by 3 is $(1 - /3)$ and and so on. So we can calculate the probability that the number is not divisible by a prime as $p_N = (1 - 1/2)(1 - 1/3) \cdots$. Now, we see

$$1/p = (1 + 1/2 + 1/2^2 + \dots)(1 + 1/3 + 1/3^2 + \dots) \cdots = 1 + 1/2 + 1/3 + 1/4 + \dots$$

Since we are only considering numbers upto N we can only look at $1 + 1/2 + 1/3 + 1/4 + \cdots + 1/N$. This number is close to $\log(N)$ for large N. It folows that p is approximately $1/\log(N)$.

The assertion that $p_N \simeq 1/\log(N)$ as N approches infinity, is called the Prime Number Theorem. It was conjectured by Gauss and eventually proved by Hadamard and de la Vallee Poussion independently. (Why is the above argument not a proof? Which parts can be "fixed" and which are really hard?) The statment $f(x) \simeq g(x)$ as x approaches infinity is to be read as $f(x)$ asymptotically approaches $g(x)$ as x approches infinity. Its precise meaning is that $f(x)/g(x)$ approaches 1 as x approaches infinity.

In a similar fashion, we would like to understand the asymptotic behaviour of certain discrete probability distributions.

Poisson Distribution

We start with a standard Binomial distributed random variable Y with parameters n and p. In other words,

$$P(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

As seen earlier, we have $E(Y_n) = np$. Suppose we *fix* $np = c$ and k, and allow n to go to infinity, and try to understand the asymptotic behaviour. Note that this means that $p = c/n$ decreases with n; in particular, it is dependent on n and could be written as p_n..

This is not merely to exercise our "Mathematical Muscles"! Here is an example that may be interesting to some people.

Suppose that an estimated c birds of a rare species have come onto IISER campus. A bunch of n students learn to recognise the birds and go to different parts of the campus looking for a bird of this species. If we assume that:

(a) there are so few birds that each one is likely to spot only one bird, and
(b) the birds no not move around so it is unlikely that the same bird is spotted by two spotters,

then the probability of a given spotter being successful is c/n. So the probability of k spotters being successful is $\binom{n}{k}(c/n)^r(1-c/n)^{n-k}$ (under various simplifying assumptions of independence etc.). We can ask what happens as we increase the number of spotters. We can also ask for an *estimate* of this probability when the number of spotters is very large (compared with k and c).

To make such an estimate, we re-write the above expression as follows:

$$\binom{n}{k}(c/n)^k(1-(c/n))^{n-k} = \frac{c^k}{k!} 1 \cdot (1-1/n) \cdots (1-(k-1)/n) \frac{(1-(c/n))^n}{(1-(c/n))^k}$$

Now, if k and c are fixed and n goes to infinity:

- $(1-(a/n))^r$ goes to 1 for all a between 1 and k and also $a = c$.
- $(1-(c/n))^n$ goes to $\exp(-c) = e^{-c}$

In other words, as n goes to infinity this probability has $(c^k/k!)e^{-c}$ as its limit. The Poisson distribution is defined by the probability mass function (we can check that this defines a distribution; see below)

$$P(X = k) = (c^k/k!)e^{-c} \text{ for } k \text{ a non-negative integer}$$

Since the word "poisson" is the French word for "fish" it is interesting that this distribution is a useful way to estimate the probability of r successes in a fishing expedition where there are a fixed number of fish in a large pond and many fishermen!

We can calculate the expectation of the Possion distribution

$$E(X) = \sum_{k=0}^{\infty} k \frac{c^k}{k!} e^{-c} = c \sum_{k=1}^{\infty} \frac{c^{k-1}}{(k-1)!} e^{c} = c$$

We can also calculate the variance by observing that

$$\sum_{k=0}^{\infty} k(k-1) \frac{c^k}{k!} e^{-c} = c^2 \sum_{k=2}^{\infty} \frac{c^{k-2}}{(k-2)!} e^{c} = c^2$$

It follows that $E(X^2) = c^2 + c$ so that $\sigma^2(X) = c$.

Given Binomial random variable Y_n with $E(Y_n) = np = c$, then for a fixed k, and for *very large* n, the value of $P(Y_n = k)$ is well approximated by $P(X = k)$; where X is a Poisson random variable with $E(X) = c$. In this case we say that the distribution of Y_n converges to the distribution of X; or that Y_n converges to X *in distribution*. Note that this does not allow us to interpret X in terms of Y_n in a probabilistic sense! We will see later how we can do that.

Poisson Density

A different kind of limit can be found via the Negative Binomial distribution for 1 success.

In this case, the random variable T_p counts the number of coin flips before getting *one* Head. Then $P(T_p = k) = p(1-p)^{k-1}$. In fact

$$P(T_p > k) = \sum_{s=k}^{\infty} p(1-p)^s = p(1-p)^k \sum_{s=0}^{\infty} (1-p)^s = p(1-p)^r \frac{1}{1-(1-p)} = (1-p)^k$$

If we replace p by c/n and r by tn, this becomes

$$P((1/n)T_{c/n} > t) = (1-(c/n))^{tn}$$

and we can take a limit as n goes to infinity. This becomes $P(W > t) = e^{-ct}$. We note that this makes sense for $t \geq 0$ being *any* number. Again, we note that this determines a probability distribution.

In fact, this distribution is the integral of a density ce^{-cs} for $s \geq 0$. This follows from the identity

$$\int_0^t ce^{-cs}ds = 1 - e^{-ct} = P(W \leq t)$$

This too may seem like a Mathematical exercise unless we give an interesting example.

Consider the situation where you are waiting for a message on your phone. Suppose p is the probability that a message arrives during a one hour interval. The probability that you waited for (at least) t hours (with hourly checks for a message) without seeing a message is $(1-p)^t$.

Because you are impatient for the message and you don't want to miss the message when it arrives, you start checking every 5 minutes. The probability that the message arrives in a given 5 minute interval is now $p/12$. The probability that you waited for t hours (where t is a fracion of the form $a/12$) is $(1-p/12)^{12t}$ (since you checked $12t$ times!). We can now imagine that you checked every minute or even more frequently. Finally, you are just staring at your phone until you see the message! The probability that you waited t hours (where t is now a real number) is e^{-pt}. In other words, if W denotes the random variable denoting the amount of time you need to wait for a message to appear, then $P(W > t) = e^{-pt}$; here p is the frequency with which messages arrive (per hour). Note that $P(W \leq t) = 1 - e^{-pt}$ which is similar to our limit above with c replaced by p (which corresponds with our idea of probability as frequency).

By using the density as given above, it is easy to use a little calculus to calculate the expectation and variance for this waiting time distribution.

$$E(W) = \int_0^\infty s(ce^{-cs})ds = \int_0^\infty ue^{-u}(du/c) = 1/c$$

If we interpret c as a frequency as in the previous paragraph, then this makes sense as $1/c$ is the expected amount of time to wait for a singe message. Similarly, we can calculate

$$E(W^2) = \int_0^\infty s^2(ce^{-cs})ds = \int_0^\infty u^2e^{-u}(du/c^2) = -u^2e^{-u}|_0^\infty + \int_0^\infty (2u)e^{-u}(du/c^2) = 2/c^2$$

It follows that $\sigma^2(T) = 1/c^2$.

Once again, it is worth pointing out that the only link between $(1/n)T_{c/n}$ and W is that the *distribution* of the former converges to the distribution of the latter. At some later stage, a stronger connection can be established.