THE ESSENCE OF THE LAW OF LARGE NUMBERS

Michael Keane

CWI P.O. Box 94079
1090 GB Amsterdam
Electronic mail: keane@cwi.nl

The law of large numbers, not really a law but a mathematical theorem, is at the same time a justification for application of statistics and an essential tool for the mathematical theory of probability. As such, it must be taught to many students. The traditional method for this, using independent and identically distributed random variables, was developed by Kolmogorov in the 1930's, and explains well what happens, and much more, at this level of generality. However, it has recently come to light that the reason for the validity of this theorem in its general setting, that of stationarity, is much simpler than was first thought. In this short article, I shall try to explain to the general audience towards whom this collection is directed, the essence of the law of large numbers. A complete treatment should certainly include many references and interesting historical comments, and I apologize for their absence here.

Let me start with the *basic law of large numbers* by considering, very simply, an infinite sequence

$$x_0,\ x_1,\ x_2, \ldots$$

each of whose elements is either 0 or 1. Perhaps it will help (or hinder!) to think of x_n as the result of the n^{th} trial of an uncertain experiment, with $x_n = 1$ designating success and $x_n = 0$ failure. Let

$$a_n = \frac{x_0 + x_1 + \ldots + x_{n-1}}{n} \qquad (n \geq 1)$$

denote then the average numbers of successes up to time n. It is very easy to see mathematically that for some sequences x,

$$\lim_{n\to\infty} a_n$$

exists, while for other sequences x, this is not the case. One can only affirm with certainty that

$$\lim_{n\to\infty} (a_{n+1} - a_n) = 0,$$

but nothing impedes the averages a_n from oscillating more and more slowly as n grows. Thus it seems that further discussion is useless, and that uncertainty here must be accepted.

Phenomenologically, however, we are faced with the fact that in certain situations, such limits seem to exist, and the society makes seemingly understandable statements concerning the percentage of smokers dying of cancer, the probability of rain tomorrow, or an industrial average yield. We are confronted with the question as to whether nature produces sequences whose averages do converge, and why. Of course, this is not a mathematical question, and in order to say something mathematically sensible, one must adopt a model.

The currently accepted model, and it is difficult to see how it could be replaced by something else, is that for a given situation in which such sequences x appear, in principle all sequences are possible, but there is also a mass distribution with total mass 1 over the set of sequences, which assigns to each "event" which might occur a *probability*, this being the total mass of those sequences for which the event occurs. If an event, for instance the existence of $\lim_{n\to\infty} a_n$, has probability 1, then one says that the event will occur *almost surely*.

The determination of such a mass distribution in different practical situations is one of the most important tasks for probabilists, and requires a good mixture of mathematics, other sciences, and good old common sense. First principles are of utmost importance, as determining such an object by experimentation resembles very much a cat chasing its own tail! One of the basic properties of such a mass distribution, already alluded to briefly above, is that of *stationarity*. We say that the probability measure (= mass distribution) is *stationary* if the events have time-homogeneous probabilities. That is, shifting any event forwards or backwards in time does not change its probability.

Perhaps a brief remark on mass distributions is in order. There is a branch of mathematics, measure theory, which deals extensively with the specification and manipulation of such objects. However, one can understand well most arguments and principles by using the intuitive notion, which is my intention here.

Now we can state the

BASIC LAW OF LARGE NUMBERS:

If $x = (x_0,\ x_1,\ldots)$ is a stationary sequence of zeroes and ones, then $lim_{n\to\infty}\, a_n$ exists almost surely.

Just to be sure that you are (mathematically) still with me: A unit mass distribution on sequences of zeroes and ones is given; it is stationary. Then the set of all sequences x for which $\lim_{n\to\infty} a_n$ exists has total mass 1. The set of sequences for which this limit does not exist has mass 0. Remember, this is a theorem, and I want to explain the proof.

To understand the proof will require the level of first-year university analysis, given the intuitive acceptance of the mass distribution notion. We begin by defining

$$\bar{a} := \limsup_{n\to\infty}\ a_n;$$

this always exists, and $0 \le \bar{a} \le 1$. It is also clear that if we had started observing x at a later time point, the value $\bar{a}$ would be the same:

$$\bar{a} = \limsup_{n\to\infty} \frac{x_k + x_{k+1} + \ldots + x_{k+n-1}}{n}$$

for any $k \geq 0$ and any sequence x.

Next, we need a way to measure how close we are to the lim sup, $\bar{a}$. Thus, let $\epsilon > 0$ be a fixed positive number, and for each $k \geq 0$, define

$$N_k := \min\left\{n \geq 1 : \frac{x_k + x_{k+1} + \ldots + x_{k+n-1}}{n} \geq \bar{a} - \epsilon\right\}.$$

By the definition of lim sup, the set on the right is non-empty and hence N_k is finite for each k. The crucial point we need to address concerns the size of the numbers N_k; to make our idea clear, let us examine the simplest case first.

CASE 1. Suppose that for each $\epsilon > 0$ there exists a (large) positive integer M such that for each k, $N_k \leq M$ almost surely. (That is, the set of sequences x for which $N_k \leq M$ has total mass 1.)

REMARK: Note that by our assumption of stationarity the events $N_k \leq M$ for different k all have the same probability.

If now x is such a sequence that for each k, $N_k \leq M$, we claim that $\lim_{n\to\infty} a_n$ exists. The idea is that, as n gets larger, a_n can only change more and more slowly, and that then wandering is impossible because the lim sup is reached again and again within M steps. Formally, one proceeds as follows. Fix $\epsilon > 0$ and choose any $n > M/\epsilon$. Then starting at the beginning of x, break x up into pieces of lengths at most M such that the average of x over each piece is at least $\bar{a} - \epsilon$. Stop at the piece containing the coordinate n. Then it is clear that

$$x_0 + x_1 + \ldots + x_{n-1} \geq (n - M)\,(\bar{a} - \epsilon),$$

so that

$$a_n = \frac{x_0 + x_1 + \ldots + x_{n-1}}{n} \geq (1 - \epsilon)\,(\bar{a} - \epsilon) \geq \bar{a} - 2\epsilon$$

for each $n > M/\epsilon$; it follows that $\lim_{n\to\infty} a_n = \bar{a}$ exists.

REMARK: Note that only the last piece is of importance; it must not become too long.

Actually, the same type of argument works in the general case, when combined with an idea coming originally from non-standard analysis.

CASE 2: General case. By the remark after Case 1, it remains true that the events $N_k \leq M$ all have the same probability, for any k and fixed M. Since N_k is finite for each x, we may not be able to find an M, for $\epsilon > 0$ given, such that these events have probability 1, but we certainly can choose M so large that for any k, the probability of $N_k \leq M$ is less than ϵ.

Fix now such an integer M, given $\epsilon > 0$. Next, we want to make the same inequality work for us, but we are impeded whenever $N_k > M$. So let us change x at those places

to insure quick arrival at the lim sup.
Namely, define

$$x_k^* := \begin{cases} x_k & \text{if} \quad N_k \leq M \\ \\ 1 & \text{if} \quad N_k > M. \end{cases} \qquad (k \geq 0)$$

Then clearly $x_k^* \geq x_k$ for each k, so that if we set

$$N_k^* := \min\left\{n \geq 1 : \frac{x_k^* + \ldots x_{k+n-1}^*}{n} \geq \bar{a} - \epsilon\right\}$$

(same $\bar{a}$), then $N_k^* \leq N_k$, and moreover if k is such that

$$N_k > M,$$

then we have

$$N_k^* = 1,$$

since setting $x_k^* = 1$ insures immediate arrival above $\bar{a} - \epsilon < 1$.

Now we are almost ready. As above, breaking x^* up into pieces yields for $n > M/\epsilon$.

$$x_0^* + x_1^* + \ldots + x_{n-1}^* \geq (n - M)\ (\bar{a} - \epsilon),$$

but now we cannot conclude anything about the sequence x because we have replaced it by x^*.

Instead, we now need to use our mass distribution to calculate the average value of each side of the inequality over all sequences x, called by probability theory the *expectation* and denoted by $\mathbb{E}(\cdot)$. Let

$$\mathbb{E}\ (x_0) =: p$$

and

$$\mathbb{E}\ (x_0^*) =: p^*;$$

by stationarity, $\mathbb{E}\ (x_p^*) = p^*$ for all k, and by the choice of M, we have

$$p^* \leq p + \epsilon.$$

Of course, p is just the probability that $x_k = 1$, and p^* the probability that $x_k^* = 1$, for any k. Now, taking expectations of each side of the inequality results in

$$n(p + \epsilon) \geq np^* \geq (n - M)(\mathbb{E}\ (\bar{a}) - \epsilon) \qquad (n \geq M/\epsilon).$$

Now divide by n, send n to infinity and then ϵ to zero, giving

$$\mathbb{E}\ (\bar{a}) = \mathbb{E}\ (\limsup_{n\to\infty}\ \frac{x_0 + \ldots + x_{n-1}}{n}) \leq p.$$

Finally, apply the entire argument above to the "mirrored" $0-1$–sequence $y_k = 1 - x_k$; an easy calculation (exercise!) shows that

$$\mathbb{E}\ (\liminf_{n\to\infty} \frac{x_0 + \ldots + x_{n-1}}{n}) \geq p.$$

But for any sequence x, certainly

$$\liminf_{n\to\infty} \frac{x_0 + \ldots + x_{n-1}}{n} \leq \limsup_{n\to\infty} \frac{x_0 + \ldots + x_{n-1}}{n};$$

it is an elementary fact of expectations or averaging that the three inequalities then must be equalities, the last one almost surely. Hence lim sup = lim inf for a set of sequences of total mass one, i.e. the limit exists almost everywhere. This concludes the proof of the basic law of large numbers.

In concluding, we state without proof that this method can be widely extended with minor, straight-forward modifications to the most general laws of large numbers based on stationarity. The above proof should, however, in my opinion be included in basic probability courses, since it so clearly shows the nature of the interplay of stationarity assumptions and the existence of statistical limits.