Repetition Of Finite Outcome Experiments

The Laws Of Large Numbers With “Detours” To Randomness

I placed the word detours in quotation marks because in my opinion these are not detours rather necessary explanations without which even the weak law of large numbers is meaningless.

Basics

The possible outcomes of an experiment are denoted as $\omega_1, \omega_2, \ . \ . \ . \ , \omega_m$.
At throwing a coin $m = 2$ and $\omega_1 = H =$ head , $\omega_2 = T =$ tail.
At throwing a die $m = 6$ and $\omega_i = i$.
Drawing cards from a deck or betting on horses at the track are also finite outcome experiments.
But there are infinite outcome experiments like throwing a dart onto a board.
The simplest finite outcome experiments have equal chances: $[\,\omega_1\,] = [\,\omega_2\,] = \ . \ . \ . = [\,\omega_m\,] = \dfrac{1}{m}$.
At a fair coin both sides have $\dfrac{1}{2}$ chance while at a fair die all sides have $\dfrac{1}{6}$.
We all heard about loaded or leaded coins or dice and these can have “false” chances.
But what remains is that the total of the chances $[\,\omega_1\,] + [\,\omega_2\,] + \ . \ . \ . + [\,\omega_m\,] = 1$.
If the tail is leaded and so the head becomes $\dfrac{2}{3}$ chanced, then the tail must be $\dfrac{1}{3}$.

The most obvious intuitive consequence of chances are its reciprocals as a “kind of expectability”.
So, we expect a 6 from every six throws but we know just as well that six consecutive trials will not assure a 6 outcome and of course we can succeed at once from a first throw.
So this “expectability” is something much more complicated to be exact.

Counting cases

The really basic and yet very general intuition that can put us into the right track is that the chance of some desired property is the number of outcomes in the property divided by the number of all possible outcomes, if these outcomes are all equal chanced.
This last condition is crucial and ignoring it can be a source of mistakes.

Simplest example of such error is regarding throws of two coins and then saying that the possible outcomes are three as two heads, two tails or one head one tail. So then we would get one third chance for all these three possible scenarios. In truth, the one head one tail means two cases because physically we can have two versions depending on which coin was which. Using two different coins this becomes evident and this shouldn’t influence the outcomes.
So we have four possible outcomes and the chances are a quarter.

Time and order

This example highlights a very important other issue. Namely, why simultaneous tries are usually replaced by repeated tries. Indeed, throwing a coin twice we would at once realize that the time order of head then tail is different from tail then head. So time brings in the trivial outcome distinctions even with actually using common objects.
In spite of this, we must realize that according to our present knowledge, time has no role in the chances! Even a whole infinite sequence of tries should be the same as trying them simultaneously.

This then becomes a paradox!
In a sequence of tries we expect the heads and tails to become half half portions of the longer and longer beginnings. This is the simplest version of the physical law of large numbers.
In spatial version the same expectation is about the beginnings as left end sections of the simultaneous tries. So we can visualize beggars in an finite row and each flipping their coins.
Then it’s natural to contemplate that such equalization also means that we can rearrange the beggars so that the outcomes would become head tail alternatingly. This as infinite outcome would of course be impossible randomly, in spite of obeying the physical law of large numbers.
The correct implication is that this law is obviously insufficient to describe randomness but a fatal error in the whole vision is that the dependence on order is much stronger than we thought!
Even if we regard 0-s and 1-s that are now outcomes for our beggars throwing dices denoting the no six or the six outcomes, we would have an infinity of 0-s and 1-s. The 0-s were of course now much more frequent and only a sixth of the beginnings should be 1-s. And yet to rearrange these into an exactly alternating 0 and 1 sequence is again possible simply by both being infinite many. So the equalization had no role in our previous trickery and we can alter even the chances. The main moral is that randomness seems to be order dependent as if the outcomes would know where they come about. This is false though. The outcomes are not position dependent! We simply rearranged one possible outcome with an other. Every single outcome scenario is a nil chanced possibility so the alternating is correctly equivalent with a random one.
This paradox is already appearing in finite cases.
Ten heads is not rarer that any fix outcome combination if we mark the coins.
So we are back to the earlier simplest mistake of outcome scenarios with two coins.
The intuition that after ten heads a tail is more likely is a false but natural initial one if someone is thinking in time. The spatial combinations will destroy this false intuition very easily.
A bigger mystery is that gamblers who play in time, will all come to this correct no time dependence and thus realization that after ten heads the eleventh head is exactly same half chanced as a tail. So time must have a deeper layer of role in our intuitions that corrects itself too.

Set paradoxes

The rearrangement paradox of random outcomes to become ruled is relating to the simpler infinity paradox of sets.
The simplest is that a sequence remains a sequence if we cut off a beginning.
A more tricky one realized by Galileo already is that the odd and even members are two halves but each a sequence on their own. So a half of an infinite set is same many.
With new attached meanings these set paradoxes become deeper.
For example, even the simplest one element cut off can have a new twist.
Imagine beggars in a row again! Steal the coin of the first and say him: “Steal your neighbors coin and tell him to do the same”! Eventually, we created a coin from nothing.

More involved case countings

First let’s see an error again with cases to calculate the chances of double sixes in throwing two dices, which was frequent at early gamblers.
The aimed property has one case { 6 , 6 } and we seem to have 21 cases by the following logic:
We have five other double outcomes and we also have fifteen mixed outcomes as :
{ 1 , 2 } , { 1 , 3 } , { 1 , 4 } , { 1 , 5 } , { 1 , 6 } , { 2 , 3 } , { 2 , 4 } , { 2 , 5 } , { 2 , 6 } ,
{ 3 , 4 } , { 3 , 5 } , { 3 , 6 } , { 4 , 5 } , { 4 , 6 } , { 5 , 6 }.
Of course, we were wrong because these mixed outcomes constitute actually two cases each if we distinguish the two dice somehow like one being black the other red. And we have to do this because the double cases are singular physically. We would not have two double six for example with a black and red die. So the actual number of cases are all possible ordered pairs:
(1 , 1) , (1 , 2) , . . . , (1 , 6) , (2 , 1) , (2 , 2) , . . . , (2 , 6) , . . . , (6 , 1) , . . . , (6 , 6).

The first members are the black die outcomes the second the red. And of course we have 36 pairs.
So the correct chance of a double six is $\frac{1}{36}$.
The really interesting situation starts if we stay with this set of possible outcomes but look for not a double six rather to throw at least one six.
The desirable outcomes are (6 , 1) , (6 , 2) , (6 , 3) , (6 , 4) , (6 , 5) , (1 , 6) , (2 , 6) , (3 , 6) , (4 , 6) , (5 , 6) , (6 , 6) eleven in number and so our chance is $\frac{11}{36}$.
Surprisingly big at first sight but then quite oppositely it seems not large enough.
Indeed, we might say that one die has $\frac{1}{6}$ chance and so at least one should be $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.
Or maybe even more because the double is "extra".
That something is wrong with this "logic" comes out if we apply the same for a coin.
Then two coins giving at least one head should have $\frac{1}{2} + \frac{1}{2} = 1$ chance which is obviously absurd. In fact, we know that we have three desired cases from the four.
Or in an even simpler way, we have one undesired case, the double tail. So our chance is $\frac{3}{4}$.
Accepting the $\frac{11}{36}$ chance for a double six, prepares us for an even stranger fact among the triple throws. Three throws is half of the six outcomes and so now a desired triple six is even more seems to be half chanced. But the calculation gives again a surprise.
The total number of cases is easy as all possible triplets 6 x 6 x 6 = 216.
The desired cases have three groups. All six has one case, two sixes has 5 + 5 + 5 = 15 cases as three possible non sixes, and finally we have 3 x 5 x 5 = 75 cases for the three possible single six because this means two five combinations.
Altogether 91 outcomes, giving the desired chance as $\frac{91}{216}$ which is clearly less than half.
Observe, that we used addition in the second case counting, while multiplication in the last.
The reason is that the three positions of the possible five outcomes were excluding cases while the five outcomes for the two positions in the last situation were simultaneous. Also, the first was for an "or" combination while the second for an "and".
This goes into a similar logic of adding the chances for excluding "or" but multiplying for independent "and". For example the already used three even sides of a die are excluding and so their "or" is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$.
But the double six is the "and" of the two independent outcomes and indeed $\frac{1}{6} \text{ x } \frac{1}{6} = \frac{1}{36}$.
Beside these two rules we often use a trick of regarding not the desired, rather the undesired cases first as we mentioned above for the double tail.
For example, the previous 91 desired cases also can be obtained much simpler as the remaining cases of the undesired ones which are the triplets without any six and so their number is simply 5 x 5 x 5 = 125. And indeed, 91 = 216 – 125.
This level that we are now was about where the early probabilitists were.
They were actually gamblers who realized that simple common sense can be misleading and started to calculate chances by counting cases. But sometimes they fell back to false intuitions.
The most famous case was of Chevalier de Mere who knew the above three dice paradox and so knew that to get at least one six from three consecutive throws is also less than half chanced.
So if we play with double paybacks then betting on at least a six from three throws is a "looser".
But he also knew from playing a lot, that from four throws with same double pay we will win.
And indeed, there the possibilities are 6 x 6 x 6 x 6 = 1296 and the desired case number is:
1296 – 5 x 5 x 5 x 5 = 1296 – 625 = 671 > 625 and so the chance of a six containing four outcome is bigger than one without a six.

But then he used a logic to transfer these to sequences of double dice throws.
Here the six possibilities are replaced by 36 many and he thought that just as four consecutive trials overcame the half chance barrier, here 24 trials should do the same because $4 : 6 = 24 : 36$.
Unfortunately, he was wrong. The proper counting shows why! The total cases of 24 throws are 36^{24} and the undesired are 35^{24} and so the desired ones are $36^{24} - 35^{24} < 35^{24}$.
This can only be seen with a computer of course.
Amazingly, he realized that he was wrong by simply loosing and this was regarded as a contradiction in math.

Borel's Monkey

Now I'll turn to a much later and deeper paradox.
To drastically increase the possible outcomes, we can use a computer keyboard that has about fifty keys. Hitting the keyboard randomly, that is with eyes closed, we will have much smaller chances of getting each particular chosen key.
Emile Borel at the turn of the century envisioned a typewriter instead of a computer and operated by a monkey instead of we hitting it with eyes closed. The result is the same, namely we get a random infinite text if we or the monkey keeps hitting the keyboard for ever.
Borel claimed that such random infinite text must contain the Bible and so the monkey will type down the Bible too if we wait long enough!
This shocking result becomes very plausible from two simple assumptions.
The first is that if something has any small chance then repeating its trials, it will come about!
For coins or dice we find this very natural and I also mentioned a "kind of average expectability" how many trials should bring about a success. But I also mentioned that this is not a guarantee.
In theory, we can have arbitrary many trials without success but as the number of trials increases it seems less and less possible. So to say that trying infinitely we must succeed, is the minimum we can claim for sure. Amazingly though, this is sure for arbitrary small chanced outcomes.
A lotto jackpot for example has miniscule chances and yet we know that if we would play week after week for trillions of years then eventually we should hit the jackpot.
This eventualization of any chance should be called the Realistic Murphy's Law.
The original skeptical Murphy's Law claimed that the buttered bread always falls on the buttered side. This realistic version merely says that if we keep on dropping our bread then sooner or later we must be unlucky.
The second assumption we'll need is that a fix n number of repetition of an experiment is itself an experiment. The gamblers as we saw used this too but sometimes with false beliefs.
We'll only combine it with our Realistic Murphy's Law and so we won't care about exact chances.
So, all we assume is that every particular n long text has a chance from all possible n long ones and thus typing n long texts again and again blindly or by a monkey, we must achieve the particular text.
But these n long segments as trials and outcomes are already there in any infinite sequence of trials and outcomes of the single keystrokes. We simply have to regard n long windows there.
Now the Bible typing monkey starts to become not so freakish at all!
Lets count how many letters and symbols are in the Bible including spaces too. Say a billion.
So lets regard billion long windows in the infinite random text the monkey types.
Every possible outcome combination must come about in our windows.
So in particular, the Bible too will appear in our windows somewhere far away.
In fact, after it comes up we can continue watching our windows and the infinite future remains and so we will encounter the Bible again. Then again and again infinite many times!
Even more strangely, the Bible probably turned up already before our "first Bible" because we only checked our exact windows but it could have been there not starting at an exact window rather in-between and overlapping into the next window.

Exponentiality

The paradoxical nature of this result is due to a simple counting illusion.
With low outcome experiments like coin or dice we accept the repetitions to come about too. Double or triple coin outcomes are not freakish. But with the large fifty keystroke possibilities, we somehow feel that long combinations have almost no chance. And this is true but actually it's not the original outcome numbers that makes the repetition outcomes miniscule rather the sheer number of repetitions.
With the lowest two outcomes of a coin, regarding ten outcome segments we obviously have $2 \times 2 \times \ldots \times 2 = 2^{10} = 1024$ possible outcomes so these are not that many.
In few hours we can perform these and so in a day we can "expect" all possible segments to come about a few times. Particular twenty long segments would require thousands of hours that is years and thirty long ones would require thousands of years to appear.

Exactifying the earliers

First, the exactification of the chance expectations:
We feel that a p chanced outcome not only has to come about and so will come about infinitely, but that it will come about in same proportions as this p chance dictates it. So from thousand coin throws we should have half half heads and tails and from thousand dice throws we should have about $\frac{1000}{6} \approx 166$ many sixes.
Most amazingly, we also have a "side" intuition of this seemingly main proportional one!
Namely, we feel that the actual outcomes of the sixes not only will not exactly be the expected 166 in numbers in every thousand but they must be sometimes less sometimes more! Not strictly alternating of course but eventually going under and above the expectations.
This "dual oscillation" or "dual fluctuation" of the outcomes from the expectable ones is actually a much better intuition than the expectation itself. It regulates a sheer number without ratios.
But most amazingly, it works for any chosen length and even for increasing lengths.
So even just using the beginnings as increasing lengths, the dual oscillation of the outcomes from the expectables must be true as counted from the start. The future must always bring about too many or too few outcomes even counted from the start. Thus, the expectable beginning ratios must be approached from above and from under too. In particular, for the simplest case of coins:
To have always more heads than tails from the start would be "fishy" that is a strangeness.
This generalized dual oscillation then implies that we must have infinite many exact crossing overs when the outcome numbers are the exact expectable proportions from the start.
And this itself backstabs the expectation law! Indeed, after these points we can place arbitrary long imaginary windows and these must contain all combinations including full heads or tails.
So the oscillations after these will be arbitrary big. This shows that a vague "equalization" of the heads and tails is not true! In fact, the opposite happens. Only the division with the total length of the beginnings causes an approach to the probabilities.
The strongest claim is that in any infinite trial sequence this divided, that is the so called "relative" success frequency is approaching the $[\omega_i]$ chance of the ω_i outcome. So this seemingly physical Law Of Large Numbers is an experimental verifier of our assumed chances.
But the meaning of it is actually very mathematical!
Namely, such approach means that to error with at least a fix ε value can only happen finite many times. So if the occurrence of the ω_i outcome is k_i in the n long beginning then :

$$\left| \frac{k_i}{n} - [\omega_i] \right| \geq \varepsilon \text{ can only happen for finite many } n.$$

This of course means that for every ε we can give an $N(\varepsilon)$ number that after N we don't have such n any more, that is: If $n > N$ then $\left| \frac{k_i}{n} - [\omega_i] \right| < \varepsilon$.

Mathematics did , can and will never prove this claim simply because the concept of such outcome sequence is not a mathematical concept. But this can be seen from two sides too.
First of all, how can we know if our $[\omega_i]$ was correct. If a coin is leaded then the outcomes will not be approaching half obviously.
Secondly, we can always create outcome possibilities that defy such approaches.
The all head or all tail sequences will have $\varepsilon = \frac{1}{2}$ errors all the way through!
Even more sadly, this physical Law Of Large Numbers is totally outside the interest of physicists!
Abstract “fairness” of a coin is not a physical concept yet.
So this strongest Law Of Large Numbers is at present in a no man’s land.

Alternative visions

Randomness tries to describe these artificial mathematical sequences that defy the chances, while Probability tries to accept them and rather claim Laws Of Large Numbers that include these seemingly defying artificial sequences.
The fundamental intuition that an infinite trial sequence is a reality is rejected by physicists.
Some Probabilitists accept it but regard the artificial outcomes just as possible as any other.
And by today, even the Randomness researchers avoid a reality of random outcomes.
But if all heads are possible as outcome sequence then of course the Realistic Murphy’s Law is false too and Borel’s Monkey may not type down the Bible either!
My own beliefs changed in time too! I don’t regard an all head outcome sequence as simply impossible anymore. It is merely drastically harder to produce in random trials.
The first step is to get rid of the naïve time concept in our sequences.
The infinite many outcomes are spatial! We can try them simultaneously and maybe trying them enough many times they can become all heads too!

Short history of Randomness

The early chance calculations soon initiated tendency claims and mathematicians realized that such claim is possible from the initial $[\omega_i]$ chances. Simply because the segment chances are calculable and so those n segments that have at least ε errors can be counted and seen to be rare. This won’t assure a stopping of these in a concrete trial sequence of course, but it says at least something. Then to realize that something even stronger can be said was a modern development. But at this same time there remained those who cared not mainly about the chances rather whether those artificial non random or “strange” sequences that defy the physical Law Of Large Numbers can be blocked out. The non strange ones then of course would be the Random ones. Unfortunately, this was not the vision, rather the expectabilities of the Random sequences were pursued. Theoretically of course, strangeness or expectability are simple negatives.
Random sequences are those that avoid all strangenesses or equivalently, they obey all expectabilities. And we all have both intuitive strangenesses and expectabilities.
So in which direction should we go? For strangenesses or expectabilities?
The early choice was expectability and even more sadly with an obsession to the Law Of Large Numbers. I could attack this mistake a bit unfairly by simply saying, look at these:

$$H , T , H , T , H , T , H , \ldots .$$

$$1 , 2 , 3 , 4 , 5 , 6 , 1 , 2 , 3 , 4 , 5 , 6 , \ldots$$

The first obeys the chances of a fair coin as far as the physical Law Of Large Numbers is concerned and the second of a fair die.
But these outcomes are obviously impossible for a random trial sequence.

Interestingly, the Law Of Dual Oscillations is violated in both and for example in our second sequence the relative frequencies of the 6-es are always under $\frac{1}{6}$ except at the sixth, twelfth, and so on beginnings where they are exactly this value.
But this attack would not be fair because the person who was the biggest Randomness initiator and was truly obsessed with the law Of Large Numbers, Von Mises was very aware of this.
In fact, his grand idea can be started from these examples.
Von Mises said that the Law Of Large Numbers must not only be true for the full sequence but for any so called "observational" sequence created from it. Indeed, why shouldn't we be allowed to look at only every second place in our first sequence. This then should be random too and thus obey the Law Of Large Numbers. Which in our example will drastically fail.
To see the problems with Von Mises grand idea, we first have to realize that it can not be merely applied to fix locations as above. Indeed, suppose that in a random sequence of die outcomes where obviously we'll see infinite many . . . , 1 , 2 , 3 , 4 , 5 , 6 , . . . segments, we place after every such, say hundred extra 6-es. Obviously we ruined the randomness and this would indeed come out from Von Mises claim too but only if we allow as observational condition these segments and as observational places the next hundred outcomes.
That's great, so what's the problem?
The problem is that with this flexibility of conditional observational sequences, we could simply say look at the 6 outcomes themselves in a truly random sequence and obtain all 6-s.
Now we might argue that this "crazy" place selection was using the outcomes directly, while before we used outcome places after some outcomes. But this is merely "semantics" because we can regard the 6 outcomes in a random sequence and then collect all those beginnings that precede these outcomes. Then the places after these will still give the 6 outcomes.
So the deeper problem is that we used the concrete outcome of a full infinite random sequence.
To film a random sequence forever is impossible and so we can not collect these beginnings.
But we can collect infinite many potential beginnings in finite methods too.
This realization came from Church who was already trying to define effective collections of objects. To use this for binary segments as beginnings and thus define these B effective beginning sets seemed as a perfect solution. Then the Law Of Large Numbers must be true for all observational sequences defined by those places that are after such b from B.
Such effective collection of B can also be claimed as done by a machine and quite amazingly an alternative machine oriented definition of effectivity by Turing was exactly using such binary segments. So all looks very promising and then we can say that a sequence is random if all the observational sequences obtained from it obey the Law Of large Numbers.

The death blow came by Ville. He created sequences that qualify as random, that is all machine observational sequences obtained from them obey the Law Of Large Numbers. But they disobey the Law Of Dual Oscillations. Namely all relative frequencies of heads are under half.
For a long moment it seemed that the whole effectivity or machine idea is useless.
But then an even stronger use of the machines emerged that defines the information contained in segments. This so called AIT or Algorithmic Information Theory is one of the most heuristic ideas of the twentieth century. It defies the common sense that all n long outcome segments with a fair coin have a same $\frac{1}{2^n}$ chance. Or rather it leaves this assumption but projects a meaning behind our intuitive paradox that special outcomes like n many all heads, all tails, or alternating ones are somehow truly special. The common chance calculation seems to say that this intuition is purely subjective but AIT says that we are wrong! The specialness of segments is objective because machines can compress such segments that is give a shorter creation of them than merely listing them. Indeed, saying "one billion alternating head-tails" is shorter than one billion.
This gradual information content of the finite segments then becomes "alive" as the cause of the sharp black and white difference of infinite sequences as random or strange. The trick is again to use the beginnings and now regard their compressibilities.
But it was not smooth sailing. There was a glitch that prevented the instant success.

The grand idea of AIT came from two people, Solomonoff and Kolmogorov and the glitch was recognized by Martin Löf, a student of Kolmogorov. So he built on a totally different earlier probabilistic direction of Kolmogorov and injected the machine idea into that.
Soon AIT succeeded and turned out to give the same sequences as random.
If that is not enough, an earlier gambling approach of Ville was effectivized too, giving a third identical definition. Finally, a major simplification of Martin Löf's definition came by Solovay.
I will start with this simplest fourth but I have to give now the bad news too:
Behind these four identical definitions alternate possible definitions appeared and became just as plausible. So we don't have a fix concept of randomness today.
The most famous champion of AIT is Chaitin who spent a lifetime to claim that this is the true and only meaning of randomness.
All these are very personal to me because in high school I was obsessed with randomness and came to the Solovay definition on my own. Then for four decades I ignored the whole subject and came back to it just ten years ago.

The Solovay definition starts again with a B set of beginnings collected by a machine.
But now we use this B not as selector for observational places as Church did to apply Von Mises idea, rather use the B directly as potential beginnings for a sequence. This means that a sequence must build up from B, that is continue in B, that is have infinite many beginning from B.
We claim that for some special B this is impossible for a random sequence!
So random sequences can not continue in these B beginning options. So the random sequences can only have finite many beginnings from such B, that is they must stop in such B.
Our first thought could be that such special restriction is unnecessary because how could a random sequence continue in any machine collected option set anyway. But we are wrong because machines can collect a wide enough B so that random sequences can build up from B.
The extreme example is the following B set of beginnings:

H , T , HH , HT , TH , TT , HHH , HHT , HTH , HTT , THH, THT , TTH , TTT , HHHH , . . .

Amazingly, we listed all possible beginnings and so every sequence can continue in this B.
So we have to narrow our machine collections and this relies on chances.
The total chances in our above full set of beginnings is :

$$\frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \ . \ . \ . \ = 1 + 1 + 1 + \ . \ . \ . \ = \infty$$

So not surprisingly the condition of "narrowness" is that the chance total should be finite.

Thus the Law Of Stopping says that any random sequence must stop in a machine collected B if it is narrow, that is has finite chance total.
A reversal regards a continuation in some narrow and machine created B as some strangeness.
This is intuitively verifiable for every concrete effective B. But the major step is to claim that these strangenesses are sufficient to dodge all strangenesses in general.
So a sequence is random if it avoids all strangeness of the forms of continuing in some narrow and machine collected B.
So strangeness is a better concept than expectability, though formally they are negatives and actually even Martin Löf's version came from statistical expectabilities.

As I said, in spite of its simple beauty, this has not become the standard definition of randomness.
Not only its equivalent Martin Löf alternative is better known but a whole arsenal of alternative and non equivalent versions of that exist too. This shouldn't be a reason to avoid the aboves.
What's more, the strengthening of the Weak Law Of Large Numbers is deeply related to the Martin Löf and alternate versions.

The Weak Law Of Large Numbers

The total opposite of randomness is not to even deal with the individual infinite sequences and only talk about the beginnings. But then we can still stay physical and so carry out M many trials of n long b outcome segments. The at least ε erroring ones are those for which:

$$| \frac{k_i(b)}{n} - [\omega_i] | \geq \varepsilon$$

Where $k_i(b)$ is the number of ω_i occurrences in b.
This $\frac{k_i(b)}{n}$ proportion of the ω_i occurrence in b can be smartly abbreviated as $[b]_i$.

So we collect the $| [b]_i - [\omega_i] | \geq \varepsilon$ kind of n long b-s from our M many trials as an $E_i(n, \varepsilon, M)$ set. These are the at least ε erroring ones and we claim that the number of elements in E relative to M become rare as both n and M increase. In other words:

$$\frac{\text{num}E_i(n,\varepsilon,M)}{M} \to 0 \quad \text{If} \quad n \text{ and } M \to \infty$$

The use of b for segments still reflects our obsession with a single sequence and indeed we could use a single infinite sequence to test this claim by regarding different n long windows and counting M many such forward. Here the b-s are truly segments and not beginnings.
We could go one step further and regard the at least ε erroring beginnings not just for a fix n length rather for all beginnings in our sequence and claim a diminishing of their relative numbers as the single M increases. But this is a very big mathematical replacement of the previous claim.
To find the purely mathematical content behind our previous claim, we should first simplify our situation for $[\omega_1] = [\omega_2] = \ldots = [\omega_m] = \frac{1}{m}$ equal chanced outcome set.
Then we expect equal chances for the n long b segments too.
This also implies that we should have equal numbered $E_i(n, \varepsilon, M)$ sets for all i.
But the really important part is not this, rather that this number can be simply counted.
Namely, using $M = m^n$ and instead of physical trials we can just list all possible n segments.
Then the $E_i(n, \varepsilon, M) = E_i(n, \varepsilon, m^n) = E_i(n, \varepsilon)$ segments can be counted and:

$$[E_i(n, \varepsilon)] = \frac{\text{num}E_i(n,\varepsilon)}{m^n} \to 0 \quad \text{could be proved.}$$

So with equal initial chances, the possible n combinations are an equal outcome set and the mere number of the sub combinations with at least ε error should be the expected number of their outcomes.
To generalize this mathematical fact to non equal chances, we must avoid M again but now the m^n many possible combinations can not be an equal field of opportunities. Indeed, for example a loaded coin tossed twice has still four combinations but with different chances.
The first rule of chances that we used already is that independent outcomes combined by "and" will multiply their chances. A die has $\frac{1}{6}$ chance for six and a coin $\frac{1}{2}$ chance for a head. So to throw them together and get a six with head has $\frac{1}{12}$ chance.
Most of the time we repeat the same basic $\omega_1, \omega_2, \ldots, \omega_m$ outcomes and then the "and"-s are all possible $\omega_i\,\omega_j$ pairs. Repeating more n times we get the possible n segments denoted as b.

And if $b_1 , b_2 , . . . , b_n$ denote the elements of b then our product law says that :

$$[b] = [b_1] [b_2] . . . [b_n]$$

We might say "eureka" because these $[b]$ chances are the expectable chances of each b from the m^n many possible combinations. That's why counting them was not correct for $E_i (n , \varepsilon)$ either and instead somehow we should adjust num $E_i (n , \varepsilon)$ accordingly. Luckily there is a much more elegant way to avoid such dubious calculation and avoid num $E_i (n , \varepsilon)$ and even division with m^n altogether. So we can get the expectable $[E_i (n , \varepsilon)]$ ratio directly.
The road to find this $[E]$ ratio for any E set is based on our second chance law that we used. This is not multiplication rather addition and it is for "or" combination of excluding outcomes. Most trivial case was inside the basic outcome set but the same goes among the repetition segments. So quite "simply" :

$$[E_i (n , \varepsilon)] = \sum_{|[b]_i - [\omega_i]| \geq \varepsilon} [b]$$

So we add all the $[b]$ chances of those b segments for which $| [b]_i - [\omega_i] | \geq \varepsilon$.
And now we'll claim the $\rightarrow 0$ diminishing of this as n increases.

Though we have now unequal chances, we would only have to prove this for say $i = 1$ because we can reorder the basic outcomes. Also simplifying ω_1 as 1 we must prove that:

$$[E_1 (n , \varepsilon)] = \sum_{|[b]_1 - [1]| \geq \varepsilon} [b] \rightarrow 0$$

The really nice thing is that this non equal chanced generalization is just as easy to prove.
In fact, it will be further generalized. And the earlier combinatorical ratio limit would be instant consequence too. Indeed, for equal chances and for any E collection of special segments:

$$\sum_{b \in E} [b] = \sum_{b \in E} [b_1][b_2]...[b_n] = \sum_{b \in E} \left(\frac{1}{m}\right)^n = \frac{\sum_{b \in E} 1}{m^n} = \frac{\text{numE}}{m^n}$$

Now back to our task and to prove the above limit we'll use a basic trick by Chebishev:
Lets multiply our sum with $\frac{\varepsilon^2}{\varepsilon^2} = 1.$
Since ε is a fix value, we can take the numerator ε^2 inside the sum so we get there

$$\sum_{|[b]_1 - [1]| \geq \varepsilon} [b]\varepsilon^2$$

This is divided of course by ε^2 which is a fix value and so it would be enough if this numerator tends to 0.
The sum would increase by the following two alterations:
First we replace ε^2 by $([b]_1 - [1])^2$ and secondly we sum not just for the b-s that satisfy $| [b]_1 - [1] | \geq \varepsilon$ but for all n long b-s. So enough to prove that:

$$\sum [b]([b]_1 - [1])^2 = \sum_n \rightarrow 0$$

Finally, this would become trivial if we could show that actually $\sum_n = \dfrac{\sum_1}{n}$.

As an example for this claim using a loaded coin with $[\,H\,] = [1] = \dfrac{2}{3}$ and $[\,T\,] = [2] = \dfrac{1}{3}$:

$$\sum_1 = [1]\,(\,[1]_1 - [1]\,)^2 + [2]\,(\,[2]_1 - [1]\,)^2 = \frac{2}{3}\,(\,1 - \frac{2}{3})^2 + \frac{1}{3}\,(\,0 - \frac{2}{3})^2 =$$

$$\frac{2}{3} \times \frac{1}{9} + \frac{1}{3} \times \frac{4}{9} = \frac{6}{27} = \frac{2}{9}.$$

$$\sum_2 =$$

$$[11]\,(\,[11]_1 - [1]\,)^2 + [12]\,(\,[12]_1 - [1]\,)^2 + [21]\,(\,[21]_1 - [1]\,)^2 + [22]\,(\,[21]_1 - [1]\,)^2 =$$

$$\frac{4}{9}\;(\,1 - \frac{2}{3})^2 + \frac{2}{9}\;(\,\frac{1}{2} - \frac{2}{3})^2 + \frac{2}{9}\;(\,\frac{1}{2} - \frac{2}{3})^2 + \frac{1}{9}\;(\,0 - \frac{2}{3})^2 =$$

$$\frac{4}{9} \times \frac{1}{9} + \frac{2}{9} \times \frac{1}{36} + \frac{2}{9} \times \frac{1}{36} + \frac{1}{9} \times \frac{4}{9} = \frac{4}{81} + \frac{1}{162} + \frac{1}{162} + \frac{4}{81} = \frac{9}{81} = \frac{1}{9}.$$

Now we'll make a big generalization of our claim.

We'll associate some $<\omega_1> \;=\; <1>$, $<\omega_2> \;=\; <2>$, . . . , $<\omega_m> \;=\; <m>$ real number values to our outcomes too.
A physical meaning could be for example to paint our coin's head white and tail black but this would be still just a visual help for us and for an optical device the actual luminescence were the measured values say $<H> \;=\; 2.5 = \dfrac{5}{2}$ and $<T> \;=\; .5 = \dfrac{1}{2}$.
Quite apart from this practical direction, such real values could be used as 1 for the observed ω_i and 0 for all the others. Then the $[\,\omega_1\,] = [1]$, $[\,\omega_2\,] = [2]$, . . . , $[\,\omega_m\,] = [m]$ chance values could be avoided because : $[1]<1> + \; [2]<2> + \; . . + \; [m]<m> \;=\; [\,i\,]$.
So this new generalization gives any i choice at once and so we don't need the previous $i = 1$ assumption. For this of course the $\dfrac{k_i(b)}{n} = [\,b\,]_i$ proportion of the ω_i occurrence in b should be also universal. And it does become as $\dfrac{<b_1> + ... + <b_n>}{n} \;=\; <b>$.
Indeed, the numerator is exactly the number of ω_i outcomes in b if $<i> \;=\; 1$ and the others are 0.
Most amazingly, the $\sum_n = \dfrac{\sum_1}{n}$ claim remains true with any new $<\;>$ value system.

Then the two tricks to get $[\,i\,]$ and $[\,b\,]_i$ will have new meanings:

$[1]<1> + \; [2]<2> + \; . . + \; [m]<m> \;=\; e$ is the weighed sum of the values or in short the expectable value. This has no chosen i meaning any more and it is a kind of average of all possible values. But amazingly, the other new meaning $<b>$, is also a kind of average of the values but given by the n trials.

The new claim of the Law Of Large Numbers is still about the total chances of some segments diminishing, but now they are not those that error from a particular [i] rather those that error in average values from the expectable value.

$$\sum_{|<b>-e|\geq\varepsilon} [b] \quad \to \quad 0$$

The trick is again the same! Multiplying with $\frac{\varepsilon^2}{\varepsilon^2}$, taking the numerator ε^2 inside, replace it with $(< b > - e)^2$ then sum for all b-s and finally realize that:

$$\sum [b] (< b > - e)^2 = \sum_n = v_n = \frac{v_1}{n} = \to 0$$

Where:

$[b] = [b_1] \ldots [b_n]$

$$< b > = \frac{< b_1 > + \ldots + < b_n >}{n}$$

$e = [1] < 1 > + [2] < 2 > + \ . \ . + [m] < m >$ and

$v_1 = [1] (< 1 > - e)^2 + \ . \ . \ . + [m] (< m > - e)^2$

With our previous $[1] = \frac{2}{3}$, $[2] = \frac{1}{3}$, $< 1 > = \frac{5}{2}$, $< 2 > = \frac{1}{2}$ example:

$$v_1 = [1] (< 1 > - e)^2 + [2] (< 2 > - e)^2 = \frac{2}{3} (\frac{5}{2} - \frac{11}{6})^2 + \frac{1}{3} (\frac{1}{2} - \frac{11}{6})^2 =$$

$$\frac{2}{3}\left(\frac{4}{6}\right)^2 + \frac{1}{3}\left(\frac{8}{6}\right)^2 = \frac{2}{3} x \frac{4}{9} + \frac{1}{3} x \frac{16}{9} = \frac{24}{27} = \frac{8}{9}$$

$$v_2 = [11] (< 11 > - e)^2 + [12] (< 12 > - e)^2 + [21] (< 21 > - e)^2 + [22] (< 22 > - e)^2 =$$

$$\frac{4}{9} (\frac{5}{2} - \frac{11}{6})^2 + \frac{2}{9} (\frac{3}{2} - \frac{11}{6})^2 + \frac{2}{9} (\frac{3}{2} - \frac{11}{6})^2 + \frac{1}{9} (\frac{1}{2} - \frac{11}{6})^2 =$$

$$\frac{4}{9}\left(\frac{4}{6}\right)^2 + \frac{2}{9}\left(\frac{2}{6}\right)^2 + \frac{2}{9}\left(\frac{2}{6}\right)^2 + \frac{1}{9}\left(\frac{8}{6}\right)^2 = \frac{4}{9} x \frac{4}{9} + \frac{2}{9} x \frac{1}{9} + \frac{2}{9} x \frac{1}{9} + \frac{1}{9} x \frac{16}{9} =$$

$$\frac{16}{81} + \frac{2}{81} + \frac{2}{81} + \frac{16}{81} = \frac{36}{81} = \frac{4}{9}.$$

The natural idea to prove our general $v_n = \frac{v_1}{n}$ claim could by showing why the $< b >$ value system becomes such gradually as n grows.
Now $< b >$ contains n additions and a division by n so we would need two laws.
One for addition of value systems and one for multiplying a value system with a c constant.
Then the original ω value system could be added successively n times and finally the resulting sum multiplied by $\frac{1}{n}$.

This idea is consistent with how in fact Set Theory regards the b segments or so called tuples too. For example an $(\omega_3\ ,\ \omega_1\ ,\ \omega_4\ ,\ \omega_2\ )$ four membered tuple is actually regarded as: $(\,(\,(\omega_3\ ,\ \omega_1)\ ,\ \omega_4\ )\ ,\ \omega_2\ )$. So we build up everything from ordered pairs.
Set Theory even reduces the ordered pairs to unordered pairs but that trick is irrelevant now.

The definitions of adding two γ , δ value systems into a $\gamma + \delta$ is then regarding all $(\gamma_i\ ,\ \delta_j\ )$ or in short $\gamma_i\,\delta_j$ pairs as outcomes and assigning:

$$[\,\gamma_i\,\delta_j\,] = [\,\gamma_i\,]\,[\,\delta_j\,] \qquad \text{and} \qquad <\gamma_i\,\delta_j> = <\gamma_i> + <\delta_j>.$$

The definition of multiplying an α value system with a c to get $c\,\alpha$ is even simpler.
We keep α_i and $[\,\alpha_i\,]$ and merely alter $<\alpha_i>$ to become $<(c\,\alpha\,)_i> \ =\ c\ <\alpha_i>$.
Corresponding to the growing of the original ω value system, we want the growing of v_n to be gradual too. So then it shouldn't be defined using the original e rather e itself should be gradually changing too. So we have the definitions as:

$$e_\alpha \ =\ \sum\ [\,\alpha_i\,] < \alpha_i > \qquad \text{and} \qquad v_\alpha \ =\ \sum\ [\,\alpha_i\,]\ (\ <\alpha_i> - e_\alpha\ )^2$$

Our earlier e and v_1 then of course are e_ω and v_ω.
Trivially $e_{c\alpha} \ =\ c\,e_\alpha$ and not so trivially but easily $e_{\gamma+\delta} \ =\ e_\gamma \ +\ e_\delta$.

$$\text{Indeed,}\qquad \sum\ [\,\gamma\ \delta\,] < \gamma\ \delta > \ = \sum\ [\,\gamma\,]\,[\,\delta\,]\,(\,<\gamma> + <\delta>\,)\ =$$

$$\underset{1}{(\sum\ [\,\delta\,])}\ \underset{e_\gamma}{(\sum\ [\,\gamma\,]<\gamma>\,)}\ +\ \underset{1}{(\sum\ [\,\gamma\,])}\ \underset{e_\delta}{(\sum\ [\,\delta\,]<\delta>\,)}$$

For example with two members in both γ and δ :

$$[\gamma_1]\,[\,\delta_1]\,(<\gamma_1> + <\delta_1>)\ +\ [\gamma_1]\,[\,\delta_2]\,(<\gamma_1> + <\delta_2>)\ + [\gamma_2]\,[\,\delta_1]\,(<\gamma_2> + <\delta_1>)\ +$$

$$[\gamma_2]\,[\,\delta_2]\,(<\gamma_2> + <\delta_2>)\ =$$

$$\underset{1}{(\,[\,\delta_1] + [\,\delta_2]\,)}\,\underset{e_\gamma}{(\,[\,\gamma_1]<\gamma_1> + [\gamma_2]<\gamma_2>\,)}\ + \underset{1}{(\,[\,\gamma_1] + [\,\gamma_2]\,)}\,\underset{e_\delta}{(\,[\,\delta_1]< \delta_1> + [\,\delta_2]< \delta_2>\,)}$$

Thus for our built up $\beta \ =\ \{\,(\,..(\ \omega\ +\ \omega)\ +\ \omega\ +\ .\ .\ )\ +\ \omega\ \}\ \dfrac{1}{n}$ value system :

$$e_\beta \ =\ \{\,(\,..(\ e + e) + e +\ .\ .\ ) + e\ \}\ \frac{1}{n}\ =\ n\ e\ \frac{1}{n}\ =\ e.$$

Thus v_n becomes v_β.

From $e_{c\alpha} \ =\ c\,e_\alpha$ also trivially $v_{c\alpha} \ =\ c^2\ v_\alpha$.

But surprisingly: $v_{\gamma+\delta} = v_\gamma + v_\delta$.

This becomes logical from writing

$$v_\alpha = \sum [\alpha_i] (< \alpha_i > - e_\alpha)^2 = \sum [\alpha_i] (< \alpha_i >^2 + e_\alpha^2 - 2 < \alpha_i > e_\alpha)$$

Because then the previous re-summations work again.

Thus for $\beta = \{ (..(\omega + \omega) + \omega + . .) + \omega \} \frac{1}{n}$ now

$$v_\beta = \{ (..(v_1 + v_1) + v_1 + . .) + v_1 \} \frac{1}{n} = n\ v_1 \frac{1}{n^2} = \frac{v_1}{n}$$

Thus $v_n = \frac{v_1}{n}$ too.

We should explain why the letter v was used all along. It stands for "variance" and indeed it has the meaning how "varied" are the values away from the expectable vale.
In fact, the crucial point in the proof above that $e_\beta = e$ causing $v_\beta = v_n$ gives two interpretations for the variance meaning. One could be simply that for any α value system v_α should give a positive value of the left and right variances and that's why it had to use squaring.
But a deeper meaning of $v_\omega = v_1$ is that it gave a rate of how well the built up or repeated experiments will bring about the average $< b >$ values close to $e_\omega = e$.

The Strong Law Of Large Numbers

We needed five pages to prove this "grand" result that the number of erroring segments tend to zero in proportion to the number all possible segments. And this was supposed to replace the unprovable claim that such erroring beginnings should stop in a "real" random sequence.
But of course we wanted to avoid Randomness as a slippery subject.
And now we'll still get a slap in our face! Because Randomness is not avoidable! Certain trivial elements of it are always with us even through simple chances. So I will show now that this whole diminishing of the erroring segments really can't mean a stopping at all. Simply because there are many B beginning properties that diminish and yet "must" occur infinitely!
Our assumption for this "must" will be of course the Realistic Murphy's Law and the Borel's Monkey consequence. Instead of the Bible we accept that in a random sequence of 0-s and 1-s, there will be arbitrary long full 1 segments. And then the B beginning property that must occur infinitely is when a new 1-champion is achieved. That is, when a beginning has more 1-s at its end than ever before consecutively.
The s "state" of a beginning from our view is of course the number of consecutive 1-s at its end.
This can be 0 if it ends with 0. The r "rank" of a beginning or its "old champion level" is the largest consecutive number of 1-s prior to the growing end section.
Being champion then simply means that $s > r$.
Now a beginning continued with a 1 will increase the state by 1 and keep the rank.
Continuing with a 0 will make the state become 0 and the new rank will become the largest of the previous state and rank. The total number of possible beginnings always double after a single continuation and the 0 continuations can never be champions. So we need to show that the 1 continuations have less and less proportional champions.

The details are not that easy but the truth of the claim is quite plausible.

But to really show that diminishing chances don't imply stopping, I will give a much simpler example too not for beginnings rather independent trials of a coin.
The trick will be that we'll use increasing number of coins too.
We start with a single coin and flip it twice. Then we use two coins and throw them four times.
Then we use three coins and cast these together eight times. And so on.
So we use as repeats exactly the already mentioned "kind of expectability".
Indeed, the two coin outcomes are 4 in number so each have $\frac{1}{4}$ chances, the three coin ones are 8 in number so have $\frac{1}{8}$ chances, and so on.
We'll aim for all heads and as we know these repeats can not guarantee such outcome in the section. Instead, we might think that there is some kind of fix chance value attached to have at least one success and we might even think of the half value.
But this is obviously false because the first section of the two single trials has $\frac{3}{4}$ chance of getting at least one head. Indeed, we have three desired possibilities or negatively we have only one undesired one namely having both tails.

Then in the next section of the four double throws we have as undesired outcome sequence those outcomes where each is not the double head that is has three possibilities and $\frac{3}{4}$ chance.
So the chance of failure is $\frac{3}{4} \text{x} \frac{3}{4} \text{x} \frac{3}{4} \text{x} \frac{3}{4}$ and the chance of success is $1 - \frac{3}{4} \text{x} \frac{3}{4} \text{x} \frac{3}{4} \text{x} \frac{3}{4}$.

If you care to calculate this, you'll see that it is again more than half.
I instead will show it in a smarter way:

$$\frac{3}{4} \text{x} \frac{3}{4} \text{x} \frac{3}{4} \text{x} \frac{3}{4} < \frac{3}{4} \text{x} \frac{4}{5} \text{x} \frac{5}{6} \text{x} \frac{6}{7} = \frac{3}{7} < \frac{4}{8} = \frac{1}{2}$$

This same trick works in any group and so we have a bigger than $\frac{1}{2}$ chance of getting at least one full head in every section.
Repeating such bigger than half chanced trials means that we must succeed and then for the future sequence the same logic applies so we must have infinite many successes.
But a success in a group means at least one all heads occurring and so we must have infinite many such in our sequence without regarding the segments too.
So the $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{8}$, ... diminishing chances of the trials still guaranteed infinite successes. The obvious fact we can see is that

$$\frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \ . \ . \ . \ = 1 + 1 + 1 + \ ... \ = \infty$$

Now comes the real big surprise!
The opposite, that is finite sum assumption can give us "some kind" of assurance that the occurrence has to be only finite!
The real point behind this result is a purely mathematical fact nothing to do with chances.
Namely, this finite sum assumption is equivalent to the fact that if we go forward in our sum and regard only the left over members still ahead, then these in total must diminish!

First of all, to have a finite sum means that the beginning sums approach this value and so the leftover indeed has to become arbitrary small. But also in reverse, if a sum is infinite then the leftovers are always infinite and so to have diminishing leftovers automatically implies a non infinite sum.
Now we will give a chance meaning to these later and later leftover sums.
As we remember, adding of chances applies to excluding outcomes and they are pretty rare.
But the amazing luck is that this addition law, unlike the multiplication law for independent trials can lead to an inequality valid for all trials.
Indeed, if we try non excluding outcomes then the outcome cases for "or" will decrease and so the chance of the "or" has to be less than the added chances:

$$[\, t_1 \text{ or } t_2 \text{ or } .\,. \text{ or } t_m \,] \le [\, t_1] + [\, t_2] + .\,. + [\, t_m]$$

These are any trials not necessarily the excluding basic outcomes we denoted with Greek letters.

We'll use this for infinite many members, namely for the later and later members in an infinite $t_1 , t_2 , . . .$ sequence. So:

$$[\, t_{n+1} \text{ or } t_{n+2} \text{ or } . . . \,] \le [\, t_{n+1}] + [\, t_{n+2}] + . . .$$

The claim to have infinite occurrence in our $t_1 , t_2 , . . .$ trial sequence of course implies all the $(t_{n+1} \text{ or } t_{n+2} \text{ or } . . .)$ claims to be true and so we use a second inequality for such claim implications.
This is simply the fact the chance of a consequence is at least as big as the cause.
The chance of the humans being wiped out is at least as big as the chance of a meteor hitting the Earth.
So the $c_n = (t_{n+1} \text{ or } t_{n+2} \text{ or } . . .)$ consequence of the c claim that we have infinite many occurrence has also at least as big chance as the c cause:

$$[\, c \,] \le [c_n] = [\, t_{n+1} \text{ or } t_{n+2} \text{ or } . . . \,]$$

Applying this for all n and combining them with the previous inequalities:

$$[\, c \,] \le [\, t_1] + [\, t_2] + . . .$$

$$[\, c \,] \le [\, t_2] + [\, t_3] + . . .$$

$$[\, c \,] \le [\, t_3] + [\, t_4] + . . .$$

. . .

But these right sides are diminishing and so $[\, c \,]$ can only have one value 0.
So the chance of infinite occurrence has nil chance.
The previously explained identity of having finite sum or having diminishing leftovers can be now even more visually expressed as calling these as a fast diminishing.
So then the earlier also shown fact that diminishing doesn't imply stopping makes also sense.
Only the fast diminishing causes a stopping.
This makes our earlier explained Simplified Solovay Randomness perfect too. The finite chance total of the effective B beginning set means that the beginning chances are fast diminishing.
Of course, within Probability Theory we can not talk about stopping at all. But we obtained that the non stopping has 0 chance.

So the sequences where a fast diminishing property would not stop are very rare.
An even more visual and precise claim is that these sequences can be “covered” by arbitrary small chance totaled beginning sets. And this “a sequence being covered by a beginning set” simply means being continuation of some members from the beginning set.
So the infinite sequences are visualized as points while the beginnings as intervals containing all those “points” that are continuations. I call this whole visualization the Kolmogorov Road.
He created this to axiomatize Probability Theory in 1933 and Martin Löf effectivized this much later. He regarded the strangenesses as “effectively nil” sets. But this has to include also that the rate of diminishing is effective too. If we just require effective beginnings that diminish in any way then we get a new concept that could be called as merely “effective nil” sets. It won’t be equivalent to the Solovay Randomness and yet in many respect it is a better strangeness concept.

The weak Law Of Large Numbers claimed that the at least ε erroring beginnings diminish in portion. The bigger truth is that the total of these portions is finite, that is they diminish fast.
So then the infinitely at least ε erroring sequences have nil chance or are a nil set.
Then as a beautiful application of the Kolmogorov Road, we get that not only these infinitely at least fix ε erroring sequences are nil in size but all the ones that have any error infinitely.

This is the strong Law Of Large Numbers.

Indeed, we can regard a diminishing ε sequence like $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ . \ . \ .$ and for each the at least that much erroring sequences are coverable by arbitrary small totals. Meaning that for arbitrary small δ there is a beginning set totaling δ and covering the sequences.
So then we can also find coverings for the sequence of ε values with $\frac{\delta}{2}, \frac{\delta}{4}, \frac{\delta}{8}, \ . \ . \ .$ total values. Then all the erroring sequences will be covered by $\frac{\delta}{2} + \frac{\delta}{4} + \frac{\delta}{8} + \ . \ . \ . = \delta$ total.

The Martin Löf Randomness is the effectivization of this strong law that leaves open alternatives for the diminishing itself. The finite sum equivalence will only be true for one version and so whether this is an absolute concept of Randomness is questionable.

The Kolmogorov Road and coverings is the well accepted vision today as opposed to fast diminishings and continuations. I have to admit that this official vision has deep merit and I am very keen for making the nil cover vision conscious right from the start by the following paradox:

How many percentage of the numbers up to a million will not use the digit 7 ?
We all feel that excluding a single digit can not be that “big thing”. But observe the followings:
Up to 10 we indeed excluded 10 percent so kept 90.
But up to 100 we’ll exclude al the “seventies” and the seven ending ones so we get only
.9 x .9 = .81. Then this logic continues, so up to 1000 we have .9x.9x.9 portion left.
These products of .9 diminish rapidly and so we hardly get any opportunities left just by excluding the digit 7.
Randomness must be regarded in all the possible variations as free. Any restriction of the possible outcomes is actually a drastic size reduction of the possible sequences! The remaining many variations then all become measly, that is a nil set. This is the basic vision that then regards the effective or effectively nil sets as the universal strangenesses.
To collect some S set of infinite sequences and attach a chance value to S is not an easy task.
But the nil sets are an easy extreme!
If for every ε we have a B beginning set that has no more than ε chance total and yet every s sequence in S is covered by B that is s is a continuation of a b in B, then the S collection of sequences truly has only a chance smaller than any value that is nil.

The chance total of B is the barbaric step here because we add up chances of b beginnings that are different in length! A ten long binary beginning has $\frac{1}{2^{10}}$ chance among the ten long ones but to add up different such chances should be insane. Yet it works because such sum is still an upper bound on the chances of continuations. A single 0 as start means half chance for all continuations. A 110 triple beginning means an eight chance. The 0 starting ones could be continued up to three length and still we have same half chance and so among these we can add up these two correctly. Of course for infinite many beginnings there is no longest so the addition can not be directly justified like this.

The importance of nil sets does not abolish the importance of continuations and distinguished fast and slow diminishings. In fact, nil sets can throw more light on these too and as consequence show how weak is the weak Law Of Large Numbers:

We saw that the diminishing property of 1-champions allowed infinite occurrence.

This was meant in a randomness vision for individual sequences and so in Probability Theory should be avoided. But actually much more turns out to be true here! Namely, that “almost all” sequences must have infinite many champions! Which means that:

The sequences that stop having champions must be a nil set!

Indeed, to have only finite many champions means that we have a maximal, say one million 1-s infinitely often. Then after a point the next digits after these segments are a predictable 0.

This strangeness again implies a nil cover though not as easily as above for the excluded digits.

So then by analogy the weak law could also allow in theory that “almost all” sequences would error infinitely at least ε. So exactly the opposite were true of what this law intended to say.

The truly amazing fact is that we all have strong intuitions about what properties must stop in a random sequence. This was how I started to deal with the whole field.

To have a repeating beginning in a binary outcome sequence is not that strange:

0 , 0 , 1 , 0 , 0 , 1 , 1 , . . . But to have this continuing infinitely is impossible. And everybody knows that. We feel that the increasing beginnings make it harder and harder to repeat again.

But in truth it’s merely the chances that force the stopping.

The open infinite future of expectabilities and restricted past of beginnings is the two opposite tendency of the random outcomes. Everything can still happen but all strangenesses must stop.

The randomness definitions as universal strangenesses or expectabilities are merely the present.

Randomness is a bigger thing that will enter Set Theory through the choice functions.

Unfortunately, we have no clue how this could happen.