A NOTE ON COUNTABLE ADDITIVITY

Z. Šikić, Zagreb

Abstract: The “definition” of probability as long run frequency does not work. But even Kolmogorov used the “definition” heuristically because it makes the proof of the axioms of probability very easy - apart from the axiom of countable additivity. The common opinion is that limiting frequencies violate countable additivity due to very simple counterexamples. We prove that it is not the case. So limiting frequencies have no problems with any of the probability axioms. Their problem is that they may not exist, i.e. it is possible that an infinite sequence of experimental results has no limiting frequency (but c.f. the random sequence concept due to martingales, Martin- Löf and Schnor).

Probability is long run frequency?

The benefit of this "definition" is that it makes the proof of the axioms of probability very easy.

Even Kolmogorov used it in his "*Empirische Deduktion der Axiome*", a part of his "*Grundbegriffe der Wahrscheinlichkeitsrechnung*", in 1933.

n (A) is the number of trials successful for A, that occur in *n* trials

(1) $0 \leq \frac{n(A)}{n} \leq 1$ i.e. $0 \leq f_n(A) \leq 1$,

(2) $\frac{n(A)}{n} + \frac{n(-A)}{n} = \frac{n}{n} = 1$ i.e. $f_n(A) + f_n(-A) = 1$,

(3) if *A* and *B* exclude each other

$\frac{n(A \vee B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n}$ i.e. $f_n(A \vee B) = f_n(A) + f_n(B)$,

(4) $\frac{n(A \wedge B)}{n(A)} = \frac{n(A \wedge B)/n}{n(A)/n}$ i.e. $f_n(B|A) = \frac{f_n(A \wedge B)}{f_n(A)}$

But, as you know, there is a big problem here. Which f_n is the probability?

Is the probability of heads given by 100 tosses, f_{100}, is it given by 1000 tosses, f_{1000}, or what? How long should be the long run?

The longest run possible could circumvent the problem and the longest run possible is the infinite run. Hence:

$$Pr = \lim_{n \to \infty} f_n$$

The limiting frequency has no empirical content, because only finite sequences are observable.

Two infinite sequences which are observably identical (i.e. identical at a beginning) could have different limiting frequencies.

There is no connection between limiting frequencies and finite observable frequencies.

But we can still be interested in an abstract mathematical theory of limiting frequencies. It is mathematical foundation of probability we could be interested in and not its applications.

Pr defined as $\lim_{n\to\infty} f_n$ satisfies the probability axioms (1) - (4), because all f_n satisfy them.

But only if $\lim_{n\to\infty} f_n$ exists.

And it is easy to construct (non-random) examples of infinite runs with non-existent limiting frequencies.

HT   HT   HHTT   HHHHTTTT   HHHHHHHHTTTTTTTT   . . .

After HT we have blocks with 2^nHs and 2^nTs, for every $n > 0$.

After the *n*-th block, the frequency of heads is 1/2 (because every block has the same number of heads and tails).

If we stop in the middle of the *n*-th block, the frequency of heads will be:

$$\frac{\left(1+1+2+\ldots+2^{n-1}\right)+2^{n}}{2\left(1+1+2+\ldots+2^{n-1}\right)+2^{n}}=\frac{2\cdot 2^{n}}{3\cdot 2^{3}}=\frac{2}{3}$$

The frequency of heads oscillates between 1/2 and 2/3 i.e. the limiting frequency of heads does not exist.

(Is this possible for random sequences?)

If an infinite sequence of heads and tails, with existing limiting frequency, is random it should not be possible to design a successful gambling system against it.

It means, for example, that a gambler's place-selections of (e.g.) heads should have the same limiting frequency.

But there are infinitely many place-selection functions, i.e. subsequences of the sequence, with whatever limiting frequency you like.

The question is whether there is some natural set of place-selection functions to use.

Wald (1937) proved that given any countable set of place-selection functions, there are sequences that preserve limiting frequency relative to that class of functions.

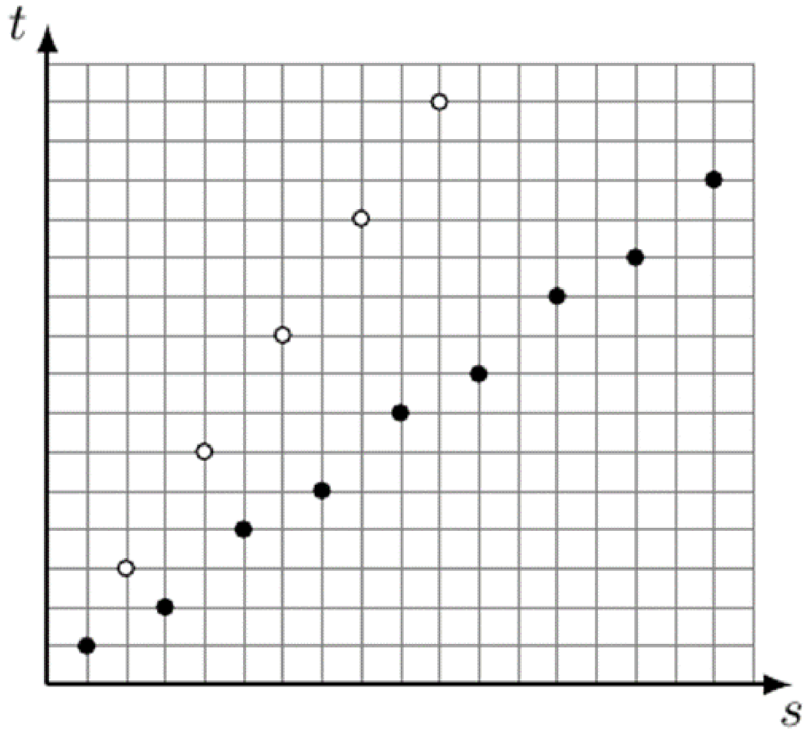
Church proposed the set of recursive subsequences of the sequence, as appropriate place-selection functions.

Hence, an infinite sequence of trials is random (von Mises' term was *Kollektiv*) only if

(1) it has limiting frequency,

(2) the limiting frequency remains the same in every recursive subsequence of the given sequence.

But suppose that results of repeated "head-tail" trials are space-time distributed in the following (non-random) way:



White: $(2,3) + (2,3) + (2,3) + (2,3) + \ldots$

Black: $(1,1) + (2,1) + (2,2) + (2,1) + (2,2) + \ldots$

If you were tossing the coin, your time sequence is:

TTH TTH TTH TTH ... $\longrightarrow$ 1/3

This is your probability of heads.

If I am inspecting the field of coins you tossed, my space sequence is:

TH TH TH TH TH TH ...$\longrightarrow$ 1/2

This is my probability of heads.

Should one answer be wright and another wrong?

I suppose that the above example, which is a folk knowledge today, was not a folk knowledge in von Mises' days. If it was, von Mises and Church would have added

(3) the limiting frequency should remain the same in every recursive reordering of the given sequence.

But why should an infinite sequence of trials produce *Kollektiv*.

Because it is a random sequence!

But do (1) – (3) define random sequences?

(Cf. martingales, Martin-Löf and Schnor.)

A further problem for frequentists is Kolmogorov's axiom of continuity which is equivalent to his theorem of countable additivity: If D_js exclude each other then

$$Pr(D_1 \vee D_2 \vee D_3 \vee \ldots) = Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots$$

"Axiom of continuity is essential for infinite fields of probability only, so it is **almost impossible to elucidate its empirical meaning**."

"Infinite fields of probability occur only as idealized models of real random processes. *We limit ourselves, **arbitrarily**, to only those models that satisfies the axiom* [because] this limitation has been found expedient in researches of most diverse sorts." Kolmogorov, 1933.

The common opinion, e.g. van Fraassen (1979) is that limiting frequencies violate countable additivity, due to counterexamples of the following kind.

Consider a lottery with ∞ tokens 1, 2, 3, 4 ... and let D_j = "token j is drawn".

Suppose that in an infinite sequence of draws none of the tokens is drawn infinitely many times. Then

$$Pr(D_j) = \lim_{n \to \infty} f_n(D_j) = 0, \text{ for every j.}$$

Hence it follows that

$$Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots = 0.$$

On the other hand

$$Pr(D_1 \vee D_2 \vee D_3 \vee \ldots) = 1.$$

Hence

$$Pr(D_1 \vee D_2 \vee D_3 \vee \ldots) = 1 \neq 0 = Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots$$

But why should

$$Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots = 0?$$

It is an indeterminate form $\infty \cdot 0$ which could be anything (if you still remember your first course in calculus).

As a matter of fact it is easy to prove that it is allways 1, as it should be according to the countable additivity.

Suppose, for example, that our infinite sequence of draws is as follows:

$D_4, D_1, D_1, D_2, D_4, D_1, D_7, D_2 \ldots$

The corresponding probabilities are:

$Pr(D_1)$ =	0/1	1/2	2/3	2/4	2/5	3/6	3/7	3/8	... → 0
$Pr(D_2)$ =	0/1	0/2	0/3	1/4	1/5	1/6	1/7	2/8	... → 0
$Pr(D_3)$ =	0/1	0/2	0/3	0/4	0/5	0/6	0/7	0/8	... → 0
$Pr(D_4)$ =	1/1	1/2	1/3	1/4	2/5	2/6	2/7	2/8	... → 0

– – – – – – – – – –

– – – – – – – – – –

(Note that I do not use the last column in the following argument.)

If we sum up all the columns, we get:

$Pr(D_1) + Pr(D_2) + Pr(D_3) + Pr(D_4) + \ldots =$

$\lim(\ 1/1,\ 1/2 + 1/2,\ 2/3 + 1/3,\ 2/4 + 1/4 + 1/4,\ \ldots) = \lim(\ 1, 1, 1, 1, \ldots)\ = 1$

It is the same for every other sequence of draws.

Namely, if F_j is the frequency of D_j, in the first n trials, then the sum in n-th column is $\sum \frac{F_j}{n}$, where $\sum F_j$ is the total number of draws in n trials. But this must be n. Hence $\sum \frac{F_j}{n} = 1$.

The same argument proves countable additivity.

Let D_1, D_2, D_3 ... exclude each other and define D as $D_1 \vee D_2 \vee D_3 \vee \ldots$.

Then $-D$, D_1, D_2, D_3 ... also exclude each other and by the preceding "column" argument (cf. the note above)

$$Pr(-D) + Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots = \lim(1, 1, 1, 1 \ldots) = 1$$

It follows that

$$Pr(-D \vee D) = Pr(-D) + Pr(D) = 1 = Pr(-D) + Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots \text{ i.e.}$$

$$Pr(D) = Pr(D_1) + Pr(D_2) + Pr(D_3) + \ldots$$

Conclusion:

Limiting frequencies satisfy the probability axioms (1) - (4), which was well known, but they also satisfy countable additivity.

Hence, limiting frequencies have no problems with probability axioms. Their problem is that they may not exist (hance random sequences; martingales, Martin-Löf, Schnor etc.).

By the way, the naive appeal to the law of large numbers: “the probability of sequences with no limiting frequencies is 0”, is not available to the frequentist. It presupposes that probability is defined independently of limiting frequencies (if it is not, the “law” becomes a complete triviality).

A. Church, 1940, *On the concept of a random sequence*, Bull. of the American Math. Society 46, pp. 130-135,

B.C. van Fraassen, 1979, *Relative frequencies*, pp. 133-166 in *Hans Reichenbach*: *Logical Empiricist*, ed. W. C. Salmon, Reidel,

A. N. Kolmogorov, 1933, *Foundations of the Theory of Probability*, Chelsea, 1956. (German original 1933.)

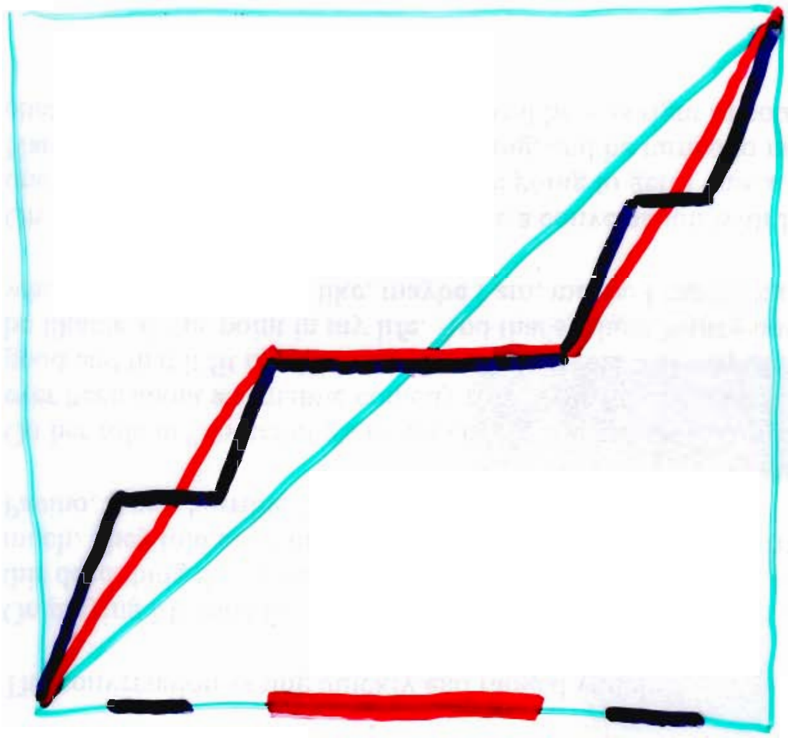
P. Martin- Löf, 1966, The Definition of Random Sequences, *Information and Control* 9, pp. 602-19.

R. von Mises, 1936, *Probability, Statistics and Truth*, Macmillan 1957. (German original 1936.)

C. P. Schnorr, 1971, *Zufälligkeit und Wahrscheinlechkeit: Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*, Springer.

Z. Šikić, 2013, *A Note on Probability, Frequency and Countable Additivity*, in *Philosophy in dialogue with sciences*, eds. L. Boršić, I. Skuhala Karasman, pp. 273-279, Institut of philosophy, Zagreb.

A. Wald, 1936, Sur la notion de collectif dans le calcul des probabilities, *Comptes rendus* 202, pp.180 -183.



f_0

f_1

f_2

$\vdots$

f mon. rost.

$$f' = \begin{cases} 0 & \overline{C} \\ \infty & C \end{cases}$$

$$f' = 0 \text{ (ss)} \to \int_0^1 f' = 0 \neq 1 = f\Big|_0^1$$

$$\boxed{\int_0^1 f' = 1 \; !!}$$

$\int_0^1 f' = 0$ <u>ale</u> $0 \cdot \infty = 0$??

196

ANALOGNO:

$$f' = \begin{cases} 0 & \bar{C} \\ \infty & C \end{cases}$$

$$|C_n| = 1 - \frac{1}{3} - \frac{2}{9} - \frac{4}{27} - \dots - \frac{2^{n-1}}{3^n} =$$

$$= 1 - \frac{1}{3} \sum \left(\frac{2}{3}\right)^{n-1} = 1 - \frac{1}{3} \frac{1 - \left(\frac{2}{3}\right)^n}{1 - \frac{2}{3}}$$

$$= \left(\frac{2}{3}\right)^n$$

$$f_n' = \frac{(1/2)^n}{(1/3)^n} = \left(\frac{3}{2}\right)^n$$

$$|C| = 0 = (1, 2/3, 4/9, 8/27, \dots)$$

$$f' = \infty = (1, 3/2, 9/4, 27/8, \dots)$$

$$0 \cdot \infty = (1, 1, 1, 1, \dots) = 1$$