The relative frequency interpretation of probability

Gábor Hofer-Szabó

Institute of Philosophy

Research Centre for the Humanities, Budapest

Email: szabo.gabor@btk.mta.hu

Project

I. Probability – concept, history, and interpretations

II. The relative frequency interpretation of probability

III. Relative frequency model

IV. Relative frequency interpretation

V. Problems

I. Probability

Probability

A brief history of probability:

- 1654: Pascal–Fermat correspondence (two *de Méré* paradoxes)
- 1665: Leibniz: *De Conditionibus* (conditional probability)
- 1670: Pascal: *Pensées* (the 'wager': maximal expected utility)
- 1713: Bernoulli: *Ars Conjectandi* (weak law of large numbers, non-additive probability)
- 1763: Bayes: *Doctrine of Chance* (bayesianism)
- 1812: Laplace: *Théorie analytique des probabilités* (central limit theorem)
- 1900: Hilbert's 6th problem
- 1933: Kolmogorov: *Grundbegriffe der Wahrscheinlichkeitsrechnung* (measure theoretical approach)

Probability

Measure theoretical probability:

- (Ω, Σ): **measurable space**
- $\mu : \Sigma \to [0, \infty]$: σ**-additive measure** on (Ω, Σ):
  - $\mu(\emptyset) = 0$
  - $\mu(\cup_i a_i) = \sum_i \mu(a_i)$, if $a_i \cap a_j = \emptyset$ for all $i \neq j$
- (Ω, Σ, μ): **measure space**
- **Probability measure**: $p(\Omega) := \mu(\Omega) = 1$
- (Ω, Σ, p): **probability measure space**

Probability

Standard interpretations of probability:

1. Classical interpretation (Laplace)
2. Logical interpretation (Keynes, Carnap)
3. Subjective interpretation (Ramsey, de Finetti)
4. Frequency interpretation (Reichenbach, von Mises)
5. *Propensity* interpretation (Popper)

Probability

What does it mean?

"The probability of getting a 6 with a fair die is 1/6."

1. **Classical:** "The ratio of the number of favorable and *equally possible* outcomes is 1/6."
2. **Logical:** "The proposition 'The dice is rolled' *confirmes* the proposition 'It comes up 6' in a degree of 1/6."
3. **Subjective:** "The *degree of rational belief* in the event that 6 will come up is 1/6."
4. **Frequency:** "The *relative frequency* of 6 in a long run of throws is 1/6."
5. ***Propensity*:** "The die has a *causal disposition* of coming up 6 in a degree of 1/6."

Probability

Salmon's criteria of interpretation:

- **Admissibility:** satisfy the probability axioms
- **Ascertainability:** be empirically accessible
- **Applicability:** serve as a 'guide to life' (Butler)

Probability

A better approach:

- Admissibility $\longrightarrow$ model
- Ascertainability $\longrightarrow$ interpretation
- ~~Applicability~~

II. Relative frequency interpretation

- “Probability is nothing else than *ratio*” (Venn, 1866)

Relative frequency interpretation

- Hans Reichenbach
  The Theory of Probability, 1949

- Richard von Mises
  Wahrscheinlichkeit, Statistik und Wahrheit, 1928
  Mathematical Theory of Probability and Statistics, 1964

Relative frequency interpretation

Von Mises' birthday paradox:

- Within a group of 366 people, the probability of there being at least two people having their birthday the same day is 1. For how many people is this probability 0.99?

Relative frequency interpretation

Von Mises' birthday paradox:

- Within a group of 366 people, the probability of there being at least two people having their birthday the same day is 1. For how many people is this probability 0.99?
- Solve

$$1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n} = 0,99$$

- Solution: $n \approx 55$

Relative frequency interpretation

The subject of probability theory:

- “is long sequences of experiments or observations repeated very often and under a set of invariable conditions. We observe, for example, the outcome of the repeated tossing of a coin or of a pair of dice; we record the sex of newborn children in a population; we determine the successive coordinates of the points at which bullets strike a target in a series of shots aimed at a bull’s-eye; or, to give a more general example, we note the varying outcomes which result from measuring “the same quantity” when “the same measuring procedure” is repeated many times. In every case we are concerned with a sequence of observations; we have determined the possible outcomes and recorded the actual outcome each time.” (von Mises, 1964 ,2)

III. Relative frequency model

Relative frequency model

Von Mises' two principles:

1. Stability of relative frequency
2. Principle of impossibility of a successful gambling system (*Prinzip vom ausgeschlossenen Spielsystem*)

Relative frequency model

1. Stability of relative frequency:

- "It is essential for the theory of probability that experience has shown that in the game of dice, as in all other mass phenomena which we have mentioned, the relative frequencies of certain attributes become more and more stable as the number of observations in increased." (von Mises, 1928, 12)

Relative frequency model

1. Stability of relative frequency:

- $x : \mathbb{N} \to \Sigma$: infinite sequence
- **Asymptotic relative frequency** of $a \in \Sigma$ in the sequence x:

$$r_x(a) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1_a(x_k)$$

(if it exists) where $1_a(x_k)$ is the **characteristic function**:

$$1_a(x_k) \;=\; \begin{cases} 1, & \text{ha } x_k \subseteq a \\ 0, & \text{ha } x_k \nsubseteq a \end{cases}$$

Relative frequency model

1. Stability of relative frequency:

- (Ω, Σ, p): probability measure space
- (Ω, Σ, p) has a **relative frequency model**: there exists a sequence $x : \mathbb{N} \to \Sigma$ such that for all $a \in \Sigma$:

$$r_x(a) \;=\; p(a)$$

Relative frequency model

Borel's theorem:

- $x \in [0,1]$
- Binary expansion: $x = 0.x_1x_2\dots$ where $x_i \in \{0,1\}$
- Relative frequency: $r_x(\{1\}) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} x_i}{n}$
- Borel's theorem (1909): $\lambda\left(\left\{x \,|\, r_x(\{1\}) = \frac{1}{2}\right\}\right) = 1$

Relative frequency model

2. Principle of impossibility of a successful gambling system (*Prinzip vom ausgeschlossenen Spielsystem*:)

- “For example, if we sit down at the roulette table in Monte Carlo and bet on red only if the ordinal number of the game is, say, the square of a prime number, the chance of winning (that is, the chance of the label red) is the same as in the complete sequence of all games. And if we bet on zero only if numbers different from zero have shown up fifteen times in succession, the chance of the label zero will remain unchanged in this subsequence . . .
  The banker at the roulette acts on this assumption of randomness and he is successful. The gambler who thinks he can devise a system to improve his chances meets with disappointment.” (von Mises, 1964, 108)

Relative frequency model

Collectives:

- Σ: algebra of properties
- $x : \mathbb{N} \to \Sigma$: infinite sequence
- x is a **collective** if
  - there exists $r_x(a)$ for all $a \in \Sigma$
  - $r_x(a)$ is invariant under place selection that is for all $a \in \Sigma$ and for all admissible place selection ϕ:

$$r_x(a) \quad = \quad r_{\phi(x)}(a)$$

Relative frequency model

Place selection:

- A typical reaction: Tornier (1933):
  “Ich glaube nicht, daß Versuche, die von Misessche Theorie rein mathematisch zu fassen, zum Erfolg führen können, und glaube auch nicht daß solche Versuche dieser Theorie zum Nutzen gereichen. Es liegt hier offentsichtlich der sehr interessante Fall vor, daß ein praktisch durchaus sinnvoller Begriff – Auswahl ohne Berücksichtigung der Merkmalunterschiede – prinzipiell jede rein mathematische, auch axiomatische Festlegung ausschließt. Wohl aber wäre es wünschenswert, das sich diesem Sachverhalt, der vielleicht von grundlegender Bedeutung ist, das Interesse weiter mathematischen Kreise zuwendet.”

Relative frequency model

Place selection:

- First idea: place selection = **Bernoulli sequence**
  - Copeland (1932), Reichenbach (1932), Popper (1935)
- $x : \mathbb{N} \to \{0, 1\}$: 0-1 sequence
- String: '01001'
- Place selection, ϕ_{01001}: if '01001' comes up in the sequence at $x_k, x_{k+1}, \ldots x_{k+l}$, then select element x_{k+l+1}
- **Bernoulli sequence:** if $r_x(1) = r_{\phi_{string}(x)}(1)$ for all strings

Relative frequency model

Place selection:

- Special case of Bernoulli sequence: **normal number**
  - Champernowne (1933)
- Let $x = 0100011011000\ldots$: binary numbers in ascending lexicographic order
- x is a Bernoulli sequence but it is constructable!
- Bernoulli sequence are not collectives in the sense of von Mises!

Relative frequency model

Place selection:

- Second idea: place selection = selection of an element x_k depends only on the elements $x_{<k}$
- $x : \mathbb{N} \to \{0, 1\}$: 0-1 sequence
- f_1, $f_2(x_1)$, $f_3(x_1, x_2)$, ... $f_{k+1}(x_1, x_2 \ldots x_k)$ ... :
  a sequence of $\mathbb{N} \to \{0, 1\}$ infinite 0-1 functions representing whether depending on $x_1, x_2 \ldots x_k$ the element x_{k+1} gets selected or not

Relative frequency model

Place selection:

- Second idea, equivalent formulation:
- $x : \mathbb{N} \to \{0, 1\}$: 0-1 sequence
- $f : \mathbb{N} \to \mathbb{R}$: arbitrary function
- Place selection, $\phi(x)$: pick the kth element of x if

$$c_k = 1, \text{ where } c_k = f(b_k), \; b_{k+1} = 2b_k + x_k, \; b_1 = 1$$

Relative frequency model

Place selection:

- Second idea, equivalent formulation:
- $x : \mathbb{N} \to \{0,1\}$: 0-1 sequence
- $f : \mathbb{N} \to \mathbb{R}$: arbitrary function
- Place selection, $\phi(x)$: pick the kth element of x if

$$c_k = 1, \ \text{where} \ c_k = f(b_k), \ b_{k+1} = 2b_k + x_k, \ b_1 = 1$$

- Kamke (1932): this definition is wrong!
  - Let $f(b_k) = x_{l(k)}$, where $l(k)$ is the least positive integer such that $2^{l(k)} > b_k$
  - In this case $\phi(x) = 1111111\ldots$, so $r_x(1) \neq r_{\phi(x)}(1)$
  - x is *not* a collective

Relative frequency model

Place selection:

- Church (1940): let f be *recursive function*
- Wald (1937): since there are countable recursive functions, therefore there are uncountable collectives
- Collectives cannot be constructed!

IV. Relative frequency interpretation

Relative frequency interpretation

Von Mises: probability theory is an **empirical science**

- “We take it as understood that probability theory, like theoretical mechanics or geometry, is a scientific theory of a certain domain of observed phenomena. If we try to describe the known modes of scientific research we may say: all exact science starts with observations, which, at the outset, are formulated in ordinary language; these inexact formulations are made more precise and are finally replaced by axiomatic assumptions, which, at the same time, define the basic concepts. Tautological (= mathematical) transformations are then used in order to derive from these assumptions conclusions, which, after retranslation into common language, may be tested by observations, according to operational prescriptions.

Relative frequency interpretation

Von Mises: probability theory is an **empirical science**

- Thus, there is in any sufficiently developed mathematical science a “middle part,” a tautological or mathematical part, consisting of mathematical deductions. Nowadays, in the study of probability there is frequently a tendency to deal with this mathematical part in a careful and mathematically rigorous way, while little interest is given to the relation to the subject matter, to probability as a science.

Relative frequency interpretation

Von Mises: probability theory is an **empirical science**

- This is reflected in the fact that today the “measure-theoretical approach” is more generally favored than the “frequency approach” presented in this book . . . Now, such a description of the mathematical tools used in probability calculus seems to us only part of the story. Mass distributions, density distributions, and electric charge are likewise additive set functions. If there is nothing specific in probability, why do we define “independence” for probability distributions and not for mass distributions? Why do we consider random variables, convolutions, chains, and other specific concepts and problems of probability calculus?” (von Mises, 1964, 43-44)

Relative frequency interpretation

Cramér's criticism:

- "The probability definition thus proposed would involve a mixture of empirical and theoretical elements, which is usually avoided in modern axiomatic theories. It would, e.g. be comparable to defining a geometrical point as the limit of a chalk spot of infinitely decreasing dimensions, which is usually not done in modern axiomatic geometry." (1946, 150)

Relative frequency interpretation

von Mises' response:

- "The 'mixture of empirical and theoretical elements' is, in our opinion, unavoidable in a mathematical science. When in the theory of elasticity we introduce the concepts of strain and stress, we cannot content ourselves by stating that these are symmetric tensors of second order. We have to bring in the basic assumptions of continuum mechanics, Hooke's law, etc., each of them a mixture of empirical and theoretical elements. Elasticity theory "is" not tensor analysis . . . the transition from observation to theoretical concepts cannot be completely mathematicized. It is not a logical conclusion but rather a choice, which, one believes, will stand up in the face of new observations." (1964, 45)

Relative frequency interpretation

The end of the frequency interpretation:

- Ville (1939): there exist gambling strategies (called *Martingales*) which cannot be represented as place selections
- von Mises: “I accept the theorem, but I do not see the contradiction.”
- 1937 Geneva Conference on the Theory of Probability: Fréchet’s criticism of the von Mises approach
- Renaissance of the frequency interpretation: Kolmogorov complexity (1965), randomicity (Martin-Löf, 1966)

V. (Alleged) problems

Problems

- **Relative frequency interpretation:**
  - Singular probability
  - The reference class problem
  - Irrelevancy of the finite relative frequency
- **Relative frequency model:**
  - σ-additivity and related issues

Problems

Singular probability:

- “’The probability of winning a battle’, for instance, has no place in our theory of probability, because we cannot think of a collective to which it belongs. The theory of probability cannot be applied to this problem any more than the physical concept of work can be applied to the calculation of the ’work’ done by an actor in reciting his part in a play.” (von Mises, 1928, 15)

Problems

Singular probability:

- “I regard the statement about the probability of the single case, not as having a meaning of its own, but as an elliptic mode of speech. In order to acquire meaning, the statement must be translated into a statement about a frequency in a sequence of repeated occurrences. The statement concerning the probability of the single case thus is given a *fictious meaning*, constructed by a *transfer of meaning from the general to the particular case*.” (Reichenbach, 1949, 376-77)

Problems

The reference class problem:

- “Let us assume, for example, that nine out of ten Englishmen are injured by residence in Madeira, but that nine out of ten consumptive persons are benefited by such a residence. These statistics, though fanciful, are conceivable and perfectly compatible. John Smith is a consumptive Englishman; are we to recommend a visit to Madeira in his case or not?” (Venn, 1866, 222-223)

Problems

The reference class problem:

- “If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result. This ambiguity has been called the problem of the reference class.” (Reichenbach, 1949, 374)

Problems

Irrelevancy:

- The finite relative frequencies are irrelevant to the asymptotic relative frequencies.
- These latter might not even exist.

Problems

von Mises' response:

- "The probability concept used in probability theory has exactly the same structure as have the fundamental concepts in any field in which mathematical analysis is applied to describe and represent reality. Consider for example a concept such as velocity in mechanics. While velocity can be measured only as the quotient of a displacement s by a time t, where both s and t are finite, non-vanishing quantities, velocity in mechanics is defined as the limit of that ratio as $t \to 0$, or as the differential quotient ds/dt. It makes no sense to ask whether that differential quotient exists 'in reality.' The assumption of its mathematical existence is one of the fundamentals of the theory of motion; its justification must be found in the fact that it enables us to describe and predict essential features of observable motions." (von Mises, 1964, 1-2)

Problems

Four claims:

(i) Frequencies do *not* form a σ-additive measure on every sequence.

(ii) Sequences with asymptotic relative frequency do *not* form a σ-algebra.

(iii) Sequences with asymptotic relative frequency do *not* form even an algebra.

(iv) Random sequences do *not* form an algebra.

Problems

Claim (i): Frequencies do not form a σ-additive measure on every sequence.

- Let $x_k \equiv k$ be the sequence of natural numbers. Here for all k the asymptotic relative frequency is 0, whereas for the countable union $\mathbb{N} = \{1\} \cup \{2\} \cup \ldots$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1_{\mathbb{N}}(x_k) \;\; = \;\; 1,$$

so frequencies do not form a σ-additive measure.

Problems

Why de Finetti did not like σ-additivity?
Dilemma:

- Either $\Sigma = \mathcal{P}(\Omega)$ but then no σ-additivity ($\longrightarrow$ mainstream)
- or σ-additivity but then $\Sigma \subset \mathcal{P}(\Omega)$ ($\longrightarrow$ de Finetti)

Problems

Non-Lebesgue measurable sets of $[0,1]$:

- Equivalence relation on $[0,1]$: $x \sim y$ iff $x - y \in \mathbb{Q}$.
- Equivalence classes: $[x] := \{x + q \in [0,1] \mid q \in \mathbb{Q}\}$
- E: one element from each equivalence class (Axiom of choice!)
- $E_q := E + q$ (modulo 1, for all $q \in \mathbb{Q}$)
- The sets E_q are countable, disjoint and their union is $[0,1]$
- Suppose that $p(E_q) = p$. Due to the translation invariance of the Lebesgue measure $p(E_{q'}) = p$ for all $q' \in \mathbb{Q}$
- If $p = 0$, then $\sum_q p(E_q) = 0$, if $p \neq 0$, then $\sum_q p(E_q) = \infty$
- But due to σ-additivity $\sum_q p(E_q) = p(\cup_q E_q) = p([0,1]) = 1$
- Hence E is not Lebesgue measurable

Problems

Claim (ii): Sequences with asymptotic relative frequency do not form a σ-algebra.

- Let $x : \mathbb{N} \to \Sigma$ be a sequence such that $r_x(a)$ does *not* exist.
- For any $n \in \mathbb{N}$ let $x^{(n)}$ be the following sequence:

$$x_{n'}^{(n)} = \begin{cases} x_n, & \text{if } n' = n, \\ \emptyset, & \text{if } n' \neq n. \end{cases}$$

- $r_{x^{(n)}}(a) = 0$ for any $x^{(n)}$
- But $x = \cup_n x^{(n)}$!

Problems

Claim (iii): Sequences with asymptotic relative frequency do not form even an algebra.

- Let $x = 110011110000111111111111000000000000\ldots$
- By the construction $r_x(1)$ does not exists. Now, let

$$\begin{aligned} y &= 101010101010101010101010101010101010\ldots \\ z &= 100110100101101010101010100101010101\ldots \\ y' &= 011001011010010101010101011010101010\ldots \\ z' &= 010101010101010101010101010101010101\ldots \end{aligned}$$

- Obviously, $r_y(1) = r_z(1) = r_{y'}(1) = r_{z'}(1) = \frac{1}{2}$
- But $x = (y \cap z) \cup (y' \cap z')$!

Problems

Claim (iv): Random sequences do not form an algebra.

- Consider any 0-1 sequence and change the 0s and the 1s.
- Independently of how randomicity is defined, the pointwise union of the two sequences will *not* be random.

Conclusions

- There are problems with the relative frequency interpretation of probability . . .

Conclusions

- There are problems with the relative frequency interpretation of probability . . . but other interpretations fair even worse!

References

- Champernowne, D. G. (1933). "The construction of decimal normal numbers in the scale of ten," *Jour. Lond. Math. Soc.*, **8**, 254-260.
- Copeland, A. H. (1932). "The theory of probability from the point of view of admissible numbers," *Ann. Math. Stat.*, **3**, 143-156.
- Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton: Princeton Un. Press.
- Kamke, E. (1932). "Über neuere Begründungen der Wahrscheinlichkeitsrechnung," in: Braithwaite R. B. (ed.), *Jahresbericht der Deutsche Mathematiker-Vereinigung*, **42**, 14-27.
- Mises, R. von (1928/51). *Wahrscheinlichkeit, Statistik und Wahrheit*, Berlin: Springer.
- Mises, R. von (1931). *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Wien: Deuticke.
- Mises, R. von (1964). *Mathematical Theory of Probability and Statistics*, NY: John Wiley.
- Reichenbach, H. (1932). "Axiomatik der Wahrscheinlichkeitsrechnung," *Math. Zeitschrift*, **34**, 568-619.
- Reichenbach, H. (1949). *The Theory of Probability*, Berkeley: University of California Press.
- Reichenbach, H. (1956). *The Direction of Time*, Berkeley: University of California Press.
- Venn, J. (1866). *The Logic of Chance*, London: Macmillan.
- Ville, J. (1939). *Étude critique de la notion de collectif*, Paris: Gauthiers-Villars.
- Wald, A. (1937). "Die Widerspruchsfreiheit des Kollektivbegriffes der

The two de Méré paradoxes

Division paradox:

- Two players are playing a fair game and they have agreed that whoever wins 6 rounds first gets the whole prize. The game stops when the first player has won 5, the second 3 rounds. How could the prize be devided?

The two de Méré paradoxes

Division paradox:

- Two players are playing a fair game and they have agreed that whoever wins 6 rounds first gets the whole prize. The game stops when the first player has won 5, the second 3 rounds. How could the prize be devided?
- Luca Pacioli, 1494: no solution
- Tartaglia, 1556: 2 : 1
- Pascal: 7 : 1

The two de Méré paradoxes

Two dice paradox:

- How can it be that
  - the probability of getting *at least one 6 in 4 rolls of a single die* is slightly less than 1/2,
  - whereas the probability of getting *at least one double 6 in 24 rolls of two dice* is slightly more than 1/2,

  since the chance of getting one 6 is six times as much as the probability of getting a double 6, and 24 is exactly six time as great as 4?

The two de Méré paradoxes

Two dice paradox:

- How can it be that
  - the probability of getting *at least one 6 in 4 rolls of a single die* is slightly less than 1/2,
  - whereas the probability of getting *at least one double 6 in 24 rolls of two dice* is slightly more than 1/2,

  since the chance of getting one 6 is six times as much as the probability of getting a double 6, and 24 is exactly six time as great as 4?
- **Solution**:
  - First case: $p = 1 - \left(\frac{5}{6}\right)^4 \approx 0,518$
  - Second case: $p = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,492$

The two de Méré paradoxes

Two dice paradox:

- **Intuition**:
  - 'Critical value': the number n such that $(1-p)^n$ exceeds $\frac{1}{2}$
    - First case: $p = 1 - \left(\frac{5}{6}\right)^4 \approx 0,518 \rightarrow$ critical value: 4
    - Second case: $p = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,492 \rightarrow$ critical value: 25
  - Intuition: proportionality rule of 'critical values'
  - De Moivre, 1718: the true law for the critical values is: $(1-p)^n = \frac{1}{2}$
  - The 'proportionality rule of critical values' holds approximately only if p is small: $n = \frac{-\ln 2}{\ln(1-p)} = \frac{-\ln 2}{p + p^2/2 + \dots}$

Relative frequency interpretation

- **Actual frequency interpretation:**
  - Probability = relative frequency in an actual sequence of trials
  - Venn, 1866: “Probability is nothing else than *ratio*”
- **Hypothetical frequency interpretation:**
  - Probability = relative frequency if the die would be tossed infinite many times
  - Reichenbach, von Mises

Relative frequency model

Three operations on collectives:

- **Mixing:** for adding probabilities
- **Partition:** for conditional probabilites
- **Combination:** for multiplying probabilities