„Analecta Cracoviensia" 54 (2022), s. 135–146
https://doi.org/10.15633/acr.5405

Adam Olszewski
Uniwersytet Papieski Jana Pawła II w Krakowie
https://orcid.org/0000-0003-3069-7518

Note on Probability of Logical Sentences and the Linda Problem[1]

The phrase 'probable sentence' usually takes on at least two meanings in natural language. According to the first meaning, a 'probable sentence' is one whose probability is greater than 1/2, in the range of real numbers from 0 to 1. According to the second meaning, a 'probable sentence' is any meaningful sentence that has a correct structure and its probability lies between 0 and 1. In the first case an improbable sentence is a sentence with probability equal to or less than 1/2. In the second case, practically 'improbable sentences' colloquially are those for which the probability is zero. Strictly, on the other hand, they are such expressions of language that are not correctly constructed meaningful sentences, i.e. they are not sentences in the logical sense. The present paper is devoted to this second notion of the 'probable sentence' applied to the formulas of the language of classical propositional logic as the specific type of sentences.[2]

---

[1] Some people have drawn my attention to the similarity of the presented conception of probability to Carnap's conception contained in R. Carnap, *Logical Foundations of Probability*, Chicago 1950. However, Carnap's conception concerns first-order language sentences and, additionally, natural language sentences, while my formulation concerns explicitly artificial language formulas of the propositional calculus. In J. Hintikka, *On Semantic Information*, in: *Information and Inference*, eds. J. Hintikka, P. Suppses, Dordrecht 1970, on pages 3–8 I found a probability function computed in a similar way to mine, but defined for a different domain, again on a set of natural language sentences as in Carnap and derived from his conceptual grid. I will address these issues in a separate article, as I reached Hintikka's work after the article was accepted for publication.

[2] The first meaning of 'probable sentence' was considered by the Scottish logician Hugh MacColl.

1. The Concept of Probability for Logical Sentences

Let us denote by the symbol Var(A) the set of all propositional variables occurring in a formula A, of the language of the Classical Propositional Calculus (PC). We denote the set of all such formulas by the symbol $Form_{PC}$. The following set of the schemas of axioms given by Jan Łukasiewicz forms a complete system of PC. The other connectives are added to this system by means of usual definitions.

(T1) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$ (Frege's syllogism)
(T2) $(A \to (B \to A))$ (simplification law);
(T3) $(\neg A \to \neg B) \to (B \to A)$ (contraposition law);
(MP) A, $(A \to B)//$ B (modus ponens rule).[3]

For any set X, the symbol $|X|$, denotes the cardinal number of this set.

Def. 2.0. [Probability Function][4]
A probability function *p* is any function defined on the formulae of a language closed under the Boolean connectives into the real numbers, satisfying the conditions (K1—K4). It will be called *finitely additive probability function* and for any A, B $\in FORM_{PC}$:

(K1) $0 \leq p(A) \leq 1$;
(K2) $p(A) = 1$, for some A;
(K3) $p(A) \leq p(B)$, if $\{A\} \vdash_{PC} B$;[5]
(K4) $p(A \vee B) = p(A) + p(B)$, if $\{A\} \vdash_{PC} \neg B$.[6] □

[3] Please note that we have no substitution rule here, instead we have an infinite set of axioms.

[4] The symbol $\vdash_{PC}A$ means that the formula A is a theorem of PC. Whereas a symbol $\models_{PC}A$ means that formula A is a tautology of PC. D. Makinson, *Bridges from Classical to Nonmonotonic Logic*, London 2005, p. 113.

[5] Symbol $\vdash_{PC}$ denotes the derivability relation for PC. Due to the completeness theorem this relation can be replaced by semantic relation of entailment i.e. $\models$ PC.

[6] Or equivalently (K4') $p(A \vee B) = p(A) + p(B)$, if $\vdash_{PC} \neg(A \wedge B)$.

Def. 2.1.[7]
Let $A \in Form_{PC}$, $m, n \in N$ and $n \neq 0$. Then a function PR called the *probability function* is defined by the following conditions:

A. PR: $Form_{PC} \rightarrow [0, 1]$, where $[0, 1]$ is the interval of real numbers from 0 to 1;
B. $PR(A) = m/n$, where m is the number of rows in the last column of the truth-table for formula A that contain 1s, and n is the number of all rows in the last column of the table.[8] □

Theorem 2.2.
The function PR is a *finitely additive probability function*. □

Proof:
We need to show that PR satisfies, for any $A, B \in Form_{PC}$, the conditions:

(P1) $0 \leq PR(A) \leq 1$;
(P2) $PR(A) = 1$, if $A \in TAUT_{PC}$;
(P3) $PR(A) \leq PR(B)$, if $\{A\} \vdash_{PC} B$;
(P4) $PR(A \vee B) = PR(A) + PR(B)$, if $\{A\} \vdash_{PC} \neg B$.

Ad. (P1) The condition follows directly from Def. 2.1.

Ad. (P2) The condition follows directly from the definition of a PC-tautology, as a formula that for any Boolean valuation takes the logical value 1.

Ad. (P3) For the proof let us suppose that $\{A\} \vdash_{PC} B$. By the deduction theorem for PC we have $\vdash_{PC} (A \rightarrow B)$. By virtue of the completeness theorem for PC we

[7] Stefan Mazurkiewcz (1888–1945), a member of the Lvov–Warsaw school of mathematics, has given two definitions of logical probability: def. 1 (1932): p(x, U) – understanding: "[P]robability of the sentence A relative to the system (deductive system AO.) U". Def. 2 (1933): p(X, Y) – understanding: "Probability of deductive system X relative to the deductive system Y". A set of formulas X is a deductive system iff X = C(X), for a given consequence operation C (iff is short for "if and only if"); cf. S. Mazurkiewicz, *Przyczynek do aksjomatyki rachunku prawdopodobieństwa. Zur Axiomatik der Warscheinlich-keitsrechnung*, "Sprawozdania z Posiedzeń Towarzystwa Naukowego Warszawskiego. Wydział III nauk matematyczno-fizycznych" 25 (1933) no 1–6, p. 1–4, and S. Mazurkiewicz, *Über die Grundlagen der Wahrscheinlichkeitsrechnung I*. "Monatshefte für Mathematik und Physik" 41 (1934), p. 343–352; https://doi.org/10.1007/BF01697866.

[8] The box symbol indicates the end of a theorem, corollary, definition or proof.

get $\models_{PC}(A \rightarrow B)$ and the formula is a tautology of PC. Let us take Var(A → B) = $\{p_1, p_2, \ldots, p_n\}$. Thus, in each row of the table for (A → B), that is, for any valuation v of the propositional variables $p_1, \ldots, p_n$: if v(A) = 1, then v(B) = 1, and from this we have: PR(A) ≤ PR(B).

Ad. (P4) Let us suppose that: $\{A\} \vdash_{PC} \neg B$. Therefore by deduction theorem for PC we have $\vdash_{PC} (A \rightarrow \neg B)$. By the virtue of the completeness theorem for PC (A → ¬B) is a tautology of PC. In each row of the table i.e. for any Boolean valuation of the formula, if v(A) = 1, then it must be that v(B) = 0, what gives $v(A \wedge B) = 0$, and $PR(A \wedge B) = 0$. Therefore $PR(A \vee B) = PR(A) + PR(B) - PR(A \wedge B) = PR(A) + PR(B)$. □

Def. 2.3. (Conditional probability)[9]
The function PR_A is called a *function of conditional probability* if: $PR_A(B) = PR(A \wedge B)/PR(A)$, if PR(A) ≠ 0. □

Corollary 2.4.
The PR function does not meet equality: $PR(p \rightarrow q) = PR_p(q)$, for different propositional variables p, q. □

Proof:
We have PR(p → q) = 0.75, while $PR(p \wedge q) = 0.25$ and PR(p) = 0.5, then $PR(p \wedge q)/PR(p)$, (0.25/0.5) = 0.5. □

Hence we have the conclusion that for the function PR it is not generally true that the probability of the conditional PR(q → q) equals the conditional probability $PR_q(p)$, but if v(A) = 1, for any v, then $PR(A \rightarrow q) = PR_A(q)$.[10] For any formulas A and B we have PR(A → (A → B)) = PR(A → B) and additionally:

Corollary 2.5.
If PR(A) ≠ 0, then $PR_A(A \rightarrow B) = PR_A(B)$. □

Proof:
$PR_A(A \rightarrow B) = PR(A \wedge (A \rightarrow B))/PR(A) = PR(A \wedge B)/PR(A) = PR_A(B)$. □

[9] Cf. D. Makinson, *Bridges from Classical to Nonmonotonic Logic*, p. 122.

[10] This problem is sometimes characterized by equation: CP = PC (*conditional probability = probability of a conditional*); cf. D. Makinson, *Bridges from Classical to Nonmonotonic Logic*, p. 118–120.

Therefore, if $PR(A), PR(B) \neq 0$, then $PR(A) \cdot PR_A(A \to B) = PR(A \wedge (A \to B)) = PR(A \wedge B) = PR_A(B) \cdot PR(A) = PR_B(A) \cdot PR(B)$.

Theorem 2.6.
For any formulas of PC-language A, B we have: $PR_A B \leq PR(A \to B)$. □

Proof:
Let $Var(A \wedge B) = n$; $PR(A \wedge B) = i/2^n$; $PR(A \to B) = k/2^n$; $PR(A) = (2^n - (k - i))/2^n$. It follows that $i \leq k$, $i \leq 2^n - (k - i)$ and $(i/2^n)/((2^n - (k - i))/2^n) \leq (k/2^n)$, because $i = (k - (k - i))$, so $(k - (k - i))/2^n)/((2^n - (k - i))/2^n) \leq (k/2^n)$. □

Corollary 2.7.
Let $|-_{PC} A$ and $B, C \in Form_{PC}$, if $A = (B \to C)$, then $PR_B(C) = PR(B \to C)$. □

Proof:
From theorem 2.5. we have that, if $k = 2^n$, where $n = |Var(A)|$, then $((2^n - (k - i)) = (2^n - (2^n - i)) = i$. and $(k - (k - i))/2^n)/((2^n - (k - i))/2^n) = (i/2^n)/(i/2^n) = 1$. □

We have also:

Corollary 2.8.
For any formulas A, B, if $PR(B) \neq 0$ and $PR(A) = 1$, then $PR_A(B) = PR(A \to B)$. □

Proof:
$PR_A(B) = PR(A \wedge B)/PR(A) = PR(A \wedge B) = PR(B) = PR(A \to B)$. □
Bayes' theorem holds for the PR function:

Corollary 2.9.
If $PR(A), PR(B) \neq 0$, then $PR_A(B) = (PR_B(A) \cdot PR(B))/PR(A)$. □

Proof:
$PR_A B = PR(A \wedge B)/PR(A) = PR(B \wedge A)/PR(A) = (PR_B(A) \cdot PR(B))/PR(A)$. □

Somewhat surprising is the following fact:

Corollary 2.10.
For any formulas A, B such that $|\{v: v(A) = 1 \text{ and } v(B) = 0\}| = |\{v: v(A) = 0 \text{ and } v(B) = 1\}|$ holds $PR(A \rightarrow B) = PR(B \rightarrow A)$. □

Proof:
The number k of valuations v, such that $v(A \rightarrow B) = 0$ equals the number of valuations v, for which $v(B \rightarrow A) = 0$. The number of all valuation of both formulas $(A \rightarrow B)$ and $(B \rightarrow A)$ equals 2^n, where $n = |Var(A \rightarrow B)|$. Then $PR(A \rightarrow B) = (2^n - k)/2^n = PR(B \rightarrow A)$. □

This concept of the logical probability can be developed in various ways. One direction for the development are the many-valued logics, for example to Łukasiewicz's logics, which seems promising. The other way is the propositional calculus with quantifiers binding the propositional variables but in a different way than it was done by Łukasiewicz and Tarski.

2. Some Properties of PR Function

We will now deal with the properties of this probability function. The following lemma holds:

Lemma 3.1.
For any $A, B \in Form_{PC}$: if $\vdash_{PC} (A \equiv B)$, then $PR(A) = PR(B)$. □

Proof:
Let us suppose that $\vdash_{PC} (A \equiv B)$. Then for any valuation of formulas v, in the truth-table for $A \equiv B$, holds: $v(A) = 1$ iff $v(B) = 1$, and formulas A and B take the same truth value precisely in the same rows of the truth-table. □

The converse implication does not hold, because for example it is not true that $\vdash_{PC} (p_1 \equiv p_2)$, but $PR(p_1) = PR(p_2)$. Let us take the relation $\cong$ defined on the set $Form_{PC}$:

Definition 3.2.
For any formulas A and B: $A \cong B$ iff $PR(A) = PR(B)$. □

The quotient algebra built on the relation is distinct from the Lindenbaum algebra built on by the relation of equivalence between formulas.[11] Let us first show that the first of these relations is indeed an equivalence.

Lemma 3.3.
Relation $\cong$ is an equivalence on the set $Form_{PC}$. □

Proof:
For any $A, B, C \in Form_{PC}$, $A \cong A$, because $PR(A) = PR(A)$. Let us suppose that $A \cong B$ and $B \cong C$, what means that $PR(A) = PR(B)$ and $PR(B) = PR(C)$, and from this it follows that $PR(A) = PR(C)$, i.e. $A \cong C$. From $A \cong B$, we get immediately $B \cong A$, because equality is symmetric. □

In the following we will denote the classes of abstraction also by small letters x, y, z, …; and the quotient algebra $Form_{PC}/\cong$ by the symbol $\mathbb{P}$. Unfortunately our equivalence relation does not give a congruence relative to the conjunction. This is shown by the following example. Let us take that $p \cong p$, $(p \wedge q) \cong (r \wedge q)$, and the operation defined $[A]\cong \wedge [B]\cong: = [A \wedge B]\cong$, then we have: $[p]\cong \wedge [p \wedge q]\cong = [p \wedge p \wedge q]\cong$ and $[p]\cong \wedge [r \wedge q]\cong = [p \wedge r \wedge q]\cong$; but $PR(p \wedge p \wedge q) = 0.25 \neq PR(p \wedge r \wedge q) = 0.125$.

Lemma 3.4.
The function PR has the following properties:

a. $PR(\neg A) = 1 - PR(A)$;
b. $PR(A \wedge \neg A) = 0$;
c. $PR(A \wedge B) \leq PR(A) \leq PR(A \vee B)$;
d. $PR(A) > PR(B)$ iff $PR(\neg B) > PR(\neg A)$;
e. $PR(A) = PR(A \wedge B) + PR(A \wedge \neg B)$;
f. $PR(A \vee B) = PR(A) + PR(B) - PR(A \wedge B)$;
g. $PR(A \vee B) + PR(A \wedge B) = PR(A) + PR(B)$;
h. $PR(A \wedge B) = PR(A) \cdot PR_A(B)$. □

Proof:
We will only give the proofs of selected points.

[11] We denote the abstraction classes of these relations by the symbols $[A]\cong$ and $[A]\equiv$ respectively.

Ad. c. $PR(A \wedge B) = PR(A) \cdot PR_A(B) \leq PR(A)$, because $PR_A(B) \leq 1$.
Ad. h. By the def. 2.3. $PR(A) \cdot PR_A(B) = PR(A) \cdot PR(A \wedge B)/PR(A) = PR(A \wedge B)$.

3. Application to the Linda problem – Reminder

Now, in order to demonstrate the usefulness of the introduced concept of probability, we will apply it to the solution of Linda's problem, a well-known one in cognitive psychology. We will first briefly recall it.[12]

The so-called Linda's experiment was carried out by Kahneman and Tversky (KT) and consisted in the following short description of a woman named Linda:

> Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. […] The subjects were asked which of the following two propositions is more probable. 'Linda is a bank teller'. (T) or 'Linda is a bank teller and is active in the feminist movement' $(T \wedge F)$[13].

The respondents were then asked which of the sentences was more likely to be a scenario for Linda's future life. The majority of the surveyed subjects indicated as more probable the conjunction of sentences:

- $(T \wedge F)$ Linda is a bank teller and a feminist.

than one of its conjuncts:

- (T) Linda is a bank teller.

That is, they considered that $P(T) < P(T \wedge F)$, contrary to the axioms of the probability calculus, which states, in this case, that $P(T \wedge F) \leq P(T)$. This became, among other things, the basis for the creation of the concept of representativeness heuristics by KT (1983).

[12] Readers interested in some critique of Kahneman's approach are referred to the papers cf. A. Olszewski *A Few Comments on the Linda Problem*, "Organon F" 24 (2017), p. 184–195 and A. Olszewski, *Linda Problem – the Tame Solution in Question*, "Analecta Cracoviensia" 51 (2019), p. 209–217.

[13] A. Tversky, D. Kahneman, *Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, "Psychological Review" 90 (1983), p. 297.

4. Linda Problem in Another Formulation

Despite the fact that Linda's characterization, quoted above, is a collection of several simple sentences, we will treat it, for simplicity, as an atomic sentence[14] and denote it with the symbol D (from description). For the other simple sentences we use respectively symbols: T (Linda is a bank teller), F (Linda is a feminist) and $T \wedge F$ (Linda is a bank teller and a feminist). According to the description of the KT experiment, subjects were asked to answer which one of the two sentences: T and $T \wedge F$; was more probable. It should be emphasized that this question was posed with Linda's description D (or Linda characteristics) as the relevant assumption. Using logically probable formulas, we can formulate the experimental question like this:

- Formulation 1. Which of the following sentences is more probable?

Sentence 1: $(D \rightarrow T)$?
Sentence 2: $(D \rightarrow T \wedge F)$?

Treating these schemas of sentences as logically probable formulas, we can calculate PR function values for them. And so we have: $PR(D \rightarrow T) = 0.75$; $PR(D \rightarrow T \wedge F) = 5/8 = 0.625$, assuming that the sentences D, T, F take one of the logical values 1 or 0. The answer, therefore, is as expected from KT, because the second case is less probable than the first.

However, when we look at the formulation 1, we see one peculiar point. The sentences T and F are treated, as if they play an equal or equivalent role in the experiment, which does not occur and contradicts the data of the problem. This is because sentence F enters into a different kind of dependency on D than sentence T, precisely because we have to take into account the implication $(D \rightarrow F)$, which is precisely to express the manner of this dependence of the sentence F on D, in contrast to the T. Intuitively this implication could mean: "It is quite probable that Linda, such as in the description D given, is a feminist." In view of this, as it seems, this implication is the premise assumed in question 2 from the above. Taking this into account we have the following formulation II:

[14] We will take D to be the conjunction of simple sentences forming D. This move does not change the substance.

- Formulation 2. Which of the sentences is more probable?

Sentence 1: (D → T)?
Sentence 2.2.: (D → F) → (D →T ∧ F)?

The logical probability of the first sentence remains the same: PR(D → T) = 0.75.[15] But the logical probability of the second sentence 2.2. is modified: PR((D → F) → (D → T ∧ F)) = 7/8 = 0.875. And the second case turned out to be more probable, as we expected.

To better understand the sense of this second formula, let us note that the following are the tautologies of PC:

- ((D → F) → (D →T ∧ F)) ≡ (D → (F → T ∧ F));
- ((D → F) → (D →T ∧ F)) ≡ ((D → F) → ((D → T) ∧ (D → F))).

Using the definition of conditional probability, we can give yet a third formulation.

- Formulation 3. Which of the sentences is more probable?

Sentence 1: (D → T)?
Sentence 2.3.: (D → T ∧ F) under the condition (D → F)?

The logical probability of the first sentence remains the same: PR(D → T) = 0.75. But the conditional probability $PR_{(D \to F)}(D \to T \wedge F)$ = PR((D → F) ∧ (D → T ∧ F))/PR(D → F) = 0.833(3). This result is inconsistent with the KT results and confirms our hypothesis that it is reasonable to consider (T ∧ F) as more probable then T.

Abstract

Note on Probability of Logical Sentences and the Linda Problem

This paper presents a logical concept of probability which seems to be obvious, as it is, but the author is not aware of any elaboration of a developed studies on the issue or of any special philosophical application of it. Such a probability of the formula A,

[15] If we take the sentence (D → F) → (D → T) instead the sentence (D → T), we get: PR((D → F) → (D → T)) = 0.875. But this extra assumption is unnecessary. We treat this case as erroneous *per excessum* and irrelevant in the aspect of Linda problem currently under consideration.

of the language of the propositional logic, is the quotient of the number of Boolean valuations of formula A of the classical propositional calculus, which takes the logical value 1, to the number of all Boolean valuations of such a formula A. An application of this concept of logical probability to the solution of the Linda problem is given.

Keywords

probability, logic, Boolean valuation, Linda problem

Abstrakt

Uwaga o prawdopodobieństwie zdań logicznych i problemie Lindy

Niniejsza praca prezentuje koncepcję prawdopodobieństwa logicznego, która choć wydaje się oczywista, to autor nie znalazł nigdzie ani jej opracowania ani tym bardziej zastosowania do rozwiązania problemów filozoficznych. To prawdopodobieństwo formuły A, języka klasycznej logiki zdaniowej, jest ilorazem liczby wartościowań boolowskich, dla których formuła A przyjmuje wartość logiczną prawdy, do liczby wszystkich wartościowań boolowskich formuły A. To nowe pojęcie prawdopodobieństwa zostało zastosowane do podania alternatywnego rozwiązania problemu Lindy.

Słowa kluczowe

prawdopodobieństwo, logika, wartościowania boolowskie, problem Lindy

Bibliography

Carnap R., *Logical Foundations of Probability*, Chicago 1950.

Hintikka J., *On Semantic Information*, in: *Information and Inference*, eds. J. Hintikka, P. Suppses, Dordrecht 1970, p. 3–27.

Lorenz D., Kooi B., Sack J., *Logic and Probability*, https://plato.stanford.edu/archives/sum2019/entries/logic-probability/ (15.12.2022).

Makinson D., *Bridges from Classical to Nonmonotonic Logic*, London 2005.

Mazurkiewicz S., *Przyczynek do aksjomatyki rachunku prawdopodobieństwa. Zur Axiomatik der Warscheinlichkeitsrechnung*, „Sprawozdania z Posiedzeń Towarzystwa Naukowego Warszawskiego. Wydział III nauk matematyczno-fizycznych" 25 (1933) no 1–6, p. 1–4.

Mazurkiewicz S., *Über die Grundlagen der Wahrscheinlichkeitsrechnung I*, "Monatshefte für Mathematik und Physik" 41 (1934), p. 343–352; https://doi.org/10.1007/BF01697866.

Olszewski A., *A Few Comments on the Linda Problem*, "Organon F" 24 (2017), p. 184–195.

Olszewski A., *Linda Problem – the Tame Solution in Question*, "Analecta Cracoviensia" 51 (2019), p. 209–217.

Tversky A., Kahneman D., *Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, "Psychological Review" 90 (1983), p. 293–315; https://doi.org/10.1037/0033-295X.90.4.293.