Jacek Malinowski

LOGIC OF SIMPSON PARADOX

Abstract. The main aim of this paper is to elucidate, from a logical point of view, the phenomenon of Simpson reversal — the paradox of a statistical reasoning. We define a binary relation of supporting in the following way: a sentence A supports a sentence B if and only if the probability of B is higher when A is true, than when A is false. It appears that a statistical argument occurring in Simpson paradox cannot be formalized by means of a binary relation. We generalize the relation of support introducing the third parameter. Then we argue that it properly mirrors main features of the statistical argument occurring in Simpson paradox.[*]

Keywords: Logical entailment, statistical inference, Bayesian inference, Simpson paradox.

1. Introduction

Let us begin with an example. Suppose that some university is trying to discriminate in favour of women when hiring staff. It advertises positions in the Department of History and in the Department of Geography, and only in those departments. Five men apply for the positions in History and one is hired, and eight women apply and two are hired. The success rate for men is twenty percent while the success rate for women is twenty-five percent. In the Geography Department eight men apply and six are hired, and five women apply and four are hired. The success rate for men is seventy-five percent and for women it is eighty percent. The Geography Department

---

[*]This paper was prepared during the author's stay at the Netherlands Institute for Advanced Study.

has favoured women over men. Yet across the University as a whole 13 men and 13 women applied for jobs, and 7 men and 6 women were hired. The success rate for male applicants is greater than the success rate for female applicants.

	Men	Women
History	1/5	2/8
Geography	6/8	4/5
Whole University	7/13	6/13

Tabular 1.

We could summarize the results above as follows:

 (i) Department of History prefer women applicants.
 (ii) Department of Geography prefer women applicants.
(iii) Both the Departments taken together prefer men applicants.

The phenomenon of the reversal above is called the Simpson paradox. First encountered by Pearson in 1899, it was investigated in Simpson [1951]. In statistics, unless used judiciously, Simpson paradox may wreak havoc. We refer a reader to Pearl [2000] for detailed investigation and references.

The aim of the present paper is to construct a formal system of propositional logic which elucidates the reasons of the above reversal.

Obviously there is a 'bias in the sampling' in the example above. The question is where exactly this bias arises? The key to this puzzling example lies in the fact that more women are applying for jobs that are harder to get. It is harder to make your way into History than into Geography. Of the women applying for jobs, more are applying for jobs in History than in Geography while the reverse is true for men. History hired only 3 out of 13 applicants, whereas Geography hired 10 out of 13 applicants. Hence the success rate was much higher in Geography, where there were more male applicants.[2] For this reason, what we can statistically deduce in the example does not depend only on the data, it depends also on the second factor - the variable which can receive one of the two values "Department of Geography" and "Department of History".

We are going to investigate from a purely logical point of view the relation of statistical inference occurring in the phenomenon of Simpson paradox.

---

[2]This example originates from the internet Stanford Encyclopedia of Philosophy.

We will define a propositional language which allows us to consider events like "Department of History prefers woman applicants" as sentences of this language. Then we construct the semantics by considering probabilities as logical valuations. This allows us to define the relation of supporting $\approx\!\!\!\mid$ – a formal counterpart of the notion of corroboration. Then we analyze to what extent the relation of supporting reflects the features of statistical entailment occurring in Simpson paradox.

We will argue that a statistical reasoning occurring in Simpson paradox is a ternary relation and cannot be reduced to a binary one. The relation of supporting is a special case of a statistical reasoning occurring in Simpson paradox.

2. Basic concepts

Let L_n denotes an n-generated sentential language with connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$, generated by n-element set $V = \{q_1, ...q_n\}$ of sentential variables. We will identify language L_n with its set of sentences (well-formed formulas). By small characters p, q, r (with or without indices) we will denote the sentential variables while the capitals A, B, C, P, Q, R will denote arbitrary sentences of L_n.

By a *classical logic* in the language L_n we mean the binary relation $\models^n$ between sets of sentences and single sentences defined in the following standard way: $X \models^n A$ iff for any *classical valuation* $v : L_n \longmapsto \{0, 1\}$, $v(A) = 1$ whenever $v(B) = 1$ for any $B \in X$. The symbol $\models^n A$ means that A is a classical tautology. If it does not lead to a misunderstanding we skip superscript n in $\models^n$.

By a *probabilistic valuation* or simply a *valuation* of the language L_n we mean any function w defined on L_n taking values in the unit interval of reals $[0, 1]$.

(W1) $0 \leq w(A) \leq 1$

(W2) $w(A) = 1$ for some sentence A

(W3) $w(A) \leq w(B)$ whenever $A \models B$

(W4) $w(A \vee B) = w(A) + w(B)$ whenever $A \models \neg B$

The conditions (W1)–(W4) correspond to Kolomogorov's axioms defining finitely additive probability function. We refer the reader to C. Howson, P. Urbach [1989] and chapter 5 of Makinson [2005] for more details.

Proposition 1. *For any valuation w the following condition hold:*

(W5) if A is a classical tautology then $w(A) = 1$

(W6) $w(\neg A) = 1 - w(A)$

(W7) $w(A_1 \vee ... \vee A_n) = w(A_1) + ... + w(A_n)$ whenever $A_i \models \neg A_j$ for all $i \neq j$, $1 \leq i, j \leq n$}.

(W8) $w(A \vee B) = w(A) + w(B) - w(A \wedge B)$

It is well known that any function $f : \{q_1, ..., q_n\} \longmapsto \{0, 1\}$ can be uniquely extended to the classical valuation of L_n. It is easy to observe that for probabilistic valuations it is more complicated. A function $f : \{q_1, ..., q_n\} \longmapsto [0, 1]$ might have many possible extensions to a probabilistic valuation.

By a *literal* of L_n we mean either a sentential variable or a negation of a sentential variable. By a *state description* in the language L_n we mean any conjunction of n different literals in a fixed order. Each i-th term of such a conjunction is either q_i or $\neg q_i$. Any function f from the set st of all of 2^n state descriptions into the unit interval $[0, 1]$ such that $\sum_{s \in st} f(s) = 1$ will be called a *probability distribution*.

Proposition 2.

(i) *In any language L_n there is 2^n state descriptions.*

(ii) *Any formula A is classically equivalent to a disjunction of a unique non-empty subset of state descriptions.*

(iii) *The conjunction of two different state descriptions is a contradiction.*

(iv) *Any probability distribution can be in a unique way extended to a valuation satisfying* (W1) - (W4). $\square$

Let w denote a probabilistic valuation w, then for any sentences A and B such that $w(A) \neq 0$ a function $p_A(B)$ defined as

$$p_A(B) = \frac{w(B \wedge A)}{w(A)},$$

will be called a *conditional probability*.

We say that A *support* B under w, in symbols $A \approx_w^n B$ if and only if either $w(A) = 0$ or $w(A) = 1$ or $w(A) \neq 0$ and $p_A(B) \geq p_{\neg A}(B)$. The fact that $A \approx_w^n B$ for any valuation v in Ln will be denoted $A \approx^n B$ or $A \approx B$ if the context clearly indicates the language. By proposition 3 (iii) below we

can express it less formally but more demonstratively in the following two equivalent ways:

$A \approxeq B$ **if and only if the probability of** B **is higher when** A **is true, than when** A **is false.**

$A \approxeq B$ **iff the fact that** A **is the case increases the probability of** B**.**

J. Malinowski [2005] contains more results on the relation $\approxeq$. The relation of supporting formalizes well known notions of corroboration Popper [1968] and confirmation Carnap [1951] considered by philosophers of science as a tool for testing scientific hypotheses. We refer the reader to Kuipers [2000] for more references on this subject.

PROPOSITION 3. *The following conditions are equivalent.*

(i) *A support B under w.*

(ii) *B support A under w.*

(iii) $w(A \wedge B) \geq w(A) \cdot w(B)$.

(iv) $w(A \wedge B) \cdot w(\neg B) \geq w(A \wedge \neg B) \cdot w(B)$.

(v) $p_B(A) \geq w(A)$.

3. Simpson paradox and its formalization

In first approximation we could describe Simpson paradox as a probability measure based on a bunch of statistical data such that a given event E is more probable than its negation on the whole population and at the same time E is less probable than its negation on all the elements of some partition of the whole population. Simpson paradox in this formulation remains the method of proving theorem by considering cases.

Also in first approximation the relation of statistical inference occurring in Simpson paradox seems to be adequately formalized by the relation $\approxeq$. As a consequence one could think that Simpson paradox consist in the failure of the following rule (Sim) for the relation $\approxeq$.

(Sim) $$\frac{P \wedge R \approxeq Q \qquad P \wedge \neg R \approxeq Q}{P \approxeq Q}$$

It appears however, that such a point is mistaken. Let us suppose that w is any valuation, $\approxeq$ is a shorthand for $\approxeq_w$, $P \wedge R \approxeq Q$ and $P \wedge \neg R \approxeq Q$. Then by Proposition 3 (iii) $w(P \wedge R \wedge Q) \geq w(P \wedge R) \cdot w(Q)$ and

$w(P\wedge\neg R\wedge Q) \geq w(P\wedge\neg R)\cdot w(Q)$. We have $w(P\wedge Q) = w(P\wedge Q\wedge(R\vee\neg R)) = w((P \wedge Q \wedge R) \vee (P \wedge Q \wedge \neg R)) = w((P \wedge Q \wedge R) + w(P \wedge Q \wedge \neg R)) \geq w(P \wedge R) \cdot w(Q) + w(P \wedge \neg R) \cdot w(Q) = w(Q) \cdot (w(P \wedge R) + w(P \wedge \neg R)) = w(Q) \cdot w((P \wedge R) \vee (P \wedge \neg R)) = w(P) \cdot w(Q)$ and then $P \not\approx Q$. As a consequence (Sim) is valid for $\not\approx_w$ for any w.

On the other hand the example from the first section shows that the reversal of inequalities in the (Sim) is possible. Since such a reversal is impossible for $\not\approx_w$ for any w, then $\not\approx_w$ is not a suitable formalization of the form of reasoning occurring in Simpson paradox.

Let us make our example more precise. Let us denote:

E - "The application is successful".
C - "Woman applies."
F - "An applicant applies to Department of Geography."

Suppose that only men and women apply, so $\neg C$ means "A man applies", and that Department of History is the only alternative for F, then $\neg F$ means "An applicant applies to Department of History". Then the data from table 1 can be expressed in the following way:

(i) History: $p_{C\wedge\neg F}(E) \geq p_{\neg C\wedge\neg F}(E)$.

(ii) Geography: $p_{C\wedge F}(E) \geq p_{\neg C\wedge F}(E)$.

(iii) University: $p_C(E) \leq p_{\neg C}(E)$.

Let us observe that only in the last line the table might be easily expressed in terms of the relation $\not\approx$ as $C \not\approx E$. It seems that the natural candidates for previous formulas are

$$C \wedge \neg F \not\approx E \qquad \text{and} \qquad C \wedge F \not\approx E.$$

It is not true however. The extended forms of the last two formulas are respectively:

$$p_{C\wedge\neg F}(E) \geq p_{\neg C\vee F}(E) \qquad \text{and} \qquad p_{C\wedge F}(E) \geq p_{\neg C\vee\neg F}(E).$$

It is easy to check that they are not equivalent to (i) and (ii). This argument shows that statistical entailment occurring in Simpson paradox cannot be, in general, formalized as a relation between two variables. It is often necessary to consider the third variable.

We will define a generalization of $\not\approx$ which seems to be more suitable as a formalization of Simpson paradox. We say that *A relatively supports C*

with respect to parameter B under valuation w, in symbols $A \vDash_w [B]C$, if and only if $w(B) \neq 0$ and

$$p_{B \wedge A}(C) \geq p_{B \wedge \neg A}(C).$$

PROPOSITION 4.

(i) *If $A \vDash C$, then for any w and B such that $w(B) \neq 0$ $A \vDash_B C$.*

(ii) *If $A \vDash_w [B]C$ and we replace A or B or C by a logically equivalent sentence then for a resulting triple of sentences A_1, B_1 and C_1 we have $A_1 \vDash_w [B_1]C_1$.*

(iii) *For any valuation w and sentence B such that $w(B) \neq 0$ there exists a valuation v such that $\vdash = \approx_v$, where $\vdash = \{(A, C) : A \vDash_w [B]C\}$.*

(iv) *If B is a classical tautology then for any valuation w and any sentences A and B, $A \vDash_w [B]C$ if and only if $A \approx C$.*

PROOF. (i), (ii) and (iv) are immediate. We will prove (iii). Lets us suppose that a valuation v and a sentence B are such that $w(B) \neq 0$ is given. Let us define a function v for any $A \in L_n$ as $v(A) = \frac{w(A \wedge B)}{w(B)}$. It it easy to check that v is a valuation. Then $A \approx_v C$ iff

$$\frac{v(A \wedge C)}{v(A)} \geq \frac{v(\neg A \wedge C)}{v(\neg A)}.$$

We transform this formula to a complex fraction using the definition of v and then cancelling by $w(B)$ the resulting fraction. In this way we obtain an equivalent formula

$$\frac{w(B \wedge A \wedge C)}{w(B \wedge A)} \geq \frac{w(B \wedge \neg A \wedge C)}{w(B \wedge \neg A)}.$$

It is equivalent to $A \vdash C$. □

Proposition 4 (iii) shows that for a fixed parameter B a binary relation between A and C, $A \vDash_w [B]C$ has the same structure as the relation $\approx_v$ for some valuation v. For this reason if the parameter B is fixed then the rule (Sim) is valid and hence Simpson paradox does not occur. Simpson paradox might occur only if different parameters occur in one argument.

References

Carnap, R. [1951], *Logical foundations of probability*, Routledge and Kegan Paul, London.

Howson, C., and P. Urbach [1989], *Scientific Reasoning: the Bayesian Approach*, Open Court Publishing Co., La Salle, Illinois.

Kuipers, T. [2000], *From Instrumentalism to Constructive Realism: On Some Relations Between Confirmation, Empirical Progress, and Truth Approximation*, Synthese Library 287, Kluwer Academic Press, Dordrecht.

Makinson, D. [2005], *Bridges from Classical to Nonmonotonic Logic*, Texts in Computing, Kings College, London.

Malinowski, J. [2005], "Bayesian propositional logic", to appear.

Pearl, J. [2000], *Causality: Models reasoning, and inference*, Cambridge University Press, Cambridge.

Popper, K. [1968], *The Logic of Scientific Discovery*, 3rd. (revised) edition, Hutchinson, London.

Simpson, E. H. [1951], "The Interpretation of Interaction in Contingency Table", *Journal of Royal Statistical Society*, Series B, vol. 13, 238–241.

Jacek Malinowski
Institute of Philosophy and Sociology,
Polish Academy of Sciences

Department of Logic,
Nicolaus Copernicus University, Toruń
jacekm@uni.torun.pl