On Simpson's Paradox and the Sure-Thing Principle

Author(s): Colin R. Blyth

Source: *Journal of the American Statistical Association*, Jun., 1972, Vol. 67, No. 338 (Jun., 1972), pp. 364-366

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2284382

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Taylor & Francis, Ltd. and *American Statistical Association* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

This content downloaded from 75.69.59.185 on Wed, 24 Nov 2021 02:47:53 UTC

© Journal of the American Statistical Association
June 1972, Volume 67, Number 338
Theory & Methods Section

On Simpson's Paradox and the Sure-Thing Principle

COLIN R. BLYTH*

This paradox is the possibility of $P(A|B)<P(A|B')$ even though $P(A|B)\geq P(A|B')$ both under the additional condition C and under the complement C′ of that condition. Details are given on why this can happen and how extreme the inequalities can be. An example shows that Savage's sure-thing principle ("If you would definitely prefer g to f, either knowing that the event C obtained, or knowing that C did not obtain, then you definitely prefer g to f.") is not applicable to alternatives f and g that involve sequential operations.

1. INTRODUCTION

Simpson's paradox is described in [3]. An occurrence in real data is given in [1, page 449]. Here is an example illustrating this paradox:

A doctor was planning to try out a new treatment on patients, mostly local (C), and a few in Chicago (C'). A statistician advised him to use a table of random numbers, and as each C patient became available, assign him to the new treatment with probability .91, leave him to the standard treatment with probability .09; and the same for each C' patient with probabilities .01 and .99 respectively. (These probabilities were expected to give him about the number of patients he could handle in each city.) When the doctor returned later with the data of Tabulation 1.1, the statistician told him that the new treatment was obviously a very bad one, and criticized him for having continued trying it on so many patients.

Tabulation 1.1

Treatment	*Standard*	*New*
Dead:	5950	9005
Alive:	5050 (46%)	1095 (11%)

The doctor replied that he had continued because the new treatment was obviously a very good one, having nearly doubled the recovery rate in both cities; and he was correct, as Tabulation 1.2 shows.

Tabulation 1.2

	C patients only		*C′ patients only*	
Treatment:	*Standard*	*New*	*Standard*	*New*
Dead:	950	9000	5000	5
Alive:	50 (5%)	1000 (10%)	5000 (50%)	95 (95%)

From this data one would expect to have had about 10,700 recoveries had all patients received the new treatment, as compared to the actual 6,145 recoveries, and to about 5,600 recoveries had all received the standard treatment. The difficulty is not one of chance variation—the observed proportions might be the true ones, or the observed numbers could be multiplied by a constant large enough to make this essentially so. As with any paradox, there is nothing paradoxical once we see what has happened: The C patients are much less likely to recover, and the new treatment was given mostly to C patients; and of course a treatment will show a poor recovery rate if tried out mostly on the most seriously ill patients.

Section 2 is a discussion of what this paradox amounts to mathematically, and how extreme it can get. Section 3 is a discussion of the sure-thing principle, in view of this paradox.

2. SIMPSON'S PARADOX

Mathematically, Simpson's paradox is this:

It is possible to have

$$P(A \mid B) < P(A \mid B'),$$

$$\text{and have at the same time both} \qquad (2.1)$$

$$P(A \mid BC) \geqq P(A \mid B'C)$$

$$P(A \mid BC') \geqq P(A \mid B'C').$$

One tends to reason intuitively that this is impossible because

$$\begin{aligned} P(A \mid B) &= \text{An average of } P(A \mid BC) \text{ and } P(A \mid BC') \\ P(A \mid B') &= \text{An average of } P(A \mid B'C) \qquad (2.2) \\ &\qquad \text{and } P(A \mid B'C'), \end{aligned}$$

which is easily seen to be true when all the conditioning events have positive probabilities:

$$\begin{aligned} P(A \mid B) &= \{P(C \mid B)\} \cdot P(A \mid BC) \\ &\quad + \{P(C' \mid B)\} \cdot P(A \mid BC') \\ P(A \mid B') &= \{P(C \mid B')\} \cdot P(A \mid B'C) \\ &\quad + \{P(C' \mid B')\} \cdot P(A \mid B'C'); \end{aligned} \qquad (2.2')$$

---

* Colin R. Blyth is professor of mathematics, Queen's University, Kingston, Ontario, Canada, on leave from the University of Illinois at Urbana. This work was supported by National Science Foundation grants GP 14786 and GP 28154 at the University of Illinois.

but the reasoning fails because these two averages have different weightings. However, in particular, if B and C are independent, these two weightings coincide and the reasoning correctly shows that (2.1) is impossible. The paradox can be said to result from the dependence or interaction of B and C.

In the introductory example, $A =$ Alive, $B =$ New treatment, $C =$ Local patient; $P(\ )$ refers to probability for a patient chosen at random from among those recorded in table (1.2), and coincides with proportions in that table, now being taken as the total available population. In that example we have

$$P(A \mid B) = .11 < P(A \mid B') = .46,$$

$$P(A \mid BC) = .10 > P(A \mid B'C) = .05,$$

$$P(A \mid BC') = .95 > P(A \mid B'C') = .50.$$

The initially surprising fact that an average of .10 and .95 is so much smaller than an average of .05 and .50 is easily explained by showing the numerical values in (2.2′):

$$.11 = \{.99\}(.10) + \{.01\}(.95)$$

$$.46 = \{.10\}(.05) + \{.90\}(.50).$$

Here the paradox could not happen if B and C were independent, i.e., if the proportion receiving the new treatment were the same for C patients and C' patients.

Here is the most extreme possibility in Simpson's paradox:

Subject to the conditions

$$P(A \mid BC) \geqq \gamma \cdot P(A \mid B'C)$$

$$P(A \mid BC') \geqq \gamma \cdot P(A \mid B'C') \qquad (2.3)$$

with $\gamma \geqq 1$, *it is possible to have*

$$P(A \mid B) \cong 0 \quad \text{and} \quad P(A \mid B') \cong 1/\gamma.$$

In particular, with $\gamma = 1$, it is possible to have $P(A \mid B) \cong 0$ and $P(A \mid B') \cong 1$. To show (2.3), recall what caused the paradox in the introductory example: the C patients were much more seriously ill, and the new treatment was given mostly to C patients. Let the probabilities of recovery for patients getting the standard, new treatments be α, $\gamma\alpha$ for C patients and β, $\gamma\beta$ for C' patients, with $\beta > \alpha$. Then, by giving the new treatment to sufficiently many C patients and sufficiently few C' patients, we can make $P(A \mid B) \cong \gamma\alpha$ and $P(A \mid B') \cong \beta$. Now, by suitable choices of α and β, and noticing the requirement $\gamma\beta \leqq 1$, we can make $P(A \mid B) \cong 0$ and $P(A \mid B') \cong 1/\gamma$, which is clearly the most extreme possibility.

It is easy to construct examples using the preceding method—make $P(A \mid C)$ small enough, $P(A \mid C')$ large enough, $P(B \mid C)$ large enough, and $P(B \mid C')$ small enough. The introductory example was constructed in just this way with $\gamma = 2$. Tabulation 2.1 shows an example constructed in the same way with $\gamma = 1$.

Tabulation 2.1

	C patients only		*C′ patients only*	
Treatment:	*Standard*	*New*	*Standard*	*New*
Dead:	100	10000	1000	9
Alive:	9 (8%)	1000 (10%)	10000 (91%)	100 (92%)

	All patients	
Treatment:	*Standard*	*New*
Dead:	1100	10009
Alive:	10009 (90%)	1100 (11%)

3. THE SURE-THING PRINCIPLE

The relatively formal statement of the sure-thing principle given by Savage, [2, pp. 21–22], is this:

> If you would definitely prefer g to f, either knowing that the event C obtained, or knowing that the event C did not obtain, then you definitely prefer g to f. (3.1)

Here f and g are two alternative possible acts of almost any sort [1, paragraph on pp. 15–16]: "What in the ordinary way of thinking might be regarded as a chain of decisions, one leading to the other in time, is to be regarded as a single decision or act;" for example, the choice of a strategy in a game, or the choice of a sequential decision procedure in statistics.

In the Simpson's paradox example of Section 1, the sure-thing principle is false as applied to

f: Draw patients at random from Tabulation 1.2 until you get one who got the standard treatment, and bet a dollar that he recovered.

g: Draw patients at random from Tabulation 1.2 until you get one who got the new treatment, and bet a dollar that he recovered.

C: The patient you bet on is a local one.

Given that C obtained, you would definitely prefer g to f because g gives you double the probability of winning a dollar. Given that C' obtained, you would definitely prefer g to f because g gives you nearly double the probability of winning a dollar. But you definitely prefer f to g because f gives you over four times the probability of winning a dollar.

Here, in detail, is the process being considered: We have the record of each of the 21,100 patients of Tabulation 1.2 on a card whose color indicates the treatment that patient got. We draw cards at random, one at a time (with replacement, say), sequentially, from this deck. We are allowed to observe the color of each card as we draw it, and to use one of the two alternative possible stopping rules f and g. Using f, the card we bet on is equally likely to belong to any one of 11,000 patients of whom 5,050 (i.e., 46%) recover; using g, the card we bet on is equally likely to belong to any one of 10,100 patients of whom 1,095 (i.e., 11%) recover (see Tabulation 1.1); therefore we definitely prefer f to g. But if we knew that C obtained, i.e., that the card we stopped on was for a C patient (or to put it another way, if the drawing were declared void and a new drawing made when the card we stop on belongs to a C' patient—of course, such knowledge would make the process a different, conditional, one), then we would know that, using f, the card we bet on would be equally likely to belong to any one of 1,000 patients of whom 50 (i.e., 5%) recover; and, using g, the card we bet on would be equally likely to belong to any one of 10,000 patients of whom 1,000 (i.e., 10%) recover (see Tabulation 1.2); therefore we would definitely prefer g to f. Similarly, if we knew that C' obtained, we would definitely prefer g to f.

The space of possible outcomes of the process just described is the familiar space of stopped sequences of sequential analysis; the probability model is that each card drawn, independent of other drawings, is equally likely to be any one of the 21,100 possible cards. Just as in sequential analysis, we can either say that a change of stopping rule changes the sample space; or else designate the sample space as the set of *all* stopped sequences, and say that a change of stopping rule changes the probability distribution on this space.

In sequential analysis, this paradox is so familiar that it is not considered paradoxical: at a stage where we can take 0, 1 or 2 more observations, and where it is inadvisable to take 1 more and inadvisable to take 2 more (and so is inadvisable to use any randomization independent of observations between 1 and 2 more), it can be advisable to use a sequential continuation that takes sometimes 1 and sometimes 2 observations. Of course, this is the whole point of sequential analysis; were it not so, sequential methods would offer no advantage.

The sure-thing principle, then, seems not applicable to situations in which any action taken within f or g, or (what can be construed as the same thing) in which the choice between f and g, is allowed to be based sequentially on events dependent with C. Such situations appear to include any in which subjective methods are used, since subjective beliefs are presumably based sequentially on events that could be dependent with C.

My thanks to L. J. Savage, for telling me of [1], and for other helpful correspondence. Similar thanks to G. W. Haggstrom, who pointed out that Simpson's paradox is the simplest form of the false correlation paradox in which the domain of x is divided into short intervals, on each of which y is a linear function of x with large negative slope, but these short line segments get progressively higher to the right, so that over the whole domain of x, the variable y is practically a linear function of x with large positive slope.

[*Received April 1970. Revised July 1971.*]

REFERENCES

[1] Cohen, M. R. and Nagel, E., *An Introduction to Logic and Scientific Method*, New York: Harcourt, Brace and Co., 1934.

[2] Savage, L. J., *The Foundations of Statistics*, New York: John Wiley and Sons, Inc., 1954.

[3] Simpson, E. H., "The Interpretation of Interaction in Contingency Tables," *Journal of the Royal Statistical Society*, Ser. B, 13, No. 2 (1951), 238–41.

Some Probability Paradoxes in Choice from Among Random Alternatives

COLIN R. BLYTH*

The probability $P(X>Y)$ can be arbitrarily close to 1 even though the random variable X is stochastically smaller than Y; the probabilities $P(X<Y)$, $P(Y<Z)$, $P(Z<X)$ can all exceed $\frac{1}{2}$; it is possible that $P(X=\min X, Y, Z)<P(Y=\min X, Y, Z)<P(Z=\min X, Y, Z)$ even though $P(X<Y)$, $P(X<Z)$, $P(Y<Z)$ all exceed $\frac{1}{2}$. This article examines these paradoxes and extensions of them, and discusses the difficulties they cause in the problem of choosing from among possible random losses or payoffs; as in choosing from among possible statistical decision procedures or from among possible wagers.

1. INTRODUCTION AND SUMMARY

This article is concerned with choosing a "best" one from a class of available random alternatives, using this simple probability model: corresponding to the ith alternative is a real-valued random variable L_i, and our choice is to be based solely on the joint distribution of the L_i's, which is known. Here the desirabilities of the alternatives depend on the outcome of a chance process, with known distribution, which generates the L_i's: on a particular realization of that process, the ith alternative delivers a numerical amount, say $L_i=u_i$, of something that is good—a payoff, such as money, time in useful life, pleasure, satisfaction, etc. (or else bad—a loss, such as money, time in a speed contest, displeasure, dissatisfaction, etc.).

Our program is this: first we choose an i, and then there

---

* Colin R. Blyth is professor of mathematics, Queen's University, Kingston, Ontario, Canada, and is on leave from the University of Illinois at Urbana. This work was supported by National Science Foundation grants GP 8727 and GP 28154 at the University of Illinois.