CONDITIONAL PROBABILITIES: INSIGHTS AND DIFFICULTIES

Ruma Falk
Department of Psychology and the Sturman Centre for Human Development
The Hebrew University
Jerusalem, Israel

Conditional probabilities play a central role in the process of inferring about the uncertain world. The formal definition of $P(A|B)$ is easy and poses no problems. However, upon careful probing into students' ideas of conditional probabilities, some misconceptions and fallacies are uncovered. In this paper I wish to discuss three issues involving conditional probabilities that I believe require serious consideration and clarification by students and by teachers of probability.

1. Interpreting conditionality as causality

Consider the following example (Falk, 1978, p. 46; 1979): An urn contains two white balls and two black balls. we blindly draw two balls, one after the other, without replacement from that urn. First we ask about $P(W_{II}|W_I)$. Students easily answer it by 1/3. Second we ask about $P(W_I|W_{II})$. This question usually triggers a lively discussion in class. Some students go the extreme of refusing to consider the problem since "it is meaningless". They claim that conditioning the probability of an outcome of a drawn on an event that occurs later is not permissible. Among those who do answer, the majority choose 1/2. These students typically argue that at that stage (before the first draw), the second draw had not yet been carried out, and "the first ball doesn't care whether the second is white or black". Therefore, they base their answer solely on the composition of the urn at the outset of the experiment, disregarding the information about the later outcome. An appropriate answer to this claim is "indeed the first ball doesn't care whether the second is white or black but we do". The heart of the problem lies in its being addressed to our state of knowledge. We have advanced beyond the initial stage when we learned that the second draw resulted in a white ball. This information removed one white ball out of the possible outcomes of the first draw, and therefore, $P(W_I|W_{II}) = 1/3$ (just like $P(W_{II}|W_I)$).

The students' verbal responses, especially their refusal to consider evidence occurring later than the judged event, reflect their causal reasoning. While the outcome of the second draw depends causally on that of the first draw, the reverse is not true. Still, the informational impact of W_{II} on W_I is the same as that of W_I on W_{II}. Psychologically however, these two problems are not perceived as symmetrical. While the first causal inference is natural and compatible with the time axis, the second "backward inference" seems to create a difficulty since its calls for probabilistic reasoning that is indifferent to temporal order. Understanding that an event's probability may be revised in light of knowledge of later occurrences, may be instructive. It can be enriched and substantiated by referring the students to familiar and unquestionable examples of situations where recently obtained information is utilized to modify prior evaluations of the probability of uncertain events. Thus, the findings of archeological excavations throw new light on historical events which took place hundreds or thousands of years earlier. In another context, although diseases are the causes and symptoms are the effect – although diseases come first and symptoms follow – the medical diagnostic process attaches probabilities to diseases on the basis of determination of their symptoms.

Tversky and Kahneman (1980) investigated the judgments of the conditional probability $P(X|D)$ of some target event X,on the basis of some data D. For a psychological analysis they distinguished between different types of relations that the judge may perceive between D and X. If D is perceived as a cause of X they referred to D as causal datum. On the other hand, if X is treated as a possible cause of D, they referred to D as a diagnostic datum. In a normative treatment of conditional probability, the distinction between the various types of relation of D to X is immaterial, and the impact of data depends solely on their informativeness. In contrast, Tversky & Kahneman showed that the psychological impact of data depends critically on their role in the causal schema. Because of the prevalence of causal schemas in our perception of the world, causal data have greater impact on our probabilistic inference than other data of equal objective informativeness. They asked subjects to compare two conditional probabilities $P(Y|X)$ and $P(X|Y)$ for a pair of positively correlated events X and Y such that (1) X is naturally viewed as the cause of Y; (2) the marginal probabilities of the two events are equal. The latter condition implies that $P(Y|X) = P(X|Y)$ (just as in the two-draws example). Most subjects judged the causal relation as stronger than the inverse (diagnostic) relation and erroneously asserted that $P(Y|X) > P(X|Y)$. Thus, the probability that a girl has blue eyes if her mother has blue eyes was judged to be greater than the probability that the mother has blue eyes if her daughter has blue eyes, although the proportions of blue eyed individuals in the two generations were regarded as equal.

Analyzing such examples of violations of normative rules may bring about illuminating psychological and didactical insights. The example of the two draws (where the two events are negatively correlated) provides an extreme case of the greater impact of causal evidence relative to diagnostic evidence via the subjects' complete denial of the relevance of the latter kind of evidence.

2. The definition of the conditioning event is often problematic

This can be best demonstrated by some notorious teasers (Gardner, 1959; Freund, 1965; Mosteller, 1965, pp. 28-29; Falk, 1978, pp. 68-69; Betteley, 1979; Bar-Hillel & Falk, 1982). Consider, for example, the well known problem of the three cards:

> Three cards are in a hat. One is blue on both sides, one is green on both sides, and one is blue on one side and green on the other. We draw one card blindly and put it on the table as it comes out. It shows a blue face up. What is the probability that the hidden side is also blue?

Most people spontaneously give an answer of 1/2. They condition their computation on the event "the double-green card is out", and they figure that each of the two remaining cards is equally likely to be the one on the table. Although the inference that the double-green card cannot be the one on the table is correct, that is not the event on which one should condition the probability of the target event. That conditioning event is not defined in the sample-space of the possible outcomes of our statistical experiment. The six elementary outcomes of the experiment are the six faces of the three cards each of which is an equally likely candidate to end up being the upper face on the table, as guaranteed by the detailed experimental procedure. We obtained a "Blue side up", denoted Bu, and this is the event on which one should condition the probability of "blue on the back-side". Two of the three outcomes in Bu have blue on the hidden side and one has green, hence the answer is 2/3.

Recently I encountered (in "real-life") a mathematically isomorphic problem in which a woman is expecting twins. A priori, the three possible combinations of twins – two boys, two girls, boy and girl – are known to be equiprobable (Stern, 1960). A chromosomal test is performed on random cells out of one (random) amnion and the results show it is a boy. What is the probability that the woman is expecting two boys? By analogy, we know that although it is true that the possibility of two girls is ruled out by the test's result, the remaining two possibilities are not equally probable any more. If the woman is carrying two boys the test's outcome is twice as likely than in the case that she is expecting a boy and a girl. Therefore, the posterior probability of "two boys" is 2/3. (The issue becomes complicated when the twins are enclosed in a single amnion. This case, however, can be neglected since only a minority of only identical twins have a single amnion – see Stern, 1960, p. 536).

The main lesson from the two versions of the example is that the probability of the target event should be conditioned on the immediate event given as datum in the problem and not on some inferred event. The conditioning event should not coincide with the valid conclusion "the double green card is out"/"two girls are out". It should be directly defined by the problem's experimental procedure, i.e., "a randomly selected side of a randomly selected card is blue"/"a randomly selected fetus out of a random twin-pregnancy is a boy". The exact method by which we obtained the given data is crucial in determining our conditioning event. What matters is not only what we know but also how do we know it. Indeed, a different experimental procedure, from which the same conclusion is desired, may result in a different conditional probability of the target event (Bar-Hillel & Falk, 1982; Glickman, 1982; Falk, 1983). These examples highlight the vital role of the basic concept of the statistical experiment, the outcomes of which define our probability space. One way to promote gaining the insights concerning such problems is to device experimental models (Glickman, 1982; Falk, 1983) in order to explicate the exact procedure that has generated the data and to uncover hidden assumptions.

3. Confusion of the inverse

That is, lack of discrimination between the two directions of conditional probability, $P(A|B)$ and $P(B|A)$. This long recognized confusion is prevalent among students and professionals at all levels. It occurs often in medical contexts in the interpretation of test results (Eddy, 1982), where the probability of disease given a positive test result is erroneously equated with that of a positive result given the disease. Pauker & Pauker (1979) demonstrate the dramatic gap between the conditional probability of a baby being affected by Down's syndrome, given a positive prenatal amniocentesis outcome, and the probability of obtaining a positive test result, given the fetus is affected. Since for women of age 30 the incidence of live-born Down's syndrome is 1/885, even if the two conditional probabilities of a correct test result, given either an affected or a normal fetus, were 99.5%, the probability of an affected child, given a positive test result, would be only 18.4%. Thus, if D denotes Down's syndrome, and POS a positive test result, we have $P(POS|D) = 0.995$, while $P(D|POS) = 0.184$.

The confusion of the inverse is especially compelling with respect to the interpretation of statistical significance tests. The level of significance of a test, denoted α, is the conditional probability of obtaining a result in the rejection region, denoted R, when H_0 is true. This means: $\alpha = P(R|H_0)$. However, when a statistical test turns out significant (which means that R has occurred) and one is asked about the probability that this has been an error (namely, that H_0 is true), one usually gives the answer " α ", although the question referred to is $P(H_0|R)$. The universal spread of that confusion (Falk, 1986) can be traced, in part to the pressing motivation to cope with the problem of chance. Whenever a random sample significantly deviates from the null hypothesis, the most natural question to ask is whether that deviation could be accounted for by random fluctuations. In fact, it is the question of all questions of statistics. Considering that one expects statistical inference to provide a decision accompanied by an evaluation of the probability of error, it is natural that, following a rejection-decision, one interprets the probability α, associated with rejection H_0, as that of having committed an error.

Another factor that probably contributes to the tendency to interpret $\alpha = P(R|H_0)$ as $P(H_0|R)$ is that the linguistic ambiguity associated with the term <u>type I error</u>. The level of significance, α, is often presented as P(type I error). That association is easily formed in students' minds and seems to be the root of subsequent confusion. Although α is a well defined <u>conditional</u> probability, the expression "type I error" is <u>not</u> conditionally phrased. Students know, however, that rejection of H_0 and making an error are involved. This noncommittal phrase leaves the exact combination of the two events (whether their conjunction or one given the other) open to different interpretations. Now, when H_0 is rejected and we wish to ascertain the probability of error, we ask ourselves what kind of error it would be. Naturally, the loose concept "type I error" comes immediately to mind. As memorized, the probability of that "event" is α, so we believe that by this we have answered our question. The crucial distinction between the two opposite directions of the conditional probabilities has been blurred via the mediation of the ill defined expression "type I error". Actually, this mediating term should not be used on its own, independently of its probability. As textbooks and teachers sometimes justly point out, a "conditional event" $A|B$, is not a legitimate concept by itself. It is only conditional probabilities that are unequivocally defined as ratios of event probabilities. "Type I error" is an unfortunate statistical term. While the equality P(type I error) = α could still be tolerated (provided it is interpreted correctly), "type I error" should by not means escape the parentheses to lead an independent life. We would have fared better without that verbal term since $P(R|H_0)$ is unequivocal.

Apparently, in other areas, functionally analogous terms play a similarly adverse role: A test's accuracy in connection with medial diagnosis, and false-positive in the context of signal detection, are likewise potential mediators of the confusion of the inverse. One way to reduce the risk of confusion would be to dispense with such short-cut terms, that are unconditionally worded. Instead, we should strictly adhere to the symbolic language of conditional probabilities. Another didactic device is to present the problem's data in a two-dimensional frequency table, so that the two orthogonal directions for the computation of the two inverse conditional probabilities will be conspicuous.

References

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. Cognition, 11, 109-122.

Betteley, I.G. (1979). Ambiguity in probability problems. The Mathematical Gazette, 63, 16-19.

Eddy, D.M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic and A. Tversky (Eds.) Judgment under uncertainty: Heuristics and biases (pp. 249-267). New York: Cambridge University Press.

Falk, R. (1978). Problems in probability and statistics, Part II (in Hebrew). Jerusalem: Akademon.

Falk, R. (1979). Revision of probabilities and the time axis. Proceedings of the third international conference for the psychology of mathematics education (pp. 64-66). Warwick, England.

Falk, R. (1983). Experimental models for resolving probabilistic ambiguities. In Proceedings of the seventh international conference for the psychology of mathematics education (pp. 319-325). Israel.

Falk, R. (1986). Misconceptions of statistical significance. Journal of Structural Learning, 9, 83-96.

Freund, J.E. (1965). Puzzle or paradox? The American Statistician, 19, 29.

Gardner, M. (1959). The Scientific American book of mathematical puzzles and diversions. New York: Simon and Schuster.

Glickman, L.V. (1982). Families, children and probabilities. Teaching Statistics, 4, 66-69.

Mosteller, F. (1965). Fifty challenging problems in probability with solutions. Reading, MA: Addison-Wesley.

Pauker, S.P., & Pauker, S.G. (1979). The amniocentesis decision: An explicit guide for parents. In C.J. Epstein et al. (Eds.), Birth defects: Original article series (Vol. 15). New York: The National Foundation. 1979.

Stern, C. (1960). Principles of human genetics. San Francisco: Freeman.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgment under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillsdale, N.J.: Lawrence Erlbaum.