1-1-2003

Statistical Probability as a Metaphor for Epistemological Probability

L. David Ritchie
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/comm_fac

Part of the Communication Commons

Citation Details

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Communication Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Running head: STATISTICAL PROBABILITY AS A METAPHOR

Statistical Probability as a Metaphor for Epistemological Probability

David Ritchie

Portland State University

*Metaphor and Symbol, 18*(1), 1-12, 2003.
Abstract

The metaphor, *epistemological probability is statistical probability* is traced to the origins of probability theory. Related metaphors appear both in everyday discourse about social processes, and in social scientific argumentation, often disguised as literal claims. Gambling provides a familiar vehicle for expressing the uncertainties associated with social interactions, and with social science research. Examples are drawn from reports of research on media effects, from everyday conversations, and from political communication.
Statistical Probability as a Metaphor for Epistemological Probability

Statistical probability (the observed frequencies of outcomes in random processes) is frequently conflated with epistemological probability (the degree of confidence in an expected outcome) and related uncertainty about social processes and interactions. One often hears phrases such as “I’m ninety percent sure I won’t be at the party,” or “the odds are better than fifty-fifty that you will be promoted.” This kind of conflation seems to stem from a metaphor that has come to be taken literally, a metaphor that is more apparent in, "Going anywhere with her is always a gamble," "I think he's a poor bet for the assignment," and the succinct response to an expressed hope or expectation, "Don't bet on it." The usage is clearly metaphoric in such cases, consistent with root metaphors such as “SOCIAL INTERACTIONS ARE RANDOM EVENTS,” or "PREDICTING SOCIAL INTERACTIONS IS A GAMBLE." As Vervaeke and Kennedy (1996) show, it is inadvisable to make the stronger induction, based on the apparent similarities, that these metaphors all stem from a particular root metaphor.

The conflation of statistical and epistemological probability, and the underlying metaphoric relationship between the two concepts, can be traced to the beginnings of modern science, when evidence based on direct observation replaced deduction from first principles as a basis for knowledge. Epistemological probability first served as a metaphor for statistical probability, but the wide-spread adoption of statistical methods as a basis for scientific argumentation has reversed our understanding, so that we now use statistical probability as a
metaphor for epistemological probability. The conflation of epistemological with statistical probability causes little difficulty in everyday usage, and in many instances may provide a useful way to express complex ideas, but when it appears in scientific writing it can lead to confusion.

**Probability: The Root Metaphor**

Hacking (1975) identifies two distinct concepts of probability, "the degree of belief warranted by the evidence," (epistemological probability) and "the tendency, displayed by some chance devices, to produce stable relative frequencies" (stochastic or statistical probability). The epistemological concept is the older by several centuries, dating back to the medieval distinction between "high sciences" and "low sciences," between "knowledge" and "opinion." The high sciences were characterized by deduction from first principles, and true knowledge could be achieved in no other way. The low sciences (e.g., alchemy, geology, astrology, and medicine) were characterized by inductive reasoning from the external evidence of signs and the external testimony of experts. Since logical demonstrations were impossible in the low sciences, they could deal only in opinion.

Originally, probability is an evaluation of opinion: The word stems from the same root as probe, probate, prove, and approval. During the Middle Ages, the opinion of experts was valued more highly than evidence from direct observation, and the opinion of experts was itself evaluated according to its basis in the authority of ancient books. Thus, according to the Medieval view, "Opinions are probable when they are approved by authority, when they are testified to, supported by ancient books" (Hacking, 1975, p. 30). But in the Renaissance, nature
itself came to be regarded as a book, and the "BOOK OF NATURE" as the testimony of the highest of all possible authorities. The signs that can be "read" in the "BOOK OF NATURE" have great probability because they come from the "author of nature," the highest of all authorities. Thus, in the "low sciences," such as alchemy and medicine (and, later, all the natural sciences), observation of signs was conceived as "reading testimony from the book of nature." (This metaphor appears in regular use as late as Darwin, and is still encountered in some contemporary scientific writings.) The concept of signs was transformed into the concept of evidence, and evidence almost entirely supplanted the approval of authority as a basis for evaluating the probability of propositions. However, signs themselves vary in their reliability, and diverse signs often lead to opposite conclusions. For example, Millman and Smith (1997) discuss an instance in which Darwin used the analogy between the actions of animal breeders and the formation of varieties in nature to illustrate opposite ideas. Darwin first emphasized the contrast between these two sources of variation within species then, as the concept of natural selection took shape, he emphasized the similarities. Given the uncertain reliability of signs, any argument from signs requires an assessment of their relative weight, that is, of their relative probability.

Randomizing devices have been associated with knowledge, in the form of Shamanic and fortune-teller’s practices, throughout history. However, the modern connection between evidence and randomizing devices appears to originate with Leibniz. The desire to gain an advantage at the gaming tables had stimulated a series of mathematical theorems about the
distribution of outcomes of randomizing devices such as dice. Leibniz, trained in the law, undertook to apply the results of these studies to the task of calculating the weight or "probability" of evidence so as to assess the relative degrees of proof in legal cases. Leibniz equated probability with possibility, and defined probability as a ratio among equally possible events. "Leibniz had learned from the law that probability is a relation between hypothesis and evidence. But he learned from the doctrine of chances that probabilities are a matter of physical propensities. Even now no philosopher has satisfactorily combined these two discoveries" (Hacking, 1975, p. 139).

Probability as a distribution of outcomes has since come to be associated with probability as a relationship between hypothesis and evidence in at least two ways. The first is in the formal logic of hypothesis testing, which follows the form of a reductio ad absurdum argument – but takes the actual form of an implicit wager. The second is in reasoning about the myriad of unknown and unknowable causal factors that contribute to any observed event, especially in research on social interactions, where the causal factor that is of interest to the researcher is conceived as only one among a multitude of causal factors.

In hypothesis testing, the convention is to pose a null hypothesis as an alternative to the expected relationship. The observed frequencies are then analyzed for consistency with the null hypothesis. If it is sufficiently improbable (statistically) that random events could produce the observed frequencies or correlations (the null hypothesis), then it is deemed more probable (epistemologically) that some alternative to the null hypothesis is correct. The statistical
probabilities associated with the null hypothesis are then used to argue in favor of the test hypothesis.

A direct example occurs in medical research: Will a flu vaccine actually prevent me from getting the flu this winter? To test the hypothesis that it will, medical researchers take a sample of patients and assign, say, 100 people to receive the vaccine, while 100 others receive a placebo. They then observe what happens during the flu season. If, say, thirty people in the treatment group and sixty people in the control group catch the flu, we conclude that I am twice as likely to catch the flu if I do not get the shot as I am if I do. But, the reasoning goes, how do I know for sure? Isn’t it possible that 90 people were going to catch the flu anyway, and 60 of them just happened to get assigned to the control group? To test this “null hypothesis” (that the observed differences were due only to chance), we calculate the mathematical probability that 60 (or more) of the 90 people who were going to catch the flu in any event would be assigned to the control group purely by random chance. It turns out to be very unlikely. By convention, we say that people in the control condition were twice as likely to come down with the flu, p < .001, which translates to: If the null hypothesis is true, and there is no relationship between getting vaccinated and catching the flu, then we would expect to observe a difference this large or larger less than one in one thousand times.

That’s why it’s not a true reductio ad absurdum argument, but rather an argument based on a gamble. In a true reductio ad absurdum argument, each of a finite set of alternatives to a proposition is considered, in turn. Each alternative is assumed to be true, and it is shown through
deductive logic that a logical contradiction (an absurdity) results. If every possible alternative leads to an absurdity, then the remaining proposition must be true. However, in the case of hypothesis testing, if the observations are taken from a sample of a very large population, then even the most extreme findings do not lead to logical contradiction. Even if all ninety flu victims were in the control condition, the null hypothesis would not lead to a logical contradiction. It is not absurd to assert that 90 subjects who were destined to catch the flu in any event may have happened, by chance, to be assigned to the control condition – it is merely very unlikely. With odds better than 1000:1, if you’re going to bet, it’s smarter to bet on the test hypothesis (the vaccine does have some effect) than on the null hypothesis. Incidentally, if the vaccine is expensive or has serious side-effects, or if the difference between the test and the control conditions is small, even at 1000:1 odds it might not seem like a good bet.

Some research questions in social science can be answered through similar experimental designs. Consider an example from the study of media effects. A researcher might wish to test a theoretical expectation that exposure to explicit depictions of sexuality in a laboratory setting will be associated with subsequent expression of negative attitudes toward women. An experiment is devised in which some men are exposed to sexual depictions (the test subjects) and others are not (the control subjects). If more of the men who were exposed to sexual depictions subsequently express negative attitudes toward women, the researcher asks, "How likely would it be that I would observe this much of a difference just because of random coincidence if there were indeed no underlying relationship (i.e., if the null hypothesis were true)?" The researcher
then calculates the statistical probability of observing a relationship this strong purely by chance, points out how improbable it would be, then argues for the epistemological probability that the preferred hypothesis is correct.

Several assumptions are imbedded in this convention. First and foremost is the concept of "random chance": The theorist explicitly acknowledges the possibility that random events could lead to results that look very convincing. These random events are most frequently described in terms of the techniques used to assign subjects to either the test or control condition. If some of the men happen to have negative attitudes that predate the experiment, and we toss a coin or use some other randomizing device to assign them to either the test or the control condition, we expect that, on average, the men with prior negative attitudes will be equally likely to end up in either condition. We recognize that even a fair coin can come up heads five, seven, or any other number of times in a row (cf. Tom Stoppard's, 1990, Rosencrantz and Guildenstern are Dead), so there is always some possibility, \( p > 0 \), that even the most extreme results are due strictly to chance. However, we know how to calculate the statistical probability of any string of coincidences, five, ten, or even a thousand heads in a row. By extension, we know how to calculate the probability that some disproportionate share of subjects who harbor negative attitudes toward women will be assigned to the test condition by chance.

Readers are expected to accept the low statistical probability of observing such results in a random process and the associated high probability that the results were due to something other than a random process, as evidence for the epistemological probability that the hypothesized
relationship really exists. As with the flu shot, the argument takes the form of *reductio ad absurdum*, but it is literally a gambler’s argument: If you’re going to bet your reputation (as a social scientist or as a policy-maker) one way or the other, you’re better off accepting the test hypothesis rather than the null hypothesis.

It can be argued that we are not interested in *truly* random events. Indeed, it is not clear that we know what a truly random event would be. (Space does not permit exploration of the fascinating literature on this question.) Nor are we concerned with quantum effects, in which what we know of a particle is itself expressed in terms of a distribution of possible states. If we accept the assumption of total causality (at the scale of everyday events such as coin tosses), then the fall of a coin is causally determined by the action of air molecules, photons, gravitational fluctuations, and so on. A similar array of events influences any other randomizing device. In this sense, the concept of “random event” simply means that we believe that it is impossible to *know* the net result of all these causal influences: It can only be *estimated*, in the aggregate, by some statistic such as the mean of a set of previous events.

Similarly, we know that subjects’ responses to questionnaires about their attitudes toward women can be influenced by many events, including snippets of conversation overheard before entering the laboratory or classroom, the form of the question itself, and so on. We are interested in differentiating the influence of theorized processes (the influence of sexually explicit media on men’s attitudes) from all the untheorized social and psychological processes that might be at work. This is the second way in which the “random process” enters social scientific thinking:
We lump all the unobservable causal factors into a category of “random noise,” or “error variance,” then use statistical techniques based on the assumption that their net result is randomly distributed to distinguish the effects of the hypothesized causal factor from the effects of all the other possible causal influences. Note that “random noise” is itself doubly metaphorical: error variance (the influence of all the unobserved causal factors in an experiment) is expressed in terms of “the random perturbations of a radio transmission” that block clear transmission of the signal, itself expressed in terms of the “distracting sounds” that make it difficult to carry on a conversation on a busy street or at a cocktail party. The “random noise” introduced by unknown and untheorized causal influences is subsumed into “sampling error” by the assumption that a subject susceptible to any particular untheorized influence is equally likely to be assigned to any condition.

“Gambling devices,” like “distracting sounds,” provide a familiar metaphor for the otherwise difficult-to-grasp concept of untheorized and unknown influences on subjects’ behavior. Gambling devices are ubiquitous and virtually everyone has had extensive experience with a variety of such devices (dice, coin tossing, shuffling decks of cards, spin-the-bottle, etc.). Comparing an outcome to drawing three aces in a poker hand or having three dice land with sixes showing enables even the less mathematically gifted to grasp the concept. The untheorized social and psychological processes that influence a subject’s responses on a questionnaire are similar to the “untheorized and unobservable physical processes that influence the fall of a tossed coin.” We summarize all of the untheorized events, some of which we could conceivably
measure but most of which we could not, within the concept of random or chance events, and understand them in terms of the familiar realm of "dice, roulette wheels, coins, and other gambling devices."

Statistical Probability as a Metaphor for Epistemological Probability

Some of the most interesting questions in social science do not lend themselves to straightforward experimental tests. In media effects research, to continue with the same example, it is often impractical, unethical, or even illegal to observe the variables that are of genuine interest. What the theorist really cares about is whether exposure to pornography leads to sexual harassment, rape, or worse - but it would be unacceptable to set up an experiment in which any of these actions could be observed. Consequently, the experimenter must think of some behavior that can legally and ethically be observed, then generate an argument linking the observable (but intrinsically uninteresting) behavior to the interesting (but ethically unobservable) behaviors.

For example, Zillmann and Bryant (1982) randomly assigned college students to watch innocuous films or no films at all (the control conditions) or a moderate to heavy diet of short, scriptless, sexually-explicit “stag films” (the experimental conditions) over a period of six weeks. Among other things, their results show that exposure to stag films increases the likelihood that college men will express agreement with a series of crude statements about women, and reduces the jail sentence recommended by college men and women for a convicted rapist, based on a newspaper account of the rape trial. There was no subject attrition, and the
results are statistically significant: The observed statistical improbability of the null hypothesis provides a basis for arguing that the observed differences in responses of subjects in the control and experimental conditions are the result of something other than random assignment. Thus far, the relationship between statistical probability and epistemological probability is straightforward – but it doesn’t answer the research question.

To bridge the gap between the results of their experiment and their research question, Zillmann and Bryant argue that watching the stag films in the laboratory is a good proxy for watching sexually explicit films in the movie theater or on TV, and that recommended jail sentences and scores on the sexual callousness scale are good indicators of callousness toward women and good predictors of how men will actually behave toward women. These arguments are supported by interpretive reasoning, based on thematic similarities and face validity, according to accepted conventions of social science. However, neither the interpretations proposed by Zillmann and Bryant nor any possible alternative interpretations can be tested statistically: There is no way to assign statistical probability to the truth of the theory itself. To provide a definitive answer to their research question, Zillmann and Bryant need to be able to make a claim such as, “Results this strong or stronger would be observed less than one time in a thousand if exposure to sexually explicit films in movie theaters or on television does not cause abusive actions toward women in real life.” However, the evidence produced by their experiment warrants no such claim. From the results reported by Zillmann and Bryant, we can estimate the statistical probability that the observed behavior (responses to pencil and paper
questionnaire items) will follow exposure to stag films in a laboratory in future circumstances similar to those in Zillmann and Bryant’s experiment. We have no way to estimate the statistical probability that actual abuse (physical or verbal) will follow exposure to stag films in real world conditions.

The experimental results do not test the theorized real-world relationship; they serve only as "signs," evidence in an argument supporting the theorized relationship. The theoretical claim cannot be evaluated, on the basis of the evidence obtained from the experimental results, in terms of statistical probability; it can only be evaluated in terms of epistemological probability. It is important to distinguish carefully between the statistical probability that results are due to something other than sampling error or random assignment error and the epistemological probability of a preferred theoretical interpretation. It would be erroneous and misleading to cite the statistical significance of the experimental observation in support of a theoretical interpretation, or indeed in support of any argument beyond the claim that the observed differences result from the experimental condition and not from sampling error.

As the controversy following publication of Zillmann and Bryant’s 1982 article demonstrates, it is often quite difficult to apply even a semblance of reductio ad absurdum argument to a theoretical interpretation. Two critics of their article, Christensen (1986) and Brannigan (1987), both offered reasonable alternative interpretations for Zillmann and Bryant’s findings. In their reply to these critics, Zillmann and Bryant (1986; 1987) defended the quality of their evidence by reiterating the statistical significance of their observations, but they made
little attempt to refute the alternative interpretations. Conceivably, the alternative interpretations
could themselves be tested through further experimentation – but the results of these further
experiments would also be susceptible to competing interpretations. Moreover, if indeed there is
an innumerably large number of potential causal influences on any social action, an innumerably
large number of reasons why a particular person reacts in a certain way, there is also an
innumerably large number of potential interpretations of any research finding. To speak of
probability in such circumstances can only be to speak epistemologically – or metaphorically.

Probability as a Metaphor

Hacking (1975) distinguishes two senses of "probability," statistical probability and
epistemological probability. The concepts of statistical and epistemological probability have a
complex history together, each has played a central role in the development of modern scientific
reasoning, and their roles have effectively reversed.

"Epistemological probability" began as a metaphor for statistical probability. But as the
role of authority in scientific argument waned and the role of observed regularities waxed,
statistical probability assumed the pre-eminent place. Indeed, in the thinking of many scientists
and philosophers of science, there was for a long time no room for any form of probability that
could not be anchored in mathematical reasoning: Statistical probability was the sole legitimate
form of probability, the sole basis for knowledge. Consequently, "statistical probability" - and
the associated world of "randomizing devices" - has become a metaphor for epistemological
probability. This is even more true in everyday language than it is in scientific discourse (where
precise use of language is more highly valued).

As noted before, people routinely say things like "The odds are better than fifty-fifty that you will be promoted," or "I stand behind Senator Eagleton one thousand percent." In these instances, there is no basis for assigning statistical probabilities. I might conceivably be able to compute the proportion, among people in like situations, who have been promoted – although in fact that has rarely been done when the phrase is used. But even if it had been done, the qualifying phrase, “among people in like situations,” is so vague as to render the statistic meaningless. On the surface, these phrases mean little more than “I think it likely but far from certain that you will be promoted,” and “As of right now, I support my running mate without qualification.”

In each case, there is a causal chain, and the speaker has in mind at least some implicit theory about a causal chain, some implicit weighing of the signs, pro and con. Or to put it in different (non-causal) terms, there is a set of reasons, and the speaker has in mind at least some implicit hierarchy of priorities by which to weigh the reasons, pro and con. Many factors enter into a decision regarding a candidate for promotion: Scores on qualifying exams, fitness evaluations, the relative merits of other candidates, time in service, how the candidate is viewed by members of a review board, etc. Some of these can be quantified, some cannot. Approximate weighting can sometimes be estimated, but are often unknown or imprecise. In an objective, “ideal” situation, all factors would be quantifiable and exact factor weights would be known; the speaker would then be in a position to give an exact estimate as to the actual probability of
promotion. In the more typical actual situation, the weighting is done in an entirely intuitive way, and “The odds are better than fifty-fifty” may express the speaker’s intuitive assessment of the composite influence of all these factors. On the other hand, the statistical phrase may serve a different rhetorical purpose altogether. In the guise of an apparently objective appraisal, it may express the speaker’s wish to be encouraging, while preparing the candidate for the possibility of disappointment.

In the case of Senator McGovern, a range of possible causal factors – or reasons – were also at play, including Senator Eagleton’s many qualifications for the position of Vice President, Senator Eagleton’s political skills and attractiveness as a candidate, the public’s reaction to the revelation that Senator Eagleton had experienced – and sought treatment for – a bout of depression following an election, the reaction of the press, and so forth. Some of this could possibly be quantified and weighted – but given the uniqueness of the situation, any computation of probabilities would have been difficult if not impossible. As with promotion, many reasons enter into selection of a vice presidential candidate, and into the decision to support or drop a controversial nomination – too many reasons for exhaustive enumeration or precise statistical calculation.

The rhetorical situation was different, however, in that the strength of Senator McGovern’s support for his vice-presidential nominee was itself a major factor in the decision. Moreover, the way Senator McGovern handled the controversy was successfully framed by the press as a test of his own character and fitness for office. In the example of the candidate for
promotion, the primary audience is the candidate, but the primary audience for Senator McGovern’s pronouncement was the press and the general public. The phrase, “one thousand percent,” by overtly exaggerating the mathematical possibilities, implicitly denied any quantitative calculations of cause and effect, while explicitly denying even the possibility that Senator Eagleton might be replaced on the ticket. If, by doing so, Senator McGovern could have dampened the excesses of journalistic speculation about the incident, his pronouncement might have proven to be self-fulfilling, and he might have been able to avoid the politically embarrassing and personally painful result of accepting Senator Eagleton’s earlier offer to withdraw from the ticket. In the event, the portrayal of the incident in the press led to the decision by Senator McGovern and his staff, only four days later, to allow the Democratic National Chairman to make a public statement that Senator Eagleton should be dropped (New York Times, 1972). The rhetorical ploy then backfired on Senator McGovern by making him appear inconstant – and by drawing attention to the illogic of the phrase “one thousand percent.”

"Statistical probability" has come to be so deeply imbedded in our thinking that we scarcely notice it. It took a gross violation of mathematical possibility, followed by a negation-by-flip-flop, to make us notice the incongruity of Senator McGovern's usage with respect to Senator Eagleton. In each example, the statistical metaphors express the degree of the speaker’s confidence in the face of the unpredictability of social processes. The statistical metaphors also serve as a kind of shorthand for the complex web of reasons that underlie social processes, and
an approximation of the way these reasons are balanced in the decision-making process: In principle, the statistical metaphors could be replaced by an enumeration of the reasons, ordered approximately according to priority. However, the statistical metaphors also serve other rhetorical functions: To encourage a candidate for promotion without cultivating an excessive degree of confidence, and to dampen journalistic speculation about a possibly damaging revelation.

Upon occasion, the known world of "gambling devices" is deliberately used to help make the unknown and difficult world of causal relationships in social processes more comprehensible. "Gambling devices" provides a useful and powerful metaphor for the uncertainty of social processes, and informs our expectations about social outcomes. It provides a convenient way to summarize the complex process of weighing reasons for our decisions, and for representing the decision-making process itself as rational and objective. It also provides rhetorical tools that can be used, for example, to influence other people’s expectations about the outcome of decision-making processes (as in the example of the candidate for promotion), or to attempt to head off speculation about a decision-making process (as in the example of Senator McGovern and Senator Eagleton).

In social science, statistical probability also plays a double role, expressing both the distribution of possible random samples from a population and the distribution of possible outcomes of a non-enumerable set of influences on the outcome of an experimental process (the “random noise” in an experiment). Our knowledge about games of chance provides a useful
metaphor for the statistics of drawing a sample. To the extent that experimental outcomes are affected by the uncertainty of social processes (i.e., by the non-enumerable causal influences summarized as “noise in the data”), our knowledge about games of chance also provides a powerful metaphor for the uncertainty of social scientific research. If social life is itself a gamble, then conducting research on social life must also be a gamble.

The everyday use of the "gambling devices" metaphor may itself be influenced by scientific usage, but it more likely derives from ordinary encounters with games of chance, beginning in most cases in early childhood. By the time a future social scientist begins formal training, a complex set of metaphoric entailments is well in place. Under such circumstances, if interpretive reasoning is to be kept free of the conflation of epistemological probability with statistical probability, it may prove helpful to acknowledge the underlying metaphor, and it will certainly prove helpful to distinguish carefully between the two meanings of “probability.” We need to be clear when we are speaking metaphorically and when we are speaking literally, when we are using statistical probability as a shorthand way of describing an array of causes or reasons, variously weighted and when we actually have the capacity to compute the probability distribution of a range of outcomes. We should especially avoid the temptation to lapse into the language of statistical probability in any formal discussion of the strength of an argument.
Probability 21
References


