Probability Logic and Probabilistic Induction

Sander Greenland

This article reviews some philosophical aspects of probability and describes how probability logic can give precise meanings to the concepts of inductive support, corroboration, refutation, and related notions, as well as provide a foundation for logically sound statistical inference. Probability logic also provides a basis for recognizing prior distributions as an integral component of statistical analysis, rather than the current misleading practice of pretending that statistics applied to observational data are objective. This basis is important, because the use of realistic priors in a statistical analysis can yield more stringent tests of hypotheses and more accurate estimates than conventional procedures. (Epidemiology 1998;9:322–332)

Keywords: Bayesian analysis, induction, inference, logic, probability, statistics.

Over the past two decades, there has been a dramatic resurgence of Bayesian philosophy and methodology in statistics, as reflected in recent textbooks[1–3] as well as journal articles. This resurgence has as yet had little impact in epidemiology, which instead has experienced lively arguments between Popperian and non-Popperian (but not necessarily Bayesian) philosophical positions (see, for example, the debates in the volumes edited by Greenland[4] and Rothman[5]). This gap between statistics and epidemiology is in part due to differing attitudes toward mathematics and computing, which often seem to be the pride and joy of statisticians but poorly connected to the underlying epidemiologic reality.

I here review some basic elements from the philosophy of probability which may be useful for bridging the gap. Central among these elements is probability logic, which provides an extension of deductive logic to reasoning under uncertainty[3,6,7] and which forms the basis of certain arguments given for inductive inference in Bayesian philosophy[6–8] and certain arguments against inductive inference in Popperian philosophy.[9,10] Because intuitive reasoning under uncertainty is poor[11,12] and because epidemiologic inference involves so many uncertainties (for example, about uncontrolled biases), one could argue that probability logic should be a centerpiece of epidemiologic training. Instead, the topic is absent from most epidemiology and statistics texts, even those that purport to address foundational issues,[13,14] while classes on probability and statistics usually cover only mathematical models for probability and the statistical methods derived from those models.

From the Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90095-1772 (address for correspondence).

Submitted February 18, 1997; final version accepted November 17, 1997.

Editors' note: See related editorial on page 233 of this issue.

Of necessity, the present review must be limited to just those elements essential for understanding the basic issues. It skips many concepts and viewpoints entirely; it is also ahistorical, even though the history of the ideas is illuminating. For more thorough coverage, one may consult any of a number of philosophic treatises.[7,15–17] Hacking[18] provides a superb history of the early origins of probabilistic concepts and controversies, while Lad[3] provides many interesting details of subjective Bayesian history.

Probability Logic

In a companion paper, I have tried to document that there are ambiguous and contradictory definitions for the word "induction."[19] Notions of probabilistic reasoning suffer from at least as many problems, as witnessed by the controversies surrounding the foundations of probability and statistics[3,7,8,14–18] (controversies long kept hidden from students, lest standard analysis methodologies be called into question[14]). Nonetheless, beyond these controversies one may discern a logical foundation for deriving uncertain conclusions from uncertain premises when certainties are measured by probabilities. The present section outlines that foundation.

Definitions of Probability

There are two major classes of probability definitions, "objective" and "subjective." Within these classes there are many variants; this is especially true of "objective probability," which subsumes frequency, propensity, fiducial, and necessarist or logical probability (confusingly, "logical probability" is only a special case of probability logic). I will represent the dichotomy among definitions prevalent today by the two most common definitions in statistics, the frequency and the subjective Bayesian definitions.

The *frequency* or *frequentist* definition asserts that probabilities are limits of sequences of relative frequencies (proportions) of events. Because relative frequencies are observable, limits of such sequences are purported to be physical properties of systems or mechanisms that generate sequences of events; hence, frequency probabilities are sometimes called *physical probabilities*. In its most pure form, frequentist theory denies any meaning to probabilities of individual events, such as the outcome of a given coin toss or patient.[20] This limitation of the theory has led to the development of theories of physical probability that allow individual probabilities, such as propensity theory.[16,21]

Individual probabilities are also allowed under the *subjective Bayesian* or *personalist* definition, which treats probabilities as constructs of an observer's mind. These constructs are supposed to correspond to the observer's "rational certainty" about a statement, where rational certainty means only that the certainties are constrained to follow the axioms of probability.[7] Because such probabilities vary from person to person, they are sometimes called *personal certainties*, *credibilities*, *personal probabilities*, or *degrees of belief*.

The terms "objective" and "subjective" confer some misleading connotations that tend to bias naive readers away from the subjective view.[8] For example, the word "subjective" suggests elements of arbitrariness or irrationality, whereas knowledgeable critics of subjective Bayesian probability often complain that it is too stringent in its demands for rational probability assignment (see, for example, the discussion in Ref 22). In contrast, the word "objective" suggests direct observability (like the height of Mount Everest); nonetheless, *limits of sequences* of relative frequencies are defined in terms of infinite sequences, which are *not* directly observable. This metaphysical property of frequentist probabilities is usually overlooked; instead, such probabilities are commonly described as referring to "the long run," which is rarely given a precise definition.[7] Other theories of objective probability share this metaphysical character, including propensity theory.[21]

It should be noted that the two definitions of probability just described are not mutually exclusive: Some authors believe that physical probabilities exist and can be estimated, but also use subjective probabilities to measure both personal degrees of uncertainty and physical probabilities.[7,8,23,24] Problems arise only because measures of physical probabilities (such as traditional *P*-values and confidence limits) are routinely misinterpreted as measures of uncertainty about hypotheses.[13,14]

Despite the compatibility of objective and subjective theories of probability, the metaphysical character of objective theories has led some Bayesians to deny the very existence of "objective" probabilities.[3,6] One such argument is roughly as follows: "Limits of sequences of relative frequencies" are really only mental constructs that are built and modified to follow observed relative frequencies; that is, so-called "objective probabilities" are only subjective probabilities constructed to mimic the magnitudes of observed proportions. A related objectivist view is that physical probabilities are never more than theories about how certain relative frequencies will unfold.[21] As such, they cannot be validly deduced from observation of event frequencies (for the same reason that no general theory can be validly deduced from observations alone). Consequently, "objective probabilities" can never be established as facts; they are instead hypothesized laws governing physical behavior, or hypothetical properties (propensities) of certain types of objects.[16,21] To complicate matters further, there are forms of Bayesian statistics and probability logic that are based on "objective" probability theory[20]; hence, there is a need to distinguish "objective" from "subjective" Bayesianism.[7] The latter has become so influential today, however, that most modern discussions of Bayesianism and probability logic (including the present one) focus on the subjective variety.[3,6–8]

Subjective probabilities can apply to statements about events, and so are often confused with physical probabilities. As an example, suppose you read a report of a randomized trial. The investigators might be 95% certain of truth of the statement: "The lower and upper 95% confidence limits for the risk difference contain the true effect." If derived from a well conducted randomized trial, that may be a perfectly reasonable subjective probability to have if there are few other data available. Nonetheless, following standard frequentist theory, the physical probability that those limits contain the true value is either one (if they do contain the true value) or zero (if they do not). There is no conflict here; the sense of conflict arises in part because, in ordinary language, the *event* in question (true effect between the 95% limits) must be described by a *statement* that the event occurred. This statement is not usually set off by quotation marks, as done here. Thus, ordinary language and ordinary thinking do not distinguish the *event* (to which the physical probability refers) and the *statement describing the event* (to which the subjective probability refers). This distinction is important, however, for avoiding misuse of frequentist statistics such as *P*-values.[7]

Axioms of Probability

Despite confusion and conflict, virtually all writers agree that probabilities should follow a few simple axioms and definitions. From those rudiments, one can validly deduce a vast body of logical consequences, known as probability theory. This axiomatic and definitional agreement results in many parallel structures in objective and subjective probabilistic systems, despite the fact that the objective system refers to the physical world and the subjective system refers to mental worlds. The key difference is that physical probabilities can apply *only* to physical events or states, whereas subjective probabilities can apply to any precise declarative statement, whether it concerns physical events or states or a hypothesis that expresses a general law of nature.

The first axiom requires probabilities to be nonnegative. The second axiom requires that every logically inevitable event (in objective terms) or tautology (in subjective terms) have a probability of one. For example, the statement "The confidence interval either will or will not contain the true value" is a tautology (that is, it is logically inevitably true, regardless of any facts), and the event it describes is logically inevitable; the statement and event cover all possibilities. Therefore, the objective probability of the event (if it exists) *must* be one; analogously, according to the subjective theory, we *should* set our subjective probabilities for the statement to one.

The third axiom requires that if A and B are mutually exclusive, the probability of "A or B" must equal the sum of the probability of A and the probability of B. Two events are mutually exclusive if no more than one of them can happen; two statements are mutually exclusive (or mutually inconsistent) if no more than one of them can be true. For example, the following two statements and the events they describe are mutually exclusive: "The observed risk difference will be greater than the true effect" and "the observed risk difference will equal the true effect." Axiom 3 asserts that the sum of the objective probabilities of these two events must equal the objective probability of the event described by "the observed risk difference will be greater than or equal to the true effect" (if these probabilities exist). In parallel, Axiom 3 asserts that we should set our subjective probabilities so that the sum of probabilities for the first two statements equals the probability of the last statement.

To summarize, let Pr stand for probability, and let A and B stand for any two events (in the objective theory) or statements (in the subjective theory). The above three axioms then assert that objective probabilities *do* satisfy and subjective probabilities *should* satisfy

A1) $\Pr(A) \geq 0$

A2) $\Pr(A) = 1$ if A is a tautology

A3) $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$

if A and B are mutually exclusive.

(Percentages can be used in place of proportions; Axiom 2 then asserts that Pr(A) = 100% if A is logically inevitable.)

Justification of the Axioms

Axioms 1 and 2 have almost no content given Axiom 3. In essence, they assert only that we should measure all probabilities (whether frequencies or certainties) on a proportion scale ranging from 0 = never or impossible to 1 = always or inevitable. We can always do so: For example, if we measure our certainties about statements using odds, we need only divide our odds by one plus the odds to transform our certainties to a 0-to-1 scale.

For frequentist theory, the above three axioms are assertions about how physical probabilities behave. When expressed as proportions or percentages of a total, common physical quantities (for example, counts, areas, weights) obey the above axioms: Proportions are nonnegative (Axiom 1), the total is 100% of itself (Axiom 2), and the proportion or percentage contributed by two nonoverlapping (exclusive) parts of the total equals the sum of the proportions or percentages contributed by each part separately (Axiom 3). For example, if 30% of the marbles in a bag are red and 20% are blue, then 20% + 30% = 50% of the marbles are red or blue. Because physical probabilities are limits of sequences of relative frequencies and the latter are proportions, such probabilities must have the same bounds and additive behavior as proportions.

For the subjective theory, the axioms are normative rules about how we *ought* to constrain our personal probabilities. Several justifications have been offered for these constraints.[3,6,7,23,25,26] An argument paralleling the frequentist justification is the following: Suppose we believe that a given event has a physical probability, and we know that probability or have a generally accepted estimate of it, as with certain games of chance and with quantum events. It has been proposed that we should then set our subjective probability for the statement of the event to the known physical probability.[7,8,23] This rule or axiom has been called the *Principal Principle* of subjective probability.[7,23] Because physical probabilities (if they exist) obey the above three probability axioms, we should make sure that our subjective probabilities do so as well in order to ensure that our predictions are as *well calibrated* as possible, that is, to ensure that our predictions of future event frequencies (for example, incidence rates) are as close as possible to the event frequencies that actually come to pass.

The Dutch Book Argument

The axiom justifications given above will not do for those who deny that physical probabilities exist. There is, however, yet another rationale for Axioms 1–3, called the "Dutch Book argument."[6,7,16] The premise of this argument is that you are willing to "put your money where your mouth is," in the following sense. Let us say you would bet on your probability assignments, up to a total stake of s dollars per assignment, if for each assignment Pr(A) you would accept either of the following bets, at my choice:

1. You would wager $\$s\Pr(A)$ on A true against my $\$s(1 - \Pr(A))$ on A false;
2. You would wager $\$s(1 - \Pr(A))$ on A false against my $\$s\Pr(A)$ on A true.

In other words, you offer betting odds of $\Pr(A)/(1 - \Pr(A))$ on A true, and will bet either way at those odds as long as the total money at stake does not exceed a certain amount. For example, suppose A is the statement that a given study will exhibit a negative association, you assign Pr(A) = 0.60, and you are willing to bet up to a total stake of a dollar per assignment. You would then be willing to wager 60 cents on a negative association against my 40 cents on a nonnegative association, and equally willing to trade sides by wagering 40 cents on a nonnegative association against my 60 cents on a negative association.

Ramsey[25] and DeFinetti[26] made the following discovery: If you are willing to bet on your assignments and your assignments violate any of the three probability axioms, it will be possible for other people to set up a system of bets against you, based on *your* probabilities, such that they will be guaranteed to win money from you *no matter what the truth of the statements in the bets*; in other words, you can be forced into sure loss (Appendix 1 gives a brief proof of this result). Conversely, if your assignments obey the probability axioms, no one will be able to force your loss with a system of bets based on your probabilities. (A system of bets that forces loss on a bettor is called a "Dutch Book.")

Some writers find the Dutch Book argument so compelling that they define a system of personal probability assignments to be *coherent* if and only if it satisfies Axioms 1–3.[3,6,7] Others dismiss the argument, commenting that it may be compelling when gambling, but science is about testing hypotheses, not gambling on them (see, for example, the discussion in Ref 22). Nonetheless, these critics tend to be believers in physical probabilities; for them, the good frequentist properties of Bayesian procedures can supply another rationale for the use of Bayesian statistics in random sample surveys and randomized trials.[22] It can also be argued that applications of science involve gambles on hypotheses; for example, in banning the asthma drug fenoterol, authorities would be gambling in favor of the hypothesis that the death rate would be lower with the ban than without.

In the objective theory, Axiom 3 is extended to include infinite sequences of events. Such a leap to the infinite has been resisted by some[3,6] but not all[7] subjective Bayesians. Fortunately, this divergence has no consequence for the present discussion and so will not be considered further.

Conditional Probability and Independence

In the objective theory, we *condition* on an event C by examining limits of relative frequencies among events of the form "A and C," where A is another event. This leads to the following axiom for the conditional probability of A given C, denoted Pr(A|C):

A4) If $\Pr(C) > 0$, $\Pr(A \mid C) = \Pr(A \text{ and } C)/\Pr(C)$.

In the objective theory, this axiom is usually referred to as a definition, but its justification is analogous to those for the other three axioms: Proportions and percentages of ordinary physical quantities will follow Eq 4. For example, if 60% of the marbles in a bag are red and 30% of the marbles are red and small, the percentage that are small among those that are red is 30/60 = 50%.

One interpretation of Eq 4 is as an axiom that shows how to modify or update our certainty about statement A if we learn that C is correct.[7] For example, suppose A is the hypothesis that "first-trimester retinol supplements can increase risk of limb reduction defects (LRD) in humans" and C is the hypothesis that "retinol supplements can increase LRD risk in rats." If you were 50% certain that rats can be affected (C) and 40% certain that the LRD risks of both humans and rats can be affected (A and C), Axiom 4 instructs you to become 0.40/0.50 = 80% certain that humans can be affected (A) if you are given that rats can, indeed, be affected (C). Like the other three probability axioms, Axiom 4 can be justified by a Dutch Book argument.[7]

An event or statement A is *independent* of another event or statement B if conditioning on B does not change the probability of A: Pr(A|B) = Pr(A). If A is independent of B, then B is independent of A, so the independence relation is symmetric. That is, if Pr(A|B) = Pr(A), then

$$\Pr(A \text{ and } B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}\Pr(B) = \Pr(A \mid B)\Pr(B) = \Pr(A)\Pr(B),$$

so

$$\Pr(B \mid A) = \Pr(A \text{ and } B)/\Pr(A) = \Pr(A)\Pr(B)/\Pr(A) = \Pr(B).$$

In the subjective theory, independence of A and B is a property of your probability assignment; it means that learning B is true will not alter your probability for A, and that learning A is true will not alter your probability for B.

Finally, A is conditionally independent of B given another event or statement C if further conditioning on B does not change the probability of A after conditioning on C; that is, if Pr(A|B and C) = Pr(A|C). This relation is also symmetric, in that it implies Pr(A and B|C) = Pr(A|C)Pr(B|C) and Pr(B|A and C) = Pr(B|C). In the subjective theory, it means that learning B is true will not alter your probability for A and learning A is true will not alter your probability for B, once you are given that C is true.

Relations between Hypotheses and Observations in Probability Logic

For the remainder of the paper, let H stand for a hypothesis and let B stand for the outcome of an observation process, with both H and B of uncertain status *a priori*; that is, $0 < \Pr(B) < 1$ and $0 < \Pr(H) < 1$. In typical problems, H is an assertion of a causal relation, whereas B is a description of a study and the data it obtained. The following definitions qualitatively characterize the dependence of H on B within a system of probability assignments Pr( ):

1. B *proves* H means that B would render H certain: Pr(H|B) = 1.
2. B *supports* H means that B would raise the probability of H: Pr(H|B) > Pr(H).
3. B is *neutral* with respect to H means that B would not change the probability of H: Pr(H|B) = P(H); that is, H is independent of B.

4. B *undermines* or *countersupports* H means that B would reduce the probability of H: $\Pr(H|B) < P(H)$.
5. B *refutes* H means that B would render H certainly wrong: $\Pr(H|B) = 0$.

It is common to see "confirms" used as a synonym for "supports" and "disconfirms" used as a synonym for "undermines,"[7] but I feel these terms have connotations too suggestive of "proves" and "refutes."

In ordinary English, one could also use "corroborates" as a synonym for "supports," but Popper[21] established another meaning for qualitative corroboration in the philosophy of science:

6. B *corroborates* H means that finding B false would refute H: $\Pr(H|\text{not } B) = 0$.

Appendix 2 shows that corroboration defined in this manner implies but is not implied by support, and neither implies nor is implied by proof; thus, corroboration is a strong form of support, though not as strong a form as proof. Corroboration is closely related to traditional notions of prediction, defined by

7. H *predicts* B means that H certainly implies B: $\Pr(H \text{ implies } B) = 1$.

Because H certainly implies B if and only if B is certain given H, we could equivalently define "H predicts B" to mean that B is certain given H: $\Pr(B|H) = 1$.

A number of authors have attempted to define measures of support or corroboration. Although no measure has prevailed in the literature, it has been recognized that an appropriate measure would have to involve comparison of probabilities.[8,21] Nonetheless, some proposed measures (such as relative likelihood) are based on comparison of observation probabilities, rather than hypothesis probabilities.

Popperian writings often emphasize the roles of refutation and corroboration in scientific research, while criticizing notions of proof of hypotheses. From the perspective of subjective probability, refutation and corroboration are as criticizable as proof because they demand an absolute certainty (probabilities of 0 or 1). A strength of the subjective theory is that it provides precise concepts of support and countersupport without invoking absolute certainty.

Probabilistic Induction and Bayes' Theorem

The general idea of probabilistic induction is that observations may somehow induce observers to make probability assignments, or at least induce observers to change their assignments. Such ideas can be traced back to the 17th century[16,18]; since then, a number of concepts in probability logic have been interpreted as probabilistic induction.

One interpretation is that probabilistic induction is the process of setting our probabilities for an unobserved outcome to the frequency of that outcome in an appropriate set of previous observations. This concept has also been called statistical induction, although it was explicated by Hume long before statistics was established as a topic distinct from probability.[27] The concept long remained imprecise because of the vagueness surrounding the concept of "an appropriate set of previous observations."[25] In modern subjective theory, however, "appropriate" is defined in terms of *exchangeability*,[3,6] a concept that will be discussed below.

A simpler interpretation is that probabilistic induction corresponds to use of any probability theorem to update (change) one's probability assignments in light of new observations; that is, to compute new assignments conditioned on the new observations. In particular, in modern subjective theory, probabilistic induction usually refers to the deductive process of updating probabilities using *Bayes' Theorem* and its consequences. The theorem, also known as Bayes' rule or Bayes' formula, is a simple formula which shows how to move from an initial or *prior* probability $\Pr(H)$ for a hypothesis H to an updated or *posterior* probability $\Pr(H|B)$ based on a new observation B.[6,7] There are several versions of the theorem; the original form derived by Rev. Thomas Bayes, a contemporary of Hume's, goes as follows:

8. A new observation B should change the probability of H via multiplication by the factor $\Pr(B|H)/\Pr(B)$; that is,

$$\Pr(H \mid B) = \Pr(H)\frac{\Pr(B \mid H)}{\Pr(B)}$$

Proof:

$$\Pr(H \mid B) = \frac{\Pr(H \text{ and } B)}{\Pr(B)}$$

$$= \frac{\Pr(H \text{ and } B)/\Pr(H)}{\Pr(B)/\Pr(H)} = \Pr(H)\frac{\Pr(B \mid H)}{\Pr(B)}.$$

An immediate corollary is that B alters the probability of H by the same proportion as H alters the probability of B:

9. $\Pr(H|B)/\Pr(H) = \Pr(B|H)/\Pr(B)$.

Many other relations between observations and hypotheses can be derived from Bayes' Theorem. Here are some simple but nonetheless statistically useful examples:

10. B corroborates H if and only if H predicts B; that is, $\Pr(H|\text{not } B) = 0$ if and only if $\Pr(B|H) = 1$.
11. B supports H if and only if B is more probable under H; that is, $\Pr(H|B) > \Pr(H)$ if and only if $\Pr(B|H) > \Pr(B)$.
12. B undermines H if and only if B is less probable under H; that is, $\Pr(H|B) < \Pr(H)$ if and only if $\Pr(B|H) < \Pr(B)$.
13. B refutes H if and only if H predicts not-B; that is, $\Pr(H|B) = 0$ if and only if $\Pr(\text{not } B|H) = 1$.

Bayes' Theorem and its consequences provide basic logical connections between probabilities of data given hypotheses, which are the outputs of standard statistical methods, and probabilities of hypotheses given data, which are routinely requested by scientists and the general public. In particular, proposition 11 provides one way of making precise the commonsense notion that successful predictions support a hypothesis.

There are theorems more elaborate than Bayes' that may also be interpreted as forms of probabilistic induction; see, for example, sec. VI, Ch. 15, of Good.[8]

The Popper-Miller Argument

In a letter to *Nature* that inspired a small literature in philosophy (Refs 28–32 provide some examples), Popper and Miller[9] claimed to prove that "probabilistic induction" was impossible. The definition of "probabilistic induction" that they used was not any of those given above, however. In fact, Popper and Miller noted that corroboration as defined above *does* imply probabilistic support; they simply argued that probabilistic support is not the same as probabilistic induction (using *their* definition of induction).[10] As several authors have pointed out, the proof depends on the idiosyncratic definitions used by Popper and Miller[7,28–32] (for example, Howson and Urbach[7] describe those definitions as "eccentric" and "strange").

What Popper and Miller actually showed was that, if $Pr(H|B) < 1$ and $Pr(B) < 1$ (that is, if the hypothesis H is not certain given the observation B, and the observation B is not certain), then Pr(B implies H|B) < Pr(B implies H). Popper and Miller implicitly defined probabilistic induction to be the reverse inequality, Pr(B implies H|B) > Pr(B implies H); hence, *under their definition*, "probabilistic induction" is impossible. This result is simply irrelevant to the definitions of probabilistic induction given earlier; at best, it is another warning that controversies often arise from semantic divergences.[19] Popper and Miller did, however, make one concluding statement that is agreed upon by all discussants: "There *is* such a thing as probabilistic support,"[9,p.688] by which they meant that proposition 11 given above is a valid relation between predictions and hypotheses. For those who *define* probabilistic induction as the process of updating probabilities in light of new observations, Popper and Miller's statement is a startling concession of the possibility and existence of such inductive processes.[28]

Bayesian Statistical Analysis

There is nothing controversial about Bayes' Theorem and its consequences as mathematical formulas. What is controversial is the use of subjective probabilities in the theorem, especially probabilities of hypotheses.[22] Even if one accepts the Bayesian philosophy, however, there are two practical obstacles to its application: Computation of the unconditional probability Pr(B) of the observation B, and specification of the prior probability Pr(H) of the hypothesis H. In typical applications, evaluation of Pr(B) requires difficult integration, although modern computing developments have greatly diminished the importance of this obstacle.[1] In contrast, proposed solutions to the specification problem remain controversial. Interestingly, Bayes was well aware of the philosophical, specification, and computational difficulties raised by the theorem, which may explain why he refrained from publishing his essay.[33]

Suppose now we wish to compare the posterior probabilities of two hypotheses H_1 and H_0. One way to do so is to take their ratio, which is called their *posterior odds*,

$$Pr(H_1 \mid B)/Pr(H_0 \mid B)$$

Applying Bayes' Theorem to both the numerator and denominator of this ratio, we obtain

$$\frac{Pr(H_1 \mid B)}{Pr(H_0 \mid B)} = \frac{Pr(B \mid H_1)Pr(H_1)/Pr(B)}{Pr(B \mid H_0)Pr(H_0)/Pr(B)} = \frac{Pr(B \mid H_1)Pr(H_1)}{Pr(B \mid H_0)Pr(H_0)}.$$

The ratio $Pr(B|H_1)/Pr(B|H_0)$ in this equation is often called the *Bayes factor* comparing the two hypotheses, while the ratio $Pr(H_1)/Pr(H_0)$ of the prior probabilities is called their *prior odds*.[1] With these definitions, the preceding equation yields

$$\text{Bayes factor} \equiv \frac{Pr(B \mid H_1)}{Pr(B \mid H_0)} = \frac{Pr(H_1 \mid B)/Pr(H_0 \mid B)}{Pr(H_1)/Pr(H_0)} = \frac{\text{Posterior Odds}}{\text{Prior Odds}},$$

so that the Bayes factor measures the change in the odds of H_1 *vs* H_0 produced by observation B. These equations bring some computational and conceptual simplifications to Bayesian analyses; for example, when H_1 and H_0 represent one-point statistical hypotheses (such as "OR = 2" and "OR = 1" in a two-by-two table), the Bayes factor is the same as the *likelihood ratio* of ordinary statistics,[7,Ch.19] which in many problems is easily derived from standard statistical outputs. Nonetheless, the equations do not supply all the statistics one might want from an analysis, such as posterior (Bayesian) interval estimates.[1]

The Specification of Prior Probabilities

There are three major approaches to the specification problem. The first and oldest, dating back to Laplace in the 18th century,[33] is not really part of the modern subjective theory. It attempts to identify and specify "noninformative," "ignorance," or "reference" prior distributions.[34] For epidemiologic analyses, this approach usually yields numerical results close or equal to standard frequentist procedures; for example, the posterior probability intervals ("Bayesian confidence intervals") obtained in this manner are usually close or equal to standard confidence intervals. Their interpretations are entirely different, however; for example, 95% posterior probability limits of 1 to 3 for a risk ratio RR are a pair of numbers such that $Pr(1 < RR < 3) = 0.95$, where Pr ( ) refers to the analyst's probability assignment. In contrast, frequentist 95% confidence limits of 1 to 3 for RR have no analogous physical probability interpretation; although commonly misinterpreted as posterior probability limits, they are simply a pair of numbers that are either known to be generated by a random mechanism (if the study involved randomization) or not (if the study was purely observational).

Methods that employ reference priors are sometimes called "objective," "logical," or "necessarist" Bayesian methods, although here "objective" means only "agreed upon by convention." Such methods are arguably less logical and less scientific than subjective Bayesian methods.[3,7,24] Consider a simple epidemiologic example, that of coffee drinking and its association with myocardial infarction. No one on any side of the controversy has ever argued that drinking one cup a day would elevate rates by more than 10% (RR = 1.1), if at all. Yet the standard "reference" prior for the coefficient of coffee cups per day (for example, in a proportional-hazard model) assigns the same prior probability density to $\ln(RR) = \ln(1.1)$, $\ln(RR) = 10^{-100}$, and every other numerical possibility, such as $\ln(RR) = 10^{100}$. If $\ln(RR) = 10^{100}$, consuming a cup of coffee would usually lead to an immediate massive coronary. No one would give $\ln(RR) = 10^{100}$ any credence if they understood its substantive meaning. Nonetheless, some statisticians continue to use and promote so-called "noninformative" priors, which correspond to precisely these kinds of scientific absurdities. Their use is sometimes rationalized on the grounds that the resulting procedures are robust (see below), but robustness against absurd possibilities is unnecessary and often costly in terms of overall accuracy and credibility of results.[24]

Another faulty rationale for noninformative prior distributions is that "they allow the data to speak for themselves." In reality, data *never* speak for themselves: Every analysis has to filter data through some set of simplifying assumptions, such as assumptions that the data were generated by a conventional probability mechanism.[8,24,35-37] This problem is recognized in Popper's theme that all observations are theory-laden.[21] In non-Bayesian analyses, all assumptions are incorporated into and often hidden by models for data probabilities [for example, models for $\Pr(B|H_0)$ and $\Pr(B|H_1)$, where H_0 and H_1 specify parameter values in a logistic model]. Bayesian methods allow one to shift the form and burden of some of the assumptions to models for prior probabilities [for example, models for $\Pr(H_0)$ and $\Pr(H_1)$]; such assumptions can and should be checked against data, just as one should check models for data probabilities.[8,24,35]

At the other extreme from noninformative specification, there have been attempts to elicit detailed quantitative specifications of prior probabilities from scientific experts.[3,34] These are laudable efforts to operationalize the spirit of Bayesian philosophy. There are many practical drawbacks to this approach, however. First, it can require extraordinary effort on the part of both the experts and the statisticians, more than practical for routine use. Second, it will not produce a convincing analysis unless either (a) there is a high degree of consensus among all experts in the field under study, so that every member of the intended audience would accept the specification used, or (b) the results are reasonably insensitive to the prior specification, in which case the effort to make the latter precise was unnecessary.[1]

A third approach, which may be viewed as somewhere between the extremes of noninformative and detailed specifications, is to focus on incorporating accepted qualitative information into the prior specification, leaving quantitative details as unknown parameters to be estimated in a higher-stage model. Such hierarchical-Bayes approaches (also known as multilevel, empirical-Bayes, or random-coefficient modeling) have expanded in parallel with recent algorithmic and computing advances[1,34] and are well suited to many epidemiologic problems.[38-41] For example, in studies of diet, nutrition, and health, nutrient measurements are constructed from diet measurements in a linear fashion using nutrient tables, and this qualitative information can be used in modeling without precisely specifying prior probabilities for any effects.[41]

As with conventional (frequentist) analysis methods, a thorough Bayesian analysis must consider many issues, including *insensitivity* and *robustness*. A *result* is insensitive if it does not change much under reasonable changes in the analysis assumptions (of which the prior specification is but one), whereas a *method* is robust if the results it produces remain valid under reasonable departures from its assumptions. Insensitivity and robustness are related but do not imply one another: a nonrobust method may yield an insensitive result, and a robust method may yield a sensitive result. Furthermore, a robust method can be much less accurate than a nonrobust method that is well tailored to the topic at hand—which is another reason why the robustness of certain "objective" Bayes and frequentist methods is not a compelling argument in their favor.[24]

Exchangeability

Central to any serious attempt at probability specification is the concept of *exchangeability*.[1,3,6,42-45] Although this concept is defined in both objective and subjective theories, I will here discuss only the subjective version. In subjective terms, you regard two unknown quantities, X and Y, as *exchangeable* (or permutable or symmetric) if, for any statement involving one or both of them, your probability assignment will not change if X and Y were interchanged. For example, for two individuals of unknown HIV status but identical values for known predictors of HIV status (age, gender, intravenous drug use, ethnicity, and sexual activities), I would regard their HIV status indicators X and Y (1 = positive, 0 = negative) as exchangeable. Thus, I would have $\Pr(X < Y) = \Pr(Y < X)$, $\Pr(X = 0) = \Pr(Y = 0)$, $\Pr(X = 1) = \Pr(Y = 1)$, and so on; with respect to my probability assignments, X and Y would be indistinguishable. More generally, you would regard the unknown quantities in a collection as exchangeable if they were indistinguishable or interchangeable with respect to your probability assignments.

Exchangeability is implied by but does not imply the common statistical assumption that the quantities under study are independent and identically distributed.[6] It is the exchangeability of characteristics of sampled persons with those of unsampled persons that justifies statistical inferences from a survey sample to the population.[1,3,6,42–44] Similarly, in comparative trials, it is the exchangeability of group outcomes under homogeneous treatment allocation that justifies inferences about causal effects.[1,43–45] Random sampling in surveys and randomization in comparative trials are perhaps the most dependable but not the only methods for inducing observers to make exchangeable probability assignments; for example, certain forms of systematic sampling or treatment allocation may suffice.

In nonexperimental comparisons, exchangeability arises in a much more conditional, partial fashion, because such comparisons require control of a "sufficient" set of selection and risk predictors. Here "sufficient" means that, within levels of controlled predictors, the outcomes of the different exposure groups would be exchangeable if every subject received the reference exposure level.[45] Note that matching in nonexperimental studies does not necessarily induce such exchangeability; it only ensures that, within each matching stratum, there will be both exposed and unexposed subjects (in cohort studies) or diseased and nondiseased subjects (in case-control studies).[13] More generally, popular strategies for control of confounding such as regression adjustment do not induce exchangeable assignments; rather, they must *assume* that the set of controlled predictors are sufficient in the sense just described, which is often not true.

Calibration

Key structures in subjective theory can be traced out in parallel with structures in objective theory. For example, in objective epidemiologic theory, "exchangeability" of exposure groups is synonymous with "comparability" of the groups, or absence of confounding.[45] I have focused instead on subjective exchangeability because no physical probabilities can be identified or even shown to exist in observational epidemiologic comparisons.[36] Nonetheless, one should not lose sight of the parallel because, if physical probabilities *do* exist in a given situation, then (generalizing from the Principal Principle) we should want our subjective probability assignments to be as closely calibrated to the physical probabilities as possible.[23]

To clarify the origin of the notion of calibration, consider meteorology, an applied physical science whose difficulties with complex observational studies and public esteem may rival those of epidemiology. I should be quite satisfied if, in my town, it rained on 70% of those days for which the forecast gave a 70% chance of rain, and so on for other percentages. I know, however, that such good calibration is well beyond current meteorology, and that I should not confuse the forecast (a subjective probability produced by the weather service) with any physical probability. Similarly, I would be quite satisfied if, in my county, there was a decline in AIDS case reports in 70% of those subgroups for which our projections gave a 70% chance of decline in the coming year, and so on. Such excellent calibration is beyond current epidemiology, and I know I should not confuse our projection (a subjective probability) with any physical probability.

In the objective theory, the "70% chance of decline" just described is referred to as an "estimate of the probability of decline," and so presumes that the physical probability exists. Whether or not the latter exists, subjective probabilities can be constructed and updated using the actual outcomes. What updating method will ensure that our subjective probabilities will approach the physical probabilities when the latter exist and data accumulate indefinitely? As with probability estimates from the objective theory, such convergence depends on whether any assumptions we have used (for example, logistic dependence of probabilities on factors) are at least approximately correct. Thus, to repeat, a Bayesian analysis does not free us from the need to check our assumptions[8,24,35]; as in Popperian philosophy, in the Bayesian philosophy espoused here the ultimate test of our hypotheses and assumptions is how well our predictions are borne out by observations.

Discussion

Unlike forecasting, risk factor epidemiology provides few opportunities to validate predictions in the manner described above. That is, risk factor epidemiology suffers from a lack of calibration opportunities, rather than a lack of theoretical testability. No philosophy, whether frequentist, Bayesian, or Popperian, can do more than point out this weakness in our science. Whether this weakness can be remedied by anything other than more randomized trials (such as the recent trials of beta-carotene) remains to be seen.

Although I have argued that subjective probability logic has value as an approach to statistical inference, like all approaches (including the Popperian approach) it should be regarded as a partial and conditional account of scientific reasoning[24]; it is not as comprehensive as some authors[3,6,7] seem to maintain. In particular, as with all statistical procedures, conclusions derived using Bayesian methods are conditional on any probability models used in the course of analysis, and well as on prior-probability assignments. Criticism of these models and assignments plays an essential role in data analysis that complements and cannot be replaced by Bayesian methods.[24] In Popperian terms, we may view a Bayesian analysis as a method to incorporate data information into a given model; it cannot substitute for *or* be replaced by the model-criticism step, in which the model is exposed to possible criticism based on conflict with observations.

It is a serious challenge for *both* objective and subjective theories to address how one should specify and justify use of probability models when there are no known physical probabilities or symmetries on which to base them.[36] By "justify," I mean deduce the probability assignments from a set of assumptions that are given high prior probabilities by everyone in the scientific community. In this sense, the standard likelihood functions used to estimate causal effects from observational epidemiologic studies appear difficult to justify in subjective terms and impossible to justify in objective terms, for their deduction assumes that the exposure was randomized in a "natural experiment," which would be a fanciful assumption in (say) a study of alcohol use and breast cancer. The Principal Principle does not help here; it asks us to model our subjective probabilities on the basis of physical probabilities that do not exist or at best are unknown. Likewise, the Dutch Book argument is of no help if we have no precise values to give to our subjective probabilities, although in such a case an argument can be made for use of interval probability assignments.[3]

It is sometimes suggested that conventional statistics should be regarded as providing the *minimal* uncertainty that one should assign to a parameter.[36] This suggestion seems too easily ignored in typical epidemiologic discussions, however, and it does not address the fact that, without justifiable probabilities, probabilistic induction and the whole body of inferential statistics (objective and subjective) is without foundation. For one must ask: Of what relevance is a statistic based on the assumption that "the data are from a perfect randomized trial" when randomization (let alone perfection) is a complete fantasy? The answer is a subjective one, in that persons with much faith in the validity of the study (typically, the study investigators and persons who are pleased with the results) will think those statistics are highly relevant, whereas persons with little faith in the validity of the study (typically, persons displeased with the results) will think those statistics are deceptive.

Skepticism about claims for objectivity should be applied to common statistical methods, such as traditional *P*-values, confidence intervals, and regression analyses, as well as to more modern approaches based on bootstrapping, Gibbs sampling, and other Monte-Carlo methods. I am aware of only two justifiable responses to this skepticism. One response is to limit analyses of observational data to pure data descriptors: graphs and tables, perhaps some means and differences and ratios, but no *P*-values or confidence intervals or standard errors or regressions. As for summary measures, only standardization could be justified; maximum-likelihood and Mantel-Haenszel estimates could not. Such a limited analysis would probably never get published and would be devoid of measures of uncertainty.

An alternative response is to expend the extra effort to propose plausible (if tentative) subjective probability assignments and use them in a Bayesian analysis; frequentist methods would sometimes be justified as approximations to Bayesian methods. Such an analysis would be "subjective" but not arbitrary, because it would be constrained by past observations and by the norms of the subjective theory. It would have several advantages over conventional "objective" analyses. First, it would yield logically derived conditional certainty statements; for example, it would allow derivation of statements of the form "Given the assumptions of this analyses, we would be 95% certain that the parameter under study lies between $\underline{RR}$ and $\overline{RR}$." In constructing such statements, the Bayesian analysis can make use of valuable information that is ignored by conventional analyses, such as probable limits on the magnitude of typical effects.[13,24,39–41]

Another Bayesian advantage is a consequence of the use of magnitude information and may be the most relevant for a Popperian epidemiologist. If the magnitude of an effect estimate is symmetrically constrained by a prior distribution or hierarchy, it will be less probable that the estimate will appear substantively or "statistically" significant than in a conventional analysis. In this sense, Bayesian analysis can provide a more stringent test of the causal hypothesis that the effect is non-negligible. ("Non-negligible" often means nonzero, but it may also mean "above a certain action threshold.") From a conventional statistical perspective, this means that the power (sensitivity) of the Bayesian analysis at a given alpha level (specificity) will be lower than that of the conventional analysis. This power loss is modest, however, compared with the dramatic reduction in Type I (false-positive) error afforded by the use of the prior information. In hypothesis-screening terms, this means that the Bayesian analysis allows one to trade a modest decrease in sensitivity for a very large increase in specificity, so that the ROC curve for Bayesian hypothesis screening lies above the curves for conventional hypothesis-screening procedures.

The preceding advantage of Bayesian analysis has been phrased in terms of hypothesis testing. Such terms, although used by Popper, have come to be abhorred by some epidemiologists because of the widespread abuse of statistical hypothesis testing.[13,14] The advantage may be rephrased in estimation terms, however: Bayesian methods facilitate the use of prior information to construct estimates of much greater accuracy (that is, with better calibration) than conventional estimates. From my perspective, this is a decisive advantage, and one that has been verified repeatedly in theory, simulation, and real applications.[1,34,38–41]

One remaining question is whether epidemiologists can be trained to employ Bayesian methods in an intelligent fashion. A cynic could point to evidence that the task is not possible, especially in the aforementioned abuse of statistical hypothesis testing. I would argue, however, that potential for abuse of Bayesian methods is insufficient grounds for denial of its benefits to competent users. Of course, widespread use of these methods will be delayed by their conceptual unfamiliarity, by lack of software, and by innate resistance to change. Fortunately, there is a growing movement in the statistics community to introduce Bayesian methods into basic statistics education[46]; this movement, along with coverage of probability logic in epidemiologic training, will hasten the day when unfamiliarity is no longer an important obstacle.

Acknowledgments

I am deeply indebted to Malcolm Maclure and Charles Poole for their extensive criticisms of this manuscript, subsequent discussions, and valuable suggestions for references and revisions.

References

1. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. New York: Chapman and Hall, 1995.
2. Berry DA, Stangl DK. Bayesian Biostatistics. New York: Marcel Dekker, 1996.
3. Lad F. Operational Subjective Statistical Methods. New York: Wiley, 1996.
4. Greenland S, ed. Evolution of Epidemiologic Ideas. Chestnut Hill, MA: Epidemiology Resources Inc., 1987.
5. Rothman KJ, ed. Causal Inference. Chestnut Hill, MA: Epidemiology Resources Inc., 1988.
6. DeFinetti B. The Theory of Probability. vols. 1 and 2. New York: Wiley, 1974.
7. Howson C, Urbach P. Scientific Reasoning: The Bayesian Approach. 2nd ed. LaSalle, IL: Open Court, 1993.
8. Good IJ. Good Thinking. Minneapolis: University of Minnesota Press, 1983.
9. Popper KR, Miller DW. A proof of the impossibility of inductive probability. Nature 1983;302:687–688.
10. Popper KR, Miller DW. Why probabilistic support is not inductive. Phil Trans R Soc Lond 1987;A321:564–591.
11. Kahneman D, Slovic P, Tversky A. Judgement under Uncertainty: Heuristics and Biases. New York: Cambridge University Press, 1982.
12. Piattelli-Palmarini M. Inevitable Illusions. New York: Wiley, 1995.
13. Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott, 1997.
14. Oakes M. Statistical Inference. Chestnut Hill, MA: Epidemiology Resources Inc., 1990.
15. Kyburg HE. Probability and Inductive Logic. London: MacMillan, 1970.
16. Cohen LJ. An Introduction to the Philosophy of Induction and Probability. Oxford: Clarendon, 1989.
17. Skyrms B. Choice and Chance: An Introduction to Inductive Logic. 3rd ed. Belmont, MA: Wadworth, 1986.
18. Hacking I. The Emergence of Probability. New York: Cambridge University Press, 1975.
19. Greenland S. Induction versus Popper: substance versus semantics. Int J Epidemiol 1998;27 (in press).
20. Von Mises R. Probability, Statistics and Truth. 2nd ed. New York: Dover, 1957.
21. Popper KR. The Logic of Scientific Discovery. New York: Harper and Row, 1968.
22. Freedman DA. Some issues in the foundations of statistics (with discussion). Found Sci 1995;1:19–83.
23. Lewis D. A subjectivist's guide to objective chance. In: Harper WL, Stalnaker R, Pearce G, eds. Ifs: Conditionals, Belief, Decision, Chance, and Time. Boston: D. Reidel, 1981;267–297.
24. Box GEP. An apology for ecumenism in statistics. In: Box GEP, Leonard T, Wu CF, eds. Scientific Inference, Data Analysis, and Robustness. New York: Academic Press, 1983.
25. Ramsey FP. Truth and probability. In: Ramsey FP. The Foundations of Mathematics and Other Logical Essays. New York: Harcourt, Brace, 1931.
26. DeFinetti B. Foresight: its logical laws, its subjective sources (original published in French). Reprinted in: Kyburg HE, Smokler HE, eds. Studies in Subjective Probability. New York: Wiley, 1937;1964.
27. Hume D. An enquiry concerning human understanding. LaSalle, IL: Open Court, 1988, Section 6 (first published 1748).
28. Levi I, Jeffrey RC, Good IJ, Popper K, Miller D. Matters arising: the impossibility of inductive probability (discussion forum). Nature 1984;310:433–434.
29. Redhead M. On the impossibility of inductive probability. Br J Phil Sci 1985;36:185–191.
30. Dunn JM, Hellman G. Dualling: a critique of an argument of Popper and Miller. Br J Phil Sci 1986;37:220–223.
31. Eells E. On the alleged impossibility of inductive probability. Br J Phil Sci 1988;39:111–116.
32. Rodriguez AR. On Popper-Miller's proof of the impossibility of inductive probability. Erkenntnis 1987;27:353–357.
33. Stigler SM. The History of Statistics. Cambridge, MA: Belknap, 1986.
34. Carlin BP, Louis TA. Bayes and Empirical Bayes Methods for Data Analysis. New York: Chapman and Hall, 1996.
35. Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. Am J Epidemiol 1986;123:392–402.
36. Greenland S. Randomization, statistics, and causal inference. Epidemiology 1990;1:421–429.
37. Greenland S. Summarization, smoothing, and inference. Scand J Soc Med 1993;21:227–232.
38. Thomas DC. The problem of multiple inference in identifying point-source environmental hazards. Environ Health Perspect 1985;62:407–414.
39. Greenland S. A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer-mortality study. Stat Med 1992;11:219–230.
40. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect 1994;102(suppl 8):33–39.
41. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology 1994;5:612–621.
42. Leamer EE. Specification Searches. New York: Wiley, 1978.
43. Draper D, Hodges JS, Mallows CL, Pregibon D. Exchangeability and data analysis. J R Stat Soc A 1993;156:9–37.
44. Greenland S, Draper D. Exchangeability. In: The Encyclopedia of Biostatistics. New York: Wiley, 1998.
45. Greenland S. Confounding. In: The Encyclopedia of Biostatistics. New York: Wiley, 1998.
46. Berry DA, Albert J, Moore DS. Teacher's corner (with discussion). Am Stat 1997;51:241–274.

Appendix 1

The Necessity of the Probability Axioms to Prevent Sure Loss

Suppose you are willing to bet on your probability assignments in the following sense: You are willing to gamble up to s monetary units in a wager against me, and, if you assign Pr(A) to a statement A (such as 0.01 to "HIV will be eradicated by 2020") then you would be willing to bet $s\mathrm{Pr}(A)$ on A being true or $s(1 - \mathrm{Pr}(A))$ on A being false, at my choice. That is, $\mathrm{Pr}(A)/(1 - \mathrm{Pr}(A))$ equals the betting odds for A that would render you indifferent to betting for or against A. Then, if your probability assignments do not obey the axioms of probability theory, I can choose my bets so that you are sure to lose money:

Axiom 1

Suppose you violate this axiom by assigning $\mathrm{Pr}(A) < 0$. Then I will bet on A. If A turns out to be false, you "win" the negative amount $s\mathrm{Pr}(A)$; this means you owe me the positive amount $-s\mathrm{Pr}(A)$. If A turns out to be true, you owe me $s(1 - \mathrm{Pr}(A))$. Thus, only by assigning $P(A) \geq 0$ (Axiom 1) can you avoid sure loss.

Axiom 2

Suppose A is logically inevitable, but you violate this axiom by assigning $\mathrm{Pr}(A) > 1$. Then I will bet against A, and you "win" the negative amount $s(1 - \mathrm{Pr}(A))$; this means you owe me $s(\mathrm{Pr}(A) - 1)$. If you assign $\mathrm{Pr}(A) < 1$, then I will bet on A and win $s(1 - \mathrm{Pr}(A))$. Only by assigning $\mathrm{Pr}(A) = 1$ (Axiom 2) can you avoid a sure loss.

Axiom 3

Suppose your assignments yield $\mathrm{Pr}(A \text{ or } B) > \mathrm{Pr}(A) + \mathrm{Pr}(B)$ for some mutually exclusive A and B. Then I will place bets against "A or B," for A, and for B. If A turns out to be true, I win

$$
\begin{aligned}
&s(1 - \mathrm{Pr}(A)) - s\mathrm{Pr}(B) - s(1 - \mathrm{Pr}(A \text{ or } B)) \\
&\quad = s\mathrm{Pr}(A \text{ or } B) - s(\mathrm{Pr}(A) + \mathrm{Pr}(B)),
\end{aligned}
\tag{A1}
$$

which is positive. Reversing A and B shows that I net the same amount if B turns out to be true. If neither occurs, I also win

sPr(A or B) − s(Pr(A) + Pr(B)). (They cannot both be true.) Parallel algebra shows I can guarantee you suffer a net loss if your assignments yield Pr(A or B) < Pr(A) + Pr(B) for some mutually exclusive A and B by betting for "A or B," against A, and against B. Only by assigning Pr(A or B) = Pr(A) + Pr(B) when A and B are mutually exclusive (Axiom 3) can you avoid a sure loss.

Appendix 2

Corroboration in the Subjective Theory

The following argument and counterexample show that corroboration implies but is not implied by probabilistic support. Suppose that finding B false was possible and would have refuted H; that is, Pr(not B) > 0 and Pr(H|not B) = 0. Then

$$\Pr(H) = \Pr(H \mid B)\Pr(B) + \Pr(H \mid \text{not } B)\Pr(\text{not } B)$$

$$= \Pr(H \mid B)\Pr(B) < \Pr(H \mid B)$$

because Pr(B) = 1 − Pr(not B) < 1.

Now suppose that Pr(H and B) = Pr(not H and not B) = 0.3 and Pr(H and not B) = Pr(not H and B) = 0.2. Then

$$\Pr(H \mid B) = 0.3/(0.3 + 0.2) = 0.6 > \Pr(H)$$

$$= \Pr(H \text{ and } B) + \Pr(H \text{ and not } B) = 0.3 + 0.2 = 0.5,$$

so B supports H, but Pr(H|not B) = 0.2/(0.3 + 0.2) = 0.4, so B does not corroborate H.

The following counterexamples show that proof and corroboration do not imply one another. First, suppose that Pr(H) = 0.6 and Pr(B) = Pr(H and B) = 0.2. Then Pr(H | B) = 0.2/0.2 = 1, so B proves H, but Pr(H | not B) = (0.6 − 0.2)/(1 − 0.2) = 0.5, so B does not corroborate H. Second, suppose that Pr(B) = 0.4 and Pr(H) = Pr(H and B) = 0.2. Then Pr(H | not B) = (0.2 − 0.2)/(1 − 0.4) = 0, so B corroborates H, but Pr(H | B) = 0.2/0.4 = 0.5, so B does not prove H.