

Probability and Statistics: 5 Questions

Carl G. Wagner

Professor of Mathematics and Adjunct Professor of Philosophy

The University of Tennessee

1. Why were you initially drawn to probability theory and/or statistics?

I first encountered the subject in Herbert Hawkes' wonderful old text, *Advanced Algebra*, in a chapter entitled "Permutations, Combinations, and Probability." Published in 1928 by Ginn and Company, this book was still in use during my senior year in high school in the autumn of 1960. In contrast to the execrable, bloated mathematics texts currently in use in American schools, this little gem, which sits on my desk as I write, measures a mere $5 \times 7 \times \frac{3}{4}$ inches. In a chapter of only fourteen pages, the author naturally stuck to the conception of probability as a ratio of favorable to total possible cases, but he managed to cover both binomial and multinomial coefficients, as well as the general form of the multiplication principle for probabilities. My interest in enumerative combinatorics, my current specialty in pure mathematics, dates back to this time.

As an undergraduate at Princeton I took a course in elementary statistics with Francis Anscombe, as well as an upper division course in probability with William Feller and in mathematical statistics with John Hartigan, who had just completed a Ph.D thesis on Bayesian theory under Anscombe's direction.¹ With such teachers, it would have been natural for me to have become enamored of probability and statistics at that point. But my real infatuation as an undergraduate was with mathematical logic, which I had the incredible good fortune to study with Carl "Peter" Hempel, Alonzo Church, James Guard, and Paul Benacerraf.

In contrast to the careful attention to given to foundational questions in my logic courses, I found several aspects of probability and statistics, as presented in our texts, to be marred by frustrating lacunae. Probabilities were said to be limiting relative frequencies in an infinite sequence. So I naturally tried to derive the Kolmogorov axioms on this basis. But it was easy to imagine an infinite sequence of observations from a countably infinite sample space Ω in which each ω in Ω occurred only finitely many times, thus violating countable additivity. I had no idea, however, how to exclude such sequences in any sort of principled way. I found equally troubling the unelaborated assertion that $p(A|E) := p(A \cap E)/p(E)$ was the appropriate revision of the prior probability $p(A)$ upon discovering the

truth of E. I tried to find a salient characterization of $p(.|E)$ among all probabilities q for which $q(E) = 1$, but again without success. It was only much later that I came to learn that these were respectable concerns that had attracted the attention of serious students of the foundations of probability.

After Princeton, I went to Duke to study logic with Joseph Schoenfield, but quickly became a convert to number theory, and wound up doing a Ph.D. thesis on p-adic function theory with Leonard Carlitz. I then joined the mathematics faculty at the University of Tennessee, which has been my academic home ever since. As a young academic mathematician I continued to pursue my undergraduate interest in philosophy during my spare time, and began reading the literature in axiomatic social choice theory. During this time I also read two intriguing papers by legal scholars Laurence Tribe (1971) and John Kaplan (1968), in which I first encountered the notion of subjective probability. I was so taken by Kaplan's application of decision theory to account for certain legal rules that, on my first sabbatical leave, I spent the autumn of 1976 attending lectures at Stanford Law School, where, among other things, I tried unsuccessfully to give a probabilistic account of proximate causation. Wandering around the halls of the philosophy department at Stanford one day, I came across an advertisement for an NEH seminar entitled "Freedom and Causality," to be directed by Keith Lehrer and Sydney Shoemaker at the Center for Advanced Study in the Behavioral Sciences during the summer of 1977. My acceptance as a participant in this seminar was nothing short of providential, for it was there that Lehrer and I began our collaboration on the consensus problem, materializing in the book, *Rational Consensus in Science and Society* (1981). Of equal importance, Keith set me on an orderly path of reading on probability and decision theory, introducing me to the work of Frank Ramsey, Richard Jeffrey, Brian Skyrms, and others, and arranging for me to spend the 1978-9 academic year as a Fellow at CASBS. On a later sabbatical at Princeton in 1990 I had the great good fortune to spend a semester in the company of Richard Jeffrey. After years of desultory engagement with the subject I found in Jeffrey's radical probabilism a congenial way of thinking about probability: *Probability is a mode of judgment, and strict conditionalization is by no means the only principled way to revise a probability assessment in the light of new evidence.* As will be evident in the remainder of this essay, this point of view has motivated all of my subsequent work in probabilistic epistemology.

2. *What is distinctive about your work in the foundations of probability or its applications?*

My work has focused almost exclusively on the problem of revising a probability assessment in the light of new evidence when the prerequisites for applying strict conditionalization are absent. This work comprises three principal themes:

I. *Consensus*

Suppose that you have assessed a probability distribution over a set Ω of possible states of the world, and you subsequently become apprised of the distributions of one or more other individuals. How should you revise your prior in light of this new information? The doctrinaire Bayesian answer is that you should update your prior by conditioning on this evidence. But, except in the simplest cases, it would be completely unrealistic to expect to be able to carry out a prior (or even retrospective) assessment of the likelihoods required to apply Bayes' rule here. Assuming that you believe the assessments of the others to have some merit, you might consider revising your prior to some sort of weighted average of all of your distributions, with weights reflecting your assessment of their individual expertise, as well as your own. Assuming that the others participated in this exercise by similarly revising their priors and that a certain pattern of mutual respect obtained in the group, Morris DeGroot (1974) and, independently, Keith Lehrer (1976) observed that under iterated weighted averaging the elementary theory of Markov chains guaranteed convergence to a consensus on the weights to be employed in averaging the priors, and hence emergence of a consensual probability distribution over Ω . In both of their models, a given individual was assumed to have uniform expertise over a hierarchy of judgmental tasks. Lehrer asked me to investigate the situation in which an individual could be assigned different weights, say, as a physicist, a judge of physicists, a judge of judge of physicists, etc. I was unaware at the time that Chatterjee and Seneta (1977) had already dealt with this problem, albeit with a different interpretation of the weights. In any case, it turned out to be fairly straightforward to identify conditions of respect under which iterated weighted averaging converged to a set of consensual weights in this extended model. See Theorems 8.2-8.5 in Lehrer and Wagner (1981).

I was intrigued, however, with the more fundamental problem of justifying the use of weighted arithmetic means in the first place, irrespective of how the weights were ultimately chosen. The first axiomatic characterization of such means appeared in Aczel and Wagner (1980), with variants appearing in McConway (1981), Lehrer and Wagner (1981), and Aczel, Ng, and Wagner (1984). At the time it was widely believed in the philosophical community (see Laddaga 1977, Laddaga and Loewer 1985, and Levi 1985) that any acceptable method of aggregating the probability distributions of several individuals should preserve

every instance of independence common to all their distributions. I argued against this restriction in Lehrer and Wagner (1983), noting that many formal cases of independence have no epistemic significance. This view received further support in Wagner (1984), where it was shown, in results reminiscent of the impossibility theorems of social choice theory, that under mild restrictions on the method of aggregation only dictatorial or imposed “consensual” probabilities satisfy universal preservation of independence. Subsequently, these limitative theorems were reproduced in Genest and Wagner (1987) under significantly weaker restrictions.

II. *Imprecise Probability*

I first encountered imprecise probabilities in Glen Shafer’s exquisite monograph, *A Mathematical Theory of Evidence* (1976) and was immediately intrigued by the prospect of non-additive numerical representations of one’s strength of belief when evidence was insufficient to support assessment of a precise probability distribution. I later came into possession of a copy of Peter Walley’s samizdat, *Coherent Lower (and Upper) Probabilities* (1981), the precursor to his magisterial *Statistical Reasoning with Imprecise Probabilities* (1991), and became thoroughly convinced of the importance of this generalization of classical probability theory.

I first tried to exploit the flexibility offered by imprecise probabilities by investigating consensus methods that mapped a “profile” (p_1, \dots, p_n) of probability distributions on the finite set Ω to a lower probability λ , hoping that this might allow for some interesting variations on weighted arithmetic averaging. Naturally, the “lower envelope” of this profile, defined by

$$(1) \quad \lambda(E) := \min \{ p_i(E) : 1 \leq i \leq n \},$$

is always available as a rough summary of group opinion. It should be noted that lower envelopes frequently fail to satisfy the highly desirable² “2-monotonicity” property,

$$(2) \quad \lambda(E \cup F) \geq \lambda(E) + \lambda(F) - \lambda(E \cap F).$$

But if $|\Omega| \geq 4$, $\lambda(E)$ depends only on the values $p_i(E)$, $p_1(E) = \dots = p_n(E) = c \Rightarrow \lambda(E) = c$, and λ is 2-monotone, then λ is a weighted arithmetic mean of the distributions p_i , whence λ is itself a probability distribution (Wagner 1989).

I had somewhat more success in generalizing Jeffrey conditionalization (henceforth, JC) to the case in which new evidence places a lower bound on the possible revisions of a prior, rather than determining new probabilities for a certain class of events. Suppose (Ω, \mathbf{A}, p) is a probability space and that new evidence calls for revision of the prior p . Suppose that the total evidence, old as well as new, determines a countable family \mathbf{E} of nonempty, pairwise disjoint events, with $p(E) > 0$ for all $E \in \mathbf{E}$, and a collection $\{\mu_E : E \in \mathbf{E}\}$ of positive real numbers summing to 1, restricting the possible revisions of p to those probability measures q satisfying

$$(3) \quad q(E) = \mu_E, \text{ for all } E \in \mathbf{E}.$$

Unless each $E \in \mathbf{E}$ is a singleton, there are infinitely many probability measures q satisfying (3). Suppose, however, that there is reason to judge that any acceptable revision q should also satisfy

$$(4) \quad q(A|E) = p(A|E), \text{ for all } A \in \mathbf{A} \text{ and all } E \in \mathbf{E}.$$

As is easily proved, (3) and (4) jointly determine a uniquely acceptable revision q , given by

$$(5) \quad q(A) = \sum_{E \in \mathbf{E}} \mu_E p(A|E),$$

a generalization of strict conditionalization first explored by Jeffrey (1965).

In what follows we take Ω to be finite and $\mathbf{A} = 2^\Omega$ for the sake of simplicity. Suppose that new evidence prompts a revision of the prior p as follows: The total evidence, old as well as new, allows us to assess a probability measure u on 2^Θ for a finite set of possible states of the world Θ , as well as a mapping $T: \Theta \rightarrow 2^\Omega \setminus \{\emptyset\}$. The set $T(\theta)$ consists of all those states ω in Ω compatible with the state θ in Θ . Volker Strassen (1964) made a thorough study of set functions m and β induced by u and T on 2^Ω , defined by

$$(6) \quad m(E) := u \{ \theta \in \Theta : T(\theta) = E \}, \text{ and}$$

$$(7) \quad \beta(E) := u \{ \theta \in \Theta : T(\theta) \subseteq E \}.$$

Since $\beta(E)$ is the sum of the probabilities of all those outcomes in Θ which entail the event E , it is clear that u and T should restrict the possible revisions of p to those q such that

$$(8) \quad \beta(E) \leq q(E) \quad \text{for all } E \subseteq \Omega.$$

Let $\mathbf{E} := \{ E \subseteq \Omega : m(E) > 0 \}$ and call members of \mathbf{E} *evidentiary focal elements*. As shown in Wagner (1992), one can articulate, in terms of a family of joint probability measures on $\Theta \times \Omega$, a generalization of the “rigidity” condition (4) that, combined with (8), determines a uniquely acceptable revision q of p , given by

$$(9) \quad q(A) = \sum_{E \in \mathbf{E}} m(E) p(A|E),$$

with (8), generalized rigidity, and (9) reducing respectively to (3), (4), and (5) precisely when the evidentiary focal elements are pairwise disjoint.

III. *Bayes Factors and the Uniformity Rule*

Suppose that q is a revision of the probability measure p , and that A and B are events. The *Bayes factor* $\beta_{q,p}(A:B)$ is the ratio

$$(10) \quad \beta_{q,p}(A:B) := q(A)/q(B) / p(A)/p(B)$$

of new to old odds. It has long been a staple of Bayesianism (Good 1985, Jeffrey 1992) that such ratios represent what is learned from new experience alone, with prior probabilities factored out. When $q = p(\cdot | E)$, then $\beta_{q,p}(A:B)$ is simply the likelihood ratio $p(E|A)/p(E|B)$. When $A = H$ and $B = H^c$, this likelihood ratio (or its logarithm) is arguably the best measure of the degree to which H is (incrementally) confirmed by E .³

In a series of papers (Wagner 1997, 1999, 2001, 2002) I have exploited Bayes factors to show that the so-called “non-commutativity” of JC is a consequence of the incorrect representation of identical new learning, and to generalize Richard Jeffrey’s (1991, 1995) solution to the old evidence problem to the case of uncertain old evidence and new statistical explanation. In pursuing this work, I ultimately came to realize that a single principle of probability revision grounded the results of all the earlier papers. An account of this principle, which I called the *Uniformity Rule*, appears in Wagner (2003), and is summarized below.

Consider the following belief revision schema, representing two possible sequential revisions of the probability measure p :

$$(11) \quad p \rightarrow q \rightarrow r \quad \text{and} \quad p \rightarrow s \rightarrow t .$$

Suppose that the revisions of p to q , and of s to t , are prompted by identical new learning, and that the revisions of q to r , and of p to s , are prompted by identical new learning. It is widely held that it should then be the case that $r = t$. As van Fraassen (1989) puts it, two persons who undergo identical learning experiences on the same day, but in a different order, ought to agree in the evening if they had exactly the same opinions in the morning. Call this the *Commutativity Principle*. It turns out that this principle is satisfied for purely atomic sigma algebras⁴ when the *Uniformity Rule* is obeyed, i.e., when identical new learning is represented by identical Bayes factors at the level of atomic events. In what follows, probability measures are assumed to take the value 0 only for the impossible event \emptyset . With the addition of certain technical conditions, however, the theorems presented here hold for arbitrary probabilities.

THEOREM 1. *Suppose that the probabilities in the revision schema (11) above are defined on a purely atomic sigma algebra \mathbf{A} , with \mathbf{A}^* denoting the set of atomic events in \mathbf{A} . If the Bayes factor identities*

$$(12) \quad \beta_{q,p}(A:B) = \beta_{t,s}(A:B) \quad \text{and} \quad \beta_{r,q}(A:B) = \beta_{s,p}(A:B)$$

hold for all $A, B \in \mathbf{A}^$, then $r = t$.*

In the case of revision by JC the Uniformity Rule admits of a coarser-grained avatar:

THEOREM 2. *Suppose that p, q, P , and Q are probability measures on the purely atomic sigma algebra \mathbf{A} , with family of atomic events \mathbf{A}^* , and suppose that q comes from p by JC on the family \mathbf{E} . Then $\beta_{Q,P}(A:B) = \beta_{q,p}(A:B)$ for all $A, B \in \mathbf{A}^*$ if and only if Q comes from P by JC on \mathbf{E} and $\beta_{Q,P}(E:E') = \beta_{q,p}(E:E')$ for all $E, E' \in \mathbf{E}$.*

Jeffrey conditionalization and commutativity. Theorems 1 and 2 imply that when identical new learning is correctly represented, then sequential revisions by JC behave just as they should:

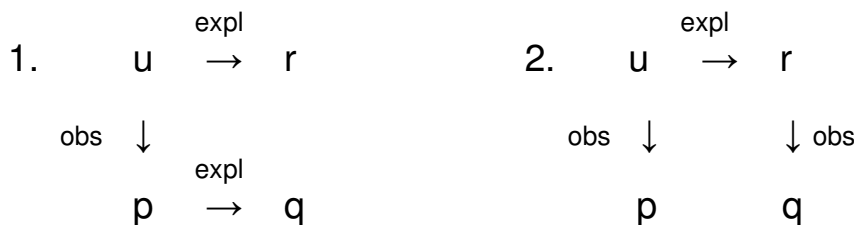
THEOREM 3. *Suppose that the probabilities in the revision schema (11) above are defined on an arbitrary sigma algebra, that q comes from p , and t from s , by JC on the countable family of events \mathbf{E} , and that r comes from q , and s from p , by JC on the countable family \mathbf{F} . If $\beta_{q,p}(E:E') = \beta_{t,s}(E:E')$ for all $E, E' \in \mathbf{E}$ and $\beta_{r,q}(F:F') = \beta_{s,p}(F:F')$ for all $F, F' \in \mathbf{F}$, then $r = t$.*

REMARK: Under mild regularity conditions (see Wagner 2001, Theorem 4.1) the identities $\beta_{q,p}(E:E') = \beta_{t,s}(E:E')$ and $\beta_{r,q}(F:F') = \beta_{s,p}(F:F')$ are also necessary for the commutativity property $r = t$.

Uncertain old evidence and new statistical explanation. The classical old evidence problem, posed by Clark Glymour (1980), asks for a probabilistic account of the confirmation of H that occurs when it is discovered that H implies the *previously known* fact E. Richard Jeffrey (1991, 1995) offered a solution to this problem using a new type of probability revision that he called *reparation*. Suppose that p denotes our distribution after observing E, but before establishing that H implies E. Jeffrey's solution involves the imaginative reconstruction of an *ur-distribution* u predating the old observation of E and the new (deductive) explanation of E by H. The natural explanation-based revision of u is then mimicked in constructing the explanation-based revision q of p, in such a way that, *ad libitur*, $q(H) > p(H)$.

Reparation may be generalized as follows: Let p denote our current distribution on the algebra **A** generated by E and H. Prior observation has yielded a degree of confidence in E reflected in p(E). Subsequent theoretical work, not relying in any way on the aforementioned observation, shows that the truth of H would confer probability α on E, and its falsity would confer probability α^* on E. What is the appropriate revision q of p in light of this new statistical explanation?

Following Jeffrey, we denote by u a notional ur-distribution, predating both our old (uncertain) observation and our new (statistical) explanation. We assume that $u(A) > 0$ for all $A \in \mathbf{A}^* := \{HE, HE^c, H^cE, H^cE^c\}$ and that p has come from u by JC on $\mathbf{E} := \{E, E^c\}$. Suppose that we made the aforementioned theoretical discovery in the conceptual state captured by u, and that this discovery left the ur-probability of H unchanged. This would warrant the revision of u to the unique probability r satisfying (i) $r(E|H) = \alpha$, (ii) $r(E|H^c) = \alpha^*$, and (iii) $r(H) = u(H)$. Two possibilities for constructing q now present themselves:



1. Construct q so that it bears the same relation to p (as dictated by the Uniformity Rule) as r does to u, since both revisions are meant to be based on identical explanatory learning.
2. Motivated by the Commutativity Principle, construct q so

that it bears the same relation to r (as dictated by the Uniformity Rule) as p does to u , since both revisions are meant to be based on identical observational learning. By Theorem 1, both of these procedures lead to identical outcomes. Moreover, by Theorem 2, q comes from r by JC on \mathbf{E} , and $\beta_{q,r}(E:E^c) = \beta_{p,u}(E:E^c)$. Denoting by β the common value of these Bayes factors, it follows that

$$(13) \quad \beta_{q,p}(H:H^c) = \frac{[(\beta - 1)r(E|H) + 1] / [(\beta - 1)r(E|H^c) + 1]}{[(\beta - 1)u(E|H) + 1] / [(\beta - 1)u(E|H^c) + 1]} .$$

From (13) one can derive several intuitively reasonable conditions under which $q(H) > p(H)$. For example, if E and H are assumed to be ur -independent,⁵ then if new statistical explanation simply amounts to the discovery that H is positively relevant to E (whence $r(E|H) > r(E|H^c)$) and if old observation simply raised our confidence in the truth of E (whence $\beta > 1$), it follows that H is confirmed. In the cases considered by Jeffrey, $p(E) = 1$ (whence $\beta = \infty$), $r(E|H) = 1$, and $r(E|H^c) = u(E|H^c)$, whence $\beta_{q,p}(H:H^c) = 1/u(E|H) > 1$, i.e., H is always confirmed.

3. *How do you conceive of the relationship between probability theory and/or statistics and other disciplines?*

The relationship between probability and other disciplines is the same as the relationship between traditional logic and those disciplines. Probability is a mode of judgment that enlarges upon the old dogmatic categories of acceptance, rejection, or suspension of judgment, and probability logic is the proper tool of uncertain inference, just as ordinary logic is the proper tool of deductive inference. On the other hand, I see probabilistic accounts of accepted principles of scientific inference and decision making not as putting those principles on firmer foundations, but rather as test cases on which to establish the bona fides of probability logic, so that this logic can then be deployed with some confidence in new normative applications.

4. *What do you consider the most neglected topics and/or contributions in probability theory and/or statistics?*

Judging from the presentations at recent meetings of the Philosophy of Science Association, the European Philosophy of Science Association, the Gesellschaft fuer Analytische Philosophie, and the annual Formal Epistemology Workshops, probabilistic epistemology and philosophical decision theory are thriving, and have attracted the attention of some of the best young philosophers in the world.

I would, however, like to see philosophers more engaged in applications of this work to areas such as legal reasoning, group decision making, evidence-based medicine, and statistical literacy. With respect to neglected contributions, I would call attention to the priority of Volker Strassen (1964) in exploring the upper and lower probabilities induced by a multi-valued mapping from a probability space to another set of possible states of the world, and of Theodore Hailperin (1965, 1984, 1996) in developing the theory of probabilistic logic. Priority has often been mistakenly attributed to Arthur Dempster (1967) in the former case, and to Nils Nilsson (1986) in the latter, especially in the artificial intelligence community. I call attention to the work of Strassen and Hailperin, not simply as a matter of just attribution, but also because their work is lucid and illuminating.

As for work by scholars outside the philosophy community, Peter Walley's work on imprecise probabilities seems finally to be attracting the attention it deserves among philosophers, a trend that I enthusiastically endorse. In addition to his *Statistical Reasoning with Imprecise Probabilities*, mentioned above, I would call attention to his beautiful survey paper (Walley 2000) on the hierarchy of models of uncertainty. To philosophers interested in the role of probability in decision theory, the economist Peter Fishburn's (1988) book on utility theory is an indispensable reference. Finally, those interested in applications of probability and statistics to legal reasoning should consult the monumental work of the legal scholar David Kaye, whose papers are too numerous to reference here.⁶

5. *What do you consider the most important open problems in probability and/or statistics and what are the prospects for progress?*

I can only offer a reply to this question from the limited perspective of someone whose interests and expertise (such as it may be) relate chiefly to probabilistic epistemology. That said, I would like to see the relation between epistemic and aleatory probability more clearly delineated. This relation is far from being clarified by the admonition to adopt "known" chances as one's epistemic probabilities. There is a certain comfort in the logically unassailable position of subjectivism. But stable relative frequencies abound, and while subjectivists can predict this fact, they cannot explain it. Indeed, laws of large numbers and even simple theorems of probability such as Chebyshev's inequality, are oddly unexciting when cast in subjectivist terms, amounting simply to the advice to place your bets on certain events in a particular way if you have already decided to place your bets on certain other events in a particular way. The simple-minded *identification* of aleatory probability with hypothetical limiting relative frequencies does no better, since its purported explanation of observed stable relative

frequencies either amounts to a tautology or involves an infinite regress, as a number of individuals have noted. Ideally, one would like to know what *assurances* follow from the knowledge that one possesses a correct probability model.⁷ Branden Fitelson (2006) has begun to explore such *epistemic bridge principles*, one goal of which is to secure the knowledge that observation E evidentially supports H on the basis of knowledge that one possesses the correct model of the stochastic process whose outcome led to knowledge of E and prior knowledge that E (incrementally) confirms H in that model. He has observed that the articulation of such principles will require a careful account of the metaphysics of stochastic processes and a “delicate story about the epistemology of probability.” The successful completion of this project would be a significant contribution to probabilistic epistemology.

NOTES

1. Remarkably, neither Anscombe nor Hartigan uttered a word about subjective probability or Bayesian statistics in their lectures.
2. Generalizations of expectation and conditionalization are particularly well-behaved for 2-monotone lower probabilities. See Walley (1981, sections 6 and 7).
3. See Fitelson (1999, p. 363, footnote 2) for a list references supporting this view.
4. A sigma algebra \mathbf{A} of subsets of Ω is *purely atomic* if the family \mathbf{A}^* of atomic events in \mathbf{A} is countable, and constitutes a partition of Ω . Every finite sigma algebra is purely atomic, and if Ω is countable, then every sigma algebra on Ω is purely atomic (Renyi 1970, Theorems 1.6.1, 1.6.2).
5. Since u is assumed, *inter alia*, to predate the discovery of a certain relevance relation between H and E, it is natural to explore the consequences of assuming that u predates knowledge of *any* such relevance relation.
6. As an especially interesting example, however, see Kaye (1982).
7. Here's the sort of thing I have in mind. Let x_1, \dots, x_n a sequence of numerical observations, with \bar{x} denoting their mean $(\sum x_i)/n$ and s^2 their variance $(\sum(x_i - \bar{x})^2)/n$. If you know only the values of \bar{x} and s , then for every $h > 0$, you can be certain, as a consequence of Chebyshev's inequality, that the fraction of those n observations that fall outside the closed interval $[\bar{x}-hs, \bar{x}+hs]$ is strictly less than $1/h^2$. Of course, in this case our probabilities are just observed relative frequencies, and the “bridge principle” simply a matter of combinatorics (though not entirely uninteresting).

REFERENCES

- J. Aczel and C. Wagner (1980), A characterization of weighted arithmetic means, *SIAM Journal on Algebraic and Discrete Methods* 1, 259-260.
- J. Aczel, C. Ng, and C. Wagner (1984), Aggregation theorems for allocation problems, *SIAM Journal on Algebraic and Discrete Methods* 5, 1-8.
- S. Chatterjee and E. Seneta (1977), Towards consensus: some convergence theorems on repeated averaging, *Journal of Applied Probability* 14, 89-97.
- M. DeGroot (1974), Reaching a consensus, *Journal of the American Statistical Association* 69, 118-121.
- A. Dempster (1967), Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics* 38, 325-339.
- B. Fitelson (1999), The plurality of Bayesian measures of confirmation and the problem of measure sensitivity, *Philosophy of Science* 66, S362-S378.

- _____ (2006), Logical foundations of evidential support, *Philosophy of Science* 73, 500-512.
- C. Genest and C. Wagner (1987), Further evidence against independence preservation in expert judgement synthesis, *Aequationes Mathematicae* 32, 74-86.
- I. Good (1985), Weight of evidence: a brief survey, in *Bayesian Statistics 2*, J. Bernardo et al, eds., North-Holland.
- T. Hailperin (1965), Best possible inequalities for the probability of a logical function of events, *American Mathematical Monthly* 72, 343-359.
- _____ (1984), Probability logic, *Notre Dame Journal of Formal Logic* 25, 198-212.
- _____ (1996), *Sentential Probability Logic*, Lehigh University Press.
- R. Jeffrey (1965), *The Logic of Decision*, McGraw-Hill.
- _____ (1991), Postscript 1991: new explanation revisited, in Jeffrey 1992, 103-107.
- _____ (1992), *Probability and the Art of Judgement*, Cambridge University Press.
- _____ (1995), Probability reparation: the problem of new explanation, *Philosophical Studies* 77, 97-102.
- J. Kaplan (1968), Decision theory and the factfinding process, *Stanford Law Review* 20, 1065.
- D. Kaye (1982), Statistical evidence of discrimination, *Journal of the American Statistical Association* 77, 773-783.
- R. Laddaga (1977), Lehrer and the consensus proposal, *Synthese* 36, 473-477.
- R. Laddaga and B. Loewer (1985), Destroying the consensus, *Synthese* 62, 79-95.
- I. Levi (1985), Consensus as shared agreement and outcome of inquiry, *Synthese* 62, 3-11.
- K. Lehrer (1976), When rational disagreement is impossible, *Nous* 10, 327-332.
- K. Lehrer and C. Wagner (1981), *Rational Consensus in Science and Society*, Reidel.
- _____ (1983), Probability amalgamation and the independence issue: a reply to Laddaga, *Synthese* 55, 339-346.
- K. McConway (1981), Marginalization and linear opinion pools, *Journal of the American Statistical Association* 76, 411-414.
- N. Nilsson (1986), Probabilistic logic, *Artificial Intelligence* 28, 71-87.
- A. Renyi (1970), *Foundations of Probability*, Holden-Day.
- G. Shafer (1976), *A Mathematical Theory of Evidence*, Princeton University Press.
- V. Strassen (1964), Messfehler und information, *Zeitschrift fuer Wahrscheinlichkeitstheorie* 2, 273-305.
- L. Tribe (1971), Trial by mathematics: precision and ritual in the legal process, *Harvard Law Review* 84, 1329.
- B. van Fraassen (1989), *Laws and Symmetry*, Oxford University Press (Clarendon).

C. Wagner (1984), Aggregating subjective probabilities: some limitative theorems, *Notre Dame Journal of Formal Logic* 25, 233-240.

_____ (1989), Consensus for belief functions and related uncertainty measures, *Theory and Decision* 26, 295-304.

_____ (1992), Generalized probability kinematics, *Erkenntnis* 36, 245-257.

_____ (1997), Old evidence and new explanation, *Philosophy of Science* 64, 677-690.

_____ (1999), Old evidence and new explanation II, *Philosophy of Science* 66, 283-288.

_____ (2001), Old evidence and new explanation III, *Philosophy of Science* 68, S165-S175.

_____ (2002), Probability kinematics and commutativity, *Philosophy of Science* 69, 266-278.

_____ (2003), Commuting probability revisions: the uniformity rule, *Erkenntnis* 59, 349-364.

P. Walley (1981), Coherent lower (and upper) probabilities, *unpublished manuscript*.

_____ (1991), *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall.

_____ (2000), Towards a unified theory of imprecise probability, *International Journal of Approximate Reasoning* 24, 125-148.

