

Comments on “Constructing a Logic of Plausible Inference: A Guide to Cox’s Theorem”, by Kevin S. Van Horn

Glenn Shafer
Rutgers Business School
180 University Avenue
Newark, New Jersey 07104
gshafer@andromeda.rutgers.edu
www.glennshafer.com

March 13, 2003

Professor Van Horn has reviewed much of the work that followed Richard Cox’s publication of his axioms for probability in 1946. My comments will emphasize work that came earlier and also merits attention. In particular, I will discuss work by Sergei Bernstein that is closely related to Cox’s but has been neglected by Cox and his commentators.

Before reviewing Bernstein’s work, I will discuss the context of Cox’s work and explain why it did not dissuade me from studying alternative representations of uncertainty. After discussing Bernstein, I will make just one comment on Van Horn’s presentation.

1 Richard Cox and his times

Richard Threlkeld Cox (1898–1991) received a doctoral degree in physics from Johns Hopkins University in 1924. He then taught at New York University until 1943, when he returned to Johns Hopkins. In addition to teaching physics there, he served as dean of the college of arts and sciences for seven

years. He worked in several areas of physics, including statistical mechanics and the scattering of electrons.

Though he was not a specialist in pure probability theory, Cox published an article on the axiomatization of probability, “Probability, Frequency, and Reasonable Expectation” [4], in the *American Journal of Physics*, a well-respected journal that specializes in pedagogical and expository articles. In 1961, he expanded the article to a small book [5]. It is the article that is usually cited when Cox’s axioms are discussed.

Cox situated his work on the axiomatization of probability in the context of debates about the meaning of probability that had been conducted in English in the preceding decades. Everyone who wrote about the interpretation of probability, Cox observed, had a different opinion, but the clearest line of division was between those who interpreted probabilities as frequencies in an ensemble and those who interpreted probabilities as degrees of reasonable expectation. Cox enlisted on the side of reasonable expectation, and he proposed axioms for reasonable expectation that imply the usual rules of the probability calculus. He did not claim to be the first to do this; he cited John Maynard Keynes [10] and Harold Jeffreys and Dorothy Wrinch [9, 22] as predecessors. But he felt that the axioms advanced by these predecessors showed “some of the tool marks of their original derivation from the study of games of chance, with the consequent implication of an ensemble” ([4], p. 5). His own axioms, he felt, escaped from this residual frequentism.

Cox’s perspective on previous work on the interpretation of probability was rather narrow. This appears to have been due, at least in part, to the influence of Keynes. We find this statement in the preface of Cox’s 1961 book:

I have tried to indicate my obligations to other writers in the notes at the end of the book. Even without any such indication, readers familiar with *A Treatise on Probability* by the late J. M. Keynes would have no trouble in seeing how much I am indebted to that work. It must have been thirty years or so ago that I first read it, for it was almost my earliest reading in the theory of probability, but nothing on the subject that I have read since has given me more enjoyment or made a stronger impression on my mind.

Although Keynes had included many continental authors in the bibliography to his book, he emphasized his English predecessors, and because of the early

date of his work,¹ he did not take into account the vigorous debate on the foundations of probability that took place in French, German, Italian, and Russian during the first third of the twentieth century [20], the fruits of which included Andrei Kolmogorov’s axiomatization of mathematical probability [11], Jean Ville’s introduction of martingales [21], and Bruno de Finetti’s personalistic formulation of subjective probability [6]. Cox followed Keynes in emphasizing English predecessors and ignoring the twentieth-century continental debate. In 1946 and again in 1961, Cox does not mention fellow subjectivists such as Henri Poincaré, Émile Borel, or Bruno de Finetti. And he does not discuss the work of Sergei Bernstein. Bernstein had been listed in Keynes’s bibliography, and Cox lists him in the bibliography of his 1961 book, but with no comment.

As we can see from Van Horn’s bibliography, this narrowness of perspective has been perpetuated in subsequent discussion of Cox’s work. I hope that my contribution here will encourage future commentators to place Cox in a richer historical perspective.

2 Why Cox did not dissuade me

When Cox wrote, there was little interest in alternatives to the standard probability calculus for the representation of evidence or belief, and so it is not clear whether Cox would have seen his axioms as excluding such alternatives. In agreement with Keynes, he wrote that “it is hardly to be supposed that every reasonable expectation should have a precise numerical value” ([4], p. 9), and so he might well have conceded some role to representations that use a pair of numbers to represent the evidence for a proposition, one representing the degree of belief in the proposition justified by the evidence, and another, possibility larger in magnitude, representing the degree to which the proposition remains plausible in light of the evidence. In recent decades, however, Cox’s axioms have been used to argue against such representations. Cox’s axioms are axiomatic, it is argued, and so only the standard probability

¹*A Treatise on Probability* was published in 1921 but appears to have been based on research done much earlier. In his preface, Keynes writes “I propound my systematic conception of this subject for criticism and enlargement at the hands of others, doubtful whether I myself am likely to get much further, by waiting longer, with a work, which, beginning as a Fellowship Dissertation, and interrupted by the war, has already extended over many years.”

calculus should be used to represent evidence.

Most of my own scholarly work has been devoted to representations of uncertainty that depart from the standard probability calculus, beginning with my work on belief functions in the 1970s and 1980s and continuing with my work on causality in the 1990s [18] and my current work with Vladimir Vovk on game-theoretic probability ([19], www.probabilityandfinance.com). I undertook all of this work after a careful reading, as a graduate student in the early 1970s, of Cox's paper and book. His axioms did not dissuade me.

As Van Horn notes, with a quote from my 1976 book [17], I am not on board even with Cox's implicit assumption that reasonable expectation can normally be expressed as a single number. I should add that I am also unpersuaded by Cox's two explicit axioms. Here they are in Cox's own notation:

1. The likelihood $\sim \mathbf{b}|\mathbf{a}$ is determined in some way by the likelihood $\mathbf{b}|\mathbf{a}$:

$$\sim \mathbf{b}|\mathbf{a} = S(\mathbf{b}|\mathbf{a}).$$

where S is some function of one variable.

2. The likelihood $\mathbf{c} \cdot \mathbf{b}|\mathbf{a}$ is determined in some way by the two likelihoods $\mathbf{b}|\mathbf{a}$ and $\mathbf{c}|\mathbf{b} \cdot \mathbf{a}$:

$$\mathbf{c} \cdot \mathbf{b}|\mathbf{a} = F(\mathbf{c}|\mathbf{b} \cdot \mathbf{a}, \mathbf{b}|\mathbf{a}),$$

where F is some function of two variables.

I have never been able to appreciate the normative claims made for these axioms. They are abstractions from the usual rules of the probability calculus, which I do understand. But when I try to isolate them from that calculus and persuade myself that they are self-evident in their own terms, I draw a blank. They are too abstract—too distant from specific problems or procedures—to be self-evident to my mind.

It may be useful, in this connection, to quote Cox's own argument for $\mathbf{c} \cdot \mathbf{b}|\mathbf{a} = F(\mathbf{c}|\mathbf{b} \cdot \mathbf{a}, \mathbf{b}|\mathbf{a})$:

Written in symbolic form, this assumption may not appear very axiomatic. Actually it is a familiar enough rule of common sense, as an example will show. Let \mathbf{b} denote the proposition that an athlete can run from one given place to another, and let \mathbf{c} denote the proposition that he can run back without stopping. The

physical condition of the runner and the topography of the course are described in the hypothesis \mathbf{a} . Then $\mathbf{b}|\mathbf{a}$ is the likelihood that he can run to the distant place, estimated on the information given in \mathbf{a} , and $\mathbf{c}|\mathbf{b} \cdot \mathbf{a}$ is the likelihood that he can run back, estimated on the initial information and the further assumption that he has just run one way. These are just the likelihoods that would have to be considered in estimating the likelihood, $\mathbf{c} \cdot \mathbf{b}|\mathbf{a}$, that he can run the complete course without stopping. In postulating only that the last-named likelihood is some function of the other two, we are making the least restrictive assumption. ([4], p. 6)

It surely does make sense to decompose problems of probability judgement into subproblems. The function F puts back together the judgements we make in the subproblems. But why should we always use the same function F to put subproblems back together? Might we not use a different F if we decompose the problem differently, say by considering first the adequacy of the runner's muscles and then the adequacy of his heart and lungs? Might we not use a different F in an entirely different problem? The only argument I see for trying always to use the same F is the example provided by the probability calculus.

Perhaps it is not out of place to recall just what is at issue when we ask, along with Cox, whether his two assumptions are "axiomatic". This word is used rather freely nowadays. When I wear a pure mathematician's hat, my calling something an axiom means only that I want to explore its logical consequences. But when Cox asks whether his assumptions are axiomatic, he is evoking an older sense of the word—the sense used by Euclid, who called an entirely self-evident assumption (things which equal the same thing also equal one another) an axiom, while calling a more debatable assumption (parallel lines never meet) a postulate. Although Van Horn's tone is sympathetic to Cox, his final conclusion, that he cannot make a compelling case for Cox's assumptions, contradicts Cox's claim that these assumptions are axiomatic.

Extended debate of this point seems futile, however. Some people are persuaded by Cox's argument. Some are not. We must leave the matter there.

3 The earlier work of Sergei Bernstein

It may be more productive to call attention to the axioms for probability published in Russian in 1917 [1] by Sergei Natanovich Bernstein (1880–1968).

Bernstein received his doctoral degree in analysis in Paris in 1904, and he is known for his work in various areas of analysis, including elliptic differential equations, approximation theory, and the theory of analytic functions. Within probability theory, he is best known for his early work on the central limit theorem for dependent random variables. He began to publish work in probability only around 1917, long after his studies in France, but his philosophical views on probability were very much in the tradition of French probabilists such as Henri Poincaré [16], Émile Borel [3], and Paul Lévy [14]. Like all these authors, and like Richard Cox after him, Bernstein believed that probability is essentially subjective and becomes objective only when there is sufficient consensus or adequate evidence.

There is an important difference, however, between Bernstein and the French probabilists on the one hand and Cox on the other. Cox wanted to divorce his subjective conception of probability from the origins of probability theory in games of chance, which he believed was tied to the concept of frequency. Bernstein and the French probabilists, on the other hand, based their subjective conception of probability squarely on the traditional concept of equally likely cases, which was the classical foundation for probability.

According to the classical foundation, first formulated by Abraham De Moivre in *The Doctrine of Chances* in 1718 [7], the probability of an event is the ratio of the number of equally likely cases that favor it to the total number of equally likely cases possible under the circumstances. From this definition, one derives the rules of probability as theorems. The *theorem of total probability* says that if A and B cannot both happen,

$$\begin{aligned} & \text{probability of } A \text{ or } B \text{ happening} \\ &= \frac{\# \text{ of cases favoring } A \text{ or } B}{\text{total } \# \text{ of cases}} \\ &= \frac{\# \text{ of cases favoring } A}{\text{total } \# \text{ of cases}} + \frac{\# \text{ of cases favoring } B}{\text{total } \# \text{ of cases}} \\ &= (\text{probability of } A) + (\text{probability of } B). \end{aligned}$$

The *theorem of compound probability* says

$$\begin{aligned}
 & \text{probability of both } A \text{ and } B \text{ happening} \\
 &= \frac{\# \text{ of cases favoring both } A \text{ and } B}{\text{total } \# \text{ of cases}} \\
 &= \frac{\# \text{ of cases favoring } A}{\text{total } \# \text{ of cases}} \times \frac{\# \text{ of cases favoring both } A \text{ and } B}{\# \text{ of cases favoring } A} \\
 &= (\text{probability of } A) \times (\text{probability of } B \text{ if } A \text{ happens}).
 \end{aligned}$$

These arguments, often associated with the name of Laplace, were still standard fare in the probability textbooks of the early twentieth century, including those by Henri Poincaré [16] and Andrei Markov [15].

In his 1917 article, “On the axiomatic foundation of the theory of probability”, Bernstein accepted equally likely cases as the starting point. But instead of arbitrarily defining numerical probability as the number of favorable cases to the total number of cases, he derived this definition from qualitative axioms. Here are his two most important axioms:

- If A and A_1 are equally likely, B and B_1 are equally likely, A and B are incompatible, and A_1 and B_1 are incompatible, then $(A \text{ or } B)$ and $(A_1 \text{ or } B_1)$ are equally likely.
- If A occurs, the new probability of a particular occurrence α of A is a function of the initial probabilities of α and A .

The first of these axioms can be thought of as a qualitative statement of the theorem of total probability, the second as a qualitative statement of the theorem of compound probability. Using the first axiom, Bernstein deduced that if A is the conjunction of m out of n equally likely and incompatible propositions, and B is as well, then A and B must be equally likely. It follows that the numerical probability of A and B is some function of the ratio m/n , and we may as well take that function to be the identity. Using the second axiom, Bernstein then deduced that the new probability of α when A occurs is the ratio of the initial probability of α to that of A .

The commonalities with Cox’s work are striking. Cox’s axioms are also qualitative statements of the theorems of total and compound probability. Like Bernstein, Cox includes a dose of convention in his argument. Bernstein says we might as well take our function of the ratio m/n to be the identity. Cox deduces that his function F must have the form

$$F(x, y) = C f(x) f(y),$$

where C is a constant and f is an arbitrary function of a single variable, and he then says we might as well take f to be the identity and C to equal one. There are also important differences. For example, Bernstein's counterpart of the rule of compound probability says that the probability of one event given another is a function of unconditional probabilities, while Cox goes in the more traditional direction, saying that the joint probability is a function of the probability of the first event and the probability of the second given the first.

The most important difference between the two authors is their attitude towards the concept of equally likely cases. Bernstein accepted this concept as his starting point. Keynes also retained a version of the concept, his "principle of indifference". This is what Cox saw as the residual frequentism in Keynes's thinking, which he claimed to eliminate. It is precisely this elimination that makes Cox's reasoning unpersuasive for me. I see the probability calculus as a special, not universal, framework for uncertain reasoning, and the concept of equally likely cases provides one way of seeing what is special about it.

Like Cox, Bernstein axiomatized the field of propositions as well as the concept of numerical probability. He also extended his theory to the case where this field is infinite. He repeated the exposition of his axioms again in a probability textbook that he published in 1927 [2], but neither the article nor the book were ever translated out of Russian into other languages. The neglect of his work is primarily due, no doubt, to its linguistic inaccessibility. His work deserves recognition, however, and I would like to see it taken into account in future discussions of Cox's work. This would enlarge both the historical and the philosophical context of these discussions.

4 Conclusion

Cox believed that the classical understanding of probability in terms of equally likely cases was hopelessly infected with frequentism. In my judgement, this belief was historically myopic. True, the frequentists of the mid-twentieth century had laid claim to games of chance. But Poincaré, Borel, and Bernstein had acknowledged the centrality of games of chance and equally likely cases without adopting frequentism.

I believe that the classical understanding of the probability calculus retains great value even today. Mathematical probability grew out of games of

chance, and game-theoretic concepts remain at its core. As Vovk and I have shown [19], these game-theoretic concepts can be generalized very substantially. But there remain contexts where game theory may not provide the most useful way of assessing evidence and belief.

In this spirit of inclusiveness and tolerance, I would like to suggest that Van Horn and other commentators on Cox reconsider one aspect of their exposition: their use of the word “plausibility”. They need a synonym for “probability” that does not presuppose the rules they set out to derive. Cox himself, as we have noted, used “likelihood” in this role. This choice is more problematic today, because so many readers will interpret “likelihood” in the sense of R. A. Fisher [8]. So Van Horn and others use “plausibility”. This choice conflicts with the way I and others have used “plausibility” in the theory of belief functions [17], for in that theory both a proposition and its negation may be plausible. As it happens, we have the dictionary on our side. In English, plausibility, even of the greatest degree, is merely an appearance of truth, which we recognize may be deceptive. When participants in a debate appropriate the other side’s terms of discourse in a way that contradicts the dictionary, the coherence and civility of the debate is imperiled. So I suggest they use some other term. How about “likeliness”?

References

- [1] Sergei N. Bernstein. Опыт аксиоматического обоснования теории вероятностей (On the axiomatic foundation of the theory of probability). Сообщения Харьковского математического общества (*Communications of the Kharkiv mathematical society*), 15:209–274, 1917.
- [2] Sergei N. Bernstein. Теория вероятностей (*Theory of Probability*). Государственное Издательство. Moscow and Leningrad. 1927.
- [3] Émile Borel. La valeur pratique du calcul des probabilités. *Revue du Mois*, 1:424–437, 1906.
- [4] Richard T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14:1–13, 1946.
- [5] Richard T. Cox. *The algebra of probable inference*. Johns Hopkins Press. Baltimore. 1961.

- [6] Bruno de Finetti. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. An English translation by Henry E. Kyburg, Jr., is included in both editions of [12].
- [7] Abraham De Moivre. *The Doctrine of Chances*. London, 1718. Third edition. (The first edition appeared in 1718, the second in 1738.)
- [8] Ronald A. Fisher. *Statistical Methods and Scientific Inference*. London. Hafner. 1956.
- [9] Harold Jeffreys *Theory of Probability*. Clarendon Press. Oxford. 1939.
- [10] John Maynard Keynes *A Treatise on Probability*. Macmillan. London. 1921.
- [11] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956. A Russian translation, by G. M. Bavli, appeared under the title *ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ* (Наука, Moscow) in 1936, with a second edition, slightly expanded by Kolmogorov with the assistance of Albert N. Shiryaev, in 1974.
- [12] Henry E. Kyburg, Jr., and Howard E. Smokler. *Studies in Subjective Probability*. Wiley. New York. 1964 This selection of readings ranges chronologically from John Venn in 1888 to Leonard J. Savage in 1961.
- [13] Henry E. Kyburg, Jr., and Howard E. Smokler. *Studies in Subjective Probability*. Second Edition. Krieger. New York. 1980. This selection of readings is slightly different from that in the first edition.
- [14] Paul Lévy. *Calcul de probabilités*. Gauthier-Villars, Paris, 1925.
- [15] Andrei Markov (1856–1922). *Исчисление вероятностей (Calculus of Probability)*. Saint Petersburg, Academy of Sciences, 1900. A second edition appeared in 1908, and a German translation of the second edition appeared in 1912.
- [16] Henri Poincaré. *Calcul des probabilités*, second edition. Gauthier-Villars, Paris. 1912. The first edition appeared in 1896.

- [17] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press. Princeton. 1976.
- [18] Glenn Shafer. *The Art of Causal Conjecture*. The MIT Press. Cambridge, Massachusetts. 1996.
- [19] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley. New York, 2001.
- [20] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. Working Paper No. 5, Game-Theoretic Probability and Finance Project. www.probabilityandfinance.com.
- [21] Jean Ville (1910–1988). *Étude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939.
- [22] Dorothy Wrinch and Harold Jeffreys. *Philosophical Magazine*. Sixth Series, 38, 1919.