

New Axioms for Rigorous Bayesian Probability

Maurice J. Dupré
Tulane University

Frank J. Tipler
Tulane University

Abstract. By basing Bayesian probability theory on five axioms, we can give a *trivial* proof of Cox’s Theorem on the product rule and sum rule for conditional plausibility without assuming continuity or differentiability of plausibility. Instead, we extend the notion of plausibility to apply to unknowns, giving them plausible values. Thus, we combine the best aspects of two approaches to Bayesian probability theory, namely the Cox-Jaynes theory and the de Finetti theory.

Keywords: Cox, de Finetti, Jaynes, Axiomatic Bayesian Probability, Sum Rule, Product Rule

1 Introduction

The problem of developing the foundations of Laplacian or Bayesian probability has led to many different approaches, which are mainly variations on two. The first, which is our main concern is due to R. T. Cox (1961). The second is due to Bruno de Finetti (1974), de Finetti (1975) and a concise general treatment is in Regazzini (1985). We find that by combining aspects of both, we can substantially weaken the assumptions and arrive at a proof of the problematic Cox theorem which is very simple, *in fact trivial*, as it should be. This triviality is of the utmost importance. The foundational axioms assumed by de Finetti for conditional expectation (his conditional prevision) require a positive normalized linear functional on a partially ordered commutative algebra (with identity element and over the reals) of objects more general than random variables which he calls “random quantities” and which we will simply call “unknowns” or “unknown numerical quantities”. Even though it is not explicitly pointed out axiomatically, de Finetti, in his discussion in the first two chapters of his book, assumes that the ordering of the lattice of idempotents is inherited from that of the algebra. To define his conditional prevision, he requires a coherence condition (see also Regazzini (1985)) equivalent to a quadratic minimization property assumed to be satisfied by the choice of previsions. We merely assume the existence of *some* form of rules and demonstrate that if the assignment of plausible value (what de Finetti calls “prevision”) is required to be logically consistent, then the actual rules themselves are uniquely determined. In particular, as a consequence, the rules of conditional expectation in Bayesian probability theory are uniquely determined by the requirement of logical consistency. Thus we find a *trivial* proof of the Cox Theorem, which properly understood, is a uniqueness theorem, and we show that the theory of de Finetti is a consequence. That is, de Finetti assumes his coherence condition, meaning existence of a positive linear functional which also satisfies a certain quadratic minimization criteria, and he uses the quadratic minimization criteria to prove the multiplication rule of conditional expectation, provided that the commutative algebra contains the indicators of a Boolean algebra of statements. But this quadratic

minimization property is essentially a consequence of the rules of conditional expectation which we prove are uniquely determined by weaker and more natural assumptions. Of course, any positive linear functional on an algebra of continuous functions can be represented as an integral against a measure Rudin (1966), and consequently it has long been known that the Kolmogorov approach can be developed from the basic rules of expectation (see for instance Whittle (2000) where the product rule of conditional probability is assumed as an axiom). But the de Finetti approach is more general in that the unknowns or random quantities are no longer assumed to be functions on some sample space. In fact there need not be any sample space. Thus, we shall generalize the Cox approach to apply to the de Finetti unknowns (Cox restricted attention to a Boolean algebra of statements — still no sample space) and more generally to any partially ordered commutative algebra (over the reals and with identity) of objects we might want to call unknowns and which contains the Boolean algebra of indicators of the statements in which we are interested. We will make all our assumptions as clearly stated axioms which are easily motivated.

Bayesian probability theory, as expressed by Cox (1961) and followed by Jaynes (2003) in his classic book, is not defined by precise axioms, but by three “desiderata”: (I) representations of plausibility are to be given by real numbers; (II) plausibilities are to be in qualitative agreement with common sense; and (III) the plausibilities are to be “consistent”, in the sense that anyone with the same information would assign the same real numbers to the plausibilities. It is known that Cox’s 1961 development (mainly copied by Jaynes (2003)) is non-rigorous, and in fact a counterexample has been found by Halpern (1999a). Attempts to circumvent this counterexample have led to very complicated approaches (Halpern (1999b), Hardy (2002), who used the work of Koopman (1940), Kraft et al. (1959), Regazzini (1985), Scott (1964), and Tribus (1969)). The assumption of differentiability or even continuity is questionable since the Boolean algebra of statements may only be finite. We shall see that our axioms are mathematically simpler and more intuitive than Cox’s desiderata. In particular, we shall not need to make any continuity or differentiability assumptions on any assumed function. Our results are special cases of more general trivial facts about retraction mappings which can be applied to give plausible value to more general objects such as unknown vectors or operators and as well there are other possible choices for the simple axioms (Dupre and Tipler (2006)).

2 The Five Axioms for Bayesian Probability Theory

If T is an algebra (over \mathbb{R} and with identity), then we denote by $E(T)$ the set of idempotents of T . We generally assume all our algebras are commutative and partially ordered, and we will assume that statements of interest form a Boolean subalgebra of $E(T)$. The set $E(T)$ of idempotents in the commutative algebra T naturally forms a Boolean algebra under a natural order (if $x, y \in T$, are idempotent where T is any algebra over \mathbb{R} , then we say $x \leq y$ if and only if $yx = x$). It is also obvious that if $e, c \in E(T)$, we have

$$e \cup c = e + c - ec. \quad (2.1)$$

As far as the Boolean algebra E itself is concerned, if we start with it alone, we can note that in general, by Stone's Theorem (discussed at length in Halmos (1963)), we can embed E as a Boolean algebra of idempotents in the algebra C_E of continuous real valued functions on the Stone space of E .

To proceed formally, we list our axioms.

AXIOM 1. STRUCTURE OF UNKNOWN REAL NUMBERS AND PLAUSIBLE VALUE. *We assume a set T of unknown numbers is a partially ordered commutative algebra over \mathbb{R} with identity, 1.*

*We assume in addition a given sub-Boolean algebra E of $E(T)$ with $0, 1 \in E$ and denote by E_0 the set of non-zero members of E . We assume that the partial ordering in $E(T)$ as a Boolean algebra coincides with the ordering that $E(T)$ inherits from the algebra T . Finally, we assume a function $PV : T \times E_0 \rightarrow \mathbb{R}$, called **PLAUSIBLE VALUE**, whose value on the pair (x, e) is denoted $PV(x|e)$.*

AXIOM 2. STRONG RESCALING FOR PLAUSIBLE VALUE. *If a, b belong to \mathbb{R} , if x belongs to T , and if e belongs to E_0 , then*

$$PV(ax + b|e) = aPV(x|e) + b. \quad (2.2)$$

AXIOM 3. ORDER CONSISTENCY FOR PLAUSIBLE VALUE. *If $x, y \in T$ and if $e \in E_0$, implies that $x \leq y$, then $PV(x|e) \leq PV(y|e)$.*

Notice that if $e \in E(T)$, then $0 \leq e \leq 1$, in T , as it is true in the lattice ordering of $E(T)$.

AXIOM 4. THE COX AXIOM FOR PLAUSIBLE VALUE: *If e, c are fixed in E , with $ec \in E_0$, if x_1, x_2 are in T , if $PV(x_1|ec) = PV(x_2|ec)$, then $PV(x_1e|c) = PV(x_2e|c)$. That is, we assume that as a function of x , the plausible value $PV(xe|c)$ depends only on $PV(x|ec)$.*

AXIOM 5. RESTRICTED ADDITIVITY OF PLAUSIBLE VALUE. *For each fixed $y \in T$ and $e \in E_0$, the plausible value $PV(x + y|e)$ as a function of $x \in T$ depends only on $PV(x|e)$, which is to say that if $x_1, x_2 \in T$ and $PV(x_1|e) = PV(x_2|e)$, then $PV(x_1 + y|e) = PV(x_2 + y|e)$.*

3 Motivations for the Axioms

We define the *plausibility* $PL(A|B) = PV(A|B)$ for $A \in E$, $B \in E_0$ by naturally regarding $E \subset E(T)$, where statements are identified with their indicator unknowns.

Since *not* $A = 1 - A$, an immediate consequence of axiom 2 is then the fact that

$$PL(\text{not } A|B) = 1 - PL(A|B), \quad A \in E, \quad B \in E_0, \quad (3.1)$$

which is the additivity part of the Cox Theorem.

We point out here that the structural properties of the algebra of unknowns related to ordering are very intuitive if we think of unknowns as having descriptions as numerical quantities in the sense of de Finetti (1974), and this structure is implicitly assumed by de Finetti without stating it. Let us consider the motivation for our last two axioms. The Cox axiom 4 for plausible value is a natural generalization to plausible value of the corresponding axiom assumed by Cox for plausibility of statements. The usual motivation for the Cox axiom begins with the idea that $PL(AB|C)$ should or could depend on all of the quantities

$$PL(A|B\&C), \quad PL(B|C), \quad PL(A|C), \quad PL(A|B), \quad PL(B|AC), \quad PL(C|A), \quad PL(C|B),$$

through some real-valued function of 7 real variables and then observing that unless the function is independent of the last 5 quantities, absurdities result in various special cases (see Jaynes (2003) and Tribus (1969)). For instance, we can immediately eliminate the fourth and last two variables since C is to be assumed in calculating $PL(AB|C)$. In our situation, PL is replaced by its generalization PV , and A is now replaced by an arbitrary unknown, so the fifth and sixth variables can be eliminated immediately, as for a general unknown X , the plausible value for an unknown given X , is nonsense. Moreover, we will not assume that our function is defined on all of \mathbb{R}^3 , since the variables in which we have only indicators appearing need only be in the range of the plausibility function, which we will denote by J . That is, $J = \{PL(A|B) : A \in E, B \in E_0\}$, and we can assume, under the same reasoning that there is

$$f : \mathbb{R} \times J \times \mathbb{R} \longrightarrow \mathbb{R}$$

with the property that

$$PV(XA|B) = f(PV(X|A\&B), PL(A|B), PV(X|B)), \quad (3.2)$$

for any choice of $A, B \in E$ and $X \in T$, for which all the inputs make sense. Next, we will drop the assumption that f is in any way universal, that is we only assume it applies to our given setup consisting of the algebra T , the Boolean algebra E , and our specific function PV . If we take the case of $X = r(\text{not } A) + s$, where $r, s \in \mathbb{R}$, then as $(\text{not } A)A = 0$, we find that $XA = sA$, so $PV(XA|B) = sPL(A|B)$, is independent of the choice of $r \in \mathbb{R}$. But $PV(X|B) = rPL(\text{not } A|B) + s$, and if $PL(\text{not } A|B) \neq 0$, then we see that the last variable can be made to assume any real value without changing

$PV(XA|B)$, from which conclude that f must be independent of its last variable. Thus we arrive at the axiom 4, which we see is still more general, because, we in fact allow the function f to depend on A and B . More specifically, notice we are fixing A and B and only requiring that for each such choice of A and B there is a function $f_{(A,B)}$ for which $PV(XA|B) = f_{(A,B)}(PV(X|AB), PL(A|B))$. Thus, in a sense, we are assuming less than the usual axiom of Cox, but for a broadened domain, namely all of the unknowns and not merely for idempotents. The final axiom 5 is only necessary to obtain general additivity of plausible value and is really not necessary for the proof of the Cox theorem for plausibility, but it seems natural to want to have general additivity for plausible value. We could simply assume additivity of plausible value as being intuitively obvious, but we prefer to keep our assumptions to a reasonable minimum, and the axiom 5 is in the same spirit as the axiom 4.

Notice that by taking the case that $b = 0$ in (2.2) we immediately obtain **HOMOGENEITY OF PLAUSIBLE VALUE**:

$$PV(ax|e) = aPV(x|e), \quad a \in \mathbb{R}, \quad e \in E_0, \quad x \in T, \quad (3.3)$$

whereas by taking the case that $a = 0$ in (2.2) we immediately find the **RETRACTION PROPERTY OF PLAUSIBLE VALUE**:

$$PV(b|e) = b, \quad b \in \mathbb{R}, \quad e \in E_0. \quad (3.4)$$

Proposition 3.1. *If $x, y \in T$ and $e \in E_0$ implies that $x = y$, then $PV(x|e) = PV(y|e)$.*

Proof. This is an immediate consequence of axiom 3, simply because for real numbers, $=$ is the same as \leq & \geq . \square

Proposition 3.2. *If $x \in T$ and $e \in E_0$ implies that $x = r \in \mathbb{R}$, then $PV(x|e) = r$.*

Proof. This follows from (3.4) and proposition 3.1. \square

Now an immediate consequence of Axioms 2 and 3 is that if a and b are real numbers and $e \in E_0$ implies that $a \leq x \leq b$, then

$$a \leq PV(x|e) \leq b.$$

As a special case, if $0 \neq e, c \in E$, then $0 \leq c \leq 1$, so by Axioms 2 and 3 we can immediately conclude that

$$0 \leq PL(c|e) \leq 1.$$

4 The General Product Rule

Theorem 4.1. *If $x \in T$, if $e, c \in E$ with $ec \in E_0$, then*

$$PV(xc|e) = PV(x|ce)PV(c|e) = PV(x|ce)PL(c|e). \quad (4.1)$$

Proof. We consider a real-valued function

$$F = F_{(c,e)} : T \rightarrow \mathbb{R}$$

defined by

$$F(x) = PV(xc|e).$$

The axiom 4 says that $F(x)$ depends only on the numerical value $PV(x|ce)$ and not on the particular $x \in T$. Put another way, this says that there is a real-valued function

$$f = f_{(c,e)} : \mathbb{R} \rightarrow \mathbb{R}$$

so that for any $x \in T$, we have

$$PV(xc|e) = F(x) = f(PV(x|ce)).$$

If we take the case of $x = r \in \mathbb{R} \subset T$, then $F(x) = F(r) = PV(rc|e) = rPV(c|e)$, by (3.3). But as $r = PV(r|ce)$, this means, $f(r) = F(r) = rPV(c|e)$, for all $r \in \mathbb{R}$, which in turn means f is simply multiplication by $PV(c|e) = PL(c|e)$. But now, if $X \in T$ is arbitrary, we conclude that $PV(xc|e) = F(x) = f(PV(x|ce)) = PV(x|ce)PL(c|e)$.

□

Corollary 4.1. *If e, c, g belong to E , with $ec \in E_0$, then*

$$PL(gc|e) = PL(g|ce)PL(c|e), \tag{4.2}$$

which is the standard product rule of Cox's Theorem.

5 Conclusion

Our approach to the Cox theorem (4.1) has two major advantages over other treatments as is seen from the proof of theorem (4.1). First it is trivial. Second it has eliminated the problems which allow the counterexample of Halpern (1999a). Specifically, his counterexample violates the homogeneity property which plausibility here inherits from plausible value. We do not need to have an associative multiplication on the real line or an interval, we do not need to assume any continuity or differentiability or divisibility, we do not need to assume that our Boolean algebra of propositions has sufficiently many plausible values to have dense range in an interval of numbers. It is thus the homogeneity, not the additivity assumed by some authors Koopman (1940), which is crucial. Also, we have assumed our set of unknowns forms an algebra with identity over the reals as seems natural, but we really only used the fact that we have a set of unknowns that is closed under multiplication by indicators and contains all indicators from E and the real numbers themselves as a subset.

We have not mentioned general additivity of PV . If we are only concerned with the additivity of plausibility for statements in E , then this is already a consequence of

(4.2) and our axioms, using (2.1) and (3.1), as in Jaynes (2003). In fact, for arbitrary $A, B \in E$, repeated applications of the multiplication rule with deMorgan's Law yields easily

$$PL(A \text{ or } B|C) = PL(A|C) + PL(B|C) - PL(A \& B|C).$$

We therefore get the usual rules of probability on defining the conditional probability, $P(A|B)$, by $P(A|B) = PV(A|B) = PL(A|B)$. Then **plausibility satisfies all the axioms of ordinary probability theory by our axioms**. One can easily use the technique of the proof of the multiplication theorem to prove general additivity of plausible value using axiom 5.

About the Authors

Maurice J. Dupré is in the Mathematics Department at Tulane University, New Orleans, LA 70118. He can be reached at mdupre@tulane.edu.

Frank J. Tipler is in the Mathematics Department and in the Physics Department at Tulane University, New Orleans, LA 70118. He can be reached at tipler@tulane.edu.

Acknowledgments

The authors wish to thank the editor of Bayesian Analysis and the referee for very helpful suggestions.

References

- Cox, R. T. (1961). *The Algebra of Probable Inference*. Baltimore, Maryland: The Johns Hopkins Press.
- de Finetti, B. (1974). *Theory of Probability, vol. I*. New York: John Wiley and Sons.
- (1975). *Theory of Probability, vol. II*. New York: John Wiley and Sons.
- Dupre, M. J. and Tipler, F. J. (2006). “The Cox theorem, unknowns and plausible value.” *arXive*, math.PR/0611795.
- Halmos, P. (1963). *Lectures on Boolean Algebras*. Princeton, New Jersey: van Nostrand.
- Halpern, J. (1999a). “A counterexample to theorems of Cox and Fine.” *Journal of Artificial Intelligence Research*, 10: 67–85.
- (1999b). “Technical addendum, Cox's theorem revisited.” *Journal of Artificial Intelligence Research*, 11: 429–435.
- Hardy, M. (2002). “Scaled Boolean algebras.” *Advances in Applied Mathematics*, 29(2): 243–292.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge, U.K: Cambridge University Press.

- Koopman, B. O. (1940). "The axioms and algebra of intuitive probability." *Annals of Mathematics*, 41: 269–292.
- Kraft, C. H., Pratt, J. W., and Seidenberg, A. (1959). "Intuitive probability on finite sets." *Annals of Mathematical Statistics*, 30: 408–419.
- Regazzini, E. (1985). "Finitely additive conditional probabilities." *Rendiconti del Seminario Matematico e Fisico di Milano*, 55: 69–89.
- Rudin, W. (1966). *Real and Complex Analysis*. New York: McGraw-Hill.
- Scott, D. (1964). "Measurement structures and linear inequalities." *Journal of Mathematical Psychology*, 1: 233–247.
- Tribus, M. (1969). *Rational Descriptions, Decisions and Designs*. New York: Pergamon Press.
- Whittle, P. (2000). *Probability Via Expectation*. New York: Springer-Verlag, 4 edition.