Bayes Rules in Finite Models

Stefan Arnborg[1] and **Gunnar Sjödin**[2]

Abstract. Of the many justifications of Bayesianism, most imply some assumption that is not very compelling, like the differentiability or continuity of some auxiliary function. We show how such assumptions can be replaced by weaker assumptions for finite domains. The new assumptions are a non-informative refinement principle and a concept of information independence. These assumptions are weaker than those used in alternative justifications, which is shown by their inadequacy for infinite domains. They are also more compelling.

1 Introduction

The normative claim of Bayesianism is that every type of uncertainty should be described as probability. Bayesianism has been quite controversial in both the statistics and the uncertainty management communities. It developed as subjective Bayesianism, in [5, 11]. Recently, the information based family of justifications, initiated in [3] and continued in [1] have been discussed in [12, 6, 13]. We will try to find assumptions that are strong enough to strictly imply Bayesianism and at the same time convincing in a subjective way (common sense). In section 2 we give a short outline of the arguments of Cox and his followers and introduce the function F relating the plausibility of a conjunction to the plausibilities of its conjuncts, and the similar functions S and G describing plausibilities of complements and disjunctions. In section 3 we discuss the problem raised in recent papers and propose assumptions weaker than the standard ones: non-informative refinability and information independence. In 3.1 we observe that our assumptions imply that F and G must be associative and symmetric, and jointly distributive. In 3.2 we show that even if there is no direct violation of strict monotonicity, associativity or symmetry, there can be problems in a model that surface after a number of refinement steps, and we describe a theorem (proved in the appendix) saying that for a finite domain, and under our assumptions, natural refinements are possible if and only if the function F is transformed to multiplication by some monotone scaling of the plausibility measure. Our argument shows that a similar scaling must exist taking the function G to addition. We outline an argument showing that a rescaling taking $G(x, y)$ to $x+y$ must also take $F(x, y)$ to xy. In 3.3 we discuss the extension to infinite but non-dense domains. We show with an example that some new assumption is required for the infinite case before Bayes rules.

2 Summary of assumptions

In 1946, R.T. Cox published his findings [3] on some properties required by any good calculus of plausibility of statements. He stated three desiderata (the following is actually from [9], but very similar):

I: Divisibility and comparability- The plausibility of a statement is a real number and is dependent on information we have related to the statement.

II: Consistency - If the plausibility of a statement can be derived in two ways, the two results must be equal.

III: Common sense - Deductive propositional logic should be the special case of reasoning with statements known to be true or known to be false, and plausibilities should vary sensibly with the assessment of plausibilities in the model.

Cox's paper is very appealing to believers in Bayesianism, but sometimes more has been put in it than there is, and sometimes less. Although the first two desiderata can be criticized, they are usually accepted and their interpretation is uncontroversial. The third desideratum is clearly open-ended, and sometimes it has been interpreted as found suitable considering the proof method used rather than by some serious consideration of what is meant by common sense.

After introducing the notation $A|C$ for the plausibility of statement A given that we know C to be true, Cox finds the governing functional equation for defining the plausibility of a conjunction: $AB|C = F(A|BC, B|C)$ must hold for some function F. There is also assumed to be a function S describing complements: $S(A|C) = \overline{A}|C$. Using the rules of propositional logic and a number of auxiliary assumptions it is possible to show that there must exist a strictly monotone scaling $w(x)$ of the plausibility measure that satisfies the rules of probabilities, *i.e.*, takes F to multiplication and S to $1 - x$ – in other words, $w(F(x, y)) = w(x)w(y)$ and $w(S(x)) = 1 - w(x)$. The assumptions used by Cox are not explicitly stated, but they can be inferred from his text[13]. They involve among others some density and differentiability assumptions that have been criticized in [12, 6]. The existence of the function $w(x)$ that translates the plausibility measure to another measure satisfying the rules of probabilities will be called *rescalability*, and the main topic of investigation in this note is under what reasonable and precise assumptions rescalability obtains. From rescalability all the machinery of Bayesian analysis follows, except the way to assign prior probabilities.

A more precise derivation of rescalability with significantly weaker assumptions was published by Aczél[1]. Aczél relaxes the differentiability assumption of Cox and introduces the function G with the use: $A \vee B|C = G(A|C, B\overline{A}|C)$. It is then only necessary to assume continuity of G with the accompanying domain denseness, strict monotonicity, associativity, symmetry, distributivity and obvious boundary conditions to prove rescalability. Paris also gives a full-fledged proof of rescalability [12] that does not use differentiability assumptions, but still contains a density assumption of the set of plausibility values (which cannot hold in a finite setting).

An example is given in Halpern[6] of a finite model where the function F is only defined on the plausibilities appearing in the

[1] Nada, Kungl. Tekniska Högskolan, SE-100 44 Stockholm, SWEDEN email: stefan@nada.kth.se

[2] ARC, Swedish Institute of Computer Science, Box 1263, SE-164 29 Kista, SWEDEN email: sjodin@sics.se

model and is not associative. Since associativity is preserved by monotone scaling, it cannot be possible to rescale the plausibility measure of this model so that F is translated to multiplication, because multiplication is associative but F is not. The plausibilities of this model can thus not be considered equivalent to probabilities.

3 Weakening assumptions

We want to see if density and associativity assumptions are necessary, or if there are more fundamental arguments in favor of Bayesianism. We thus assume that the function F can be different in different models. But we do not regard a model as fixed and finished. Instead, we think of a model as being worked out incrementally by refinement, *i.e.*, by breaking down the situation in sub-cases to the degree desired in an application, and our method should not be such that we are unnaturally and arbitrarily constrained in how far and how our model can be refined. We adapt an informative refinement method, sometimes known as 'extending the discussion'[14, 7]. When a user is not confident about his assessment of $A|C$, he divides the context C into two subcases C_1 and C_2, and assesses $A|C_1$, $A|C_2$, $C_1|C$ and $C_2|C$. Now $A|C = G(AC_1|C, AC_2|C) = G(F(A|C_1, C_1|C), F(A|C_2, C_2|C))$. A weaker form of model refinement is a non-informative refinement where we do split cases into subcases, but do not assign new plausibilities of existing events contingent on the new subcases. Such refinements should never change the information obtainable from the model, and should not make it inconsistent. As pointed out in [6], a user interested in finite models is probably not impressed by an assumption implying that models are infinite. Although refinability assumes that an infinite number of new symbols are available, it is probaly acceptable to a user who realizes that models even if finite must be refinable when worked out, and such a user would probably accept the unbounded nature of refinability as an approximation to what is required by practicality considerations. We also want to have at our disposal the possibility of claiming that two new statements are independent in a given context, so that knowledge of one does not change the plausibility of the other. This condition we call information independence. It is closely related to statistical independence, but we do not use this term since we have not yet introduced probabilities. Finally, we introduce a strict monotonicity requirement, which is also introduced in most previous analyses without a lot of discussion. Our proposed assumptions, which should reflect one possible interpretation of Cox's and Jaynes's common sense desideratum, are:

- **Refinability:** *If we have already made a particular splitting of a statement into sub-cases, by adding new statements implying it, it should always be possible to refine another statement in the same way, and with the same plausibilities in the new refinement. As an example, if we defined A' with $A' \rightarrow A$ and $A'|A = a$, it should for any existing statement B be allowed to define B' as a new symbol with $B' \rightarrow B$ and $B'|B = a$.*
- **Information Independence:** *If a statement is refined by several new symbols, it should be possible to state that they are information independent, so that knowledge of one does not affect the plausibility of the other. As an example, if A and B are introduced as refinements of C, we should be permitted to claim that $A|BC = A|C$ and $B|AC = B|C$.*
- **Strict Monotonicity:** *The plausibility of a conjunction is always strictly less than those of the conjuncts, if these are independent and their plausibilities are not 0 or 1. The plausibility of a disjunction of exclusive statements with non-zero plausibilities is always strictly larger than those of its disjuncts. Moreover $F(x, y)$ and $G(x, y)$ are strictly increasing in both arguments for non-zero plausibility values of x and y.*

The main purpose of this note is to show that these assumptions entail Bayesianism for finite models.

3.1 Associativity, symmetry and joint distributivity

Halpern takes his example [6] as an indication that the conclusion of Cox does not apply to finite models. It is true that density and continuity cannot be dropped without introduction of some new assumption. However, if we accept refinability as a reasonable assumption, things change. If we have worked out a model where $F(a, F(b, c)) \neq F(F(a, b), c)$ for some plausibilities a, b and c, then we take an arbitrary statement S (not false) and refine with S_a, S_b and S_c: $S_a \rightarrow S_b$, $S_b \rightarrow S_c$ and $S_c \rightarrow S$, $S_a|S_b = a$, $S_b|S_c = b$ and $S_c|S = c$. Now the value $S_aS_bS_c|S$ can be computed in two ways giving different results, as $(S_aS_b)S_c|S = F(S_aS_b|SS_c, S_c|S) = F(F(S_a|S_b, S_b|S_c), c) = F(F(a, b), c)$ and as $S_a(S_bS_c)|S = F(S_a|SS_bS_c, S_bS_c|S) = F(S_a|S_b, F(S_b|S_c, S_c|S)) = F(a, F(b, c))$. So our model is not consistently non-informatively refinable if F violates associativity.

A similar argument can be used to show that F is bound by common sense to be symmetric: If we have defined some function value $F(a, b) = c$, it is reasonable that one should be able to refine any statement S into two information independent statements A and B, with $A|BS = A|S = a$ and $B|AS = B|S = b$ and thus $AB|S = BA|S = c$. If we allow this refinement operation, symmetry of F follows from commutativity of conjunction. The above is meant to suggest that it is quite reasonable to require that the function F is associative and symmetric for all values that appear in a given model. If the function G of [1] were used, the corresponding refinement principle would be that if $G(a, b) = c$ is assumed, then we should be allowed to take any statement S in the model and define a sub-case C with $C|S = c$, and two disjoint and exhaustive subcases A and B of C with plausibilities $A|S = a$ and $B|S = b$. With this refinement operation it is easy to show that G must be symmetric and associative, since it follows readily from commutativity and associativity of disjunction.

Finally we must consider the distributive law assumed in [1] and satisfied by $*$ and $+$, which we will soon require for showing that a scaling taking F to $*$ will also take G to $+$. For any three plausibilities a, b and c occurring in the model, and where $G(a, b)$ has been defined, we require that it should be possible to refine any statement S into three statements $A \rightarrow S$, $B \rightarrow S$ and $C \rightarrow S$ with $A|S = a$, $B|S = b$ and $C|S = c$, and such that AB is false and A, B and $A \vee B$ are all information independent of C. This requirement leads to the distributive law for F and G on the domain, since $(A \vee B) \wedge C|S = F(G(a, b), c)$ and $(A \vee B) \wedge C|S = (A \wedge C) \vee (B \wedge C)|S = G(F(a, c), F(b, c))$.

In summary:

Observation 1 *In order to satisfy natural requirements on consistency being preserved by non-informative refinements of models, we must work with models where F and G are partially specified in such a way that they satisfy the laws of associativity and symmetry, as well as joint distributivity.*

3.2 Equivalence of refinability and rescalability

We can now dismiss the example of [6] by observing that non-associativity inevitably violates the principle of consistent refinability. It remains to consider whether any partially specified function can be extended to an associative function if it is associative on its range of definition. This is not generally the case, even if it also satisfies the other properties that will be required from the completed function: strict monotonicity and symmetry. If an appropriate rescaling to probabilities exists, we can find it by solving a finite linear system of equations and inequalities for the log probabilities $l_i = \log w(x_i)$ excluding the value for falsity. The system has an equation $l_i + l_j = l_k$ for each triple $x_k = F(x_i, x_j)$ and an inequality $l_i < l_j$ for every pair with $x_i < x_j$, and an equality $l_i = l_j$ when $x_i = x_j$.

If a partially specified function can be completed to a full function over the support points (and some more points) satisfying associativity, symmetry and strict monotonicity, then the system is solvable. Its solution set is either unbounded (because the system is homogeneous) or empty, and it is empty only if there is no completion satisfying associativity, symmetry and strict monotonicity. A simple case where the partially specified function triples satisfy the laws, but no completion over the support points does so, is the following: Assume the partial specification satisfies

$$F(x_4, x_4) = a \tag{1}$$
$$F(x_3, x_5) = a \tag{2}$$
$$F(x_2, x_4) = b \tag{3}$$
$$F(x_1, x_5) = b \tag{4}$$
$$F(x_4, x_6) = c \tag{5}$$
$$F(x_3, x_7) = c \tag{6}$$
$$F(x_2, x_6) = d \tag{7}$$
$$F(x_1, x_8) = d \tag{8}$$

Here we have assumed that the x_i quantities are ordered increasingly in the open interval $(0, 1)$, but the quantities a, b, c and d can have any values. If the plausibilities were scalable to log probabilities l_i, there should be a solution to the system:

$$l_4 + l_4 = l_a \tag{9}$$
$$l_3 + l_5 = l_a \tag{10}$$
$$l_2 + l_4 = l_b \tag{11}$$
$$l_1 + l_5 = l_b \tag{12}$$
$$l_4 + l_6 = l_c \tag{13}$$
$$l_3 + l_7 = l_c \tag{14}$$
$$l_2 + l_6 = l_d \tag{15}$$
$$l_1 + l_8 = l_d, \tag{16}$$

together with the conditions $l_i < l_{i+1}$.

If we now add together the equations (9-16) multiplied with the coefficient sequence $(1, -1, -1, 1, -1, 1, 1, -1)$, we find after cancelling that they imply $l_7 = l_8$, contrary to the condition $l_7 < l_8$.

But if it were possible to complete the partially specified F so that it satisfies symmetry and associativity, we can reach the same conclusion by composing, with the function F, equations (1-8), after first swapping the equations with negative coefficient. The resulting equation is $F(F(x_4, x_4), F(a, F(F(x_1, \ldots)))) = F(F(x_3, x_5), F(a, F(F(x_2, \ldots))))$,
and thus by symmetry and associativity we can rearrange it to

$$F(x_7, F(a, F(b, \ldots))) = F(x_8, F(a, F(b, \ldots))), \tag{17}$$

where the omitted (dotted) parts of the left and right sides of (17) are equal. This entails, because of strict monotonicity and because no variable is zero, that $x_7 = x_8$, contrary to the assumption that $x_7 < x_8$. This also means that it is possible to add a finite set of statements by refinement with plausibilities that leads to inconsistency in the plausibility assignment. In this example we can add statements $\{A_i\}_{i=1}^7$, B_4 and C, with $A_i|C = x_i$ and $B_4|C = x_4$. If the A_i and B_4 are independent given C, the statement $A_1A_2A_3A_4B_4A_5A_6A_7|C$ can be shown to have two different plausibilities, $F(q, x_7)$ and $F(q, x_8)$ for $q = F(a, F(b, F(c, F(d, x_1, F(x_2, F(x_3, F(x_4, F(x_4, F(x_5, x_6) \cdots)$.

We are now ready to state that rescalability of the F function follows from our assumptions. The argument goes as follows: If rescalability obtains, it is trivial to extend F to an associative, symmetric and strictly monotone function over the dense interval $(0, 1)$ which covers any refinement. If rescalability does not hold, then this is equivalent to non-solvability of a linear program. But this means that a dual program has a solution (from which the coefficient vector in the example is obtained), and it so happens that this solution defines a refinement that is a proof of non-compliance of F with strict monotonicity, assuming it is symmetric and associative - but these properties have already been shown to follow from refinability. The concept of *extension base* is a formalization of the definedness of expressions such as those in (17) and is a set of multiplicities of its domain elements. If we have extended F to an extension base it means that all expression like those in (17) are defined, if the number of occurrences of each domain element in the expression is no greater than its multiplicity in the extension base. A more precise definition is given in the appendix.

Theorem 2 *Let $X = (x_i)_{i=1}^L$ be a strictly increasing sequence of real values. Let $S = \{1, \ldots, L\}$ and $T_\circ \subset S^3$ be a finite set of triples. The partial function $\circ$ satisfies $x_i \circ x_j = x_k$, for all $(i, j, k) \in T_\circ$. Then the following conditions (i) and (ii) are equivalent:*

(i) There is a finite extension base B of X to which $\circ$ cannot be extended as a symmetric, associative and strictly increasing function.

(ii) There is no increasing sequence of numbers $(f_i)_{i=1}^L$ such that if $(i, j, k) \in T_\circ$, then $f_i + f_j = f_k$.

This theorem (proved in the appendix) applies both to the function F and the function G of a consistently refinable plausibility model, since these functions are both, by the preceding analysis, associative, symmetric and strictly increasing. For the function F, the numbers f_i must be negative, since we assumed $F(x, y) < \min(x, y)$ and we have an equation $f_L + f_i < f_j$, $f_j < f_i < f_L$, in our system. The f_i can thus be taken as log probabilities. For the function G, the f_i must for analog reasons be positive. They can be taken as probabilities after some normalizing linear scaling. There are thus a number of rescalings transforming the function F to $*$ and a number of rescalings transforming G to $+$, but we do not yet know if there is one rescaling satisfying both criteria.

What can we say about the joint effect of rescaling on the functions F and G? The analysis of [1] takes the joint distributivity of F and G as a starting point, and then we can draw on results for the extremely well studied Cauchy equation. We outline an argument, which is very much based on the classical analysis of Cauchy's equation $f(x + y) = f(x) + f(y)$[1], the difference being that we do not have a dense domain, but instead we have a refinability principle that allows us to construct a finite inconsistency proof for any given non-linear solution. If we scale G to $+$, the distributivity equation is transformed to a family of Cauchy equations, $F(x + y, z) = F(x, z) + F(y, z)$. We claim that the general solution has the form

$F(x,z) = xc(z)$ for some monotone function c. Indeed, the refinability principle for complement, given that G is $+$, lets us define two cases of a statement S with $S'|S = a$ and $S''|S = 1-a$. By repeated refinement, an arbitrarily fine grid of statements in the model, related by addition, can be created. Thus, if $F(x_0, z_0) = sx_0$, $F(x_0, z_0) = sx_0 = F(F(x_0, a), z_0) + F(F(x_0, 1-a), z_0)$. We can thus by repeating refinement find points $x_0^{(i)}$ arbitrarily close to 0 with both $F(x_0^{(i)}, z_0) \leq sx_0^{(i)}$ and $F(x_0^{(i)}, z_0) \geq sx_0^{(i)}$. If we had $F(x_0, z_0) = sx_0$ and $F(x_1, z_0) = s'x_1$, with $s > s'$, we can find a refinement with two plausibilities a and a' such that $a < a'$ but $F(a, z_0) \geq sa > s'a' \geq F(a', z_0)$, a violation of strict monotonicity of F. So $F(x,z)$ must vary linearly with x for constant z, $F(x,z) = xc(z)$. But since $F(1,z) = z$ we must have $c(z) = z$ on the domain, *i.e.*, $F(x,y) = xy$. We have outlined an argument for:

Theorem 3 *Let $X = (x_i)_{i=1}^L$ be an increasing sequence of distinct values in the open interval $(0,1)$, and $S = \{1, \ldots, L\}$. Given two sets of triples $T_F, T_G \subset S^3$ interpreted as specifications of two partial functions F and G satisfying also $F(1, x_i) = x_i$, $F(0, x_i) = 0$ and $G(0, x_i) = x_i$.*

The following are equivalent:

- *(i) There is a finite extension base B of X to which F and G cannot be jointly extended as symmetric, associative and strictly increasing functions satisfying joint distributivity.*
- *(ii) There is no increasing sequence of real numbers $(p_i)_{i=1}^L$ such that if $(i,j,k) \in T_F$, then $p_i * p_j = p_k$, and if $(i,j,k) \in T_G$, then $p_i + p_j = p_k$.*

The above means that we can assign plausibilities in two different ways: either we choose functions F and G that have the required properties and use them for assigning plausibilities of conjunctions and disjunctions, or else we assign plausibilities on the fly but check always (by solving linear and non-linear constraint problems) that no newly defined triple (arguments and function value) violates the required properties. In both cases it would be better to work with probabilities.

We finally note in which way our argument is new. We have completely dropped the continuity requirements used previously. We have also removed density assumptions. However, refinements add to the domain because of strict monotonicity. There is no bound on how far we may have to refine a non-rescalable model in order to show its deficiency, but for every non-rescalable model there is a finite argument showing its deficiency. One could of course still wonder whether refinability is strictly weaker than those assumptions used in earlier work like [1, 12]. These assumptions were characterized in [6] as assuming a dense domain. Our assumptions are really weaker, for the somewhat surprising reason that our assumptions until now do not suffice for the infinite case, as we will show in the next section.

3.3 Infinite models

The replacement of denseness, continuity and associativity assumptions by refinability entails Bayesianism in finite domains. Now it remains to consider non-finite domains. There seem to be no principled reason that Theorem 2 should not work in infinite domains. However, there is a problem in how we interpret strict monotonicity, and in particular we do not think that finite refinability is sufficient for infinite domains, as shown by the following consideration: in a probability model, if $x < y$ then the union of the intervals $[x^i, y^i]$ is a finite set of disjoint intervals, since the intervals will overlap for large i. But the number of intervals is invariant under strictly monotone rescaling. So a model where the union of such intervals (exponent now denoting iteration of F, so that $x^1 = x$ and $x^{n+1} = F(x, x^n)$) is an infinite set of disjoint intervals cannot be rescalable.

As an example with an infinite number of intervals thus not being rescalable, consider a domain generated from two statements with plausibilities $y = 1/4$ and $x = 1/5$. Let exponents of plausibilities denote iteration of the F function. The model is defined by: $F(y^j, x^k) = 1/(3(j+k) + (j+2k)/(j+k))$. Now $x^p = 1/(3*p+2)$, $y^p = 1/(3*p+1)$, and separation is not obtained, because no y^{p+1} is larger than x^p for any positive integer p, and therefore all intervals are disjoint. We have not at all used the function G, so we have no means to even talk about values of G in refinements. An interesting observation on this model is that each of its finite subsets is rescalable, and of course the function F is associative, symmetric and strictly monotone. There appears to be no finite argumentation for its inadequacy, at least not using reasonable refinability arguments. In a forthcoming note we will show that our interpretation of Cox assumptions for infinite models entails rescalability to infinitesimal or extended probability[2, 15], where probabilities take values in an ordered algebraic field of reals and infinitesimals. The example above would correspond to extended probabilities a for x and $a + \epsilon$ for y, where a is some probability value and ϵ is an infinitesimal.

4 Conclusions

Bayesianism can be motivated by successful application and by several different and more or less convincing arguments[8, 12, 9]. We proposed to weaken the common sense assumptions used previously from domain denseness and continuity of F to refinability and allowing information independence, and showed such assumptions sufficient for finite models. That our proposal uses truly weaker assumptions is shown by its inadequacy for the infinite case. The requirement of strict monotonicity is suggested by common sense and proposed in most previous justifications, although it is in no way inevitable.

One could note that our discussion is rather neutral as to how unavoidable our assumptions are in real applications, and thus we cannot claim inevitability of Bayesianism. An overwhelming number of claims for and against Bayesianism can be found in the literature and it is clearly beyond the scope of any reasonable length paper to sort out these claims. Many are criticized, e.g., in [8, 11, 9]. Our refinability and information independence principles, which are not completely new [7], but not until now shown to be central for uncertainty management foundations, appear to be somewhat more compelling than strict monotonicity and insistence on a single real number for plausibility, although maybe less compelling than consistency and (non-strict) monotonicity. We conjecture that arguments similar to those above can be used to relax the continuity assumption made by Lindley[11].

REFERENCES

[1] J. Aczél, *Lectures on Functional Equations and their Applications*, Academic Press, 1966.

[2] E. Adams, 'Probability and the logic of conditionals', in *Aspects of Inductive Logic*, eds., J. Hintikka and P. Suppes, pp. 265–316. North Holland, Amsterdam, (1966).

[3] R.T. Cox, 'Probability, frequency, and reasonable expectation', *Am. Jour. Phys.*, **14**, 1–13, (1946).

[4] A. Dax, 'An elementary proof of Farkas' lemma', *SIAM Rev.*, **39**, 503–507, (1997).

[5] B. de Finetti, *Theory of Probability*, London:Wiley, 1974.

[6] J. Halpern, 'A counterexample to theorems of Cox and Fine', *Journal of AI research*, **10**, 67–85, (1999).

[7] D. Heckerman and H. Jimison, 'A Bayesian perspective on confidence', in *Uncertainty in Artificial Intelligence*, eds., L. N. Kanal, T.S. Levitt, and J.F. Lemmer, pp. 149–160. Elsevier Science Publishers, (1989).

[8] C. Howson and P. Urbach, *Scientific Inference: The Bayesian Approach*, Open Court Publishing Company, 1989.

[9] E. T. Jaynes, *Probability Theory: The Logic of Science*, Preprint: Washington University, 1996.

[10] H. W. Kuhn, 'Dual systems of homogeneous linear relations', in *Linear Inequalities and Related Systems*, eds., H. W. Kuhn and A. W. Tucker, pp. 3–18. Princeton University Press, (1956).

[11] D. V. Lindley, 'Scoring rules and the inevitability of probability (with discussion)', *Internat. Stat. Rev*, **50**, 1–26, (1982).

[12] J.B. Paris, *The Uncertain Reasoner's Companion*, Cambridge University Press, 1994.

[13] Paul Snow, 'On the correctness and reasonableness of Cox's theorem for finite domains', *Computational Intelligence*, **14**, 452–459, (1998).

[14] M. Tribus, 'Bayes's equation and rational inference', in *Rational Descriptions, Decisions and Designs*, pp. 73–106. Pergamon Press, (1969).

[15] N. Wilson, 'A logic of extended probability', in *Proceedings of the First International Symposium on Imprecise Probabilities and their Applications*, eds., Gert de Cooman, Fabio G. Cozman, Serafin Moral, and Peter Walley, pp. 397–404. Gent University, (1999).

A Rescalability Theorem

First some definitions: We use the infix operator notation $x \circ y$ for $F(x, y)$, which is convenient for associative and symmetric functions. An *extension base* B of a sequence X of length L is a sequence (n_i) of length L of non-negative integers. A partial function that is associative and symmetric on X^2, where $X = (x_i)$, can be *extended to extension base* B if it can be extended to an associative and symmetric function on a domain such that the expression $v_1 \circ v_2 \circ \cdots \circ v_n$ and all its subexpressions have values if every v_i is equal to some x_j, and for all i, the number of occurrences of x_i is not larger than the corresponding number n_i in B.

The following is a result in duality theory of linear programming, [10, Corollary 1A, case (i)] (we could also have used the slightly less suitable but better known Farkas's Lemma [4] - there appears to be no really simple proof of either):

Lemma 4 (Kuhn) *The system of equations $Ax = 0$ has a positive solution $x > 0$ if there is no u such that $A^T u \geq 0$ and $u \neq 0$.*

We can now prove a lemma that is obvious for dimension two and three, but not in general:

Lemma 5 *Let F be a linear subspace of R^n. The following conditions are equivalent:*

(i) There is no element in F with all components positive.

(ii) There is a nonzero vector d with non-negative components that is orthogonal to F.

Proof.

(ii)→(i): This direction is obvious, since a vector orthogonal to a non-zero and non-negative one cannot have all components positive.
(i)→(ii): Assume (i): There is no element in F of R^n with all components positive. Let F be the space spanned by the rows of matrix B, $F = \{B^T y : y \in R^k\}$. Let the rows of A be a base for the orthogonal co-space of F, $AB^T = 0$. Thus, $F = \{x : Ax = 0\}$ and $Ax = 0$ has no positive solution x by our assumption that (i) is the case. Since $Ax = 0$ has no positive solution, by Lemma 4 there is a u such that $A^T u \geq 0$ and $u \neq 0$. Now $u^T A$ is a non-negative vector, and it is orthogonal to every vector in F because $AB^T = 0$ and thus $(u^T A)(B^T x) = 0$ for all $x \in R^k$. So (ii) applies, *i.e.*, (i)→(ii).

Conditions (i) and (ii) are thus equivalent.

□

We can now prove Theorem 2:

Proof.(of Theorem 2)
(i)→(ii) : If (ii) is not the case, there exist appropriate f_i. Define $l(x)$ by interpolation to a strictly increasing function between the constraints $l(x_i) = f_i$. The function $x \circ y = l^{-1}(l(x) + l(y))$ is associative, symmetric and strictly increasing. So also (i) is not the case, which shows (i)→(ii).
(ii)→(i): Assume (ii) is the case. Define the $|T_\circ|$ by L matrix M to have one row for each tuple in $T_\circ$. For such a tuple (i, j, k), the row has the value 1 in columns i and j, the value -1 in column k, and zero otherwise. Matrix D is $L - 1$ by L and has value $D = I' - I''$ where I' and I'' is the L by L unit matrix with the first and last row, respectively, deleted. From now on we regard f as a sequence of variables f_i. Since (ii) is the case, there is no L-vector solution f to $Mf = 0$ that also satisfies $Df > 0$, since such a solution would contradict non-existence of the f_i.

The solution space F of $Mf = 0$ is such that the linear subspace DF is orthogonal to some non-zero vector d with non-negative components, by Lemma 5. In other words, a linear equation $d^T Df = 0$ for f can be derived from $Mf = 0$ only, *i.e.*, the null space $\{f : Mf = 0\}$ of M is included in the null space $\{f : d^T Df = 0\}$ of $d^T D$, and $d^T D = c^T M$ for some vector c. Since M and D have integer elements, and the condition is homogeneous in d, we can assume that d consists of natural numbers and c of integers. Thus, a linear equality $d^T Df = 0$ for f can be obtained as a linear combination with integer coefficients of the linear equalities given by the rows of the system $Mf = 0$. But each row r of M is derived from a constraint $x_k = x_i \circ x_j$ for the function $\circ$. By composing these constraints with the associative and commutative operator $\circ$ in the pattern indicated by c we can derive a functional constraint on $x \circ y$, and at last obtain a functional constraint corresponding to the linear constraint $d^T Df = 0$. We compose the constraints coded by a triple of $T_\circ$ a number of times given by the magnitude of the corresponding coefficient c_i of the linear combination, reversing the equation if the coefficient is negative. In this way we derive a functional constraint:

$$a_1 \circ a_2 \circ \cdots \circ a_m = b_1 \circ b_2 \circ \cdots \circ b_n. \tag{18}$$

The corresponding linear constraint $d^T Df = 0$ can be written as

$$d_1 f_1 + d_2 f_2 + \cdots d_{L-1} f_{L-1} = d_1 f_2 + d_2 f_3 + \cdots d_{L-1} f_L, \tag{19}$$

where no d_i is negative and at least one is positive. But (19) results from the linear form of (18) by cancelling certain elements in both sides. Thus, $n = m$ and either $a_i = b_i$ (for quantities cancelling in the linear combination) or $a_i < b_i$ (for quantities remaining in (19), with at least one strict inequality since at least one d_i is non-zero.

But then, from strict monotonicity, we must also have: $a_1 \circ a_2 \circ \cdots \circ a_m < b_1 \circ b_2 \circ \cdots \circ b_n$.

There can thus not be a strictly increasing extension of $\circ$ to an extension base defined by the union of the (a_i) and (b_i) sequences, in other words (i) is the case.

So (i) and (ii) are equivalent. □