Knuth K.H. 2007. Lattice theory, measures and probability. K.H. Knuth, A. Caticha, J.L. Center, A. Giffin, C.C. Rodriguez (eds.),
Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 2007,
AIP Conference Proceedings 954, American Institute of Physics, Melville NY, 23-36.

Lattice Theory, Measures and Probability

Kevin H. Knuth

Department of Physics
Department of Informatics
University at Albany (SUNY)
Albany NY, 12222, USA
(e-mail: kknuth@albany.edu, http://knuthlab.rit.albany.edu)

Abstract. In this tutorial, I will discuss the concepts behind generalizing ordering to measuring and apply these ideas to the derivation of probability theory. The fundamental concept is that anything that can be ordered can be measured. Since we are in the business of making statements about the world around us, we focus on ordering logical statements according to implication. This results in a Boolean lattice, which is related to the fact that the corresponding logical operations form a Boolean algebra.

The concept of logical implication can be generalized to degrees of implication by generalizing the zeta function of the lattice. The rules of probability theory arise naturally as a set of constraint equations. Through this construction we are able to neatly connect the concepts of order, structure, algebra, and calculus. The meaning of probability is inherited from the meaning of the ordering relation, implication, rather than being imposed in an ad hoc manner at the start.

Keywords: lattice, order, poset, measure, probability
PACS: 01.70.+w, 02.10.-v, 02.50.-r

INTRODUCTION

We begin by considering one of the most basic of concepts: *ordering*. Two objects are ordered by simply comparing them to one another. This basis of comparison relies on some sort of rule that decides which item is 'bigger'. This rule for comparing two objects is called a *binary ordering relation*. We can then extend this activity to a set of objects by considering all pairwise comparisons. These pairwise comparisons will result in a *partially-ordered set*. We say that the set is partially-ordered since it may be the case that two objects may be incomparable according to the binary ordering relation we have chosen. Whatever the case may be, given a set or elements and a binary ordering relation, we can generate a partially-ordered set, or *poset* for short. This abstraction is important because the concept of ordering is fundamental. In hindsight, we will not be surprised at the deep connections that we will uncover here.

Measuring is a more advanced concept as it takes ordering one step further. Rather than just saying that one element is bigger than another, measuring specifies how much bigger. Measuring is based on ordering, and we shall see that the rules for manipulating measures originate as constraints imposed by the partial order. Another way to think about it is that the partial order dictates the algebra and the measure dictates the calculus.

At this point we have kept this discussion purposefully abstract. The reason is to highlight the fundamental nature of the concepts. In this paper, we will use ordering relations to derive probability theory. This methodology is far more fundamental than

that introduced by Richard Cox [1, 2], who was the original inspiration for this work. We will leave no room for arguing about poorly-defined concepts such as belief. Instead, we will work with the set of logical statements and apply the straightforward binary ordering relation known as implication.

What we will discover is applicable to measures on *all* partially-ordered sets isomorphic to our resulting poset of logical statements ordered by implication. We will also see relationships shared by other partially-ordered sets, and find that the commonalities are due to the underlying ordering relations.

MATHEMATICAL PRELIMINARIES

My intention is to keep this tutorial as basic as possible without delving deep into the mathematics. However, some introduction to order theory is necessary to properly visualize these partially-ordered sets. I encourage the reader to work through this section as the results are extremely fruitful. Remember, we are merely formalizing the extremely familiar notion of ordering.

Consider a set of elements S, and a binary ordering relation denoted $\leq$. Note that $\leq$ may mean the usual 'less than or equal to', but here we use the symbol to represent any binary ordering relation. We say that z *includes* x if $x \leq z$. If it is true that x is not equal to z, then we can write $x < z$. Last, if $x < z$, and there exists no element y in the set S such that $x < y$ and $y < z$ then we say that z *covers* x, and write $x \prec z$. In this case we can view z as succeeding x in a hierarchy. We will use this concept of covering to construct diagrams of partially-ordered sets (Figure 1). It is important to stress that inclusion, as described above, is another way of looking at ordering.

Once we have a partially-ordered set, we can consider an element of the set and note where it lies in the ordering. In this sense, the perspective of a partially-ordered set leads to the concept of structure. The set of elements that include a given element x is called the upper bound of x. The set of elements included by x is called the lower bound of x. We can also talk about upper and lower bounds of a subset of S, so that the set of all elements of S that include a subset U of S is the upper bound of U. Likewise for the lower bound.

When it happens that both the upper and lower bound exist for all pairs of elements in the partially-ordered set, then we say that the set is a lattice. For two elements x and y in the lattice $\mathcal{S}$, composed of the set S and a binary-ordering relation $\leq$, one can denote the upper bound of x and y as $x \vee y$ and the lower bound of x and y as $x \wedge y$. The symbols $\vee$ and $\wedge$ can be seen to act as operators that take a pair of elements and map them to a third element. These operators are called the *join* and *meet*, respectively. This is the algebraic perspective. Thus a lattice can be viewed as either a structure or an algebra.

Some lattices have special elements called the top $\top$ and bottom $\bot$. The top is the unique greatest upper bound, whereas the bottom is the unique least lower bound. Note that not all posets have tops and bottoms. In Figure 1 neither poset (a) nor (b) has a top element.

All lattices share similar structural features that can be expressed as algebraic rules. We will find later that these rules impose constraints on the measures that we assign, and appear as constraint equations that are found universally throughout the sciences.

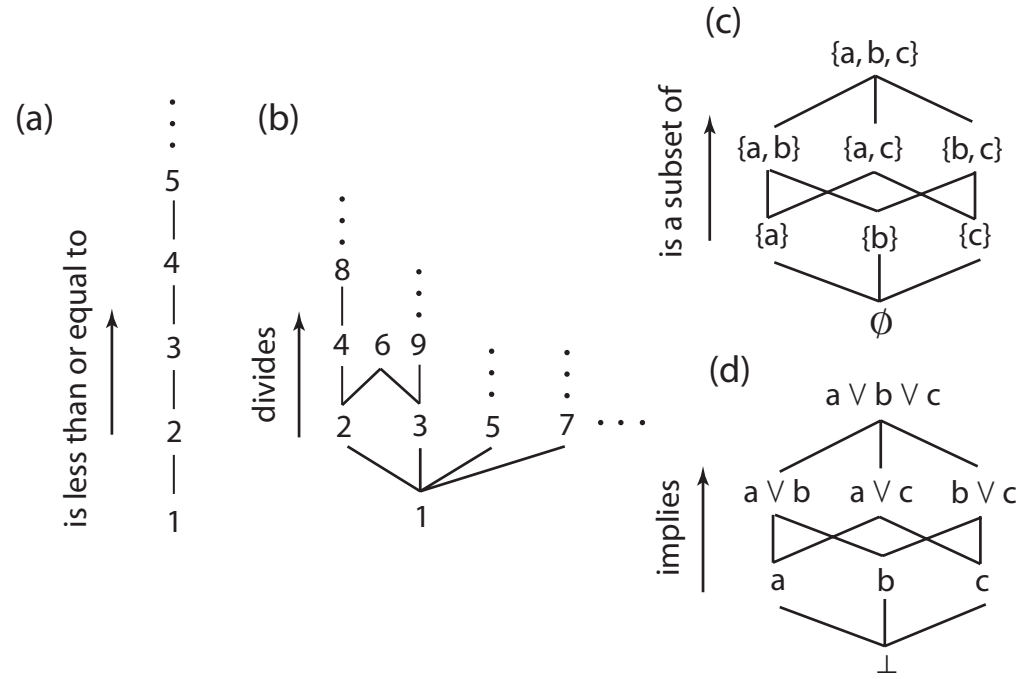

FIGURE 1. Four examples of posets. (a) The poset of integers ordered according to the usual '*less than or equal to*'. (b) The poset of integers ordered according to '*divides*'. (c) The poset of of subsets of $\{a, b, c\}$ ordered according to '*is a subset of*'. (d) The poset of logical statements ordered according to '*implies*'. Note that even though posets (a) and (b) both involve integers, the ordering relation results in distinct structures. Conversely, posets (c) and (d) are isomorphic despite the fact that both the sets and ordering relations are different.

We will find that some elements cannot be expressed as the join of any two other elements in the poset. Such elements are called *join-irreducible*. As an example, the elements covering the bottom in all four posets of Figure 1 are join-irreducible. In the lattice of logical statements, these are the exhaustive set of mutually-exclusive statements. In the set of integers ordered by '*divides*', these are the prime numbers. Already we are seeing strange connections. The poset of integers ordered by '*less than or equal to*' is composed entirely of join-irreducible elements. Incidentally, this poset is called a *chain*.

By formalizing the notion of ordering, we will see that there are general features that are shared by a wide array of partially ordered sets. These general features are the prize that awaits those who are willing to back away from the specifics and to generalize the operations involved in ordering. In this sense we are unifying all things ordered.

MEASURING

By focusing on the concepts underlying the process of ordering, we can now see measuring in a new light. Measuring is a generalization of ordering. Since ordering is represented by inclusion, one way to generalize ordering is to generalize the binary concept

TABLE 1. Features of Lattices

ORDERING	
Order	$\leq$
Reflexivity	For all A, $A \leq A$
Antisymmetry	If $A \leq B$ and $B \leq A$ then $A = B$
Transitivity	If $A \leq B$ and $B \leq C$ then $A \leq C$
OPERATIONS	
Join	$\vee$
Meet	$\wedge$
Idempotency	$A \vee A = A$
	$A \wedge A = A$
Commutativity	$A \vee B = B \vee A$
	$A \wedge B = B \wedge A$
Associativity	$A \vee (B \vee C) = (A \vee B) \vee C$
	$A \wedge (B \wedge C) = (A \wedge B) \wedge C$
Absorption	$A \vee (A \wedge B) = A \wedge (A \vee B) = A$
Distributivity	$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$
	$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$
CONSISTENCY	
$A \leq B \quad \Leftrightarrow \quad A \wedge B = A \quad \Leftrightarrow \quad A \vee B = B$	

of inclusion to *degrees of inclusion*. Since we are required to preserve *transitivity*, we can use Real Numbers to quantify degrees of inclusion. There are other mathematical quantities that we could use as well, but for now we will focus on real numbers.

Inclusion can be encoded in terms of the zeta function, which quantifies whether the lattice element y includes the lattice element x, that is

$$\zeta(x,y) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{if } x \not\leq y. \end{cases} \qquad (\textit{zeta function}) \tag{1}$$

Here I define the dual of the zeta function $\zeta^{\partial}(x,y)$, which quantifies whether x includes y, that is

$$\zeta^{\partial}(x,y) = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{if } x \not\geq y. \end{cases} \qquad (\textit{dual of the zeta function}) \tag{2}$$

Note that if x does not include y, we group the ordered pair into an equivalence class where the answer is zero.

We now introduce a generalization of the function above where

$$z(x,y) = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{if } x \wedge y = \perp \\ z & \text{otherwise, where } 0 < z < 1. \end{cases} \qquad (\textit{degrees of inclusion}) \tag{3}$$

The motivation here is that if x does not include y, we would like to consider the degree to which x includes y. It is here that the generalization takes place. It should be noted that

TABLE 2. The function $z(x,y)$ in equation (3). The x-values are aligned in the rows, and the y-values are in the columns. Note that this is a generalization of the dual of the zeta function (2) where a subset of the zero values are now unknown values between zero and one. These are indicated by the question marks. This function describes the degree to which y implies x. Clearly, we need more information to assign these values.

	$\bot$	a	b	c	$a\vee b$	$a\vee c$	$b\vee c$	$\top$
$\bot$	1	0	0	0	0	0	0	0
a	1	1	0	0	?	?	0	?
b	1	0	1	0	?	0	?	?
c	1	0	0	1	0	?	?	?
$a\vee b$	1	1	1	0	1	?	?	?
$c\vee a$	1	1	0	1	?	1	?	?
$b\vee c$	1	0	1	1	?	?	1	?
$\top$	1	1	1	1	1	1	1	1

this generalization is not a measure per se, which takes one argument to a real number, but rather a bi-valuation, which takes two lattice elements to a real number. The notation we are using here $z(x,y)$ emphasizes this fact, and we shall be careful not to lose sight of this as we adopt other notations specific to certain lattices.

Now it is important to note that this generalization does not tell us which particular values our measure should take for each pair of lattice elements. Some of the values of this bi-valuation are inherited directly from the zeta function. These are the situations when $x \geq y$, so that $z(x,y) = \zeta^{\partial}(x,y) = 1$, or when $x \wedge y = \bot$, so that $z(x,y) = \zeta^{\partial}(x,y) = 0$. However, we now need to determine the values of the function $z(x,y)$ when $x \ngeq y$ and $x \wedge y \neq \bot$. This is similar to the idea of analytic continuation when extending real-valued functions to complex-valued functions. To accomplish this, we need another principle.

CONSISTENCY

If our bi-valuation is to represent the lattice structure, the value we assign to any given pair of lattice elements must not be independent of the assignments we make to all the other pairs. If the assignments were independent of the lattice structure, then the measure cannot possibly represent the ordering. Dependence is necessary.

To derive these dependencies among the bi-valuation assignments, we turn to the lattice algebra. For element, $a \vee b$, one would expect that $z(a \vee b, c)$ must somehow depend on $z(a,c)$ and $z(b,c)$. Our goal is to determine the functional relationship by considering the lattice structure. Skilling's Principle of Generality [3] suggests that if a function is to be generally applicable, then it must work in special cases. If such a function exists, and we can constrain its form with a sufficient number of special cases, then we will be able to determine the functional form by relying on these constraints.

The most effective way to impose a special case is to use the algebra to write an expression two different ways. Consistency requires that two distinct expressions of the

same quantity must give the same result when used as an argument in the bi-valuation. If this were not the case, then we would have two different measures assigned to the same pair of lattice elements, and we would not have a function. Let us see where this leads.

Constraints Imposed by the Lattice

Associativity of the Join

We first consider associativity of the join operation. The idea behind this derivation is that we consider an expression such as $u \vee (v \vee w)$ and use associativity to write this expression a second way $(u \vee v) \vee w$. Since these two expressions represent the same element, the degree to which $u \vee (v \vee w)$ includes a fourth statement t must equal the degree to which $(u \vee v) \vee w$ includes t.

We now consider the special case where $u \wedge v = v \wedge w = u \wedge w = \perp$. This is helpful since $z(x, \perp) = 0$ for all elements x, and we can neglect any dependence on the various meets, such as $u \wedge v$. If the assignments of the bi-valuation are to be consistent with one another and if $u \wedge v = \perp$ then $z(u \vee v, t)$ must be a related to $z(u,t)$ and $z(v,t)$. We express this idea by introducing a function $S(\cdot,\cdot)$ such that

$$z(u \vee v, t) = S[z(u,t), z(v,t)]. \tag{4}$$

The next step is to rewrite the two bi-valuations $z(u \vee (v \vee w), t)$ and $z((u \vee v) \vee w), t)$ in terms of the function $S[\cdot,\cdot]$

$$z(u \vee (v \vee w), t) = z((u \vee v) \vee w, t), \tag{5}$$

which results in

$$S[z(u,t), z((v \vee w), t)] = S[z((u \vee v), t), z(w,t)]. \tag{6}$$

Using the relation (4) again, we find that

$$S[z(u,t), S[z(v,t), z(w,t)]] = S[S[z(u,t), z(v,t)], t), z(w,t)]. \tag{7}$$

Substituting $a = z(u,t)$, $b = z(v,t)$ and $c = z(w,t)$ we obtain

$$S[a, S[b,c]] = S[S[a,b], c], \tag{8}$$

which is a functional equation known as the *associativity equation*. The general solution [4] is

$$S(a,b) = f(f^{-1}(a) + f^{-1}(b)), \tag{9}$$

where f is an arbitrary function. This is simplified by letting $g = f^{-1}$

$$g(S(a,b)) = g(a) + g(b). \tag{10}$$

Writing this in terms of the original expressions we find that,

$$g(z(u \vee v, t)) = g(z(u,t)) + g(z(v,t)), \tag{11}$$

which reveals that there exists a function $g : \mathbb{R} \to \mathbb{R}$ that re-maps these numbers to a more convenient representation [5, 6]. Defining $p(u,t) \equiv g(z(u,t))$ we get

$$p(u \vee v, t) = p(u,t) + p(v,t), \tag{12}$$

which is the sum rule for the join of two atomic elements.

In general, for any pair of lattice elements x and y one can show [6] that the constraint equation becomes the familiar *sum rule*

$$p(u \vee v, t) = p(u,t) + p(v,t) - p(u \wedge v, t). \tag{13}$$

This 'rule' is a constraint equation that ensures that associativity of the join is preserved by the bi-valuation assignments across the lattice.

This result is ubiquitous throughout mathematics and the sciences; so much so, that the relations often appear obvious [7]. For instance, the sum rule appears in many contexts, including probability theory, measure theory, geometry, quantum mechanics, information theory, etc. Here are some familiar examples:

$$p(x \vee y|\top) = p(x|\top) + p(y|\top) - p(x \wedge y|\top) \tag{14}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{15}$$

$$max(A,B) = A + B - min(A,B) \tag{16}$$

where the fact that these functions are bi-valuations has been suppressed by the standard notation. In the last example, A and B are Real numbers. Keep in mind that this relation is a constraint equation that enforces associativity. The constraint of associativity appears everywhere, once you know to look for it.

Associativity of Order and Distributivity

Given a chain of lattice elements $u \leq v \leq w \leq t$, we can consider the degree to which u includes t, $z(u,t)$. For the lattice of logical statements ordered by implication this is the degree to which t implies u. Since it is clear that t includes v, consistency requires that $z(u,t)$ be related to the degree to which u includes v, $z(u,v)$, and the degree to which v includes t, $z(v,t)$. We express this relationship as an unknown function $P(\cdot,\cdot)$

$$z(u,t) = P[z(u,v), z(v,t)]. \tag{17}$$

Using the four elements in the chain, we can use associativity to write this two ways

$$z(u,t) = P[z(u,v), z(v,t)] = P[z(u,w), z(w,t)]. \tag{18}$$

We can now write $z(v,t)$ and $z(u,w)$ using the same function. This results in

$$P[z(u,v), P[z(v,w), z(w,z)]] = P[P[z(u,v), z(v,w)], z(w,t)]. \tag{19}$$

Substituting $a = z(u,v)$, $b = z(v,w)$ and $c = z(w,z)$ we again obtain the associativity equation

$$P[a,P[b,c]] = P[P[a,b],c]. \tag{20}$$

The solution is of the form given in (9). The fact that there is an arbitrary function involved means that we can obtain a solution to this functional equation, which is commensurate with our previous sum rule. By exponentiating, we obtain the *product rule*

$$p(u,t) = p(u,v)p(v,t). \tag{21}$$

This may look a bit more familiar if we use the fact that $u \leq v$ to write $u = u \wedge v$, and the fact that $v \leq t$ to write $v = v \wedge t$.

$$p(u \wedge v,t) = p(u,v \wedge t)p(v,t). \tag{22}$$

This constraint equation was derived in the special case of a chain. If a constraint equation is to hold in general, it must hold in special cases. It is straightforward to show that this form does indeed hold in general when the lattice elements do not lie in a chain. This can be done by considering the constraint of distributivity of $\wedge$ over $\vee$ [5, 6].

Commutativity

The fact that $u \wedge v = v \wedge u$ allows us to write the product rule two ways. This gives us our final constraint equation, which is known as *Bayes' Theorem*

$$p(v,u)p(u,t) = p(u,v)p(v,t). \tag{23}$$

In the case of our chain, the first term $p(v,u) = 1$ and this reduces to our original equation (21) as a triviality. The product rule and Bayes' Theorem are much more interesting when the elements do not lie in a chain

$$p(v,u \wedge t)p(u,t) = p(u,v \wedge t)p(v,t). \tag{24}$$

Bayes' Theorem is so easily derived that one might be tempted to dismiss the constraint of commutativity as being a triviality. However, one can imagine ordered structures that are not commutative, and hence not lattices. In such structures associativity would give the usual sum and product rules, but the lack of commutativity would deny us a Bayes' Theorem. This is suggestive of the structure of quantum mechanics.

PROBABILITY THEORY

The derivations of the previous sections are applicable to distributive lattices in general. One such lattice is the Boolean lattice of logical statements. The bi-valuation $p(x,t)$, which describes the degree to which the statement x includes the statement t, or equivalently the degree to which t implies x. The *meaning* of the function $p(\cdot,\cdot)$ is inherited from the *meaning* of the ordering relation. Meaning cannot be assigned arbitrarily. For

this reason, we will not talk about degrees of belief. Nor shall we require the notion of truth. The game we are playing involves only computations regarding the degree to which one statement implies another.

To highlight how these results relate to probability theory, we simply change notation

$$p(x|y) \doteq p(x,y). \tag{25}$$

Note that $p(x|y)$ is a bi-valuation that takes two logical statements as arguments. I will call this function *probability*, since it performs exactly the way we have come to expect probabilities to perform.

These logical statements are not some ill-defined prior information I that Jaynes introduces. Instead they are logical statements that belong to the lattice. While this function is traditionally read as 'the probability of x given y', we read 'the degree to which y implies x'. The prior information I that acts to externally constrain the values of the bi-valuation is not an argument of this function. Instead, this prior information I acts as an additional constraint. This may be an excellent place to adopt a new notation introduced by Skilling to represent the external information that goes into constraining the probability assignments

$$p(x|y) \qquad\qquad || \quad I. \tag{26}$$

This careful analysis highlights the fact that probability is a bi-valuation, and not a measure in the usual sense. Expressions like $p(x)$ have no meaning in this context. Of course one is free to define a new function that takes one argument by fixing the value of the second argument, such as

$$p_t(x) \doteq p(x|t). \tag{27}$$

This is usually done for $t = \top$ and written by overloading the function p

$$p(x) \doteq p(x|\top). \tag{28}$$

This notation is unfortunate as it obscures the fact that this function is a bi-valuation.

Product Spaces

We rarely work in a single space. Instead we form product spaces, and it is essential to look at this explicitly to clear up some confusion that has persisted. We begin with the *Model Space*, $\mathcal{M}$, (also called the *state space*) which consists of a set of atomic statements that describe our physical system. If we are describing fruit, the atomic statements might be:

$$\begin{aligned} `a &= \text{It is an apple!'} \\ `b &= \text{It is a banana!'} \\ `c &= \text{It is a citrus fruit!'}. \end{aligned}$$

One may choose other statements, perhaps some that refine the notion of a citrus fruit to oranges, grapefruit, lemons, limes and kumquats. That is your model; you construct

the space! The only requirement is that the atoms be mutually exclusive and exhaustive. If they are not, then the 'garbage in, garbage out' rule applies.

A lattice can be formed from the atomic statements by considering statements that consist of all possible disjunctions of the atomic statements. The resulting space is known as the *hypothesis space*, $\mathcal{H}$. The hypothesis space entertains statements such as 'It is an apple or a banana!' Clearly the fruit can never be in such a state. However an observer can make such a statement and thus convey the information that he or she possesses about the system. All of our calculations are performed in the hypothesis space. In this sense our results do not describe the universe so much as they describe the predictive power of one who might make statements about the universe. Since statistical mechanics is based on inferences in the hypothesis space, statistical mechanics is a theory describing the statements one might make about a thermodynamic system rather than being a theory about the thermodynamic system itself. It is an important, but subtle distinction. One might go as far as suggesting that the laws of physics are actually laws of inference [8].

For the remainder of this section, I will consider a more limited model space by considering just apples and bananas: *a* and *b*. The lattice describing the hypothesis space is shown in the upper left-hand corner of Figure 2 and is isomorphic to $\mathbf{2}^M$ where M is the number of atomic statements in the model space $\mathcal{M}$. In the text that follows, I will make reference to probabilities defined in the hypothesis space, and will denote this function as $p_{\mathcal{H}}(\cdot|\cdot)$ to differentiate it from probabilities defined in the other spaces we will consider.

Other statements can be made about our fruit. Imagine that we have a color sensor that can generate two statements:

$$\begin{aligned} `r &= \text{It is red!'} \\ `y &= \text{It is yellow!'} \end{aligned}$$

These atomic statements form a second Boolean lattice, isomorphic to $\mathbf{2}^N$, with N being the number of atomic statements about the measurements. I will call this space the *Data Space* $\mathcal{D}$. Probabilities defined in this space will be denoted by $p_{\mathcal{D}}(\cdot|\cdot)$.

If we choose to consider these statements that the color sensor produces along with statements in the hypothesis space, we can create the product space $\mathcal{H}\times\mathcal{D}$, which results in a more complicated lattice isomorphic to $\mathbf{2}^{(M+N)}$. This is the space in which the Bayesian does his or her work. Probabilities defined in this joint space will be denoted $p_{\mathcal{H}\times\mathcal{D}}(\cdot|\cdot)$, or just $p(\cdot|\cdot)$ for short.

The structure of the joint lattice imposes constraints on the values of the probability. These constraints are given by the three constraint equations: the sum rule, the product rule, and Bayes theorem. Our job is to assign the unknown values of $p(\cdot|\cdot)$, (for example, refer to Table 2). To do this, we need to find the values for all combinations of arguments, which amounts to $(2^{(M+N)})^2$ values. Many of these are trivial, and several are constrained to be related to one another by our three constraint equations, but a few remain that are left unconstrained. This is where external constraints in the form of prior information come into the problem. If one possesses prior information about how these statements imply one another, then this information acts as an external constraint and must be used to guide the assignments of the values for our function $p(\cdot|\cdot)$. There

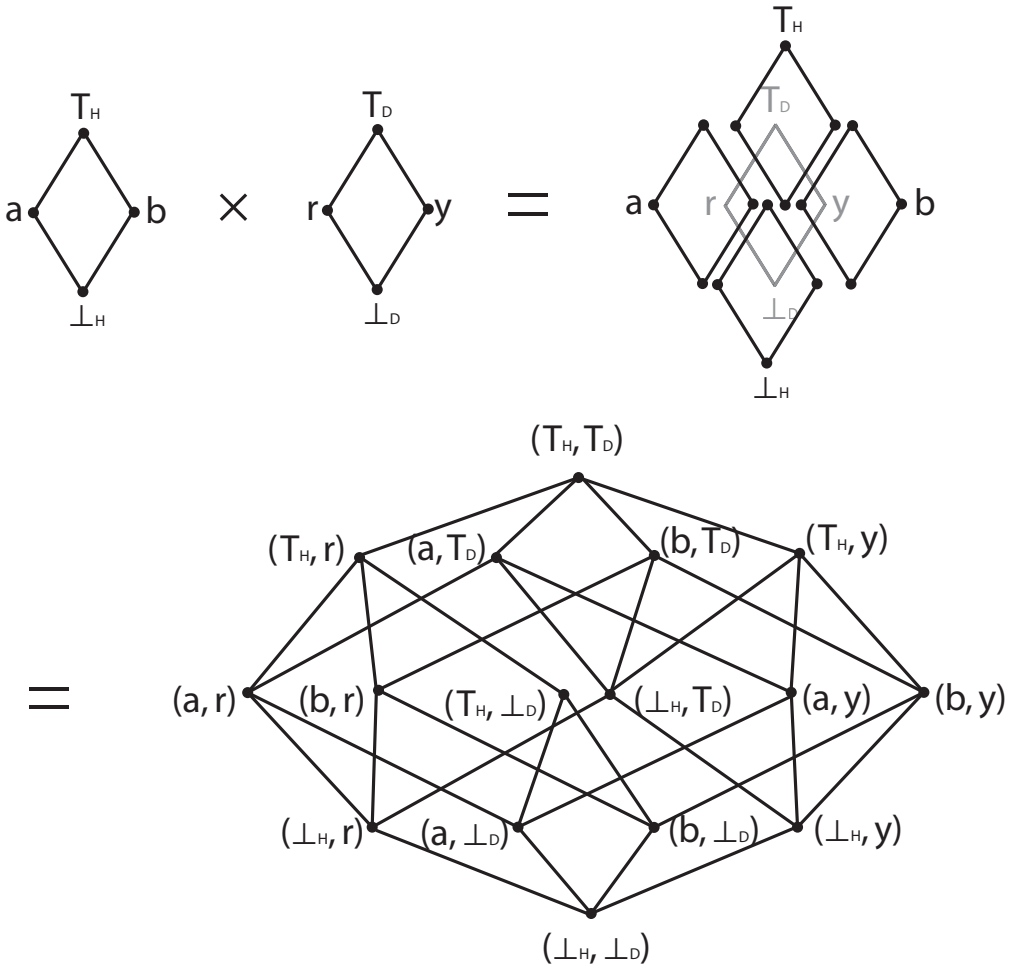

FIGURE 2. This figure illustrates the generation of a product space from a hypothesis space $\mathcal{H}$ and a data space $\mathcal{D}$. The lattice product is formed from the Cartesian product of their respective elements. Our task is to assign bi-valuations to each pair of elements in this joint space. Refer to the text for a complete description of the spaces.

are many ways to do this. One could assign the degree to which the joint atomic statements, such as (a,r), are implied by the joint top $\top \equiv (\top_H, \top_D)$. This degree is written $p((a,r)|\top)$.

One way to do this is by defining all of the prior probabilities in the hypothesis space. In this case, we would assign $p_{\mathcal{H}}(a|\top_H)$ and $p_{\mathcal{H}}(b|\top_H)$ in the hypothesis space. For consistency, these bi-valuations must be equal to their counterparts in the joint space where all possible data statements are considered. That is:

$$p_{\mathcal{H}\times\mathcal{D}}((a,r)\vee(a,y)|\top) \doteq p_{\mathcal{H}}(a|\top). \tag{29}$$

The probability on the left can be written in a more compact notation by defining $(a,\top_D) \doteq (a,r)\vee(a,y)$, and by dropping the messy joint subscript on p

$$p((a,\top_D)|\top) \doteq p_{\mathcal{H}}(a|\top). \tag{30}$$

The values of the prior probabilities in the hypothesis space $p_{\mathcal{H}}(a|\top)$ carry over into the joint space $p((a,\top_D)|\top)$ when we consider the top element in the data space, $\top_D$, which in this case is $\top_D = r \vee y$.

The next assignments to be made are commonly called likelihoods. These are the bi-valuations of the form $p((\top_H, r)|(a,\top_D))$. These probabilities are only defined in the joint space as they consider the degree to which a statement about the hypothesis implies a statement about the data. Note that $(\top_H, r)$ is a statement in the joint space that considers all the possible statements, so that

$$(\top_H, r) \equiv (a,r) \vee (b,r) \tag{31}$$

and that

$$(a,\top_D) \equiv (a,r) \vee (a,y). \tag{32}$$

The posterior probabilities, which are again just a subset of the possible values of the bi-valuation $p(\cdot|\cdot)$, are found by applying the constraint equation known as Bayes Theorem. These are the probabilities, such as $p((a,\top_D)|(\top_M,r))$. The evidence, or the marginal likelihood $p((\top_M,r)|\top)$ can be found using the constraint that all probabilities should sum to one.

An important, and perhaps surprising, fact to note is that the prior probability Π, the likelihood L, the evidence E, and the posterior P in Bayes Theorem

$$P = \Pi \frac{L}{E} \tag{33}$$

are not separate functions! They are all the generalized zeta function $p(\cdot|\cdot)$ with different classes of statements considered in its two arguments. There is no need to use different symbols for this function. It is one function defined on one joint lattice.

That these four terms are each manifestations of the same probability function defined on the joint space, highlights the fact that the likelihood is indeed a probability. It is the degree to which a statement describing a hypothesis implies a statement about the data. It does not matter if one data statement is realized, or even if anyone ever performs the experiment. The likelihood is not just a probability, but it is the same function as the posterior, the prior, and the evidence. This is important since the prior information that we use to constrain the prior probability assignments can also be used to constrain the likelihood assignments and vice versa. They are the same function, and the same information can be used to constraint the values of that function. One can even conceive that one can examine the likelihood assignment, identify the constraints that were applied and then apply them to the prior probability assignments. There is absolutely no problem with this since the entire game is about using information to constrain the probability assignments. All that is required is that this information be applied consistently. This is essentially the approach taken by Rodríguez [9] and Caticha and Preuss [10] to assign entropic priors.

The job of the Bayesian is to extend the definition of this bi-valuation across the joint lattice. From this perspective, there is no updating—only assignment. This is a ‘timeless’ perspective roughly analogous to the four-dimensional view of space-time in a many-worlds picture. All possible models and all possible data are represented, along with all possible combinations of their joins. The hypothesis space is potentially vast.

When a data value is observed, a logical statement about the data is made. The set of all other complementary data statements were not made and no longer need to be part of the lattice. They are the worlds that did not happen. These unspoken statements are grouped into the same equivalence class with the bottom element, and the lattice collapses. Furthermore, if a hypothesis is ruled out with certainty, that statement too can be grouped into an equivalence class with the bottom, and the lattice experiences further collapse. In both cases the information provided by these statements has constrained a specific class of bi-valuations to be zero. We can carry these values on in a large lattice, or simplify the lattice by adopting an equivalence class as described above and effectively adopt a smaller, yet functionally equivalent, lattice.

CONCLUSION

The most basic binary comparisons lead to the concept of ordering. The particular set of elements being compared, and the nature of the comparison give rise to a partially-ordered set (poset). We can then generalize inclusion on this poset to degrees of inclusion, which is implemented by generalizing the zeta function which encodes inclusion on the lattice. This process is analogous to the procedure of analytic continuation of real-valued functions into the complex plane. The end result is that there are a host of undefined values of this function. However, these undefined values are related to one another via a set of constraint equations that encode the structural, or algebraic, properties of the lattice. Three constraint equations arise in distributive lattices: the sum rule, the product rule and Bayes' Theorem. The universal nature of these lattices explains why these three constraint equations appear throughout mathematics and the sciences.

The game we Bayesians are playing is based on assigning the undefined values of this new function, which encodes degrees of inclusion. In most cases, we work with product spaces where one set of statements refer to what we call the hypothesis space, and the other set/s of statements refer to what we call the data space. We artificially divide this function into classes based on which space the statements forming the arguments originate. This gives rise to the notion of the prior, the likelihood, the evidence, and the posterior. However, each of these 'functions' are merely manifestations of this generalized function defined across the joint lattice. They are one and the same function—one and the same probability. The process of updating is simply that of assignment. Furthermore, if one can examine the likelihood assignments and extract the information being used to further constrain their values, one can use this information to constrain the values of the priors. This is just another example of the applications of consistency and honesty, which are fundamental to these theories.

This new perspective is timeless with respect to data acquisition. It is fundamentally distinct from Cox's approach, which directly relied on controversial and poorly-defined concepts such as belief. Furthermore, this perspective is unifying as it applies to measuring in general.

ACKNOWLEDGMENTS

I am deeply indebted to Ariel Caticha, John Skilling, Carlos Rodríguez, Janos Aczél, Deniz Gençağa and Jeffrey Jewell for insightful and inspiring discussions, and many invaluable remarks and comments.

REFERENCES

1. R. T. Cox, *Am. J. Physics* **14**, 1–13 (1946).
2. R. T. Cox, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, 1961.
3. J. Skilling, "The Axioms of Maximum Entropy," in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, edited by G. J. Erickson, and C. R. Smith, Kluwer, Dordrecht, 1988, pp. 173–187.
4. J. Aczél, *Lectures on Functional Equations and Their Applications*, Academic Press, New York, 1966.
5. A. Caticha, *Phys. Rev. A* **57**, 1572–1582 (1998).
6. K. H. Knuth, *Neurocomputing* **67**, 245–274 (2005).
7. K. H. Knuth, "Deriving laws from ordering relations.," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole WY, USA, August 2003*, edited by G. J. Erickson, and Y. Zhai, AIP Conference Proceedings 707, American Institute of Physics, New York, 2004.
8. A. Caticha, and C. Cafaro, "From Information Geometry to Newtonian Dynamics," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs NY, USA, July 2007*, edited by K. H. Knuth, A. Caticha, J. Center, J. L., G. A., and C. C. Rodríguez, AIP Conference Proceedings, ????
9. C. C. Rodríguez, Entropic priors, Tech. rep., Department of Mathematics, University at Albany, Albany NY (1991).
10. A. Caticha, and R. Preuss, *Phys. Rev. E* **70**, 046127 (2004).