Bayesian Inference and Maximum Entropy Methods in Science and Engineering, São Paulo, Brazil, 2008

The Origin of Probability and Entropy

Kevin H. Knuth

Departments of Physics and Informatics
University at Albany
Albany NY 12222
USA

Abstract. Measuring is the quantification of ordering. Thus the process of ordering elements of a set is a more fundamental activity than measuring. Order theory, also known as lattice theory, provides a firm foundation on which to build measure theory. The result is a set of new insights that cast probability theory and information theory in a new light, while simultaneously opening the door to a better understanding of measures as a whole.

Keywords: order, lattice, inference, inquiry, fundamentals, probability, entropy, information
PACS: 02.50.Cw, 89.70.Cf

INTRODUCTION

Probability theory is traditionally built upon the framework of measure theory. However, measuring is the quantification of ordering and is not fundamental. By relying on order theory, which also known as lattice theory, I demonstrate that there are three spaces of interest: the state space, the hypothesis space, and the inquiry space. These three spaces enable us to describe a system, to describe what we know about a system, and to describe what can potentially be known about a system. Defining measures on the hypothesis space and the inquiry space gives rise to probability and relevance, which is a natural generalization of information theory. It should not be expected that these insights will change the application of probability theory to inference problems. Bayes works and works well. However, the relationships between probability and entropy are clarified, which may pave the way for new applications of these techniques.

In this paper, I will describe the important concepts behind this new foundation, and will leave many of the mathematical details to previous works published during the development of these ideas [1, 2, 3, 4]. Although, it should be noted that in [3] equation 31 and the results that follow are incorrect.

CONCEPTS

This development relies on a handful of basic concepts:

* **Order is fundamental**
  The basic properties of ordering impose strong constraints on measures.
* **Sum and product rules arise as *constraint equations***
  The constraints imposed by ordering manifest themselves as constraint equations.

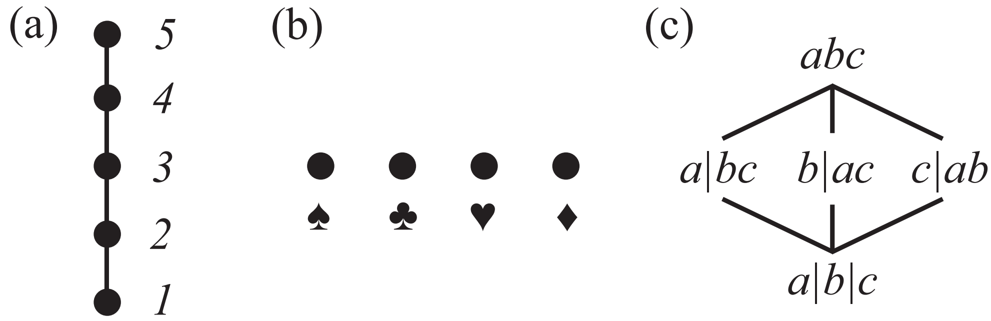

FIGURE 1. Three basic examples of posets. (a) The integers ordered by the usual $\leq$ form a *chain*. The element 2 is drawn above 1 since $1 \leq 2$, and they are connected by a line because 2 covers 1 in the sense that there is no integer x between 2 and 1 such that $1 \leq x \leq 2$. (b) The four card suits are incomparable under a wide variety of card game rules and we draw them side-by-side to express this. This configuration is called an *antichain*. (c) The set of partitions of three elements a, b and c ordered by partition containment forms a more complex poset that exhibits both chain and antichain behavior. One chain consists of the elements $a|b|c$, $a|bc$, and abc since each successive partition contains the previous. The elements $a|bc$, $b|ac$, and $c|ab$ form an antichain because none of these three partitions contains another.

These equations are more commonly known as the sum and product rules.

* **States, statements and questions are elements of three fundamental spaces**
  We work in three spaces: the space of states, the space of statements, and the space of questions. Much confusion arises when these three spaces are not explicitly distinguished.
* **Probability is a measure in the space of statements**
  Probability theory arises as a measure on the space of statements. This space is ordered by implication, and the measure inherits its meaning from this ordering relation. Thus probability is a measure of degrees of implication. The sum rule and product rule follow naturally from the constraints imposed by implication.
* **Relevance is a measure in the Space of Questions**
  Relevance is the measure on the space of questions. By relating the relevance of questions to the probability of their answers, one can recover a natural generalization of information theory based on Shannon entropy.

The result of this work is that probability theory and information theory can be *derived* in this order-theoretic formulation! Moreover, this approach is generally applicable to arbitrary state spaces based on partially ordered sets. The result is that the sum and product rules are ubiquitous.

ORDER

Before one can measure, one must be able to order. This makes the process of ordering the more fundamental activity. Ordering is performed by taking two elements of a set and comparing them according to a binary ordering relation, generically denoted $\leq$. One of the first examples that comes to mind may be the ordering of the integers according to

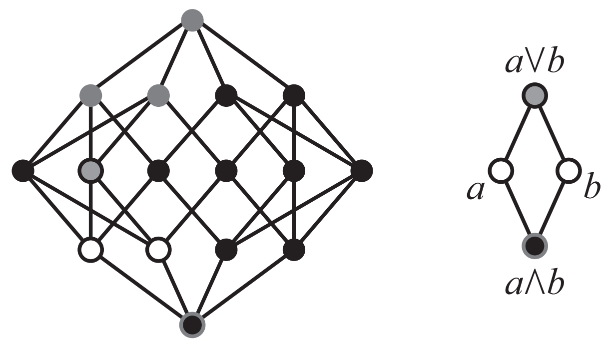

FIGURE 2. This figure illustrates the concept of the join and meet. The upper bound of the two elements denoted by the white circles are the elements denoted by gray circles. Their least upper bound, or the join, is the gray circle with the black outline. In this example, their meet is the bottom element. The four key elements in the lattice on the left are isolated and illustrated on the right.

the usual meaning of the symbol $\leq$ '*is less than or equal to*'. This is one of the simplest orderings called a *chain* (Fig. 1a). To make this figure we simply draw element B above element A if $A \leq B$ and connect them with a line if there does not exist an element X in the set such that $A \leq X \leq B$.

Partially Ordered Sets

The simplicity of the chain may lead one to think that ordering is always simple. However, things become very interesting when some elements of the set are incomparable with respect to the binary ordering relation. Think of the traditional example of incomparability: apples and oranges. Another example is that of card suits: hearts, clubs, spades, and diamonds, where no suit is greater than any other. The result is an *antichain* (Fig. 1b) where the elements are placed side-by-side to indicate the fact there is not one element that includes any other.

A more complex example that involves both inclusion and incomparability is the structure that one obtains when considering partitions. Consider the problem of putting three pieces of fruit on a table (Fig. 1c). One could put all three pieces of fruit on the same plate *abc*, or each on separate plates $a|b|c$. Of course, we could put two pieces of fruit on one plate and the third piece of fruit on a second plate, which can be represented by these partitions: $ab|c$, $bc|a$ and $ca|b$. These partitions can be compared according to a relation that decides whether one partition includes another. The partition *abc* includes the partition $a|b|c$ since I can easily take the fruit on one plate and sub-divide the plate into three sections, each with one piece of fruit. However, the partitions $ab|c$ and $bc|a$ are incomparable since there is no way to sub-divide the plates in $ab|c$ to obtain $bc|a$. Since not all the partitions of the set can be compared, we call this a *partially ordered set* or a *poset* for short.

Lattices

We begin by introducing the concept of an *upper bound*. Given a set of elements, their upper bound is the set of elements that contain them. For example, the upper bound of the integer 3 in Fig. 1a is the set $\{4,5\}$; whereas the upper bound of the partition $ab|c$ in Fig. 1c is the set $\{abc\}$. Given a pair of elements a and b, the least element of their upper bound is called the *join*, which we denote as $a \vee b$. This concept is illustrated in Fig. 2. We can define the *lower bound* of a pair of elements similarly by considering all the elements that pair of elements contain. The greatest element of the lower bound is called the meet, which we denote as $a \wedge b$. A lattice is a partially ordered set where each pair of elements has a unique meet and a unique join. Graphically, the join can be found by starting at both elements and following the lines upward until they first intersect. The meet is found similarly. Not all elements can be formed from the join of two other elements. Such elements are called *join-irreducible elements*. For example, the partitions $ab|c$, $bc|a$ and $ca|b$ cannot be formed by joining any other pair of partitions.

Algebras

The join and meet defined in terms of the upper bound and lower bound emphasize the hierarchical structure of the poset. We can also choose to view the join and meet as algebraic operations that take any two lattice elements to a unique third lattice element. In this context, the lattice is an algebra. This results in both a structural and operational perspective of the poset, and their relationship is given by

$$x \leq y \Leftrightarrow \begin{array}{c} x \vee y = y \\ x \wedge y = x \end{array} \tag{1}$$

Given different posets, we find that this leads to different algebraic identities. For example, the integers ordered by the usual '*less than or equal to*' leads to

$$x \leq y \Leftrightarrow \begin{array}{c} \max(x,y) = y \\ \min(x,y) = x \end{array} \tag{2}$$

whereas the positive integers ordered by '*divides*' leads to

$$x|y \Leftrightarrow \begin{array}{c} \mathrm{lcm}\ (x,y) = y \\ \gcd(x,y) = x \end{array} \tag{3}$$

Sets ordered by the usual '*is a subset of*' leads to

$$x \subseteq y \Leftrightarrow \begin{array}{c} x \cup y = y \\ x \cap y = x \end{array} \tag{4}$$

Written generally, the identities for each of these posets can be expressed quite simply as in (1).

QUANTIFICATION

We can generalize the concept of inclusion by introducing valuations. Valuations are functions that take lattice elements to real numbers. The concept of a measure arises from considering a function that takes one lattice element to a real number. Valuations must have an agreed-upon context for them to inherit meaning from the ordering relation. Bi-valuations, which take two lattice elements to a real number can rely on the additional lattice element to supply the context. This is why probabilities can be written as valuations $p(x)$ or bi-valuations $p(x|y)$. In the first instance, the context is assumed to be implicitly understood, whereas in the second instance the context is made explicit.

Constraint Equations

If the valuations are to reflect the lattice structure in any meaningful way, we cannot have complete freedom to assign the valuations any way we wish. The lattice structure must impose constraints. Since the join and meet locally define the lattice structure, these constraints can only manifest themselves in terms of a relationship between the valuations of a pair of elements and the valuations assigned to their join and meet.

The procedure for deriving these constraints is straightforward. Using the lattice properties, we write an expression describing a lattice element in two different ways. We then require that the valuation assigned to that lattice element (or that bi-valuation assigned to the relevant pair) does not depend on the particular expression we choose. There are many ways that one can view this. One way is to look at the basic lattice properties, and consider what constraints they place on valuation assignments. It turns out that one need only consider three properties. All additional lattice properties can be shown to result in constraint equations that reproduce these results.

Associativity of the Join

The first constraint we consider is associativity of the join. This property is possessed by all lattices, and the constraint equation that results, will be universally valid. Using associativity of the join, we can write the following expression two ways $a \vee (b \vee c) = (a \vee b) \vee c$. Any valuation assignment involving this lattice element must not depend on the way we express the element mathematically. That is, both expressions refer to the same element.

We begin by considering a special case where the meet of two elements a and b is the bottom element of the lattice. If the valuations are to convey the lattice structure, the valuations assigned to a and b must somehow be related to the valuations assigned to their join $a \vee b$. We do not assume that we know what this relation is, merely that it exists. If it did not exist, then the valuations we assigned would not reflect the lattice structure. For simplicity, within a particular context t we write this relationship as

$$v(a \vee b) = S(v(a), v(b)), \tag{5}$$

or more explicitly, using bi-valuations

$$w(a \vee b, t) = S(w(a,t), w(b,t)), \tag{6}$$

where S is an unknown function.

We now consider a more complicated expression for a lattice element $a \vee (b \vee c)$. Associativity allows us to rewrite the same expression as $(a \vee b) \vee c$. Writing the valuation for this element in terms of the function S leads to two expressions, with the function S appearing nested within itself in two different ways. Since both expressions must hold, the result is a functional equation which can be solved for the function S. This functional equation is known as the associativity equation, and the solution gives rise to the *sum rule*

$$\begin{aligned} w(a \vee b, t) &= S(w(a,t), w(b,t)) & (7) \\ &= w(a,t) + w(b,t). & (8) \end{aligned}$$

In the case of two general lattice elements, this rule becomes [3]

$$w(a \vee b, t) = w(a,t) + w(b,t) - w(a \wedge b, t), \tag{9}$$

which relates the valuations assigned a and b to the valuations assigned to their join and meet. While I have left out many details of the proof [3], what is important to realize is that this result holds for all valuations on all lattices. Rather than thinking of the sum rule as a rule for manipulating valuations, it is instead a constraint equation that constrains the values that the valuation function can take on the lattice. This rule appears in many contexts. Here are a few of my favorites:

$$\begin{aligned} p(a \vee b|t) &= p(a|t) + p(b|t) - p(a \wedge b|t) & (10) \\ I(a;b) &= H(a) + H(b) - H(a,b) & (11) \\ max(a,b) &= a + b - min(a,b) & (12) \\ \log(\text{lcm}(a,b)) &= \log(a) + \log(b) - \log(\text{gcd}(a,b)). & (13) \end{aligned}$$

The first is the standard sum rule of probability. The second is the relation between mutual information $I(a;b)$ and the joint entropy $H(a,b)$. The third is an interesting example from Pölya and Szegö [5, Vol II., p. 121]. The last is a formula that I recently derived by considering the integers ordered by division. By working with lattices in general, we have not only derived probability theory, but we have also derived measure theory since the sum rule really is the property of countable additivity or σ-additivity.

Distributivity of Join over Meet

Since many the spaces on which we will be assigning valuations are limited to distributive lattices, we can consider the additional constraints that distributivity imposes. The derivation [3] proceeds similarly to the previous case. Consider a lattice element written as $a \wedge (b \vee c)$. Using the distributivity property, this element can also be written

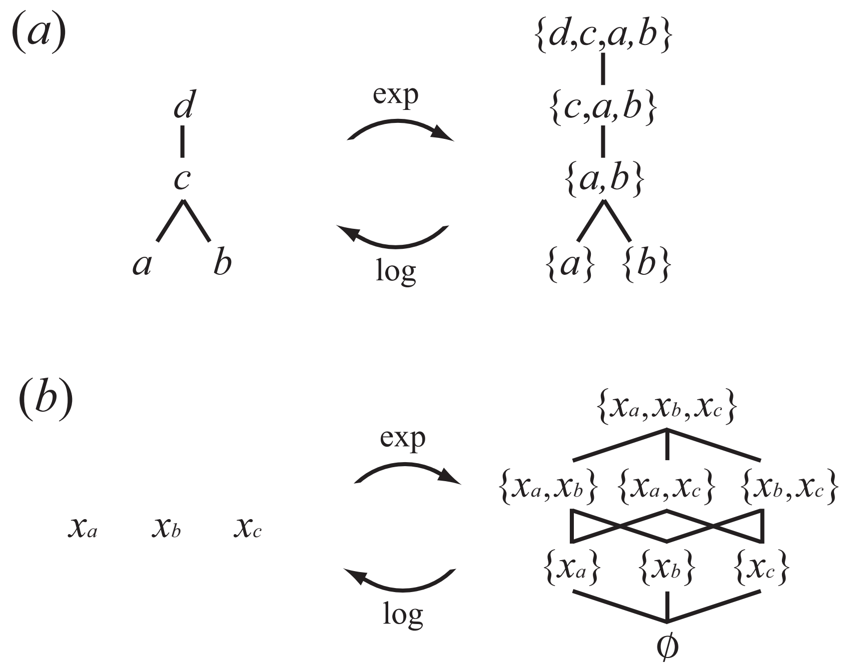

FIGURE 3. Exponentiation of the antichain leads to a Boolean lattice. The join-irreducible elements of the Boolean lattice are the three mutually exclusive elements just above the bottom. This set of join-irreducibles is isomorphic to the original antichain.

as $(a \wedge b) \vee (a \wedge c)$. Assuming that the valuations we assign to a and b are related to the valuations we assign to the meet $a \wedge b$, via some unknown function P, we find that this function gives rise to the constraint equation known as the *product rule*

$$w(a \wedge b, t) = w(a, t) w(b, a \wedge t). \tag{14}$$

Commutativity of Meet

The last constraint equation comes from the fact that the meet is commutative. Through a very straightforward proof following the same procedure as above, commutativity in conjunction with the product rule gives rise to Bayes theorem

$$w(a, b \wedge t) = w(a, t) \frac{w(b, a \wedge t)}{w(b, t)}. \tag{15}$$

This result is so trivial that it is easy to take it for granted. However, there are situations that lack commutativity and this constraint does not hold. This lack of commutativity is the reason that quantum wavefunction amplitudes obey both a sum rule (Feynman path integrals) and a product rule [6, 2], but not Bayes Theorem [2].

GENERATING LATTICES

Here we consider the spaces that we work in when we describe a system, describe what we know about a system, and describe what we can potentially know about a system. Each space is a lattice (or poset), with each successive lattice deriving from the previous one via order-theoretic exponentiation. Exponentiation relies on generating new lattice elements from old by grouping the old elements into sets called downsets [1]. Downsets are constructed so that they contain their lower bound. That is, given any element in the downset, all elements included by that element are also members of the set.

Figure 3 shows two examples of the downset construction. The upper image (Fig. 3a) shows an arbitrary lattice and the lattice formed by taking all of the possible downsets and ordering them according to subset inclusion. Note that there can be no element $\{c\}$ since this set must include all of the elements that c includes, such as a and b, which gives $\{c,a,b\}$. For this reason, even though there are four elements in the original lattice, there are fewer than $2^4 = 16$ elements in its exponentiation. It is important to note that the join-irreducible elements $\{a\}$, $\{b\}$, $\{c,a,b\}$, and $\{d,c,a,b\}$ form a sublattice that is isomorphic to the original. The procedure of selecting the join-irreducible elements is the order-theoretic analog of the logarithm. This exp-log relationship is known as *Birkhoff's Representation Theorem* [7], since the simpler lattice can act as a representation of the more complex one.

The second example (Fig. 3b) is more relevant to our purposes here. The exponentiation of the antichain of three elements leads to the Boolean lattice of $2^3 = 8$ elements. In this lattice, we see that all possible combinations of subsets of the three elements in the antichain occur since in the antichain there are no elements which include any others. Note again that the join-irreducible elements of the Boolean lattice, $\{x_a\}$, $\{x_b\}$, and $\{x_c\}$, form a structure isomorphic to the antichain, x_a, x_b, and x_c.

We now examine how the downset construction enables us to generate the three spaces that we work in when performing inference.

THREE SPACES

The ***State Space*** is the most fundamental as it describes the system. In general, the state space is a poset. Its elements are called *states*, and each state constitutes a description of system. While states are typically unordered and form an *antichain* (Fig. 4, left), one can construct descriptions that form more complex posets or lattices.

The ***Hypothesis Space*** is a lattice that is constructed by taking downsets of states and ordering them according to the usual set inclusion $\subseteq$. I use the term *statements* to describe sets of states. Statements can be thought of as a set of potential states of the system. In this sense, statements can describe what we know about a system, or equivalently our states of knowledge about the system. Statements ordered by set inclusion are equivalent to the usual logical statements ordered by logical implication. Since the state space is an antichain, the hypothesis space is a *Boolean lattice* (Fig. 4, left). Note that I am using a different notation here than in Fig. 3b where the statement $a \equiv \{x_a\}$, $\vee \equiv \cup$ (set union), and the ordering relation is $\subseteq$ (subset inclusion).

The ***Inquiry Space*** is constructed by taking sets of statements and ordering them

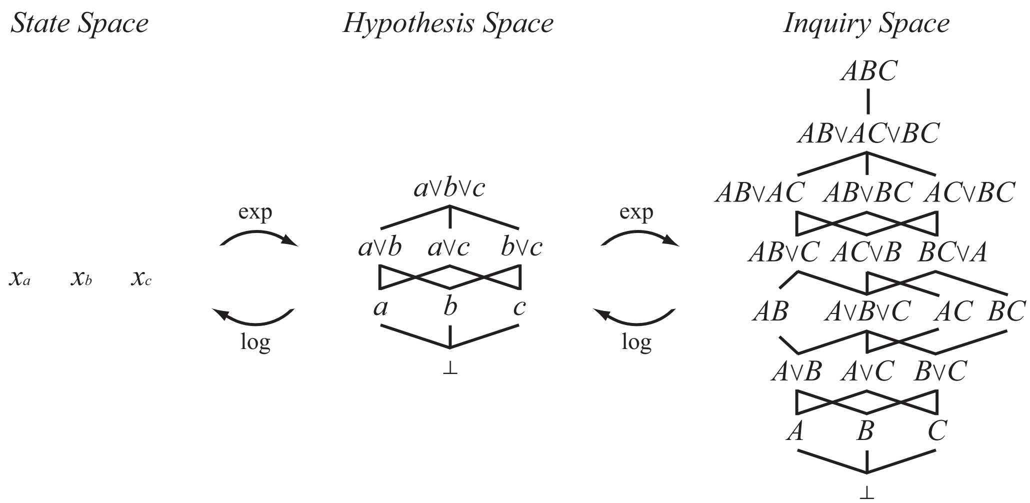

FIGURE 4. Three spaces generated by successive order-theoretic exponentiation. Explanation of the notation can be found in the text.

according to the usual set inclusion $\subseteq$ in the same way that the hypothesis space is constructed from the state space. Each element describes what can be known about a system, or equivalently, the set of potential states of knowledge about a system. For this reason, its elements are called *questions*. Questions ordered by set inclusion are equivalent to the usual concept of logical questions ordered according to whether answering one question answers another. If the hypothesis space is Boolean, the inquiry space is a *free distributive lattice* (Fig. 4, right). Here I use a compact notation where $A \equiv \{a, \bot\}$, $AB \equiv \{a \vee b, a, b, \bot\}$, etc., and $\vee \equiv \cup$ (set union), so that $C \vee AB \equiv \{a \vee b, a, b, c, \bot\}$.

The Inquiry Space

While we are relatively familiar with the Boolean hypothesis space, the space of questions is perhaps more mysterious. First, from the downset construction, one can see that a question is defined in terms of the set of statements that answer it. These sets of answers must include all statements that imply them, which comes naturally through the use of downsets. It follows that two questions are equivalent if they are both answered by the same set of statements, such as '*Is it raining*?' and '*Is it not raining*?'. Answering is implemented by subset inclusion so that if question X is a subset of question Y, $X \subseteq Y$, then by answering X one will have also answered Y. If $X \subseteq Y$, we say that X answers Y.

There is a special element of the lattice called the center, which in this case is $A \vee B \vee C \equiv \{a, b, c, \bot\}$. Aside from the bottom element (the falsity), this question entertains only the join-irreducible (and mutually exclusive) answers a, b, c in the hypothesis space, and for this reason I call it the *central issue*. The questions above the central issue are real questions as defined by Cox [8] since they entertain each of the join-irreducible answers and thus are answered always by a true statement. This is the space of interest and I call it the real sublattice or real question lattice.

In the real sublattice, there are questions that neatly partition the answers. In this example, they are $A \vee B \vee C, A \vee BC, B \vee AC, C \vee AB$, and ABC. These elements form a sublattice that is isomorphic to the partition lattice (Fig. 1c). For this reason, I call these elements partition questions. The central issue $A \vee B \vee C$ is the smallest partition question, and can be read as 'Is it a, b, or c?'; whereas the partition question $C \vee AB$ can be read as 'Is it c or not c?'

Not all questions are easily verbalized. One reason for this is that we tend to ask relevant questions, and this means that we lack the language to describe questions with lesser relevance. One should keep in mind that these questions are mathematical abstractions, which are useful for computation; much like the utility of negative numbers when balancing a check book.

Valuations on the Three Spaces

We have the freedom to assign valuations to the join-irreducible elements of a lattice [9]. The constraint equations determine the remaining valuation assignments. In the case of an antichain state space, the poset structure itself imposes no constraints. Instead, problem-specific constraints must be considered to make objective assignments.

Since these three spaces are merely representations of one another, valuations on one lattice should be related to valuations on another. That is, the valuations assigned to one lattice should be a function of the valuations assigned to another lattice. However, since we have the freedom to assign valuations to the join-irreducible elements of each lattice, there are no constraints imposed by the lattice structure on this function. The choice of this function is arbitrary, however, it will serve not only to determine the valuation assignments, but also the *context* of the bi-valuations. Again, application-specific constraints are critical. Different problems will result in different valuations.

We begin by considering a valuation assigned to the join-irreducible elements of the state space. Here I consider mono-valuations since there is no obvious common context. Imagine that the system is a sheet of paper, which has been divided into three sections: the first x_a encompassing 1/2 of the area of the paper, the second x_b encompassing 1/3 of the area of the paper, and the third x_c encompassing 1/6 of the area of the paper. We may choose to assign the following valuations to these states

$$m(x_a) = 0.5 \quad (16)$$

$$m(x_b) = 0.33 \quad (17)$$

$$m(x_c) = 0.17 \quad (18)$$

to represent the fractional areas that each state encompasses.

The bi-valuation assigned to the hypothesis space represents the degree to which one statement implies another, which is the probability. The natural context in this space is the top element $\top$, so that $p(a) \equiv p(a|\top)$. Since the hypothesis space is generated by the state space, the valuations assigned to the join-irreducible elements of the hypothesis space must be a function of the corresponding valuations in the state space. That is

$$p(a) = f(m(x_a)) \tag{19}$$

where x_a is the state with valuation $m(x_a)$, a is the statement with probability $p(a)$ and $f(\cdot)$ is an arbitrary function. Given the application, we may choose the function $f(\cdot)$ to be the identity, so that

$$p(a) = 0.5 \tag{20}$$

$$p(b) = 0.33 \tag{21}$$

$$p(c) = 0.17. \tag{22}$$

Bi-valuations in the inquiry space represent the degree to which one question answers another. For this reason, I call this bi-valuation the *relevance*. The natural context is the central issue, denoted I, so that $d(I|A)$ represents the degree to which A answers I. Since the inquiry space is generated by the hypothesis space, we again assume that the valuation assigned to a join-irreducible question is a function of the probability of the corresponding statement. For example, the valuation $d(A)$ assigned to question A must be a function of the probability of the statement a, where $A = \{a\}$,

$$d(A) = g(p(a)). \tag{23}$$

We could indeed choose $g(\cdot)$ arbitrarily, but we will see that there is an additional consideration that imposes a strong constraint on $g(\cdot)$.

Relevance of Questions

One choice is to assign the valuations on the space of questions by letting $g(\cdot)$ be the identity so that $d(A) = p(a)$. Doing so leads to the result that every question R in the real sublattice will have a valuation of $d(R) = 1$. This valuation indicates the degree to which the central issue I answers the question R, $d(R|I)$. While this is a perfectly valid measure, it may not be what we want.

By imposing a new constraint, we can change the context of the bi-valuation assignment. Since we are focused on the real sublattice, we consider a constraint that explicitly indicates this. The join-irreducible elements of the real sublattice can be shown to be partition questions. We instead declare that the degree to which a partition question P resolves the central issue I is a function of the probabilities of the N greatest statements $x_1, x_2, \ldots, x_N$ that define the question. That is, the relevance of the question $A \vee BC$, which asks '*Is it A or not A*?', depends on the probability of the statement a and the probability of the statement $\sim a = b \vee c$. This can be expressed as

$$\begin{aligned} d(I|P) &= K(p(x_1|\top), p(x_2|\top), \ldots, p(x_N|\top)) \\ &\equiv K(p_1, p_2, \ldots, p_N), \end{aligned} \tag{24}$$

where the $K(\cdot)$ is a function to be determined.

Due to the structure of the lattice, the relevance satisfies four important properties: *additivity* and *subadditivity* which are assured by the sum rule (9), *symmetry* with respect to disjunction of the second argument, and *expansibility* which describes how the lattice collapses when a statement is found to be false. An important result from Aczél et al. [10] enables us to find the unique form of K to be related to the Shannon entropy

$$K(p_1, p_2, \cdots, p_N) = aH(p_1, p_2, \cdots, p_N) + b \tag{25}$$

where a, b are arbitrary non-negative constants and the Shannon entropy [11] is

$$H(p_1, p_2, \cdots, p_N) = -\sum_{i=1}^{N} p_i \log_2 p_i. \tag{26}$$

The result is that relevance is a generalization of information theory with Shannon entropy as the basis. This is an interesting generalization as we have the machinery of the sum and product rules to construct assignments that are more complex than usually considered. This includes many of the higher-order informations that have been previously suggested [12, 13], but due to the product rule we also see that ratios of information theoretic quantities are also important. Furthermore, since the coefficients a and b are arbitrary, we have the ability to normalize the relevance between zero and one solving an important problem where higher-order informations become negative in information theory [13].

MaxEnt and Assigning Probabilities

We can also use the relationships between valuations in these different spaces to assign probabilities in a different way. A great deal of information goes into selecting the poset that describes the state space. This poset ultimately dictates the nature of the central issue in the question lattice. We can use this information about the chosen state space structure to assign probabilities in the hypothesis space. This is accomplished by maximizing the relevance of the central issue with respect to any problem-specific constraints and computing the corresponding probabilities. The reason why we should maximize the relevance of the central issue can be seen by considering what would happen if the relevance of the central issue were not maximumized. In this case, it would suggest that there exists a more relevant state space for the problem. If that were the case, then one ought to use that state space.

This procedure of maximizing the relevance of the central issue is known as the *Principle of Maximum Entropy* [14]. It works because we use information that went into the state space specification to assign probabilities to statements in the hypothesis space.

Updating Probabilities

The last case I consider is the situation where we have already assigned both probabilities and relevances, but wish to update them using information about a new additional constraint. One can consider a set of hypothesis spaces each with different probability assignments. To select a new probability assignment, we simply rank the possibilities according to preference. Relying this strategy, Caticha [15], expanding on the work of Skilling [16], has shown that the preferred probability assignment can be found by maximizing the relative entropy. The relative entropy is the valuation on this space of statement lattices endowed with a valuation.

CONCLUSION

Order theory is the fundamental picture that gives rise to measure theory in general, and more specifically probability theory, and information theory. The result is that the poset that describes the state space gives rise to a lattice of statements, called the hypothesis space, and a lattice of questions, which is called the inquiry space, via order-theoretic exponentiation. When we solve problems we work in these three spaces. These spaces describe the system, what we know about the system, and what we can potentially know about a system. Much confusion has resulted from a lack of understanding of these three spaces, the valuations they support and the roles that they play in problem-solving. This work comprises a novel and unifying foundation for both inference and inquiry.

ACKNOWLEDGMENTS

I would like to thank John Skilling for his encouragement and strong support of this work, as well the new ideas and directions that he has introduced. I would also like to thank Janos Aczél, Ariel Caticha, Philip Goyal, Steve Gull, Jeffrey Jewell, Vassilis Kaburlasos, and Carlos Rodríguez for inspiring discussions, invaluable remarks and comments, and much encouragement. This work was supported in part by the College of Arts and Sciences and the College of Computing and Information of the University at Albany (SUNY), the NASA Applied Information Systems Research Program (NASA NNG06GI17G) and the NASA Applied Information Systems Technology Program (NASA NNX07AD97A). I would like to thank the organizers Carlos Pereira, Julio Stern and Marcelo Lauretto for their efforts in organizing this meeting and their kind invitation. Last, I would like to extend a special thank you to Julio Stern and his wife Marisa for their generous hospitality before, during and after the meeting.

REFERENCES

1. K. H. Knuth, "What is a question?," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Moscow ID, USA, 2002*, edited by C. Williams, AIP Conference Proceedings 659, American Institute of Physics, New York, 2003, pp. 227–242.
2. 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, pp. 204–235.
3. K. H. Knuth, *Neurocomputing* **67C**, 245–274 (2005).

4. K. H. Knuth, “Lattice theory, measures, and probability.,” in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 2007*, edited by K. H. Knuth, A. Caticha, J. L. Center, A. Giffin, and C. C. Rodríguez, AIP Conference Proceedings 954, American Institute of Physics, New York, 2007, pp. 23–36.
5. G. Pölya, and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis, 2 vols.*, Springer, Berlin, 1964, 3rd edn.
6. A. Caticha, *Phys. Rev. A* **57**, 1572–1582 (1998).
7. B. A. Davey, and H. A. Priestley, *Introduction to Lattices and Order*, Cambridge Univ. Press, Cambridge, 2002.
8. R. T. Cox, “Of inference and inquiry, an essay in inductive logic,” in *The Maximum Entropy Formalism*, edited by R. D. Levine, and M. Tribus, The MIT Press, Cambridge, 1979, pp. 119–167.
9. D. A. Klain, and G.-C. Rota, *Introduction to Geometric Probability*, Cambridge Univ. Press, Cambridge, 1997, ISBN 0-521-59654-8.
10. J. Aczél, B. Forte, and C. T. Ng, *Adv. Appl. Prob.* **6**, 131–146 (1974).
11. C. F. Shannon, and W. Weaver, *The Mathematical Theory of Communication*, Univ. of Illinois Press, Chicago, 1949.
12. W. J. McGill, *IEEE Trans. Info. Theory* **4**, 93–111 (1955).
13. A. J. Bell, “The co-information lattice,” in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA 2003*, edited by S. Amari, A. Cichocki, S. Makino, and N. Murata, 2003.
14. E. T. Jaynes, “Where do we stand on maximum entropy,” in *The Maximum Entropy Formalism*, edited by R. D. Levine, and M. Tribus, The MIT Press, Cambridge, 1979, pp. 15–118.
15. A. Caticha, “Relative Entropy and Inductive Inference,” 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, pp. 75–96.
16. J. Skilling, “Classic maximum entropy,” in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, Kluwer, Dordrecht, Netherlands; Boston, U.S.A., 1989, pp. 45–52.