Probability Space for First-Order Predicate Logic

Part of the ordinary meaning of any idiom of quantification consists of susceptibility to restrictions; and that restrictions come and go with the pragmatic wind[1].

I. Introduction

This work demonstrates a new method for defining and calculating probability values for quantified expressions over a finite domain $\boldsymbol{D}$. A model on $\boldsymbol{D}$ gives every first-order formula ϕ an extension [ϕ] in the space of assignment sequences $\boldsymbol{D}^{\omega}$, thus $[\phi] \in 2^{\wedge}(\boldsymbol{D}^{\omega})$. These extensions generate a sigma-algebra $\boldsymbol{F_D}$. Probability is the infinite product of the normalized counting measure, $P[\phi] = |\phi|/|\boldsymbol{D}|$. The novelty here is to extend the familiar first-order model adjoining the power set $2^{\boldsymbol{D}}$, and re-interpreting bound variables as if they are free variables of a distinct type, denoting elements not of $\boldsymbol{D}$ but of $2^{\boldsymbol{D}}$. The resulting system is intermediate between a two-sorted elementary logic, and a second order logic.

In standard models[2], sentences (closed quantified formulas) are true or false according to their satisfaction either by all sequences, or by none. In the extended models here, bound variables in quantified expressions can be satisfied with subsets in $2^{\boldsymbol{D}}$. Compared to previous suggestions for probability logic, the resulting system is perhaps more complex, but far more powerful.

II. Homage to Gaifman

A foundational contribution to the study of probability for first-order logic was Haim Gaifman's "Concerning Measures in First Order Calculi." To define an addition law for any measure μ() on sentences rather than sets, Gaifman uses two proof-theoretic axioms:

(1) if $\vdash \phi$ and $\vdash \psi$ then $\mu(\phi) = \mu(\psi)$
(2) if $\vdash \sim(\phi \wedge \psi)$ then $\mu(\phi \vee \psi) = \mu(\phi) + \mu(\psi)$

which together entail:

(3) if $\vdash \phi \equiv \psi$ then $\mu(\phi) = \mu(\psi)$

His construction depends on a pre-existing assignment of probability measure m() to quantifier-free sentences, then extending that measure to m*() on compound and quantified sentences, subject to the following condition:

---

[1] David Lewis (1986), On the Plurality of Worlds, p. 164
[2] E.g., Hilbert and Ackerman 1928, Wehmeier 2018

(4) $m^*(\exists x\ \phi x) = \sup\{ m^*(V\{ \phi a_i \}_n \}$[3]

with a corresponding condition using the infimum for universal expressions. This definition is required because there may be no assumption that distinct atomic sentences 'ϕa' and 'ϕb' are "disjoint" in the sense of (2). In the case of universal quantification and compound conjunctions, it is similarly unjustified to assume independence. Gaifman proves by an argument similar to Caratheodory's extension theorem that m*() is a unique extension of m() if it satisfies (4).

To contrast the proposal here with Gaifman's, we now cherry-pick two different measures and apply the condition (4). First, assume an $m_1()$ which in effect applies the condition (4) to the conjunction & disjunction of any two formulas.

(5) $m_1(\phi \vee \psi) = \sup\{ m_1(\phi), m_1(\psi) \}$
(6) $m_1(\phi \wedge \psi) = \inf\{ m_1(\phi), m_1(\psi) \}$

Assume m_1 is a purely subjective probability distribution on a language with just two predicates, Ra and Wa, representing colored balls in an urn. Also assume each ball is labeled or named with numeral constants 'a' ε {0…9}). Next, tell a story: Starting with ten white ping-pong balls. You yourself give them each their own number 0-9 and paint the first six, 0-5, red. You leave the others, 6-9, plain white. Now you have certain knowledge of the color of each ball, and your subjective probability for each atomic assertion R0…R9 is either one or zero. Same for the subjective probability of each W0…W9. Call this probability $m_1()$.

Now let's ask you the probability of the generalizations (∀x)Rx and (∃x)Rx. The Gaifman condition says that your $m_1()$ must respect all the conjunctions and disjunctions of finite subsets of assertions in {R0…R9}. Call the collection of subsets of the atoms 2**{Ra}, where any particular subset $\{Ra\}_n$ ε 2**{Ra}, and the logical combinations $V\{Ra\}_n$:= Ră, and $\Lambda\{Ra\}_n$:= Râ[4].

Clearly the inf{Râ} over those finite conjunctions will be zero, since each of the atoms R6…R9 give P(Ra) = 0. Mutatis mutandis, the sup{Ră} must be one, since each R0…R5 give P(Ra) = 1. This is in accordance with intuition, since in this scenario it's foreknown that (∃x)Rx is true while (∀x)Rx is false – you painted more than one yourself, but deliberately not all. But what if you have incomplete information?

In a second scenario, suppose that a colleague Nancy prepares the same domain, and informs you only that six of the balls were painted red, while the other four were left white. Suppose also that she claims to have permuted the balls by a randomizing device before choosing six to paint. You trust her on both reports. This may well change the subjective probability of your judgements on the atomic sentences

[3] In a compact notation analogous to computer languages, $\{ \phi a_i \}_n$ denotes a finite set of atomic formulas, each with a name 'a_i' where i<n; V{…} denotes the formula which is the disjunction of all the members in the indicated set; sup{…} denotes the numerical supremum of all values in the indicated set, from all finite formulas.

[4] To clarify notation, $\{Ra\}_n$ is a set of atomic formulas, V(Rã) is the single compound formula which is their disjunction, and 'Ră' is a notational abbreviation for V(Rã). Finally, let {Ră} denote the set of all such disjunctions, while {Râ} is the set of conjunctions.

R0…R9. Call your new distribution $m_2()$. You might reasonably assign $m_2(Ra) = 60\%$ each 'a', $m_2(Wa) = 40\%$, again for every 'a'. Now, since each individual atom is one among the finite disjunctions of atoms,

(7) $m_2((\forall x)Rx) = \inf\{ m_2\{R\hat{a}\}\} = m_2(Ra) = 60\%$ and
(8) $m_2((\exists x)Rx) = \sup\{ m_2\{R\check{a}\}\} = m_2(Ra) = 60\%$

But isn't this odd, given that you already know from Nancy that $(\exists x)Rx$ is certainly true, while $(\forall x)Rx$ is certainly false? Perhaps it's a bigger problem that both generalizations have equal probability, which is equal also to the probability of the open formula 'Rx'.

III. Apologia - Extensions and Bound Variables

Notwithstanding the successes of Gaifman's construction, there is an important contrary intuition to acknowledge. Bound variables in quantified expressions are indispensable to the basic function of first-order models – giving truth-values to sentences. The new alternative here may appear to modify radically the semantics of quantifiers. Rather, the modification proposed is a natural option for, and a normal evolution of the standard semantics. Again, unlike other attempts at probability logic[5], this is not a proposal for a multi-valued logic. It is a proposal to interpret free and bound variables as ranging over distinct domains, and expressions as ranging over a product space.

Two Semantics for Propositional Logic

- Can be interpreted either with binary truth values - truth tables
- Or with set extensions: (A & B) → C

A	B	C	A & B	(A & B) → C
T	T	T	T	T
T	T	F	T	F
T	F	T	F	T
T	F	F	F	T
F	T	T	F	T
F	T	F	F	T
F	F	T	F	T
F	F	F	F	T

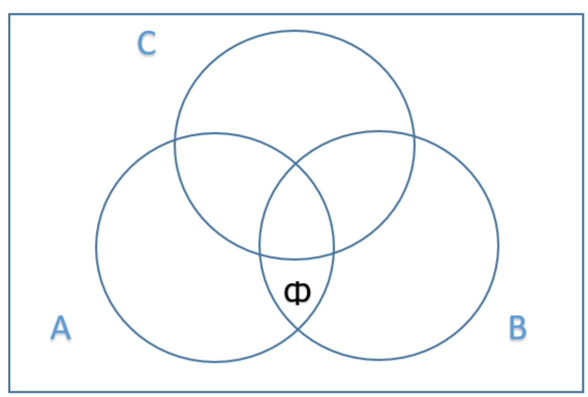


Figure 1

[5] For historical surveys and reviews, see Stanford Encyclopedia of Philosophy, Halpern, and Proceedings of ACM

First, consider the foundational relationship between logic and set theory in Figure 1.

In the propositional calculus, a logical variable 'A' is interpretable either with a binary truth-value A ε {**T**, **F**}, or with an extension [A] ⊆ ***D***. The extensional interpretation directly supports a probability interpretation, by defining a Kolmogorov space (***D***, ***F***, P) where ***F*** is a sigma-algebra generated by the extensions [A_i]. In the case of a finite domain, the normalized counting measure, P[A] = |A| / |***D***|, where |A| indicates cardinality of A, is a natural probability.

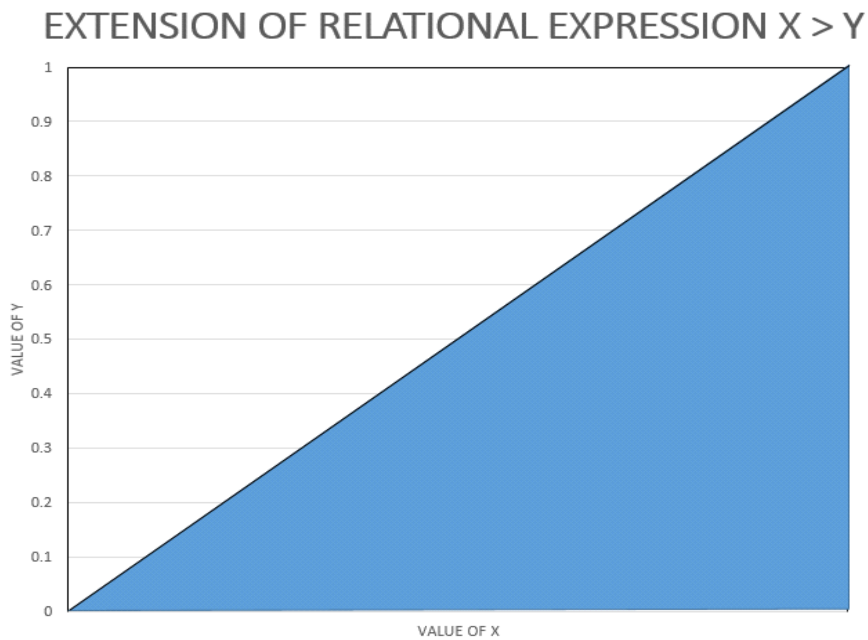


Figure 2

We are all familiar with the extension of quantifier-free relational expressions, ie, open formulas with free variables only[6]. These are common in algebra as graphical expression, and simultaneous equations. Open relational expressions have semantic possibilities similar to propositional expressions – their extensions are subsets of a domain. They can also have binary truth-values on specific tuples, which relate in the usual way to their extension – the set of points where the relational expression is true.

Bound variables, in the usual Tarski semantics of a first order model, do not obviously have any extension at all, except perhaps ***D*** itself or the empty set Ø. This is how the standard model gives truth-values to closed expressions, sentences. They are true exactly when every sequence in the model satisfies the sentence. However, there is an intuitive, natural connection between bound variables and subsets of the domain. A material conditional does not quite express a subset relation, but a quantified material conditional expresses it precisely.

[6] Cf. Monk

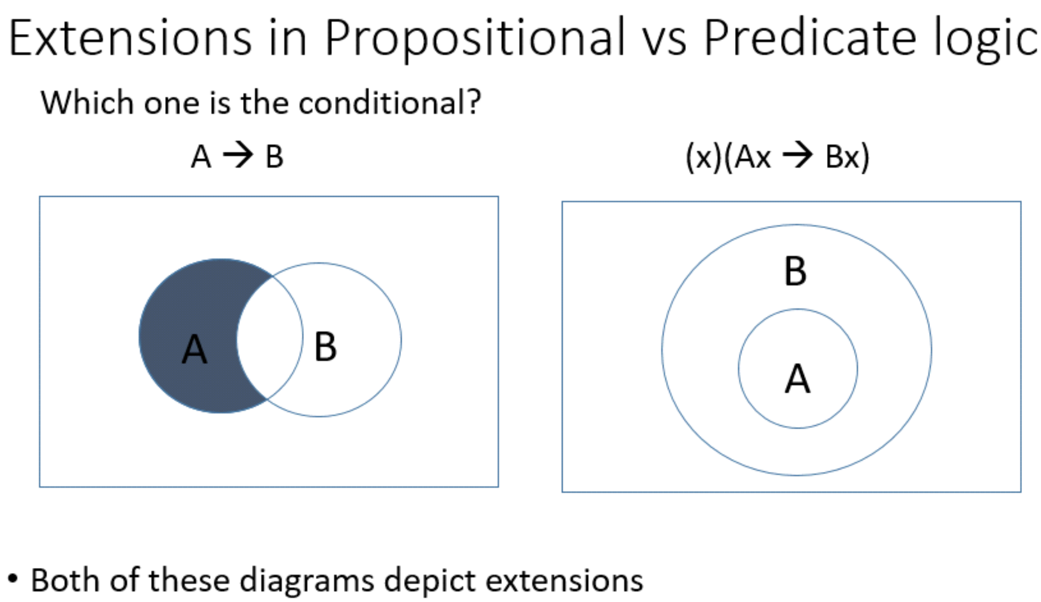


Figure 3

Looking at the Euler diagram on the right, it is clear that the quantified conditional (x)(Ax → Bx) is true whenever [Ax] is *any* subset of [Bx].

It is no accident that predicate calculus is an alternative representation for the basic relations of set theory. First, note that axioms such as comprehension and specification (along with their pitfalls) are as old as Cantor and Frege. These axioms connect the membership of a set to the satisfaction of a logical expression. More to the point is the basic grammar of predication, wherein the formula '***fx***' represents set membership, **x** ε *f*. Meanwhile, the quantified conditional is the essence of the subset relation. All that is missing from first-order logic is a class of symbols to denote subsets.

The remedy is to let a bound variable in a quantified expression denote a subset of the domain. To give a closed first-order sentence an extension, interpret the bound variables as satisfiable not only in relation to the domain ***D*** as a whole, but instead in relation to an arbitrary subset of ***D*** which it denotes – an element in the power set 2^{D}. Take the case of simple closed sentences with only one bound variable, (x)Ax. Analogous to the definition of satisfaction in standard models, take the expression '(x)Ax' to be satisfied by a subset *a* ε 2^{D} when *a* ⊆ [A] – in other words, when every x ε *a* satisfies Ax.

Under this interpretation, a generic closed expression (x)ϕx has the extension [(x)ϕx] = 2^[ϕx]. In other words, quantification is interpreted as exponentiation. If we consider this a "localization" in the sense of our epigraph, then that localization is extrapolated to denotations of subsets of the domain.

IV. Probability of a simple quantified sentence

Before presenting a complete model theory for arbitrary WFFs, consider a simple case of a formula with one quantifier and one variable, (x)(Ax → Bx).

Let ($2^{\boldsymbol{D}}$, ***F***, P) be a probability space, where ***D*** is a non-empty finite set, ***F*** = 2^($2^{\boldsymbol{D}}$), and the P is the normalized counting measure P(A) = |A| / |$2^{\boldsymbol{D}}$| for any A in ***F***. Then using the semantics suggested above, we can calculate the probability of a subset membership assertion as follows:

Calculate extension of [(y)Ay → By] =

= 2^[Ay → By]	Definition of satisfaction
= 2^[~Ay v By]	definition of material conditional
= 2^[~Ay v ByAy]	Tautology to get disjoint union
= 2^(***D*** \ [Ay] U [ByAy])	definition of complement

Calculate measure, |(y)Ay → By| =

= 2^|***D*** \ [Ay] U [ByAy]|	definition of counting measure
= 2^(|***D***| - |Ay| + |ByAy|)	difference and disjoint union
= 2^|***D***| * 2^-|Ay| * 2^|ByAy|	distribution of exponents
= (2^|***D***| * 2^|ByAy|) / 2^|Ay|	rearrange to show divisor

Calculate probability, P((y)Ay → By) =

= **(**2^|***D***| * 2^|ByAy| / 2^|Ay|**)** / 2^|***D***|	normalized counting measure
= 2^|ByAy| / 2^|Ay|	cancel divisor with factor

The calculation above is illustrative of a general pattern for probability in quantified sentences. In a more general WFF, the free variables are interpreted as denoting individual elements of the domain ***D*** while bound variables are interpreted as denoting subsets of ***D***. This treatment will be expanded below.

Next, consider the final expression for P((y)Ay → By) and compare that to the following calculation of the conditional probability P((y)By | (y)Ay) :

P((y)By | (y)Ay) = P((y)ByAy} / P(Ay)	by definition of CP
= 2^|ByAy|*2^-|***D***| / 2^|Ay|*2^-|***D***|	normalized counting measure
= 2^|ByAy| / 2^|Ay|	cancel divisors

Here we have an observation that contravenes a well-established consensus in philosophical logic, that only a trivial probability distribution can allow the probability of a conditional expression to equal its conditional probability. That consensus developed in response to a divergence between the research programs of David Lewis and Robert Stalnaker and has spawned an extensive literature[7]. That literature

[7] An encyclopedic survey of failures is in Khoo and Santorio. For a systematic characterization of possible modal successes, see Bacon.

is outside the scope of the current project, except to note that none of it discusses the probability of expressions in first-order predicate calculus.

The detailed arithmetic above is instructive. The exponential interpretation maps the additive operations of the measure space on ***D***, to multiplication operations in 2^***D***. This matches the arithmetic of the conditional probability formula.

V. Probability Space for Mixed Multivariate Expressions

Probably the closest in spirit to the preset work is the cylindrical algebra of Tarski[8]. These algebraic systems are intended to serve a role corresponding to the Boolean algebras for propositional calculus, including the distinction between the abstract algebra and concrete algebra of sets. The specific contribution of cylindrical algebra is to represent the extension of relational expressions including identity, order, functions, operations, and multivariate polynomials.

The present work differs from a purely cylindrical algebraic treatment. The fundamental distinction between the two approaches is in the preservation or loss of information contained in the bound variables. Simplistically put, cylindrical algebra simply erases the bound variables, while the present treatment strenuously preserves that information.

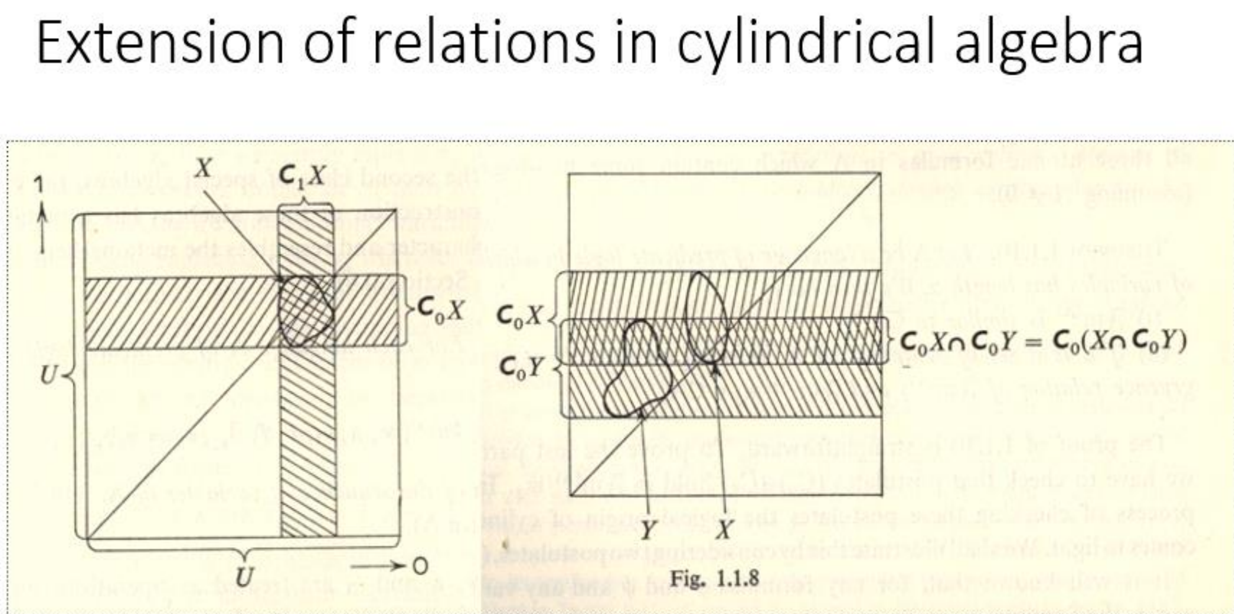


Figure 4-from Henkin, Monk, and Tarski

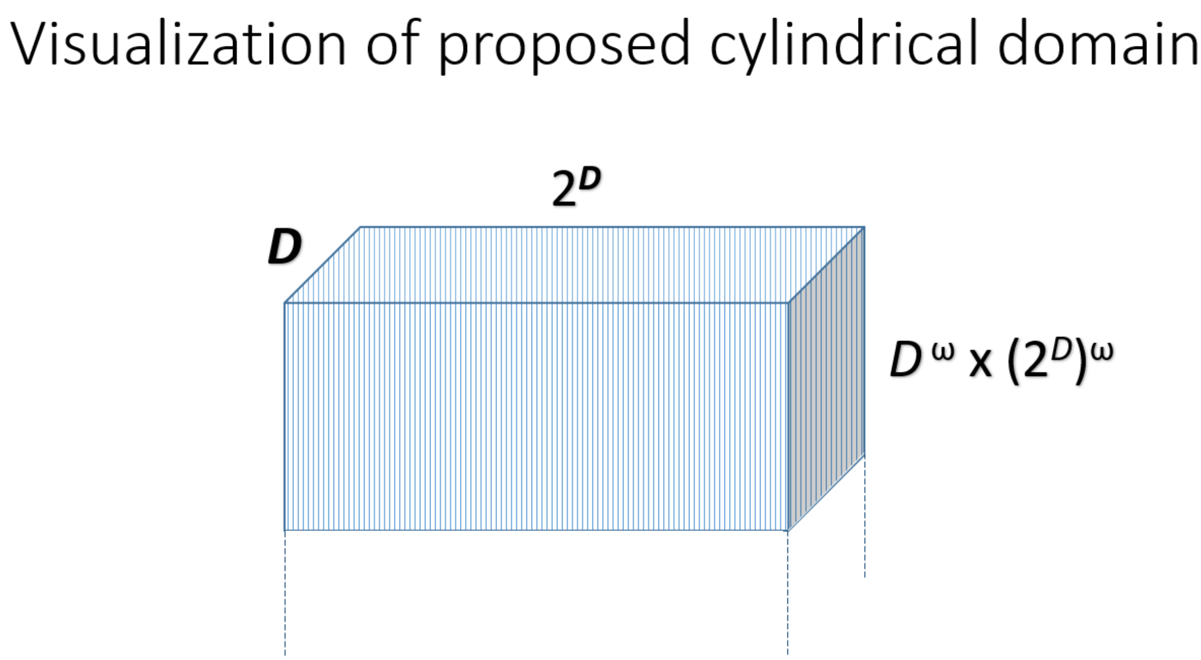
The diagrams in Figure 4 show two interesting facts. First, the cylinder associated with any specific extension 'X' for a certain two-place relation is definable in terms of only one variable – here depicted as '$\boldsymbol{C}_1$X' or '$\boldsymbol{C}_0$X'. Second, the extension of the two-place relation 'X' when projected into either the 0-

---

[8] Cf. Henkin, Monk, and Tarski, "Cylindrical Algebra"

cylinder or the 1-cylinder is indistinguishable from both, the extension of a rectangle bounding the extension of ‘X’, and from a diagonal of that rectangle. Any information about the interior or the boundary of ‘X’ itself is erased by the cylindrification operation.

The present work shares with the cylindrical algebra approach its emphasis on extensions of formulas in a product space. The difference is that bound variables are not removed from the dimensional structure of a formula’s extension. Adding a quantifier formula to a quantifier-free formula does change its extension but does not reduce its dimensionality. Instead, a quantifier converts a variable from denoting individuals within the domain ***D*** to denoting subsets of the domain, elements of $2^{\boldsymbol{D}}$. The semantics proposed here replaces quantifier logic using single-typed variables, with quantifier-free logic having variables of two types.

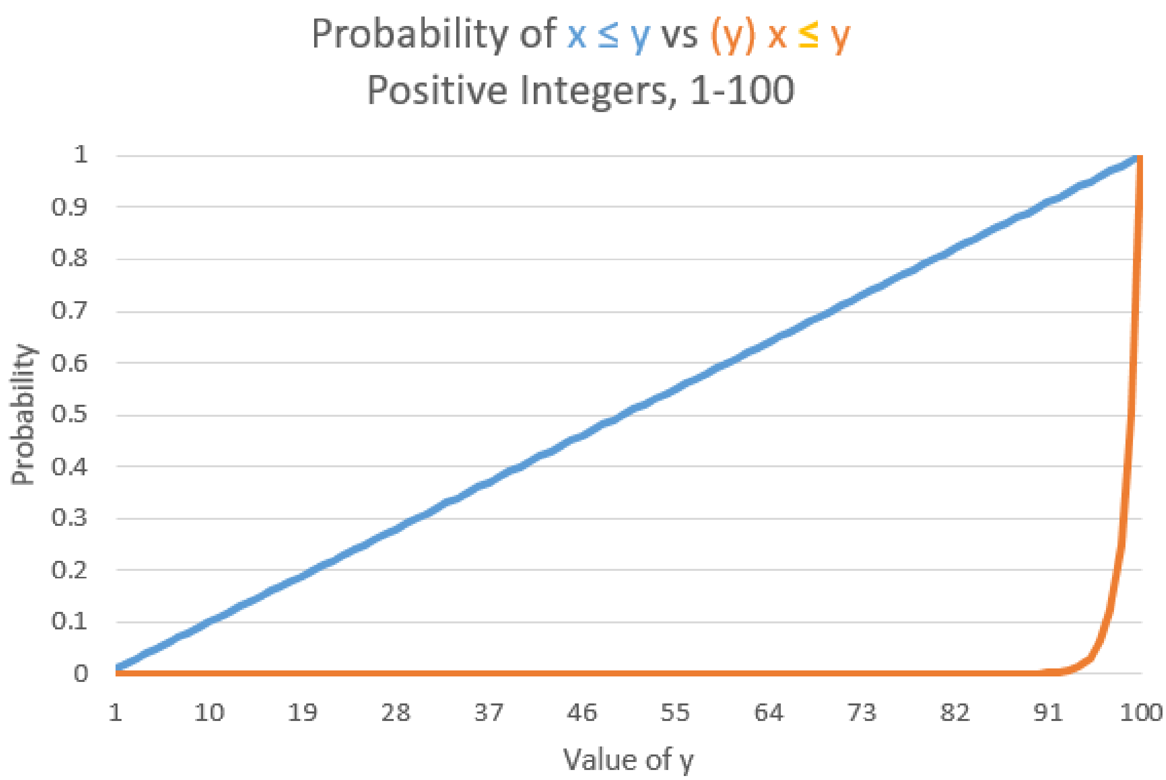


One distinctive aspect of the current proposal is to avoid the commonplace identification of logical truth with probability 1. This may have been natural in the classical theory of finite domains, but it is rather contrary to modern probability theory based on measure. Instead, here the logical concept of truth is intact, especially as applied to the extension of formulas. Every formula is simply true or false for any assignment of values to its variables. The true assignments are the extension of the formula, and the measure of that extension is its probability. For finite domains with counting measure, only the empty set has measure zero, and only the whole domain ***D*** has probability 1, but the distinction remains.

Just as a standard model of first-order logic has a set of sequences of domain elements, an extensional probability model has sequences of ordered pairs. This is (as usual) one place in the sequence for every allowed variable in the language, with the new feature of providing for each variable a subset as its denotation as well as an individual. Note that a variable may appear in a formula both within the scope of a quantifier, and outside such a scope. That variation is handled here exactly as in a standard model – the variable has a different contribution to the satisfaction of the formula in each case.

The case of empty domains has sometimes been controversial in first-order model theory. So-called free logics or inclusive logics may allow for an empty domains or non-referring terms to count as satisfied. In this proposal a leaf is taken from such books. While the domain itself is assumed to be non-empty, the empty subset is taken to satisfy a universally quantified expression such as (x)A. This convention is assumed in the calculation for (x)(Ax Bx) above.

For example, take the expression $(x)\ x \leq y$. In a standard model, this would assert for some individual ‘y’ that it is the maximum element in the domain. This it is has a truth-value, either true or false. But here it asserts that ‘y’ is an upper bound of certain subsets, and thus has an extension of all the subsets containing no elements larger than ‘y’. This is an example of three interesting cases in probability logic. First, of probability for a mixed expression with both free and bound variables. Second, of probability for both relational expressions $x \leq y$ and $(x)\ x \leq y$. Third, of probability for a singular assertion about an individual object – that ’y’ is the maximum of the domain.



There is an important exception to this convention for the identity relation, which is treated as a logical constant with the special property that it cannot be satisfied by an empty set in a bound variable. Thus an expression $(x)\ x = a$, asserting that ‘a’ is the only individual in a subset, is satisfied only by the singleton subset {a}, not by the empty set, and not by any other or larger set. This corresponds to the common mathematical usage in identity theorems that for partial functions f() and g(), the claim “f(x) =

g(x)" asserts that when one value exists, then both do, and the values are equal. It is also necessary for calculations involving the law of total probability, as will be seen below.

To close this section, note the significance of the product measure on the Cartesian product structure in the set of infinite sequences. This is an essential feature in practical calculations of probability. Because the logical expressions are finite and have finitely many variables, each extension has a finite number of "information-bearing" dimensions, and then an infinity of unit factors in its cylindrical extension. This allows every probability calculation to proceed as if the extension has a finite dimensionality. This is the standard practice for calculations within parameter spaces of finite dimension. The infinite sequence domain supports the same calculations with no special qualifications.

VI. Second Example – An Optimal Stopping Problem

This is the Secretary Hiring Problem. The goal is to hire the best secretary from a finite queue of applicants, with no prior knowledge of the range of skills among applicants. The optimal procedure[9] is to fix a certain number of applicants to interview for the purpose of reference only, and thereafter hire the first applicant who is superior to the reference class.

To formalize this problem in predicate calculus, take ***D*** to be a finite set of positive integers, |***D***| = N, and let a function *f*(): ***D*** |→ ***D*** be a permutation, mapping the domain to itself. An argument 'i' of *f*(i) represents the position of an applicant in the queue, while the value of *f*(i) represents the desirability of applicant 'i'.

Formalized in predicate calculus, the problem conditions are:

(1)	(i)(j)(f(i) = f(j) → i = j)	-- f() is injective := Every applicant has a unique value
(2)	r < n	-- r := reference class size, n := successful applicant
(3)	(i)(j) (n < i → r < j .→ f(j) < f(n))	-- defines success := not last → better than ref class
(4)	(i)(r < i & f(n) < f(i) .→ n < i)	-- 'n' is the earliest success

The goal is to maximize the probability of the following expression:

(5) (i)(f(i) ≤ f(n)) -- 'n' is the true max of f(n).

What we see here is that even the most elementary and familiar statistical optimization problems require the calculation of a probability for a quantified first-order expression. It is a mixed blessing to see that our project is not new. To see this, write (5) as:

(6) $\max_{i \in \boldsymbol{D}} f(i) = f(n)$

Which illustrates that optimizing the probability of a max() or min() has *always* depended on assigning probability to first-order expressions. Maximizing a probability or an expectation has been essential to probability theory since the days of Pascal and Fermat. Optimization problems are the soul of both

[9] Cf. Bruss

operations research and game theory, and always depend on quantified relational expression. Optimization of a probability resembling (6) are commonly all through the statistical literature.

Another important difference between past treatments of probability logic vs the present work is that there is no attempt here to add an expression denoting probability to the language. This is a departure from previous work, but it is parallel to almost all discussions of probability for measurable spaces, for propositional logic, for set theory, and for random variables. In these formats, probability expressions appear in a metalanguage, and calculation of the probability is secondary to identifying some other phenomenon – the measure of a set, the distribution of a variable, a degree of belief, or the extension of an expression. That is the method followed here and applied specifically to (5) and (6) – which is characteristic of every contribution to the literature on this problem.

The general strategy of analysis for optimization problems is to obtain a function from one parameter to another. In this case, to compute the probability P(5) as a function of 'r', the size of the reference class. The first step toward that probability is to observe that the assumptions (1)-(4) do define, deterministically, a single valued function 'r' | 'n' for each specific permutation *f*() of the applicant queue. The problem becomes statistical when one must deal with all possible permutations. For example, Rao[10] takes the set of possible permutations to be the co-domain of a random variable. In the present treatment that is not necessary, because the extension of assumption (1) includes precisely all the permutations of the domain, which take either 'i' or 'j' as their dependent variable. The statistical analysis then proceeds, as usual, by partitioning this extension into events B_k:

(7) $B_k := (i)(\, f(i) \leq f(j)\,)$ The maximum of $f(i)$ occurs at $i = j$

The extensions of the formulas $[B_k]$ are mutually disjoint, and their union is the extension of all the permutations *f*() allowed in the expression (1). Now write (5) as 'B_n' – the selected applicant at index 'n' is in fact the best. Then we can apply the law of total probability to say that the probability $P[B_n]$ is the sum:

(8) $P[B_n] = \sum_{k \varepsilon \boldsymbol{D}} P[B_n \mid B_k] * P[B_k]$

Because each $[B_k]$ is equal in number to all the others, and $|\boldsymbol{D}| = N$, the probabilities $P[B_k]$ are all equal for every 'k':

(9) $P[B_k] = 1/N$ and therefore

(10) $P[B_n] = \sum_{k \varepsilon \boldsymbol{D}} P[B_n \mid B_k] * (1/N)$ by total probability as in (8)

Meanwhile, since the applicants are examined in order, and by (4) selection stops with the earliest success, the conditional probability of k=n given any B_k, is a decreasing function of 'k'. The conditional probability needed in (10) is:

(11) $P[B_n \mid B_k] = r / (k - 1)$ for $k > r$

---

[10] M. M. Rao, 1993, Conditional Measures and Applications

Because by (3) B_n cannot be true unless some B_k is true for all $k < r$, Thus

(12) $P[B_n] = \sum_{k>r} (r/(k-1) * (1/N)$ Substitute (11) into (12)

(13) $= (r/N) \sum_{k>r} 1/(k-1)$ Exporting the constants 'r' and 'N'

At this point the familiar expositions all observe that by a change of variables and extrapolating to infinity, (13) becomes:

(14) $P(x) = x \int_x^1 1/t\, dt = -x \ln(x)$

Which has a maximum when x = 1/*e*, the reciprocal of Euler's constant. Note that the role of logic itself in this calculation of probability is complete with (13), both here and in standard treatment of the problem.