Whence the Laws of Probability?

Anthony J.M. Garrett

Byron's Lodge, 63 High Street
Grantchester
Cambridge CB3 9NF
United Kingdom

Abstract

A new derivation is given of the sum and product rules of probability. Probability is treated as a number associated with one binary proposition conditioned on another, so that the Boolean calculus of the propositions induces a calculus for the probabilities. This is the strategy of R.T. Cox (1946), with a refinement: a formula is derived for the probability of the NAND of two propositions in terms of the probabilities of those propositions. Because NAND is a primitive logic operation from which any other can be synthesised, there are no further probabilities that the NAND can depend on. A functional equation is then set up for the relation between the probabilities and is solved. By synthesising the non-primitive operations NOT and AND from NAND the sum and product rules are derived from this one formula, the fundamental 'law of probability'.

1 Introduction: Probability

This paper sets out a novel derivation of the sum and product rules for probabilities – the 'laws of probability' – by deriving them from a single simpler equation, to which they are jointly equivalent; and by deriving that simpler equation from more primitive axioms. In deriving laws of probability from more fundamental ideas we must engage with what 'probability' means. This is a contentious issue; fortunately, if you disagree with the definition proposed here, there is a get-out that allows other definitions to be preserved.

The symbol $p(X|Y)$ denotes the probability that X is true upon supposing that Y is true, where X and Y are statements that can be true or false. More formally, when X and Y are binary propositions that have truth value either TRUE or FALSE, then $p(X|Y)$ denotes the probability that the truth value of proposition X is TRUE, supposing that the truth value of proposition Y is TRUE. Probability is a logical measure of the extent to which truth of X is implied by (supposed) truth of Y; it is an 'if... then...' concept.

The arguments of a probability are truth values of propositions, not propositions themselves. (This is in contrast with utility.) The distinction is easily lost since propositional

logic uses a shorthand whereby the truth value of a proposition X is itself denoted by X. Through the truth values, probabilities depend on the ontological relations between the things to which the propositions refer. For example, if Y is "there is a die with sides labelled '1' to '6'" and X is "the face labelled '2' is uppermost" then $p(X|Y)=1/6$, because it is an ontological assertion that there are six faces, between which Y discriminates solely according to (arbitrary) label; hence their probabilities are equal.

This idea of probability as a degree of partial implication, or *implicability*, is emphasised in the textbook of probability by Keynes (1921), the economist. It dates back to Leibniz, and implicitly to mediaeval law through the extent to which guilt is implied by evidence (Franklin 1991). This view is operationally equivalent to the Bayesian idea that probability is the degree to which it is appropriate to believe that X is true in the light of Y being true – the *believability* of X, supposing Y. The Keynesian view carries less psychological baggage.

It is degree of implication to which the word 'probability' refers throughout this paper. If, however, you have a different idea of what probability means, you can just replace 'probability' by 'implicability' from here on, and reserve 'probability' for whatever use you wish. It is implicability, however named, that is required in any real problem in which 'probability theory' is deployed. For the aim is always to work out how strongly it is implied, from the information available including any experimental data – "the measured value was so-and-so" – that a parameter takes a particular value, or that a hypothesis is true, or whatever. Equally empty in operational terms is debate over the merits of words such as probability, credibility, verisimilitude, likeliness, plausibility, expectation, confidence, surprise and so on.

2 The Single Law of Probability

We shall now find a single equation which implies both the sum and product rules. The product rule is

$$p(AB|I) = p(A|BI)p(B|I) \tag{1}$$

where AB denotes the truth value of the logical product ('AND') of two propositions, defined by its truth table to be true only if both of A and B are true, and false otherwise. The sum rule is

$$p(A|I) + p(\overline{A}|I) = 1, \tag{2}$$

where $\overline{A}$ ('NOT A'; the negation of A) denotes the proposition that is false if A is true, and vice-versa (since $\overline{\overline{A}}=A$). Every logical operation can be synthesised from the logical product together with negation, and both of these operations are needed to do so. Therefore these two rules are necessary and sufficient to decompose the probability of any proposition.

We now use the sum and product rules to find the decomposition of $p(A\uparrow B|I)$, where $A\uparrow B$ is the logic function defined to be false only if A and B are both true, and true otherwise. It may therefore be recognised as the operation NOT AND, or NAND. We seek this particular decomposition because every logical operation can be synthesised from NAND alone: given only enough NAND gates, any logic circuit can be constructed from them. Correspondingly we should be able to derive both of the product and sum rules, and decompose any probability, using the decomposition of $p(A\uparrow B|I)$ alone. NOT and AND are given by

$$\overline{A} \;=\; A\uparrow A$$

$$AB \;=\; \overline{A{\uparrow}B} \;=\; (A{\uparrow}B){\uparrow}(A{\uparrow}B). \tag{3}$$

(A comprehensive reference on Boolean algebra is Kuntzmann, 1967.) From the sum rule

$$p(A{\uparrow}B|I) \;=\; p(\overline{AB}|I) \tag{4}$$

$$\;=\; 1 - p(AB|I) \tag{5}$$

whence, using the product rule, the required decomposition is

$$p(A{\uparrow}B|I) = 1 - p(A|BI)p(B|I). \tag{6}$$

To verify that (6) is more fundamental than either the sum rule or product rule alone, we reverse the analysis and derive both of them from it. We can also recover the fact that the probability of a proposition implied to be true is constant and equal to unity, and the probability of a proposition implied to be false is zero. Begin by putting $B=I$ in (6). Now $A{\uparrow}B=\overline{AB}=\overline{A}+\overline{B}$ by de Morgan's theorem of Boolean algebra, so that the arguments of the probability on the left-hand side become $A{\uparrow}I|I=(\overline{A}+\overline{I})|I=\overline{A}|I$ and (6) reduces to

$$p(\overline{A}|I) = 1 - p(A|I)p(I|I). \tag{7}$$

In this result, send $A \to \overline{A}$ to give

$$p(A|I) = 1 - p(\overline{A}|I)p(I|I) \tag{8}$$

and eliminate $p(\overline{A}|I)$ between these equations. The result is

$$[1 - p(I|I)^2]\, p(A|I) = 1 - p(I|I). \tag{9}$$

Since $p(A|I)$ is free to vary with proposition A, each side of this equation must be zero, so that $p(I|I)=1$ for any proposition I. By substituting this result back into (7) the sum rule (2) follows. Now it is easy to recover the product rule, by sending $A \to AB$ in the sum rule just derived to give

$$p(\overline{AB}|I) = 1 - p(AB|I). \tag{10}$$

Comparison of this with (6) (recall that $A \uparrow B = \overline{AB}$) gives the product rule (1). The product rule shows how to move propositions across the conditioning solidus. Finally, by putting $A=I$ in the sum rule (2) and using $p(I|I)=1$, we have $p(\overline{I}|I)=0$.

The end-point relations $p(I|I) = 1$ and $p(\overline{I}|I) = 0$ generalise in an intuitive way. In $p(I|I)=1$ let $I \to AI$; since $AI|AI=A|AI$ (both sides are true), we have $p(A|AI)=1$. Truth of AI implies truth of A, so that $p(A|AI)$ is the probability of a proposition implied by the conditioning proposition to be true. It is constant, so that we may denote it p_t (for 'true'): $p_t=1$. Also,

$$0 = p(\overline{AI}|AI) = p(\overline{A}+\overline{I}|AI) = p(\overline{A}|AI) \tag{11}$$

since $\overline{I}|AI$ is false. We denote $p(\overline{A}|AI)$ by p_f (for 'false'), so that $p_f=0$.

We have derived the end-point relations

$$p_f = 0\ , \qquad p_t = 1 \tag{12}$$

and the sum and product rules from the single relation

$$p(A{\uparrow}B|I) = 1 - p(A|BI)p(B|I). \tag{13}$$

If, now, we can derive this relation from more basic ideas, we shall have a tighter derivation of the conventional laws of probability. Such a derivation now follows.

3 Deriving the Calculus of Probability from the Calculus of Propositions

A probability is a number associated with a conditioned proposition. Truth values of propositions obey a calculus known as Boolean algebra, familiar in logic circuitry. The Boolean calculus of propositions induces a corresponding calculus for the numbers associated with the propositions, *i.e.* for the probabilities. We shall derive the calculus of probabilities from the calculus of propositions.

This strategy is due to R.T. Cox (1946), who, two centuries after the sum and product rules became commonplace, rederived them by this means and so provided formal justification that 'probability' can be used to mean implicability. The present derivation is a refinement of Cox's.

The derivation begins by proposing a relationship for the decomposition of $p(A\uparrow B|I)$. It is

$$p(A\uparrow B|I) = F(p(A|BI), p(B|I), p(B|AI), p(A|I)). \tag{14}$$

Since the conditioning proposition I represents what is unquestioned, it remains everywhere on the right of the conditioning solidus. (Except where stated, it is supposed that I does not include any logical factor of A, $\overline{A}$, B or $\overline{B}$.) And since, crucially, all logical operations, including NOT and AND, can be synthesised from NAND (defined by its truth table – ignore the etymology), no logical operation need appear in the right-hand side apart from the logical product with I in the conditioning, which is known to be what is required. Hence (14) exhausts the possibilities.

Having justified (14) we proceed to reduce the number of probabilities in the right-hand side. Look first at the special case of (14) obtained when $B=AC$. There is loss of generality in B, which is false if A is false, but no loss of generality in A or C. On the left-hand side $A\uparrow B$ becomes

$$A\uparrow B = A\uparrow(AC) = \overline{AAC} = \overline{AC} = A\uparrow C. \tag{15}$$

The right-hand side simplifies, since

$$p(A|ACI) = p_t, \qquad p(AC|AI) = p(TC|AI) = p(C|AI)\ . \tag{16}$$

Equations (16) follow from the dependence of probabilities specifically on truth values of propositions. We find using (16) that

$$p(A\uparrow C|I) = F(p_t, p(AC|I), p(C|AI), p(A|I)) \tag{17}$$

or, relabelling $C \to B$,

$$p(A\uparrow B|I) = F(p_t, p(AB|I), p(B|AI), p(A|I)). \tag{18}$$

This is to be compared with (14).

Look now at the special case of (14) for which $B=A$: from (3), it reduces to

$$p(\overline{A}|I) = F(p_t, p(A|I), p_t, p(A|I)), \tag{19}$$

which tells us that there is a direct relation connecting the probabilities of truth and falsehood of any proposition:

$$p(\overline{A}|I) = \chi(p(A|I)) \tag{20}$$

where

$$\chi(z) = F(p_t, z, p_t, z). \tag{21}$$

By eliminating $p(\overline{A}|I)$ between (20) and itself with $A \to \overline{A}$ and since $\overline{\overline{A}} = A$, we have (in operator notation) $\chi^2 = \mathcal{I}$, the unit operator; the function χ will be studied later. The next step is to replace A by $A\uparrow B$ in (20), giving, from (3),

$$p(AB|I) = \chi(p(A\uparrow B|I)), \tag{22}$$

which, upon substitution in the right-hand side of (18), yields

$$p(A\uparrow B|I) = F(p_t, \chi(p(A\uparrow B|I)), p(B|AI), p(A|I)). \tag{23}$$

At this point there are two possibilities: that the right-hand side of (23) is identically equal to $p(A\uparrow B|I)$ and has no dependence upon $p(B|AI)$ or $p(A|I)$, so that (23) is tautologous and (18) is merely a disguised form of (22); or that (23) expresses in implicit form a relation between $p(A\uparrow B|I)$, $p(B|AI)$ and $p(A|I)$. In that case we would have succeeded in simplifying (14) to dependence on just two probabilities. To investigate whether this is possible we shall use (14) to eliminate $p(A\uparrow B|I)$ in both sides of (23). Define for convenience the new function $\psi(\ldots) = \chi(F(\ldots))$ and

$$u = p(A|BI), \quad v = p(B|I), \quad x = p(B|AI), \quad y = p(A|I). \tag{24}$$

Then we have

$$p(AB|I) = \psi(u, v, x, y) \tag{25}$$

and, substituting (14) into both sides of (23) and taking χ of both sides, we obtain the functional equation

$$\psi(u, v, x, y) = \psi(p_t, \psi(u,v,x,y), x, y). \tag{26}$$

Inspection reveals that there is indeed a solution for $\psi(u, v, x, y)$ which is an arbitrary function of the last two arguments; this is what we were hoping for. However it is necessary to check for other solutions, and we must therefore solve (26) systematically. We must also be careful, since independence lies not in the four algebraic variables u, v, x, y that represent probabilities, but in the three logical variables A, B, I.

By defining $J(z, x, y) = \psi(p_t, z, x, y)$ we can rewrite this equation, in the spirit of (23), as

$$\psi = J(\psi, x, y). \tag{27}$$

In this form there are three probability variables, ψ, x and y; and, since there are three degrees of freedom (A, B and I), we may take ψ, x and y to vary independently. There are four possibilities: that J is a constant; that it is non-constant and varies with its first argument alone; that it is non-constant and independent of its first argument, varying with the others; or that it definitely varies with both its first and its other arguments. If it varies with its first argument alone, then for (27) to make sense the function $J(\psi', x', y') = \psi'$, where a prime attached to an algebraic variable denotes that it is a dummy and not a shorthand for any probability. This corresponds to the tautologous case in (23), but in (26) it yields a family of solutions for $\psi(u, v, x, y)$ satisfying $\psi(u'=p_t, v', x', y') = v'$. Next, if J is independent of its first argument we have $\psi(u, v, x, y) = f(x, y)$, the solution we anticipated. Since f

is arbitrary this case can be taken to include that of constant J. The remaining case is when J varies with both its first argument and the others, so that (27) expresses a relation among ψ, x and y. This may always be reverted to give ψ as a function of x and y, so that $\psi(u,v,x,y)=f(x,y)$ again; even if the function $J(\psi',x',y')$ is multivalued, this arbitrariness can be absorbed in that of f.

Therefore the solution must be either the one we want, $\psi(u,v,x,y)=f(x,y)$, or a 'tautology' solution satisfying $\psi(u'=p_t,v',x',y')=v'$. To generate further constraints on solutions we look back to where we put $B=AC$ in (14), and instead put $A=BC$. The analysis runs exactly parallel and yields the further functional equation for ψ,

$$\psi(u,v,x,y) = \psi(u,v,p_t,\psi(u,v,x,y)). \tag{28}$$

Correspondingly this has solutions $\psi(u,v,x,y)=f(u,v)$ independent of x and y, and solutions satisfying $\psi(u',v',x'=p_t,y')=y'$. Valid solutions must fall into one or other of these categories. Of the solutions we have already found for (26), the class $f(x,y)$ are clearly not those independent of x and y (unless f is a constant); therefore to be valid they must satisfy $\psi(u',v',x'=p_t,y')=y'$. It remains to distribute those solutions of (26) satisfying $\psi(u'=p_t,v',x',y')=v'$ between these categories; some will satisfy $\psi(u,v,x,y)=f(u,v)$, some will satisfy $\psi(u',v',x'=p_t,y')=y'$ (and some will satisfy neither and be ruled out). Collecting, solutions must be: constant; or of the form $\psi(u,v,x,y)=f(u,v)$ or $f(x,y)$ where $f(p_t,z)=z$; or must satisfy both tautology conditions $\psi(u'=p_t,v',x',y')=v'$ and $\psi(u',v',x'=p_t,y')=y'$. Constant solutions are useless, while no solutions satisfy both tautology conditions, for upon putting u' and x' equal to p_t and equating the resulting expressions for $\psi(p_t,v',p_t,y')$ we obtain $v'=y'$, which is a contradiction since these are independent variables. We are left with only the desired solutions $f(u,v)$ or $f(x,y)$.

Before proceeding, we must settle a question. Since the logical product is commutative we may equate decompositions of $p(AB|I)$ and $p(BA|I)$ to give

$$\psi(u,v,x,y) = \psi(x,y,u,v). \tag{29}$$

Upon substituting our solution $\psi(u,v,x,y)=f(u,v)$ into this equation we find that $f(u,v)=f(x,y)$; were u, v, x and y four independent variables, this would imply that f is just a useless constant. However, we have seen that independence lies in the three logical variables A, B and I, one fewer degree of freedom, with symmetry between A and B. The equation $f(u,v)=f(x,y)$ is properly understood as a relation among the four probabilities u, v, x and y. If symmetry is required we can now rewrite either of our solutions $f(u,v)$, $f(x,y)$ as $\frac{1}{2}(f(u,v)+f(x,y))$, demonstrating their equivalence.

We henceforth work with the solution $\psi(u,v,x,y)=f(u,v)$ or, in probabilities,

$$p(AB|I) = f(p(A|BI),p(B|I)). \tag{30}$$

By substituting this into (22), taking χ of both sides, and using $\chi^2=\mathcal{I}$, it follows that

$$p(A\uparrow B|I) = F(u,v,x,y) = \chi(f(u,v)) \tag{31}$$

which, upon substitution in (21), gives

$$\chi(z) = \chi(f(p_t,z)). \tag{32}$$

Since we have shown that $f(p_t, z) = z$, this does not tell us anything more about the function χ; for that we must travel further. Our two-variable solution f allows us, as we hoped, to write $p(A\uparrow B|I)$ as a function of just $p(A|BI)$ and $p(B|I)$:

$$p(A\uparrow B|I) = F(p(A|BI), p(B|I)) \tag{33}$$

where F refers now to a new function $\chi(f)$ of just two arguments. This relation for NAND is a big simplification over the four arguments in (14).

Finally, it is a useful exercise to verify that if further probabilities, such as $p(AB|I)$, are tacked on as extra arguments of F in (14), the same simplification (33) can be effected.

4 Derivation and Solution of a Functional Equation Relating Probabilities

We now wish to find the particular function F in (33) which expresses the probability of NAND of two propositions in terms of probabilities of those propositions. To do this, we repeat the above synthesis of the probability of the logical product (AND) of two propositions and of the negation of a proposition, using the result we have established, (33). R.T. Cox (1946) has investigated the functional relations for AND and NOT, so that from there on we can borrow his analysis. By putting $B = A$ in (33) we have

$$p(\overline{A}|I) = F(p_t, p(A|I)). \tag{34}$$

The fact that $p(\overline{A}|I)$ depends only on $p(A|I)$ and (hence) vice-versa was an assumption of Cox's, but here it has been derived (originally at (19)). Define for convenience

$$\chi(v) = F(p_t, v) \tag{35}$$

so that

$$p(\overline{A}|I) = \chi(p(A|I)) \tag{36}$$

where χ is the same function defined earlier, so that $\chi^2 = \mathcal{I}$. Any inverse of χ must therefore be equal to χ itself: $\chi = \chi^{-1}$. We require the mapping between the probabilities $p(A|I)$ and $p(\overline{A}|I)$ to be unique, so that χ^{-1} exists and χ is self-inverse. Any function which is its own reflection in the line $y = x$ is self-inverse, and if $S(x, y) = 0$ where S is an exchangeable function admitting an invertible relation between x and y then these are identical self-inverse functions of each other.

For the logical product we use a short cut: NAND is NOT AND, and we use (22) to remove the NOT. Upon defining as before $\psi = \chi(F)$ (only with two arguments), we have by substituting (33) into (22) the decomposition

$$p(AB|I) = \psi(p(A|BI), p(B|I)). \tag{37}$$

This was derived earlier at (30), and was assumed by Cox (1946). We now bring in the main innovation of Cox (1946) and apply (37) twice, to decompose the logical product of three propositions. Partitioning of the logical product can be done in more than one way but, since the logical product is associative, the results must coincide. By equating the decompositions

a functional equation is generated for ψ, and its solution gives information about F. We shall find that the same procedure would not work for a decomposition of $p(A+B|I)$ in terms of $p(A|BI)$ and $p(B|I)$ even though the logical sum is also associative; Cox's assumption of (37) was correct. A more heuristic justification of (37) than the present one, starting from $p(AB|I)$ as a function of the four elemental probabilities, is given by Tribus (1969, p14).

Begin by writing

$$\begin{aligned} p(XYZ|I) &= p((XY)Z|I) && (38)\\ &= \psi(p(XY|ZI), p(Z|I)) && (39)\\ &= \psi\Big(\psi(p(X|YZI), p(Y|ZI)), p(Z|I)\Big). && (40) \end{aligned}$$

Alternatively,

$$\begin{aligned} p(XYZ|I) &= p(X(YZ)|I) && (41)\\ &= \psi(p(X|YZI), p(YZ|I)) && (42)\\ &= \psi\Big(p(X|YZI), \psi(p(Y|ZI), p(Z|I))\Big). && (43) \end{aligned}$$

By equating these two decompositions and defining

$$x = p(X|YZI), \qquad y = p(Y|ZI), \qquad z = p(Z|I), \tag{44}$$

we obtain the functional equation

$$\psi(x, \psi(y, z)) = \psi(\psi(x, y), z) \tag{45}$$

where x, y and z can clearly be taken as independent variables.

Equation (45) is called the *associativity equation* and has been known since Abel studied it in the 19th century (see Aczél, 1966). Its solution is a matter of mathematics, and we quote the result. There is a general solution

$$\psi(u, v) = \phi^{-1}(\phi(u)\phi(v)) \tag{46}$$

(in dummy variables u, v), where ϕ is a continuous function of a single argument and ϕ^{-1} its inverse, which are both supposed to be unique; ϕ is otherwise arbitrary. Though this solution is lengthy to derive it is easy to verify: each side of (45) is equal to $\phi^{-1}(\phi(x)\phi(y)\phi(z))$. The solution (46) was derived in an appendix by Cox (1946) and more neatly in a later exposition (Cox 1961, chapter 3). It is derived using an iterative method, and with a discussion of the conditions of validity of the solution, by Aczél (1966, section 6.2). A summary of the iteration theory for this equation is given by Smith and Erickson (1990).

Equation (45) also has particular solutions

$$\psi(u, v) = f(u), \qquad \psi(u, v) = f(v) \tag{47}$$

where, by direct substitution in (45), $f^2 = f$ (in operator language), so that f is either the identity operator ($f(w) = w$) or a projection which may be discontinuous (an extreme case is the solution $f =$ constant). The identity solution $f(w) = w$ is deducible by Cox's method, although Cox did not state it. Back in (37), the solution $\psi(u, v) = f(v)$ cannot be valid

for arbitrary A. The solution $\psi(u,v)=f(u)$ fails if $A=B$ since the left-hand side then depends on B, which is arbitrary, but the right-hand side does not. For the remaining, more general solution of the associativity equation, (46), the relation (37) may be written, dropping brackets for convenience, as

$$\phi p(AB|I) = \phi p(A|BI)\,.\,\phi p(B|I)\,. \tag{48}$$

This looks uncannily like the product rule; but let us not anticipate.

Now we undo the transformations to recover F and the relation between $p(A\uparrow B|I)$, $p(A|BI)$ and $p(B|I)$. Since $\psi=\chi(F)$ and since $\chi^2=\mathcal{I}$ it follows that $F=\chi(\psi)$, so that from (46)

$$F(u,v) = \chi\phi^{-1}(\phi(u)\phi(v)) \tag{49}$$

where further information is obtained by substituting this result for F into (35), to give

$$\chi(v) = \chi(\phi^{-1}(\phi p_t\,.\,\phi(v))). \tag{50}$$

Since χ possesses an inverse it can be peeled away from either side; it then follows that

$$\phi p_t = 1. \tag{51}$$

The next step is to substitute (49) into (33) to give, upon taking χ of both sides and simplifying,

$$(\phi\chi\phi^{-1})\phi p(A\uparrow B|I) = \phi p(A|BI)\,.\,\phi p(B|I) \tag{52}$$

or, on defining $\Theta\equiv\phi\chi\phi^{-1}$,

$$\Theta\,\phi p(A\uparrow B|I) = \phi p(A|BI)\,.\,\phi p(B|I) \tag{53}$$

where, since $\chi=\phi^{-1}\Theta\phi$ and $\chi^2(v)=v$,

$$\Theta^2(v) = v. \tag{54}$$

It is useful to define

$$p' \equiv \phi p \tag{55}$$

so that $p'_t=1$, and (53) becomes

$$\Theta\,p'(A\uparrow B|I) = p'(A|BI)p'(B|I). \tag{56}$$

The last three equations encapsulate our progress so far. Equation (48) becomes

$$p'(AB|I) = p'(A|BI)\,p'(B|I), \tag{57}$$

and is readily derived by running through the decomposition of the logical product, set out at the start of this section, using (54) and (56) rather than the original equations in which the functions were undetermined. In this process we find a relation for negation by putting $B=A$ in (56):

$$\Theta\,p'(\overline{A}|I) = p'(A|I). \tag{58}$$

This explains why Θ satisfies (54) and must be self-inverse.

For $A=\overline{B}$ equation (57) becomes $p'_f = p'_f p'(B|I)$, which implies that $p'_f = 0$ (or ∞, which we shall demonstrate gives a logically equivalent formalism). Hence we have shown that

$$p'_f = 0 \ , \qquad p'_t = 1 \tag{59}$$

and, by putting $I = ABJ$ and $I = \overline{A}J$ respectively in (56), we find that

$$\Theta(0) = 1 \ , \qquad \Theta(1) = 0. \tag{60}$$

Equations (60) anchor the function Θ at the end points of its argument.

We now use (56) to synthesise a relation for the logical sum (inclusive OR); this will generate an equation for Θ. It is more convenient to synthesise the logical sum from negation and the logical product, using de Morgan's theorem, than to derive it directly from NAND. De Morgan's theorem is

$$\overline{A+B} = \overline{A}\,\overline{B} \tag{61}$$

so that

$$p'(A+B|I) = p'(\overline{\overline{A}\,\overline{B}}|I). \tag{62}$$

Our strategy will be to remove all negations using Θ, as in (58). First,

$$p'(A+B|I) = \Theta(p'(\overline{A}\,\overline{B}|I)). \tag{63}$$

Now decompose this using the 'product rule' (57):

$$\begin{aligned} p'(A+B|I) &= \Theta(p'(\overline{A}|I)p'(\overline{B}|\overline{A}I)) &(64)\\ &= \Theta\Big(\Theta(p'(A|I))\Theta(p'(B|\overline{A}I))\Big). &(65)\end{aligned}$$

The remaining negation in (65) is to the right of the conditioning solidus. To undo it using Θ, it must be moved to the left. This is done by exploiting commutativity of NAND: we equate the decompositions of $\Theta p'(\overline{A}\uparrow B|I)$ and $\Theta p'(B\uparrow\overline{A}|I)$ given by (56), to find

$$p'(B|\overline{A}I) = p'(B|I)p'(\overline{A}|BI)/p'(\overline{A}|I), \tag{66}$$

which is a prototype of Bayes' theorem. When this is substituted into (65) and the further negations are undone using Θ, the result is

$$p'(A+B|I) = \Theta\left(\Theta(p'(A|I))\Theta\Big(\frac{p'(B|I)\Theta(p'(A|BI))}{\Theta(p'(A|I))}\Big)\right) \ . \tag{67}$$

Define now

$$x \equiv p'(A|I) \ , \quad y \equiv p'(B|I) \ , \quad z \equiv p'(AB|I). \tag{68}$$

These are new definitions of x, y, z unrelated to (44). Now (67) becomes

$$p'(A+B|I) = \Theta\left(\Theta(x)\Theta\Big(\frac{y\Theta(\frac{z}{y})}{\Theta(x)}\Big)\right). \tag{69}$$

Since the logical sum is commutative the left-hand side is exchangeable in A and B; but, for arbitrary Θ, the right-hand side is not. If we therefore equate this expression to itself

with A and B – and hence x and y – exchanged, we will gain a new equation which Θ must satisfy. On taking Θ of both sides and using (54), this equation is

$$\Theta(x)\Theta\Big(\frac{y\Theta(\frac{z}{y})}{\Theta(x)}\Big) = \Theta(y)\Theta\Big(\frac{x\Theta(\frac{z}{x})}{\Theta(y)}\Big)\ . \tag{70}$$

This functional equation must be solved for Θ subject to the boundary conditions (60) and also (54).

The solution of this equation subject to these constraints is again a matter of mathematics. In the special case $z=0$ the equation reduces without restriction on x or y to

$$\Theta(x)\Theta\Big(\frac{y}{\Theta(x)}\Big) = \Theta(y)\Theta\Big(\frac{x}{\Theta(y)}\Big) \tag{71}$$

since $\Theta(0)=1$. Redefine $x \to \Theta(x)$ and $y \to \Theta(y)$ and use (54) to obtain

$$x\,\Theta\Big(\frac{\Theta(y)}{x}\Big) = y\,\Theta\Big(\frac{\Theta(x)}{y}\Big)\ , \tag{72}$$

which incorporates (54): to see this, put either of x, y to unity and use the boundary conditions.

Equation (72) was derived from (58) in a different way by Cox (1946), who also solved it. It has been solved without assuming differentiability by Aczél (1963); another method of solution is given by Jaynes (in preparation: see references). The solution is

$$\Theta(u) = (1-u^k)^{1/k} \tag{73}$$

where k is arbitrary but positive (to satisfy the boundary conditions). It is easy to verify that this solution also satisfies the full equation (70): each side is equal to $(1-x^k-y^k+z^k)^{1/k}$. The solution remains real – allowing an ordering as we require – provided that its domain of validity is $[0,1]$, the interval between the boundary conditions. The solution of a functional equation often depends on the domain, which may in turn depend on that solution since in the equation the function may itself appear as an argument. Here the domain is determined along with the solution by the equation and boundary conditions.

The irregular solution $p'_f=\infty$ referred to following equation (58) corresponds to $k<0$ and domain $[1,\infty)$; this solution transforms into the present one under $k \to -k$, or equivalently $p' \to 1/p'$, and it is therefore a matter of convention which to prefer; we stick with the usual one for familiarity.

Upon substituting the solution for Θ into (56) and rearranging, we have

$$p'(A{\uparrow}B|I)^k = 1 - p'(A|BI)^k p'(B|I)^k \tag{74}$$

while (58) becomes

$$p'(A|I)^k + p'(\overline{A}|I)^k = 1 \tag{75}$$

and (69) reduces to

$$p'(A+B|I)^k = p'(A|I)^k + p'(B|I)^k - p'(AB|I)^k. \tag{76}$$

By defining

$$p'' \equiv p'^k = \phi(p)^k \tag{77}$$

these equations become

$$p''(A\uparrow B|I) = 1 - p''(A|BI)p''(B|I) \tag{78}$$

and

$$p''(A|I) + p''(\overline{A}|I) = 1 \tag{79}$$

(which is the sum rule) and

$$p''(A+B|I) = p''(A|I) + p''(B|I) - p''(AB|I), \tag{80}$$

which, in combination with the sum and product rules, tells us that $p''(A+\overline{A}|I)=1$ whatever I has to say about A. Equation (80) is easily derived by running through the derivation of (69) with the given form of Θ. When advance is made in expressing the solution of a functional equation it is often useful to rerun the analysis up to that point with the new version of the solution in place; this trick has has been used more than once in the present analysis. By moving the logical product to the left-hand side, equation (80) has the happy phrasing that the probability of the logical sum of two propositions and the probability of their logical product add up to the sum of probabilities of the propositions. This relation can be used to eliminate logical sums in favour of logical products or vice-versa. Finally, (57) is invariant under power-law transformations, so that

$$p''(AB|I) = p''(A|BI)p''(B|I), \tag{81}$$

which is the product rule. Hence p'' satisfies the laws of probability: the sum and product rules. Both follow from (78).

5 Uniqueness

We have already absorbed arbitrariness of the function ϕ and the index k by defining $p'' \equiv p'^k = \phi(p)^k$. Before discussing the significance of this freedom we must learn whether there is any more arbitrariness in the system. This is done by investigating the invariance properties of equation (78). Suppose that a transformed function $\Lambda(p'')$ satisfies the same equation:

$$\Lambda(p''(A\uparrow B|I)) = 1 - \Lambda(p''(A|BI))\Lambda(p''(B|I)). \tag{82}$$

Then, by eliminating $p''(A\uparrow B|I)$ using (78) and (re)defining $x \equiv p''(A|BI)$, $y \equiv p''(B|I)$,

$$\Lambda(1-xy) = 1 - \Lambda(x)\Lambda(y). \tag{83}$$

By construction $\Lambda(z)=z$ is a solution; but are there others? In (83) put $y=1$ (which is an allowed value) and write xy for x in the result, to give

$$\Lambda(1-xy) = 1 - \Lambda(xy)\Lambda(1). \tag{84}$$

By comparison of the last two equations it follows that

$$\Lambda(1)\Lambda(xy) = \Lambda(x)\Lambda(y) \tag{85}$$

which is a functional equation of Cauchy type. To solve it, differentiate it with respect to y and then put $y=1$, giving

$$\Lambda(1)x\Lambda'(x) = \Lambda'(1)\Lambda(x) \tag{86}$$

(a prime here denotes differentiation), which integrates to give solutions proportional to an arbitrary power of x. (The same solutions are isolated without assuming differentiability in Aczél, 1966, chapter 2.) Back in (83) the only solutions that survive are $\Lambda(z)=z$ and the (useless) constant roots of the quadratic $\Lambda^2+\Lambda-1=0$. Hence we have proved that there is no further arbitrariness.

The choice of ϕ and k remains arbitrary. If p is a numerical representation of the extent to which truth of one proposition is implied on supposing the truth of another, then so also is any monotonic increasing function of p. (Monotonicity is necessary to preserve ordering of probabilities.) We have learned that it is useful to write such a function as ϕ^k, where the function ϕ and the index k reflect the structure of the theory. We shall choose to work directly with p'' (and drop the primes) in order to retain a representation which satisfies the familiar two laws of probability rather than a transformed version of them. That this is possible confirms we may indeed interpret probability as a partial degree of implication. The systematic derivation of these rules from the Boolean calculus of propositions reveals that, in any problem involving logical relations between propositions, they (or their transformed equivalents) are compulsory. Anything inequivalent is wrong.

We mention some specific transformations of the laws. The theory may be re-expressed using percentages, on a scale from 0 to 100, by defining $q=100p$. Next is the transformation $q\propto -\ln p$. The product rule transforms to additive form

$$q(AB|I)=q(B|AI)+q(A|I). \tag{87}$$

Conversely, under $q\propto\exp(-p)$, the sum rule transforms to product form. The product rule and sum rule are named because they involve the product and the sum of probabilities. Since we are free to transform at will, this justification is weak. More descriptive names would be the *conditioning rule* and the *negation rule* (though the product rule at least decomposes the probability of the logical product). Finally, a useful transformation is to the *odds*, defined by

$$q \equiv p/\overline{p} = p/(1-p). \tag{88}$$

Since the sum rule and product rule separately are less general than (78) they may have their own separate invariances, and we have seen that the product rule is invariant under power-law transformations of p. The sum rule is invariant under any transformation of $(p-\frac{1}{2})$ into an odd function of itself; to see this, write $q=p-\frac{1}{2}$ so that the sum rule becomes $q(A|I)=-q(\overline{A}|I)$, and demand that the transformed function $Q(q)$ satisfy the same relation, so that

$$Q(q(A|I)) = -Q(q(\overline{A}|I)) = -Q(-q(A|I)), \tag{89}$$

whence Q is odd.

6 Conclusion

The idea of deriving the calculus of probability from the calculus of propositions, since the arguments of probabilities are propositions, is due to R.T. Cox (1946). The sum and product

rules have been known for centuries to apply to proportions (relative frequencies), but this derivation places them on a different, deeper foundation. Our theory is a generalisation of deductive, Boolean logic, the new ingredient being the notion of the extent to which truth of one proposition is implied by truth of another, and its quantification on a continuous scale.

The present starting point uses NAND to reduce the assumptions in the analysis, since any logic function can be constructed from NAND. Upon extending this idea to probabilities from propositions, the result is the equation

$$p(\overline{AB}|I) = 1 - p(A|BI)p(B|I), \tag{90}$$

where we have now written NAND as NOT AND. From this equation both the sum and product rules can be derived, as in section 2. This equation deserves the title 'the law of probability'. It may also be derived via NOR which, like NAND, is a primitive logical operation from which all others can be derived.

By using (90) we are able to generate relations between probabilities of propositions from the logical relations between the propositions themselves. In combination with techniques for assigning probabilities $p(X|Y)$ from the interrelation of X and Y, this technique provides a complete theory of probability. We have seen from the example of the die in the introductory section that symmetry principles provide a method of assigning probabilities. When that symmetry is broken by 'testable information' – information about a probability distribution, such as "the mean is 3.9" for throws of a standard six-sided die – then a variational principle known as maximum (information) entropy generates the distribution; the entropy functional is Claude Shannon's (1948) and is applied to probability assignment by Jaynes (1983, and in preparation). However probability assignment is not fully understood yet.

We finish by bringing our work closer to application and deriving Bayes' theorem, which through its relation between $p(A|BI)$ and $p(B|AI)$ is the central theorem in tackling inverse problems. Since NAND is commutative we have immediately from (90)

$$p(A|BI)p(B|I) = p(B|AI)p(A|I). \tag{91}$$

When this is added to itself with $A \to \overline{A}$, and combined with the sum rule, the result is

$$p(B|I) = p(A|I)p(B|AI) + p(\overline{A}|I)p(B|\overline{A}I) \tag{92}$$

so that we may rewrite (91) in the form

$$p(A|BI) = K^{-1}p(A|I)p(B|AI), \qquad K = p(A|I)p(B|AI) + p(\overline{A}|I)p(B|\overline{A}I). \tag{93}$$

This is Bayes' theorem.

Consider now an exclusive and exhaustive set of propositions $\{A_i\}$, so that one of the propositions is true and the rest are false. This corresponds to a discrete variable, for example A_i="the ith face of the die shows". Equivalently, the logical product of every pair of propositions is false and their logical sum is true. It now follows from the sum and product rules that $\sum_i p(A_i|I)=1$ and that

$$p(A_i|BI) = K^{-1}p(A_i|I)p(B|A_iI), \qquad K = p(B|I) = \sum_i p(A_i|I)p(B|A_iI). \tag{94}$$

Generalisation to the continuum proceeds routinely by defining propositions such as "the tree is between heights h and $h + dh$", and by defining a probability density function which on multiplication by dh generates the probability of this proposition. We also define densities for continuous variables to the right of the conditioning solidus, although these are not probability densities.

These immediately applicable formulae provide a good place to stop.

Acknowledgement

I thank John Skilling for tightening section 3.

References

Aczél, J. 1963. Remarks on probable inference. *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae, Sectio Mathematica* **6**, 3–11.

Aczél, J. 1966. *Lectures on Functional Equations and Their Applications.* Academic Press, New York, USA.

Cox, R.T. 1946. Probability, frequency and reasonable expectation. *American Journal of Physics* **14**, 1–13.

Cox, R.T. 1961. *The Algebra of Probable Inference.* Johns Hopkins Press, Baltimore, Maryland, USA.

Franklin, J. 1991. The ancient legal sources of seventeenth-century probability. In: *The Uses of Antiquity*, editor S. Gaukroger, Kluwer, Dordrecht, Netherlands, pp.123–144.

Jaynes, E.T. 1983. *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics.* Synthese Library **158**. Editor R.D. Rosenkrantz, Reidel, Dordrecht, Netherlands.

Jaynes, E.T. In preparation. *Probability Theory: The Logic of Science.* Cambridge University Press, Cambridge, UK. Provisional versions available on the World Wide Web at `http://bayes.wustl.edu/`

Keynes, J.M. 1921. *A Treatise on Probability.* Macmillan, London, UK.

Kuntzmann, J. 1967. *Fundamental Boolean Algebra* (English translation). Blackie, London, UK.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–423 & 623–659. Reprinted in: *The Mathematical Theory of Communication*, editors C.E. Shannon & W.W. Weaver, University of Illinois Press, Urbana, Illinois, USA, 1949.

Smith, C.R. & Erickson, G.J. 1990. Probability theory and the associativity equation. In: *Maximum Entropy and Bayesian Methods, Dartmouth, USA, 1989*, editor P.F. Fougère, Kluwer, Dordrecht, Netherlands, pp.17–30.

Tribus, M. 1969. *Rational Descriptions, Decisions and Designs.* Pergamon Press, New York, USA.