Inference and probability

Sourendu Gupta

TIFR, Mumbai, India

Methods in Science 2013
March 19, 2013

Outline

Logical: the calculus of certainty

Probability: the calculus of uncertainty

- Setting up the logic
- Conjunction of statements
- Negations of statements
- Simple applications of probability

Keywords and References

Keywords

logic, syllogism, Boolean algebra, uncertain propositions, plausibility, information and implication, common sense, consistency, fairness, equivalence, probability, Bayesian reasoning, jumping to conclusions, Bernoulli's urn, randomness

Books

Probability Theory, by E. T. Jaynes

The one rule: syllogism

Proposition	A	It is sunny outside
Proposition	B	It is day
Relation	A⇒B	If A then B
Necessary	Syllogism	A is true; B is true
Impossible	Syllogism	A is false; B is true
Necessary	Syllogism	B is false; A is false
Possible	Syllogism	B is true; A is false
Plausible	Weak syllogism	B is true; A is plausible
Plausible	Weak syllogism	A is false; B is less plausible

The one rule: syllogism

Proposition	A	It is raining
Proposition	B	It was cloudy earlier
Relation	A⇒B	If A then B
Necessary	Syllogism	A is true; B is true
Impossible	Syllogism	A is false; B is true
Necessary	Syllogism	B is false; A is false
Possible	Syllogism	B is true; A is false
Plausible	Weak syllogism	B is true; A is plausible
Plausible	Weak syllogism	A is false; B is less plausible

The one rule: syllogism

Proposition	A	It is raining
Proposition	B	It was cloudy earlier
Relation	A⇒B	If A then B
Necessary	Syllogism	A is true; B is true
Impossible	Syllogism	A is false; B is true
Necessary	Syllogism	B is false; A is false
Possible	Syllogism	B is true; A is false
Plausible	Weak syllogism	B is true; A is plausible
Plausible	Weak syllogism	A is false; B is less plausible

Logic is not about causal relations, but science is.

Boolean logic

1. A, B, C, etc stand for propositions. They can be true or false.
2. We say $A, B, C, \cdots \in 0, 1$.
3. There is a unary relation of negation: $\overline{A}$.
4. Equality of propositions, $A = B$, means that the values of the propositions are equal; both are true or both are false.
5. There are two more binary relations; A or B is written $A + B$, whereas A and B is written AB.

Idempotence	$A = A + A$, $A = AA$
Commutativity	$A + B = B + A$, $AB = BA$
Associativity	$(A + B) + C = A + (B + C)$, $(AB)C = A(BC)$
Distributivity	$A(B + C) = AB + AC$, $A + BC = (A + B)(A + C)$
Duality	$\overline{(A + B)} = \overline{A}\,\overline{B}$, $\overline{(AB)} = \overline{A} + \overline{B}$

Logical operations

- We have already met the binary relation $A \Rightarrow B$. This is false if A is true and B is false, and true for all other values of A and B.
- Are the operations of and and or completely independent?
- How many different binary operations can one have?
- Can some of these binary operations be written in terms of others?
- How many n-ary operations can one have?
- How many of these are independent?
- How many operations are necessary in order to write a computer program?
- If we decide to build a trinary logic, then how many unary operations can we have? How many binary operations? How many are independent? Can we reduce this to binary logic?

Plausibility

Rule 1

Degrees of plausibility of the truth of propositions are represented as real numbers.

Some conventions make it easier to manipulate plausibilities. First we will assume that greater plausibility corresponds to a larger number. Second we will assume continuity in the form that an infinitesimally greater plausibility corresponds to an infinitesimally larger number.

We will take into account that the plausibility of A may depend on another proposition B. The plausibility of A given that B is true is denoted by the symbol

$$A|B$$

Plausibility

Rule 1

Degrees of plausibility of the truth of propositions are represented as real numbers.

Some conventions make it easier to manipulate plausibilities. First we will assume that greater plausibility corresponds to a larger number. Second we will assume continuity in the form that an infinitesimally greater plausibility corresponds to an infinitesimally larger number.

We will take into account that the plausibility of A may depend on another proposition B. The plausibility of A given that B is true is denoted by the symbol

$$(A|B)$$

Plausibility

Rule 1

Degrees of plausibility of the truth of propositions are represented as real numbers.

Some conventions make it easier to manipulate plausibilities. First we will assume that greater plausibility corresponds to a larger number. Second we will assume continuity in the form that an infinitesimally greater plausibility corresponds to an infinitesimally larger number.

We will take into account that the plausibility of A may depend on another proposition B. The plausibility of A given that B is true is denoted by the symbol

$$(A|B)$$

Standard Boolean functions will yield meanings for $(A|BC)$ or $(A+B|C)$ and so on.

Physics and information

There is a causal relation between it being day and the day being sunny. However, the causal relation cannot be turned into a logical implication without further information. Logical implication in the other direction is straightforward.

Logic deals with relations between propositions based on knowledge as stated, *i.e.*, on given information. Plausibilities are an extension of logic: they deal with relations between propositions. These relations measure the state of our knowledge, and not the actual properties of systems.

How often does a tossed coin land on its tail? The answer depends on the state of our knowledge about the mechanics of the toss. The frequency with which a coin falls on its tail is not intrinsic to the coin.

Common sense

We have chosen the convention that if A makes B more plausible than C, then

$$(B|A) > (C|A).$$

Rule 2

Common sense must not be violated.

If the plausibility of A is increased if the information C changes to C', then

$$(A|C) < (A|C') \qquad \text{and} \qquad (\overline{A}|C) > (\overline{A}|C').$$

If the plausibility of B given A is not changed by this, then

$$(B|AC) = (B|AC') \qquad \text{and} \qquad (AB|C) < (AB|C').$$

Consistency and all that

Rules 3,4,5

- Consistency: if a conclusion can be reached in more than one way, then each way must give the same plausibility.
- Fairness: rules of combining propositions cannot arbitrarily ignore any evidence.
- Equivalence: if in two problems the state of knowledge is the same, then the same plausibility must be assigned to these outcomes.

One can determine $(AB|C)$ in two ways. First by determining $(B|C)$ followed by $(A|BC)$. Second by determining $(A|C)$ followed by $(B|AC)$. These two ways must be consistent. In other words,

$$(AB|C) = F[(B|C),(A|BC)] = F[(A|C),(B|AC)].$$

One of the big problems in setting up the theory is to find the functional form of F. We do this next.

A precise definition of common sense

The common sense rule says that if the prior C changes to C' so that $(B|C') > (B|C)$ and $(A|BC') = (A|BC)$, then $(AB|C') \geq (AB|C)$. The equality occurs only of AB is impossible. Then with the previous notation $(AB|C) = F[(B|C),(A|BC)]$, we find that

$$F_1(x,y) \equiv \frac{\partial F}{\partial x} \geq 0.$$

The common sense rule also says that if the prior C changes to C'' so that $(B|C'') = (B|C)$, but $(A|BC'') > (A|BC)$ then $(AB|C'') \geq (AB|C)$. As a result,

$$F_2(x,y) \equiv \frac{\partial F}{\partial y} \geq 0.$$

Using the associative law

Let us expand our search by trying to evaluate $(ABC|D)$. Now, we can find this by compounding in two ways, corresponding to associativity. Consistency then gives a restriction on F:

$$F[F(x,y),z] = F[x,F(y,z)].$$

One solution is obvious, $F(x,y) =$ constant. However, it assigns equal plausibility to certainty and impossibility; also, it never satisfies the inequalities imposed by common sense.

Define $u = F(x,y)$ and $v = F(y,z)$, so that the associativity equation becomes $F(u,z) = F(x,v)$. Imposing common sense and the chain rule gives

$$F_1(x,v) = F_1(u,z)F_1(x,y),$$
$$F_2(x,v)F_1(y,z) = F_1(u,z)F_2(x,y),$$
$$F_2(x,v)F_2(y,z) = F_2(u,z).$$

Simplifying the problem

Dividing the second equation by the first, and writing $G = F_2/F_1$ gives

$$G(x,v)F_1(y,z) = G(x,y) \quad \text{and} \quad G(x,v)F_2(y,z) = G(x,y)G(y,z).$$

The second equation is obtained by multiplying both sides of the first by $G(y,z)$. We find

$$\begin{aligned} \frac{\partial}{dz} G(x,v)F_1(y,z) &= F_2(y,z)\frac{\partial G(x,v)}{\partial v}F_1(y,z) + G(x,v)F_{12}(y,z) \\ \frac{\partial}{dy} G(x,v)F_2(y,z) &= F_1(y,z)\frac{\partial G(x,v)}{\partial v}F_2(y,z) + G(x,v)F_{12}(y,z). \end{aligned}$$

Now the derivative in the first line vanishes, since the right hand of the first equation has no dependence on z. So the derivative in the second line must also vanish, showing that $G(x,y)G(y,z)$ is independent of y.

Almost the end

As a result, we must be able to write $G(x, y) = rH(x)/H(y)$ for constant r. Since F was shown to be monotonic, we must have $r > 0$ and H must be of fixed sign. The two equations give $F_1(y, z) = H(v)/H(y)$ and $F_2(y, z) = rH(v)/H(z)$. As a result, the relation $dv = F_1 dy + F_2 dz$ gives the differential equation

$$\frac{dv}{H(v)} = \frac{dy}{H(y)} + r\frac{dz}{H(z)}.$$

This gives almost the complete solution in the form

$$\begin{aligned} p[F(y,z)] &= p(v) = p(y)[p(z)]^r \\ \text{where} \quad & p(x) = \exp\left[\int^x \frac{dx}{H(x)}\right]. \end{aligned}$$

Here p is increasing or decreasing according to the sign of H. The indeterminate lower limit means that up to now p has an overall freedom of normalization.

Completing the solution

The associativity equation says

$$p(x)[p(y)]^r[p(z)]^r = p(x)[p(y)]^r[p(z)]^{r^2}.$$

This implies that $r = 1$. So, the functional equation for F reduced to the product rule

$$p(AB|C) = p(A|BC)p(B|C) = p(B|AC)p(A|C).$$

In the special case when $C \Rightarrow A$, it is clear that $(AB|C) = (B|C)$. Also, clearly $(A|BC) = (A|C)$, as long as $B \equiv \overline{C}$. So, the first equality in the product rule states $p(B|C) = p(A|C)p(B|C)$, so

$$p(1) = 1.$$

On the other hand, if $C \Rightarrow \overline{A}$, then $(AB|C) = (A|C) = (A|BC)$. Then $p(A|C) = p(A|C)p(B|C)$, independent of $(B|C)$. This implies either $p(A|C) = 0$ or ∞. It is sufficient to choose

$$p(0) = 0.$$

Setting up

Given B, the more plausible A becomes, the less plausible is $\overline{A}$. Writing $u = p(A|B)$ and $v = p(\overline{A}|B)$, there must be a relation

$$v = S(u) \quad \text{with} \quad S(1) = 0,\ S(0) = 1 \quad \text{and} \quad 0 \le S(u) \le 1.$$

In order to find S, we start from the product rule written using S: $p(AB|C) = p(A|C)S[p(\overline{B}|AC)]$. Again from a product rule, we find $p(\overline{B}|AC) = p(A\overline{B}|C)/p(A|C)$, and we can substitute this into the argument of S. We can do the same again after commuting A and B in $(AB|C)$. This gives the identity

$$p(A|C)S\left[\frac{p(A\overline{B}|C)}{p(A|C)}\right] = p(B|C)S\left[\frac{p(B\overline{A}|C)}{p(B|C)}\right].$$

We will use this to set up a functional equation for S.

A functional equation

Examine a compound statement $\overline{B} = AD$, for some new D. By idempotency $A\overline{B} = \overline{B}$. Also $B\overline{A} = \overline{A}$. The previous identity must be true also for this B. So we get

$$p(A|C)S\left[\frac{p(\overline{B}|C)}{p(A|C)}\right] = p(B|C)S\left[\frac{p(\overline{A}|C)}{p(B|C)}\right].$$

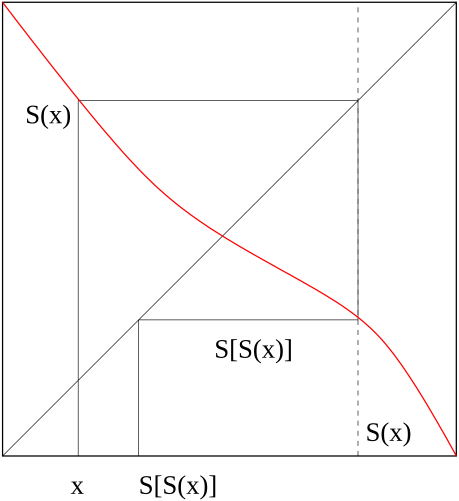
Writing $x = p(A|C)$ and $y = p(B|C)$, this gives us the functional equation

$$xS\left[\frac{S(y)}{x}\right] = yS\left[\frac{S(x)}{y}\right].$$

If $y = 1$ then $S[S[x]] = x$, so S is its own inverse.
Examine the functional equation for $y = S(x) + \epsilon$. Expanding both sides to order ϵ gives $S'(x) = -1$. Which yields $S(x) = 1 - x$. The solution of the equation is the sum rule

$$p(A|B) + p(\overline{A}|B) = 1.$$

The functional equation



Probability as good as plausibility

We have mapped plausibilities of statements, $(A|B)$, into a one-to-one function $p(A|B)$ which we may call a probability. Since every statement about plausibility can be converted into a statement of probability, it is sufficient to work with the latter. The product rule is called Bayes' theorem,

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)}.$$

Using probabilities, we have a very useful form of the sum rule:

$$\begin{aligned} p(A+B|C) &= 1 - p(\overline{AB}|C) = 1 - p(\overline{A}|C)p(\overline{B}|\overline{A}C) \\ &= 1 - p(\overline{A}|C)[1 - p(B|\overline{A}C)] = p(A|C) + p(\overline{A}B|C) \\ &= p(A|C) + p(B|C)[1 - p(A|BC)] \\ &= p(A|C) + p(B|C) - p(AB|C). \end{aligned}$$

These two rules suffice to examine all questions about probabilities.