

FOUNDATIONS?

John Skilling

Maximum Entropy Data Consultants Ltd
Kenmare, County Kerry, Ireland
(e-mail: skilling@eircom.net)

Abstract

The deeper we delve into the foundations of probability theory, the clearer and simpler it gets.

Following the approach of Richard Cox (1946), the basic structure we work with is a Boolean lattice of propositions (*aka* hypotheses). In the world of finite reasoning, the lattice is always finite, so we can avoid the infinite. “Arbitrarily large” is good enough for us. For an acceptable calculus of plausibility, we need to adhere to the symmetries of the lattice, which imply the rules of proportion known in inference as probability calculus. There is no alternative: “you *must* do it this way”. More precisely, any alternative will fall foul of symmetry arguments so compelling as to be convincing counter examples: “you are advised to do it this way, *or else!*”. Weaker justifications are not needed.

Beneath the notion of probability $P(x | t)$ lies a simpler measure $m(x)$ in terms of which

$$P(x | t) = \frac{m(x \wedge t)}{m(t)}$$

The frequentists of old would recognise this, but we know to avoid the intellectual confusion that arose when m was required to be an actual count. Regardless of interpretation, this definition leads to the ordinary sum and product rules of probability calculus, within which we do all our inference.

The search for a variational principle beneath the calculus leads to the entropy

$$S(\tilde{m} | m) = \sum_a \left(\tilde{m}(a) - m(a) - \tilde{m}(a) \log \frac{\tilde{m}(a)}{m(a)} \right)$$

for a measure \tilde{m} constrained somehow away from the primary measure m . As applied to probabilities, the entropy reduces to the standard $-\sum P \log(P/\pi)$ form, which is the directed distance from starting distribution π to final distribution P .

It's all very simple.