

#Paradoxes of probability theory

← Back to Chapters

### Comments.

Here is a mathematical take: Since no contradiction in mainstream mathematics has been found yet, there are no actual paradoxes. There are two types of things that get called “paradoxes”: 1) unintuitive results (a la Banach-Tarski theorem) and 2) fallacious arguments, resulting in incorrect conclusions.

The “paradoxes” of the first type are useful in order to refine one’s intuition: they shed light on the framework one is operating in (in the case of Banach-Tarski, roughly speaking, measure theory), elucidating this framework’s strengths or weaknesses (in the case of Banach-Tarski, restriction to measurable sets). One is then free to reject the framework or to keep it – augmented with appropriate warning labels about the range of its validity.

The “paradoxes” of the second type (like inappropriately “summing” infinite sequences to prove  $0 = 1$ ) are useful in highlighting non-obvious errors and, hopefully, preventing one from committing similar errors later. Thus, one learns not to do unjustified things. One then may also learn, separately, how to do things in a more appropriate way. In the specific example of “summing infinite sequences” we say that one learns mathematical analysis – a more appropriate and contradiction-free (as far as we know) framework. Failure to use analysis properly does not constitute an indictment of methods of analysis – an ironic parallel with the statements Jaynes makes about critics of Bayesianism.

### Comments on 15.3 – 15.6.

What Jaynes is arguing, in effect, is that finitely-additive axiomatization is too weak, as manifest by the fact that it allows pathologies like “nonconglomerability”. This means that some other axiomatization is needed; the mainstream mathematicians of today seem to agree, as it is the countably-additive axioms that are universally taught to students.

It is a curious linguistic phenomenon: who is “more careful” 1) the mathematician who works with weaker axioms, so as to not put in more assumptions than they deem justified OR 2) the mathematician works with more restrictive axioms, so as to avoid “pathological” examples allowed by the more permissive axioms. I’ll leave the judgement up to linguists (though Jaynes seems to have a particular position).

Indeed, “nonconglomerability” is a phenomenon that does not appear in mainstream probability theory – the paper of Kadane, Schervish and Seidenfeld, to which this section addresses itself, is explicitly set in the context of finitely-additive, but not countably additive, probability. In fact, one of the points of the KSS paper is to advocate for finitely-additive measures (despite their

“pathologies” like nonconglomerability) as good fit for Bayesian framework (in particular for their ability to handle improper non-informative priors).

One way of reading section 15.3 is to see it as an argument against both the finitely-additive and countably-additive approaches, pointing out that some alternative, third way of thinking about probability is needed, perhaps more directly based on the “taking the limit in the end” idea. Coming up with a rigorous (which is to say, fully specified) version of such an alternative theory is an interesting, but unsolved and probably difficult question.

### Exercise 15.1

Let  $N(l, a)$  be the number of sequences of results producing a specific record  $x$  of length  $l$  in with  $a$  annihilations (that is, in  $l + 2a$  tosses). We will find a recursion relation for  $N(l, a)$  (which will also establish that this number depends on  $x$  only via  $l$ , and so is well-defined). Now, clearly,

$$N(l, 0) = 1.$$

Observe that  $N(1, a)$  is independent of the one character  $x = \alpha$  – given any sequence generating some other  $\hat{x} = \hat{\alpha}$  we can relabel the characters to get a sequence generating  $\alpha$ , and vice versa, establishing a bijection between the sequences generating  $\alpha$  and  $\hat{\alpha}$ .

(In group theoretic language, we are studying paths in the Cayley graph of the free group on 2 generators  $e, \mu$ ; we will call all four of the  $e = e^+, e^{-1} = e^-, \mu = \mu^+, \mu^{-1} = \mu^-$  “generators”. The group admits an automorphism sending any generator  $\alpha$  to any other generator  $\hat{\alpha}$ ; this automorphism produces an automorphism of the Cayley graph and hence bijections on paths from the identity to  $\alpha$  and paths from identity to  $\hat{\alpha}$ . )

Now, observe that

$$N(0, a) = 4N(1, a - 1),$$

because the way to obtain empty word in the end is to start with some symbol  $\alpha$  (4 options), and then produce a word that is equal to  $\alpha^{-1}$ , which one can do in  $N(1, a - 1)$  ways.

Finally, for all  $l \geq 1$  one we have

$$N(l, a) = N(l - 1, a) + 3N(l + 1, a - 1),$$

because the first toss either produces the first symbol in  $x$  – after which one has to generate the rest of the symbols of  $x$  with  $a$  annihilations; or it produces one of 3 other symbols, after which one has to produce the inverse of that symbol and then  $x$ , with  $a - 1$  annihilations (one annihilation being taken up by cancelling the first tossed symbol with its inverse).

Thus we have our recurrence relation and boundary conditions:

$$N(l, 0) = 1$$

$$N(0, a) = 4N(1, a - 1)$$

$$N(l, a) = N(l - 1, a) + 3N(l + 1, a - 1)$$

The problem asks for  $N(20, 10)$ . A small dynamic programming script gives  $38192689856872 \approx 0.4 \times 10^{14}$ .