

What Conditional Probability Also Could Not Be

(Reply to Kenny Easwaran)

Alan Hájek

1. The story so far

In “What Conditional Probability Could Not Be”, I argued for several theses concerning conditional probability. The main ones were:

1. Kolmogorov’s ratio formula, $P(A \mid B) = P(A \ \& \ B)/P(B)$ (provided $P(B) > 0$) should not be regarded as a stipulative definition of conditional probability, but rather as an *analysis* of that notion. I called it ‘the ratio analysis of conditional probability’.

2. The ratio analysis is inadequate. I argued that conditional probabilities can be well defined, and can have intuitively obvious values, in cases where the ratio analysis goes silent. The main cases I adduced were ones in which:

- i) $P(B) = 0$, what I called *the problem of the zero denominator*.
- ii) $P(A \ \& \ B)$ and $P(B)$ are vague.
- iii) $P(A \ \& \ B)$ and $P(B)$ are undefined.

The problem of the zero denominator is well-known to mathematicians and statisticians, although less known to philosophers. Part of my mission, then, was to draw philosophers’ attention to it, and also to give new kinds of examples of it. I went on to discuss what I called “Kolmogorov’s elaboration of the ratio formula”, his more sophisticated treatment of probability conditional on the outcome of a random experiment. I appealed to this treatment to vindicate intuitions that various conditional

probabilities, conditional on probability zero antecedents, could be well defined—for example, those that we find in the Borel paradox.

At this point, you might have the reaction: “All is well with conditional probability, then—philosophers just need to do their homework and catch up with the state of the art.” To be sure, various well-known philosophical tomes on probability, by Fine, Skyrms, Earman, Howson and Urbach, etc. make no mention of Kolmogorov’s elaboration. (Teddy Seidenfeld is a notable exception among philosophers in this regard.) But I went on to argue that there were further problems for the ratio analysis that were also problems for Kolmogorov’s elaboration. These were cases of type ii) and type iii). I offered various problematic examples from science and philosophy, especially of type iii) cases.

My punch line at this point was negative: the ratio analysis (and even the elaboration, although that was less my concern) is inadequate. I concluded more positively:

3. Conditional probability should be taken as the primitive notion in probability theory. Here I advocated a return to the views of some of the founding fathers in the philosophy of probability: among others, Johnson (1921), Keynes (1921), Carnap (1952), Jeffreys (1961), and de Finetti (1974, reprinted 1990). I thus favored following Popper, Renyi, Roeper and Leblanc, van Fraassen and others in axiomatizing conditional probability directly.

In his impressive paper, Kenny Easwaran grants my negative arguments against the ratio analysis. But he is less happy with my positive conclusion. Primitive conditional probabilities, he argues, will not rid us of various problems associated with zero probability antecedents. Instead, he favors adopting what he calls “Kolmogorov’s extended analysis” (what I called “Kolmogorov’s elaboration of the

ratio formula”). I will respond to Kenny’s arguments, conceding to him that primitive conditional probabilities are not the panacea that I took them to be; but I will conclude that Kolmogorov’s extended analysis is not the panacea that Kenny takes it to be.

2. The problem of the zero denominator, revisited

Central to Kenny’s paper is the problem of the zero denominator, so we will do well to remind ourselves what the problem is. It arises from the conjunction of two facts. The first fact, familiar from elementary school arithmetic, is that you can’t divide by 0. The second fact is that contingent propositions may be assigned probability 0—in a slogan, ‘probability 0 does not imply impossible’. Put these facts together, and you have the recipe for trouble: by the second fact, it should be legitimate to form conditional probabilities with probability zero conditions; but by the first fact, these would be undefined if calculated by the ratio formula. This trouble might be dismissed as a ‘don’t care’, a minor curiosity at the edge of probability theory. I argued that we should care about such cases, summoning various intuitions about the values of conditional probabilities that seemed non-negotiable to me.

Kenny writes that “the intuitions Hájek appeals to are unreliable” (4), and “[h]is intuitions lead us into trouble if we try to work with cases much more complicated than the ones he discusses.” (9) For the record, here is a complete list of the intuitions that I called upon in my discussion of the problem of the zero denominator. For any Z that is possible, but that has probability 0:

(1) $P(Z, \text{ given } Z) = 1.$

(2) $P(Z^c, \text{ given } Z) = 0.$

(3) $P(T, \text{ given } Z) = 1,$ where T is a necessarily true proposition (e.g., “everything is self-identical”).

- (4) $P(F, \text{ given } Z) = 0$, where F is a necessarily false proposition (e.g., not-T).
- (5) $P(\text{this coin lands heads, given } Z) = 1/2$ (where the coin's landing heads is independent of Z).
- (6) $P(\text{a dart randomly thrown at } [0, 1] \text{ lands on the point } 1/4, \text{ given that it lands on either } 1/4 \text{ or } 3/4) = 1/2$.
- (7) When a point is randomly chosen from the Earth's surface, $P(\text{point lies in the western hemisphere, given that it lies on the equator}) = 1/2$.

These claims come in three flavors:

(1) – (4) are alike in being *extreme* cases, and in being not merely true but *necessarily* true. That is, *all* probability functions should agree on them. They are non-negotiable in the strongest sense.

(5) turns on a judgment of *independence*. When (as we say) two propositions are independent of each other, the unconditional probability of one should agree with its conditional probability given the other. But not all probability functions will concur on such an independence judgment. The standard practice of saying that two propositions are independent of each other is a little careless, suggesting as it does that independence is an intrinsic, two-place relation between such propositions; in fact it is a *three*-place relation between a proposition, another proposition, and a probability function. Thus, (5) is only contingently true in the sense that it holds for some probability functions but not others. Still, one can easily come up with examples in which it *is* true, and I purported to give such examples. Note that if we heed the moral of Kenny's paper, independence judgments require still further relativity: independence is then a *four*-place relation between a proposition, another proposition,

a probability function, *and a sigma algebra representing the set of alternatives of an experiment.*

(6) and (7) are based on *symmetry* judgments. Such judgments are in the eye of the beholder, and again they will not be shared by all probability functions. Still, I argued that in certain cases they are compelling. Since Kenny's paper focuses on (7), let us do so too. I argued that, since half of the equator lies in the western hemisphere, when a point is randomly chosen from the surface of the earth, intuition tells us that $P(\text{western hemisphere, given equator}) = 1/2$.

My seven officially stated intuitions concerning the values of conditional probabilities are before you. Now, all I need is *one* of them to be correct for the ratio analysis to be refuted. For the ratio analysis regards all of these conditional probabilities to be undefined. But in fact, I gather that Kenny *concedes all seven of these intuitions*. Yes, even (7)—for while he shows how two methods of calculation yield disagreeing values for various conditional probabilities, he concedes that both methods yield (7)'s answer of $1/2$.

So when Kenny says that “the intuitions Hájek appeals to are unreliable”, he can't have *these* intuitions in mind. Rather, he considers more complicated cases than mine, such as the probability that a random point on Earth has a probability of lying between 30 and 45 degrees north, given that it has longitude 0. Now, in order to extrapolate from the one intuition regarding the sphere that I provided, to guide intuitions about his new cases, he must be attributing some rule to me. I take him to be showing that this rule is unreliable. To be fair to him, I probably would have been

tempted to respond to his problems as he attributes. But to be fair to myself, I never gave such a rule in print.

Never mind the Hájek exegesis; Kenny's point is interesting whether or not it addresses what I actually said. I appealed to a certain intuition based on a certain symmetry in the sphere example. But there are symmetries, and then there are symmetries; there are intuitions, and then there are intuitions. There are competing symmetries, apparently equally compelling, that yield very different answers in the more complicated cases that he considers. Notice, though, that none of this undercuts my negative argument against the ratio analysis. The fact that there are *answers*—albeit competing ones—to his questions about conditional probabilities shows that the ratio analysis is inadequate. For it yields no answer at all. He goes on to argue that in such cases conditional probabilities need to be further relativized to sets of alternatives. This is only more bad news for the ratio analysis, which has no truck with any of that. In this sense Kenny is my ally.

3. Primitive conditional probabilities

Where he disagrees with me is over my positive thesis. I went on to eulogize primitive conditional probabilities. I argued that conditional probability is the true primitive of probability theory, and that it should be axiomatized directly, rather than derived from unconditional probabilities. Kenny contends that primitive conditional probabilities are liable to yield conflicting answers to his questions about the sphere. Here he is in a way playing Bertrand to my Laplace—or what would be my Laplace if I did not add further constraints on the sphere set-up before consulting intuitions about conditional probabilities.

So Kenny is right to press me that even primitive conditional probabilities give at best an incomplete account—they need to be completed with something else. He suggests that this should be a specification of a set of alternatives, as represented by a sigma algebra. This is to say that more of the underlying probability model must be incorporated into the conditional probabilities themselves. I wonder whether instead the further specification could be given, say, in English prose in the very formulation of the problem at hand; then conditional probabilities alone can carry the day. Perhaps sometimes not even the prose is needed, if there is a natural default assumption at work. I would like to think that the equator example was like that: it was natural, I maintain, to imagine a uniform distribution around the equator.

Be that as it may, Kenny believes that all the tools that we need were already at hand when Kolmogorov provided his extended analysis: “While Kolmogorov’s strict ratio definition doesn’t quite suffice, I claim that Kolmogorov did provide a formulation that does” (11). To be sure, that analysis seems to handle the Borel paradox with aplomb. But is it the final word on conditional probability? I think not.

4. What conditional probability also could not be

For as I said, I gave further arguments against the ratio analysis, which I took to be equally telling against the extended analysis. These were cases in which the requisite unconditional probabilities were what I called *vague*, or *undefined*, yet the corresponding conditional probabilities were perfectly well defined.

Actually, I think that Kenny slightly misconstrues my argument from vague probabilities, but the fault probably lies with my labeling of it (although there is precedent for it—e.g. in van Fraassen). And I am grateful for this, for it leads me to distinguish *another* kind of argument against the ratio analysis. The example that he

offers in the spirit of my argument from vague probabilities is this: “Consider two infinitely thin darts thrown at the real line, with independent uniform probability distributions over the interval $[0, 100]$... Given that the first dart hits a large value, what is the probability that the second dart hits a value below 50?” He rightly considers this a case of *vagueness*; I would have done better, then, to follow Levi in calling mine cases of *imprecision*. Think of the difference between my saying that I am “forty-something” and my saying that I am “fortyish”. The former conveys the definite, determinate interval $[40, 50]$, which admits of no borderline cases; the latter conveys a fuzzy, indeterminate region centered around 40 for which there are borderline cases. Call the former case one of imprecision, the latter one of vagueness. In Kenny’s example, the content of the unconditional probability $P(\text{first dart hits a large value})$ is vague because ‘large value’ is vague, admitting of borderline cases; this vagueness is plausibly then inherited in turn by the probability itself (say, ‘0.3-ish’, again with borderline cases). I had in mind instead a case of an unconditional probability with perfectly determinate content—e.g. ‘the Democrats win the next election’ (no borderline cases there); then, an *imprecise* assignment to that content (say, $[0.4, 0.5]$, again without borderline cases). But I think that *both* kinds of indeterminacy will cause trouble for the ratio analysis. For we can easily come up with sharp corresponding conditional probabilities, as I did and as Kenny did.

And both kinds of indeterminacy will cause trouble for the extended analysis. For that analysis equates a certain integral in which the relevant conditional probability figures, to the probability of a conjunction; but when this latter probability is either vague or imprecise, the analysis goes silent. Likewise for cases where the probability of the conjunction fails to exist at all, as in my cases of undefined unconditional

probabilities. So I believe that my original paper already cast some doubt on the cogency of the extended analysis.

Here, perhaps, remedies are easily found: simply import whatever devices you favor for handling imprecise or vague unconditional probability. For example, you may be able to supplement or generalize the extended analysis in a supervaluational spirit, much as Levi, Jeffrey, and van Fraassen did with imprecise unconditional probabilities, using sets of sharp probabilities. Of course, this is to concede that the extended analysis itself is inadequate, in need of supplementation or generalization. (Compare my (2003) remarks on p. 294 regarding the ratio analysis.) But you could reasonably say that at least the spirit of the extended analysis is maintained. The full story about conditional probability, you then say, is this generalization.

Or is it? I also wonder how the extended analysis, without further assumptions, will deliver even certain results *concerning probability zero conditions* that we would like. Fixing the distribution over the two-dimensional surface of the sphere puts no constraint on the distributions over sets of measure 0, such as the equator. Indeed, it is consistent with the 2-d distribution that there are gaps in the distribution over the equator – e.g. over the western hemisphere. So merely given the 2-d distribution, I don't see how *any* analysis, extended or otherwise, can answer questions about probabilities conditional on the equator, such as the one I asked, without some further tacit assumption—notably, an assumption of symmetry around the equator. But then granting *me* the same tacit symmetry assumption (which I did tacitly make), my preferred conditional probabilities could also deliver the desired answers.

Moreover, there is trouble for the extended analysis on other fronts. I am thinking especially of results by Teddy Seidenfeld, Mark Schervish and Jay Kadane in their paper “Improper Regular Conditional Distributions”. (“Regular conditional

distributions” are another name for the extended analysis.) Let $P(\cdot | \mathcal{A})(\omega)$ denote the regular conditional distribution for the measure space (Ω, \mathcal{B}, P) given the conditioning sub- σ -field \mathcal{A} . Following Blackwell et al., say that a regular conditional distribution is *proper at ω* if it is the case that whenever $\omega \in A \in \mathcal{A}$,

$$P(A | \mathcal{A})(\omega) = 1.$$

The distribution is *improper* if it is not everywhere proper. Improperity seems to be disastrous. In my original paper, I held this truth to be self-evident: the conditional probability of anything consistent, given *itself*, should be 1. I considered this to be about as fundamental fact about conditional probability as there could be, on a par with the fundamental fact in logic that any proposition implies itself. So the possibility of improperity, however minimal and however localized it might be, is a serious defect in an account of conditional probability. But Teddy and his collaborators show just how striking the problem is. They give examples of regular conditional distributions that are *maximally* improper. They are cases in which

$$P(A | \mathcal{A})(\omega) = 0$$

(as far from the desired value of 1 as can be), and this improperity holds *almost everywhere* according to P , so the improperity is maximal both locally and globally.¹

(When \mathcal{A} is atomic, the event A which is assigned the embarrassing conditional probability of 0, evaluated at ω , is the very atom of \mathcal{A} that contains ω .)

Kenny writes that calculating conditional probabilities according to the extended analysis is “the practice of statisticians and mathematicians working in the field” (11).

¹ A necessary condition for this is that the conditioning sub-sigma algebra is not countably generated.

Indeed it is; but I do not recognize anything that behaves improperly, let alone *so* improperly, as capturing the pre-theoretical notion of conditional probability.²

² I am grateful to Kenny Easwaran, Branden Fitelson, Teddy Seidenfeld, and especially George Kahramanis for generous and helpful feedback.