Declarations of Independence

Branden Fitelson and Alan Hájek

Abstract: According to orthodox (Kolmogorovian) probability theory, conditional probabilities are by definition certain ratios of unconditional probabilities. As a result, orthodox conditional probabilities are regarded as undefined whenever their antecedents have zero unconditional probability. This has important ramifications for the notion of probabilistic *independence*. Traditionally, independence is defined in terms of unconditional probabilities (the factorization of the relevant joint unconditional probabilities). Various "equivalent" formulations of independence can be given using conditional probabilities. But these "equivalences" break down if conditional probabilities are permitted to have conditions with zero unconditional probability. We reconsider probabilistic independence in this more general setting. We argue that a less orthodox but more general (Popperian) theory of conditional probability should be used, and that much of the conventional wisdom about probabilistic independence needs to be rethought.

1. Introduction

According to orthodox (Kolmogorovian) probability theory, conditional probabilities are by *definition* certain ratios of unconditional probabilities. As a result, orthodox conditional probabilities are regarded as undefined whenever their antecedents have zero unconditional probability.[1] Such zero probability cases are typically glossed over or ignored altogether in standard treatments of probability, especially those found in the philosophical literature. The orthodox treatment of conditional probabilities goes hand in hand with the orthodox treatment of probabilistic *independence*. Traditionally, independence is defined in terms of unconditional probabilities (the factorization of the relevant joint unconditional probabilities). Various "equivalent" formulations of independence can be given using conditional probabilities. But these "equivalences" presuppose orthodoxy about conditional probabilities, and they break down if conditional probabilities are permitted to have conditions with zero unconditional probability. In this paper we reconsider the nature of probabilistic independence in this more general setting. We will argue that when a less orthodox but more general (Popperian) theory of conditional probability is used, much of the conventional wisdom about probabilistic independence will need

---

[1] To be sure, Kolomogorov was well aware of this problem, and he went on to offer a more sophisticated treatment of probability conditional on a sigma algebra, P(A‖𝓕), in order to address it. We will return to this point later; as we will see, this approach also faces some serious problems.

to be rethought. Because independence is such a central notion in probability theory, we will conclude that the entire orthodox mathematical framework for thinking about probabilities should be reconsidered at a fundamental level.

2. The Orthodox Theory of Probability: Kolmogorov

Almost every textbook or article about probability theory since 1950 follows Kolmogorov's approach.[2] That is to say, almost all contemporary probabilists characterize probabilities by: (i) taking unconditional probabilities $P(\bullet)$ as primitive, (ii) using Kolmogorov's (1933, 1956) axioms for $P(\bullet)$, and (iii) defining conditional probabilities $P(\bullet \mid \bullet)$ in terms of unconditional probabilities $P(\bullet)$ in the standard way, using Kolmogorov's ratio definition. Specifically, most probabilists adopt the following framework for probabilities:

Definition. Let $\mathcal{F}$ be a field on a set Ω. An unconditional Kolmogorov probability function $P(\bullet)$ is a function from $\mathcal{F}$ to $[0,1]$ such that, for all $A, B \in \mathcal{F}$:

K1. $P(A) \geq 0$.

K2. $P(\Omega) = 1$.

K3. If $(A \cap B) = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.[3]

We call a triple $\mathcal{M} = (\Omega, \mathcal{F}, P)$ satisfying these axioms a *Kolmogorov probability model*. The conditional Kolmogorov probability function $P(\bullet \mid \bullet)$ in $\mathcal{M}$ is now defined in terms of $P(\bullet)$ as the following partial function from $\mathcal{F} \times \mathcal{F}$ to $[0,1]$:

K4. $P(A \mid B) =_{def} \dfrac{P(A \cap B)}{P(B)}$, provided $P(B) > 0$.

Thus, on the orthodox account, all probabilities are unconditional probabilities or ratios thereof.

---

[2] To name just a few of the most famous of these, we have Parzen (1960), Papoulis (1965), Feller (1968), Rozanov (1977), Loève (1977), Billingsley (1995), Ross (1998), and, of course, Kolmogorov (1950) himself.

[3] Axiom K3 (finite additivity) is often strengthened to require additivity over denumerably many mutually exclusive events. There is considerable controversy over the issue of countable additivity. Both Savage and de Finetti urged against the assumption of countable additivity in the context of personalistic probability. Moreover, the assumption

In his classic textbook, Billingsley (1995) writes that "the conditional probability of a set *A* with respect to another set *B* … is defined of course by $P(A|B) = P(A \cap B)/P(B)$, unless $P(B)$ vanishes, in which case it is not defined at all" (427). Three things leap out at us here: the ratio is regarded as a *definition* of conditional probability; its being so regarded is *obvious* ("of course"); and it is regarded as *not defined at all* when $P(B) = 0$. The very *meaning* of conditional probability is supposed to be given by the ratio definition; whether this definition is adequate is not questioned; and the conditional probability has no value ("defined" has another meaning here) when its condition has probability 0. On another understanding, the account *is silent* about such conditional probabilities. We need not choose one understanding over the other here, since either way such conditional probabilities are problematic when they should be well-defined, and indeed constrained to have particular values and not others.

Now if 'conditional probability' were a purely formal notion, stipulatively defined by K4, then there would be no worrying objection here. You can introduce a new theoretical term and stipulate it to be whatever you like, and if your readers and interlocutors are cooperative, they will adhere to the stipulation. If you want to stipulate that by the 'schmonditional probability of A, schmiven B' you mean the cube root of $P(A \cup B) - P(\neg B)$, go right ahead. You will not find much enthusiasm for schmonditional probability, however, unless you can convince your audience that it does some distinctive *work*, that it captures some concept of use to us.

Conditional probability is clearly intended to be such a concept. 'The probability of A, given B' is not merely *stipulated* to be such-and-such a ratio; rather, it is meant to capture the familiar notion of the probability of A in the light of B, or informed by B, or relative to B. It is a useful concept, because even if Hume was right that there are no necessary connections between distinct existences, still it seems there are at least some non-trivial probabilistic relations between them. That's just what we mean by saying things like 'B supports A', or 'B is evidence for A', or 'B is counterevidence for A', or 'B disconfirms A'. Presumably it is a concept that will guide

---

of countable additivity has some surprising and paradoxical consequences in Kolmogorov's more general theory of conditional probability (Seidenfeld et al. 2001). We will briefly comment on this issue below.

such judgments that Kolmogorov is seeking to capture. Then it is an open question (somewhat in the sense of Moore) whether he has succeeded. If he has, then that just means that he has given us a good piece of conceptual analysis. By analogy, Tarski's account of logical consequence is not merely a stipulative definition, a new piece of jargon for beleaguered students of logic to learn; rather, it is intended to capture the strongest form of 'support ' or 'evidence for' relation that could hold between two sentences. It is intelligible to ask whether he has succeeded (and Etchemendy 1990 answers the question in the negative).

So let us call the identification of conditional probabilities with ratios of unconditional probabilities the *ratio analysis* of conditional probability. In the next section, we discuss some of its peculiarities. These are reasons for thinking that the ratio analysis is unsuccessful.

3. Some Peculiarities of Kolmogorov's Ratio Analysis of Conditional Probability

Let us look in more detail at the peculiarity that conditional probabilities are undefined whenever their antecedents have zero unconditional probability. K4 has its proviso for a reason. Now, perhaps the proviso strikes you as innocuous. To be sure, we could reasonably dismiss probability zero antecedents as 'don't cares' if we could be assured that all probability functions of any interest are *regular*—that is, they assign zero probability only to logical impossibilities. Unfortunately, this is not the case. As probability textbooks repeatedly drum into their readers, probability zero events need not be impossible, and indeed can be of real significance. It is curious, then, that some of the same textbooks glide over K4's proviso without missing a beat.

In fact, interesting cases of probability zero antecedents are manifold. Consider an example due to Borel: A point is chosen at random from the surface of the earth (thought of as a perfect sphere); what is the probability that it lies in the Western hemisphere, given that it lies on the equator? 1/2, surely. Yet the probability of the antecedent is 0, since a uniform probability measure over a sphere must award probabilities to regions in proportion to their area, and the equator has area 0. The ratio analysis thus cannot deliver the intuitively correct answer. Obviously there are uncountably many problem cases of this form for the sphere.

Nor is the problem limited to uncountable, or even infinite domains. Finite probability models can contain propositions that receive zero probability. While all contradictions must have zero probability, the converse is not true—even in finite spaces. For instance, rational agents (who may have credences over finite spaces) may—and arguably must[4]—have irregular probability functions, and thus assign probability 0 to non-trivial propositions.

Probability theory and statistics are shot through with non-trivial zero-probability events. Witness the probabilities of continuous random variables taking particular values (such as a normally distributed random variable taking the value 0). Witness the various 'almost sure' results—the strong law of large numbers, the law of the iterated logarithm, the martingale convergence theorem, and so on. They assert that certain convergences take place, not with certainty, but 'almost surely'. This is not merely coyness, since these convergences may fail to take place—genuine possibilities that receive probability 0, and interesting ones at that. A fair coin may land tails forever.

And so it goes; indeed it could not have gone otherwise. For necessarily a probability distribution over uncountably many outcomes is irregular (still assuming Kolmogorov's axiomatization). More than that: such a distribution must accord uncountably many outcomes probability zero. For each such outcome, we have a violation of a seeming platitude about conditional probability: that the probability of a proposition *given that very proposition*, is 1. Surely that is about as fundamental a fact about conditional probability as there could be—on a par, we would say, with the platitude about logical consequence that every proposition entails itself.

Equally platitudinous, we submit, is the claim that the conditional probability of a proposition, given something else that entails that proposition, is 1: for all distinct X and Y, if X entails Y then the probability of Y, given X, is 1. But clearly this too is violated by the ratio analysis. Suppose that X and Y are distinct, X entails Y, but both have probability 0 (e.g., X = 'the randomly chosen point lies in the western hemisphere of the equator', and Y = 'the point

---

[4] See Hájek (2003), Pruss (2013), Easwaran (2014).

lands on the equator'). Then P(Y|X) is undefined; but intuition demands that the conditional probability be 1.

The difficulties that probability zero antecedents pose for the ratio formula for conditional probability are well known (which is not to say that they are unimportant). Indeed, Kolmogorov himself was well aware of them, and he offered a more sophisticated account of conditional probability as a random variable conditional on a sigma algebra, appealing to the Radon-Nikodym theorem to guarantee the existence of such a random variable. But our complaints about the ratio analysis are hardly aimed at a strawman, since (RATIO) is by far the most commonly used analysis of conditional probability, especially in philosophical applications of probability. (Indeed, most philosophers—even mathematically literate ones—would be unable to state the more sophisticated account.) Moreover, the move to the more sophisticated account does not solve all the problems raised in this paper. In particular, even that account does not respect the platitudes of conditional probability stated above, as evidenced by the existence of so-called *improper* conditional probability random variables. Seidenfeld et al. (2001) show just how extreme and how widespread violations of the platitudes can be. And the more sophisticated machinery does not help with zero probability conditions in finite domains. Above all, the more sophisticated theory still delivers unintuitive verdicts regarding independence, as we will see in section 7. So that theory is no panacea.

Hájek (2003) goes on to consider further problems for the ratio formula: cases in which the unconditional probabilities that figure in the ratio are *vague* (imprecise) or are *undefined*, and yet the corresponding conditional probabilities are defined. (These prove to be problematic also for the more sophisticated account, so therein lies no solution either.) For our purposes in this paper, however, cases of probability zero antecedents are problematic enough.

We conclude that it is time to rethink the foundations of probability. It is time to question Kolmogorov's axiomatization, and in particular the conceptual priority it gives to unconditional probability. It is time to consider taking conditional probability as the fundamental notion in probability theory.

4. An Unorthodox Account of Conditional Probability: Popper

There are various ways to define conditional probabilities as *total* functions from $\mathcal{F} \times \mathcal{F}$ to [0,1].[5] Popper (1959) presents a general account of such conditional probabilities. The functions picked-out by Popper's axioms are typically called *Popper functions*. Here is a simple axiomatization of Popper functions. (We follow Roeper and Leblanc's 1999 set-theoretic axiomatization.[6])

Definition. Let $\mathcal{F}$ be a field on a set Ω. A Popper function $\Pr(\bullet, \bullet)$ is a total function from $\mathcal{F} \times \mathcal{F}$ to [0,1] such that:

P1. For all $A, B \in \mathcal{F}, 0 \le \Pr(A, B)$.

P2. For all $A \in \mathcal{F}, \Pr(A, A) = 1$.

P3. If there exists a $C \in \mathcal{F}$ such that $\Pr(C, B) \ne 1$, then $\Pr(A, B) + \Pr(\neg A, B) = 1$.

P4. For all $A, B, C \in \mathcal{F}, \Pr(A \cap B, C) = \Pr(A, B \cap C) \Pr(B, C)$.

P5. For all $A, B, C \in \mathcal{F}, \Pr(A \cap B, C) = \Pr(B \cap A, C)$.

P6. For all $A, B, C \in \mathcal{F}, \Pr(A, B \cap C) = \Pr(A, C \cap B)$.

P7. There exist $A, B \in \mathcal{F}$ such that $\Pr(A, B) \ne 1$.

---

[5] For instance, Carnap (1950, 1952), Popper (1959), Kolmogorov (1950), Renyi (1970) and several others have proposed axiomatizations of conditional probability (as primitive). See Roeper and Leblanc (1999) for a very thorough survey and comparison of these alternative approaches (in which it is shown that Popper's definition of conditional probability is the most general of the well-known proposals).

[6] This axiomatization is slightly different (syntactically) from Popper's original axiomatization. But, the two are equivalent (as Roeper and Leblanc show). Moreover, we are defining Popper functions over sets rather than statements or propositions, which is non-standard. If you prefer, think of our sets as sets of possible worlds in which the corresponding propositions are true. This is an inessential difference (since it doesn't change the formal consequences of the axiomatization), and we will use the terms "entailment" and "set inclusion" interchangeably. Our aim here is to frame the various axiomatizations as generally (and commensurably) as possible. We don't want to restrict some axiomatizations (e.g., Popper's) to logical languages or other structures that have limited cardinality. Popper's aim was to provide a *logically autonomous* axiomatization of conditional probability. Ours is simply to compare various axiomatizations in various ways, with an eye toward independence judgments. So we don't mind *interpreting* the connectives in both Popper's axiomatization and Kolmogorov's axiomatization, and doing so in the same (*non*-autonomous, set-theoretic) way. Given our set-theoretic reading of the connectives, axioms P5 and P6 above are redundant. We include them so that the reader can easily cross-check the above axiomatization with the (autonomous) axiomatic system given in Roeper and Leblanc. We also recommend that text for various key lemmas and theorems that are known to hold for Popper functions.

We call a triple $\mathcal{M} = (\Omega, \mathcal{F}, \text{Pr})$ satisfying these axioms a *Popper probability model*. It can be shown (Roeper and Leblanc 1999, chapter 1) that the Popper function $\text{Pr}(X, \Omega)$, thought of as a unary function of X, is just the Kolmogorovian unconditional probability function P(X) defined on $(\Omega, \mathcal{F})$. In this sense, Popper functions can be thought of as an (conservative) extension of the Kolmogorovian theory of unconditional probability. Indeed, Popper and Kolmogorov agreed on the nature of unconditional probability. They only disagreed about the nature of conditional probability. The following important fact about Popper functions will be used in subsequent sections:

For all A and B, $\text{Pr}(A, A \cap B) = 1$.

This codifies our second platitude about conditional probability: the conditional probability of a proposition, given something else that entails that proposition, is 1. (P2 codifies the first: the probability of a proposition *given that very proposition,* is 1.)

Now, we do not want to insist that Popper's axiomatization is necessarily definitive. If some other simpler, more intuitive, or more powerful axiomatization of conditional probability can be given, so much the better for our cause. But we will happily work with Popper's axiomatization in the meantime, mainly because it is the one best known to philosophers.

5. Some preliminary reflections on independence

We now come to the notion of independence—or rather, to the *notions* of independence, since it will soon emerge that there are many.

On the Kolmogorovian theory of probability, A and B are said to be *independent (in a model $\mathcal{M} = (\Omega, \mathcal{F}, P))$* just in case

(FACTORIZATION) $P(A \cap B) = P(A) P(B)$.

Here we arrive at that part of probability theory that is distinctively *probabilistic*. Axioms K1 – K3 are general measure-theoretic axioms that apply equally to length, area and volume (suitably

normalized), and even to mass (suitably understood). It is this characterization of independence that we find only in probability theory.

Note that (FACTORIZATION) generalizes nicely to *n* events, as follows:

$A_1, \ldots, A_n$ are *independent (in a model $\mathcal{M} = (\Omega, \mathcal{F}, P)$)*

iff

(GENERALIZED FACTORIZATION) $P\left(\bigcap A_i\right) = \prod P(A_i)$.

The right hand side is to be understood as a shorthand for many simultaneous conditions: all probabilities of two conjuncts factorize, all probabilities of three conjuncts factorize, … , the probability of the conjunction of all n events factorizes.

K1 – K3 presuppose nothing but the real interval [0, 1] and elementary set theory, things that are antecedently well understood. But mathematics gives us no independent purchase on 'independence'. For that we must look elsewhere. Paralleling our discussion of conditional probability in section 2: If (FACTORIZATION) were simply a stipulative definition of a new technical term, then there could be no objection to it (nor to a stipulative definition of 'schmindependence', should anyone be moved to give one).

But 'independence' is clearly a concept with which we were familiar before Kolmogorov arrived on the scene, even more so than the concept of 'conditional probability'. The choice of word, after all, is no accident. Indeed, it has such a familiar ring to it that we are liable to think that (FACTORIZATION) captures what we always meant by the word 'independence' in plain English. The concept of 'independence' of A and B is supposed to capture the idea of the *insensitivity* of (the probability of) A's occurrence (truth) to B's, or the *uninformativeness* of B's occurrence (truth) to A's. The probability of A is unmoved by how things stand with respect to B. Once again, it is an open question whether Kolmogorov has succeeded in capturing this idea. If he has, then that just means that he has given us a good piece of conceptual analysis.

We think that he has not succeeded. For starters, note that unlike 'conditional probability', 'independence' is obviously a multifarious notion. We use the same word when we speak of

logical independence, counterfactual independence, metaphysical independence (the Humean absence of any 'necessary connection' between distinct existences in virtue of which some proposition X is true in every possible world in which another proposition Y is true), evidential independence, and now, probabilistic independence. So *at best*, Kolmogorov has given us an analysis of just *one* concept of independence among many. Still, that would be quite an achievement. Von Neumann and Morgenstern certainly did us a service in giving us 'game theory', even if solitaire doesn't fall within its scope.

So let us focus on probabilistic independence—and from now on when we speak simply of 'independence', that's what we will mean. It is usual to speak simply of one proposition or event being independent of another, as in "A is independent of B". This is somewhat careless, encouraging one to forget that independence is a *three*-place relation among two propositions *and a probability model*. Here we part from ordinary English, and from all the other notions of 'independence' listed above, in which independence is a two-place relation. Nevertheless, when a particular probability function is for some reason salient, the usual practice might seem to be innocuous enough (and in such cases we will sometimes follow it ourselves).

According to the orthodox account, probabilistic independence is symmetric: if A is independent of B, then B is independent of A (with respect to a given probability model). This is obvious from the symmetric role that A and B play in (FACTORIZATION). Causal independence, on the other hand, is not symmetric: A can be causally independent of B without B being causally independent of A: your survival depends causally on a healthy distribution of air molecules in your vicinity while such a distribution is causally independent of your survival. Similarly, A can be counterfactually independent of B without B being counterfactually independent of A. We can have:

B □→ A and ¬B □→ A (A is counterfactually independent of B)

while having

A □→ B and ¬A □→ ¬B (B is counterfactually *dependent* on A).

(Lewis 1979 made much of such asymmetries in counterfactual dependence in his work on time's arrow: the past, he argued, is counterfactually independent of the future but the future is counterfactually dependent on the past.) And regarding supervenience as a species of metaphysical dependence, this notion of dependence is also not symmetric: mental states may be dependent on physical states without physical states being dependent on mental states (think of multiple realizability of a given mental state). To be sure, *evidential* independence appears to be symmetric. Perhaps, however, that appearance should be questioned once we question the usual account of probabilistic independence, as we are now.

Kolmogorov's notion of independence may be the odd one out in various ways among several notions of independence, but that does not yet mean that it is *odd*. Let us now see just how odd it is.

6. Further peculiarities of the factorizing construal of independence

We have reflected at some length on the orthodox treatment of independence, defined in terms of the factorization of joint probabilities as given by (FACTORIZATION). We will call this the *factorizing construal* of independence.

According to the factorizing construal of independence, anything with extreme probability has the peculiar property of being probabilistically independent of *itself*:

If $P(X) = 0$, then $P(X \cap X) = 0 = P(X)P(X)$.

If $P(X) = 1$, then $P(X \cap X) = 1 = P(X)P(X)$.

Yet offhand, we would have thought that identity is the ultimate case of *de*pendence (with one exception that we are about to note). Every possible proposition X is:

- logically dependent on itself (since X entails X);

- counterfactually dependent on itself (since X $\square\!\rightarrow$ X and ¬X $\square\!\rightarrow$ ¬X);

- supervenient on itself (since in every possible world in which X is the case, X is the case).[7]

To be sure, we should be cautious about drawing morals for probabilistic dependence from other dependence relations. That said, offhand one would expect the probabilistic dependence of X on X to be *maximal*, not minimal. Much as we took it to be a platitude that

for every proposition X, the probability of X given X is 1,

so we take it to be a platitude that for every proposition X, X is *de*pendent on X. What better support, or evidence, for X could there be than X itself? What proposition could X's truth value be more sensitive to than X's itself?

Well, perhaps we should allow exactly one exception: the case where X has probability 1. Perhaps then X *is* probabilistically insensitive to itself, since its probability is already maximal. It seems right that its probability is unmoved by its own occurrence—it has nowhere higher to move! Very well then; let us admit the exception. But this very consideration only drives home how serious the problem is at the other end: the case where X has probability *0*. The alleged self-independence of probability *zero* events is disastrous by these lights, for their sensitivity to their own occurrence should be maximal, not minimal.

In any case, perhaps the notion of probabilistic independence is not univocal. We may well have more than one such concept. Then it should come as no surprise if our intuitions are sometimes pulled in different ways—different concepts may be pulling them. *X* is logically *dependent* on itself even if it has probability 1; perhaps we should admit a notion of probabilistic independence that yields the same verdict? We may well want inductive logic, understood as probability theory, to be continuous with deductive logic.[8] However, the Kolmogorovian orthodoxy treats probabilistic independence as if it is univocal. And we insist that there is *no* concept of independence that should regard probability zero events as independent of themselves. They could not be more hostage to their own occurrence, probabilistically speaking!

---

[7] The case of causal dependence is a little different because there are built-in logical or mereological 'no-overlap' constraints on the relata of the causal relation. See Arntzenius (1992) for discussion.

[8] Thanks to Leon Leontyev for this way of expressing the point.

More generally, according to the factorizing construal of independence, any proposition with extreme probability has the peculiar property of being probabilistically independent of *anything*. This includes anything that entails the proposition, and anything that the proposition entails. Much as we took it to be a platitude that

for all X and Y, if X entails Y then the probability of Y, given X, is 1,

so we take it to be a platitude that for every proposition X, X is *de*pendent on anything that entails X. Again, we should perhaps grant probability 1 events exceptional status. But all the more we should *not* grant exceptional status to probability 0 events, as orthodoxy would have it. They could not be further from deserving exemption from the platitude—even allowing for more than one conception of probabilistic independence.

We could perhaps tolerate these unwelcome consequences of the factorizing construal of independence if every probability function were regular. Then the extreme-probability propositions would be confined to logical truths and contradictions, and they could reasonably be dismissed as *'don't-cares'* (much as the result that in classical logic everything follows from a contradiction might be dismissed as a 'don't-care'). But, just as we saw in our discussion of problems for the ratio analysis, many of the propositions in question are *'cares'*: non-trivial propositions of genuine interest to probability theory and statistics.

7. "Equivalent" formulations of independence

As long as all relevant conditional probabilities are well-defined, the orthodox theory has many other ways — *via* conditional probabilities — of saying that A and B are independent. Here are four of the simplest and most common of these ways:[9]

   i. $P(A \mid B) = P(A)$

   ii. $P(B \mid A) = P(B)$

   iii. $P(A \mid B) = P(A \mid \neg B)$

---

[9] It is shown in Fitelson (1999, 2001) that, despite the unified nature of the Kolmogorovian theory of probabilistic (in)dependence, there are many (radically) *non*-equivalent Kolmogorovian measures of *degree* of dependence (*i.e.*, degree of correlation among propositions).

iv. $P(B \mid A) = P(B \mid \neg A)$

It is crucial that we add the caveat about all salient conditional probabilities being *well-defined*. Otherwise the equivalence between the factorizing construal and these alternative, conditional forms of independence breaks down. Some textbooks on probability neglect to mention this caveat.[10] This is unfortunate since, as we just saw, according to the factorizing construal, anything with extreme probability is probabilistically independent of itself, whereas this is *not* the case for all the conditional probability formulations of independence. For if $P(B) = 0$, then the orthodox theory cannot deliver the mandatory verdict that the conditional probability of B, given *itself*, is 1. Thus, whereas the factorizing construal gives a verdict on the self-independence of such zero probability propositions, all of i – iv go silent. Similarly, if $P(B) = 1$, then the orthodox theory goes silent on the value of $P(B \mid \neg B)$. So iii and iv go silent for probability one propositions.

In the remaining sections, we will consider what happens when conditional probabilities are defined even when their antecedents have probability zero. We will show that when conditional probabilities are taken as primitive (and defined as *total* functions), a new and different theory of probabilistic independence emerges.

8. Conditional Probability and Independence — Revisited

When we adopt a Popper—and we suggest, a *proper*—definition of conditional probability, we are forced to rethink the notion of probabilistic independence. Perhaps the easiest way to see this is to reconsider the case in which the unconditional probability of Z is zero. The Popper analogue of the factorizing construal will say that A and B are independent (in $(\Omega, \mathcal{F}, \text{Pr})$) iff:

(POPPER FACTORIZATION) $\text{Pr}(A \cap B, \Omega) = \text{Pr}(A, \Omega)\text{Pr}(B, \Omega).$

---

[10] See, for example, Pfeiffer (1990, pp. 73–84) who states 16 "equivalent" renditions of "A and B are probabilistically independent" (including our four, above) without mentioning that this "equivalence" depends on the assumption that the conditional probabilities are well-defined. He does the same thing in his discussion of *conditional* independence (pp. 89–113). Moreover, in the very same text (pp. 454–462), he discusses Kolmogorov's more sophisticated definition of conditional probability. So, he is clearly well aware of the problem of zero probability conditions in the general case. This is not atypical.

But if $Pr(Z, \Omega) = 0$, then $Pr(Z \cap Z, \Omega) = 0 = Pr(Z, \Omega) \, Pr(Z, \Omega)$. That is, the Popper analogue of the factorizing construal says that Z is independent of itself in this case. However, if $Pr(Z, \Omega) = 0$, then $Pr(Z, Z) = 1 > Pr(Z, \Omega) = 0$. So, on this conditioning construal of probabilistic independence, Z is *maximally de*pendent on itself (*i.e.*, Z is *maximally* positively relevant to itself). So, not only are the factorizing construal and this conditioning construal of independence not equivalent; they are incompatible in the strongest possible sense. This means that we are now forced to be more precise about what we mean when we say that "X and Y are probabilistically independent."

We take it as intuitively clear that the factorizing construal is, in fact, an *incorrect* account of probabilistic independence—indeed, we claim to have just shown this. After all, it seems clear that nothing could be more relevant to Z than Z itself. The fact that Z has (the Popper analogue of) unconditional probability zero does nothing to undermine this intuition. Indeed, on *any* reasonable measure of incremental support, Z's incremental support for itself will be maximal.

A simple example makes this more general point quite clearly. Let H be the hypothesis that a coin generated by a certain machine will land heads. The background evidence Ω includes the information that the machine in question spits out coins so that their biases are uniformly distributed on [0,1]. Let B(b) be the proposition that the bias of the coin is b. It is reasonable to assign zero probability to each of the B(b), since the governing distribution is continuous, and it is also reasonable to assign conditional probabilities so as to satisfy $Pr(H, B(b)) = b$. In any case, for some values of b, B(b) will be *highly relevant to* H. That is, for some values of b—namely, high and low ones—we will surely have $Pr(H, B(b)) \gg Pr(H, \Omega)$, or $Pr(H, B(b)) \ll Pr(H, \Omega)$. Moreover, it seems intuitively obvious to us that, in such cases, we should say that B(b) and H are probabilistically *dependent*. Alas, the orthodox account of independence cannot say this. The conditioning construals of independence in which B(b) is the condition do not speak at all.

Worse still, the other conditioning construals and the factorizing construal do speak, but one wishes they wouldn't, because what they say is false: they judge B(*b*) and H to be independent.

Moving to Kolmogorov's more sophisticated treatment of conditional probability will not save the day. For starters, his FACTORIZATION construal is stated purely in terms of unconditional probabilities, so any finessing of conditional probability will make no difference there. But even the more sophisticated versions of construals i – iv will fail. It follows from the results of Seidenfeld et al. (2001) that probability zero events can be *minimally* rather than *maximally* dependent on themselves. Not only do such events violate the platitude that every event is maximally self-dependent; they could not violate it more.

So it seems that we must abandon Kolmogorov's factorizing and conditioning construals of independence. We propose replacing them with Popper-style conditioning construals. But, *which* conditioning construals? As we mentioned above, there are *many* possible candidates. We have already rejected the Popper analogue of the factorizing construal. The following two candidates for analyzing 'A is independent of B (in $(\Omega, \mathcal{F}, \mathrm{Pr}))$' naturally come to mind:

(Ω CONSTRUAL) $\qquad$ $\mathrm{Pr}(A, B) = \mathrm{Pr}(A, \Omega)$

(NEGATION CONSTRUAL) $\mathrm{Pr}(A, B) = \mathrm{Pr}(A, \neg B)$

Interestingly, in the theory of Popper functions, (NEGATION CONSTRUAL) is strictly logically stronger than (Ω CONSTRUAL), and (Ω CONSTRUAL) is strictly logically stronger than (POPPER FACTORIZATION). To see that (POPPER FACTORIZATION) does not entail either (Ω CONSTRUAL) or (NEGATION CONSTRUAL), suppose that $\mathrm{Pr}(A \cap B, \Omega) = 0$, $\mathrm{Pr}(B, \Omega) = 0$, $\mathrm{Pr}(A, \Omega) < 1$, $\mathrm{Pr}(A, \neg B) < 1$, and B entails A, so that $\mathrm{Pr}(A, B) = 1$. Then, $\mathrm{Pr}(A \cap B, \Omega) = 0 = \mathrm{Pr}(A, \Omega)\,\mathrm{Pr}(B, \Omega)$, but $\mathrm{Pr}(A, B) > \mathrm{Pr}(A, \Omega)$, and $\mathrm{Pr}(A, B) > \mathrm{Pr}(A, \neg B)$. To see that (Ω CONSTRUAL) does not entail (NEGATION CONSTRUAL), just let $A = \neg\Omega = \varnothing$ and $B = \Omega$. Then, $\mathrm{Pr}(A, B) = 0 = \mathrm{Pr}(A, \Omega)$, since $B = \Omega$. But $\mathrm{Pr}(A, \neg B) = 1 \neq \mathrm{Pr}(A, B) = 0$, since $\neg B = A$.

See the Appendix for axiomatic proofs of the chain of entailments (NEGATION CONSTRUAL) $\Rightarrow$ (Ω CONSTRUAL) $\Rightarrow$ (POPPER FACTORIZATION).

Moreover, both conditioning construals are superior to (POPPER FACTORIZATION), since neither has the consequence that probability zero (conditional on Ω) events are independent of themselves; they are rightly judged as self-dependent (indeed, maximally so). However, (NEGATION CONSTRUAL) disagrees with (Ω CONSTRUAL) in judging the self-dependence of probability *one* events (conditional on Ω). (NEGATION CONSTRUAL) says that they are self-dependent, but (Ω CONSTRUAL) says that they are self-independent. Given our earlier discussion, in which we considered giving probability one events exceptional status in being self-independent, this seems to be a point in favor of (Ω CONSTRUAL). Then again, we also allowed that our concept of probabilistic independence may not be univocal. Perhaps (Ω CONSTRUAL) and (NEGATION CONSTRUAL) codify different senses in which the probability of A is unmoved by how things stand with respect to B. It could be a matter of whether that probability is unmoved by *the information that B*; or a matter of whether it is unmoved by *the answer to the question of which of B or not-B is the case*.[11] It is perhaps surprising that this distinction can make a difference; more power to the Popper formalism that it brings out this difference.

In any case, we need not settle here the issue of whether one of (Ω CONSTRUAL) and (NEGATION CONSTRUAL) is superior to the other. What matters here is that *both* (Ω CONSTRUAL) and (NEGATION CONSTRUAL) are superior to their Kolmogorovian counterparts.

There are further interesting consequences of our proposals. Probabilistic independence, understood either as (Ω CONSTRUAL) or (NEGATION CONSTRUAL), is *not* symmetric. Let Z have probability **z**ero (conditional on Ω). According to (Ω CONSTRUAL), Z is independent of $\neg$Z:

---

[11] Thanks to Hanti Lin for helpful discussion here.

$$Pr(Z, \neg Z) = 0 = Pr(Z, \Omega).$$

But ¬Z is *dependent* on Z:

$$Pr(\neg Z, Z) = 0 \neq Pr(\neg Z, W) = 1.$$

Similarly, we can have $Pr(A, B) = Pr(A, \neg B)$ without having $Pr(B, A) = Pr(B, \neg A)$. Think of a random selection of a point from the [0, 1] interval. Let

A = the point is 1/2.

B = the point lies in [1/4, 3/4].

Then, intuitively, we should have:

$$Pr(A, B) = 0 = Pr(A, \neg B).$$

$$Pr(B, A) = 1 \neq 1/2 = Pr(B, \neg A).$$

So on either construal, we can no longer just make claims of the form "A and B are dependent". We must now say things like "A is dependent on B" (as opposed to "B is dependent on A") to make clear *which direction* of dependence we have in mind. (This brings probabilistic independence closer to causal, counterfactual, and supervenience, which are similarly not symmetric, as we have seen.) On either construal, we can distinguish two (asymmetric) notions of dependence in the obvious ways:

- (Ω CONSTRUAL):

  o   A is positively dependent on B iff $Pr(A, B) > Pr(A, \Omega)$

  o   A is negatively dependent on B iff $Pr(A, B) < Pr(A, \Omega)$

- (NEGATION CONSTRUAL):

  o   A is positively dependent on B iff $Pr(A, B) > Pr(A, \neg B)$

  o   A is negatively dependent on B iff $Pr(A, B) < Pr(A, \neg B)$

The generalization to $n \geq 3$ propositions is not so straightforward for either the (Ω CONSTRUAL) or (NEGATION CONSTRUAL). We can't just say that "$A_1, \ldots, A_n$ are mutually independent." We need to say which of the A_i are independent of (or dependent on) which.

To sum up: on a Popperian account of independence, we must specify a *direction* of independence. Claims about probabilistic (in)dependence are now of the form: "A is (in)dependent on (of) B, relative to a probability model $\mathcal{M}$". Properly understood, probabilistic independence is an asymmetric, three-place relation.

A defender of the Kolmogorov orthodoxy might say at this point that we have done his work for him: we have exposed how complicated independence becomes when it is given a Popperian gloss. Doesn't Ockham's razor bid us to prefer the simpler theory of independence? Of course not. We ought not multiply senses of independence (or of anything else) *beyond need*. But in demonstrating the inadequacies of the Kolmogorov treatment, we have demonstrated that patently the need is there. Thales' theory that everything is made of water is spectacularly simpler than modern chemistry. That's hardly a reason to prefer it. As Einstein said: "Everything should be made as simple as possible, but not simpler."

9. Conditional Independence — The Plot Thickens

Things get even more interesting when we consider the notion of *conditional* independence. On the orthodox theory, this is again understood in terms of factorizing of probabilities. Propositions A and B are said to be *conditionally independent, given* C (according to a probability model $(\Omega, \mathcal{F}, P)$) iff

$$P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C).$$

Call this the *factorizing construal of conditional independence*. But, once again, on the orthodox view, this means that any proposition with extreme probability is conditionally *and* unconditionally probabilistically independent of all other propositions, for all conditions (with positive probability). This is incorrect for reasons we have given.

Moreover, the "equivalence" between factorizing and conditioning construals of independence breaks down even more easily in cases of conditional independence. One conditioning construal of conditional independence of A and B, given C (or 'C screening-off A

from B') is that $P(A \mid B \cap C) = P(A \mid \neg B \cap C)$. Of course, this expression is regarded as undefined on the orthodox theory whenever $P(B \cap C) = 0$, so the theory goes silent in such a case.[12] This is not a happy result. As an illustration, consider the following simple twist on the coin example. Let X be the proposition that the bias of the coin is on $[0, a]$, and Y be the proposition that the bias of the coin is on $[a, 1]$, with uniform distributions in each case. Intuitively, $Pr(H, X \cap Y) = a$, even though $Pr(X \cap Y) = 0$. And for all values of $a < 1$, we surely want to say that H and X are dependent, *given* Y. But, the orthodox theory cannot say this. The conditioning construal of conditional independence just mentioned does not say anything: $P(H \mid X \cap Y)$ is undefined. The factorizing construal of conditional independence says that H and X are independent, given Y: $P(H \cap X \mid Y) = 0 = P(H \mid Y) P(X \mid Y)$. These construals come apart from each other, and neither delivers the verdict that intuition demands.

Note that this is a case in which *none* of the three propositions (taken *individually*) has zero unconditional probability. It is a case in which the conditioning event "interacts" with one of the other events, so as to undermine the putative "equivalence" of the factorizing and conditioning construals of conditional independence in the orthodox theory. This is only more bad news for the orthodox accounts of conditional probability and independence.

But it is only more good news for the Popper-style accounts that we advocate, for again they can handle all the requisite conditional probabilities with ease. Again, we have two construals of 'H is conditionally independent of X, given Y', based respectively on the (Ω CONSTRUAL) and (NEGATION CONSTRUAL) respectively:

$$Pr(H, X \cap Y) = Pr(H, Y).$$

$$Pr(H, X \cap Y) = Pr(H, \neg X \cap Y).$$

---

[12] An important special case occurs when C itself has zero unconditional probability. When this happens, *no event* can be conditionally independent (or dependent) of any other event, given C. The example below is even more compelling than this special case, since none of its individual propositions has zero probability.

There is no impediment to imposing the constraint that $\Pr(H, X \cap Y) = a$. So we have (where $a < 1$):

$$\Pr(H, X \cap Y) = a \neq \Pr(H, Y) = \Pr(H, \text{bias of coin is on } [a, 1]) = (1 + a)/2.$$

H is dependent on X, given Y, based on the (Ω CONSTRUAL). Moreover,

$$\Pr(H, X \cap Y) = a \neq \Pr(H, \neg X \cap Y) = \Pr(H, \text{bias of coin is on } (a, 1]) = (1 + a)/2.$$

H and X are adjudicated as conditionally dependent, given Y (for $a < 1$), based on the (NEGATION CONSTRUAL). Either way, our intuition is upheld.

10. A call to arms

We conclude that it is time to bring to an end the hegemony of Kolmogorov's axiomatization, and with it, the Kolmogorovian account of independence. We seek independence, as it were, from that account of independence. Popper's axiomatization, and the conditioning construal of independence that it inspires, represent more promising alternatives. Long live the revolution![13]

Appendix: Proof of (NEGATION CONSTRUAL) $\Rightarrow$ (Ω CONSTRUAL) $\Rightarrow$ (POPPER FACTORIZATION).

First, we prove that (Ω CONSTRUAL) $\Rightarrow$ (POPPER FACTORIZATION). Indeed, we'll prove the following more general result (the result in question is a special case of the following, with C $= \Omega$):

$$\Pr(A, B \cap C) = \Pr(A, C) \Rightarrow \Pr(A \cap B, C) = \Pr(A, C)\,\Pr(B, C)$$

Assume $\Pr(A, B \cap C) = \Pr(A, C)$. Then, by Popper's product axiom P4, we have

$$\Pr(A \cap B, C) = \Pr(A, B \cap C)\,\Pr(B, C) = \Pr(A, C)\,\Pr(B, C). \text{ QED.}$$

Now, we prove that (NEGATION CONSTRUAL) $\Rightarrow$ (Ω CONSTRUAL). That is, by logic,

$Pr(A, B \cap \Omega) = Pr(A, \neg B \cap \Omega) \Rightarrow Pr(A, B \cap \Omega) = Pr(A, \Omega)$.

Now, $Pr(A, B \cap \Omega) = Pr(A, \neg B \cap \Omega)$ [Assumption]

Thus,

$Pr(A, B \cap \Omega) Pr(B, \Omega) = Pr(A, \neg B \cap \Omega) Pr(B, \Omega)$ [algebra]

But also,

$Pr(A, B \cap \Omega) Pr(B, \Omega) = Pr(A \cap B, \Omega)$ [Popper's product axiom P4]

Thus,

$Pr(A \cap B, \Omega) = Pr(A, \neg B \cap \Omega) Pr(B, \Omega)$,

and so

$Pr(A \cap B, \Omega) = Pr(A, \neg B \cap \Omega) (1 - Pr(\neg B, \Omega))$ [Popper's additivity axiom P3]

> (This axiom implies $Pr(B, \Omega) + Pr(\neg B, \Omega) = 1$,
>
> since it is not that case that for all X, $Pr(X, \Omega) = 1$. The fact that
>
> there exists a X such that $Pr(X, \Omega) \neq 1$ is proven as lemma 4(t)
>
> in Roeper and Leblanc (1999, page 198).)

$Pr(A \cap B, \Omega) = Pr(A, \neg B \cap \Omega) - Pr(A, \neg B \cap \Omega) Pr(\neg B, \Omega)$ [algebra]

$Pr(A \cap B, \Omega) = Pr(A, \neg B \cap \Omega) - Pr(A \cap \neg B, \Omega)$ [Popper's product axiom P4]

$Pr(A \cap B, \Omega) + Pr(A \cap \neg B, \Omega) = Pr(A, \neg B \cap \Omega)$ [algebra]

$Pr(A, \Omega) = Pr(A, \neg B \cap \Omega)$

> [It can be shown that Popper's axioms imply $Pr(A, \Omega) = Pr(A \cap B, \Omega) + Pr(A \cap \neg B, \Omega)$,
>
> since it is not the case that for all X, $Pr(X, \Omega) = 1$ (as above).
>
> This is proved as Lemma 4(i) in Roeper and Leblanc (1999, page 197).]

$Pr(A, \Omega) = Pr(A, B \cap \Omega)$ [by our assumption] *QED*

References

Arntzenius, Frank (1992) "The common cause principle", *Proceedings of the 1992 PSA conference*, Vol 2, 227-37.

Billingsley, Patrick (1995) *Probability and Measure*, 3rd ed., New York: John Wiley & Sons.

Carnap, Rudolf (1950) *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Carnap, Rudolf (1952) *The Continuum of Inductive Methods*. Chicago: The University of Chicago Press.

Easwaran, Kenny (2014) "Regularity and Hyperreal Credences", *The Philosophical Review* 123 (1), 1-41.

Etchemendy, John (1990) *The Concept of Logical Consequence*, Harvard University Press.

Feller, William (1968) *An Introduction to Probability Theory and Its Applications*, New York: John Wiley & Sons.

Hájek, Alan (2003) "What Conditional Probability Could Not Be", *Synthese* 137 (3), 273-323.

Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse Der Mathematik*. Translated as *Foundations of Probability*. New York: Chelsea Publishing Company, 1950.

Lewis, David (1979) "Counterfactual Dependence and Time's Arrow", *Noûs,* Vol 13, 455-76.

Loève, Michel (1977) *Probability theory*. *I*, *4th edition*, New York: Springer–Verlag.

Papoulis, A. (1965) *Probability, random variables, and stochastic processes*, New York McGraw-Hill.

Parzen, E. (1960) *Modern probability theory and its applications*, New York-London: John Wiley & Sons, Inc.

Pfeiffer, P. (1990) *Probability for Applications*, New York: Springer-Verlag.

Popper, Karl (1959) *The Logic of Scientific Discovery*. London: Hutchinson & Co.

Pruss, Alexander (2013) "Probability, regularity and cardinality", *Philosophy of Science* 80, 231–240.

Rényi, Alfréd (1955) "On a New Axiomatic Theory of Probability." *Acta Mathematica Academiae Scientiarum Hungaricae,* Vol 6, 285–335.

Roeper, Peter, and Leblanc, Hugues (1999) *Probability Theory and Probability Logic*. Toronto: University of Toronto Press.

Ross, Sheldon (1998) *A First Course in Probability, 5th edition*, Upper Saddle River, NJ: Prentice Hall.

Rozanov, Y. A. (1977) *Probability theory*, Revised English edition, Translated from the Russian and edited by Richard A. Silverman, New York: Dover.

Seidenfeld, Teddy, Mark J. Schervish and Joseph B. Kadane (2001) "Improper Regular Conditional Distributions", *The Annals of Probability,* Vol. 29, No. 4, 1612-1624.