Kolmogorov's Axiomatisation and its Discontents

AIDAN LYON[1]

January 30, 2012.

For *The Oxford Handbook of Probability and Philosophy*

(Draft — please do not cite without permission.)

> "[... T]here is little disagreement about the truth of the theory—indeed, it would not be an exaggeration to say that the theory of probability is commonly regarded as though it were necessarily true." Humphreys [1985], p. 568.

1 Introduction

Discontent with Kolmogorov's axioms? Surely discontent with The Axioms is as insane as discontent with The Truths of Logic. Indeed, some have argued that Kolmogorov's axioms *are* logic, the logic of partial belief, a natural—and the only—generalisation of classical logic, the logic of full belief (e.g., Jaynes [2003]).

Of course, many philosophers and logicians have expressed much discontent with classical logic. Admittedly, many philosophers and logicians are perhaps not quite bastions of sanity; however, their reasons for revising classical logic are perfectly sane. There are significant issues with vagueness, indeterminacy, inconsistency, and the like that all cause classical logic problems. Any such revision of classical logic can underpin a corresponding revision of the probability axioms. The essay, *Probability and Non–Classical Logics*, in this volume, focuses on such non–classical probability theories and reasons for adopting them. This chapters focuses mostly on other sources of discontent. Discontent with Kolmogorov's axioms does not necessarily stem from discontent with classical logic.

Even though it is orthodoxy that Kolmogorov's axioms are correct, and perhaps even necessarily correct, they appear to be incompatible with the most common so–called "interpretations" of probability. Finite actual frequencies, infinite hypothetical frequencies, propensities, degrees of entailment, and even rational partial belief all appear to fail to satisfy Kolmorogov's axiomatisation of probability. Or, to put it from another perspective: Kolmogorov's

[1] aidanlyon@gmail.com

axioms appear to fail to satisfy each of the most common theories of what probabilities are. Each of these failures is a source of possible discontent. In addition to this, there are reasons for discontent that seem to be independent of any interpretation of probability, and there is reason for discontent from the sciences too.

This paper will survey the most common reasons to be discontented with Kolmorogov's axiomatisation. Along the way, in the axiomatisation's defence, I'll discuss possible responses that one can make to each expression of discontent. First, though, we should be clear as to what all the discontent is about.

2 Kolmogorov's Axiomatisation of Probability

Let Ω be a non–empty set, over which $\mathscr{F}$ is an algebra. Let P be a function from $\mathscr{F}$ to $\mathbb{R}$. If P satisfies the axioms:

$$\begin{aligned}
&\text{(K1)} \quad P(A) \geq 0 \\
&\text{(K2)} \quad P(\Omega) = 1 \\
&\text{(K3)} \quad P(A \cup B) = P(A) + P(B), \text{ if } A \cap B = \varnothing
\end{aligned}$$

for every A and B in $\mathscr{F}$, then P is a *probability function*, and $(\Omega, \mathscr{F}, P)$ is a *probability space*. These axioms are often called *non–negativity*, *normalization*, and *finite additivity*, respectively.

If $\mathscr{F}$ is a σ–algebra, finite additivity is extended to so–called *countable addivity*:

$$\text{(K3')} \quad P\left(\bigcup_{i=1}^{i=\infty} A_i\right) = \sum_{i=1}^{i=\infty} P(A_i)$$

where the A_i are mutually disjoint.

Kolmogorov then adds a definition of conditional probability to the above axioms:

$$\text{(CP)} \quad P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ where } P(B) > 0$$

for every A and B in $\mathscr{F}$. I shall call K1–3′ + CP *Kolmogorov's axiomatisation of probability*.

Before we move on to reasons why one might be unhappy with Kolmogorov's axiomatisation, it will be worthwhile to pause for a moment, and think about the relation between the formal theory of probability and the concept of probability.

3 Where to Place the Blame?

Komogorov's axiomatisation is quite remarkable in many ways. It situates probability theory as a branch of a more general mathematical theory, the theory of measures. And its three simple axioms have resulted in numerous remarkable theorems, which have countless important applications throughout the sciences. It is simultaneously incredibly simple, unified, and powerful. And yet, there are reasons to be dissatisfied with it nonetheless.

It seems, *prima facie*, strange to be discontented with a piece of mathematics. One can be discontented with all sorts of things—politicians, the dinner menu, software, society, etc.—but these are all things that are changeable in some sense. The facts of mathematics, on the other hand, are apparently unchangeable. Being discontented with a piece of mathematics would therefore seem to be on par with wishing that "up" was "down".

However, discontent with mathematics can and has led to significant progress. Mathematicians have expressed discontent with their theories for a variety reasons. They sought the Zermelo–Fraenkel axioms to avoid the inconsistency of naïve set theory. And then there was—and still is—the Axiom of Choice controversy, with some arguing we should accept it based on its necessity for several important theorems, and others arguing that we shouldn't because it leads to counterintuitive objects or that it isn't acceptable on constructivist grounds.

Such expressions of discontent arise when the truth of the relevant mathematics is clearly in question. Inconsistent set theory cannot possibly be true (according to classical logic), and there was *a priori* epistemic uncertainty surrounding the Axiom of Choice. Sometimes the uncertainty surrounding the mathematics in question has an empirical or "external" flavour to it. Perhaps the clearest example of this comes from Geometry. Initially, the Parallel Postulate was called into doubt, as an axiom, because of its lack of self–evidence. Moreover, once it was realised that the postulate is false for the actual physical space that we happen to inhabit, it was quickly rejected, and so–called non–Euclidean geometries quickly flourished.

As we'll see, the reasons for discontent with Kolmogorov's axiomatisation are often external in this way—i.e., they come from particular applications of the theory. Applications include capturing the epistemic norms of graded belief, the behaviour of chance, and statistical reasoning. It is when the axiomatisiation is used in these applications that problems arise and discontent may ensue. Just as the truth of the Parallel Postulate for our actual physical space was called into question, we can ask if, e.g., countable additivity is true for the epistemic norms of peoples' degrees of belief, or if the definition of conditional probability is true for conditional physical chance, etc.

In any particular application of Kolmogorov's axiomatisation, it is important to keep in mind what exactly the link is between the formalism and the target of the formalism. I began this article somewhat dramatically, with an analogy between probability and logic. This is a useful analogy to keep in mind as we consider various sources of discontent with Kolmorogov's axiomatisation. As already mentioned, the sources of discontent often stem from particular applications of the axioms. In any such application, there has to be some link between the purely abstract axioms and the target of the application. For example, we will often want P to represent an agent's *degrees of belief*. Call such a link a *bridge principle*. Often, when things go wrong, there are two places where the blame can be laid: with the bridge principle in question or with the axiomatisation.

Compare this with a potential source of discontent with classical logic. According to classical logic, $A \wedge \neg A$ entails everything—this is known as *explosion*. Why is explosion bad? Suppose a person has inconsistent beliefs. If classical entailment models the normative closure of belief, then that person ought to believe everything. Some authors have concluded from this absurdity (along with other considerations) that we need to revise classical logic, and adopt some non–classical logic, in which explosion does not hold (e.g., Meyer [1971], pp. 814–5). However, as Harman has pointed out, we could equally well question the bridge principle involved: that classical entailment models the normative closure of belief (Harman [1986], p. 6).

Even when we're sure that the axioms deserve the blame, it may not be clear which axiom deserves that blame. No axiom is an island, entire of itself; the axioms work in tandem to produce the results of Kolmogorovian probability theory—whether those results are preferable or not. And so when trouble arises, more than just one axiom can often be blamed. Moreover, the blame can even be placed with something that the axioms presume—e.g., that probability values are real numbers, that the objects of probability are sets, that they form an algebra, etc.

Finally, there is a third place where the blame can be laid: with the application itself. As we'll see, some interpretations of probability have been criticised for not satisfying Kolmogorov's axiomatisation. Is this a problem for the interpretation, or for the axiomatisaion—or for the bridge principle that links them? It's hard to say, but as we'll see, almost *every* interpretation has some degree of discomfort with Kolmogorov's axioms—even finite frequentism!

4 Discontent with Countable Additivity

Consider a fair lottery with denumerably many tickets. Since the lottery is fair, each ticket has equal probability of being drawn. But there are only two ways in which this can happen, and on both ways, trouble looms. On the first way, each ticket has some positive probability of winning, and any positive probability added to itself denumerably many times is (much!) larger than one, and so a violation of the axioms. On the second way, each ticket has zero probability of being drawn. Zero added to itself denumerably many times is zero, which is less than one, and so also a violation of the axioms. It's a case of too much or too little: on both ways of the lottery being fair, the probability that *some* ticket is drawn is either greater than one or less than one, and so there is a violation of the axioms. It seems, then, that Kolmogorov's axiomatisation rules out—*a priori*—the possibility of a fair, countably infinite lottery. (Incidentally, the problem is reminiscent of Zeno's Paradox of Plurality. See e.g., Salmon [2001], pp. 13–15.)

What's causing the problem? The argument—originally due to de Finetti [1974]—is typically used as an objection to K3′, the axiom of countable additivity. However, one might also be use it to object to K2, the axiom of normalisation, which is clearly involved in the generation of the problem. Both of these options would be to blame the axioms. One can also try to blame the application. Some have argued that there is no problem here because there is no physical device that could set up such chances (Spielman [1977], Howson and Urbach [1989]). However, a response to this is that it doesn't matter whether such a physical device exists or not; a rational agent should be allowed to assign equal credence to each ticket winning (Williamson [1999]).

Can one blame the bridge principle? Yes. To set up the problem in a complete and precise way, one typically has an algebra that includes denumerably many events, one of each corresponding to each of the denumerably many tickets winning. One can then prove that there is no uniform probability distribution over those events. An alternative way to approach the problem, and so an alternative bridge principle, would be to start with a finite algebra, and a uniform distribution (which is unproblematic), and let the algebra increase to any arbitrary–but finite–size. (Jaynes [ref].) This is an unconventional line of thought, but it shows that there is a way to question the bridge principle involved, i.e., how one should model the situation (whether it be chances or someone's ideal credences) in the mathematics.

Another way to blame the bridge principle is to say that, in this particular situation, there is *no* bridge principle between our credences and Kolmogorov's axioms. As Bartha [2004] points out, one may argue that we don't have real–valued degrees of belief in this situation (pp. 306–7). If, instead, we only have relative probabilities[2], then one can maintain that the axiom of countable additivity is true because it doesn't apply in the case of de Finetti's lottery (pp. 309-10). (See Bartha [2004] for a more detailed discussion and constraints on relative probabilities.)

5 Discontent with Finite Additivity

Although on much safer ground, finite additivity has its discontents too. The structure it imposes on probability makes it difficult for probability to represent states of pure ignorance.

Consider a situation in which we are reasoning about whether some event, E, will occur or not. If we know nothing about E, we have no evidence for or against it, we have no convictions or intuitions about it, then it it seems to be unreasonable to presume E to be more likely than $\neg E$, or vice versa. So if we are to assign probabilities to E and $\neg E$, we have to give them *equal* probabilities. It then follows, from finite additivity, that they both have to be given a probability of $1/2$.[3] Compare this with a situation in which we know lots about E and $\neg E$—e.g., they are the outcomes of a fair flip of your lucky dime. In both situations, the probability calculus represents you as equally confident of E. And yet this seems not to be so.

Worse still is that the above reasoning can lead to paradox. If a cube factory produces cubes and those cubes can have side lengths between 0 and 1m, what is your probability that the next cube has its side length between 0 and 1/2m? Either it has this side length, E, or it does not, $\neg E$. By the reasoning above, the probability of E is therefore $1/2$. But if the cube factory produces cubes and those cubes have face areas between 0 and 1m^2, what is your probability that the next cube has its face area between 0 and 1/4m^2? Here the reasoning is slightly different, but similar enough. It seems there are four events for which we cannot suppose any to be more probable than any other: the face area is between (i) 0 and $1/4$, (ii) $1/4$ and $1/2$, (iii) $1/2$ and $3/4$, and (iv) $3/4$ and 1. If these all have equal probability, then, by finite additivity, they all have probability of $1/4$. But now we have a contradiction, for $E \equiv$ (i) (a side length of $1/2$ is the same as a face area of $1/4$), and yet we have given them different probabilities. This example is taken from van Fraassen [1989], and is representative of a larger class of paradoxes—see also e.g., Keynes [1921] and Jaynes [1973].

As usual, there are many places where the blame for these problems can be laid, and

[2] Quick explanation of relative probabilities here.

[3] From equal probabilities, we have $P(E) = P(\neg E)$; from finite additivity we have $P(E) + P(\neg E) = 1$; and putting these together we get: $2P(E) = 1$, $2P(\neg E) = 1$, and so $P(E) = 1/2$ and $P(\neg E) = 1/2$.

one place that can be blamed is finite additivity. If we drop finite additivity, then from $P(E) = P(\neg E)$ it does not follow that $P(E) = P(\neg E) = 1/2$. It is now permissible that, say, $P(E) = P(\neg E) = 0$. And perhaps that is the right probability assignment in a case of total ignorance. If probability is credence and you have no evidence or prior knowledge, then there is nothing to lend any credence to E or $\neg E$; you therefore have no credence in E or $\neg E$. Assigning vacuous probabilities to all of the possibilities for which you are completely ignorant about avoids the above contradiction, and it also distinguishes you from the person who is well–informed about the options, but happens to have uniform credences.

One popular theory that drops finite additivity is Dempster–Shafer theory, sometimes called the theory of *belief functions* (see e.g., Shafer [1976]). See also Ghirardato [2001] for reasons for adopting non–additive probabilities due to cases of ignorance.

6 Discontent with Conditional Probability

According to Kolmogorov's definition of conditional probability, $P(A|B)$ is undefined whenever $P(B) = 0$. As many authors have noted there are countless situations where this appears to be false.

Suppose we choose a point randomly from the surface of the Earth, where we assume the Earth is a perfect sphere. What is the probability that the point is in the Western Hemisphere, given that it is on the Equator? The answer is surely $1/2$, and yet according to Kolmogorov, it is undefined because the probability of the point being on the Equator is 0. (In this example, unconditional probability corresponds to relative area of the surface of the Earth. The event of the point being on the Equator gets a probability of 0 because the Equator is a line and therefore has area of 0.)

What's the cause of this problem? The natural response is that the problem is due to the overly restrictive definition of conditional probability. Indeed, Kolmogorov himself was fully aware of this sort of problem and developed are more general, and complicated, definition of conditional probability (Kolmogrov [1933], p. X). However, this complicated definition of conditional probability runs into problems of its own (see e.g., Seidenfeld [2001] or Hájek [2003], p. X).

Other authors have developed alternative axiomatisations that take conditional probability as primitive and define unconditional probability in terms of it (e.g., Rényi [1955], Popper [1959], pp. X–Y). Such axiom systems allow $P(A|B)$ to be given a definite value even if $P(B|\Omega) = 0$, which in these contexts is typically understood as the unconditional probability of B.

Another common response to the problem is to argue that in cases like the Equator example, $P(B)$ is not really equal to 0. This response places the blame with the assumption that probabilities are real–valued, and so I'll postpone discussion of it until section 7, where I discuss discontent with this assumption more generally.

Also according to Kolmogorov's definition of conditionality probability, $P(A|B)$ is undefined whenever the unconditional probabilities in which it is defined are undefined. What is the probability that Joe get "heads", given he flips a fair coin? Answer: 1/2. But what is the probability that Joe flips a fair coin? Answer: Who knows?! Whether Joe flips the coin may be a matter of free will, in which case there may be no chance associated with it (Hájek [2003], p. X). Never mind matters of free will, there just may be no unconditional probability of Joe flipping the coin. There may be a *conditional* probability of Joe flipping *given* that someone asks Joe nicely to flip it, or pays him, or if he is in a coin flipping mood, or if he has sworn to never flip coins again, etc. But a free–floating unconditional probability of Joe flipping the coin? It's not at all clear that such a probability exists; and even if it does, if by probability we mean rationally required degree of belief, then there may nevertheless be no such thing anyway. It seems you can have a conditional degree of belief of 1/2 that the coin lands "heads" given Joe flips the coin fairly and not be rationally required to have some degree of belief that Joe flips the coin fairly.

The problem seems to be due to Kolmogorov's definition of conditional probability in terms of unconditional probability. Perhaps the most natural response, then, is that we should axiomatise conditional probability directly, instead of defining it in terms of unconditional probability. This approach also seems to have the advantage of also solving the problem from the Equator example. Hájek, for example, concludes that we should move to an alternative axiom system that takes conditional probability as primitive (Hájek [2003], pp. 315–6). Unfortunately, many (if not all) such alternative axiom systems don't really solve the problem. Many of the alternative axiom systems require that probabilities like P(I flip a fair coin | I flip a fair coin or I don't) be defined. But for the same reasons as before, it seems there is intuitively no such probability. The problem is therefore not Kolmogorov's alone and it seems that something more drastic has to be done.

So far I have been focusing on problems where the conditional probabilities are undefined when they should be defined. But the opposite also happens. Just as problematic are cases where conditional probabilities are defined when they should be undefined, or defined in some other way. For example, according to the propensity interpretation of probability,

$P(E|C)$ is the propensity of C to produce, or cause, E. However, Humphreys' [1985] has shown that in many situations Kolmogorov's axiomatiation requires that if $P(E|C)$ has a value, then so does $P(C|E)$. This seems strange since presumably causes must precede effects, and yet Kolmogorov's axioms seem to tell us that an effect can have a propensity to produce its cause. Surely that is wrong: the lighting of a match in the evening has a propensity to burn down the factory at night, but the burning down of the factory at night doesn't have a propensity to light the match in the evening. This problem for Kolmogorov's understanding of conditional probability for the propensity interpretation is known as Humphreys Paradox. The standard response to this problem is that it is trouble for the propensity interpretation—i.e., the application gets the blame. Humphreys' own conclusion, however, is different: he blames the axiom system (Humphreys [1985], 568–9).

7 Discontent with Positive Real Numbers

When faced with undefined conditional probabilities or infinite lotteries, there is yet a another aspect of the axioms that we could blame: the assumption that probabilities are *real numbers*.

In both situations, we want a probability value that is really small, but not 0. The problem with the real numbers is that there is no number that is small enough but not so small that it is 0. This is easiest to see in de Finetti's lottery example: as soon as the probability of each ticket winning is greater than 0, then no matter how small that number is, the total probability will sum to more than 1. One response, then, is to give up on the real number system, and move to a richer system that has numbers that are smaller than every real number, but that are greater than 0. Such numbers are called *infinitesimals*, and the most cited number system that contains them is the system of hyperreals (e.g., Robinson [1966]).

If probability values are hyperreals, then we can say that each ticket in de Finetti's lottery has some infinitesimal probability of winning (e.g., Bartha and Hitchcock [1999]). Similarly, we can say that the event of choosing a point on the Equator has infinitesimal probability, and so the conditional probability of choosing a point in the Western Hemisphere given the chosen point is on the Equator doesn't go undefined (e.g., Lewis [1980], p. 267–8). It would seem that infinitesimals provide an elegant solution to two different problems.

In fact, there is even more good news for infinitesimal fans. It seems to be a dictum of rationality that one should be open minded: one should not assign zero probability to any event that one considers to be possible. This principle often goes by the name of *Regularity*.

According to Kolmogorov's axioms, though, it is mathematically impossible to satisfy this principle when dealing with uncountably many events (e.g., Hájek [2003], pp. 281–2). One is forced to assign zero probability to uncountably many events that one considers to be possible. However, this is not true if probability values can take on infinitesimal values.

So infinitesimals seem to solve three distinct sources of discontent with Kolmogorov's axioms. (Moreover, there is a sense in which adopting infinitesimals does not amount to a revision of the axioms, for the hyperreal system is a non–standard model *of the reals*. One could read Kolmogorov's axioms as leaving the model of the reals as unspecified, in which case we don't really have a case of discontent with the axiomatisation.) Infinitesimals have their problems, though. It's impossible to name one of them, for example. So when we say that each ticket in de Finetti's lottery has an infinitesimal probability of winning, *which* infinitesimal is it? It's strange that we can't answer that question (see Hájek [2003], pp. 292–3 for more details). More seriously, though, is that even if we allow probability values to be infinitesimals, we are *still* forced to assign zero probability to possible events (Williamson [2007]).

Instead of moving to a richer number system, some may want to move to a *poorer* number system. According to Kolmogorov's axiomatisation, $P(A) = 1/\pi$ is a legitimate probability assignment—i.e., there are probability functions that make such an assignment. Those who believe in the finite actual frequency interpretation of probability have to disagree. Nothing can occur with a $1/\pi$ relative frequency in a finite number of trials, and so $P(A) = 1/\pi$ cannot be a legitimate probability assignment. Finite frequentists have to insist that P is a function from $\mathscr{F}$ to $\mathbb{Q}$, and not $\mathbb{R}$. (Of course, this point is often used as an objection to finite frequentism (e.g., Hájek [1997], pp. 224–5)—but as usual, the sword can cut both ways.)

So far I have been discussing discontent with the assumption that probability values are real numbers. Kolmogorov's axiomatisation says something stronger, though: the axiom of non–negativity, KP1, makes sure that they are *positive* real numbers. Why the ban on negative numbers? One reason is that it is not at all clear how one would interpret them: if $Pr(A) = -0.5$, should we expect A to happen about minus fifty percent of the time? (!)

Despite our inability to make sense of negative probabilities, they nevertheless appear in the sciences. They appear in quantum mechanics, and even in classical physics (see Muckenhaim *et al.* [1986]). They also make an appearance in financial mathematics (Haug [2007], Burgin and Meissner [2011]) and machine learning (Lowe [2004]). Negative probabilities typically only appear as intermediate steps in calculations—much like $3 = 4 + (-1)$—or as probabilities of unobservables. However, see Burgin [2010] for a frequency interpretation of negative probabilities. Some physicists have sought ways to avoid negative probabilities by changing their equations, without changing the physics (e.g., [ref]). This amounts to changing the bridge principle between the physics and the mathematical theory of probability.

Another source of discontent with the probability values of Kolmogorov's axiomatisation is that it assumes that probabilities are point values. Many authors have argued for interval valued probabilities. For example, it seems rationally permissible to not have a point-valued credence in all propositions. Does your confidence that it will rain tomorrow have to be precise to infinitely many decimal points for you to be rational? Arguably, no it doesn't. Instead of someone having precise credences, an agent may have lower and upper bounds on their credences—forming so–called *indeterminate credences*. See Walley [1991], Hájek and Smithson [201X], and Levi [2000] for more details.

One final reason to be discontent with the probability values of Kolmorogov's axiomatisation is with the very assumption that there are probability *values*. Fine [1973] argues that a comparative probability axiom system (i.e., one that axiomatises "A is more probable than B") is more general and powerful than Kolmogroov's axiom system.

8 Discontent with Sets and Algebras

According to Kolmogorov's axiomatisation the objects of probabilities are sets. This immediately rules out any interpretation of probability that thinks the objects are sentences, or events, or even propositions, if propositions are not sets of something (e.g., worlds). For these reasons, Popper thought that a formal theory of probability should make no assumption regarding what the bearers of probability are:

> "In Kolmogorov's approach it is assumed that the objects a and b in $p(a, b)$ are sets (or aggregates). But this assumption is not shared by all interpretations: some interpret a and b as states of affairs, or as properties, or as events, or as statements, or as sentences. In view of this fact, I felt that in a formal development, no assumption concerning the nature of the 'objects' or 'elements' a and b should be made [...]." Popper [1959], p. 40. [check ref.]

Popper goes on to develop an axiom system that makes no such assumptions and has the virtue—in Popper's eyes—that the algebraic structure of the objects of probability emerge as a consequence from the axioms, rather than being built into them (e.g., Popper [1959], p. X).

However, some have argued that the objects of probability should not necessarily form an algebra, or a σ–algebra, so this property should be neither built into the axioms nor a consequence of them. For example, consider the following:

> "A class of photographs may be such that the probability of a predominantly dark pictures is $\frac{1}{2}$ and the probability of a grainy texture is also $\frac{1}{2}$. However, we may have no data on, and little interest in, whether grainy patterns tend to be dark or not. Should we then be prevented from using probability theory to model this sample source when designing pattern classifiers for this problem?" Fine [1973], p. 62.

Fine's point is that there are situations where we can assign probabilities to A and B but not to their union $A \cup B$ or intersection $A \cap B$. And yet Kolmogorov's axioms, because they require probabilities to be defined over an algebra, require these probabilities to be defined.

One alternative to the regular algebra is the Λ*–field*. The closure properties of a Λ–field allow that A and $B \in \Lambda$, but $A \cap B \notin \Lambda$. An example would be $\Lambda = \{\emptyset, \Omega, (a, b), (c, d), (a, d), (b, c)\}$, which could be interpreted as:

> A photograph is
> (a) dark colour and grainy texture
> (b) dark colour and smooth texture
> (c) light colour and smooth texture
> (d) light colour and grainy texture

Note that the event of a photograph being dark and grainy is not in Λ, and so it doesn't get a probability (Fine [1973], p. 63). The neat thing about using a Λ–field is that it only requires the introduction of events whose probabilities can be calculated from the given probabilities using K1–K3. A downside is that it doesn't allow us to define many conditional probabilities that seemingly ought to be defined—e.g., we can't say that $P(\text{dark colour}|\text{dark colour and grainy texture}) = 1$.

Another alternative to the σ–field is the von Mises field, the $\mathscr{V}$*–field*. The precise definition of the $\mathscr{V}$–field is complicated and the interested reader should see e.g., Fine [1973], pp. 64–5 for a definition. One important difference between the $\mathscr{V}$–field and the σ–field is that the former only contains events that can be measured using finite data sets. For example, for an indefinitely long series of coin flips, the event "heads occurs only finitely often" does not appear in the $\mathscr{V}$–field (*ibid*). Another important difference is that the set of events with hypothetical limiting frequencies form a $\mathscr{V}$–field but not a σ–field (nor even a field). (von Mises was a hypothetical frequentist.)

9 Conclusion

Who could be discontented with Kolmogorov's axiomatisation of probability? The answer seems to be: almost everyone! We have seen reasons to be discontent with almost every aspect of Kolmogorov's axiomatisation: that the objects of probabilities are *sets*, and that they form an *algebra* of sets; that probability values are *numbers*, that they are *real* numbers, that they are *positive* real numbers, that they have an *upper bound*, and that there are even probability *values*; that probabilities are *countably additive*, and even that they are *additive*; and finally that conditional probabilities are *ratios of unconditional probabilities*.

With so much discontent, how has Kolmogorov's axiomatisation achieved its level of orthodoxy? Perhaps the reason is that the axiomatisation is the best compromise between many competing demands on the formal theory of probability. It's a wonder that it has such a wide domain of applicability to a concept that we understand so little and use so widely. Probability is everywhere, the guide to life as some have (often) said, and indispensable to the sciences; and yet it is notoriously difficult to analyse. It's somewhat amazing, then, that there is a such a successful formal theory of something that is so conceptually messy.

Nevertheless, we might wonder why probability shouldn't go the way of geometry, with different axiomatisations appropriate for different applications. Put this way, Kolmogorov's axiomatisation is the Euclidean geometry of probability. It's still useful for many purposes and often a good first approximation, but there are other axiomatisations that are better suited for particular tasks.

References

Bartha, P. (2004). Countable Additivity and the de Finetti Lottery. *British Journal for the Philosophy of Science 55*, 301–321. 5, 6

Bartha, P. and C. Hitchcock (1999). The Shooting-Room Paradox and Conditionalising on "Measurably Challenged" Sets. *Synthese 118*(403–37). 9

Burgin, M. (2010). Interpretations of negative probabilities. *Arxiv preprint arXiv:1008.1287*. 10

de Finetti, B. (1974). *Theory of Probability*, Volume 1. (1990) Wiley Classics Library, John Wiley and Sons. 5

Fine, T. (1973). *Theories of Probability*. New York: Academic Press. 11, 12

Ghirardato, P. (2001). Coping with Ignorance: Unforseen Contingencies and Non–Additive Uncertainty. *Economic Theory 17*, 247–276. 7

Hájek, A. (1997). 'Mises Redux'–Redux: Fifteen Arguments Against Finite Frequentism. *Erkenntnis 45*, 209–227. Available from: http://www.springerlink.com/content/k7j628r482648205/. 10

Hájek, A. (2003). What Conditional Probability Could Not Be. *Synthese 137*(3), 273–323. 7, 8, 10

Hájek, A. and M. Smithson (201X). Rationality and Indeterminate Credences. *Synthese (forthcoming)*. 11

Harman, G. (1986). *Change in View: Principles of Reasoning*. MIT Press. 4

Haugh, E. G. (2007). *Derivatives Models on Models*. John Wiley & Sons, New York. 10

Howson, C. and P. Urbach (1989). *Scientific Reasoning* (1 ed.). La Salle, Ill : Open Court. 5

Humphreys, P. (1985). Why Propensities Cannot Be Probabilities. *Philosophical Review 94*, 557–570. 1, 9

Jaynes, E. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press Cambridge. 1

Jaynes, E. T. (1973). The Well Posed Problem. *Foundations of Physics 4*(3), 477–492. 6

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan. 6

Kolmogorov, A. N. (1933). *Foundations of Probability*. (1950) Chelsea Publishing Company. 7

Levi, I. (2000). Imprecise and Indeterminate Probabilities. *Risk, Decision and Policy 5*, 111–22. 11

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*, Volume II. Berkeley: University of California Press. 9

Lowe, D. (2004). Machine learning, uncertain information, and the inevitability of negative 'probabilities'. In *Machine Learning Workshop 2004*, Volume (http://videolectures.net/mlws04$_lowe_mluii/).Sheffield,England.$10

Meissner, G. and M. Burgin (2011). Negative probabilities in financial modeling. 10

Meyer, R. (1971). Entailment. *The Journal of Philosophy*, 808–818. 4

Muckenheim, W., G. Ludwig, C. Dewdney, P. Holland, A. Kyprianidis, J. Vigier, N. Petroni, M. Bartlett, and E. T. Jaynes (1986). A Review of Extended Probability. *Physics Reports 133*, 337–401. 10

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Basic Books: New York. 7, 11

Rényi, A. (1955). On a New Axiomatic Theory of Probability. *Acta Mathematica Academiae Scientiarum Hungaricae 6*, 286–335. 7

Robinson, A. (1966). *Non-Standard Analysis*. North Holland Publishing Company, Amsterdam. 9

Salmon, W. C. (2001). Introduction. In W. C. Salmon (Ed.), *Zeno's Paradoxes*. Hackett Publishing Company. 5

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (2001). Improper Regular Conditional Distributions. *The Annals of Probability 29*(4), 1612–1624. 7

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press. 7

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Oxford University Press. 6

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman Hall. 11

Williamson, T. (2007). How Probable is an Infinite Sequence of Heads? *Analysis 67*(295), 173–180. 10