The Borel-Kolmogorov paradox is your paradox too:

A puzzle for conditional physical probability*

Alexander Meehan and Snow Zhang†

November 23, 2020

Abstract

The Borel-Kolmogorov paradox is often presented as an obscure problem that certain mathematical accounts of conditional probability must face. In this paper, we point out that the paradox arises in the physical sciences, for physical probability or chance. By carefully formulating the paradox in this setting, we show that it is a puzzle for everyone, regardless of one's preferred probability formalism. We propose a treatment which is inspired by the approach that scientists took when confronted with these cases.

1   Introduction

A projectile is about to be launched at an obstacle. Suppose the collision is a chancy process, such that X, the speed of the projectile right after impact, might range anywhere from zero

---

†Contact: alexandermeehan@princeton.edu, xueyinz@princeton.edu. Affiliation: Department of Philosophy, Princeton University, 1879 Hall, Princeton, NJ 08544, USA.

to one meters per second, with each value equally likely. For instance, the probability of *fast*, the event that X exceeds half a meter per second, is $1/2$. At the same time, a second projectile is to be launched toward another obstacle. The possible values of Y, the speed of this projectile right after impact, also range from zero to one meters per second, with each value equally likely. Suppose X and Y are causally and probabilistically independent. Given this set-up, what is $P(X > 1/2|X = Y)$? In other words, what is the probability of *fast* conditional on *same speed*?

A natural thought is that the probability of *fast* should not be affected by *same speed*: $P(X > 1/2|X = Y) = P(X > 1/2) = 1/2$. Perhaps you don't share this thought, however, and are unready to commit to an answer until you have performed some calculations. In that case, you may still be willing to endorse the following weaker claim:

Uniqueness In this set-up, the physical probability of *fast* given *same speed* is well-defined and determinable based on the information given above.[1]

In support of **Uniqueness**, note that *fast* and *same speed* are consistent, well-defined physical events. It would be odd to insist that there is no fact of the matter regarding the chance of one given the other. Moreover, it seems we have fully specified the relevant probabilistic and causal structure of the set-up. What further information is needed?[2]

Uniqueness is an attractive assumption. However, consider the following equally, if not more, attractive claim:

Equivalence If B is necessarily equivalent to *same speed*, then the physical probability of *fast* given B equals the probability of *fast* given *same speed* (if well-defined).[3]

The thought is that if A, B are necessarily equivalent then one cannot happen without the

---

[1]At least for now, we stay neutral on whether determinability is understood in a metaphysical sense (e.g. supervenience) or an epistemic sense (e.g. derivability).

[2]To lend credence to the second clause, if $P(X = Y) > 0$, then the probability of *fast* given *same speed* is indeed determinable based on the information specified above.

[3]Here we have in mind nomological (or physical) necessity, though similar puzzles will arise even if we strengthen to metaphysical or logical necessity.

other, and so (for example) if the occurrence of A raises the chance of C, then the occurrence of B must raise the chance of C, and vice-versa.

Unfortunately, **Uniqueness** and **Equivalence** are jointly untenable. Moreover, the problem is not restricted to this case. The projectile example was designed to be mathematically simple, but there are less artificial examples where the same issue arises. Later in the paper, we will see an example from statistical mechanics (Kac and Slepian, 1959).

Although this formulation in terms of physical probability and **Uniqueness** and **Equivalence** is new, the mathematical aspect of this puzzle is familiar. It is a *Borel-Kolmogorov-type paradox*, in which conditioning on the same event B, where $P(B) = 0$, seems to lead to different values of $P(A|B)$ depending on how B is 'viewed' (Kolmogorov, 1933; Rao, 1988). How to interpret this phenomenon is a controversial question. For instance, Kolmogorov suggests that this type of paradox "shows that the concept of a conditional probability with regard to an isolated given hypothesis whose probability equals 0 is inadmissible". Others see the paradox as a problem internal to Kolmogorov's probability formalism (Hájek, 2003; Kadane et al., 1999; cf. Gyenis et al., 2017).

Why rehash this well-known paradox in the physical probability setting? We believe the paradox gains new force and significance when considered in the context of the physical sciences. After presenting the conflict between **Uniqueness** and **Equivalence**, we point out that the paradox arises regardless of whether we use the standard Kolmogorov formalism or axiomatize conditional probability directly. We also point out that the paradox has a new urgency, because it calls into question the reliability of several of scientists' probability judgments. We then propose a way to make headway on resolving the paradox. The proposal is inspired by the approach that scientists took toward an instance in statistical mechanics.

2 Uniqueness and Equivalence

In this section we survey multiple methods for determining the probability of *fast* given *same speed* and show that, given **Equivalence**, they each yield incompatible answers depending on how they are applied, suggesting a conflict between **Uniqueness** and **Equivalence**.

2.1 Limits

In elementary probability theory, there is a standard method for computing conditional probabilities given probability zero events, which we will call the *limits* method. Roughly speaking, one calculates $P(A|B)$ as the limit of $P(A|B_n)$ where the B_n are an appropriately chosen sequence of positive-probability events that converge to B.

Recall the projectile set-up above. The idea is to consider the positive probability event that X is *close* to Y, $|X - Y| \leq \epsilon$, and then compute the limit of the conditional probability as ϵ shrinks. Under our assumption that X and Y are independent and both uniformly distributed, the calculation,

$$
\begin{aligned}
P(X > 1/2 | X = Y) &= \lim_{\epsilon \to 0} P(X > 1/2 \,|\, |X - Y| \leq \epsilon) \\
&= \lim_{\epsilon \to 0} \frac{P(X > 1/2, |X - Y| \leq \epsilon)}{P(|X - Y| \leq \epsilon)},
\end{aligned}
$$

yields $1/2$, as expected. To see why, we can picture $|X - Y| \leq \epsilon$ as a band of thickness ϵ around the diagonal line $X = Y$, as shown in Figure 1a. The conditional probability of *fast* given this event "*close speed*", one takes the proportion of this band that overlaps with the right-hand side of the square (the area corresponding to $X > 1/2$). By the symmetry of the band, this proportion is clearly $1/2$, and will stay at $1/2$ as the thickness of the band shrinks.

Crucially, however, different ways of quantifying the "closeness" between the two speeds leads to different results when the same method is applied.[4] We have considered whether the

---

[4]The mathematical backbone of this case is due to Lindley (1982) and Rescorla (2015).

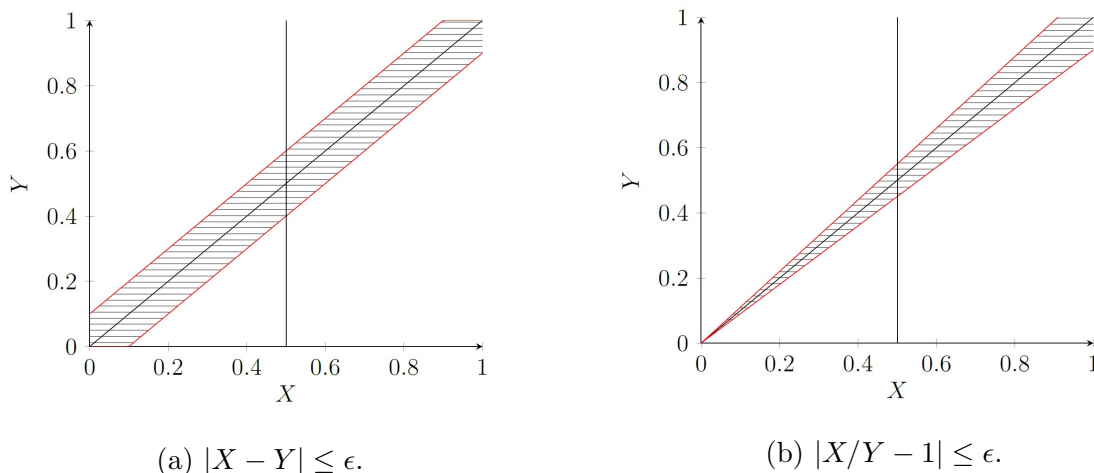(a) $|X - Y| \leq \epsilon$.   (b) $|X/Y - 1| \leq \epsilon$.

Figure 1: Two measures of closeness.

difference between the two recoil speeds is small, $|X - Y| \leq \epsilon$. But we might also consider whether their ratio is almost 1, $|X/Y - 1| \leq \epsilon$. And if we instead consider the limit as this ϵ shrinks, the calculation,

$$P(X > 1/2 | X = Y) = \lim_{\epsilon \to 0} P(X > 1/2 \,|\, |X/Y - 1| \leq \epsilon)$$

yields a value greater than $1/2$. Why? Note that smaller speeds are less likely to have smaller ratios (for instance, 0.01 and 0.001 have a ratio of 10, whereas 0.91 and 0.901 have a ratio near 1), and so X/Y being small raises the probability X is large. The situation is represented in Figure 1b. The lopsided band corresponds to the event $|X/Y - 1| \leq \epsilon$. The proportion of the band that overlaps with the right-hand side is clearly greater than the proportion that overlaps with the left-hand side.

Let's make the tension between these two results more precise. On the one hand, defining $Z = X/Y$, the limits method implies that:

1. $P(X > 1/2 \,|\, |X - Y| = 0) = \lim_{\epsilon \to 0} P(X > 1/2 \,|\, |X - Y| \leq \epsilon) = 1/2$, and

2. $P(X > 1/2 \,|\, |Z - 1| = 0) = \lim_{\epsilon \to 0} P(X > 1/2 \,|\, |Z - 1| \leq \epsilon) > 1/2$.

On the other hand, since $X = Y$ is necessarily equivalent to $|X - Y| = 0$ and $|Z - 1| = 0$,

Equivalence implies:

3. $P(X > 1/2|X = Y) = P(X > 1/2|\,|X - Y| = 0) = P(X > 1/2|\,|Z - 1| = 0)$.

And these three claims are inconsistent, assuming **Uniqueness** (in particular, assuming $P(X > 1/2|X = Y)$ is well-defined).

Now, one might insist that the difference $|X - Y| \leq \epsilon$ is the correct or more natural way to quantify closeness. First note, though, that we do not always take difference as the intuitive measure of closeness. Consider the population data of two counties C and D in years 1918 and 2018 shown in Table 1. The absolute difference between the populations is the same in both years, but there is an intuitive sense in which the population of C is "closer" to the population of D in 2018 than it was in 1918. Second, note that even if we grant that absolute difference is the correct measure of closeness, whether the response succeeds is sensitive to one's choice of units. Imagine the Logarians are in every aspect like their Earthian counterparts, except that they measure lengths in logameters rather than meters, and speeds in logameters per second rather than meters per second. The Logarians might agree that absolute difference is the natural measure of closeness. However, because of their choice of scale, they will judge two speeds as close in absolute difference whenever we judge them as close in ratio. For the Logarians, the closer the two speeds are stipulated to be, the more likely it is for them to be fast. And so in the limit, when two speeds are so close that they are equal, the current method will yield the verdict that the probability of $X > 1/2$ is greater than 1/2. In fact, by choosing a suitable rescaling in units, one can make the limiting conditional probability any value one wants (Arnold and Roberston, 2002).

	1918	2018
C	10	100
D	20	110

Table 1: Population of C and D by year.

2.2 Independence reasoning

Examining the limits method and reflecting on its failure, it seems it does not make full use of a key piece of information about the set-up, namely that X and Y are both probabilistically and causally independent. Perhaps these independence features can allow us to determine that the answer is $1/2$ without **Equivalence** raising trouble.

As a first attempt, one might propose the following principle:

Naive Independence If two physical quantities X and Y are causally and probabilistically independent of each other, then the chance that X takes a specific value or range of values is unaffected by whether or not Y happens to sync up with X. In particular, for any x, the probability of $X > x$ given $X = Y$ equals the unconditional chance of $X > x$.

Since X is causally and probabilistically independent of Y in the projectile scenario, Naive Independence yields $P(X > 1/2 | X = Y) = P(X > 1/2) = 1/2$ as desired.

As the name suggests, however, Naive Independence is false. It even gives the wrong verdict in simple finite cases. Imagine Jonny tosses a coin with bias a on the North Pole and Donny tosses a coin with bias b on the South Pole, where the two coin tosses are causally and probabilistically independent of one another. Let *heads* denote the event that Jonny's coin lands heads, and let *same side* denote the event that Jonny and Donny's coins land on the same side. The unconditional chance of *heads* is a. What is the probability of *heads* given *same side*? The Independence Principle entails the answer is a. But suppose, for example, $a = b = 0.6$. Then:

$$P(heads | same \ side) = \frac{P(heads \ \& \ same \ side)}{P(same \ side)} = \frac{0.6^2}{0.6^2 + 0.4^2} > 0.6.$$

In general, one can check that $P(heads | same \ side) = P(heads)$ if and only if Donny's coin is unbiased ($b = 0.5$). Intuitively, this is because if Donny's coin is biased towards heads (tails), then *same side* means it is more likely that Jonny's coin lands heads (tails). An upshot here

is that causal and probabilistic independence of physical quantities is not always a direct guide to conditional chances involving those quantities.

Nevertheless, the observation that *heads* is independent of *same side* when Donny's coin is unbiased also suggests a charitable amendment to Naive Independence. Crucially, in the projectile example, Y is uniformly distributed, not 'biased' toward low or high speeds. At least in that special case, the thought goes, the occurrence or truth of $X = Y$ should not affect the probability of $X \geq x$:

Independence Principle If two physical quantities X and Y are causally and probabilistically independent of each other, have the same (marginal) distribution and are uniformly distributed over their support,[5] then the distribution of X is unaffected by whether or not X happens to sync up with Y. In particular, for any x the chance of $X > x$ given $X = Y$ equals the unconditional chance of $X > x$.

This principle also implies the desired verdict of $1/2$ in the projectile case. And it fares much better than Naive Independence. Indeed, one can prove that if chances are probabilities that satisfy the standard Kolmogorov axioms, then the principle is true whenever $P(X = Y) > 0$:

Fact 1. *Suppose X and Y are probabilistically independent with identical, uniform distributions. If $P(X = Y) > 0$, then $P(X > x | X = Y) = P(X > x)$ for any x.*

It would be very odd, one might think, if the Independence Principle were false only if $P(X = Y) = 0$. Intuitively, the truth of this principle should not be sensitive to the chance of $X = Y$.

Remarkably, though, trouble looms even for this revised principle. First notice that analogous reasoning supports the following *de*pendence principle:

Dependence Principle If two physical quantities W and V are causally and probabilistically independent of each other, have the same (marginal) distribution and are *not* uniformly

---

[5]The support of a real-valued random variable X is the set $\{x \in \mathbb{R} : P(X \in B(x, r)) > 0 \text{ for all } r > 0\}$ where $B(x, r)$ denotes the ball with radius r centered at x.

distributed over their support, then there exists a w such that the probability of $W > w$ given $W = V$ does *not* equal the unconditional chance of $W > w$.

Indeed, one can prove that this principle is also true in the case where $P(W = V) > 0$:

Fact 2. *Suppose W and V are probabilistically independent and have the same (marginal) distribution, and are not uniform over their support. If $P(W = V) > 0$, then there is an w such that $P(W > w|W = V) \neq P(W > w)$.*

However, now we can derive a conflict, as follows. Consider the quantities $E_X = \frac{1}{2}mX^2$ and $E_Y = \frac{1}{2}mY^2$ which represent the kinetic energy of the projectiles after impact, where we suppose the two projectiles have the same mass m. Note that E_X and E_Y are not uniformly distributed and, more generally, they satisfy the conditions of the Dependence Principle, with $W = E_X$ and $V = E_Y$. Thus, by the Dependence Principle,

1. There exists an a such that $P(E_X > a|E_X = E_Y) \neq P(E_X > a)$.

On the other hand, by the Independence Principle:

2. For all x, $P(X > x|X = Y) = P(X > x)$.

But since $X > \sqrt{\frac{2a}{m}} \Leftrightarrow E_X > a$, and $E_X = E_Y \Leftrightarrow X = Y$, **Equivalence** implies:

3. $P(X > \sqrt{\frac{2a}{m}}|X = Y) = P(E_X > a|E_X = E_Y)$.

Again these three claims are inconsistent assuming **Uniqueness**, in particular assuming the conditional probabilities are well-defined. To see this, note that $P(X > \sqrt{\frac{2a}{m}}) = P(E_X > a)$, so if $P(E_X > a|E_X = E_Y) \neq P(E_X > a)$ then $P(X > x|X = Y) \neq P(X > x)$ for some x.

2.3 Other methods

There are other methods to derive $P(X > 1/2|X = Y) = 1/2$ that one might attempt. For instance, consider the *discretization* method. Unlike the limits method, which considers $|X - Y| \leq \epsilon$ as ϵ shrinks, the discretization method begins by modeling the projectile

speeds as discrete quantities $\tilde{X}_n$ and $\tilde{Y}_n$ with n distinct possible values, and then taking $\lim_{n\to\infty} P(\tilde{X}_n > 1/2 | \tilde{X}_n = \tilde{Y}_n)$.

However, as one might predict, this method yields different results depending on how the quantities are discretized. For example, depending on whether we discretize state space into increments of equal speed or increments of equal kinetic energy, we get different results even in the limit.

We need not belabor the point. Reflecting on all these cases, **Uniqueness** and **Equivalence** seem in conflict. Crucially, this is not to say there is *no* good argument for the answer of 1/2. Perhaps there are good reasons to discretize X in a certain way, or to take the difference approximation $|X - Y| \leq \epsilon$ rather than the ratio. Our point is simply that such arguments would plausibly have to appeal to physical considerations beyond those supplied in the specification of the chance set-up. It seems doubtful that the physical probability of *fast* given *same speed* supervenes on, or can be determined from, only the facts about X, Y and the situation that were explicitly provided. And this is what **Uniqueness** requires.

3 Is the formalism to blame?

It is tempting to chalk up the puzzles and difficulties of the last section to a limitation of our mathematical representation, and in particular a limitation of standard Kolmogorov probability theory, according to which conditional probabilities are defined derivatively rather than axiomatized directly. As Gyenis et al. (2017, p. 2595) write, "According to [a certain popular view], the Borel–Kolmogorov Paradox poses a serious threat for the standard measure theoretic formalism of probability theory, in which conditional probability is a defined concept, and this is regarded as justification for attempts at axiomatizations of probability theory in which the conditional probability is taken as the primitive rather than a defined notion." Proponents of this view include Kadane et al. (1999), who see the paradox as a defect of Kolmogorov's theory (224–5, notation adapted): "The seeming contradiction is often

resolved by claiming that the transformation of variables only yields conditional probability given the sigma field of events determined by the random variable [...] not given individual events in the sigma field. This approach is unacceptable from the point of view of the statistician who, when given the information that $X = Y$ has occurred, must determine the conditional distribution of X."

We think this view is mistaken for two reasons. First, the tension between **Uniqueness** and **Equivalence** is formalism-neutral. It exists insofar as one grants that $P(A|B)$ equals the limit of $P(A|B_n)$, or that both the Independence Principle and the Dependence Principle are true. And these principles seem attractive even if one takes conditional probability to be primitive rather than derivative.

Second, none of the existing alternative probability formalisms really solve the puzzle, conceptualized as a tension between **Uniqueness** and **Equivalence**. For instance, in the projectile example, if we model $P(\cdot|\cdot)$ as a primitive two-place function (as in Popper (1959) and Rényi (1955)), then there will be many $P(\cdot|\cdot)$'s that are compatible with the set-up, but differ on the value assigned to $P(X > 1/2|X = Y)$. In fact, one can prove that for any answer obtained from the limits method discussed in Section 2.1, there exists a Popper function $P(\cdot|\cdot)$ which yields that answer and is otherwise consistent with the set-up.[6] Similarly, if we allow P to range over infinitesimals in addition to positive reals between 0 and 1, then the value of $P(X > 1/2|X = Y)$ typically depends on the choice of ultrafilters that specify how to sum up infinitesimal probabilities.[7] In general, these non-standard formalisms do not solve whether $P(X > 1/2|X = Y) = 1/2$; they agree that this question can only be answered with respect to an appropriate Φ, where Φ is a primitive conditional probability function, an ultrafilter, a limiting procedure, a conditioning sigma-field, etc., but they leave open how a particular choice of Φ can be justified.[8]

---

[6]This follows from a result due to Dubins (1975).

[7]See for example Benci et al. (2012).

[8]That being said, one's choice of formalism is not entirely irrelevant to the paradox. Different formalisms may fit more or less naturally with certain approaches to resolving the paradox. Kolmogorov's formalism, for example, lends itself very naturally to approaches that reject **Equivalence**, whereas if one models probabilities using primitive conditional probability functions, **Equivalence** follows as a theorem of their

4 The urgency of the paradox

We have been arguing that the paradox is formalism-neutral, and in this sense, even if you don't ascribe to the Kolmogorovian formalism, "the Borel-Kolmogorov paradox is your paradox too" (cf. Hájek (2007)). But it is worth mentioning that even if you doubt this conclusion, the paradox should still concern you for a further reason.

Scientists use the standard formalism. They use the formalism to represent continuous chancy quantities and, crucially, to infer claims about conditional probabilities involving those quantities. For example, Sauve et al. (1999) consider a gamma ray shot toward an obstacle and allowed to scatter. They ask for the probability that the ray's out-of-plane scatter angle ϕ is greater than a, given that the ray's Compton scatter angle θ equals b, and reason from the premise that ϕ and θ are independent to the answer $P(\phi > a|\theta = b) = P(\phi > a)$, which they then use in further calculations. However, our discussion in Section 2 casts into doubt the reliability of this independence-based reasoning. Thus a question arises whether to try to recover the scientists' judgment, or to somehow explain away the role of these conditional probabilities in the model, not only in this case but also many other cases where theorists and practitioners apply the methods from Section 2. We think discarding the judgments is a last resort, so we should explore strategies for resolving the paradox that might vindicate them.

5 Toward a resolution

Since **Uniqueness** and **Equivalence** conflict, the first step in a proposed resolution to the paradox is to reject at least one of these premises.

We think that rejecting **Equivalence** is a last resort. In the setting of rational credence, $Cr(A|B)$ may plausibly differ from $Cr(A|B')$ even when B and B' are metaphysically necessarily equivalent. This may happen, for instance, if rational credence is sensitive to

---

axioms.

intensional or hyperintensional differences; for example, our credences conditional on Hesperus is Phosphorus may differ from our credences conditional on Phosphorus is Phosphorus.[9]
However, we find this kind of failure of **Equivalence** much less plausible in the case of physical probability. Physical probability, it seems, should not be sensitive to intensional and hyperintensional differences, at least not in ordinary scenarios like the projectile example.

We propose instead rejecting **Uniqueness**, and in particular the determinability clause. In the projectile scenario, while there is a single true value of the physical probability $P(X > 1/2|X = Y)$ which does not change depending on how $X = Y$ is viewed or represented, this value is not determinable from the information provided. Either $P(X > 1/2|X = Y)$ taking the value it does is a primitive matter or, more plausibly, it takes the value it does in virtue of physical features of the chance set-up that were not explicitly specified.[10]

What kind of physical features could those be? On our treatment, this is the central question. While we do not have a general answer to this question, we think an example from statistical mechanics (Kac and Slepian, 1959) offers us an important clue. We conclude the paper by examining this case.

Consider a particle whose trajectory is a Gaussian process $\{x(t)\}_{t\in\mathbb{R}}$. In particular, fixing a time t, the displacement $x(t)$ follows a Gaussian distribution with a mean of 0 meters and a standard deviation of, say, 1 meter from the origin. Let $v(t)$ denote the velocity of the particle at t, which we suppose also follows a Gaussian distribution and is independent of $x(t)$. Given this set-up, what is $P(v(0) > 1/2|x(0) = a)$? In other words, what is the

---

[9]Rescorla (2015) analyzes the Borel-Kolmogorov paradox in the subjective setting, and pursues this kind of idea.

[10]A related treatment of the paradox is the 'relativization' approach (Kolmogorov, 1933; Jaynes, 2003; Easwaran, 2008; Gyenis et al., 2017). The idea is that conditional probability is a three-place relation; the probability of A given B is always relativized to a sigma-field that represents some "hypothetical experiment" from which B is drawn. Our proposal goes beyond this approach in two respects. First, while we think conditional (physical) probabilities are sensitive to contextual information about the chance set-up, we are agnostic here about (i) whether this information is exhausted by information about the experiment, if there is one, and (ii) whether it can always be mathematically represented as a sigma-field of events. Second, as Rescorla (2015) points out, the relativization approach does not "explain why conditional probabilities are relativized" but merely "stipulates that they are". Our proposal gives a partial explanatory story: as one might expect, physical probabilities are sensitive to physical features of the chance set-up; what the paradox shows is that these features may not always be fully specified in textbook toy examples.
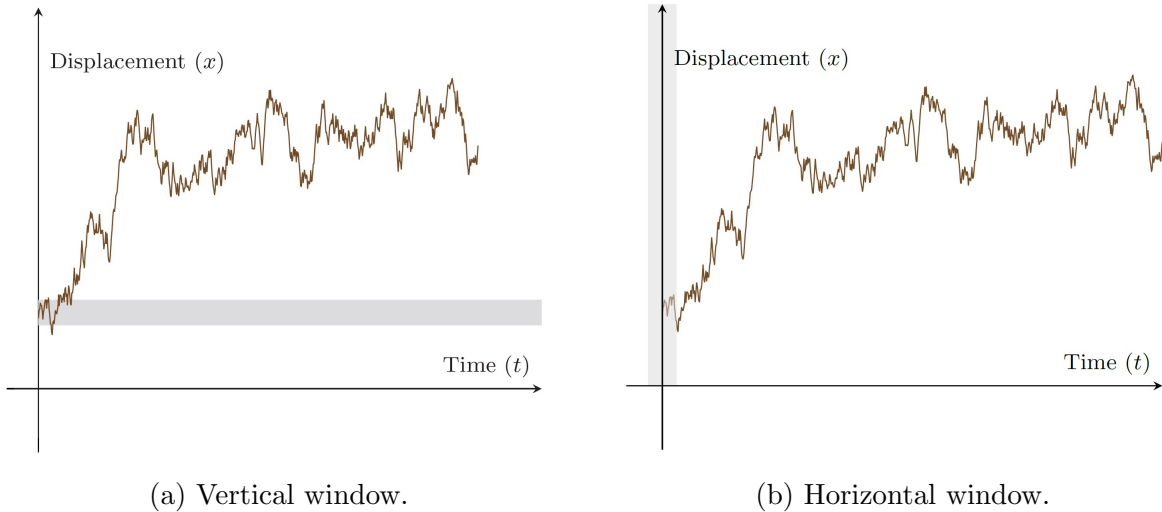
(a) Vertical window.   (b) Horizontal window.

Figure 2: The vertical and horizontal windows.

probability that the particle is fast at $t = 0$, given it is located at point a?

Independence considerations yield the answer:

$$P(v(0) > 1/2|x(0) = a) = P(v(0) > 1/2) \approx 0.31.$$

This is equivalent to the answer we obtain by fixing the time $t = 0$, and the considering the limit as $x(0)$ gets closer to the point a:

$$\lim_{\epsilon \to 0} P(v(0) > 1/2 | |x(0) - a| \leq \epsilon) = P(v(0) > 1/2) \approx 0.31.$$

This procedure, depicted in Figure 2a, is called taking the 'vertical window'.

However, suppose instead we consider the 'horizontal window' where we hold fixed the point a, and consider a shrinking interval of *time* toward $t = 0$ (Figure 2b):

$$\lim_{\epsilon \to 0} P(v(0) > 1/2|x(t) = a \text{ for some } t \in [-\epsilon, 0]) \approx 0.47.$$

This second procedure yields a different answer, similarly to the situation in Section 2.

So far, these findings parallel our earlier findings in the projectile case. However, in this

14

case there is a more obvious physical consideration that breaks the stalemate between these two answers—and surprisingly, not in favor of the independence-based verdict. The consideration is that the process governing the particle is *ergodic*. Roughly speaking, ergodicity means the probabilistic properties of the process $x(t)$ can be deduced from its frequency properties in a relevant time ensemble. It turns out that the horizontal window procedure is the only one that respects this equality between the probability and time ensemble limiting frequency. As Cramér and Leadbetter (2004, p. 221) write (notation adapted):

> Further, one of the main properties which we shall require in many cases is that the conditional probability function should admit an appropriate interpretation in terms of a limit of relative frequencies. In the simple example given above, suppose that t_i denotes the time points in $(0, T)$ at which crossings of the level a [by the variable x] occur. Then, one may wish to show that the proportion of those t_i for which $v(t_i) > 1/2$ converges (with probability one) to $P(v(0) > 1/2 | x(0) = a)$ as $T \to \infty$, under an appropriate definition of the conditional probability. The definition to be given will be such that "ergodic properties" of this nature may be demonstrated under appropriate conditions. Specifically, suppose we replace the conditioning event $a - \epsilon < x(0) \leq a$ in $P(v(0) > 1/2 | a - \epsilon < x(0) \leq a)$ by the event "$x(t) = a$ for some t in the interval $[-\epsilon, 0]$." A conditional probability $P_2(v(0) > 1/2 | x(0) = a)$ may then be defined by [the horizontal window definition above].

Granted, these specific considerations will not generalize to all chancy processes where the paradox arises. But it is a useful illustration of how, rather than rejecting **Equivalence**, we might appeal to further physical features of the chance set-up to determine the answer. The example also serves to further illustrate that the paradox discussed in this paper, concerning the conflict between **Equivalence** and **Uniqueness**, is not just a contrived mathematical puzzle. It arises in routine scientific modeling, and should be taken seriously.

A   Proofs

To prove Fact 1, we use the following lemma:

Lemma 1. *Let X and Y be two random variables that are independent and identically distributed. Let P be a countably additive probability function. If $P(X = Y) > 0$, then both X and Y are discrete random variables that only take on finitely many values.*

Proof of Lemma 1. Since Y is uniformly distributed, it cannot be neither discrete nor continuous, for then there are y, y' such that $P(Y = y) > 0$ but $P(Y = y') = 0$. We claim that Y must be discrete. Suppose toward contradiction that Y is continuous, i.e. for all y in the range of Y, $P(Y = y) = 0$. There are two cases: either X is continuous, or it is not. Suppose X is continuous. Let $Z = X - Y$. Then Z is a continuous variable. So $P(Z = 0) = P(X = Y) = 0$, which contradicts the assumption that $P(X = Y) > 0$. Suppose X is not continuous. Let $A = \{w : P(X = x) > 0\}$, i.e. A is the discrete component of X. Let $Z_1 = Z \cdot 1_{X \in A}$ (the conditional variable that equals Z on the set $X \in A$, and 0 otherwise). Let $Z_2 = Z \cdot 1_{X \notin A}$. Then $P(Z = 0) = P(Z_1 = 0) + P(Z_2 = 0)$. Since both X and Y are continuous on the set $X \notin A$, Z_2 is a continuous variable. So $P(Z_2 = 0) = 0$. On the other hand, $P(Z_1 = 0) = \sum_{x \in A} P(X = x, Y = x) = \sum_{x \in A} P(X = x)P(Y = x) = 0$ since $P(Y = x) = 0$ for all $x \in A$. So $P(Z_1 = 0) = 0$. But then again we have a contradiction with $P(Z = 0) = P(X = Y) > 0$. Therefore Y cannot be continuous. Since a distribution that is neither discrete nor continuous cannot be uniform, it follows that Y (and therefore X) must be discrete. Moreover, since Y is uniform and P is countably additive, both X and Y can only take on finitely many values. $\square$

Proof of Fact 1. Let X, Y be two independent random variables with identical, uniform marginal distributions. Suppose $P(X = Y) > 0$. By Lemma 1, both X and Y are discrete. Thus, to show that $P(X > x | X = Y) = P(X > x)$, it suffices to show that $P(X = x | X = Y) = P(X = x)$ for all x. Clearly if $P(X = x) = 0$, then $P(X = x | X = Y) = 0 = P(X = x)$.

Suppose $P(X = x) = P(Y = x) = \alpha > 0$. Then

$$P(X = x | X = Y) = \frac{P(X = x, X = Y)}{P(X = Y)} = \frac{P(X = x, Y = x)}{\sum_{x'} P(X = x', Y = x')}$$
$$= \frac{\alpha P(X = x)}{\alpha \sum_{x' : P(X = x') > 0} P(X = x')} = P(X = x),$$

as claimed. $\qquad\square$

Proof of Fact 2. For each interval $(a, b]$, we have

$$P(a < W \le b | W = V) = \frac{P(a < W \le b, W = V)}{P(W = V)} \le \frac{P(a < W \le b, a < V \le b)}{P(W = V)}$$
$$= \frac{P(a < W \le b) \cdot P(a < W \le b)}{P(W = V)} = P(a < W \le b) \cdot \frac{P(a < W \le b)}{P(W = V)}.$$

We consider two cases: either there is $(a, b]$ such that $0 < P(a < W \le b) < P(W = V)$, or there is not. In the first case, for that $(a, b]$, if $P(W > b | W = V) \ne P(W > b)$, then we are done. If $(W > b | W = V) = P(W > b)$, then $P(a < W \le b | X = Y) < P(a < W \le b)$. So

$$P(W > a | W = V) = P(a < W \le b | W = V) + P(W > b | W = V)$$
$$< P(a < W \le b) + P(W > b)$$
$$= P(W > a).$$

Now suppose that for all intervals $(a, b]$, $P(a < W \le b)$ equals 0 or is greater than $P(W = V)$. This implies that W is discretely distributed, with its atoms at least of weight $P(W = V)$.[11] Since W and V are identically distributed, the same holds for V. Now since V is not uniform over its support, there are v, v' such that $P(V = v) > P(V = v')$ where

---

[11]*Proof.* It suffices to show that X has a countable support. Suppose towards a contradiction that the support of W is uncountable. Then there exists w in the support of W such that $P(W = w) = 0$ but for all r, $P(W \in B(x, r)) > 0$. It follows that there is r such that $0 < P(W \in B(x, r)) < P(W = V)$. Let $(a, b] = [w - r, w + r]$. Then. $0 < P(a < W \le b) < P(W = V)$, contradicting our assumption that for all $(a, b]$, $P(a < W \le b) = 0$ or $P(a < W \le b) > P(W = V)$.

both these events have positive probability. Thus

$$\frac{P(W = v'|X = Y)}{P(W = v|X = Y)} = \frac{P(W = v', V = y')}{P(W = v, V = v)} = \frac{P(W = v')P(V = y')}{P(W = v)P(V = v)}$$
$$= \frac{P(W = v')}{P(W = v)} \cdot \frac{P(V = v')}{P(V = v)} < \frac{P(V = v')}{P(V = v)},$$

and so either $P(W = v|V = V) \neq P(W = v)$ or $P(W = v'|W = V) \neq P(W = v')$ or both. Since a change in W's posterior distribution implies a change in its cdf, there must exist an x which satisfies the desired condition. □

References

Arnold, B. and C. Roberston (2002). The conditional distribution of X given X=Y can be almost anything! In N. Balakrishnan (Ed.), *Advances on Theoretical and Methodological Aspects of Probability and Statistics*, pp. 75–81. New York: Taylor and Francis.

Benci, V., L. Horsten, and S. Wenmackers (2012). Axioms for non-archimedean probability (NAP). In D. V. J. and D. L. (Eds.), *Future Directions for Logic. Proceedings of PhDs in Logic III*, Volume 2. College Publications.

Cramér, H. and M. R. Leadbetter (2004). *Stationary and related stochastic processes: Sample function properties and their applications*. New York: Dover Publications.

Dubins, L. E. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability 3*(1), 89–99.

Easwaran, K. K. (2008). *The foundations of conditional probability*. University of California, Berkeley.

Gyenis, Z., G. Hofer-Szabó, and M. Rédei (2017). Conditioning using conditional expectations: The Borel–Kolmogorov paradox. *Synthese 194*(7), 2595–2630.

Hájek, A. (2003). What conditional probability could not be. *Synthese 137*(3), 273–323.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese 156*(3), 563–585.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge University Press.

Kac, M. and D. Slepian (1959). Large excursions of Gaussian processes. *The Annals of Mathematical Statistics 30*(4), 1215–1228.

Kadane, J. B., M. J. Schervish, and T. Seidenfeld (1999). Statistical implications of finitely additive probability. In J. B. Kadane, M. J. Schervish, and T. Seidenfeld (Eds.), *Rethinking the Foundations of Statistics*, pp. 211–232. Cambridge University Press.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin.

Lindley, D. (1982). The Bayesian approach to statistics. In J. de Oliveria and B. Epstein (Eds.), *Some advances in statistics.* New York: Academic Press.

Popper, K. R. (1959). *The Logic of Scientific Discovery.* Routledge.

Rao, M. (1988). Paradoxes in conditional probability. *Journal of Multivariate Analysis 27*(2), 434–446.

Rényi, A. (1955). On a new axiomatic theory of probability. *Acta Mathematica Hungarica 6*(3-4), 285–335.

Rescorla, M. (2015). Some epistemological ramifications of the Borel–Kolmogorov paradox. *Synthese 192*(3), 735–767.

Sauve, A. C., A. Hero, W. L. Rogers, S. Wilderman, and N. Clinthorne (1999). 3d image reconstruction for a compton spect camera model. *IEEE Transactions on Nuclear Science 46*(6), 2075–2084.