

Regularity and infinitely tossed coins

Colin Howson¹

Received: 13 January 2016 / Accepted: 22 March 2016 / Published online: 13 May 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Timothy Williamson has claimed to prove that regularity must fail even in a nonstandard setting, with a counterexample based on tossing a fair coin infinitely many times. I argue that Williamson's argument is mistaken, and that a corrected version shows that it is not regularity which fails in the non-standard setting but a fundamental property of shifts in Bernoulli processes.

Keywords Probability · Hyperreal · Williamson

1 Introduction

As is well-known, the expansion of the real line by the addition of infinitesimals¹ is no longer a Leibnizean pipe-dream: halfway through the last century Abraham Robinson showed that there exists an embedding of the first-order structure of the ordered field of real numbers into an extension containing infinitesimal and infinitely large numbers, satisfying all the first-order properties of real numbers, and of which the infinitely small and infinitely large numbers form completely ordered subrings. The members of such a nonstandard extension, of which there are infinitely many, are known as *hyperreals*.²

One of the more persuasive arguments for probabilistic *regularity*, i.e. the principle that only the impossible event should receive zero unconditional probability, involves an appeal to probability functions whose domain is a standard algebra of events but whose range is the nonstandard unit interval in a hyperreal extension. Probably the most powerful objection to regularity is the mathematical fact that in uncountable outcome

¹Infinitesimals are numbers with absolute value smaller than every positive real number.

²The hyperreals can also be constructed as members of an ultrapower of the real numbers. Łoś's Theorem shows that the embedding is elementary, justifying the so-called Transfer Principle.

✉ Colin Howson
howson.colin@gmail.com

¹ London School of Economics and Political Science, London WC2A 2AE, UK

spaces, for example that of a fair coin tossed infinitely often (it has the cardinality of the real numbers) in which all singletons are assigned a probability, all but a countable number must have zero probability. Once the range of a probability function consists of hyperreals, however, the objection no longer has force because all those outcomes can be assigned a positive infinitesimal probability, even the same infinitesimal probability.³

Thus may - and often does - the up-to-date defender of regularity argue. But the defence has been challenged by Timothy Williamson (2007), who used the model of a fair coin tossed infinitely many times to argue that at least one outcome sequence must have probability 0, and hence that regularity must fail. What is most striking about Williamson's argument is his claim that it is valid even in the context of hyperreal probabilities themselves. In what follows I will argue that it is not valid.⁴ That is the programme. I will start by reviewing Williamson's argument.

2 The argument

A coin is flipped infinitely many times. Suppose $H(1 \dots)$ represents the event 'all the outcomes are heads', and $H(2 \dots)$ 'all the outcomes after the first are heads' (2007, p.4). Then according to Williamson,

$H(1\dots)$ and $H(2\dots)$ are isomorphic events. More precisely, we can map the constituent single-toss events of $H(1\dots)$ one-one onto the constituent single-toss events of $H(2\dots)$ in a natural way that preserves the physical structure of the set-up just by mapping each toss to its successor. $H(1\dots)$ and $H(2\dots)$ are events of exactly the same qualitative type; they differ only in the inconsequential respect that $H(2\dots)$ starts one second after $H(1\dots)$. That $H(2\dots)$ is preceded by another toss is irrelevant, given the independence of the tosses. Thus $H(1\dots)$ and $H(2\dots)$ should have the same probability. (2007, p.5)

But if we assume that the singleton sequences of possible outcomes are assigned the same probability, we quickly infer that that probability must be 0. For

$$P(H(1\dots)) = P(H(2\dots)|H(1))P(H(1)) \quad (1)$$

where $H(1)$ is the event 'the first toss lands heads'. So by independence

$$P(H(1\dots)) = P(H(2\dots))P(H(1))$$

³ Bernstein and Wattenberg 1969. How useful such assignments are in practice when not used simply to elicit results about standard mathematics is of course another question.

⁴ Williamson is not the only one to argue that appealing to hyperreals fails to save regularity. Pruss (2013) has a cardinality-based argument (but see Hofweber 2014), and another (2012) claiming that since infinitesimally-valued probabilities are non-conglomerable in some countable partition, one can know beforehand with near-certainty that one's updated hyperreal-valued probability of a suitable event A will differ from the current one.

i.e.

$$P(H(1\dots)) = P(H(2\dots))/2.$$

But because of the presumed isomorphism of $H(1 \dots)$ and $H(2 \dots)$, we have

$$P(H(1\dots)) = P(H(2\dots)),$$

whence

$$2P(H(1\dots)) = P(H(1\dots)),$$

and so, noting that infinitesimals obey the same field rules as the real numbers themselves, we must have $P(H(1 \dots))=0$, contradicting regularity.

Williamson reinforces his claim that $H(1 \dots)$ and $H(2 \dots)$ are isomorphic (in a way that demands their being assigned the same probability) by extending the example to include a second coin, in all relevant physical respects identical to the first:

To make the point vivid, suppose that another fair coin, qualitatively identical with the first, will also be tossed infinitely many times at one second intervals, starting at the same time as the second toss of the first coin, all tosses being independent. Let $H^*(1\dots)$ be the event that every toss of the second coin comes up heads, and $H^*(2\dots)$ the event that every toss after the first of the second coin comes up heads. Then $H(1\dots)$ and $H^*(1\dots)$ should be equiprobable, because the probability that a coin comes up heads on every toss does not depend on when one starts tossing, and there is no qualitative difference between the coins. But for the same reason $H^*(1\dots)$ and $H(2\dots)$ should also be equiprobable. These two infinite sequences of tosses proceed in parallel, synchronically, and there is no qualitative difference between the coins; in particular, that the first coin will be tossed once before the $H(2\dots)$ sequence begins is irrelevant. By transitivity, $H(1\dots)$ and $H(2\dots)$ should be equiprobable: [hence] $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots))$ (2007, p.6)

The argument, notes Williamson, ‘is neutral between standard and non-standard probabilities. Even when infinitesimal probabilities are allowed, the nature of the case still yields the conclusion that the probability of an infinite sequence of heads is 0.’

Williamson’s argument sounds very plausible. Nevertheless I will argue that it is mistaken, and that the mistake is not one about hyperreals but a confusion about what he calls ‘isomorphic events’, assisted by an inadequate notation. That his argument is indeed wrong is anyway strongly suggested by the fact that in the reference he himself supplies to the classic paper by Bernstein and Wattenberg (cited above), it is shown that there is a uniform positive infinitesimal probability distribution over the members of the standard closed unit interval $[0,1]$.⁵ It is well known that the real numbers in this

⁵ The authors show that there is a hyperreal-valued probability measure which assigns the same positive infinitesimal probability to each point in the standard unit interval $[0,1]$ and which is infinitely close to Lebesgue measure on the Lebesgue-measurable subsets.

interval are representable (mod equivalent sequences) by the infinite binary sequences in $\{0,1\}^{\mathbb{N}}$, $\mathbb{N} = \{1, 2, \dots, n, \dots\}$,⁶ a set which can clearly also represent all possible infinite sequences of heads and tails. It would be very strange indeed if as soon as one interprets the set of sequences in this last way the positive infinitesimal distribution over its members suddenly becomes impossible. So what is Williamson's mistake (if there is a mistake)?

3 The mistake

There is indeed a mistake, and it is a fairly elementary error in Williamson's purely probabilistic reasoning. To identify it clearly however a little more notation is required. In the usual parlance of probabilists, a *probability space* is a triple (S, F, P) , where S is the outcome-set of some experiment (understood loosely), F is an algebra of subsets of S containing S and the empty set, and P is a probability function on F . I shall follow Williamson in assuming that P is hyperreal-valued. In his example of the single coin, S is the set $\{0,1\}^{\mathbb{N}}$ of all possible infinite sequences of outcomes of tossing the coin (1 for heads, 0 for tails). Here, F is generated by the cylinder sets, i.e. sets whose member-sequences are identified by a finite number of indices, and is closed under set difference, countable intersections and unions. Therefore F will also contain all the singleton sequences, and Bernstein and Wattenberg's result implies that we can assign infinitesimal probabilities to each of these.

The stage is now set to recapitulate Williamson's argument. His $H(1 \dots)$ is the singleton of $(1, 1, 1, \dots, 1, \dots)$. But in this space, a fact essential to step (1) of Williamson's derivation, $H(2 \dots)$ unlike $H(1 \dots)$ is a *compound event*, the pair-set whose members are the two outcome-sequences $(1, 1, 1, \dots, 1, \dots)$ and $(0, 1, 1, \dots, 1, \dots)$ representing the disjunction 'a head occurred first followed by all heads *or* a tail occurred first followed by all heads'. Clearly, however, it is strictly nonsensical to say that the singleton $\{(1, 1, 1, \dots, 1, \dots)\}$ is isomorphic to a pair (a necessary condition for isomorphism is cardinal equivalence) and so the step to $P(H(1 \dots)) = P(H(2 \dots))$, supposedly justified by the appeal to isomorphism, fails and with it Williamson's argument.⁷

Similar considerations apply to his addition of the second coin to his example. We can represent the two coin-tossing experiments in two probability spaces, all the elements of the second of which, following Williamson's example of $H^*(1 \dots)$, I shall

⁶ 0 is sometimes included in \mathbb{N} .

⁷ Ruth Weintraub has also challenged Williamson's argument, by arguing that the isomorphism to which he thinks he is appealing does not preserve all the physical features of the situation:

Williamson's example shows that isomorphism doesn't preserve *all* basic physical properties. He claims that the two 'sequences of events are of *exactly* the same qualitative type' ... But although all the physical properties of the *constituent events* are preserved by the mapping, as are the temporal intervals between adjacent tosses, there is a *global* property (of the complex event) which is *not* preserved. The second sequence is a *proper subset* of the first. (2008, p.249)

But this is wrong: there are not two *sequences* of events (the quote is inaccurate on this point) but two events, and far from the second, $H(2 \dots)$, being a proper subset of the first, $H(1 \dots)$, the reverse is actually the case as we have seen.

label with an asterisk (and the outcomes of the tosses are also recorded with an asterisk appended).⁸ Other than the labelling-difference, the two spaces are identical, and so we can agree with Williamson that $P(H(1\dots)) = P^*(H^*(1\dots))$ ‘because the probability that a coin comes up heads on every toss does not depend on when one starts tossing, and there is no qualitative difference between the coins.’ But when he continues ‘But for the same reason $H^*(1\dots)$ and $H(2\dots)$ should also be equiprobable’ he is *not* correct: $H(2\dots)$ is not the original sequence truncated at the second toss but a set containing two sequences, one the original sequence and the other the same sequence whose first member is changed to 0. Hence there is no justification for equating $P(H(1\dots))$ and $P(H(2\dots))$ and Williamson’s argument fails.

It might be objected that Williamson has merely made an inappropriate choice of coin-tossing model, since there is one (it is claimed) in which his argument clearly succeeds: that of the same coin being tossed from minus infinity to plus infinity. Formally, this has the outcome space $W = \{0,1\}^{\mathbb{Z}}$ where \mathbb{Z} is the set of integers,⁹ and where 1 stands for a head and 0 for a tail. Let $\mathbf{x} = \langle x_i \rangle$ be a sequence in W and suppose T is the so-called shift transformation of W , i.e. the mapping of W into itself such that $(T\mathbf{x})_i = x_{i+1}$. Williamson’s argument can now go through straightforwardly (claims the objection) in the following way. Consider the measurable event (set) S of sequences \mathbf{x} where $x_i = 1$ for $i \geq 1$. The shift transformation takes S to the set S' of T -images of members of S , where each sequence in S' is the isomorphic image of one in S under T , and conversely since T is invertible. Hence (the objection proceeds) S' and S are isomorphic and so, if we follow the Williamson doctrine about isomorphism, should receive the same positive infinitesimal probability p . Now we just replicate Williamson’s argument making the appropriate changes.

Unfortunately the conclusion that regularity must fail here is as unwarranted as it was in Williamson’s original argument, with the mistake again lying in the appeal to isomorphism. For S and S' are not themselves sequences but unordered *sets of sequences*, and are isomorphic only in the trivial sense that any two cardinally-equal unstructured sets are isomorphic (relations can of course be defined on them – e.g. the full and empty n -ary relations for any n – relative to which the sets are isomorphic, but that is beside the point). To use that fact to justify setting their probabilities equal would however clearly be absurd. It cannot be objected that S and S' are physically indistinguishable, since they are different events, nor that they are indistinguishable with regard to their probabilistic properties, for that would simply beg the question. Rather than Williamson’s argument showing that regularity fails, which it does not, what it implicitly does reveal is that a fundamental theorem of probabilistic dynamics, that shift transformations in Bernoulli processes are measure-preserving, can fail in this hyperreal context. This will probably not worry those working in nonstandard probability theory overmuch, since the object there is not so much, or at all, to regard hyperreal probabilities as on the same footing as real-valued ones¹⁰ but to use the nonstandard universe simply as an aid to the standard theory by translating standard problems into nonstandard ones by means of the Transfer Principle, where they are often more tractable – which is exactly how Leibniz regarded his ‘ideal elements’, the infinitesimals and their reciprocals.

⁸ It is of course possible to represent the two experiments within one probability space, differentiating them by means of appropriate random variables, but the device of two spaces is simpler.

⁹ Together with a binomial probability parameter, this set-up is called a two-sided Bernoulli process.

¹⁰ The real numbers are unique up to isomorphism in the standard model of set theory, but there are infinitely many non-isomorphic fields of hyperreals.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bernstein, A. R., & Wattenberg, F. (1969). Nonstandard measure theory. In W. A. J. Luxemburg (Ed.), *Applications of model theory to algebra, analysis and probability* (pp. 171–186). New York: Holt, Rinehart and Winston.
- Hofweber, T. (2014). Cardinality arguments against regular probability measures (forthcoming in *Thought*).
- Pruss, A. R. (2012). Infinite lotteries, perfectly thin darts and infinitesimals. *Thought*, 1, 81–89.
- Pruss, A. R. (2013). Probability, regularity, and cardinality. *Philosophy of Science*, 80, 231–240.
- Weintraub, R. (2008). How probable is an infinite sequence of heads? *Analysis*, 68, 247–250.
- Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, 67, 173–180.