How Probable is an Infinite Sequence of Heads?*

(*Analysis* **67**, 3 (2007): 173-180)

Isn't probability 1 certainty? If the probability is objective, so is the certainty: whatever has chance 1 of occurring is certain to occur. Equivalently, whatever has chance 0 of occurring is certain not to occur (it has no chance of occurring). If the probability is subjective, so is the certainty: if you give credence 1 to an event, you are certain that it will occur. Equivalently, if you give credence 0 to an event, you are certain that it will not occur (it has no weight in your calculations of expected outcomes). And so on for other kinds of probability, such as evidential probability.

The formal analogue of this picture is the regularity constraint: a probability distribution over sets of possibilities is *regular* just in case it assigns probability 0 only to the null set, and therefore probability 1 only to the set of all possibilities. For convenience, restrict the term 'possibility' to those maximally specific in relevant respects. Thus possibilities are mutually exclusive and jointly exhaustive. The probability of a possibility is just the probability of its singleton. Assume that each possibility has a well-defined probability. Then regularity is equivalent to the constraint that every possibility has a probability greater than 0.

Regularity runs into notorious trouble when the set of possibilities is infinite, given the standard mathematics of probabilities, on which they are real numbers between 0 and 1. Indeed, when the set of possibilities is uncountable, no probability distribution is regular. For let A_n be the set of all possibilities of probability at least $1/n$, (n is a natural number). A_n has at most n members, otherwise its probability exceeds 1. If a possibility has probability greater than 0, it belongs to A_n for some natural number n. But the union of the A_n over all natural numbers n is a countable union of finite sets and therefore itself countable. Thus only countably many possibilities have probability greater than 0, so uncountably many possibilities have probability 0. For example, suppose that a rotating pointer can stop at any point on a circle. As space is usually conceived, the circle comprises uncountably many points. For each point, it is neither objectively nor subjectively certain that the pointer will not stop at it. Yet on any real-valued probability distribution, for almost every point on the circle, the real-valued probability that the pointer will stop at it is 0.

The trouble is not confined to uncountable cases. Consider a countable infinity of points on the circle. Call them the *select* points. If we treat all the points as the real numbers between 0 and 1 (including 0 but excluding 1), we can treat the select points as the rational numbers between 0 and 1. Since nothing favours one point over another, the probability of the pointer's stopping at a given point conditional on its stopping at a select point should be the same for each select point; let it be x. By the Archimedean principle, for any real number x greater than 0, nx exceeds 1 for some natural number n. Thus if x were greater than 0, for any n select points the probability of the pointer stopping at one of them conditional on its stopping at a select point would exceed 1. Therefore the probability of the pointer's stopping at a given select point conditional on its stopping at a select point is 0, even though it is neither objectively nor subjectively certain that it will not stop at that point, conditional on its stopping at a select point.

An attractive response is to deny the assumption that probabilities must be real-valued. In non-standard analysis, the numbers between 0 and 1 include infinitesimals, greater than 0 but less than $1/n$ for each (standard) natural number n. That avoids the troubles above. Each point can have the same infinitesimal probability of being stopped at (Bernstein and Wattenberg 1969). Thus infinitesimal probabilities appear to rehabilitate the equation of probability 1 with certainty and the equivalent conception of probability 0. For example, David Lewis appeals to them in requiring that the credence of a proposition is zero only if it is 'the empty proposition, true at no worlds'; he justifies the requirement 'as a condition of reasonableness' on the grounds that one who started out by violating it and then learned from experience by conditionalizing 'would stubbornly refuse to believe some propositions no matter what the evidence in their favour' (1986: 88). In the case of objective probability, he was also 'inclined to think that […] there are no worlds where anything with zero chance happens; the contrary opinion comes of mistaking infinitesimals for zero' (1986: 333).

Do infinitesimal probabilities really rehabilitate the equation of probability 1 with certainty or the equivalent conception of probability 0? Consider one of the simplest cases. A fair coin will be tossed infinitely many times at one second intervals. The tosses are independent. Let H(1…) be the event that every toss comes up heads. It is one of uncountably many possible outcomes. It is not subjectively certain that H(1…) will not occur. Given a suitable indeterminism, it is not objectively certain that H(1…) will not occur. Henceforth, we need not specify what kind of probability is in play, because the argument is the same for all kinds. In standard probability theory, the only probability H(1…) can have is 0, since for each natural number n its probability is no greater than the probability of an initial sequence of *n* tosses, $1/2^n$. That argument fails in the non-standard setting, since it does not exclude the assignment of an infinitesimal probability to H(1…).[1] Let us reason in this non-standard framework.

Let H(1) be the event that the first toss comes up heads and H(2…) the event that every toss after the first comes up heads. Thus H(1…) is equivalent to the conjunction H(1) & H(2…). Prob(X) is the probability of X and Prob(X | Y) the probability of X conditional on Y. By the nature of conditional probabilities:

(1) Prob(H(1…)) = Prob(H(1) & H(2…)) = Prob(H(1)).Prob(H(2…) | H(1))

Since the coin is fair:

(2) Prob(H(1)) = ½

Since the tosses are independent:

(3) Prob(H(2…) | H(1)) = Prob(H(2…))

By (1)-(3):

(4) Prob(H(1…)) = Prob(H(2…))/2

But H(1…) and H(2…) are isomorphic events. More precisely, we can map the constituent single-toss events of H(1…) one-one onto the constituent single-toss events of H(2…) in a natural way that preserves the physical structure of the set-up just by mapping each toss to its successor. H(1…) and H(2…) are events of exactly the same qualitative type; they differ only in the inconsequential respect that H(2…) starts one second after H(1…). That H(2…) is preceded by another toss is irrelevant, given the independence of the tosses. Thus H(1…) and H(2…) should have the same probability. 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…) be the event that every toss of the second coin comes up heads, and H*(2…) the event that every toss after the first of the second coin comes up heads. Then H(1…) and H*(1…) 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…) and H(2…) 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…) sequence begins is irrelevant. By transitivity, H(1…) and H(2…) should be equiprobable:

(5) Prob(H(1…)) = Prob(H(2…))

By (4) and (5):

(6) Prob(H(1…)) = Prob(H(1…))/2

But even in non-standard analysis the principle $x = x/2 \rightarrow x = 0$ holds universally, for non-standard analysis concerns non-standard models of the very same first-order theory as standard analysis. Thus (6) yields:

(7) Prob(H(1…)) = 0

This argument for (7) 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. The same goes for every other specific outcome. Yet they are all possible, and one of them will be actual. Each outcome has probability 1 of not occurring, but it is not certain that it will not occur. Regularity fails. Infinitesimal probabilities may be fine in other cases, but they do not solve the present problem.

Is the underlying problem the attempt in such cases to measure probabilities by numbers, standard or non-standard? If so, we might fall back on ordinal relations of comparative probability, such as 'is more probable than' (>) and 'is at least as probable as' (≥). Let us not even assume that comparative probability involves a total ordering. Thus 'X > Y' and 'Y ≥ X' may be contraries rather than contradictories. If ⊥ is a contradiction (corresponding to the null set of possibilities), 'X > ⊥' replaces the claim that X has probability greater than 0. Thus regularity becomes the constraint that X > ⊥

(and $\neg\bot > \neg X$) unless X corresponds to the null set of possibilities. Can the claim that H(1…) is more probable than a contradiction be defended in this purely qualitative setting?

A promising principle for comparative probability is this:[2]

(!) If X and Y are each incompatible with Z, then

(a) $X > Y$ if and only if $X \vee Z > Y \vee Z$

(b) $X \geq Y$ if and only if $X \vee Z \geq Y \vee Z$

(Unless comparative probability is a total ordering, (!a) and (!b) are not equivalent.) One motivation for (!) is that if X and Y are each incompatible with Z, then the possibilities of X without Y are exactly the possibilities of $X \vee Z$ without $Y \vee Z$ and the possibilities of Y without X are exactly the possibilities of $Y \vee Z$ without $X \vee Z$; since those possibilities should determine any probabilistic difference between X and Y or between $X \vee Z$ and $Y \vee Z$ respectively, the comparative probability relations within the two pairs should be the same. Without (!), what substance is there to regarding $>$ and $\geq$ as relations of comparative *probability*?

Now let X be H(1…), Y be $\bot$ and Z be $\neg$H(1) & H(2…). Clearly, H(1…) and $\bot$ are both incompatible with $\neg$H(1) & H(2…). Thus (!a) yields:

(8) H(1…) $> \bot$ if and only if H(1…) $\vee$ ($\neg$H(1) & H(2…)) $> \bot \vee$ ($\neg$H(1) & H(2…))

But $H(1\ldots) \vee (\neg H(1) \,\&\, H(2\ldots))$ is logically equivalent to $H(2\ldots)$ (under appropriate definitions), and $\perp \vee (\neg H(1) \,\&\, H(2\ldots))$ to $\neg H(1) \,\&\, H(2\ldots)$. Given the natural constraint that comparative probability relations are invariant across logical equivalents, (8) simplifies thus:

(9) $H(1\ldots) > \perp$ if and only if $H(2\ldots) > \neg H(1) \,\&\, H(2\ldots)$

Since $\neg H(1) \,\&\, H(2\ldots)$ and $H(1\ldots)$ differ only on the outcome of the first toss, which is equally likely to be heads or tails, and the other tosses are independent of it, the two events are equiprobable. Consequently:

(10) $\neg H(1) \,\&\, H(2\ldots) \geq H(1\ldots)$

Since $X > Y$ and $Y \geq Z$ entail $X > Z$ by the logic of comparatives, (10) and the left-to-right direction of (9) yield:

(11) $H(1\ldots) > \perp$ only if $H(2\ldots) > H(1\ldots)$

The considerations above in favour of (5) concerned only comparative probability relations. Thus in the present notation they become considerations in favour of:

(12) $H(1\ldots) \geq H(2\ldots)$

Since (12) is inconsistent with the consequent of (11):

(13) Not $H(1\ldots) > \perp$

A sequence of infinitely many heads is not more probable than a contradiction; regularity fails.

We can reach a similar conclusion using (!b) in place of (!a). Let X be $\perp$, Y be $H(1\ldots)$ and Z be $\neg H(1)\ \&\ H(2\ldots)$. Thus (!b) yields:

(14) $\perp\ \geq H(1\ldots)$ if and only if $\perp\ \vee\ (\neg H(1)\ \&\ H(2\ldots)) \geq H(1\ldots)\ \vee\ (\neg H(1)\ \&\ H(2\ldots))$

By the same simplifications as before, (14) becomes:

(15) $\perp\ \geq H(1\ldots)$ if and only if $\neg H(1)\ \&\ H(2\ldots) \geq H(2\ldots)$

From (10) and (12) by transitivity:

(16) $\neg H(1)\ \&\ H(2\ldots) \geq H(2\ldots)$

Thus from (16) and the right-to-left direction of (15):

(17) $\perp\ \geq H(1\ldots)$

A contradiction is at least as probable as an infinite sequence of heads.

Thus (!a) and (!b) each separately generate the analogue for comparative probability of the conclusion that an infinite sequence of heads has probability 0. The problem for regularity does not lie in the use of numerical rather than comparative probability, given natural constraints on the latter.

On some regular non-standard probability distributions, the difference in probability between H(1…) and H(2…) is infinitesimal. For any standard probability distribution is approximated to within infinitesimal differences by a regular non-standard distribution (Kraus 1968, McGee 1994). Thus the standard probability distribution on which H(1…) and H(2…) both have probability 0 is approximated by a non-standard distribution on which they both have infinitesimal probabilities. By (4) these probabilities will differ, by an infinitesimal. However, anyone who thinks regularity worth defending by appeal to infinitesimals must think that the difference between probability 0 and an infinitesimal probability matters, and therefore that at least some infinitesimal differences in probability matter. Moreover, the argument that H(1…) has probability 0 is not merely an argument that its probability is at least approximately 0. In particular, the considerations in favour of the crucial claim (5) that H(1…) and H(2…) are equiprobable do not merely favour the claim that they have at least approximately the same probability; they favour the claim that H(1…) and H(2…) have *exactly* the same probability. For the relevant sequences of events are of *exactly* the same qualitative type. Recall that H*(1…) is the event that the second coin comes up heads on every toss. H(1…) has exactly the same chance as H*(1…) and H*(1…) has exactly the same chance as H(2…). It would be unreasonable to give H(1…) more or less credence than H*(1…) or H*(1…) more or less credence than H(2…). Likewise for other kinds of probability. Sometimes the problem with regular non-standard distributions is that too many are eligible: the assignment of one infinitesimal probability rather than another to a given possibility seems arbitrary (Elga 2004). Here, by contrast, the problem is that too few are eligible: none satisfies the non-arbitrary constraints.

One advantage of regularity is that it makes sense of the definition of the conditional probability P(X | Y) as the ratio P(X & Y)/P(Y) whenever Y is possible. But this is not decisive, for conditional probability can be treated as primitive, subject to appropriate axioms, rather than being defined as a ratio of unconditional probabilities (Popper 1955, Renyi 1955). The arguments above do not tell against primitive conditional probabilities on H(1…) ((1) above concerns the conditional probability Prob(H(2…) | H(1)), but since Prob(H(1)) is ½, (1) does not depend on whether conditional probability is primitive or defined). However, such probabilities do not dispel all the present mysteries.

Let the ticket t(X) pay a wonderful prize if X obtains and otherwise nothing (the prize is the same whatever X is). You should be indifferent between t(H(2…)) and the pair of tickets t(H(1…)) and t(¬H(1) & H(2…)), for the outcome is exactly the same in every case, since H(2…) obtains if and only if either H(1…) or ¬H(1) & H(2…) obtains (they cannot both obtain). You should prefer that pair of tickets to t(H(1…)) alone, for to throw away t(¬H(1) & H(2…)) is simply to throw away one chance of a wonderful prize. Thus you should prefer t(H(2…)) to t(H(1…)). But since H(1…) and H*(2…) are isomorphic independent events as already discussed, you should be indifferent between t(H(1…)) and t(H*(2…)). Therefore you should prefer t(H(2…)) to t(H*(2…)). By an exactly symmetric argument, you should prefer t(H*(2…)) to t(H(2…)). What has gone wrong?

Cantor showed that some natural, apparently compelling forms of reasoning fail for infinite sets. This moral applies to forms of probabilistic and decision-theoretic reasoning in a more radical way than may have been realized. Infinitesimals do not solve the problem. Can we do better by weakening (!)?

Notes

* Thanks to Frank Arntzenius and Adam Elga for very helpful discussion.

1 Lewis appeals to infinitesimals in defence of regularity in this very case (1986: 90). Vann McGee uses it in a similar capacity (1994: 179-80).

2 (!) corresponds to de Finetti's fourth axiom of comparative probability (1964: 100). His other three axioms correspond to regularity and the linearity and transitivity of comparative probability. His first three axioms hold on *any* linear ordering of contingencies, with the necessities added at the top and the impossibilities at the bottom.

References

Bernstein, A. R., and Wattenberg, F. 1969. 'Non-standard measure theory', in W. A. J. Luxemburg, ed., *Applications of Model Theory of Algebra, Analysis, and Probability*. New York: Holt, Rinehard and Winston.

Elga, A. 2004. 'Infinitesimal chances and the laws of nature', *Australasian Journal of Philosophy* 82: 67-76.

Finetti, B. de. 1964. 'Foresight, its logical laws, its subjective sources', in H. Kyburg and H. Smokler, eds., *Studies in Subjective Probability*. Huntington, N.Y.: Krieger.

Kraus, P. H. 1968. 'Representation of conditional probability measures on Boolean algebras', *Acta Mathematica Academiae Scientiarum Hungaricae* 19: 229-241.

Lewis, D. K. 1986. *Philosophical Papers: Volume II*. Oxford: Oxford University Press.

McGee, V. 1994. 'Learning the impossible', in E. Eells and B. Skyrms, eds., *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press.

Popper, K. R. 1955. 'Two autonomous systems for the calculus of probabilities', *British Journal for the Philosophy of Science* 6: 51-57.

Renyi, A. 1955. 'On a new axiomatic theory of probability', *Acta Mathematica Academiae Scientiarum Hungaricae* 6: 285-335.