LSE Research Online

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Colin Howson

What probability probably isn't

Article (Accepted version) (Refereed)

Original citation: Howson, Colin (2015) *What probability probably isn't.* Analysis, 75 (1). pp. 53-59. ISSN 0003-2638

DOI: 10.1093/analys/anu111

© 2014 The Author

This version available at: http://eprints.lse.ac.uk/64397/
Available in LSE Research Online: November 2015

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.lse.ac.uk

Appeared in *Analysis*, 2015, 75(3), 413-424.

What Probability Probably Isn't[1]

Abstract

Joyce and others have claimed that degrees of belief are estimates of truth-values and that the probability axioms are conditions of admissibility for these estimates with respect to a scoring rule penalising inaccuracy. In this paper I argue that the claim that the rules of probability are truth-directed in this way depends on an assumption which is both implausible and lacks any supporting evidence, strongly suggesting that the probability axioms have nothing intrinsically to do with truth-directedness.

Key words: probability, truth, admissibility, Joyce, de Finetti.

1 Introduction

A question much discussed, possibly over-discussed, in the annals of philosophical probability is why a rationally-constrained numerical measure of belief should obey the laws of (at least finitely additive) probability. Several answers have been offered, but each of these at some point seems to beg questions. Probably the best-known and certainly the simplest, based on the so-called Dutch Book argument, begs at least two: why an agent will accept either side of a bet at their true degree of belief, even at 'small' stakes, and why they will accept a finite sum of bets (one needs to assume only two) at their corresponding degrees of belief. To some extent these questions are answered by appealing to a more general, and certainly more complex, theory of utility, but that strategy raises further questions which are well-known and which it is not necessary to go into here.

Joyce (1998, 2009) takes a radically different approach, one grounded on what he claims to be an objective *epistemological* criterion that he calls the *Norm of Truth*:

> An epistemically rational agent must strive to hold a system of full beliefs that strikes the best attainable overall balance between the epistemic good of fully believing truths and the epistemic evil of fully believing falsehoods (where fully believing a truth is better than having no opinion about it, and having no opinion about a falsehood is better than fully believing it). (1998, p.577)

Joyce's theory has aroused a great deal of interest and indeed inaugurated a whole programme of research which still flourishes. Since his work is well-known I shall confine myself to a very brief description of it, and then say why I disagree with both its main thrust and its conclusions.

[1] I would like to thank Tim Childers for very helpful comments on an earlier draft of this paper.

2 Gradational Accuracy

Joyce's theory has two main components. One, which he calls 'the Alethic Principle', is the interpretation of 'an epistemically rational' agent's degree of belief in a proposition A as their best estimate of the truth-value of A (2009, p.269). Since a truth-value isn't obviously a number, some numerical coding is required to give workable content to the Principle, and Joyce follows an established convention in coding the truth-values of propositions by their indicator functions, where for any proposition A the indicator function, sometimes also called the characteristic function, $I_A(w)$ of A is equal to 1 if state w makes A true and equal to 0 if not. Given that coding, something like a formal version of the Alethic Principle seems to be embedded in mathematical probability theory (as a consequence of the additivity of probability functions), by virtue of the fact that the probability of A is equal to the expected value of I_A, though clearly this fact can't without circularity be used in support of the Principle if the latter itself is enlisted, as it is in Joyce's theory, in support of an argument for the probability axioms themselves. I shall come back to this point later.

The second component of Joyce's theory is a demonstration that the rules of probability are what decision-theorists call a condition of admissibility, or non-dominance, relative to a measure of the inaccuracy of these estimates: hence

> Since epistemically rational agents must strive to hold accurate beliefs, this establishes conformity with the axioms of probability as a norm of epistemic rationality whatever its prudential merits or defects might be (Joyce 1998, p.575).

What is a measure of inaccuracy? Formally Joyce takes it to be a non-negative real-valued function of belief functions crossed with distributions of truth-values over their arguments (propositions from a Boolean algebra), and he shows that if certain conditions are imposed on it (dominance, continuity, normality, extensionality, convexity and symmetry) then any belief function not satisfying the probability axioms is dominated by one that does.

The precise details of Joyce's argument will be less my concern here than his programme of justifying the probability axioms in terms of admissibility with respect to truth-accuracy, central to which, it will turn out, is the Alethic Principle. I will argue that this alleged principle lacks any adequate supporting argument, is indeed arguably false, and that without it the programme collapses. First, however, I will show that despite its apparent novelty, Joyce's argument is already implicit in some seminal work of the eminent probabilist Bruno de Finetti.

.

3 de Finetti's quadratic scoring rule

In his (1974) de Finetti described a scenario for evaluating the closeness of an agent's estimates ('previsions' in his terminology) of the values of a finite sequence of random quantities $X_1, \dots, X_n$ to their actual observed values. In this scenario, if the estimates are $q_1, .. ,q_n$ and the true values (in the given state) are $x_1, ... ,x_n$, the agent suffers a penalty equal to $\Sigma_i [(x_i - q_i)/k_i]^2$, where the k_i are decided in advance and are dimensionally homogeneous with the X_i (so the penalty is a pure number). De Finetti called the estimates q_i 'coherent' just in case the penalty cannot be reduced across the set of all possible states by choosing any other sequence q'_i : i.e., the original choice is admissible[2]. If the X_i are the indicator functions of n propositions A_i, as will be assumed in what follows, the penalising rule, often called a scoring rule, is *proper* in the sense that the expected value of the penalty is uniquely minimised when the q_i are equal to the agent's corresponding degrees of belief in the A_i . If, as de Finetti did and Joyce does, we code truth-values by indicator functions, and any proper scoring rule is used, we obtain in effect Joyce's Alethic Principle: it is in the agent's interest to set their degrees of belief in the A_i equal to their estimates of the corresponding truth-values.

De Finetti points out that his quadratic penalty is the square of the distance between vectors **x** and **q** in the linear space R^n endowed with the metric $\Sigma(x_i/k_i)^2$ (1974, p.90), yielding the Euclidean metric when the k_i are all equal to unity; the corresponding penalty is the so-called Brier score $\Sigma(x_i-q_i)^2$. De Finetti then proceeded to show, by a simple geometric argument, that a necessary and sufficient condition for an agent's estimates to be coherent is that they satisfy the finitely additive probability axioms (ibid.). For example, suppose n=3, that the A_i are a partition and the penalty is the Brier score. The A_i can then be represented as the three points (1,0,0), (0,1,0), (0,0,1) at unit distance from the origin in 3-dimensional Euclidean space. The simplex (in this case plane triangle) whose vertices are these three points is their convex hull, i.e. the set of all convex combinations of the extreme points and hence the set of triples (p_1,p_2,p_3) such that $p_1+p_2+p_3 = 1$, $p_i \geq 0$. Now suppose P* is any point outside the simplex. Its projection onto the simplex clearly has a shorter distance to each of the vertices than that from P*. Notice that while any probabilistically incoherent estimate is dominated by a probabilistically coherent one, no probabilistically coherent estimate will dominate any other.

Joyce discusses de Finetti 's theory, but rejects it because according to him (Joyce) it justifies probabilistic coherence on purely prudential grounds, lacking, Joyce claims, any epistemic rationale – the sort of rationale that he believes is afforded by his inaccuracy-theoretic account (1998, pp.584-585). However, the objection seems to have overlooked the fact that, despite de Finetti's drawing attention to it and indeed exploiting it, the additive quadratic scoring rule is both a penalty and a distance in a finite-dimensional linear space. In its second guise, it seems to perform the function Joyce demands of a cognitively adequate justification of the probability axioms: given the coding of truth-values by indicator functions, a probabilistically coherent forecast cannot be dominated in terms of an overall reduction in distance from the truth by any incoherent one[3]; indeed, in a section of his (2009) headed 'In praise of

[2] de Finetti gave two formal characterisations of coherence in (1974) and proved their equivalence. One is that described above; the other relates to an agent forced to give odds on a set of propositions when the stakes and the directions of the bets are set by an opponent: coherence in this sense means that no finite sum of such bets will yield a positive gain or loss. Since then the word 'coherent' has largely come to mean 'consistent with the probability calculus'. This last I will call 'probabilistic coherence'.

[3] After de Finetti published his result the question was raised to what extent it depended on the precise form of his scoring rule. Several generalisations have since been proved, among them Joyce's, culminating in the paper by Predd et al. (2009) who, in an analogue of de Finetti's geometrical argument but based on a distance function

the Brier rule' Joyce commends it precisely on the ground that it represents what he calls 'pure accuracy' (p.293)[4].

4 The Catch

The finitely additive probability axioms therefore seem to be demonstrable conditions of admissibility for the 'pure' truth-inaccuracy measured by the Brier rule (and any proportional to it like its arithmetical average, also often referred to simply as 'the Brier rule'). Perhaps surprisingly, that conclusion does not follow. It does not follow, because the indicator function $I_A(w)$ for A is just one method of coding numerically the truth-value of A in state w (though any other in which the number representing 'true' is greater than that representing 'false' is inessentially different, and there is another coding, no less faithful, for which the conclusion is false.

That other coding is the complementary function $C_A = 1-I_A$. Despite certain practical advantages, the choice of I_A is of course merely conventional, as Joyce himself concedes, and some systems of multi-valued logic do in fact choose C_A. So suppose we do so also in the present context. The coordinates of the A_i in the C_A coding are (0,1,1), (1,0,1), (1,1,0). The convex hull *K'* of these points lies in a plane parallel to that of the convex hull *K* of the A_i in the old system. *K* is the set of all points (p_1, p_2, p_3), $p_i \geq 0$, $\Sigma p_i = 1$, and *K'* is the set of all points (p'_1,p'_2,p'_3) where $p'_i = 1-p_i$. It follows that every probability distribution over the algebra generated by the A_i is dominated in terms of distance from the truth by some forecast in *K'*, namely the projection of that probability distribution onto *K'*. The identification of probabilistic coherence with gradational accuracy thus appears to be broken under this coding. Indeed, where P is a probability function, it is the forecast $f(A_i) = 1-P(A_i)$ of the 'truth-value' of A_i which cannot be uniformly improved on. True, Joyce actually *identifies* a proposition A with I_A (2009, p.268), but this simply pre-empts and does not answer a potential charge of arbitrariness.[5]

It is true that the penalty $(C_A(w)-f)^2$ for estimate f is not proper: its expected value for the belief distribution P(A) = P(A is true) = $P(C_A=0)$ = p, P(~A) = $P(C_A=1)$ = 1-p is minimised at f = 1-p, not p. This means that your forecast of the value of the truth-value of A (under the C_A coding) records your degree of belief in the *falsity* of A. Joyce's response is to invert the statement above into the claim that the coding 'true' = 0, 1 = 'false' implies the adoption of a *belief function* ***b*** such that ***b***(A) has the formal properties of P(~A): setting the code of the truth-value of a contradiction greater than that of a logical truth, he tells us,

---

called Bregman divergence, show that relative to *any* continuous proper scoring rule (de Finetti's rule is continuous), if a forecast f is probabilistically coherent then it is not weakly dominated by any forecast, and if f is probabilistically incoherent then it is strongly dominated by some probabilistically coherent forecast (ibid., Theorem 1). I strongly suspect, however, that de Finetti's seeming lack of interest in generalising his result to a wider class of penalty functions was because of the direct geometrical significance he gave to the quadratic penalty.

[4] Though in (2009) he deems a single norm of accuracy 'unduly restrictive' (p.264), favouring a broader constraint of 'epistemic utility'.

[5] Leitgeb and Pettigrew's development (op. cit.) of Joyce's ideas is also based on determining inaccuracy in terms of distance in Euclidean n-space from the points (1, 0, 0, ...), (0,1, 0, ...), ... representing n propositions A_1, A_2, ...

> requires us to interpret larger ***b***-values as signaling *less* confidence in the *truth* of a proposition and *more* in confidence in its *falsehood*.' (2009, p.268, footnote 7)

This seems on the face of it a very strange claim: why should the choice of a conventional coding of truth-values tell us anything about the properties of a belief function?[6] Joyce himself lists as basic properties of a belief function that it sets the degree of belief in the impossible proposition at a value strictly less than that of the certain proposition, and the degrees of belief in these two propositions are the endpoints of the interval in which belief is measured[7] (and he also adds finite additivity). But none of these properties appears to depend on the way truth-values are coded. The mystery is compounded by Joyce's remark, in this same context, that 'If one mistakenly tries to retain the idea that ***b*** measures confidence in truths while setting ***b***(~T) > ***b***(T), one ends ups with nonsense' (2009, p.264, footnote 2; T is the logical truth in an algebra of propositions). But that is so obviously true that one might well wonder why Joyce bothers to mention it.[8]

The mystery is dispelled by recalling that Joyce includes the Alethic Principle among the basic properties of belief functions of 'epistemically rational agents'. As best estimates of truth-values (according to the Principle), these belief functions must reflect the latter's ordering under whatever numerical coding is chosen. Since belief functions must also 'measure confidence in truths' they must satisfy the condition ***b***(~T) < ***b***(T). Thus the C_A coding is ruled out and I_A is uniquely determined once the end-points of ***b*'s** scale are fixed at 0 and 1. So everything depends on accepting the Alethic Principle in addition to pretty well universally accepted properties of belief functions. The only possible justification for doing so, however, would seem to be that the Principle is somehow implicit in the requirement of 'epistemic rationality'. But even if the aim of the epistemically rational agent is truth, it hardly follows that their degrees of belief should be interpreted as best estimates of numerically-coded truth-*values*. To claim otherwise seems only to conflate 'degree of belief in A's *truth*' with 'best estimate of A's truth-*value*', and I see no reason to suppose that these should necessarily be identical. On the contrary, if I choose to represent A's truth-value by C_A then clearly the identity fails and one can't argue, without obvious circularity, that I am forbidden to make that choice on the ground that it conflicts with the conjunction

---

[6] We find the same sort of thing in Leitgeb and Pettigrew (op. cit.): 'assigning a degree of belief 1 to a proposition A would mean that the agent believes that A is true rather than false, *since the degree of belief 1 is closer – in fact identical – to the real number 1 that represents truth than it is to the real number 0 representing falsity*' (2010, p.9; my emphasis).

[7] One could even take these end-points as plus and minus infinity.

[8] The remark is aimed at me. Joyce continues: '[that mistake] seems to be the basis of the worries raised in Howson (2008, pp. 20–21).' This is not true: here is what I said "it is not clear that it is inaccuracy with respect to *truth* that Joyce's measure represents, depending as it seems to do on the (purely conventional) use of 1 as the numerical proxy for 'true' rather than 0. Indeed, by changing these values round one gets a very different result. A perfectly accurate belief function b with respect to [C_A] is now only *dually* probabilistic, with b assigning the value 0 to a tautology, etc., and Joyce's proof would show that for any probabilistic belief function there is a non-probabilistic one strictly less inaccurate than it with respect to all [C_A]-valuations. Of course, the definition of dominance ... could be transformed correspondingly so that larger numerical discrepancies between b-values and [C_A]-values correspond to smaller 'distances' from the truth, but that would hardly *explain* the probability axioms as taking the form they do."

of the Alethic Principle and the condition $\boldsymbol{b}(\perp) < \boldsymbol{b}(\mathrm{T})$. Neither can one appeal without circularity to the fact that the probability that A is true is equal to the expected value of I_A, since that assumes both that truth-values are canonically coded by indicator functions, and that degrees of belief should obey the probability axioms, which is what ultimately is to be proved. Nor, again without begging the question, can one adduce the fact that it is in my interest, judged by expected utility, to make my degree of belief in A equal to my forecast of I_A when a proper scoring rule is used to reckon distance from the truth of such forecasts: the question is begged, of course, by identifying I_A with the truth-value of A.

5 Conclusion

This is not the place for a detailed assessment of other ways of trying to link the rules of probability to some form of truth-directness. *Calibration* at one time seemed an attractive option, proof against Hume's deadly arguments for inductive scepticism, but Joyce himself in (1998) added to the already powerful battery of arguments in the literature against it. Unfortunately, if the foregoing is correct, his own account fares little better, depending as it does on the Alethic Principle for which there appears to be no non-question-begging supporting argument. But without that postulate there is nothing to link belief functions with the choice of truth-value coding by indicator functions, and consequently, given the symmetry between I_A and C_A as such codes, nothing to deliver the probability axioms as admissibility conditions on truth-accuracy. Together with the apparent failure of the calibration programme, this seems to me to raise a large question mark against the claim that the probability axioms have anything intrinsically to do with truth-accuracy at all.

References

de Finetti, B. 1974. *Theory of Probability*, vol. 1, Wiley.

Howson, C. 2008. 'De Finetti, Countable Additivity, Consistency and Coherence', *British Journal for the Philosophy of Science,* **59**, 1–23.

Joyce, J. 1998. 'A Non-Pragmatic Vindication of Probabilism', *Philosophy of Science,* **65**, 575–603.

Joyce, J. 2009. 'Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief'**,** F. Huber, C. Schmidt-Petri eds., *Degrees of Belief*, Synthese Library 342, Springer, 263-297.

Leitgeb, H. and Pettigrew, R. 2010. 'An Objective Justification of Bayesianism I: Measuring Inaccuracy', *Philosophy of Science*, **77**, 201-235.

Predd, J.B., Seiringer, R., Lieb, H.L., Osherson, D.N., Poor, H.V., Kulkarni, S.R. 2009. 'Probabilistic Coherence and Proper Scoring Rules', *IEEE Transactions on Information Theory*, **55**, 4786-4792.