Jeanne Peijnenburg and David Atkinson

Biased Coins
A model for higher-order probabilities

Abstract

Is it coherent to speak of the probability of a probability, and the probability of a probability of a probability, and so on? We show that it is, in the sense that a regress of higher-order probabilities can lead to convergent sequences that determine all these probabilities. By constructing an implementable model which is based on coin-making machines, we demonstrate the consistency of our regress.

Keywords: Higher-order probability, infinite regress, consistency.

1. Introduction

If it makes sense to express the probability that a proposition is true as a specific number, it also makes sense to doubt whether that number itself is accurate. One can further consider the probability that the probability of the proposition is equal to the number in question. Hence we are led ineluctably into an infinite regress, for then we must consider the probability of the probability of the probability that the proposition is true – and so on *ad infinitum*. It has indeed been argued that this sounds the death knell of the idea of higher-order probabilities (Hume 1738/1961, Rescher 2010). However, in this paper we shall show that the regress in question is generally benign. Far from producing a *reductio ad absurdum*, as we will see, the regress usually engenders a convergent sequence that leads to a well-defined probability of the truth of the original proposition.

Some have seriously entertained the notion of a second-order probability (Uchii 1973, Skyrms 1980, Domotor 1981, Kyburg 1987, Gaifmann 1988). Skyrms in particular discusses and demolishes a number of attempts aimed at showing that the concept is inconsistent. One of those attempts has to do with Miller's paradox, to which Skyrms gives short shrift, dismissing it as "simply a fallacy of equivocation".[1] Skyrms continues by presenting a short proof in the form of a model, which makes it clear that second-order probabilities are formally consistent. The model uses relative frequencies, but Skyrms, tongue-in-cheek, argues

1 Skyrms 1980, p. 111. For the original statement of this paradox, if it deserves the name, see Miller 1966.

M.C. Galavotti et al. (eds.), *European Philosophy of Science – Philosophy of Science in Europe and the Viennese Heritage*, Vienna Circle Institute Yearbook 17, DOI 10.1007/978-3-319-01899-7_17,

that it is implicit in the work of that champion of subjective probabilities, Bruno de Finetti:

> I would say nothing more about formal inconsistency were it not that some reputable philosophers continue to have suspicions (if not arguments) ... Though it may be a case of bringing out a cannon to swat a fly, I therefore feel obliged to point out that there is implicit in de Finetti's work a proof of formal consistency for a theory of second order probabilities: simply interpret $[Pr]$ as relative frequency probability. ... This is not the intended interpretation, but it suffices to settle the question of consistency.[2]

Here Skyrms is interpreting a relation like $PR[Pr(E) = x]$, where Pr is a first-order probability, and PR a second-order probability, by making the former an objective, relative-frequency chance, and the latter a subjective probability about a proposition concerning that chance.

Skyrms' model of a second-order probability is assuredly successful. However he remained silent on the question whether the same goes for an infinite regress of higher-order probabilities: does the concept of a regress of probabilities of probabilities, and so on, make sense? In a recent paper we have explained how to set up such a regress, and we also showed how to calculate the probability of the original proposition by summing a convergent series[3]. What we did not do, however, was to consider the formal question of the *consistency* of our equations. In the present paper we fill this lacuna, i.e. we construct a model of an infinite regress of probabilities of probabilities, where 'model' is used in the logical sense of a structure that makes all the sentences of a theory true. Our model is based on coin-making machines and can in principle be implemented. We show that when the coin-making structure is inserted into the abstract equations defining a regress of higher-order probabilities, the resulting statements are indeed all true.

We start in Section 2 by sketching the relevant results of our earlier paper.[4] In Section 3 we work out a numerical example of a regress of higher-order probabilities. Finally in Section 4 we offer our model, which involves coin-making machines. We show that, in this model, all the formulas of the abstract theory developed in Section 2 are true; thus the regress is consistent.

---

2 Skyrms 1980, p. 112.

3 Peijnenburg and Atkinson 2012.

4 As in that paper, we assign a particular number (rather than an interval) to the probability that a given proposition is true. This is meaningful only if the probabilities in question are discrete. The generalization to a continuous probability distribution will be given elsewhere – Atkinson and Peijnenburg, to appear.

2. PROBABILITIES OF PROBABILITIES

In this section we set up a regress of higher-order probabilities in general terms. For convenience we will talk about the probabilities as if they were subjective; but in fact some could be objective and some subjective: it is only the abstract system that is of importance here.

Let q_0 stand for some proposition of which we doubt the truth. Perhaps we think the probability that q_0 is true is merely v_0, this being some number between 0 and 1. Define q_1 to be the proposition $P(q_0) = v_0$. If we were quite sure that q_1 itself is true, we would effectively be asserting that the probability of q_0 is simply equal to the conditional probability of q_0, given q_1. Let us designate this, our first estimate of the probability of q_0, as $P^{(1)}(q_0)$, where

$$P^{(1)}(q_0) = P(q_0|q_1) = P(q_0|P(q_0) = v_0) = v_0\,. \tag{1}$$

The last equality in (1), namely $P(q_0|P(q_0) = v_0) = v_0$, is warranted by what Skyrms has called Miller's principle. In words, it says that the probability of q_0, given that the probability of q_0 is v_0, is v_0.[5]

However it may be that we are not at all sure that q_1 is true, but only think that its probability is v_1. This thought can be expressed by the proposition $P(q_1) = v_1$, which we will dub q_2. If were sure that q_2 is true, then we would be asserting that the probability of q_1 is v_1. In that case, our next estimate of the probability of q_0 would be $P^{(2)}(q_0)$, which, with use of the rule of total probability, is given by

$$\begin{aligned} P^{(2)}(q_0) &= P(q_0|q_1)P(q_1) + P(q_0|\neg q_1)P(\neg q_1) \\ &= v_0 v_1 + w_0(1 - v_1)\,. \end{aligned} \tag{2}$$

Here we have used Miller's principle as above, and we have abbreviated the conditional probability $P(q_0|\neg q_1)$ by the symbol w_0. Clearly this new estimate of the value of the probability of q_0 is *not* equal to v_0; it will be somewhere between v_0 and w_0. The new estimate of the probability of q_0 can be thought of as an update on the previous estimate.

The update followed from the provisional assumption that q_2 is true; but suppose next that q_2 is not known to be true, and let q_3 be the proposition $P(q_2) = v_2$. On the assumption that q_3 is true, we would set $P(q_2)$ equal to v_2, so we could

---

5 As Skyrms has it: "Those who have followed the development of modal logic will already know that we invite no additional difficulty by universally generalising Miller's principle to

$$\text{for any } x,\, x = PR[E \text{ given that } Pr(E) = x]\text{"}$$

(Skyrms 1980, p. 112). As we have noted, Skyrms uses the symbol Pr for the first-order probability and PR for the second-order probability. Miller's principle has the same form as David Lewis's Principal Principle, but Lewis would interpret Pr as an objective chance, PR as a subjective credence (Lewis 1980).

then write, at this level of estimation for the probability of q_1,

$$\begin{aligned} P(q_1) &= P(q_1|q_2)P(q_2) + P(q_1|\neg q_2)P(\neg q_2) \\ &= v_1 v_2 + w_1(1 - v_2)\,, \end{aligned} \tag{3}$$

where $P(q_1|\neg q_2)$ has been designated by the symbol w_1. The next update for the probability of q_0 is $P^{(3)}(q_0)$, which has the form

$$\begin{aligned} P^{(3)}(q_0) &= P(q_0|q_1)P(q_1) + P(q_0|\neg q_1)P(\neg q_1) \\ &= v_0 P(q_1) + w_0[1 - P(q_1)]\,, \end{aligned} \tag{4}$$

where for $P(q_1)$ we are to understand the expression on the right of (3), namely $v_1 v_2 + w_1(1 - v_2)$. When this is inserted into (4), $P^{(3)}(q_0)$ is thereby expressed as a function of v_0, w_0, v_1, w_1, and v_2.

It should be clear now how to continue the sequence of updates. The nth step leads to $P^{(n)}(q_0)$, which can be evaluated in an analogous way. By way of illustration, a numerical example is worked out in the next section. The sequence of updates $P^{(n)}(q_0)$ converges to a limit, except in extreme cases.[6]

3. A Numerical Example

Consider the uniform situation in which $P(q_0|q_1) = P(q_1|q_2) = P(q_2|q_3) = \ldots = 0.9$ and $P(q_0|\neg q_1) = P(q_1|\neg q_2) = P(q_2|\neg q_3) = \ldots = 0.5$. This assumption of uniformity is introduced purely to facilitate the calculation. The method of Section 2 also works when the conditional probabilities do differ from step to step in the regress, on condition that the constraint (5) in footnote 6 is respected.

We begin with

$$P^{(1)}(q_0) = P(q_0|q_1) = 0.9\,,$$

The next update is calculated from Eq.(2), where in this uniform example $P(q_1)$ is the same as $P^{(1)}(q_0)$, so it is equal to 0.9. Thus

$$\begin{aligned} P^{(2)}(q_0) &= 0.9 \times 0.9 + 0.5 \times (1 - 0.9) \\ &= 0.86\,. \end{aligned}$$

6 Convergence is guaranteed if the difference, $v_n - w_n$, does not approach one too rapidly as n increases. To be precise, the sequence converges so long as the difference between $v_n - w_n$ and 1 does not tend to zero faster than does $1/n$. In mathematical notation

$$1 - v_n + w_n = O\,(1/n) \quad \text{as } n \text{ tends to } \infty\,, \tag{5}$$

where $O\,(1/n)$ means "of order $1/n$". For a proof of this convergence, see Atkinson and Peijnenburg (2010), Appendices A and B.

To calculate $P^{(3)}(q_0)$ we first need $P(q_1)$ at order 2, and because of uniformity this is the same as $P^{(2)}(q_0)$, namely 0.86. Hence

$$\begin{aligned} P^{(3)}(q_0) &= 0.9 \times 0.86 + 0.5 \times (1 - 0.86) \\ &= 0.844\,. \end{aligned}$$

For update number 4 we need $P(q_1)$ at order 3, which, again thanks to uniformity, is the same as $P^{(3)}(q_0)$, which we have just calculated to be 0.844, so

$$\begin{aligned} P^{(4)}(q_0) &= 0.9 \times 0.844 + 0.5 \times (1 - 0.844) \\ &= 0.8376\,. \end{aligned}$$

and so on.

Here is an overview of the values of $P^{(n)}(q_0)$ after an increasing number of updates:

n	1	2	3	4	5	6	10	∞
$P^{(n)}(q_0)$	0.9	0.86	0.844	0.8376	0.8351	0.8340	0.83334	$\frac{5}{6}$

Probability of q_0 after n updates

Note that $P^{(\infty)}(q_0)$ is precisely equal to $\frac{5}{6} = 0.83333...$, and that $P^{(10)}(q_0)$ is already very close indeed to this limiting value.

4. Model of higher-order probabilities

In this section we set up a model of the abstract system of equations in Section 2. Although we talked about the system in terms of subjective probabilities of subjective probabilities, it is indeed abstract in its form. The only requirement is that the probabilities satisfy Kolmogorov's axioms. The probabilities in the model we are about to describe are however all objective chances, which we will couch in the language of relative frequencies. The purpose is twofold: (a) to show that the equations governing the model are precisely those of Section 2, and (b) to show that all the implications of the abstract equations are true in the model. This demonstrates that the abstract system of higher-order probabilities is consistent.

Suppose there are two machines, each of which produces trick coins. Machine V_0 makes coins each of which has bias v_0, by which we will mean that each has probability v_0 of falling heads when tossed; whereas machine W_0 makes coins each of which has bias w_0. An experimenter tosses a coin from machine V_0. We identify the propositions q_0 and q_1 as follows:

q_0 is the proposition "this coin will fall heads"

q_1 is the proposition "this coin comes from machine V_0" .

The probability that q_0 is true is based on the datum that the coin comes from machine V_0, in other words that proposition q_1 is true. Clearly

$$P(q_0|q_1) = P\big(\text{“this coin will land heads”}\,|\,\text{“this coin comes from machine } V_0\text{”}\big)$$

is the same thing as $P\big(q_0|P(q_0) = v_0\big)$, for if the coin has come from machine V_0, the probability of a head is v_0; and conversely, if the probability is v_0, the coin must have come from machine V_0. As in Section 2, we have $P(q_0|q_1) = v_0$.

The experimenter is now instructed to take many coins from *both* machines, and to mix them thoroughly in a large pile. Moreover, the numbers of coins that have been added to the pile from machines V_0 and W_0 are not entirely left to the whim of our experimenter, for their relative number is determined by a second experiment, which is performed by a supervisor. This second experiment is much like the first one, but it involves two new machines, V_1, which produces trick coins with bias v_1, and W_1, which produces trick coins with bias w_1. The supervisor extracts a coin from machine V_1; and he instructs the experimenter to make sure that the relative number of coins that she takes from her machine V_0 is equal to the probability that his coin falls heads when tossed. That is to say, the number of coins that she must add to the pile from machine V_0 is equal to v_1 multiplied by the total number of coins removed from machines V_0 and W_0.

The experimenter takes one coin at random from her pile and, understanding q_0 now to refer to this coin, we can deduce the probability of q_0 in the new situation. Indeed, if

$$q_2 \quad \text{is the proposition} \quad \text{“the supervisor's coin comes from machine } V_1\text{”}\,,$$

and q_2 is true, then $P(q_1) = v_1$. We conclude that, if the experimenter were to repeat the procedure of tossing a coin from her pile many times (with replacement), the resulting relative frequency of heads would be approximately equal to $P^{(2)}(q_0)$, as given by (2) (and the approximation would get better and better as the number of tosses increases – more carefully: the probability that the relative number of heads will differ by less than any assigned $\varepsilon > 0$ from v_0 will tend to unity as the number of tosses tends to infinity). This concludes the description of the model of the first iteration of the regress, constrained by the condition that the supervisor's coin comes from machine V_1, that is by the veridicality of q_2.

In the next iteration, the supervisor receives instructions from an AI (artificial intelligence) that simulates the working of yet another duo of machines, V_2 and W_2, which produce simulated coins with biases v_2 and w_2, respectively. The supervisor makes a large pile of coins from his machines V_1 and W_1; and he adjusts the relative number of coins that he takes from V_1 to be equal to the probability that a simulated coin from V_2 would fall heads when tossed. That is to say, the number of coins in the instructor's pile that have been taken from V_1, divided by the total number that have been taken from V_1 and W_1, is equal to v_2.

If the supervisor were to repeat the procedure of tossing a coin from his pile many times (with replacement), then the resulting relative frequency of heads would be approximately equal to $P(q_1)$, as given by (3) (with the usual probabilistic proviso). This value of $P(q_1)$ is handed down to the experimenter, and she runs her experiment as above, but with the updated value of $P(q_1)$. The relative frequency of heads that she would observe will be approximately equal to $P^{(3)}(q_0)$, as given by (4). The above constitutes a model of the second iteration of the regress, constrained by the condition that the AI's simulated coin comes from the simulated machine V_2, that is by the veridicality of q_3, where

$$q_3 \quad \text{is the proposition} \quad \text{“this simulated coin comes from simulated machine } V_2\text{”} .$$

Of course this procedure must be repeated *ad infinitum*. A subprogram must simulate the working of yet another duo of machines, V_3 and W_3, which program the production of coins with biases v_3 and w_3, and so on. In this way an implementable model has been produced for an arbitrary number of iterations of the abstract system of Section 2, thereby showing that the whole regress is consistent.

Acknowledgements
We would like to thank the audience of the conference “European Philosophy of Science and the Viennese Heritage” for their helpful remarks, in particular Dennis Dieks, John Norton, Wlodek Rabinowicz and Miklós Rédei.

References

Atkinson, David, and Jeanne Peijnenburg. 2010. “The Solvability of Probabilistic Regresses. A Reply to Frederik Herzberg”, *Studia Logica*, vol. 94, no. 3, 347-353.

Domotor, Zoltan. 1981. “Higher Order Probabilities”, *Philosophical Studies*, vol. 40, no. 1, 31-46.

Gaifmann, Haim. 1988. “A Theory of Higher Order Probabilities”, in B. Skyrms and W.L. Harper (eds.), *Causation, Chance and Credence*. London: Kluwer, pp. 191-219.

Hume, David. 1738/1961. *A Treatise of Human Nature, Volume 1*. London: J.M. Dent.

Kyburg, Henry E. 1987. *Higher Order Probabilities and Intervals.* Volume 36 of Technical Report Department of Computer Science, University of Rochester.

Lewis, David. 1980. “A Subjectivist's Guide to Objective Chance”, in R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability. Vol II.* Berkeley: University of California Press, 263-293.

Miller, David. 1966. “A Paradox of Information”, *British Journal for the Philosophy of Science*, vol. 17, 59-61.

Peijnenburg, Jeanne, and David Atkinson. 2012. “An Endless Hierarchy of Probabilities”, *American Philosophical Quarterly*, vol. 49, no. 3, 267-276.

Rescher, Nicholas. 2010. *Infinite Regress. The Theory and History of Varieties of Change*. New Brunswick, NJ: Trabsaction.

Skyrms, Brian. 1980. “Higher Order Degrees of Belief”, in D.H. Mellor (ed.), *Prospects for Pragmatism: Essays in Memory of F.P. Ramsay*. Cambridge: Cambridge University Press, 109-137.

Uchii, Soshichi. 1973. “Higher Order Probabilities and Coherence”, *Philosophy of Science*, vol. 40, no. 3, 373-381.

Faculty of Philosophy
University of Groningen
Oude Boteringestraat 52
9712 GL Groningen
The Netherlands
d.atkinson@rug.nl
jeanne.peijnenburg@rug.nl