

University of Groningen

A consistent set of infinite-order probabilities

Atkinson, David; Peijnenburg, Jeanne

Published in:
International Journal of Approximate Reasoning

DOI:
[10.1016/j.ijar.2013.04.015](https://doi.org/10.1016/j.ijar.2013.04.015)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Atkinson, D., & Peijnenburg, J. (2013). A consistent set of infinite-order probabilities. *International Journal of Approximate Reasoning*, 54(9), 1351-1360. <https://doi.org/10.1016/j.ijar.2013.04.015>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



ELSEVIER

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar



A consistent set of infinite-order probabilities



David Atkinson*, Jeanne Peijnenburg

University of Groningen, Faculty of Philosophy, Oude Boteringestraat 52, 9712 GL Groningen, Netherlands

ARTICLE INFO

Article history:

Received 18 January 2013
 Received in revised form 24 April 2013
 Accepted 26 April 2013
 Available online 6 May 2013

Keywords:

Model
 Higher-order probability
 Infinite regress

ABSTRACT

Some philosophers have claimed that it is meaningless or paradoxical to consider the probability of a probability. Others have however argued that second-order probabilities do not pose any particular problem. We side with the latter group. On condition that the relevant distinctions are taken into account, second-order probabilities can be shown to be perfectly consistent.

May the same be said of an infinite hierarchy of higher-order probabilities? Is it consistent to speak of a probability of a probability, and of a probability of a probability of a probability, and so on, *ad infinitum*? We argue that it is, for it can be shown that there exists an infinite system of probabilities that has a model. In particular, we define a regress of higher-order probabilities that leads to a convergent series which determines an infinite-order probability value. We demonstrate the consistency of the regress by constructing a model based on coin-making machines.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Let q_0 be a proposition with a probability v_0 :

$$P^1(q_0) = v_0, \quad (1)$$

where P^1 stands for an ordinary unconditional probability of the first order, and where v_0 is some number between 0 and 1. We now might try to assert a second-order probability, P^2 , to the effect that the probability of q_0 , given (1), is v_0 :

$$P^2(q_0|P^1(q_0) = v_0) = v_0. \quad (2)$$

Is this coherent? Some have denied that it is. Bruno de Finetti famously claimed that second-order probabilities are devoid of meaning, whereas David Miller has argued that they lead to an absurdity [1,11]. According to Miller, if we substitute $P^1(\neg q_0)$ for v_0 in Eq. (2), we obtain

$$P^2(q_0|P^1(q_0) = P^1(\neg q_0)) = P^1(\neg q_0),$$

which is the same thing as

$$P^2(q_0|P^1(q_0) = \frac{1}{2}) = P^1(\neg q_0).$$

However, with $v_0 = \frac{1}{2}$, we see that (2) yields $P^2(q_0|P^1(q_0) = \frac{1}{2}) = \frac{1}{2}$. Therefore $P^1(\neg q_0) = \frac{1}{2}$, and thus $P^1(q_0) = \frac{1}{2}$. So if Eq. (2) were unrestrictedly valid, we could prove that the probability of an arbitrary proposition q_0 is equal to one-half, which is absurd. This absurdity is known as the Miller paradox.

* Corresponding author.

E-mail address: d.atkinson@rug.nl (D. Atkinson).

Brian Skyrms pointed out that the Miller paradox “rests on a simple *de re-de dicto* confusion” [18, p. 111]. One and the same expression is used both referentially and attributively, so that a scalar or number (here v_0) is wrongly put on a par with a random variable (here $P^1(\neg q_0)$) that takes on a range of possible values [5, pp. 399–400]. So long as we recognize this confusion and keep the two levels apart, the notion of a second-order probability is harmless.

Another objection to higher-order probabilities can be discerned in de Finetti’s work. As is well known, de Finetti holds that probability judgements are expressions of attitudes that lack truth values. However, as Skyrms has pointed out, de Finetti’s work is less hostile to a theory of higher-order probabilities than might at first appear [18, p. 113]:

“For a given person and time there must be, after all, a proposition to the effect that that person then has the degree of belief that he might evince by uttering a certain probability attribution.

De Finetti grants as much:

The situation is different of course, if we are concerned not with the assertion itself but with whether ‘someone holds or expresses such an opinion or acts according to it,’ for this is a real event or proposition. [1, p. 189]

With this, de Finetti grants the existence of propositions on which a theory of higher order personal probabilities can be built, but never follows up this possibility.”¹

It seems, then, that de Finetti was not opposed to higher-order personal probabilities in such an uncompromising way as might at first sight seem to be the case.

Skyrms is not alone in having seen that second-order probabilities need not pose any particular problem. Several others recognized that, when the relevant distinctions are taken into account, second-order probabilities can be shown to be formally consistent [19,10,2,8,3]. This is not to say that such probabilities are mandatory. As Pearl has explained, second-order probabilities, although consistent, can be dispensed with, for one can always express them by using a richer first-order probability space [12].

These findings on second-order probabilities can be easily extended to probabilities of any finite order. But what about a hierarchy of higher-order probabilities that is infinite? Is the idea of a probability of a probability *ad infinitum* just as consistent as the idea of a probability of a finite order? As far as we know, nobody has ever claimed that it is, let alone has anyone given a proof of consistency. The reason is not difficult to find. A proof of consistency should involve a demonstration that the infinite series is convergent, and it is not immediately clear how that can be done.

As early as 1738, David Hume objected to an infinite hierarchy of probabilities of probabilities, on the grounds that it would tend to zero in the end:

“Having thus found in every probability ... a new uncertainty ... and having adjusted these two together, we are oblig’d ... to add a new doubt This is a doubt ... of which ... we cannot avoid giving a decision. But this decision, ... being founded only on probability, must weaken still further our first evidence, and must itself be weaken’d by a fourth doubt of the same kind, and so on in infinitum: till at last there remain nothing of the original probability, however great we may suppose it to have been, and however small the diminution by every new uncertainty.”²

Doubts about the consistency of an infinite regress of higher-order probabilities have also been expressed in more recent times. Savage wrote that such a hierarchy is beset by “insurmountable difficulties” and his worries were anticipated by Russell ([17, p. 58]; [16, p. 385 ff]). Modern objections against infinite-order probabilities are however not always rooted in Humean worries. For example, Nicholas Rescher claims that an infinite hierarchy of probabilities is inconsistent, not because it goes to zero, but because it will forever remain indeterminate [15, pp. 36–37].

Against these objections, we shall argue in the present paper that an infinite hierarchy of probabilities can indeed be consistent. The above-mentioned difficulty of finding a consistency proof will be overcome in two stages. First we construct a structure for which the proof *can* be given; we do this by making sure that this structure is subject to a Markov condition. Then we demonstrate that this structure is a model of an abstract system of infinite-order probabilities; in other words, it is a structure that makes all the sentences of the abstract system true. In this way we show that one can calculate a definite probability value from an endless hierarchy of probability statements. *Pace* Hume, this probability is in general not zero.

It should be noted that our purpose in this paper is not to give general conditions on probability distributions over probability distributions *ad infinitum*. Spelling out the details of those conditions would be very interesting; but we shall not attempt such an ambitious project in the present paper. Rather we content ourselves with a simple Bernoulli distribution, and with single events at each order.

This paper is set up as follows. We will start, in Section 2, by describing our model for a regress of higher-order probabilities; the model in question consists in an infinite set of machines that produce biased coins. In Section 3 we set up an abstract system of equations that produces the same probabilities as those in the model of Section 2, thereby showing that

¹ Skyrms [18, pp. 113–114].

² Hume [7, Book I, Part IV, Section I]. See also Lehrer [9].

this system is consistent. We will abstain from any detailed discussion of how the abstract system should be interpreted. As we see it, different interpretations are possible: in particular the first-order probability might stand for a chance or for a credence, depending on the application one has in mind, whereas the higher-order probabilities would typically be credences that reflect degrees of uncertainty about the lower-order credences or chances. Of course, the higher the credences are, the less realistic the interpretation of the system becomes. But here we are not so much interested in practical matters; our focus is on the theoretical consistency of the system.

In Section 4 we specify a numerical example in which the various probabilities of the system of Section 3 are explicitly calculated. Section 5 covers the case in which the probabilities are not given as points (that is, as precise numbers), but are merely specified to lie within certain intervals. In Section 6 we show how to modify the coin-making scenario to mimic the equations of Section 5, thereby ensuring their consistency as well.

2. A model of infinite-order probability

In this section we describe a model for an infinite hierarchy of probability statements. The probabilities in this model are objective, but that is not the essential point. What is essential is that the structure to be described is a genuine model, which implies that two desiderata are met. First, the model must be well-defined and free from contradictions. Second, it must map into the infinite hierarchy of probabilities. As we will see, both desiderata can be satisfied by making use of what Reichenbach called 'screening off', which is nowadays known as a Markov condition. This condition allows us to give a proof of convergence for the model and thereby for the abstract system of equations that we will set up in Section 3.

Suppose there are two machines which produce 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 . In an initial experiment that serves as a calibration, an experimenter repeatedly tosses one coin from machine V_0 . We define 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 ".

What is the probability of q_0 , on the assumption that the coin comes from machine V_0 ? We shall use the symbol $P_1(q_0)$ for the probability of a head in this calibration experiment; evidently it is the conditional probability of q_0 , given q_1 :

$$P_1(q_0) \stackrel{\text{def}}{=} P(q_0|q_1) = P(\text{"this coin will land heads"}|\text{"this coin comes from machine } V_0\text{"}).$$

Clearly $P_1(q_0) = v_0$, for if the coin comes from machine V_0 , the probability of a head is indeed v_0 , for that is the bias produced by machine V_0 . Note that P_1 is not the same as P^1 . The former is a conditional probability, in this case the probability of q_0 given q_1 ; the latter is the first-order unconditional probability of q_0 .

Next, the experimenter is instructed to take many coins from *both* machines, and to mix them thoroughly in a large pile. The numbers of coins that are to be added to the pile from machines V_0 and W_0 are regulated 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 (although machine W_1 will play no role in this first iteration, it will be relevant for the second, and subsequent iterations, as will become clear below). 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 from V_1 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 .³

The experimenter takes one coin at random from her pile and she tosses it. 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 is the proposition "the coin of the supervisor comes from machine V_1 ",

then we can ask what the probability is that the experimenter's coin falls heads, given that q_2 is true. We use the symbol $P_2(q_0)$ for this probability. It is equal to the conditional probability of q_0 , given q_2 , which can be calculated from the following variation of the rule of total probability⁴:

³ For the sake of this story, we limit v_1 to be a rational number, so it makes sense to say that the number of coins to be taken from V_0 is equal to v_1 times the total number taken from V_0 and W_0 . Similarly, in the subsequent discussion, the biases should all be considered to be rational numbers. Since the rationals are dense in the reals, this is not an essential limitation.

⁴ The proof of Eq. (3) goes as follows:

$$\begin{aligned} P(q_0 \wedge q_2) &= P(q_0 \wedge q_1 \wedge q_2) + P(q_0 \wedge \neg q_1 \wedge q_2) \\ &= P(q_0|q_1 \wedge q_2)P(q_1 \wedge q_2) + P(q_0|\neg q_1 \wedge q_2)P(\neg q_1 \wedge q_2). \end{aligned}$$

On dividing both sides of this equation by $P(q_2)$ we obtain (3).

$$P_2(q_0) \stackrel{\text{def}}{=} P(q_0|q_2) = P(q_0|q_1 \wedge q_2)P(q_1|q_2) + P(q_0|\neg q_1 \wedge q_2)P(\neg q_1|q_2). \quad (3)$$

By definition, $P(q_0|q_1 \wedge q_2)$ is the probability that the experimenter's coin will fall heads, on condition that this coin has come from machine V_0 , and that the supervisor's coin has come from machine V_1 . Similarly $P(q_0|\neg q_1 \wedge q_2)$ is the probability that the experimenter's coin will fall heads, on condition that this same coin has *not* come from machine V_0 , and that the supervisor's coin has come from machine V_1 .

This series of experiments gives rise to a Markov chain. For the condition that the experimenter's coin has come from machine V_0 is already enough to ensure that the probability that this coin will fall heads is v_0 ; and that situation is not affected by the provenance of the supervisor's coin, so $P(q_0|q_1 \wedge q_2) = P(q_0|q_1) = v_0$. Likewise, the condition that the experimenter's coin has not come from machine V_0 guarantees that it has come from machine W_0 , and therefore ensures that the probability of a head is w_0 ; again, that is not affected by the provenance of the supervisor's coin, so $P(q_0|\neg q_1 \wedge q_2) = P(q_0|\neg q_1) = w_0$. In Reichenbach's locution, q_1 is said to screen off q_0 from q_2 [14, pp. 159–167]. The screening-off or Markov condition will turn out to be an essential part of our model. For as we will see later, it is by virtue of this condition that we can provide a proof of convergence. We thereby show that the model, as well as the abstract system of which it is a model, are consistent, even if the abstract system does not itself satisfy the Markov system.

The Markov condition enables us to simplify (3) as follows:

$$\begin{aligned} P_2(q_0) &\stackrel{\text{def}}{=} P(q_0|q_2) = P(q_0|q_1)P(q_1|q_2) + P(q_0|\neg q_1)P(\neg q_1|q_2) \\ &= v_0v_1 + w_0(1 - v_1). \end{aligned} \quad (4)$$

We conclude that, if the experimenter repeats the procedure of tossing a coin from her pile many times (with replacement and randomization), the resulting relative frequency of heads would be approximately equal to $P_2(q_0)$, as given by (4). 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 $v_0v_1 + w_0(1 - v_1)$ will tend to unity as the number of tosses tends to infinity. However, one should not think that $P_2(q_0)$ is merely a correction of $P_1(q_0)$. It is rather that they refer to two different experiments. In the first experiment it was certain that the experimenter took a coin from machine V_0 , whereas in the second experiment it was certain that the supervisor took a coin from machine V_1 : as a consequence it was no longer sure that the experimenter took a coin that had come from V_0 . Instead of being only a correction, $P_2(q_0)$ is the result of a longer, and more sophisticated experiment than is $P_1(q_0)$.

So much for 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 that he must add to the pile from machine V_1 is equal to v_2 multiplied by the total number of coins removed from machines V_1 and W_1 .

Let

q_3 be the proposition “this simulated coin comes from simulated machine V_2 ”.

If q_3 is true, then the probability of q_2 is equal to v_2 , that is to say $P(q_2|q_3) = v_2$. Again, screening off is essential here: q_2 screens off q_1 from q_3 . So we may write

$$\begin{aligned} P(q_1|q_3) &= P(q_1|q_2 \wedge q_3)P(q_2|q_3) + P(q_1|\neg q_2 \wedge q_3)P(\neg q_2|q_3) \\ &= P(q_1|q_2)P(q_2|q_3) + P(q_1|\neg q_2)P(\neg q_2|q_3) \\ &= v_1v_2 + w_1(1 - v_2). \end{aligned} \quad (5)$$

This value of $P(q_1|q_3)$ is handed down to the experimenter, and she reruns her experiment, but with $P(q_1|q_3)$ in place of $P(q_1|q_2)$. Since q_1 screens off q_0 from q_3 (and from all the higher q_n for that matter), we calculate

$$\begin{aligned} P_3(q_0) &\stackrel{\text{def}}{=} P(q_0|q_3) = P(q_0|q_1 \wedge q_3)P(q_1|q_3) + P(q_0|\neg q_1 \wedge q_3)P(\neg q_1|q_3) \\ &= P(q_0|q_1)P(q_1|q_3) + P(q_0|\neg q_1)P(\neg q_1|q_3) \\ &= v_0P(q_1|q_3) + w_0[1 - P(q_1|q_3)], \end{aligned} \quad (6)$$

in which we are to replace $P(q_1|q_3)$ by $v_1v_2 + w_1(1 - v_2)$, in accordance with Eq. (5). This yields

$$P_3(q_0) \stackrel{\text{def}}{=} P(q_0|q_3) = w_0 + (v_0 - w_0)w_1 + (v_0 - w_0)(v_1 - w_1)v_2. \quad (7)$$

The relative frequency of heads that the experimenter would observe will be approximately equal to $P_3(q_0)$, as given by (7) – with the usual probabilistic proviso. 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 .

This procedure must be repeated *ad infinitum*. A subprogram simulates the working of yet another duo of virtual machines, V_3 and W_3 , which simulate the production of coins with biases v_3 and w_3 , and so on. At the n th step of the iteration one finds

$$\begin{aligned}
 P_n(q_0) \stackrel{\text{def}}{=} P(q_0|q_n) &= w_0 + (v_0 - w_0)w_1 + (v_0 - w_0)(v_1 - w_1)w_2 \dots \\
 &+ (v_0 - w_0)(v_1 - w_1) \dots (v_{n-3} - w_{n-3})w_{n-2} \\
 &+ (v_0 - w_0)(v_1 - w_1) \dots (v_{n-2} - w_{n-2})v_{n-1}.
 \end{aligned} \tag{8}$$

In Appendix A it is shown that, under a weak condition on the conditional probabilities, the sequence $P_1(q_0), P_2(q_0), P_3(q_0) \dots$ converges to a limit, $P_\infty(q_0)$, that is well-defined.⁵ Moreover, under the same condition the last term in (8), namely $(v_0 - w_0)(v_1 - w_1) \dots (v_{n-2} - w_{n-2})v_{n-1}$, tends to zero as n tends to infinity, so finally

$$P_\infty(q_0) = w_0 + (v_0 - w_0)w_1 + (v_0 - w_0)(v_1 - w_1)w_2 \dots, \tag{9}$$

the infinite series being convergent.⁶

In this way we have designed a set of experiments that is well-defined in the sense that it could in principle be performed to any finite number of steps, where the successive results for the probability that the experimenter throws a head get closer and closer to a limiting value that can be calculated. To be precise, for any $\varepsilon > 0$, and for any set of conditional probabilities that satisfies the weak condition given in Appendix A, one can calculate an integer, N , such that $|P_N(q_0) - P_\infty(q_0)| < \varepsilon$, and one could actually carry out the experiments to determine $P_N(q_0)$. That is, one can get as close to the limit of the infinite regress of probabilities as one likes.

3. Abstract system of infinite-order probability

In this section we will set up an abstract system of equations for the determination of probabilities of infinite order, for which we will use the symbol P^∞ . We shall give the basic equations governing the regress for $P^\infty(q_0)$, the infinite-order unconditional probability of q_0 . At first sight, it might seem that the regress for $P^\infty(q_0)$ can never get off the ground. After all, one needs $P^\infty(q_1)$ to calculate $P^\infty(q_0)$, but to calculate $P^\infty(q_1)$ one needs $P^\infty(q_2)$, and so on. Hence the question arises as to the consistency of the final, infinite expression for $P^\infty(q_0)$. The answer to this question was however anticipated in the specification of the coin-making machines of the preceding section. In that model there was no infinitely postponed calculation. Instead we defined a sequence of probabilities, $P_1(q_0), P_2(q_0), P_3(q_0)$, and so on, each one involving a finite number of steps; and it was shown that the sequence has a limit. We have seen that the screening-off conditions were essential in this procedure, and the coin-making scenario was tailor-made to guarantee this Markov constraint. The consistency of the abstract system will be underwritten when we show that the formula for $P^\infty(q_0)$ precisely matches the final formula for the limit, $P_\infty(q_0)$, in the model of the previous section.

Let q_1 be the proposition $P^1(q_0) = v_0$, which is Eq. (1) in Section 1. In the place of Eq. (2) of Section 1 we write

$$P^\infty(q_0|q_1) = v_0. \tag{10}$$

We write P^∞ , the infinite-order probability function, rather than P^2 , because the whole infinite regress stands behind the computation of the unconditional probability, $P^\infty(q_0)$, as we shall shortly see. One cannot be sure whether q_1 is true or not; $P^\infty(q_1)$ will in general lie somewhere between 0 and 1. The rule of total probability yields

$$\begin{aligned}
 P^\infty(q_0) &= P^\infty(q_0|q_1)P^\infty(q_1) + P^\infty(q_0|\neg q_1)P^\infty(\neg q_1) \\
 &= v_0P^\infty(q_1) + w_0P^\infty(\neg q_1) \\
 &= w_0 + (v_0 - w_0)P^\infty(q_1),
 \end{aligned} \tag{11}$$

with use of Eq. (10), and where we have abbreviated the conditional probability $P^\infty(q_0|\neg q_1)$ by the symbol w_0 .

⁵ The Chapman-Kolmogorov theorem ensures convergence in the homogeneous case (in which all the v_n are equal to one another, and all the w_n are equal to one another). However, we are interested in the more general inhomogeneous situation, and for that we provide a special proof. For a good overview of Markov chains, see Gardiner [4].

⁶ Since we do not exclude the possibility that w_n could be greater than v_n , for some, or all n , it follows that the series (9) may not be one of positive terms only. One might therefore be worried that $P_\infty(q_0)$ could be zero; and this would raise the specter of an ill-defined conditional probability. A sufficient (but not a necessary) condition to rule out this possibility is by requiring that neither v_0 nor w_0 vanishes. For by analogous reasoning to the above, we find

$$P_\infty(q_1) = w_1 + (v_1 - w_1)w_2 + (v_1 - w_1)(v_2 - w_2)w_3 \dots$$

from which it follows that $P_\infty(q_0) = w_0 + (v_0 - w_0)P_\infty(q_1)$, and therefore that $P_\infty(q_0)$ lies within the interval between v_0 and w_0 . By imposing the strict inequalities $0 < v_n < 1$ and $0 < w_n < 1$, for all n , we similarly ensure that none of the unconditional probabilities, $P_\infty(q_n)$ or $P_\infty(\neg q_n)$, are zero, and therefore that all conditional probabilities are well-defined (this being a consistency requirement).

To calculate $P^\infty(q_0)$ from Eq. (11), we evidently need, beside the conditional probabilities v_0 and w_0 , the unconditional probability $P^\infty(q_1)$. This is obtained by the analogous expression

$$\begin{aligned} P^\infty(q_1) &= P^\infty(q_1|q_2)P^\infty(q_2) + P^\infty(q_1|\neg q_2)P^\infty(\neg q_2) \\ &= w_1 + (v_1 - w_1)P^\infty(q_2), \end{aligned} \quad (12)$$

where q_2 is the proposition $P^1(q_1) = v_1$, and $P^\infty(q_1|\neg q_2)$ has been designated by the symbol w_1 .

Now substitute $w_1 + (v_1 - w_1)P^\infty(q_2)$ for $P^\infty(q_1)$ in Eq. (11):

$$P^\infty(q_0) = w_0 + (v_0 - w_0)w_1 + (v_0 - w_0)(v_1 - w_1)P^\infty(q_2). \quad (13)$$

It should be clear now how to continue. At the n th step, namely

$$\begin{aligned} P^\infty(q_n) &= P^\infty(q_n|q_{n+1})P^\infty(q_{n+1}) + P^\infty(q_n|\neg q_{n+1})P^\infty(\neg q_{n+1}) \\ &= w_n + (v_n - w_n)P^\infty(q_{n+1}), \end{aligned} \quad (14)$$

we can eliminate $P^\infty(q_n)$ in favour of $P^\infty(q_{n+1})$. The generalization of Eq. (13) may be written

$$P^\infty(q_0) = \Delta_n + \Gamma_n P^\infty(q_{n+1}), \quad (15)$$

where

$$\begin{aligned} \Gamma_n &= (v_0 - w_0)(v_1 - w_1) \dots (v_n - w_n), \\ \Delta_n &= w_0 + \Gamma_0 w_1 + \Gamma_1 w_2 + \dots + \Gamma_{n-1} w_n. \end{aligned}$$

It is shown in Appendix A that, if $|v_n - w_n|$ tends less quickly to one than $1/n$ tends to zero, then Γ_n tends to zero as n tends to infinity, with the result that Eq. (15) reduces to the infinite series

$$P^\infty(q_0) = \Delta_\infty = w_0 + \Gamma_0 w_1 + \Gamma_1 w_2 + \Gamma_2 w_3 \dots \quad (16)$$

In Appendix A it is also shown that this series is convergent, so that the infinite-order probability, $P^\infty(q_0)$, has been expressed as a function of the conditional probabilities alone. This is precisely the same as the series (9) for the limit, $P_\infty(q_0)$, of the probabilities associated with the coin-making machines of the previous section. The probability function P^∞ is a function of infinite order, in the sense that it involves an unending sequence of probabilities of probabilities. The system of Section 2 provides a model of the equations of the present section, thereby showing that the latter are consistent.

At this juncture, one question might be raised. Why could we not have set up the abstract system of higher-order probabilities as a sequence of successive approximations, as we did for the coins? The answer is that, in this section, we considered an abstract system, irrespective of whether there is a Markov condition. This proved possible through the use of the unconditional infinite-order probabilities, but the question of consistency was thereby left open. Precisely by finding a model that has the Markov constraint built into its very structure were we able to show the consistency of the abstract system. A model is after all a structure that makes all sentences of a theory true: in our case this translates as a structure (the coin-making scenario in Section 2) that validates all sentences of the abstract system (the equations in the present section).

4. An example

In this section we will give a numerical example in order to explain how exactly the value of $P^1(q_0)$ differs from the value of $P^\infty(q_0)$. The example serves to show that the equations of Section 3 work out properly in one case: it illustrates, as it were, how *one* sentence of the system is true, whereas the model in Section 2 amounted to a demonstration that *all* sentences are true.

An explicit example of an infinite set of conditional probabilities that leads to a well-defined $P^\infty(q_0)$ is the following:

$$v_n = 1 - \frac{1}{n+2} + \frac{1}{n+3}, \quad w_n = \frac{1}{n+3}, \quad (17)$$

so that

$$v_n - w_n = 1 - \frac{1}{n+2} = \frac{n+1}{n+2}.$$

'Telescoping' occurs, and we find

$$\begin{aligned} \Gamma_n &= \frac{1}{2} \times \frac{2}{3} \times \dots \times \frac{n}{n+1} \times \frac{n+1}{n+2} = \frac{1}{n+2}, \\ \Gamma_{n-1} w_n &= \frac{1}{n+1} \times \frac{1}{n+3} = \frac{1}{2} \left(\frac{1}{n+1} - \frac{1}{n+3} \right), \\ \Delta_n &= \frac{1}{3} + \frac{1}{2} \left(\frac{1}{2} - \frac{1}{4} \right) + \frac{1}{2} \left(\frac{1}{3} - \frac{1}{5} \right) + \dots + \frac{1}{2} \left(\frac{1}{n+1} - \frac{1}{n+3} \right) \\ &= \frac{1}{3} + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{3} - \frac{1}{n+2} - \frac{1}{n+3} \right), \end{aligned}$$

so Eq. (15) becomes

$$P^\infty(q_0) = \frac{1}{3} + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{3} - \frac{1}{n+2} - \frac{1}{n+3} \right) + \frac{1}{n+2} P^\infty(q_{n+1}). \tag{18}$$

On letting n tend to infinity we are left with

$$P^\infty(q_0) = \frac{1}{3} + \frac{1}{2} \left(\frac{1}{2} + \frac{1}{3} \right) = \frac{3}{4}. \tag{19}$$

This is to be compared with

$$P^1(q_0) = v_0 = 1 - \frac{1}{2} + \frac{1}{3} = \frac{5}{6}. \tag{20}$$

Thus the infinite-order probability of q_0 , as given by (16), is perfectly well-defined, and it is different from the first-order probability.

5. Imprecise first-order probabilities

In the foregoing sections it was assumed that the first-order unconditional probabilities were precise numbers, resulting from conditional probabilities that were likewise precise. More realistically, however, these first-order probabilities might well be specified imprecisely as merely lying within some specified intervals (cf. Pearl [13]). After all, it might be objected that a precise specification of the conditional probabilities would be a travesty of what reasonably could be expected. For precise conditional probabilities lead to precisely determined unconditional probabilities; and such precision seems alien to the probabilistic enterprise.

In order to meet this objection we will replace the points of the previous sections by intervals. More particularly, we will change the point definitions of the q_n along the following lines. Suppose again that q_0 is some proposition, the truth value of which is in doubt; but now let

- q_1 be the proposition $P^1(q_0) \in [v_0 - \varepsilon_0, v_0 + \varepsilon_0]$
- q_2 be the proposition $P^1(q_1) \in [v_1 - \varepsilon_1, v_1 + \varepsilon_1]$
- q_3 be the proposition $P^1(q_2) \in [v_2 - \varepsilon_2, v_2 + \varepsilon_2]$,

and so on. The restrictions $\varepsilon_n < v_n < 1 - \varepsilon_n$ are imposed, thus guaranteeing that all the $P^1(q_n)$ lie strictly within the unit interval. The task is once more to determine $P^\infty(q_0)$. In the first instance we think that q_0 has a certain probability lying between $v_0 - \varepsilon_0$ and $v_0 + \varepsilon_0$; and as before we update our estimate according to the rule of total probability:

$$P^\infty(q_0) = P^\infty(q_0|q_1)P^\infty(q_1) + P^\infty(q_0|\neg q_1)P^\infty(\neg q_1).$$

In full this would read

$$\begin{aligned} P^\infty(q_0) &= P^\infty(q_0|P^1(q_0) \in [v_0 - \varepsilon_0, v_0 + \varepsilon_0])P^\infty(P^1(q_0) \in [v_0 - \varepsilon_0, v_0 + \varepsilon_0]) \\ &\quad + P^\infty(q_0|P^1(q_0) \notin [v_0 - \varepsilon_0, v_0 + \varepsilon_0])P^\infty(P^1(q_0) \notin [v_0 - \varepsilon_0, v_0 + \varepsilon_0]). \end{aligned}$$

In the new situation

$$P^\infty(q_0|q_1) = P^\infty(q_0|P^1(q_0) \in [v_0 - \varepsilon_0, v_0 + \varepsilon_0]) = v_0 + \eta_0 \varepsilon_0,$$

where η_0 is some undetermined number between -1 and $+1$.

Similar considerations apply for further steps in the regress, the n th one being

$$P^\infty(q_n) = (v_n + \eta_n \varepsilon_n)P^\infty(q_{n+1}) + w_n P^\infty(\neg q_{n+1}), \tag{21}$$

where, as before, we write w_n for the conditional probability $P^\infty(q_n|-q_{n+1})$. For fixed values of the $v_n, \eta_n, \varepsilon_n$ and w_n , and on condition that both $|v_n + \varepsilon_n - w_n|$ and $|v_n - \varepsilon_n - w_n|$ tend less quickly to one than $1/n$ tends to zero, the iterative relation (21) can be solved to yield a well-defined solution for $P^\infty(q_0)$. The two extreme values within which $P^\infty(q_0)$ must lie are obtained by allowing all the η_n to vary independently between -1 and $+1$.

To get a feeling for how this works, consider the uniform situation⁷ in which the v_n, ε_n and w_n are the same for all n , so we may drop the suffix on these variables (but not on η_n , which is random). For any fixed $\eta_n \in [-1, 1]$, a convergent series like Eq. (16), but with $v + \eta_n \varepsilon$ in place of v_n , yields the following well-defined value for the infinite-order probability:

$$P^\infty(q_0) = w[1 + (v - w + \eta_0 \varepsilon) + (v - w + \eta_0 \varepsilon)(v - w + \eta_1 \varepsilon) + (v - w + \eta_0 \varepsilon)(v - w + \eta_1 \varepsilon)(v - w + \eta_2 \varepsilon) + \dots].$$

If $0 \leq v - w + \varepsilon < 1$, we find that this probability must satisfy

$$P^\infty(q_0) \leq w[1 + (v - w + \varepsilon) + (v - w + \varepsilon)(v - w + \varepsilon) + \dots] = \frac{w}{1 - v + w - \varepsilon}. \tag{22}$$

If $0 \leq v - w - \varepsilon < 1$, we have further

$$P^\infty(q_0) \geq w[1 + (v - w - \varepsilon) + (v - w - \varepsilon)(v - w - \varepsilon) + \dots] = \frac{w}{1 - v + w + \varepsilon}. \tag{23}$$

So, under the restriction $\varepsilon \leq v - w < 1 - \varepsilon$, we find

$$\frac{w}{1 - v + w + \varepsilon} \leq P^\infty(q_0) \leq \frac{w}{1 - v + w - \varepsilon}. \tag{24}$$

In this way a constraint on the conditional probabilities has led to a constraint on the infinite-order probability, $P^\infty(q_0)$. If the above restriction is not satisfied, and in the more general, nonuniform case that the v_n, ε_n and w_n are not the same for all n , the calculation of the bounds on $P^\infty(q_0)$ is more laborious, but it can still be carried out.

We propose in the following section to extend our model of the coin-making machines to cover this case of imprecisely specified conditional probabilities, the purpose being to demonstrate the consistency of the equations of this section.

6. A model for imprecise first-order probabilities

Suppose once more that there are two machines, each of which produces trick coins. Machine V_0 now makes coins that have variable bias, in such a way that any bias between $v_0 - \varepsilon_0$ and $v_0 + \varepsilon_0$ may be produced. Machine W_0 still makes coins each of which has bias w_0 , where however w_0 does *not* lie in the interval $[v_0 - \varepsilon_0, v_0 + \varepsilon_0]$. Our experimenter takes many coins from each machine, as before, and mixes them. She picks one coin at random from the heap and tosses it. As before, q_0 and q_1 are formally defined as in Section 2:

- q_0 is the proposition “this coin will land heads”
- q_1 is the proposition “this coin comes from machine V_0 ”;

but, since machine V_0 has now different properties,

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

will now be equivalent to $P_1(q_0) \stackrel{\text{def}}{=} P(q_0 | P(q_0) \in [v_0 - \varepsilon_0, v_0 + \varepsilon_0]) = v_0 + \eta_0 \varepsilon_0$, with η_0 an undetermined number in the interval $[-1, 1]$.

It should be clear enough now how to proceed in the construction of a model that matches the equations of the previous section. Our experimenter receives her instructions about the relative numbers of coins to take from machines V_0 and W_0 from her supervisor on the basis of a coin that he takes from machine V_1 . This leads to

$$P_2(q_0) \stackrel{\text{def}}{=} P(q_0 | q_2) = (v_0 + \eta_0 \varepsilon_0)P(q_1 | q_2) + w_0 P(\neg q_1 | q_2),$$

where q_2 is once more the proposition

- q_2 is the proposition “the coin of the supervisor comes from machine V_1 ”;

and where again use has been made of the fact that q_1 screens off q_0 from q_2 .

The AI prints out relative numbers of coins for the supervisor from simulated machines that produce biases between $v_2 - \varepsilon_2$ and $v_2 + \varepsilon_2$, and w_2 , respectively, and so on and so forth. As in the previous section, the whole regress has to be repeated many times: each run corresponds to the same values of v_n, w_n and ε_n , but the stochastic variables η_n will be

⁷ In the Markov chain language, this would be called a homogeneous chain – see footnote 4.

different for each run. The extreme values correspond to the lower and upper bounds on the values of $P_n(q_0)$ that would be obtained by our experimenter, were she to repeat the whole set of experiments many times (with the same values of v_n , ε_n and w_n , but not η_n , over which she has no control).

We have thereby shown the consistency of the equations of Section 5, since the machines that we have just described exactly mimic them.

7. Conclusion

Although several philosophers have argued that second-order probabilities are consistent, serious doubts have been raised concerning probabilities of infinite order. At first sight it would seem that the idea of a probability of a probability, *ad infinitum*, is simply incoherent – not merely from a practical, but also from a theoretical point of view. The reason for this pessimism is not difficult to discern. For how could such an infinite-order probability ever get off the ground? We need a proof of convergence in order to show that an infinite sequence of probabilities of probabilities is consistent: in the absence of such a proof, scepticism is warranted.

In the present paper we have tackled this difficulty by clearing two hurdles. First we have described a scenario in which, thanks to the Markov condition, convergence could be demonstrated. Second we showed that this scenario is a genuine model for an abstract system of infinite-order probabilities, in the sense that it is a structure that makes the sentences of this system true. The abstract system itself does not need to satisfy a Markov condition, it is enough that the model, which does satisfy the condition, reproduces the final expression for the infinite-order probability.

The demonstrations that we have given apply in the first place to the standard situation in which the probabilities in question are given as specific numbers. However, we also sketched how to extend these demonstrations to the case in which the probabilities are only specified to lie in certain intervals.

Acknowledgements

We would like to thank three anonymous referees for their helpful comments.

Appendix A

Suppose that $v_n \in (0, 1)$ and $w_n \in (0, 1)$, for all integers n , and further suppose that for some number $A > 1$, and some integer $N > A$,

$$1 - |v_n - w_n| \geq \frac{A}{n} \tag{25}$$

for all $n > N$. So either $|v_n - w_n|$ remains less than some number less than one, or it tends to one, but less quickly than $\frac{1}{n}$ tends to zero, as n tends to infinity. We may rewrite (25) in the form

$$|v_n - w_n| \leq 1 - \frac{A}{n}. \tag{26}$$

This constraint is very weak: it only excludes cases of what we call quasi-bi-implication.⁸

In terms of the quantities

$$\Gamma_n = (v_0 - w_0)(v_1 - w_1) \dots (v_n - w_n),$$

$$\Delta_n = w_0 + \Gamma_0 w_1 + \Gamma_1 w_2 \dots + \Gamma_{n-1} w_n,$$

we shall prove

$$(i) \quad \Gamma_\infty = \prod_{n=0}^{\infty} (v_n - w_n) = 0,$$

$$(ii) \quad \Delta_\infty = \sum_{n=0}^{\infty} \Gamma_{n-1} w_n < \infty,$$

with the understanding that Γ_{-1} is to be set equal to 1.

⁸ If $|v_n - w_n|$ were equal to 1, then we would have either $v_n = 1$ and $w_n = 0$, or $v_n = 0$ and $w_n = 1$. In the former case the relation between q_n and q_{n+1} in Eq. (14) would be one of bi-implication; in the latter case the relation between q_n and $\neg q_{n+1}$ would be one of bi-implication. The constraint (26) excludes situations in which the relation between q_n and q_{n+1} approaches that of bi-implication very rapidly as n tends to infinity.

Proof of (i). From inequality (26) we see that

$$-\log |v_n - w_n| \geq -\log \left(1 - \frac{A}{n}\right) = \frac{A}{n} + \frac{1}{2} \left(\frac{A}{n}\right)^2 + \frac{1}{3} \left(\frac{A}{n}\right)^3 + \dots > \frac{A}{n}, \quad (27)$$

for all $n > N > A$. Therefore, for any integer M greater than N ,

$$\begin{aligned} -\log \prod_{n=N}^M |v_n - w_n| &= -\sum_{n=N}^M \log |v_n - w_n| \\ &> A \sum_{n=N}^M \frac{1}{n}, \end{aligned}$$

where (27) was used to obtain the last line. Since the last sum tends to $+\infty$ as M tends to infinity, it follows that $\log \prod_{n=N}^{\infty} |v_n - w_n| = -\infty$, and hence that $\prod_{n=N}^{\infty} |v_n - w_n| = 0$. This implies that $\Gamma_{\infty} = \prod_{n=0}^{\infty} (v_n - w_n) = 0$. Note that the restriction $A > 1$ is not needed for this part of the proof.

Proof of (ii). We shall show that the series Δ_{∞} is absolutely convergent, i.e., that the series $\sum_{n=0}^{\infty} |\Gamma_{n-1} w_n|$ is convergent. Since $|w_n| \leq 1$, it follows that $\sum_{n=0}^{\infty} |\Gamma_{n-1} w_n| \leq \sum_{n=0}^{\infty} |\Gamma_{n-1}|$, so it will suffice to show that the last series is convergent. The ratio of successive terms in this last series satisfies

$$\frac{|\Gamma_n|}{|\Gamma_{n-1}|} = |v_n - w_n| \leq 1 - \frac{A}{n},$$

for all $n > N > A > 1$, so, by Raabe's test,⁹ $\sum_{n=0}^{\infty} |\Gamma_{n-1}|$ is convergent, thereby demonstrating the absolute convergence of $\sum_{n=0}^{\infty} \Gamma_{n-1} w_n$, that is the finitude of Δ_{∞} .

References

- [1] Bruno de Finetti, Probabilities of probabilities: A real problem or a misunderstanding? in: F.A. Aykaç, C. Brumat (Eds.), *New Developments in the Applications of Bayesian Methods*, North-Holland, Amsterdam, 1977, pp. 1–10.
- [2] Zoltan Domotor, Higher order probabilities, *Philosophical Studies* 40 (1981) 31–46.
- [3] Haim Gaifmann, A theory of higher order probabilities, in: B. Skyrms, W.L. Harper (Eds.), *Causation, Chance and Credence*, Kluwer, London, 1988, pp. 191–219.
- [4] Crispin Gardiner, *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*, Springer Series in Synergetics, vol. 13, Springer, Berlin, 1983.
- [5] Colin Howson, Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, second edition, Open Court Publishing, Peru, IL, 1993, pp. 399–400.
- [6] James M. Hyslop, *Infinite Series*, fifth edition, Oliver and Boyd, Edinburgh, 1942/1954.
- [7] David Hume, *A Treatise of Human Nature*, vol. 1, J.M. Dent, London, 1738/1961.
- [8] Henry E. Kyburg, Higher order probabilities and intervals, *International Journal of Approximate Reasoning* 2 (1988) 195–209.
- [9] K. Lehrer, The evaluation of method: A hierarchy of probabilities among probabilities, *Grazer Philosophische Studien* 12–13 (1981) 131–141.
- [10] D. Lewis, A subjectivist's guide to objective chance, in: R.C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*, vol. 2, University of California Press, Berkeley, 1980, pp. 263–293.
- [11] David Miller, A paradox of information, *British Journal for the Philosophy of Science* 17 (1966) 59–61.
- [12] Judea Pearl, Do we need higher order probabilities and, if so, what do they mean? in: *Proceedings of the AAAI Workshop on Uncertainty in AI*, Seattle, Washington, 1987, pp. 170–179.
- [13] Judea Pearl, On probability intervals, *International Journal of Approximate Reasoning* 2 (1988) 211–216.
- [14] Hans Reichenbach, *The Direction of Time*, University of California Press, Berkeley, 1956.
- [15] Nicholas Rescher, *Infinite Regress. The Theory and History of Varieties of Change*, Transaction, New Brunswick, NJ, 2010.
- [16] Bertrand Russell, *Human Knowledge. Its Scope and Limits*, George Allen and Unwin, London, 1948/1966.
- [17] Leonard J. Savage, *The Foundations of Statistics*, second revised edition, Dover, New York, 1954/1972.
- [18] Brian Skyrms, Higher order degrees of belief, in: D.H. Mellor (Ed.), *Prospects for Pragmatism: Essays in Memory of F.P. Ramsey*, Cambridge University Press, Cambridge, 1980, pp. 109–137.
- [19] Soshichi Uchii, Higher order probabilities and coherence, *Philosophy of Science* 40 (3) (1973) 373–381.

⁹ See Hyslop [6, p. 48].