



Probabilistic truth approximation and fixed points

David Atkinson¹ · Jeanne Peijnenburg¹ 

Received: 1 September 2020 / Accepted: 23 November 2020 / Published online: 19 December 2020
© The Author(s) 2020

Abstract

We use the method of fixed points to describe a form of probabilistic truth approximation which we illustrate by means of three examples. We then contrast this form of probabilistic truth approximation with another, more familiar kind, where no fixed points are used. In probabilistic truth approximation with fixed points the events are dependent on one another, but in the second kind they are independent. The first form exhibits a phenomenon that we call ‘fading origins’, the second one is subject to a phenomenon known as ‘the washing out of the prior’. We explain that the two phenomena may seem very similar, but are in fact quite different.

Keywords Probability · Truth approximation · Fixed points · Genetics

1 Introduction

We shall consider two kinds of systems in which one can say that there is an ‘approach to the probabilistic truth’.¹ In the first kind, one event is made more likely by another event, which in turn is made more likely by still another event, and so on. In the second kind, which is more familiar, the successive events are not made more likely by their precursors; on the contrary, they are independent of one another. While the first kind relies on the method of fixed points, the second one does not.

The prime example of the first kind is a biological population in which the relative number of individuals with some property (like having blue eyes or hammer toes or an aptitude for mathematics) approaches a fixed value; this value corresponds to a steady state of affairs or a stable ratio which we call *the probabilistic truth* for that system. The prime example of the second kind is the simple tossing of a coin in which the probability of a head is unknown. As more and more tosses are made, the relative

¹ We shall sometimes speak of ‘approach to the truth’, but we always mean ‘approach to the probabilistic truth’.

✉ Jeanne Peijnenburg
jeanne.peijnenburg@rug.nl

¹ University of Groningen, Groningen, The Netherlands

frequency of heads approaches a particular number (except on a set of measure zero), which is the probabilistic truth for *this* system.

The example about the biological population exhibits a phenomenon that we here dub ‘fading origins’: the probabilistic truth does not depend on whether the primal ancestor or ancestors have or lack the property in question.² The coin tossing scenario is subject to what Bayesians call the ‘washing out of the prior’: the probabilistic truth does not depend on the value of the prior. Although these two effects, fading origins and the washing out of the prior, seem superficially to be similar, they are in fact very different, as we shall explain.

We present probabilistic truth approximation with fixed points in Sects. 2, 3 and 4 by giving three examples. These examples are taken from genetics and they are of increasing complexity. The simplest example has a one-dimensional structure (Sect. 2), the next example is two-dimensional (Sect. 3), and the most complicated one is three-dimensional (Sect. 4). While the first example may appear rather straightforward to readers familiar with fixed points, the second and the third example are surprising in view of their connection with two acclaimed scientific breakthroughs, namely the Mandelbrot fractal and the celebrated Hardy–Weinberg equilibrium.

Section 5 is devoted to the more familiar way of approaching the truth, as exemplified in the coin tossing example. Here we show how Bayes’ theorem leads to the basic formula of Carnap’s λ -system. In Sect. 6 we explain the difference between fading origins and the washing out of the prior. Technical details have been relegated to four appendices.

2 A mitochondrial trait

Suppose we want to know how likely it is that Mary has a particular trait, T . The trait in question is mitochondrial: it is inherited along the female line. We know it to be more likely that Mary has T if her mother had the trait than if her mother lacked it. Similarly, it is more likely that Mary’s mother has T if her mother had it, and so on. Here the events (having or lacking T) are clearly dependent upon one another.

Consider the following hypotheses:

h_0 : Mary has T

h_{-1} : Mary’s mother has T

h_{-2} : Mary’s (maternal) grandmother has T

and so on, for all negative integers $-3, -4$, et cetera. These hypotheses can be represented in a one-dimensional chain, as in Fig. 1, where the solid lines indicate that a hypothesis left of a line is made more likely by the one right of the line. Thus h_0 is made more likely by h_{-1} , which is made more likely by h_{-2} , and so on.

² Elsewhere we have talked about ‘fading foundations’ (Atkinson and Peijnenburg 2017), but here we follow the advice of an anonymous reviewer according to whom ‘origins’ is better suited to our biological and genetic examples.



Fig. 1 One-dimensional chain

We assume that empirical research has provided us with the numerical values of the conditional probabilities that satisfy the following inequalities:

$$\begin{aligned}
 P(h_0|h_{-1}) &> P(h_0|\neg h_{-1}) \\
 P(h_{-1}|h_{-2}) &> P(h_{-1}|\neg h_{-2}) \\
 P(h_{-2}|h_{-3}) &> P(h_{-2}|\neg h_{-3}) \\
 &\dots\dots
 \end{aligned}$$

We further assume that the conditional probabilities are uniform. That is, all the conditional probabilities on the left of the inequality signs are numerically identical, and the same goes for all the conditional probabilities on the right of the signs.³ The uniformity assumption enables us to adopt the succinct notation

$$\alpha = P(h_i|h_{i-1}) \qquad \beta = P(h_i|\neg h_{i-1}), \tag{1}$$

for $i = 0, -1, -2, \dots$

It is a theorem of the probability calculus that:

$$P(h_0) = \alpha P(h_{-1}) + \beta(1 - P(h_{-1})). \tag{2}$$

What is the value of $P(h_0)$, the probability that Mary has T ? Eq. (2) does not tell us. For though we have assumed that empirical research has provided the conditional probabilities, α and β , we still need to know the value of $P(h_{-1})$ in order to use (2) for the calculation of $P(h_0)$. We could simply guess the value of $P(h_{-1})$, but that of course is not very satisfactory. A better idea might be to eliminate $P(h_{-1})$ by using the theorem

$$P(h_{-1}) = \alpha P(h_{-2}) + \beta(1 - P(h_{-2})), \tag{3}$$

and replacing $P(h_{-1})$ in (2) by the right hand side of (3). We thereby obtain:

$$P(h_0) = \beta + (\alpha - \beta)\beta + (\alpha - \beta)^2 P(h_{-2}),$$

from which the unconditional probability $P(h_{-1})$ has disappeared. However, this manoeuvre does not seem to help much. True, we got rid of the unknown $P(h_{-1})$, but we are now saddled with another unknown unconditional probability, viz. $P(h_{-2})$.

³ This uniformity assumption is a natural one for biological systems in which the external conditions are steady from generation to generation. When these conditions change appreciably with time, uniformity of the conditional probabilities will no longer be true and our fixed point method is then inapplicable. There exists however an alternative method, which does not require uniformity but also works with nonuniform conditional probabilities, and which gives analogous results. We have used this alternative method elsewhere in another context (Atkinson and Peijnenburg 2017).

We could try to remove $P(h_{-2})$ by essentially repeating the procedure, only to find that we are then stuck with still another unknown, $P(h_{-3})$. And so on. With each replacement, an unknown unconditional probability disappears and a new one comes in. We seem to have made no progress at all.

But appearances are deceptive: we *did* make headway. Every time that an unconditional probability disappears and another one enters, our estimation of $P(h_0)$ has improved. This is because the greater the distance between h_0 and h_{-n} , the smaller is the impact of $P(h_{-n})$ on our estimate of $P(h_0)$. If n is very large, the contribution of $P(h_{-n})$ to $P(h_0)$ is very small, and if n goes to infinity, the contribution peters out completely. This is shown in detail in “Appendix A”, but the gist can be explained as follows.

Whatever the value is of the unknown $P(h_{-n})$, it turns out that

$$\frac{\beta}{1 - \alpha + \beta} - (\alpha - \beta)^n \leq P(h_0) \leq \frac{\beta}{1 - \alpha + \beta} + (\alpha - \beta)^n.$$

For small values of n , corresponding to only a few eliminations of the unconditional probabilities, $(\alpha - \beta)^n$ may not be small, so the above inequality will be lax; but since $\alpha - \beta$ is positive and less than one, $(\alpha - \beta)^n$ will be tiny if n is very large. As a result the probability that Mary has the trait in question, $P(h_0)$, will be squeezed more and more around the value of $\frac{\beta}{1 - \alpha + \beta}$ as n becomes larger. In the limit of infinite n ,

$$P(h_0) = \frac{\beta}{1 - \alpha + \beta} \equiv p_*.$$

Here p_* is called a fixed point, and its value is fully determined by the conditional probabilities α and β . It is only when n is infinite that the values of $P(h_0)$ and p_* coincide. With finite n , the value of $P(h_0)$ typically differs from that of p_* . For in that case our calculation of $P(h_0)$ will depend not only on α and β , but *also* on an unconditional probability $P(h_{-n})$, which represents the probability that a female ancestor in the n th generation has T .

In “Appendix A” we prove that the fixed point p_* is not repelling but *attracting*. This means that the larger n is, and thus the longer the chain of Mary’s ancestors, the closer our estimate of $P(h_0)$ will be to p_* . In this sense we are dealing with a form of *probabilistic truth approximation*: as we go further and further back into Mary’s ancestry, the chain of conditional probabilities increases, and the probability that Mary has the trait in question, $P(h_0)$, approaches a final value, namely p_* , the probabilistic truth of the system in question.

The idea that an attracting fixed point is a form of probabilistic truth approximation can be made more intuitive as follows. Imagine three medical doctors, each of whom tries to estimate the probability that Mary has T . All three are aware of the fact that T is mitochondrial, but since Mary is a foundling whose parents are unknown, the doctors have no information about Mary’s ancestors. In particular they do not know whether any of her female progenitors had T . They agree that if one is totally ignorant about whether a woman has T or not, one should set the probability that this woman has T equal to one half. The first doctor now calculates the probability that Mary has

T , $P(h_0)$, by taking into account only the probability that Mary's mother has T , that is $P(h_{-1}) = \frac{1}{2}$. The second doctor, however, also takes into consideration the probability pertaining to Mary's grandmother, $P(h_{-2}) = \frac{1}{2}$. The third one decides to go back to Mary's great-grandmother, and performs his calculations on the basis of $P(h_{-3}) = \frac{1}{2}$.

Each of the three doctors will now come up with a different estimate for $P(h_0)$. The estimation of the second doctor will be better than that of the first, but the estimation of the third one will be the best. For that estimation will be the closest to a definite value, the fixed point p_* , which would be the result of an imaginary doctor who went back an infinite number of generations. The calculation of this imaginary doctor is free of any guessing or any ignorance. It is based solely on the known values of the conditional probabilities α and β , and for that reason the outcome of this calculation can be called the probabilistic truth for this system.

Of course in practice no doctor can go back infinitely far. In determining the probability that Mary has T , a doctor will normally stop after two or three steps, at the unconditional probability that Mary's grandmother or great-grandmother has the trait, and for all practical purposes this is enough.

This means that for finite n , the estimate of $P(h_0)$ will in general differ from p_* . There is however a natural way to measure the distance between the estimate of $P(h_0)$ based on finite n and the value of the probabilistic truth p_* . Let $P^{(n)}(h_0)$ be the estimate of $P(h_0)$ based on going back n steps in Mary's ancestry. The distance to the probabilistic truth is then simply the absolute difference between $P^{(n)}(h_0)$ and p_* :

$$|P^{(n)}(h_0) - p_*|,$$

and this difference can never be greater than $(\alpha - \beta)^n$. Thus the so-called 'logical problem' of truthlikeness or of truth approximation is handled here simply in terms of the absolute difference between a current value and the true value.⁴

Our method with fixed points not only implies that $P(h_0)$ gets closer and closer to the probabilistic truth as the number of generations increases. It also entails that the incremental change in $P(h_0)$ brought about by going one step further into Mary's ancestry is a decreasing function of the length of the chain. For example, if we add to a chain with ten ancestors of Mary an eleventh one, $P(h_0)$ will be changed by a certain amount. If we now add another ancestor, thus bringing the tally to twelve, the new change in $P(h_0)$ will be less than it was when the eleventh ancestor was included in the calculation.⁵ This tendency goes on: every time we add an ancestor, the change in $P(h_0)$ is less. Here we call this effect *fading origins*: the influence of Mary's maternal ancestors on the probability that she has T diminishes as the ancestors are further and further away. In the limit this influence disappears completely.⁶

⁴ In Sects. 3–5, we will tackle the logical problem in the same way. As to the epistemic problem of truthlikeness, that is dealt with by identifying the probabilistic truth and explaining how it is approached by estimates based on empirical evidence. Thanks to an anonymous reviewer for pressing us to make this explicit.

⁵ A proof of this statement can be found at the end of "Appendix A".

⁶ Atkinson and Peijnenburg (2017). There we call the effect 'fading foundations', and h_0 was labelled A_0 whereas h_{-i} was labelled A_i .

So far we have assumed that the conditional probabilities α and β have been measured with perfect accuracy. This assumption is however not very realistic. Nor is it necessary: our argument also works if α and β are subject to error. Suppose that empirical studies indicate that the values of α and β lie in the following intervals:

$$\alpha^{\min} \leq \alpha \leq \alpha^{\max} \qquad \beta^{\min} \leq \beta \leq \beta^{\max}. \quad (4)$$

The fixed point p_* is an increasing function of both α and β .⁷ Therefore the minimum and maximum values of p_* are

$$p_*^{\min} = \frac{\beta^{\min}}{1 - \alpha^{\min} + \beta^{\min}} \qquad p_*^{\max} = \frac{\beta^{\max}}{1 - \alpha^{\max} + \beta^{\max}}.$$

If the error bounds in (4) correspond to two standard deviations, we can say that it is 95% probable that the probabilistic truth lies between these two values: $p_*^{\min} \leq p_* \leq p_*^{\max}$.

Note that in the above method with fixed points, truth approximation is not a separate epistemic utility associated with hypotheses. It is not something that we take into account beside the hypotheses' probabilities. Rather it is a *function* of the probabilities. More particularly, it is a function of the conditional probabilities: the more conditional probabilities we have been able to estimate empirically, the better will be the approach to the truth. As has often been remarked, it is not rational to believe a proposition *merely* because it has high probability. This is because the probability value of a proposition might change under further investigations and discoveries—it might go up or down, both are possible. The form of truth approximation presented here, however, does not go up or down as the chain lengthens. On the contrary; although it is a function of the probabilities alone, truth approximation via fixed points always improves as the chain gets longer.

In the present section we have restricted ourselves to one-dimensional chains. The same points can however be made for multi-dimensional networks: the greater and more complex the network is, the more precise the value of $P(h_0)$ will be. In the limit that the network spreads out to infinity, $P(h_0)$ will tend to its real or final value. We discuss two-dimensional networks in Sect. 3 and three-dimensional ones in Sect. 4. Interestingly, as we will see, the two-dimensional networks will take us to the famous Mandelbrot fractal while the three-dimensional ones resemble the structure of the Hardy–Weinberg equilibrium.

3 Inheritance in two dimensions

In this section we extend the one-dimensional system to a structure with two dimensions: rather than considering a trait that can be inherited only from the mother, we examine one that can be inherited from the mother as well as the father. We will see that nevertheless the same argument applies and the same conclusion follows.

⁷ The reason is that $\frac{\partial}{\partial \alpha} \frac{\beta}{1 - \alpha + \beta} = \frac{\beta}{(1 - \alpha + \beta)^2}$ and $\frac{\partial}{\partial \beta} \frac{\beta}{1 - \alpha + \beta} = \frac{1 - \alpha}{(1 - \alpha + \beta)^2}$ are both positive.

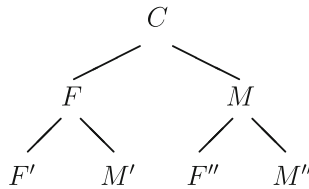


Fig. 2 Two-dimensional network

Like the one-dimensional system, the two-dimensional structure exhibits an attracting fixed point which is determined exclusively by known conditional probabilities, and therefore can be regarded as the probabilistic truth of the system.

Consider the tree structure of Fig. 2. Here C stands for some child in a particular generation, F and M for the child’s father and mother, and F', M', F'' and M'' for the four grandparents. Since each node arises from two forebears, the structure is two-dimensional. The child stems from two parents, so the probability that she has T is a function of the characteristics of her mother and her father. Rather than two reference classes (the mother having or not having T), we now have four reference classes, and four conditional probabilities:

$$\begin{aligned} \alpha &= P(\text{child has } T \mid \text{both parents have } T) \\ \beta &= P(\text{child has } T \mid \text{both parents lack } T) \\ \gamma &= P(\text{child has } T \mid \text{the father has, and the mother lacks } T) \\ \delta &= P(\text{child has } T \mid \text{the mother has, and the father lacks } T). \end{aligned}$$

As before, the conditional probabilities are supposed to be known through empirical research. We again assume them to be uniform, that is the same from generation to generation.

Consider the following propositions:

- c : A child in generation $i + 1$ has T .
- m : The child’s mother in generation i has T .
- f : The child’s father in generation i has T ,

where $i = -1, -2, -3$, and so on. Then the probability calculus prescribes the following relation between the unconditional probability that the child has T , and the unconditional probabilities relating to its parents:

$$P(c) = \alpha P(m \wedge f) + \beta P(\neg m \wedge \neg f) + \gamma P(\neg m \wedge f) + \delta P(m \wedge \neg f). \tag{5}$$

We further simplify matters by making two assumptions (which are just made for convenience, they are not essential for our argument):

- the probabilities that mother and father have T are the same;
- with respect to T mother and father are independent.

Note that the assumed independence of mother and father with respect to T has to do with unconditional probabilities in one and the same generation. Between generations

there is still dependence of course, just as there was in the one-dimensional chain of the previous section.

Accordingly $P(m \wedge f) = (P(h_i))^2$, where $P(h_i)$ stands for the unconditional probability that an individual in the i th generation has T . Similar considerations apply to $P(\neg m \wedge \neg f)$, $P(\neg m \wedge f)$ and $P(m \wedge \neg f)$, so (5) becomes

$$P(h_{i+1}) = \alpha(P(h_i))^2 + \beta(1 - P(h_i))^2 + (\gamma + \delta)P(h_i)(1 - P(h_i)), \quad (6)$$

where $i = -1, -2, -3, \dots$ et cetera, and $P(h_{i+1})$ is the probability that an individual in generation $i + 1$ has the trait T .

Although this scenario is more complicated than the one where T is mitochondrial, it likewise provides an illustration of truth approximation with fixed points. For in the present scenario, too, the probability that a child has T tends to a definite value. Interestingly, to find this value we can transform the iteration (6) into the one that generates the celebrated Mandelbrot fractal,⁸ so we can press Mandelbrot's analysis into our service. In "Appendix B" we explain how to transform (6) into Mandelbrot's form and how to find the fixed point, p_* , which turns out to have the following expression:

$$p_* = \frac{\beta}{\beta + \frac{1}{2}(1 - \gamma - \delta) + \sqrt{\beta(1 - \alpha) + \frac{1}{4}(1 - \gamma - \delta)^2}}. \quad (7)$$

Here p_* is the probabilistic truth, approached as the number of generations increases without bound. It depends on the conditional probabilities only, as did the limit in the one-dimensional example of the previous section. In particular it in no way relies on the unconditional probabilities that the infinitely remote ancestors had T .

As in the one-dimensional case, a distance measure presents itself in a natural way. Let h_0 again be the hypothesis that a child called Mary has T . And let $P^{(n)}(h_0)$ be the probability that Mary has T , as calculated by going back n generations in Mary's family tree. Then the distance measure is given by the absolute difference

$$|P^{(n)}(h_0) - p_*|.$$

In general this difference will not be zero, but it will decrease as we follow Mary's family tree further back into the past. In the limit $n \rightarrow \infty$, that is when infinitely many forebears are taken into account, this distance shrinks to zero so that the influence of the remote ancestors fades away completely.

Again, as in the one-dimensional case, it is not a problem if the conditional probabilities α , β , γ and δ are only approximately known. For also in the two-dimensional case it is possible to replace precise values by intervals and work out the corresponding interval within which the fixed point p_* will lie, say at the 95% confidence level.

⁸ Mandelbrot (1977). See also Atkinson and Peijnenburg (2017).

4 Genetics in three dimensions

In this section we are going to look at a system in three dimensions. Once more we will argue that the fixed point method is a way of approaching the probabilistic truth. However, the three-dimensional structure is considerably richer than the systems in one or two dimensions. We will now have no less than three fixed points, and we will specify the conditions under which each of them is attracting rather than repelling.⁹

In describing a three-dimensional structure, we will leave the level of phenotypes of the previous sections, and descend to the level of the genes that are responsible for phenotypical traits like eye colour or hammer toes. Since at the genotypical level it proves more convenient to look at the ‘descendants’ of a gene rather than at its ‘ancestors’, the direction in this section will be reversed. Rather than starting with a child called Mary and going back into her past by looking at her progenitors, as we did in the one- and two-dimensional cases, we shall now concentrate on a gene and focus on how it reproduces itself in the future.

To simplify matters, we look at one gene only. It is true that most phenotypical traits are affected by many genes that act in concert. Nevertheless, some traits are determined by just one gene, and those are the ones that we will concentrate on here.

A trait might be determined by one gene that occurs in two variants or *alleles*, which we call *A* and *a*. There are three genotypes: the two homozygotes *AA* and *aa*, and the one heterozygote *Aa*. Let N_i^{AA} , N_i^{Aa} and N_i^{aa} be the number of individuals in generation *i* of genotype *AA*, *Aa*, *aa* respectively, so the total number of individuals is $N_i = N_i^{AA} + N_i^{Aa} + N_i^{aa}$. The number of alleles of type *A* in generation *i* is $N_i^A = 2N_i^{AA} + N_i^{Aa}$, and similarly the number of alleles of type *a* is $N_i^a = 2N_i^{aa} + N_i^{Aa}$. Therefore the relative numbers of alleles of types *A* and *a* in generation *i* are respectively

$$\begin{aligned} p_i &= \frac{N_i^A}{N_i^A + N_i^a} = \frac{2N_i^{AA} + N_i^{Aa}}{2N_i} \\ q_i &= \frac{N_i^a}{N_i^A + N_i^a} = \frac{2N_i^{aa} + N_i^{Aa}}{2N_i}, \end{aligned} \quad (8)$$

with $p_i + q_i = 1$. Here the relative number p_i is the probability that an arbitrarily selected allele in the *i*th generation is of type *A*. Similarly, q_i is the probability that such a random allele is of type *a*.

The probability that the father contributes *A* while the mother contributes *a* is $p_i q_i$; conversely, the probability that the father contributes *a*, and the mother *A* is $q_i p_i$. Therefore there will be $N_i p_i^2$ organisms of genotype *AA*, $N_i q_i^2$ organisms of genotype *aa*, and $2N_i p_i q_i$ organisms of genotype *Aa*.

If mating is random, with females showing no preference at all for the genotype of the male with whom they copulate, the alleles in the next generation will be randomly paired. Suppose that the average number of alleles passed on to generation *i* + 1 by

⁹ Whereas the one- and two-dimensional examples in the previous sections are variations of cases described in Atkinson and Peijnenburg (2017), the present three-dimensional system has not been considered by us before.

an organism of genotype AA is n^{AA} . For simplicity we assume that this so-called reproductive fitness of the organism (not to be confused with a fixed point) does not change from generation to generation. Since there are $N_i p_i^2$ organisms of this genotype, on average $N_i p_i^2 n^{AA}$ organisms of genotype AA will be produced in generation $i + 1$. Each of these organisms carries two A alleles, therefore the number of A alleles passed on to generation $i + 1$ is twice this number, namely $2N_i p_i^2 n^{AA}$.

Analogously, suppose that the average number of alleles passed on to generation $i + 1$ by an organism of genotype Aa is n^{Aa} . There are $2N_i p_i q_i$ organisms of this genotype, and there will on the average be $4N_i p_i q_i n^{Aa}$ alleles passed on to generation $i + 1$. However, only half of these will be A alleles, i.e. there will be only $2N_i p_i q_i n^{Aa}$ of them. The other half will be a alleles. Thus the total number of A alleles in generation $i + 1$ is $2N_i (p_i^2 n^{AA} + p_i q_i n^{Aa})$. Similarly the total number of a alleles in generation $i + 1$ is $2N_i (q_i^2 n^{aa} + p_i q_i n^{Aa})$. The total number of alleles is therefore $2N_i (p_i^2 n^{AA} + 2p_i q_i n^{Aa} + q_i^2 n^{aa})$. So the relative numbers of alleles in generation $i + 1$ of types A and a are respectively

$$\begin{aligned}
 p_{i+1} &= \frac{p_i^2 n^{AA} + p_i q_i n^{Aa}}{p_i^2 n^{AA} + 2p_i q_i n^{Aa} + q_i^2 n^{aa}} \\
 q_{i+1} &= \frac{q_i^2 n^{aa} + p_i q_i n^{Aa}}{p_i^2 n^{AA} + 2p_i q_i n^{Aa} + q_i^2 n^{aa}}.
 \end{aligned}
 \tag{9}$$

If n^{AA} , n^{Aa} and n^{aa} are all equal, that is to say if the average number of children per parent does not depend on the parent’s genotype, then from (9) we deduce that $p_{i+1} = p_i$ and $q_{i+1} = q_i$, i.e. the relative numbers of the two alleles remain the same from generation to generation.¹⁰ This was the conclusion of Hardy and Weinberg, who in 1908 independently obtained this result, now known as the Hardy–Weinberg law or equilibrium.¹¹ However, if the average number of children per parent *does* depend on the parent’s genotype, the Hardy–Weinberg equilibrium will be broken, and the relative numbers of the alleles A and a in (9) will in general change as time goes on.¹²

We turn now to such deviations from the Hardy–Weinberg equilibrium. The connections between the genotypes are shown in Fig. 3, which represents a three-dimensional network because some nodes give rise to three successors. As we explain in “Appendix C”, the first of Eq. (9) can be rewritten in the form

$$p' = p + pq \frac{p n^{AA} + (q - p) n^{Aa} - q n^{aa}}{p^2 n^{AA} + 2pq n^{Aa} + q^2 n^{aa}},
 \tag{10}$$

¹⁰ When $n^{AA} = n^{Aa} = n^{aa}$,

$$p_{i+1} = \frac{p_i^2 + p_i q_i}{p_i^2 + 2p_i q_i + q_i^2} = \frac{p_i(p_i + q_i)}{(p_i + q_i)^2} = p_i.$$

¹¹ Hardy (1908), Weinberg (1908).

¹² More details can be found in Okasha (2012).

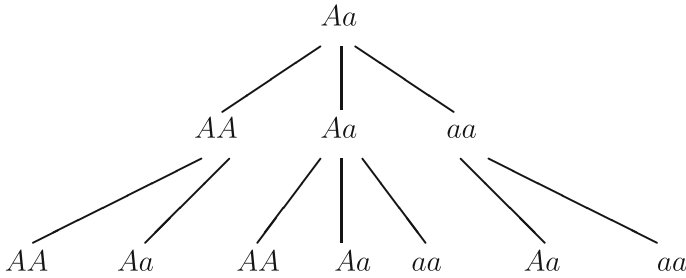


Fig. 3 Genetic transmission

where we have replaced p_i by p , q_i by q , and p_{i+1} by p' . Are there fixed points of the mapping $p \rightarrow p'$, that is, are there values of p such that $p' = p$? At such points there will be no variation from generation to generation.

When we look at (10) we see that there are three fixed points

1. $p = 0$
2. $q = 0$ which is equivalent to $p = 1$, since $q = 1 - p$
3. $p = p_*$, where $p_*n^{AA} + (q_* - p_*)n^{Aa} - q_*n^{aa} = 0$, which is equivalent to

$$p_* = \frac{n^{Aa} - n^{aa}}{2n^{Aa} - n^{AA} - n^{aa}}. \tag{11}$$

All of these fixed points are such that if p_0 , the relative number of A alleles in the first generation, is equal to one of them, then this relative number will not change from generation to generation. However, if p_0 is not equal to one of the fixed points, p_i will change as i increases. In general it will tend to one or other of the fixed points as i increases; but which fixed point it will depend on the relative sizes of the average numbers of progeny of the three genotypes.

For specified reproductive fitnesses of the genotypes only one of the three fixed points is attracting, and only this one corresponds to the probabilistic truth that will be approached. The effect of iterating (10) is to draw p' closer to this attracting fixed point. A repelling fixed point, on the other hand, is a solution of the equation $p' = p$ such that any tiny deviation from this value will result in repulsion from it towards an attracting fixed point.

As indicated, which fixed point is attracting depends on the relative sizes of n^{AA} , n^{Aa} and n^{aa} . If n^{AA} is greater than n^{Aa} and n^{aa} , then there is only one attracting fixed point, namely $p = 1$. In words, if the homozygote AA is more successful in producing progeny than the heterozygote Aa , and also more successful than the homozygote aa , then in the long run the a allele will disappear from the population. Similarly, if n^{aa} is greater than n^{Aa} and n^{AA} , then the A allele will disappear in the long run. In this case the attracting fixed point is $q = 1$, i.e. $p = 0$.

If, on the other hand, the heterozygote is more successful than either of the homozygotes, which can happen in some cases, then neither allele dies out in the long run: the attracting fixed point in this case is p_* . In Fig. 4 the axes represent the two homozygotic reproductive fitnesses, n^{AA} and n^{aa} ; and the dotted square at the bottom, where

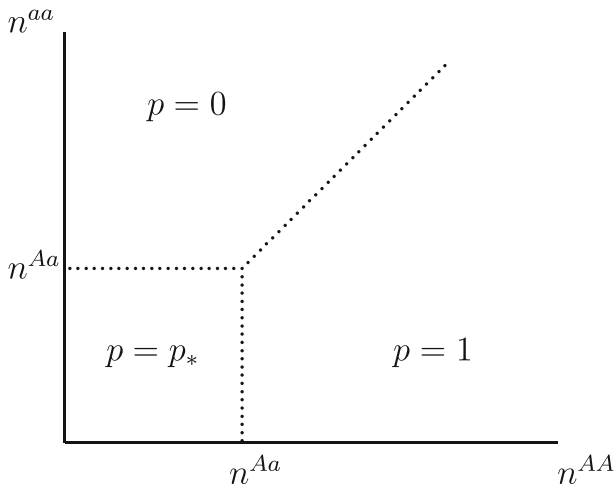


Fig. 4 Fixed points

$n^{AA} < n^{Aa}$ and $n^{aa} < n^{Aa}$, is the region in which the attracting fixed point is p_* . We demonstrate in “Appendix C” that in this domain p' is closer to the fixed point (11) than is p , which means that the iteration (10) will converge to p_* . This fixed point represents the probabilistic truth for the system, in which the relative numbers of the three zygotes (AA , Aa and aa) no longer change. In practice, in finite systems there will be small fluctuations; but the salient point is that fading origins can be discerned here at the genetic level. Whatever the original distribution of genotypes was, if the heterozygote Aa is more successful than either of the homozygotes AA or aa , the steady state after a large number of generations will be the probabilistic truth p_* .

To summarize, we have looked in this section at a three-dimensional structure, where the dimensions are represented by three genotypes: two homozygotes AA and aa , and one heterozygote Aa . The structure has three fixed points. Which of these points is attracting, and hence can function as a probabilistic truth that we gradually approach, depends on the reproductive fitnesses of the zygotes, i.e. the average number of alleles that a zygote passes on to a next generation. If one of the homozygotes has the greatest reproductive fitness, then the attracting fixed point turns out to be a trivial one: either $p = 0$ or $p = 1$. But if the heterozygote is the most successful, then the fixed point has a definite value between 0 and 1, which in turn tells us the relative number of the three zygotes in the steady state.

As in the one- and two-dimensional cases, in the three-dimensional structure, too, a distance measure emerges naturally, but we will not spell out the details here. Similarly, we will not explain in detail the fact that also in the three-dimensional structure, the fixed points can attract even if the conditional probabilities are imprecise.

5 Bayes and Carnap

Approaching the probabilistic truth by means of fixed points can be easily confused with another kind of truth approximation. The latter kind is familiar among Bayesians, but as we will show it is already implicit in Carnap (1952).

Suppose we have some evidence, e , for a hypothesis, h , and suppose we are able to calculate $P(e|h)$, the probability with which e would obtain if h were true. We are interested in the inverse probability, i.e. the probability that h is correct, given e , and this we obtain from Bayes' formula:

$$P(h|e) = \frac{P(e|h)P_0(h)}{P(e)}. \tag{12}$$

Here $P_0(h)$ is the prior probability of h , to be superseded by the posterior probability, $P(h|e)$. The denominator in (12) can be computed from the rule of total probability:

$$P(e) = P(e|h)P_0(h) + P(e|\neg h)P_0(\neg h), \tag{13}$$

on condition that the likelihood $P(e|\neg h)$ can also be calculated. More generally, if $\{h_1, h_2, \dots, h_n\}$ is a partition of the space of hypotheses, then (13) is replaced by

$$P(e) = P(e|h_1)P_0(h_1) + P(e|h_2)P_0(h_2) + \dots + P(e|h_n)P_0(h_n). \tag{14}$$

More generally still, suppose we have, rather than a discrete set $\{h_1, h_2, \dots, h_n\}$, a *continuum* of hypotheses $\{h_r\}$, one for each value of r between 0 and 1. Then in place of (14) we must use an integral:

$$P(e) = \int_0^1 P(e|h_r)P_0(h_r) dr, \tag{15}$$

in which $P_0(h_r)$ is a prior probability *density*, and $P_0(h_r) dr$ is an infinitesimal prior probability.

Here is an example. We are about to toss a coin and we want to know the probability that a head will come up.¹³ Let h_r be the hypothesis that this probability is r , and suppose e to be the evidence that n tosses with this coin resulted in m heads. The outcomes of the tosses are independent and identically distributed. The likelihood of e , given h_r , is

$$P(e|h_r) = \frac{n!}{m!(n-m)!} r^m (1-r)^{n-m}, \tag{16}$$

the factor involving the factorials being the number of different ways that m heads can turn up in n tosses.

¹³ The example is based on a binomial probability distribution, but the same sort of analysis applies to other discrete distributions.

We substitute (16) into (15) and into

$$P(h_r|e) = \frac{P(e|h_r)P_0(h_r)}{P(e)},$$

which is (12), with h replaced by h_r . This yields the posterior probability density, $P(h_r|e)$, and this will in general be non-zero for all values of r . On the one hand, this is as it should be: one single value for r is not singled out as the only possibility. On the other hand, that is not quite what we want: we are looking for the correct value of r , given e , not a whole spread of values. We therefore calculate the mean, or expected value of r , which will give the most likely value for the sought-for probability. The result is:

$$\bar{r} = \int_0^1 r P(h_r|e) dr, \quad (17)$$

which tells us that, given the prior density and the evidence, the hypothesis with the best credentials is $h_{\bar{r}}$. This hypothesis can be seen as the probabilistic truth that we approach by means of the Bayesian method, for it is the most likely value of the probability that heads will come up when the coin is tossed.

Following Pollard, Howson and Urbach inserted for the prior density $P_0(h_r)$ a so-called beta distribution:

$$P_0(h_r) = \frac{(u+v-1)!}{(u-1)!(v-1)!} r^{u-1} (1-r)^{v-1}, \quad (18)$$

where u and v are free parameters.¹⁴ As we explain in detail in “Appendix D”, (17) and (18) lead to

$$\bar{r} = \int_0^1 r P(h_r|e) dr = \frac{m+u}{n+u+v}. \quad (19)$$

The reason Howson and Urbach gave for using beta distributions was to stress the malleability of the Bayesian method, which can model any belief system:

... beta distributions take on a wide variety of shapes, depending on the values of two positive-valued parameters, u and v , enabling you to choose a beta distribution that best approximates your actual distribution of beliefs.¹⁵

This is true perhaps, but still rather arbitrary and therefore not so interesting. More interesting, in our view, is the observation that Howson and Urbach’s free parameters u and v acquire an interpretation if we derive a formula of Carnap’s λ -calculus from the Bayes result. The simple transformation $u = \frac{\lambda}{k}$ and $v = \lambda - \frac{\lambda}{k}$ does the trick, for it turns (19) into

¹⁴ Howson and Urbach (2006, p. 242), Pollard (1985, pp. 190–192).

¹⁵ Howson and Urbach, *ibid.*

$$\bar{r} = \frac{m + \frac{\lambda}{k}}{n + \lambda},$$

and the right-hand side is the basic formula of Carnap's λ -system. Here k is taken to be the number of conceivable elementary outcomes, and λ is interpreted as a measure of the speed with which we want to learn from experience.¹⁶

The phenomenon known as 'the washing out of the prior' can now be explained in terms of Carnap's formula, for

$$\bar{r} - \frac{m}{n} = \frac{\lambda}{n + \lambda} \left(\frac{1}{k} - \frac{m}{n} \right).$$

When n goes to infinity, the right-hand side clearly tends to zero, so in this limit

$$\bar{r} \sim \frac{m}{n}.$$

In other words, the dependence of \bar{r} on the prior parameters λ and k washes out as n increases indefinitely. This means that we are thrown back on the relative frequency as the best estimate of the sought-for probability. The Bayesian method has thus provided us with a simple answer to our question: the probability that the next toss with our coin will result in heads if m heads have turned up in n tosses is the limit of $\frac{m}{n}$ (except on a set of measure zero) as n tends to infinity. This limit is the probabilistic truth for the system, which we approach by means of Carnap's formula.

One might of course ask what is the point of the Bayesian treatment if the final answer is just the relative frequency of heads? The best answer seems to be a pragmatic one. If one has some believable information to motivate the choice of a prior, then this treatment can influence the choice of k and hopefully speed up the convergence to the limiting value.

6 Concluding remarks: mind the gap

In the first part of this paper (Sects. 2–4) we have been concerned with approaching the probabilistic truth in a system where the various events depend on one another. We have given three examples. The first one concerned the event of Mary having a mitochondrial trait T , which was made more likely by the event of Mary's mother having T , and so on. What typically happens in this one-dimensional chain is that the probability of Mary having T approaches a fixed point—the probabilistic truth—as the chain becomes infinitely long. The second and third examples involved networks in two and three dimensions. We showed that in these networks, too, the probability that we are interested in tends to a fixed point as the network expands without bound. The distance measure in the three examples is the absolute numerical difference between the fixed point and the precise or imprecise probability of interest, and in the limit this difference tends to zero.

¹⁶ Carnap (1952, p. 33). The Carnapian connection is also noted in Kuipers (1978, pp. 57–59), and in Festa (1992, p. 60).

In all three cases the contribution of the distant parts of the system to the fixed point diminishes and peters out in the limit. We call this phenomenon ‘fading origins’ to indicate that the influence of remote events, such as Mary’s ancestors having or lacking T , decreases as the distance between Mary and her ancestors increases. In the limit that the probability of Mary having T reaches the final value, the effect of the infinitely remote event on Mary has faded away completely.

In Sect. 5 we looked at a different sort of system, exemplified by the binomial process of tossing a coin. Here the events were independent of one another. Nevertheless there is here too a probabilistic truth which is reached in the limit of an infinite number of tosses (except on a set of measure zero). From the Bayes formula, equipped with a suitably flexible, but not too flexible prior probability, it proved possible to derive the expected value of the probability that the coin lands heads. As the number of tosses increases, so this mean value tends toward the probabilistic truth. Moreover this probabilistic truth is independent of the parameters in the specification of the Bayesian prior probability. This phenomenon goes under the epithet ‘the washing out of the prior’.

The two phenomena, fading origins and washing out, bear a superficial resemblance to one another. They are however very different and it is essential to mind the gap between them. Washing out of the prior flags the fact that in Bayesian reasoning the posterior is in the limit unaffected by the choice of prior. Fading origins flags the fact that the event we are interested in (e.g. Mary having T) is in the limit unaffected by the nature of the original event (e.g. her infinitely distant primal grandmother having T). These effects are very different. In Bayesian updating the Bayes formula involves the computation of $P(h_r|e)$ in terms of the inverse conditional probability, $P(e|h_r)$. In our example the evidence, e , grows as more tosses of the coin are made. This is quite different from our calculation of $P(h_0)$ in for example the one-dimensional case of the mitochondrial trait. There was no inversion *à la* Bayes in that case, but rather a sequence of events that follow one another in a linear chain, and as we have seen, similar considerations apply in the many-dimensional situations.

Acknowledgements We thank the editors of this Topical Collection on Approaching Probabilistic Truths, Ilkka Niiniluoto, Theo Kuipers and Gustavo Cevolani. We are also indebted to two anonymous reviewers: their comments made the paper much better.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendices

A One-dimensional iteration

Consider the iteration

$$P(h_{i+1}) = \alpha p_i + \beta(1 - P(h_i)),$$

where $i = -1, -2, -3, \dots$ and so on. The conditional probabilities α and β were defined in (1), and they satisfy $0 < \beta < \alpha < 1$. The case $i = -1$ corresponds to (2). In this appendix we show that the iteration has an attracting fixed point, and we discuss some of its features.

The above equation can be rearranged to read

$$P(h_{i+1}) = \beta + (\alpha - \beta)P(h_i). \tag{20}$$

Define

$$p_* = \frac{\beta}{1 - \alpha + \beta}.$$

If $P(h_i) = p_*$, we see from (20) that

$$\begin{aligned} P(h_{i+1}) &= \beta + (\alpha - \beta) \frac{\beta}{1 - \alpha + \beta} \\ &= \beta \left(\frac{1 - \alpha + \beta + \alpha - \beta}{1 - \alpha + \beta} \right) \\ &= \frac{\beta}{1 - \alpha + \beta} = p_*. \end{aligned}$$

The mapping $p \rightarrow p'$, where

$$p' = \beta + (\alpha - \beta) p \tag{21}$$

is said to have a fixed point at $p = p_*$. Since

$$p_* = \beta + (\alpha - \beta) p_*;$$

it follows from (20) that the distance between the probability of Mary having the trait, $P(h_0)$, and the truth p_* is:

$$\begin{aligned} |P(h_0) - p_*| &= (\alpha - \beta) |P(h_{-1}) - p_*| \\ &= (\alpha - \beta)^2 |P(h_{-2}) - p_*| \\ &= \dots \\ &= (\alpha - \beta)^n |P(h_{-n}) - p_*|, \end{aligned} \tag{22}$$

which goes to zero as n tends to infinity, *irrespective* of the value of $P(h_{-n})$, since $0 < \alpha - \beta < 1$. Therefore $P(h_0)$ tends to p_* in the limit. The relation (21) is said to be a contraction mapping.

Since $|P(h_{-n}) - p_*| \leq 1$, it follows from (22) that $|P(h_0) - p_*| \leq (\alpha - \beta)^n$, or equivalently

$$\frac{\beta}{1 - \alpha + \beta} - (\alpha - \beta)^n \leq P(h_0) \leq \frac{\beta}{1 - \alpha + \beta} + (\alpha - \beta)^n.$$

Consider next what would happen if one were to extend the chain back only n generations. Then

$$P^{(n)}(h_0) - p_* = (\alpha - \beta)^n (\hat{p} - p_*),$$

where $P^{(n)}(h_0)$ is the probability of h_0 that would be calculated if $P(h_{-n})$ were equal to some arbitrary number \hat{p} in $[0, 1]$. It follows that

$$\left| P^{(n)}(h_0) - P^{(n+1)}(h_0) \right| = (\alpha - \beta)^n (1 - \alpha + \beta) |\hat{p} - p_*|,$$

where in order to calculate $P^{(n+1)}(h_0)$ the same value \hat{p} has been assumed for $P(h_{-n-1})$. Continuing this procedure one more step, we find

$$\begin{aligned} & \left| P^{(n)}(h_0) - P^{(n+1)}(h_0) \right| - \left| P^{(n+1)}(h_0) - P^{(n+2)}(h_0) \right| \\ &= (\alpha - \beta)^n (1 - \alpha + \beta)^2 |\hat{p} - p_*| > 0. \end{aligned}$$

Thus the incremental change in $|P^{(n)}(h_0) - P^{(n+1)}(h_0)|$ is a decreasing function of n .

B Two-dimensional iteration

With p in place of p_i and p' in place of p_{i+1} , the iteration (6) turns into the mapping

$$p' = \alpha p^2 + \beta(1 - p)^2 + (\gamma + \delta)p(1 - p). \quad (23)$$

Here we will show that there are *two* fixed points, but only one of them is attracting. As we will see, the form of the iteration is isomorphic to the one that generates the well-known Mandelbrot fractal.

On condition that $\alpha + \beta \neq \gamma + \delta$, define

$$\begin{aligned} q &= (\alpha + \beta - \gamma - \delta) p - \beta + \gamma + \delta \\ q' &= (\alpha + \beta - \gamma - \delta) p' - \beta + \gamma + \delta. \end{aligned}$$

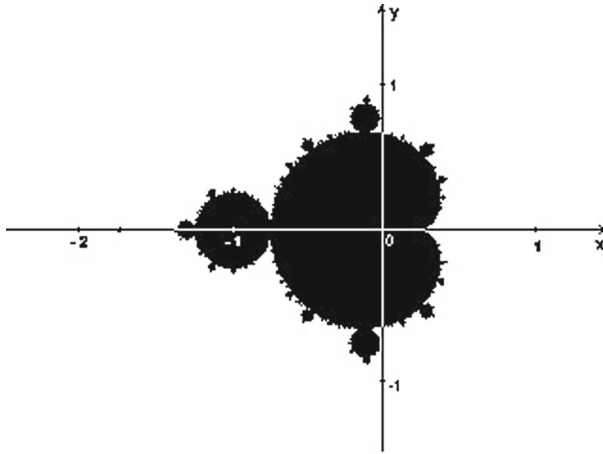


Fig. 5 The Mandelbrot fractal, associated with the complex number $c = x + iy$, is generated by the quadratic iteration $q_{n+1} = c + q_n^2$

After some algebra, one finds that (23) is equivalent to

$$q' = c + q^2, \tag{24}$$

where

$$c = \frac{1}{4}(\gamma + \delta)(2 - \gamma - \delta) - \beta(1 - \alpha). \tag{25}$$

We recognize (24) as the mapping that leads to the Mandelbrot fractal (see Fig. 5). For any point in the black region of Fig. 5, if c is the corresponding complex number, the Mandelbrot iteration does not diverge to infinity. Some of these numbers correspond to a fixed point, others to an iteration that switches back and forth between two values (a two-cycle), yet others to a three-cycle, and so on. This mapping has two fixed points. As can be verified by substitution, one of these is $\frac{1}{2} + \sqrt{\frac{1}{4} - c}$, but this is repelling, and so is of no interest, while the other is

$$q_* = \frac{c}{\frac{1}{2} + \sqrt{\frac{1}{4} - c}}, \tag{26}$$

and we will now show that this one is attracting.

We can rearrange (25) as follows:

$$c = \frac{1}{4} - \frac{1}{4}(1 - \gamma - \delta)^2 - \beta(1 - \alpha) < \frac{1}{4},$$

since $0 < \beta < \alpha < 1$. A different rearrangement is

$$c = \frac{1}{4}(\gamma + \delta)(2 - \gamma - \delta) + (\alpha - \beta)(1 - \alpha) + \left(\frac{1}{2} - \alpha\right)^2 - \frac{1}{4} > -\frac{1}{4},$$

since also $0 < \gamma + \delta < 2$. From these inequalities it follows that

$$\left|1 - \sqrt{1 - 4c}\right| < 1. \tag{27}$$

In order to find the domain in which the fixed point (26) is attracting, we define $s = q - q_*$ and $s' = q' - q_*$, so that (24) is transformed into

$$s' = s \left(1 - \sqrt{1 - 4c} + s\right). \tag{28}$$

In view of (27), for s not too large $|1 - \sqrt{1 - 4c} + s|$ will be less than 1, which proves that the fixed point is attracting. The proviso that s be not too large implies that q and q_* are close to one another, and thus that the attraction of q to q_* is true only in a so-called ‘basin of attraction’, i.e. in a restricted neighbourhood of q_* .

Going back to the original form (23) of the mapping, we find that the fixed point (26) corresponds to the attracting fixed point

$$p_* = \frac{\beta}{\beta + \frac{1}{2}(1 - \gamma - \delta) + \sqrt{\beta(1 - \alpha) + \frac{1}{4}(1 - \gamma - \delta)^2}},$$

which is (7).

C Genes

With p in place of p_i , q in place of q_i , and p' in place of p_{i+1} , we see that the first of Eq. (9) implies that

$$\begin{aligned} p' - p &= \frac{p^2 n^{AA} + pq n^{Aa}}{p^2 n^{AA} + 2pq n^{Aa} + q^2 n^{aa}} - p \\ &= \frac{p^2 n^{AA} + pq n^{Aa} - p^3 n^{AA} - 2p^2 q n^{Aa} - pq^2 n^{aa}}{p^2 n^{AA} + 2pq n^{Aa} + q^2 n^{aa}} \\ &= pq \frac{p n^{AA} + (q - p) n^{Aa} - q n^{aa}}{p^2 n^{AA} + 2pq n^{Aa} + q^2 n^{aa}}, \end{aligned} \tag{29}$$

which is equivalent to (10). This iteration has *three* fixed points, $p = 0$, $p = 1$ and a nontrivial one $p = p_*$, and we investigate the conditions under which the last one is attracting.

If $p = p_*$, where $p_* n^{AA} + (q_* - p_*) n^{Aa} - q_* n^{aa} = 0$, and $q_* = 1 - p_*$, we see from (29) that $p' = p$, i.e. p_* is a fixed point of the iteration.

We have $p_* n^{AA} + (1 - 2p_*) n^{Aa} - (1 - p_*) n^{aa} = 0$, from which it follows that

$$p_* = \frac{n^{Aa} - n^{aa}}{2n^{Aa} - n^{AA} - n^{aa}}.$$

By expanding $p' - p_*$ in powers of $(p - p_*)$, one can show that

$$p' - p_* = \lambda(p - p_*) + O((p - p_*)^2),$$

where

$$\lambda = \frac{n^{AA}(n^{Aa} - n^{aa}) + n^{aa}(n^{Aa} - n^{AA})}{(n^{Aa})^2 - n^{AA}n^{aa}},$$

and O means ‘of order’. So if $(p - p_*)$ is small, something of order $(p - p_*)^2$ is very small. If $n^{Aa} > n^{AA}$ and $n^{Aa} > n^{aa}$, it is evident that $\lambda > 0$. Moreover, one can rearrange this expression for λ as follows:

$$\lambda = 1 - \frac{(n^{Aa} - n^{aa})(n^{Aa} - n^{AA})}{(n^{Aa})^2 - n^{AA}n^{aa}},$$

so $\lambda < 1$. We conclude that, when $n^{AA} < n^{Aa}$ and $n^{aa} < n^{Aa}$, the fixed point p_* is attracting: p' is closer to the fixed point (10) than is p , which means that the iteration (8) will converge to p_* .

D Bayesian truth approximation

In this appendix we gather together some basic results concerning the Bayesian treatment of coin tossing. With the likelihood

$$P(e|h_r) = \frac{n!}{m!(n - m)!} r^m (1 - r)^{n-m},$$

and the prior

$$P_0(h_r) = \frac{(u + v - 1)!}{(u - 1)!(v - 1)!} r^{u-1} (1 - r)^{v-1},$$

the Bayes formula yields

$$\begin{aligned} P(h_r|e) &= \frac{P(e|h_r)P_0(h_r)}{\int_0^1 P(e|h_r)P_0(h_r) dr} \\ &= \frac{r^{m+u-1} (1 - r)^{n-m+v-1}}{\int_0^1 r^{m+u-1} (1 - r)^{n-m+v-1} dr} \end{aligned}$$

$$= \frac{(n + u + v - 1)!}{(m + u - 1)!(n - m + v - 1)!} r^{m+u-1} (1 - r)^{n-m+v-1}.$$

The mean value of r is

$$\begin{aligned} \bar{r} &= \int_0^1 r P(h_r|e) dr \\ &= \frac{(n + u + v - 1)!}{(m + u - 1)!(n - m + v - 1)!} \int_0^1 r^{m+u} (1 - r)^{n-m+v-1} dr \\ &= \frac{(n + u + v - 1)!}{(m + u - 1)!(n - m + v - 1)!} \frac{(m + u)!(n - m + v - 1)!}{(n + u + v)!} \\ &= \frac{m + u}{n + u + v}, \end{aligned}$$

which justifies (17).

References

- Atkinson, D., & Peijnenburg, J. (2017). *Fading foundations. Probability and the regress problem*. Dordrecht: Springer. <https://doi.org/10.1007/978-3-319-58295-5>.
- Carnap, R. (1952). *The continuum of inductive methods*. Chicago: University of Chicago Press.
- Festa, R. (1992). *Optimum inductive methods*. Ph.D. thesis of the University of Groningen.
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, 28, 49–50.
- Howson, C., & Urbach, P. (2006). *Scientific reasoning. The Bayesian approach* (3rd ed.). Chicago: Open Court.
- Kuipers, T. A. F. (1978). *Studies in inductive probability and rational expectation*. Dordrecht: D. Reidel Publishing Company.
- Mandelbrot, B. B. (1977). *The fractal geometry of nature*. Second printing with update, 1982. New York: W.H. Freeman and Co.
- Okasha, S. (2012). Population genetics. In: E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/fall2012/entries/population-genetics/>.
- Pollard, W. E. (1985). *Bayesian statistics for evaluation research: An introduction*. Beverly Hills: Sage Publications.
- Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen (On evidence concerning heredity in humans). *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, 64, 368–382.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.