1.7 Probability

It is very common to start with probabilities, and then define expectations in terms of probabilities. I have done the opposite, because I think that expectations are a better 'primitive'. I find that I can often have beliefs about a collection of random quantities $\boldsymbol{X}$ that do not involve probabilities, but which obey the axioms of Expectation.[23] It is also the case that many of the core topics in statistics, such as Decision Theory (Chapter 3), are naturally expressed in terms of expectations rather than probabilities.

[23] JCR: A later chapter, not yet written, takes this notion much further.

1.7.1 Definition

If we start with expectations, then we need to define probabilities in terms of expectations. It turns out that there is no choice in how to do this, if the resulting probabilities are to obey the Laws of Probability.

Definition 4 (Laws of Probability)**.**

1. *For any proposition* A, $\Pr(A) \geq 0$;
2. *If* A *if certain, then* $\Pr(A) = 1$;
3. *If* A *and* B *are mutually exclusive, then* $\Pr(A \vee B) = \Pr(A) + \Pr(B)$.

Theorists have a slightly stronger requirement for (3.). As it stands, (3.) can be extended to finite disjunctions of mutually-exclusive propositions, by recursion, termed *finite additivity*. But theorists require a stronger property, to account for non-finite disjunctions, termed *countable additivity*. A few people get worked up about the different between these two conditions, and one, Bruno de Finetti, was famous for rejecting countable additivity (see, e.g. de Finetti, 1972, 1974/75). Others have risen to the challenge of working within the more general but less tractable framework of finite additivity (e.g., Dubins and Savage, 1965, this is not an easy read). I doubt it matters at our level of generality, but I personally have a preference for finite additivity, when reasoning about the real world.

Here is the definition of probability in terms of expectation, which ensures that the Laws of Probability hold. Probability is defined on the domain of random propositions. The definition makes this clear.

Definition 5 (Probability, Pr)**.** *Let* $\boldsymbol{X}$ *be a set of random quantities. Let* $q(\boldsymbol{x})$ *be any sentence from first-order logic*[24]*, termed a* proposition*. Define* $Q := q(\boldsymbol{X})$*, termed a* random proposition*. Then*

$$\Pr(Q) := \mathrm{E}(\mathbb{1}_Q),$$

where $\mathbb{1}$ *is the indicator function.*[25]

[24] That is, a statement about $\boldsymbol{x}$ that evaluates to either FALSE or TRUE.

[25] That is, the function of the proposition p for which $\mathbb{1}_p = 0$ if p is FALSE, and $\mathbb{1}_p = 1$ if p is TRUE.

It is straightforward to check that complete coherence implies that probabilities defined in this way satisfy the Laws of Probability. (1.) follows by lower-boundedness, because $\mathbb{1}_A \geq 0$. (2.) follows by normalisation, because $\mathbb{1}_A = 1$ if A is certain. (3.) follows by additivity, because if A and B are mutually exclusive, then $\mathbb{1}_{A\vee B} = \mathbb{1}_A + \mathbb{1}_B$. Complete coherence is required to ensure that these pairwise properties hold for all possible propositions.

Here is the result which shows that this is the only way to define probability in terms of expectation.

Theorem 1.7. *Suppose that* $\Pr(Q) = \mathrm{E}\{g(Q)\}$, *where*

$$g : \{\mathit{FALSE}, \mathit{TRUE}\} \to \mathbb{R},$$

for some choice of g. *The only choice of* g *which is compatible with both complete coherence and the Laws of Probability is* $g(Q) := \mathbb{1}_Q$.

Proof. For complete coherence, the FTP (Thm 1.2) asserts that there is a $p \in \mathcal{P}$ such that

$$\Pr(Q) = \mathrm{E}\{g(Q)\} = \sum_{\omega\in\Omega} g(q(\boldsymbol{x}(\omega))) \cdot p(\omega)$$

for every first-order sentence $q(\boldsymbol{x})$. The Laws of Probability imply that if $q(\omega) = \mathsf{FALSE}$ for all ω then $\Pr(Q) = 0$, and if $q(\omega) = \mathsf{TRUE}$ then $\Pr(Q) = 1$. Since $p(\omega) \geq 0$ and $\sum_\omega p(\omega) = 1$, it follows that $g(\mathsf{FALSE}) = 0$ and $g(\mathsf{TRUE}) = 1$, i.e. $g(Q) = \mathbb{1}_Q$, as was to be shown. □

1.7.2 *Probability mass functions*

The definition of probability provides a straightforward interpretation of $p \in \mathcal{P}$ from the FTP (Thm 1.2).

Theorem 1.8. $\Pr(\boldsymbol{X} \doteq \boldsymbol{x}) = p\big(\omega^{-1}(\boldsymbol{x})\big)$.

Proof.

$$\begin{aligned}
\Pr(\boldsymbol{X} \doteq \boldsymbol{x}) &= \mathrm{E}(\mathbb{1}_{\boldsymbol{X}\doteq\boldsymbol{x}}) \\
&= \textstyle\sum_\omega \mathbb{1}_{\boldsymbol{x}(\omega)\doteq\boldsymbol{x}} \cdot p(\omega) && \text{by the FTP} \\
&= \textstyle\sum_\omega \mathbb{1}_{\omega\doteq\omega^{-1}(\boldsymbol{x})} \cdot p(\omega) && \text{because } \omega \mapsto \boldsymbol{x} \text{ is bijective} \\
&= p\big(\omega^{-1}(\boldsymbol{x})\big).
\end{aligned}$$

□

1.7.3 *Foundational issues*

This section tackles the profound question: *Why these Laws of Probability and not some others?* The answer to this question must involve some desire on our part to adopt exactly these Laws and no others; that is, a common agreement that probabilities which obey these Laws are sensible, and probabilities which do not obey them are not-sensible. This requires us to provide a practical definition of

probability which can be used to distinguish between sensible sets of probabilities and not-sensible ones. And then show that, according to this definition, the sensible sets of probabilities are exactly the ones which obey the Laws of Probability. Ressuringly, this is possible.

This strand of reasoning about probabilities goes back to Ramsey (1931)[26] and Savage (1954). The basic idea that $p := \Pr(Q)$ is an expression of my indifference between having £p with certainty, and owning a bet which pays £0 if Q is FALSE, and £1 if Q is TRUE. Call this the *betting interpretation* of probability.

[26] JCR: sort out this reference.

Under the betting interpretation, I would pay £p to buy one unit of bet on Q, or I would accept £p to sell one unit of bet. In general I would exchange $w \cdot$ £p for an outcome of $w \cdot$ £$\mathbb{1}_Q$, where w is the number of units, with $w > 0$ indicating buying w units of bet (paying $w \cdot$ £p to win $w \cdot$ £$\mathbb{1}_Q$) and $w < 0$ indicating selling w units of bet (receiving $|w| \cdot$ £p to pay out $|w| \cdot$ £$\mathbb{1}_Q$). All together, I am prepared, notionally if not in practice, to enter into contracts of the form

$$w \cdot (\mathbb{1}_Q - p) \quad \text{for any } w, \text{ negative or positive.}$$

This is always accepting that $|w|$ is not outlandishly large. There is a generalisation, which goes back to Ramsey (1931), which swaps £ for a more general preference-based currency, which can be thought of as tickets in a lottery.

No one would disgree that the probability of an impossible proposition is 0, and the probability of a certain proposition is 1. This is implied by the betting interpretation (under conditions to be made clear below). What the betting interpretation does is provide a way for us to attach probabilities to propositions that are neither impossible nor certain. In some situations this is straightforward. The situations of classical probability, for example, where we roll dice or toss coins, or sample randomly from a population. In this case, the betting interpretation should give the same answer as classical probabilities. But the betting interpretation extends to arbitrary propositions. For example, proposition A might be "sea level in 2100 is at least 0.5 m higher than today". One can bet on this proposition, but not embed it in a classical situation: it represents a one-off event which, come 2100, we will know to be false or true.

What would be a *not-sensible* situation according to the betting interpretation of probability? It would be one where, in a set of probabilities $p_1, \ldots, p_k$ on propositions $A_1, \ldots, A_k$, it is possible to find a set of amounts $\boldsymbol{w} := (w_1, \ldots, w_k)$ such that I can never win. More precisely, there is no outcome where I will make money, and at least one outcome where I will lose money. In the vernacular, with these probabilities I could be turned into a 'money pump'. People would bet with me for as large a $|\boldsymbol{w}|$ as I could stand, confident that they could never lose money, and on at least one outcome they will make money. Sets of probabilities where this is possible are termed *incoherent*, otherwise they are *coherent*. It seems fundamentally irrational to have incoherent probabilities; indeed, if it

were pointed out to me that my probabilities were incoherent, I would definitely want to change them. So not-sensible = incoherent, and sensible = coherent.

Definition 6 (Coherent probabilities)**.** *Let $A_1, \ldots, A_k$ be a set of propositions, with probabilities $p_1, \ldots, p_k$. These probabilities are coherent exactly when there is no set of amounts $(w_1, \ldots, w_k)$ for which the agent holding these probabilities will not make money on any outcome, and will lose money on at least one outcome.*

Now for the exciting result. A set of probabilities is coherent *if and only if* the probabilities obey the Laws of Probability given in Def. 4. This result knits together the betting interpretation of probability and the Laws of Probability; it is sometimes termed the *Dutch Book* argument, which is the name I will use below. There is a proof of this result using expectations, which I do not like; see Howson (1997) and Kadane (2011, sec. 1.7). I will provide a better proof based on a standard mathematical result.

Before this next result, a brief clarification on vector inequalities: $\boldsymbol{x} \geq \mathbf{0}$ indicates that $x_i \geq 0$ for all i; $\boldsymbol{x} > \mathbf{0}$ indicates that $\boldsymbol{x} \geq \mathbf{0}$ and $x_i > 0$ for at least one i; $\boldsymbol{x} \gg 0$ indicates that $x_i > 0$ for all i. There are lots of variants on the following result; my reference for this one (and its name) is the *Encyclopedia of Mathematics*, `https://www.encyclopediaofmath.org/index.php/Motzkin_transposition_theorem`.

Theorem 1.9 (Stiemke's Theorem)**.** *Let A be an $m \times n$ matrix of reals. Then exactly one of these two alternatives is true:*

1. *There exists an $\boldsymbol{x} \gg \mathbf{0}$ for which $A\boldsymbol{x} = \mathbf{0}$,*
2. *There exists a $\boldsymbol{y}$ for which $A^T\boldsymbol{y} > \mathbf{0}$.*

Theorem 1.10 (Dutch Book Theorem)**.** *Probabilities are coherent if and only if they obey the Laws of Probability (Def. 4).*

Proof. Consider any two propositions which are mutually exclusive, and label them P and Q. Let $p := \Pr(P)$, $q := \Pr(Q)$, and $r := \Pr(P \vee Q)$. Construct the outcome matrix of a set of one-unit bets, where each row is one possible outcome. There are three outcomes in total: P is true and Q is false, P is false and Q is true, or P is false and Q is false. Each column is the pay-off for one unit on one of the three bets, on P, on Q, and on $P \vee Q$. Thus

$$M := \begin{array}{r} \\ P \wedge \neg Q \\ \neg P \wedge \ Q \\ \neg P \wedge \neg Q \end{array} \begin{array}{c} \begin{array}{ccc} P & Q & P \vee Q \end{array} \\ \begin{pmatrix} 1-p & -q & 1-r \\ -p & 1-q & 1-r \\ -p & -q & -r \end{pmatrix} \end{array}.$$

The outcomes for a set of amounts $\boldsymbol{w} := (w_1, w_2, w_3)$ is $M\boldsymbol{w}$.

The presence of all three rows in M indicates that none of the three outcomes are impossible. If an outcome such as '$\neg P, \neg Q$' is impossible under the definition of P and Q then its row is dropped

from M. This is part of the operational nature of the betting interpretation: when we think about outcomes, we only think about outcomes that are possible, because these are the only ones where it matters whether we gain or lose money.[27]

[27] As Matthew records, "sufficient unto the day is the evil thereof" (ch6, v34).

Consider Stiemke's Theorem (Thm 1.9), with $A \leftarrow -M^T$ and $\boldsymbol{y} \leftarrow \boldsymbol{w}$. The second alternative now reads $M\boldsymbol{w} < \mathbf{0}$. This is the definition of incoherence. Therefore coherence is equivalent to the first alternative, which now reads "there exists a $\boldsymbol{x} \gg \mathbf{0}$ for which $M^T\boldsymbol{x} = \mathbf{0}$." We have to show that this condition is equivalent to $p \geq 0$, $q \geq 0$, $p + q = r$, and, if P and Q are exhaustive, $r = 1$. We will prove the equivalent conditions that, if P and Q are not impossible, then $p > 0$, $q > 0$, the other conditions being the same.

Let $s := x_1 + x_2 + x_3$, where $\boldsymbol{x} \gg \mathbf{0}$ implies that $s > 0$. Multiply out $M^T\boldsymbol{x} = \mathbf{0}$ to derive the three equations

$$\begin{aligned} x_1 - p \cdot s &= 0 \\ x_2 - q \cdot s &= 0 \\ x_1 + x_2 - r \cdot s &= 0. \end{aligned}$$

We infer immediately that $p > 0$, $q > 0$, and $p + q = r$. Now let P and Q be exhaustive, as well as mutually exclusive. The third outcome is now impossible, so M has two rows and three columns, and $\boldsymbol{x} = (x_1, x_2)$. Multiplying out as before gives the equations

$$\begin{aligned} x_1 - p \cdot s &= 0 \\ x_2 - q \cdot s &= 0 \\ (1 - r) \cdot s &= 0, \end{aligned}$$

from which $r = 1$ (and, as before, $p > 0$, $q > 0$, and $p + q = r$).

To check the converse, we show that if the Laws of Probability are violated then $M^T\boldsymbol{x} = \mathbf{0}$ does not have a solution $\boldsymbol{x} \gg \mathbf{0}$, in which case we would be in the second alternative of Stiemke's Theorem and the probabilities would be incoherent. This is straightforward. Briefly: if $p = 0$ and P is not impossible, then $x_1 = 0$; likewise, if $p < 0$ then x_1 and s would be of different signs. If $p + q \neq r$, then $s = 0$. If $r \neq 1$ when P and Q are exhaustive, then $s = 0$. In all of these cases the condition $\boldsymbol{x} \gg \mathbf{0}$ is contradicted.

□

The proof of Thm 1.10 is quite clear about the equivalence of 'Q is impossible' and $\Pr(Q) = 0$. 'Impossible' means 'logically impossible', not merely 'almost inconceivable'. Impossible outcomes get removed from M, but almost inconceivable ones do not, because one can still lose money if an almost inconceivable outcome occurs. Thus not-impossible outcomes have positive probabilities under coherence, even though they may be tiny. It is a mistake to think that tiny probabilities can be set to zero. Interesting propositions can be constructed as disjunctions of billions of mutually exclusive atomic propositions (see below). If all tiny probabilities were set to zero, then we could end up with the probability of the certain event

being less than 1, and that would be incoherent. Dennis Lindley (1985) made this into a Principle.

Definition 7 (Cromwell's Rule)**.** *Reserve* $\Pr(Q) = 0$ *for cases where* Q *is logically impossible.*

* * *

There are a huge number of additional relations that are implied by the Laws of Probability; this is the topic of Probability Theory. If $A_1, \ldots, A_k$ were a rich set of propositions, then it would be almost impossible for me to specify coherent probabilities for all of the propositions that could be constructed from $A_1, \ldots A_k$. This is why, in practice, it is better to build probabilities by applying the Laws of Probability, achieving probabilities for complicated propositions by combining simpler ones.

The most primitive strategy for doing this is to break all of the propositions down into a set of mutually exclusive and exhaustive 'atoms', so that every proposition can be expressed as a disjunction of atoms. For a finite set of propositions, this takes the form of expanding out the tautology[28]

$$\mathsf{TRUE} = (A_1 \vee \neg A_1) \wedge \cdots \wedge (A_k \vee \neg A_k) = \bigvee_{j=1}^{2^k} A^{(j)}$$

where each atom $A^{(j)}$ has the form $(\tilde{A}_1 \wedge \cdots \wedge \tilde{A}_k)$, where $\tilde{A}_i$ is either A_i or $\neg A_i$. Many of these atoms will be impossible and have zero probabilities. For example, if A_i implies A_j, then all atoms with A_i and $\neg A_j$ in them will have zero probabilities. The rest must have positive probabilities which sum to 1.

[28] Remember the distributive rule that $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$.

This comment is not as abstract as it seems. In Statistics, when the propositions concern random quantities, the atoms are associated with the elements of the joint realm of $\boldsymbol{X}$ represented by the set Ω. We have

$$\mathsf{TRUE} = \bigvee_{\omega \in \Omega} (\boldsymbol{X} \doteq \boldsymbol{x}(\omega)).$$

The probabilities on the atoms are represented by the function $p \in \mathcal{P}$, according to Thm 1.8. According to Thm 1.10, the two conditions $p(\omega) \geq 0$ and $\sum_\omega p(\omega) = 1$ are necessary and sufficient for probabilistic coherence. When theorists write "Let Ω be a set, let $\mathcal{F}$ be a σ-algebra over Ω, and let p be a non-negative, finite, σ-additive measure on $\mathcal{F}$, normalised so that $p(\Omega) = 1$" they are doing exactly this, but using concepts that allow generalisation to non-countable Ω, for which the notion of an atom is more tricky.

1.8 Conditional probabilities

The stunning result of the Dutch Book Theorem prompts us to go further, and consider conditional probabilities. We need to find a betting interpretation of the conditional probability 'P given Q', and

then verify that bets involving conditional probabilities are coherent if and only if

$$\Pr(P, Q) = \Pr(P \mid Q) \cdot \Pr(Q) \tag{1.39}$$

which is accepted as the defining property of the conditional probability '$\Pr(P \mid Q)$'. Note the convention in Probability and Statistics of writing a comma in place of the conjunction '$\wedge$', i.e.

$$(P, Q) := (P \wedge Q).$$

Some authors write that $\Pr(P \mid Q)$ is undefined when $\Pr(Q) = 0$; this is a mistake. The relation has the form $0 = \Pr(P \mid Q) \cdot 0$, and hence $\Pr(P \mid Q)$ is arbitrary in this case, not undefined.

The interpretation that works is a 'called off bet'. Let $r := \Pr(P \mid Q)$. Then I am indifferent between having r with certainty, and owning a bet with pay-off

$$\mathbb{1}_Q \cdot \mathbb{1}_P + (1 - \mathbb{1}_Q) \cdot r.$$

In this bet I get $\mathbb{1}_P$ if Q is true, and my money back if Q is false. Thus the bet is 'called off' if Q is false. All together, I am prepared to enter into contracts of the form

$$w \cdot \mathbb{1}_Q(\mathbb{1}_P - r) \quad \text{for any } w, \text{ positive or negative.}$$

Theorem 1.11 (Conditional Dutch Book Theorem)**.** *Let P and Q be any two propositions. Then the conditional probability $\Pr(P \mid Q)$ is coherent if and only if $\Pr(P, Q) = \Pr(P \mid Q) \cdot \Pr(Q)$.*

Proof. It's the same proof as Thm 1.10. Let $p := \Pr(P, Q)$, $q := \Pr(Q)$, and $r := \Pr(P \mid Q)$. We must show that $p = r \cdot q$.

Assume that all four outcomes concerning P and Q are possible. Now the outcome matrix is

$$M := \begin{array}{r} \\ \neg P, \neg Q \\ P, \neg Q \\ \neg P, \quad Q \\ P, \quad Q \end{array} \begin{array}{c} \begin{array}{ccc} P \wedge Q & Q & P \mid Q \end{array} \\ \begin{pmatrix} -p & -q & 0 \\ -p & -q & 0 \\ -p & 1-q & -r \\ 1-p & 1-q & 1-r \end{pmatrix} \end{array}.$$

Multiply out $M^T \boldsymbol{x} = \mathbf{0}$ to give the three equations

$$\begin{aligned} x_4 - p \cdot s &= 0 \\ x_3 + x_4 - q \cdot s &= 0 \\ x_4 - (x_3 + x_4) \cdot r &= 0, \end{aligned}$$

where $s := x_1 + x_2 + x_3 + x_4 > 0$, as before. Because $\boldsymbol{x} \gg \mathbf{0}$, so $r > 0$ and $x_3 + x_4 = x_4/r$, from the third equation. Substituting into the first and second equations gives

$$\begin{aligned} x_4 - p \cdot s &= 0 \\ \frac{x_4}{r} - q \cdot s &= 0 \end{aligned}$$

from which it follows immediately that

$$p = r \cdot q,$$

as was to be shown (also $0 < p < q$).

To check the converse, note that $s > 0$ can be taken without loss of generality, because of the free variables x_1 and x_2. $p = 0$ would imply $x_4 = 0$. If $p > 0$,

$$\frac{p}{q} = \frac{x_3}{x_3 + x_4}, \quad r = \frac{x_3}{x_3 + x_4}.$$

If $p \neq r \cdot q$ then there is no solution in $\boldsymbol{x}$. Thus the second alternative in Stiemke's Theorem would hold, i.e. the probabilities p, q, and r would be incoherent. □

1.9 *Further thoughts on subjective probabilities*

Here are some more general comments, which apply as much to expectations as they do to probabilities.

First, how I or anyone else produces a value $\Pr(Q)$ is mysterious. Through my life I have been exposed to information which may be relevant to the truth of Q; some of this information I have remembered more-or-less intact, other information has done no more than leave a vague impression. I may go and seek out new information. In the end, I reach for a probability that 'seems right' to me, and I test out my probability on myself by asking whether I would be willing to buy or sell a bet at price £p. The Laws of Probability say no more than $\Pr(Q) > 0$ if Q is not impossible, and $\Pr(Q) < 1$ if Q is not logically certain. If I have a second proposition R, and Q and R happen to be mutually exclusive, then the Laws have something further to say. If it turns out that my probabilities are incoherent, the Laws do not tell me how to modify them. This is down to me.

On this basis, the impression that we often agree, approximately, about probabilities deserves some thought. Likewise the related impression that we are often willing to accept someone else's probabilities as our own. In fact this latter impression is not so hard to understand. There are some domains, future weather for example, where some people have hard-earned expertise. A meteorologist knows a lot more about future weather than I do, and it would be sensible of me to accept a meteorologist's probabilities as my own, once I have satisfied myself that her probabilities are coherent.[29] I am not accepting her probabilities because they are 'right', a concept which makes no sense. I am accepting them, and sometimes paying for them, because I believe that my decisions made on the basis of her probabilities about future weather will work out better than decisions made on the basis of my own probabilities.

[29] This is the practical definition of an expert: 'someone whose probabilities you accept as your own'.

But what to make of the impression that we often agree, approximately, about probabilities? The simplest explanation is that we humans tend to think alike, and, in many cases where we agree, it is because we have been exposed to similar models and similar evidence. Here is a cute result on this topic. I'm not claiming much more for it than this!

Suppose there is a sequence of experimental outcomes, $E_1, E_2, \ldots,$

all of which are implied by a scientific model M. Represent this as

$$\Pr(E_{\mathcal{A}} \mid M) = 1 \quad \text{for all } \mathcal{A},$$

where $E_{\mathcal{A}}$ denotes the conjunction of any subset $\mathcal{A}$ of the experimental outcomes.[30] Then we have the following remarkable result, termed the *First Induction Theorem* by Good (1975), and originally proved by Wrinch and Jeffreys (1921).

[30] I.e., $E_{\mathcal{A}} := \wedge_{i \in \mathcal{A}} E_i$ where $\mathcal{A} \subset \mathbb{N}$.

Theorem 1.12 (First Induction Theorem)**.** *Let* $\Pr(E_{\mathcal{A}} \mid M) = 1$ *for all* $\mathcal{A}$*. If* $\Pr(M) > 0$ *then*

$$\lim_{n \to \infty} \Pr(E_n \mid E_1, \ldots, E_{n-1}) = 1.$$

Proof. Under the conditions of the theorem,

$$\begin{aligned} \Pr(E_{\mathcal{A}}) &= \Pr(E_{\mathcal{A}} \mid M) \Pr(M) + \Pr(E_{\mathcal{A}} \mid \neg M) \Pr(\neg M) \\ &\geq \Pr(E_{\mathcal{A}} \mid M) \Pr(M) \\ &= \Pr(M) \end{aligned}$$

for al $\mathcal{A}$. Now let $\mathcal{A} \leftarrow \{1, \ldots, n\}$ and write the lefthand side as

$$p_n := \Pr(E_1, \ldots, E_n) = \Pr(E_1) \prod_{i=2}^{n} \Pr(E_i \mid E_1, \ldots, E_{i-1}).$$

$p_1, p_2, \ldots$ is a monotone decreasing sequence bounded below by $\Pr(M)$. Since $\Pr(M) > 0$ it converges to a positive limit, in which case $\Pr(E_n \mid E_1, \ldots, E_{n-1})$ converges to 1. □

The remarkable thing about this result is that the displayed equation in Thm 1.12 makes no reference to model M at all. It indicates that anyone who believes that M implies the E's and that M is not logically impossible is bound, sooner or later, on the accumulation of enough evidence, to act as though M is true, in terms of their probabilities for other implications of M. Relaxing the conditions of the result, to allow for 'fuzziness' in the definition of M and in the nature of the evidence, we can still infer that probabilities will tend to be similar, because we will be channeled by exposure to similar evidence into probabilistically similar models for the world.