Internal laws of probability, generalised likelihoods and Lewis' infinitesimal chances – a response to Adam Elga

Frederik Herzberg[*†]

Abstract

Elga's rejection [5] of an infinitesimal solution to the zero-fit problem does not seem to appreciate the opportunities provided by the use of internal finitely additive probability measures. Indeed, internal laws of probability can be used to find a satisfactory infinitesimal answer to many zero-fit problems, not only to the one suggested by Elga, but also to Markov chain (that is, discrete and memory-less) models of reality. Moreover, the generalisation of likelihoods that Elga has in mind is not as hopeless as it appears to be in his article. In fact, for many practically important examples, through the use of likelihoods one can succeed in circumventing the zero-fit problem.

1 The zero-fit problem on infinite state spaces

In general, the zero-fit problem arises when two candidates for a probabilistic law of nature assign probability ("fit") zero to an observation and it is therefore not possible to prefer one over the other. This is the case for instance on any state space that is an infinite power of a nontrivial probability space (i.e. a probability space without an atom[1] of probability 1) [2] – such as Elga's example $2^\omega = \{0,1\}^{\mathbb{N}}$, the space of all countable sequences of coin tosses – and holds as obviously for every atomless probability space (for example the Lebesgue measure on the unit interval $[0,1]$ or the Gaussian distribution on the field of the reals).

---

*Institut für Angewandte Mathematik, Universität Bonn, Germany

†Mathematical Institute, University of Oxford, United Kingdom

[1] An atom of a measure space $(\Omega, \mathcal{A}, \mu)$ is a set $A \in \mathcal{A}$ of strictly positive μ-measure such that any measurable subset of A has either measure $\mu(A)$ or zero.

[2] The proof for this assertion is straightforward: Let the probability space be $\Omega = {\Omega_0}^\kappa$ with a probability measure P_0 defined on the σ-algebra generated by the singletons of elements of Ω_0 and an infinite κ. Every singleton $\{(a_i)_{i\in\kappa}\} \subset \Omega$ has probability less or equal to the product $\Pi_{i\in I} P_0\{a_i\}$ for any finite $I \subset \kappa$. Since $P_0\{a_i\} < 1$ holds for every $i \in \kappa$ (because of the nontriviality of (Ω, P_0)), $P_0\{(a_i)_{i\in\kappa}\}$ is less than any positive number, thus equal to zero.

One should realise that the zero-fit problem will never be encountered when actually *modelling* physical data in practice: Due to constraints on the instruments, there will always be bounds on the range of any observation variable that can be measured. In view of the quantisation of physical observables, this entails the finiteness of the state spaces associated with the respective experiment. Moreover, as we shall see in section 6, in the most prominent case of probability measures (or *probability distributions*, as we will sometimes call them) on finite-dimensional real vector spaces, statisticians have a tool to deal with the zero-fit problem: Likelihoods, a special case of what Elga rejects as "integrating over densities" in generality.

Notwithstanding this remark, whenever *thought experiments* are conducted for the investigation of theoretical hypotheses, there is of course no reason why the state spaces should be assumed to be finite. Indeed, thought experiments on criteria for the adequate description of an infinite history of observations have been the framework in which the zero-fit problem was phrased for the first time by David Lewis [16].

Now, if one was allowing for infinitesimal probabilities – which is precisely what Lewis suggested [ibid.] to address the zero-fit problem as an obstacle to a successful "best-system analysis" – , one might be able to distinguish between the two competing probability distributions of *standard* (as opposed to nonstandard in the sense of nonstandard analysis[3]) fit zero by choosing the one assigning the bigger (infinitesimal) probability to the "observation" obtained in the thought experiment (in Elga's paper [5] referred to as the "actual world"). Any attempt to address and settle the zero-fit problem thus entails a built-in bias to take a nonstandard notion of probability into consideration.

2 Elga's critique of the infinitesimal approach to the zero-fit problem

Before summarising Elga's reasoning, let us introduce some convention from "standard" measure-theoretic terminology. The term "measure" in this paper will always mean "σ-additive measure defined on a σ-algebra", unless it is used in the combination "finitely-additive measure"[4] or "internal [probability] measure".

Elga's main line of thought against the use of infinitesimals to solve the zero-fit problem can be rendered as follows:

1. The only way to obtain a *nonstandard probability measure* (whether finitely additive or σ-additive) is by approximating a standard probability measure

[3]Nonstandard analysis is the part of analysis that uses infinitesimal elements and model-theoretic techniques to formalise heuristic combinatorial reasoning. Its modern shape dates back to Robinson's classical monograph [23]. There are many accessible introductions to nonstandard analysis, one might mention e.g. [17] and the introductory parts of [2].

[4]A *finitely-additive measure* is a function $\mu : \mathcal{A} \to [0, \infty]$ defined on an algebra $\mathcal{A}$ assigning to the union of any two disjoint sets the sum of the measure of these two sets, assigning measure zero to the empty set and being monotone with respect to $\subseteq$. A *measure* μ, on the other hand, has the additional property that for every sequence $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ of disjoint sets the sum $\sum_n \mu(A_n)$ converges and equals $\mu\left(\bigcup_n A_n\right)$. A *$\sigma$-algebra* is an algebra which is closed under forming countable unions (and therefore under countable intersections as well). A *(finitely additive) probability measure* is a (finitely additive measure) assigning measure 1 to the whole space. In [5], finitely-additive probability measures are called "probability functions".

on the same (σ-)algebra.

2. The only possibility to perform this approximation is the method proposed in [5, Appendix A].

3. The approximation in Appendix A is not unique in that every nonstandard probability assigned to an event could be changed by an infinitesimal.

   Conclusion:

4. There is never a canonical nonstandard probability measure for the formulation of a probablistic law of nature.

The first presumption in his argument is flawed[5] – notwithstanding that Elga has proposed a convincing reasoning in favour of the second and third point in the main argument. It is the main objective of this article to show that Elga's first premise is inaccurate.

There is no reason why the nonstandard probability measure we are looking for should be defined on the state space itself, as long as we can find a measurable transformation from the probability space where the nonstandard measure is defined on to the state space. Therefore, it is sensible to allow for nonstandard probabilistic laws of nature that are defined on *internal* probability spaces (different from the state space). This prompts a new possibility to devise nonstandard probability measures (which in turn may give the ultimately successful candidates for probabilistic natures in the best system analysis suggested by Lewis): The construction of (finitely-additive) *internal probability measures* on internal algebras that can be measurably mapped to an algebra containing all the singletons of the state space.

Since standard probability measures are extremely well understood, it is desirable to associate a standard probability measure with any such internal probability measure. In fact, one is lucky: Due to a theorem by Loeb [18], any internal probability measure gives rise to a standard σ-additive probability measure extending the standard part[6] of the internal measure.

One can even define finitely-additive nonstandard (i.e. hyperreal-valued) probability measures (*nonstandard probability functions* in the terminology of [5]) satisfying Elga's "regularity condition"[7] very naturally on certain internal subalgebras of internal (i.e. in the model-theoretic sense: *definable*) algebras[8] by taking a generating system of this algebra and removing all algebra elements to which the internal probability measure assigns zero. And every *internal probability space* gives – via the Loeb measure construction outlined in Appendix

[5] Elga himself appears to realise that this point might be problematic, when he writes "To my knowledge, there are only two ways of cooking up regular nonstandard probability functions.“ [5, p. 70, footnote 10]

[6] For every hyperreal number $r \in {}^*\mathbb{R}$, the *standard part* r *of* r is defined to be the unique real number s (if existent) such that $r = s + \varepsilon$ for an infinitesimal ε. If no such number s exists, one sets $r = +\infty$ if $r > 0$ and $r = -\infty$ if $r < 0$.

[7] This condition says that the empty set is the only event with probability zero.

[8] Probably the most frequently used example is the normalised hyperfinite counting measure on the hyperfinite time line $\mathbb{T} := \left\{0, \frac{1}{H}, \ldots, \frac{H-1}{H}, 1\right\}$, a subset of the hyperreal unit interval ${}^*[0,1]$ for an infinite hyperfinite number $H \in {}^*\mathbb{N} \setminus \mathbb{N}$. The normalised hyperfinite counting measure assigns to an internal subset A of $\mathbb{T}$ the internal cardinality of A (i.e. the smallest hyperfinite number h such that there is an internal bijection between $\{0, \ldots, h-1\}$ and A) divided by H.

A – rise to a σ-additive measure on the completion of the σ-algebra generated by the internal algebra.

Summarising and paraphrasing the previous remarks, the pivotal difference between Elga's construction and ours is the order in which nonstandard and standard measures are introduced.

3 Two examples for infinitesimal solutions to the zero-fit problem

Now we will look for canonical candidates for such an internal (finitely additive) probability measure in two settings.

First, we shall turn our attention to the example that Elga himself has presented to illustrate the ambiguities that one will invariably get from merely asserting the existence of nonstandard probability measures.

Let us fix a real number $\bar{p} \in (0,1)$ and denote the Bernoulli distribution of parameter p by $B_p = p\delta_0 + (1-p)\delta_1$ for any $p \in (0,1)$. For each $h \in {}^*\mathbb{N}$ and $p \in (0,1)$, we can assign an infinitesimal probability to each element a of the sample space 2^ω by defining

$$\forall a \in 2^\omega \quad \nu_p^{(h)}\{a\} = \prod_{i<h} B_p \left\{ {}^*a(i) \right\},$$

where the right hand side is an element of ${}^*(0,1)$ by the transfer principle. If h is finite – that is, just a standard natural number – , then, after choosing an $a \in 2^\omega$, the number $\nu_p^{(h)}\{a\}$ is the probability to get the sequence $(a(0), \ldots, a(n-1))$ after n coin tosses, under the assumption that the coin gives 0 with probability p. In this setting, the function $p \mapsto \nu_p^{(h)}\{a\}$ has an isolated maximum in $\bar{p}$. Combining this and other analytic properties of the function $p \mapsto \nu_p^{(h)}\{a\}$ with various instances of the transfer principle, we get that the function $p \mapsto \nu_p^{(h)}\{a\}$ is maximised by $\bar{p}$ among all standard real numbers $p \in [0,1] \cap \mathbb{R}$. This is proven in Appendix B as Theorem B.1.

Let us next turn to a model of (one part of) what we perceive as the history of "physical reality" that – while still being rather simplistic – comes closer to what an actual scientific description would look like than a mere sequence of coin tosses.

In order to keep things simple, we will exploit the quantisation of time and we will therefore model the time-line as ω, the set of natural numbers. We shall look at a model of the world that only knows finitely many states of the world, and we shall impose the assumption that this description be *instantaneously complete* in the sense that the conditional probability for a transition from one state A to another one B will only depend on A. Technically speaking, we will thus deal with a finite-state Markov chain.

We can view this as the partial description of a finite set of particles, endowed with mass and electromagnetic charge, obeying the laws of classical Newtonian mechanics (interpreted appropriately to take our time quantisation premise into account) and Coulomb's law – except when a collosion of two particles occurs, which will stochastically result either in an elastic or a non-elastic collision (the latter one forming a new particle, called a two-component particle), or in the case of one of the particles being two-component, this two-component particle will always be breaking up into its components or the collision will be elastic.[9] Then, if we assume that for each collision the stochastic distribution of possible outcomes will only depend on the current state of the world (as opposed to its history), a Markov chain model can be employed to describe the *configuration* of the particles, the set of all singletons and pair-sets which correspond to one or two-component particles, respectively. The decription through a finite-state Markov chain will make sense since the set of all possible configurations has finite cardinality.

Let us enumerate the possible configurations by $0, \ldots, N-1$ and let $p_{i,j} := p(i,j)$ for $i,j \in \{0, \ldots, N-1\}$ denote the *conditional probability* that, given the current configuration is i, the configuration in the next moment will be j. Then $p := (p_{i,j})_{i<N, j<N} \in [0,1]^{N \times N}$ is commonly referred to as the *matrix of transition probabilities* and satisfies the two conditions

$$\forall j < N \quad \sum_{i=0}^{N-1} p_{i,j} = 1 \tag{1}$$

(as we assume not to have omitted in our enumeration any possible configuration that might occur in the next moment) as well as the dual condition

$$\forall i < N \quad \sum_{j=0}^{N-1} p_{i,j} = 1 \tag{2}$$

(meaning that we have counted in all the possible configurations from which the current one might have come from). The set of possible histories of (configurations which characterise one aspect of) the world will now equal the set $(N \times N)^\omega = (\{0, \ldots, N-1\} \times \{0, \ldots, N-1\})^{\mathbb{N}}$.

The zero-fit problem now appears in the following guise: If $a \in (N \times N)^\omega$ is the actual history of the world, we are inclined to think that the matrix $p = (p_{i,j})_{i<N, j<N} \in [0,1]^{N \times N}$ of conditional probabilities that best fit our observations should be given by the empirical limiting conditional frequencies defined as

$$\forall i,j < N \quad q_{i,j}(a) = \lim_{n \to \infty} \frac{|\{\ell \leq n \ : \ a(\ell) = i, \quad a(\ell+1) = j\}|}{|\{\ell \leq n \ : \ a(\ell) = i\}|}.$$

But the probability for the course that history has taken will always be

$$\mathbb{P}_p^\xi \{a\} = \xi\{a(0)\} \cdot \prod_{n=0}^{\infty} p_{a(n), a(n+1)}$$

(where ξ denotes the start distribution, that is the distribution of possible configurations at time 0) and this probability will – regardless of the particular choice of p – always equal zero, except in the trivial case of $\forall n < \omega \quad a(n) = a(0)$ and $p_{a(0),a(0)} = 1$ (corresponding to a world that does not change at all) where it is going to be $\xi\{a(0)\}$.

However for the sake of circumventing this zero-fit problem we can now employ the method outlined in Appendix B to canonically assign for arbitrary

[9]Studying this simple model of physical reality was suggested by an anonymous referee to whom the the author would like to take this opportunity and express his sincere gratitude.

hypernatural $h \in {}^*\mathbb{N} \setminus \mathbb{N}$ an infinitesimal chance $\nu_p^{(h)}\{a\}$ to a for every $p \in [0,1]^{N \times N}$ satisfing (1) and (2) by

$$\nu_p^{(h)}\{a\} = \xi\{a(0)\} \cdot \prod_{n<h} p_{{}^*a(n),{}^*a(n+1)}$$

(where * is an elementary embedding – in the model-theoretic sense – from the standard mathematical universe to a nonstandard universe). One can then show, in the same manner as in the proof of Theorem B.1 that for any infinite h, $\nu_p^{(h)}\{a\}$ as a function of p becomes maximal when $p = q_{i,j}(a)$ – which is nothing else but to say that when we truncate history after finitely many steps, the transition matrix p that fits the observation best will simply be the one obtained from evaluating the observation. Using the transfer principle and the nonstandard characterisation of limits, this will also hold for all $h \in {}^*\mathbb{N}$ when $p = q_{i,j}(a)$.

A more rigorous presentation of these deliberations is provided in Theorem C.1 and its proof.

4 Mathematical modelling in nonstandard universes?

The only remaning difficulty in the route of the previous section, which can be but very serious, is to find the appropriate internal probability space related to the state space. To a certain extent, this obstacle should not be too surprising. On the one hand, roughly speaking, this is what nonstandard analysis is all about[10]. On the other hand, the probabilistic modelling of a phenomenon in the physical world is not a trivial task either – apart from certain toy problems that are only used to emphasise methodoligical complications.

Given the task of finding an appropriate mathematical framework for a discourse involving probabilistic statements, there is no obvious reason to conceive of probabilities as elements of the real unit interval $[0,1]$ rather than its nonstandard counterpart ${}^*[0,1]$[11] Basically, the notion of probability is a generalisation and idealisation of the notion of relative frequency.

Therefore, there are not that many restrictions on what a suitable notion of probability should be. Let us assume that we agree that a notion of probability should always be realised by a function $\mu : \mathcal{A} \to S$, where $\mathcal{A}$ is an algebra (reflecting the canonical correspondence between Boolean algebras[12] and sets of sentences of a first order propositional language) on some space Ω and S is a certain structure. The requirements on this structure are more or less immediate

[10] Written from a mathematical physicist's point of view, [1] is a non-technical article highlighting related advantages and challenges of nonstandard analysis.

[11] A "radical" work by Nelson [20] concentrates on the latter rather than the former, proving nonstandard versions of the Central Limit Theorem and the Law of Large Numbers for internal stochastic processes and random variables. This is not about using nonstandard methods to obtain results in standard stocahstic analysis (which has also been successful, witnessed e.g. in [3],[21],[8, 9],[22], [14],[2],[25], [15], and more recently in [6]), but about a quite different conception of what probability is.

[12] As was already pointed out, the usual convention in mathematical probability theory is to consider, by definition of the spaces of events, only σ-additive measures on σ-algebras. This is largely a consequence of the intention to study stochastic processes and their limiting behaviour.

from how we use natural language to talk about probabilities: One should be able to do basic arithmetic in S; it should be possible to compare two elements of S (by means of a linear order $\prec$); there should be two elements $o, \mathbf{1} \in S$ such that $o \prec p \prec \mathbf{1}$ for all S; the rational numbers between 0 and 1 should be embeddable into S (in order to include references to relative frequencies). Since every ordered field can only have characteristic zero, one would technically require:

1. S is a subset of an ordered field $(F, \prec, +, \cdot, o, \mathbf{1})$;
2. For all $p \in S$, $o \prec p \prec \mathbf{1}$

Obviously, ${}^*[0,1]$ satisfies these criteria for the field $({}^*\mathbb{R}, \leq, +, \cdot, 0, 1)$.

5 Are nonstandard models unnatural?

One objection to the use of any nonstandard analysis at all might read: "The proof for the existence of nonstandard models of the reals is nonconstructive – using the ultrafilter existence theorem – , therefore there is no way to call any nonstandard (finitely additive) probability measure canonical." In fact, there is a definable nonstandard model of the reals, as proven by Kanovei and Shelah [13]. Also note that the method of constructing nonstandard finitely additive probability measures proposed in this paper does not rely on features of a particular class of nonstandard models of the reals (apart from $\aleph_1$-saturation, which holds regardless of the ultrafilter one chooses to extend the filter of co-finite sets on ω in the usual existence proof of ${}^*\mathbb{R}$).

6 Likelihoods and densities

Elga argues against another possible solution of the zero-fit problem which he refers to as "integrating over densities". Basically, this is nothing but a straightforward generalisation of what is well-known as "comparing likelihoods" to statisticians. The common terminology in statistics [4], which we shall adopt, is to address as "likelihood" what Elga calls "density", and to refer to Elga's "density functions" as "densities" or "likelihood functions".

What is remarkable about a rejection of this method (or its canonical generalisation) is that this approach is one of the most frequently-used and basic tools in statistics, although its usual formulation can only be applied successfully if one has to compare probability distributions on finite-dimensional real vector spaces. On the other hand, this is certainly the most important example for a zero-fit problem arising in practice.

The theoretical background for the use of densities or likelihood functions is the *Radon-Nikodym Theorem*: If μ and ν are measures on the same measure space, then ν has a μ-density if and only if all null sets with respect to μ (these are called *μ-null sets*) are ν-null sets as well [7].

For example, all Gaussian measures on finite-dimensional real vector spaces have, of course, a density with respect to the Lebesgue measure. In the case of the space 2^ω of countable coin tossing sequences, Elga has shown in his paper [5] that no such measure μ can exist, no matter what powers of Bernoulli distributions one might be looking at.[13]

So the deeper reason why a generalised likelihood ("integrating over densities") approach fails in the case of the countable coin tossing example is that the powers of Bernoulli distributions on 2^ω do not have all the same null sets. But this already provides a new way of distinguishing two candidates for the best-system analysis, and possibly a comparison of the ideals of null sets for powers of different Bernoulli measures highlights a difference that vindicates the preference of one over the other.

Generalising our previous remarks, we may maintain that the idea of using generalised likelihoods is not automatically deemed to failure, but might well be a promising approach to deal with zero-fit problems.

When looking for explicit formulae for the densities of the candidate distributions, there are certain probability measures that can be regarded as the canonical ones to be used as the measure with respect to which the respective densities are to be found: If the measure space on which the candidate measures are defined has a topological group structure, there is the Haar measure [7], the essentially unique translation-invariant measure – e.g. in the case of the reals with their order topology and the addition as group operation, the Lebesgue measure is the Haar measure. But even if there is no structure at all apart from the probability space structure, one can still find a canonical measure by means of the Maharam spectrum[14] [19].

In order to talk meaningfully about the value of a density function with respect to this canonical measure when applied to a particular element of the probability space, we will again have to single out one particular version of this density function – as any density function will *a priori* only be defined up to changes on null sets. One particular approach will be to look for continuous densities: In an appendix, we shall show that if we assume the probability space $(\Gamma, \mathcal{G}, \nu)$ to be a complete separable metric space and $\mathcal{G}$ to be the Borel σ-algebra generated by the topology of Γ, then a continuous version of a density function on $\Omega = \sum_{\kappa \in I} \alpha_\kappa \cdot [0,1]^\kappa$ will be well-defined everywhere. In that case a pointwise comparison of two continuous versions of density functions makes sense and can be applied to a best system analysis in the following vein. Suppose we are given a set of equivalent measures $\{\mu_j : j \in J\}$ of equal complexity (for instance if $j \in J$ is simply a parameter occurring in the definition of the μ_j) – possible models of (one aspect of) reality – on Ω, and suppose ν_j have continuous densities f_j. Then our observations in the preceding paragraph yield the uniqueness of f_j. Therefore, fixing the actual world $a \in \Omega$, it does make sense to look, generalising the maximum likelihood method, for a parameter

[13]One might suspect that Elga's proof is incomplete because he does not show the measurability of those subsets of 2^ω that yield the *reductio ad absurdum*. However, this is not a serious gap in this particular case (cf. Appendix D).

[14]The *measure algebra* of a measure space is the σ-algebra of the measure space modulo the ideal of null sets. One can show [19] that for any atomless probability space there is a countable set of cardinals, the *Maharam spectrum* I, and a set of positive numbers $\{\alpha_\kappa\}_{\kappa \in I}$ satisfying $\sum_\kappa \alpha_\kappa = 1$ such that the measure algebra of that probability space is algebra-isomorphic to the measure algebra of $\Omega = \sum_{\kappa \in I} \alpha_\kappa \cdot [0,1]^\kappa$ – where the unit interval $[0,1]$ is assumed to be equipped with the Lebesgue measure – , the sum being direct and calculated event-wise. The measure algebras derived from the measure space $[0,1]^\kappa$ for some cardinal κ are called *homogeneous measure algebras*. On the importance of Maharam spectra for the model theory of stochastic processes, cf. Fajardo and Keisler [6].

or index $j \in J$ such that $f_j(a)$ is maximal. However, one should note once again that this way of comparing two atomless probability measures first of all presupposes having found a density with a continuous version on the associated direct sum of homogeneous measure algebras.

Appendix

A Internal probability measures and the Loeb measure construction

As a summary of the following, somewhat technical paragraph: One can define finitely-additive nonstandard (i.e. hyperreal-valued) probability measures (*nonstandard probability functions* in the terminology of [5]) satisfying Elga's "regularity condition"[15] often very naturally on internal (i.e. in the model-theoretic sense definable) algebras[16]. And such an *internal probability space* gives rise to a σ-additive measure on the completion of the σ-algebra generated by the internal algebra.

Starting from an *internal probabilty space* $(\Omega, \mathcal{A}, \mu)$ – i.e. a triple consisting of an internal set Ω, an internal algebra $\mathcal{A}$ of subsets of Ω and an *internal probability measure*[17] $\mu : \mathcal{A} \to {}^*[0,1]$ – one can, as Loeb [18] discovered, construct in the canonical way the Carathéodory extension [7] of $(\mathcal{A}, \mu)$, since $\aleph_1$-saturation of the nonstandard universe implies the $\emptyset$-continuity of the finitely-additive measure μ. Thus, one obtains a real-valued (as opposed to hyperreal-valued), σ-additive, probability measure extending μ to the completion $\overline{\sigma(\mathcal{A})} = \sigma(\mathcal{A}) \vee \mathcal{N}_\mu$ (with respect to μ) of the σ-algebra $\sigma(\mathcal{A})$ generated by the algebra $\mathcal{A}$. The σ-algebra $L(\mathcal{A}) := \overline{\sigma(\mathcal{A})}$ is called the *Loeb σ-algebra* of $\mathcal{A}$ (with respect to μ) and the extension of μ to $L(\mathcal{A})$ is called the *Loeb measure* $L(\mu)$ of μ.

One could even envisage a more extended application of Loeb's work [18] to our problem: Due to Loeb [18], every element of the Loeb σ-algebra can be approximated by an element of the internal algebra[18]. Since it is possible to obtain this internal set C for every element of the Loeb σ-algebra A, one can assign, using the axiom of choice, to A the nonstandard probability $\mu(C)$ and arrives at a nonstandard set function which is finitely-additive up to an infinitesimal on $L(\mathcal{A})$ (and finitely-additive on $\mathcal{A}$). This shows that it is even possible to extend the nonstandard finitely additive probability measure in a suitable way to the σ-algebra generated by $\mathcal{A}$. However, given the non-constructive manner in which this extension is achieved, one is not immediately able to avoid the problem that Elga's approximation in [5, Appendix A] has: Our "requirements

[15] This condition says that the empty set is the only event with probability zero.

[16] Probably the most frequently used example is the normalised hyperfinite counting measure on the hyperfinite time line $\mathbb{T} := \left\{0, \frac{1}{H}, \ldots, \frac{H-1}{H}, 1\right\}$, a subset of the hyperreal unit interval ${}^*[0,1]$ for an infinite hyperfinite number $H \in {}^*\mathbb{N} \setminus \mathbb{N}$. The normalised hyperfinite counting measure assigns to an internal subset A of $\mathbb{T}$ the internal cardinality of A (i.e. the smallest hyperfinite number h such that there is an internal bijection between $\{0, \ldots, h-1\}$ and A) divided by H.

[17] A function $\mu : \mathcal{A} \to {}^*[0,1]$ defined on an internal algebra $\mathcal{A}$ is an *internal probability measure* if and only if it is monotone with respect to $\subseteq$, assigns to the union of two disjoint sets the sum of the measures of these two sets, and assigns 0 to $\emptyset$.

[18] More precisely: For every $B \in L(\mathcal{A})$ there is a $C \in \mathcal{A}$ such that $L(\mu)(B \Delta C) = L(\mu)((B \setminus C) \cup (C \setminus B)) = 0$.

only very weakly constrain the probabilities those functions assign to any individual outcome" [5, p. 70]. This illustrates why one needs to accomplish that the probability measure on the algebra generated by the singletons stems from an internal (finitely-additive) probability measure.

As soon as one can assign an element of the internal space to every element of the state space and provided the measure on the internal space is chosen in such a way that singletons have always positive (possibly infinitesimal) probability, one can easily define a nonstandard finitely-additive probability measure on the algebra generated by the singletons from elements of the state space such that one ends up with a "regular nonstandard probability function" in the sense of [5]. It is crucial to note that the use of a choice function in the extension of the internal finitely additive probability measure is now causing no harm any more, for the space of the singletons of the individual worlds (which we would like assign non-zero infinitesimal probabilities to) is internal. Moreover, singletons of elements of internal sets are always internal. And for elements of the internal algebra, the nonstandard probability is fixed by the internal finitely additive measure μ.

B The (countable) coin tossing sequence revisited

We are going to show how there can be, for every Bernoulli distribution on $2 = \{0, 1\}$, a canonical nonstandard probability assigned to any possible world $a \in 2^\omega$.

Let $p \in (0, 1)$, $h \in {}^*\mathbb{N} \setminus \mathbb{N}$, and set $\Omega := 2^h$, which is the space of internal functions from $\{0, \ldots, h-1\}$ to 2. The algebra $\mathcal{A}$, consisting of all internal subsets of Ω, should be the algebra of events. There shall be an internal finitely additive probability measure $\mu : \mathcal{A} \to {}^*[0, 1]$, defined by

$$\mu_p : \{b\} \mapsto \Pi_{i<h} B_p\{b(i)\},$$

where $B_p := B_{1,p}$ is the Bernoulli measure $p \cdot \delta_0 + (1-p)\delta_1$ on 2, assigning probability p to $\{0\}$. This product on the right hand side always exists because of the transfer principle, and since $\mathcal{A}$ contains only hyperfinite elements (for 2^h is hyperfinite), this set function can be uniquely extended from the set of all singletons from Ω to the entire algebra $\mathcal{A}$. Observe that any possible world $a \in 2^\omega$ has a canonical image ${}^*a \in 2^{{}^*\mathbb{N}}$ in the nonstandard universe (which is usually identified with a).

Now one notes that $\bigcap_{i<h} \{\cdot(i) = a(i)\}$ is hyperfinite and defines the chance of a possible world a to be

$$\nu_p^{(h)}\{a\} := \mu_p \left[\bigcap_{i<h} \{\cdot(i) = a(i)\} \right],$$

which is infinitesimal, since every factor in the product of the definition of μ_p for singletons is less than one. This gives a finitely additive nonstandard probability measure in the algebra generated by singletons of the state space 2^ω, just as required.

The only arbitrarity in this construction is the choice of an $h \in {}^*\mathbb{N} \setminus \mathbb{N}$. However, this has no effect on for which worlds a a particular Bernoulli distribution

B_p should be preferred over other ones. Hence, it has no impact on the outcome of the best-system analysis.

Most importantly, one can solve the zero-fit problem in this case by proving

Theorem B.1. *Using the previous paragraph's notation and fixing an $a \in 2^\omega$, the limiting relative frequency $\bar{p}$ of 0's in a maximises $\nu_p^{(h)}\{a\}$ amongst the standard real numbers $p \in [0,1]$ for all infinte hypernatural h.*

Proof. Observe that

$$\begin{aligned} \forall a \in 2^{\mathbb{N}} \forall p \in [0,1] \quad f(a,h,p) &:= \mu_p \{a \restriction h\} \\ &= \prod_{i<h} B_p \{a(i)\} \\ &= p^{\mathrm{card}\{a(\cdot)=0\}} (1-p)^{\mathrm{card}\{a(\cdot)=1\}} \\ &= p^{\phi(a,h)} (1-p)^{h-\phi(a,h)}, \end{aligned}$$

where $\phi(a,h) = \sum_{i<h} a(i)$ is the absolute frequency of 0's in $a \restriction h$. Then by transfer

$$\begin{aligned} \forall b \in {}^*\left(2^{\mathbb{N}}\right) \forall p \in {}^*[0,1] \quad {}^*f(b,h,p) &= \mu_p \{b \restriction h\} \\ &= \prod_{i<h} B_p \{b(i)\} \\ &= p^{\mathrm{card}\{b(\cdot)=0\}} (1-p)^{\mathrm{card}\{b(\cdot)=1\}} \\ &= p^{\phi(b,h)} (1-p)^{h-\phi(b,h)}. \end{aligned}$$

In particular, for all $a \in 2^{\mathbb{N}}$, and $p \in {}^*[0,1]$

$$\begin{aligned} \nu_p^{(h)}\{a\} &= \mu_p \{{}^*a \restriction h\} = {}^*f\left({}^*a, h, p\right) \\ &= p^{\phi({}^*a,h)} (1-p)^{h-\phi({}^*a,h)}. \end{aligned}$$

where we do not distinguish between ${}^*\phi$ and ϕ. Now, for finite h and standard $a \in 2^{\mathbb{N}}, h \in \mathbb{N}$, one easily checks, via computing the derivative of the function $f(a,h,\cdot)$ as

$$(f(a,h,\cdot))' : p \mapsto p^{\phi(a,h)-1} (1-p)^{h-\phi(a,h)-1} \left(\phi(a,h) - p \cdot h\right),$$

that the (isolated) maximum of f on $[0,1]$ is attained when $p = \phi(a,h)/h$. Moreover, this is the only local extremum, and the farther p is from $\phi(a,h)/h$, the less is $f(a,h,p) = \nu_p^{(h)}\{a\}$. This is to say that the function $f(a,h,\cdot) : r \mapsto r^{\phi(a,h)} \cdot (1-r)^{1-\phi(a,h)}$, where

$$\phi(a,h)/h = \frac{1}{h} \sum_{i<h} a(i),$$

is strictly monotonely decreasing on $\left[\frac{\phi(a,h)}{h}, 1\right]$ and strictly monotonely increasing on $\left[0, \frac{\phi(a,h)}{h}\right]$. Note that all this can be expressed in the language of ordered fields: $<$ and $=$ are the only relations, and $+, -, \cdot, \div$ the only functions that need to be employed to formalise the last sentence. By transfer to the nonstandard universe, the assertion of the penultimate sentence therefore also holds for infinite hypernatural h and $a \in {}^*\left(2^{\mathbb{N}}\right)$: $\nu_{\cdot}^{(h)}\{a\} = {}^*f\left({}^*a, h, \cdot\right)$ strictly monotonely decreasing on $\left[\frac{\phi({}^*a,h)}{h}, 1\right]$ and strictly monotonely increasing on $\left[0, \frac{\phi({}^*a,h)}{h}\right]$.

Let us now fix such an $h \in {}^*\mathbb{N} \setminus \mathbb{N}$ and let us identify a with ${}^*a$ for $a \in 2^{\mathbb{N}}$. If we denote the limiting relative frequency of $a \in 2^{\mathbb{N}}$ by $\bar{p}$, i.e.

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i<n} a(i) = \bar{p},$$

we have – by the nonstandard characterisation of limits – the approximate equality

$$\phi(a,h)/h = \frac{1}{h}\sum_{i<h} a(i) \approx \bar{p}$$

(where h could be any infinite hypernatural number). Note that the function $f^{1/h}$ has the same monotonicity properties as f, but it is also *-differentiable with non-infinitesimal difference quotients $\frac{\left(\nu^{(h)}_{\frac{\phi(a,h)}{h}+\varepsilon}\{a\}\right)^{\frac{1}{h}} - \left(\nu^{(h)}_{\frac{\phi(a,h)}{h}}\{a\}\right)^{\frac{1}{h}}}{\varepsilon}$ and $\frac{\left(\nu^{(h)}_{\frac{\phi(a,h)}{h}}\{a\}\right)^{\frac{1}{h}} - \left(\nu^{(h)}_{\frac{\phi(a,h)}{h}-\varepsilon}\{a\}\right)^{\frac{1}{h}}}{\varepsilon}$ for non-infinitesimal $\varepsilon > 0$. (To see this, note that we have found an expression for $\left(\nu^{(h)}_p\{a\}\right)^{\frac{1}{h}}$ which only depends on the *relative* frequency $\frac{\phi(a,h)}{h}$ of 0's in the sequence, and on p; furthermore, this expression is continuous in $\frac{\phi(a,h)}{h}$, hence we can use the non-zero lower bounds on the absolute values of these difference quotients that we get for finite h also for hyperfinite h.) Therefore we can prove that whenever q is a standard real number and $\bar{p} = {}^\circ\frac{\phi({}^*a,h)}{h}$ (where ${}^\circ : {}^*\mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ denotes the standard part function), then $f^{1/h}(\bar{p}) > f^{1/h}(r)$, which immediately gives $f(p) > f(r)$.

□

C Solution to the zero-fit problem for a finite-state model without memory

We shall now keep the promise made in Section 3 and prove rigorously that in discrete memory-less models of reality, one can solve the zero-fit problem by means of infinitesimal probabilities.

Theorem C.1. *Consider any infinite hypernatual number h. Adopting this section's notation and considering an $a \in (N \times N)^\omega$, the function*

$$f := f(a,h,\cdot) : p \mapsto \prod_{n=0}^{h-1} p\left({}^*a(n), {}^*a(n+1)\right)$$

attains its maximum on the set

$$Q := \left\{ (p(i,j))_{i<N,j<N} \in [0,1]^{N\times N} \;:\; \begin{array}{ll} \forall j<N & \sum_{i=0}^{N-1} p(i,j) = 1, \\ \forall i<N & \sum_{j=0}^{N-1} p(i,j) = 1 \end{array} \right\} \subset \mathbb{R}^{N\times N}$$

(the set of transitions matrices) in $\left(q_{i,j}(a)\right)_{i,j<N}$.

Proof. In analogy to the proof of Theorem B.1, it suffices to prove that for finite $h > 0$, the function $f(a,h)$ has an isolated global maximum on Q in $\left(\frac{|\{\ell \le h \ : \ a(\ell)=i, \quad a(\ell+1)=j\}|}{|\{\ell \le h \ : \ a(\ell)=i\}|}\right)_{i,j<N}$, – which is nothing else but to say that when we truncate history after finitely many steps, the transition matrix p that fits the observation best will simply be the one obtained from evaluating the observation. A formal proof for this assertion can be given by introducing the abbreviations

$$\phi(i,a,h) = |\{\ell \le n \ : \ a(\ell) = i\}|$$

as well as

$$\phi(i,j,a,h) = |\{\ell \le n \ : \ a(\ell) = i, a(\ell+1) = j\}|$$

for all $i,j < N$ and then writing f as

$$\begin{aligned} & f(p) \\ = & \prod_{i<N-1, j<N-1} p(i,j)^{\phi(i,j,a,h)} \\ & \cdot \prod_{i<N-1} \left(1 - \sum_{j<N-1} p(i,j)\right)^{\phi(i,N-1,a,h)} \cdot \prod_{j<N-1} \left(1 - \sum_{i<N-1} p(i,j)\right)^{\phi(N-1,j,a,h)} \\ & \cdot \left(1 - \sum_{i<N-1} \left(\sum_{j<N-1} p(i,j)\right)\right)^{\phi(N-1,N-1,a,h)}, \end{aligned}$$

which is a function that only depends on the coordinates (i,j) with $i,j < N-1$ and therefore is well-defined as a function on $\mathbb{R}^{(N-1)\times(N-1)}$, as one can then, via computing the partial derivatives of the smooth function f, show that $f : [0,1]^{(N-1)\times(N-1)} \to [0,1]$ attains its global maximum on the compact set $[0,1]^{(N-1)\times(N-1)}$ in $\left(\frac{|\{\ell \le h \ : \ a(\ell)=i, \quad a(\ell+1)=j\}|}{|\{\ell \le h \ : \ a(\ell)=i\}|}\right)_{i,j<N-1}$, and that for all $i_0, j_0 < N-1$, the function $f\left(\left(\frac{|\{\ell \le h \ : \ a(\ell)=i, \quad a(\ell+1)=j\}|}{|\{\ell \le h \ : \ a(\ell)=i\}|}\right)_{i,j<N-1, i\ne i_0, j\ne j_0}, \cdot\right)$ will be strictly monotone on each of the intervals $\left[\frac{|\{\ell \le h \ : \ a(\ell)=i_0, \quad a(\ell+1)=j_0\}|}{|\{\ell \le h \ : \ a(\ell)=i_0\}|}, 0\right)$ and $\left(\frac{|\{\ell \le h \ : \ a(\ell)=i_0, \quad a(\ell+1)=j_0\}|}{|\{\ell \le h \ : \ a(\ell)=i_0\}|}, 0\right]$.

Since this is a first-order predicate logic expression in the language of ordered fields, we can use the transfer principle to obtain that the same assertion also holds for infinite hypernatural h and the corresponding function $f : {}^*[0,1]^{(N-1)\times(N-1)} \to {}^*[0,1]$. But then $f^{1/h}$ must have the same monotonicity properties and, what is more, it will be *-differentiable with non-infinitesimal difference quotients outside the monad of $\left(\frac{|\{\ell \le h \ : \ a(\ell)=i, \quad a(\ell+1)=j\}|}{|\{\ell \le h \ : \ a(\ell)=i\}|}\right)_{i,j<N-1}$. Hence whenever $r \in [0,1]^{(N-1)\times(N-1)} \cap \mathbb{R}^{(N-1)\times(N-1)}$ with

$$r \ne {}^\circ \left(\frac{|\{\ell \le h \ : \ a(\ell) = i, \quad a(\ell+1) = j\}|}{|\{\ell \le h \ : \ a(\ell) = i\}|}\right)_{i,j<N-1} =: \bar{p} \in \mathbb{R}$$

(${}^\circ : {}^*\mathbb{R}^{(N-1)\times(N-1)} \to (\mathbb{R} \cup \{\pm\infty\})^{(N-1)\times(N-1)}$ denoting the multidimensional standard part function), we will have that $f^{1/h}(\bar{p}) > f^{1/h}(r)$ and therefore $f(\bar{p}) > f(r)$.

However, due to the nonstandard characterisation of limits, for all $h \in {}^*\mathbb{N}\backslash\mathbb{N}$,

$$\begin{aligned}\bar{p} &= {}^\circ\left(\frac{|\{\ell \leq h \ : \ a(\ell) = i, \quad a(\ell+1) = j\}|}{|\{\ell \leq h \ : \ a(\ell) = i\}|}\right)_{i,j<N-1} \\ &= \lim_{n\to\infty}\left(\frac{|\{\ell \leq n \ : \ a(\ell) = i, \quad a(\ell+1) = j\}|}{|\{\ell \leq n \ : \ a(\ell) = i\}|}\right)_{i,j<N-1} = (q_{i,j}(a))_{i,j<N-1}\,.\end{aligned}$$

Thus we finally arrive at $f(r) < f\left((q(a)\right)$ for all $r \neq q(a)$. □

D An additional note on "Integrating over densities"

Elga's opposition to the generalisation of a likelihood approach to the zero-fit problem for countable coin tossing sequences is based on the assumption that, in Elga's notation, the subset $L_x \subset 2^\omega$ of all those sequences where the limiting frequence of tails (or heads) is x should be measurable with respect to any power of Bernoulli measures on 2^ω. Hence, in our set of possible worlds 2^ω, it should be considered an event, i.e. a measurable[19] set, that the limiting frequency equals such a number $x \in [0,1]$. In case there is any doubt about this,

$$L_x = \bigcap_{\varepsilon\in\mathbb{Q}_{>0}} \bigcup_{N\in\omega} \bigcap_{n\geq N} \left\{ \left| \frac{1}{n}\sum_{i=0}^{n-1} \chi_{\pi_i{}^{-1}\{0\}} - x \right| \leq \varepsilon \right\},$$

if $\pi : 2^\omega \to 2$ is the projection on the i-th toss.

E Well-defined continuous versions of density functions

We shall now prove that if we make the topological assumption on the probability space Γ described in Section 6, then a continuous version of a density function (if existent) will be well-defined everywhere on the direct sum of homogenous measure spaces associated with Γ.

So let us assume that the probability space $(\Gamma, \mathcal{G}, \nu)$ (whose measure algebra is isomorphic to $\Omega := \sum_{\kappa\in I} \alpha_\kappa \cdot [0,1]^\kappa$) will be an "ordinary" probability space in the sense of Fajardo and Keisler's [6], that is a separably metrisable complete space with ν being the completion of a probability measure on the Borel σ-algebra of Γ. In that case the Maharam spectrum will only consist of finite and countable cardinals[20], thus the Maharam spectrum will be a subset of $\omega + 1$ (and therefore it cannot be the direct sum of a set of probability spaces that

[19] Maybe it is worth reminding the reader that measurability cannot always be expected from an arbitrary subset of a measure space. Assuming the Axiom of Choice, it is for example by no means necessary that every subset of the reals is Lebesgue-measurable, as one knows from the standard example of the Vitali set. Although the Vitali set is "constructed" by means of the Axiom of Choice, it was not clear until the late sixties whether it would be possible to obtain such an object without use of the Axiom of Choice. Solovay showed, using forcing techniques, in the late Sixties that this is impossible [24]. Both [10] and [12] contain more detailed, as well as historical, information on the so-called "measure problem".

[20] To see this, first recall that according to Fajardo and Keisler [6, Theorem 3B.7] the Maharam spectrum of a saturated probability space is a set of uncountable cardinals. However,

contains a Loeb space[21]). Due to separability, this is a huge advantage, as. Let us now assume without loss of generality that our model was built on Ω (instead of Γ) in the first place.

The topology with respect to which our density function should be continuous is the one inherited from the Maharam spectrum. Starting from the order topology on $[0,1]$, we obtain – by choosing the coarsest topology such that projections are continuous – a topology on $[0,1]^\alpha$ for all α in the Maharam spectrum $I \subset \omega+1$, enabling us to endow $\sum_{\kappa\in I} \alpha_\kappa \cdot [0,1]^\kappa$ with a topology. Under this construction, we can now employ the assumption $I \subset \omega+1$ to see that all continuous functions $f : [0,1]^\alpha \to \mathbb{R}$ will be measurable, since all closed sets in $[0,1]^\omega$ will be measurable[22]. Continuing to assume as little topological measure theory as possible on the part of the reader, we now remark that two continuous functions $f_0, f_1 : [0,1]^\omega \to \mathbb{R}$ which we know to agree on all of $[0,1]^\kappa$ except possibly on a null set will in fact be equal[23]. Thus, if a density function on the direct sum Ω of homogeneous measure spaces associated with Γ has a continuous version, it is going to be well-defined everywhere on Ω.

Acknowledgements. Professor Adam Elga's article was communicated to the author by Markus Schrenk. The author gratefully acknowledges financial support from the German National Academic Foundation (*Studienstiftung des deutschen Volkes*) and a PhD grant from the German Academic Exchange Service (*Doktorandenstipendium des Deutschen Akademischen Austauschdienstes*). Last not least, the author would like to thank Adam Elga as well as two anonymous referees for tremendously helpful comments that enhanced both the content and presentation of this paper.

References

[1] S. Albeverio. Some personal remarks on nonstandard analysis in probability theory and mathematical physics. In M. Mebkhout and R. Sénéor, editors, *VIIIth International Congress on Mathematical Physics – Marseille 1986*. World Scientific, 1987.

[2] S. Albeverio, R.J. Høegh-Krohn, J.E. Fenstad, and T. Lindstrøm. *Nonstandard methods in stochastic analysis and mathematical physics*, volume 122 of *Pure and Applied Mathematics*. Academic Press, Orlando, FL, 1986.

any direct sum of a set of probability spaces containing a universal (e.g. saturated, cf. [6, Theorem 3A.3(a)]) space is itself universal. Hence, as ordinary probability can never be universal [6, Theorem 1E.11, Lemma 1E.6], their Maharam spectra must be sets of finite or countable cardinals.

[21] The Maharam spectrum of a Loeb space is, according to Jin and Keisler [11], always a set of uncountable cardinals.

[22] Suppose a closed set $A \subset [0,1]^\omega$ was not measurable, thereby assuming the existence of an $i < \omega$ such that $\pi_i(A)$ is not measurable; because of Tikhonov's Theorem (which is equivalent to the Axiom of Choice), $[0,1]^\omega$ is compact, thus so is A and its image under the continuous map π_i must be compact, too for all π_i – a contradiction.

[23] Note that $\{f_0 \neq f_1\} = \{f_1 - f_0 > 0\} \cup \{f_1 - f_0 < 0\}$ is open. On the other hand, the topology of $[0,1]^\omega$ is generated by the inverse images under projections of open subsets of $[0,1]$, and – since nonempty open subsets of $[0,1]$ all have positive measure – these have positive measure. However, the intersection of n such sets will then, after an appropriate rearrangement of the coordinate axes, either be empty or the inverse image under the projection onto $[0,1]^n$ of an open subset of $\mathbb{R}^n$ – again a set of positive measure. Hence, all open nonempty subsets of $[0,1]^\omega$ have positive measure. Exploiting the openness of $\{f_0 \neq f_1\}$, we get that $f_1 = f_0$ on the whole space $[0,1]^\omega$.

[3] R.M. Anderson. A non-standard representation for Brownian motion and Itô integration. *Israel Journal of Mathematics*, 25(1-2):15–46, 1976.

[4] D.R. Cox and D.V. Hinkley. *Theoretical statistics*. Chapman and Hall, London, 1974.

[5] A. Elga. Infinitesimal chances and the laws of nature. *Australasian Journal of Philosophy*, 82(1):67–76, 2004.

[6] S. Fajardo and H.J. Keisler. *Model theory of stochastic processes*, volume 14 of *Lecture Notes in Logic*. A.K. Peters, Natick, MA, 2002.

[7] P.R. Halmos. *Measure theory*, volume 18 of *Graduate Texts in Mathematics*. Springer, Berlin, 1974.

[8] D.N. Hoover and E. Perkins. Nonstandard construction of the stochastic integral and applications to stochastic differential equations. I. *Transactions of the American Mathematical Society*, 275:1–29, 1983.

[9] D.N. Hoover and E. Perkins. Nonstandard construction of the stochastic integral and applications to stochastic differential equations. II. *Transactions of the American Mathematical Society*, 275:30–58, 1983.

[10] Th. Jech. *Set theory. The third millennium edition*. Springer Monographs in Mathematics. Springer, Berlin, 2000.

[11] R. Jin and H.J. Keisler. Maharam spectra of Loeb spaces. *Journal of Symbolic Logic*, 65(2):550–566, 2000.

[12] A. Kanamori. *The higher infinite*. Springer Monographs in Mathematics. Springer, Berlin, 2 edition, 2003.

[13] V. Kanovei and S. Shelah. A definable nonstandard model of the reals. *Journal of Symbolic Logic*, 69(1):159–164, 2004.

[14] H.J. Keisler. An infinitesimal approach to stochastic analysis. *Memoirs of the American Mathematical Society*, 297, 1984.

[15] H.J. Keisler. Infinitesimals in probability theory. In N. Cutland, editor, *Nonstandard analysis and its applications (Hull, 1986)*, volume 10 of *London Mathematical Society Student Texts*, pages 106–139. Cambridge University Press, Cambridge, 1988.

[16] D. Lewis. Humean supervenience debugged. *Mind*, 103(412):473–490, 1994.

[17] T. Lindstrøm. An invitation to nonstandard analysis. In N. Cutland, editor, *Nonstandard analysis and its applications (Hull, 1986)*, volume 10 of *London Mathematical Society Student Texts*, pages 1–105. Cambridge Univ. Press, Cambridge, 1988.

[18] P.A. Loeb. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, 211:113–122, 1975.

[19] D. Maharam. On homogeneous measure algebras. *Proceedings of the National Academy of Sciences of the United States of America*, 28:108–111, 1942.

[20] E. Nelson. *Radically elementary probability theory*, volume 117 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1987.

[21] E. Perkins. A global intrinsic characterization of Brownian local time. *Annals of Probability*, 9:800–817, 1981.

[22] E. Perkins. Stochastic processes and nonstandard analysis. In *Nonstandard analysis—recent developments (Victoria, B.C., 1980)*, volume 983 of *Lecture Notes in Mathematics*, pages 162–185. Springer, Berlin, 1983.

[23] A. Robinson. *Non-standard analysis*. North-Holland, Amsterdam, 1966.

[24] R.M. Solovay. A model of set theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics*, 92(1):1–56, 1970.

[25] K.D. Stroyan and J.M. Bayod. *Foundations of infinitesimal stochastic analysis*, volume 119 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam, 1986.