Explanatory Chance

msuareza@ucm.es

Versions #1: September 2017
#2: July 2019
(9,330 words)

Acknowledgements: For comments I want to thank audiences at Paris (IHPST Panthéon-Sorbonne, 2017 and 2018), Turin (conference on the models of explanation in 2018), and Pittsburgh (HPS annual lecture series 2018-9). Financial support is acknowledged from the Spanish government projects FFI2014-57064-P and PGC2018-099423-B-100.

Explanatory Chance

Abstract: Debates in philosophy of probability over the nature and ontology of objective chance by and large remain inconclusive. No reductive account of chance has ultimately prospered. This article proposes a change of focus towards the functions and roles that chance plays in our cognitive practices. Its starting philosophical point is pluralism about objective probability. The *complex nexus of chance* is the interlinked set of roles in modelling practice of i) parametrized probabilistic dispositions ("propensities"); ii) distribution functions ("probabilities"); and iii) statistical finite data ("frequencies"). It is argued that the modelling literature contains sophisticated applications of the chance nexus to both deterministic and indeterministic phenomena. These applications may be described as lying on a spectrum between what I call 'pure probabilistic', and 'pure stochastic' models. The former may be found in the tradition of the method of arbitrary functions; the latter in present-day techniques for stochastic modelling in the complex sciences, as well as some orthodox approaches to quantum mechanics. These modelling practices provide positive arguments for the irreducible complexity of the chance nexus.

1. Interpreting and Applying Chance

Philosophers of probability have discussed the issue of the reality of chance or objective probability extensively. The discussion has often been framed as part of a debate or dispute about the ontology and epistemology of chance. Realists have tended to focus on issues of metaphysical constitution, and semantic reference; anti-realists have been concerned with evidence and epistemic accessibility. Hence empiricist-minded philosophers of science have attempted to reduce objective chances to frequencies, or ratios of observable outcomes in experimental sequences of events. By contrast, metaphysically minded philosophers have attempted to interpret chance in light of propensities or dispositional properties. In either case the assumption is that chance is an obscure - or at any rate contested - concept that must be defined in terms of other simpler, more fundamental, accessible, or acceptable concepts. While chances are not supposed to be accessible, frequencies are meant to be directly accessible through observation. And while chances are not supposed to be real and fundamental, dispositional properties are understood to be genuine and fundamental properties of chance set-ups. It makes sense from the perspective of this debate about the ontology and epistemology of chance to attempt to reduce chance – objective probability – to either frequency or propensity.

Nevertheless, as is by now well known, both the frequency and the propensity interpretations of objective probability are ultimately unviable. The former encounters insurmountable difficulties associated to the so-called reference class problem; while the latter confronts a family of problems related to the notorious Humphreys' paradox. In the next section of this paper I briefly review these conclusive arguments. In addition, I present a further explanatory argument that cuts against both frequency and long run propensity interpretations. Chances are employed in practice for explanatory purposes, in science and in ordinary life alike; but I shall argue that this explanatory function remains elusive on any interpretation of objective probability that identifies it with ratios in sequences of experimental outcomes. The conclusion of this section is that it is time to consider alternative philosophical projects in our understanding of chance, going beyond sterile ontological disputes about its 'interpretation'.

In the remaining sections of the paper I consequently explore a different philosophical approach, which focuses instead on the role that chance plays in scientific modelling practice. In a brief slogan I aim to shift philosophical attention from the 'metaphysics and epistemology of chance' towards the 'methodology of chance modelling'. In particular I shall argue that chance plays an essential explanatory role in that practice, which already militates against any reductive analysis. Section 3 outlines chance pluralism in the form of what I call the nexus of chance: The tripartite conceptual distinction between propensities, probabilities, and frequencies. Let it be noted outright that this methodological turn does not resolve or finish off ontological or epistemological disputes in the metaphysics of chance literature. On the contrary, it is not intended to foreclose any option in the ontology of chance, and it certainly does not resolve or alleviate epistemic worries or concerns, which are likely to endure. For while realists may want to take the explanatory power of chance as further evidence in favour of the reality of chances, anti-realists will deny the inference from the explanatory power of the chance nexus to its reality. I do not here pretend to resolve such quarrels. Rather, my aim is more modestly to show that the epistemological debate goes under different terms in this new methodological arena, one that bypasses the ontological and epistemological disputes. What I propose is a new take on the question over the nature of chance, one that starts from the standpoint of methodological practice, and which raises problems and issues of its own – including possibly new versions of the perennial debates in the ontology and epistemology of chance.

Indeed, in the main sections 4 and 5 I argue that once this methodological outlook is adopted, new and interesting philosophical questions arise for a philosophical study of chance. I will focus in particular on the role that dynamical equations play in stochastic modelling, and I shall argue that all stochastic or statistical modelling in the natural sciences can be placed on a spectrum that goes from purely deterministic to purely indeterministic dynamics. Most models will use a mixture of both, and the regular assumption that all stochastic modelling must invoke a thoroughly stochastic dynamics is disproven by the fact that many statistical models for macroscopic phenomena assume underlying deterministic dynamics. In the distinguished tradition of the method of arbitrary functions, which originates in the writings of Von Kries and Poincaré at the turn of the 19th century, the stochasticity is brought about by the ingenious way in which a deterministic dynamics (in the traditional, Newtonian sense) can lead from a probability density or distribution defined over initial micro-variables to a probability distribution over the relevant macro-variables at the end of the process. I call this kind of modelling purely probabilistic, since the probabilities that it prescribes do not originate in the dynamics in or by itself. In contrast, what I call purely stochastic modelling requires no initial probability distributions over any dynamical variables – the final probability distributions arise naturally out of the dynamics. Most statistical modelling, I argue, is <u>impure</u>, i.e. it is neither pure probabilistic nor pure stochastic, but rather a mixture of both, and thus lies somewhere along the spectrum.

The concluding section 6 of the paper wraps things up by pointing out that the tripartite account of the nexus of chance is in some ways a trivial consequence of the application of general philosophical lessons regarding the modelling attitude or methodology in general to the particular case of statistical modelling. There is little mystery to chance – to the ways in which the chance nexus is employed in practice – beyond whatever mysteries remain in scientific modelling in general.

2. Two Projects in Ontology Revisited

The philosophy of probability throughout the 20th century has been centrally concerned with providing an appropriate semantics for objective statements of probability, or chance. This in turn has been seen as requiring a description of what the world may be like for such statements to be true. And such a description at the very least requires the provision of some ontology for chance, or more technically, a stipulation of the truth-makers for our probability statements – those things in the world (objects, properties, facts, events) in virtue of which our probability statements can be true or false. There have two broad approaches to this question throughout the 20th century, answering to a markedly empiricist school earlier on, and to more broadly realist leanings later on.

2.1. The frequency interpretation

The empiricist tradition, in particular the logical empiricists, understood an appropriate semantics for probability statements to entail reductions to observable facts or events - ascertainable directly by inspection or observation. The most likely candidate for such reduction would – in Laplacean fashion – take the form of ratios of outcome-types, or attributes, in regular series or successions of observable events in repeatable experiments. Thus, Von Mises (1928) and

Reichenbach (1934/49) formulated empiricist interpretations of probability by insisting that the range of possible cases be observable outcomes in repeatable sequences of experimental trials. This is sometimes known as the "frequency interpretation of probability", and it takes the form of some sort of conceptual identity along the lines of: "P (A) is the probability of outcome A if and only if there is an appropriate sequence S of outcomes such that P (A) is the frequency or ratio of outcomes of type A in S."

I shall refer to this generic statement as the *frequency identity* of probability. Hence, for example, P (heads) = ½ is the probability of heads in a coin tossing experiment if and only if the ratio of heads to tails in the appropriate sequence S is exactly ½. It is easy to see how the *frequency identity* would generalise to more complex discrete or continuous probability distributions over a larger set of possible outcomes (a larger "outcome space"). Thus, a probability of some attribute in some population can easily be identified with a frequency ratio of the attribute in a representative sample of the population.

The strategy seems at first sight the most natural way to deliver us from any unverifiable metaphysical commitments often understood to be the holy grail of any empiricist philosophy. Thus, one traditional goal of empiricism ever since Hume, if not before, has been to analytically reduce unverifiable statements about unobservable or inaccessible entities or matters of fact, so as to "transform" them, into verifiable statements about observable, or at any rate accessible, matters of fact. The empiricist tradition has attempted such reductions on problematic concepts such as lawhood (often, as in Mill or Mackie identified with nomological regularity or non-accidental generalization); causation (which at least since Hume has been thought to be reducible to regular continuous succession, or a projection thereof); psychological time and personhood, etc. In the context of chance and probability this has often translated into a requirement to express probability statements as claims regarding series or sequences of events that can be verified if not in practice at least in principle; and the frequency interpretation seems to readily deliver on just such a requirement.

However, the "frequency interpretation" is not really one single interpretation but a family of interpretations, generated by diverse renditions of the frequency identity. More precisely, all frequency interpretations obey the *frequency identity* as expressed above, but they differ as to what they take to be the 'appropriate' sequences in its defining statement. Very generally, we may classify frequency interpretations into two families: the finite frequency (FF), and the hypothetical frequency (HF) interpretations. Roughly, finite frequencies are ratios of outcomes endowed with the given attribute (A) in the *actual* (and hence necessarily finite) frequencies of experimental outcomes of real experiments performed on chance systems or set ups. The FF interpretations have thus the virtue of reducing probability to an empirically accessible quantity. Since most if not all frequency interpretations are motivated by empiricism, this is clearly perceived to be an advantage. There is no doubt, on any FF interpretation, that probability is epistemically accessible, since it is constituted by events that can be directly ascertained by the senses. And it can be empirically or experimentally measured, as long as the populations that it pertains to are themselves accessible, and the different attributes can be empirically distinguished by inspection.

Nevertheless, FF interpretations have severe problems or deficiencies; I will emphasise here only the two problems that are most relevant to my purposes (the reader can find a full list of arguments against FF interpretations in Hajek, 1997). First of all, it is clear that for any actual finite sequence, no matter how large, the ratio of the appropriate attribute can in fact diverge from the probability. One need not consider weird situations as those described in Tom Stoppard's play "Rosencrantz and Guildenstern are Dead" (in which the characters repeatedly toss a coin that increasingly unnervingly always falls on heads). For a coin toss, any given odd numbered finite frequency will necessarily diverge, however minimally, from ½. And we all intuitively understand that it is perfectly *possible* for any frequency, no matter how large, to diverge (and even to diverge maximally, as in the weird Rosencrantz and Guildenstern scenario). This sense of 'possible divergence' is built into the very judgement of the probability of any event in a series, as long as any outcome event is genuinely independent of any other (i.e. in the coin case: as long as the outcome at each toss does not alter the probability of any outcome at any later toss).

The capacity of any finite frequency to diverge from the probability it is intended to analytically reduce comes under a variety of names in the literature. I shall refer to it here as "frequency divergence": for every finite frequency exhibited in a regular experimental trial, there are myriad ways in which it may diverge from the underlying probability it can at best approximate. There are purposes in the study of frequencies for which frequency divergence comes in handy, such as in assessing or ranking the 'faithfulness' or 'representativeness' of possible frequencies in sequences. However, as regards the finite frequency (FF) version of the frequency identity, it is ultimately lethal. For if any finite frequencies can as a matter of principle diverge arbitrarily from the probabilities they are supposed to conceptually reduce, then the frequency identity is surely false: there can be no sequence S that exhibits the appropriate frequency, with any certainty, and this would entail that there is not really a probability P (A) of the attribute in question, contrary to the assumption (of divergence). Hence finite frequencies do not analytically reduce probabilities. Or, to put it conversely, probabilities are in themselves never finite frequencies, contrary to what strict empiricism would require.

2.2. The propensity interpretation

An obvious way to circumvent these objections is thus precisely to give up on the strict empiricist commitments, and to adopt a realist interpretation of objective probability instead – in terms of propensities. Propensity interpretations themselves come in a considerable variety. For example, it is customary at least since Gillies (2000) to distinguish *long-run* from *single-case* propensities. In a single case version of the propensity interpretation, the underlying propensity is manifested in every single run of the experiment as a probability P (A). In the long run version of the propensity interpretation, by contrast, the function P (A) is rather identified with the long run frequency – thus Gillies (2000, ch. 8), asserts the single case probabilities can only if anything be subjective. One may then substitute a propensity identity in place of the frequency identity, roughly as follows: "P (A) is the probability of some event type A if and only if A is a possible outcome of a chance set up S endowed with a certain propensity P (A) to generate outcome A in the long run."

However, any propensity interpretation - of either variety - that adopts a strict identity between probabilities and propensities fails too for reasons that have been explored extensively in the literature (Eagle (2004), Humphreys (1985), Salmon (1979), Suárez (2013)). First of all, it is well-known that many well-defined objective probabilities cannot be identified with any propensities. In fact, many conditional probabilities that have a straightforward propensity interpretation also often have well-defined inverse conditional probabilities that fail to have any propensity interpretation. Salmon's (1979) original example involved shooting's propensity to kill, which precludes any interpretation of killing as a propensity to shoot. But one can think of a myriad other examples: smoking has a certain propensity to produce lung cancer, while lung cancer does not have a propensity to generate smoking – yet, for any control population, if the conditional probability of lung cancer given smoking is well defined, so is the inverse conditional probability of smoking given lung cancer. And so on.

The underlying problem is that propensities exhibit an asymmetry akin to cause and effect, and this is an asymmetry lacking in probabilities. Any 'propensity identity' that identifies the two will ensue in contradiction: this shows that they cannot be the same thing. Humphrey's paradox (Humphreys, 1985) provides the definitive objection, since it shows that many *bona fide* propensities are not interpretable or identifiable with conditional probabilities, on pain of contradiction with the Kolmogorov classical calculus. There are a number of possible resolutions to this "paradox", but they all involve giving up the propensity identity in some respect or other. As has been shown elsewhere the only rendition of the propensity view that ultimately works is not an identity or analytical reduction of any sort. Thus a defensible statement relates propensities to probabilities for outcomes in experimental trials or set ups, but it does not identify them: "P (A) is the objective probability or chance of outcome A if and only if A is produced by a chance set up S endowed with a certain propensity to generate each A with some probability P (A)." In this statement, a propensity is ascribed to a set up when an objective probability obtains for some outcome of that set up – yet the propensity and the probability are not identified, but rather kept entirely distinct.

We may conclude that the propensity identity is a flawed but necessary presupposition in a long run version of the propensity interpretation. It is also commonly adopted for single case versions, but it does not turn out to be in fact necessary for a single case propensity view. On the contrary, as shall be pointed out in section 3, there are notions of single case propensity that do not entail or require any identification of probability with propensity.

2.3. The explanatory argument

There is a further argument against any analytical reduction of probability by means of any 'identity thesis'. It is related to the explanatory power of chances or objective probabilities – so we may refer to it as the 'explanatory argument'. The point is best made in the context of attempts to reduce probability to frequency by means of the 'frequency identity' (although it applies to a 'propensity identity' too). Objective probabilities in practice often explain regular occurrences of types of outcomes in different kinds of sequences. For instance, we explain the relative frequencies of a game of roulette, or dice, in virtue of the chances that are presumably operative in the game in question. If you ask me why I got 5 heads and 5 tails in tossing a coin, I can legitimately offer the explanation that it is a fair coin, i.e. that it is built so as to display such a probability. More generally, science will often invoke theoretically grounded probabilities in the explanation of observed frequencies. The difference between the observed decay rates of two pieces of radioactive material may be explained by reference to the half-life of each of them, that is by noting their different atomic structure. The difference between the recovery rates of two sorts of patient afflicted by the same condition may be explained by reference to the efficiency of the different kinds of treatment they have been subjected to. Etc.

This is a basic explanatory fact, which I will at this point in the argument take as primitive: probabilities are often invoked and used – in ordinary cognition and scientific practice alike – in order to explain frequencies. Yet, if the FF interpretation is correct, probabilities are frequencies, and it is impossible for a frequency to explain itself. It may be objected that the frequency explained is not the frequency involved in the explanation, so that the situation is not as blatantly circular as it may at first seem. Or perhaps not all frequencies are explanatorily on a par, but as long as we restrict ourselves to finite frequencies (as the FF interpretations do), it is very hard to see what the explanatory power of some frequencies over others could be. They are after all just the same *kind* of thing, and explanatory power requires some distinct property to be doing the explaining. It cannot serve to explain the actual finite frequency ratio in a chancy experiment to merely point out to another actual finite frequency ratio in that experiment: to do so would seem to merely expand the demand for explanation.

It may be thought at this point that the finiteness of the frequencies is generating the explanatory circularity. Finite frequencies (FF) may be expanded into what is known as hypothetical frequencies (HP). An HP interpretation will identify probabilities with hypothetical frequencies over ideal infinite sequences of experimental outcomes that contain the appropriate finite sub-sequences whose properties are to be explained. For instance, in the case of a tossing coin, the appropriate frequency that identifies the objective probability or chance is supposed to be only definable in the abstract as the limiting frequency lying at the limit of the hypothetical infinite sequence of tosses. Presumably if the coin is fair the limiting frequency in the hypothetical infinite sequence is precisely ½. This may answer the first objection in section 2.1. from frequency divergence, since any finite frequency is allowed to diverge from the much larger (hypothetical and infinite) frequency. The large number theorem shows that the degree of divergence is inversely proportional to the length of the sequence, or in other words the finite frequencies will approach the limiting frequency as the finite sequence grows – and the finite frequency will become the actual probability in the infinite limit. However, that is just another way to concede that for any finite frequency, no matter how large, there will always be a degree of divergence.

The move to HF interpretations is not really successful in resolving the problems generated by the reference class problem, and in fact raises additional and important difficulties. There are at least two reasons why HF interpretations fail. First of all, for any given finite frequency there is a large number of consistent hypothetical frequencies, since for any finite subsequence, there are a large number of sequences that would include the subsequence as their initial segment. The large number theorem is no retort, since it presupposes that there is an actual probability, and then goes on to show that the limiting frequency will arbitrarily approach it. However, the point of a frequency interpretation of probability is not to *presuppose* the existence of an actual probability that frequencies can be shown to approach in the infinite limit. The point of any frequency interpretation is to *identify* the probability itself *as the frequency* in accordance with the *frequency identity* discussed above. In other words, the large number theorem cannot really help to define probabilities as limiting frequencies in hypothetical sequences – rather the theorem only works on the assumption that there are probabilities independent of any frequencies, or their limiting character.

At any rate, the second, additional, explanatory argument remains; for the explanatory power of a frequency in a *hypothetical* sequence remains elusive. If the explanatory power relies on merely subsuming the finite sequences whose frequencies are to be explained within the hypothetical sequences that are to explain them, we have the recurrent problem above with FF interpretations: we seem to have merely expanded the demand for explanation. If on the other hand the appeal is to antecedent explanatory probabilities, as in the large number theorem, we restore the explanatory power but at the expense of postulating probabilities over and above any frequencies, contrary to the frequency identity.

3. Pluralism about Chance

The failure of reductionist projects suggests pluralism regarding chance. Rudolf Carnap (1945), Frank Ramsey (1928), and Ian Hacking (1975), amongst others, already argued that we must carefully distinguish objective probability (chance) from subjective probability (credence, or partial logical entailment). Carnap moreover argued that the conflation of these forms of probability leads to contradictions, confusions, and / or paradoxes, which only the correct formal explication of the concept would be able to resolve. Ramsey and Hacking did not explicitly embrace such hopes of co-existence, but all their views are united in the rejection of the view that there is one single thing that probability is – or measures. Thus, all these authors emphasise distinct uses and historical origins in subjective and objective probabilities[1]. This is not yet pluralism about chance or objective probability, but it suggests that such a pluralism is one natural step on the pluralist road.

In scientific modelling practice, chance appears as a dynamical nexus of properties that typically includes probabilistic dispositions or propensities; distribution functions or formal probabilities; and frequencies in actual or imagined data. In adopting the terminology of 'propensities' as the dispositional properties of chance setups that ground probabilities (Suárez, 2017; see also Mellor, 2005), the view moves decisively away from long run propensity theories that ultimately identify propensities with either finite, infinite, or hypothetical frequencies. Instead it proposes a sui generis explanatory relation between propensities and probabilities, which is in some cases a relation of cause and effect - although not necessarily always so. In distinguishing formal probability functions from either propensities or frequencies, any interpretational identification of the former in terms of the latter is precluded. (Thus, the nexus of chance is compatible with views that reject any need to interpret the concept of probability, such as Sober, 2010). Finally, in emphasising the role of finite experimental frequency

[1] This glosses over the undeniable differences between Ramsey's, Carnap's and Hacking's accounts of probability.

data, the nexus of chance restores an empiricist outlook, which grounds probability statements in the actual data collected in genuine experimental and observational contexts. On this view there is no need for any recourse to hypothetical frequencies or their presumed limiting character: the only frequencies are proportions or ratios of outcome-types within the actual sequences of experimental outcome events.

While the nexus of chance has been defended already at a theoretical level (Suárez, 2017; see also Mellor, 2005, and Humphreys, 1989, for related views), it remains to be studied at the level of modelling practice. This paper is a first attempt to establish the elements of a methodological research programme into the workings and operations of the tripartite nexus of chance. The distinctions that I present here, while preliminary, are grounded in modelling practice, and have application to a range of cases in the natural sciences. Together, they amount to a research programme into the methodological foundations of chance-based explanations in scientific practice.

4. The Nexus of Chance in Action

The modelling practice that I would like to focus on in this paper is known as *statistical modelling*. This type of modelling involves not merely formal descriptions of correlation phenomena, but it is typically employed with explanatory purposes. In other words, the typical explananda are already prepared descriptions of statistical correlation phenomena between a set of inter-related 'observable' variables – which may indeed be observational variables in a data model, but may also represent properties of an underlying phenomenon in a controlled experiment, or the results of various interventions carried out in laboratory conditions. The basic explanatory tool in chance explanations of such statistical phenomena is then a model featuring probabilities evolving in accordance to some dynamical law.

4.1. Parameter and Sample Spaces

The probabilistic models that ensue deserve some attention. A common assumption in the philosophical literature is to suppose that a *probability model* is simply a probability distribution function defined over the observable variables. To take the common - and apparently most simple - illustration of the coin toss: if a coin is tossed repeatedly, under identical conditions, the series of outcomes would constitute the observable data. Suppose the finite data exhibits a 47% incidence of heads and 53% incidence of tails. A *probability model* is then, in accordance to this common view, an ascription of a probability distribution that can account for, or make sense of, this distribution. It is obvious that a 47-53% probability distribution is the one that best accounts for, and makes sense of, this distribution, but others may do too within acceptable margins of experimental and systematic error. (What 'acceptable' margins of error are is an eminently relevant question, and the object of considerable debate.)

I don't have any fundamental quarrel with this simple definition of a *probability model* – as long as it is clearly understood that it is not the same notion as the more sophisticated *statistical model* that statisticians and scientists use in their everyday modelling practice. To illustrate the difference, with regards to coin tossing again, it is worth considering what the model of a fair coin would be. On the philosophical notion of a probability model, this could only be the ascription of a flat probability distribution ρ (i.e. equal 50-50 probabilities) over the head (h) and tail (t) outcome events: $\rho: \{h, t\} \longrightarrow \left\{\frac{1}{2}\right\} \in \mathbb{R}$. But as we shall see, the statistical model of the phenomenon of coin tossing, even for a fair coin, turns out to be a much more complex and interesting entity.

There are two critical differences between *probability* and *statistical* models, as I shall present them, and they have significant philosophical implications. The first distinguishing feature of a statistical model is that it is not a single probability distribution function but a *parametrized family of functions*, in a sense to be specified. The second distinguishing feature is that a statistical model is dynamical in a way that a probability model need not be. In other words, as we shall see, the probability functions in a statistical model either evolve in time, apply to different stages of a dynamical process, or both evolve and apply at different stages in a dynamical process. It is the conjunction of these two distinguishing features (multiple parametrization, dynamics) that endows a statistical model with explanatory power. I shall argue that the tripartite nexus of chance provides an understanding – if not the only understanding available – of the explanatory role of statistical models.

As regards the first distinguishing feature, an influential article by McCullough (2002, p. 1225) expresses it aptly as follows: "A statistical model is *a set of probability distributions* on the sample space $\mathcal{S}$" (my italics). What is more relevant is that a statistical model requires an antecedent parametrization of the phenomenon, that is: a typically dynamical description of the phenomenon under some set of parameters. It is only once the phenomena to be modelled is so described that a properly parametrized statistical model can be provided for it, by ascribing to each parameter a distinct probability function over the sample space: "A parametrized statistical model is a parameter set Θ *together with* a function $P\colon \Theta \longrightarrow \rho(\mathcal{S})$, which assigns to each parameter point $\theta \in \Theta$ a probability distribution P_θ *on* $\mathcal{S}$" (ibid, p. 1225). Hence a statistical model is most abstractly defined as a function that ascribes to a specific element in some antecedent parametrization of the phenomenon a probability function from a family defined over the sample space. In other words, not only does a statistical model involve a family or set of probability functions, it is in fact a hybrid entity of parameters and probabilities. As I shall point out, in many statistical models at least, this hybrid complexity already entails some of the critical distinctions characteristic of what I called the nexus of chance in the previous section 3.

In applying or building a *probability model*, the only sensitive judgement concerns the selection of the sample space. And indeed, it is a well-known philosophical lesson that choosing the appropriate sample space – i.e. selecting the outcome events or types that are to go into the space – is critical, and that the choice may importantly alter the properties of the model description. Statisticians by contrast, see this selection as the final and simpler stage in a more complex modelling process, one that requires first of all to judiciously choose an appropriate parametrization of the phenomenon, secondly to choose the probability distributions that best correspond to each parameter, and only thirdly, and consequently, to choose the sample spaces. The 'art of statistical modelling' concerns all of these stages, and it is mainly the most sensitive first two stages that David Cox has in mind when he writes (Cox, 2006, p.197): "Formalization […] is clearly of critical importance. It translates a subject-matter question into a formal statistical question and that translation must be reasonably faithful and, as far as is feasible, the consistency of the model with the data must be checked. How this translation from subject-matter problem to statistical model is done is often the most critical part of the analysis."

There is more to say about the critical first parametrization space; in particular, in many cases of statistical modelling in the natural sciences, it is the stage at which considerations regarding dispositional properties in the chance setups – or propensities – enter. Thus, the relationship between the parameter and the sample spaces $(\Theta, \mathcal{S})$ is at the heart of the distinct roles of propensities and probabilities in the nexus of chance in practice. It makes sense to discuss those roles in the light of the second distinguishing feature of statistical models, namely their dynamical character.

In a typical statistical model in the natural sciences, the relevant parameters include time, and the parametrized description will be time-dependent. As a result, the probability functions will be dynamical and evolve in time, in accordance with some law, often described in a differential or master equation. Statistical models differ on account of the kind of laws that they employ, and I shall in particular distinguish two kinds, reserving the term '*pure probabilistic model'* for those that are endowed with a deterministic dynamics only; while employing "*pure stochastic model*' for those that obey exclusively an indeterministic dynamics. Many models are hybrid from this point of view, and include a variety of different laws, both deterministic and indeterministic. Hence statistical models lie on a spectrum from pure probabilism to pure stochasticity. By investigating both pure types we also investigate the end extremes of this spectrum.

4.2. Pure probabilism: The method of arbitrary functions

The main aim of many statistical models is to generate probability distributions over the outcome space that to a good approximation match those frequencies observed in experiments run on the modelled systems. This amounts to a type of explanation of the resulting frequencies (this type will be addressed in section 5 of this article). If the dynamics in the model is deterministic, the laws on their own cannot provide those probabilities – a deterministic law may only generate probabilities out of probabilities. Hence a *probabilistic* model (a statistical model with a deterministic dynamics) can only dynamically explain statistical phenomena if it acts on a set of probability distributions over the initial conditions of the system. Many systems generating statistical phenomena at the macro-level (including most well-known games of chance, such as dice, roulette, etc.) are on the face of it, indeterministic, since they obey classical mechanical or Newtonian laws. How can probabilistic models account for such phenomena?

A long and distinguished tradition in mathematical physics has endeavoured to provide a template for such models. The most articulate and developed version (the method of arbitrary functions, or MAF) begins with Von Kries and Poincaré at the turn of the 19th century, and it remains relevant today in important work in mathematical statistics. [2] The central idea in the MAF is the thought that some systems are dynamically stable or invariant under permutations (within some range given by some formal constraints) of the initial probability distributions over the initial conditions of the system. In other words, the probability distributions over the outcome events are independent of the initial distribution over the initial conditions; they rather mainly depend only on the precise form of the deterministic dynamics. The phenomena modelled by MAF are thus in some sense the converse of chaotic phenomena: while the latter exhibit extreme sensitivity to (small variations in) initial conditions, the former display extreme resilience from (changes in the probability distributions over the) initial conditions.

[2] See Von Plato (1985) for a historical review, and Engel (1992) for state-of-the-art methodology.

The MAF is a method that falls well within the kind of more complex parametrized dynamical modelling practice that I have here referred to as 'statistical' modelling. Another, more specific reason to discuss it, is that it furnishes a genuine dynamic probabilistic model for the paradigmatic example that I have been employing of a chance system, the coin toss. Keller (1986) provides the most sophisticated treatment, which employs a highly idealised parametrized description of the phenomenon – what this essay argues is the essential first stage in any statistical model. There are a number of idealising assumptions involved because the model purports to reduce the set of free parameters to just two: the initial upwards velocity at which the coin is spun at its ejection (ν), and the angular momentum through its trajectory (ω). To achieve such a reduction of the relevant dynamical variables a streamlined parametrized description is needed of what is in reality a more complex phenomenon. These idealisations allow the modeller to neglect every other dynamical variable for the purposes at hand (see figure 1 from Keller, 1986, p X), and include:

- The coin's radius is a and it remains constant throughout its motion.
- The coin is assumed to be of negligible thickness, or infinitely flat.
- Hence the coin's geometrical centre is its centre of gravity.
- At every instant t through its motion the coin's centre of gravity finds itself at $y(t) = x$.
- At the initial stage t=0 the coin finds itself at precisely height a: $y(0) = a$.
- At the end of the motion, the coin's landing position is final (no rebound).
- Air friction is negligible; and the coin is not slowed down as a result.
- The coin's angular velocity is constant throughout its motion: $\frac{d^2\theta(t)}{dt^2} = 0$, where θ is the angle subtended to the upwards motion.

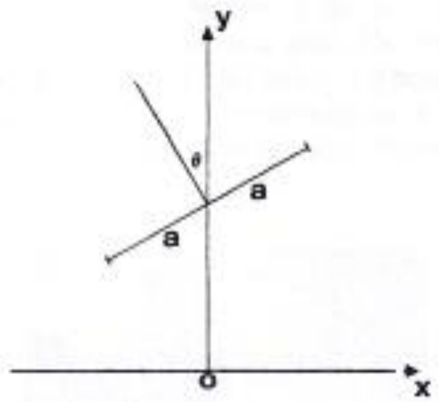

FIG. 1. The x, y plane intersects the coin along a diameter of length $2a$. The normal to the side of the coin marked heads makes the angle θ with the positive y-axis.

It is then possible to show, by applying classical mechanical equations of motion, that any arbitrary distribution over the initial upwards velocity ν and angular velocity ω, as long as it fulfils minimal requirements, yields a final probability distribution over the Heads and Tails outcomes, which in the case of a fair coin (i.e. one not bent), is the equiprobable: Prob (H)= Prob (T) = ½. The requirements receive different names in the literature, and they have been the object of a considerable and intricate discussion. [3] What matters for our purposes is that a parametrisation is implicit already in the selection of the relevant quantities that the initial probability functions shall range over.

In other words, the ascription of probabilities to the possible Heads (H) and Tails (T) outcomes of a given coin tossing experiment is not the result of a simple application of the principle of indifference, or any other variety of a principle of sufficient reason. Philosophers sometimes assume that the principle of indifference on its own will yield probability ½ for each possible outcome of the toss of a fair coin. A 'probability' model is just such an ascription of probabilities (i.e. prob(H) = prob (T) = 0,5) on the basis of indifference. There is not any need for any dynamical model of coin tossing in order to arrive at the conclusion: a simple inspection of the geometrical properties of the coin would do. By contrast, a

[3] Poincaré (1912) and Hopf (1932) explicitly require that the initial distribution functions be "continuous"; Strevens (2003) and Marshall (2012) invoke "micro-constancy": a slightly different requirement applying to the dynamics as much as the initial distribution functions, but has identical consequences for our purposes.

'statistical' model of the phenomenon of coin tossing will necessarily be much more involved. A coin may well be perfectly symmetrical, and fair in the sense that its outcomes are equiprobable; but the reason for the fairness of the coin is not in a statistical model to be found in the symmetries of the object. It is rather to be found, as in Keller's model, in the complex dynamics of the entire coin tossing phenomenon under a suitable idealised parametrisation. The system as modelled is not a thing, or entity, at a given time, but a rather complex dynamical process evolving in time, as described under a set of relevant parameters.

To sum up, the MAF employs what I have called *pure probabilistic* models. These are models of systems that yield a stable or resilient probability distribution over macroscopic variables of their chance setup solely out of some deterministic dynamics acting on a range of distribution functions over initial microscopic variables of the system:

$$p_i \begin{cases} s_1 \\ s_2 \\ s_3 \end{cases} \rightarrow Law_{deterministic} \rightarrow p_f \begin{cases} o_1 \\ o_2 \end{cases}$$

The critical feature of MAF models is their ability to generate resiliently the same probability function, given the same parametrisation of the phenomenon, as ideally described. A different probability function p'_i would result only out of a different parametrisation, with a distinct set of initial conditions $\{s'_1, s'_2, \ldots., s'_n\}$, in turn resulting from a different set of idealisations in the model:

$$p'_i \begin{cases} s'_1 \\ s'_2 \\ s'_3 \end{cases} \rightarrow Law_{deterministic} \rightarrow p'_f \begin{cases} o_1 \\ o_2 \end{cases}$$

For instance, in the case of coin tossing, this entails relaxing the idealisation that the coin is fair, e.g. because the coin is no longer modelled as infinitely flat, or as having its centre of gravity at the geometrical centre, or because it is assumed to be experiencing precession, and hence its angular velocity is far from constant. [4]

[4] See Diaconis et al. (2007) for some of the relevant de-idealisations.

Most games of chance may be modelled in this fashion – and the methodology extends further to complex systems with underlying deterministic dynamics. I return, in section 5, to the implications of both the invariance under changes in initial conditions, given a parametrisation; and the breakdown in invariance elicited by a new parametrisation introduced in response to changes in the system's physical properties.

4.3. Pure stochasticity: Indeterministic dynamical modelling

By contrast, a *pure stochastic* statistical model is one where the probabilities emerge out of the dynamics by itself, without recourse to any initial probability distributions over initial micro or macroscopic conditions:

$$\begin{cases} s_1 \\ s_2 \\ s_3 \end{cases} \rightarrow Law_{stochastic} \rightarrow p_f \begin{cases} o_1 \\ o_2 \end{cases}$$

The laws in pure stochastic models are indeterministic or stochastic and generate objective probability distributions over the outcome events out of very precise specifications of the actual initial state of the system. The probability functions P_f predicted by such models are hardly ever invariant under changes in initial conditions – they are therefore not just sensitive to the parametrisation entailed by the idealised description of a system, but also to the initial probability functions themselves, including their sample spaces. Since the underlying dynamics is not deterministic, these cases tend to lie outside the domain of ordinary macroscopic phenomena. Two examples include collapse interpretations in quantum mechanics; and stochastic models for genetic variance in evolutionary theory.

Collapse theories in quantum mechanics assume an indeterministic change of the state of a quantum system (its wavefunction) either as a result of interaction with the open environment (as in quantum state diffusion or QSD theory) or spontaneously with a certain frequency (as in the so-called Ghirardi-Rimini-Weber or GRW theory). There is no deterministic equation of motion; the changes are rather sudden and stochastic: one can only determine their probability, in the form of either transition probabilities or relaxation times. Thus, for example, a model for a quantum state diffusion process is a statistical model that yields continuous probability distributions for the evolution of the state in an abstract space, such as a Bloch space. As such the motion of the state vector in the space appears random when as a matter of fact it is highly constrained by the probabilistic equations of motion. Gisin and Percival (1992, p. 5679) make it clear that these equations derive from a master equation including a drift term and stochastic fluctuations, and they are therefore irreducibly indeterministic: "[…] there can be no general deterministic equation for the pure states $|\psi\rangle$. But there are stochastic equations, as might be expected from the probabilistic nature of the interaction with the environment. In time dt the variation $|d\psi\rangle$ in $|\psi\rangle$ is then given by the Itô form: $|d\psi\rangle = |v\rangle dt + \sum_j |u_j\rangle d\xi_j$, where $|v\rangle dt$ is the drift term and the differential stochastic fluctuations are represented by a sum over independent Wiener processes." The process may be understood as a sort of random walk on the Bloch sphere where states are represented.

The Ghirardi-Rimini-Weber (GRW) theory is similar except that it does not require open systems in constant interaction with the environment but rather postulates stochastic and spontaneous 'shocks' on the wavefunction which bring it regularly into the eigenstates of macroscopically well-defined observables. The relaxation times are construed in such a way that any finite-time observation on any macroscopic composite typically yields a definite outcome. There is no macroscopic superposition due to the aggregate of the non-linear stochastic terms added to the Schrödinger dynamics. The GRW modification of the dynamics in effect "leaves things unchanged for microscopic objects, while, for macroscopic objects, transforms quantum mechanics into a stochastic mechanics in phase space exhibiting the classical features" (Ghirardi et al., 1986, p. 34). This somehow inverts the traditional picture, since the Schrödinger equation is a deterministic equation on the wavefunction; while the GRW theory presupposes that the fundamental stochastic collapses it postulates for the wavefunction manifest themselves at the macrolevel. It is thus plausible to think of GRW as providing a mixed statistical model, which is not purely probabilistic nor purely stochastic, but a mixture of both.

Another field that illustrates statistical modelling in its stochastic variety is evolutionary biology – particularly population genetics, but more generally in the study of variability across populations, or in ecosystems. As for the former, consider the notorious Wright-Fisher model for genetic drift. [5] The model describes the time-evolution of a population of N genes, under considerably strong idealising conditions. For instance, it assumes that populations are of finite and do not vary in size from one to the next generation, and that the generations do not overlap – i.e. they are replaced wholesale every time. According to this model the number of alleles in generation g+1 is obtained by drawing with replacement from the gene population in the previous generation g. Thus if there are i alleles of type A in generation g, then then number of type A alleles in generation g+1 has a binomial distribution yielding a Markov process or chain with a transition matrix given as: $P_{ij} = \binom{N}{j}\left(\frac{i}{N}\right)^{j}\left(1-\frac{i}{N}\right)^{N-j}$, for $0 \leq i, j \leq N$. Each expression for i,j provides a transition probability for the number of alleles in a later generation.

The model can be refined and extended by suitably weakening the idealisations and varying the range of parameters. Kimura introduced the hypothesis of neutrality: some gene mutations have no effect whatever on fitness, and hence such alleles cannot vary out of natural selection; so genetic drift must account for a larger share of gene pool variability than previously thought. This invites the thought that the idealisations in the original Wright-Fisher model may be too strong, particularly non-overlapping generations. A new stochastic model then developed allowing for overlaps amongst generational populations. Once again, a parametrisation of the phenomenon, under some idealised description, is critical in order to establish the appropriate probability functions and their domains. Many models in evolutionary biology are neither purely probabilistic, nor purely stochastic, but lie somewhere in the spectrum. [6]

---

[5] For an exposition see e.g. Blythe and McKane (2007), or the seminal Fisher (1930).
[6] Rice (2008).

5. Towards a Methodology of Chance Explanation

Statistical modelling is a complex activity that centres around providing explanatory models for observed or presumed correlation phenomena. The models invoke dynamical laws and employ particular parametrisations, often describing the phenomena in a highly idealised form. Whether the laws employed are deterministic or stochastic, the models appear to have an explanatory role. This often reflects the fact that the idealised parametrisations represent to some degree the underlying mechanisms, causal powers, or capacities operating in the system. But it is more generally and minimally related to the models' essential posits in the nexus of chance.

The explanans employs an idealised description of the propensities – or probabilistic dispositions – inherent in the system. As the idealisations change, so do the required parametrisations, and the ensuing description of the propensities in the system. A biased or precessing coin has distinct propensities to land heads or tails if tossed, and it must be modelled so; an open quantum system interacting with the environment displays a propensity to localise as a result; gene populations possess certain propensities to pass on types of alleles to the next generation; and so on. In all these cases, there is a complex relation between i) the propensities in the systems or chance set ups, as revealed in the parametrisation employed; ii) the probabilities that ensue over the outcome events, often at a macroscopic level; and iii) the frequencies that are presumed or observed in experimental runs, which provide the empirical basis for our probability claims, and which are ultimately the object of our models' explanation.

This is the nexus of chance in action; its distinct parts (propensities, probabilities, frequencies) are all required in order to make sense of the methodology employed in statistical modelling. The order of explanation seems to entail a distinct hierarchy, with the propensities at the highest level of explanation, the probabilities as the formal representation of the dynamical consequences of the propensities, and the finite frequencies as the putative consequences or explananda. Most minimally, the explanation is a variety of the model explanations that have been recently discussed in the literature (Bokulich, 2008). The essential explanatory posits in these models are precisely the components in the nexus of chance: propensities, understood as probabilistic dispositions, give rise within the highly idealised model descriptions to probability distributions over the outcomes; together propensities and probability distributions entail certain finite frequencies in particular experimental set-ups. This is to say that all of them together provide explanations for the finite frequencies that are observed, or postulated in the phenomena. To the extent that a phenomenon P is minimally explained by the essential posits of a successful model representation for it, it follows that the nexus of chance is involved essentially in all of these explanations.

Where does this discussion leave the philosophical debate regarding the nature of chance or objective probability? Is objective probability reducible to propensity or frequency? As was mentioned earlier on, this discussion does not address the debate in any way decisively; it only confronts it indirectly, by showing that all three play a distinct and irreducible role in the complex practice of statistical modelling. The underlying assumption is that chance is a complex and plural notion, requiring us to consider the interaction in modelling practice of its distinct and complex parts – while refusing to reduce any of the parts to the rest, or indeed the whole complex nexus to just one of its parts. This plural attitude to chance goes a much longer way in understanding the practice. Does it also provide the foundations for a very different enquiry into the nature of chance? Some have certainly thought so – including Hans Reichenbach in his doctoral dissertation (Reichenbach, 1915/2008). That would turn what I have presented here as a project in methodology into another inquiry into the ontology of chance.

Whilst there is no doubt that some new avenues open up for such an inquiry, the safe and more limited conclusion of this paper is this: regardless of what the ontology of chance is, the methodology of chance explanations via statistical models is undeniably plural and irreducible. No serious philosophical inquiry into the nature of chance can start from different assumptions.

6. Conclusion: To (Statistically) Save the (Statistical) Phenomena

I have defended the view that chance is a plural tripartite notion involving propensities, formal probability distributions, and frequencies. There are arguments in favour of this conclusion coming from the irreducibility of chance to either propensity or frequency. But most significantly, the main explanatory argument for chance pluralism derive from scientific modelling practice. The nexus of chance, as I have called it, is the interlinked set of practices that employ dispositional probabilities – or propensities – as the grounds for the formal probability distributions over outcome spaces typical of chancy phenomena. Both jointly have explanatory roles in practice. Thus statistical models differ from simpler descriptive 'probability' models in that they are deeply stepped into explanatory considerations relative to the idealisations that they employ. They also significantly differ on account of their dynamics, and so they lie on a spectrum from 'pure probabilistic' models – those endowed only with deterministic dynamics – to what I have called 'pure stochastic' models – those which are governed only by stochastic dynamics.

Together propensities and probabilities can be employed to account for, or to explain (in a minimal sense of model explanation), the kind of finite frequency data so common in experimental runs on chance setups. In this respect, statistical modelling is no different from any other form of scientific modelling practice. A large part of what is required in understanding chance is related to understanding the practice of statistical modelling.

REFERENCES

Abrams, M. (2012), "Mechanistic probability", *Synthese*, 187 (2), pp. 343-375.

Blythe, R. A. and A. J. McKane (2007), "Stochastic models of evolution in genetics, ecology and linguistics", *Journal of Statistical Mechanics: Theory and Experiment*.

Bokulich, A. (2008), “Explanatory Fictions” in M. Suárez, ed., *Fictions in Science: Philosophical Essays on Modelling and Idealization*, Routledge.

Carnap, R. (1945), “The two concepts of probability: The problem of probability”, *Philosophy and Phenomenological Research*, 5(4), pp. 513-532.

Cox, D. R. (2006), *Principles of Statistical Inference*, Cambridge: Cambridge University Press.

Diaconis, P. et al (2007), “Dynamical bias in the coin toss”, *Society for Industrial and Applied Mathematics* (SIAM review), 49 (2), pp. 211-235.

Eagle, A. (2004), “Twenty-one arguments against propensity interpretations of probability”, *Erkenntnis* 60 (3), pp. 371-416.

Engel, E. (1992), *A Road to Randomness in Physical Systems*, Springer-Verlag, Berlin.

Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*, Oxford: Clarendon Press.

Gillies, D. (2000), *Philosophical Theories of Probability*, London: Routledge.

Hacking, I. (1975), *The Emergence of Probability*, Oxford: Oxford University Press.

Hajek, A. (1997), “’Mises Redux’ – redux: Fifteen arguments against finite frequentism”, *Erkenntnis* 45 (2/3), pp. 209-227.

Hopf, E. (1932), “On Causality, Statistics and Probability”, *Journal of Mathematics and Physics*, MIT 13, pp. 51-102.

Humphreys, P. (1985), “Why propensities cannot be probabilities”, *The Philosophical Review* 94 (4), pp. 557-570.

Humphreys, P. (1989), *The Chances of Explanation*, Princeton: Princeton University Press.

Keller, J. (1986), "The probability of heads", *American Math. Monthly*, 93, pp. 191-197.

Kimura, M. (1968), *The Neutral Theory of Molecular Evolution*, Cambridge: Cambridge University Press.

McCullough, P. (2002), "What is a statistical model?", *Annals of Statistics* 30 (5), pp. 1225-1310.

Mellor, H. (2004), *Probability: A Philosophical Introduction*, London: Routledge.

Moran, P. (1959), "Random processes in genetics", *Proc. Cambridge Philos. Society*, 54, pp. 60-72.

Poincaré, H. (1912), Calcul de Probabilités, 2nd edition, Paris: Gauthier Villars. (1st edition, 1898).

Ramsey, F. (1928), "Truth and Probability", in R. B. Braithwaite, *The Foundations of Mathematics and other Logical Essays*, London: Kegan Paul.

Reichenbach, H. (1915 / 2008), *The Concept of Probability*, Open Court Publishers.

Reichenbach, H. (1935/49), *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*, University of California Press.

Rice, S. H. (2008), "A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution", *BMC Evolutionary Biology*, 8, 262.

Salmon, W. (1979), "Propensities: A Discussion Review", *Erkenntnis*, 14, pp. 183-216.

Sober, E. (2010), "Evolutionary Theory and the Reality of Macro-Probabilities", in Eels and Fetzer (eds.), *The Place or Probability in Science*, Springer, pp. 133-160

Strevens, M. (2003), *Bigger than Chaos: Understanding Complexity through Probability*, Cambridge, MA: Harvard University Press.

Suárez, M. (2013), "Propensities and Pragmatism", *The Journal of Philosophy*, CX (2), pp. 61-92.

Suárez, M. (2017), "Propensities, Probabilities, and Experimental Statistics", in Massimi, M., J. W. Romeijn, and G. Schurz (eds.), *EPSA15: Selected Papers: European Studies in Philosophy of Science*, vol. 5, pp. 335-345.

Von Mises, R. (1928/1959), *Probability, Statistics, and Truth*, Dover Publications.

Von Plato, J. (1985), "The Method of Arbitrary Functions", *British Journal for the Philosophy of Science*, 34 (1), pp. 37-47.