archives-ouvertes.fr

Philosophical Foundations of Probabilities

Pierre Dangauthier

To cite this version:

Pierre Dangauthier. Philosophical Foundations of Probabilities. [Research Report] RR-5988, INRIA. 2006, pp.14. <inria-00102049v2>

HAL Id: inria-00102049

https://hal.inria.fr/inria-00102049v2

Submitted on 29 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Philosophical Foundations of Probabilities

Pierre Dangauthier

N° 5988

Septembre 2006

Thème NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5988--FR+ENG

Philosophical Foundations of Probabilities

Pierre Dangauthier*

Thème NUM — Systèmes numériques
Projet E-motion

Rapport de recherche n° 5988 — Septembre 2006 — 11 pages

Abstract: This document presents some views on the semantic of the probability notion. It's mainly a summary of different readings, in particular of a *Stanford Encyclopaedia of Philosophy* article. Probabilities are daily used in real world application, often without a clear understanding of their mathematical foundation and of the meaning we attach to them. This lack of understanding often leads to bad interpretations of statistical/probabilistic procedures (*e.g.* p-values of tests), and is an obstacle for research. Nowadays there is a debate opposing *frequentists* to *bayesians*, but we'll see that the problem of interpreting probabilities is broader, even if we'll present some strong clues in favour of bayesianism.

Key-words: probability, bayes, statistics, philosophy, axiomatic

* *pierre.dangauthier@imag.fr*

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Fondations philosophiques des probabilités

Résumé : Ce document présente quelques fondements mathématiques et philosophiques de la notion de probabilité. Les probabilités sont utilisées dans de nombreuses théories et applications scientifiques, parfois sans une claire et complète compréhension de leur définition mathématique, ni de leur interprétation sémantique. Cette mauvaise compéhension peut conduire à de nombreuses erreurs d'interprétation des résultats statistiques, comme dans le cas des tests d'hypothèses. Bien que le débat philosophique ne se limite pas à l'opposion entre l'approche fréquenciste et l'approche bayésienne, nous présentons plusieurs avantages de cette dernière.

Mots-clés : probabilité, statistique, bayes, philosophie, axiomatique

1 Introduction

The tautology *probabilities are things that obey the axioms of probability* raises two questions. Firstly what are those axioms, why are they what they are? And secondly what's the link with physical reality, what interpretation can we do of those numbers?

Indeed when we want to argue on the meaning of probabilities, we firstly need to have a common definition of the word probability. This is usually done by relying Kolmogorov theory which is basically a mathematically consistent formulation of our intuitive desiderata. Actually this theory just says that a probability is a non-negative real number between 0 and 1 associated with a "proposition" (or "event", "outcome"...). It also adds that probabilities should be additive for disjointed events.

After quickly reviewing Kolmogorov definition, we'll present several quite different interpretations presented in [HÓ3]: classical, logical, frequency, propensity and subjective probabilities. They have to be regarded through several criteria: do they obey to the axioms, do they provide a way to assign a number to the probability of an event, do they make a link with long term behaviour, with rational belief, and do they make sense when applied to physics.

In the last section, we'll see several good properties of the subjective bayesian approach.

2 Axiomatic

Independently from the meaning of probabilities, Kolmogorov defined an axiomatic for probabilities based on measure theory. It's important to notice that it's just a mathematical definition of what a probability is, and that it's *a priori* independent of any interpretation[1]. In particular, the "random variable" concept doesn't have to be a taboo word, even for the most bayesian scientist.

2.1 Details

A probability is a positive function from a σ-algebra to $[0,1]$ which sum to one, and which verifies countable additivity. The element of the σ-algebra are called events or propositions, and a probability assigns a real number to each of them. The σ-algebra condition just means that the set of propositions taken into consideration have to be suitable in some sense. Those conditions are really intuitive: existence of the complementary proposition, and stability by numerable union.

A probabilized space is a space Ω with a σ-algebra and a probability function. A random variable is just a function from a probabilized space to a measurable space T. For instance the meaning of $P(a < X < b)$ is actually $P(X^{-1}([a,b])) = P(\{\omega \in \Omega / X(\omega) \in [a,b]\})$. At this stage, for each probability on Ω, each random variable defines its own probability distribution on T. We're used to forget about Ω and just speak about the induced probability on T by X, which is simply a normalized measure.

[1] We'll see that it's not completely true because some interpretations fail to satisfy some desiderata

Then, a "random variable" is just a mathematical tool simplifying our notation, by allowing to speak about $P(X = x)$ as a function of x instead of having to take into consideration a huge number of unlinked events. It's important to note that the notion of "random variable" doesn't imply any notion of "randomness", neither a particular

Under certain conditions on T (Radon-Nikodym theorem [Wik06]), it's possible to define a probability density on T, $p(X)$. Again it's common to confuse p with P, the density with the probability, but it's important to remember that this is not the same thing. In particular the density is a distribution (in the sense of Schwartz), which can be irregular, as a real Dirac is. On top of that, in the bayesian interpretation, what is relevant is the probability, which is an integration of the density around a point. Generally speaking, maximizing a density has no meaning and often leads to over fitting. "Bayesian methods don't over fit, because they don't fit anything !(Z. Ghahramani)."

2.2 Why this?

Some of Kolmogorov axioms can be contested to bring some new theories (countable additivity, or probability domain). But in general, authors admit that they form a healthy base. Nevertheless we'll see that some interpretations have problems to match exactly theses axioms.

There are two strong arguments supporting to those axioms. Firstly Cox's theorem [Cox61] asserts that, if you consider probability as a degree of subjective belief verifying some simple common sense desiderata, then they must obey to the rules of probability calculus (really close to Kolmogorov axioms). The other argument is the existence of Dutch books if your probability assignments don't follow the axioms. A Dutch book is a set of bets you'll accept according to your probability assignments, but with which you'll loose with probability one.

2.3 Other axiomatics

Some research have been done by allowing probabilities to live in a broader space than just the real $[0, 1]$ interval. Complex and quaternion probabilities are strongly linked with quantum mechanics. [You01] argues that Cox requirements still holds when you consider measuring subjective beliefs with complex numbers (or quaternions). Then he derives a new branch of bayesian inference, and argues that this simplifies a lot the interpretation of quantum mechanics. Despite appealing claims, this approach isn't widespread and we can read form Jefferys:

> "There is no need to modify probability theory from any perspective in order to do quantum mechanics. Bayesianism uses standard, unmodified probability theory. Bayesian interpretations of QM use standard, unmodified probability theory. [...] Such [complex, quaternion probabilities] approaches are also not necessary and in my opinion they confuse more than they illuminate."

3 Classical probabilities

Let's start with the first, classical interpretation of probabilities. It's due to the seminal work of Laplace, and has been continued by Pascal, Bernoulli, Huygens and Leibniz, mainly in the context of games of chance and during the beginning of nineteenth century.

It's based on the **principle of indifference**, stating that in the absence of any evidence (or presence of balanced evidence), probability is shared equally among (a finite number of) outcomes. $P(A) = \frac{Card\, A}{Card\, \Omega}$. It wasn't Laplace's intention to produce some *subjective* probabilities, but rather an objective assignment with respect to a set of "equally possible" cases. His classical theory has been extended to countably infinite spaces thanks to the **maximum entropy principle**. To bridge the gap with frequencies, Laplace added he's **law of succession** $P(\text{success on } N+1 \text{ trial } N \text{ consecutive successes }) = \frac{N+1}{N+2}$. Thus, inductive learning becomes possible.

This interpretation raises several questions:

- Classical theory implies rational numbers. It's a problem in science, in particular we need irrationals for quantum mechanics.
- More fundamentally, the idea to assign equal probabilities to "equally possible" cases is somehow a circular definition. It's not obvious that "equal possibilityness" can be defined without the notion of probability (or conditional probability in the presence of balanced evidence).
- Pragmatically, the principle of indifference can be self contradictory, when assigning different probabilities to the same events with two different parametrisation (Betrand's paradoxes).
- The principle of indifference somehow extracts information form ignorance. If I know nothing about the result of a coin tossing, I even don't known that the coin is fair.

4 Logical probabilities

Studied by Johnson, Keynes, Jeffreys and **Carnap** during first half of the century. This approach tries to **extend classical logic** in order to allow **induction**. This means providing a notion of "degree of implication" that a piece of evidence confers upon a given hypothesis. It needs a finite formal language composed of logically independent predicates, of variables and of the usual logical connectives. It extends the classical interpretation by allowing non uniform probabilities, and by taking into account unbalanced pieces of evidence. It allows inductive learning.

This interpretation encounters several difficulties. It needs some arbitrary choices (confirmation function"), this is a purely syntactic approach, unrelated to the predicates meaning and the probability of an event depends on the choice of the language in which this event is described.

5 Frequency interpretation

This is the most widespread and taught approach to probabilities, and its results are daily used by numerous scientists and applied statisticians. The ideas are coming from Venn (1876) "probability is nothing but that proportion", Reichenbach (1949), von Mises (1957), Church (1940),De Finetti (1972) and greatly benefited from the Kolmogorov formalization.

The simplest case is the finite frequentist approach, in which the probability of an event is simply its **actual frequency**. It's close to classical interpretation at the difference that it counts actual outcomes instead of all possible outcomes. With this finite approach, we face immediately to the so-called **problem of the single case**. If an event is unrepeatable, its probability could be 0 or 1, nothing else. This problem is extendable to every problem with a finite number of realizations, and unfortunately they are quite numerous in the real world.

The solution was to move the definition to the **limiting relative frequencies in a hypothetical infinite sequence of trials**, but it has still several problems:

- For the limiting process, it's always possible to find a reordering of an infinite sequence in order to make the frequencies converge to an arbitrary number between 0 and 1. Why should an ordering be privileged over others ?
- The frequency of an event depends upon the class of considered events, and on the considered sequence. Von Mises and Church precises things in order to make this problem disappear, and at the end the notion of probability lost all its meaning except for infinite sequences.
- Limiting frequencies can violate Kolmogorof axiomatic, in particular countable additivity.
- The connection between this definition of probability and frequency (identity) is too tight, facts about finite relative frequencies should serve as evidence, not conclusive evidence.

6 Propensity

Studied by Popper (1959), Hacking (1965), Giere (1973), Fetzer(80's), Miller (90's) Gillies (2000).

In this propensity interpretation, the probability represents a quality of the physical word. It represents a **tendency, a disposition of the world to behave in a certain way**. There are two kinds of propensities:

- the long run propensities, assessing that, the relative frequencies have a tendency to converge to a certain number. The strength of this tendency is the propensity. This interpretation is closed to but different of the frequency approach.
- the single case propensity, saying that an event as a tendency of, say 90% to occur. In this case 90% is the propensity.

Fundamentally, the propensity isn't well defined. This seems to be more a metaphysical concept than a scientific one, they are untestable, we don't have any cue of what experiment do in order to test if they're positive, normalized or additive. Moreover, propensities suppose *a priori* a kind of stability, of repeatability of the world, and nothing presuppose that propensities verify probability calculus rules. Actually they don't in the single outcome case. The last problem with propensities is that Bayes' rule is symmetric, but inverting causalities makes no sense in the propensity interpretation, there is an asymmetric notion of casual tendency. That's why propensities also violate Kolmogorov's axioms.

7 Subjective

To finish, we'll focus on the subjective approach, which main contributors were Ramsey (1926), De Finetti (1980), Kemeny (1955); von Neumann (1944), Savage (1954), Jeffrey (1966), Jaynes, Skilling, Lindley, Good, Cox. This approach is making a coming back recently, partly because modern computers allow the heavy computations often implied.

Probability is the subjective **degree of belief** of a **suitable** agent. The definition is "suitable" is crucial, because if we tolerate all agents, then there is no reason that probabilities satisfy Kolmogorov's axioms. But it has been shown that the **rationality** of an agent is equivalent to the respect of the probability calculus rules. An agent is rational if he's coherent, if he won't accepted a set of individually acceptable bets with which he'll surely loose. This situation (Dutch book) won't happen if and only if the subjective probabilities obey to Kolmogorov rules.

How to measure a degree of belief ? An operational definition can be the price of a bet the agent would place. More precisely, it's the price that makes the game become fair in the eyes of the agent. Now it's possible to extract the probability values of an agent from a series of experiments, presenting him different bets and seeing if he accepts them or not. Actually probabilities are linked to rational choices through decision theory. In this theory, a rational choice maximizes expected utility. then, we don't have access to the degree of belief itself, but to the product belief-utility. Ramsey, Savage and Jeffrey derived both probability and utility functions from rational preferences, but all these temptatives are controversial, because of the non identity between belief and desire.

The desire for a *rational* agent has been discussed, and several other criteria have been added to constraint the possible values of probability. For instance *regularity* which assert that only falsehoods can get a zero probability. Other recent approach noticed that a rational agent evaluates his degree of belief according how they match observed frequencies, how they match "truth", and also according to knowledge coming from experts. The main difficulties of this approach are:

- Measuring the degree of belief of an agent isn't easy.
- Then, when we have to pick a prior, we may not be certain what prior reflect our true state of prior knowledge, especially on continuous spaces, on infinite space, on high dimensional space and on infinite dimensional spaces. Some schools (ex: maximum

entropy) prefer expressing their belief in term of constraints on theses priors, instead of real valued functions.

- The non-objectivity is sometimes seen as an obstacle to scientific communication, and scientific comparison of hypothesis.
- Inference in the bayesian framework is computationally challenging, and even if some good approximate methods are available, computations in high dimensional spaces is really difficult.

8 Subjective/objective debate

After those philosophical considerations, we see that the recurrent debate between orthodox and Bayesians (frequentists and subjectivists) has some profound roots. Briefly a frequentist is someone who thinks the "probability" of an event means nothing other than its frequency of occurrence under specified repeatable conditions, while a Bayesian is someone who thinks the "probability" of an event is a measure of the belief he has on the truth of it[2].

And those philosophical differences make some **difference in practical** cases.

Because of their limiting frequencies approach, objectivists are far more restricted in the set of variable they can consider as random variables. If the noise model of a location measurement is a gaussian centred on the true position of the object, you can't consider put a probability distribution on this parameter. They'll them employ ad-hoc estimation technique to get an idea of the true position.

According William Jefferys [Jef02]:

> The basic difference between Bayesians and frequentists is this: Bayesians condition on the data actually observed, and consider the probability distribution on the hypotheses; they believe it reasonable to put probability distributions on hypotheses and they behave accordingly. Frequentists condition on a hypothesis of choice and consider the probability distribution on the data, whether observed or not; they do not think it reasonable to put probability distributions on hypotheses (in their opinion, one hypothesis is true, the rest are false, even if we do not know which is the case), and they behave accordingly.

and

> Bayesians condition on the data actually observed, and put probability distributions on hypotheses. If you aren't willing to put probability distributions on hypotheses, then you can't be a Bayesian.

[2] This distinction based on different interpretation of probability isn't totally uncontroversial. For instance some authors define bayesianism by the fact of estimating prior parameters on data. But as this is impossible in a frequentist approach, we think this issue isn't really important. On the other hand, bayesian people like Bernardo try to define non-subjective, or weak, priors [Ber97], but in fact, this doesn't conflict with the subjective approach.

> Frequentists condition on hypotheses, and put probability distributions on data sets. They consider it reasonable to consider data sets other than the one actually observed. If you don't think it reasonable to consider data sets other than the one actually observed, then you can't be a frequentist.

This difference has some really practical consequences. Indeed, a frequentist makes his conclusion by looking at the actual observed data and also at all possible other outcomes. In certain experiments, his conclusion will depend on those other non observed data, which depend on his experiment plan, on his intention. Several simple, practical examples of such pathologies are presented in [Min01].

Actually classical hypothesis testing violates the likelihood principle, which says that all the information relevant to the estimation of a parameter is contained in the likelihood of the observed data. In the frequentist method, it's possible that the likelihood function related to an observed data set also depends on unobserved data. Then your belief on those unobserved data, on the way you could have done experiments (but didn't) can influence an estimation. That's a strong argument in favour of bayesian approach, it seems unreasonable that our conclusions depend on the unspecified intention of the experimentalist.

The applicability to science, and in particular to quantum mechanics, is another point that makes the subjective approach more convenient. Quoting John Baez [Bae03] :

> In the jargon of probability theory, the frequentist interpretation of probability is wrong.[...]
>
> The basic idea behind the Bayesian interpretation is [that] we must start by assuming some probabilities. Then we can use these to calculate how likely various events are. Then we can do experiments to see what actually happens. Finally, we can use this new data to update our assumptions.[...]
>
> Subjective Bayesians argue that ones choice of prior is unavoidably subjective. Objective Bayesians try to find rules for choosing the "right" prior [see [Ber97]].[...]
>
> This seems to imply that the "wave function" of the electron is just a summary of our assumptions about it, not someone we can ever measure.[...]
>
> Probability theory is the special case of quantum mechanics in which ones algebra of observable is commutative.[...]
>
> Any computation of probabilities is a computation w.r.t. some probability measure [...]
>
> I believe the frequentist interpretation just isn't good enough for understanding the role of probability in quantum theory. This is especially clear in quantum cosmology, where we apply quantum theory to the entire universe. We can't prepare a large number of identical copies of the whole universe to run experiments on!

Finally, a common reproach to bayesianism is its lack of objectivity, but actually frequentist statistics is also subjective, it's just that the subjectivity is better hidden in the choice of p-values and pivotal functions. We recognize that we have some difficulties to believe in true objectiveness in statistics, as in all other scientific domains[3].

9 Conclusion

Probability theory and its statistical aspects play a special role in science, mainly because they are a necessary tool to understand and interact with the complex real world. When using probabilities to speak about experiments, one has to provide a meaning, a physical interpretation of this concept. Depending on this interpretation, he should use the theory in different ways, for instance by assigning probabilities to different kind of objects. Despite the fact that all interpretations face some difficulties, bayesianism proposes paradox-free, intuitive, consistent and rational methods for inference in a variety of practical cases. Moreover, bayesianism is a promising approach to understand quantum mechanics, rational decision making, psychology, and more generally, it has been presented as an extension of Aristotle's logic [Jay03].

References

[Bae03] John Baez. Bayesian probability theory and quantum mechanics. On line, September 2003.

[Ber97] José-Miguel Bernardo. Noninformative priors do not exist: A discussion. *Journal of Statistical Planning and Inference*, 65:159–189, 1997.

[Cox61] Richard Threlkeld Cox. *The Algebra of Probable Inference.* The Johns Hopkins Press, Baltimore, 1961.

[HÓ3] Alan Hájek. Interpretations of probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Summer 2003.

[Jay03] E. T. Jaynes. *Probability Theory : The Logic of Science.* G. Larry Bretthorst, 2003.

[Jef02] Bill Jefferys. Re: Bayesian & frequentist probability theory. sci.physics.research 2002-03 Threads, 2002. [Online].

[Min01] Thomas Minka. Pathologies of orthodox statistics. 2001.

[Wik06] Wikipedia. Radon nikodym theorem, wikipedia, the free encyclopedia, 2006. [Online; accessed 3-March-2006].

[3] In particular we do believe that the second law of thermodynamics, and entropy is deeply subjective notions.

[You01] S. Youssef. Physics with exotic probability theory. *ArXiv High Energy Physics - Theory e-prints*, October 2001.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399