

**FINITE ADDITIVITY VERSUS COUNTABLE ADDITIVITY:
De FINETTI AND SAVAGE**

N. H. BINGHAM

Electronic J. History of Probability and Statistics **6.1** (2010), 33p

Résumé

L'arrière-plan historique, d'abord de l'additivité dénombrable et puis de l'additivité finie, est étudié ici. Nous discutons le travail de de Finetti (en particulier, son livre posthume, de Finetti (2008)), mais aussi les travaux de Savage. Tous deux sont surtout (re)connus pour leurs contributions au domaine des statistiques; nous nous intéressons ici à leurs apports du point de vue de la théorie des probabilités. Le problème de mesure est discuté – la possibilité d'étendre une mesure à *tout* sous-ensemble d'un espace de probabilité. La théorie des jeux de hasard est ensuite présentée pour illustrer les mérites relatifs de l'additivité finie et dénombrable. Puis, nous considérons la cohérence des décisions, où un troisième candidat apparaît – la *non-additivité*. Nous étudions alors l'influence de différents choix d'axiomes de la théorie des ensembles. Nous nous adressons aux six raisons avancés par Seidenfeld (2001) à l'appui de l'additivité finie, et faisons la critique des plusieurs approches à la fréquence limite.

Abstract

The historical background of first countable additivity, and then finite additivity, in probability theory is reviewed. We discuss the work of the most prominent advocate of finite additivity, de Finetti (in particular, his posthumous book de Finetti (2008)), and also the work of Savage. Both were most noted for their contributions to statistics; our focus here is more from the point of view of probability theory. The problem of measure is then discussed – the possibility of extending a measure to *all* subsets of a probability space. The theory of gambling is discussed next, as a test case for the relative merits of finite and countable additivity. We then turn to coherence of decision making, where a third candidate presents itself – *non-additivity*. We next consider the impact of different choices of set-theoretic axioms. We address six reasons put forward by Seidenfeld (2001) in favour of finite additivity,

and review various approaches to limiting frequency.

Keywords: finite additivity, countable additivity, problem of measure, gambling, coherence, axiom of choice, axiom of analytic determinacy, limiting frequency, Bruno de Finetti, L. J. Savage.

2000 Math. Subject Classification: Primary 60-02, Secondary 60-03, 01A60

§1. Background on countable additivity

In the nineteenth century, probability theory hardly existed as a mathematical subject – the area was a collection of special problems and techniques; for a view of the position at that time, see Bingham (2000). Also, the mathematics of length, area and volume had not developed beyond Jordan content. This changed around the turn of the twentieth century, with the path-breaking thesis of Henri Lebesgue (1875-1941), Lebesgue (1902). This introduced both measure theory, as the mathematics of length, area and volume, and the Lebesgue integral, as the more powerful successor and generalization of the Riemann integral. Measure-theoretic ideas were introduced early into probability theory by Emile Borel (1871-1956) (Borel's normal number theorem of 1909 is a striking early example of their power), and Maurice Fréchet (1878-1973), and developed further by Paul Lévy (1886-1971). During the first third of the last century, measure theory developed beyond its Euclidean origins. Fréchet was an early advocate of an abstract approach; one of the key technical developments was the extension theorem for measures of Constantin Carathéodory (1873-1950) in 1914; another was the Radon-Nikodym theorem, proved by J. Radon (1887-1956) in the Euclidean case in 1913, O. M. Nikodym (1887-1974) in the general case in 1930. For accounts of the work of the French, Polish, Russian, Italian, German, American and English schools during this period, see Bingham (2000).

Progress continued during this period, culminating in the publication in 1933 of the *Grundbegriffe der Wahrscheinlichkeitsrechnung* (always known as the *Grundbegriffe*) by A. N. Kolmogorov (1903-1987), Kolmogorov (1933). This classic book so successfully established measure-theoretic probability that it may serve as a major milestone in the history of the subject, which has evolved since along the lines laid down there. In all of this, probability and measure are (understood) σ -additive or countably additive. For appreciations of the *Grundbegriffe* and its historical background, see Bingham (2000), Shafer and Vovk (2006). For Kolmogorov's work as a whole, includ-

ing his highly influential paper of 1931 on Markov processes, see Shiryaev (1989), Dynkin (1989), Chaumont et al. (2007).

Now there is inevitably some tension between the countability inherent in countable additivity and the uncountability of the unit interval, the real line or half-line, or any other time-set used to index a stochastic process in continuous rather than discrete time. The first successful systematic attempt to reconcile this tension was Doob's theory of separability and measurability, which found its textbook synthesis in the classic work Doob (1953). Here, only twenty years after the *Grundbegriffe*, one finds measure-theoretic probability taken for granted; for an appreciation of the impact of Doob's book written in 2003, see Bingham (2005).

Perhaps the most important subsequent development was the impact of the work of Paul-André Meyer (1934-2003), the general theory of (stochastic) processes, the work of the Strasbourg (later Paris, French, ...) school of probability, and the books, Meyer (1966), Dellacherie (1972), Dellacherie and Meyer (1975, 1980, 1983, 1987), Dellacherie, Maisonneuve and Meyer (1992). For appreciations of the life and work of Meyer and the achievements of his school, see the memorial volume Emery and Yor (2006).

The upshot of all this is that probability theory and stochastic processes find themselves today admirably well established as a fully rigorous, modern, respected, accepted branch of mathematics – pure mathematics in the first instance, but applied mathematics also in so far as the subject has proved extremely successful and flexible in application to a vast range of fields, some of which (gambling, for example, to which we return below) motivated its development, others of which (probabilistic algorithms for factorizing large integers, for example) were undreamt of in the early years of probability theory. For appreciations of how matters stood at the turn of the millenium, see Accardi and Heyde (1998), Bingham (2001a).

§2. Background on finite additivity

In parallel with all this, other approaches were developed. Note first that before Lebesgue's work on length, area and volume, such things were treated using Jordan content (Camille Jordan (1838-1922) in 1892, and in his three-volume *Cours d'Analyse* of 1909, 1913 and 1915)¹, which is finitely but not countably additive. The concept is now considered outmoded, but

¹Jordan was anticipated by Giuseppe Peano (1858-1932) in 1887: Bourbaki (1994), p.221.

was still in use in the undergraduate literature a generation ago – see e.g. Apostol (1957) (the book from which the present writer learned analysis). The Banach-Tarski paradox (Stefan Banach (1892-1945) and Alfred Tarski (1902-1983)), of which more below, appeared in 1924. It was followed by other papers of Banach, discussed below, and the development of functional analysis, the milestone book on which is Banach (1932).

The essence of the relevance of functional analysis here is duality. The dual of the space L_1 of integrable functions on a measure space is L_∞ , the space of bounded functions. The dual of L_∞ in turn is ba , the space of bounded, *finitely additive* measures absolutely continuous with respect to the measure of the measure space (Hildebrandt (1934)). For a textbook exposition, see Dunford and Schwartz (1958), Ch. III.

The theory of finitely additive measures is much less well known than that of countably additive measures, but is of great interest as mathematics, and has found use in applications in several areas. The principal textbook reference is Rao and Rao (1983), who refer to them as *charges*. The terminology is suggested by the first application areas of measure theory after length, area and volume – gravitational mass, probability, and electrostatic charge. While mass is non-negative, and probability is non-negative of total mass 1, electrostatic charge can have either sign. Indeed, the Hahn-Jordan theorem of measure theory, decomposing a signed measure into its positive and negative parts, suggests decomposing an electrostatic charge distribution into positively and negatively charged parts. The authors who made early contributions to the area include, as well as those already cited, R. S. Phillips, C. E. Rickart, W. Sierpinski, S. Ulam – and Salomon Bochner (1899-1982); see Bochner (1992), Part B. Indeed, Dorothy Maharam Stone opens her Foreword to Rao and Rao (1983) thus:

”Many years ago, S. Bochner remarked to me that, contrary to popular mathematical opinion, finitely additive measures were more interesting, more difficult to handle, and perhaps more important than countably additive ones. At that time, I held the popular view, but since then I have come round to Bochner’s opinion.”

Stone ends her foreword by mentioning that the authors plan to write a book on finitely additive probability also. This has not yet appeared, but (personal communication to the present writer) the book is still planned.

At this point it becomes necessary for the first time to mention the measure space. If this is purely atomic, the measure space is at most countable, the measure theory as such becomes trivial; all subsets of the measure

space have a probability, and for each set this may be calculated by summing over the probabilities of the singleton sets of the points it contains. (This is of course in the countably additive case; in the finitely additive case, one can have all finite sets of zero probability, but total mass 1.) It also becomes necessary to specify our axioms. As always in mathematics, we assume Zermelo-Fraenkel set theory, or ZF (Ernst Zermelo (1871-1953), Adolf Fraenkel (1891-1965) in 1922). As usual in mathematics, we assume the Axiom of Choice (AC) (Zermelo, 1904) unless otherwise specified – that is, we work with ZFC, for ZF + AC. Then, if the measure space contains a continuum, non-measurable sets exist. See Oxtoby (1971), Ch. 5 (“The oldest and simplest construction is due to Vitali in 1905”). It would be pleasant if one did not have to worry constantly about measurability problems ...

At this point, the figure of Bruno de Finetti (1906-1985), the centenary of whose birth led to the present piece and of whom more below, enters the picture. From 1930 on, de Finetti in his extensive writings, and in person, energetically advocated a view of probability as follows:

1. Probability should be finitely additive but not countably additive.
2. *All* sets should have a probability.
3. Probability is personal (or personalist, or subjective). That is, it does not make sense to ask for the probability of something *in vacuo*. One can only ask a person what *his or her* assessment of a probability is, and require that their assessments of the probabilities of different events be mutually consistent, or *coherent*.

Because the countably additive approach is so standard, it is perhaps as well to mention here some other approaches, if only to illustrate that de Finetti is not alone in rejecting the conventional approach. Measure and integral go hand in hand. Whereas the Lebesgue integral in the Euclidean case, and the measure-theoretic integral in the general case – the expectation, in the case of a probability measure – dominate, Kolmogorov (1933) is preceded by the integrals of Denjoy and Perron, and succeeded by those of Henstock and Kurzweil (see e.g. Saks (1937) and Henstock (1963) for background). Further, Kolmogorov himself (Kolmogorov (1930); Kolmogorov (1991), 100-143 in translation, and comments, 426-429) developed an integration concept, the refinement integral or Kolmogorov integral, which differs from the measure-theoretic one (see Gogvadze (1979) for a textbook account). In his later work, Kolmogorov also developed an algorithmic approach to probability, which is quite different from the standard measure-theoretic one (see Kolmogorov (1993)). Thus one need not expect uniformity of approach in

these matters.

The conventional wisdom is that finite additivity is harder than countable additivity, and does not lead to such a satisfactory theory. This view does indeed have some substance; see e.g. Dudley (1989), §3.1, Problems, and Edwards (1995), who comments (p.213) ‘Finitely additive measures can exhibit behaviour that is almost barbaric’. Furthermore, countable additivity allows one to take in one’s stride such things as, for a random variable X with the Poisson distribution $P(\lambda)$,

$$P(X \text{ odd}) = \sum_{k=0}^{\infty} P(X = k, k \text{ odd}) = \sum_{k \text{ odd}} \frac{e^{-\lambda} \lambda^k}{k!} = \frac{1}{2}(1 - e^{-2\lambda}),$$

using the power series for $e^{\pm\lambda}$. Many, including myself, are not prepared to deny themselves the freedom to proceed as above.

In his review of Rao and Rao (1983), Uhl (1984) points out that there are three ways of handling finitely additive measures. One can proceed directly, as in Rao and Rao (1983). He discusses two methods of reduction to the countably additive case (due to Stone and to Drewnowski – see below). He comments “Both the Stone representation and the Drewnowski techniques allow one to reduce the finitely additive case to the countably additive case. They both show that a finitely additive measure is just a countably additive measure that was unfortunate enough to have been cheated on its domain.” The Stone approach involves the Stone-Čech compactification (M. H. Stone (1903-1987), Eduard Čech (1893-1960), both in 1937; see e.g. Walker (1974), Hindman and Strauss (1998)), and amenability, to which we turn in the next section. The Drewnowski approach involves a subsequence principle ².

§3. The problem of measure

At the very end of the classic book Hausdorff (1914), one finds posed what Lebesgue (1905), VII.II and Banach (1923) call the *problem of measure*: is it possible to assign to *every* bounded set E of n -dimensional Euclidean space a number $m(E) \geq 0$ which adds over (finite) disjoint unions, has $m(E_0) = 1$ for some set E_0 , and has $m(E_1) = m(E_2)$ when E_1 and E_2 are congruent (superposable by translation and rotation)? Hausdorff (1914) proves that this is not possible in dimension 3 or higher. By contrast, Banach (1923) proves

²Drewnowski (1972): “The Nikodym boundedness theorem for finitely additive measures can be deduced directly from the countably additive case by this technique”: Uhl (1984), 432.

that it is possible in dimensions 1 and 2 – the line and the plane. That is, volume or hypervolume – Lebesgue measure in dimension 3 or higher – cannot be defined for *all* sets without violating either additivity over disjoint unions or invariance under Euclidean motions – both minimal requirements for using the term volume without violating the language. In other words, the mathematics of volume necessarily involves non-measurable sets. By contrast, length and area – Lebesgue measure on the line and the plane – *can* be defined for all sets, with invariance under translation (and rotation, in the plane), *provided* one is prepared to work with finite rather than countable additivity. That is, length and area need not involve non-measurable sets, if one works with finite additivity. This gives strong support to de Finetti’s programme of §2 above – *but only in dimensions 1 and 2*.

Hausdorff proves the following. The unit sphere in 3-space can, to within a countable set D , be decomposed into three disjoint sets A , B and C , congruent to each other *and* to $B \cup C$. Were it possible to assign volumes, this would give each one third the volume of the sphere by symmetry, and also two-thirds, by finite additivity. The resulting contradiction is called the *Hausdorff paradox*; see Wagon (1985), Ch. 2. He deduces the non-existence of extensions of volume to all sets from this ³.

Similar, and more famous, is the Banach-Tarski paradox (Banach and Tarski (1924); Wagon (1985), Ch. 3; see also Székely (1990), V.2). This states that in Euclidean space of dimension $n \geq 3$, any two bounded sets with non-empty interior (for example, two spheres of different radii) can be decomposed into finitely many parts, the parts of one being congruent to those of the other. A similar statement applies on the surface of a sphere. This statement is astonishing, and violates our geometric intuition. For example, we could take a sphere of radius one, decompose it into finitely many parts, move each part by a Euclidean motion (translation and rotation), and reassemble them to form a sphere of radius two. Of course, such sets must be non-measurable, or such ”volume doubling” would be a practical proposition. The Banach-Tarski result depends on the Axiom of Choice – as the construction of a non-measurable set does also; see §8 below.

The corresponding statements, however, are false in dimension one or

³Hausdorff’s classic book *Grundzüge der Mengenlehre* has three German editions (1914, 1927, 1937) and an English translation of 1957 of the third edition. Only the 1914 edition contains the result above – but the Collected Works of Felix Hausdorff are currently being published by Springer-Verlag; Volumes II (the 1914 edition above), IV, V (including probability theory, Hausdorff (2006)) and VII have already appeared.

two.

The question arises, of course, as to what it is that discriminates so sharply between dimensions $n = 1$ or 2 and $n \geq 3$. The answer is group-theoretic. Write G_n for the group of Euclidean motions in n -dimensional Euclidean space. Then G_1 and G_2 are solvable, and so can contain no free non-abelian subgroup. By contrast, G_3 does contain a free non-abelian subgroup, and similarly for higher dimensions; see Wagon (1985), Ch. 1, Ch. 2 and Appendix A. This explains the dimension-split observed above. Wagon (1985) summarizes this by saying that G_n is *non-paradoxical* for $n = 1, 2$ but *paradoxical* for $n \geq 3$.

Rather more standard than this terminology is that of *amenability*. In the above, ‘amenable’ is the negation of ‘paradoxical’; thus G_n is amenable only for $n = 1, 2$. The usual definition of amenability is in terms of the existence of an invariant mean – a left-invariant, finitely additive measure of total mass 1 *defined on all the subsets of the group*. The idea and the basic results are due to von Neumann (1929); see Wagon (1985), Ch. 10. For background, see Greenleaf (1969), Pier (1984), Paterson (1988) ⁴.

The question of extension of Lebesgue measure is so important that we mention here the results of Kakutani (1944) and Kakutani and Oxtoby (1950); see Kakutani (1986), 14-18, 36-46 and commentaries by J. C. Oxtoby (379-383) and K. A. Ross (383-384), Hewitt and Ross (1963), Ch. 4, §§16, 17 ⁵.

As we saw in §2, de Finetti’s programme involves finitely additive mea-

⁴Von Neumann used the term *meßbar*, or measurable. The modern terms are amenable in English – combining the usual connotation of the word with ‘meanable’, a pun due to Day (1950), (1957) – *mittelbar* in German, *moyennable* in French.) The subject has ramified extensively (the standard work, Paterson (1988), contains a bibliography of 73 pages). For applications to statistics (e.g. the Hunt-Stein theorem), see e.g. Bondar and Milnes (1981), Strasser (1985), §§32, 48, and, using finitely additive probability, Heath and Sudderth (1978), §4.

⁵The *character* of a measure space is the smallest cardinality of a family by which all measurable sets are approximable. Then, although there is no countably additive extension of Lebesgue measure to *all* subsets of n -space, there is an invariant extension (under left and right translations and reflection) of Lebesgue measure to a measure space of character 2^c , where c here denotes the cardinality of the continuum; similarly for Haar measure on infinite compact metric groups. The σ -field here is vastly larger than that of the measurable sets, but vastly smaller than that of all sets.

There is also a dimension-split in this area. To repeat the title of Sullivan (1981): for $n > 3$ there is only one finitely additive rotationally invariant measure on the n -sphere defined on all Lebesgue-measurable subsets.

asures defined on all sets. This needs an extension procedure for finitely additive measures, parallel to the standard Carathéodory extension procedure in the countably additive case. The standard procedure of this type is due to Los and Marczewski (1949), and is expounded in Rao and Rao (1985), §3.3. For a recent application, see Meier (2006) ⁶.

The Stone-Čech compactification of §2 above is always denoted by the symbol β . We quote (Paterson (1988), p.11): "The whole philosophy is simple: we shift from studying a bad (finitely additive) measure on a good set (G) to studying a good (that is, countably additive) measure on a complicated space βG ."

One area where the distinction between finite and countable additivity shows up most clearly is in the question of a uniform distribution over the integers. In the countably additive case, no such distribution can exist (the total mass would be infinity or zero depending on whether singletons had positive or zero measure). In the finitely additive case, such distributions do exist (all finite sets having zero measure); see e.g. Schirokauer and Kadane (2007). Three relevant properties here are extending limiting frequencies, shift invariance, and giving each residue class modulo m mass $1/m$. Calling the classes of such measures L , S and R (for limit, shift and residue), Schirokauer and Kadane show that $L \subset S \subset R$, both inclusions being proper.

Such results are relevant to at least two areas. One is Bayesian statistics, and the representation of prior ignorance. If the problem has shift invariance, so should a prior; improper priors (priors of infinite mass) are known to lead to problems such as the so-called marginalisation paradox. The other is probabilistic number theory. The Erdős-Kac central limit theorem, for example (Paul Erdős (1913-1996), Mark Kac (1914-1984) in 1939), deals with the prime divisor functions $\Omega(n)$, $\omega(n)$ (one needs two, to be able to count with and without multiplicity), and asserts that, roughly speaking, for a large integer n each is approximately normally distributed with mean and variance $\log \log n$. See e.g. Tenenbaum (1995), III.4.4 for a precise statement (and for the refinement of Berry-Esseen type, due to Rényi and Turán). Kac memorably summarized the result as saying that 'primes play a game of chance' (Kac (1959), Ch. 4). Of course, in the conventional approach via countable additivity one needs quotation marks here. Using finite additivity, one would

⁶The renewal theorem exhibits a similar dimension-split, this time between $d = 1$ (where Blackwell's theorem applies) and $d \geq 2$ (where the limit is always zero). In the group case, amenability again plays a crucial (though not on its own a decisive) role. For details and references, see Bingham (2001b), I.9 p.180.

not; we raise here the question of deriving the Erdős-Kac and Rényi-Turán results using finite additivity.

§4. De Finetti

Bruno de Finetti was born on 13.6.1906 in Innsbruck, Austria of Italian parents. De Finetti enrolled at Milan Polytechnic, where he discovered his bent for mathematics, transferred to the then new University of Milan in 1925, and graduated there in 1927 with a dissertation on affine geometry. He worked from then till 1931 – the crucial years for his intellectual development, as it turned out – at the Italian Central Statistical Institute under Gini, and then worked for an insurance company, also working part-time in academia. He returned to full-time academic work in 1946, becoming professor at Trieste, then moved to La Sapienza University of Rome in 1954, from where he retired; he died on 20.6.1985. For more on de Finetti's life, see Cifarelli and Regazzini (1996), Lindley (1986) and the autobiographical account de Finetti (1982).

De Finetti's first work of major importance is his paper de Finetti (1929a), written when he was only twenty-three, on processes with independent increments. De Finetti, with Kolmogorov, Lévy and Khintchine, was one of the founding fathers of the area of infinite divisibility and stochastic processes with stationary independent increments, now known as Lévy processes. See Bertoin (1996), Sato (1999) for modern textbook accounts.

Almost simultaneous was the work for which, first and foremost, de Finetti's name will always be remembered (at least by probabilists): *exchangeability*. A sequence $\{X_n\}_{n=1}^{\infty}$ is *exchangeable* (or interchangeable) if its distribution is invariant under permutation of finitely many coordinates. Then – *de Finetti's Theorem*, de Finetti (1929b), (1930a), (1937) – a sequence is exchangeable if and only if it is obtainable from a sequence of independent and identically distributed (iid) random variables with a random distribution function, *mixed* by taking expectations over this random distribution. That, is, exchangeable sequences have the following structure: there exists a *mixing distribution* F such that, conditional on the value Y obtained by sampling from F , the X_n are conditionally iid given Y .

Exchangeability has proved profoundly useful and important. Aldous (1985) gives an extended account of subsequent work; see also Diaconis and Freedman (1987). Kallenberg (2005) uses exchangeability (symmetry under finite permutations) as one of the motivating examples of his study of symmetry and invariance. An earlier textbook account in probability is Chow

and Teicher (1978). On the statistics side, Savage (of whom more below) was one of the first to see the importance of de Finetti's theorem for subjective or personalist probability, and Bayesian statistics. For further statistical background, see Lindley and Novick (1981).

We note that de Finetti's theorem has a different (indeed, harder) proof for finite than for infinite exchangeable sequences (see e.g. Kallenberg (2005), §§1.1, 1.2). As a referee points out, this relates to differences between finite and countable additivity.

From de Finetti (1930b), (1937) on (Dubins and Savage cite eight references covering 1930-1955)), he advocated the view of probability outlined in §2: *finitely additive, defined on all subsets, personalist*. De Finetti was fond of aphorisms, and summarized his views in the famous (or notorious) aphorism **PROBABILITY DOES NOT EXIST**.

Lindley comments in his obituary of de Finetti (Lindley (1986)), of the period 1927-31: "These were the years in which almost all his great ideas were developed; the rest of his life was devoted to their elaboration".

De Finetti and Savage both spoke at the Second Berkeley Symposium in 1950, and became friends (they were already in contact – Fienberg (2006), §5.2 and footnote 21). Savage became a convert to the de Finetti approach, and used it in his book *The foundations of statistics*, Savage (1954). Lindley writes "Savage's greatest published achievement was undoubtedly *The foundations of statistics* (1954)", and again, "Savage was the Euclid of statistics" (Savage (1981), 42-45). De Finetti credited Savage with saving his work from marginalization – whether because his views were unorthodox, or because he wrote in Italian, or both. Savage, once convinced, was able to proselytize effectively because he was a first-rate mathematician, a beautiful stylist, and possessed of both a powerful intellect and a powerful personality (W. A. Weaver, Savage (1981), 12).

De Finetti wrote over 200 papers (most in Italian and many quite hard to obtain), and four books. The first, *Probability, induction and statistics: The art of guessing* (de Finetti (1972) – 'PIS'), written in memory of Savage, gives translations into English and revisions of a number of his papers from the forties to the sixties. (Incidentally, de Finetti was aware of the dimension-split of §3: see PIS, p.122, footnote.) The next two, *The theory of probability: A critical introductory treatment*, Volumes 1 and 2 (de Finetti (1974), (1975) – 'P1', 'P2') provide a full-length treatment of his views on probability and statistics. The dedication reads: "This work is dedicated to my colleague Beniamino Segre who about twenty years ago pressed me

to write it as a necessary document for clarifying one point of view in its entirety”.

The reader will form his own view regarding finite v. countable additivity. Regarding the mathematical level of P1, P2: see the end of §8.9.9, on path-continuity of Brownian motion. The treatment is loose and informal, indeed lapsing into countable additivity at times, in contradiction of his main theme (and note that de Finetti was himself one of the pioneers of Lévy processes, of which the Brownian motion and Poisson process are the prime examples).

De Finetti’s last book, *Philosophical lectures on probability*, de Finetti (2008) (‘PLP’), is posthumous. It is an edited English translation of the text of a twelve-lecture course he gave in Italian in 1979, and was published in honour of his centenary. The tone is aggressively anti-mathematical, and in particular attacks the axiomatic method (indeed, ridicules it). This is surprising, as de Finetti was trained as a mathematician and loved the subject as a young man, and his thesis subject was in geometry, a subject linked to the axiomatic method since antiquity. But perhaps it was not written for a mathematical audience, and perhaps it was not intended for publication.

One of the leading exponents of conventional, axiomatic, measure-theoretic probability was J. L. Doob (1910-2004) (see Snell (2005), Bingham (2005) for appreciations). The contrast between Doob’s approach and de Finetti’s is well made by the following quotation from Doob: “I cannot give a mathematically satisfactory definition of non-mathematical probability. For that matter, I cannot give a mathematically satisfactory definition of a non-mathematical chair” (Snell (1997), p. 305). For further views on de Finetti, see e.g. Cifarelli and Regazzini (1996), Dawid (2004). For recent commentaries on Bayesian and other approaches to statistics, see Bayarri and Berger (2004), Howson and Urbach (2005), Howson (2008), Williamson (1999), (2007), (2008a), and on PLP, Williamson (2008b).

De Finetti had his own approach to integration, but had no particular quarrel with measure theory and integration as analysis. What he objected to was the orthodox, or Kolmogorov, use of measures of mass one and integration with respect to them as probability and expectation. It is worth remarking in this connection that, where there is a group action naturally present, one is led to the Haar measure (Lebesgue measure, in the Euclidean case). The de Finetti approach is then faced with either the problem of measure of §3 (whether or not all subsets can be given a probability will depend on the probability space), or with violating the natural invariance of the problem under the group action. Of course, invariance and equivariance

properties are very important in statistics (as well as in probability and in mathematics); see e.g. Eaton (1989).

§5. Savage

Leonard Jimmie Savage – always known as Jimmie – was born in Detroit on 20.11.1917. His great intelligence was clear very early. He had very poor eyesight, and wore glasses with thick lenses; all his life he read voraciously, despite this. He studied mathematics at the University of Michigan, graduating in 1938 and taking his doctorate in 1941, in geometry. After war work and several academic moves, during which he came into contact with John von Neumann (1903-1957), Milton Friedman (1912-2006) and Warren Weaver (1894-1978), he went to the University of Chicago in 1946, spending 1949-60, his best years, in the Department of Statistics. He returned to Michigan for 1960-64, then to Yale, where he spent the rest of his life; he died on 1.11.1971.

Savage's mathematical power and versatility are well exemplified by the results for which he is best known to mathematicians and probabilists: the Halmos-Savage theorem on sufficient statistics (Halmos and Savage (1949)), the Hewitt-Savage zero-one law (Hewitt and Savage (1955), Kallenberg (2005)) and the Dubins-Savage theorem on bold play (§6) – and for his classic book *Dubins and Savage* (1965). To statisticians, he is best known for his book *Savage* (1954) (recall Lindley's praise of this in §4 above), and for his life-long, persuasive advocacy of a subjective approach to probability and statistics⁷. A posthumous illustration of his broad range and open-ness to the ideas of others is *Savage* (1976), giving his reactions to re-reading the work of the great statistician R. A. (Sir Ronald) Fisher (1890-1962).

Note. The contrast between the first (1954) and second (1972) editions of Savage's *The Foundations of Statistics* is interesting and relevant here. Savage wrote in the Preface to the second edition, and in his 1962 paper 'Bayesian statistics' (reprinted in *Savage* (1981), p.416-449), of how he failed in the attempt to amend classical statistical practice that motivated the first edition, yet failed to acknowledge this.

Fienberg (2006), in a well-referenced paper whose viewpoint and bibliography rather complement ours here, gives a history and overview of Bayesian

⁷De Finetti is not the only figure whose work Savage promoted in the USA. He also promoted that of Louis Bachelier (1870-1946). See the Foreword by Paul A. Samuelson (1915-2009) to Davis and Etheridge (2006).

statistics. In particular, he identifies (§5.2) the 1950s as the decade in which the term "Bayesian" emerged. It was apparently first used by Fisher (1950), 1.2b, pejoratively, but came into use by adherents and opponents alike around 1960 in its modern sense, of a subjective or personal view of probability, harnessed to statistical inference via Bayesian updating from prior to posterior views. For further background here, we refer to the special issue of *Statistical Science* **19** Number 1 (February 2004), the paper Bellhouse (2004) in it celebrating the tercentenary of Bayes' birth, and for Bayesian statistics generally to a standard work such as Berger (1985) or Robert (1994).

Treatments of Bayesian and non-Bayesian (or frequentist, or classical – see §8) statistics together are given by DeGroot and Schervish (2002), Schervish (1995). This is in keeping with our own preference for a pluralist approach; see the Postscript below.

§6. Gambling

Gambling plays a special role in the history of probability theory; see e.g. David (1962), Hacking (1975), Ch. 2. We confine ourselves to one specific problem here, as it bears on the theme of our title.

Suppose one is playing an unfair gambling game, in a situation in which success (or to be more dramatic, survival) depends on achieving some specific gain against the odds. This is the situation studied by Dubins and Savage, in their classic book *How to gamble if you must, or Inequalities for stochastic processes*, depending on the edition, Dubins and Savage (1965) or (1976). The objective is to select an optimal strategy – a strategy that maximizes the probability of success (which is less than $1/2$, as the game is unfavourable). Folklore and common sense suggest that *bold play is optimal* – that is, that the best strategy is to stake the most that one can at each play, subject to the natural constraints: one cannot bet more than one has, and should not bet more than one needs to attain one's goal. The problem, with an imperfect solution, is in Coolidge (1908-9). In 1956, Dubins and Savage set themselves the task of finding a complete, rigorous proof of the optimality of bold play (at red-and-black, say, or roulette, ...). They achieved their goal, but found it surprisingly difficult. It still is.

By the early fifties (see below), Savage had come to accept the de Finetti approach, of finitely additive probability defined on all sets. How they came to adopt this approach is related by Dubins in the Preface to the 1976 edition of their book, p. iii. They discuss their approach, and reasons for it, in §2.3 – essentially, to assume less and obtain more, and to avoid technical problems

involving measurability.

The Dubins-Savage theorem – bold play is optimal in an unfair game – is a major result in probability theory, and a very attractive one. The fact that its authors derived it using finite additivity was a standing challenge to the orthodox approach using countable additivity, and its adherents. The extraordinarily successful series *Séminaire de Probabilités* began in 1966-67 in the University of Strasbourg under the academic leadership of Paul-André Meyer (1934-2003) (the most recent issue is the forty-second). Meyer, with his student M. Traki, solved the problem of proving the Dubins-Savage theorem by orthodox (countably additive) means in 1971; its publication came later, Meyer (1973). The key tool Meyer used was the *réduite* – least excessive majorant (reduction, in the terminology of Doob (1984), 1.III.4); see Meyer (1966), IX.2, Dellacherie and Meyer (1983), X.1. This concept originates in potential theory (as may be guessed from the term ‘excessive’ above and seen from the titles of the three books just cited); for a probabilistic approach, see El Karoui *et al.* (1992). It is of key importance in the theory of optimal stopping, the Snell envelope, and American options in finance.

Maitra and Sudderth (1996) (pupils of Blackwell and of Dubins) ‘provides an introduction to the ideas of Dubins and Savage, and also to more recent developments in gambling theory’ (p. 1). Dubins and Savage work with a continuum of bets; thus their sample space is the set of sequences with coordinates drawn from some closed interval (which we can take as $[0, 1]$, as the gambler has some starting capital and is aiming to achieve some goal, which we can normalize to 1). Maitra and Sudderth use countable additivity, but restrict themselves to a discrete set of possible bets, to avoid measurability problems. This decision is sensible; the book is still quite hard enough. See Bingham (1997) for a review of both books ⁸.

One thus has a choice of approach to the Dubins-Savage theorem and related results – by finite additivity or by countable additivity. Most authors have a definite preference for one or the other. In comparing the two, one should be guided by difficulty (apart from preference). This is not an easy subject, however one approaches it.

The technicalities concerning measurability mentioned above are of various kinds; one of the principal ones concerns *measurable selection theorems*.

⁸The Dubins-Savage theorem entered the textbook literature in Billingsley (1979), §7. The main tool Billingsley uses is a functional equation (his (7.30)). Such functional equations are considered further in Kairies (1997).

For an exposition, see Dellacherie (1980), §2 of his *Un cours sur les ensembles analytiques*. One’s preference in choice of approach here is likely to be swayed by one’s knowledge of, or attitude to, analytic sets. These are of great mathematical interest in their own right. They date back to M. Ya. Souslin (Suslin) (1894-1919)⁹ and his teacher N. N. Lusin (Luzin) (1883-1950) in 1916; see Phillips (1978), Rogers(1980), Part 1, §1.3 for history, Lusin (1930) for an early textbook account. They are very useful in a variety of areas of mathematics; see, for example, Hoffmann-Jorgensen’s study of automatic continuity in Rogers (1980), Part 3, and that by Martin and Kechris (1980) of infinite games and effective descriptive set theory in Part 4. Suffice it to say that analytic sets provide ”all the sets one needs” (the dictum of C. A. Rogers (1920-2005), one of their main exponents). For a probabilist, analytic sets are needed to ensure, for example, that the hitting times by suitable processes of suitable sets are measurable – that is, are random variables (*théorèmes de début*); see Dellacherie (1972), (1980) for precise formulations and proofs. Thus to a conventional probabilist, analytic sets are familiar in principle but finite additivity is not; to adherents of finite additivity the situation is reversed; ultimately, choice here is a matter of personal preference.

Analytic sets are closely related to Choquet capacities (Choquet (1953); Dellacherie (1980), §1), to which we return below. We note here that analytic sets include both measurable sets and sets with the property of Baire (Kechris (1995), 29B) – that is, sets which are ‘nice’ measure-theoretically or topologically – and that measurability and the property of Baire are preserved under the Souslin operation of analytic set theory (Rogers (1980), Part 1, §2.9).

For some purposes, one needs to go beyond analytic sets to the *projective sets*. For these, and the *projective hierarchy*, see e.g. Kechris (1995) Ch. V; we return to the projective sets in §8 below.

§7. Coherence

F. P. Ramsey (1906-1930) worked in Cambridge in the 1920s. His paper *Truth and Probability*, Ramsey (1926), was published posthumously in Ramsey (1931). Its essential message is that to take decisions coherently (that is, avoiding self-contradictory behaviour), we should maximize our expected utility with respect to some chosen utility function. Similar ideas were put forward by von Neumann in 1928; see p.1 of von Neumann and Morgenstern

⁹Suslin died tragically young of typhus aged 24

(1953). Lindley (1971), §4.8 compares Ramsey's work with Newton's, and says "Newton discovered the laws of mechanics, Ramsey the laws of human action". Whether or not one chooses to go quite this far, the work is certainly important. The principle of maximizing expected utility is widely used in decision theory under uncertainty, in statistics and in economics; see e.g. Fishburn (1970). This classical paradigm has been increasingly questioned in recent years, however; we return to alternatives to it below.

For small amounts of money (relative to the resources of the individual, or organization), utility may be equated to money. For large amounts, this is not so, the law of diminishing returns sets in, and utility functions show curvature. Indeed, the difference between the utility functions for a customer and a company is what makes insurance possible, a point emphasized in, e.g., Lindley (1971).

In mathematical finance, the most important single result is the famous Black-Scholes formula of 1973. This tells one how (under admittedly idealized assumptions and an admittedly over-simplified model) one can price financial options (European calls and puts, for instance). One reason why this result came so late was that, before 1973, the conventional wisdom was that there could be no such formula; the result would necessarily depend on the utility function of the economic agent – that is, on his attitude to risk. But *arbitrage* arguments suffice: one need only assume that agents prefer more to less (and are insatiable) – that is, that utility is money. For an excellent treatment of the mathematics of arbitrage, see Delbaen and Schachermayer (2006).

In risk management (for background on which see e.g. McNeil, Frey and Embrechts (2005)), the risk managers of firms attempt to handle risks in a coherent way – again, so as to avoid self-contradictory or self-defeating behaviour. A theory of coherent measures of risk was developed by Artzner, Eber, Delbaen and Heath (1999); see also Föllmer and Schied (2002), Föllmer and Penner (2006). The coherent risk measures of these authors are mathematically equivalent to submodular and supermodular functions (one can restrict to one of these, but it is convenient to use both here). Now "Submodular functions are well known and were studied by Choquet (1953) in connection with the theory of capacities" (Delbaen (2002), Remark, p. 5; recall from §6 the link between capacities and analytic sets). Furthermore, with Choquet capacities comes the Choquet integral, which is a *non-linear* integral; for a textbook treatment, see Denneberg (1997).

Delbaen (2002) gives a careful, thorough account of coherent risk mea-

sures. The treatment is functional-analytic (as is that of Delbaen and Schachermayer (2006)), and makes heavy use of duality arguments, and the spaces L_1 , L_∞ and ba of integrable random variables, bounded random variables and bounded finitely additive measures mentioned earlier. The point to note here is the interplay, not only between the countably additive and finitely additive aspects as above, but also between both and the *non-additive* aspects.

The most common single risk measure is Value at Risk (VaR), which has become an industry standard since its introduction by the firm J. P. Morgan in their internal risk management. This is not a coherent risk measure (Delbaen (2002), §6)¹⁰.

Much of the current interest in finite additivity is motivated by economic applications. See for example the work of Gilles and LeRoy (1992), (1997), Huang and Werner (2000) on asset price bubbles, Rostek (2010).

Choquet capacities and non-linear integration theory also find economic application, the idea being that a conservative, or pessimistic, approach to risk sees one give extra weight to unfavourable cases; see e.g. Fishburn (1988), Bassett, Koenker and Kordas (2004). For applications in life insurance – where the long time-scale means that undiversifiable risk is unavoidable, and so that a pessimistic approach, at least at first, is necessary for the security of the company – see e.g. Norberg (1999) and the references cited there. This may be viewed as a change of probability measure, in a context that predates Girsanov’s theorem and use of change of measure (to an equivalent martingale measure) in mathematical finance (Bingham and Kiesel (2004), Delbaen and Schachermayer (2006)). I thank Ragnar Norberg for this comment. As a referee points out, a general account of non-additive probabilities, sympathetic to de Finetti’s theory of coherence, is given in Walley (1991).

§8. Other set-theoretic axioms

As mentioned in §3, de Finetti’s approach differs from the standard one by using finite additivity and measure defined on *all* sets, rather than countable additivity and measure defined only on *measurable* sets. To probe the difference between these further, we must consider the axiom systems used.

To proceed, we need a digression on game theory. For a set X , imagine two players, I and II, alternately choosing an element of X , player I mov-

¹⁰Both Arzner, Eber, Delbaen and Heath (1999) and Delbaen (2002) are available on Freddy Delbaen’s home page at ETH Zürich; the reader is referred there for detail, and warmly recommended to download them.

ing first; write x_n for the n th element chosen. For some target set A of the set of sequences (finite or infinite, depending on the game) on X , I wins if the sequence $\{x_n\} \in A$, otherwise II wins. The game is *determined* if I has a winning strategy. Then, with the product topology on the space of sequences, if A is open, the game is determined; similarly if A is closed. One summarizes this by saying that all open games are determined, and so are all closed games (Gale and Stewart (1953); Martin and Kechris (1980), §1.4). In particular, all finite games (where the topology is discrete) are determined – for example, the game of Nim (Hardy and Wright (1979), §9.8). Further, all Borel games are determined (Martin and Kechris (1980), Th. 1.4.5 and §3).

To go beyond this, one must work conditionally – that is, one must specify the set-theoretic axioms one assumes. Assuming the existence of measurable cardinals as a set-theoretic axiom, analytic games are determined (Martin and Kechris (1980), Th. 1.4.6 and §4 – we must refer to §1.3 and §4.2 there for the definition of measurable cardinals). The assumption that all analytic games are determined – analytic determinacy – may thus itself be used as a set-theoretic axiom (it cannot be proved within ZF: Martin and Kechris (1980), §4.1). So too may its strengthening, projective determinacy (Martin and Steel (1989)).

One can assume more – that all sets are determined. This is the Axiom of Determinacy (AD); it is inconsistent with the Axiom of Choice AC (Mycielski and Steinhaus (1962); Mycielski (1964), (1966)). Under ZF + AD, *all sets of the line are Lebesgue measurable* (Mycielski and Swierczkowski (1964)). Thus, the problem of measure of §3 evaporates, *provided* that one is prepared to pay the price of replacing the Axiom of Choice AC by the Axiom of Determinacy AD – a price that most mathematicians, most of the time, will not be prepared to pay ¹¹.

§9. Seidenfeld’s Six Reasons

A recent study by Seidenfeld (2001) addresses the same area – comparison between finite and countable additivity – from the point of view of statistics (particularly Bayesian statistics) rather than probability theory as here. Seidenfeld advances six reasons ‘for considering the theory of finitely additive probability’. We give these below, with our responses to them.

¹¹There is a range of set-theoretic axioms involving the existence of large cardinals; see e.g., Kleinberg (1977), Kanamori (2003), Solovay (1970), (1971). For a recent survey of their relations to determinacy, see Neeman (2007).

(Note: This section and the next are included for their relevance, and at the suggestion of a referee. In other sections, I have endeavoured to write even-handedly and ‘above the fray’. But in these two, I have no choice but to give my own viewpoint, based on my experience as a working probabilist.)

Reason 1. Finite additivity allows probability to be defined on *all* subsets, even when the sample space Ω is uncountable; countable additivity does not.

Response. This is not necessarily an advantage! It is true that this deals with the problem of measure, and so with problems of measurability (though at higher cost than conventional probability theory is prepared to pay). However:

(a) Conventional (countably additive) probability theory is in excellent health (§1), despite the problem of measure (§3).

(b) It is no more intuitively desirable that one should be able to assign a probability to *all* subsets of, say, the unit square than that one should be able to assign an *area* to all such subsets. Area has intuitive meaning only for rectangles, and hence triangles and polygons, then circles by approximation by rectangles, ellipses by dilation of circles, etc. But here one is using ad hoc methods, and one is pressed to take this very far. In any degree of generality, the only method is approximation, by, say, the squares of finer and finer sheets of graph paper. This procedure fails for sets which are ‘all edge and no middle’ – which exist in profusion. Indeed, sets of complicated structure are typical rather than pathological. Our insight that ‘roughness predominates’ is deepened by the subjects of geometric measure theory and fractals; for a good introduction, we refer to Edgar (1990).¹²

Reason 2. Limiting frequencies are finitely additive but not countably additive. Seidenfeld gives the example of a countable sample space $\Omega = \{\omega_n\}_{n=1}^{\infty}$, and a sequence of repeated trials in which each outcome occurs only finitely often. Then each point ω_n has frequency zero, and these sum to 0, not 1.

Response. Write $p_i := P(\{\omega_i\})$. The most important case is that of independent replications. Then by the strong law of large numbers, there is probability 1 that $\{\omega_i\}$ has limiting frequency p_i for each i separately, and hence so too for all i simultaneously. Then for $A \subset \Omega$, the limiting frequency is, with probability 1,

$$\mu(A) = \sum_{i:\{\omega_i\} \subset A} p_i,$$

¹²A referee draws our attention here to Lévy, who wrote (in a letter to Fréchet of 29.1.1936) that probability involving an infinite sequence of random variables can only be understood through finite approximation. See Barbut et al. (2004).

and this is countably additive.

One can generalize. There are two ingredients:

- (a) some form of the strong law of large numbers, giving existence of limiting frequencies with probability one;
- (b) the Hahn-Vitali-Saks theorem – the setwise limit of a (countably additive) measure is a (countably additive) measure (Doob (1994), III.10, IX.10; Dunford and Schwartz (1958), III.7.2).

Thus, where limiting frequencies exist, they will be countably additive almost surely ¹³.

Reason 3. ‘Some important decision theories require no more than finite additivity, e.g. de Finetti (1974), and Savage (1954). However, these theories require a controversial assumption about the state-independent utility for *consequences* (see Seidenfeld and Schervish (1983) and Schervish et al. (1990))’.

Response. The first part is not at issue, nor is how far de Finetti and Savage were able to go with finite additivity (see Sections 4 and 5). The second part partially offsets the first; we must refer to the cited sources for detail.

Reason 4. ‘Two-person, zero-sum games with bounded payoffs that do not have (minimax) solutions using σ -additive probabilities do, when finitely additive mixed strategies are permitted’. Reference is made to Schervish and Seidenfeld (1996), and to the natural occurrence of “least favourable” priors from the statistician’s point of view, corresponding to finitely additive strategies for Nature (Berger (1985), 350).

Response. Again, this point is not in contention – see the Postscript below. Indeed, the point could be reinforced by referring to Kadane, Schervish and Seidenfeld (1999). From the Preface: “... if one player is limited to countably additive mixed strategies while the other is permitted finitely additive strategies, the latter wins. When both can play finitely additive strategies, the winner depends on the order in which the integrals are taken.”

Reason 5. ‘Textbook, classical statistical methods have (extended) Bayesian models that rely on purely finitely additive prior probabilities’. The use of improper priors for location- scale models in Jeffreys (1971) is cited.

¹³Seidenfeld (2001) points out that the class of events for which limits of frequencies exist need not even form a field; an example, from number theory, is in Billingsley (1979), Problem 2.15. He cites the paper Kadane and O’Hagan (1995) for ‘an interesting discussion of how such a finite but not countably additive probability can be used to model selecting a natural number “at random”’. See the discussion of Schirokauer and Kadane (2007) and of probabilistic number theory at the end of §3.

Response. It is not in contention that Bayesian statisticians often use finitely additive probabilities. The use of improper priors has often been criticized, even from within Bayesian statistics.

Reason 6. This – which occupies the bulk of Seidenfeld’s interesting paper – concerns conditioning on events of probability zero. Recall that Kolmogorov’s definition of conditioning on σ -fields is the central technical innovation of the *Grundbegriffe*, Kolmogorov (1933), and has been called ‘the central definition of modern probability theory’ (Williams (1991), 84).

The Kolmogorov approach is to work with *conditional expectations* $E(X|\mathcal{A})$, for a random variable given a σ -field \mathcal{A} (which includes conditional probabilities $P(A|\mathcal{A})$, taking X the indicator function of the event A). These are Radon-Nikodym derivatives, whose existence – guaranteed by the Radon-Nikodym theorem of §1 – is not in doubt. Seidenfeld discusses *regular conditional probabilities*, which are ‘nice versions’ of these Radon-Nikodym derivatives – but these need *not* exist. For background here, see Blackwell and Ryll-Nardzewski (1963), Blackwell and Dubins (1975), Seidenfeld et al. (2001).

Mathematically, this question is linked with the theory of *disintegration*, for which see e.g. Kallenberg (1997), Ch. 5. This is a rather technical subject; it is not surprising that treatments of it differ between the finitely and countably additive theories, nor that attitudes to it differ between the probability and statistics communities (de Finetti uses the term *conglomerability* in his approach to conditioning – see de Finetti (1974), Ch. 4, Dubins (1975), Hill and Lane (1985)).

As Seidenfeld points out, the simplest way to construct examples of non-existence of regular conditional probabilities is to take the standard (Lebesgue) probability space on $[0, 1]$ and ‘pollute’ it by adjoining one non-measurable set. Now as a general rule in probability theory, one aims to keep the measure space decently out of sight. This can usually be done. As above, it cannot be done when one uses regular conditional probabilities. This is one reason why – useful and convenient though they are when they exist – they are not often used. To a probabilist, this diminishes the force of arguments against countable additivity based on regular conditional probabilities. (Note that Seidenfeld (2001), p.176, balances these difficulties in the countably additive theory against the corresponding difficulties with finite additivity – failure of disintegration, or conglomerability.)

Note. 1. Recall that we had to mention the measure space in §3, where amenability (or existence of paradoxical decompositions) depends on the dimension.

2. Aspects depending on the measure space may be hidden. An example concerns convergence in probability and with probability one. The first implies the second, but not conversely (or the two concepts would coincide, and we would use only one). However, they *do* coincide if the measure space is purely atomic – as then there are no non-trivial null sets. Recall a standard example of convergence in probability but not with probability one on the Lebesgue space $[0, 1]$ (see e.g. Bogachev (2007), Ex. 2.2.4).

§10. Limiting frequency.

This section, which again arose from a referee’s report, aims to summarize some of the various viewpoints on limiting frequency. It may be regarded as a digression from the main text, and may be omitted without loss of continuity.

1. *Countably additive probability.*

In conventional probability (that is, using countable additivity and the Kolmogorov axiomatics), our task is two-fold. When doing probability as pure mathematics, we harness the mathematical apparatus of (countably additive) measure theory, by restricting to mass 1. When doing probability as applied mathematics, we have in mind some real-world situation generating randomness, and seek to use this apparatus to analyze this situation. To do this, we must assign probabilities (to enough events to generate the relevant σ -fields, and thence to all relevant events – i.e., all relevant measurable sets – by the standard Carathéodory extension procedure of measure theory). To do this we must understand the phenomenon well enough to make a sensible assignment of probabilities. *Assigning probabilities is an exercise in model-building.* One can adequately model only what one adequately understands.

This done, we can use the Kolmogorov strong law of large numbers, as in §9 Reason 2 above, to conclude, in the situation of independent replication, that sample means converge to population means as sample size increases, subject to the mildest conceivable restriction – “with probability 1”, or “almost surely”, or “a.s.” Of course, *some* such qualification is unavoidable. A fair coin may fall tails for the first ten tosses, or first thousand tosses, or whatever; the observed frequency of heads is then 0 in each case; the limit of 0 is 0; the expected frequency is $1/2$. The point of the strong law is that *all such exceptional cases together* carry zero probability, and so may be neglected for practical purposes.

The conventional view of the strong law is (in my own words of 1990) as follows:

”Kolmogorov’s strong law [of large numbers] is a supremely important result, as it captures in precise form the intuitive idea (the ‘law of averages’ of the man in the street) identifying probability with limiting frequency. One may regard it as the culmination of 220 years of mathematical effort, beginning with J. Bernoulli’s *Ars Conjectandi* of 1713, where the first law of large numbers (weak law for Bernoulli trials) is obtained. Equally, it demonstrates convincingly that the Kolmogorov axiomatics of the *Grundbegriffe* have captured the essence of probability.” (Bingham (1990), 54). Indeed, it is my favourite theorem.

The only point open to attack here is the dual use of ‘probability’, as shorthand for ‘measure in a measure space of mass 1’ on the one hand, and as an ordinary word of the English language on the other. The gap between the two is the instance relevant to probability of the gap between any pure mathematical theory and the real-world situations that motivate it. The prototypical situation here is, of course, that of Euclidean geometry. This deals with idealized objects – ‘points’ with zero dimensions, ‘lines’ of zero width etc., so that we cannot draw them, and if we could, we couldn’t see them. So the gap is there. But so too is the dual success, over two and a half millennia, of geometry as an axiomatic mathematical system on the one hand and as the key to surveying, navigation, engineering etc. (and now to such precision modern exotica-turned-necessities as geographical position systems or GPS). This is just one of the innumerable instances of what Wigner famously called *the unreasonable effectiveness of mathematics*.

Within its mathematical development, one does not ”define probability”, any more than one defines any other of the constituent entities of a mathematical system. One defines a probability space – or a vector space, or whatever – by its properties, or defining axioms. In particular, one most definitely does not ”define probability as limiting frequency”. That would be circular – and doomed to failure anyway. Lévy famously remarked that it is as impossible to build a mathematically satisfactory theory of probability in this way as it is to square a circle ¹⁴.

2. *Finitely additive probability.*

Most of this applies also in the finitely additive case, but here there are two changes:

¹⁴The von Mises theory of collectives, and Kolmogorov’s work on algorithmic information theory, are relevant here. See e.g. Kolmogorov (1993), Kolmogorov and Uspensky (1987), Vovk (1987), and for commentary, Bingham (1989), §7, Bingham (2000), §11.

(a) The nature of probability is now different, and so the qualification ‘a.s.’ means something different.

(b) The proofs of the theorems – strong law, law of the iterated logarithm, etc. – are different.

For strong (almost sure) limit theorems in a finitely additive setting, see e.g. Purves and Sudderth (1976), Chen (1976a), (1976b), (1977). A monograph treatment is lacking, but is expected (in the sequel to Rao and Rao (1983)).

3. *Bayesian statistics.*

Here, one represents uncertainty by a probability distribution – prior before sampling, posterior after sampling, and one updates from prior to posterior by using Bayes’ theorem. Whether a countably or a finitely additive approach is used is up to the statistician; de Finetti advocated finite additivity.

One may discuss the role of finite versus countable additivity in the context of de Finetti’s theorem (§4). But much of the thrust of laws of large numbers in this context is transferred to the sense in which, with repeated sampling, the information in the data swamps that in the prior. See e.g. Diaconis and Freedman (1986), Robert (1994), Ch. 4.

4. *Non-Bayesian statistics.*

We confine ourselves here to aspects dominated by the role of the likelihood function. This is hardly restrictive, since from the time of Fisher’s introduction of it (he used the term as early as 1912, but his definitive paper is in 1922), likelihood has played the dominant role in (parametric) statistics, Bayesian or not. Early results included *first-order asymptotics* – large-sample theory of maximum likelihood estimators, etc. More recent work has included refinements – *second-order asymptotics* (see e.g. Barndorff-Nielsen and Cox (1989), (1994)). The term ‘neo-Fisherian’ is sometimes used in this connection; see e.g. Pace and Salvan (1997), Severini (2000).

The real question arising out of the above is not so much on one’s attitude to probability, or to limit theorems for it such as laws of large numbers, as to one’s attitude to the *parametric model* generating the likelihood function. We recall here Box’s dictum: ”All models are wrong. Some models are useful.” Whether it is appropriate to proceed parametrically (in some finite – usually and preferably, small – number of dimensions), non-parametrically (paying the price of working in infinitely many dimensions, but avoiding the committal choice to a parametric model, which will certainly be at best an approximate representation of the underlying reality), or semi-parametrically, in a model with aspects of both, depends on the problem (and, of course, the

technical apparatus and preferences of the statistician).

Note. (i) One's choice of approach here will be influenced, not so much by one's attitude to Bayesian statistics as such, as to the *Likelihood Principle*. It would take us too far afield to discuss this here; we content ourselves with a reference to Berger and Wolpert (1988).

(ii) Central to non-parametric statistics is the subject of *empiricals*. This gives powerful limit theorems generalizing the law of large numbers, the central limit theorem, etc., to appropriate classes of sets and functions in higher dimensions. To handle the measurability problems, however, one needs to step outside the framework of conventional (countably additive) probability, and work instead with inner and outer measures and integrals. For textbook accounts, see e.g. van der Vaart and Wellner (1996), Dudley (1999). The point here is that the standard approach to probability does not suffice for the problems naturally arising in statistics.

(iii) The terms usually used in distinction to Bayesian statistics are 'classical' and 'frequentist', rather than non-Bayesian as here. The term classical may be regarded as shorthand for reflecting the viewpoint of, say, Cramér (1946), certainly a modern classic and the first successful textbook synthesis of Fisher's ideas in statistics with Kolmogorov's ideas in probability – as well as, say, the work of Neyman and Pearson. Against this, much of statistics pre-Fisher used prior probabilities (not always explicitly), and so may be regarded as Bayesian (but not by that name) – see Fienberg (2006) for historical background. On the other hand, the term frequentist may be regarded as justifying, say, use of confidence intervals at the 95% level by the observation that, were the study to be replicated independently many times, the intervals would succeed in covering the true value about 95% of the time, by (any version of) the law of large numbers. In real life, such multiple replication does not occur; the statistician must rely on himself and his own study, and use of the language of confidence intervals is accordingly open to question – indeed, it is rejected by Bayesians.

A rather different problem (which led to this section) is that loose use of the term frequentist may suggest that probability is being *defined* as limiting frequency – a lost endeavour, advocated by no one nowadays.

§11. Postscript

The theme of this paper is that whether one should work in a countably additive, a finitely additive or a non-additive setting depends on the problem. One should keep an open mind, and be flexible. Each is *an* approach, not

the approach. A subsidiary theme is that differing choices of set-theoretic axioms are possible, and that these matters look very different if one changes one's axioms. Again, it is a question of *an* approach, not *the* approach. Thus we advocate here a *pluralist* approach.

Bruno de Finetti was one of the profound and original thinkers of the last century, was – with Savage – one of the two founding fathers of modern Bayesian statistics, and deserves credit for (among many other achievements, particularly exchangeability) ensuring that finite additivity is of interest today to a broader community than that of functional analysts. It is a tribute to the depth of his ideas that in order to subject them to critical scrutiny, one must address foundational questions, not only in probability and in measure theory, but in mathematics itself.

De Finetti's principal focus was on the foundations of statistics, and his posthumous book PLP is on the philosophy of probability. Since such areas are even less amenable to definitive treatment than mathematics, there even more there will always be room for differences of approach.

Acknowledgements. I thank David Fremlin, Laurent Mazliak, Ragnar Norberg, Adam Ostaszewski, René Schilling and Teddy Seidenfeld for their comments. I am most grateful to the referees for their constructive, scholarly and detailed reports, which led to many improvements.

References

- ACCARDI, L. and HEYDE, C. C. (1998). *Probability towards 2000*. Lecture Notes in Statistics **128**, Springer, New York.
- ALDOUS, D. J. (1985). Exchangeability and related topics. *Sém. Probabilités de Saint-Flour XIII – 1983. Lecture Notes in Math.* **1117**, 1-198.
- APOSTOL, T. M. (1957). *Mathematical analysis: A modern approach to advanced calculus*, Addison-Wesley, Reading, MA.
- ARTZNER, P., EBER, J.-M., DELBAEN, F. and HEATH, C. (1999). Coherent measures of risk. *Math. Finance* **9**, 203-228.
- BANACH, S. (1923). Sur le problème de la mesure. *Fundamenta Mathematicae* **4**, 7-33 (reprinted in Banach, S. (1967), 66-89, with commentary by J. Mycielski, 318-322).
- BANACH, S. (1932). *Théorie des Opérations Linéaires*. Monografie Matematyczne **1**, Warsaw.
- BANACH, S. (1967). *Oeuvres, Volume I* (ed. S. Hartman and E. Marczewski), PWN, Warsaw.

- BANACH, S. and TARSKI, A. (1924). Sur la décomposition des ensembles des points en parties respectivement congruentes. *Fundamenta Mathematicae* **6**, 244-277 (reprinted in Banach, S. (1967), 118-148, with commentary by J. Mycielski, 325-327).
- BARBUT, M., LOCKER, B. and MAZLIAK, L. (2004). *Paul Lévy – Maurice Fréchet, 50 ans de correspondance mathématique*. Hermann.
- BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic techniques for use in statistics*. Chapman and Hall.
- BARNDORFF-NIELSEN, O. E. and COX, D. R. (1994). *Inference and asymptotics*. Chapman and Hall.
- BASSETT, G. W., KOENKER, R. and KORDAS, G. (2004). Pessimistic portfolio allocation and Choquet expected utility. *J. Financial Econometrics* **2**, 477-492.
- BAYARRI, M. J. and BERGER, J. O. (2004). The interplay between Bayesian and frequentist analysis. *Statistical Science* **19**, 59-80.
- BELLHOUSE, D. R. (2004). The Reverend Thomas Bayes FRS: A biography to celebrate the tercentenary of his birth. *Statistical Science* **19**, 3-43.
- BERGER, J. O. (1985). *Statistical decision theory and Bayesian analysis*, 2nd ed. Springer, New York.
- BERGER, J. O. and WOLPERT, R. L. (1988). *The likelihood principle*, 2nd ed. Institute of Mathematical Statistics, Hayward, CA (1st ed. 1984).
- BERTOIN, J. (1996). *Lévy processes*. Cambridge University Press, Cambridge (Cambridge Tracts in Math. **121**).
- BILLINGSLEY, P. (1979). *Probability and measure*. Wiley, New York (2nd ed. 1986, 3rd ed. 1995).
- BINGHAM, N. H. (1989). The work of A. N. Kolmogorov on strong limit theorems. *Th. Probab. Appl.* **34**, 129-139.
- BINGHAM, N. H. (1990). Kolmogorov's work on probability, particularly limit theorems. Pages 51-58 in Kendall (1990).
- BINGHAM, N. H. (1997). Review of Maitra and Sudderth (1996). *J. Roy. Statist. Soc. A* **160**, 376-377.
- BINGHAM, N. H. (2000). Measure into probability: From Lebesgue to Kolmogorov. *Studies in the history of probability and statistics XLVII. Biometrika* **87**, 145-156.
- BINGHAM, N. H. (2001a). Probability and statistics: Some thoughts on the turn of the millenium. Pages 15-49 in Hendricks et al. (2001).
- BINGHAM, N. H. (2001b). Random walk and fluctuation theory. Chapter 7 (p.171-213) in Shanbhag and Rao (2001).

- BINGHAM, N. H. (2005). Doob: a half-century on. *J. Appl. Probab.* **42**, 257-266.
- BINGHAM, N. H. and KIESEL, R. (2004). *Risk-neutral valuation. Pricing and hedging of financial derivatives*, 2nd ed. Springer, London (1st ed. 1997).
- BLACKWELL, D. and DUBINS, L. E. (1975). On existence and non-existence of proper, regular, conditional distributions. *Ann. Probabl.* **3**, 741-752.
- BLACKWELL, D. and RYLL-NARDZEWSKI, C. (1963). On existence and non-existence of proper, regular, conditional distributions. *Ann. Probab.* **3**, 741-752.
- BOCHNER, S. (1992). *Collected papers of Salomon Bochner*, Volume 1 (ed. R. C. Gunning), AMS, Providence, RI.
- BOGACHEV, V. I. (2007). *Foundations of measure theory*, Volumes 1,2, Springer.
- BONDAR, I. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups. *Probability Theory and Related Fields* **57**, 103-128.
- BOURBAKI, N. (1994). *Elements of the history of mathematics*. Springer.
- CHAUMONT, L., MAZLIAK, L. and YOR, M. (2007). Some aspects of the probabilistic work. *Kolmogorov's heritage in mathematics* (ed. Eric Charpentier et al.), Springer, 41-66.
- CHEN, R. W. (1976a). A finitely additive version of Kolmogorov's law of the iterated logarithm. *Israel J. Math.* **23**, 209-220.
- CHEN, R. W. (1976b). Some finitely additive versions of the strong law of large numbers. *Israel J. Math.* **24**, 244-259.
- CHEN, R. W. (1977). On almost sure convergence theorems in a finitely additive setting. *Z. Wahrschein.* **39**, 341-356.
- CHOQUET, G. (1953). Theory of capacities. *Ann. Inst. Fourier* **5**, 131-295.
- CHOW, Y.-S. and TEICHER, H. (1978). *Probability theory: Independence, interchangeability and martingales*. Springer, New York.
- CIFARELLI, D. M. and REGAZZINI, E. (1996). De Finetti's contributions to probability and statistics. *Statistical Science* **11**, 253-282.
- COOLIDGE, J. L. (1908-09). The gambler's ruin. *Annals of Mathematics* **10**, 181-192.
- CRAMÉR, H. (1946). *Mathematical methods of statistics*. Princeton University Press.
- DAVID, F. N. (1962) *Games, gods and gambling*. Charles Griffin, London.

- DAWID, A. P. (2004). Probability, causality and the empirical world: A Bayes-de Finetti-Popper-Borel synthesis. *Statistical Science* **19**, 58-80.
- DAY, M. M. (1950). Means for the bounded functions and ergodicity for the bounded representations of semigroups. *Trans. American Math. Soc.* **69**, 276-291.
- DAY, M. M. (1957). Amenable semigroups. *Illinois J. Math.* **1**, 509-544.
- DeGROOT, M. H. and SCHERVISH, M. J. (2002). *Probability and statistics*, 3rd ed., Addison-Wesley, Redwood City CA.
- de FINETTI, B. (1929a). Sulle funzioni a incremento aleatorio. *Rend. Acad. Naz. Lincei Cl. Fis. Mat. Nat.* **10**, 163-168.
- de FINETTI, B. (1929b). Funzione caratteristica di un fenomeno aleatorio. *Att. Congr. Int. Mat. Bologna* **6**, 179-190.
- de FINETTI, B. (1930a). Funzione caratteristica di un fenomeno aleatorio. *Mem. R. Acc. Lincei* **4**, 86-133.
- de FINETTI, B. (1930b). Sulla proprietà conglomerativa della probabilità subordinate. *Rend. Ist. Lombardo* **63**, 414-418.
- de FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1-68.
- de FINETTI, B. (1972). *Probability, induction and statistics: The art of guessing*. Wiley, London.
- de FINETTI, B. (1974). *Theory of Probability: A critical introductory treatment, Volume 1*. Wiley, London.
- de FINETTI, B. (1975). *Theory of Probability: A critical introductory treatment, Volume 2*. Wiley, London.
- de FINETTI, B. (1982). Probability and my life. P. 1-12 in Gani (1982).
- de FINETTI, B. (2008). *Philosophical Lectures on Probability*. Collected, edited and annotated by Alberto Mura. Translated by Hykel Hosni, with an Introduction by Maria Carla Galavotti. Synthèse Library **340**, Springer (Italian version, 1995).
- DELBAEN, F. (2002). Coherent risk measures on general probability spaces. *Advances in finance and stochastics: Essays in honour of Dieter Sondermann* 1-37, Springer, Berlin.
- DELBAEN, F. and SCHACHERMAYER, W. (2006). *The mathematics of arbitrage*. Springer, Berlin.
- DELLACHERIE, C. (1972). *Capacités et processus stochastiques*, Springer, Berlin.
- DELLACHERIE, C. (1980). Un cours sur les ensembles analytiques. Part 2 (p. 183-316) in Rogers (1980).

- DELLACHERIE, C. and MEYER, P.-A. (1975). *Probabilités et potentiel*, Ch. I-IV, Hermann, Paris (English translation *Probabilities and potential*, North-Holland, Amsterdam, 1978).
- DELLACHERIE, C. and MEYER, P.-A. (1980). *Probabilités et potentiel*, Ch. V-VIII, Hermann, Paris (English translation *Probabilities and potential B*, North-Holland, Amsterdam, 1982).
- DELLACHERIE, C. and MEYER, P.-A. (1983). *Probabilités et potentiel*, Ch. IX-XI, Hermann, Paris (English translation *Probabilities and potential C. Potential theory for discrete and continuous semigroups*, Ch. 9-13, North-Holland, Amsterdam, 1988).
- DELLACHERIE, C. and MEYER, P.-A. (1987). *Probabilités et potentiel*, Ch. XII-XVI, Hermann, Paris (see above for translation of Ch. 12, 13).
- DELLACHERIE, C., MAISONNEUVE, B. and MEYER, P.-A. (1992) *Probabilités et potentiel*, Ch. XVII-XXIV, Hermann, Paris.
- DENNEBERG, D. (1997). *Non-additive measure and integral*. Kluwer, Dordrecht.
- DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimators. *Ann. Statist.* **14**, 1-26.
- DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti style theorems in search of a theory. *Ann. inst. H. Poincaré* **23**, 397-423.
- DOOB, J. L. (1953). *Stochastic processes*, Wiley, New York.
- DOOB, J. L. (1984). *Classical potential theory and its probabilistic counterpart*. Grundle. math. Wiss. **262**, Springer, New York.
- DOOB, J. L. (1994). *Measure theory*. Springer, New York.
- DREWNOWSKI, L. (1972). Equivalence of the Brooks-Jewett, Vitali-Hahn-Saks and Nikodym theorems. *Bull. Acad. Polon. Sci.* **20**, 725-731.
- DUBINS, L. E. (1975). Finitely additive conditional probabilities, conglomerability, and disintegrations. *Ann. Probab.* **3**, 89-99.
- DUBINS, L. E. and SAVAGE, L. J. (1965). *How to gamble if you must: Inequalities for stochastic processes*. McGraw-Hill, New York (2nd ed. *Inequalities for stochastic processes: How to gamble if you must*, Dover, New York, 1976).
- DUDLEY, R. M. (1989). *Real analysis and probability*. Wadsworth & Brooks/Cole, Pacific Grove CA (2nd ed. Cambridge University Press, 2002).
- DUDLEY, R. M. (1999). *Uniform central limit theorems*. Cambridge University Press.
- DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear operators. Part I: General theory*. Interscience, Wiley, New York (reprinted 1988).

- EATON, M. L. (1989). *Group invariance applications in statistics*. Institute of Mathematical Statistics/American Statistical Association, Hayward, CA.
- EDGAR, G. A. (1990). *Measure, topology and fractal geometry*. Springer, New York.
- EDWARDS, R. E. (1995). *Functional analysis: Theory and applications*. Dover, New York.
- EL KAROUI, N., LEPELTIER, J.-P. and MILLET, A. (1992). A probabilistic approach to the réduite. *Prob. Math. Stat.* **13**, 97-121.
- EMERY, M. and YOR, M. (2006). In Memoriam: Paul-André Meyer. *Séminaire de Probabilités XXXIX. Lecture Notes in Mathematics* **1874**, Springer, Berlin.
- FIENBERG, S. A. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Analysis* **1**, 1-40.
- FISHER, R. A. (1950). *Contributions to mathematical statistics*. Wiley, New York.
- FISHBURN, P. C. (1970). *Utility theory for decision making*. Wiley, New York (reprinted, R. E. Krieger, Huntington NY, 1979).
- FISHBURN, P. C. (1988). *Non-linear preference and utility theory*. Johns Hopkins University Press, Baltimore MD.
- FÖLLMER, H. and SCHIED, A. (2002). *Stochastic finance. An introduction in discrete time*. Walter de Gruyter, Berlin.
- FÖLLMER, H. and PENNER, I. (2006). Convex risk measures and the dynamics of their penalty functions. *Statistics and Decisions* **24**, 61-96.
- GALE, D. and STEWART, F. M. (1953). Infinite games with perfect information. *Ann. Math. Statist.* **28**, 245-256.
- GANI, J. (1982). *The making of statisticians*. Springer, New York.
- GILLES, C. and LEROY, S. (1992). Bubbles and charges. *International Economic Review* **33**, 323-339.
- GILLES, C. and LEROY, S. (1997). Bubbles as payoffs at infinity. *Economic Theory* **9**, 261-281.
- GOGUADZE, D. F. (1979). *On Kolmogorov integrals and some of their applications* (in Russian). Metsniereba, Tbilisi.
- GREENLEAF, F. P. (1969). *Invariant means on topological groups*. Van Nostrand, New York.
- HACKING, I. (1975). *The emergence of probability*. Cambridge University Press.
- HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20**, 225-241.

- HARDY, G. H. and WRIGHT, E. M. (1979). *An introduction to the theory of numbers*, 5th ed., Oxford University Press, Oxford.
- HARPER, W. E. and WHEELER, G. R. (ed.) (2007). *Probability and inference. Essays in honour of Henry E. Kyburg Jr.*, Elsevier.
- HAUSDORFF, F. (1914). *Grundzüge der Mengenlehre*. Leipzig (reprinted, Chelsea, New York, 1949; *Gesammelte Werke Band II*, Springer, 2002).
- HAUSDORFF, F. (2006). *Gesammelte Werke Band V: Astronomie, Optik und Wahrscheinlichkeitstheorie*, Springer, Berlin.
- HEATH, D. C. and SUDDERTH, W. D. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6**, 333-345.
- HENDRICKS, V. F., PEDERSEN, S. A. and JORGENSEN, K. F. (2001). *Probability Theory: Philosophy, Recent History and Relations to Science* (Synthèse Library **297**). Kluwer, Dordrecht.
- HENSTOCK, R. (1963). *Theory of integration*. Butterworth, London.
- HEWITT, E. and ROSS, K. A. (1963). *Abstract harmonic analysis, Volume I*. Springer, Berlin.
- HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. American Math. Soc.* **80**, 470-501.
- HILDEBRANDT, T. H. (1934). On bounded linear functional operators. *Trans. American Math. Soc.* **36**, 868-875.
- HILL, B. M. and LANE, D. (1985). Conglomerability and countable additivity. *Sankhya A* **47**, 366-379.
- HINDMAN, N. and STRAUSS, D. (1998). *Algebra in the Stone-Čech compactification*. Walter de Gruyter.
- HOLGATE, P. (1997). The late Philip Holgate's paper 'Independent functions: Probability and analysis in Poland between the Wars'. Studies in the history of probability and statistics **XLV**. *Biometrika* **84**, 159-173.
- HOWSON, C. (2008). De Finetti, countable additivity, consistency and coherence. *British J. Philosophy of Science* **59**, 1-23.
- HOWSON, C. and URBACH, P. (2005). *Scientific reasoning: the Bayesian approach*. Open Court (1st ed. 1989, 2nd ed. 1993).
- HUANG, K. H. D. and WERNER, J. (2000). Asset price bubbles in Arrow-Debreu and sequential equilibrium. *Economic Theory* **15**, 253-278.
- JEFFREYS, H. (1971). *The theory of probability*, 3rd ed. Oxford University Press.
- KAC, M. (1959). *Statistical independence in probability, analysis and number theory*. Carus Math. Monographs **12**, Math. Assoc. America.
- KADANE, J. B. and O'HAGAN, A. (1995). Using finitely additive proba-

- bility: uniform distributions on the natural numbers. *J. Amer. Math. Soc.* **90**, 626-631.
- KADANE, J. B., SCHERVISH, M. J. and SEIDENFELD, T. (1999). *Rethinking the foundations of statistics*. Cambridge University Press.
- KAIRIES, H. H. (1997). Functional equations for peculiar functions. *Aequationes Mathematicae* **53**, 207-241.
- KAKUTANI, S. (1944). Construction of a non-separable extension of the Lebesgue measure space. *Proc. Imperial Acad. Japan* **20**, 115-119.
- KAKUTANI, S. (1986). *Shizuo Kakutani: Selected Papers, Volume 2* (R. Kallman, ed.). Birkhäuser, Boston MA.
- KAKUTANI, S. and OXTOBY, J. C. (1950). Construction of a non-separable invariant extension of the Lebesgue measure space. *Ann. Math.* **52**, 580-590.
- KALLENBERG, O. (1997). *Foundations of modern probability* (2nd ed. 2002). Springer, New York.
- KALLENBERG, O. (2005). *Probabilistic symmetries and invariance principles*. Springer, New York.
- KANAMORI, A. (2003). *The higher infinite. Large cardinals in set theory from their beginnings*, 2nd ed. Springer, New York (1st ed. 1994).
- KECHRIS, A. S. (1995). *Classical descriptive set theory*. Graduate Texts in Math. **156**, Springer.
- KENDALL, D. G. (1990). Obituary: Andrei Nikolaevich Kolmogorov (1903-1987). *Bull. London Math. Soc.* **22**, 31-100.
- KLEINBERG, E. M. (1977). *Infinite combinatorics and the axiom of determinacy. Lecture Notes in Mathematics* **612**, Springer, Berlin.
- KOLMOGOROV, A. N. (1930). Untersuchungen über den Integralbegriff. *Math. Annalen* **103**, 654-696.
- KOLMOGOROV, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin.
- KOLMOGOROV, A. N. (1991). *Selected Works. Volume I: Mathematics and Mechanics*. Kluwer, Dordrecht.
- KOLMOGOROV, A. N. (1992). *Selected Works. Volume II: Probability Theory and Mathematical Statistics*. Kluwer, Dordrecht.
- KOLMOGOROV, A. N. (1993). *Selected Works. Volume III: Information Theory and the Theory of Algorithms*. Kluwer, Dordrecht.
- KOLMOGOROV, A. N. and USPENSKY, V. A. (1987). Algorithms and randomness. *Th. Probab. Appl.* **32**, 425-455.
- LEBESGUE, H. (1902). Intégrale, longueur, aire. *Annali di Mat.* **7.3**, 231-

256.

LEBESGUE, H. (1905). *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthier-Villars, Paris (Collection de monographies sur la théorie des fonctions) (2nd ed. 1928).

LINDLEY, D. V. (1971). *Taking decisions*. Wiley, London (2nd ed. 1985).

LINDLEY, D. V. (1986). Obituary: Bruno de Finetti, 1906-1985. *J. Roy. Statist. Soc. A* **149**, 252.

LINDLEY, D. V. and NOVICK, M. R. (1981). The role of exchangeability in inference. *Ann. Statist.* **9**, 45-58.

LOS, J. and MARCZEWSKI, E. (1949). Extensions of measures. *Fundamenta Mathematicae* **36**, 267-276.

LUSIN, N. (1930). *Leçons sur les ensembles analytiques*. Gauthier-Villars, Paris (Collection de monographies sur la théorie des fonctions) (2nd ed. Chelsea, New York, 1972).

MAITRA, A. D. and SUDDERTH, W. D. (1996). *Discrete gambling and stochastic games*. Springer, New York.

MARTIN, D. A. and KECHRIS, A. S. (1980). Infinite games and effective descriptive set theory. Part 4 (p. 403-470) in Rogers (1980).

MARTIN, D. A. and STEEL, J. R. (1989). A proof of projective determinacy. *J. Amer. Math. Soc.* **2**, 71-125.

McNEIL, A. J., FREY, R. and EMBRECHTS, P. (2005). *Quantitative risk management*. Princeton University Press, Princeton NJ.

MEIER, M. (2006). Finitely additive beliefs and universal type spaces. *Ann. Probab.* **34**, 386-422.

MEYER, P.-A. (1966). *Probability and potentials*, Blaisdell, Waltham MA.

MEYER, P.-A. (1973). Réduite et jeux de hazard. *Sém. Prob. VII. Lecture Notes in Math.* **321**, 155-171.

von MISES, R. (1928). *Wahrscheinlichkeit, Statistik und Wahrheit*. Springer, Berlin (3rd ed., reprinted as *Probability, statistics and truth*, Dover, New York, 1981).

MYCIELSKI, J. and STEINHAUS, H. (1962). A mathematical axiom contradicting the axiom of choice. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Aston. Phys.* **10**, 1-3.

MYCIELSKI, J. (1964). On the axiom of determinateness. *Fund. Math.* **54**, 67-71.

MYCIELSKI, J. (1966). On the axiom of determinateness II. *Fund. Math.* **66**, 203-212.

MYCIELSKI, J. and SWIERCZKOWSKI, S. (1964). On the Lebesgue mea-

- surability and the axiom of determinateness. *Fund. Math.* **54**, 67-71.
- NEEMAN, I. (2007). Determinacy and large cardinals. *Proc. Internat. Congr. Math. Madrid 2006, Vol. II*, 27-43. European Math. Soc. Publishing House.
- von NEUMANN, J. (1929). Zur allgemeinen Theorie des Maßes. *Fundamenta Mathematicae* **13**, 73-116 (reprinted as p. 599-642 in *John von Neumann: Collected works, Volume 1: Logic, theory of sets and quantum mechanics* (ed. A. H. Taub), Pergamon Press, Oxford, 1961).
- von NEUMANN, J. and MORGENSTERN, O. (1953). *Theory of games and economic behaviour*, 3rd ed. Princeton University Press (1st ed. 1944, 2nd ed. 1947).
- NORBERG, R. (1999). A theory of bonus in life insurance. *Finance and Stochastics* **3**, 373-390.
- OXTOBY, J. C. (1971). *Measure and category*. Springer, New York (2nd ed. 1980).
- PACE, L. and SALVAN, A. (1997). *Principles of statistical inference from a neo-Fisherian perspective*. World Scientific, Singapore.
- PATERSON, A. L. T. (1988). *Amenability*. American Math. Soc., Providence RI.
- PHILLIPS, E. R. (1978). Nicolai Nicolaevich Luzin and the Moscow school of the theory of functions. *Historia Mathematica* **5**, 275-305.
- PIER, J.-P. (1984). *Amenable locally compact groups*. Wiley, New York.
- PURVES, R. A. and SUDDERTH, W. D. (1976). Some finitely additive probability. *Ann. Probab.* **4**, 259-276.
- RAMSEY, F. P. (1926). Truth and probability. Ch. VII in Ramsey (1931).
- RAMSEY, F. P. (1931). *The foundations of mathematics*. Routledge & Kegan Paul, London.
- RAO, K. P. S. Bhaskara and RAO, M. Bhaskara (1983). *Theory of charges*. Academic Press, London.
- ROBERT, C. P. (2001). *The Bayesian choice: A decision-theoretic motivation*, 2nd ed. Springer, New York (1st ed. 1994).
- ROGERS, C. A. (1980). *Analytic sets*. Academic Press, London.
- ROSTEK, Marzena J. (2010). Quantile maximization in decision theory. *Review of Economic Studies* **77**, 339-371.
- SAKS, S. (1937). *Theory of the integral*, 2nd ed., Hafner, New York (1st ed. Monografie Matematyczne, Warsaw, 1933).
- SATO, K.I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

- SAVAGE, L. J. (1954). *The foundations of statistics*. Wiley, New York (2nd ed. 1972).
- SAVAGE, L. J. (1976). On re-reading R. A. Fisher (with discussion) (ed. J. W. Pratt). *Annals of Statistics* **4**, 441-500.
- SAVAGE, L. J. (1981). *The writings of Leonard Jimmie Savage: A memorial selection*. American Statistical Association, Washington DC.
- SCHERVISH, M. J. (1995). *Theory of statistics*. Springer, New York.
- SCHERVISH, M. and SEIDENFELD, T. (1996). A fair minimax theorem for two-person (zero-sum) games involving finitely additive strategies. In *Bayesian analysis in statistics and econometrics* (ed. D. Berry, C. Chaloner and J. Geweke). Wiley, New York.
- SCHERVISH, M., SEIDENFELD, T. and KADANE, J. B. (1990). State-dependent utilities. *J. Amer. Math. Soc.* **85**, 840-847.
- SCHIROKAUER, O. and KADANE, J. B. (2007). Uniform distributions on the natural numbers. *J. Theoretical Probability* **20**, 429-441.
- SEIDENFELD, T. (2001). Remarks on the theory of conditional probability: Some issues of finite versus countable additivity. Pages 167-178 in Hendricks et al. (2001).
- SEIDENFELD, T. and SCHERVISH, M. (1983). A conflict between finite additivity and avoiding dutch book. *Phil. Sci.* **50**, 398-412.
- SEIDENFELD, T., SCHERVISH, M. and KADANE, J. B. (2001). Improper regular conditional distributions. *Ann. Probab.* **29**, 1612-1624.
- SEVERINI, T. A. (2000). *Likelihood methods in statistics*. Oxford University Press.
- SHAFER, G. and VOVK, V. (2001). *Probability and finance: It's only a game!* Wiley, New York.
- SHAFER, G. and VOVK, V. (2006). The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science* **21**, 70-98.
- SHANBHAG, D. N. and RAO, C. R. (2001). *Stochastic processes: Theory and methods*. Handbook in Statistics **19**, North-Holland, Amsterdam.
- SNELL, J. L. (1997). A conversation with Joe Doob. *Statistical Science* **12**, 301-311.
- SNELL, J. L. (2005). Obituary: Joseph Leonard Doob. *J. Applied Probability* **42**, 247-256.
- SOLOVAY, R. M. (1970). A model for set theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics* **92**, 1-56.
- SOLOVAY, R. M. (1971). Real-valued measurable cardinals. *Proc. Symp. Pure Math. XIII Part I* 397-428, American Math. Soc., Providence RI.

- STRASSER, H. (1985). *Mathematical theory of statistics: Statistical experiments and asymptotic decision theory*. De Gruyter Studies in Mathematics **7**, Walter de Gruyter, Berlin.
- SULLIVAN, D. (1981). For $n > 3$ there is only one finitely additive rotationally invariant measure on the n -sphere defined on all Lebesgue-measurable subsets. *Bull. Amer. Math. Soc.* **4**, 121-123.
- SZÉKELY, G. J. (1990). *Paradoxa: Klassische und neue Überraschungen aus Wahrscheinlichkeitsrechnung und mathematischer Statistik*. Akadémiai Kiadó, Budapest.
- TENENBAUM, G. (1995). *Introduction to analytic and probabilistic number theory*. Cambridge University Press.
- UHL, J. J. (1984). Review of Rao and Rao (1983). *Bull. London Math. Soc.* **16**, 431-432.
- van der VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes, with applications to statistics*. Springer.
- VOVK, V. G. (1987). The law of the iterated logarithm for Kolmogorov or chaotic sequences. *Th. Probab. Appl.* **32**, 412-425.
- VOVK, V. G. (1993). A logic of probability, with applications to the foundations of statistics (with discussion). *J. Roy. Statist. Soc. B* **55**, 317-351.
- WAGON, S. (1985). *The Banach-Tarski paradox*. Encyclopaedia of Mathematics and its Applications **24**, Cambridge University Press.
- WALKER, R. C. (1974). *The Stone-Ćech compactification*. Ergebnisse Math. **83**, Springer.
- WALLEY, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London.
- WILLIAMS, D. (1991). *Probability with martingales*. Cambridge University Press.
- WILLIAMSON, J. (1999). Countable additivity and subjective probability. *British J. Philosophy of Science* **50**, 401-416.
- WILLIAMSON, J. (2007). Motivating objective Bayesianism: From empirical constraints to objective probabilities. In Harper and Wheeler (2007).
- WILLIAMSON, J. (2010a). *In defence of objective Bayesianism*. Oxford University Press.
- WILLIAMSON, J. (2010b). Review of de Finetti (2008). *Philosophia Mathematica* **18**, 130-135.

Department of Mathematics, Imperial College London, London SW7 2AZ
n.bingham@ic.ac.uk nick.bingham@btinternet.com