Countable additivity in the philosophical foundations of probability

Colin Elliot

Abstract

In this thesis, I study an open problem in the current philosophy of science: should countable additivity be an axiom of probability? We say that a probability function is finitely additive if the probability of a union of two (or any finite number of) events is equal to the sum of the single probabilities of each event. This accepted by all schools of thought on probability. Countable additivity just extends this property to countably infinite unions of events. That this should hold is a hotly debated issue. In mathematics, probability is defined axiomatically, its properties prescribed without need of justification. Countable additivity is used in almost all modern mathematical probability, because of the powerful integration technique, and convergence theorems it makes possible. Many philosophers object that it is hard to justify this adoption, and that the principle makes it impossible, amongst other things, to model a Humean scepticism towards induction, and impossible to follow some very basic intuitions which regard uniform distribution of probability over all possible events. Having examined the available philosophical arguments, I reach the conclusion that they all, on both sides of the debate, sometimes openly but sometimes not, crucially rely on two deep intuitions which are simply incompatible: one regards additivity, the other regards (the possibility of) uniformity between probability values. Given that any argument for or against the principle of countable additivity must contrast one of these two intuitions, this explains why the debate is still open, and will most likely stay that way. Finally, I examine a recent attempt at solving the deadlock, which makes use of non-standard analysis, at the price of losing real-valued probabilities and our usual idea of sum.

Contents

1 Introduction

The debate which I explore in my thesis touches upon a number of different areas of mathematical and philosophical reasoning. The dilemma around which it is centred, however, is very simple. As many have done before, I present it by an idealised example first introduced by Bruno de Finetti, a pioneer of probability theory, active from the late 1920s to the 1980s, who discussed it in many of his works. Strikingly, solutions to this example are still sought to this day: I present in Chapter 6 an approach to it published in 2013. The idealised scenario is often known as the 'de Finetti lottery', and it is as follows. Let us imagine a lottery over all the natural numbers, each number representing a ticket. We know that one, and only one, ticket will be picked, but of course we do not know which one. What are the chances of one given ticket of winning? We would obviously say the chances are very small. We might even say each ticket has, in fact, no chances of winning. Let us represent these chances numerically, as is convention in probability, with 1 for events which are *certain*, or *almost certain*, and 0 for events which are *impossible*, or *almost impossible*.[1] Informally, the commonly accepted axioms, or rules that govern probability, are the following. For now, let E represent a proposition whose truth value is uncertain, or an event whose occurrence is uncertain, and interpret P as an informal measure of this uncertainty; let $A \cup B$ be the union of two events, or disjunction of two propositions: this represents the event that A is the case, or that B is the case, or that both are, if this is possible. We have:

(1) $P(E) \geq 0$, with $P(E_0) = 0$ if E_0 is an (almost) impossible event or proposition;

(2) $P(E_1) = 1$, if E_1 is an (almost) certain event or proposition;

(3) If A and B are two mutually exclusive events, or incompatible propositions, then
$$P(A \cup B) = P(A) + P(B).$$

The first two rules are just the convention that certainty is represented by 1, impossibility by 0, and all other uncertain values are somewhere in between. Note also the following: suppose we have an uncertain event H_1 which 'includes' another one, H_2; these could be, for example, the event that the roll of a die will result in an even number, compared to the event that exactly 2 will be the result. Then in general $P(H_1) = P(H_1 \cup H_2) = P(H_1 \setminus H_2) + P(H_2) \geq P(H_2)$, since all probabilities are non-negative.[2] Hence we also see from the rules that an event or proposition which encompasses more than another event, is also more likely than it. This seems common sense. We obtained it by using rule (3). This rule is called **finite additivity**, and is accepted as a desirable feature of probability by all schools of thought on the matter. It says that the probability of two mutually exclusive events is the sum of their probabilities. For example, if the number 2 has probability $\frac{1}{6}$ of resulting from the roll of a die, and 5 has the same probability, then the event that either 2 or 5 will come out is $\frac{2}{3}$. This extends easily to any finite number of events: indeed, now consider the event that '2 or 5' will come out, together with the event that 3 will. We already

[1] We will see in Chapter 2 why we introduce the 'almost'.

[2] The set $H_1 \setminus H_2$ is the set of all events, or elements, which are in H_1 but not in H_2.

know the probability of the first event, and also know that it is incompatible to the second, so we use rule (3) again. These matters will be treated in a more complete manner below. For now, let us return to our lottery.

Suppose we attach a probability to each number in the infinite lottery. What should it be? Should all numbers be considered equally probable? Suppose we answer yes to the second question. The problem is that any real number is too big. Recall rules (3) and (2) above: the sum of any finite number of probabilities should be less than 1. But however small we choose these positive probabilities, by summing them we will eventually, in a finite number of steps, achieve a number greater than 1. So, if we want to preserve the fact that all numbers are equally probable, they must all be assigned probability 0. This is fine by the three rules above: nowhere do they state that an *infinite* union of exclusive events must have the same probability as the infinite sum of the probabilities of the single events. But precisely such a rule is adopted in all modern mathematical probability: it is called **countable additivity**. We write it as follows:

(4) Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of mutually exclusive events, all of which have a well-defined probability value; then:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

The immediate consequence of adopting countable additivity is that any uniform distribution of probabilities over all tickets in the infinite lottery is now impossible. As above, any small number is too large; and now a uniform distribution of 0s is also ruled out, for their countable sum is just 0, when this value should be 1: all tickets together make up the certain event, because one ticket *will* be picked. For an infinite number of probabilities to add up to 1, they must form a convergent series. This means that the probabilities must form a sequence converging to 0. This in turn means that, whatever sequence we choose to adopt, we will always have the vast majority of the probability assigned to a finite set of numbers in the lottery. This contradicts this seemingly obvious intuition:

> "We should be able to assign equal probability to all events, including in an infinite setting"

On the other hand, it is also clear that the solution of assigning 0 probability to all events, when their union has probability 1, is also highly counter-intuitive: we lose the idea that the total probability is the sum of its composing parts; and we would have a sum of impossible events making up a certain event. The intuition it contradicts can be expressed thus:

> "The probability of a union of events, should be equal to the sum of the probability of each event, including in an infinite setting"

These two intuitions are plainly in contrast: if we have a countably infinite number of events, they cannot be both valid at the same time. Hence we must choose which one to adopt and which one to drop. Perhaps the most influential writer on this subject was Bruno de Finetti (1906-1985), who was a firm, vocal and life-long opponent of adopting countable additivity as an axiom. His main motivation in this was his belief in the first intuition above, and the solution to the infinite lottery which this allows. More discussion of his work and ideas is to come. In the chapters that follow, I explore some consequences of adopting or not adopting countable additivity, and what the philosophical foundations can tell us about this choice. By philosophical foundations I mean, broadly speaking, what we understand probability to be: for example, (i) the frequencies of successes of an experiment in the long run; (ii) a personal expression of ignorance or uncertainty; (iii) something which derives directly from intuitive axioms and logic; (iv) or something really existing in the world. Interpretations (ii) and (iii) will be treated in some detail, because they have the potential to shine a light on the matter. We can ask, for example: if probabilities are nothing but an agent's degrees of uncertainty, should they then be countably additive? If probability derives somehow from logic, should it be countably additive? These questions are addressed in Chapter 4 and Chapter 5 respectively. Chapter 2 serves as a more general background to the issue, where we see just why countable additivity is so important for mathematics, we briefly address interpretation (i) above, and we study some immediate consequences of the axiom, together with some proposed solutions; in Chapter 3 we see the consequences that adopting the axiom can have in epistemology; and Chapter 6 contains a recent proposal on how to solve the deadlock surrounding the adoption of countable additivity. My conclusions are in Chapter 7. I can anticipate, at the cost of ruining the surprise at the end of what I hope is an interesting and wide-ranging debate, that I found the available foundational arguments, for and against the adoption of the principle, unable to decide the case. What is more, I found the two intuitions outlined above to be ever-present, as an implicit or explicit backdrop to nearly all arguments. Therefore, I conclude that the very structure of the problem does not allow for a solution which will satisfy both intuitions.

2 Mathematical probability and immediate consequences of countable versus finite additivity

2.1 Setting

This chapter collects information about the discussion surrounding the axiom, as a preliminary of sorts to the rest of the discussion I present. It forms an extended 'setting of the scene' for the problem I address in my work. The chapters that follow contain a more detailed analysis of some of the perceived problems in adopting the axiom, some foundational arguments for or against it, and a proposed way out of the deadlock. Before this, however, I think it is good to know the following: why the axiom is so important for mathematical probability; some immediate consequences of the adoption of the axiom for an intuitive property of probability, and an intuitive understanding of probability; and why it is relevant to think about the consequences of the axiom when we consider a probabilistic view of induction. These matters are treated in this order in the sections below.

2.2 Measure-theoretic probability, countable additivity and integration

In this section, I outline some basics of measure-theoretic probability, which is by far the dominating approach in modern mathematical probability (see Bingham in [7], or the textbooks used in the present work, [3] and [9], but in general most graduate texts on the subject). Far from aiming at formal completeness, I just wish to explain why countable additivity is so fundamentally important for this discipline. In this section I follow a recent textbook by Cohn [9]. (For a full treatment, I refer the reader to the textbooks just cited, or any other on measure theory or measure-theoretic probability.) As we will see, if we give up on countable additivity, it is not only advanced applications that we must forsake: the principle is needed for the very definition of *Lebesgue integral.* Most often, the role of countable additivity is not highlighted—simply because the principle is given in the axiomatic definition of probability, and holds no 'special' position with respect to the other axioms. In the context of my thesis, following the thread of countable additivity in the foundations of measure-theoretic probability will help explain why the vast majority of mathematicians adopt the principle, and why the finitely additive probability demanded by de Finetti has not achieved a great deal of success. De Finetti's own position with regards to integration, however, helps put his position into perspective: he was not *against* countably additive probability; he merely thought finitely additive measure should be considered probability measures too. In this section I first give some definitions which are needed for the general discussion; then I explain why the definition of integral does not make sense without countable additivity; finally, I give de Finetti's view on integration.

Definitions. The following are all taken from [9]. Here is the definition of σ**-algebra**. This is a collection $\mathcal{A}$ of subsets of an arbitrary set S, such that:

1. $S \in \mathcal{A}$;

2. If $A \in \mathcal{A}$, then $A^C \in \mathcal{A}$, where A^C is the complement of A;
3. For any infinite sequence $\{A_n\} \in \mathcal{A}$, we have that the union of all the sets in the sequence is also in the σ-algebra: $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$.

An important σ-algebra for probability theory is the Borel σ-algebra in $\mathbb{R}$. It is the σ-algebra generated by the collection of open sets of $\mathbb{R}$ (where the σ-algebra 'generated by G' is the smallest σ-algebra which contains the collection of sets G).

A **measure** is a function μ which maps sets of the σ-algebra $\mathcal{A}$ to values in $[0, \infty]$, has $\mu(\emptyset) = 0$, and is **countably additive**, which means

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n), \tag{1}$$

for any sequence of disjoint sets $\{A_n\} \in \mathcal{A}$. A **measure space** is $(S, \mathcal{A}, \mu)$, or: a set S, a σ-algebra defined upon the subsets of S, and a measure μ defined upon the sets of such σ-algebra. A **probability space** is a measure space $(\Omega, \mathcal{A}, P)$ such that $P(\Omega) = 1$. A set E with $E \in \mathcal{A}$ is called an event, and $P(E)$ is called the probability of such event. Because, in this context, an event with probability 0 need not be an empty set, we call such event 'almost impossible'. Similarly, we call en event having probability 1 'almost certain', as it need not be the whole sample space. A **measurable function**, is a function f between two measurable spaces $(S_1, \mathcal{A}_1)$ and $(S_2, \mathcal{A}_2)$ (sets which are associated with a σ-algebra of their subsets), such that for all sets $E_2 \in \mathcal{A}_2$, we have that $f^{-1}(E_2) \in \mathcal{A}_1$. In the context of probability, such functions are called **random variables**, and are the functions $X : \Omega \to \mathbb{R}$, where the σ-algebra associated to $\mathbb{R}$ is the Borel σ-algebra defined above. This is actually intuitive: a random variable can be understood as a numerical observation (it has values in $\mathbb{R}$), whose probability we know because each observation can be mapped back to a set which has a probability value. The **distribution** of a random variable X is a probability measure defined directly on the Borel sets B of $\mathbb{R}$ by taking directly $P(X^{-1}[B])$ as a value [9, pp.307-308]. Next I will describe finite additivity, and the common attitude taken by mathematicians towards the issue of finite versus countable additivity. After that, an outline of Lebesgue integration will follow, to explain how this relies on countable additivity.

Finite additivity. Finite additivity is the following property, valid for any finite $N \in \mathbb{N}$:

$$\mu\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu(A_n). \tag{2}$$

Any countably additive measure is also finitely additive: we simply take the finite sequence $\{A_n\}_{n=1}^{N}$ and extend it by an infinite sequence of empty sets; the measure of an empty set is 0, so the countable sum will be just be $\mu(A_1) + \cdots + \mu(A_N) + 0 + 0 + \ldots$. However, a finitely additive probability might not be countably additive. Suppose the mutually exclusive sets $\{A_n\}_{n=1}^{\infty}$ form a partition of the set S (upon which the σ-algebra is defined). Then if μ is only finitely additive, we can consistently have that for each i, $\mu(A_i) = 0$, so that for any finite N,

$$\mu\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu(A_n) = 0 \tag{3}$$

while if we take a *countable* union we have

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = 1. \tag{4}$$

This particular aspect of finite additivity can be said to be at the centre of the philosophical debate on which I focus in my thesis. The question is whether we should allow measures that are only finitely additive to be considered 'probabilities' or not. Cohn exemplifies what I found to be the typical response of mathematicians toward the issue:

> Finite additivity might at first seem to be a more natural property than countable additivity. However, countably additive measures on the one hand seem to be sufficient for almost all applications and, on the other hand, support a much more powerful theory of integration than do finitely additive measures. Thus we will follow the usual practice and devote almost all of our attention to countably additive measures [9, p.7].

Here is Halmos, in his classical textbook on the subject:

> Countable additivity is [...] a restriction without which modern probability theory could not function. It is a tenable point of view that our intuition demands infinite additivity just as much as finite additivity. At any rate, however, infinite additivity does not contradict our intuitive ideas, and the theory built on it is sufficiently far developed to assert that the assumption is justified by its success [24, p. 187].

It was perhaps put most succinctly and famously by Kolmogorov, who was among the first to provide measure-theoretic axioms for probability, and is considered the father of such approach in mathematics; we will discuss his view in more detail in Section 5.4, so we omit it here.

Integration. Integration has a fundamental importance in probability theory, and perhaps the main reason probability is studied as a part of measure theory, is that we can use the powerful technique of Lebesgue integration, which I will outline below. Integration is used to calculate the expected value of a random variable (indeed, the expectation is *defined* as the Lebesgue integral of a random variable (as defined above), with respect to the measure given by its distribution), and related values such as the variance. Lebesgue integration of measurable functions is defined in a natural way, starting from the concept of measure we described above. Here is a brief outline, mostly taken from [3]. We start from the integral of a **simple function**, which is a function that only assumes a finite number (say n) of values. A simple function can be written as

$$g(x) = \sum_{k=1}^{n} y_k \mathbf{1}_{A_k}(x),$$

where $\mathbf{1}_{A_k}(x)$ is the indicator, or characteristic, function of the set A_k, defined as

$$\mathbf{1}_{A_k}(x) = \begin{cases} 1 & \text{if } x \in A_k \\ 0 & \text{if } x \notin A_k \end{cases}$$

and y_k are the values assumed by the function g. The Lebesgue integral of such a function, with respect to the probability measure μ, is defined as follows:

$$\int g \, \mathrm{d}\mu = \sum_{k=1}^{n} y_k \mu(A_k).$$

This is a natural definition, which preserves the intuition that the integral is the area under the graph of a function: we are multiplying the values assumed by the function g, by the measure of the intervals over which these values are assumed. The Lebesgue integral of a non-negative measurable function f is the following:

$$\int f \, \mathrm{d}\mu = \sup \left\{ \int g \, \mathrm{d}\mu \; : g \leq f \text{ and } g \text{ is a non-negative simple function} \right\} \quad (5)$$

The integral of a general measurable function is the integral of the positive part of the function, minus the integral of the negative part (if such an expression makes sense, i.e. if they are not both infinite).

Continuity and countable additivity. How does the above rely on countable additivity? As we shall see shortly, definition 5 needs a property sometimes called **continuity** of measures in order to make sense. This is equivalent to countable additivity. I show a proof for this, before going on to explain how it is used in the definition of integral. We are again working in the measure space $(S, \mathcal{A}, \mu)$. The continuity of measures is defined as follows: let $\{A_k\}$ be an increasing sequence of sets in the σ-algebra $\mathcal{A}$. This means that for all k, $A_k \subset A_{k+1}$, and so $\lim_{k\to\infty} A_k = \bigcup_{k=1}^{\infty} A_k$. By definition of σ-algebras, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$. Continuity is the following property:

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \mu(A_k). \quad (6)$$

We now show it is equivalent to countable additivity. The proof is similar in [9], [24] and [33].

Countable additivity $\Rightarrow$ continuity. Define the sequence of sets $\{B_i\}$ as follows: $B_1 = A_1$, and $B_i = A_i - A_{i-1}$ for $i > 1$. Then the sets B_i are disjoint, and are in $\mathcal{A}$. We have that $A_k = \bigcup_{i=1}^{k} B_i$, and so $\bigcup_{k=1}^{\infty} A_k = \bigcup_{i=1}^{\infty} B_i$. Therefore we can write:

$$\begin{aligned}
\mu\left(\bigcup_{k=1}^{\infty} A_k\right) &= \mu\left(\bigcup_{i=1}^{\infty} B_i\right) \\
&= \sum_{i=1}^{\infty} \mu(B_i) \quad \text{by countable additivity} \\
&= \lim_{k\to\infty} \sum_{i=1}^{k} \mu(B_i) \\
&= \lim_{k\to\infty} \bigcup_{i=1}^{k} B_i \quad \text{(by finite additivity)} \\
&= \lim_{k\to\infty} \mu(A_k).
\end{aligned}$$

Continuity ⇒ countable additivity. Now let $\{E_n\}$ be a disjoint sequence of sets in $\mathcal{A}$. Of course, $\bigcup\limits_{n=1}^{\infty} E_n \in \mathcal{A}$. Now define $F_n = \bigcup\limits_{i=1}^{n} E_i$. This is an increasing sequence of sets, and so we can apply continuity:

$$\begin{aligned}
\mu\left(\bigcup_{i=1}^{\infty} E_i\right) &= \lim_{n\to\infty} \mu(F_n) \quad \text{by continuity} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(E_i) \quad \text{(by finite additivity)} \\
&= \sum_{i=1}^{\infty} \mu(E_i).
\end{aligned}$$

This shows that continuity implies countable additivity, and vice versa. We now go back to the definition of integral to see why this is relevant.

Integration and continuity of measures. As I anticipated above, hidden in the definition of integral (5) is the requirement for continuity. I now explain why this is the case: suppose, in that definition, that f is also a simple function. But we already had a definition of integral of a simple function, and here we are claiming that this integral is equal to the following: the largest integral of a non-negative simple function g, such that $g \leq f$. In particular, suppose $\{g_n\}$ is a non-decreasing sequence of non-negative simple functions, such that $\lim\limits_{n\to\infty} g_n = f$. By a property of integrals of simple functions (which I do not prove here), we have that $g_n \leq g_{n+1}$ implies $\int g_n \,\mathrm{d}\mu \leq \int g_{n+1} \,\mathrm{d}\mu$. Therefore, in this case definition 5 dictates that $\int f \,\mathrm{d}\mu = \lim\limits_{n\to\infty} \int g_n \,\mathrm{d}\mu$. That this is the case is guaranteed by continuity of the measure μ, and hence countable additivity. We see this in the proof of the following theorem, which Cohn introduces before giving the full definition of Lebesgue integral, explaining it is necessary for such definition. It is Proposition 2.3.2 in Cohn's book [9, pp.54-55]. With all premises and notation of this paragraph, the theorem says the following: if $\lim\limits_{n\to\infty} g_n = f$, meaning g_n converges to f point-wise for each $x \in S$, then indeed $\int f \,\mathrm{d}\mu = \lim\limits_{n\to\infty} \int g_n \,\mathrm{d}\mu$. The proof by Cohn is as follows: because of the property of integrals mentioned above, we know that, since $\{g_n\}$ is a non-decreasing sequence, $\int g_n \, \mathrm{d}\mu \leq \int g_{n+1} \, \mathrm{d}\mu \leq \cdots \leq \int f \, \mathrm{d}\mu$. Hence $\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$. If we prove that the reverse inequality is also true, then equality must hold between the two expressions. This is what we prove now. We want to construct another sequence of non-negative simple functions, $\{h_n\}$, such that for all n it holds that $h_n \leq g_n$, and with $\lim_{n\to\infty} h_n = (1-\varepsilon) \int f \, \mathrm{d}\mu$, where ε is an arbitrary number number in $(0,1)$. Because $\int h_n \, \mathrm{d}\mu \leq \int g_n \, \mathrm{d}\mu$, we will have $(1-\varepsilon) \int f \, \mathrm{d}\mu \leq \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu$, and this in turn means $\int f \, \mathrm{d}\mu \leq \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu$, because ε is arbitrary. We now need to construct such sequence $\{h_n\}$. Recall that f is a non-negative simple function, and suppose $a_1, \ldots, a_k$ are the non-zero values that f assumes. Suppose it takes on these values in the sets $A_1, \ldots, A_k$ respectively. Then we can write $f = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}$. Now for each n and i we define:

$$A(n,i) = \{x \in A_i : g_n(x) \geq (1-\varepsilon)a_i\}$$

Each set $A(n,i) \in \mathcal{A}$, and for each i the sequence $\{A(n,i)\}$ is non-decreasing, meaning that we have $A(n,i) \subseteq A(n+1,i)$. This is because g_n converges to f point-wise at each $x \in S$, and so for each i, g_n reaches the value a_i in the limit. But this just means that, because $(1-\varepsilon)a_i < a_i$, eventually $g_n(x) \geq (1-\varepsilon)a_i$. This is true at all points $x \in S$, and so $\bigcup_{n=1}^{\infty} A(n,i) = A_i$. And now we recognise that we have the necessary conditions to apply the continuity of the measure μ to the sequence $\{A(n,i)\}$.[3] We will indeed apply continuity shortly. First we need to define the elements of the sequence $\{h_n\}$:

$$h_n = \sum_{i=1}^{k} (1-\varepsilon) a_i \mathbf{1}_{A(n,i)}.$$

This is a non-negative simple function, and we clearly have that $h_n \leq g_n$ for each n. Finally, we have that:

$$\begin{aligned}
\lim_{n\to\infty} \int h_n \, \mathrm{d}\mu &= \lim_{n\to\infty} \sum_{i=1}^{k} (1-\varepsilon) a_i \mu(A(n,i)) \quad \text{by definition of integral} \\
&= \sum_{i=1}^{k} (1-\varepsilon) a_i \lim_{n\to\infty} \mu(A(n,i)) \quad \text{because the sum is finite} \\
&= \sum_{i=1}^{k} (1-\varepsilon) a_i \mu(A_i) \quad \text{applying continuity of } \mu \text{ to } \{A(n,i)\} \\
&= (1-\varepsilon) \int f \, \mathrm{d}\mu.
\end{aligned}$$

This completes the proof. Summing up, we saw that for a consistent definition of Lebesgue integral we need continuity of the measure with respect to which we integrate, and we showed that continuity is equivalent to countable additivity. Hence we can conclude that, given the importance of integration in probability, measure-theoretic probability relies heavily on countable additivity.

[3] The sequence of sets defined for equation 6 was actually increasing, rather than non-decreasing; it is simple, however, to re-label a non-decreasing sequence of sets so that it becomes increasing.

De Finetti's integration and countable additivity. For de Finetti's full outline of integration see his [14, pp.115-128]. Here I only remark that he draws an informal parallel between the usual continuous functions of analysis, and the special interest they have there, and random variables which have the property above with respect to Lebesgue integration. He shows that such property is equivalent to continuity of the measure, which in turn is equivalent to countable additivity. Hence the random variables for which the theorem above holds, could be called 'continuous'. De Finetti suggests then, that we consider these random variables, the ones that induce a measure which is countably additive, as a special class, but not the *only* one. He explains that of the various properties such random variables have, continuity would be the most important. By not enforcing that all probability measures be countably additive, we avoid the counter-intuitive aspects that countable additivity brings [14, p.121]. From this point of view, it could seem that the disagreement is purely a matter of terms and names: measure-theoretic probability could be seen as studying a certain class of probability measures, which is the most important one, but not the only one. However, this leaves open some difficult questions, such as if we should accept all the results obtained through Lebesgue integration, or if we should consider the powerful convergence results of measure theoretic probability as true in general. If we claim that all relevant probabilities are countably additive, and use the results that this gives, we should come up with some convincing explanation of why we ignore probabilities which do not have this property. On the other hand, if we limit ourselves to finitely additive probabilities, we should accept some seemingly paradoxical results as being genuine characteristics of what 'probability' is. We see here at play the contrast between the two intuitions I outlined in the introduction. The rest of this chapter is devoted to exploring some characteristics and consequences of finite versus countable additivity.

2.3 Conglomerability

Conglomerability is a long name to describe a very natural characteristic of probabilities, if we write them as weighted averages: namely, that the weighted average is always within the range defined by the smallest and the largest value in the average. What follows is a more careful definition. Below, I will follow the definition and example given in [31], who in turn attribute it to [14]. I also critically examine a possible solution to *nonconglomerability*, or the failure of this natural property, given by Jaynes in his [30]. We have an event space S and a finitely additive probability measure P defined on it; suppose we have an exhaustive partition of S into n mutually exclusive subsets, $\bigcup_{i=1}^{n} h_i = S$. Then we can write any event E as

$$E = \bigcup_{i=1}^{n} E \cap h_i,$$

and by finite additivity of P

$$P(E) = \sum_{i=1}^{n} P(E \cap h_i).$$

Now, the conditional probability of E, given h_i is defined, if the denominator is non-zero, as follows:

$$P(E|h_i) = \frac{P(E \cap h_i)}{P(h_i)},$$

and so we can write $P(E \cap h_i) = P(E|h_i)P(h_i)$, so that now we indeed have a weighted average expression for $P(E)$:

$$P(E) = \sum_{i=1}^{n} P(E|h_i)P(h_i).$$

Conglomerability is just the natural property of $P(E)$, the weighted average, to be within the range of the members $P(E|h_i)$ of the sum. Since the coefficients $P(h_i)$ of the weighted average sum up to 1, this is actually a special kind of weighted average known as *convex combination*. We express conglomerability as the following property: for all constants k_1, k_2, if $k_1 \leq P(E|h_i) \leq k_2$ for all h_i, then $k_1 \leq P(E) \leq k_2$.

Here is a practical, albeit completely fictional, example, to ground the idea that this is a very natural and intuitive property for probabilities. Suppose we are estimating how likely it is to have an accident, if we go through a busy crossing without paying any attention to the traffic lights. In this fictional scenario, the light stays red (event R) for 60% of the time, green (G) for 30% of the time, and yellow (Y) for 10% of the time; we write $P(R) = 0.6$, $P(G) = 0.3$, $P(Y) = 0.1$. We give the chances of being in an accident (A), if it happens that we cross with a red light, as $P(A|R) = 0.7$; for the other lights we give $P(A|G) = 0.01$, and $P(A|Y) = 0.05$. Then the overall probability of being in an accident if we go through this crossing while paying no attention to the lights, is

$$P(A) = P(A|R)P(R) + P(A|G)P(G) + P(A|Y)P(Y) = 0.428$$

A failure of conglomerability in this context (although we would never really have it in a finite case such as this) would mean that: either we would affirm that the *overall* risk of being in accident at that crossing is *higher* than the specific probability of crossing when it is most dangerous, namely when the light is red; or that the overall risk is *lower* than crossing when it is safest, when the light is green. Neither option makes any sense. Unfortunately, as anticipated above, there exist known failures of conglomerability, cases which are allowed by a merely finite probability measure, but not by a countably additive one. If we only impose that probabilities be finitely additive, we must accept that they might fail on conglomerability.

A finitely additive probability measure which is not conglomerable over all partitions. Here is the example given in [31]. Let P be a finitely additive probability measure, defined over all sets of couples of positive integers, i.e. the set $\{(i, j) : i, j \text{are positive integers}\}$. This is also the set of integer coordinate points of the first quadrant. We define

$$P((i, j)) = 0$$

for all single points (this is allowed by finite additivity), and

$$P((i, j)|B) = 0$$

if B is an infinite set. We now want to look at the probability of the set $A = \{(i, j) : j \geq i\}$, or the set of all points above the diagonal line $i = j$, including the diagonal itself. We first define the partition π_1 of the first quadrant as follows, for all finite i, j:

$$\pi_1 = \{h_i : h_i = \{(i, j)\}\} .$$

So h_k is just the set of all points with i-coordinate equal to k: the vertical line of points at $i = k$. It is clear that π_1 is an exhaustive partition of the first quadrant, made of mutually exclusive subsets. Hence by conglomerability, $P(A)$ must be within whatever constants constrain all values $P(A|h_i)$. We note the following, however. For any i, $P(A|h_i) + P(A^C|h_i) = 1$, by finite additivity. But we also know that:

$$P(A^C|h_i) = P((i, 1)|h_i) + P((i, 2)|h_i) + \cdots + P((i, i-1)|h_i) = 0$$

where the first equality is true by finite additivity of P, and the second one by the definition of P and the fact that we only ever have a finite number of points $(i, 1), \ldots, (i, i-1)$ under the diagonal line $i = j$, and so not in A. It follows that we have:

$$P(A|h_i) = 1$$

for all finite i, and so, by conglomerability, $P(A) = 1$, since $P(A)$ must be within any constant which constrain the values $P(A|h_i)$.

We now partition the first quadrant in horizontal lines, rather than vertical ones:

$$\pi_2 = \left\{h'_j : h'_j = \{(i, j)\}\right\} .$$

We still have for all j that $P(A|h'_j) + P(A^C|h'_j) = 1$, but this time, by the same reasoning as above, we see that

$$P(A|h'_j) = 0$$

for all j, and so, again by conglomerability, we see that $P(A) = 0$. But $P(A)$ cannot be 0 and 1 at the same time; and hence, over one of the two partitions, conglomerability must fail.

A suggested solution for the nonconglomerability example. This has been the argument above: one partition imposes a certain value for $P(A)$, while another partition imposes a different one. P is single-valued and so we must choose which value of $P(A)$ applies. Having made this choice, it must result that with respect to one of the two partitions, P is no longer conglomerable, since the value for $P(A)$ will now be out of the range of the values of $P(A|h_n)$. I repeat it here because I think it will make clearer my criticism of the solution proposed by Jaynes to this example, in [30, pp.453-455]. I present this solution below, with some slight modifications in order to adapt it to the example I described just above.

The idea, in this solution, is to start from a finite $M \times N$ array of points in the first quadrant, deduce an explicit expression for $P(A)$ in the finite case, and then observe the behaviour in the limit of this expression. What we will see, is that, having obtained an explicit formula for $P(A)$, it is absurd to think that this could change according to how we partition the first quadrant; it can only change according to how the limit is approached, or how the $M \times N$ arrays increase in size as M and N go to infinity. Moreover, we will see that the expression for $P(A)$ is bounded by the expressions $P(A|h_n)$ both in the case of horizontal partitions, and of vertical partitions, for all finite M and N. Hence it will be bounded by these expressions in the limit too. Jaynes claims he has solved the riddle, and that it arose simply from a bad understanding of mathematical infinity and of limits. We should never operate on infinite sets directly, but only on finite sets, and then observe the limiting behaviour. I think Jaynes misses the point in two ways. The first is that it is pointless to show, as he does, that $P(A)$ is conglomerable with respect to one or another partition, because even in the example above this was trivially the case. The problem was that the two resulting probabilities were in disagreement with each other. Here this cannot happen, since we have an explicit expression for $P(A)$. The second way Jaynes, in my opinion, misses the point, is contained in how he concludes the section on this example:

> Thus, nonconglomerability on a rectangular array, far from being a phenomenon of probability theory, is only an artifact of failure to obey the rules of probability theory as developed in Chapter 2 [30, p.455].

Jaynes' Chapter 2 in [30], is the derivation of quantitative probability rules from qualitative axioms, following Cox's [10]. As I will discuss in detail in my Chapter 5, there is nothing in that derivation which includes countable additivity. Hence the point here is that probability theory as developed in Jaynes' approach *does permit* examples of nonconglomerability, and this is what is problematic. In order to solve this example Jaynes must modify it. I give this attempted solution in what follows.

We start from a finite $M \times N$ array in the first quadrant, and we take $A = \{(i,j) : j \geq i\}$ as before (this is a very slight deviation from Jaynes, who takes the set of points with $j > i$; it is not a substantial difference, of course). We assume now that each point (i,j) has probability $\frac{1}{MN}$. This fundamental assumption changes the whole nature of the example, because in the case above, all points had individual probability of 0. I will take Jaynes' point through anyway, and display the necessary calculations, for completeness, and in order to explore whether we have any insights into the example above. By counting, we see that the following holds:

$$P(A) = \begin{cases} \frac{N+1}{2M} & \text{if } M \geq N \\ 1 - \frac{M-1}{2N} & \text{if } M \leq N \end{cases} \tag{7}$$

We also calculate the probabilities over the vertical and horizontal partitions of the first quadrant. The notation is like above. For the vertical partitions we have:

$$P(A|h_i) = \begin{cases} 1 - \frac{i-1}{N} & \text{for } 1 \leq i \leq M \leq N \text{ and } 1 \leq i \leq N \leq M \\ 0 & \text{for } N \leq i \leq M \end{cases}$$

Whereas for the horizontal partitions we have:

$$P(A|h'_j) = \begin{cases} 1 - \frac{j}{M} & \text{for } 1 \le i \le M \le N \text{ and } 1 \le i \le N \le M \\ 1 & \text{for } N \le i \le M \end{cases}$$

We can now check if the extreme values of $P(A|h_i)$ and $P(A|h'_j)$ include all possible values of $P(A)$, for all finite M and N. In other words, we are checking conglomerability. This indeed holds, because for the vertical partitions we have the following:

Vertical partitions, case $M \ge N$. The upper bound of $P(A|h_i)$, with $1 \le i \le N \le M$, is 1. The lower bound is 0. Taking the relevant expression from equations 7, we see that indeed conglomerability holds, because when $M \ge N$:

$$0 \le \frac{N+1}{2M} \le 1$$

Vertical partitions, case $M \le N$. Here the upper and lower bounds for $P(A|h_i)$ are, respectively, 1 and $1 - \frac{M-1}{N}$. Hence, with $M \le N$, we indeed have conglomerability:

$$1 - \frac{M-1}{N} \le 1 - \frac{M-1}{2N} \le 1.$$

For the horizontal partitions equivalent reasoning show that conglomerability holds in all finite cases. I only write the expressions for the lower and upper bounds on $P(A|h'_j)$, to show how they indeed constrain the value of $P(A)$, given in the relevant expression from equations 7.

Horizontal partitions, case $M \ge N$.

$$\frac{1}{M} \le \frac{N+1}{2M} \le \frac{N}{M}$$

Horizontal partitions, case $M \le N$.

$$\frac{1}{M} \le 1 - \frac{M-1}{2N} \le 1$$

I showed the solution displayed by Jaynes because, I believe, it has some appeal. It seems that we have solved the issue of nonconglomerability by just being careful with how we approach infinity. However, as I pointed out above, we now see that the problem is not really solved at all. In fact, if we take the same finitely additive probability measure as was in the original example by [31], which is the one Jaynes is actually replying to, we simply end up in the same situation, since for all finite arrays $M \times N$, $P(A)$ would be zero, and we could attach a different value to an infinite union of points, because finite additivity, which is all Jaynes has proved, does not forbid this. I also mentioned the superfluousness of showing that $P(A)$ is conglomerable with respect to both partitions, and we see this now explicitly. We have an expression for $P(A)$, so it is impossible that with the lower and upper bounds of $P(A|h_i)$ and $P(A|h'_j)$ we can 'bound values of $P(A)$ away from each other', which is what happened in the first problematic example. However, Jaynes had to modify the problem in order to solve it, and his own rules allow such problems as the original to exist—therefore, on this evidence, nonconglomerability must be accepted as a possible phenomenon if we accept finitely additive probability.

Admissibility. As is explained in [31, pp.213-214], we see that nonconglomerability leads to a failure of a principle of *decision theory* called admissibility. Without delving into decision theory, we can simply say the following: suppose we have an event space partitioned into sub-events, and we have to make a decision on which action to take. Suppose we can take either of two options, each of which has a cost, associated to which event in the partition occurs. If one action has a better cost profile over all possible events (represented by the partitions of the event space), then this action is strictly preferred to the other one, which is called *inadmissible* [31, p.214]. Suppose, however, that we take this reasoning and apply it to the first example above. If we partition the event space over the vertical lines, we would bet on A in all partitions, since it has probability 1. However, if we partition the event over horizontal lines, we would bet against A, since it now has probability 0 in all sub-events (the partitions). Hence betting for and against A can both be either the preferred option or inadmissible, according to which partition of the event space we consider. Briefly put: finite additivity permits failures of admissibility. Next, we will see the relation between a popular way to view chance, and countable additivity.

2.4 Frequency interpretation of probability

Perhaps the most intuitive understanding of probability is gained by thinking of it as a frequency: in a repeated experiment, we count how many instances of a certain event occurred, and we take the ratio of that number over the total number of experiments. This, then, will perhaps give us an indication of how likely an event is to occur in the future. For the 'true' probability of a certain phenomenon, we can take the limit, for the number of trials going to infinity, of the finite ratios of number of successes over number of trials. We could see two potential problems here, in the context of countable additivity: (1) the argument that sustains the claim that through a large number of trials we will reach the true probability of a phenomenon is often based on 'laws of large numbers', of which there exist many kinds, but in many cases they rely on countable additivity. This is only really a potential issue because of problem (2), which is, very simply, that limiting frequencies, as we will see again in Chapter 6, are not countably additive. We readily see this with the infinite lottery example: each ticket has, in the limit, a frequency of 0, and yet their probabilities must add up to 1. Different authors have positioned themselves differently with respect to this issue. Van Fraassen simply asserts that, since limiting relative frequencies fail countable additivity, they cannot be considered a suitable definition of probability [19, pp.133-135]. Others, such as Schurz and Leitgeb [34, pp.257-259], suggest that this might mean that countable additivity is too strong a requirement, since it deprives us of an intuitive and seemingly well-grounded understanding of probability. They underline that abandoning countable additivity can pave the way for the "[d]evelopment of a *genuinely frequentistic* probability theory" [34, p.259] [emphasis in the original]. De Finetti is opposed to the interpretation of probabilities as limiting relative frequencies, but still points out that on this interpretation, as well as on his own, countable additivity is not a valid requirement [14, pp.89-90].

2.5 The infinite lottery and Humean scepticism towards induction

In [34, p.258], Schurz and Leitgeb explain that "a *Humean skeptical view* of induction requires a non-[countably-additive] probability measure" (emphasis in original). This is closely linked to the argument by Kelly which I explore in Chapter 5, so I will not go over the general claim. Here I go back to the infinite lottery, to see in practice what this claim means. To make the link with induction, we take an infinite lottery 'in time' (this is in some respects similar to the example by Howard in [26, pp.133-134]). Thus we imagine a situation in which we start examining the tickets one by one, starting from tickets $1, 2, 3, \ldots$, and we check whether each one is a winning ticket. We can visualise this as an (infinite) urn containing numbered balls. However, in order to have two simple hypotheses to compare, we further modify the infinite lottery scenario: now we do not know whether we are in one of the following two situations: (1) either all balls are black, and there is no winning number, or (2) there is exactly one white ball, which represents the winning number. As we examine one ball after another, we must decide whether we are in situation (1) or (2). We model this example as it would be viewed by a Bayesian observer, which in one case will employ countable additivity, and in the other only finite additivity. The observer will update her degrees of belief according to the evidence she has seen up to a given moment, and decide which hypothesis, (1) or (2), she deems more likely. How do we represent Humean scepticism towards induction, in this context? A Humean sceptic will refuse to predict the future according to the necessarily finite number of balls she has examined. Hence, as we will see in this example, she will need to adopt finite additivity only. Note that in the countable additivity case, this is a numerical example, to see in practice the effect of the principle, but the result is general, as stressed throughout my thesis: for a countable sum of probabilities to converge to 1, the individual probabilities must, from some finite point onwards, form a decreasing sequence. Hence I show also the general result (general insofar as the two hypotheses we consider are equi-probable to start with).

These are the two possibilities in the situation described above, and labelled with the letter we will adopt for them:

- $B :=$ all balls are black,
- $W :=$ there is one white ball.

As our observer takes balls from the urn, she records their colour in the following way: she writes 1 if the ball is black, and 0 if the ball is white. Hypothesis B is made true by an infinite sequence of 1s. All sequences that end in a 0 make W true. We assume that the lottery ends if the winning ticket is picked. $P(10)$, for example, is the probability of picking first a black ball, then the white one. $P(10|B)$ is the probability of observing such a sequence, given that hypothesis B is true. I assume throughout finite additivity of this kind:

$$P(1) = P(10) + P(11).$$

Note that in the distribution of probabilities $P(\cdot|W)$, or the probability of a ball being black or white given that there exists one white ball, we are back to the de Finetti infinite lottery. I give simple numerical examples of what happens when we impose that these conditional probabilities add up to 1 (respecting countable additivity) and what happens if we apply a de Finetti-style fix to the problem, i.e. assign 0 to all these conditional probabilities. In what follows I suppose the prior probabilities for B and W are the same, namely $P(B) = P(W) = \frac{1}{2}$. Clearly, we also have $P(1\ldots0|W) = 1$, $P(1\ldots0|B) = 0$. Lastly, I will write $1^{(n)}$ for a sequence of n 1s.

First case: imposing countable additivity. Suppose the probabilities $P(\cdot|W)$ follow this sequence: $P(0|W) = \frac{1}{2}$, $P(10|W) = \frac{1}{4}$, $P(110|W) = \frac{1}{8}$, and so on: $P\left(1^{(n-1)}0|W\right) = \frac{1}{2^n}$. These values add up to 1 because $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$, so countable additivity is ensured. First I explain a numerical example, then I show the general case. We want to know:

$$P\left(B|1^{(n)}\right) = \frac{P(B)P\left(1^{(n)}|B\right)}{P\left(1^{(n)}\right)} \tag{8}$$

and

$$P\left(W|1^{(n)}\right) = \frac{P(W)P\left(1^{(n)}|W\right)}{P\left(1^{(n)}\right)}, \tag{9}$$

since if the white ball is extracted there is no uncertainty over which hypothesis is correct. Suppose we extracted 2 balls. We have

$$P(11|W) + P(10|W) + P(0|W) = 1$$

and $P(10|W) = \frac{1}{4}$, and $P(0|W) = \frac{1}{2}$, so $P(11|W) = \frac{1}{4}$. We also can find $P(11)$:

$$P(11) = P(B)P(11|B) + P(W)P(11|W) = \frac{1}{2} + \frac{1}{8} = \frac{5}{8}.$$

Also, $P(11, B) = 1$. Then we get $P(B|11) = \frac{4}{5}$ and $P(W|11) = \frac{1}{5}$. This suggests that the probabilities for B being true will converge towards one as we extract more black balls. This is what Kelly points out, and it is indeed the case. For a general number n of balls extracted we have (I omit calculations since they are entirely similar to the simple 2-ball case just seen):

$$P\left(B|1^{(n)}\right) = \frac{2^n}{2^n + 1},$$

and

$$P\left(W|1^{(n)}\right) = \frac{1}{2^n + 1}.$$

We also have, as expected, $P\left(W|1^{(m)}0\right) = 1$ and $P\left(B|1^{(m)}0\right) = 0$. As we see, the probability of hypothesis B, that there is no winning ticket, increases as we observe more black balls, whereas hypothesis W becomes less probable on this evidence. Away from this numerical example, we can show that this is the case in general. Suppose that for $M > N$ we have that $P\left(1^{(M)}|W\right) \leq P\left(1^{(N)}|W\right)$; note that it must be possible to find such a case, since

$$P(1^{(n)}|W) + P(1^{(n-1)}0|W) + \cdots + P(0|W) = 1,$$

and countable additivity imposes that the values $P(1^{(m)}0|W)$ cannot all be 0. If we show that it follows that

$$\frac{P\left(1^{(M)}|W\right)}{P\left(1^{(M)}\right)} \leq \frac{P\left(1^{(N)}|W\right)}{P\left(1^{(N)}\right)},$$

then we are done, because we would have $P\left(W|1^{(M)}\right) \leq P\left(W|1^{(N)}\right)$ by equation 9. Now, we know that

$$P\left(1^{(M)}\right) = P(B)P\left(1^{(M)}|B\right) + P(W)P\left(1^{(M)}|W\right) = \frac{1}{2} + \frac{1}{2}P\left(1^{(M)}|W\right),$$

and similarly for $P\left(1^{(N)}\right)$. So in fact,

$$\frac{P\left(1^{(M)}|W\right)}{P\left(1^{(M)}\right)} = \frac{2P\left(1^{(M)}|W\right)}{P\left(1^{(M)}|W\right)+1}.$$

Let $x := P\left(1^{(M)}|W\right)$ and note that $\frac{2x}{x+1}$ has derivative $\frac{2}{(x+1)^2}$ and thus is an increasing function on $(0,1)$. Therefore, we can conclude that $P\left(1^{(M)}|W\right) \leq P\left(1^{(N)}|W\right)$ implies that

$$\frac{2P\left(1^{(M)}|W\right)}{P\left(1^{(M)}|W\right)+1} \leq \frac{2P\left(1^{(N)}|W\right)}{P\left(1^{(N)}|W\right)+1},$$

and this in turn means that $P\left(W|1^{(M)}\right) \leq P\left(W|1^{(N)}\right)$. So we see that because of countable additivity, seeing a larger number of black balls implies that we *must* assign a lower probability to the hypothesis that there a white ball in the urn. This could be seen as a Bayesian solution to the problem of induction: but the role of countable additivity in this solution is what many authors find suspect.

Second case: de Finetti-style solution Now suppose $P(0,W) = P(10,W) = \cdots = 0$. This is de Finetti's desired distribution for the 'infinite lottery'. Having extracted two balls, we have

$$P(11,W) + P(10,W) + P(0,W) = P(11,W) + 0 = 1$$

and

$$P(11) = P(B)P(11,B) + P(W)P(11,W) = \frac{1}{2} + \frac{1}{2} = 1.$$

This stops the convergence which Kelly and other authors find suspect, since we see plugging in these values that $P(B,11) = \frac{1}{2}$ and $P(W,11) = \frac{1}{2}$. This could represent Humean scepticism towards induction. It is also obviously true for any finite number of extracted balls:

$$P\left(B|1^{(n)}\right) = \frac{1}{2},$$

and

$$P\left(W|1^{(n)}\right) = \frac{1}{2}.$$

Unfortunately, however, if we try to evaluate $P\left(W, 1^{(m)}0\right)$ and $P\left(B, 1^{(m)}0\right)$, which should be 1 and 0 respectively, we immediately see we would need to divide by 0, since $P\left(1^{(m)}0\right) = 0$:

$$P\left(1^{(m)}0\right) = P(B)P\left(1^{(m)}0|B\right) + P(W)P\left(1^{(m)}0|W\right) = 0 + 0.$$

This is quite strange: Bayesian updating would fail upon viewing the white ball, just when the hypothesis W would be confirmed as true.

Summing up: (i) we saw in practice how countable additivity can give convergence in a somewhat artificial way; (ii) we saw that finite additivity has the means to stop any sort of convergence, but also has some disconcerting aspects, like not being able to compute, in our example, the probability of there being a winning ticket, upon having seen such winning ticket.

2.6 Remarks

I conclude this chapter by summing up what we saw so far, and adding some brief remarks. Firstly, countable additivity is an essential part of modern mathematical probability, in which this is defined as a measure. This view made the study of probability a fully formal branch of mathematics, and has dominated, indeed defined, this branch since its axiomatisation by Kolmogorov in 1933 in [33] (see [37, pp.1-26, 198-199]). We will have occasion to comment on the consequences of this throughout this text. Secondly, we saw that the *failure* of countable additivity can have very counter-intuitive consequences (nonconglomerability), and that a fix to this does not really seem possible. Thirdly, we also saw how the *adoption* of countable additivity can have counter-intuitive properties, namely that we cannot think of probabilities as long-run relative frequencies of experiments. Lastly, we witnessed how the adoption of the axiom is relevant in a probabilistic treatment of induction: because of the convergence it enforces, a universal hypothesis (all balls are black) became more likely than its existential complement (in the context, that there exist one white ball), the more positive instances we saw. We keep with the scenario of the infinite lottery in time in the next chapter, in which we examine hypotheses more complex than those seen here, W and B.

3 Countable additivity and Kelly's Formal Learning Theory

3.1 Setting

We are in the scenario described in Chapter 2: an agent is extracting balls from an urn, and writing a 0 or a 1 if the ball is respectively white or black. We noted in that chapter that countable additivity gave us a tangible 'advantage', or bias, towards one of the two hypotheses. Kelly, in his 1996 book [32], has been a vocal critic of just this characteristic, and has been influential in keeping open the debate surrounding countable additivity. He is mentioned by most of the contemporary authors I cite in this text, whether they agree with his conclusions or not. Kelly's book is on *formal learning theory*, a kind of formal epistemology. One of the reasons his arguments have been so influential in the debate on countable additivity, is that, in his framework, we see explicitly just what epistemological advantage countable additivity gives us.

3.2 Convergence-to-the-truth theorems

We could describe Kelly's argument schematically thus: countable additivity is crucial to the mathematics which underpins a class of results known as *convergence to the truth theorems* (or convergence of opinion theorems), which could be perceived as important for Bayesian philosophy of science. But countable additivity is a suspect principle, because it gives an epistemological advantage (because of its convergence properties) which finite additivity does not. Hence, it is not an innocuous technical assumption, but one of utmost philosophical relevance: if we adopt the principle to obtain convergence theorems, we should be prepared to somehow defend this adoption on philosophical grounds. Broadly speaking, a convergence of opinion theorem is a kind of theorem which ensures that, as evidence accumulates, degrees of belief converge towards the confirmation of a correct hypothesis and refutation of fallacious ones. I give some more detail about subjective Bayesianism in Chapter 4. Briefly put, according to this influential school of thought, probability is a personally held degree of belief in a hypothesis, or occurrence of an event, or the like. Coherent agents must have degrees of belief that respect the rules of probability calculus, or they could suffer certain loss from betting (or perhaps relying) on their beliefs. A possible problem with this view could be how to explain convergence of opinion in science, and the objectivity of science: after all, rational agents are free to choose *any* coherent collection of degrees of belief. And a possible answer to this problem, is a convergence of opinion theorem: whatever the starting degree of belief of different rational agents, as they see more and more of the same evidence, they will converge to the same, correct, hypothesis. Kelly takes explicit aim at this kind of theorem. In the opening of his chapter in [32] on probability, he says:

> Of particular interest are the limiting reliability claims made for probabilistic methods. For example, it is often said that the process of updating probabilities by Bayes' theorem will almost surely approach the truth in the limit [...]. In the light of the many negative results in the preceding chapters [of [32]], such claims sound

too good to be true. Are they? Or do they illustrate the triumph of modern, probabilistic thinking over scepticism?

Kelly's book is relatively recent, but there seems to be some consensus now that this sort of confirmation of a hypothesis is no longer considered important or worthy of exploring. Weisberg (in 2011) [38, p.23], says: "Many do not think these theorems provide any real vindication, and would not miss them if they were lost". There existed misgivings about such theorems already, for example, in the earlier works of Glymour (1980) [21, pp.72-74] and Earman (1992) [18, pp.147-149]. It may well be that Kelly's was a decisive blow for such interpretation of probabilistic convergence theorems as convergence-to-the-truth. In the positions above, however, the characteristics most criticised of these theorems is that they are true *in the limit*, which means, to put it crudely, when we are all dead; and that, as Weisberg puts it,

> the theorems only show that the convergence will happen "almost everywhere" i.e. on a set of models with probability 1, where that certainty is judged by the probability function whose success is in question. From an impartial perspective, one that does not assume that the agent's initial probabilities have any bearing on the truth, this guarantee is no guarantee at all [38, p.23].

We should note that these perceived problems do not depend on countable additivity. Weisberg mentions a convergence theorem by Hawthorne, presented in [25], that uses only finite additivity, but comments that in it too the convergence is relative to the probability function we start out with. Furthermore, although we mentioned in Chapter 2 that countable additivity is adopted for its convergence properties, versions of some of the convergence theorems (versions of the strong law of large numbers, for example) exist also in finitely additive setting (see [8]). Obviously, being limit theorems, they are open to the first criticism above, if we wish to interpret them as applicable to human agents.

Here is an example of a theorem which can be interpreted as a convergence to the truth theorem, which is cited by Kelly. It is in Halmos' textbook on measure theory [24, p.213]. Its proof makes use, of course, of countable additivity. I do not wish to write it out completely formally, as this would require introducing a lot of notation for which we will have no further use. We can explain it thus: suppose we have an infinite sequence of sets $\{X_i\}$, and we take as our event space their Cartesian product, which we call X. This is the set of all the points $x = (x_1, x_2, \dots)$ such that $x_1 \in X_1$, $x_2 \in X_2$, and so on. Then for any measurable set $E \in X$ (a measurable set is a set in the σ-algebra—see Chapter 2), we consider the probability of such set E, conditional on the first n coordinates of a point in $x \in X$ (this point is an infinite sequence of coordinates). As $n \to \infty$, this conditional probability converges to either 0 or 1, according to whether x is in E or not; this convergence is valid for all $x \in X$, except for on a subset of X of measure 0. The interpretation as 'convergence to the truth' is clear: suppose E is our hypothesis—that all balls in the urn are black, say; interpret x as the infinite sequence of balls which in principle we could extract from the urn; we modify our probability for E by conditioning on the first n balls we have seen. Then, as $n \to \infty$, we see that our hypothesis will converge to having probability 1 or 0 according to whether the sequence of balls is one that, at infinity, makes the hypothesis E true, or not. (The sequence of balls in question is the uninterrupted sequence of black balls.) It will converge, that is, except for on a set of sequences of measure 0. From this example we see both why it is tempting to give a real-world interpretation to such a theorem, and its clear shortcomings in this. I believe the scepticism towards such an interpretation expressed in the remarks in the last paragraph, are enough to conclude that countable additivity need not even enter in the discussion of these theorems. Attacking convergence to the truth theorems now would even seem pointless. Nonetheless, I think Kelly's remarks on countable additivity are interesting in themselves, because they allow us another perspective on the issue, from an epistemological point of view. I sketch his approach in what follows.

3.3 Kelly's framework

I give a sketch of Kelly's framework, enough to understand his point about countable additivity. We keep working with our running example of an infinite urn containing black balls and, perhaps, a white one. But now we generalise this situation by stating that there need not be only one white ball. We just know there are white balls and black balls in the urn. Our hypothesis, previously, was that either all balls are black, or there exists one white ball. Now we wish to study two more generalised hypotheses:

W′: there exist only a finite number of black balls, meaning after a certain point, all balls in the infinite sequence will be white;

B′: there exist an infinite number of black balls, meaning there is no such last black ball.

As in the previous treatment of the example, our agent writes down 0 when she encounters a white ball, and 1 when she encounters a black one. I will mostly talk of these sequences of numbers directly, as they allow an easier treatment than constantly mentioning coloured balls. We can visualise the situation as in Figure 1 below.

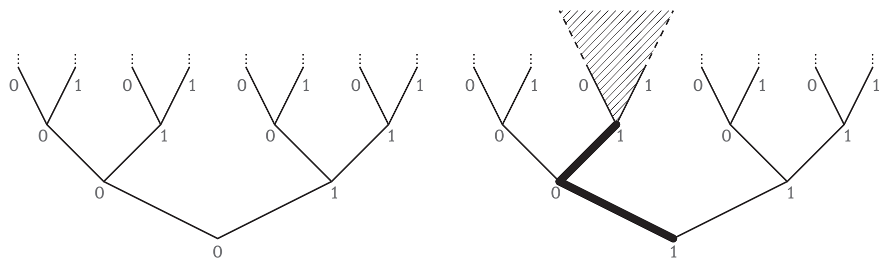

Figure 1: Possible sequences of 0s and 1s. In this example the agent picked balls of colours: black, white, black.

Now, we interpret what she is writing as the initial segment of an infinite sequence of 0s and 1s. We have no further information about the contents of the infinite urn, so in principle all infinite sequences of 0s and 1s are possible. The set of all infinite sequences of 0s and 1s is called the **Cantor space**. If we understand the finite sequences that our agent writes down as the initial segments of infinite sequences, then we see that in fact the sequence could continue in any direction after the initial one registered (i.e. it could continue as any combination of 0s and 1s). We call **fan** the set of all infinite sequences that share the same finite initial sequence. Hence what our agent records is always a fan. We can see a fan represented in 1 as a thicker line, which could then continue in any direction. Kelly characterises **hypotheses** as open, closed and clopen (both open and closed) sets of the Cantor space. In the topology defined on the Cantor space, the open sets are arbitrary unions of fans, and the closed sets are just the complements of such open sets.

We associate hypotheses to types of sets as follows. The simple hypothesis that there exists (at least) one white ball is made true by all the sequences that contain at least one 0. That is, all fans which contain at least a 0 confirm our hypothesis. This is clearly an infinite collection of fans, and hence it is an open set of the Cantor space. Because it is possible to *verify* this hypothesis with a finite number of observations (we just need to see a 0), we call such a hypothesis **verifiable with certainty**. The complementary hypothesis to this is that all balls are black. This is made true by an infinite, uninterrupted sequence of 1s. This is a closed set, being the complement of the open set which made true its complementary hypothesis. Because it is possible to prove this hypothesis *wrong* with a finite number of observations (we just need to see a 0), we call such hypotheses **refutable with certainty**.

So far we have established these connections:

- hypothesis verifiable with certainty $\leftrightarrow$ open set
- hypothesis refutable with certainty $\leftrightarrow$ closed set

But our hypotheses above, W' and B' do not fall in either of these descriptions. They regard the number of black balls: is it finite or infinite? We will never get an answer to this from a finite number of observations. We must stop observing at some point, and even if we observed white balls all our life, the day after we retire, a black ball could emerge. But not all is lost, at least we can still classify these hypotheses. This is how: suppose we are examining hypothesis W' (that there are only finitely many black balls, so only finitely many 1s), by looking at successive balls. Then we guess an assessment of the hypothesis simply by saying it is *correct* every time we see a white ball, and *incorrect* every time we see a black ball. Suppose there are only finitely many black balls. Then there must be a final one. After the black ball, and onwards to infinity, our assessment will stabilise to the right answer: it will say W' is *correct*. We then call the hypothesis W' **verifiable in the limit**. Now take the complementary hypothesis B' that there exist infinitely many black balls. This hypothesis is **refutable in the limit** by virtually the same assessment method, only this time the method says *correct* every time we see a black ball, and *incorrect* whenever we see a white ball. Then it will successfully stabilise to *incorrect* if the sequence has a final black ball, followed by white balls up to infinity. But is B' also verifiable in the limit? The answer is no. Suppose we had an ideal method which tries to decide if B' is true. However the method might work, suppose we have the following: whenever it decides on *correct* (presumably after having seen many black balls), a succession of white balls start appearing; if the method sticks to the assessment *correct*, the white balls keep going forever, making the assessment false; if the method switches to *incorrect*, a succession of black balls starts. This time, if the method switches again to *correct*, the balls start appearing white, while if the method switches to incorrect, the balls keep appearing black. Suppose this goes on forever: then this ideal method is never able to decide if B' is true. We say the hypothesis is **not decidable in the limit**. This hypothesis is also **not verifiable in the limit**: if a sequence makes B' false, we have a method that will stabilise to *incorrect* forever; but there is no method that can decide if B' is true. A hypothesis is decidable in the limit if it is both refutable and verifiable in the limit. Note that it could seem as if there was a malicious agent changing the colours of the balls according to the decision we made about our hypothesis. Indeed, Kelly calls these 'demonic arguments'. But of course a demon is just a mathematical artifice: the fact is that such sequences could exist; and so we are not guaranteed verifiability in the limit. Note also that even if a demon knew our first, simple, assessment method, there would be no chance of failing: if the number of black balls is finite, then they will stop eventually, and our method will return the right assessment. There is nothing even a demon could do about this. By a specular demonic argument we show that W' is **not refutable in the limit** (for this would just mean B' would be verifiable in the limit) [32, pp.51-55]. These kinds of hypothesis can again be characterised as sets in the Cantor space (for the proof see [32, pp.92-94]). We have:

- hypothesis verifiable in the limit $\leftrightarrow$ countable union of closed sets
- hypothesis refutable in the limit $\leftrightarrow$ countable intersection of open sets

Finally, countable additivity comes into play. Recall the convergence theorem at the end of the last section. There, we were guaranteed, except for on a set of measure 0, to receive an answer, in the limit, on whether the sequence x confirmed or refuted our hypothesis E, if E was a measurable set. This means that if we only consider, as valid hypotheses, sets from a σ-algebra, and we accept 'probability-1 decidability', rather than decidability over all possible sequences, we have decidability in the limit for any hypothesis. We simply assess E as *correct* if the limit in the theorem is 1, and *incorrect* if it is 0. But we have not changed the experimental situation: problematic sequences did not suddenly disappear. It is our way of looking at them which changed, by using measure theory (and thus countable additivity). Kelly notes that if we take away countable additivity, this probability-1 decidability fails, as we would expect given the comments about the role of the axiom in measure theory. And hence, he argues, it is artificial to believe that these problematic sequences, over which convergence to *correct* or *incorrect* fails, are unimportant because they have probability 0: they only have probability 0 thanks to countable additivity, and relative to the initial probability measure we adopted (this echoes the criticisms of convergence theorems above). He actually concludes that adopting countable additivity constitutes a realist bias, and a dogmatism:

> It is a fine thing for the realist that alleged canons of inductive rationality entitle him to claim sufficient knowledge to solve all inductive problems in the limit. A competent sceptic, on the other hand,

will reject this blanket entitlement to knowledge as a technically disguised form of dogmatism [32, p.320].

That the realism versus anti-realism debate can be understood in the simple choice on whether to adopt countable additivity or not, or whether to apply convergence-to-the-truth to agents or not, is an extremely interesting remark. It is perhaps qualified by the fact that such interpretation of said theorems seems to have died out, but the same cannot be said of realism; and that it could make some thinkers unknowing realists. Neither remark is necessarily knock-down, of course. We end this chapter by summing up and commenting on what we saw so far.

3.4 Remarks

We will see in Chapter 6 that the choice between countable and finite additivity is not the only one: we could use non-standard analysis to get out of the deadlock and define a different type of summation. It would be very interesting to see how this sort of solution stands to Kelly's remarks on additivity. Because the interest in non-standard probability seems to have resurfaced very recently (see [39] from 2013, after similar solutions were suggested earlier, see [2] from 1999), this might be a little-explored path. I also remark that there exist theorems in [8] which appear to be similar to the one by Halmos described above, but proved in a finitely additive setting. I do not know if they evade Kelly's criticism, but the area is too unfamiliar to me, for me to comment. With regards to Kelly's critique in general, it is very interesting in that it shows (somewhat) explicitly what we gain by using measure theory: convergence on almost all data sequences, at the cost of having eliminated the problematic data sequences to start with. The critique appears tempered, today, by the seemingly universal refusal to interpret converge theorems literally as applying to agents in time. But Kelly adds something to the two criticisms generally levelled at such application of convergence theorems, which are (i) that they are relative to the chosen probability measure, and (ii) that they only work in the infinitely long run. In fact, since Kelly studies ideal assessment methods and their behaviour in the limit, the latter would not a relevant criticism from him. What Kelly highlights is that probabilistic convergence theorems give 'selective' information about this behaviour in the long run. We saw above how the hypothesis that there exist only finitely many black balls is not decidable, even in the infinite limit. But we also saw that the probabilistic convergence theorem guarantees to decide any hypothesis in the limit, given the necessary premises (including countable additivity). Note that neither approach is mathematically wrong, of course. It is just the *interpretation* of the probabilistic convergence theorems which might, in a sense, give us something of a false sense of security with regards to what future data sequences hold.

Another possible comment on Kelly's position is that it might seem artificial to suppose that what we are viewing is the finite initial segment of something infinite. This assumption will not apply in any finitary updating of our beliefs, but of course neither will convergence theorems or countable additivity: we see then that this assumption is, in fact, something that runs throughout the debate presented in my work. Lastly, is Kelly's an argument against the adoption of countable additivity? I think the answer is no: the point is rather to make sure we realise what the cost of applying measure theory in philosophy of science is. Kelly calls *hybrid probabilism* the following philosophical view: chance exists in nature (the so-called propensity of something to occur or not), and probability, for agents, is just an assignment of degrees of belief. These degrees of belief should be coordinated to propensities, if an agent is rational (this is the well-known Principal Principle by David Lewis). Kelly explains that propensities, in this view, respect the measure-theoretic definition of probability, and so countable additivity can be freely used. A quick survey (see [23]) actually shows that there is no agreement with regards to whether propensities are additive or not. But this need not change the message of the interesting passage by Kelly which follows: we simply read ‘a probabilistic view which adopts countable additivity’ for his ‘hybrid probabilism’:

> I do not mean to suggest that hybrid probabilism is a mistake. I only wish to make it clear that the apparent advantages of the view (e.g., its correspondence with statistical practice and its ability to overcome global underdetermination) come at expense of a clear, logical account of what science is about, of what its import is for observation, and of how scientific method leads to the truth [32, p.337].

Here Kelly is criticising the Principal Principle as an unclear link between observation, degrees of belief and science, and the way in which probabilistic theorems lead to the truth through the mathematical artifice of countable additivity. He also remarks that these criticisms do not affect views of probability which are only finitely additive, because these lack the strong convergence results which seem at odds with Kelly’s epistemology. We will mention this view again in the final conclusions, because criticisms in a way similar to this are often brought against measure-theoretic probability: namely that it shifts probability away from our intuitive, real-world idea of it. For now, we devote the next chapter to the subjective probability of individual agents: should a coherent agent have countably additive degrees of belief? We address this question in what follows.

4 Dutch Book arguments for countable additivity

4.1 Setting

De Finetti, whose infinite lottery scenario is the backdrop to this whole work, was one of the main proponents of what came to be called 'subjective Bayesianism'. His philosophical position was that probabilities 'do not exist', except for as a numerical way for rational beings to express their epistemic uncertainty about some event, or hypothesis, or any aspect that requires uncertain reasoning. Hence probability is 'subjective', and probabilities are personal 'degrees of belief'. He was also one of the first modern Bayesians, in that he showed that if a decision procedure in a situation of uncertainty is rational, then it must follow the rules of Bayesian updating [15]. The merits and faults of such school of thought are discussed to this day, but what is interesting in the context of infinite lotteries is de Finetti's definition of a coherent thinker.[4] In de Finetti's subjective Bayesianism, it is pleasingly practical and simple:

> [...] once an individual has evaluated the probabilities of certain events, two cases can present themselves: either it is possible to bet with him in such a way as to be assured of gaining, or else this possibility does not exist. In the first case one clearly should say that the evaluation of the probabilities given by this individual contains an incoherence, an intrinsic contradiction; in the other case we will say that the individual is coherent. It is precisely this condition of coherence which constitutes the sole principle from which one can deduce the whole calculus of probability: this calculus then appears as a set of rules to which the subjective evaluation of probability of various events by the same individual ought to conform if there is not to be a fundamental contradiction among them [12, p.103].

A collection of bets which guarantee a loss in either direction (so a certain loss on one of the two sides of the bet) is often called a Dutch book. A coherent agent, by de Finetti's standards, must avoid being the subject of a Dutch book. Since, as de Finetti remarks, we can obtain the rules of probability through a study of how to avoid Dutch books, this technique is often called 'the Dutch book argument' for probability.

Given that de Finetti himself came up with both this concept and the infinite lottery, it might be surprising to learn that there exists a Dutch book for countable additivity, and one which is in fact very straightforward. I present this argument below, in a version by Williamson [41]. Reading de Finetti's [14] one realises that he was aware, of course, that such an argument could be made. In that book he provides an answer to it, which he clearly thought was obvious enough, since the matter is dismissed in one short paragraph. A reading of it, however, shows how puzzling it is and why it has failed to put off successive writers from making an explicit Dutch book argument against merely finite additivity. Howson, in [27], provides a detailed study of just what de Finetti could have meant. Before presenting his reconstruction, I propose another, simpler

[4] or a consistent one: there is disagreement with regards to how to translate de Finetti's *coerenza* from the Italian original—se Howson's [27].

reading of de Finetti's passage, based on the reading of his original manuscript in Italian. It is ultimately equivalent to what Howson concludes. There is a small but crucial difference between what de Finetti wrote in Italian and what is reported in the translated version in [14], which is what Howson quotes. Namely, where de Finetti wrote *serie*, or series (an infinite sum), the translator writes *sequence*. I believe this reading strengthens de Finetti's position on the matter. In what follows, the probability measure P is always assumed to have the usual properties of being positive, real-valued and less than or equal to 1 for all events. Additivity properties will be mentioned explicitly in each case.

4.2 A Dutch book argument for the adoption of countable additivity

Suppose we have a countable infinity of mutually exclusive and exhaustive propositions. To each proposition we attach a degree of belief. If these subjective probabilities do not add up to 1, Williamson [41, pp.411-412] shows that we are guaranteed a loss whatever happens. In particular, if in the de Finetti lottery we assign probability 0 to each number being picked, then we are open to an unfair system of bets. Let us call the event 'the number n is picked in the de Finetti lottery' simply 'n'. Then $P(3)$, for example, is the probability that 3 will be picked. Let us now give a definition, following de Finetti and Williamson [41, p.402] of **degrees of belief** in the betting set up. Suppose we are betting on an event occurring (or not). If we are willing to pay $P \cdot S$ on the event occurring, where S is the prize if it does occur, then we say that P is our degree of belief in such occurrence. A bet on an event E is very simple: we pay PS, and we get S back if E occurs, and nothing if it does not. (Recall that $P(E) \leq 1$.) Crucially, we only know P, and perhaps information about how likely the event E is. We do not know the direction of the bet (S could be positive or negative), and neither do we know the (real-valued) amount $|S|$ we are betting. Both values could depend on our chosen degree of belief. In other words, P has to be chosen so that, even if someone were betting against us and could modify magnitude and direction of S, we would not be *guaranteed* a loss. Williamson wants to prove that, in the de Finetti lottery, we are not guaranteed a loss if and only if $\sum_{n=1}^{\infty} P(n) = 1$. Here is the proof by Williamson (adapted slightly from [41, pp.411-412] for the sake of consistent notation in my work). Suppose $\sum_{n=1}^{\infty} P(n) < 1$. Now suppose the stake is the same, S, for all numbers. Then clearly we would make a certain gain if S were positive, as we would spend $\sum_{n=1}^{\infty} P(n)S$, less than what we would gain from the winning number, S. We would make a certain loss if the direction of the bet were reversed. So if we wish to avoid a Dutch book, our countable degrees of belief must sum up to 1.

Now for the other direction of the double implication. We assume that $\sum_{n=1}^{\infty} P(n) = 1$, and that all amounts of money exchanged are finite. The latter is equivalent to the requirement that all stakes and all losses be finite. Let L_h be the loss we make if the number h is extracted and let S_i be the stake associated with the bet over the occurrence of event i, or the extraction of i in the infinite lottery. Then:

$$L_h = \sum_{n=0}^{\infty} P(n)S_n - S_h.$$

We want to avoid *all* possible losses being positive, and we assumed that $L_h < \infty$ for all h, which is equivalent to assuming that $\sum_{n=1}^{\infty} P(n)S_n < \infty$, since we already assumed stakes to be finite.Note the following:

$$\begin{aligned}\sum_{i=0}^{\infty} P(i)L_i &= \sum_{i=0}^{\infty} P(i)\left[\sum_{j=0}^{\infty} P(j)S_j - S_i\right] \\ &= \sum_{i=0}^{\infty} P(i)\sum_{j=0}^{\infty} P(j)S_j - \sum_{i=0}^{\infty} P(i)S_i \quad \text{as all money amounts are finite} \\ &= 1\sum_{j=0}^{\infty} P(j)S_j - \sum_{i=0}^{\infty} P(i)S_i = 0.\end{aligned}$$

So we have $\sum_{i=0}^{\infty} P(i)L_i = 0$, but we know that $P(i) \geq 0$ for all i, and we have assumed that these values add up to 1, so it must be $P(k) > 0$ for some k. Therefore, for the series to add up to 0, it must be the case that for some j, $L_j < 0$, and hence the Dutch book is avoided.

Williamson argues that his Dutch book argument provides a normative reason to abandon uniform distributions of degrees of belief in countable cases. The fact that we are instinctively attracted to such positions, he remarks, is no argument for privileging uniformity over countable additivity; in this case, our intuitions must simply be wrong. Let us now see what de Finetti (writing a few decades before) made of such arguments.

4.3 De Finetti's dismissal of the Dutch book argument for countable additivity

As anticipated above, de Finetti considers the argument above (or, more precisely, one direction of it: that we are open to a Dutch book if we do not adopt countable additivity), but he quickly dismisses it. Here is the passage in full, as translated for the book [14], which is a collection of essays by de Finetti. Note that de Finetti calls countable additivity 'complete additivity'; also, he writes p_n where we would write $P(n)$, the probability of picking n in the infinite lottery (or any event n which forms part of an exhaustive partition of the event space into mutually exclusive events). The betting set-up is identical to that described above.

> The first argument in favour of complete additivity is the following: if the sum of probabilities p_n is $p < 1$, it would be possible, by entering the infinite number of available bets, to receive 1 in any event for a total payment of amount p, and this is clearly unreasonable. But in reality the argument is circular, for only if we know that complete additivity holds can we think of extending the notion of combinations of fair bets to combinations of an *infinite* number of bets, with

the corresponding sequence of betting odds [14, p.91] [emphasis in original].

What does de Finetti mean by this? As Howson remarks, this has puzzled authors for a long time [27, p.5]. The first impression must be, I think, that Williamson escapes such circularity, although the question is not treated explicitly by him. It does not seem correct that we need the idea of countable additivity in order to envisage an infinite sequence of bets. By finite additivity, we must have that

$$P\left(\bigcup_n n\right) \geq P(1) + P(2) + \dots$$

because $P\left(\bigcup_n n\right) = 1$, whereas all finite sums of probabilities of events are less than or equal to 1. Countable additivity imposes equality in this equation. Williamson shows (but so does de Finetti) that if $P(1)+P(2)+\dots < 1$ a Dutch book is possible. Merely talking about this infinite sum, however, it seems that we are not yet committing to countable additivity—the resulting number could still be different from $P\left(\bigcup_n n\right)$. But, by properties of probability, and if we wish to avoid a Dutch book, we have that

$$P\left(\bigcup_n n\right) \geq P(1) + P(2) + \dots = 1 = P\left(\bigcup_n n\right),$$

so that indeed $P\left(\bigcup_n n\right) = P(1) + P(2) + \dots$. But it seems strange that de Finetti would have missed such a simple reasoning. Looking at his original Italian manuscript containing the passage above (the published version is not easily available), could hold the key to understanding what de Finetti meant. I add the Italian original in a footnote below; here is the relevant passage (as translated by me), which is the second half of the passage above:

> But this is something of a vicious circle, because only if I knew complete additivity to be valid could I think of extending the notion of 'fair combination of bets' to combinations of <u>infinite</u> bets, and base them on the <u>series</u> of the betting odds [13, p.12] [emphasis as in original].[5]

It is quite remarkable that *serie* was translated as *sequence*, when the mathematical meaning of the two terms is very different—a series is the sequence of partial sums of elements of a sequence. 'Sequence' does not imply any addition: it is just a list of elements. I think that now what de Finetti meant can be interpreted and assessed more easily: it is not that we cannot conceive of an infinite number of bets if we do not have countable additivity. In fact, I think we can paraphrase de Finetti thus: we simply would not associate an infinite

[5] Un motivo che tenderebbe ad avvalorare l'additività completa: se le probabilità p_n hanno somma $p < 1$, stipulando tutte le infinite scommesse posso ricevere in ogni caso 1 pagando p, e quindi avrei un'incongruenza. Ma è un po' un circolo vizioso, perchè solo se sapessi valida l'additività completa potrei pensare di estendere la nozione di 'combinazione di scommesse equa' a combinazioni di <u>infinite</u> scommesse, e di basarle sulla <u>serie</u> delle quote di scommessa.

combination of bets with the infinite sum of the underlying single odds—unless, that is, we use countable additivity. Or, to use another formulation: it is true that the infinite sum of the single probability values, in an infinite lottery, opens us to a Dutch book if it is not equal to one; however, if we do not use countable additivity, we are never committed to saying that our degree of belief, *in an infinite sum of bets*, should be the infinite sum of the single odds, or degrees of belief. Indeed, this is just countable additivity. If we do use countable additivity, then our degrees of belief must adhere to such rule. If we give weight to this observation by de Finetti, Williamson's argument breaks down right at the start. Of course, if the price we pay to enter the bet, which is just our degree of belief, is less than 1 then we are ensured a win, and this reveals an incoherence. This applies if our degree of belief corresponds to the infinite sum of our single degrees of belief. But this is not the case: finite additivity allows us to consistently bet 0 on each single number, while betting one on the infinite union of all numbers, regardless of what the infinite sum of our degrees of belief might be (as long as it is less than or equal to 1). To my knowledge, the above interpretation of this passage of de Finetti, based on a re-translation of the original Italian manuscript, has not been made in the literature in these simple terms. I find the argument simple and convincing, which makes it plausible as an interpretation of what he had meant to say, and would also explain the apparent confidence with which he dismisses the issue. As mentioned above, other debates about the existing translations of de Finetti's work exist in the literature, so a (possible) mistake in this passage would not be so extraordinary. I next Howson's interpretation of the passage, based on the translation of de Finetti in [14]. I believe it ends up at the same conclusion I present here, but takes a longer path to it. Howson invites us to take a closer look to what de Finetti actually intended to show with his Dutch book argument for finite additivity.

4.4 Howson's reading of de Finetti's rebuttal.

Rigidity and finite additivity. Howson explains that de Finetti's quote above (in the opening paragraph of this section) is based on the following assumption, which de Finetti makes silently in [14, p.77] and explicitly in [16]. The assumption is the following: 'A finite sum of bets is fair with respect to P just in case each is fair with respect to P' [27, p.7]. Howson calls this assumption (A). He can thus explain de Finetti's passage as follows:

> The explanation is in two parts: (i) by 'extending the notion of combinations of fair bets to combinations of an infinite number of bets' de Finetti actually means *extending* (A) *to include countably infinite sums*; (ii) the extension of (A) to countable sums *entails* countable additivity, and conversely [27, pp.7-8].

Hence when de Finetti claims that the Dutch book argument for countable additivity is circular, according to Howson, he is pointing out that it would amount to the above extension of the principle (A). Howson has a simple proof of the fact that if we extend (A) to countable bets we get countable additivity—but in de Finetti's own terms the reasoning is even simpler, as I will note below. To understand how this works, we need to see what relation the principle (A) has to finite additivity. It might seem obvious that a sum of fair bets is fair if and only if each one is fair. However, consider the following scenario: suppose that, instead of paying P to bet, as above, we were buying a good for that price. The 'fair' price is the price we would be prepared to pay. Here is de Finetti:

> In general, it is not true that if one is prepared to buy an article A at the price $P(A)$, and an article B at a price $P(B)$, one must be prepared to buy both of them together at a price $P(A) + P(B)$. It may happen that the purchase of one of them affects, in various ways, the desirability of the offer [16, p.74].

In fact, many of us are risk-adverse, and we would rather pay a small sum, rather than incur the 50/50 chance of winning, or losing, a larger sum. However, for simplicity it is assumed that the property above holds. This is called *rigidity* in the face of risk [16, pp.77-78]. What is slightly disconcerting, is that this assumed property immediately gives finite additivity—or even, in this framework, it just *is* finite additivity: the price for $A \cup B$ is $P(A \cup B)$ by definition, but we also just affirmed that the price for $A \cup B$ is $P(A) + P(B)$, so that

$$P(A \cup B) = P(A) + P(B).$$

Then why should we assume rigidity? That the betting scenario is not representative of how many of us view risk, is a well-known criticism of this approach to probability (see for example Glymour's [20, pp.70-71]). But it is also, for its simplicity, one of its strong points. With rigidity, De Finetti is supposing that 'utility' is just equal to monetary value. For this he assumes that all the bets considered are on small amounts—we would baulk at betting on a million Euro, say, but we might be willing to bet 5 or 10 Euro because those amounts do not have a big impact on our finances. If we do not set this rigid scale, we would have to somehow describe how our risk-aversion increases as the sums involved increase, and how our judgement of utility fluctuates according to countless circumstantial factors. Money, however, ensures a certain generality: we can speak of a bet being unfair if it causes us to lose money whatever happens. However we happen to value such loss is a separate matter—the amount of money lost is objective.

However, the problem is this: seeing de Finetti's explanation, we could simplify Howson's principle (A), because, as remarked, the principle is just finite additivity. And note that if we assume that for an *infinite* collection of bets rigidity still holds, we are simply assuming countable additivity: rigidity is the assumption that we will pay, for a number of items or bets taken together, the sum of the individual prices. Then why should we not extend rigidity to a countable number of bets? Well, the answer must be, because we should not force agents to be countably additive. In order to avoid, in other words, countable additivity. This is circular, as is circular the Dutch book 'justification' of finite additivity. We seem to use them to prove finite additivity, but only after having assumed finite additivity itself (or a principle that, as de Finetti explicitly says, is entirely equivalent to it). And so, two questions seem very legitimate: (1) why do we assume rigidity, since it is equivalent to finite additivity; (2) what is the Dutch book argument for finite additivity supposed to show. If finite additivity of a collection of probability estimates, which is assumed, is entirely equivalent to not being open to a Dutch book (given finite additivity), then surely the second characterisation can add nothing new. It is important to note that de Finetti does not hide from this circularity; he discusses the hypothesis of rigidity at some length (see [16, pp.77-82, 92-93]) but, as is evident from the quote above, he does not pretend that it is a feature which is generally true in all real-world situations. De Finetti is perfectly aware that his Dutch book argument is a characterisation, rather than a justification, of probability functions; and he wishes not to impose that all probability functions be countably additive, hence why he does not assume rigidity in the face of infinite bets. The answer to question (2) is the following:

> In order to give an effective meaning to a notion—and not merely an appearance of such in a metaphysical-verbalistic sense—an operational definition is required. By this we mean a definition based on a criterion which allows us to measure it. We will therefore be concerned with giving an operational definition to the prevision of a random quantity, and hence to the probability of an event [16, p.76].

Hence just defining probabilities abstractly is clearly not enough for de Finetti: he deems it necessary to be able to elicit actual probabilistic statements from agents, even if in a slightly idealised situation, to ground our concept of probability. And we had also asked, in question (2): why stop at probabilities which are finitely additive? Or, which is the same, why assume that we will consider fair the sum of two fair prices? In de Finetti's words,

> let us turn to the other reasons for preferring this approach: these are essentially concerned with simplicity. The separation of probability from utility [i.e. the adoption of rigidity], of that which is independent of risk aversion from that which is not, has first of all the same kind of advantages as result from treating geometry apart from mechanics, and the mechanics of so-called rigid bodies without taking elasticity into account [. . .] [16, p.81].

Therefore, a Dutch book argument is *not*, for de Finetti, a proof that finite additivity is necessary in order to avoid certain loss; this is only true if rigidity—which, as stressed above, is in this case another word for finite additivity—is assumed. Having assumed rigidity, a Dutch book argument is just the following simple observation: since we have deemed that a fair price for article $A \cup B$ is $P(A) + P(B)$, and the price for $A \cup B$ is by definition $P(A \cup B)$, we should not accept to pay more than that fair price. If we had the situation in which $P(A)+P(B) > P(A\cup B)$ we know would be paying more than we deemed fair. What follows is finally a practical example of a Dutch book argument, inspired from the passage by de Finetti which Howson cites most, in [14, p.77].

The Dutch book argument 'for finite additivity'. We are looking at finite sums of bets, which we write as:

$$X = k_0\mathbf{1}_{E_0} + k_1\mathbf{1}_{E_1} + \cdots + k_n\mathbf{1}_{E_n}.$$

We should interpret this as follows: the E_i are events, (where E_0 is the certain event, so that we just write k_0 omitting the indicator function; note that also $P(E_0) = 1$); for the sum of bets above, we pay $k_0 + k_1P(E_1) + \cdots + k_nP(E_n)$, and we get back a combination of $k_0 + k_i + k_j + \ldots$, namely we receive k_i for every event E_i that occurs, and 0 for the ones that do not occur. This is how a **fair sum of bets** is defined: X, as written above, is fair with respect to the function P if we can have the following property:

$$k_0 = -\left(k_1 P(E_1) + \cdots + k_n P(E_n)\right). \tag{10}$$

What this means in practice is best explained by taking two mutually exclusive and exhaustive events E_1 and E_2. The bet is now the following:

$$X = k_0 \mathbf{1}_{E_0} + k_1 \mathbf{1}_{E_1} + k_2 \mathbf{1}_{E_2}.$$

Either E_1 or E_2 must happen, and not both at the same time; E_0 will happen in any case. So the outcomes of this bet can only be:

$$k_0 + k_1 \quad \text{if } E_1 \text{ occurs,} \tag{11}$$

and

$$k_0 + k_2 \quad \text{if } E_2 \text{ occurs.} \tag{12}$$

Recall that we only have partial control over the price we pay for these bets, which is $k_1 P(E_1) + k_2 P(E_2)$, as we choose $P(E_1)$ and $P(E_2)$, but we know nothing of the k_i. The definition of fairness given in 10 means that we should be able to rewrite the two possible results of the bet as follows:

$$k_1 - k_1 P(E_1) - k_2 P(E_2) \quad \text{if } E_1 \text{ occurs,} \tag{13}$$

and

$$k_2 - k_1 P(E_1) - k_2 P(E_2) \quad \text{if } E_2 \text{ occurs.} \tag{14}$$

Now, it is up to us to avoid a Dutch book: what this means in practice, is adjusting the parameters in our control, $P(E_1)$ and $P(E_2)$, so that both quantities above cannot be negative whatever happens. How could they both be negative? Very simply, as follows:

$$\begin{cases} k_1 - k_1 P(E_1) - k_2 P(E_2) < 0 \\ k_2 - k_1 P(E_1) - k_2 P(E_2) < 0 \end{cases}$$

which is equivalent to

$$\begin{cases} k_1 < \frac{k_2 P(E_2)}{1 - P(E_1)} \\ \\ k_2 < \frac{k_1 P(E_1)}{1 - P(E_2)} \end{cases}$$

This has no solutions if and only if $P(E_1) + P(E_2) = 1$, for only in that case the system reduces to:

$$\begin{cases} k_1 < k_2 \\ k_2 < k_1 \end{cases}$$

Therefore, fairness, or the avoidance of certain loss, is equivalent to finite additivity. Given the comments above, this is not surprising. Rigidity comes into it as the presupposition that we were able to simply sum bets numerically, without making adjustments for the fact that a larger bet might be less desirable than a smaller one. The assumption is contained in the definition of fairness 10, which makes it clear that the loss we incur from a sum of bets is the sum of the individual losses.

This, of course, works in general. For a general proof in this style for a number n of bets see de Finetti's [12, pp.103-104] (it is very similar in spirit to the one above, except for the slight complications which arise from solving larger systems of equations). It is intuitively clear that, if the total loss we suffer from a collection of bets is the sum of the individual bets, then in general this condition:

$$\min_i k_i \leq k_1 P(E_1) + \cdots + k_n P(E_n) \leq \max_i k_i,$$

ensures that at least two results from the bet will be of opposite sign, thus making a Dutch book impossible. It holds if the quantity in the middle is a weighted average (in this case a convex combination, which means the coefficients $P(E_i)$ add up to 1), because in that case such quantity will definitely be within the range defined by the two extreme values for k_i.

I find that in the passage quoted by Howson the concepts above are less clear than in other writings by de Finetti (such as in [12] and in [16, pp.69-90]), however, I report it for completeness. Howson's point is that in [14, p.77], what de Finetti *actually* shows is this: if the events form an algebra, and if finite additivity fails, we could have a finite sum of bets, *each of which is fair*, with each one giving certain loss; how de Finetti words it, however, is that, if the events form an algebra, we would have a *fair sum of bets*, which collectively give rise to a loss. The issue is that we need assumption (A) (rigidity) to pass from the second claim to the first. (De Finetti clearly gives it for granted at this more advanced stage of his treatment—the main point in those pages is to discuss the consequences of imposing that probability is defined only on a σ-algebra.) And if we only have the first, we perhaps we would need not worry, because we would not necessarily have a loss from a sum of bets which individually give losses—unless, that is, we assume principle (A). Put schematically: a failure of finite additivity implies a sum of unfair bets. That this sum is itself unfair overall, is an extra assumption.

Let us see how the concept of fairness, in the passage cited by Howson, underpins the claim that a failure of finite additivity can bring about a certain loss (see [14, p.77]). We first prove this direction of the claim: assuming fairness as defined above in 10, if we have finite additivity, then a sum of bets, each of which is positive (we noted that a bet in which all possible results have the same sign opens us to a Dutch book), is not fair. Now, suppose finite additivity of P holds; then we can write the expectation of the sum of bets X as follows:

$$\mathbb{E}(X) = X \cdot P(X) = k_0 + k_1 P(E_1) + \cdots + k_n P(E_n).$$

By our definition of fairness, we must have $\mathbb{E}(X) = 0$. If all of

$$k_0, k_1 P(E_1), \ldots, k_n P(E_n) > 0$$

we would have $\mathbb{E}(X) \neq 0$, and so the overall bet is not fair. Note that this is precisely where assumption (A) comes into play: by the definition of fairness, we have 'for free' that a sum of unfair bets is itself unfair. For the other direction, we want to show the following: the fact that a sum of bets, each of which is positive, gives rise to an unfair bet, implies finite additivity. This is equivalent to showing that if finite additivity fails, then we would consider as fair a sum of unfair bets. So we suppose that finite additivity fails, namely that we have two incompatible events E_1, E_2 such that

$$P(E_1 \cup E_2) \neq P(E_1) + P(E_2).$$

Now consider the following sum of bets:

$$k(P(E_1) + P(E_2) - P(E_1 \cup E_2)) - k\mathbf{1}_{E_1} - k\mathbf{1}_{E_2} + k(\mathbf{1}_{E_1} + \mathbf{1}_{E_2})$$

where we note that the first member of the sum, $k(P(E_1)+P(E_2)-P(E_1 \cup E_2))$ is just a constant, what we previously called k_0. Then we immediately see that this sum of bets is fair, with respect to the definition above. But we can also rewrite it as

$$k(P(E_1) + P(E_2) - P(E_1 \cup E_2))$$

because $-\mathbf{1}_{E_1} - \mathbf{1}_{E_2} + \mathbf{1}_{E_1} + \mathbf{1}_{E_2} = 0$, since the events are incompatible. And now we see that this can easily be made positive, by picking k appropriately. Hence a failure of finite additivity would make us consider this bet fair, by the definition of fairness given above, while we see that it could bring certain loss. The reader might be wondering why de Finetti embarked on this more convoluted argument, and why the explicit mention of algebras was made above. The fact is that, as anticipated above, in these passages de Finetti is not seeking justifications for finite additivity, but actually discussing how we would define probability if the events did not form an algebra. For example, in the case where we only had three events E_1, E_2 and E_3, with $E_3 = E_1 \cup E_2$. How would we say that $P(E_3) \leq P(E_1) + P(E_2)$, as must be correct? We would not have the intersections of the sets defined to make this precise. It is in this context that de Finetti defines linear combinations of events (linear combinations of bets) and gives the notion of fairness. He comments that this fairness is equivalent to finite additivity if we are operating within an algebra, as we would expect. If we are operating in algebra of sets, if E_1 and E_2 are in the algebra, then so are $E_1 \cup E_2$ and $E_1 \cap E_2$ and we can always define the probability of $E_1 \cup E_2$. However, de Finetti's aim is to show that even if the set is not an algebra, we can extend a probability function defined on such set to the whole algebra generated by that collection. We do this by using the definition of fairness given in equation 10 and the condition that we do not accept Dutch books [14, pp.76-79]. I do not treat this further, and pass now to some concluding remarks for this chapter.

4.5 Remarks

I claimed above that a closer reading of de Finetti's work shows that he did not think that Dutch book arguments were the *proof* of the necessity of finite additivity. This seems the case especially given the fact, readily admitted by de Finetti, that they are circular, in that they rely on finite additivity of degrees of belief in order to work. Therefore, the avoidance of Dutch books is rather a *characterisation* of probability. This was especially important in de Finetti's philosophical view, because, something of a positivist, he absolutely refused to see probability as the description of something which is actually existing in the world. His view is expressed perhaps most extensively in his essay on *Probabilism* [17], and here is an emblematic passage:

> Probability exists for me only as a function of the degree of ignorance in which I find myself at the time; it would be absurd, even if it were not meaningless, to consider probability as a mysterious and unreachable metaphysical entity, existing in abstraction, on which the occurrence of an event somehow or other depends [17, p.178].

Hence he needs an operational definition of probability in order to properly ground what he is talking about. A coherent assignment of probabilities is an assignment that would not open us to certain loss, supposing we were to bet on the events in question occurring, and assuming rigidity. Probability, for de Finetti, is nothing else than this.

Since this is the case, just as a Dutch book argument does not really 'show' finite additivity, because it must assume it in the first place, it also cannot show countable additivity, for the same reason. De Finetti is fully aware of this, but he would refuse to make the same assumption of rigidity for infinite combinations of probabilities, *because* he is opposed to countable additivity (for reasons such as the infinite lottery scenario). Hence a Dutch book argument for countable additivity does not really show anything, unless we assume that the betting odds for an infinite combination of bets are the sum of the odds of each bet. If we do not assume countable additivity, we also need not assume this. This seems to be what de Finetti had meant in his rebuttal of such arguments, and it is underpinned by what Howson concludes, after his closer look at Dutch book arguments for additivity in general. Howson's next step in reasoning is this: if Dutch book arguments cannot independently support either finite or countable additivity, then what can? For an answer, he turns to a work by Cox. It is important to note that Howson turns to Cox's construction, which I will study in some detail in the next chapter, in his search for "a type of completeness theorem telling us that the rules of probability extend to finite but not countable additivity " [27, p.17]. This because "De Finettis argument that a purely 'formal'principle [i.e. coherence] should not forbid *in principle* a uniform distribution over the elementary events (atoms) in a power set algebra has, I believe, a very strong intuitive pull" [27, p.17]. In his [10], Cox shows that starting with just two qualitative axioms, which do not mention additivity properties, we are able to arrive to the known probability rules, but only up to finite additivity. This is especially interesting for the present work since Jaynes, a respected author on Bayesian probability, takes exactly the same starting point but comes to opposite conclusions, namely that probability functions which are not countable additive should not be considered 'probabilities'.

4.6 Appendix: A proposed solution to the infinite lottery

Here is a proposal on how to solve the dilemma of the de Finetti lottery, by Bartha [1]. I present an outline of it here because it uses the betting terminology introduced in this chapter, but I argue that unfortunately it seems a reformulation of, rather a solution to, the problem. Bartha agrees with Williamson that if we have real-valued degrees of belief, then we are subject to a Dutch book in the infinite lottery, unless we adopt countable additivity. To this he proposes different solutions: one is we could have non-standard degrees of belief; he outlines this solution in his [2], and I present a similar one in Chapter 6. The other solution he proposes is to adopt *relative betting quotients*. Instead of having degrees of belief for all outcomes of a random event (say, the tickets in the infinite lottery), we could express only how likely we consider events to be relative to each other. This works as follows. Taking fair betting odds as defined above, suppose event E_1 has betting odds p, and event E_2 has odds q. Then define their relative betting quotient as $k = \frac{p}{q}$. Drawing up the table of possible gains and losses for bets *for* E_1 and *against* E_2 (meaning we win if

E_2 does not occur), Bartha notes that p and q do not appear any longer. The payouts, then, depend only on k. This suggests that we could use this k to define relative bets even if one or both the single odds are 0, or for betting on two events whose probability is not defined in general, but only with respect to each other [1, pp.307-308]. Hence he defines a general simultaneous bet, for E_1 and against E_2, such that we receive kS if E_1 occurs and E_2 does not, and $-S$ if E_2 occurs and E_1 does not.[6] Note that k is a non-negative number, but that is the only restriction. It should reflect 'how much more or less likely' we consider E_1 to be with respect to E_2, because, as in the sections above, S could be positive or negative. Postulating that no money is exchanged if neither event happens, the results of the bet just described are set in Table 1 below.

E_1	E_2	Payoff
T	F	kS
F	F	0
T	T	$(k-1)S$
F	T	$-S$

Table 1: The relative betting quotients for E_1 against E_2.

The relative betting quotient for k is the non-negative number that makes the bet described in Table 1 fair [1, p.308]. This approach, however, only solves the de Finetti lottery in the following sense: there is no real number k that would make the above bet fair, if the two events are: a particular number is picked (very unlikely); versus: some number is picked (which is certain). On the other hand, it is easy to express that two given tickets have the same probability of being picked: we set $k = 1$ for the relative bet between those two tickets. For Bartha, this shows that the de Finetti lottery does not constitute a counter-example to countable additivity, because countable additivity only applies if we have well defined degrees of belief [1, 309-310]. It seems to me that this is not a strong argument, and not a real solution to the deadlock. Having defined relative betting quotients, Bartha goes on to explain that in the infinite lottery these relative betting quotients do not exist. He claims that this tells us something about the infinite lottery; I think a more natural reading would be that it tells us the method is not adequate to the problem in question.

6 To 'receive' $-S$ means to pay out S.

5 Deriving probability from qualitative axioms

5.1 Setting

In the previous chapter I argued, with Howson, that Dutch book arguments cannot be the deciding factor in judging whether to take countable additivity as an axiom or not: in both cases, the additivity principles need to be assumed in order for the Dutch book argument to work. In this chapter I present a completely different argument for the foundations of probability. Howson turns to this foundational argument *because* it stops short of justifying countable additivity, since he deems this not a desirable axiom. As anticipated above, Jaynes (in his recent book [30]), also uses this very same foundational argument for probability, but he considers the very concept of merely finite additivity quite absurd, and he maintains that we get countable additivity naturally if we approach the limit of a sequence of events in the proper manner; and infinite sequences that do not achieve countable additivity in the limit should not be considered objects of probability theory [30, p.465].

The aim of this chapter is straightforward: I wish to explore how a principle of additivity is obtained in Cox's framework, and which kind of additivity this is; and examine whether Jaynes' position on countable additivity can be grounded in this framework, or if it must come from additional assumptions. I present Cox's own derivation of the rules of probability in some detail, and follow it step-by-step, except for some minor changes in notation, and some passages in which I try to make the reasoning more immediately understandable; in one passage I follow Jaynes's approach in [30, pp.24-38].

5.2 Cox's axioms, and a derivation of quantitative rules for probability

Cox's aim in [10] is to obtain *quantitative* rules for uncertain reasoning, starting from *qualitative* principles which, it seems clear, must be respected in this context. Hence he stipulates two "axioms of probable inference"; from these, using the rules of Boolean algebra and mathematical manipulation, we will be able to obtain the usual rules of probability. Here are the two axioms, quoted in full:

I The probability of an inference on given evidence determines the probability of its contradictory on the same evidence [10, p.3].

II The probability on given evidence that both of two inferences are true is determined by their separate probabilities, one on the given evidence, the other on this evidence with the additional assumption that the first inference is true [10, p.4].

Suppose we accept the two above facts as correct characterisations of what probability should be; suppose further that we accept the formal correctness of Boolean algebra and standard mathematical manipulation; then we will accept the results of these elements combined. These results are the known probability axioms. Hence we have a 'justification' of probability, or at least a characterisation of it, which might seem better grounded than the usual axiomatic definition. This can represent a justification in subjectivist Bayesian terms because we are compelled to obey, in rational thinking, rules resulting from axioms we accept and truth-preserving manipulations. Therefore it is very interesting from the point of view of countable additivity: here we have an alternative characterisation of probability; it too is normative, if we accept the starting axioms and the formal correctness of mathematics and Boolean algebra; does it imply countable additivity? As we will see below, it does not, at least not directly. Inspecting Cox's arguments, it appears that the only way of extending these results to include countable additivity would be to explicitly include the principle itself. This could then, perhaps, be motivated by some other assumption or intuition, but these would need to be added on to Cox's axioms.

Looking at the axioms themselves, we notice that the first one immediately seems correct and basic. The second axiom, on the other hand, while reasonable, hardly seems something we would intuitively take as a *foundation* for quantitative uncertain reasoning. Cox [10, pp.3-4] and Howson [27, p.18] provide some justification on why we should consider the axiom reasonable. Here is a concrete example by Cox, on the plausibility of Sir John Maundeville's assertion that "Noah's Ark may be seen on a clear day [. . .] on the top of Mount Ararat" [10, p.3]. This depends on the likelihood that Maundeville (a) made the assertion based on memory rather than invention. (b): given (a), the likelihood that his memory is correct. (c): given (a) and (b), that what Maundeville saw on the top of Mount Ararat actually was Noah's Ark. Howson prefers another way of justifying the second axiom. Suppose the plausibility of an assertion depends only on the plausibility of (a) and (b). If I only know how likely (b) is given (a) and the given evidence, I would not know how likely (b) is in general. But if I also knew how likely (a) is on the given evidence, then it seems that I should know how likely (a) and (b) are taken together [27, p.18].

I am not sure if it is worth trying to emphasise the role of Cox's axioms as intuitively fundamental for probability. Howson and Cox succeed in making the second axiom seem *reasonable*, but of course this is not why it was chosen: the point is to derive the known axioms of probability. Other aspects of conditioning or of uncertain thinking may very well seem more fundamental than these two, the second one especially. Adding these notions as 'axioms', however, would be redundant, since we will derive all relevant aspects of probability from axioms I and II anyway. I think it is natural to wonder, however, what we gain by stipulating these two assertions as axioms, rather than stipulating the usual axioms of probability directly. Howson emphasises that Cox's result shows "how strikingly little in the way of constraints on a numerical measure suffice to yield the finitely additive probability functions as canonical representations" [27, p.17]. He adds: "Cox believed, I think correctly, that these three rules deserve to be regarded as fundamental " [27, p.19].[7] More discussion of the merits of this approach will follow. First, I will explain concretely how we can get from qualitative axioms as the ones above to the desired rules of probability. My aim is to show, step-by-step, how we arrive to the sum rule for probabilities of unions of exclusive events. We will need to express the axioms slightly more formally. In doing so, I adopt Cox's convention of writing 'probabilities' (even if we have not yet defined what probability is) simply as the letters representing the

[7]Howson adds a third axiom to Cox's two; the third is that if two assertions are equivalent, or have the same truth value, and two given evidence instances are equivalent, then the two conditional probabilities (of either assertion conditional on either evidence) should be the same. I follow Cox in keeping only the two rules above as axioms and noting explicitly when I use this third, arguably fundamental, fact.

propositions themselves (using brackets when necessary for clarity); for example, $a|b \wedge c$ is the probability of a given b and c.[8] As we progress in the reasoning below, we will find quantitative constraints for this expression, that will justify our calling it a probability. It is more usual to denote the probability of an event by a function of the letter representing that event, rather than by the letter itself. However, we will need to write functions of such probability functions, and the notation would quickly become messy. Some further remarks about notation: since Cox and Jaynes treat probability as attached to a *proposition*, logical notation is used, but this is entirely equivalent to attaching probabilities to sets and using set-theoretic notation. We could rewrite everything that comes below in set-theoretic terms, with the following translations: for propositions a, b we would commonly write A, B for the sets with that name; $\vee$ is the logical 'or', where $a \vee b$ means 'a is the case or b is the case or both are the case', and it is equivalent to the set-theoretic union $A \cup B$. The symbol $\wedge$ is the logical 'and', where $a \wedge b$ means 'a is the case and b is the case simultaneously'. It is the same as the set-theoretic intersection, $A \cap B$. Finally, $\neg$ is the negation: $\neg a$ means 'it is not the case that a'; it is equivalent to the complement of a set, which we write as A^C. I use 'proposition' and 'event' (which is what we usually call subsets of the whole range of possibilities, in a set-theoretic context) interchangeably throughout this chapter. Also throughout the chapter, we read, for example, $a|b \wedge \neg c$ as $a|(b \wedge \neg c)$, and $a \vee b|h$ as $(a \vee b)|h$; placing these brackets explicitly every time would result in notation which is hard to read. We can now start working directly with the axioms.

Axiom II transformed into a quantitative rule. Axiom II says the following:

$$i \wedge j|h = F\left((i|h), (j|h \wedge i)\right), \tag{15}$$

where F is some function, as yet unspecified. We want to find out more about this function, using the rules of Boolean algebra and calculus. We apply Equation 15 to $b \wedge (c \wedge d)|a = b \wedge (c \wedge d)|a$ and get

$$b \wedge c \wedge d|a = F\left((b|a), (c \wedge d|a \wedge b)\right). \tag{16}$$

We can label $x := b|a$ and rewrite equation 16 as

$$b \wedge c \wedge d|a = F\left(x, (c \wedge d|a \wedge b)\right).$$

Now we apply 15 again to $c \wedge d|a \wedge b$:

$$c \wedge d|a \wedge b = F\left((c|a \wedge b), (d|a \wedge b \wedge c)\right).$$

[8] Note that this conception of probability is always in conjunction with, or conditioned on, a given hypothesis (say H). We do not have the concept of the 'pure' probability of an event E, say $P(E)$, but only $P(E|H)$, the probability of E given H. Whether $P(E)$ or $P(E|H)$ should be considered the more fundamental concept in probability, with one defined in terms of the other, is object of some discussion in the philosophy of probability (see [12], [22]). In the present context, however, we need not worry too much about this. In the infinite lotteries considered, we always condition upon the same hypothesis (namely H = 'one, and only one, number will be picked') and when this is the case we leave it unwritten and take it for granted. This is to say, that the results obtained here are immediately applicable to the other lotteries seen in my thesis, where we would only need to make the hypothesis H explicit, in order to have completely identical notation.

Now call $y := c|a \wedge b$ and $z := d|a \wedge b \wedge c$ and we have that

$$b \wedge c \wedge d|a = F\left(x, F(y,z)\right).$$

But now apply 15 again, to $b \wedge c \wedge d|a$:

$$b \wedge c \wedge d|a = F\left((b \wedge c|a), (d|a \wedge b \wedge c)\right),$$

where we note that

$$b \wedge c|a = F\left((b|a), (c|a \wedge b)\right) = F(x,y).$$

So we also have

$$b \wedge c \wedge d|a = F\left(F(x,y), z\right).$$

And therefore we have the important equality:

$$F\left(F(x,y), z\right) = F\left(x, F(y,z)\right). \tag{17}$$

We now assume F is differentiable. Howson [27, p.19, note 14] and Jaynes [30, p.27] point out that this condition is not necessary, but it makes for a much shorter derivation; I make the assumption so I can follow Cox's proof. In any case, it does not make a substantial difference to the general argument. We call $F_i(x_1, x_2) := \frac{\partial F(x_1,x_2)}{\partial x_i}$, whatever the arguments x_1 or x_2 may be. Now we differentiate both sides of 17, first with respect to x then y to obtain two new equations. We just need the chain rule and the definition of F_i above. For example:

$$\begin{aligned}\frac{\partial F(F(x,y),z)}{\partial x} &= \frac{\partial F(F(x,y),z)}{\partial F(x,y)} \frac{F(x,y)}{\partial x} \\ &= F_1(F(x,y),z) F_1(x,y).\end{aligned}$$

The equations we obtain are:

$$\begin{aligned}F_1(x, F(y,z)) &= F_1(F(x,y),z) F_1(x,y) \\ F_1(F(x,y),z) F_2(x,y) &= F_2(x, F(y,z)) F_1(y,z).\end{aligned}$$

We can take a ratio of the two equations and write:

$$\frac{F_2(x, F(y,z))}{F_1(x, F(y,z))} F_1(y,z) = \frac{F_2(x,y)}{F_1(x,y)}. \tag{18}$$

We can multiply equation 18 by $\frac{F_2(y,z)}{F_1(y,z)}$ to obtain an equally valid equation:

$$\frac{F_2(x, F(y,z))}{F_1(x, F(y,z))} F_2(y,z) = \frac{F_2(x,y)}{F_1(x,y)} \frac{F_2(y,z)}{F_1(y,z)}. \tag{19}$$

Now call $G(u,v) = \frac{F_2(u,v)}{F_1(u,v)}$ to rewrite the above equations as:

$$G(x, F(y,z)) F_1(y,z) = G(x,y), \tag{20}$$

$$G(x, F(y,z)) F_2(y,z) = G(x,y) G(y,z). \tag{21}$$

Now we compute the derivative of equation 20 with respect to z and the derivative of equation 21 with respect to y. Clearly $\frac{\partial G(x,y)}{\partial z} = 0$. Simple computations show that

$$\frac{\partial[G(x, F(y,z))F_1(y,z)]}{\partial z} = \frac{\partial[G(x, F(y,z))F_2(y,z)]}{\partial y}$$

if it holds that

$$\frac{\partial F_1(y,z)}{\partial z} = \frac{\partial F_2(y,z)}{\partial y}.$$

This in turn holds, by what is known as Schwartz's theorem, if such second partial derivatives are continuous. This extra condition on the otherwise arbitrary function F is not noted explicitly by Cox or Jaynes, but I do not think it weakens the derivation in a significant way. Therefore, granting continuity of the second partial derivatives of F, we have that the left hand side of equations 20 and 21 are identical when differentiated with respect to z and y respectively. So, we have:

$$\frac{\partial G(x,y)G(y,z)}{\partial y} = 0.$$

Hence the product $G(x,y)G(y,z)$ must be constant in y, or in other words y must not appear. This means that in general, $G(u,v)$ must be of the form $a\frac{H(u)}{H(v)}$, with a an arbitrary constant. Using this, we can rewrite equations 20 and 21 as:

$$F_1(y,z) = \frac{H(F(y,z))}{H(y)}$$
$$F_2(y,z) = a\frac{H(F(y,z))}{H(z)}.$$

By property of differentials:

$$\mathrm{d}F(y,z) = F_1(y,z)\mathrm{d}y + F_2(y,z)\mathrm{d}z,$$

and using the identities just above:

$$\frac{\mathrm{d}F(y,z)}{H(F(y,z))} = \frac{\mathrm{d}y}{H(y)} + a\frac{\mathrm{d}z}{H(z)}.$$

Therefore we can write:

$$\int \frac{\mathrm{d}F(y,z)}{H(F(y,z))} = \int \frac{\mathrm{d}y}{H(y)} + a\int \frac{\mathrm{d}z}{H(z)}. \tag{22}$$

Thus we have

$$\exp\left(\int \frac{\mathrm{d}F(y,z)}{H(F(y,z))}\right) = \exp\left(\int \frac{\mathrm{d}y}{H(y)}\right) + \left[\exp\left(\int \frac{\mathrm{d}z}{H(z)}\right)\right]^a.$$

We can write this more efficiently if we call $w(u) := \exp\left(\int \frac{\mathrm{d}u}{H(u)}\right)$, and $w(u)$ will be an arbitrary function, since F was arbitrary (albeit with the two properties mentioned above, differentiability and continuity of second partial derivatives).

Here I depart slightly from Cox's account and I continue with the proof by Jaynes [30, p.28]. We have, from the preceding steps,

$$w(F(y,z)) = w(y)(w(z))^a.$$

And we apply this to write:

$$w(F(x,F(y,z))) = w(x)(w(F(y,z)))^a \tag{23}$$
$$w(F(F(x,y),z)) = w(F(x,y))(w(z))^a. \tag{24}$$

But recalling the result in equation 17, it must be:

$$w(x)(w(F(y,z)))^a = w(F(x,y))(w(z))^a$$

and so

$$w(x)(w(y))^a(w(z))^{a^2} = w(x)(w(y))^a(w(z))^a,$$

which is non-trivial only if $a = 1$. So now we have $w(F(y,z)) = w(y)w(z)$. Applying this to $i|h$ and $j|h \wedge i$, and recalling our initial definition at 15, we finally have a functional relation between the two plausibilities which we wanted:

$$w(i \wedge j|h) = w(i|h)w(j|h \wedge i). \tag{25}$$

Of all the arbitrary functions that w could be, we make here an assumption: that it is simply the function $w(u) = u$. This assumption is greatly simplifying, because the relation above becomes a direct relation between probabilities, and this is the main result of this section:

$$i \wedge j|h = (i|h)(j|h \wedge i) = (j|h)(i|h \wedge j), \tag{26}$$

where the second equality is obtained by recalling that $i \wedge j|h = j \wedge i|h$. The assumption above is warranted, Cox points out, because choosing to represent probability by another function would only amount to a change in notation, because we would never use the symbol $i|h$ except for in the expression $w(i|h)$ [10, p.16]. We can now obtain that the certain event has probability 1, and the impossible event has probability 0. I follow [30, pp.29-30]. It is simple to see that if we admit the basic principle that propositions with the same truth value have the same probability, and if j is certain, then $i \wedge j|h = i|h$ and also $j|h \wedge i = j|h$. (The second assertion is obvious if we consider a drawing such as the one below). From this we get that, for j certain (on given evidence h),

$$i|h = (i|h)(j|h),$$

and so $j|h = 1$. On the other hand, if $i|h$ is impossible, then $i \wedge j|h$ is also impossible, and so by the principle just invoked that propositions having the exact same plausibility must have the same probability value, $i|h = i \wedge j|h$. We also have that $i|j \wedge h$ must be impossible, supposing $j \wedge h$ itself is not a contradiction. Hence $i|h = i|j \wedge h$, and so we can rewrite equation 26, in this case, again as:

$$i|h = (i|h)(j|h),$$

but this time we want it to be true for general values of $j|h$. Hence $i|h$, the probability of an impossible event, must be either 0 or infinite. We choose 0 by convention. Hence all probability values are between 0 and 1, with 0 for the impossible event, and 1 for the certain event.

Axiom I and genesis of the sum rule. We now can use the results above to trace the genesis of the sum rule for a conjunction of events. Axiom I says that the probability of a proposition and that of its contradiction must be in functional relation. Then it must be, for some function f,

$$\neg i|h = f(i|h). \tag{27}$$

Again, the idea is to use substitutions obtained by Boolean algebra and the previous result in equation 26 to gain an insight in what this function f must be. The first important property f must have, is the following:

$$f(f(x)) = x. \tag{28}$$

To see this, we only need to apply equation 27 with $j := \neg i$. Then $j|h = f(\neg j|h) = f(f(j|h))$. We can say much more about f. We now apply it again, with $i \wedge j$ instead of i:

$$\begin{aligned} f(i \vee j|h) &= \neg(i \vee j)|h = (\neg i \wedge \neg j)|h \\ &= (\neg i|h)(\neg j|h \wedge \neg i) \\ &= f(i|h)f(j|h \wedge \neg i), \end{aligned}$$

applying the product rule obtained above in equation 26 then equation 27 again. So:

$$f(j|h \wedge \neg i) = \frac{f(i \vee j|h)}{f(i|h)},$$

which, taking f on each sides, becomes

$$j|h \wedge \neg i = f\left(\frac{f(i \vee j|h)}{f(i|h)}\right).$$

We can now use the product rule 26 again to see that

$$j|h \wedge \neg i = \frac{\neg i \wedge j|h}{\neg i|h} = \frac{\neg i \wedge j|h}{f(i|h)}.$$

And so:

$$\neg i \wedge j|h = f(i|h)f\left(\frac{f(i \vee j|h)}{f(i|h)}\right).$$

We apply the product rule once again to the left hand side to get:

$$\neg i \wedge j|h = j \wedge \neg i|h = (j|h)(\neg i|h \wedge j) = (j|h)f(i|h \wedge j) = (j|h)f\left(\frac{i \wedge j|h}{j|h}\right),$$

and finally the equality:

$$(j|h)f\left(\frac{i \wedge j|h}{j|h}\right) = f(i|h)f\left(\frac{f(i \vee j|h)}{f(i|h)}\right). \tag{29}$$

Suppose we now apply this equation to propositions i, j such that $i \wedge j = i$ and $i \vee j = j$. If we think in terms of sets and set operations, i is a subset of j; in yet other terms, j implies i. Then the equation above becomes:

$$(j|h)f\left(\frac{i|h}{j|h}\right) = f(i|h)f\left(\frac{f(j|h)}{f(i|h)}\right).$$

A little clarification might be helpful here. We will use this equation to narrow down the properties that function f must have, which is what we are looking for. However, this does not mean that the results are limited to events of the type just described above—but because they *could* be, f must be valid in this case too. Hence the result we obtain for f must satisfy this equation. We make the following substitutions to rewrite more simply: $y := f(i|h)$, and $z := j|h$. Here is the equation which must hold:

$$zf\left(\frac{f(y)}{z}\right) = yf\left(\frac{f(z)}{y}\right). \tag{30}$$

We can call $u := \frac{f(y)}{z}$ and $v := \frac{f(z)}{y}$. If we differentiate equation 30 with respect to y, to z and to y and z we obtain three new equations (I omit the standard manipulations); combined with equation 30, we obtain the following equation:

$$\frac{uf''(u)f(u)}{(uf'(u) - f(u))f'(u)} = \frac{vf''(v)f(v)}{(vf'(v) - f(v))f'(v)}. \tag{31}$$

Cox here simply comments that u and v are 'mutually independent' [10, p.21], and so, since the function on both sides is the same, this function must be constant for an arbitrary constant x. Of course he does not mean probabilistic independence, for we have not defined yet. Expanding briefly on what he means, we can say the following. Suppose f is of the type $f(x) = k-x$. This satisfies the only condition on f we have so far, $f(f(x)) = x$. Then if $u = \frac{f(y)}{z} = \frac{i|h}{j|h} = M$, say, we have $v = \frac{f(z)}{y} = \frac{f(j|h)}{f(i|h)} = \frac{k-(j|h)}{k-(i|h)}$. This can take arbitrary values, regardless of the value of M, which does not appear. And now suppose we know that $v = \frac{f(z)}{y} = \frac{f(j|h)}{f(i|h)} = \frac{k-(j|h)}{k-(i|h)} = N$. Again, $u = \frac{f(y)}{z} = \frac{i|h}{j|h}$ can take arbitrary values, as it does not depend on N. Therefore, we can now conclude as follows: suppose we take the value of equation 31 for an arbitrary x; then this must be equal to a constant c, since equation 31 takes the same value for arbitrary values of its argument. Therefore we can write:

$$xf''(x)f(x) = cf'(x)(xf'(x) - f(x)),$$

or

$$x\frac{\mathrm{d}f'}{\mathrm{d}x}f(x) = cf'(x)(x\frac{\mathrm{d}f}{\mathrm{d}x} - f(x)),$$

which Cox [10, p.21] rewrites, with a little manipulation, as:

$$\frac{\mathrm{d}f'}{f'} = c\left(\frac{\mathrm{d}f}{f} - \frac{\mathrm{d}x}{x}\right).$$

Integrating on both sides, we get

$$\log|f'| = c\log|f| - c\log|x| + A',$$

(A' is just a integration constant). Cox derives directly the following conclusion:

$$f' = A\left(\frac{f}{x}\right)^c, \tag{32}$$

where A is another constant of integration. But to get rid of the absolute value we need to assume the arguments are positive. This assumption has already been made for probability values, and this extends to x and $f(x)$, since the latter is also a probability value by equation 27. However, for $f'(x)$ I think we must make the additional assumption now, namely that the derivative f' is positive, and hence f is in fact a non-decreasing function. Now we have that equation 32 is a differential equation, which is in fact separable. Solving in the usual way, we have

$$f^{1-c} = Ax^{1-c} + B, \tag{33}$$

B another constant of integration. Now we can apply the two constraints we had found for f to narrow down what it should be. Firstly, applying equation 33 to equation 30 we get easily that $(A^2 - B)y = (A^2 - B)z$. So if we want the equation to be valid for general y, z, it must be $B = A^2$. Note that in equation 30 there is in fact a relation between the variables, namely that i is a sub-event, or is implied by j, and the variables y and z are defined in terms of conditional probabilities of i and j. However, the relation between i and j only dictates that one probability will be lesser than or equal to the other; we still require equation 30 to be valid for otherwise arbitrary values. Secondly, f, as defined in equation 33, should satisfy equation 28. Therefore, the following third-degree equation in A must be satisfied: $A^3 + (x+1)A^2 - x = 0$. The only solution that does not involve x (keeping in mind that A is constant) is $A = -1$. Neither of these two conditions put restraints on the exponent $1 - c$. Therefore we now have:

$$f^{1-c}(x) = -x^{1-c} + 1.$$

And recalling the definition of the function f:

$$(i|h)^{1-c} + (\neg i|h)^{1-c} = 1.$$

This is nearly the equation we expect for the sum of the probabilities of an event and its complement, except for the exponent $1 - c$. The conditions on f do not specify what this should be; as a matter of convenience, then, we are free to take it equal to 1 to finally obtain:

$$(i|h) + (\neg i|h) = 1. \tag{34}$$

These results are the first in which a sum appears. As we will see shortly, it will be possible to derive the desired sum rule for a conjunction of events. First note that applying the product rule at equation 26 we can get

$$(i \wedge j|h) + (i \wedge \neg j|h) = (i|h)((j|h \wedge i) + (\neg j|h \wedge i)),$$

which, when combined with equation 34, gives the expected property:

$$(i \wedge j|h) + (i \wedge \neg j|h) = (i|h). \tag{35}$$

Now recall that for any a, b $a \vee b = \neg(\neg a \wedge \neg b)$, so we can apply equation 34 as follows:

$$\begin{aligned} 1 &= ((\neg a \wedge \neg b)|h) + (\neg(\neg a \wedge \neg b)|h) \\ &= ((\neg a \wedge \neg b)|h) + ((a \vee b)|h). \end{aligned}$$

We can apply equations 34 and 35 again to the first member on the right hand side above:

$$((\neg a \wedge \neg b)|h) = \neg a|h - (\neg a \wedge b|h) \quad \text{by equation 35} \tag{36}$$
$$= 1 - (a|h) - (\neg a \wedge b|h) \quad \text{by equation 34.} \tag{37}$$

And so

$$(a \vee b)|h = (a|b) + (\neg a \wedge b|h).$$

But by equation 35,

$$(\neg a \wedge b|h) = (b|h) - (b \wedge a|h),$$

and so finally we have the sum rule for a conjunction of events:

$$a \vee b|h = (a|h) + (b|h) - (b \wedge a|h). \tag{38}$$

Cox proves the general version of this rule for any number n of propositions [10, pp.25-28]. Here we are only interested in mutually exclusive propositions, such that $b \wedge a|h$ is impossible and so has probability 0. The generalisation to any number n is immediate. Suppose $a_1, \ldots, a_n$ are n mutually exclusive propositions. Then:

$$(a_1 \vee \cdots \vee a_n|h) = ((a_1 \vee \cdots \vee a_{n-1}) \vee a_n|h) \tag{39}$$
$$= ((a_1 \vee \cdots \vee a_{n-1})|h) + (a_n|h) \tag{40}$$
$$\vdots \tag{41}$$
$$= (a_1|h) + (a_2|h) + \cdots + (a_n|h). \tag{42}$$

This is the result we were after, and this is precisely finite additivity.

5.3 Jaynes and countable additivity

I mentioned in the introductory discussion of this chapter that it would be interesting to see whether we can obtain countable additivity from Cox's derivation, by some modification which would not just be countable additivity itself. We quickly see this seems impossible. Adding countable additivity to the condition just above, is precisely that, just an addition of countable additivity as a further property. In Cox's argument I explained above, additivity derived from the relation expressed in Axiom II, which is between an event and its complement; this enables us to interpret the result of the differential equation 32, in equation 33, as a functional relation between the probability of an event and that of its complement. We then develop this into the relation between the probability of a conjunction of two events, and that of their individual probabilities (equation 38). This is a binary relation, and as such can be extended easily to a finite number of events, as is done in equations 39 and following, but to extend the relation to an infinite number of events would be simply to assume countable additivity. It is not easy to see how we could modify the derivation 'earlier', before assuming infinite additivity outright, in order to get it as a final result without explicitly demanding it from the start. By this I mean a way to obtain countable additivity which is akin to the way we obtained finite additivity above, without it being mentioned in the initial axioms. Somewhat tentatively, I would say that proceeding to additivity from the binary relation between the probabilities of an event and its complement, can only bring us as far as finite additivity. Even if the event in question, or its complement, were infinite unions of events, we still would only have the definition of probability of a *finite* union between events, which happen to be themselves infinite unions of sub-events; this would tell us nothing about breaking down their probability into the single probabilities of these sub-events. This is what we can conclude, then: if we accept the interpretation of probability which arises from Cox's derivation, we should also accept that countable additivity cannot be one of our axioms—it may or may not hold—unless we add it 'manually'. This is why Jaynes' position in his [30] is extremely interesting: while basing his approach on Cox's, he also clearly thinks that either a countable number of events have probabilities which are countable additive, or probability on such set should not be allowed to exist and be considered absurd [30, p.464]. But I find that his argument unfortunately does not clarify the situation. He denies that countable additivity is a stronger assumption than finite additivity [30, p.466], when it clearly is: all countably additive probability measures are also finitely additive; the opposite is not true. He reports an example by Feller in which, without countable additivity, we could have a measure that assigns 0 to all finite intervals and 1 to the whole line [30, p.465]. He comments that "We are trying to make a probability density that is everywhere zero, but which integrates to unity. But *there is no such thing*, according not only to all the warnings of classical mathematicians from Gauss on, but according to our own elementary common sense" [30, p.465](emphasis in original). But there clearly *is* such a probability measure, it seems to me, since he has just defined it. And, as we will see shortly, it satisfies all rules that we could derive from Cox's proof. So if we want to exclude such probability measures, we must add some rule to do this—this is countable additivity. What is particularly weak, in my view, in Jaynes' argument against merely finitely additive measures, is that, by his own assertion, 'our own elementary common sense' is no argument for or against a mathematical result. Jaynes bemoans the "psychological phenomenon"in which

> someone asserts a principle that seems to him intuitively right, and when probability analysis reveals the error, instead of taking this opportunity to educate his intuition, he reacts by rejecting the probability analysis. For him, his intuitive *ad hoc* principle takes precedence over the rules of probability theory. [...]. One can be so deeply committed to his position that mathematical proof to the contrary, and any number of counter-examples, carry no weight for him [30, p.488] (emphasis in the original).

But it seems to me that Jaynes himself is victim of the very same phenomenon. Here is the probability measure which is viewed as "weird" [30, p.465]: we let $F((a, b)) = 0$ for all $b < \infty$, and $F((a, \infty)) = 1$. We suppose $a > 0$. Let us check that Cox's rules for probability measures are satisfied. The additivity reported by Feller only regards the case in which we divide an interval I into a finite number of non-overlapping intervals I_n with $\bigcup_n I_n = I$. Firstly, all events have probability between 0 and 1, trivially. Secondly, the certain event, as required in [10, p.17], has probability 1:[9]

$$\begin{aligned} i \vee \neg i | h = i \vee \neg i &= F((a,b) \cup (a,b)^C) \\ &= F\left((a,b) \cup (0,a] \cup [b,\infty)\right) \\ &= F((0,\infty)) = 1 \end{aligned}$$

Finally, take a finite sequence $(I_n)_{1 \leq n \leq N}$ of disjoint intervals $I_n = (a_n, b_n)$, such that $\bigcup\limits_{n=1}^{N} I_n = I$ for some interval I. Then I is either a finite interval (a,b) or has an infinite right end-point, (a,∞). Suppose the former is the case. Then

$$F(I) = F\left(\bigcup_{n=1}^{N} I_n\right) = 0.$$

The sub-intervals $I_n = (a_n, b_n)$ must clearly all be finite, giving

$$\sum_{n=1}^{N} F(I_n) = 0 + \cdots + 0 = 0,$$

so that $F\left(\bigcup\limits_{n=1}^{N} I_n\right) = \sum\limits_{n=1}^{N} F(I_n)$, and finite additivity holds in this case. Now suppose $I = (a,\infty)$. Then

$$F(I) = F\left(\bigcup_{n=1}^{N} I_n\right) = 1.$$

Now, there must be at least one sub-interval I_k with $I_k = (a_k, \infty)$, because if this were not the case, there would be a largest $b_j < \infty$ such that $\bigcup\limits_{n=1}^{N} I_n = (a, b_j) \subsetneq I$. On the other hand, there must be only one such interval with infinite right end-point. To see this, suppose this were not the case, and we had $(a_{k_1}, \infty), (a_{k_2}, \infty) \in \bigcup\limits_{n=1}^{N} I_n$. Then $(a_{k_1}, \infty) \cap (a_{k_2}, \infty) = (a_K, \infty) \neq \emptyset$, where $a_K = \max\{a_{k_1}, a_{k_2}\}$, contradicting the fact that all sub-intervals are disjoint. Hence we have

$$\sum_{n=1}^{N} F(I_n) = 0 + \cdots + 0 + 1 + 0 + \cdots + 0 = 1,$$

giving again $F\left(\bigcup\limits_{n=1}^{N} I_n\right) = \sum\limits_{n=1}^{N} F(I_n)$. Therefore, finite additivity is ensured, and F satisfies all of Cox's quantitative rules for probability. What happens if I is divided into a countably infinite number of intervals, whose union is equal to I? If I is finite, then $F(I) = F\left(\bigcup\limits_{n=1}^{\infty} I_n\right) = 0$. Of course, all sub-intervals must be finite too, and so $\sum\limits_{n=1}^{\infty} F(I_n) = 0$.[10] If $I = (a, \infty)$, we know that

[9] I ignore the hypothesis h because a truism is certain on any hypothesis [10, p.17].

[10] To divide a finite interval into an infinite number of sub-intervals, it is enough that each sub-interval has a decreasing width $b - a$, with the widths forming a converging series. For example, we could divide the interval $(0,1)$ in the sub-intervals $(0, \frac{1}{2}], (\frac{1}{2}, \frac{1}{2} + \frac{1}{4}], \ldots, (a_n, a_n + \frac{1}{2^n}], \ldots$, with the widths adding up to 1.

$F(I) = F\left(\bigcup_{n=1}^{\infty} I_n\right) = 1$. However, since we allow this interval to be divided into an infinite number of sub-intervals, we have two cases: either one (and only one, by the argument above) of the sub-intervals is of the type (a_k, ∞), in which case $\sum_{n=1}^{\infty} F(I_n) = 1$. But it could also be that all sub-intervals are finite. (For example, the infinite union of all intervals of unit length, $(0, 1] \cup (1, 2] \cup \ldots$, clearly is equal to the whole positive real line.) In that case, $\sum_{n=1}^{\infty} F(I_n) = 0$. Hence in general we can only affirm the following:

$$F\left(\bigcup_{i=1}^{\infty} I_i\right) \geq \sum_{i=1}^{\infty} F(I_i)$$

and so we do not have countable additivity in general. What Jaynes requires, to avoid this 'weird' example, as he calls it, is an additional condition on infinite sums: he requires countable additivity. But we see now that his argument in favour of the principle is circular: he requires countable additivity to avoid cases of probability measures which are not countably additive. Jaynes' argument relies on something which he considers obvious, but clearly many others do not, and is not prescribed by Cox. Curiously, Jaynes' point can be reduced to Kolmogorov's well-known and concise remark on the subject. Here is what Jaynes says:

> [...] it is a trivial remark that our probabilities have 'finite additivity'. As $n \to \infty$ it seems rather innocuous to suppose that the sum rule goes in the limit into a sum over a countable number of terms, forming a convergent series; whereupon our probabilities would be called countably additive. Indeed (although we do not see how it could happen in a real problem), if this should ever fail to yield a convergent series we would conclude that the infinite limit does not make sense, and we would refuse to pass to the limit at all [30, p.465].

And here is what Kolmogorov famously wrote: "*We limit ourselves, arbitrarily, to only those models which satisfy [countable additivity]*" [33, p.15] (emphasis in the original). The word Jaynes would dispute, I believe, is 'arbitrarily'; I have argued, however, that his arguments are not convincing, and thus his choice to consider only countably additive measures may indeed be viewed as arbitrary.

5.4 Remarks

As explained above, Jaynes and Howson use the same foundational argument for probability, but reach opposite conclusions regarding countable additivity. What can we conclude from this? What were the arguments from either side? The only argumentations that emerged, amounted to claiming the obviousness, or the desirability, of the *intuition* which should be respected in the issue of additivity.[11] The problem, of course, is that these obvious intuitions are in conflict. Now Jaynes' and Howson's opposite conclusions, reached from the

[11] Howson in [27] also points to some similarities between logic and finitely additive probability. This does not seem a strong argument in favour of finite additivity in general, nor does

same starting point (Cox's axioms and derivation), can be explained. Other interesting points emerge from this discussion.

Ad-hockery of countable additivity. Firstly, we could see a partial balancing-out of the ad-hockery accusations levelled at the assumption of countable additivity: namely, that it is a principle that has nothing to do with the nature of probability, and is only assumed for technical convenience (see, for example [34], [39], [14] and [11]). It is often highlighted how even Kolmogorov himself apparently supported this idea—and if the very person who introduced countable additivity thought it was somewhat extraneous to probability, the arguments seem to suggest, then surely that must be the case. We see this, an interpretation of what Kolmogorov meant by conceding that the requirement was arbitrary, in Schurz and Leitgeb's [34, p.258], and Wenmackers and Horsten's [39, p.59]. The former say: "Already Kolmogorov has emphasized that the condition of σ-additivity is merely a useful assumption of idealization but is not contained in the meaning of 'probability'" [34, p.258]; an identical feeling is expressed in the latter. I think this is too strong an interpretation of the passage which both papers cite. It is the famous passage I quoted partially above; here it is in full:[12]

> Since the new axiom [i.e. countable additivity] is essential for infinite fields of probability only, it is almost impossible to elucidate its empirical meaning, as has been done, for example, in the case of Axioms I - V [. . . ; the axioms regulating finitary probability]. For, in describing any observable random process we can obtain only finite fields of probability. Infinite fields of probability occur only as idealized models of real random processes. *We limit ourselves, arbitrarily, to only those axioms which satisfy Axiom VI.* This limitation has been found expedient in researches of the most diverse sort [33, p.15].

Kolmogorov only says that we cannot give an *empirical meaning* to a rule regarding infinite sets; this seems straightforward. He does not refer here to the meaning of probability itself, and the interpretation above seems forced. Reading a few pages after this passage, the impression is strengthened. Kolmogorov presents an *extension theorem*, which is a theorem to the effect that a finitely additive measure μ_0 defined on an algebra $\mathcal{A}_0$, can always be uniquely extended to a countably additive measure μ on a σ-algebra $\mathcal{A}$, with $\mu(E) = \mu_0(E)$ for all $E \in \mathcal{A}_0$ (see Section 2.2 for a reminder of these definitions). After the theorem, he comments that while the sets of an algebra can be considered as ideally corresponding to actual random events, this is not the case with a σ-algebra, which is a purely mathematical concept. And this is how he continues (in our notation):

> Thus the sets of $\mathcal{A}$ are generally merely ideal events to which nothing corresponds in the outside world. However, if reasoning which utilizes the probabilities of such ideal events leads us to a determination of the probability of an actual event of $\mathcal{A}_0$, then, from an

the author, it seems to me, present it as such. We would need a further justification of why it is vital to have these particular similarities between logic and probability, such that countable additivity should be excluded.

[12] An infinite field is a σ-algebra; a field is an algebra; see Section 2.2.

empirical point of view also, this determination will automatically fail to be contradictory [33, p.18].

This comment is a simple consequence of the fact that $\mu(E) = \mu_0(E)$ for all $E \in \mathcal{A}_0$, and that we are allowed to consider sets in $\mathcal{A}_0$ as corresponding to real random events. It seems wrong to claim Kolmogorov emphasised countable additivity to be extraneous to the idea of probability: we would ignore that he called the axiom 'essential for infinite fields', and that he said we can correctly transfer the results obtained in such infinite fields back to reality. Be that as it may, let us return to the partial balancing-out of the adhockery accusations. Why were the two axioms chosen as they were by Cox? Presumably, in order to obtain a derivation of the laws of probability and provide a logical justification for them. While these two axioms are definitely reasonable, it would be quite a stretch to claim that they are obviously fundamental to uncertain reasoning. So if we do not want to commit to giving an exceptional weight to these two particular axioms (and all the mathematical manipulation needed to derive the probability axioms from them), we would have to admit that these too are chosen *ad hoc* for the purpose of justifying probability. It seems to me that in principle, this is not too different from assuming countable additivity to obtain powerful integration techniques, for example. We assume a principle which seems reasonable (probabilities are additive in a countably infinite setting), in order to justify the techniques we wish to use for uncertain reasoning. It is not obvious why assuming axioms which are supposedly 'more fundamental' (Howson believes they are [27, p.19]), *then* deriving the rules we wanted in the first place, would put us on more solid ground. Firstly, we would still need to justify the first axioms (with other axioms?). And secondly, that these other axioms are more fundamental could be questionable. For this see the point below.

On which axioms are more fundamental. With regards to which principles should be seen as more fundamental, between the usual definition of probability or Cox's axioms, we note the following: in Cox's derivation we needed to have a pre-existing, well-formed concept of integral, both in equation 22 and the subsequent ones, and in the solution to the differential equation 32. In mathematical probability, there is no need for this. As we saw in Chapter 2, important properties of integrals are not assumed as already existing, in order to define probability: rather, we define probability axiomatically as a measure, and then construct a powerful and flexible theory of integration from it. Given the importance of integration in probability, this could be seen as more natural. This, I believe, is an argument against the perceived position of Cox's qualitative axioms (I) and (II) as more 'fundamental' than the direct axiomatic definition of probability. It highlights that the 'naturalness' of qualitative axioms for probability could be disputed, for it relies critically on non-qualitative, pre-existing concepts in order for one of its most important instruments to work. An axiomatic definition of probability is not qualitative, and stipulates directly what features a probability measure must have. However, we can use this as a starting point for a natural definition of integral. We leave this discussion at that, to return to infinite lotteries. This time, a different kind of solution is sought, one which hopes to go beyond the deadlock we have explored thus far between arguments for and against countable additivity.

6 α-theory and solutions to the infinite lottery problem

6.1 Setting

Being extremely concise, the problem of the infinite lottery can be explained thus: we would like to assign equal probability to each number, but any real number is too big for this, and 0 is too small. Perhaps, then, it could be useful to go first back to *finite* lotteries for a moment, to try and extrapolate what makes them fair, and apply that somehow to an infinite case. In this chapter I present such an approach, given by Wenmackers and Horsten in [39]. I follow their setting, but I refer the reader to their article and to [4] for the complete formal approach using ultrafilters. Not being familiar with ultrafilters, I limit myself to explaining their approach by using α-theory (as is also done in that article, [39]), as presented in [5]. α-theory is an axiomatic approach to *non-standard analysis*, and the numbers involved are objects of this branch of mathematics. I will make little use of non-standard analysis in general, however, being more concerned with finding a 'practical' solution to the infinite lottery problem. Of course, this practical solution is backed up and given plausibility by its formal theoretical background, but for this I refer the reader to the articles I cite above. I will introduce some elements of α-theory shortly, insofar as they are needed for the discussion in this chapter. The explanations for the result I present are equivalent if we use ultrafilters or α-theory, and I find the latter approach easier to follow. Before presenting this result, I propose an approach of my own. It is very similar to the one in [39], and indeed identical in spirit and in the results of the simple worked-out examples given in that paper. I follow the main ideas of Wenmackers and Horsten, except for some small changes, and the fact that I derive a probability measure entirely from the instruments of α-theory. I do not wish to claim that my approach is better. However, I found α-theory intuitively more approachable, and perhaps a solution that uses only its instruments can help in the argument that this type of solution (using α-theory or ultrafilters, but the bottom line is that they all use non-standard analysis) can be understandable and desirable. Moreover, the proof of additivity resulted different, in my construction, than that in [39]. After having presented both probability functions, I show that they are indeed equivalent.

6.2 Finite lotteries, asymptotic densities and where they fail

Let us start, then, from finite fair lotteries. Throughout this chapter the sample space will be the set of natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$. I simply assume each ticket is numbered according to the natural numbers, and call the tickets by their numbers: $P(k)$ is the probability that the ticket labelled by the number k has of winning. I consider lotteries in which there is only one winning ticket, and so all events are mutually exclusive, and one ticket must win, so the events are exhaustive. This gives a simplified additivity, as we do not need to subtract the probability of more than one ticket winning simultaneously, from the probability of a union of tickets. For any lottery with n tickets, the fair probability of each ticket is $\frac{1}{n}$. For any finite subset A of tickets, the probability, respecting finite additivity, should be $\frac{\#(A)}{n}$, where $\#(A)$ is the function that maps a finite set to its number of elements. What is clearly tempting, then, is to take the limit at infinity of this fraction. This is called asymptotic density ad as is defined as just described:

$$ad(A) = \lim_{n\to\infty} \frac{\#(A \cap \{1, \ldots, n\})}{n}, \tag{43}$$

if such limit exists. I will call $ad_n(A) := \frac{\#(A\cap\{1,\ldots,n\})}{n}$. Clearly the numerator will (eventually, or in the limit) give the size of the set A in question. However, this limit does not exist for all subsets of $\mathbb{N}$, and the sets for which it does exist do not constitute an algebra (so it could happen that we know the probability of two sets, for example, but not of their union). Therefore, such limit is not a good solution to the infinite lottery problem [39, p.42]. In the article just cited, there is mentioned the following example: take the set of natural numbers that, when written in binary notation have an even number of digits. Then, for this set, the limit ad does not exist, and hence this set has no probability value according to the formula above (note 6 in [39, p.42]). The authors do not elaborate further on this example. I will treat it in some detail, because it seems perfect to test whether, and in what sense, a non-standard solution can be better than ad. Hence I will describe it in the section that follows, then take it up again below, after I have described a new probability function.

Example of a set with no *ad* value. How many numbers have a binary expansion with an even number of digits? Rather than thinking from the natural numbers, it is simpler to reason directly in terms of binaries. Suppose we have k digits, and each can be 0 or 1. Then we can form 2^k different combinations, but half of them must start with a 0. Since we ignore these initial 0s in the notation of an integer, we actually have 2^{k-1} different numbers. Then, supposing we start from the number 1, we can get one number from one digit (1 in binary, which is 1 also in decimal notation); from $k = 2$ digits, we have $2^1 = 2$ numbers (10 and 11, which correspond to 2 and 3 in base 10); from $k = 3$ digits we have $2^2 = 4$ numbers, and so on. We call A the set of all those natural numbers which arise from even values of k. We want to find an expression for $ad_n(A)$ for finite n. Dividing the natural numbers according to the number of digits of their binary expansion, we will have the following scenario: we have 1 number with one digit; followed by 2^1 numbers with 2 digits; followed by 2^2 numbers with 3 digits; and so on. Integers with an odd number of binary digits (I will call them 'odd-digit numbers') are in the sets $[2^{2k}, 2^{2k+1})$, while numbers with an even number of binary digits ('even-digit numbers') are in the sets $[2^{2j+1}, 2^{2j})$. The way to see which of these sets a natural number n belongs to, is to take its logarithm in base 2—or rather, the floor function (the largest integer smaller than a given number) of its logarithm, $\lfloor \log_2(n) \rfloor$. However, we see that we will have different formulae for the cases in which n falls in an interval $[2^{2k}, 2^{2k+1})$ or an interval $[2^{2j+1}, 2^{2j})$, i.e. whether $\lfloor \log_2(n) \rfloor$ is respectively odd or even. This is because we are counting, in the numerator of ad_n, only how many numbers, less than or equal to n, fall in intervals $[2^{2j+1}, 2^{2j})$. The denominator of ad_n will of course always be n, but the numerator is seen to be, respectively: the number of even-digit numbers up to $2^{\lfloor \log_2(n) \rfloor}$, if $\lfloor \log_2(n) \rfloor$ is even; and n minus the number of odd-digit numbers up to $2^{\lfloor \log_2(n) \rfloor}$, if $\lfloor \log_2(n) \rfloor$ is odd. Hence

$$\#(A \cap \{1, \ldots, n\}) = \sum_{i=0}^{\frac{\lfloor \log_2(n) \rfloor}{2} - 1} 2^{2i+1} \quad \text{if } \lfloor \log_2(n) \rfloor \text{ is even,}$$

and

$$\#(A \cap \{1, \ldots, n\}) = n - \sum_{i=0}^{\frac{\lfloor \log_2(n) \rfloor - 1}{2}} 2^{2i} \quad \text{if } \lfloor \log_2(n) \rfloor \text{ is odd.}$$

These are sums of powers, so the function for the probability of A for each n can be written as:

$$ad_n(A) = \frac{2^{\lfloor \log_2(n) \rfloor + 1} - 2}{3n} \quad \text{if } \lfloor \log_2(n) \rfloor \text{ is even} \tag{44}$$

and

$$ad_n(A) = 1 - \frac{2^{\lfloor \log_2(n) \rfloor + 1} - 1}{3n} \quad \text{if } \lfloor \log_2(n) \rfloor \text{ is odd.} \tag{45}$$

Inspection of the function immediately shows that it does not have a limit. The function grows in intervals in which $\lfloor \log_2(n) \rfloor$ is odd, and decreases when this number is even. The first assertion follows from the fact that $ad_n(A) < 1$, and in general if $\frac{a}{b} < 1$, then $\frac{a+1}{b+1} = \frac{a}{b} + \frac{b-a}{b(b+1)} > \frac{a}{b}$. The second follows from the fact that in intervals in which $\lfloor \log_2(n) \rfloor$ is even, we add 1 to the denominator and 0 to the numerator at each step. We also see that for all n which have $n = 2^{2k}$ for some k, $ad_n(A) = \frac{2}{3} - \frac{2}{3n}$, and for any n that has $n = 2^{2j+1}$ for some j, we have $ad_n(A) = \frac{1}{3} + \frac{1}{3n}$. We also notice that it is bounded above—by a smaller number than the obvious bound 1. To see this for the $\lfloor \log_2(n) \rfloor$ even case (just 'the even case' for the rest of the treatment of this example), we can rewrite the expression in equation 44 as follows:

$$ad_n(A) = \frac{2^{\lfloor \log_2(n) \rfloor - \log_2(n) + 1}}{3} - \frac{2}{3n} < \frac{2}{3}, \tag{46}$$

while the case for $\lfloor \log_2(n) \rfloor$ odd ('the odd case') gives:

$$ad_n(A) = 1 - \frac{2^{\lfloor \log_2(n) \rfloor - \log_2(n) + 1}}{3} + \frac{1}{3n} \leq \frac{2}{3}. \tag{47}$$

So $ad(A) \leq \frac{2}{3}$. We can also find an expression in terms of n for the lower bound. For the even case, we look again at expression 46 and note that this is smallest when $\lfloor \log_2(n) \rfloor - \log_2(n) + 1$ has its minimums, which is in the cases where n is the number before $2^{\lfloor \log_2(n) \rfloor + 1}$, so that $n = 2^{\lfloor \log_2(n) \rfloor + 1} - 1$. Substituting this into expression 44 gives $ad_n(A) = \frac{1}{3} - \frac{1}{3n}$. For the odd case, the lowest value attained is just the value at any n that has $n = 2^{2j+1}$ for some j. This is because the function is increasing for n from $n = 2^{2j+1}$ to $n = 2^{2j+2} - 1$. In this case $ad_n(A) = \frac{1}{3} + \frac{1}{3n}$. We saw, then, that $\liminf ad_n(A) = \frac{1}{3}$, while $\limsup ad_n(A) = \frac{2}{3}$. This means that the limit $\lim_{n \to \infty} ad_n(A) = ad(A)$ does not exist, and the set A cannot be assigned a probability value. We note that we know a range in which the value of the probability of A must be, however it really seems a contradiction in terms to say (as we must, since the relevant limit does not exist), 'the set A does not have a probability' and at the same time affirm that 'the probability of the set A lies between such-and-such extremes'. Hence we may want to relax this definition of probability to include difficult cases like this one. For example, we could modify $ad(A)$ to fix this as follows. The limit inferior and limit superior of a sequence always exist for any sequence, so we could redefine $ad(A)$ to at least take an interval value, $ad(A) \in (\limsup ad_n(A), \liminf ad_n(A))$ if $\limsup ad_n(A) - \liminf ad_n(A) > 0$, and $ad(A) = \lim_{n\to\infty} ad_n(A)$ if the limit exists (i.e. $\limsup ad_n(A) - \liminf ad_n(A) = 0$). While this is in a sense more promising, we still do not get countable additivity. In fact, if we take any singleton $\{m\}$, $m \in \mathbb{N}$, then $ad(\{m\}) = \lim_{n\to\infty} \frac{1}{n} = 0$. However, while $\mathbb{N} = \bigcup_{m=1}^{\infty} m$, we have that $ad(\mathbb{N}) = \lim_{n\to\infty} \frac{n}{n} = 1$. Note that any finite subset A of natural numbers would also have $ad(A) = 0$. This suggests that if we want a probability function which: (1) takes on a value for any subset of $\mathbb{N}$; and (2) has additive properties, both for finite unions and for infinite unions of sets; then we should look at other ways of approaching the concept of values 'at infinity'. This is the basic idea of α-calculus (and of non-standard analysis in general), which I introduce below. As will become apparent, it is possible to satisfy the two properties above; however, we have to define a new concept of infinite sum; and we have to give up on real-valued probabilities.

6.3 α-theory and a different concept of size

Looking at equation 43, it is clear that the numerator will give the size of any finite set A, by giving the number of its elements. But what is the 'size' of $\mathbb{N}$, over which, in the limit, we will be dividing? When we are applying the counting function to $A \cap \{1, \ldots, n\}$, for increasing numbers of n, we are effectively 'adding a 1' for each element in the intersection. We can put this more clearly as follows. Let $\mathbf{1}_A$ be the characteristic function of the set A, defined as

$$\mathbf{1}_A(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{if } n \notin A \end{cases}$$

Then we can create a sequence of 0s and 1s as we apply $\mathbf{1}_A$ to an initial section of $\mathbb{N}$ (the complete sequence of 0s and 1s is what is called a 'characteristic bit-string' in [39]). I call the partial bit-string up to $n \in \mathbb{N}$ simply $T(n)$:

$$T_A(n) = \bigcup_{i=1}^{n} \{\mathbf{1}_A(i)\},$$

where the curly brackets are used to indicate that we consider the *set* $\{1\}$, say, rather than just the number 1. As an example, suppose A is the set of odd integers; then $T_A(6) = \{1, 0, 1, 0, 1, 0\}$. (I omit the curly bracket notation in future writings of $T(n)$, giving the meaning of the notation for granted.) Now call $S(n)$ the partial sum of elements of $T(n)$:

$$S_A(n) = \sum_{x \in T(n)} x = \sum_{i=1}^{n} \mathbf{1}_A(i).$$

We have simply written an equivalent form of the counting function #, but written in this way we will be able to apply to it the rules of α-calculus directly. The idea for our probability measure is to consider the sequence of partial sums $S(n)$ and its value 'at infinity', and this will be its assigned size; we then divide this number by the size assigned to $\mathbb{N}$. The size of $\mathbb{N}$ is where α makes its first appearance. Suppose we take the bit-string $T(n)$ of $\mathbb{N}$. Then this is $T_{\mathbb{N}}(n) = \{1, 1, \ldots, 1\}$ for all n. Therefore, $S_{\mathbb{N}}(n) = n$, for all n. We can think of α as an infinitely large integer, and of the α-limit of a sequence as its 'value at infinity'. We *define* the α-limit of the identity sequence, such as $S_{\mathbb{N}}(n)$, to be: $S_{\mathbb{N}}[\alpha] = \alpha \notin \mathbb{N}$. The main advantage of taking α-values instead of conventional limits is that, while the limit of a sequence may or may not exist, *the α-value of a sequence always exists.* Here are the five axioms which determine the properties of α. I quote them nearly *verbatim* from [5], except for some notation which I found more convenient. Note that *atoms* here are those primitive objects which are not sets. Numbers, for example are considered to be atoms. (We could consider all numbers to be 'sets', but this is avoided in [5] for the sake of clarity.)

1 *Extension Axiom.* For every sequence φ there is a unique element $\varphi[\alpha]$, called the α-limit or α-value of φ.

2 *Composition Axiom.* If φ and ψ are sequences and if f is any function such that the compositions $f \circ \varphi$ and $f \circ \psi$ make sense, then

$$\varphi[\alpha] = \psi[\alpha] \Rightarrow f \circ \varphi[\alpha] = f \circ \psi[\alpha]$$

3 *Number Axiom.* If $c_r : n \mapsto r$ is the constant sequence with value $r \in \mathbb{R}$, then $c_r[\alpha] = r$. If $I_{\mathbb{N}} : n \mapsto n$ is the identity sequence on $\mathbb{N}$ then $I_{\mathbb{N}}[\alpha] = \alpha \in \mathbb{N}$.

4 *Pair axiom.* For all sequences φ, ψ and υ:

$$\upsilon(n) = \{\varphi(n), \psi(n)\} \text{ for all } n \Rightarrow \upsilon[\alpha] = \{\varphi[\alpha], \psi[\alpha]\}$$

5 *Internal Set Axiom.* If ψ is a sequence of atoms, then $\psi[\alpha]$ is an atom. If $c_\emptyset : n \mapsto \emptyset$ with constant value the empty set, then $c_\emptyset[\alpha] = \emptyset$. If ψ is a sequence of non-empty sets, then

$$\psi[\alpha] = \{\varphi[\alpha] \,|\, \varphi \in \psi(n) \text{ for all } n\}.$$

Benci and Di Nasso formulate these postulates as an axiomatic approach to non-standard analysis. However, here I will only use the results as a 'practical' solution to the infinite lottery problem, without introducing or treating non-standard analysis fully. We immediately notice, however, that α is not in $\mathbb{N}$. It is actually in $\mathbb{N}^*$, the non-standard extension of $\mathbb{N}$. I will define this more carefully below. I remarked above that the sequence $S_A(n)$ is the same as the function $\#(A \cap \{1, \ldots, n\})$. For our probability function, we want to consider the α-limit of the sequence $T(n)$, and take its sum $S(n)$. All sets $T(n)$ are finite for all n (indeed they have all precisely n elements), and the α-value of a sequence of finite sets is called a *hyper-finite set* [5, p.364]. Axiom 5 above says that the elements of $T[\alpha]$ are all the α-values of all sequences with values in $T(n)$

for all n. However, $T(n)$ only contains 0s and 1s at each n, so by the Pair Axiom (number 4), the α-values of such sequences must be 0 or 1. However, because of how the hyper-finite sum, or the sum over elements of $T[\alpha]$ is defined, we need not worry about these α-values of the elements of $T(n)$. We only remark that they are trivially *hyper-natural numbers*, where the set of hyper-naturals is defined as $\mathbb{N}^* = \{\psi[\alpha] \,|\, \psi : \mathbb{N} \to \mathbb{N}\}$. This is the mould for the general definition of the star-transform, a crucial concept in non-standard analysis; I will come back to this below. The last definition we need for current purposes is that of *hyper-finite sums*: if $\varphi[\alpha]$ is a hyper-finite set of hyper-natural numbers, then its hyper-finite sum $\sum_{x\in\varphi[\alpha]} x$ is the α-value of the sequence of finite sums $\sum_{x\in\varphi[n]} x$ [5, p.381].

We are now nearly ready to give the definition of a probability measure obtained thanks to α-theory. First, however, I would like to redefine $T(n)$ and $S(n)$ given above, to make additivity properties more obvious for unions of sets. The probability of a single set will be seen to remain unchanged. As we will see, additivity of hyper-finite sums follows easily from finite additivity. This is why the probability measure I propose is hyper-finitely additive, rather than hyper-countably additive (I define the hyper-countable sum below—intuitively, it is just the non-standard extension of a countable sum). The measure obtained in [39] is hyper-countably additive, but only in a vacuous sense, as they explain. There will be more on this in the discussion at the end of this chapter.

Now, suppose we are interested in the probability value of the union of k sets, A_1 to A_k. Recall that all sets represent (sets of) lottery tickets and are mutually exclusive. We wish to keep using the hyper-finite sum as defined above, and this applied to summing the members of sequences of finite sets. If we first take the characteristic bit-string $T(n)$ of the sets A_1 to A_k, and then consider their union, we would have a union of k hyper-finite sets, and we could not use the definition of hyper-finite sum directly. But this is desirable, since, as mentioned above and as we will see shortly, the additivity of infinite sets (or hyper-finite sets), follows very naturally if we use hyper-finite sums. Therefore, we redefine $T(n)$ as follows, for each n,

$$T(n) = \bigcup_{j=1}^{n} \bigcup_{i=1}^{n} \mathbf{1}_{A_j}(i),$$

with the understanding that for all $n > k$, the characteristic bit-string of A_n will simply be $\{0, 0, \ldots, 0\}$. This is reasonable since we may consider these to be empty sets. The partial sum $S(n)$ of elements of T(n) will be as follows:

$$\begin{aligned} S(n) &= \sum_{i=1}^{n} \mathbf{1}_{A_1}(i) + \cdots + \sum_{i=1}^{n} \mathbf{1}_{A_n}(i) \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} \mathbf{1}_{A_j}(i), \end{aligned}$$

where we could also stop the first summation in the equation above after $n > k$, if we are considering the union of a finite number k of sets, since all elements of the bit-strings for such A_n are just 0. This means that for the case of a single set A, the definition of the partial sum $S(n)$ effectively remains the same. For a union of countable sets, however, there will be a set for each n. Visually it might be helpful to understand this new definition of $S(n)$ as the sum of all the 1s and 0s present in an $n \times n$ array which contains the value of $\mathbf{1}_{A_j}(i)$ at position (j, i). The redefined $T(n)$ can be seen as a two-dimensional 'characteristic bit-array'. We remark that the sequence $T(n)$ is still a sequence of finite sets (each $T(n)$ has n^2 elements) and that, since the sets A_j are disjoint, $S(n)$ can attain a maximum value of α but not more. For the probability function I propose, we take the following steps:

1 we associate each subset of $\mathbb{N}$, or union of such subsets, to its characteristic bit-array given by $\mathbf{1}_{A_j}(i)$, and we further associate this sequence to its value 'at infinity', the unique α-value of $T(n)$ (which we write as $T[\alpha]$);

2 we obtain a probability function P by taking the α-value of the ratio $\frac{S(n)}{n}$, where the numerator is the hyper-finite sum of $T[\alpha]$, and the denominator the α-value of the sequence $c_{\mathbb{N}}(n) = n$, which is just α.

Writing this out as equations we have, for a single set:

$$P(A) = P_\alpha(T[\alpha]) = \frac{S[\alpha]}{\alpha}. \tag{48}$$

Here P_α is simply an intermediate function that maps a set $T(n)$ to the sum of its elements $S(n)$ for all n, so that the α-value of $S(n)$ is the sum of the α-value of the sequence of bit-strings (or bit-arrays) $T(n)$. The probability function P has the following desirable properties

1 $P(\emptyset) = 0$, since $T_\emptyset(n) = 0$ and so $S_\emptyset(n) = 0$ for all n, and so $S_\emptyset[\alpha] = 0$ by Axiom 1;

2 $P(\mathbb{N}) = 1$, since $T_{\mathbb{N}}(n) = n$ and so $S_{\mathbb{N}}(n) = \frac{n}{n} = 1$ and so $S_{\mathbb{N}}[\alpha] = 1$ again by Axiom 1.

3 Finite additivity follows from the definition, the additivity of the counting function # (which is the same as the partial sums $S(n)$) and the properties of α-values of sequences. We write '$|_\alpha$' for 'evaluated at α', or the α-limit of the sequence.

$$\begin{aligned}
P\left(\bigcup_{j=1}^{k} A_j\right) &= \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{A_1}(i) + \cdots + \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{A_k}(i)\right)\Bigg|_\alpha \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{A_1}(i)\right)\Bigg|_\alpha + \cdots + \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{A_k}(i)\right)\Bigg|_\alpha \\
&= \sum_{j=1}^{k} P(A_j),
\end{aligned}$$

since for all sequences of natural numbers, if $\psi(n) = \phi(n) + \varphi(n)$ for all n, then $\psi[\alpha] = \phi[\alpha] + \varphi[\alpha]$ [5, p.362].

Unfortunately, it is not possible to define a countable sum over non-standard numbers [39, p.50], and the α-limits of the sums $S(n)$ will in general be non-standard numbers. Hence we cannot work out the probability of a single ticket (which is non-zero, as we will see shortly) then add this number over the natural numbers. However, this probability function does preserve additivity for infinite unions of disjoint sets: it is **hyper-finitely additive**. This means that the probability of a countable union of disjoint sets, which is just $P\left(\bigcup_{j=1}^{\infty} A_j\right) = P_\alpha\left(\bigcup_{j=1}^{n}\bigcup_{i=1}^{n} \mathbf{1}_{A_j}(i)\Big|_\alpha\right)$, is equal to the hyper-finite sum of the probabilities of each set. We show this now, starting from the latter hyper-finite sum.

$$\sum_{j=1}^{\alpha} P(A_j) = \sum_{j=1}^{\alpha} P_\alpha\left(\bigcup_{i=1}^{n} \mathbf{1}_{A_j}(i)\Bigg|_\alpha\right) = \sum_{j=1}^{\alpha}\left(\frac{1}{\alpha}\sum_{x\in T_{A_j}[\alpha]} x\right). \tag{49}$$

So far we have only written out the definition of $P(A_j)$. Let us relabel these sums for simpler notation. Let $\varphi_j(n) := \frac{1}{n}\sum_{x\in T_{A_j}(n)} x = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{A_j}(i)$, so that $\varphi_j[\alpha] = \frac{1}{\alpha}\sum_{x\in T_{A_j}[\alpha]} x$. Then we can rewrite the last member of equation 49 more simply as follows:

$$\sum_{j=1}^{\alpha} \varphi_j[\alpha] = \sum_{y\in\{\varphi_1[\alpha],\ldots,\varphi_\alpha[\alpha]\}} y. \tag{50}$$

Now, $\{\varphi_1[\alpha],\ldots,\varphi_\alpha[\alpha]\}$ is a hyper-finite set, because it is the α-limit of the sequence of finite sets defined by $\{\varphi_1(n),\ldots,\varphi_n(n)\}$ for each n. Therefore, by the definition of hyper-finite sum, equation 50 is actually the α-limit of the finite sums of members $x \in \{\varphi_1(n),\ldots,\varphi_n(n)\}$. Thus we write:

$$\begin{aligned}
\sum_{y\in\{\varphi_1[\alpha],\ldots,\varphi_\alpha[\alpha]\}} y &= \sum_{y\in\{\varphi_1(n),\ldots,\varphi_n(n)\}} y\Bigg|_\alpha &(51)\\
&= [\varphi_1(n) + \cdots + \varphi_n(n)]|_\alpha &(52)\\
&= \left[\frac{1}{n}\left(\sum_{i=1}^{n} 1_{A_1}(i) + \cdots \sum_{i=1}^{n} 1_{A_n}(i)\right)\right]\Bigg|_\alpha &(53)\\
&= \left[\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} 1_{A_j}(i)\right]\Bigg|_\alpha &(54)\\
&= P_\alpha\left(\bigcup_{j=1}^{n}\bigcup_{i=1}^{n} \mathbf{1}_{A_j}(i)\Bigg|_\alpha\right) &(55)\\
&= P\left(\bigcup_{j=1}^{\infty} A_j\right). &(56)
\end{aligned}$$

Therefore, we have defined a probability function which respects all axioms of finitely additive probability, and on top of that, is hyper-finitely additive. The difference with conventional probability measures, however, is that this P is not real valued, but rather maps all subsets of $\mathbb{N}$ to ratios of numbers in $\mathbb{N}^*$, the non-standard extension of the natural numbers. These numbers are in $\mathbb{Q}^*$, the hyper-rational numbers, which are the non-standard extension of the rational numbers $\mathbb{Q}$. Hence $P : \mathcal{P}(\mathbb{N}) \to \mathbb{Q}^*$. This might be puzzling at first, but, as is argued in [39], just as the real numbers are obtained as the limiting values of sequences of rational numbers, the same is true for the hyper-rational numbers (as we see in the function P), albeit with a different idea of limit, namely the α-value. So there seems no *intrinsic* reason, at least, why a real-valued probability function should be more intuitive than a hyper-rational-valued one. There will be more about this in the discussion at the end of this chapter. In the following paragraphs I present some examples of probabilities of subsets of $\mathbb{N}$. We will need to define a few more concepts from non-standard analysis to interpret these numbers; however, I think it greatly aides intuition to first see some actual cases of probabilities in practice, so I intersperse the definitions between the examples.

6.4 Examples and more definitions.

Singletons and finite sets. We saw above that $P(\mathbb{N}) = 1$ and $P(\emptyset) = 0$. Now take a **singleton** $m \in \mathbb{N}$. Then $P(\{m\}) = P_\alpha\left(\bigcup_{i=1}^{n} \mathbf{1}_{\{m\}}(i)\right) = \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{m\}}(i)\right]\Big|_\alpha$. Consider the sequence of partial sums in the equation, which we called $S(n)$. For all $M > m$, $M \in \mathbb{N}$, we have that $S(M) = S(m)$. By Proposition 1.3 in [5], if a sequence $\varphi(n) = \psi(n)$ for all but finitely many n, then $\varphi[\alpha] = \psi[\alpha]$. But $S(m) = 1$ and so $S[\alpha] = 1$, because $S(n) = 1$ for all but a finite number (precisely $m-1$) of n. Then $P(\{m\}) = \frac{1}{\alpha}$, hence the probability of a single ticket is 1, over the size of the whole set. This is exactly the sort of value one has in mind for a fair lottery. The reasoning is similar for any **finite subset** A of $\mathbb{N}$. Any such subset must have a largest member, say $N \in \mathbb{N}$. Then for all $M > N$, $M \in \mathbb{N}$, we have that $S(M) = S(N) = \#(A) \in \mathbb{N}$. Hence, $S[\alpha] = \#(A)$ and $P(A) = \frac{\#(A)}{\alpha}$. Again, this is exactly as expected. We have already shown the hyper-finite additivity of P, but perhaps it would be nice to check that the probabilities of *all* single tickets add up (or rather 'hyper-finitely add up') to 1. Indeed they do:

$$\sum_{j=1}^{\alpha} \frac{1}{\alpha} = \sum_{x \in \underbrace{\left\{\frac{1}{\alpha}, \ldots, \frac{1}{\alpha}\right\}}_{\alpha}} x = \sum_{x \in \underbrace{\left\{\frac{1}{n}, \ldots, \frac{1}{n}\right\}}_{n}} x\Big|_\alpha$$
$$= 1|_\alpha = 1.$$

Star-transforms, infinitesimals and shadows. Before treating other examples, it will be useful to define a few more concepts from non-standard analysis. The **star-transform** of any entity should be understood as its non-standard extension; in the context of α-theory, which is the one that interests us, the star-transform of any non-empty set is the set of the α-values of all sequences with values in A [5, p.361]:

$$A^* = \{\varphi[\alpha] \mid \varphi : \mathbb{N} \to A\}.$$

By the manner in which we obtain probabilities according to the function P, namely taking α-values of rational numbers, we immediately see that these probabilities are numbers in the set $\mathbb{Q}^*$. As anticipated above, this set is called the set of hyper-rational numbers. Without aiming to be rigorous, we nonetheless note the following properties of the star-transform: it preserves all basic properties of sets, so that for example $A \subseteq B \Rightarrow A^* \subseteq B^*$; and, because of Axiom 3 (the Number Axiom), $A \subset A^*$. That this inclusion is proper is readily understood by taking the natural numbers. By the Number Axiom, the α-value of all constant sequences $c_r(n) = r$, is simply $c_r[\alpha] = r$. Hence all natural numbers are also hyper-natural numbers, so $\mathbb{N} \subseteq \mathbb{N}^*$. However, we also know that the identity sequence $I_{\mathbb{N}}(n) = n$ has $I_{\mathbb{N}}[\alpha] = \alpha \notin \mathbb{N}$, and hence actually $\mathbb{N} \subset \mathbb{N}^*$ [5, p.361-363]. We now see that $\frac{1}{\alpha}$, seen above as the probability for a singleton, is a hyper-rational number. In this context we may want to know the following: 'how big' a number such as $\frac{1}{\alpha}$ is; and how, if at all, we could associate this number with a standard (say real, or rational) number, in order to translate this solution back to a standard infinite lottery. For the first question we introduce **infinitesimal numbers**. A number $\xi \in \mathbb{R}^*$ is infinitesimal if $-r < \xi < r$ for all $r \in \mathbb{R}$. Note that this definition applies to hyper-rational numbers too, as $\mathbb{Q} \subset \mathbb{R} \Rightarrow \mathbb{Q}^* \subset \mathbb{R}^*$. Note also that the only infinitesimal number in $\mathbb{R}$ is 0. The number $\frac{1}{\alpha} \in \mathbb{Q}^*$, however, is indeed infinitesimal. The relation $<$ considered in the definition of infinitesimals is actually the star-transform of the relation $<$ in $\mathbb{R}$, so we should actually write it as $<^*$. A possible definition of $<^*$ is the following: for $\zeta, \xi \in \mathbb{R}^*$, $\zeta <^* \xi \Leftrightarrow \xi - \zeta \in (\mathbb{R}_+)^*$, with $\mathbb{R}_+$ the set of positive real numbers. However we follow [5] in writing this relation simply as $<$, and we see that $\frac{1}{\alpha}$ is indeed infinitesimal by using the following sufficient condition given in that paper [5, p.363]: if φ and ψ are real sequences, with $\varphi(n) < \psi(n)$ eventually, then $\varphi[\alpha] < \psi[\alpha]$. Hence, because $\frac{1}{n}$ is eventually less than any real number, $\frac{1}{\alpha}$ is indeed infinitesimal. Similarly to the definition of infinitesimals, we have *finite* hyper-real numbers, which are those $\xi \in \mathbb{R}^*$ such that $-r < \xi < r$ for some positive real number r; and *infinite* hyper-real numbers for which no such r exists. These definitions obviously carry over to hyper-rational and hyper-natural numbers too. Other infinitesimal numbers are the following: (i) if ζ and ξ are infinitesimal, then $\zeta \cdot \xi$ and $\zeta + \xi$ are infinitesimal; (ii) if ζ is finite and ξ is infinite, then $\frac{\zeta}{\xi}$ is infinitesimal [5, p.365]. The last important definition we need in this section is that of **shadow** of a hyper-real number. First this important theorem: every finite hyper-real number ξ is infinitely close to a unique real number r; hence it can be written uniquely as the unique sum of a real number and an infinitesimal number: $\xi = r + \varepsilon$ [5, p.366]. The real number r in this representation is called the *shadow*. Now we can answer the second point above, namely how we can associate our non-standard valued probabilities with real or rational numbers. The answer now is very simple: we take the shadow of the obtained result. We see, then, that the shadow of $\frac{1}{\alpha}$ is just 0. This indicates that we are not able to simply 'translate back' the values obtained in the non-standard-valued probability, to something more recognisable but that will maintain its additivity properties. We can now examine other examples, and see what their shadow would be.

Odd and even numbers. Because the probability function is the α-value of the sequence of partial sums $\frac{S(n)}{n}$, if we can write a formula for the partial sums in terms of n, then we easily have its α-value. Take the set of all odd integers as an example (let us call it ODD). If n is even, exactly half of the integers in $\{1,\ldots,n\}$ are odd. Hence $\frac{S(n)}{n} = \frac{\frac{n}{2}}{n} = \frac{1}{2}$. If n is odd, the number of odds in $\{1,\ldots,n\}$ is the same as the number of odds in $\{1,\ldots,n+1\}$, so $\frac{S(n)}{n} = \frac{\frac{n+1}{2}}{n} = \frac{n+1}{2n}$. Using the function

$$\frac{1-(-1)^k}{2} = \begin{cases} 1 & \text{if } k \text{ is odd} \\ 0 & \text{if } k \text{ is even} \end{cases}$$

we can write this for all n as:

$$S_{ODD}(n) = \frac{1}{2} + \frac{1-(-1)^n}{4n},$$

so

$$S_{ODD}(n)\big|_\alpha = \frac{1}{2} + \frac{1-(-1)^\alpha}{4\alpha}.$$

This highlights what could seem the biggest drawback of this approach: *the α-limit always exists, but it is not unique.* We see that:

$$P(ODD) = \begin{cases} \frac{1}{2} & \text{if } \alpha \text{ is even} \\ \\ \frac{1}{2} + \frac{1}{2\alpha} & \text{if } \alpha \text{ is odd} \end{cases}$$

Benci and Di Nasso remark that we can consistently postulate these kind of properties about α [5, p.367]. We could simply add to the axioms describing α, an additional one deciding whether it is even or odd. A similar situation, of course, arises for the **set of even integers** ($EVEN$). We have

$$P(EVEN) = \begin{cases} \frac{1}{2} & \text{if } \alpha \text{ is even} \\ \\ \frac{1}{2} - \frac{1}{2\alpha} & \text{if } \alpha \text{ is odd} \end{cases}$$

A first answer could be to simply postulate that α should be even. However, as [39, p.54-55] point out, the shadow is the same in either case, namely $\frac{1}{2}$, so in this case we could argue that it makes little difference—or infinitesimal difference. A condition on α being even or odd is not the only one we could want, however, as we will see in the next example.

Multiples of k. Let A_k be the set of all $m \in \mathbb{N}$ such that $m = 0 \mod k$, or all the **multiples of** k. Then the partial sums $S_{A_k}(n)$ follow the formula

$$S_{A_k}(n) = \frac{\lfloor \frac{n}{k} \rfloor}{n}.$$

Therefore,

$$P(A_k) = \frac{\lfloor \frac{\alpha}{k} \rfloor}{\alpha} = \begin{cases} \frac{1}{k} & \text{if } \alpha = 0 \mod k \\ \\ \frac{1}{k} - \frac{r}{\alpha k} & \text{if } \alpha = r \mod k \end{cases}$$

Just as we could postulate that α is even, by the same reasoning we should perhaps say that α is a multiple of all prime numbers, so that we always get a 'nice' result in cases such as these. This would represent an infinite number of constraints on α. Again, however, we could simply ignore the different cases and call them equivalent, since the shadow of both solutions, for α a multiple of k or not, is simply $\frac{1}{k}$, because $r < k$ and k is finite, so $\frac{r}{\alpha k}$ is infinitesimal. So far, so good, then: the α-limit always exists, and it is not unique but this does not really represent any serious issue, since the solutions so far all have the same value in real numbers. The next example, however, shows that while the probability function P is in many ways a good solution, it is not 'miraculous'.

Even-numbered binary expansions. We can now go back to example 6.2, where we found a formula for the probability of the set of numbers which have an even number of digits in their binary notation (we call this set A). Recall that we had (using that the functions $\#(A \cap \{1, \dots, n\})$ and $S(n)$ are equivalent):

$$S(n) = \frac{2^{\lfloor \log_2(n) \rfloor + 1} - 2}{3n} \quad \text{if } \lfloor \log_2(n) \rfloor \text{ is even}$$

and

$$S(n) = 1 - \frac{2^{\lfloor \log_2(n) \rfloor + 1} - 1}{3n} \quad \text{if } \lfloor \log_2(n) \rfloor \text{ is odd.}$$

We saw that the conventional limit did not exist. Can the function P, obtain through the α-calculus, do better? Firstly, the α-value, like the limit superior and the limit inferior, always exists, so this particular issue is solved. We do not, however, get a unique value for an α-limit. The α-limit is as follows:

$$P(A) = \begin{cases} \frac{2^{\lfloor \log_2(\alpha) \rfloor + 1} - 2}{3\alpha} & \text{if } \lfloor \log_2(\alpha) \rfloor \text{ is even} \\ \\ 1 - \frac{2^{\lfloor \log_2(\alpha) \rfloor + 1} - 1}{3\alpha} & \text{if } \lfloor \log_2(\alpha) \rfloor \text{ is odd} \end{cases}$$

This constitutes a range of hyper-rational numbers, and it is not possible to write this as a real number plus an infinitesimal, unless we make further assumptions on α. However, assumptions such as the ones above will not necessarily help. Assuming α is a multiple of 2, for example, does not narrow down a value for the probability of A. Even numbers obviously have values of $\lfloor \log_2(n) \rfloor$ all across the range. We could say, then, that α could be an integer power of 2, but even this gives different results: $P(A)$ is either $\frac{2}{3} - \frac{2}{3\alpha}$ (if $\log_2(\alpha)$ is even), or $\frac{1}{3} + \frac{1}{3n}$ (if $\log_2(\alpha)$ is odd). These have shadows of $\frac{2}{3}$ and $\frac{1}{3}$ respectively. Arbitrarily, we could say that $\log_2(\alpha)$ is even, or odd, and get equivalently valid, but different solutions. The upper and lower bounds on $P(A)$, as worked out in the first discussion of this example, become respectively $\frac{2}{3}$ and $\frac{1}{3} - \frac{1}{3\alpha}$, the latter being $\frac{1}{3}$ plus an infinitesimal. Hence using the function P we can still talk of intervals for the probability function, making use of the information we have about this. However, this example shows that P is not unique in an important way, and not only in the sense above that different solutions only have an infinitesimal difference between them. This is in disagreement with what is claimed in [39, p.54]. I believe, however, that the construction given in that paper suffers from the exact same problem we have just seen, since it is essentially the same as what I have described thus far. I outline such construction below. As a last example, we go back to Humean induction applied to the infinite lottery, as done in section 2.5.

Humean induction and infinitesimal degrees of belief. Recall that, in this example, we had an agent picking numbered balls from an infinite urn, and she did not know whether she was in one of these two situations: either all balls are black, or there is one white ball in the urn. As she picks balls from the urn, she writes 1 for black ones and 0 for the white one. We call the two hypotheses

- $B :=$ all balls are black,
- $W :=$ there is one white ball,

and we are interested in computing how the degrees of belief of the agent change in time as she picks ball after ball from the urn. We had noted that the problem of assigning a probability to the event 'ball number n is white, given that there exists a white ball', which is needed for the Bayesian updating of the agent's probabilities, is precisely an infinite lottery as the ones considered thus far. Hence, given the results of this chapter, we assign such events the following probabilities: $P(0|W) = P(10|W) = \cdots = P\left(1^{(n)}0|W\right) = \cdots = \frac{1}{\alpha}$. Then we have the following:

$$P\left(B|1^{(n)}\right) = \frac{\alpha}{2\alpha - n},$$

and

$$P\left(W|1^{(n)}\right) = \frac{\alpha}{2\alpha - n} - \frac{n}{2\alpha - n}.$$

Hence, for all finite n, $P\left(W|1^{(n)}\right)$ is only infinitesimally smaller than $P\left(B|1^{(n)}\right)$. If we take the α-limit of these expressions, however, we obtain that B has probability 0, and W probability 1, as they should. We also get the correct results in the case of a the white ball being picked, $P\left(W|1^{(m)}0\right) = 1$ and $P\left(B|1^{(m)}0\right) = 0$. Moreover, writing out the following,

$$\begin{aligned} P\left(W|1^{(n+1)}\right) &= \frac{\alpha - n}{2\alpha - n} - \frac{\alpha}{(2\alpha - n)(2\alpha - n - 1)} \\ &= P\left(W|1^{(n)}\right) - \frac{1}{4(\alpha - n) - 2 + \frac{n^2+n}{\alpha}} \end{aligned}$$

we see that $P\left(W|1^{(n+1)}\right) = P\left(W|1^{(n)}\right) - \varepsilon$, where ε is an infinitesimal number for all finite n. So $P\left(W|1^{(n)}\right)$ decreases only an infinitesimal amount each time another black ball is picked. All these observations point to the fact that this seems a reasonable solution.

6.5 Another construction for P and hyper-countable additivity.

In the construction I presented above I essentially followed the one given in [39]. The main difference is that I omitted to use ultrafilters, and I sought to show directly that the probability function obtained is hyper-finitely additive, rather than hyper-countably additive. I now describe the probability measure given in that paper, after having introduced two more concepts which are needed. We defined $\mathbb{N}^*$ above; a *hyper-countable sum* is simply a sum over all members of this set. We also defined the star-transform of a set. We now give the definition of the *star-transform of a function*, taken from [5]: let $f : A \to B$ be a function.

Then its star-transform $f^* : A^* \to B^*$ is a function such that, for every sequence $\varphi : \mathbb{N} \to A$,

$$f^*(\varphi[\alpha]) = (f \circ \varphi)[\alpha]. \tag{57}$$

We are now ready to present P_{num}, the probability measure given in [39]. Although in that paper ultrafilters are used, as remarked above, we do not need to introduce them to present the additivity proof or for a comparison with my construction described in the sections above. First we define the *numerosity* of a set A by:

$$num(A) = \#^*(A^* \cap \{1, \dots, \alpha\}). \tag{58}$$

Then the probability measure in [39, p.49] is just:

$$P_{num} : \mathcal{P}(\mathbb{N}) \to [0,1]^*_{\mathbb{Q}^*} \tag{59}$$

$$A \mapsto \frac{num(A)}{\alpha}. \tag{60}$$

Before presenting the additivity proof, some remarks: I asserted above that the star-transform preserves all the basic properties of sets (except the power set [5, p.361]); this property is called 'Transfer'. Another property used in the proof is the following:

$$\left(\bigcup_{n\in\mathbb{N}} A_n\right)^* = \bigcup_{N\in\mathbb{N}^*} A^*_N \tag{61}$$

The proof in [39, p.52] regards the function *num* only. The aim is to show that the *num* of a countable sequence of disjoint sets, is equal to the hyper-countable sum of the individual *num* of each set.

Proof of hyper-countable additivity of *num* ([39, p.52]).

$$\begin{aligned} num\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \#^*\left(\left(\bigcup_{n\in\mathbb{N}} A_n\right)^* \cap \{1,\dots,\alpha\}\right) \\ &= \#^*\left(\bigcup_{N\in\mathbb{N}^*} A^*_N \cap \{1,\dots,\alpha\}\right) \quad \text{by equation 61} \\ &= \#^*\left(\bigcup_{N\in\mathbb{N}^*} (A^*_N \cap \{1,\dots,\alpha\})\right) \quad \text{by properties of union and intersection and Transfer} \\ &= \sum_{N\in\mathbb{N}^*} \#^*\left(A^*_N \cap \{1,\dots,\alpha\}\right) \quad \text{by CA of } \# \text{ and Transfer} \end{aligned}$$

I now follow the passages explained in [39] in order to make the last line explicitly about A_n. The *num* of a sequence of sets is defined as

$$num(\langle A_n \rangle) = \#^*(\langle A_n \rangle \cap \{1, \dots, \alpha\}), \tag{62}$$

where the star-map of a sequence in $\mathbb{N}$ is defined as:

$$(\langle A_n \rangle_{n\in\mathbb{N}})^* = \langle \langle A^*_N \rangle \rangle_{N\in\mathbb{N}^*}. \tag{63}$$

The double angled parentheses are used to indicate hyper-sequences, or sequences over the hyper-natural numbers. We will need to intersect this sequence with $\{1, \ldots, \alpha\}$, and apply the star-transform of the counting function, $\#^*$, to such intersection. These operations are defined component-wise: $\langle\langle A^*_N \rangle\rangle \cap S = \langle\langle A^*_N \cap S \rangle\rangle$ for any $S \in (\mathcal{P}(\mathbb{N}))^*$; and we define: $\#^*(\langle\langle A^*_N \rangle\rangle) = \langle\langle \#^*(A^*_N) \rangle\rangle$. Finally, using this we can write, for the *num* of a sequence of disjoint sets:

$$num(\langle A_n \rangle) = \langle\langle \#^*(A^*_N \cap \{1, \ldots, \alpha\}) \rangle\rangle. \tag{64}$$

This is a hyper-sequence of hyper-natural numbers. If we call the N^{th} element of this hyper-sequence $(num(\langle\langle A_n \rangle\rangle))_N$, we can conclude the additivity proof above by rewriting the last line:

$$num\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{N\in\mathbb{N}^*} (num(\langle\langle A_n \rangle\rangle))_N. \tag{65}$$

Differences between P and P_{num}. I intend to show that my approach is entirely equivalent to the one in [39] I just described, which of course guided my own. Having shown this, however, it must follow that the more serious non-uniqueness which affects my way to the solution, must also affect P_{num}. The two obvious differences between the approaches, which I believe are only superficial, are the following: (1) for the numerator of P_{num} we take $num(A) = \#^*(A^* \cap \{1, \ldots, \alpha\})$, while in P we take the α-limit of the sequence $S(n)$, which is just the α-limit of the *sequence* $S(n) = \#(A \cap \{1, \ldots, n\})$; (2) P_{num} is hyper-countably additive, while P is only hyper-finitely additive. I will treat these two matters in this order, followed by some remarks on the non-uniqueness of P and P_{num}.

α-limit versus star-transform. I wish to show that $\#^*(A^* \cap \{1, \ldots, \alpha\}) = (\#(A \cap \{1, \ldots, n\}))[\alpha]$, which is to say, $num(A) = S_A[\alpha]$. Recall the definition of A^* for any set A: $A^* = \{\varphi[\alpha] | \varphi : \mathbb{N} \to A\}$. Hence $A^* \cap \{1, \ldots, \alpha\}$ is the set of all α-limits of all sequences which have all values in A, and have α-limits less than or equal to α. We also have, by definition of α-limit of a set, that:

$$(A \cap \{1, \ldots, n\})[\alpha] = \{\varphi[\alpha] | \varphi(n) \in (A \cap \{1, \ldots, n\}) \text{ for all } n\}.$$

Hence clearly $(A \cap \{1, \ldots, n\})[\alpha] \subset (A^* \cap \{1, \ldots, \alpha\})$, because $(A \cap \{1, \ldots, n\})[\alpha] \subset A^*$ and we know that all sequences $A \cap \{1, \ldots, n\}$ have α-limits less than or equal to α. On the other hand, take $\psi : \mathbb{N} \to A$ to be a sequence with $\psi[\alpha] \in A^* \cap \{1, \ldots, \alpha\}$. This sequence may or may not have the property $\psi(n) \in A \cap \{1, \ldots, n\}$ for all n. If it does, then $\psi[\alpha] \in (A \cap \{1, \ldots, n\})[\alpha]$. Suppose the sequence $\psi(n)$ does not have this property. It must still take on values in $A \cap \{1, \ldots, m_n\}$, for some finite m_n for all $\psi(n)$. In particular, it must have a lowest value $\psi(L) \in A \cap \{1, \ldots, m_L\}$ for some finite m_L. Hence, for all $n \geq m_L$, we indeed have that $\psi(n) \in A \cap \{1, \ldots, n\}$. Therefore, any sequence $\psi : \mathbb{N} \to A$ with $\psi[\alpha] \in A^* \cap \{1, \ldots, \alpha\}$, must be equal, except for in a finite number of positions, to a sequence $\psi' : \mathbb{N} \to A$ with $\psi'(n) \in A \cap \{1, \ldots, n\}$ for all n. And therefore any such sequence has $\psi[\alpha] = \psi'[\alpha] \in (A \cap \{1, \ldots, n\})[\alpha]$. Therefore $(A^* \cap \{1, \ldots, \alpha\}) \subset (A \cap \{1, \ldots, n\})[\alpha]$, and so we can conclude that

$(A^* \cap \{1, \ldots, \alpha\}) = (A \cap \{1, \ldots, n\})[\alpha]$.[13] Now, using the definition of the star-transform of a function, we have the following:

$$\begin{aligned} S[\alpha] &= (\#(A \cap \{1, \ldots, n\}))[\alpha] \\ &= \#^*((A \cap \{1, \ldots, n\})[\alpha]) \quad \text{by definition 57} \\ &= \#^*(A^* \cap \{1, \ldots, \alpha\}) \quad \text{by steps above} \\ &= num(A). \end{aligned}$$

Therefore, P and P_{num} are completely equivalent. Why, then, is the latter hyper-countably additive, while the former merely hyper-finitely additive? I treat this next.

Additivity of P and P_{num}. Considering what we discussed above, the different additivity of the two functions can only be on a formal level. And indeed, it is best explained as follows:

> [T]he lottery on ℕ is HCA [hyper-countably additive] in a very specific sense: [...] there can always be found a hyper-natural number, $K \in \mathbb{N}^*$ (α in the example [of the union of all single tickets], but possibly larger in other cases), such that the hyper-countable sum decomposes in a hyper-finite sum and a hyper-countable tail with zero-terms only. Thus we may call *num* and P_{num} hyper-finitely additive (HFA) [39, p.52].

Hence we could say that the hyper-countable additivity of P_{num} is only a result of how the proof was approached; P_{num} is in fact only vacuously hyper-countably additive, much in the same way as a finite lottery is 'countably additive' [39, p.52]. Even the assertion that we could sum up to a hyper-natural number $K \in \mathbb{N}^*$ which is larger than α, could be qualified, in my opinion: in fact, we would need exactly enough members of the sum to be 0, in order for *num* to be less than or equal to α. Hence we would be, in truth, adding a maximum of α members.

Non-uniqueness of P and P_{num}. Wenmackers and Horsten, in [39, p.54], address the worries that the non-standard solution to the infinite lottery is not unique, as follows: "[I]n the present context the accusation of arbitrariness boils down to the choice of a free ultrafilter $\mathcal{U}$. A different choice of free ultrafilter produces a different value of α and hence a probability function with the same standard part but infinitesimal differences". (Note that the standard part is what I called the shadow of a hyper-rational number.) We need not worry about ultrafilters here (I quote the passage in full only for completeness), but only about different values of α. In that article, the results obtained for the probability of the sets of even and odd numbers, and of multiples of a number, are the same as the ones I worked out with the function P. This is not surprising, since the two functions P and P_{num} are the same, as seen above. No other examples are treated in [39]. Now, even without considering ultrafilters, we note that in [39] it is explicitly allowed to consider α even. In my example 6.4 above,

[13] Note that this is a special case and I am not claiming that $E[\alpha] = E^*$ for general sequences of sets E_n. A counterexample is $E_1 = \{1, 2\}$, $E_2 = E_3 = \cdots = \{1\}$, where the sequence $\langle 2, 2, \ldots \rangle$ has α-limit $2 \in E_1^*$ but $2 \notin E_\alpha$.

however, we saw that even postulating that α be even, we still have a range of results whose shadow differ in more than just infinitesimals. In particular, we had that the shadow of $P(A)$ was either $\frac{2}{3}$ or $\frac{1}{3}$ according to whether α was an even or odd power of 2. This is in direct contradiction with the claim by Wenmackers and Horsten quoted above. I presently do not see how, if they were to calculate the probability in this example, they could obtain a different result from mine. Especially, that is, given that the functions P and P_{num} appear to be entirely equivalent. If my workings are correct, and if my understanding of the passage above is sound, then there is a considerably more substantial degree of non-uniqueness than the authors claim, in the non-standard solution.

6.6 Remarks

A solution to a different problem. In [39], there is an interesting remark about infinite lotteries defined on the hyper-natural numbers, rather than defined on the natural numbers *then* assigned hyper-rational probability values; the remark is simply that this approach "does not solve the original problem. Instead, it is a solution to a different problem" [39, p.56]. It would seem legitimate to make the same remark about the solutions presented in this chapter, in the context of my thesis. We wanted an answer to the problem that we are forced to choose between fairness and additivity in infinite lotteries, whereas this choice does not need to be made in any finite case. We came up with a solution that changes another usual axiom of probability, namely that probabilities are real-valued; and yet we *still* did not achieve countable additivity. Also, we saw that we cannot work in the non-standard universe, and then translate back our results into something more usual: if we use infinitesimal numbers, we must add them either in finite sums, or in hyper-finite sums; most disappointing, perhaps (although entirely to be expected), is that when we do this translation back to real values, we find that the probability of each ticket simply becomes 0. Nonetheless, the discussion in the previous chapters showed that there simply is no solution that will include both countable additivity and absolute fairness in an infinite lottery. From this point of view, then, this constitutes a viable solution, a way around the deadlock, if not through it.

Intuitions and the non-standard solution. Wenmackers and Horsten [39] specifically set out to seek a solution for a infinite lottery which will respect some intuitions we have with regards to lotteries in general. I quote them in full, with the names given in [39, p.40]:

- **Fair**. The lottery is fair.
- **All**. Every ticket has a probability of winning.
- **Sum**. The probability of a combination of tickets can be found by summing the individual probabilities.
- **Label**. The labelling of the tickets is neutral with respect to the outcome.

The solution above respects **Fair** (all tickets have a $\frac{1}{\alpha}$ chance of winning), **All** and, the authors in [39, p.54] claim, **Sum**. It has to be pointed out, however, that the non-standard solutions described above only respect **Sum** if we redefine it to mean 'hyper-finite sum' or 'hyper-countable sum'. As the preceding chapters amply demonstrate, it is often quite tricky to talk about intuitions: contrasting arguments can seem equally well justified by different, but maybe equally deep-seated, intuitions we may have. But I can surely say this: *to me* hyper-finite sums, and hyper-countable sums even more so, are not very intuitive. However, perhaps we should not resign to thinking of the non-standard solution as something counter-intuitive and abstruse. Benci and Di Nasso make the appealing case that some concepts in calculus are actually much more natural if we are allowed to operate with infinitesimals and infinite numbers, rather than banish this concept, as is done in standard analysis:

> The view of many working mathematicians in non-standard analysis is that infinitesimal numbers actually do exist and that the notion of a limit is just an awkward way to indirectly talk about infinitesimals without explicitly mentioning them [5, p.378].

They further remark that, given a simple axiomatic approach like their α-theory, we would have no need to translate over all the standard concepts of analysis, to re-write them in non-standard terms. We could learn what is currently called non-standard theory, just as we learnt standard calculus in school and university. Hence, in the context of the infinite lottery, we can assert that there is no inherent reason why the non-standard solution should be less intuitive than a solution using standard calculus only. There is even a sense, however, in which this argument is too strong; for which intuitions are inherently *better* than others? And how much are intuitions based on and conditioned by our existing knowledge? These seem to be difficult questions. If we can axiomatically replace any intuition we have with another set of rules, and claim that there is no inherent difference between the newly obtained approach and the previous intuitive one, then, in a sense, *anything goes*. We wanted real-valued degrees of belief—we ended up with hyper-rational probability values, and the deeper problem of why we should favour one set of axioms over another, equally powerful, one; we wanted countable additivity—we ended up defining a new way of summing altogether. And this brings us back to the original doubt, namely that we might have described the solution to a slightly different problem to the one we started out with. However, since the original problem seems unsolvable without giving up on some intuition or another, the non-standard solution is definitely one way to approach the issue, and a valuable and interesting one at that.

7 Conclusions

It is time to recollect all the threads we discussed in the remarks which concluded each chapter, and assess what we can conclude from this study. The main theme that emerged was certainly that of the two contrasting intuitions which I outlined in the introduction: on the one hand, we would want it to be possible to have a uniform distribution of probabilities even in infinite cases, especially if we see no reason not to; on the other hand, it would also seem reasonable that our probability for a union of incompatible events be equal to the sum of each single probability, as it is in the finite case. That these two intuitions are incompatible when adopting real-valued probabilities is clear; what I tried to show in the present work is that most available positions in the debate ultimately rely upon one of these two intuitions. This points to the fact that it is not possible to conclude the debate in a way which is satisfactory to both sides, unless we abandon the sphere of the real numbers and standard analysis. As I remarked in Chapter 6, doing so forces us to abandon yet other intuitions about numbers and about summation. As I also remarked there, with Benci and Di Nasso, our intuition is influenced by what we happen to know already, and as such it can be educated.

De Finetti puts this matter, of intuitions and defining probability, very eloquently in the passage below, written in 1930 as part of his academic correspondence with the probabilist Frechét (who took probability as a countably additive measure). I quote it in full because, strikingly, notwithstanding the fact that it was written so long ago, it brings up clearly many of those issues which are still hotly debated today:

> Every concept, mathematical ones included, is more or less directly and clearly suggested by intuition: however, its definition is totally arbitrary, as long as the consequences that we wish to draw from it are purely formal: as long as, that is, they are propositions in which that concept acts in the convened sense which it assumed by its definition. This is the case of measure; we would have a different case, on the other hand, for weight, because we cannot impose to the scale to work according to our definition; in the same way, it seems to me that probability too is a different case. In the case of probabilities the trouble is that when, *based on a convention*, we conclude, for example, that the probability of an infinite sum of incompatible events with null probability has null probability, we intuitively think that this sum is an *almost impossible* event, while the definition only allows us to conclude, completely rigorously, that what has a value equal to zero is that numerical function which we have, by convention, called "probability"[11, pp.4-5] [emphasis in the original; my translation].

The points raised here can guide us in a number of comments. Firstly, Bingham, for example, agrees that calling the measure we defined in Section 2.2 'probability' is an artefact, or a double use of the same word: in one sense it is a word of natural language, and in another sense it is that measure we defined rigorously. And by giving a formal, conventional definition, we have indeed distanced ourselves from its real-world meaning (whatever that is)[7, p.22]. We can read Kelly's warning in the same light: we should be fully aware that theorems that seem to guarantee convergence to the truth can only be viewed as formal results of what we, by pure convention, called probability. Kelly argues that we cannot justify countable additivity from its consequences (e.g. adopt it because it allows Lebesgue integration), and then interpret these very consequences for an epistemological advantage (the convergence to the truth results). I think this point must be conceded, and, whether thanks to Kelly or not, it largely has been. So far so good, then: it seems that measure-theoretic probability is formally powerful, but unfortunately distant from the real-world, natural-language significance of probability.[14] But the following problem arises: we would need to prove that *finitely* additive probability is closer to the real-world thing, as de Finetti believes; but this will be difficult, because as seen on different occasions above, we *also* have intuitions that tell us that the real-world probability should be countably additive. Perhaps Cox's derivation could persuades us that probability, if we see it as a kind of extension of logic that deals with uncertain reasoning, really is only finitely additive. But, as I argued above, this is quite problematic; we would need to be convinced of the qualitative axioms' more fundamental status, and this is not at all obvious. And on the other hand, for example, assuming countable additivity we can prove the Strong Law of Large Numbers, which says the following: if we have a sequence of independent and identically distributed random variables, then the average of the first n random variables, converges almost surely to the common expectation of the random variables. Almost sure convergence means that this convergence occurs for all sequences, except for on a set of probability 0. What this law guarantees in practice is that the long run relative frequency must, with very high probability, converge to the 'real' probability of the event. This constitutes a vindication and explanation of a very important intuition about probability, which is of course encountered in practice: namely that the past relative frequency of an experiment, if performed a large number of times, gives us a very good indication of its probability of success in future repetitions. Bingham, citing an earlier paper by himself, remarks that this fact "demonstrates convincingly that the Kolmogorov axiomatics of the Grundbegriffe [which is [33]] have captured the essence of probability" [7, p.22].

De Finetti rejected axiomatic definitions of probability; as we saw above, he thought that only an operational definition, in terms of risk, decisions and losses, could have a meaning which is not merely "an appearance of such in a metaphysical-verbalistic sense" [16, p.76]. As suggested by Howson, a reading of how de Finetti actually used Dutch book arguments, shows their purpose was precisely this grounding in meaning, rather than the 'proving' of finite, or countable, additivity. Hence, Dutch book arguments do not really tell us anything new about additivity either way. By setting himself the task of describing formally what probability *really is*, de Finetti embarked on something very difficult, which the vast majority of mathematicians simply have no interest in, or avoid (see, for example, the comments by Doob in footnote 14, but also the remarks by Bingham in that paper, or also Littlewood's chapter on *The Dilemma of Probability Theory* in his popular [35]). His theory of subjective

[14]Note that this conclusion would most likely be serenely accepted by most mathematicians: Bingham reports this emblematic quote by Doob, a mathematician who did much seminal work in probability: "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" [7, p.11].

probability went on to play vital roles in decision theory, game theory and laid the foundations for Subjective Bayesianism in the philosophy of science (see [37, pp.275-276]). Evaluating his position on countable additivity is more difficult. Von Plato, for example, says that "de Finetti remained faithful to the principles of his foundational program and his radically empiricist philosophy, such as the requirement that unverifiable infinitary events cannot in general have well-determined probability values" [37, p.244]. But in [14, p.6] de Finetti does not shy away from having his agent choose from an infinite (indeed, uncountable) number of admissible choices, in order to model a decision problem. Also, he was not *opposed* to countable additivity as such, as we remarked in Section 2.2. Rather, he was in favour of considering the study of countably additive probabilities as a special case, much like continuous functions are studied in analysis. Hence there is a sense in which de Finetti's is a purely 'verbalistic' dispute, in that it could be solved by adding a few lines to all probability textbooks to the effect that, while the concept of probability is broader, we nearly always study countably additive probabilities.

To sum up my position, I would say the following. I agree with Howson when he argues that Dutch book arguments have nothing to say about finite versus countable additivity; I disagree with him when he goes on to claim that arguing from qualitative axioms puts us on more solid ground. Indeed, it is not clear what such argument really adds to a simple axiomatic stipulation of the properties of probability functions. With regards to these properties, and the question of their additivity, there is nothing wrong, in principle, with de Finetti's proposal: we could consider countably additive measures as a special case of probability, albeit by far the most important one. This would probably fail to raise mathematical interest in the matter, especially given that, by most accounts, the finitely additive theory is much less accessible. On the other hand, it would allow us to model, in a countably infinite setting, philosophically interesting concepts such as uniformity of preference and, closely linked, Humean scepticism towards induction. From a philosophical point of view, it does seem arbitrary to exclude such cases from being modelled, especially if such exclusion comes by arguing from the perceived obviousness of countable additivity. I found Jaynes' comments on the principle a good example of this sort of argument—and Kelly's arguments should serve as a warning against the easy application of convergence theorems to epistemology that could follow (although we note that Jaynes does not make this step). Of course, in the countably infinite setting, finitely additive probability has counter-intuitive aspects: we can believe each single event to be impossible and their infinite union to certain; but so does countable additivity: we must assign nearly all probability to some finite subset of events, possibly against our better judgement. We can finally address this seemingly simple question for subjective probability: should an agent's beliefs be countably additive? But the answer will be terribly disappointing: *it depends*—on whether we consider uniformity and scepticism important, or whether we deem additivity to be more natural. Foundational arguments fail to settle the question, and because of the mathematical features of the problem, and the counter-intuitive consequences of either solution, cannot possibly do so.

References

[1] Paul Bartha, *Countable Additivity and the de Finetti Lottery*, British Journal for the Philosophy of Science, Volume: 55, Issue: 2 (June 1, 2004), pp: 301-324

[2] Paul Bartha, *The Shooting-Room Paradox and Conditionalizing on Measurably Challenged Sets*, Synthese, Vol. 118, No. 3 (1999), pp. 403-437;

[3] Alexandr A. Borovkov, *Probability Theory*, Springer-Verlag, London: 2013;

[4] Vieri Benci and Mauro Di Nasso, *Numerosities of Labelled Sets: A New Way of Counting*, Advances in Mathematics 173 (2003) 50-67;

[5] Vieri Benci and Mauro Di Nasso, *ALpha-Theory: An Elementary Axiomatics for Nonstandard Analysis*, Expositiones Mathematicae 21 (2003): 355-386;

[6] J. Cornfield, *Bayes Theorem*, Review of the International Statistical Institute, Vol. 35, No. 1 (1967), pp. 34-49

[7] Nicholas Bingham, *Finite additivity versus countable additivity*, Electronic Journal for History of Probability and Statistics, Vol.6, n1, June 2010;

[8] Robert Chen, *On Almost Sure Convergence in a Finitely Additive Setting*, Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, 37, 341-356 (1977);

[9] Donald L. Cohn, *Measure Theory*, Second edition, Birkhäuser Advanced Texts, Springer: 2013;

[10] , Richard T. Cox, *The Algebra of Probable Inference*, Baltimore, The John Hopkins Press: 1961;

[11] Bruno de Finetti, *A proposito dell'estensione del teorema delle probabilita totali alle classi numerabili*, Rendiconti del R. Istituto di Scienze e Lettere, 1930, vol.LXIII, fasc.11-14, pp.1-5

[12] Bruno de Finetti, *Foresight: Its Logical Flaws, Its subjective Sources*, pp.93-158 in *Studies in Subjective Probability*, Henry E. Kyburg Jr. & Howard E. Smokler (eds.), John Wiley & Sons: 1964

[13] Bruno de Finetti, *Sull'impostazione assiomatica del calcolo delle probabilità*, 1942, 1947 Box 5, Folder 26, Bruno de Finetti Papers, 1924-2000, ASP.1992.01, Archives of Scientific Philosophy, Special Collections Department, University of Pittsburgh

[14] Bruno de Finetti, *Probability, Induction and Statistics: The Art of Guessing*, John Wiley & Sons: 1972

[15] Bruno de Finetti, *Bayesianism: Its Unifying Role for Both the Foundations and Applications of Statistics*, International Statistical Review, Vol. 42, No. 2(Aug., 1974), pp. 117-130

[16] Bruno de Finetti, *Theory of Probability: A Critical Introductory Treatment (Volume 1)* John Wiley & Sons: 1974;

[17] Bruno de Finetti, *Probabilism: A Critical Essay on the Theory of Probability and on the Value of Science*, Erkenntnis (1975-), Vol. 31, No. 2/3, Bruno de Finetti's Philosophy of Probability (Sep., 1989), pp. 169-223;

[18] John Earman, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, MIT Press: 1992

[19] Bas C. Van Fraassen, *Relative Frequencies*, Synthese 34 (1977) 133-166;

[20] Clark Glymour, *Theory and Evidence*, Princeton University press: 1980;

[21] Clark Glymour, K. T. Kelly, *Why Probability Does Not Capture the Logic of Scientific Justification*, in Christopher Hitchcock, ed., Contemporary Debates in the Philosophy of Science, London: Blackwell: 2004

[22] Alan Hájek, *What Conditional Probability Could Not Be*, Synthese 137: 273323, 2003.

[23] Alan Hájek, *Interpretations of Probability*, The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL = ¡http://plato.stanford.edu/archives/win2012/entries/probability-interpret/¿.

[24] Paul R. Halmos, *Measure Theory*, Springer-Verlag: 1974

[25] James Hawthorne, *Inductive Logic*, 2008. URL http://plato.stanford.edu/entries/ logic-inductive/;

[26] John V. Howard, *Countable Additivity and the Foundations of Bayesian Statistics*, Theory and Decision, Volume: 60, Issue: 2-3 (May 1, 2006), pp: 127-135

[27] Colin Howson, *De Finetti, Countable Additivity, Consistency and Coherence*, The British Journal for the Philosophy of Science, No. 59 (2008), pp.1-23

[28] Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, Open Court, Chicago and La Salle, Illinois: 2006

[29] Franz Huber, *What is the point of confirmation?*, Philosophy of Science, Vol.72, No.5: 2005

[30] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge: 2003

[31] Joseph B. Kadane, Mark J. Schervish and Teddy Seidenfeld, *Statistical Implications of Finitely Additive Probability*, in *Rethinking the Foundations of Statistics*, Cambridge University Press: 1999

[32] Kevin T. Kelly, *The Logic of Reliable Inquiry*, Oxford University Press: 1996

[33] Andrey N. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea Publishing Company: 1956

[34] G. Schurz and H. Leitgeb, *Finitistic and Frequentistic Approximation of Probability Measures with or without σ-Additivity*, Studia Logica (2008) 89: 257283;

[35] John E. Littlewood, *Littlewood's Miscellany*, Cambridge University Press; Revised edition, 1986

[36] P. Mancosu, *Measuring the Size of Infinite Collections of Natural Numbers: Was Cantor's Theory of Infinite Numbers Inevitable?*, The Review of Symbolic Logic 2/4 (2009), 612-646;

[37] Jan von Plato, *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*, Cambridge University Press: 1994;

[38] Jonathan Weisberg, *Varieties of Bayesianism*, in Dov Gabbay, Stephan Hartmann & John Woods (eds.), *Handbook of the History of Logic*, vol. 10. Elsevier: 2011

[39] Sylvia Wenmackers and Leon Horsten, *Fair Infinite Lotteries*, Synthese (2013) 190:3761;

[40] Sylvia Wenmackers, *Philosophy of Probability: Foundations, Epistemology, and Computation*, PhD Thesis, University of Groningen, 2011;

[41] Jon Williamson, *Countable Additivity and Subjective Probability*, British Journal for the Philosophy of Science, Volume: 50, Issue: 3 (September 1, 1999), pp: 401-416