Uniform Probability Measures and
Epistemic Probability

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op donderdag 16 november 2017 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Timber Kerkvliet

egeboren te Leiden
promotor: prof.dr. R.W.J. Meester
ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Ronald. I am very grateful for the opportunity that you gave me to do research in areas of probability that are not particularly mainstream, sometimes even have an overlap with philosophy. I always felt like you had a great trust in the quality of my work, despite the indecipherable pieces of minimalistic writing that you sometimes found in your inbox. I enjoined our many conversations and discussions. I hope we will continue to have many of those in the future.

I want to thank the reading committee, Sylvia, Michel, Fabio and Klaas, for taking the time to read this thesis and for your useful feedback.

I would like to thank my parents for your continuous support and encouragement throughout the years.

A special gratitude goes out to my wife for dealing with all the little quirks of the mathematician you are married to.

I would also like to thank Zeno of Elea. At age 14, my interest in science and philosophy was sparked by your paradoxes. Who knows what would have become of me without your discovery.

And last but not least, thank you Bram for helping me with the cover.

Timber
## Contents

### I Uniform probability measures

1 Introduction

2 Uniform probability measure on $\mathbb{N}$
   2.1 Introduction and main results
   2.2 Weak thinnability
   2.3 The pair $(A_{uni}, \alpha)$
   2.4 Generalization to metric spaces
   2.5 Proofs
   2.6 Discussion

3 The supertask of an infinite lottery
   3.1 The supertask
   3.2 The mathematical model
   3.3 Discussion

### II Epistemic probability

4 Introduction

5 Conditioning and independence
   5.1 Contingent conditioning
   5.2 Necessary conditioning
   5.3 Mixed conditioning
   5.4 Independence
   5.5 Dempster’s rule of combination
   5.6 A law of large numbers

6 A betting interpretation
   6.1 Betting functions and degrees of belief
   6.2 Adding B-consistency
Part I

Uniform probability measures
CHAPTER 1

Introduction

Within the bounds of the Kolmogorov axioms [36], a probability measure on $\mathbb{N} = \{1, 2, 3, \ldots\}$ cannot assign the same probability to every singleton and therefore, a uniform probability measure on $\mathbb{N}$ does not exist.\(^1\) The crux of the problem is countable additivity: since every subset of $\mathbb{N}$ is countable, every subset should have probability zero, which is impossible because the probability of $\mathbb{N}$ itself should be 1. Despite this, we have some intuition about what a uniform probability measure on $\mathbb{N}$ should look like. According to this intuition, for example, we would assign probability $1/2$ to the subset of all odd numbers. If we want to capture this intuition in a mathematical framework, we have to violate at least one of the axioms of Kolmogorov.

One suggestion by De Finetti [19] is to relax countable additivity of the measure to finite additivity.\(^2\) To see why this suggestion is reasonable, we must first understand why it is possible, within the axioms of Kolmogorov, to set up uniform (Lebesgue) measure on $[0, 1]$. The type of additivity we require plays a crucial role here. In the standard theory one always requires countable additivity. If every singleton has the same probability, in an infinite space, every singleton must have probability zero. With countable additivity this means that every countable set must have probability zero. This is no problem if we are working on the uncountable $[0, 1]$, since we still have freedom to assign different probabilities to different uncountable subsets of $[0, 1]$. The interval $[0, 1/2]$, for example, has Lebesgue measure $1/2$, while it is equipotent with $[0,1]$, which has Lebesgue measure 1. The crux is that the cardinality of the index set is smaller than the cardinality of the space itself.

On $\mathbb{N}$, in analogy with Lebesgue measure, we want finite subsets to have

\(^1\)If we allow probability measures to take hyperreal values, then a uniform probability measure on $\mathbb{N}$ that satisfies a form of $\sigma$-additivity (and hence satisfies what one might call a non-standard version of the Kolmogorov axioms) does exist [64].

\(^2\)There is a controversy about the type of additivity that should be required for probability measures. We take the side of finite additivity in this debate, much in line with De Finetti. Notable arguments for $\sigma$-additivity in this debate are found in the work of, for example, Skyrims [52] and Williamson [65].
probability zero and we want to be able to assign different probabilities to countable subsets. To do this, we should change the type of additivity to finite additivity. In short: since the cardinality of the space changes from uncountable to countable, the additivity should change from countable to finite.

Schirokauer and Kadane [44] study three different collections of finitely additive probability measures on \( \mathbb{N} \) which may qualify as uniform. The first is the collection \( L \) of measures that extend natural density, i.e. measures \( \mu \) such that

\[
\mu(A) = \lim_{n \to \infty} \frac{|A \cap \{1, 2, \ldots, n\}|}{n} \tag{1.0.1}
\]

for all \( A \) for which this limit exists. The second collection is \( S \), consisting of all shift-invariant measures, i.e. measures \( \mu \) such that

\[
\mu(A) = \mu(\{k + 1 : k \in A\}) \tag{1.0.2}
\]

for all \( A \). The third collection is the collection \( R \) of measures that measure residue classes uniform, i.e. measures \( \mu \) such that

\[
\mu(\{m + kn : k \in \mathbb{N}\}) = \frac{1}{n} \tag{1.0.3}
\]

for all \( 0 \leq m < n \). They show that \( L \subset S \subset R \), where the inclusions are strict.

Some sets \( A \subseteq \mathbb{N} \), for instance

\[
A = \bigcup_{k \text{ odd}} \{2^k, 2^k + 1, \ldots, 2^{k+1} - 1\}, \tag{1.0.4}
\]

are without natural density, i.e.

\[
\frac{|A \cap \{1, 2, \ldots, n\}|}{n} \tag{1.0.5}
\]

does not converge as \( n \to \infty \). Given some sequence \( n_1 < n_2 < n_3 < \ldots \), measures \( \mu \) such that

\[
\mu(A) = \lim_{k \to \infty} \frac{|A \cap \{1, 2, \ldots, n_k\}|}{n_k} \tag{1.0.6}
\]

for all \( A \) for which this limit exists, are in \( L \). Given any set \( A \) without natural density, by choosing \( n_k \) appropriately, we can obtain different measures in \( L \) that assign different probabilities to \( A \). So even the smallest collection discussed by Schirokauer and Kadane does not lead to a uniquely determined uniform probability for sets which do not have a natural density.

In Chapter 2, that is based on [31], our aim is to find a natural notion of uniformity, such that the collection \( C \) of all probability measures that are uniform under this notion, is a subcollection of \( L \). In particular, we want that

\[
\{A : \mu_1(A) = \mu_2(A) \text{ for all } \mu_1, \mu_2 \in C\} \tag{1.0.7}
\]
is larger than the collection of sets having a natural density. The notion we propose, which we call weak thinnability, is a generalization of the property of uniform probability measures on $[0, 1]$ or on a finite spaces that if we condition on any suitable subset, the resulting conditional probability measure is again uniform on that subset. (The actual definition of weak thinnability is given later, and will also involve two technical conditions.)

Instead of trying to come up with probability measures that qualify as ‘uniform’, one can also try to come up with a chance experiment that has a natural number as outcome and qualifies as ‘fair’ and then determine which probability measures describe this experiment. The collection of probability measures that do this, naturally can be considered as ‘uniform’. In other words, from such an experiment emerges a notion of uniformity by considering the probability measures that describe the experiment.

Hansen [26, 27] introduces an experiment in which an infinity of gods together select a random natural number by each randomly removing a finite number of balls from an urn, leaving one final ball. This is what called a supertask: a task that consists of countably many tasks that are carried out within a finite time interval. In this case, every task consists of randomly removing a finite number of balls and is probabilistic in nature. This means that by the very nature of the supertask, what will happen is undetermined. However, the probability distribution of what happens in every task, given what has happened up to that task, is completely determined. This means we need a mathematical model of the supertask that specifies all the (a priori) possible distributions and the constraints on those distributions as dictated by the supertask. Within such a model, one can try to determine the collection of all distributions consistent with the supertask, i.e., the collection of possible distributions that satisfy the constraints.

At face value, it seems perfectly reasonable to call this supertask ‘fair’ and all probability distributions of the final ball consistent with the supertask ‘uniform’. Hansen conflates this notion of uniformity that emerges from the supertask, with notions of uniformity proposed on completely different grounds, like the ones of Schirokauer and Kadane discussed above. It is a priori not clear that any of these proposed notions is equivalent with the notion of uniformity that emerges from the supertask. A mathematical model followed by an analysis of that model, is necessary to conclude that.

In Chapter 3, that is based on [29], we present a mathematical model for the supertask of Hansen to study the notion of uniformity that emerges from the supertask. Within this model, we only insist on the finite additivity of candidate probability measures. We prove that given any $p \in [0, 1]$ and set $A \subseteq \mathbb{N}$ such that $A$ and $A^c$ are infinite, there is a probability measure consistent with the supertask that assigns probability $p$ to the final ball being in $A$. We also show that if $A$ or $A^c$ is finite, then every probability measure consistent with the supertask has to assign respectively probability 0 or 1 to the final ball being in $A$.

Our analysis shows that the description of the supertask is highly underdetermined: there are many probability measures consistent with the supertask,
all giving different probabilities of the final ball being in certain sets. As a consequence, we cannot speak of this supertask as having an uniquely determined underlying probability measure. The underdetermination makes the notion of uniformity emerging from the supertask unreasonably weak: it, for example, violates that every residue class modulo $m$ (e.g. in case $m = 2$ the odd and even numbers) has probability $1/m$. Hence it casts doubt on calling the supertask a ‘uniform’ lottery on $\mathbb{N}$. 
CHAPTER 2

Uniform probability measure on $\mathbb{N}$

2.1 Introduction and main results

In this chapter, we are interested in uniquely determined probability. Collections of sets that have a uniquely determined probability, are always closed under complements and finite disjoint unions. They are, however typically not algebras. Therefore, we allow probability measures to be defined on collections of sets that are closed under complements and finite disjoint unions. We should, however, be cautious when allowing domains that are not necessarily algebras, for the following reason. De Finetti [19] shows that if the domain of the probability measure is an algebra, the finite additivity of the probability measure implies coherence. On domains only closed under complements and finite disjoint unions, however, this implication no longer holds. Therefore, we would like to add coherence as additional constraint. For completeness, we study both the case with and the case without coherence as additional constraint on the probability measure.

Definition 2.1.1. Let $X$ be a space and write $\mathcal{P}(X)$ for the power set of $X$. An $f$-system on $X$ is a nonempty collection $\mathcal{F} \subseteq \mathcal{P}(X)$ such that

- $A, B \in \mathcal{F}$ with $A \cap B = \emptyset$ implies that $A \cup B \in \mathcal{F}$,
- $A \in \mathcal{F}$ implies that $A^c \in \mathcal{F}$.

A probability measure on an $f$-system $\mathcal{F}$ is a map $\mu : \mathcal{F} \rightarrow [0, 1]$ such that

- $A, B \in \mathcal{F}$ with $A \cap B = \emptyset$ implies that $\mu(A \cup B) = \mu(A) + \mu(B)$,
- $\mu(X) = 1$.

A coherent probability measure is a probability measure $\mu : \mathcal{F} \rightarrow [0, 1]$ such that for all $n \in \mathbb{N}, \alpha_1, ..., \alpha_n \in \mathbb{R}, A_1, ..., A_n \in \mathcal{F}$

$$\sup_{x \in X} \sum_{i=1}^{n} \alpha_i(I_{A_i}(x) - \mu(A_i)) \geq 0.$$ (2.1.1)
A probability pair on $X$ is a pair $(F, \mu)$ such that $F$ is an $f$-system on $X$ and $\mu$ is a probability measure on $F$.

**Remark 2.1.2.** Schurz and Leitgeb [46, p. 261] call an $f$-system a pre-Dynkin system, since in case of closure under countable unions of mutually disjoint sets, such a collection is called a Dynkin system.

**Remark 2.1.3.** Expression 2.1.1 has the following interpretation. If $\alpha_i \geq 0$, we buy a bet on $A_i$ that pays out $\alpha_i$ for $\alpha \mu(A_i)$. If $\alpha_i < 0$, we sell a bet on $A_i$ that pays out $|\alpha_i|$ for $|\alpha_i| \mu(A_i)$. Then (2.1.1) expresses there is no guaranteed amount of net loss.

We aim at uniquely determining the probability of as many sets as possible. In particular, we are interested in probability pairs with an $f$-system consisting only of sets with a uniquely determined probability. So we are not only interested in probability pairs satisfying our stronger notion of uniformity, but in the canonical ones, where “canonical” is to be understood in the following way.

**Definition 2.1.4.** Let $P$ be some collection of probability pairs. A pair $(F, \mu) \in P$ is canonical with respect to $P$ if for every $A \in F$ and every pair $(F', \mu') \in P$ with $A \in F'$ we have $\mu(A) = \mu'(A)$.

Before we give a more detailed outline of this chapter, we need the following definition. Set

$$M := \left\{ \bigcup_{i=1}^{\infty} [a_{2i-1}, a_{2i}) : 0 \leq a_1 \leq a_2 \leq a_3 \leq \cdots \right\}. \quad (2.1.2)$$

Note that $M$ is an algebra on $[0, \infty)$. It turns out that by working on $[0, \infty)$ instead of $\mathbb{N}$, where we restrict ourselves to sub-$f$-systems of $M$, we can formulate and prove our claims much more elegantly. Here, we view the elements of $P(\mathbb{N})$ embedded in $M$ by the injection

$$A \mapsto \bigcup_{n \in A} [n-1, n). \quad (2.1.3)$$

We should emphasize, however, that conceptually there is no difference between $[0, \infty)$ and $\mathbb{N}$ and that the work we do in Sections 2.2 and 2.3 can be done in the same way for $\mathbb{N}$. After working on $M$, we explicitly translate our result to $\mathbb{N}$ and other metric spaces in Section 2.4.

For $A \in M$ we define $\rho_A : [0, \infty) \to [0, 1]$ by $\rho_A(0) := 0$ and

$$\rho_A(x) := \frac{1}{x} \int_{0}^{x} 1_{A}(y) \, dy \quad (2.1.4)$$

for $x > 0$. Also set

$$C := \{ A \in M : \rho_A(x) \text{ converges} \}, \quad (2.1.5)$$

14
which are the elements of \( \mathcal{M} \) that have natural density and let \( \lambda : \mathcal{C} \rightarrow [0,1] \) be given by

\[
\lambda(A) := \lim_{x \to \infty} \rho_A(x).
\] (2.1.6)

We write \( L^* \) for the collection of probability pairs \( (\mathcal{F},\mu) \) on \([0,\infty)\) such that \( \mathcal{C} \subseteq \mathcal{F} \subseteq \mathcal{M} \) and \( \mu(A) = \lambda(A) \) for \( A \in \mathcal{C} \). Our earlier observation about the indeterminacy of probability under \( L \) gets the following formulation in terms of \( L^* \): a pair \( (\mathcal{F},\mu) \in L^* \) is canonical with respect to \( L^* \) if and only if \( \mathcal{F} = \mathcal{C} \). We write \( WT \) for the collection of probability pairs that are a weakly thinnable pair (WTP), that is, a probability pair that satisfies the condition of weak thinnability. The collection \( WT \) is a proper subset of \( L^* \) and contains pairs \( (\mathcal{F},\mu) \) canonical with respect to \( WT \) such that \( \mathcal{F} \setminus \mathcal{C} \neq \emptyset \). In other words, with restricting \( L^* \) to \( WT \) we are able to assign a uniquely determined probability to some sets without natural density. Finally, we write \( WTC \subseteq WT \subset L^* \) for the elements \( (\mathcal{F},\mu) \in WT \) such that \( \mu \) is coherent.

The structure of this chapter is as follows. In Section 2.2, we discuss weak thinnability and motivate why this is a natural notion of uniformity. In Section 2.3, we introduce the probability pair \( (A^{uni},\alpha) \) where

\[
A^{uni} = \left\{ A \in \mathcal{M} : \exists L \lim_{D \to \infty} \sup_{x \in (1,\infty)} \left| \frac{1}{\log(D)} \int_{x}^{xD} \frac{1}{y} A(y) \, dy - L \right| = 0 \right\}
\] (2.1.7)

and

\[
\alpha(A) = \lim_{D \to \infty} \frac{1}{\log(D)} \int_{1}^{D} \frac{1}{y} A(y) \, dy.
\] (2.1.8)

Remark 2.1.5. The expression in (2.1.8) is sometimes called the logarithmic density of \( A \) [60, p. 272].

We end Section 2.3 with the following theorem, which is the main result of this chapter.

**Theorem 2.1.6 (Main theorem).** The following holds:

- The pair \( (A^{uni},\alpha) \) is a WTP, is extendable to a WTP \( (\mathcal{F},\mu) \) with \( \mathcal{F} = \mathcal{M} \) and \( \alpha \) is coherent.
- The pair \( (A^{uni},\alpha) \) is canonical with respect to both \( WT \) and \( WTC \).
- If a pair \( (\mathcal{F},\mu) \) is canonical with respect to \( WT \) or \( WTC \), then \( \mathcal{F} \subseteq A^{uni} \).

In Section 2.4, we derive from \( (A^{uni},\alpha) \) analogous probability pairs on certain metric spaces including Euclidean space. The proofs of the results in Sections 2.2-2.4 are given in Section 2.5.

We write \( \mathbb{N}_0 := \{0,1,2,\ldots\} \). For real-valued sequences \( x,y \) or real-valued functions \( x,y \) on \([0,\infty)\) we write \( x \sim y \) or \( x_i \sim y_i \) if \( \lim_{i \to \infty} (x_i - y_i) = 0 \). Since we work only on \([0,\infty)\) in Sections 2.2 and 2.3, every time we speak of an \( f \)-system, probability pair or probability measure it is understood that this is on \([0,\infty)\).
2.2 Weak thinnability

Let $m$ be the Lebesgue measure on $\mathbb{R}$. For Lebesgue measurable $Y \subseteq \mathbb{R}$ with $0 < m(Y) < \infty$ the uniform probability measure on $Y$ is given by

$$ \mu_Y(X) := \frac{m(X)}{m(Y)} \quad (2.2.1) $$

for all Lebesgue measurable $X \subseteq Y$. Let $A \subseteq B \subseteq C$ be all Lebesgue measurable with $m(B) > 0$ and $m(C) < \infty$. Observe that

$$ \mu_C(A) = \mu_C(B) \mu_B(A). \quad (2.2.2) $$

We want to generalize this property to a property of probability pairs on $[0, \infty)$. For $A \in \mathcal{M}$ define $S_A : [0, \infty) \to [0, \infty)$ by

$$ S_A(x) := m(A \cap [0, x)). \quad (2.2.3) $$

Write

$$ \mathcal{M}^* := \{ A \in \mathcal{M} : m(A) = \infty \}. \quad (2.2.4) $$

Consider for $A \in \mathcal{M}^*$ the map $f_A : A \to [0, \infty)$ given by $f_A(x) := S_A(x)$. The map $f_A$ gives a one-to-one correspondence between $A$ and $[0, \infty)$. If $A \in \mathcal{M}^*$ and $B \in \mathcal{M}$, we want to introduce notation for the set

$$ \{ f_A^{-1}(b) : b \in B \}, \quad (2.2.5) $$

that gives the subset of $A$ that corresponds to $B$ under $f_A$. Inspired by van Douwen [13], we introduce the following operation.

**Definition 2.2.1.** For $A, B \in \mathcal{M}$, define

$$ A \circ B := \{ x \in [0, \infty) : x \in A \land S_A(x) \in B \}. \quad (2.2.6) $$

Note that if $A, B \in \mathcal{M}$, then $A \circ B \in \mathcal{M}$ and that for $A \in \mathcal{M}^*$ we have

$$ A \circ B = \{ f_A^{-1}(b) : b \in B \}. \quad (2.2.7) $$

We can view this operation as thinning $A$ by $B$ because we create a subset of $A$, where $B$ is “deciding” which parts of $A$ are removed. We also can view the operation $A \circ B$ as thinning out $B$ over $A$, since we “spread out” the set $B$ over $A$. Taking for example

$$ A = \bigcup_{i=0}^{\infty} [2i, 2i + 1) = [0, 1) \cup [2, 3) \cup [4, 5) \cup [6, 7) \cup \ldots \quad (2.2.8) $$

and

$$ B = \bigcup_{i=1}^{\infty} [i^2 - 1, i^2) = [0, 1) \cup [3, 4) \cup [8, 9) \cup [15, 16) \cup \ldots \quad (2.2.9) $$
we get

\[ A \circ B = [0, 1) \cup [6, 7) \cup [16, 17) \cup [30, 31) \cup [48, 49) \cup [70, 71) \cup ... \]  

(2.2.10)

and

\[ B \circ A = [0, 1) \cup [8, 9) \cup [24, 25) \cup [48, 49) \cup [80, 81) \cup [120, 121) \cup ... \]  

(2.2.11)

Let \((F, \mu)\) be a probability pair and let \(A \in F \cap M^*\). If \(B \in M\), the set \(A \circ B\) is the subset of \(A\) corresponding to \(B\). We can use this to transform \(\mu\) into a measure on \(A\) as follows. We set \(F_A := \{ A \circ B : B \in F \}\) and then define \(\mu_A : F_A \to [0, 1] \) by

\[ \mu_A(A \circ B) := \mu(B). \]  

(2.2.12)

Given \(B \in F\) such that \(A \circ B \in F\), the condition that

\[ \mu(A \circ B) = \mu(A)\mu_A(A \circ B) \]  

(2.2.13)

is a natural generalization of (2.2.2). Using (2.2.12) this translates into

\[ \mu(A \circ B) = \mu(A)\mu(B). \]  

(2.2.14)

We now have the restriction that \(A \in F \cap M^*\). However, if \(A \in F \setminus M^*\), then any uniform probability measure should assign 0 to \(A\) and since \(A \circ B \subseteq A\) (2.2.14) still holds. In Section 2.6, we show that the condition that (2.2.14) holds for all \(A, B \in F\) is so strong that only probability pairs with relatively small \(f\)-systems satisfy it. Since it is our goal to find a notion of uniformity that allows for a canonical pair with a large \(f\)-system, we choose to use a weakened version of this property which asks that \(\mu(C \circ A) = \mu(C)\mu(A)\) for every \(C \in C\) and \(A \in F\).

Weak thinnability also involves two technical conditions. Let \((F, \mu)\) be a probability pair, let \(A, B \in F\) and suppose it is true for every \(x \in [0, \infty)\) that

\[ S_A(x) \geq S_B(x). \]  

(2.2.15)

Since this inequality is true for every \(x\), the set \(B\) is “sparser” than \(A\). Therefore, it is natural to ask that \(\mu(A) \geq \mu(B)\). We call this property “preserving ordering by \(S\)”.

Since we have \(C \subseteq F\), it seems natural to also ask \(\mu|_C = \lambda\), but it turns out to be sufficient to ask the weaker property that \(\mu([c, \infty)) = 1\) for every \(c \in [0, \infty)\). So, to reduce redundancy we require the latter and then prove that \(\mu|_C = \lambda\). Putting everything together, we obtain the following definition.

**Definition 2.2.2.** A probability pair \((F, \mu)\) with \(F \subseteq M\) is a WTP if it satisfies the following conditions:

(WT1) For every \(C \in C\) and \(A \in F\) we have \(C \circ A \in F\) and \(\mu(C \circ A) = \mu(C)\mu(A)\),

(WT2) \(\mu\) preserves ordering by \(S\),
(WT3) $\mu([c, \infty)) = 1$ for every $c \in [0, \infty)$.

That every WTP extends natural density is implied by the following result.

**Theorem 2.2.3.** Let $(\mathcal{F}, \mu) \in WT$. Then for $A \in \mathcal{F}$ we have

$$\liminf_{x \to \infty} \rho_A(x) \leq \mu(A) \leq \limsup_{x \to \infty} \rho_A(x).$$

(2.2.16)

### 2.3 The pair $(\mathcal{A}^{\text{uni}}, \alpha)$

For $A \in \mathcal{M}$ set $\sigma_A : (0, \infty)^2 \to [0, 1]$ given by

$$\sigma_A(D,x) := \frac{1}{D} \int_x^{x+D} 1_A(y) dy,$$

(2.3.1)

which is the average of $1_A$ over the interval $[x, x + D]$. Then set for any $A \in \mathcal{M}$

$$U(A) := \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \sigma_A(D,x)$$

(2.3.2)

and

$$L(A) := \liminf_{D \to \infty} \inf_{x \in (0, \infty)} \sigma_A(D,x).$$

(2.3.3)

Define

$$\mathcal{W}^{\text{uni}} := \{ A \in \mathcal{M} : L(A) = U(A) \}.$$  

(2.3.4)

It is easy to check that $(\mathcal{W}^{\text{uni}}, \lambda|_{\mathcal{W}^{\text{uni}}})$ is a probability pair. For any $A \in \mathcal{M}$, we set

$$\log(A) := \{ \log(a) : a \in A \cap [1, \infty) \}.$$  

(2.3.5)

**Definition 2.3.1.** We define

$$\mathcal{A}^{\text{uni}} := \{ A \in \mathcal{M} : \log(A) \in \mathcal{W}^{\text{uni}} \}$$

(2.3.6)

and $\alpha : \mathcal{A}^{\text{uni}} \to [0, 1]$ by

$$\alpha(A) := \lambda(\log(A)).$$

(2.3.7)

Notice that Definition 2.3.1 gives a definition of $(\mathcal{A}^{\text{uni}}, \alpha)$ that is slightly different from (2.1.7) and (2.1.8). For a justification of equations 2.1.7 and 2.1.8, see the proof of Lemma 2.5.2. Our first concern is that $\alpha$ coincides with natural density.

**Theorem 2.3.2.** We have $\mathcal{C} \subseteq \mathcal{A}^{\text{uni}}$ and for every $A \in \mathcal{C}$

$$\alpha(A) = \lambda(\log(A)) = \lambda(A).$$

(2.3.8)
A typical example of a set in $A^\text{uni}$ that is not in $C$, is

$$A = \bigcup_{n=0}^{\infty} [e^{2n}, e^{2n+1}).$$  

(2.3.9)

It is easy to check that $A \notin C$, but

$$\log(A) = \bigcup_{n=0}^{\infty} [2n, 2n + 1),$$  

(2.3.10)

so $\log(A) \in W^\text{uni}$ with $\lambda(\log(A)) = 1/2$. Hence $A \in A^\text{uni}$ with $\alpha(A) = 1/2$.

That $(A^\text{uni}, \alpha)$ is a probability pair follows directly from the fact that $(W^\text{uni}, \lambda|_{W^\text{uni}})$ is a probability pair. The pair $(A^\text{uni}, \alpha)$ is also a WTP.

**Theorem 2.3.3.** We have $(A^\text{uni}, \alpha) \in WTC \subseteq WT$ and we can extend $(A^\text{uni}, \alpha)$ to a WTP with $M$ as f-system.

**Remark 2.3.4.** We use free ultrafilters in the proof of Theorem 2.3.3 to show there exists an extension to a WTP ith $M$ as f-system. The existence of free ultrafilters is guaranteed by the Boolean Prime Ideal Theorem, which can not be proven in ZF set theory, but is weaker than the axiom of choice [25]. The existence of a atomfree or nonprincipal (i.e. every singleton has measure zero) finite additive measure defined on the power set of $\mathbb{N}$ cannot be established in ZF alone [59]. Consequently, a version of the axiom of choice is always necessary to construct a probability measure on $M$ that assigns measure zero to all bounded intervals.

We do not only want an element of $WTC$, but a canonical one. This is guaranteed by the following theorem.

**Theorem 2.3.5.** The pair $(A^\text{uni}, \alpha)$ is canonical with respect to both $WT$ and $WTC$.

The pair $(A^\text{uni}, \alpha)$ is maximal in the sense that it contains every pair that is canonical with respect to $WT$ or $WTC$.

**Theorem 2.3.6.** If $(F, \mu) \in WT$ is canonical with respect to $WT$, then $F \subseteq A^\text{uni}$. If $(F, \mu) \in WTC$ is canonical with respect to $WTC$, then $F \subseteq A^\text{uni}$.

### 2.4 Generalization to metric spaces

In this section we derive probability pairs on a class of metric spaces that are analogous to $(A^\text{uni}, \alpha)$. Of course one could also try to construct such a probability measure by working more directly on these metric spaces, instead of constructing a derivative of $(A^\text{uni}, \alpha)$. Since probability pairs on $[0, \infty)$, motivated from the problem of a uniform probability measure on $\mathbb{N}$, is the priority of this chapter, we do not make such an effort here.
Let us first sketch the idea of the generalization. Let $A \in \mathcal{M}$. Whether $A$ is in $\mathcal{A}^\text{uni}$ depends completely on the asymptotic behavior of $\rho_A$ (Lemma 2.5.2). If $A \in \mathcal{A}^\text{uni}$, then also $\alpha(A)$ only depends on the asymptotic behavior of $\rho_A$ (Lemma 2.5.2). Now suppose that on a space $X$, we can somehow define a density functions $\tilde{\rho}_B : [0, \infty) \to [0, 1]$ for (some) subsets $B \subseteq X$ in a canonical way. Then, by replacing $\rho$ by $\tilde{\rho}$, we get the analogue of $(\mathcal{A}^\text{uni}, \alpha)$ in $X$. The goal of this section is to make this idea precise.

Let $(X, d)$ be a metric space. For $x \in X$ and $r \geq 0$, write $B(x, r) := \{y \in X : d(y, x) < r\}$. (2.4.1)

Write $\mathcal{B}(X)$ for the Borel $\sigma$-algebra of $X$. We need a “uniform” measure on this space to measure density of subsets in open balls. It is clear that the measure of an open ball should at least be independent of where in the space we look, i.e. it should only depend on the radius of the ball. This leads to the following definition.

**Definition 2.4.1.** We say that a Borel measure $\nu$ on $X$ is uniform if for all $r > 0$ and $x, y \in X$ we have

$$0 < \nu(B(x, r)) = \nu(B(y, r)) < \infty.$$ (2.4.2)

On $\mathbb{R}^n$ with Euclidean metric, the standard Borel measure as obtained by assigning to a product of intervals the product of the lengths of those intervals, is a uniform measure. In general, on normed locally compact vector spaces, the invariant measure with respect to vector addition, as given by the Haar measure, is a uniform measure.

A result by Christensen [5] tells us that uniform measures that are Radon measures are unique up to multiplicative constants on locally compact metric spaces. This, however, does not cover all cases. The set of irrational numbers, for example, is not locally compact, but the Lebesgue measure restricted to Borel sets of irrational numbers is a uniform measure and unique up to a multiplicative constant. We give a slightly more general version of the result of Christensen.

**Theorem 2.4.2.** If $\nu_1$ and $\nu_2$ are two uniform measures on $X$, then there exists some $c > 0$ such that $\nu_1 = c \nu_2$.

Proposition 2.4.2 gives us uniqueness, but not existence. To see that there are metric spaces without a uniform measure, consider the following example. Let $X$ be the set of vertices in a connected graph that is not regular. Let $d$ be the graph distance on $X$. If we suppose that $\nu$ is a uniform measure on $X$, from (2.4.2) with $r < 1$ it follows that for some $C > 0$ we have $\nu(\{x\}) = C$ for every $x \in X$. But then $\nu(B(x, 2)) = C(1 + \deg(x))$ for every $x \in V$, which implies (2.4.2) cannot hold for $r = 2$ since the graph is not regular. A characterization of metric spaces on which a uniform measure exist, does not seem to be present in the literature.

We now assume $X$ has a uniform measure $\nu$ and that $\nu(X) = \infty$. In addition to that, we write $h(r) := \nu(B(x, r))$ for $r \geq 0$ and assume that

$$\forall C > 0 \lim_{r \to \infty} \frac{h(r + C)}{h(r)} = 1,$$ (2.4.3)
which is equivalent with amenability in case \((X, d)\) is a normed locally compact vector space [41]. For the importance of this assumption, see Remark 2.4.4 below.

Set
\[
\begin{align*}
r^-(u) &:= \sup\{r \in [0, \infty) : h(r) \leq u\}, \\
r^+(u) &:= r^- + 1
\end{align*}
\] (2.4.4)

for \(u \in [0, \infty)\). Note that \(h(r^-) \leq u\) and \(h(r^+) \geq u\). Write \((X, L(X), \bar{\nu})\) for the (Lebesgue) completion of \((X, B(X), \nu)\). Fix some \(o \in X\). For \(A \in L(X)\) define the map \(\bar{\rho}_A : [0, \infty) \to [0, \infty)\) given by
\[
\bar{\rho}_A(u) := \frac{\bar{\nu}(B(o, r^-) \cap A)}{h(r^-)}
\] (2.4.5)

for \(r > 0\). The value \(\bar{\rho}_A(u)\) is the density of \(A\) in the biggest open ball around \(o\) of at most measure \(u\). Notice that \(\bar{\rho}_A\) is independent of the choice of \(\nu\) as a result of Proposition 2.4.2. The function \(\bar{\rho}_A\) does depend on the choice of \(o\), but in Proposition 2.4.3 we show that the asymptotic behavior of \(\bar{\rho}_A\) does not depend on the choice of \(o\). We also show in Proposition 2.4.3 that the asymptotic behavior of \(\bar{\rho}_A\) is not affected if we replace \(r^-\) by \(r^+\) in (2.4.5).

**Theorem 2.4.3.** Fix \(x, y \in X\) and \(A \in L(X)\). Then
\[
\frac{\bar{\nu}(B(x, r^-) \cap A)}{h(r^-)} \sim \frac{\bar{\nu}(B(y, r^+) \cap A)}{h(r^+)}.
\] (2.4.6)

**Remark 2.4.4.** Proposition 2.4.3 is not necessarily true if we do not assume (2.4.3), as illustrated by the following example. Suppose \(X\) is the set of vertices of a 3-regular tree graph and \(d\) is the graph distance. Let \(\nu\) be the counting measure, which is a uniform measure on this metric space. Then clearly (2.4.3) is not satisfied. Now pick any \(x \in X\) and let \(y\) be a neighbor of \(x\). Let \(A \subseteq P(X)\) be the connected component containing \(y\) in the graph where the edge between \(x\) and \(y\) is removed. Then
\[
\lim_{r \to \infty} \frac{\bar{\nu}(B(x, r) \cap A)}{h(r)} = 1/3 \quad \text{and} \quad \lim_{r \to \infty} \frac{\bar{\nu}(B(y, r) \cap A)}{h(r)} = 2/3.
\] (2.4.7)

Proposition 2.4.3 justifies the use of \(\bar{\rho}\) to determine the density, since its asymptotic behavior is canonical. So, we define for \(A \in L(X)\) the map \(\xi_A : (1, \infty)^2 \to [0, 1]\) given by
\[
\xi_A(D, x) := \frac{1}{\log(D)} \int_x^D \frac{\bar{\rho}_A(y)}{y} \, dy.
\] (2.4.8)

Then we set
\[
\mathcal{A}^{\text{uni}}(X) := \left\{ A \in L(X) : \limsup_{D \to \infty} \sup_{x > 1} \xi_A(D, x) = \liminf_{D \to \infty} \inf_{x > 1} \xi_A(D, x) \right\}
\] (2.4.9)
and $\alpha^X : \mathcal{A}^{\text{uni}}(X) \to [0, 1]$ by

$$\alpha^X(A) := \limsup_{D \to \infty} \sup_{x \in (1, \infty)} \bar{\xi}_A(D, x) = \liminf_{D \to \infty} \inf_{x \in (1, \infty)} \bar{\xi}_A(D, x).$$

(2.4.10)

The pair $(\mathcal{A}^{\text{uni}}(X), \alpha^X)$ gives us the analogue of $(\mathcal{A}^{\text{uni}}, \alpha)$ in $X$. In particular, it gives for $X = \mathbb{N}$ the corresponding uniform probability measure on $\mathbb{N}$ we initially searched for. In case of Euclidean space, we have the following expression for $(\mathcal{A}^{\text{uni}}(X), \alpha^X)$, which in the special case of $X = \mathbb{R}$ gives us an extension of $\alpha$ ($\mathcal{A}^{\text{uni}}(\mathbb{R})$ is the maximal sub-$f$-system of $\mathcal{L}(\mathbb{R})$, where $\mathcal{A}^{\text{uni}} \subseteq \mathcal{A}^{\text{uni}}(\mathbb{R})$ is the maximal sub-$f$-system of $\mathcal{M}$).

**Theorem 2.4.5.** Suppose $X = \mathbb{R}^n$ and $d$ is Euclidean distance. Let $\sigma$ be the surface measure on the unit sphere in $\mathbb{R}^n$. Then for $A \in \mathcal{L}(\mathbb{R}^n)$ we can replace $\bar{\xi}_A(D, x)$ in (2.4.9) and (2.4.10) by

$$\frac{1}{\log(D)} \int_x^{Dx} \frac{K_A(y)}{y} dy,$$

(2.4.11)

where $K_A : [0, \infty) \to [0, 1]$ is given by

$$K_A(r) := \frac{\Gamma(n/2)}{2\pi^{n/2}} \int_{S^{n-1}} 1_A(ru) \sigma(du).$$

(2.4.12)

### 2.5 Proofs

First we show that every $f$-system of a WTP is closed under translation and that every probability measure of a WTP is invariant under translation.

**Lemma 2.5.1.** Let $(\mathcal{F}, \mu)$ be a WTP. Let $A \in \mathcal{F}$ and $c \in [0, \infty)$. Then

$$A' := \{c + a : a \in A\} \in \mathcal{F}$$

(2.5.1)

and $\mu(A) = \mu(A')$.

**Proof.** Let $(\mathcal{F}, \mu)$ be a WTP. Let $A \in \mathcal{F}$ and $c \in [0, \infty)$. Set $B := [c, \infty)$. We have $B \in \mathcal{C} \subseteq \mathcal{F}$ and by (WT3) we have $\mu(B) = 1$. Therefore, $A' = B \circ A \in \mathcal{F}$ and

$$\mu(A') = \mu(B) \mu(A) = \mu(A)$$

(2.5.2)

by (WT1).

**Proof of Theorem 2.2.3.** Let $(\mathcal{F}, \mu)$ be a WTP and $A \in \mathcal{F}$. Set $u := \limsup_{x \to \infty} \rho_A(x)$.

If $u = 1$ there is nothing to prove, so assume $u < 1$. Let $\epsilon > 0$ be given. Let $u' \in [0, 1] \cap \mathbb{Q}$ such that $u' > u$ and $u' - u < \epsilon$. The idea is to construct a $Y \in \mathcal{M}$ such that we can easily see that $\mu(Y) = u'$ and $\rho_A(x) \leq \rho_Y(x)$ for all $x$, so that with (WT2) we get $\mu(A) \leq u'$. 

22
First we observe that there is a \( K > 0 \) such that for all \( x \geq K \) we have \( \rho_A(x) \leq u' \). We can write \( u' \) as \( u' = \frac{p}{q} \) for some \( p, q \in \mathbb{N}_0 \) with \( p \leq q \). Now we introduce the set \( Y \) given by

\[
Y := [0, K) \cup \bigcup_{i=0}^{\infty} [iq, iq + p).
\]

Note that \( Y \in C \subseteq \mathcal{F} \). Lemma 2.5.1 and the fact that \( \mu \) is a probability measure, gives us that \( \mu(Y) = u' \). Further, observe that for each \( x \in [0, \infty) \) we have \( \rho_A(x) \leq \rho_Y(x) \), so with (WT2) we get

\[
\mu(A) \leq \mu(Y) = u' < u + \epsilon.
\]

Letting \( \epsilon \downarrow 0 \) we find

\[
\mu(A) \leq u = \limsup_{x \to \infty} \rho_A(x).
\]

By applying this to \( A^c \) we find

\[
\mu(A) = 1 - \mu(A^c) \geq 1 - \limsup_{x \to \infty} \rho_A(x) = \liminf_{x \to \infty} \rho_A(x).
\]

Before we prove Theorem 2.3.2 and Theorem 2.3.3, we present the following alternative representation of \((A^{\text{uni}}, \alpha)\). We define for \( A \in \mathcal{M} \) the map \( \xi_A : (1, \infty)^2 \to [0, 1] \) given by

\[
\xi_A(D, x) := \frac{1}{\log(D)} \int_x^{Dx} \frac{\rho_A(y)}{y} \, dy.
\]

(2.5.3)

Set

\[
S := \left\{ s \in (1, \infty)^N : \lim_{n \to \infty} s_n = \infty \right\}.
\]

(2.5.4)

If \( s \in S \) and \( f \in (1, \infty)^N \), then we can interpret the pair \((s, f)\) as the sequence \((s_1, f_1), (s_2, f_2), \ldots \) in \((1, \infty)^2\). Write

\[
\mathcal{P} := \{(s, f) : s \in S, \ f \in (1, \infty)^N\}
\]

(2.5.5)

for the collection of all such sequences.

For every \((s, f) \in \mathcal{P}\) we set

\[
A^{s,f} := \{ A \in \mathcal{M} : \lim_{n \to \infty} \xi_A(s_n, f_n) \text{ exists} \}
\]

(2.5.6)

and

\[
\alpha^{s,f}(A) := \lim_{n \to \infty} \xi_A(s_n, f_n).
\]

(2.5.7)
Lemma 2.5.2 (Alternate Representation). We have

\[ A^\text{uni} = \bigcap_{(s,f) \in \mathcal{P}} A^{s,f} \]  

(2.5.8)

with for any \((s,f) \in \mathcal{P}\) and \(A \in A^\text{uni}\)

\[ \alpha(A) = \alpha^{s,f}(A). \]  

(2.5.9)

Proof. Let \(A \in \mathcal{M}\). We start to relate \(\sigma_{\log(A)}\) and \(\xi_A\). If \(D, x \in (1, \infty)\), then

\[
\begin{align*}
\sigma_{\log(A)}(\log(D), \log(x)) &= \frac{1}{\log(D)} \int_{\log(x)}^{\log(D)} 1_A(e^y)dy \\
&= \frac{1}{\log(D)} \int_x^{Dx} \frac{1_A(u)}{u} du \\
&= \frac{1}{\log(D)} \int_x^{Dx} \frac{S'_A(u)}{u} du \\
&= \frac{1}{\log(D)} \left( \frac{S_A(u)}{u} \bigg|_{u=x}^{Dx} + \int_x^{Dx} \frac{S_A(u)}{u^2} du \right) \\
&= \frac{\rho_A(Dx) - \rho_A(x)}{\log(D)} + \xi_A(D, x).
\end{align*}
\]  

(2.5.10)

This implies that for \((s, f) \in \mathcal{P}\) we have

\[ A^{s,f} = \left\{ A \in \mathcal{M} : \lim_{n \to \infty} \sigma_{\log(A)}(\log(s_n), \log(f_n)) \text{ exists} \right\} \]  

(2.5.11)

with for \(A \in A^{s,f}\)

\[ \alpha^{s,f}(A) = \lim_{n \to \infty} \sigma_{\log(A)}(\log(s_n), \log(f_n)). \]  

(2.5.12)

Since for any \(A \in \mathcal{M}\) and \((s, f) \in \mathcal{P}\)

\[ L(\log(A)) \leq \liminf_{n \to \infty} \sigma_{\log(A)}(\log(s_n), \log(f_n)) \]  

(2.5.13)

and

\[ \limsup_{n \to \infty} \sigma_{\log(A)}(\log(s_n), \log(f_n)) \leq U(\log(A)), \]  

(2.5.14)

we find that if \(\log(A) \in \mathcal{W}^\text{uni}\), then \(A \in A^{s,f}\) with \(\alpha^{s,f}(A) = \alpha(A)\).

The only thing left to show is that

\[ \bigcap_{(s,f) \in \mathcal{P}} A^{s,f} \subseteq A^\text{uni}. \]  

(2.5.15)

24
So assume $A \in \bigcap_{(s,f) \in P} A^{s,f}$. Suppose we have $(s, f) \in P$ such that $\alpha^{s,f}(A) = L(\log(A))$ and $(s', f') \in P$ such that $\alpha^{s',f'}(A) = U(\log(A))$. Then we can create a new sequence given by

$$(2.5.16)$$

$$s'' := (s_1, s'_1, s_2, s'_2, \ldots) \text{ and } f'' := (f_1, f'_1, f_2, f'_2, \ldots).$$

Because by assumption $A \in A^{s'',f''}$, we then have $\alpha^{s',f'}(A) = \alpha^{s'',f''}(A)$. Hence $A \in A^{\text{uni}}$. So it is sufficient to show that we can choose $(s, f)$ and $(s', f')$ in the desired way.

Choose $s \in S$ such that

$$\lim_{n \to \infty} \inf_{x \in (1, \infty)} \sigma_{\log(A)}(\log(s_n), \log(x)) = \lim_{D \to \infty} \inf_{x \in (1, \infty)} \sigma_{\log(A)}(\log(D), \log(x)).$$

(2.5.17)

Choose $f \in (1, \infty)^N$ such that

$$\left| \inf_{x \in (1, \infty)} \sigma_{\log(A)}(\log(s_n), \log(x)) - \sigma_{\log(A)}(\log(s_n), \log(f_n)) \right| < \frac{1}{n}$$

(2.5.18)

for every $n \in \mathbb{N}$. Then $(s, f) \in P$ with

$$\alpha^{s,f}(A) = \lim_{D \to \infty} \inf_{x \in (1, \infty)} \sigma_{\log(A)}(\log(D), \log(x)) = L(\log(A)).$$

(2.5.19)

In the same way choose $(s', f') \in P$ such that

$$\alpha^{s',f'}(A) = \lim_{D \to \infty} \sup_{x \in (1, \infty)} \sigma_{\log(A)}(\log(D), \log(x)) = U(\log(A)).$$

(2.5.20)

Proof of Proposition 2.3.2. Let $A \in C$ and $(s, f) \in P$. Since $\rho_A(y) \to \lambda(A)$, we have $\xi_A(s_n, f_n) \sim \lambda(A)$, so $\alpha^{s,f}(A) = \lambda(A)$. The result now follows by Lemma 2.5.2.

Proof of Theorem 2.3.3. Notice that any intersection of $f$-systems closed under weak thinning is again closed under weak thinning. Therefore, if we show that $(A^{s,f}, \alpha^{s,f})$ is a WTP for every $(s, f) \in P$, it follows from Lemma 2.5.2 that $(A^{\text{uni}}, \alpha)$ is a WTP.

Let $(s, f) \in P$. It immediately follows that $(A^{s,f}, \alpha^{s,f})$ is a probability pair and that (WT2) and (WT3) hold, so we have to verify (WT1). Note that for
every $A, B \in \mathcal{M}$ and $x > 0$ we have
\[
\rho_{A \circ B}(x) = \frac{1}{x} \int_0^x 1_{A \circ B}(y)dy \\
= \frac{1}{x} \int_0^x 1_A(y)1_B(S_A(y))dy \\
= \frac{1}{x} \int_0^{S_A(x)} 1_B(u)du \\
= \frac{S_A(x)}{x} S_A(x) \int_0^{S_A(x)} 1_B(u)du \\
= \rho_A(x) \rho_B(S_A(x)) = \rho_A(x) \rho_B(x \rho_A(x)).
\] (2.5.21)

Let $A \in C$ and $B \in \mathcal{A}^{s,f}$. Then
\[
\xi_{A \circ B}(s_n, f_n) = \frac{1}{\log(s_n)} \int_{f_n}^{s_n f_n} \rho_A(y) \frac{\rho_B(y \rho_A(y))}{y} dy \\
\sim \lambda(A) \frac{1}{\log(s_n)} \int_{f_n}^{s_n f_n} \rho_B(\lambda(A)y) dy.
\] (2.5.22)

If $\lambda(A) = 0$ it is clear that $A \circ B \in \mathcal{A}^{s,f}$ with $\alpha^{s,f}(A \circ B) = 0 = \lambda(A) \alpha^{s,f}(B)$. If
\[
\lambda(A) > 0, \text{ then we see that}
\int_{f_n}^{s_n f_n} \frac{\rho_B(\lambda(A)y)}{y} dy = \int_{\lambda(A)f_n}^{\lambda(A)s_n f_n} \frac{\rho_B(u)}{u} du.
\] (2.5.23)

Since
\[
\left| \int_{\lambda(A)f_n}^{\lambda(A)s_n f_n} \frac{\rho_B(u)}{u} du - \int_{f_n}^{s_n f_n} \frac{\rho_B(u)}{u} du \right| \leq \int_{f_n}^{s_n f_n} \frac{1}{u} du + \int_{\lambda(A)f_n}^{f_n} \frac{1}{u} du \\
= 2 \log \left( \frac{1}{\lambda(A)} \right),
\] (2.5.24)

we have
\[
\frac{1}{\log(s_n)} \int_{f_n}^{s_n f_n} \frac{\rho_B(\lambda(A)y)}{y} dy \sim \frac{1}{\log(s_n)} \int_{f_n}^{s_n f_n} \frac{\rho_B(u)}{u} du \sim \alpha^{s,f}(B).
\] (2.5.25)

Thus $A \circ B \in \mathcal{A}^{s,f}$ and since $\lambda(A) = \alpha^{s,f}(A)$ (see the proof of Theorem 2.3.2), we have
\[
\alpha^{s,f}(A \circ B) = \lambda(A) \alpha^{s,f}(B) = \alpha^{s,f}(A) \alpha^{s,f}(B).
\] (2.5.26)

We have showed that $(\mathcal{A}^{\text{uni}}, \alpha) \in WT$. To show that $(\mathcal{A}^{\text{uni}}, \alpha)$ can be extended, let $\mathcal{U}$ be any free ultrafilter on $\mathbb{N}$ and $(s, f) \in \mathcal{P}$. Then consider $\mu : \mathcal{M} \to [0, 1]$ given by
\[
\mu(A) := \mathcal{U}- \lim_{n \to \infty} \xi_A(s_n, f_n).
\] (2.5.27)
2.5. PROOFS

Since the $U$-limit is multiplicative it follows completely analogous that $(\mathcal{M}, \mu)$ is a WTP. Hence every $(A^{s,f}, \alpha^{s,f})$ can be extended to a WTP with $\mathcal{M}$ as its $f$-system. In particular, by Lemma 2.5.2, this means that $(A^{uni}, \alpha)$ can be extended to a WTP with $\mathcal{M}$ as its $f$-system.

From de Finetti [18] it follows that if $\alpha$ can be extended to a finitely additive probability measure on an algebra, then $\alpha$ is coherent. Since we have showed that $\alpha$ can be extended to $\mathcal{M}$, which is an algebra, it follows that $(A^{uni}, \alpha) \in WTC$. Notice that we showed that $\alpha^{s,f}$ can be extended to $M$ for every $(s,f) \in P$, so we also have $(A^{s,f}, \alpha^{s,f}) \in WTC$ for every $(s,f) \in P$. 

For our proof of Theorem 2.3.5, we need an alternate expression for $U(\log(A))$. For $A \in M$ set $\tau_A : (1, \infty) \times \mathbb{N} \to [0, 1]$ given by

$$
\tau_A(C, j) := \sigma_A(C^{j-1}(C - 1), C^{j-1}) = \frac{1}{C^{j-1}(C - 1)} \int_{C^{j-1}}^{C^j} 1_A(y)dy.
$$

(2.5.28) (2.5.29)

Also set for $C > 1$ and $A \in M$

$$
U^*(C, A) := \lim_{n \to \infty} \sup_{k \in \mathbb{N}} \frac{1}{n} \sum_{j=k}^{k+n-1} \tau_A(C, j).
$$

(2.5.30)

**Lemma 2.5.3.** For every $A \in M$ we have

$$
\lim_{C \downarrow 1} U^*(C, A) = U(\log(A)).
$$

(2.5.31)

**Proof.** Let $A \in M$ and fix $C > 1$.

**Step 1** We show that

$$
U(\log(A)) = \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{1}{D} \sum_{j=P(x)+1}^{Q(D, x)} \int_{C^{j-1}}^{C^j} \frac{1_A(u)}{u}du,
$$

(2.5.32)

where

$$
P(x) := \left\lceil \frac{x}{\log(C)} \right\rceil \quad \text{and} \quad Q(D, x) := \left\lceil \frac{D + x}{\log(C)} \right\rceil
$$

(2.5.33)

for $D, x \in (0, \infty)$. Define

$$
E(D, x) := \sigma_{\log(A)}(D, x) - \frac{1}{D} \int_{C^{P(x)}}^{C^{Q(D, x)}} \frac{1_A(u)}{u}du.
$$

(2.5.34)

Since

$$
\sigma_{\log(A)}(D, x) = \frac{1}{D} \int_{x}^{x+D} 1_A(e^y)dy = \frac{1}{D} \int_{e^x}^{e^{x+D}} \frac{1_A(u)}{u}du,
$$

(2.5.35)
we have

\[ |E(D, x)| \leq \frac{1}{D} \int_{C^P(x)}^{C^Q(D, x) + 1} \frac{1}{u} \, du + \frac{1}{D} \int_{C^Q(D, x)}^{C^Q(x)} \frac{1}{u} \, du = \frac{2}{D} \log(C). \quad (2.5.36) \]

This implies

\[ U(\log(A)) = \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \sigma_{\log(A)}(D, x) \]

\[ = \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{1}{D} \int_{C^P(x)}^{C^Q(x)} \frac{1}{u} \, du \quad (2.5.37) \]

\[ = \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{1}{D} \sum_{j=P(x)+1}^{C^Q} \int_{C^{j-1}}^{C^j} \frac{1}{u} \, du. \]

**Step 2** We give an upper and lower bound for

\[ \int_{C^{j-1}}^{C^j} \frac{1}{u} A(u) \, du \quad (2.5.38) \]

in terms of \( \tau_A(C, j) \).

If we set for \( j \in \mathbb{N} \)

\[ \zeta(j) := \int_{C^{j-1}}^{C^j} 1_A(y) \, dy = \tau_A(C, j)(C - 1)C^{j-1} \quad (2.5.39) \]

then

\[ \int_{C^{j-1}}^{C^j} \frac{1}{u} \, du \leq \int_{C^{j-1}}^{C^j} \frac{1}{u} A(u) \, du \leq \int_{C^{j-1}}^{C^{j-1} + \zeta(j)} \frac{1}{u} \, du. \quad (2.5.40) \]

We now observe that

\[ \int_{C^{j-1}}^{C^{j-1} + \zeta(j)} \frac{1}{u} \, du = \log \left( \frac{C^{j-1} + \zeta(j)}{C^{j-1}} \right) = \log(1 + (C - 1)\tau_A(C, j)) \quad (2.5.41) \]

and

\[ \int_{C^{j-1}}^{C^j} \frac{1}{u} \, du = \log \left( \frac{C^j}{C^j - \zeta(j)} \right) = \log(C) - \log (1 + (C - 1)(1 - \tau_A(C, j))). \quad (2.5.42) \]

The fact that \( \log(1 + y) \leq y \) for every \( y \geq 0 \), combined with (2.5.40), (2.5.41) and (2.5.42) gives

\[ \log(C) - (C - 1)(1 - \tau_A(C, j)) \leq \int_{C^{j-1}}^{C^j} \frac{1}{u} A(u) \, du \leq (C - 1)\tau_A(C, j). \quad (2.5.43) \]
Step 3 We combine Step 1 and Step 2 to finish the proof.

Observe that

\[
\limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{1}{Q(D, x) - P(x)} \sum_{j = P(x) + 1}^{Q(D, x)} \tau_A(C, j) = \limsup_{n \to \infty} \sup_{k \in \mathbb{N}} \frac{1}{n} \sum_{j = k}^{k + n - 1} \tau_A(C, j) = U^*(C, A). \tag{2.5.44}
\]

We use (2.5.43) and (2.5.44) to find an upper bound for the expression in (2.5.37), giving us

\[
U(\log(A)) = \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{1}{D} \sum_{j = P(x) + 1}^{Q(D, x)} \int_{C j}^{C(j - 1)} \frac{1_A(u)}{u} \, du \\
\leq \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{C - 1}{D} (Q(D, x) - P(x)) \gamma(D, x) \\
= \frac{C - 1}{\log(C)} \limsup_{D \to \infty} \sup_{x \in (0, \infty)} \frac{1}{Q(D, x) - P(x)} \sum_{j = P(x) + 1}^{Q(D, x)} \tau_A(C, j) \\
= \frac{C - 1}{\log(C)} U^*(C, A). \tag{2.5.45}
\]

Analogously, we find that

\[
U(\log(A)) \geq 1 - \frac{C - 1}{\log(C)} (1 - U^*(C, A)). \tag{2.5.46}
\]

Combining (2.5.45) and (2.5.46) we obtain

\[
\frac{\log(C)}{C - 1} U(\log(A)) \leq U^*(C, A) \leq 1 - \frac{\log(C)}{C - 1} (1 - U(\log(A))), \tag{2.5.47}
\]

which implies

\[
\lim_{C \downarrow 1} U^*(C, A) = U(\log(A)). \tag{2.5.48}
\]

We also need the following lemma.

**Lemma 2.5.4.** Let $(\mathcal{F}, \mu)$ be a WTP. Then for any $A \in \mathcal{F}$ and $C > 1$

\[
\mu(A) \leq C \sup_{j \in \mathbb{N}} \tau_A(C, j). \tag{2.5.49}
\]
Proof. Let \((\mathcal{F}, \mu)\) be a WTP with \(A \in \mathcal{F}\). Fix \(C > 1\) and write
\[
S := \sup_{j \in \mathbb{N}} \tau_A(C, j). \tag{2.5.50}
\]
The idea is to introduce a set \(B \in \mathcal{M}\) for which we have \(\limsup_{x \to \infty} \rho_B(x) \leq CS\) and \(\rho_A(x) \leq \rho_B(x)\) for all \(x\). Set
\[
B := \bigcup_{j=1}^{\infty} [C^{j-1}, C^{j-1} + SC^{j-1}(C - 1)]. \tag{2.5.51}
\]
By construction of \(B\) we have \(\rho_A(x) \leq \rho_B(x)\) for every \(x \in (0, \infty)\). So
\[
\limsup_{x \to \infty} \rho_A(x) \leq \limsup_{x \to \infty} \rho_B(x)
= \limsup_{n \to \infty} \rho_B(C^n + SC^n(C - 1))
= \limsup_{n \to \infty} \sum_{j=1}^{n+1} \frac{SC^{j-1}(C - 1)}{C^n + SC^n(C - 1)}
= \limsup_{n \to \infty} \frac{S(C^{n+1} - 1)}{C^n + SC^n(C - 1)}
= \limsup_{n \to \infty} \frac{CS}{C - C^n(1 + S(C - 1))}
= \frac{CS}{1 + S(C - 1)} \leq CS. \tag{2.5.52}
\]
By Proposition 2.2.3 we then find
\[
\mu(A) \leq \limsup_{x \to \infty} \rho_A(x) \leq CS. \tag{2.5.53}
\]

We are ready to give the proof of Theorem 2.3.5.

Proof of Theorem 2.3.5. Let \((\mathcal{F}, \mu)\) be a WTP and \(A \in \mathcal{F}\). It is sufficient to show that
\[
L(\log(A)) \leq \mu(A) \leq U(\log(A)). \tag{2.5.54}
\]
We give the following example to give an idea of the proof that follows. Set
\[
Z_1 := \bigcup_{i=1}^{\infty} [2i, 2i + 1) = [2, 3) \cup [4, 5) \cup [6, 7) \cup \ldots,
\]
\[
Z_2 := \bigcup_{i=1}^{\infty} [4i + 1, 4i + 2) = [5, 6) \cup [9, 10) \cup [13, 14) \cup \ldots,
\]
\[
Z_3 := \bigcup_{i=1}^{\infty} [4i + 3, 4i + 4) = [7, 8) \cup [11, 12) \cup [15, 16) \cup \ldots.
\]

Note that \( Z_1, Z_2, Z_3 \in C \) are pairwise disjoint. Now, we set
\[
A' := Z_1 \circ A + Z_2 \circ A + Z_3 \circ A. \tag{2.5.55}
\]
Observe that for \( j \geq 3 \)
\[
\tau_{A'}(2, j) = \frac{1}{2} (\tau_A(2, j - 1) + \tau_A(2, j - 2)). \tag{2.5.56}
\]
So we constructed a set \( A' \) that on each interval \([2^{j-1}, 2^j)\) with \( j \geq 3 \) has an average that equals the average of the averages of \( A \) on two consecutive intervals.

**Step 1** We construct a \( \hat{A} \in F \).

Fix \( C > 1 \) and \( n \in \mathbb{N} \). We split up \([C^{j-1}, C^j)\) into intervals of length 1 plus a remainder interval for every \( j \). Set for \( j \in \mathbb{N} \)
\[
N_j := [C^{j-1}(C - 1)] \tag{2.5.57}
\]
and for \( j \in \mathbb{N} \) and \( l \in \{1, \ldots, N_j\} \)
\[
I(j, l) := [C^{j-1} + l - 1, C^{j-1} + l), \tag{2.5.58}
\]
so that for every \( j \in \mathbb{N} \) we have
\[
[C^{j-1}, C^j) = [C^{j-1} + N_j, C^j) \cup \bigcup_{l=1}^{N_j} I(j, l). \tag{2.5.59}
\]
Choose \( u \in \mathbb{N} \) such that for every \( j \in \mathbb{N} \) we have
\[
N_{u+j} \geq (N_{n+j} + 1) \sum_{p=0}^{n} [C^p], \tag{2.5.60}
\]
which can be done since \( N_j \) is asymptotically equivalent with \( C^{j-1}(C - 1) \). For \( p \in \{0, \ldots, n\} \), \( k \in \{1, \ldots, [C^p]\} \) and \( j \in \mathbb{N} \) we set
\[
I^{p,k}(j) := \bigcup_l I \left( j, l \sum_{i=0}^{n} [C^i] + \sum_{i=0}^{p-1} [C^i] + k \right). \tag{2.5.61}
\]
For \( l \leq T \) set
\[
\zeta(l, T) := \bigcup_{i=0}^{[l]-1} \left( \frac{T_i}{|l|}, \frac{T_i + l}{|l|} \right) \tag{2.5.62}
\]
that ‘evenly’ distributes mass $l$ over the interval $[0, T)$. Note that (2.5.60) guarantees that
\[ m(I^{p,k}(u + j)) \geq C^{n+j-1}(C-1) \geq C^{n-p+j-1}(C-1) \quad (2.5.63) \]
for every $j \in \mathbb{N}$, so
\[ Z(p,k) := \bigcup_{j=1}^{\infty} (I^{p,k}(u + j) \circ \zeta (C^{n-p+j-1}(C-1), m(I^{p,k}(u + j)))) \quad (2.5.64) \]
is well defined. Note that by construction $Z(p,k) \in \mathcal{C}$ and
\[ m(Z(p,k) \cap I^{p,k}(u + j)) = C^{n-p+j-1}(C-1). \quad (2.5.65) \]
From this it directly follows that
\[ \lambda(Z(p,k)) = \frac{C^n}{C^{p+u}}. \quad (2.5.66) \]

We now introduce
\[ \hat{A} := \bigcup_{p=0}^{n} \bigcup_{k=1}^{\lfloor C^p \rfloor} Z(p,k) \circ A. \quad (2.5.67) \]

Observe that all the $Z(p,k)$ are disjoint. So (WT1) and the fact that $\mathcal{F}$ is an $f$-system imply that $\hat{A} \in \mathcal{F}$.

**Step 2** We give an upperbound for $\mu(A)$ by first giving an upperbound for $\mu(\hat{A})$ and then relating $\mu(A)$ and $\mu(\hat{A})$.

A crucial property of $\hat{A}$ is that for $j \in \mathbb{N}$
\[ m([C^{u+j-1}, C^{u+j}] \cap \hat{A}) = \sum_{p=0}^{n} [C^n] m([C^{j+n-p-1}, C^{j+n-p}] \cap A). \quad (2.5.68) \]
Hence
\[ \tau_{\hat{A}}(C, u + j) = C^{n-u} \sum_{p=0}^{n} [C^p] C^{-p} \tau_A(C, j + n - p) \]
\[ \leq C^{n-u} \sum_{p=0}^{n} \tau_A(C, j + n - p) \quad (2.5.69) \]
\[ \leq C^{n-u} \sup_{k \in \mathbb{N}} \sum_{j=k}^{k+n} \tau_A(C, j). \]

We apply Lemma 2.5.4 for $\hat{A}$ and find with (2.5.69) that
\[ \mu(\hat{A}) \leq C^{n-u+1} \sup_{k \in \mathbb{N}} \sum_{j=k}^{k+n} \tau_A(C, j). \quad (2.5.70) \]
The weak thinnability of \( \mu \) gives that
\[
\mu(\hat{A}) = \sum_{p=0}^{n} \sum_{k=1}^{a_p} \mu(Z(p,k))\mu(A) = \mu(A)C^{n-u} \sum_{p=0}^{n} \lfloor C^p \rfloor C^{-p}.
\] (2.5.71)

Combining (2.5.70) and (2.5.71) gives
\[
\mu(A) = C^{u-n} \frac{C^{u-n}}{\sum_{p=0}^{n} (C^p - 1) C^{-p} \mu(\hat{A})}
\]
\[
\leq \frac{C^{u-n}}{n+1} \frac{1}{1-1/C} \mu(\hat{A})
\]
\[
\leq C \frac{n+1}{n+1-1/C} \sup_{k \in \mathbb{N}} \frac{1}{n+1} \sum_{j=k}^{k+n} \tau_A(C, j).
\] (2.5.72)

**Step 3** We take limits in (2.5.72).

Unfix \( n \) and \( C \). We first take the limit superior for \( n \to \infty \) in (2.5.72), giving
\[
\mu(A) \leq C \limsup_{n \to \infty} \frac{1}{n+1} \sum_{j=k}^{k+n} \tau_A(C, j) = CU^*(C, A).
\] (2.5.73)

Then we take the limit superior for \( C \downarrow 1 \) and find by Lemma 2.5.3 that
\[
\mu(A) \leq \limsup_{C \downarrow 1} U^*(C, A) = U(\log(A)).
\] (2.5.74)

The lower bound we can now easily obtain by applying our upper bound for the complement of \( A \). Doing this, we see that
\[
1 - \mu(A) = \mu(A^c)
\]
\[
\leq U(\log(A^c))
\]
\[
= 1 - L(\log(A)),
\] (2.5.75)
giving that \( \mu(A) \geq L(\log(A)) \).

**Proof of Theorem 2.3.6.** We prove the contrapositive. Let \((\mathcal{F}, \mu)\) be a WTP with \( \mathcal{F} \setminus \mathcal{A}^{uni} \neq \emptyset \). Let \( A \in \mathcal{F} \setminus \mathcal{A}^{uni} \). By Lemma 2.5.2, this means that there is a \((s, f)\) in \( \mathcal{P} \) such that
\[
I := \liminf_{n \to \infty} \xi_A(s_n, f_n) \neq \limsup_{n \to \infty} \xi_A(s_n, f_n) =: S.
\]

Clearly, we can find \( m, l \in \mathbb{N}^\infty \) such that \( \xi_A(s_{m_n}, f_{m_n}) \) tends to \( I \) and \( \xi_A(s_{l_n}, f_{l_n}) \) tends to \( S \). Now set \( s_n' := s_{m_n}, f_n' := f_{m_n}, s_n'' := s_{l_n} \) and \( f_n'' := f_{l_n} \). Then we see that \( A \in \mathcal{A}^{s', f'} \) and \( A \in \mathcal{A}^{s'', f''} \) with
\[
\alpha^{s', f'}(A) = I \quad \text{and} \quad \alpha^{s'', f''}(A) = S.
\]
In the proof of Theorem 2.3.3 we showed that \((A^{s',j'},\alpha^{s',j'})\) and \((A^{s'',j''},\alpha^{s'',j''})\) are both in WTC. Thus \((\mathcal{F},\mu)\) is not canonical with respect to WT and in case \(\mu\) is coherent, \((\mathcal{F},\mu)\) is not canonical with respect to WTC.

\[\square\]

**Proof of Theorem 2.4.2.** We give a proof along the lines of Mattila [38, p. 45], with small adaptations for completeness and more generality.

Let \((X,d)\) be a metric space and \(\nu_1,\nu_2\) uniform measures on \(X\). Write \(h_1(r) := \nu_1(B(x,r))\) and \(h_2(r) := \nu_2(B(x,r))\) for \(r > 0\), which are well defined since \(\nu_1\) and \(\nu_2\) are uniform. We show that \(\nu_1 = c\nu_2\) for some \(c > 0\). It is sufficient to show that \(\nu_1 = c\nu_2\) on all open sets.

First let \(A\) be an open set of \((X,d)\) with \(\nu_1(A) < \infty\) and \(\nu_2(A) < \infty\). Suppose that \(r > 0\) is such that \(h_2\) is continuous in \(r\). Then

\[
|\nu_2(A \cap B(x,r)) - \nu_2(A \cap B(y,r))| \leq \nu_2(B(x,r) \triangle B(y,r)) \\
\leq \nu_2(B(x,r + d(x,y)) \setminus B(x,r)) \\
= h_2(r + d(x,y)) - h_2(r).
\]

Hence \(x \mapsto \nu_2(A \cap B(x,r))\) is a continuous mapping from \(X\) to \([0,\infty)\). Since \(h_2\) is nondecreasing, it can have at most countable many discontinuities. So we can choose \(r_1, r_2, r_3, \ldots\) such that \(\lim_{n \to \infty} r_n = 0\) and \(h_2\) is continuous in every \(r_n\).

For \(n \in \mathbb{N}\) let \(f_n : X \to [0,1]\) be given by

\[
f_n(x) := 1_A(x) \frac{\nu_2(A \cap B(x,r_n))}{h_2(r_n)}.
\]

Notice that by our previous observation \(f_n\) is continuous on \(A\), hence \(f_n\) is measurable. Because \(A\) is open, we have \(\lim_{n \to \infty} f_n(x) = 1\) for every \(x \in A\). With Fatou’s Lemma we find

\[
\nu_1(A) = \int_A \lim_{n \to \infty} f_n(x) \nu_1(dx) \\
\leq \liminf_{n \to \infty} \frac{1}{h_2(r_n)} \int_A \nu_2(A \cap B(x,r_n)) \nu_1(dx) \\
\leq \liminf_{n \to \infty} \frac{1}{h_2(r_n)} \int_X \int_A 1_{B(x,r_n)}(y) \nu_2(dy) \nu_1(dx).
\]

Note that any uniform measure is \(\sigma\)-finite. Applying Fubini’s theorem we obtain

\[
\nu_1(A) \leq \liminf_{n \to \infty} \frac{1}{h_2(r_n)} \int_X \int_A 1_{B(x,r_n)}(y) \nu_1(dx) \nu_2(dy) \\
= \liminf_{n \to \infty} \frac{1}{h_2(r_n)} \int_A \nu_1(B(y,r_n)) \nu_2(dy) \\
= \liminf_{n \to \infty} \frac{h_1(r_n)}{h_2(r_n)} \nu_2(A).
\]

34
By interchanging $\nu_1$ and $\nu_2$ we get

$$\nu_2(A) \leq \lim \inf_{n \to \infty} \frac{h_2(r_n)}{h_1(r_n)} \nu_1(A). \quad (2.5.80)$$

Note that $\lim \inf_{n \to \infty} \frac{h_2(r_n)}{h_1(r_n)} > 0$ since (2.5.80) would otherwise imply that all open balls are null sets. So we may rewrite (2.5.80) as

$$\nu_1(A) \geq \frac{1}{\lim \inf_{n \to \infty} \frac{h_2(r_n)}{h_1(r_n)}} \nu_2(A)$$

$$= \lim \sup_{n \to \infty} \frac{h_1(r_n)}{h_2(r_n)} \nu_2(A) \quad (2.5.81)$$

$$\geq \lim \inf_{n \to \infty} \frac{h_1(r_n)}{h_2(r_n)} \nu_2(A).$$

Hence $\nu_1(A) = c \nu_2(A)$ with

$$c := \lim \inf_{n \to \infty} \frac{h_1(r_n)}{h_2(r_n)} > 0. \quad (2.5.82)$$

Now let $A$ be any open set of $(X,d)$. Let $x \in X$ and set $A_n := A \cap B(x,n)$ for $n \in \mathbb{N}$. Note that $A_n$ is open with $\nu_1(A_n) \leq \nu_1(B(x,n)) < \infty$ and $\nu_2(A_n) \leq \nu_2(B(x,n)) < \infty$. Hence, by the first part of the proof, we find $\nu_1(A_n) = c \nu_2(A_n)$. But then

$$\nu_1(A) = \lim_{n \to \infty} \nu_1(A_n) = \lim_{n \to \infty} c \nu_2(A_n) = c \nu_2(A). \quad (2.5.83)$$

\[ \square \]

\textbf{Proof of Theorem 2.4.3.} Fix $A \in \mathcal{L}(X)$ and $x, y \in X$. By (2.4.3) we have

$$\lim_{u \to \infty} \frac{h(r^+(u))}{h(r^-(u))} = \lim_{u \to \infty} \frac{h(r^-(u) + 1)}{h(r^-(u))} = \lim_{r \to \infty} \frac{h(r + 1)}{h(r)} = 1. \quad (2.5.84)$$

Hence

$$\frac{\nu(B(x,r^-(u)) \cap A)}{h(r^-)} \sim \frac{\nu(B(x,r^+(u)) \cap A)}{h(r^+)} \quad (2.5.85)$$

Observe that for any $r \in [0, \infty)$ we have

$$\left| \frac{\nu(B(x,r) \cap A)}{h(r)} - \frac{\nu(B(y,r) \cap A)}{h(r)} \right| = \frac{1}{h(r)} \left| \nu(A \cap B(x,r)) - \nu(A \cap B(y,r)) \right|$$

$$\leq \frac{1}{h(r)} \nu(B(x,r) \Delta B(y,r))$$

$$\leq \frac{1}{h(r)} \nu(B(x,r + d(x,y)) \setminus B(y,r))$$

$$= \frac{h(r + d(x,y)) - h(r)}{h(r)}. \quad (2.5.86)$$
By (2.4.3), it follows that
\[
\frac{\tilde{\nu}(B(x, r^{-}(u)) \cap A)}{h(r^{-}(u))} \sim \frac{\tilde{\nu}(B(y, r^{-}(u)) \cap A)}{h(r^{-}(u))} \quad (2.5.87)
\]
Combining (2.5.85) and (2.5.87) gives the desired result.

Proof of Theorem 2.4.5. Suppose \( X = \mathbb{R}^n \) with \( d \) Euclidean distance. Set
\[
\delta_n := \frac{2\pi^{n/2}}{\Gamma(n/2)}. \quad (2.5.88)
\]
Let \( \nu \) be the Borel measure on \( \mathbb{R}^n \). Note that \( h(r) = n^{-1}\delta_n r^n \). If we set \( u = \sqrt[n]{n\delta_n^{-1}} y \), then
\[
\int_x^D \frac{\tilde{\rho}_A(y)}{y} dy = \int_x^D \delta_n \frac{\sqrt[n]{n\delta_n^{-1}}}{y^2} \int_0^{\sqrt[n]{n\delta_n^{-1}}y} r^{n-1} K_A(r) dr dy
= \int \frac{\sqrt[n]{n\delta_n^{-1}}}{\sqrt[n]{n\delta_n^{-1}}x} \int_0^u r^{n-1} K_A(r) dr du. \quad (2.5.89)
\]
Now observe that by partial integration
\[
\int \frac{n^2}{u^{n+1}} \int_0^u r^{n-1} K_A(r) dr du = -\frac{n}{u^n} \int_0^u r^{n-1} K_A(r) dr + n \int \frac{K_A(u)}{u} du. \quad (2.5.90)
\]
If we set for \( D, x \in (1, \infty) \)
\[
\zeta_A(D, x) := -\frac{1}{\log(D) u^n} \int_0^u r^{n-1} K_A(r) dr \bigg|_{u=x}^{x^D}, \quad (2.5.91)
\]
then
\[
\bar{\zeta}_A(D^n, n^{-1}\delta_n x^n) = \zeta_A(D, x) + \frac{1}{\log(D)} \int_x^{x^D} \frac{K_A(u)}{u} du. \quad (2.5.92)
\]
Since \(|\zeta_A(D, x)| \leq \frac{1}{\log(D)}\), the desired result follows.

2.6 Discussion

Thinnability

Suppose that in Definition 2.2.2 we replace (WT1) by the property that for every \( A, B \in F \) we have \( A \circ B \in F \) and \( \mu(A \circ B) = \mu(A)\mu(B) \). Instead of weak thinnability, we call this thinnability. Now consider the set
\[
A = \bigcup_{n=0}^{\infty} [2^{2n}, 2^{2n+1}). \quad (2.6.1)
\]
2.6. DISCUSSION

We have \( A, A^c \in \mathcal{A}^{uni} \) with \( \alpha(A) = \alpha(A^c) = 1/2 \). But also, we have \( A \circ A^c \in \mathcal{A}^{uni} \) with

\[
\alpha(A \circ A^c) = \alpha \left( \bigcup_{n=0}^{\infty} \left[ 2^{2n} + \frac{1}{6} \cdot 2^{2n}, 2^{2n} + \frac{2}{3} \cdot 2^{2n} \right] \right) = \lambda \left( \bigcup_{n=0}^{\infty} [2n \log(2) + \log(1 + 1/6), 2n \log(2) + \log(1 + 1/3)] \right) = \frac{\log(1 + 2/3) - \log(1 + 1/6)}{2 \log(2)} \neq \frac{1}{4} = \alpha(A)\alpha(A^c). \tag{2.6.2}
\]

So \( (\mathcal{A}^{uni}, \alpha) \) is not a thinnable pair. Since every thinnable pair is also a WTP, by Theorem 2.3.5 we see that a thinnable probability measure on \( \mathcal{A}^{uni} \), does not exist.

Notice that we are not necessarily looking for the strongest notion of uniformity, but for a notion that allows for a canonical probability pair with a “big” \( f \)-system. This is the reason why we are interested in weak thinnability rather than thinnability. There may, of course, be other notions of uniformity that lead to canonical pairs with bigger \( f \)-systems than \( \mathcal{A}^{uni} \). At this point, we can not see any convincing motivation for such notions.

Weak thinnability

In this chapter, we only studied the notion of weak thinnability from the interest in canonical probability pairs. There are, however, interesting open questions about the property of weak thinnability itself, that we did not address here. Some examples are:

- Is every probability pair that extends \( (\mathcal{A}^{uni}, \alpha) \) a WTP?
- Is every WTP coherent?
- Can every WTP be extended to a WTP with \( \mathcal{M} \) as \( f \)-system?
- How do the sets \( \{\mu(A) : (\mathcal{F}, \mu) \in WT \text{ and } A \in \mathcal{F}\} \) and \( \{\mu(A) : (\mathcal{F}, \mu) \in WTC \text{ and } A \in \mathcal{F}\} \) look like for \( A \notin \mathcal{A}^{uni} \)?
- Is (WT2) redundant? If no, what probability pairs are not a WTP, but do satisfy (WT1) and (WT3)?
- How does weak thinnability relate to the property \( \mu(cA) = \mu(A) \), where \( cA := \{ca : a \in A\} \) and \( c > 1 \)?
Size of $\mathcal{A}^{\text{uni}}$

A typical example of a set in $\mathcal{M}$ that does not have natural density, but is assigned a probability by $\alpha$, is

$$A := \bigcup_{n=0}^{\infty} [e^{2n}, e^{2n+1}),$$

for which we have $\alpha(A) = 1/2$. It is, however, unclear how “many” of such sets there are, i.e. how much “bigger” the $f$-system $\mathcal{A}^{\text{uni}}$ is than $\mathcal{C}$ and how much “smaller” it is than $\mathcal{M}$. If we could construct a uniform probability measure on $\mathcal{M}$ by the method of Section 2.4, we could determine the probability of $\mathcal{A}^{\text{uni}}$ if $\mathcal{A}^{\text{uni}} \in \mathcal{A}^{\text{uni}}(\mathcal{M})$. To construct such a probability measure, we need to equip $\mathcal{M}$ with a metric $d$ such that $(\mathcal{M}, d)$ has a uniform measure. It is, however, not at all clear how we should choose $d$. So at this point, it is not clear if there is a useful way of measuring the collections $\mathcal{C}$ and $\mathcal{A}^{\text{uni}}$. 
CHAPTER 3

The supertask of an infinite lottery

3.1 The supertask

We start to give our own compressed description of the supertask:

An infinite number of gods, all identified by an unique natural number, gather to hold a lottery. To do that, for every natural number \(k\), they produce a ball of size \(2^{-k}\) (so that the total volume of the balls is finite) with the number \(k\) and put them in an urn of finite size. Then they follow the following instruction: at time \(t_k = 1/k\) god \(k\) takes the urn. If god \(k\) finds \(k + 1\) balls in the urn, he randomly removes one. If god \(k\) not finds \(k + 1\) balls in the urn, he first empties the urn and then puts balls 1 to \(k\) in the urn.\(^1\)

In the version of Hansen god \(k\) removes \(2^{k-1}\) balls. We simplified this by letting each god remove only one ball. This only changes the supertask cosmetically, because the only crucial part is that every god removes a finite number of balls. More precisely, the task of god \(k\) in the description of Hansen (reducing the number of balls from \(2^k\) to \(2^{k-1}\)) is precisely the task that gods \(2^k - 1\) to \(2^{k-1}\) perform together in our description. The main reason for this adjustment is that it turns out to simplify the mathematics in Section 3.2.

We write an outcome of the process as \{3\}, \{3, 7\}, \{3, 4, 7\}, ..., where every set in the sequence adds one (new) natural number to its precursor and every natural number is added at some point in the sequence. The first set represents the set of balls after god 1 has removed a ball, followed by the set of balls after god 2 has removed a ball, etcetera. So in this case god 2 found balls 3, 4 and 7 and removed ball 4. After that, god 1 removed ball 7 leaving ball 3 as the final ball. On the outcome space consisting of all such sequences, we want to define our probability measure. The constraint on this probability measure that we get.

\(^1\)The explicit instruction for the case that god \(k\) finds not \(k + 1\) balls is given to ensure that this part of the instruction is actually never used: since the instruction is also followed by god \(k + 1\), god \(k\) necessarily finds \(k + 1\) balls in the urn.
from the supertask is that for every $k$, the conditional probability of the set of balls after god $k$ being in some collection $S$ (consisting of sets of size $k$), given that after god $k + 1$ there are precisely $j$ balls that upon removal give an element of $S$, is $j/(k + 1)$.

Within our model, we prove that for any $p \in [0, 1]$ and any set $A \subseteq \mathbb{N}$ such that $A$ and $A^c$ are infinite, there is a probability measure satisfying the constraint, assigning probability $p$ to the final ball being in $A$. To show the idea behind the proof, we fix the outcome

$$\{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots \tag{3.1.1}$$

Consider the probability measure that assigns probability $1/2$ to the two outcomes that are identical to our fixed outcome up to god 1, i.e.

$$\{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots \text{ and}$$
$$\{2\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots \tag{3.1.2}$$

Since what happens up to god 1 is deterministic ($\{1, 2\}, \{1, 2, 3\}, \ldots$ with probability 1) under this probability measure, the constraint for $k = 1$ entails that god 1 removes ball 1 or 2 both with probability $1/2$, which is exactly what happens. So this probability measure satisfies the constraint for $k = 1$, but not for all other $k$. Now we consider the probability measure that assigns probability $1/6$ to the outcomes that are identical to our fixed outcome up to god 2, i.e.

$$\{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots,$$
$$\{2\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots,$$
$$\{2\}, \{2, 3\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots,$$
$$\{3\}, \{2, 3\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots,$$
$$\{1\}, \{1, 3\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots \text{ and}$$
$$\{3\}, \{1, 3\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \ldots \tag{3.1.3}$$

What happens up to god 2 is now deterministic and since every possibility for the actions of gods 1 and 2 have the same probability, the constraint is satisfied for $k = 1$ and $k = 2$. By continuing this process, we get probability measures that satisfy the constraint for more and more $k$. In this way, in the limit, we get a probability function that satisfies the constraint for every $k$ (see Section 3.2 for the precise construction of this limit probability measure). If what happens up to god $k$ is deterministic and everything after that happens with the same probability, then the probability of the final ball being even is given by the fraction of even balls after god $k + 1$ has removed a ball. Since our fixed outcome is $\{1\}, \{1, 2\}, \{1, 2, 3\}, \ldots$, the probability of the final ball being even converges to $1/2$ as $k$ grows to infinity. Hence the probability that the final ball is even is $1/2$.

The crux is that we can start with any fixed outcome. Suppose we want a probability measure that assigns probability $1/3$ to the final ball being even. Then we construct an appropriate fixed outcome in the following way. We start with
3.2. THE MATHEMATICAL MODEL

The mathematical model

Because the fraction of even numbers is now 0, we add an even number for the next element of the sequence: \( \{1\}, \{1, 2\} \). The fraction of even numbers is now 1/2, which is greater than 1/3, so we add an odd number: \( \{1\}, \{1, 2\}, \{1, 2, 3\} \). Now the fraction of even numbers is exactly 1/3, so we can add any number: \( \{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\} \). Because the fraction is now again greater than 1/3, we add an odd number: \( \{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\} \).

3.2 The mathematical model

In this section, we make everything mathematically precise. To do that, we start with some definitions. Write for \( k \in \mathbb{N} \)

\[
\mathbb{N}_k := \{ A \subseteq \mathbb{N} : \left| A \right| = k \}
\]

\[
\Omega_k := \{ (B_k, B_{k+1}, \ldots) : B_j \in \mathbb{N}_j, B_j \subset B_{j+1} \text{ and } \bigcup_j B_j = \mathbb{N} \}.
\]

(3.2.1)

Elements of \( \Omega_k \) represent what happens up until god \( k \). The space \( \Omega := \Omega_1 \) is our outcome space. On this space, we define the following two functions. For
$B = (B_1, B_2, \ldots) \in \Omega$ with $B_1 = \{n\}$, we define
\[
H_k(B) := (B_k, B_{k+1}, \ldots) \in \Omega_{k+1}
\]
\[
R(B) := n.
\]
(3.2.2)

The function $H_k$ maps an outcome to what happened up until god $k$ and $R$ gives us the remaining ball at the end of the process.

For $B = (B_{k+1}, B_{k+2}, \ldots) \in \Omega_{k+1}, S \subseteq \Omega_k$ and $1 \leq j \leq k + 1$ we set
\[
N(S; B) := |\{b \in B_{k+1} : (B_{k+1} \setminus \{b\}, B_{k+1}, B_{k+2}, \ldots) \in S\}|
\]
\[
P_j(S) := \{B \in \Omega_{k+1} : N(S; B) = j\} \subseteq \Omega_{k+1}
\]
(3.2.3)

In words, the set $P_j(S)$ gives us everything that can happen up until god $k + 1$ such that there are precisely $j$ of the $k + 1$ balls that upon removal give us an element of $S$.

We consider the measure space $(\Omega, \mathcal{P}(\Omega))$. On this measure space we consider probability measures as candidates. The constraint on probability measure $\mu$ that we obtain from the description of the supertask is that for every $k$, $S \subseteq \Omega_k$ and $T \subseteq P_j(S)$ such that $\mu(H_{k+1} \in T) > 0$ we have
\[
\mu(H_k \in S | H_{k+1} \in T) = \frac{j}{k + 1}.
\]
(3.2.4)

We note that using (3.2.4) we can determine $\mu(H_k \in S | H_{k+1} \in T)$ for every $k$, $S \subseteq \Omega_k$ and $T \subseteq \Omega_{k+1}$ such that $\mu(H_{k+1} \in T) > 0$, by writing
\[
\mu(H_k \in S | H_{k+1} \in T) = \sum_{j=0}^{k+1} \frac{\mu(H_k \in S; H_{k+1} \in T \cap P_j(S))}{\mu(H_{k+1} \in T)}
\]
\[
= \sum_{j=1}^{k+1} \frac{j}{k + 1} \frac{\mu(H_{k+1} \in T \cap P_j(S))}{\mu(H_{k+1} \in T)}.
\]
(3.2.5)

We write $\mathcal{M}$ for the set of probability measures that satisfy (3.2.4) and prove the following theorem.

**Theorem 3.2.1.** Let $A \subseteq \mathbb{N}$. Then
\[
\{\mu(R \in A) : \mu \in \mathcal{M}\} = \begin{cases}
\{0\} & \text{if } A \text{ is finite} \\
\{1\} & \text{if } A^c \text{ is finite} \\
[0, 1] & \text{otherwise}
\end{cases}
\]
(3.2.6)

We present the proof of Theorem 3.2.1 in three steps.

**Step 1** We construct for every $Z \in \Omega$, a probability measure $\alpha_Z \in \mathcal{M}$ based on $Z$.

To construct a probability measure such that (3.2.4) is satisfied, we fix a $Z := (Z_1, Z_2, Z_3, \ldots) \in \Omega$ and set
\[
F_n := \{H_n = (Z_n, Z_{n+1}, \ldots)\} \subseteq \Omega.
\]
(3.2.7)
3.2. THE MATHEMATICAL MODEL

For $S \subseteq \Omega = \Omega_1$ we define the sequence $x(S) \in [0, 1]^\infty$ by

$$x_n(S) := \frac{|S \cap F_n|}{|F_n|} = \frac{|S \cap F_n|}{n!}, \quad (3.2.8)$$

which gives the density of $S$ with respect to $F_n$. Now, we let $n$ go to infinity. To do that, we extend the limit operator on all convergent sequences to a linear operator $L$ on all bounded sequences. Such an extension can be constructed by using a free ultrafilter on $\mathbb{N}$. The existence of such a free ultrafilter is guaranteed by the Boolean Prime Ideal Theorem, which cannot be proven in ZF set theory, but is weaker than the axiom of choice [25].

We define $\alpha_Z : \mathcal{P}(\Omega) \to [0, 1]$ by

$$\alpha_Z(S) := L(x(S)). \quad (3.2.9)$$

Now we show that $\alpha_Z \in \mathcal{M}$. For $S \subseteq \Omega_k$, $T \subseteq P_j(S)$ and $k < n$, we have

$$j|\{H_{k+1} \in T\} \cap F_n| = \sum_{B \in T} N(S, B)|\{H_{k+1} = B\} \cap F_n| = \sum_{B' \in S} \sum_{B \in T} N(\{B'\}, B)|\{H_{k+1} = B\} \cap F_n| = \sum_{B' \in S} \sum_{B \in T} (k+1)|\{H_k = B'; H_{k+1} = B\} \cap F_n| = (k+1)|\{H_k \in S; H_{k+1} \in T\} \cap F_n|. \quad (3.2.10)$$

Dividing by $n!(k+1)$ on both sides in (3.2.10) we find that

$$x_n(\{H_k \in S; H_{k+1} \in T\}) = \frac{j}{k+1} x_n(\{H_{k+1} \in T\}). \quad (3.2.11)$$

Since $L$ is linear, we get that

$$\frac{\alpha_Z(H_k \in S; H_{k+1} \in T)}{\alpha_Z(H_{k+1} \in T)} = \frac{j}{k+1} \quad (3.2.12)$$

for every $k$, $S \subseteq \Omega_k$ and $T \subseteq P_j(S)$ with $\alpha_Z(H_{k+1} \in T) > 0$, which is what we wanted to show.

**Step 2** We show that for every infinite $A \subseteq \mathbb{N}$ such that $A^c$ is also infinite and every $p \in [0, 1]$ there is a $Z \in \Omega$ such that $\alpha_Z(R \in A) = p$.

Let $A \subseteq \mathbb{N}$ be such that both $A$ and $A^c$ are infinite. Let $p \in [0, 1]$ be given. If we choose $Z \in \Omega$ such that

$$\frac{|Z_k \cap A|}{k} \to p \quad (3.2.13)$$

2The existence of a atomfree or nonprincipal (i.e. every singleton has measure zero) probability measure defined on the power set of $\mathbb{N}$ cannot be established in ZF alone [59]. Consequently, a version of the axiom of choice is always necessary to construct an element of $\mathcal{M}$.
as $k \to \infty$, then $\alpha_Z(R \in A) = p$. We construct $Z$ by recursion, using as basic rule that we add an element from $A$ if the density is too high and add an element from $A^c$ if the density is too low. To make sure that we add every natural number somewhere in the process and thus $Z \in \Omega$, we make one exception to this rule: for every $j$, we add an element from $A$ in the $j^2$-th step and an element from $A^c$ in the $j^2 + 1$-th step. This exception, however, does not influence the limit density we obtain since the fraction of square numbers in $[0,k]$ converges to zero as $k \to \infty$. We set

$$Z_1 := \{1\}$$

$$a_k := \min A \cap Z_k^c$$

$$b_k := \min A^c \cap Z_k^c$$

$$Z_{k+1} := \begin{cases} 
Z_k \cup \{a_k\} & \text{if } k \notin \{j^2, j^2 + 1\} \text{ for some } j \text{ and } \frac{|Z_k \cap A|}{k} \leq p \\
Z_k \cup \{b_k\} & \text{if } k \notin \{j^2, j^2 + 1\} \text{ for some } j \text{ and } \frac{|Z_k \cap A|}{k} > p \\
Z_k \cup \{a_k\} & \text{if } k = j^2 \text{ for some } j \\
Z_k \cup \{b_k\} & \text{if } k = j^2 + 1 \text{ for some } j 
\end{cases} \quad (3.2.14)$$

Notice that $a_k$ and $b_k$ are well defined for every $k$ because $A$ and $A^c$ are both infinite. Also note that $Z = (Z_1, Z_2, Z_3, \ldots) \in \Omega$ and (3.2.13) is satisfied, as desired.

**Step 3** We show that for all $\mu \in \mathcal{M}$ and finite $A \subseteq \mathbb{N}$ we have $\mu(R \in A) = 0$.

Let $a \in \mathbb{N}$ and $\mu \in \mathcal{M} \neq \emptyset$. If we take

$$S_k := \{(B_k, B_{k+1}, \ldots) \in \Omega_k : a \in B_k\}, \quad (3.2.15)$$

we get from (3.2.4) that

$$\mu(H_k \in S_k) = \frac{k}{k+1} \mu(H_{k+1} \in S_{k+1}). \quad (3.2.16)$$

This implies that $\mu(R = a) = \mu(H_1 \in S_1) = \frac{1}{k} \mu(H_k \in S_k)$ for every $k$, so $\mu(R = a) \leq \frac{1}{k}$ for every $k$. Hence $\mu(R = a) = 0$. Since $\mu$ is finitely additive, we have $\mu(R \in A) = 0$ for every finite $A \subseteq \mathbb{N}$.

### 3.3 Discussion

We have given a mathematical model for the supertask of Hansen. Our conclusions depend of course on the choice for this model: another model could give different results. We think, however, that the present model is not controversial. From our model, we have concluded that there are many probability measures that are consistent with the supertask. We have expressed this indeterminacy in terms of the different probabilities the probability measures assign to the final ball being in some subset $A \subseteq \mathbb{N}$. If $A$ is infinite and its complement is also infinite, then any probability $p$ is assigned to the final ball being in $A$ by some probability measure consistent with the supertask.
This conclusion shows that the description of the supertask as presented, just does not give us enough information to say anything about the probability of the final ball being in a set of the type described above. As a consequence, we should be very careful about this indeterminacy when thinking about the supertask. At one point, Hansen mentions the ‘empirical distribution’ obtained by the gods if they keep repeating the supertask. Assuming there is the same underlying probability measure every time they perform the supertask, of course one can, at least in theory, estimate this probability measure through performing experiments. However, the description of the supertask does not give enough information to decide what that underlying probability measure is.

One would hesitate to call a probability measure assigning probability 1 to the even numbers and probability 0 to the odd numbers, a ‘uniform’ probability measure. A reasonable notion of uniformity at least includes that every residue class modulo \(m\) has probability \(1/m\) (giving for \(m = 2\) that the even numbers have probability \(1/2\)).\(^3\) Our model shows, however, that the final ball being even can get any probability. This means that the notion of uniformity emerging from the supertask is weaker than any reasonable notion of uniformity.

Our analysis also shows that for any \(n \in \mathbb{N}\), a probability measure consistent with the supertask necessarily has to assign probability zero to the final ball being \(n\). Although this property of being atomfree is not a reasonable notion of ‘uniformity’, it is some form of ‘fairness’ nonetheless. This form of fairness is sufficient for the point Hansen wants to make about the supertask, namely that given any possible final ball \(k\), the probability is 1 that the final ball is bigger than \(k\). In other words: the final ball seems always unexpectedly low. This is certainly not an inconsistency, but Hansen does call it an ‘absurdity’.

We do not want to present an argument here about whether this is indeed an absurdity or not, but only to add clarity to that discussion by a proper analysis of the supertask. We do, however, want to point out that the alleged absurd property of the probability measure on the final ball is not unique. Also it is not restricted to probability measures. A typical \(\sigma\)-additive probability measure with this property is the following. Let \(\omega_1\) be the first uncountable ordinal number and let \(\mathcal{F}\) be the \(\sigma\)-algebra of subsets \(F \subseteq \omega_1\) such that either \(F\) or \(F^c\) is countable. Let \(\mu: \mathcal{F} \rightarrow [0, 1]\) be given by

\[
\mu(F) := \begin{cases} 
0 & \text{if } F \text{ is countable} \\
1 & \text{if } F \text{ is uncountable}
\end{cases} \quad (3.3.1)
\]

Then \(\mu\) is a \(\sigma\)-additive probability measure. If \(\mu\) models a lottery on \(\omega_1\), then completely analogous to the situation on \(\mathbb{N}\), for any outcome \(\alpha \in \omega_1\) we have \(\mu(\{\beta \in \omega_1 : \beta > \alpha\}) = 1\). Accepting the Axiom of Choice, this property also translates to picking a random number from \([0, 1]\). By the well ordering theorem, there exists a well ordering \(\preceq\) of \([0, 1]\). If we model the experiment with Lebesgue measure \(\lambda\) on \([0, 1]\), for any outcome \(x \in [0, 1]\), we again have \(\lambda(\{y \in [0, 1] : y \succ x\}) = 1\).

\(^3\)This is the weakest notion of uniformity considered by [44].
Part II

Epistemic probability
CHAPTER 4

Introduction

When interpreting probability, two broad categories of interpretations can be distinguished: aleatory probability and epistemic probability [24]. Aleatory probability concerns uncertainty as a result of a random experiment, such as flipping a coin or rolling a die. If (1) the experiment is repeatable and (2) for each event, the relative frequency of that event occurring converges to some limit frequency if we have more and more repetitions, then we can interpret the probability of an event $A$ occurring as the limit frequency of $A$.

Epistemic probability, on the other hand, concerns uncertainty as a result of lacking knowledge. Epistemic uncertainty can also stem from a random experiment in case we do not know the outcome of the experiment. But also something that is considered completely deterministic can be epistemically uncertain, either in general (e.g. whether there is life on Jupiter’s moon Europa) or to specific individuals (e.g. whether the Amazon or the Nile is longer, which may be known by some and not by others). In epistemic probability, the probability that $A$ occurs is usually interpreted as the degree of belief an agent has in $A$, making probability assignments subjective. The degree of belief in $A$ can be quantified as the price for which the agent is willing to buy a bet that pays out 1 if $A$ turns out to be true. We notice that this interpretation in terms of betting behavior is very general. In particular, if epistemic uncertainty is stemming from a random experiment, we do not need to assume repeatability or relative frequencies converging, to interpret probability in this way.

In modern mainstream probability theory, both aleatory and epistemic probability are almost exclusively based on the axioms of Kolmogorov [36], and we start our exposition with quoting these axioms. We fix a finite set $\Omega$ of outcomes.

**Definition 4.0.1** (Kolmogorov axioms). A function $P : 2^{\Omega} \to [0,1]$ is a probability distribution if it satisfies:

(P1) $P(\Omega) = 1$ and $P(\emptyset) = 0$;
(P2) For every every $A, B \subseteq \Omega$ we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$  (4.0.1)

From an aleatory point of view, the justification of the Kolomorov axioms is straightforward. Axiom (P1) is trivially true in the frequentistic interpretation. Axiom (P2) is also true in the frequentistic interpretation, since it then says that the frequency that either $A$ or $B$ occurs, equals the frequency that $A$ occurs plus the frequency that $B$ occurs, minus the frequency that $A$ and $B$ both occur.

From an epistemic point of view, the Dutch Book argument is used to justify the Kolmogorov axiomatization. The crux of this argument is the following. If an agent buys bets on $A_1, \ldots, A_N$ for respectively $P(A_1), \ldots, P(A_N)$ and sells bets on $B_1, \ldots, B_M$ for respectively $P(B_1), \ldots, P(B_M)$, then his or her net result, if the outcome is $\omega \in \Omega$, equals

$$\sum_{i=1}^{N} 1_{A_i}(\omega) - P(A_i) + \sum_{j=1}^{M} P(B_j) - 1_{B_j}(\omega).$$  (4.0.2)

The argument now says that reasonable agents avoid sure loss, by which we mean that they assign prices in such a way, that there is always an outcome in which they do not have a net loss. In other words, there must be a $\omega \in \Omega$ such that (4.0.2) is non-negative. The Dutch Book Theorem then shows us (see for instance De Finetti [19]) that this requirement is equivalent with the Kolmogorov axioms.

**Theorem 4.0.2** (Dutch Book Theorem). A function $P : 2^\Omega \rightarrow [0, 1]$ is a probability distribution if and only if for all $A_1, \ldots, A_N \subseteq \Omega$ and $B_1, \ldots, B_M \subseteq \Omega$, there is a $\omega \in \Omega$ such that

$$\sum_{i=1}^{N} 1_{A_i}(\omega) - P(A_i) + \sum_{j=1}^{M} P(B_j) - 1_{B_j}(\omega) \geq 0.$$  (4.0.3)

Despite this, degrees of belief cannot always be satisfactorily described with the classical axioms of Kolmogorov, something which has been recognized and confirmed by many researchers from different disciplines [47, 48, 49, 6, 61, 23, 12, 15]. These authors have argued that the classical axioms of probability are too restrictive and the most important point is that the classical theory cannot distinguish between lack of belief and disbelief. By disbelief we mean the willingness to bet on the negation of a proposition, whereas by lack of belief we mean not be willing to bet on the proposition. As Shafer [47] puts it, the classical theory does not allow one to withhold belief from a proposition without according that belief to the negation of the proposition. When we want to apply a theory of probabilities to, for example, legal issues, this becomes a relevant issue.\(^1\)

\(^{1}\)It has been debated for several decades as to what extent theories of probability, are useful and/or suitable for assessing the value of evidence in legal and forensic settings, see e.g. [20, 8, 42, 1, 51]. The current dominant view proposes that we should use the classical probability axioms in court.
Indeed, if certain exculpatory evidence in a case is dismissed, then this may result in less belief in the innocence of the suspect, but it may give no further indication for guilt.

A particular consequence of not distinguishing between lack of belief and disbelief, is the impossibility for an agent to suspend all judgment, that is, assigning $P(A) = P(A^c) = 0$. To see how this follows, suppose for a moment that we indeed do not distinguish between lack of belief and disbelief. Suspending judgment means not betting on $A$ and not betting on $A^c$, so under our assumption, this is the same as betting on both $A^c$ and $A$. Paying a high amount for the bets on both $A$ and $A^c$, however, is clearly unreasonable for an agent to do since either $A$ or $A^c$ must occur. Hence suspending judgment is indeed unreasonable if we do not distinguish between lack of belief and disbelief. But it can, however, be very reasonable for an agent to suspend judgment in some situations. Consider, for example, the situation in which an agent has no supporting evidence at all for either $A$ or $A^c$.

**Example 4.0.3.** Suppose that one morning, we find a big, sealed jar of gumballs\(^2\) at our doorstep. On it is a little note that says: ‘I can sell you a bet that pays out 100 euro if the number of gumballs is even and a bet that pays out 100 euro if the number of gumballs is odd. Just tell me your price.’ Without any reliable way to count the gumballs or having other evidence, we are completely ignorant about the number being even or odd. We decide that the cautious thing to do is not buy any of the two bets.

There are also more serious situations.

**Example 4.0.4.** In The Netherlands, a well known court case concerned a traffic accident caused by a car with two passengers. Although it was not disputed that the car caused the accident, it was unclear which of the two passengers was driving. In reality we know that one of the two passengers drove, but we are otherwise ignorant.

This example can be generalized into the well known and classical island problem:

**Example 4.0.5.** In the classical version of the island problem (see e.g. [53] and [4]) a crime has been committed on an island, making it a certainty that an inhabitant of the island committed it. In the absence of any further information, the classical point of view is to assign a uniform prior probability over all inhabitants concerning the question who is the culprit. The combination of assigning probability 1 to the collection of all inhabitants and probability 0 to each individual is impossible under the classical axioms of probability, although this may be exactly the prior one needs and wants to impose.

The classical solution to deal with this, is to impose a uniform prior on all individuals. The problem is, however, that although uniform distributions represent equal belief in all possibilities, they do not represent complete ignorance.

---

\(2\text{This example was inspired by the ‘gumball analogy’ of Matt Dillahunty}\)
When we look at the island problem in Example 4.0.5, it is simply the case that we do not have information pointing to any individual. We do have group information, but no individual information. With a uniform distribution over the group, you nevertheless make a statement about each individual. This is very relevant in legal cases since these are against individuals, not against a whole population.

That uniform distributions do not represent complete ignorance is also illustrated in the following example, where an agent that is ignorant and decides to buy bets according to uniform probabilities, is easily exploited.

**Example 4.0.6.** Suppose Agent A has some machine that takes a coin and puts the coin out with heads facing up and Agent A knows that the machine does this. Agent A tells another agent, Agent B, that (s)he is going to put a coin in the machine, but tells Agent B nothing more than that the machine will put the coin on one of its two sides. Now Agent A offers Agent B a bet that pays out 1 if the coin shows tails (knowing that this will never happen) and asks Agent B how much (s)he is willing to pay for that bet. Steeped in ignorance about what side of the coin will be facing up, Agent B decides that the best (s)he can do is split the probability mass fifty-fifty. So Agent B tells Agent A that (s)he is willing to pay 1/2 for the bet. Agent A happily takes 1/2 for the bet, knowing that (s)he never has to pay anything.

The crux of the problem here is that Agent B, by using a uniform distribution, implicitly uses information that (s)he does not have, namely that half of the time coin will face up and half of the time tails will face up. Of course this information is not correct and this is precisely the reason that Agent B is doomed to lose money.

Royall [43] formulates the problem of representing ignorance in yet another way: ‘The reason why pure ignorance cannot be represented by a probability distribution is that every probability distribution represents a particular state of uncertain knowledge; none represents the absence of knowledge’. His point is illustrated by the next example.

**Example 4.0.7.** Suppose a number between 1 and 100 is picked by some unknown mechanism. If uniform distributions represent ignorance, then it would be reasonable to pay 1/100 for the bet that \( n \) is the number picked. This would have as a consequence that we would pay 1/4 for the bet that the picked number is prime and 3/4 for the bet that the picked number is not prime, since 25 numbers between 1 and 100 are prime. We are, however, just as ignorant about which number is picked as we are about whether or not the picked number is prime. Hence, with the exact same argument, we could argue that we should pay 1/2 for both the bet that the picked number if prime and the bet that the picked number is not prime. In fact, we could have framed the whole experiment as picking between a prime and non-prime, suggesting an outcome space of only two elements, and have a probability of 1/2 to see a prime. In other words, the choice of the outcome space determines what ‘uniform’ means.

More generally, Example 4.0.7 suggests that any uninformative measure \( P \),
should have the property that, for any surjective \( X : \Omega \to \Omega' \), the map \( A' \mapsto P(X^{-1}(A')) \) for \( A' \subseteq \Omega' \), gives again an uninformative measure on \( \Omega' \). Clearly, uniform distributions do not have this property.

Further, we would like to point out that an uninformative prior actually leads to different results than a uniform prior in our theory, as we will see in Section 7.1. The fact that these priors lead to different results confirms that these priors are really distinct: a uniform prior is not a prior representing ignorance, and using a uniform prior does not lead to the same results as using a prior that does represent ignorance.

The examples suggest that the usual axioms of probability may not always be appropriate if we interpret probability as a degree of belief. This begs the question what goes wrong in the Dutch Book argument, that is supposed to justify these axioms. The answer lies in the assumption of the Dutch Book argument that the price for which an agent is willing to buy a certain bet is equal to the price for which an agent is willing to sell the same bet. The difference between buying and selling, however, is precisely the difference between lack of belief and disbelief. Willingness to buy a bet on \( A \) only for a small amount, indicates lack of belief in \( A \), whereas willingness to sell a bet on \( A \) for a small amount, indicates disbelief in \( A \). Hence the crux of the problem is right at the starting point of the Dutch Book argument, by not making the distinction between buying and selling prices.

One might defend the assumption of the Dutch Book argument, by pointing out that there are situations in which not betting on both \( A \) and \( A^c \) (suspending judgment) is either not allowed or has unacceptable consequences. In these situations an agent is forced to put one price on the bet on \( A \), meaning (s)he should be willing to both buy and sell the bet for that price. So, in such scenarios, the assumption of matching buying and selling prices seems justified. However, in a situation where an agent is forced to buy and sell the bet on \( A \) for the same price, although this price may give a measure of how likely the agent thinks \( A \) is relative to \( A^c \), it does not represent an agent's actual belief in \( A \). To put the same point differently: the price for which an agent is willing to buy a bet on \( A \) only is representative of his or her belief in \( A \) if (s)he is completely free in buying any bet.

These and similar considerations motivated Glenn Shafer [47] to introduce belief functions, which were supposed to better capture the nature of epistemic probability. To see how belief functions differ from classical probability distributions, we note that it is well known that (P2) can be expanded, for any collection of events \( A_1, \ldots, A_N \), into the well-known inclusion-exclusion formula

\[
P \left( \bigcup_{i=1}^{N} A_i \right) = \sum_{I \subseteq \{1, \ldots, N\}} (-1)^{|I|+1} P \left( \bigcap_{i \in I} A_i \right).
\] (4.0.4)

In the Shafer axiomatization of belief functions, (4.0.4) is replaced by a corresponding inequality.
Definition 4.0.8 (Shafer axioms). A function $\text{Bel} : 2^\Omega \to [0,1]$ is a belief function provided that

(B1) $\text{Bel}(\emptyset) = 0$ and $\text{Bel}(\Omega) = 1$;

(B2) For all $A_1, A_2, \ldots, A_N \subseteq \Omega$, we have

$$\text{Bel}\left( \bigcup_{i=1}^{N} A_i \right) \geq \sum_{\substack{I \subseteq \{1,\ldots,N\} \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bel}\left( \bigcap_{i \in I} A_i \right).$$

(4.0.5)

Immediately from the definition, there are a few things to notice. First, since (B2) is a relaxation of (P2), every probability distribution is a belief function. Secondly, belief functions do distinguish lack of belief in $A$, i.e. $1 - \text{Bel}(A)$, and disbelief in $A$, i.e. $\text{Bel}(A^c)$, since

$$\text{Bel}(A^c) \leq 1 - \text{Bel}(A),$$

(4.0.6)

but there is in general no equality. In particular, it is possible to represent complete ignorance with a belief function by setting $\text{Bel}(A) = 0$ for all $A \neq \Omega$ and $\text{Bel}(\Omega) = 1$. We notice that these belief functions have the property mentioned in relation to Example 4.0.7: for any surjective $X : \Omega \to \Omega'$, we find again that $\text{Bel}(X \in A') = 0$ for every $A' \neq \Omega'$ and $\text{Bel}(X \in \Omega') = 1$.

We want to emphasize that the underlying logic, that manifests itself in Definition 4.0.8 in the used set theory, is still classical. So although we can have $\text{Bel}(A) = \text{Bel}(A^c) = 0$, the law of the excluded middle still holds since $\text{Bel}(A \cup A^c) = 1$. In terms of Examples 4.0.4 and 4.0.5, this means that although we do not know who is guilty, we do know that someone is guilty. This means this approach is fundamentally different from approaches, e.g. Weatherson’s intuitionistic probability [63], in which the underlying logic is non-classical.$^3$

Shafer showed that belief functions have the following characterization.

Definition 4.0.9. A function $m : 2^\Omega \to [0,1]$ is called a basic belief assignment if $m(\emptyset) = 0$ and

$$\sum_{S \subseteq \Omega} m(S) = 1.$$

(4.0.7)

Theorem 4.0.10. A function $\text{Bel} : 2^\Omega \to [0,1]$ is a belief function if and only if there exists a basic belief assignment $m$ such that

$$\text{Bel}(A) = \sum_{S \subseteq A} m(S).$$

(4.0.8)

There is a one-to-one correspondence between belief functions and basic belief assignments, and $\text{Bel}$ is a probability distribution if and only if $m$ concentrates on singletons.

$^3$An analysis of the relation between intuitionism and belief functions is carried out by Iourinski [28].
We note that Theorem 4.0.10 not only tells us that a belief function of the form given in (4.0.8) is indeed a belief function, but also tells us that every belief function is of this form. However, we must be careful here. In the betting interpretation of belief functions that we have used so far, the basic belief assignment is only a (mathematical) derivative of an agent’s belief function, that itself is not interpreted.

In this thesis, we will consider a second interpretation, in which we interpret the basic belief assignment as the limiting frequency that our information is “$S$ occurs” (or “the truth is in $S$”). This means we consider $\mathcal{F} := 2^\Omega$ as an outcome space of a random experiment, where the outcome $\Omega \in \mathcal{F}$ represents complete ignorance (or no information) and a singleton $\{\omega\} \in \mathcal{F}$ represents a situation in which we know precisely what happens. Since a probability measure on a finite space is determined by its values on singletons, $m$ naturally defines a probability measure $P_m$ on $2^\mathcal{F}$ via

$$P_m(\{S\}) := m(S). \quad (4.0.9)$$

The maximum price for which, if the bet on $A$ is bought repeatedly in a series of repetitions of the experiment, it is guaranteed that there is no loss in the limit, is

$$P_m(\{S : S \subseteq A\}) = \sum_{S \subseteq A} m(S) = \text{Bel}(A), \quad (4.0.10)$$

which is the probability that the outcome of the experiment implies that $A$ occurs.

The interpretation of the belief in $A$ as the probability that $A$ is implied, goes back to Pearl [39], who calls this belief ‘the probability of provability’. We note, however, that the interpretation we offer here, is slightly different: we do not interpret belief in $A$ as the probability of provability, i.e., limiting frequency at which $A$ can be proved, but as the maximum price for which, if the bet on $A$ is bought repeatedly in a series of repetitions of the experiment, it is guaranteed that there is no loss. This subtle distinction plays a role when it comes to conditioning, as we will see in Section 5.1.

Shafer interprets beliefs as partial knowledge about an underlying probabilistic mechanism, and also explicitly connects belief functions to repeatable experiments, as becomes clear from the following citation [47], page 16:

“The chances governing an aleatory experiment may or may not coincide with our degrees of belief about the outcome of the experiment. If we know the chances, then we will surely adopt them as our degrees of belief. But if we do not know the chances, then it will be an extraordinary coincidence for our degrees of belief to be equal to them. The second case is the typical one. When we first conceive of an experiment as random, we typically have little idea about what chance density governs it. And though we may eventually form some opinion about the true chance density on the basis of actual observations of the experiment, we may never obtain any very exact or certain values for the true chances.”
Unlike Shafer, though, we do not want to assume that the underlying uncertainty in the experiment can or be satisfactorily modelled with a probability distribution. Apart from Shafer in his original 1976 book, other authors have also interpreted belief functions in a repeatable and frequentistic setting, but as far as we are aware, the interpretation we offer has not been written down explicitly. In [10] for instance, a method is developed to build a sequence of belief functions based on i.i.d. trials, where the question is what the most suitable belief function is about the underlying distribution given the outcomes of the first \( n \) experiment. See also [62] and [37] for frequentistic interpretations including laws of large numbers.

We will refer to this interpretation as the ‘frequentist interpretation’, since the values of \( m \) are interpreted as limiting frequencies. This label should, however, be used with some caution since only \( m \) is interpreted frequentistically: \( \text{Bel} \) is interpreted as a price, not as a limiting frequency. In addition, we would like to emphasize that this interpretation is still within the realm of epistemic probability, since \( m(S) \) is interpreted as the limiting frequency that our information is “\( S \) occurs”, which is of course not the same as the limiting frequency that \( S \) occurs.

We now give an example of this interpretation.

**Example 4.0.11.** Suppose Agent A flips two fair coins. Now Agent A agrees with Agent B that, if the second coin is heads, Agent A gets the opportunity to flip the first coin on the other side if (s)he wants. Agent A asks Agent B what (s)he is willing to pay for the bet that the first coin is heads. With probability 1/2 the second coin is heads and since Agent B does not know if Agent A will use the opportunity to flip the coin, (s)he is completely ignorant about the first coin in that case. If the second coin is tails, however, Agent B simply knows that half of the time the first coin will be heads and the other half of the time the first coin will be tails. If we write \( \Omega = \{h, t\} \) for the possible outcomes of the first coin, this corresponds to the basic belief assignment

\[
m(\{h\}) = m(\{t\}) = 1/4, \quad m(\Omega) = 1/2.
\] (4.0.11)

The maximum price for which Agent B can many times buy the bet that the first coin is heads, such that it is guaranteed that (s)he will do not lose money, is

\[
\text{Bel}(\{h\}) = m(\{h\}) = 1/4.
\] (4.0.12)

Although the theory of belief functions has been expanded and applied since its introduction (see [11] for a recent overview and [7] for a recent text), researchers in mainstream mathematics have essentially stayed away from it. An important factor in this lack of mainstream acceptance, is the so called Demspeter’s rule of combination. This rule, that is supposed to combine two belief function into a new one, is used (and even ‘celebrated’ [14]) by many advocates of the theory, even within the ‘probability of provability’ interpretation [55]. The rule, however, has met a lot of criticism: it would not be well founded or motivated, and it would
lead to unacceptable results [17, 45]. Because the rule of combination appears intertwined with the theory of belief functions, criticism of the rule automatically becomes criticism of the theory.

In Chapter 5, that is based on [33], we start by addressing the calculus of belief functions. We do this by carefully picking from the already quite rich number of proposals regarding the calculus of belief functions (see e.g. [67, 68, 16]). We will use the frequentist interpretation introduced above since it turns out to motivate the calculus in a more straightforward from this interpretation and postpone discussion from a betting interpretation to Chapter 6. With respect to conditioning, instead following the distinction of Dubois and Prade [14] between ‘focusing’ and ‘updating’, we make a distinction between learning that something is contingently true (or has occurred contingently) and learning that something is necessarily true (or has necessarily occurred). Contingent conditioning coincides with what Dubois and Prade call ‘focusing’ and corresponds to a rule of conditioning that was already described by Dempster [9]. The rule that corresponds to necessary conditioning can be seen as the special case of Dempster-Shafer conditioning in which we have Bel($H^c$) = 0, if $H$ is the set on which we condition. Having introduced these two notions of conditioning, we explain how we can consider (general) Dempster-Shafer conditioning, the rule of conditioning that corresponds to what Dubois and Prade call ‘updating’, as a ‘mix’ between contingent and necessary conditioning.

After conditioning, we discuss independence. We first discuss independence concepts related to the notions of conditioning we discussed. Then we show that the notion of ‘strong independence’, a renaming of what Shafer called ‘evidential independence’, is a very natural concept within the interpretation. Having interpreted both Dempster-Shafer conditioning and strong independence, we are able to explain why we do not see any justification for Dempster’s rule of combination within the interpretation we use. We do not do this by pointing to results of the rule, but by pointing to problems with the interpretation of the rule itself. We conclude Chapter 5 by presenting a law of large numbers.

In Chapter 6, that is based on [32] and [33], the goal is to justify the proposed calculus of Chapter 5 from a betting interpretation. The first step is, to justify belief functions themselves, i.e. justify Definition 4.0.8 from a betting interpretation, in an argument analogous to the Dutch Book argument. The theory of Peter Walley [61], which already provides betting schemes with a distinction between buying and selling prices, serves as a natural starting point here. We then argue that on top of the constraint of coherence that Walley gives, reasonable agents should adhere to an additional constraint. We show that requiring this additional constraint, leads to Definition 4.0.8. This leads to a betting interpretation of belief functions, distinct from other interpretations that have been offered [57, 54, 48, 50]. The remainder of Chapter 6 is dedicated to interpreting contingent conditioning, necessary conditioning and strong independence.

In Chapter 7, that is based on [30], we apply the theory developed in Chapter
5 to forensic problems. We do this for two variants of the island problem, already mentioned in Example 4.0.5 and a case of parental identification. In these cases, a crucial problem of the classical approach is that one needs to assign prior probability to the criminal (in case of the island problem) and father (in case of parental identification). A uniform distribution is what is typically used, but as already explained, this does not really represent ignorance. With the theory of belief functions, we are able to take an actual uninformative prior. We show that doing so, leads to different answers than the classical answers and we compare these answers. We end the chapter by commenting on the consequences for legal practice.

In Chapter 8, that is based on [34], we use the theory to study an infinite regress problem. In this problem, we have an infinite sequence $E_0, E_1, E_2, \ldots$ of propositions. The central question is: if we know the conditional beliefs of $E_n$ given $E_{n+1}$ of some agent for every $n$, what can we infer about his or her (unconditional) beliefs about the individual propositions, and proposition $E_0$ in particular. To analyze this problem, we develop some basic theory of belief functions on infinite spaces in this chapter. The result of this turns out to be that, contrary to certain claims in the literature, this regress problem does in general not have a unique solution.
CHAPTER 5
Conditioning and independence

5.1 Contingent conditioning

In this chapter, we use the frequentist interpretation introduced in Chapter 4. In the classical frequentist approach, conditioning on an event \( H \) means that we ignore all outcomes which are not in \( H \). The conditional probability of an event \( A \) given \( H \) is then the relative frequency of \( A \) in the resulting subsequence.

This idea can be used in our context of belief functions as well. The problem is, however, that we do not always know whether or not \( A \) or \( H \) actually occur, since we do not have complete information. To deal with this, we introduce the plausibility of an event \( A \).

**Definition 5.1.1.** The plausibility of an event \( A \) is defined as

\[
\text{Pl}(A) := 1 - \text{Bel}(A^c) = \sum_{S : S \cap A \neq \emptyset} m(S) = P_m(\{S : S \cap A \neq \emptyset\}).
\]  

(5.1.1)

In the current interpretation, the plausibility of \( A \) is the minimum price for which, if the bet on \( A \) is bought in a series of repetitions of the experiment, it is guaranteed that in the limit, there is no loss.

If \( \text{Bel}(H) = \text{Pl}(H) \), then we know that \( H \) occurs with limiting frequency \( \text{Bel}(H) = \text{Pl}(H) \), since there is no set with positive \( m \)-mass which intersects both \( H \) and \( H^c \). This means that for \( A \subseteq \Omega \), the conditional belief in \( A \) can now be defined as the relative frequency of the number of times that we know \( A \) occurs in the subsequence consisting of all occurrences of \( H \). This relative frequency is given by

\[
\frac{\text{Bel}(A \cap H)}{\text{Bel}(H)},
\]

(5.1.2)

which, therefore, can be taken as the conditional belief in \( A \) given \( H \).

If, however, \( \text{Bel}(H) < \text{Pl}(H) \) with strict inequality, then there is no known limiting frequency at which \( H \) occurs, since there is at least one set \( B \) with
m(B) > 0 such that B has non-empty intersection with both H and $H^c$. There may not even exist a limiting frequency of H, we simply do not know. As a consequence, there is no (known) limiting frequency at which A occurs relative to H.

To overcome this problem, instead of looking at the limiting frequency, we can look at the limit inferior of the frequency at which A occurs relative to H. Of course this limit inferior always exists, although we can not in general determine it. We can, however, determine the largest lower bound of this limit inferior we can be sure of, which clearly is the conditional maximum price for which one can buy the bet on A such that it is guaranteed that there will be no loss (in the limit). To do this, we partition $2^Ω$ into four collections as follows.

(1) The collection of outcomes of the experiment that imply $A \cap H$, i.e.

$$C_1 = \{C : C \subseteq A \cap H\}. \quad (5.1.3)$$

(2) The collection of outcomes of the experiment that imply $H^c$, i.e.

$$C_2 = \{C : C \subseteq H^c\}. \quad (5.1.4)$$

(3) The collection of outcomes of the experiment that do not rule out H and $A^c$ occurring simultaneously, i.e.

$$C_3 = \{C : C \cap (H \setminus A) \neq \emptyset\}. \quad (5.1.5)$$

(4) The collection of outcomes of the experiment that do not rule out $H^c$, but do guarantee that if H occurs, then so does A, i.e.

$$C_4 = \{C : C \cap H^c \neq \emptyset; \emptyset \neq C \cap H \subseteq A\}. \quad (5.1.6)$$

Now what can we say about the frequency of A occurring relative to the occurrence of H? Let us first assume that $P_m(C_3) > 0$. In Case (1) it is clear that both A and H occur and Case (2) can be discarded since H does not occur. In Case (4) A occurs whenever H does, but we cannot be sure that H occurs. Hence we play safe if we simply ignore these outcomes, since leaving out situations in which both A and H occur can only decrease the relative frequency A. Finally, in Case (3) H may occur while A does not. We cannot ignore these situations when we aim for minimizing the relative frequency of A given H. The best we can do is count them in the denominator (that is, assume H occurs) but not in the numerator (since A may not occur). Taking all the considerations into account, the greatest lower bound of the limit inferior of the frequency at which A occurs relative to H that we can give is given by the formula

$$\frac{P_m(C_1)}{P_m(C_1) + P_m(C_3)}. \quad (5.1.7)$$
5.1. CONTINGENT CONDITIONING

If $P_m(C_3) = 0$, we know that $H$ and $A^c$ occurring simultaneously is ruled out with probability 1. If $H$ occurring is not ruled out, i.e., if $P(H) > 0$, the relative frequency of $A$ occurring relative to the occurrence of $H$, should be 1. This is exactly what (5.1.7) gives for $P_m(C_3) = 0$ if it is defined, that is, if $P_m(C_1) > 0$. This then finally leads to the following definition.

**Definition 5.1.2** (Contingent conditioning). Assume that $P(H) > 0$. We define the conditional belief in $A$ given $H$ by

$$Bel_H(A) := \frac{Bel(A \cap H)}{Bel(A \cap H) + Pl(H \setminus A)}$$

if $Pl(H \setminus A) > 0$ and $Bel_H(A) = 1$ if $Pl(H \setminus A) = 0$.

This notion of conditioning is already described by, among others, Demspter [9] and proposed by Fagin and Halpern [15]. Although we have defined it directly in terms of the original belief function, Fagin and Halpern define the conditional belief function by

$$Bel_H(A) = \inf \{P(A|H) : P \text{ prob. distr., } \forall A P(A) \geq Bel(A)\}$$

if $Bel(H) > 0$ and then show that this equals (5.1.8). The corresponding plausibility of $A$ is given by

$$Pl_H(A) = \frac{Pl(A \cap H)}{Pl(A \cap H) + Bel(A^c \cap H)}.$$  \hspace{1cm} (5.1.10)

Note that if Bel is a probability distribution, we find

$$Bel_H(A) = Pl_H(A) = \frac{Bel(A \cap H)}{Bel(H)} = \frac{Pl(A \cap H)}{Pl(H)},$$ \hspace{1cm} (5.1.11)

making contingent conditioning a generalization of classical conditioning. Also note that although we have also motivated the definition based on determining a greatest lower bound, the identity (5.1.9) suggests that there is an (unknown) underlying probability distribution on $\Omega$. We do not assume this, and we only used the distribution $P_m$ defined on $2^{2^3}$.

Fagin and Halpern have showed that $Bel_H$ is again a belief function and we write $m_H$ for the corresponding basic belief assignment. They only showed this for the case $Bel(H) > 0$, since they only defined $Bel_H$ for $Bel(H) > 0$. We notice, however, that in the case $Bel(H) = 0$, it can be easily seen that $Bel_H$ is the belief function corresponding to the basic belief assignment $m_H$ given by $m_H(U) = 1$, where $U$ is the set

$$U = \bigcap_{C : Pl(H \setminus A) = 0} C.$$ \hspace{1cm} (5.1.12)

This basic belief assignment $m_H$, however, is in general not easily determined and does not seem to have a clear interpretation. The belief function $Bel_H$ is
obtained by computing the beliefs directly, and \( m_H \) simply happens to be the basic belief assignment corresponding to \( \text{Bel}_H \). In particular, this means that we can not interpret \( m_H(S) \) as a limiting frequency that is related to the original distribution \( P_m \), at least not in a straightforward way.

**Example 5.1.3.** Let \( \Omega = \{a, b, c, d\} \) and set

\[
m(a) = m(b,c) = m(c,d) = \frac{1}{3}. \tag{5.1.13}
\]

Suppose we condition on \( H = \{a, b, c\} \) in a contingent way, then we find

\[
\text{Bel}_H(a) = \frac{\text{Bel}(a)}{\text{Bel}(a) + \text{Pl}(b,c)} = \frac{1}{3}. \tag{5.1.14}
\]

and

\[
\text{Bel}_H(b,c) = \frac{\text{Bel}(b,c)}{\text{Bel}(b,c) + \text{Pl}(a)} = \frac{1}{2}. \tag{5.1.15}
\]

From these beliefs we can reconstruct the corresponding basic belief assignment:

\[
m_H(a) = \frac{1}{3}, \ m_H(b,c) = \frac{1}{2}, \ m_H(a,b,c) = \frac{1}{6}. \tag{5.1.16}
\]

If we would like to apply contingent conditioning again, considering beliefs contingent upon some \( G \subseteq \Omega \), we should interpret \( \text{Bel}_H \) as a belief function based on the chance experiment corresponding to \( m \) rather than \( m_H \). After all, that is precisely how we defined \( \text{Bel}_H \). This means that we would simply get \( \text{Bel}_{H \cap B} \) as new belief function. If we would interpret \( \text{Bel}_H \) as a belief function based on the chance experiment corresponding to \( m_H \), we get \( (\text{Bel}_H)_G \) as our new belief function. A natural question is if \( \text{Bel}_{H \cap G} \) and \( (\text{Bel}_H)_G \) are the same and if contingent conditioning is commutative. For classical probability this is of course true and indeed if \( \text{Bel} \) is a probability distribution, then \( \text{Bel}_{H \cap G} \) and \( (\text{Bel}_H)_G \) are the same, as is easy to see from (5.1.11). This turns out not to be true in general, as illustrated by the following example.

**Example 5.1.4.** We consider the situation of Example 5.1.3 again. A crucial observation is that the belief in \( \{b,c\} \) is higher after conditioning, while the belief in \( \{a\} \) remained the same. Note that this can not occur if \( \text{Bel} \) is a probability distribution: in that case the ratios between the probabilities remain the same after conditioning. The reason that it happens in this example is that there is a chance of \( \frac{1}{3} \) on \( \{c,d\} \), which is consistent with both \( H \) and \( H^c \). Precisely for that reason, as described before, the beliefs in \( \{a\} \) and \( \{b,c\} \) are treated differently. For \( \{a\} \) the set \( \{c,d\} \) falls into category (3) and is counted in the computation. For \( \{b,c\} \) the set \( \{c,d\} \) falls into category (4) and is not counted. The latter leads to some loss of information and thus one would indeed not expect contingent conditioning to be commutative.

Indeed, if we set \( G := \{a,b\} \), we find

\[
(\text{Bel}_H)_G(a) = \frac{\text{Bel}_H(a)}{\text{Bel}_H(a) + \text{Pl}_H(b)} = \frac{1}{3}. \tag{5.1.17}
\]
and
\[(\text{Bel}_G)_H(\{a\}) = \text{Bel}_G(\{a\}) = \frac{\text{Bel}(\{a\})}{\text{Bel}(\{a\}) + \text{Pl}(\{b\})} = \frac{1}{2}.\] (5.1.18)

We do not see this phenomenon as a weakness or drawback of the notion of contingent conditioning, but as a reasonable and realistic property.

Another point that we want to make is that there no equivalent of the law of total probability for contingent conditioning. Classical conditioning satisfies the following elementary property:
\[
\min\{P(A|H), P(A|H^c)\} \leq P(A) \leq \max\{P(A|H), P(A|H^c)\},
\] (5.1.19)
of which the law of total probability is a special case. The next examples shows that contingent conditioning lacks this property.

**Example 5.1.5.** Suppose we have two coins and we write
\[\Omega = \{(h,h), (h,t), (t,h), (t,t)\},\] (5.1.20)
where the first and second coordinate represent the side of respectively the first and second coin facing up (here ‘h’ stand for heads and ‘t’ for tails). Instead of flipping the coins, someone rolls a fair die and follows the following instruction. If he rolls 1 or 2, he puts both coins with heads facing up. If he rolls 3 or 4, he puts both coins with tails facing up. If he rolls 5 or 6, we have no information about what the agent does. This translates into the following basic belief assignment:
\[m(\{(h,h)\}) = m(\{(t,t)\}) = m(\Omega) = \frac{1}{3}.\] (5.1.21)

We are now interested in \(A = \{(t,t), (h,h)\}\), which is the event that both coins have the same side facing up. We compute \(\text{Bel}(A) = \frac{2}{3}\).

Now suppose that we are given the information that the second coin is heads. We write \(H = \{(t,t), (h,h)\}\) and compute
\[
\text{Bel}_H(A) = \frac{\text{Bel}(\{(h,h)\})}{\text{Bel}(\{(h,h)\}) + \text{Pl}(\{(t,h)\})} = \frac{1}{2}.
\] (5.1.22)

In exactly the same way, we compute
\[
\text{Bel}_{H^c}(A) = \frac{\text{Bel}(\{(t,t)\})}{\text{Bel}(\{(t,t)\}) + \text{Pl}(\{(h,t)\})} = \frac{1}{2}.
\] (5.1.23)

This might be somewhat surprising: if the chance is \(\frac{1}{2}\) in case we learn the second coin is heads and in case we learn the second coin is tails, one might think that the unconditional chance should also be \(\frac{1}{2}\), since we know beforehand that the second coin is either heads or tails. However, we must be very careful with interpretation here. The quantity \(\text{Bel}_H(A)\) gives the belief in \(A\) contingent upon \(H\) and \(\text{Bel}_{H^c}\) the belief in \(A\) contingent upon \(H^c\). While it is true that either \(H\) occurs or \(H^c\) occurs, it is clearly not true that either every time that \(H\) occurs the outcome is ignored or every time that \(H^c\) occurs the outcome is ignored. So we do not know at all beforehand that the belief in \(A\) should be considered either contingent upon the second coin being heads or contingent upon the second coin being tails.
5.2 Necessary conditioning

Now we consider updating according to the information that some set $H \subseteq \Omega$ necessarily occurs. This must be understood in the context of the chance experiment: we learn that in every repetition of the experiment $H$ occurs. This is information about the experiment in the sense that it narrows down the information we get in the experiment. The means, however, that this type of information can only be used for updating if it does not contradict the chance experiment: every outcome of the experiment $C \subseteq \Omega$ that has positive chance (i.e. $P(\{C\}) > 0$) is consistent with $H$ (i.e. $C \cap H \neq \emptyset$). Hence we must have $\text{Pl}(H) = 1$. If this is the case, the chance experiment can be updated: the outcome $C \subseteq \Omega$ now becomes the outcome $C \cap H$. So the updated belief in $A$ should become

$$P_m(\{C : C \cap H \subseteq A\}) = P_m(\{C : C \subseteq A \cup H^c\}).$$ \hspace{1cm} (5.2.1)

This leads to the following definition.

**Definition 5.2.1** (Necessary conditioning). Assume that $\text{Pl}(H) = 1$. We define

$$\text{Bel}_{H,\text{ nec}}(A) := \text{Bel}(A \cup H^c).$$ \hspace{1cm} (5.2.2)

The corresponding plausibility and basic belief assignment are given by respectively

$$\text{Pl}_{H,\text{ nec}}(A) = 1 - \text{Bel}_{H,\text{ nec}}(A^c)$$

$$= 1 - \text{Bel}(A^c \cup H^c)$$

$$= \text{Pl}(A \cap H)$$ \hspace{1cm} (5.2.3)

and

$$m_{H,\text{ nec}}(A) = \sum_{S : S \cap H = A} m(S).$$ \hspace{1cm} (5.2.4)

Necessary conditioning is a special case of what is known as Shafer-Dempster conditioning (general Shafer-Dempster conditioning is discussed in the next subsection). It is also a special case of the unnormalized conditioning of Smets [56], who defined (5.2.1) as a conditional belief for any subset $H \subseteq \Omega$. Necessary conditioning is, technically speaking, not a generalization of classical conditioning because it is only defined in case $\text{Pl}(H) = 1$. If Bel is a probability distribution and $\text{Bel}(H) = \text{Pl}(H) = 1$, then necessary conditioning trivially coincides with classical conditioning, with $\text{Bel}_{H} = \text{Bel}$.

Contrary to contingent conditioning, necessary conditioning is commutative. Indeed, for $H,G \subseteq \Omega$ such that $\text{Pl}(H) = \text{Pl}(G) = 1$, we have

$$(\text{Bel}_{H,\text{ nec}})_G,\text{ nec}(A) = \text{Bel}((A \cup H^c) \cup G^c)$$

$$= \text{Bel}(A \cup (H \cap G)^c)$$

$$= \text{Bel}_{H \cap G,\text{ nec}}(A)$$

$$= (\text{Bel}_G,\text{ nec})_H,\text{ nec}(A).$$ \hspace{1cm} (5.2.5)
5.3. MIXED CONDITIONING

However, similarly to contingent conditioning, it does not satisfy
\[
\min\{\text{Bel}_{H,\text{ Nec}}(A), \text{Bel}_{H^c,\text{ Nec}}(A)\} \leq \text{Bel}(A) \\
\leq \max\{\text{Bel}_{H,\text{ Nec}}(A), \text{Bel}_{H^c,\text{ Nec}}(A)\},
\]
as the following example illustrates.

**Example 5.2.2.** Suppose we have two coins and we write
\[
\Omega = \{(h,h), (h,t), (t,h), (t,t)\},
\]
where the first and second coordinate represent the side of respectively the first and second coin facing up (here ‘h’ stand for heads and ‘t’ for tails). The first coin is flipped randomly, and we receive information about the outcome. No information about the second coin is given. So we have
\[
m(\{(h,h),(h,t)\}) = m(\{(t,h),(t,t)\}) = \frac{1}{2}.
\]
If we set \(A = \{(t,t),(h,h)\}\), then \(\text{Bel}(A) = 0\).

Now suppose that we are given the information that the second coin is heads. We write \(H = \{(t,h),(h,h)\}\) and compute
\[
\text{Bel}_{H,\text{ Nec}}(A) = \text{Bel}(\{(t,t),(h,h),(h,t)\}) = \frac{1}{2}.
\]
In exactly the same way, we compute
\[
\text{Bel}_{H^c,\text{ Nec}}(A) = \text{Bel}(\{(t,t),(h,h),(t,h)\}) = \frac{1}{2}.
\]
This example is very similar to Example 5.1.5: the conditional belief under \(H\) and \(H^c\) are both \(\frac{1}{2}\), but the unconditional belief is not \(\frac{1}{2}\). The explanation is also similar: while it is true that either \(H\) occurs or \(H^c\) occurs, it is clearly not true that either \(H\) *always* occurs or \(H^c\) *always* occurs. So, if the belief in \(A\) is equal to \(\frac{1}{2}\) under both these scenarios, one can simply not conclude that it always has to be \(\frac{1}{2}\), since they are not the only two scenarios.

### 5.3 Mixed conditioning

The type of information corresponding to mixed conditioning, consists of two steps. The first step is that \(H\) occurs in every situation we already know that is consistent with \(H\). The second step is that we ignore any outcome where \(H^c\) occurs.

Perhaps this description appears somewhat unnatural. The reasons to include it here are (1) that it precisely leads to the classical notion of Shafer-Dempster conditioning [47], and (2) that in certain circumstances mixed conditioning can
be seen as a concatenation of contingent and necessary conditioning, see Theorem 5.3.2 below.

In the context of the chance experiment, mixed conditioning means that every outcome \( C \subseteq \Omega \) of the experiment that is consistent with \( H \) (i.e. \( C \cap H \neq \emptyset \)), now becomes \( C \cap H \) and that every outcome \( C \subseteq \Omega \) that is inconsistent with \( H \), is ignored. So the updated belief in \( A \) should equal

\[
\frac{P_m(\{C : \emptyset \neq C \cap H \subseteq A\})}{1 - P_m(\{C : A \cap H^c \subseteq \Omega\})} = \frac{P_m(\{C : C \subseteq A \cup H^c\}) - P_m(\{C : C \subseteq H^c\})}{1 - P_m(\{C : C \subseteq H^c\})}
\] (5.3.1)

and this leads to the following definition.

**Definition 5.3.1** (Mixed conditioning). Assume that \( \text{Bel}(H^c) < 1 \). We set

\[
\text{Bel}_{H, \text{mix}}(A) = \frac{\text{Bel}(A \cup H^c) - \text{Bel}(H^c)}{1 - \text{Bel}(H^c)}.
\] (5.3.2)

The corresponding plausibility and basic belief assignment are given by respectively

\[
\text{Pl}_{H, \text{mix}}(A) = 1 - \text{Bel}_{H, \text{mix}}(A^c)
\] 
\[
= 1 - \frac{\text{Bel}(A^c \cup H^c) - \text{Bel}(H^c)}{1 - \text{Bel}(H^c)}
\] 
\[
= \frac{1 - \text{Bel}(A^c \cup H^c)}{1 - \text{Bel}(H^c)}
\] (5.3.3)

\[
= \frac{\text{Pl}(A \cap H)}{\text{Pl}(H)}.
\]

and

\[
m_{H, \text{mix}}(A) = \frac{\sum_{S : \emptyset \neq S \cap H = A} m(S)}{\sum_{S : \emptyset \neq S \cap H} m(S)}.
\] (5.3.4)

This notion of conditioning is known in the literature as Dempster-Shafer conditioning [47]. Mixed conditioning is a generalization of both classical and necessary conditioning. Mixed conditioning (and therefore also necessary conditioning) always results in a higher conditional belief than contingent conditioning. Indeed, since

\[
\text{Pl}(A \cap H) + \text{Bel}(A^c \cap H) \leq \text{Pl}(H),
\] (5.3.5)

we find

\[
\text{Pl}_{H, \text{mix}}(A) = \frac{\text{Pl}(A \cap H)}{\text{Pl}(H)} \leq \frac{\text{Pl}(A \cap H)}{\text{Pl}(A \cap H) + \text{Bel}(A^c \cap H)} = \text{Pl}_H(A).
\] (5.3.6)

Or, equivalently:

\[
\text{Bel}_{H, \text{mix}}(A) \geq \text{Bel}_H(A).
\] (5.3.7)

In the following special case, mixed conditioning can be expressed in terms of necessary and contingent conditioning.
Theorem 5.3.2. If $\Pi(N) = 1$ and $\text{Bel}(H) = \Pi(H) > 0$, then

$$\text{Bel}_{H \cap N, \text{mix}} = (\text{Bel}_{N, \text{nec}})_H = (\text{Bel}_H)_{N, \text{nec}}.$$  \hfill (5.3.8)

Proof. Suppose $\Pi(N) = 1$ and $\text{Bel}(H) = \Pi(H)$. We prove the identity for the plausibility. Note because of $\text{Bel}(H) = \Pi(H)$ we have

$$\Pi(A \cap H) + \text{Bel}(A^c \cap H) = \Pi(H).$$ \hfill (5.3.9)

Hence

$$\Pi_H(A) = \frac{\Pi(A \cap H)}{\Pi(H)}. \hfill (5.3.10)$$

Applying necessary conditioning on $N$ we find

$$(\Pi_H)_{N, \text{nec}}(A) = \frac{\Pi(A \cap H \cap N)}{\Pi(H)} = \Pi_{N \cap H, \text{mix}}(A). \hfill (5.3.11)$$

If we apply the necessary conditioning on $N$ first, we have

$$\Pi_{N, \text{nec}}(A) = \Pi(A \cap N). \hfill (5.3.12)$$

We now observe that

$$\text{Bel}_{N, \text{nec}}(H) = \text{Bel}(H \cup N^c)$$

$$= \text{Bel}(H)$$

$$= \Pi(H) \hfill (5.3.13)$$

$$= \text{Bel}_{N, \text{nec}}(H).$$

So contingent conditioning on $H$, gives

$$(\Pi_{N, \text{nec}})_H(A) = \frac{\Pi_{N, \text{nec}}(A \cap H)}{\Pi_{N, \text{nec}}(H)}$$

$$= \frac{\Pi(A \cap H \cap N)}{\Pi(H)} \hfill (5.3.14)$$

$$= \Pi_{N \cap H, \text{mix}}(A).$$

\hfill \Box

5.4 Independence

Before we discuss different notions of independence, we setup the necessary mathematical framework. To do that, we set $\Omega := \Omega_1 \times \Omega_2$ and let $\text{Bel} : 2^\Omega \to [0, 1]$ be a belief function and $m$ be its corresponding basic belief assignment. Let $\pi_1 : \Omega \to \Omega_1$ and $\pi_2 : \Omega \to \Omega_2$ be the projections of $\Omega$ onto respectively $\Omega_1$ and $\Omega_2$, i.e.

$$\pi_1((\omega_1, \omega_2)) := \omega_1 \hfill (5.4.1)$$
and
\[ \pi_2((\omega_1, \omega_2)) := \omega_2. \]

We now define basic belief assignments \( m_1 : 2^{\Omega_1} \to [0, 1] \) and \( m_2 : 2^{\Omega_2} \to [0, 1] \) by
\[ m_1(A) := \sum_{S \subseteq \Omega : \pi_1(S) = A} m(S). \quad (5.4.3) \]
and
\[ m_2(B) := \sum_{S \subseteq \Omega : \pi_2(S) = B} m(S). \quad (5.4.4) \]

The belief function \( \text{Bel}_1 : 2^{\Omega_1} \to [0, 1] \) corresponding to \( m_1 \) is then given by
\[ \text{Bel}_1(A) = \sum_{S \subseteq \pi_1^{-1}(A)} m_1(S) \]
\[ = \sum_{S \subseteq \pi_1^{-1}(A)} \sum_{C \subseteq \Omega : \pi_1(C) = S} m(C) \]
\[ = \sum_{S \subseteq \pi_1^{-1}(A)} m(S) \]
\[ = \text{Bel}(\pi_1^{-1}(A)) \]
\[ = \text{Bel}(A \times \Omega_2). \quad (5.4.5) \]

A similar identity holds for the belief function \( \text{Bel}_2 \) corresponding to \( m_2 \).

We start by discussing the three notions of independence that follow from the three notions of conditioning that we have discussed, by requiring that conditioning on \( \Omega_1 \times B \) does not change the belief in \( A \times \Omega_2 \) and vice versa. Sometimes this kind of independence notion is called an ‘irrelevance notion’. We start with contingent conditioning. We require that
\[ \text{Bel}_{\Omega_1 \times B}(A \times \Omega_2) = \text{Bel}(A \times \Omega_2) \quad (5.4.6) \]
and
\[ \text{Bel}_{A \times \Omega_2}(\Omega_1 \times B) = \text{Bel}(\Omega_1 \times B) \quad (5.4.7) \]
whenever these are well defined. This requirement comes down to
\[ \text{Bel}(A \times B) = \text{Bel}_1(A) (\text{Bel}(A \times B) + \text{Pl}(A^c \times B)) \quad (5.4.8) \]
and
\[ \text{Bel}(A \times B) = \text{Bel}_2(B) (\text{Bel}(A \times B) + \text{Pl}(A \times B^c)) \quad (5.4.9) \]
for all \( A \subseteq \Omega_1 \) and \( B \subseteq \Omega_2 \).

Clearly, the notion of contingent independence is a generalization of independence for probability distributions. The problem of this notion is that it is very strong: only in a few circumstances one has this kind of independence.
5.4. INDEPENDENCE

Example 5.4.1. Suppose we have two coins and we write
\[ \Omega_1 = \Omega_2 = \{h, t\}, \]  
(5.4.10)
where \( \Omega_1 \) and \( \Omega_2 \) represent the side of respectively the first and second coin facing up (here ‘h’ stand for heads and ‘t’ for tails). Suppose that we have
\[ m(\{(h, h)\}) = m(\{(h, h), (h, t)\}) = m(\{(h, h), (t, h)\}) = m(\Omega) = \frac{1}{4}. \]  
(5.4.11)
It is easy to check that
\[ m_1(\{h\}) = m_1(\Omega_1) = \frac{1}{2}, \quad m_2(\{h\}) = m_2(\Omega_2) = \frac{1}{2} \]  
and that
\[ m(A \times B) = m_1(A)m_2(B). \]  
(5.4.12)
We cannot rule out that there is an unknown probability distribution that describes the phenomenon which makes the two components not independent. In case of this example, it can not be ruled out that we have \((h, h)\) precisely \(1/4\) of the time, \((h, t)\) precisely \(1/2\) of the time and \((t, h)\) precisely \(1/4\) of the time. Precisely because such a scenario can not be ruled out, we do not have independence with respect to contingent conditioning. Indeed, we find
\[ \text{Bel}_{\Omega_1 \times \{h\}}(\{h\} \times \Omega_2) = \frac{\text{Bel}(\{(h, h)\})}{\text{Bel}(\{(h, h)\}) + \text{Pl}((t, h))} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} = \frac{1}{3}, \]  
(5.4.13)
while
\[ \text{Bel}_1(\{h\}) = \frac{1}{2}. \]  
(5.4.14)

The case in which \( \text{Bel}_1 \) and \( \text{Bel}_2 \) are probability distributions is one of the few circumstances under which this form of independence is possible, as the next theorem illustrates.

Theorem 5.4.2. If (5.4.8) and (5.4.9) hold, then either \( m_1(\{\omega\}) = 1 \) for some \( \omega \in \Omega_1 \), or \( m_2(\{\omega\}) = 1 \) for some \( \omega \in \Omega_2 \), or
\[ m_1(A) + \sum_{\omega \in A} m_1(\{\omega\}) = 1 \]  
(5.4.15)
for some \( A \subseteq \Omega_1 \) with \(|A| > 1 \) and
\[ m_2(B) + \sum_{\omega \in B} m_2(\{\omega\}) = 1 \]  
(5.4.16)
for some \( B \subseteq \Omega_2 \) with \(|B| > 1 \).

Proof. Set
\[ A := \bigcap_{\text{Bel}_1(A') = 1} A' \quad \text{and} \quad B := \bigcap_{\text{Bel}_2(B') = 1} B'. \]  
(5.4.17)
If \( A = \{\omega\} \), then \( m_1(\{\omega\}) = 1 \) and we are done. The same holds if \( B = \{\omega\} \), so assume that \( |A| > 1 \) and \( |B| > 1 \). Clearly, every \( S \subseteq \Omega_1 \times \Omega_2 \) with \( m(S) > 0 \) must be a subset of \( A \times B \). Let \( S \subseteq A \times B \) be an arbitrary set with \( |S| > 1 \) and \( \pi_1(S) \neq A \). To prove that
\[
m_1(A) + \sum_{\omega \in A} m_1(\{\omega\}) = 1, \tag{5.4.18}
\]
it is sufficient to show that \( m(S) = 0 \).

We distinguish two cases. First, we consider the case \( |\pi_2(S)| = 1 \). Set \( V := \pi_2(S) \). Since \( |B| > 1 \), we find that \( \text{Bel}_2(V) < 1 \). Further, since \( |S| > 1 \), there is a \( U \subseteq \Omega_1 \) such that \( S \) intersects both \( U \times V \) and \( U^c \times V \). We now set
\[
\Delta := \text{Bel}_2(V) - \text{Bel}(U \times V) - \text{Bel}(U^c \times V)
\]
\[
D_1 := \text{Bel}(U \times V) + \text{Pl}(U \times V^c); \tag{5.4.19}
\]
\[
D_2 := \text{Bel}(U^c \times V) + \text{Pl}(U^c \times V^c).
\]

We find
\[
\text{Bel}_2(V) = \text{Bel}(U \times V) + \text{Bel}(U^c \times V) + \Delta
\]
\[
= \text{Bel}_2(V) (D_1 + D_2) + \Delta
\]
\[
\geq \text{Bel}_2(V) (\Delta + D_1 + D_2)
\]
\[
= \text{Bel}_2(V) (\text{Bel}_2(V) + \text{Pl}(U \times V^c) + \text{Pl}(U^c \times V^c))
\]
\[
\geq \text{Bel}_2(V) (\text{Bel}_2(V) + \text{Pl}(\Omega_1 \times V^c))
\]
\[
= \text{Bel}_2(V). \tag{5.4.20}
\]

In particular this implies that the third line must be equal to the second line and since \( \text{Bel}_2(V) < 1 \), this can only be the case if \( \Delta = 0 \). Thus
\[
\text{Bel}_2(V) = \text{Bel}(U^c \times V) + \text{Bel}(U \times V), \tag{5.4.21}
\]
and this implies that \( m(S) = 0 \).

Now consider the case \( |\pi_2(S)| > 1 \). Set \( U := \pi_1(S) \). Since \( U \neq A \), we find that \( \text{Bel}_1(U) < 1 \). Since \( |\pi_2(S)| > 1 \), there is a \( V \subseteq \Omega_2 \) such that \( S \) intersects both \( U \times V \) and \( U \times V^c \). By a completely analogous argument as before we find
\[
\text{Bel}_1(U) = \text{Bel}(U \times V) + \text{Bel}(U \times V^c), \tag{5.4.22}
\]
implying that \( m(S) = 0 \).

The property for \( m_2 \) follows in exactly the same way.

The notion of independence following from necessary conditioning, boils down to the requirement that
\[
\text{Pl}(A \times B) = \text{Pl}_1(A) \text{Pl}_2(B) \tag{5.4.23}
\]
if either $\Pr_1(A) = 1$ or $\Pr_2(B) = 1$. This notion of independence does not generalize classical independence. In fact this requirement is so weak that if $\Bel$ is any probability distribution, it is satisfied.

The more general requirement that follows from mixed conditioning, is

$$\Pr(A \times B) = \Pr_1(A)\Pr_2(B) \quad (5.4.24)$$

for all $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$. This notion (as already recognized by Shafer) does generalize classical independence. However, requiring this independence in combination with the requirement that $\Bel_1$ and $\Bel_2$ are certain fixed belief functions, does not uniquely determine $\Bel$.

We now consider a fourth notion of independence that is not directly based on a notion of conditioning, but is a very natural candidate giving the provability interpretation, is one based on classical independence. Let $P$, $P_1$, $P_2$ the chance experiments corresponding to respectively $\Bel$, $\Bel_1$ and $\Bel_2$. The chance that $A \times B$ happens, should be the product of chances of $A$ and $B$, i.e.

$$P(\{A \times B\}) = P_1(\{A\})P_2(\{B\}) \quad (5.4.25)$$

for all $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$. This notion of independence is called ‘evidential independence’ by Shafer. Since we do not approach the theory of belief functions as a theory of evidence, we prefer to call it ‘strong independence’.

**Definition 5.4.3.** We say $\Bel_1$ and $\Bel_2$ are **strongly independent** if

$$m(A \times B) = m_1(A)m_2(B) \quad (5.4.26)$$

It follows directly from the definition that if we require $\Bel_1$ and $\Bel_2$ to be certain fixed belief functions (and hence $m_1$ and $m_2$ to equal certain fixed basic belief assignments), that strong independence then completely determines $\Bel$ since (5.4.26) determines $m$.

The following theorem goes back to Shafer.

**Theorem 5.4.4.** $\Bel_1$ and $\Bel_2$ are **strongly independent** if and only if

$$\Bel(A \times B) = \Bel_1(A)\Bel_2(B) \quad (5.4.27)$$

and

$$\Pr(A \times B) = \Pr_1(A)\Pr_2(B) \quad (5.4.28)$$

for all $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$.

Theorem 5.4.4 shows that strong independence implies independence with respect to mixed and necessary conditioning. However, it does not imply irrelevance with respect to contingent conditioning, as can be seen from Example 5.4.1, where we do have strong independence, but not independence with respect to contingent conditioning.

71
5.5 Dempster’s rule of combination

In this section, we explain why we do not see any justification for Dempster’s rule of combination in general within the current interpretation.

We first describe how Dempster’s combined belief function is constructed. The combination rule deals with the situation in which there are two (different) belief functions on the same outcome space $\Omega$. We write $Bel_1$ and $Bel_2$ for these two belief functions. Shafer considers the auxiliary product space $\Omega \times \Omega$ as an outcome space, where the first coordinate is associated with the first belief function and the second coordinate is associated with the second belief function.

The construction of the “combined” belief function $Bel^* : 2^\Omega \rightarrow [0,1]$ now proceeds in two steps. First, we construct an auxiliary belief function $Bel$ on $\Omega \times \Omega$ by setting the basic belief assignment

$$m(A \times B) = m_1(A)m_2(B),$$

where $m_1$ and $m_2$ are the basic belief assignments corresponding to respectively $Bel_1$ and $Bel_2$. This means that $Bel_1$ and $Bel_2$ are considered strongly independent.

Then we condition in the mixed sense on $H = \{(\omega, \omega) : \omega \in \Omega\}$. We set the belief in $S \subseteq \Omega$ to be

$$Bel^*(S) := Bel_{H,\text{mix}}(\{(\omega, \omega) : \omega \in S\}).$$

It is easy to check that the basic belief assignment $m^*$ corresponding to $Bel^*$ is

$$m^*(S) = \frac{\sum_{A \cap B = S} m_1(A)m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A)m_2(B)}$$

for every $S \neq \emptyset$, matching Shafer’s original definition of combined belief.

After this factual description of the construction, we address the problems associated to it within the current interpretation. The fact that $Bel_1$ and $Bel_2$ are different belief functions on the same space, is interpreted as there being two different chance experiments $P_1$ and $P_2$ whose outcomes give information about the same phenomenon. The interpretation of $Bel_1$ and $Bel_2$ being strongly independent, as explained in the previous subsection, is that the chance experiments $P_1$ and $P_2$ are independent. The chance experiments give information about the same phenomenon, but since we want to condition on $H$, we must treat the fact that they are about the same phenomenon as something we initially do not know. After all, if we simply know a priori that $H$ must happen, this means that $m(S)$ can only be positive if $S \subseteq H$. Since $P_1$ and $P_2$ are independent, this is only possible in the trivial case that $m_1(\{\omega\}) = m_2(\{\omega\}) = 1$ for some $\omega \in \Omega$.

So we approach the chance experiments as experiments giving information about two phenomena that a priori may or may not be the same, of which we then learn that they are actually the same. The crux of the problem is the type of conditioning that is used. Learning that the two phenomena are actually the
5.5. DEMPSTER’S RULE OF COMBINATION

same, clearly corresponds to conditioning on $H$ in a necessary way, and not in a mixed way. Applying necessary conditioning instead of mixed conditioning, is only possible if $\Pr(H) = 1$. We have $\Pr(H) = 1$ precisely if for every $A, B \subseteq \Omega$ such that $m_1(A) > 0$ and $m_2(B) > 0$, we find that $A$ and $B$ are not disjoint. Only in this rather trivial circumstance, we can apply the combination rule and find a belief in $S \subseteq \Omega$ of

$$\text{Bel}^*(S) = \sum_{A, B : A \cap B \subseteq S} m_1(A)m_2(B).$$  \hfill (5.5.4)

The constraint of course heavily undermines the generality of the combination rule, because all cases in which there is actual conflict between the belief functions are excluded. Of course it are precisely these cases that are interesting and are emphasized when applying the rule. We end with an example in which the combination rule does makes sense.

**Example 5.5.1.** Suppose person A has three coins. Let us write $C_1$ and $C_2$ for the first two coins and let us write both $C_X$ and $C_Y$ for the third coin. So there are three physical coins, but four labels since we introduced two labels for the third coin. Person A flips $C_1$ and $C_2$ independently of each other. If one of them is heads, he puts $C_X = C_Y$ with heads facing up. In case they are both tails, he puts $C_X = C_Y$ with the side of his preference facing up. Suppose person B knows that (1) $C_1$ and $C_2$ are independently flipped, are distinct from each other and both distinct from $C_X$ and $C_Y$, (2) if $C_1$ is heads then $C_X$ is also heads, (3) if the $C_2$ is heads then $C_Y$ is also heads. But person B does initially not know that $C_X$ and $C_Y$ are labels for the same coin. Write $\Omega = \{h, t\}$ and let the outcome space $\Omega \times \Omega$ represent the side of $C_X$ (first coordinate) that is facing up and the side of $C_Y$ (second coordinate) that is facing up. The chance experiment that describes the information person B gets, is

$$m(\{(h, h)\}) = \frac{1}{4},$$

$$m(\{(h, h), (h, t)\}) = \frac{1}{4},$$

$$m(\{(h, h), (t, h)\}) = \frac{1}{4},$$

$$m(\Omega \times \Omega) = \frac{1}{4}. \hfill (5.5.5)$$

Note that this is precisely the same basic belief assignment as in Example 5.4.1. If person A now tells person B that $C_X$ and $C_Y$ are actually the same coin, the combined belief in that coin being heads now becomes

$$\text{Bel}^*(\{h\}) = m(\{(h, h)\}) + m(\{(h, h), (h, t)\}) + m(\{(h, h), (t, h)\}) = \frac{3}{4}. \hfill (5.5.6)$$
5.6 A law of large numbers

Since we only have developed our theory for finite outcome spaces, we only present a ‘weak’ law of large numbers. The result is quite similar to the weak law of large numbers of Marinacci [37], the key difference being that Marinacci proves it for convex capacities (a further generalization of belief functions) and random variables that satisfy a regularity condition that is not satisfied by all the cases we discuss here.

Let \( X : \Omega \rightarrow \mathbb{R} \), \( m : 2^\Omega \rightarrow [0, 1] \) be a basic belief assignment and \( \text{Bel} \) the corresponding belief function. To state our theorem, we need to generalize the concept ‘expectation’ to our setting.

**Definition 5.6.1.** The *expectation* of \( X \) (with respect to \( m \)) is
\[
\text{Exp}(X) := \sum_{C \subseteq \Omega} m(C) \min_{\omega \in C} X(\omega). \tag{5.6.1}
\]

The expression in (5.6.1) is known as the Choquet integral of \( X \) with respect to the belief function that corresponds to \( m \) (see for instance [22]), and is also called the *lower* expectation of \( X \) (see for instance [49]). The expectation can be seen as a generalization of belief in the sense that \( \text{Exp}(X) \) is the maximum price an agent basing his or her belief on \( m \) as a chance experiment, would pay for the bet that has \( X(\omega) \) as a net result if \( \omega \) turns out to be the outcome. Indeed, we find that if we take for \( A \subseteq \Omega \) the \( X = 1_A \), then we find
\[
\text{Exp}(1_A) = \sum_{C \subseteq \Omega} m(C) \min_{\omega \in C} 1_A(\omega) = \sum_{C \subseteq A} m(C) = \text{Bel}(A). \tag{5.6.2}
\]
Further, we find that in case \( \text{Bel} = P \) is a probability distribution, we have
\[
\text{Exp}(X) = \sum_{\omega \in \Omega} m(\{\omega\}) X(\omega) = \sum_{\omega \in \Omega} P(\{\omega\}) X(\omega) = E(X), \tag{5.6.3}
\]
and thus Definition 5.6.1 is consistent with the concept of expectation for probability distributions.

First, we want to show that in the long run, \( \text{Exp}(A) \) is an lower bound for the average of \( n \) independent ‘copies’ of \( X \). Secondly, we want to show that there is no bigger lower bound than \( \text{Exp}(A) \). We make this precise. On \( \Omega^n \) we define the basic belief assignment \( m_n : 2^{\Omega^n} \rightarrow [0, 1] \) as
\[
m_n(A_1 \times \cdots \times A_n) := \prod_{j=1}^n m(A_j), \tag{5.6.4}
\]
making all projections strongly independent. Let \( \text{Bel}_n \) be the corresponding belief function. We set \( X_j : \Omega^n \rightarrow \mathbb{R} \) by
\[
X_j((\omega_1, \omega_2, \ldots, \omega_n)) := X(\omega_j). \tag{5.6.5}
\]
Theorem 5.6.2. For every $\epsilon > 0$ we have

$$\lim_{n \to \infty} \text{Bel}_n \left( \frac{1}{n} \sum_{j=1}^{n} X_j \geq \text{Exp}(X) - \epsilon \right) = 1$$  \hspace{1cm} (5.6.6)

and

$$\lim_{n \to \infty} \text{Bel}_n \left( \frac{1}{n} \sum_{j=1}^{n} X_j \geq \text{Exp}(X) + \epsilon \right) = 0.$$  \hspace{1cm} (5.6.7)

Proof. Let $\epsilon > 0$ be given. Let $P : 2^{2^{\Omega}} \to [0,1]$ the probability distribution corresponding to $m$, i.e. $P(\{C\}) = m(\{C\})$ and define $P_n : (2^{\Omega})^n \to [0,1]$ by

$$P_n(\{(C_1, C_2, \ldots, C_n)\}) := \prod_{j=1}^{n} m(C_j) = \prod_{j=1}^{n} P(\{C_j\}).$$  \hspace{1cm} (5.6.8)

We define the random variable $\hat{X} : 2^{\Omega} \to \mathbb{R}$ by

$$\hat{X}(C) := \min_{\omega \in C} X(\omega)$$  \hspace{1cm} (5.6.9)

for $C \neq \emptyset$ and set $\hat{X}(\emptyset) = 0$. Let $\hat{X}_j : (2^{\Omega})^n \to \mathbb{R}$ be given by

$$\hat{X}_j((C_1, C_2, \ldots, C_n)) := \hat{X}(C_j).$$  \hspace{1cm} (5.6.10)

Observe that for any $\alpha \in [0,1]$ we have

$$\text{Bel}_n \left( \frac{1}{n} \sum_{j=1}^{n} X_j \geq \alpha \right) = P_n \left( \left\{ (C_1, C_2, \ldots, C_n) : \min_{\omega_j \in C_j} \frac{1}{n} \sum_{j=1}^{n} X(\omega_j) \geq \alpha \right\} \right)$$

$$= P_n \left( \left\{ (C_1, C_2, \ldots, C_n) : \frac{1}{n} \sum_{j=1}^{n} \hat{X}(C_j) \geq \alpha \right\} \right)$$

$$= P_n \left( \frac{1}{n} \sum_{j=1}^{n} \hat{X}_j \geq \alpha \right).$$  \hspace{1cm} (5.6.11)

The (classical) expectation of the $\hat{X}_j$ is

$$E(\hat{X}_j) = E(\hat{X}) = \sum_{C \subseteq \Omega} P(\{C\}) \hat{X}(C) = \sum_{C \subseteq \Omega} m(C) \min_{\omega \in C} X(\omega) = \text{Exp}(X)$$  \hspace{1cm} (5.6.12)
and by the definition of $P_n$ all the $\hat{X}_1, \ldots, \hat{X}_n$ are (classically) independent. With the classical weak law of large numbers, we then find that

$$\lim_{n \to \infty} \text{Bel}_n \left( \frac{1}{n} \sum_{j=1}^{n} X_j \geq \text{Exp}(X) - \epsilon \right) = \lim_{n \to \infty} P_n \left( \frac{1}{n} \sum_{j=1}^{n} \hat{X}_j \geq E(\hat{X}) - \epsilon \right) = 1$$ (5.6.13)

and

$$\lim_{n \to \infty} \text{Bel}_n \left( \frac{1}{n} \sum_{j=1}^{n} X_j \geq \text{Exp}(X) + \epsilon \right) = \lim_{n \to \infty} P_n \left( \frac{1}{n} \sum_{j=1}^{n} \hat{X}_j \geq E(\hat{X}) + \epsilon \right) = 0.$$ (5.6.14)
6.1 Betting functions and degrees of belief

We approach belief behaviorally, working under the assumption that the degree to which someone believes an event, can be measured by his or her willingness to accept bets. To go towards a definition that we can work with mathematically, we introduce betting functions.

**Definition 6.1.1.** A bet (or gamble) on \( \Omega \) is a function \( X : \Omega \to \mathbb{R} \). We write \( \mathcal{X} = \mathbb{R}^{X} \) for the collection of all bets on \( \Omega \). A betting function is a function \( R : \mathcal{X} \to \{0, 1\} \) such that for each \( X \in \mathcal{X} \), there is an \( \alpha_0(X) \in \mathbb{R} \) such that \( R(X + \alpha) = 0 \) for \( \alpha < \alpha_0 \) and \( R(X + \alpha) = 1 \) for \( \alpha \geq \alpha_0 \).

We interpret \( R \) as a function that indicates, for each bet \( X \in \mathcal{X} \), whether or not an agent is willing to accept \( X \), where we interpret \( R(X) = 1 \) as ‘willing to accept the bet’ and \( R(X) = 0 \) ‘not willing to accept the bet’. The definition of a betting function justifies the following definition.

**Definition 6.1.2.** Let \( R \) be a betting function. The **buy function** \( \text{Buy}_R : \mathcal{X} \to \mathbb{R} \) is given by

\[
\text{Buy}_R(X) := \max\{\alpha \in \mathbb{R} : R(X - \alpha) = 1\}. \tag{6.1.1}
\]

This is the maximum price an agent is willing to pay for a bet which pays out \( X \). The **sell function** \( \text{Sell}_R : \mathcal{X} \to \mathbb{R} \) is given by

\[
\text{Sell}_R(X) := \min\{\alpha \in \mathbb{R} : R(\alpha - X) = 1\}. \tag{6.1.2}
\]

This is the minimum price an agent is willing to sell the bet which pays out \( X \).

The buy and sell function have the following relation:

\[
\text{Buy}_R(X) = \max\{\alpha \in \mathbb{R} : R(X - \alpha) = 1\} = \max\{-\alpha \in \mathbb{R} : R(X + \alpha) = 1\} = -\min\{\alpha \in \mathbb{R} : R(\alpha - (-X)) = 1\} = -\text{Sell}_R(-X). \tag{6.1.3}
\]
This shows how buy and sell functions are dual in the sense that \(\text{Buy}_R\) is completely determined by \(\text{Sell}_R\) and vice versa. This relation between Buy and Sell is precisely the relation between lower and upper previsions in [61], but in our case the relation follows from the underlying concept of a betting function. Note that we can recover \(R\) from \(\text{Buy}_R\), since

\[
R(X) = \begin{cases} 1 & \text{if } \text{Buy}_R(X) \geq 0, \\ 0 & \text{if } \text{Buy}_R(X) < 0. \end{cases}
\]  

(6.1.4)

**Definition 6.1.3.** Let \(R : \mathcal{X} \to \{0, 1\}\) be a betting function. Then \(R\) is said to be *coherent* if

(R1) If \(X > 0\), then \(R(X) = 1\);

(R2) \(R(\lambda X) = R(X)\) for every \(\lambda > 0\);

(R3) If \(R(X) = R(Y) = 1\), then \(R(X + Y) = 1\).

These conditions do not necessarily capture everything a reasonable agent should adhere to, and for rational behavior more is needed as we will see. However, the conditions formulate the very basics. Condition (R1) says that agents should be willing to accept bets with guaranteed positive results. Condition (R2) says that willingness to accept bets should only depend on the relative sizes of the results, but not on their absolute sizes. Condition (R3) says that if an agent is willing to accept two bets, (s)he should be willing to accept bets simultaneously. We want to point out that both condition (R2) and (R3) have a debatable consequence: if an agent is willing to accept a bet in which (s)he wins 1 euro if \(A\) is true and loses 1 euro if \(A\) is false, then (s)he should be willing to accept a bet in which (s)he wins 1 million euro if \(A\) is true and loses 1 million euro if \(A\) is false. In the real world, one could think of all kinds of reasons why an agent would not want to accept the second bet, even if (s)he is willing to accept the first one. By imposing coherence, we thus consider highly idealized agents.

The following result says that \(\text{Buy}_R\) is coherent as a lower prevision in the sense of Walley [61] if and only if \(R\) is coherent in the sense of Definition 6.1.3.

**Theorem 6.1.4.** Let \(R : \mathcal{X} \to \{0, 1\}\) be a betting function. Then \(R\) is coherent if and only if

- \(\text{Buy}_R(X) \geq \min X\);
- \(\text{Buy}_R(\lambda X) = \lambda \text{Buy}_R(X)\) for every \(\lambda > 0\);
- \(\text{Buy}_R(X + Y) \geq \text{Buy}_R(X) + \text{Buy}_R(Y)\).

**Proof.** Suppose \(R\) is coherent. By (R1), we have that \(R(X - \min X + \epsilon) = 1\) for every \(\epsilon > 0\). Hence

\[
\text{Buy}_R(X) = \max\{\alpha \in \mathbb{R} : R(X - \alpha) = 1\} \geq \min X - \epsilon
\]  

(6.1.5)
for every $\epsilon > 0$, so $\text{Buy}_R(X) \geq \min X$. By (R2), we have for every $\lambda > 0$ that

$$\text{Buy}_R(\lambda X) = \max\{\alpha \in \mathbb{R} : R(\lambda X - \alpha) = 1\}$$

$$= \max\{\alpha \in \mathbb{R} : R\left(\frac{X - \alpha}{\lambda}\right) = 1\}$$

$$= \max\{\lambda \alpha \in \mathbb{R} : R(X - \alpha) = 1\}$$

$$= \lambda \max\{\alpha \in \mathbb{R} : R(X - \alpha) = 1\}$$

$$= \lambda \text{Buy}_R(X).$$

Suppose that, for some $\alpha, \beta \in \mathbb{R}$, we have $R(X - \alpha) = 1$ and $R(Y - \beta) = 1$. Then, by (R3), we have $R(X + Y - \alpha - \beta) = 1$ and thus

$$\text{Buy}_R(X + Y) = \max\{\gamma \in \mathbb{R} : R(X + Y - \gamma) = 1\} \geq \alpha + \beta. \quad (6.1.7)$$

It follows that

$$\text{Buy}_R(X + Y) \geq \max\{\alpha : R(X - \alpha) = 1\} + \max\{\beta : R(X - \beta) = 1\}$$

$$= \text{Buy}_R(X) + \text{Buy}_R(Y). \quad (6.1.8)$$

For the converse, suppose that $\text{Buy}_R$ has the listed properties. We set $\psi : \mathbb{R} \to \{0, 1\}$ by

$$\psi(x) = \begin{cases} 
1 & \text{if } x \geq 0, \\
0 & \text{if } x < 0,
\end{cases} \quad (6.1.9)$$

and we note as before that

$$R(X) = \psi(\text{Buy}_R(X)). \quad (6.1.10)$$

If $X > 0$, then $\text{Buy}_R(X) \geq 0$ by the first property. Hence $R(X) = \psi(\text{Buy}_R(X)) = 1$. So (R1) holds.

The second property tells us that, for every $\lambda > 0$, we have

$$R(\lambda X) = \psi(\text{Buy}_R(\lambda X)) = \psi(\lambda \text{Buy}_R(X)) = \psi(\text{Buy}_R(X)) = R(X), \quad (6.1.11)$$

so (R2) holds.

By the third property, we have

$$R(X + Y) = \psi(\text{Buy}_R(X + Y))$$

$$\geq \psi(\text{Buy}_R(X) + \text{Buy}_R(Y))$$

$$\geq \min\{\psi(\text{Buy}_R(X)), \psi(\text{Buy}_R(Y))\}$$

$$= \min\{R(X), R(Y)\}. \quad (6.1.12)$$

Hence if $R(X) = R(Y) = 1$, it follows that $R(X + Y) = 1$, so (R3) holds.}

The second and third property of Theorem 6.1.4 tell us that $\text{Buy}_R$ is a super-linear functional if $R$ is coherent. Coherence of $R$ can of course also be captured in terms of $\text{Sell}_R$:
Theorem 6.1.5. Let $R : \mathcal{X} \to \{0, 1\}$ be a betting function. Then $R$ is coherent if and only if

- $\text{Sell}_R(X) \leq \max X$;
- $\text{Sell}_R(\lambda X) = \lambda \text{Sell}_R(X)$ for every $\lambda > 0$;
- $\text{Sell}_R(X + Y) \leq \text{Sell}_R(X) + \text{Sell}_R(Y)$.

Proof. This follows directly from Theorem 6.1.4 and the relation between $\text{Buy}_R$ and $\text{Sell}_R$.

We want to measure the degree to which an agent believes $A \subseteq \Omega$ by the willingness to accept bets with a desirable (i.e., positive) result if $A$ is true and an undesirable (i.e., negative) result if $A$ is false. A natural choice is the bet $1_A - \alpha$ for $0 \leq \alpha \leq 1$, which is the bet an agent is willing to accept if and only if (s)he wants to buy the bet $1_A$ for a price of $\alpha$. It is willingness to buy $1_A$ for a high price and willingness to sell $1_{A^c}$ for a low price, that shows a high degree of belief in $A$. This leads to our definition of degree of belief.

Definition 6.1.6 (Degree of belief). Let $R$ be the coherent betting function of an agent. We define the degree to which this agent believes an event $A \subseteq \Omega$ as $\text{Buy}_R(1_A) = 1 - \text{Sell}_R(1_{A^c})$.

6.2 Adding B-consistency

In this section, we introduce an additional condition for betting functions and show that this constraint precisely leads to buy functions which are belief functions when restricted to bets of the form $1_A$, with $A \subseteq \Omega$. Before we do this, however, we will briefly discuss how our setup relates to the Dutch Book argument for probability distributions.

The Dutch Book argument is centered around the principle that betting behavior of agents should not lead to sure loss. The following theorem tells us that not having a sure loss is already implied by coherence. This theorem is well known within Walley’s theory, but we give our version of the proof as a service to the reader.

Theorem 6.2.1. If $R$ is a coherent betting function, then

$$\max_{\omega \in \Omega} \left( \sum_{i=1}^{N} (X_i - \text{Buy}_R(X_i)) + \sum_{j=1}^{M} (\text{Sell}_R(Y_j) - Y_j) \right) \geq 0 \quad (6.2.1)$$

for all $X_1, \ldots, X_N \in \mathcal{X}$ and $Y_1, \ldots, Y_M \in \mathcal{X}$.

Proof. We first show that $\text{Buy}_R(X) \leq \max X$ for each $X \in \mathcal{X}$. Suppose that $\text{Buy}_R(X) > \max X$. This means there is an $\epsilon > 0$ such that

$$R(X - \max X - \epsilon) = 1. \quad (6.2.2)$$
Note that for every $Y \in \mathcal{X}$, there is a $\lambda > 0$ such that $\lambda(X - \max X - \epsilon) < Y$. Since $R(\lambda(X - \max X - \epsilon)) = 1$ by (R2) and $R(Y - \lambda(X - \max X - \epsilon)) = 1$ by (R1), it follows with (R3) that $R(Y) = 1$. Hence $R(Y) = 1$ for all $Y \in \mathcal{X}$, but then there is no maximum $\alpha \in \mathbb{R}$ such that $R(Y - \alpha) = 1$. This is a contradiction, and it follows that $\text{Buy}_R(X) \leq \max X$.

Now let $X_1, \ldots, X_N \in \mathcal{X}$ and $Y_1, \ldots, Y_M \in \mathcal{X}$. By using Theorem 6.1.4 and the property we just proved, we find

$$\sum_{i=1}^{N} \text{Buy}_R(X_i) - \sum_{j=1}^{M} \text{Sell}_R(Y_j) = \sum_{i=1}^{N} \text{Buy}_R(X_i) + \sum_{j=1}^{M} \text{Buy}_R(-Y_j) \leq \text{Buy}_R\left(\sum_{i=1}^{N} X_i + \sum_{j=1}^{M} -Y_j\right) \leq \max \left(\sum_{i=1}^{N} X_i + \sum_{j=1}^{M} -Y_j\right).$$

(6.2.3)

Theorem 6.1.5 tells us that coherence is stronger than having no sure loss. At the same time, even coherence does not imply that $A \mapsto \text{Buy}_R(1_A)$ is a probability distribution: it is clear that the collection of maps $A \mapsto \text{Buy}_R(1_A)$ for coherent $R$, is much richer than only probability distributions. This tells us that the property that $\text{Buy}_R(1_A) = \text{Sell}_R(1_A)$ for every $A \subseteq \Omega$, which is usually implicitly assumed when the Dutch Book argument is laid out, is crucial for the restriction to probability distributions. The next theorem confirms this.

**Theorem 6.2.2.** A function $P : 2^\Omega \rightarrow \mathbb{R}$ is a probability distribution if and only if there exists a coherent betting function $R$ such that $P(A) = \text{Buy}_R(1_A) = \text{Sell}_R(1_A)$.

**Proof.** Suppose there is a coherent $R$ such that $P(A) = \text{Buy}_R(1_A) = \text{Sell}_R(1_A)$. We have $P(\Omega) = \text{Buy}_R(1_\Omega) = 1$ and $P(A) = \text{Buy}_R(1_A) \geq 0$ by coherence. Then for disjoint $A, B \subseteq \Omega$ we find

$$P(A \cup B) = \text{Buy}_R(1_A + 1_B) \geq \text{Buy}_R(1_A) + \text{Buy}_R(1_B) = P(A) + P(B) \quad (6.2.4)$$

and

$$P(A \cup B) = \text{Sell}_R(1_A + 1_B) \leq \text{Sell}_R(1_A) + \text{Buy}_R(1_B) = P(A) + P(B). \quad (6.2.5)$$

Hence $P$ is additive.

For the converse, suppose that $P$ is a probability distribution. Then define

$$\text{Buy}_R(X) = \sum_{\omega \in \Omega} P(\{\omega\})X(\omega). \quad (6.2.6)$$

Clearly, $R$ is coherent and we have $\text{Buy}_R(X) = \text{Sell}_R(X)$. \qed
CHAPTER 6. A BETTING INTERPRETATION

As we already mentioned, the constraint that \( \text{Buy}_R(1_A) = \text{Sell}_R(1_A) \) for every \( A \subseteq \Omega \), is precisely one we do not want to impose. This property, however, is not easily weakened in a reasonable way. Therefore, we will work towards another characterization of probability distributions (Theorem 6.2.7), from which we can derive our constraint. We start with the definition of a belief valuation.

**Definition 6.2.3.** A function \( \mathcal{B} : 2^\Omega \rightarrow \{0, 1\} \) is called a belief valuation provided that

- \( \mathcal{B}(A^c) = 0 \) if \( \mathcal{B}(A) = 1 \);
- If \( \mathcal{B}(A) = 1 \) and \( A \subseteq B \), then \( \mathcal{B}(B) = 1 \);
- If \( \mathcal{B}(A) = 1 \) and \( \mathcal{B}(B) = 1 \), then \( \mathcal{B}(A \cap B) = 1 \);
- \( \mathcal{B}(\Omega) = 1 \).

A belief valuation is also called a categorical belief function or a 0-1 necessity measure in the literature. In practice, we use the characterization that \( \mathcal{B} \) is a belief valuation if and only if there is a nonempty set \( E_{\mathcal{B}} \subseteq \Omega \) such that

\[
\mathcal{B}(A) = \begin{cases} 
1 & \text{if } E_{\mathcal{B}} \subseteq A \\
0 & \text{otherwise} 
\end{cases} \tag{6.2.7}
\]

We can also describe belief valuations in terms of filters, since \( \mathcal{B} \) is a belief valuation if and only if

\[
\{ A \subseteq \Omega : \mathcal{B}(A) = 1 \} \subseteq 2^\Omega \tag{6.2.8}
\]

is a proper filter of subsets of \( \Omega \). This filter can be interpreted as the collection of sets in which an agent has full belief. The next result makes this precise. For \( R \) a coherent betting function, we denote by \( \mathcal{B}_R : 2^\Omega \rightarrow \{0, 1\} \) the function that satisfies \( \mathcal{B}_R(A) = 1 \) if and only if \( \text{Buy}_R(1_A) = 1 \).

**Theorem 6.2.4.** A function \( \mathcal{B} : 2^\Omega \rightarrow \{0, 1\} \) is a belief valuation if and only if there is a coherent betting function \( R \) such that \( \mathcal{B} = \mathcal{B}_R \).

**Proof.** Suppose first that \( \mathcal{B} \) is a belief valuation. We define \( R \) by

\[
\text{Buy}_R(X) := \min_{\omega \in E_{\mathcal{B}}} X(\omega).
\]

Clearly \( R \) is coherent and it follows from the definition of \( \mathcal{B}_R \) that \( \mathcal{B}_R = \mathcal{B} \).

For the converse, suppose that \( R \) is a coherent betting function. We check that \( \mathcal{B}_R \) is a belief valuation. Since \( \text{Buy}_R(1_\Omega) = 1 \) we have \( \mathcal{B}_R(\Omega) = 1 \). The second property in Definition 6.2.3 follows immediately from the monotonicity of \( \text{Buy}_R \). If \( \text{Buy}_R(A) = \text{Buy}_R(B) = 1 \), then

\[
\text{Buy}_R(1_{A \cap B}) = \text{Buy}_R(1_A + 1_B - 1_{A \cup B}) \\
\geq \text{Buy}_R(1_A) + \text{Buy}_R(1_B) + \text{Buy}_R(1_{A \cup B}) \geq 1,
\]

82
so \( \text{Buy}_R(1_{A \cap B}) = 1 \). Finally, if \( B_R(A) = 1 \), then since

\[
1 = \text{Buy}_R(1_A + 1_{A^c}) \geq \text{Buy}_R(1_A) + \text{Buy}_R(1_{A^c}),
\]

it follows that \( \text{Buy}_R(1_{A^c}) = 0 \).

A belief valuation should be compared with the classical notion of a truth valuation. The definition of a truth valuation \( T : 2^\Omega \to \{0, 1\} \) is similar to the definition of a belief valuation, the only difference being that in the first bullet, ‘if’ is replaced by ‘if and only if’. Hence a truth valuation is a special belief valuation, namely one that corresponds to a proper ultrafilter of sets. Truth valuations are precisely those belief valuations \( B \) for which \( E_B \) is a singleton. A major difference between truth and belief is that if an agent does not believe in \( A \), this does not imply that (s)he does believe in \( A^c \). As a result, the implication in the first bullet in the definition of a belief valuation goes in one direction only.

Given a belief valuation \( B \) and a set \( S \subseteq \Omega \) for which \( B(S) = 1 \), the agent fully believes that a bet \( X \in \mathcal{X} \) has a revenue of at least

\[
\min_{\omega \in S} X(\omega). \tag{6.2.9}
\]

Since this holds for all \( S \) for which \( B(S) = 1 \), this leads to the definition of guaranteed revenue.

**Definition 6.2.5.** For any belief valuation \( B \), the guaranteed revenue \( G_B : \mathcal{X} \to \mathbb{R} \) is defined as

\[
G_B(X) = \max_{A : B(A) = 1} \min_{\omega \in A} X(\omega).
\]

Since we have that \( B(A) = 1 \) if and only if \( E_B \subseteq A \), we can express the guaranteed revenue as

\[
G_B(X) = \min_{\omega \in E_B} X(\omega). \tag{6.2.10}
\]

To benchmark and motivate our main result Theorem 6.2.10 below we now first show how classical probability distributions can be characterized with the notion of guaranteed revenue.

**Definition 6.2.6.** A betting function \( R \) is \( P \)-consistent if and only if, for all \( X_1, \ldots, X_N \) and \( Y_1, \ldots, Y_M \) such that

\[
G_B \left( \sum_{i=1}^{N} X_i \right) \leq G_B \left( \sum_{j=1}^{M} Y_j \right) \tag{6.2.11}
\]

for every belief valuation \( B \), we have

\[
\sum_{i=1}^{N} \text{Buy}_R(X_i) \leq \sum_{j=1}^{M} \text{Buy}_R(Y_j). \tag{6.2.12}
\]
Theorem 6.2.7. $P$ is a probability distribution if and only if there exists a coherent and $P$-consistent $R$ such that $P(A) = \text{Buy}_R(1_A)$.

In words, this result says that if the guaranteed revenue of one collection of bets is larger than the guaranteed revenue of a second collection of bets, then the agent should be willing to pay more for the second collection.

The proof of the theorem below reveals that the statement of the theorem is, strictly speaking, overly complicated. Indeed, if we left out $G_B$ everywhere, the ensuing result would still be true, and probably easier to interpret: the theorem would say that if one collection of bets always pays out more than a second collection, an agent would be willing to pay more for the first collection. We have chosen for the current formulation since this formulation points the way for the necessary changes.

**Proof.** (of Theorem 6.2.7.) First suppose that $R$ is coherent, P-consistent and that $P(A) = \text{Buy}_R(1_A)$. Suppose $A \cap B = \emptyset$. Then

$$G_B(1_A + 1_B) = G_B(1_{A \cup B}) \quad (6.2.13)$$

for every $B$, so by P-consistency we have

$$\text{Buy}_R(1_A + 1_B) = \text{Buy}_R(1_A) + \text{Buy}_R(1_B). \quad (6.2.14)$$

Hence $P(A \cup B) = P(A) + P(B)$. Since $R$ is coherent, we have $P(\Omega) = \text{Buy}_R(1_\Omega) = 1$ and it follows that $P$ is a probability distribution.

Now suppose that $P$ is a probability distribution. We define $R$ by

$$\text{Buy}_R(X) := \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}). \quad (6.2.15)$$

Clearly $\text{Buy}_R(1_A) = P(A)$, and it follows from Theorem 6.1.4 that $R$ is coherent. Now suppose that

$$G_B \left( \sum_{i=1}^{N} X_i \right) \leq G_B \left( \sum_{j=1}^{M} Y_j \right) \quad (6.2.16)$$

for every belief valuation $B$. For every $\omega \in \Omega$, the map $B_\omega(A) := 1_A(\omega)$ is a belief valuation. Note that $G_{B_\omega}(X) = X(\omega)$, so (6.2.16) implies that

$$\sum_{i=1}^{N} X_i \leq \sum_{j=1}^{M} Y_j. \quad (6.2.17)$$
Hence, by our definition of \( \text{Buy}_R \):

\[
\sum_{i=1}^{N} \text{Buy}_R(X_i) = \text{Buy}_R \left( \sum_{i=1}^{N} X_i \right) \\
\leq \text{Buy}_R \left( \sum_{j=1}^{M} Y_j \right) \tag{6.2.18}
\]

\[
= \sum_{j=1}^{M} \text{Buy}_R(Y_j).
\]

So \( R \) is P-consistent. \( \square \)

Although we do not deny that the notion of P-consistency is in some sense a reasonable requirement for collections of bets, there is, from the point of view of epistemic probability, a problem with it. Whereas in (6.2.11) the guaranteed revenues of the combined bets are compared, in (6.2.12) the sums of the buy prices of the individual bets are compared. This observation goes back to the heart of the problem with the use of classical probability distributions for epistemic purposes: the guaranteed revenue of the combined bets \( 1_A \) and \( 1'_A \) is - of course - the same as the guaranteed revenue of the bet 1, under any belief valuation. But this should not have any direct implication for the maximum price for which an agent would be willing to buy \( 1_A \) or \( 1'_A \) individually.

This suggests that for epistemic purposes, we should change P-consistency in one of the following two ways: in (6.2.11) the sums of the guaranteed revenues of the individual bets should be compared, or in (6.2.12) the buy prices of the combined bets should be compared. The first option leads to our definition of B-consistency.

**Definition 6.2.8.** A betting function \( R \) is **B-consistent** if for all \( X_1, \ldots, X_N \) and \( Y_1, \ldots, Y_M \) such that

\[
\sum_{i=1}^{N} G_B(X_i) \leq \sum_{j=1}^{M} G_B(Y_j) \tag{6.2.19}
\]

for every belief valuation \( B \), we have

\[
\sum_{i=1}^{N} \text{Buy}_R(X_i) \leq \sum_{j=1}^{M} \text{Buy}_R(Y_j). \tag{6.2.20}
\]

There is a simple reason for this choice. Indeed, the alternative would result in the constraint that if \( G_B(\sum_i X_i) \leq G_B(\sum_j Y_j) \) for every belief valuation \( B \), then \( \text{Buy}_R(\sum_i X_i) \leq \text{Buy}_R(\sum_j Y_j) \). This constraint, however, is simply B-consistency for \( N = M = 1 \).

Hence, the notion of B-consistency differs from P-consistency in the sense that we compare the correct quantities: we compare sums of guaranteed revenues of individual (collective) bets to sums of buy prices of individual (collective) bets.
The next example illustrates that not every coherent betting function is $B$-consistent.

**Example 6.2.9.** Suppose $\Omega = \{1, 2, 3, 4\}$ and that $R$ is given by

$$
\text{Buy}_R(X) = \min \left\{ \frac{1}{2} \sum_{i=1}^{2} X(i), \frac{1}{4} \sum_{i=1}^{4} X(i) \right\}.
$$

(6.2.21)

It is easy to check that $R$ is coherent, and that for all belief valuations $B$ we have

$$
G_B(1_{\{2,3,4\}}) + G_B(1_{\{2\}}) \geq G_B(1_{\{2,3\}}) + G_B(1_{\{2,4\}}),
$$

(6.2.22)

but

$$
\text{Buy}_R(1_{\{2,3,4\}}) + \text{Buy}_R(1_{\{2\}}) = \frac{3}{4} < 1 = \text{Buy}_R(1_{\{2,3\}}) + \text{Buy}_R(1_{\{2,4\}}).
$$

(6.2.23)

With the notion of $B$-consistency we can now state and prove our main result. The following theorem constitutes our behavioral interpretation of belief functions. It shows that only $B$-consistency is needed to guarantee that a lower prevision in the sense of Walley [61] is in fact a belief function and that every belief function can be obtained this way. The result gives an argument for the use of belief functions for modeling epistemic probability.

**Theorem 6.2.10.** Bel is a belief function if and only if there exists a coherent and $B$-consistent $R$ such that $\text{Bel}(A) = \text{Buy}_R(1_A)$ for each $A$.

This result follows immediately from the following theorem which is interesting in its own right and characterizes $B$-consistency for coherent betting functions. It tells us that $R$ is $B$-consistent if and only if $\text{Bel}(A) := \text{Buy}_R(1_A)$ is a belief function and $\text{Buy}_R(X)$ is given by expectation of $X$. We now give the proof.

**Theorem 6.2.11.** Let $R$ be a coherent betting function. Then $R$ is $B$-consistent if and only if there is a basic belief assignment $m$ such that

$$
\text{Buy}_R(X) = \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} X(\omega)
$$

(6.2.24)

for all $X \in \mathcal{X}$.

**Proof of Theorem 6.2.11.** First suppose that there is a basic belief assignment $m$ such that

$$
\text{Buy}_R(X) = \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} X(\omega)
$$

for all $X \in \mathcal{X}$. Also suppose that for $X_1, \ldots, X_N, Y_1, \ldots, Y_M \in \mathcal{X}$ we have

$$
\sum_{i=1}^{N} \min_{\omega \in S} X_i(\omega) \leq \sum_{j=1}^{M} \min_{\omega \in S} Y_j(\omega)
$$

(6.2.25)
for every nonempty $S \subseteq \Omega$. Then

$$
\sum_{i=1}^{N} \text{Buy}_R(X_i) = \sum_{i=1}^{N} \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} X_i(\omega)
$$

$$= \sum_{S \subseteq \Omega} m(S) \sum_{i=1}^{N} \min_{\omega \in S} X_i(\omega)
$$

$$\leq \sum_{S \subseteq \Omega} m(S) \sum_{j=1}^{M} \min_{\omega \in S} Y_j(\omega)
$$

$$= \sum_{j=1}^{M} \text{Buy}_R(Y_j).
$$

Hence $R$ is B-consistent.

For the converse of the theorem, suppose that $R$ is B-consistent. First we show that Bel$(A) := \text{Buy}_R(1_A)$ is a belief function. Clearly we have

$$G_B(1_A) = B(A),
$$

and since $B$ is a belief function, we have

$$B\left(\bigcup_{i=1}^{N} A_i\right) + \sum_{\{I \subseteq \{1, \ldots, n\} \mid I \neq \emptyset, |I| \text{ even}} B\left(\bigcap_{i \in I} A_i\right) \geq \sum_{\{I \subseteq \{1, \ldots, n\} \mid I \neq \emptyset, |I| \text{ odd}} B\left(\bigcap_{i \in I} A_i\right).
$$

So by B-consistency, we have

$$\text{Bel}\left(\bigcup_{i=1}^{N} A_i\right) + \sum_{\{I \subseteq \{1, \ldots, n\} \mid I \neq \emptyset, |I| \text{ even}} \text{Bel}\left(\bigcap_{i \in I} A_i\right) \geq \sum_{\{I \subseteq \{1, \ldots, n\} \mid I \neq \emptyset, |I| \text{ odd}} \text{Bel}\left(\bigcap_{i \in I} A_i\right),
$$

and it follows that Bel is a belief function.

Since Bel is a belief function, there is a $m : 2^\Omega \to [0, 1]$ with $m(\emptyset) = 0$ such that

$$\text{Bel}(A) = \sum_{S \subseteq A} m(S).
$$

Now let $X \in \mathcal{X}$. We write $X(\Omega) = \{y_1, \ldots, y_K\}$ where $y_1 < y_k < \ldots < y_K$. Set $y_0 := 0$. It is well known that the Choquet integral of $X$ can be expressed as

$$\sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} X(\omega) = \sum_{j=1}^{K} (y_j - y_{j-1}) \text{Bel}(\{X \geq y_j\}).
$$
We have
\[
\sum_{k=1}^{K} \min_{\omega \in S} (y_k - y_{k-1}) 1_{\{X \geq y_k\}}(\omega) = \sum_{k=1}^{K} (y_k - y_{k-1}) 1(S \subseteq \{X \geq y_k\}) = \min_{\omega \in S} X(\omega).
\] (6.2.32)

So by B-consistency we find that
\[
\sum_{k=1}^{K} \text{Buy}_R((y_k - y_{k-1}) 1_{\{X \geq y_k\}})) = \text{Buy}_R(X).
\] (6.2.33)

Now it follows with (6.2.31), (6.2.33) and coherence of \( R \) that
\[
\text{Buy}_R(X) = \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} X(\omega).
\] (6.2.34)

\[
\square
\]

6.3 Discussion

We have first argued that the classical axioms of Kolmogorov are not suitable for modeling epistemic probability. This has been known for a long time, and researchers like Glenn Shafer and Peter Walley have developed a general theory of belief functions, respectively lower provisions, as an alternative for classical probability theory in an epistemological context.

In Sections 6.1 and 6.2 we have developed a behavioral interpretation of belief functions. We have embedded belief functions in a betting context, and we have shown that belief functions arise precisely when we add B-consistency to the coherent lower previsions in the theory of Walley. In this way, not only do we provide a behavioral interpretation of belief functions, but we also make a natural connection between the theory of Walley and the theory of Shafer.

Of course adding B-consistency calls for an argument why a rational agent should adhere to it. We think it is not controversial to say that no guaranteed losses (Theorem 6.1.5) is the bottom line for any reasonable constraint. Beyond that things are, of course, debatable. Our argument to restrict the lower previsions of Walley to B-consistent lower previsions, is derived from the way we obtained it, namely by altering P-consistency in a very reasonable way. If a collection of bets is in some sense guaranteed to be better than another collection of bets, then the total price should be higher. The point is now how “better” should be formulated. When compared to P-consistency, B-consistency allows for all the flexibility of epistemic probability that we asked for in the introduction. Indeed, note that it is possible to set \( m(\Omega) = 1 \), which corresponds to total ignorance apart from the fact that the outcome is in \( \Omega \). We think it is rational to compare collection of bets from the viewpoint of total guaranteed revenue, and to be willing
6.4 CONTINGENT CONDITIONING

to pay more when this quantity increases. As such, we think that B-consistency is a very reasonable constraint for a rational agent to adhere to.

But there is more than philosophy, of course. We also want a theory that is practical and relatively easy to apply. Belief functions are close to classical probabilities since they are determined by basic belief assignments. As such, we think they are much more practical and easier to use than coherent lower previsions. Indeed, in practical situations, one does not directly use constraints as coherence or B-consistency to construct a buy function. In case of belief functions, one typically proceeds by constructing a suitable basic belief assignment, see for instance Chapter 7, in which we apply belief functions in the classical forensic context of the so called island problem, precisely by setting up appropriate basic belief assignments. This is very much akin to the classical situation: not many people use characterizations like P-consistency (or related characterizations) to set up a probability measure, but it is reassuring that such characterizations exist. Hence we should view B-consistency as a theoretical underpinning for why to use belief functions, but not as a tool that is used in practice to construct belief functions.

We will now discuss contingent conditioning, necessary conditioning and strong independence in the following sections. We note that by interpreting these notions, we are able to mimic exactly the critique of Demspter’s rule within the frequestist interpretation laid down in Section 5.5, in the betting interpretation. Hence we do not explicitly copy that in this chapter.

6.4 Contingent conditioning

We assume here that \( \text{Sell}_R(1_H) > 0 \). We already know how to condition in a contingent way on \( H \) from the frequentist perspective of Chapter 5. If \( m \) is the basic belief assignment corresponding to \( R \), we argue in this section that the conditional betting function \( R_H \) should satisfy

\[
\text{Buy}_{R_H}(X) := \sum_{S \subseteq \Omega} m_H(S) \min_{\omega \in S} X(\omega),
\]

where \( m_H \) is the basic belief assignment from Section 5.1.

In the betting context, we mean by learning \( H \) in a contingent way that the agent learns that any result of the bet \( X \) is ignored if \( H^c \) occurs. We quite literally translate that into the betting function \( R^*(X) := R(1_H X) \), conform the conditional lower previsions of Walley [61]. The following theorem shows that \( R_H \) does coincide with \( R^* \) for an important set of bets.

**Theorem 6.4.1.** We have

\[
R_H(X) = R^*(X)
\]

for all \( X \in \mathcal{X} \) such that \( X = x1_A + y1_{A^c} \) for some \( A \subseteq \Omega \).
**Proof.** Let $X \in \mathcal{X}$ such that $X = x1_A + y1_{A^c}$ for some $A \subseteq \Omega$. Assume, without loss of generality, that $x \geq y$. First we compute

$$
Buy_{R^*}(1_A) = \max\{\alpha : R^*(1_A - \alpha) = 1\} = \max\{\alpha : R(1_H(1_A - \alpha)) = 1\}.
$$

(6.4.3)

We observe that $R(1_H(1_A - \alpha)) = 1$ if and only if

$$
Buy_R(1_H(1_A - \alpha)) = \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} 1_H(\omega)(1_A(\omega) - \alpha) \geq 0.
$$

(6.4.4)

We also observe, since $Buy_R(1_H) > 0$, that

$$
\alpha \mapsto \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} 1_H(\omega)(1_A(\omega) - \alpha)
$$

is a strictly decreasing function. Hence the maximum $\alpha \in [0, 1]$ for which $R(1_H(1_A - \alpha)) = 1$ is the solution of the equation

$$
\sum_{S \subseteq \Omega} m(S) \inf_{\omega \in S} (1_H(\omega)(1_A(\omega) - \alpha)) = 0,
$$

(6.4.6)

which is equivalent to

$$
Buy_R(1_H \cap A)(1 - \alpha) - \alpha Sell_R(1_H \cap A^c) = 0.
$$

(6.4.7)

Hence

$$
\alpha = \frac{Buy_R(1_H \cap A)}{Buy_R(1_H \cap A) + Sell_R(1_H \cap A^c)} = Buy_{R_H}(1_A).
$$

(6.4.8)

And thus

$$
Buy_{R^*}(1_A) = Buy_{R_H}(1_A).
$$

(6.4.9)

Hence

$$
\begin{align*}
Buy_{R^*}(X) &= Buy_{R^*}(y + (x - y)1_A) \\
&= y + (x - y)Buy_{R^*}(1_A) \\
&= y + (x - y)Buy_{R_H}(1_A) \\
&= Buy_{R_H}(X).
\end{align*}
$$

(6.4.10)

So $R^*(X) = R_H(X)$. \qed

Since coherent and B-consistent betting functions are determined by their buy values on indicator functions, we find that $R_H$ is the only coherent and B-consistent betting function such that $R_H(1_A) = R^*(1_A)$. However, the next example shows that $R^*$ is in general not B-consistent, implying that $R^*$ and $R_H$ are in general not the same.
Example 6.4.2. Set $\Omega = \{a, b, c, d\}$ and suppose that
\[
m(\{a, b, c\}) = m(\{a\}) = m(\{b, d\}) = \frac{1}{3}.
\] (6.4.11)

We set $H = \{a, b, c\}$ and consider the bets $X = 2 \cdot 1_{\{a\}}$ and $Y = 1_{\{a, b\}}$. We know from Theorem 6.4.1 that
\[
Buy_{R^*}(X) = Buy_{R_H}(X) = 2Bel_H(\{a\}) = \frac{2m(\{a\})}{m(\{a\}) + m(\{b, d\}) + m(\{a, b, c\})} = \frac{2}{3}.
\] (6.4.12)

and
\[
Buy_{R^*}(Y) = Buy_{R_H}(Y) = Bel_H(\{a, b\}) = \frac{m(\{a\})}{m(\{a\}) + m(\{a, b, c\})} = \frac{1}{2}.
\] (6.4.13)

Write $Z = X + Y$. To find $Buy_{R^*}(Z)$, we rewrite the equation
\[
Buy_R(1_H(Z - \alpha)) = 0
\] (6.4.14)
as
\[
-\frac{1}{3} \alpha + \frac{1}{3}(3 - \alpha) + 1_{\{\alpha \geq 1\}} \frac{1}{3}(1 - \alpha) = 0.
\] (6.4.15)

Now $\alpha$ cannot be smaller or equal to 1, since this would imply that $-\frac{1}{3} \alpha + \frac{1}{3}(3 - \alpha) = 0$ and thus $\alpha = \frac{3}{2} > 1$, giving a contradiction. So $\alpha > 1$ and thus
\[
-\frac{1}{3} \alpha + \frac{1}{3}(3 - \alpha) + \frac{1}{3}(1 - \alpha) = 0.
\] (6.4.16)

This gives $\alpha = \frac{4}{3}$. So
\[
Buy_{R^*}(Z) = \frac{4}{3} > \frac{2}{3} + \frac{1}{2} = Buy_{R^*}(X) + Buy_{R^*}(Y).
\] (6.4.17)

It is easy to check that
\[
\min_{\omega \in S} X(\omega) + \min_{\omega \in S} Y(\omega) = \min_{\omega \in S} Z(\omega)
\] (6.4.18)
for every nonempty $S \subseteq \Omega$. This means $R^*$ is not B-consistent.

Although $R_H$, which is B-consistent, does not coincide with $R^*$, which is not necessarily B-consistent, the next theorem tells us that there are no bets that are accepted under $R_H$ that are not accepted under $R^*$. □
CHAPTER 6. A BETTING INTERPRETATION

**Theorem 6.4.3.** We have

\[ R_H(X) \leq R^*(X) \]  

(6.4.19)

for all \( X \in X \).

**Proof.** Let \( X \in X \) and write \( X(\Omega) = \{y_1, \ldots, y_K\} \), where \( y_i < y_{i+1} \) and set \( y_0 := 0 \). We have

\[
\text{Buy}_{R_H}(X) = \sum_{j=1}^{K} (y_j - y_{j-1}) \text{Buy}_{R_H}(1_{\{X \geq y_j\}})
\]

(6.4.20)

\[
= \sum_{j=1}^{K} (y_j - y_{j-1}) \text{Buy}_{R^*}(1_{\{X \geq y_j\}})
\]

\[
\leq \text{Buy}_{R^*}(X),
\]

where we used that \( R^* \) is coherent in the last step.

Theorem 6.4.3 guarantees that \( R_H \) is more conservative than \( R^* \). We now also show that the conservatism of \( R_H \) is minimal in the following sense.

**Theorem 6.4.4.** If \( T \) is a coherent and \( B \)-consistent betting function such that

\[ T(X) \leq R^*(X) \]  

(6.4.21)

for all \( X \in X \), then

\[ T(X) \leq R_H(X) \]  

(6.4.22)

for all \( X \in X \).

**Proof.** Let \( X \in X \) and write \( X(\Omega) = \{y_1, \ldots, y_K\} \), where \( y_i < y_{i+1} \) and set \( y_0 := 0 \). We have

\[
\text{Buy}_T(X) = \sum_{j=1}^{K} (y_j - y_{j-1}) \text{Buy}_T(1_{\{X \geq y_j\}})
\]

\[
\leq \sum_{j=1}^{K} (y_j - y_{j-1}) \text{Buy}_{R^*}(1_{\{X \geq y_j\}})
\]

(6.4.23)

\[
= \sum_{j=1}^{K} (y_j - y_{j-1}) \text{Buy}_{R_H}(1_{\{X \geq y_j\}})
\]

\[
= \text{Buy}_{R_H}(X).
\]

Theorem 6.4.3 and 6.4.4 together show that if we take \( R^* \) and want to create a \( B \)-consistent betting function that is more conservative than \( R^* \), that \( R_H \) is the best we can do in the sense that it is the least conservative.
6.5 Necessary conditioning

We suppose here that \( \text{Sell}_{R}(1 \cap \text{H}) = 1 \) and aim to show that the conditional betting function \( R_{\text{H}} \) should satisfy

\[
\text{Buy}_{R_{\text{H}, \text{nec}}}(X) := \sum_{S \subseteq \Omega} m_{\text{H}, \text{nec}}(S) \min_{\omega \in S} X(\omega),
\]

(6.5.1)

where \( m_{\text{H}, \text{nec}} \) is the basic belief assignment given in (5.2.4).

In Section 5.2, learning that \( \text{H} \) necessarily occurs, was translated as: \( \text{H} \) always occurs. In the betting interpretation, we are no longer in the situation in which we have a series of repetitions. In the betting context, we translate learning that \( \text{H} \) necessarily occurs as learning that the bet on \( \text{H}^{c} \) is worth nothing. This should not be conflated with the bet on \( \text{H}^{c} \) not paying out. Learning that the bet on \( \text{H}^{c} \) is worth nothing, clearly implies that the bet on \( \text{H}^{c} \) will not pay out, but is stronger than that.

The reason for interpreting ‘necessary conditioning’ in this way, is that the bet on \( \text{H}^{c} \) being worth nothing, resembles the philosophical concept of the proposition ‘\( \text{H} \) occurs’ being necessarily true [35]. This means that one learns that \( \text{H} \) does not only occur, but that is is impossible for \( \text{H}^{c} \) to occur. In other words, the outcome space \( \Omega \) could have been chosen to be \( \text{H} \) from the beginning. An agent can revise his betting function \( R \) according to this information, only if \( \text{Buy}_{R}(1 \cap \text{H}^{c}) = 0 \). If \( \text{Buy}_{R}(1 \cap \text{H}^{c}) > 0 \), this means the agent believes that \( \text{H}^{c} \) occurring is possible, which conflicts with learning that \( \text{H}^{c} \) occurring is not possible at all. Notice, however, that \( \text{Sell}_{R}(1 \cap \text{H}^{c}) > 0 \) does not imply that the agent believes that \( \text{H}^{c} \) occurring is possible, but only that the agent does not believe that \( \text{H}^{c} \) occurring is impossible.

Example 6.5.1. Suppose that an agent, quite ignorant about astronomy, pays some positive amount for the bet that the morning star (the bright object that comes up in the morning) and evening star (the bright object that comes up in the evening) do have very different temperatures. Only then the agent learns that the morning and evening star are the same object, namely the planet Venus (which is, of course, not a star but a brightly reflective planet).\(^1\) Of course this means that the morning star and the evening star have exactly the same temperature, namely the temperature of Venus. But it is stronger than that: it means that the morning star and the evening star necessarily have the same temperature. In other words, it is impossible for the morning and evening star to have different temperatures, because they are the same object. The agent is really struck by finding this out and realizes that his or her prior idea about what were actually possible outcomes, was incorrect: (s)he had concluded that it was possible for the two stars to have different temperatures because (s)he had wrongly assumed that the morning and evening star were different objects.

Example 6.5.1 intentionally echoes the situation of the combining beliefs (see Section 5.5), where an agent learns that two distinct labels are actually referring

\(^{1}\)The ancient Greeks did in fact believe that the morning star and evening star were different objects. Only later it was discovered that they were both Venus.
to the same phenomenon. As a consequence, any property of the two phenomena, just like any property of the morning and evening star, are necessarily the same. This shows that the critique of the Section 5.5 remains valid in the general betting setting: learning that two phenomena are the same, corresponds to learning that their outcomes are necessarily the same, which we now interpret as learning that the bet on their outcomes being different is worth nothing. Hence necessary conditioning should be applied in this situation, and not mixed (Dempster-Shafer) conditioning.

To see how an agent should revise his or her betting function, let $X \leq 0$ and consider an agent buying the bet $1_{H X}$. Clearly the buy price must be negative, since $1_{H X} \leq 0$. However, the outcome 0 is the best outcome of the bet $1_{H X}$, so $H^c$ occurring is a relative good outcome for an agent buying $1_{H X}$. Hence, if the agent learns that $H$ necessarily occurs, (s)he certainly would not pay more for $X$. In other words, if we write $\alpha$ for the maximum price the agent is willing to pay for $X$ after (s)he learns that $H$ necessarily occurs, we should find

$$\alpha \leq \text{Buy}_R(1_{H X}).$$ (6.5.2)

However, we assumed that $\text{Sell}_R(1_H) = 0$, or equivalently, $\text{Buy}_R(1_{H^c}) = 0$. This means that the agent has no belief in $H^c$ to begin with, and thus nothing is really lost if (s)he learns that $H$ necessarily occurs. Thus (6.5.2) should be an equality:

$$\alpha = \text{Buy}_R(1_{H X}).$$ (6.5.3)

Indeed, we find that this is true for $\alpha = \text{Buy}_{R_{H, \text{nec}}}(X)$:

$$\text{Buy}_R(1_{H X}) = \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} 1_H(\omega) X(\omega)$$
$$= \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S \cap H} X(\omega)$$
$$= \sum_{A \subseteq H} \left( \sum_{S : S \cap H = A} m(S) \right) \min_{\omega \in A} X(\omega)$$
$$= \sum_{A \subseteq H} m_{H, \text{nec}}(A) \min_{\omega \in A} X(\omega)$$
$$= \text{Buy}_{R_{H, \text{nec}}}(X).$$ (6.5.4)

It is now only a little exercise to show that $R_{H, \text{nec}}$ is the only betting function with this property.

**Theorem 6.5.2.** If $T$ is a betting function such that

$$\text{Buy}_T(X) = \text{Buy}_R(1_{H X})$$ (6.5.5)

for all $X \leq 0$, then $T = R_{H, \text{nec}}$. 

94
6.6. STRONG INDEPENDENCE

Proof. Let $X \in \mathcal{X}$ be arbitrary. Then

$$
\text{Buy}_T(X) = \text{Buy}_T(X - \max(X)) + \max(X)
$$

$$
= \text{Buy}_R(1_H(X - \max(X))) + \max(X)
$$

$$
= \text{Buy}_{R_{H,\text{sec}}}(X - \max(X)) + \max(X)
$$

$$
= \text{Buy}_{R_{H,\text{sec}}}(X),
$$

(6.5.6)

where we used (6.5.4) in the third step.

\[ \square \]

6.6 Strong independence

In this subsection, we aim at interpreting strong independence from the betting perspective.

We first introduce some notation. Set $\Omega := \Omega_1 \times \Omega_2$. Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be the bets on $\Omega_1$ and $\Omega_2$ respectively, i.e. $\mathcal{X}_1 = \mathbb{R}^{\Omega_1}$ and $\mathcal{X}_2 = \mathbb{R}^{\Omega_2}$. Let $R$ be a coherent betting function on $\Omega$. For $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$ we define

$$
(X_1 \cdot X_2)(\omega_1, \omega_2) := X_1(\omega_1) \cdot X_2(\omega_2).
$$

(6.6.1)

We define betting functions $R_1$ and $R_2$ on respectively $\Omega_1$ and $\Omega_2$ by defining

$$
R_1(X_1) := R(X_1 \cdot 1)
$$

(6.6.2)

and

$$
R_2(X_2) := R(1 \cdot X_2).
$$

(6.6.3)

We assume that $R$ is a $B$-consistent betting function and let $m$ be the basic belief assignment corresponding to $R$, i.e.

$$
\text{Buy}_R(X) = \sum_{S \subseteq \Omega} m(S) \min_{\omega \in S} X(\omega)
$$

(6.6.4)

for all $X \in \mathcal{X}$. Let $m_1$ and $m_2$ be the marginals of $m$ (cf. Equation 5.4.3 and 5.4.4), then we see that

$$
\text{Buy}_{R_1}(X_1) = \text{Buy}_R(X_1 \cdot 1)
$$

$$
= \sum_{S \subseteq \Omega_1} m(S) \min_{(\omega_1, \omega_2) \in S} X_1(\omega_1)
$$

$$
= \sum_{S \subseteq \Omega_1} m(S) \min_{\omega_1 \in \pi_1^{-1}(S)} X_1(\omega_1)
$$

$$
= \sum_{A \subseteq \Omega_1} \left( \sum_{S \subseteq \Omega : \pi_1^{-1}(S) = A} m(S) \right) \min_{\omega_1 \in A} X_1(\omega_1)
$$

$$
= \sum_{A \subseteq \Omega_1} m_1(A) \min_{\omega_1 \in A} X_1(\omega_1).
$$

(6.6.5)
Hence $R_1$ is a coherent and $B$-consistent betting function that corresponds to $m_1$. In the same way $R_2$ is a coherent and $B$-consistent betting function that corresponds to $m_2$. We now start by giving a characterization of strong independence.

**Theorem 6.6.1.** We have

$$\text{Buy}_R(X_1 \cdot X_2) = \text{Buy}_{R_1}(X_1)\text{Buy}_{R_2}(X_2)$$

(6.6.6)

and

$$\text{Sell}_R(X_1 \cdot X_2) = \text{Sell}_{R_1}(X_1)\text{Sell}_{R_2}(X_2)$$

(6.6.7)

for all $X_1, X_2 \geq 0$ if and only if

$$m(A \times B) = m_1(A)m_2(B)$$

(6.6.8)

for every $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$.

**Proof.** First assume that $m(A \times B) = m_1(A)m_2(B)$. This implies that for $X_1, X_2 \geq 0$

$$\text{Buy}_R(X_1 \cdot X_2) = \sum_{S \subseteq \Omega} m(S) \min_{(\omega_1, \omega_2) \in S} X_1(\omega_1)X_2(\omega_2)$$

$$= \sum_{A \subseteq \Omega_1} \sum_{B \subseteq \Omega_2} m_1(A)m_2(B) \min_{(\omega_1, \omega_2) \in A \times B} X_1(\omega_1)X_2(\omega_2)$$

(6.6.9)

$$= \sum_{A \subseteq \Omega_1} m_1(A) \min_{\omega_1 \in A} X_1(\omega_1) \sum_{B \subseteq \Omega_2} m_2(B) \min_{\omega_2 \in B} X_2(\omega_2)$$

$$= \text{Buy}_{R_1}(X_1)\text{Buy}_{R_2}(X_2).$$

The other way, we have

$$\text{Buy}_R(1_A \cdot 1_B) = \text{Buy}_{R}(1_A)\text{Buy}_{R}(1_B)$$

(6.6.10)

and

$$\text{Sell}_R(1_A \cdot 1_B) = \text{Sell}_{R}(1_A)\text{Sell}_{R}(1_B).$$

(6.6.11)

It follows from Theorem 5.4.4 that this is equivalent with having

$$m(A \times B) = m_1(A)m_2(B)$$

(6.6.12)

for every $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$. □

Since $\text{Buy}_R(X_1 \cdot X_2) = \text{Exp}(X_1 \cdot X_2)$ by Theorem 6.2.11, Theorem 6.6.1 generalizes the classical result that the expectation of the product of independent variables equals the product of the expectations of the variables. To interpret the characterization of Theorem 6.6.1, let $X_1, X_2 \geq 0$ and define $Y : \Omega_1 \to \mathbb{R}$ by

$$Y(\omega_1) := \text{Buy}_R(X_1(\omega_1) \cdot X_2) = X_1(\omega_1)\text{Buy}_{R}(1 \cdot X_2).$$

(6.6.13)
Since $X_2 \geq 0$ and thus $\text{Buy}_{R_2}(X_2) \geq 0$, we find

$$\text{Buy}_{R_1}(Y) = \text{Buy}_{R_1}(X_1) \text{Buy}_{R_2}(X_2).$$  \hspace{1cm} (6.6.14)$$

This means that we can rewrite (6.6.6) as

$$\text{Buy}_R(X_1 \cdot X_2) = \text{Buy}_{R_1}(Y).$$  \hspace{1cm} (6.6.15)$$

The result of the bet $Y$ depends on $\omega_1$ only in the scaling factor $X_1(\omega_1)$ for the bet $1 \cdot X_2$. Above identity says that we have strong independence only if the price the agent is willing to pay for $Y$ equals the price (s)he is willing to pay for $X_1 \cdot X_2$. The equality for sell can of course be interpreted completely analogous.
CHAPTER 7

Assessing forensic evidence

7.1 The island problems

The context of the well known island problems is classical. A crime has been committed on an island with \( N + 1 \) inhabitants, so that we can be sure that one of them is the culprit. Now some characteristic of the criminal, e.g. a DNA profile, is found at the scene of the crime and we may assume that this profile originates from the culprit. Then we somehow select an individual \( s \) from the island, who happens to have the same characteristic as the criminal. The question is what we can say about the probability or belief in the event that \( s \) is in fact the criminal. This is not a well defined question yet, as it depends on the way \( s \) was found. We distinguish between two cases: the cold case in which we randomly select an inhabitant, and the search variant in which we consider the inhabitants one by one in a random order, until we find an inhabitant with the characteristic found at the crime scene.

With the island problems, our belief functions allow us to assign a zero prior belief in the guilt of any of the individuals of the island, while at the same time assigning belief one to the full populations. This seems better suited to the legal context than the classical Bayesian setting since assigning a non-zero prior probability to an individual, without any evidence against the individual itself other than belonging to the population, seems unreasonable. Of course, we have to make modeling assumptions, as in the classical case and we will discuss these below. But as we shall see, the outcomes are different from the classical outcome: if we assume total prior ignorance the belief that \( s \) is the culprit is in our setting different from the classical probability that he is guilty under a uniform prior. We turn to the examples now.

7.1.1 The cold case

Let \( X = \{1, \ldots, N + 1\} \) be the population of the island. At the scene of the crime a DNA profile \( \Gamma \) is found which we know has frequency \( p \in (0, 1] \) in the
population. This means that a randomly chosen person has probability $p$ to have the characteristic, independent of the other individuals. We remark that this assumption is in the realm of classical probability theory. This is reasonable since the frequency interpretation of classical probability works well within the context of DNA profiles.

Our basic belief assignment should capture our prior knowledge, that is, prior to the fact that we find an individual in the population of the island with the characteristic. Our model has three assumptions: (1) some individual $S$ from the population of the island is selected at random, (2) the population frequency of the DNA profile is $p$, and (3) the criminal is in the part of population of the island that has the characteristic. Most importantly, we do not need any arbitrary prior assumptions about who the criminal is. This leads to the following basic belief assignment, prior to finding $S$.

We set $\Omega = X \times \times X \times \{0,1\}^{N+1}$ and let $C : \Omega \to X, S : \Omega \to X$ and $\Gamma_i : \Omega \to \{0,1\}$ be projections on respectively the first, second and $i + 2$-th coordinate. $C$ represents the criminal, $S$ the selected individual from the island population, and $\Gamma_i = 1$ indicates that the $i$-th individual on the island has characteristic $\Gamma$. We now define the following basic belief assignment on $\Omega$. We set

$$m(C \in \{i \in X : y_i = 1\}, S = x, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1}) = \frac{1}{N+1} \frac{p^k(1-p)^{N+1-k}}{1 - (1-p)^{N+1}}$$

(7.1.1)

if $\sum_{i=1}^{N+1} y_i = k > 0$. Note that the normalizing factor $1 - (1-p)^{N+1}$ in (7.1.1) comes from our third assumption which implies that at least one person on the island has the characteristic.

Now we want to compute the posterior beliefs by incorporating that we have chosen $s \in X$ and that $s$ has the characteristic $\Gamma$. In other words, we want to condition in the contingent sense on $H_s = \{S = s, \Gamma_S = 1\}$. Now observe that $\text{Bel}(H_s)$ and $\text{Pl}(H_s)$ are both given by the sum of all basic belief of events with $S = s$ and $y_s = 1$, i.e.

$$\text{Bel}(H_s) = \text{Pl}(H_s)$$

$$= \sum_{y_s=1} m(C \in \{i : y_i = 1\}, S = s, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1})$$

$$= \frac{1}{N+1} \frac{1}{1 - (1-p)^{N+1}} \sum_{j=0}^{N} \binom{N}{j} p^{j+1}(1-p)^{N-j}$$

(7.1.2)

$$= \frac{p}{(N+1)(1 - (1-p)^{N+1})}.$$ 

Let $B \subseteq X$ with $s \in S$. Since $\text{Bel}(H_s) = \text{Pl}(H_s)$ contingent conditioning boils
down to
\[ \text{Bel}_{H_s}(C \in B) = \frac{\text{Bel}(\{C \in B\} \cap H_s)}{\text{Bel}(H_s)} \]
\[ = \frac{(N + 1)(1 - (1 - p)^{N+1})}{p} \text{Bel}(C \in B \mid S = s \mid \Gamma_s = 1) \]
\[ = 1 \frac{1}{p} \sum_{k=0}^{\lfloor |B| - 1 \rfloor} \binom{|B| - 1}{k} p^{k+1}(1 - p)^{N+1-k} \]
\[ = (1 - p)^{N+1-|B|}. \quad (7.1.3) \]

An interesting special case occurs when \( B = \{ s \} \). The conditional belief that \( C = s \) is apparently given by
\[ \text{Bel}_{H_s}(C = s) = (1 - p)^N, \quad (7.1.4) \]
simply take \( |B| = 1 \). This formula has a simple interpretation: the belief that \( s \) is the criminal is just the probability that all other members of the population are excluded since they have the wrong profile.

It is interesting to compare this answer to the classical one, in which a uniform prior is taken. In the classical case, the posterior probability that \( C = s \) is equal to
\[ \frac{1}{1 + Np}, \quad (7.1.5) \]
see e.g. [53] or [4]. We observe that
\[ \frac{1}{1 + Np} > (1 - p)^N. \quad (7.1.6) \]

Hence in our setting, the belief that \( C = s \) is always smaller than in the classical case, something we can intuitively understand by recalling that we did not use any assumptions about the criminal other than that the criminal has the characteristic. To give some indication of the difference between the two answers, if \( p \sim N^{-1} \) (for \( N \to \infty \)), then (7.1.5) \( \sim \frac{1}{2} \), while (7.1.4) \( \sim e^{-1} \).

Since belief functions generalize probability distributions, we should be able to re-derive the classical result (7.1.5) using our approach, and we now show that this is indeed the case. If we want to take a uniform prior for the criminal, then the basic belief assignment, denoted by \( m^c \), is as follows. We set
\[ m^c(C = t, S = x, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1}) = \frac{1}{(N + 1)^2} p^k(1 - p)^{N-k}, \quad (7.1.7) \]
if \( y_t = 1 \) and \( \sum_{i=1}^{N+1} y_i = k + 1 \). Note that we do not have the normalizing factor \( 1 - (1 - p)^{N+1} \) here, since the probability mass is now on the event that implies \( y_t = 1 \). Note that the corresponding belief function is a probability distribution,
since only singletons have positive basic belief. Next we condition on the same event $H_s = \{ S = s, \Gamma_S = 1 \}$ as before. We compute

$$\text{Bel}^c(H_s \cap \{ C = s \}) = \frac{1}{(N + 1)^2} \quad (7.1.8)$$

and

$$\text{Pl}^c(H_s \cap \{ C \neq s \}) = \frac{Np}{(N + 1)^2}. \quad (7.1.9)$$

Hence

$$\text{Bel}^c_{H_s}(C = s) = \frac{\text{Bel}^c(H_s \cap \{ C = s \})}{\text{Bel}^c(H_s \cap \{ C = s \}) + \text{Pl}^c(H_s \cap \{ C \neq s \})} = \frac{1}{1 + Np}, \quad (7.1.10)$$

as required.

This example illustrates that we lose nothing by working with belief functions, and that belief functions only add flexibility. If a certain classical prior is reasonable then we can take that prior and work with it. If there are reasons to have a non-classical prior, for instance complete ignorance within a given population, then belief functions are flexible enough to deal with this.

### 7.1.2 The search case

In the search variant, we do not choose a random individual $S$ but we check the inhabitants one by one in a random order, until an individual with the relevant characteristic is found. However, we only take into account the result of the search and not any information about the search itself. As a consequence, the search case boils down to changing the first assumption of the cold case into: some individual $S$ is selected at random from the subset of the population that has the characteristic. We consider the same space $\Omega$ as in the cold case. We set

$$m(C \in \{ i \in X : y_i = 1 \}, S = x, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1}) = \frac{1}{k} \frac{p^k(1-p)^{N+1-k}}{1 - (1-p)^{N+1}} \quad (7.1.11)$$

if $y_x = 1$ and $\sum_i y_i = k > 0$. Note that there are two differences compared to the cold case: we have to assume that $y_x = 1$, and we have to divide by $k$ rather than by $N + 1$.

Next we incorporate that we found a suspect with the characteristic by conditioning on $H_s = \{ S = s \}$. Note that the conditioning does not contain information about the length of the search or the identity of searched individuals. We only know that $s$ was the first one to be found with the characteristic. We
7.1. THE ISLAND PROBLEMS

again have $\text{Bel}(H_s) = \text{Pl}(H_s)$ with

$$\text{Bel}(H_s) = \text{Pl}(H_s)$$

$$= \sum_{y_i = 1} m(C \in \{i : y_i = 1\}, S = s, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1})$$

$$= \frac{1}{1 - (1 - p)^{N+1}} \sum_{j=0}^N \binom{N}{j} \frac{1}{j + 1} p^{j+1}(1 - p)^{N-j} \quad \text{(7.1.12)}$$

$$= \frac{1}{N + 1} \sum_{j=0}^N \binom{N + 1}{j + 1} p^{j+1}(1 - p)^{N-j}$$

$$= \frac{1}{N + 1}.$$  

Notice that we can also derive (7.1.12) by observing that $M := \text{Bel}(H_s) = \text{Pl}(H_s)$ does not depend on $s$ and thus we have $(N+1)M = 1$.

To compute the conditional belief in $\{C = s\}$, we first compute

$$\text{Bel}(H_s \cap \{C = s\}) = \text{Bel}(C = S = s) = \frac{p(1 - p)^N}{1 - (1 - p)^{N+1}} \quad \text{(7.1.13)}$$

and then conclude that

$$\text{Bel}_{H_s}(C = s) = (N + 1)\text{Bel}(H_s \cap \{C = s\})$$

$$= \frac{p(1 - p)^N(N + 1)}{1 - (1 - p)^{N+1}}, \quad \text{(7.1.14)}$$

which is different compared to the cold case.

There is a very natural interpretation of the expression in (7.1.14). In the numerator, we have the probability that a random variable with a binomial distribution with parameters $N + 1$ and $p$ is equal to 1. The denominator is the probability that this random variable is positive, so (7.1.14) is the conditional probability for such a random variable to be 1 given it is positive. This makes sense, since we can only know for sure that $C = s$ when $s$ is the only one with the characteristic. Notice that (7.1.14) equals zero if $p = 1$ and if $p < 1$ we can rewrite (7.1.14) as

$$\frac{N + 1}{\sum_{k=0}^N (1 - p)^{-k}}. \quad \text{(7.1.15)}$$

In the classical case, starting with a uniform probability distribution, the posterior probability that $C = s$ is equal to

$$\frac{1 - (1 - p)^{N+1}}{(N + 1)p} = \frac{1}{N + 1} \sum_{j=0}^N (1 - p)^j, \quad \text{(7.1.16)}$$

see e.g. [4]. Note that (7.1.16) is the arithmetic mean of $1, 1 - p, (1 - p)^2, \ldots, (1 - p)^N$, while (7.1.15) is the harmonic mean of the same sequence. Hence the answer
using our approach is - as it was in the cold case - smaller than the classical answer.

We finally demonstrate that we can also derive this classical result with our technology. In the classical case, we set the basic belief assignment to be

\[
m^c(C = t, S = x, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1}) = \frac{1}{N+1} \frac{1}{k+1} p^k (1 - p)^{N-k}
\]

if \( y_t = y_x = 1 \) and \( \sum_{i=1}^{N+1} y_i = k + 1 \).

Note that the corresponding belief function is a probability measure, since only singletons have positive basic belief assignments. We condition on \( H_s = \{S = s\} \). In the same way as before, we find

\[
\text{Bel}(H_s) = \text{Pl}(H_s) = \frac{1}{N+1}.
\]

We then compute the belief

\[
\text{Bel}(H_s \cap \{C = s\}) = \frac{1}{N+1} \sum_{k=0}^{N} \binom{N}{k} \frac{1}{k+1} p^k (1 - p)^{N-k}
\]

\[
= \frac{1}{(N+1)^2} \frac{1}{p} \sum_{k=0}^{N} \binom{N+1}{k+1} p^{k+1} (1 - p)^{N-k}
\]

\[
= \frac{1}{(N+1)^2} \frac{1}{p} (1 - (1 - p)^{N+1}),
\]

(7.1.19)

to conclude that

\[
\text{Bel}_{H_s}(C = s) = \frac{\text{Bel}(H_s \cap \{C = s\})}{\text{Bel}(H_s)} = \frac{1 - (1 - p)^{N+1}}{(N+1)p},
\]

(7.1.20)
as required.

### 7.2 Parental identification

Our next example concerns the situation in which we have a known mother and a known child, but we do not know who the father is. We assume that there is a set \( X = \{1, \ldots, N+1\} \) of potential fathers. We would like to make belief statements about the possible fatherhood of someone chosen from the population, based on the DNA profile of this chosen person. In order to keep things as simple as possible, we assume that we only consider one specific locus of the DNA. Furthermore, we assume that the alleles of mother and child at that locus are such that we know what the paternal allele must be. Every potential father \( i \) in \( X \) has two alleles at the locus. We write \( \Gamma_i = 0 \) if they both do not match the paternal allele, \( \Gamma_i = 2 \) if they both match the paternal allele. If precisely one of the two matches the paternal allele, we write \( \Gamma_i = 1^+ \) if this allele is passed on
by the potential father and $\Gamma_i = 1^-$ if this allele is not passed on by the potential father.

In order to set up our prior belief function, we set $\Omega = X \times X \times \{0, 1^-, 1^+, 2\}^{N+1}$ and let $F : \Omega \to X, S : \Omega \to X$ and $\Gamma_i : \Omega \to \{0, 1^-, 1^+, 2\}$ be projections on respectively the first, second and $2 + i$-th coordinate. $F$ represents the father, and $S$ the selected individual which is the putative father. Let $p_0, p_1$ and $p_2$ be the probabilities that an individual has respectively 0,1 or 2 alleles of the right type. We set the prior basic belief assignment to be

$\begin{align*}
m(F \in \{i : y_i \in \{1^+, 2\}\}, S = x, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1}) &= \frac{1}{N + 1} \frac{p_0^{k_0} (\frac{1}{2} p_1)^{k_1^- + k_1^+} p_2^{k_2}}{1 - (p_0 + \frac{1}{2} p_1)^{N+1}}
\end{align*}$

if $k_0 = \sum_{i=1}^{N+1} 1(y_i = 0)$, $k_1^- = \sum_{i=1}^{N+1} 1(y_i = 1^-)$, $k_1^+ = \sum_{i=1}^{N+1} 1(y_i = 1^+)$, $k_2 = \sum_{i=1}^{N+1} 1(y_i = 2)$ and $k_1^+ + k_2 > 0$. As before, this basic belief assignment is a summary of what we know, and is formulated in terms of items that can be well described by classical probabilities. Note that the normalizing factor $1 - (p_0 + \frac{1}{2} p_1)^{N+1}$ comes from the fact that we know that at least one individual $i$ (namely the actual father) has either $\Gamma_i = 2$ or $\Gamma_i = 1^+$. The factor $\frac{1}{2}$ comes from the fact that the father passes a randomly chosen allele to his child, hence the probability that a potential father has precisely one allele of the right type and passes it on, is $\frac{1}{2} p_1$. The probability that he has precisely one allele of the right type and not passes it on, is of course also $\frac{1}{2} p_1$.

Now we have to distinguish between two scenarios: the suspect in question has one or two alleles of the right type. We look at the case he has two such alleles first. This means we want to condition on

$H_{s,2} = \{S = s, \Gamma_S = 2\}.$

We write $E(x, y_1, \ldots, y_{N+1})$ for the event in (7.2.1). We compute

$Bel(H_{s,2}) = Pl(H_{s,2})$

$= \sum_{x=s, y_s=2} m(E(x, y_1, \ldots, y_{N+1}))$

$= \frac{p_2}{N + 1} \frac{1}{1 - (p_0 + \frac{1}{2} p_1)^{N+1}}$  \hspace{1cm} (7.2.3)

and

$Bel(\{F = s\} \cap H_{s,2}) = \sum_{x=s, y_s=2, \forall i \neq s \ y_i \in \{0,1^-\}} m(E(x, y_1, \ldots, y_{N+1}))$

$= \frac{p_2}{N + 1} \frac{(p_0 + \frac{1}{2} p_1)^N}{1 - (p_0 + \frac{1}{2} p_1)^{N+1}}.$

From this, it follows that

$Bel_{H_{s,2}}(F = s) = (p_0 + \frac{1}{2} p_1)^N.$  \hspace{1cm} (7.2.5)
Next we consider the case in which the putative father has one allele of the right type. We of course do not know if the putative father passes on this allele, hence we condition on

\[ H_{s,1} = \{ S = s, \Gamma S \in \{ 1^-, 1^+ \} \}. \]  

(7.2.6)

We compute

\[
\text{Bel}(H_{s,1}) = \text{Pl}(H_{s,1}) = \sum_{x=s, y_i=1^-} m(E(x, y_1, \ldots, y_{N+1})) + \sum_{x=s, y_i=1^+} m(E(x, y_1, \ldots, y_{N+1})) = \frac{1}{2}p_1(1 - (p_0 + \frac{1}{2}p_1)^N) \\
\frac{(N + 1)(1 - (p_0 + \frac{1}{2}p_1)^{N+1})}{(N + 1)(1 - (p_0 + \frac{1}{2}p_1)^N)}
\]  

(7.2.7)

and

\[
\text{Bel}(H_{s,1} \cap \{ F = s \}) = \sum_{x=s, y_i=1^+, \forall i \neq s, y_i \in \{0, 1^-\}} m(E(x, y_1, \ldots, y_{N+1})) = \frac{1}{N + 1} \frac{p_1(1 - (p_0 + \frac{1}{2}p_1)^{N+1})}{1 - (p_0 + \frac{1}{2}p_1)^N}.
\]  

(7.2.8)

From this, it follows that

\[
\text{Bel}_{H_{s,1}}(F = s) = \frac{1}{2}(p_0 + \frac{1}{2}p_1)^N \frac{1}{1 - (p_0 + \frac{1}{2}p_1)^N},
\]  

(7.2.9)

which is clearly smaller than the answer in (7.2.5).

The classic answers.

\[
m^c(F = t, S = x, \Gamma_1 = y_1, \ldots, \Gamma_{N+1} = y_{N+1}) = \frac{1}{(N + 1)^2} \frac{p_0^{k_0} (\frac{1}{2}p_1)^{k_1^+ + k_1^-} p_2^{k_2}}{\frac{1}{2}p_1 + p_2} \]  

(7.2.10)

if \( k_0 = \sum_{i=1}^{N+1} 1(y_i = 0) \), \( k_1^+ = \sum_{i=1}^{N+1} 1(y_i = 1^+) \), \( k_1^- = \sum_{i=1}^{N+1} 1(y_i = 1^-) \), \( k_2 = \sum_{i=1}^{N+1} 1(y_i = 2) \) and \( y_i \in \{1^+, 2\} \).

We write \( E(x, t, y_1, \ldots, y_{N+1}) \) for the event in (7.2.1). We compute

\[
\text{Bel}^c(H_{s,2} \cap \{ F = s \}) = \sum_{x=t=s, y_i=2} m(E(x, t, y_1, \ldots, y_{N+1})) = \frac{p_2}{(N + 1)^2(\frac{1}{2}p_1 + p_2)}.
\]  

(7.2.11)
7.3. DISCUSSION AND LEGAL PRACTICE

and

\[ \text{Pl}^c(H_{s,2} \cap \{ F \neq s \}) = \sum_{t \neq s, x=s, y_s=2} m(E(x, t, y_1, \ldots, y_{N+1})) \]
\[ = \frac{N p_2}{(N + 1)^2}. \quad (7.2.12) \]

Thus

\[ \text{Bel}^c_{H_{s,2}}(F = s) = \frac{1}{1 + N \left( \frac{1}{2} p_1 + p_2 \right)}. \quad (7.2.13) \]

We compute

\[ \text{Bel}^c(H_{s,1} \cap \{ F = s \}) = \sum_{x=t=s, y_s=1^+} m(E(x, t, y_1, \ldots, y_{N+1})) \]
\[ = \frac{1}{2} p_1 \left( N + 1 \right)^2 \left( \frac{1}{2} p_1 + p_2 \right). \quad (7.2.14) \]

and

\[ \text{Pl}^c(H_{s,1} \cap \{ F \neq s \}) = \sum_{t \neq s, x=s, y_s \in \{1, 2\}} m(E(x, t, y_1, \ldots, y_{N+1})) \]
\[ = \frac{N p_1}{(N + 1)^2}. \quad (7.2.15) \]

Thus

\[ \text{Bel}^c_{H_{s,1}}(F = s) = \frac{1}{1 + 2N \left( \frac{1}{2} p_1 + p_2 \right)}. \quad (7.2.16) \]

7.3 Discussion and legal practice

How does the theory of belief functions as developed in this thesis relate to the legal practice when expert witnesses are called to testify? It is to this question that we now turn.

We start by describing the current practice when classical probability theory is used. In this current practice, the legal representative (judge or jury member for instance) and expert witness play a very different role, and their contributions are well-separated. This is theoretically backed up by Bayes’ rule. To be precise, writing \( G \) for the event, say, that a certain suspect is the donor of a DNA profile found at the scene of the crime, and \( E \) for the available evidence, the so called odds form of Bayes’ rule states that

\[ \frac{P(G|E)}{P(G|\bar{E})} = \frac{P(E|G)}{P(E|\bar{G})} \cdot \frac{P(G)}{P(\bar{G})}, \quad (7.3.1) \]

where \( \bar{G} \) denotes the negation or complement of \( G \). Whatever the interpretation of the classical probability measure \( P \), subjective, frequentistic or otherwise, \( P(G|E) \)
represents the posterior probability of $G$ conditioned on the available evidence, while $P(G)$ denotes the prior probability of $G$, before taking the evidence $E$ into account.

Formula (7.3.1) describes how the prior odds $P(G)/P(\bar{G})$ are transformed into the posterior odds $P(G|E)/P(\bar{G}|E)$, by multiplying the prior odds with the so-called likelihood ratio $P(E|G)/P(E|\bar{G})$. This likelihood ratio is in the provenance of the forensic expert. This expert can, in certain circumstances at least, compute, estimate or assign a probability to the evidence under various hypothesis, and thereby compute the likelihood ratio. The expert does not express how likely these hypotheses themselves are, that is, the expert says nothing about the prior probabilities $P(G)$ and $P(\bar{G})$. Hence, the expert does not make a statement about the posterior probabilities $P(G|E)$ and $P(\bar{G}|E)$. Indeed, for the latter probabilities, one in addition needs the priors and these are not in the provenance of the expert.

One can also look at the posterior odds from the perspective of the joint distribution of $G$ and $E$. Once we have the full joint distribution, the posterior odds (and any other quantity relating $E$ and $G$, for that matter) can be computed. But it is only through the joint effort of expert and legal representative that the joint distribution can be computed. In other words, the complete joint distribution of $G$ and $E$ can only be determined as a joint effort of both expert and legal representative. They both contribute to the joint distribution, albeit in different ways, and as such both contribute to the posterior odds as well.

Obviously, the legal practice is more complicated than this simple procedure suggests. Although Bayes’ rule is mathematically not difficult, confusion between likelihood ratio and posterior odds arises easily, and there are many examples in which such confusion has had serious consequences see e.g. [8] and the references therein. But also, as already articulated in the introduction, there are many situations possible where a prior in the context of classical probability theory is not possible or at least not appropriate from a legal perspective. Furthermore, the clean theory about division of responsibilities between expert and legal representative does not always work this way: there are situations in which the prior actually enters the likelihood ratio, which complicates matters significantly, see e.g. [53], equations (4.8) and (5.27). Finally, it is highly questionable whether all available evidence is amenable to numerical manipulation.

The theory as developed in this thesis suggests a slightly different procedure which does not suffer from some of the problems mentioned above and in the introduction. In our setup, the legal representative still decides on the prior information, be it informative or not. Given this prior, the forensic expert can determine the joint belief structure of $E$ and $G$, and this joint belief structure contains the posterior belief in $G$. Determining the posterior belief, therefore, still is a joint effort of expert and legal representative, but the way in which they execute their roles, is slightly different from the classical case.

We can illustrate this with the cold case in the island problem discussed in Section 7.1.1. In the classical case, the expert witness can only deliver the
likelihood ratio, which in this case would be $1/p$. Indeed, in the case that the suspect the donor, the probability to find the evidence is 1, and in case he is not, the probability to find this particular profile is nothing but the frequency $p$ of that profile in the population. With uniform prior this leads to the posterior in (7.1.5), but note that other priors would lead to different result. Note that in any case, a prior must be made explicit in order to arrive at posterior probabilities.

In our setting, if the legal representative confirms that there is no prior information, the expert could simply report (7.1.4) which in that case contains all the available information in the case. If we have no prior information, then no unfounded or subjective choice for a prior needs to be made, apart from the choice of the relevant population. We do not need a uniform prior. If there is prior information, then the expert can set up the corresponding prior belief function, and compute the posterior according to the theory explained above.

We have seen in the examples in the island problem that the numbers obtained are certainly less impressive than in the classical case. This is of course not surprising: starting out with an uninformative prior instead of a uniform prior gives us less information to start with. Is this a weakness of the theory of belief functions as we have set out in this thesis? We do not think so. It is, in our opinion, better to have a less impressive number which is well founded and not so easy to challenge, than to have a more impressive number which may depend on unfounded arguments or assumptions, and which is easy to dismiss by, say, the defense. As such, our theory may even help to convict actual criminals, even though we insist that if there is no prior information about a certain quantity, we should not pretend there is.

In this thesis we have described the basic principles of our new theory, with basic forensic examples. This gives a solid foundation to work on, but obviously there is a lot of work to be done in the theoretical development and in applications to forensic examples.
CHAPTER 8

Infinite regress

8.1 Introduction and motivation

In this chapter we discuss probabilistic interpretations of the infinite epistemic regress problem. One locus classicus of the epistemic regress problem is Sextus Empiricus’ Outlines of Phyrronism, where five “Modes” are distinguished, i.e. five lines of reasoning, that sceptics have availed themselves of as a safeguard against dogmatism. The second of these modes has to do with an infinite regress. Says Sextus: “The Mode based upon regress ad infinitum is that whereby we assert that the thing adduced as a proof of the matter proposed needs a further proof, and this again another, and so on ad infinitum, so that the consequence is suspension [of assent], as we posses no starting-point for our argument.” Numerous formulations of this epistemic regress problem can be found in the literature, see [66] for a discussion, an overview, and for references.

In [40, 3, 2], Peijnenburg and Atkinsen interpret the epistemic regress problem probabilistically, as follows. Consider a proposition $E_0$ and suppose that we are interested in justified belief in $E_0$. Suppose further that there is a proposition $E_1$ that makes $E_0$ probable. Imagine that $E_2$ is in turn made probable by $E_3$, and that $E_3$ is made probable by $E_4$, and so on, ad infinitum. The question that Atkinson and Peijnenburg ask themselves is whether or not these conditions, with a suitable interpretation of the “make probable” relation, uniquely determine the belief in $E_0$.

They proceed as follows. Define $a$ as the probability of proposition $E_n$ given that proposition $E_{n+1}$ is true, and $b$ as the probability of $E_n$ given that $E_{n+1}$ is false. In formulas, this reads

$$P(E_n|E_{n+1}) = a \text{ and } P(E_n|E_{n+1}^c) = b,$$

(8.1.1)

for $n = 0, 1, \ldots$ (For simplicity, we only discuss the case in which neither of these two conditional probabilities depend on $n$.)

Although (8.1.1) does not determine the full joint probability distribution, Peijnenburg and Atkinson show that the conditions in (8.1.1) imply that $P(E_0)$
is uniquely determined and equals

\[ P(E_0) = \frac{b}{1 - a + b}. \]  

(8.1.2)

This is the “completion” of the infinite regress that Peijnenburg and Atkinson claim: despite the fact that there is an infinite chain of conditional probabilities, the unconditional probability of \( E_0 \) is uniquely determined.\(^1\)

The uniqueness of \( P(E_0) \) does not imply that a solution exists. However, it is an elementary fact that in the classical case, many solutions exist for given \( a \) and \( b \). The contribution of Peijnenburg and Atkinson was to show that whatever solution one takes, \( P(E_0) \) is uniquely determined for all such solutions.

Peijnenburg and Atkinson seem to take for granted that epistemic probability in general, and epistemic infinite regress in particular, should or can be modeled with classical probability theory, that is, with the classical axioms of Kolmogorov. It is precisely this implicit assumption that we have contested in this thesis. The main contributions of this chapter are the following:

1. We show that the belief in \( E_0 \) should not be uniquely determined by (8.1.1), by showing that rational reasoning does not lead to a unique value of one’s belief in \( E_0 \). The examples are framed as dialogues and do not involve any mathematics. It follows that rational reasoning in regress settings can not always be described by classical probabilities, since if this would be the case, the answer would have to be unique. This argument can be seen as a new argument against the use of classical probability theory for epistemic purposes.

2. We show that belief functions allow for enough flexibility as to model our examples correctly. In particular, modeling infinite epistemic regress with belief functions does not lead to a uniquely determined belief in \( E_0 \). All our examples can be correctly described by belief functions.

The second point reinforces earlier arguments to use belief functions instead of classical probabilities when modeling epistemic probability.

It is rewarding and useful to explain why the solution of Peijnenburg and Atkinson in (8.1.2) is unique, because this will tell us which aspects of classical probability theory may not be suitable for epistemic purposes. It is the rules of classical probability theory, or more precisely the law of total probability, which forces \( P(E_0) \) to take the unique value in (8.1.2). Indeed the law of total probability reads

\[ P(E_0) = P(E_1)P(E_0|E_1) + P(E_0^c)P(E_0|E_0^c), \]  

(8.1.3)

from which it also follows that if \( P(E_0|E_1) = P(E_0|E_0^c) = \gamma \), say, then \( P(E_0) = \gamma \).

---

\(^1\)We do not address here whether or not such a “completion” can be viewed as an answer to the infinite epistemic regress problem.
Now, writing \( x_k = P(E_k) \) and \( \gamma = a - b \), we find, using (8.1.3), that
\[
\begin{align*}
x_0 &= ax_1 + b(1 - x_1) \\
&= b + x_1 \gamma.
\end{align*}
(8.1.4)
\]
More generally, we have
\[
x_k = b + x_{k+1} \gamma.
(8.1.5)
\]
Equation (8.1.6) constitutes a system of infinitely many equations, and the question is whether or not the system has a (unique) solution. Note that a solution \( x_0, x_1, \ldots \) is only acceptable if all \( x_i \) satisfy \( 0 \leq x_i \leq 1 \), since the \( x_i \)'s represent probabilities. Iterating (8.1.6) leads to
\[
x_0 = b + b \gamma + b \gamma^2 + \cdots + b \gamma^n + \gamma^{n+1} x_{n+1},
(8.1.7)
\]
for all \( k \). Writing \( a_k := b \gamma^k \) for \( k = 0, 1, 2, \ldots \), and \( r_k := \gamma^{k+1} x_{k+1} \), we see that
\[
x_0 = \sum_{k=0}^{n} a_k + r_n.
(8.1.8)
\]
By letting \( n \to \infty \) it is not difficult to see that \( x_0 \) must take the value \( \sum_{k=0}^{\infty} a_k \) which coincides with (8.1.2).

As can be seen from these computations, the only probabilistic ingredient is the law of total probability, the rest is algebra. This law describes a relation between unconditional and conditional probabilities. This relation can be interpreted as a theorem once conditional probabilities are defined, but can also serve as an axiom in itself, see e.g. [21] for such an approach. We will see that the law of total probability is not always suitable in epistemic probability.

We end this introduction by an overview of the current chapter. In Section 8.2 we start by considering just two propositions \( E_0 \) and \( E_1 \), where we are interested in one’s belief in \( E_0 \) given \( E_1 \). As we will see by example, the law of total probability fails in a very natural way when we interpret probabilities epistemically. After the example, we show how the examples can be satisfactorily modeled with the theory of belief functions.

Before we turn to the case of infinite epistemic regress, we belief functions on infinite spaces. Some literature of belief function on continuous spaces exists (see e.g. [58]), but we want to discuss the concept of a belief function in the most general setting. Hence Section 8.3 serves as an intermezzo, where we give a definition of a belief function on an infinite space and prove a characterization.

After that, in Section 8.4, we turn to infinite epistemic regress. The earlier examples with only finitely many propositions suggest that there should be no unique solution to the problem introduced by Peijnenburg and Atkinson when we replace classical probabilities by belief functions, both with contingent and necessary conditioning. We first show by examples (very much in the same way as in the finite case) that a unique solution should indeed not be expected. We show that also these examples can be described by belief functions in a satisfactory
way. As announced, no unique solution exists anymore, and we end Section 8.4 by giving some bounds of the possible numerical values of someones belief in $E_0$ given a collection of conditional beliefs as in (8.1.1) (with probabilities replaced by beliefs).

8.2 Finite epistemic regress

As mentioned before, we will argue that when we deal with epistemic probability, the rule of total probability as formulated in (8.1.3) is not always reasonable, and we will start with an example, framed as a dialogue. In this example a bet is offered to T., and initially he is willing to pay 2/3 for it. However, when given any information about the outcome of one of the coins involved, he would be only willing to pay 1/2, for very good reasons. This phenomenon, naturally arising here, cannot be captured by classical probabilities because of (8.1.3).

Example 8.2.1. T. receives a call. It is R. on the other end of the line.
R: ‘I have an experiment for you. I have two coins and a fair die. I roll the die. If I roll 1 or 2, I put both coins with heads face up. If I roll 3 or 4, I put both coins with tails face up. And I do not tell you what I will do with the coins if I roll a 5 or a 6. How much would you pay for a bet that pays out 1 if the coins have the same side facing up?’
T: ‘I would pay 2/3 for that, since I know that at least 2 out of 3 times, namely when you roll 4 or less, the coins have the same side facing up. And since I have no idea what you do when you roll an 5 or 6, I cannot accept any bet on that part, so 2/3 is the right amount.’
R: ‘If you are ignorant about my actions when I roll a 5 or 6, should it not be fifty-fifty on that part and should you not be willing to pay 1/6 extra, so 5/6, for that bet?’
T: ‘I would certainly do so say if I would know that, for example, you flip both coins if you roll a 5 or 6. However, I simply do not know that. Suppose that we would repeat this experiment many times, and suppose that every time, I would be willing to pay more than 2/3 for the bet that pays out 1 if both coins have the same side facing up. I simply cannot rule out the possibility that you are cheating and that the coins are guaranteed to have opposite sides facing up every time that you roll 5 or 6. And if that is indeed the case, then paying anything more than 2/3 for that bet, would cost me a fortune in the long run.’
R: ‘Of course it could be the case that I put the coins on opposite sides all the time. However, what are the odds of that happening?’
T: ‘That is just shifting the problem. Look, I do not even know whether you have a choice in the matter. There is just no argument or evidence available to me to reach any conclusion whatsoever. Hence I insist that I do not pay anything more than 2/3 for the bet you offer me.’
R: ‘But this bet may actually pay you out if I roll 5 or 6, how can you say that is worth nothing?’
T: ‘I do not say that I think that is worth nothing. I just say that I am not paying
anything more than the 2/3 I was already willing to pay. Those two things are
not the same: if I would think that is worth nothing, I would also be willing to
give the away bet away for 2/3. I certainly would not do that, precisely because
of the potential pay out of the bet if you roll 5 or 6.’
R: ‘For what amount would you be willing to sell the bet then?’
T: ‘I would not be willing to sell the bet for less than 1.’
R: ‘I do not understand this.’
T: ‘Well, by the same token, really. Since I have no information about the relation
between the outcomes of the coins if you roll 5 or 6, I do not want to take any
risk that might ruin me if we would repeat the experiment many times. So the
only possible prize to sell the bet for would be 1.’
R: ‘OK that makes sense. Now suppose that I would tell that the second coin is
in fact heads. Would that change your mind about buying the bet?’
T: ‘Then I would know that you did not roll 3 or 4. If you rolled a 1 or 2, both
coins are heads. If you rolled 5 or 6, then the second coin must be heads, but I still
do not know anything about the first coin. So, of four the remaining possibilities,
which are that you rolled a 1, 2, 5 or 6, only the first two guarantee that both
coins have the same side facing up. So I would now pay 1/2 for the bet.’
R: ‘And what if I would tell you that the second coin is in fact tails?’
T: ‘That is a similar situation, but with the roles of heads and tails being reversed.
I would know that you did not roll 1 or 2. Of the remaining four possibilities,
again only two guarantee that both coins have the same side facing up. So, again,
I would pay 1/2 for the bet.’
R: ‘But this is weird! There are only two cases: either the second coin shows
heads or the second coin shows tails. You act exactly the same in case I tell you
the second coin is heads and in case I tell you the second coin is tails. So, now
really you could have concluded yourself that you had to end up in one of those
two cases, and the price you would be willing to pay for the bet should have been
1/2 to begin with.’
T: ‘No, I disagree.
R: ‘I think you need to explain this.’
T: ‘The context and the precise circumstances matter, and the way I receive
information from you has an effect on the price I am willing to pay for a bet.’
R: ‘How?’
T: ‘Well, suppose we agree that you are going to repeat the experiment many
times, and that you promise me that each time you offer me the bet given the
information that the second coin is either heads or tails, depending on the actual
outcome which is known to you.’
R: ‘Yes.’
T: ‘In that case I would be willing to pay 2/3 for every bet you offer me. The
law of large numbers will tell me that on average I will not loose money, and
depending on what you do when you roll a 5 or 6, I might even win some money
but I can of course not be sure about that.’
R: ‘That sounds completely reasonable to me.’
T: ‘Yes, but this is not the situation. I interpret you giving me the information that the second coin is heads in the same way conditioning in a frequentist setting is interpreted: it means that you only offer me the bet if the second coin is in fact heads, and ignore all repetitions where the second coin is tails. Then the situation changes.’
R: ‘How does this make things different?’
T: ‘Well, the problem is that although every time you roll a 5 or 6 the second coin is either heads or tails, I cannot ignore any roll of a 5 or a 6 since for all I know, the second coin might always be heads in these cases. So if you tell me that the second coin shows heads, then the above reasoning leading to 1/2 is the correct one.’
R: ‘I see your point, but it remains strange that the unconditional belief is 2/3, but that the conditional beliefs, given both heads or tails for the second coin, are both equal to 1/2.
T: ‘Perhaps it is unusual, but it is totally rational. It all depends on the circumstances under which I come to know that the second coin shows heads.’
R: ‘Very well, but now we are in trouble, since the usual law of total probability does not seem to work in this situation.’
T: ‘Exactly! This example shows that rational beliefs should not always be based on the usual axioms of probability theory and that we need a more refined theory.’

Let $E_0$ be the proposition that the two coins have the same side facing up in the example above, and $E_1$ the proposition that the second coin is heads. The example illustrates that under the constraint that the belief in $E_0$ given either $E_1$ or $E_1^c$ is equal to 1/2, it is possible to have that the initial belief in $E_0$ is 2/3. This is of course impossible to describe with probability theory, since $P(E_0|E_1) = P(E_0|E_1^c) = 1/2$ would imply that $P(E_0) = 1/2$. Hence classical probability theory draws the wrong conclusion here that the belief in $E_0$ would in fact be uniquely determined and equal to 1/2.

To show how this example can be modeled with a belief function, we set

$$
\Omega = \{(h, h), (h, t), (t, h), (t, t)\}.
$$

We interpret the first coordinate as the outcome of the first coin (where ‘h’ denotes heads and ‘t’ denotes tails) and the second coordinate as the outcome of the second coin. T. knows that 1/3 of the time the two coins are heads, so

$$m(\{(h, h)\}) = \frac{1}{3}, \quad (8.2.1)$$

He also knows that another 1/3 of the time, the two coins are tails, so

$$m(\{(t, t)\}) = \frac{1}{3}. \quad (8.2.2)$$

T. is ignorant about the remaining 1/3 of instances, and this translates into

$$m(\Omega) = \frac{1}{3}. \quad (8.2.3)$$
This is precisely all he knows, so \( m(C) = 0 \) for all other sets \( C \). T. is now interested in \( E_0 = \{(h,h), (t,t)\} \). We compute

\[
\text{Bel}(E_0) = m(\{(h,h),(t,t)\}) + m(\{(h,h)\}) + m(\{(t,t)\}) = \frac{2}{3},
\]

(8.2.4)

and this agrees with the viewpoint of T. in the dialogue to pay 2/3 for the bet on \( E_0 \).

In the example, after the initial information, R. gives additional information on which T. changes his views on what he wants to pay for the bet. The kind of information T. receives, is the information that the truth is in some set \( H \subseteq \Omega \) in a contingent way. Let us apply this form of conditioning to our example. When R. informs T. that the second coin is heads, T. should apply contingent conditioning on \( E_1 = \{(h,h), (t,h)\} \). The set \( \{(t,t)\} \) is the only outcome with positive chance that is consistent with \( E_1 \). Hence we find

\[
\text{Bel}_{E_1}(E_0) = \frac{m(\{(h,h)\})}{1 - m(\{(t,t)\})} = \frac{1}{2}.
\]

(8.2.5)

In exactly the same way we find

\[
\text{Bel}_{E_1^c}(E_0) = \frac{m(\{(t,t)\})}{1 - m(\{(t,t)\})} = \frac{1}{2}.
\]

(8.2.6)

**Example 8.2.2.** R. is calling T. again.

R: ‘I would like to offer you a bet. I have two coins. In a moment, I will flip the first one in a fair way and tell you if it is heads or tails. I do not tell you anything else. How much are you willing to pay for a bet that pays out 1 if the two coins have the same side facing up?’

T: ‘Let me see. I am completely ignorant about the relation between the two coins, right? Well then, obviously I am not willing to pay anything for that bet. Perhaps you do not flip the second coin at all, and you simply always put the opposite side of the first flip up.’

R: ‘All right, I can see that. But what if I would tell you that the second coin is necessarily heads?’

T: ‘In that case I would be willing to pay 1/2 for the bet that pays out 1 if both coins have the same side facing up.’

R: ‘How does information about the second coin nullify your earlier objections to pay anything for that bet?’

T: ‘By telling me that the second coin is necessarily heads, you reduce my ignorance about the coins having the same side up, to my ignorance about the flip of the first coin: if the first coin is heads, the coins have the same side up and if the flip is tails, the coins do not have the same side up.’

R: ‘And I assume you come to the same conclusion if I would tell you that the second coin is necessarily tails?’

T: ‘That is correct. By a completely analogous argument, I would be willing to pay 1/2 for the bet.’
R: ‘But then I have the same objection as in the previous experiment: you act exactly the same in case I tell you the second coin is necessarily heads and in case I tell you the second coin is necessarily tails. Of course you could have concluded yourself that you had to end up in one of those two cases, so the price you would be willing to pay for the bet should have been 1/2 instead of zero, after all.’
T: ‘Here, the response to that is very simple. You seem to suggest that the second coin being necessarily heads and the second coin being necessarily tails, are the only two cases. However, that is not true. The third possibility is that the second coin is *sometimes* tails and *sometimes* heads. By telling me that we are in one of the first two cases, I know that this third possibility is ruled out. And this enables me to rule out scenarios like the one in which the coins are opposite all the time. Because if the two coins were each others opposites all the time, the second coin has to be heads half of the time and tails the other half of the time.’

Let us see how Example 8.2.2 can be modeled with belief functions. We set \( \Omega = \{ (h,h), (h,t), (t,h), (t,t) \} \) as before. T. knows that half of the time the first coin is heads, so

\[
m((h,h), (h,t)) = \frac{1}{2}. \tag{8.2.7}
\]

He also knows that the other half of the time, the first coin is tails, so

\[
m((t,h), (t,t)) = \frac{1}{2}. \tag{8.2.8}
\]

This is precisely all he knows, so \( m(C) = 0 \) for all other sets \( C \). T. is now interested in the proposition ‘The two coins have the same side facing up’, which we call

\[
E_0 = \{(h,h), (t,t)\}. \tag{8.2.9}
\]

We compute

\[
\text{Bel}(E_0) = m((h,h), (t,t)) + m((h,h)) + m((t,t)) = 0 \tag{8.2.10}
\]

and see that this agrees with the viewpoint of T. in the dialogue not to pay anything for the bet on \( E_0 \).

When R. tells that the second coin is necessarily heads, T. applies necessary conditioning to \( E_1 = \{(h,h), (t,h)\} \). The sets \( \{(h,h), (h,t)\} \) and \( \{(t,h), (t,t)\} \) are the only outcomes with positive chance. Clearly, they are both consistent with \( E_1 \), so necessary conditioning applies. Intersected with \( E_1 \) they become respectively \( \{(h,h)\} \) and \( \{(t,h)\} \) so we find

\[
\text{Bel}_{E_1, \text{ nec}}(E_0) = m((h,h), (h,t)) = \frac{1}{2}. \tag{8.2.11}
\]

In exactly the same way we find

\[
\text{Bel}_{E_1^c, \text{ nec}}(E_0) = m((t,h), (t,t)) = \frac{1}{2}. \tag{8.2.12}
\]
in agreement with the conclusion of T. in the dialogue.

So far we have encountered two types of conditioning: contingent and necessary. In the following dialogue, there is a situation in which one should apply mixed conditioning.

Example 8.2.3. R. is calling T. a final time
R: ‘I have another variation of the experiment for you. Again, I have two coins and a fair die. I roll the die and flip the first coin. If I roll 1, I put both coins with heads face up. If I roll 2, I put both coins with tails face up. I do not tell you what I will do with the coins if I roll 3. If I roll 4, 5 or 6, I leave the first coin untouched, but I do not tell you what I will do with the second coin.’
T: ‘This is a mix of the last two experiments!’
R: ‘Precisely! So how much would you pay for a bet that pays out 1 if the coins have the same side facing up?’
T: ‘Analogous to the arguments I made before, I would pay 1/3 for that, since I know that 2 out of 6 times, namely when you roll 1 or 2, the coins have the same side facing up.’
R: ‘What if I tell you that the second coin is \textit{necessarily} heads except for when I roll a 2 and that the second coin is in fact heads?’
T: ‘This tells me that you did not roll 2. So, of the 12 combinations of die rolls and flips of the first coin, only 10 remain. If you rolled 1, it is guaranteed that both coins are the same. That are 2 of the remaining 10 options. It is also guaranteed if you rolled 4, 5 or 6 and flipped heads for the first one. That are 3 of the remaining 10 options. So, of the 10 remaining options, there are 5 in which I know that the two coins are the same. Hence I would be willing to pay 1/2 for the bet.’
R: ‘And I assume this would be the same if I tell you that the second coin is \textit{necessarily} tails except for when I roll a 1 and that the second coin is in fact tails?’
T: ‘Correct.’

We find the model corresponding to Example 8.2.3. We consider the same $\Omega$ again. T. knows that 1/6 of the time, the two coins are heads, so

\[ m(\{(h,h)\}) = \frac{1}{6}. \]  \hfill (8.2.13)

He also knows that another 1/6 of the time, the two coins are tails, so

\[ m(\{(t,t)\}) = \frac{1}{6}. \]  \hfill (8.2.14)

T. knows that 1/4 of the time, the first coin is heads, so

\[ m(\{(h,h),(h,t)\}) = \frac{1}{4}. \]  \hfill (8.2.15)

Similarly, he knows that 1/4 of the time, the first coin is tails, so

\[ m(\{(t,h),(t,t)\}) = \frac{1}{4}. \]  \hfill (8.2.16)
This is precisely all he knows, so \( m(\Omega) = 1/6 \) and \( m(C) = 0 \) for all other sets \( C \). T. is now again interested in \( E_0 = \{(h, h), (t, t)\} \). We compute

\[
\text{Bel}(E_0) = m(\{(h, h), (t, t)\}) + m(\{(h, h)\}) + m(\{(t, t)\}) = \frac{1}{3}. \tag{8.2.17}
\]

and see that this agrees with the viewpoint of T. in the dialogue to pay 1/3 for the bet on \( E_0 \).

When R. tells that the second coin is \textit{necessarily} heads except for when R. rolls a 2 and that the second coin is in fact heads, T. should apply mixed conditioning for \( E_1 = \{(h, h), (h, t)\} \). The set \( \{(t, t)\} \) is the only set with positive chance that is inconsistent with \( H \). The sets \( \{(h, h)\}, \{(h, h), (h, t)\} \) and \( \{(t, h), (t, t)\} \), when we conjunct them with \( E_1 \), become \( \{(h, h)\}, \{(h, h)\} \) and \( \{(t, h)\} \) respectively. So, we find

\[
\text{Bel}_{E_1,\text{mix}}(E_0) = \frac{m(\{(h, h)\}) + m(\{(h, h), (h, t)\})}{1 - m(\{(t, t)\})} = \frac{1}{2}. \tag{8.2.18}
\]

In exactly the same way we find

\[
\text{Bel}_{E_1^c,\text{mix}}(E_0) = \frac{m(\{(t, t)\}) + m(\{(t, h), (t, t)\})}{1 - m(\{(h, h)\})} = \frac{1}{2}. \tag{8.2.19}
\]

### 8.3 Belief functions on infinite spaces

Let \( \Omega \) be an arbitrary, infinite outcome space. Since the inclusion-exclusion inequality is very easy to generalize, we choose this as our definition of a belief function on \( \Omega \).

**Definition 8.3.1.** Let \( \mathcal{A} \subseteq 2^\Omega \) be an algebra of sets. Then Bel : \( \mathcal{A} \rightarrow [0, 1] \) is a belief function if

- Bel(\( \emptyset \)) = 0 and Bel(\( \Omega \)) = 1;
- For all \( A_1, \ldots, A_N \in \mathcal{A} \), we have

\[
\text{Bel} \left( \bigcup_{i=1}^N A_i \right) \geq \sum_{I \subseteq \{1, \ldots, N\}, I \neq \emptyset} (-1)^{|I|+1} \text{Bel} \left( \bigcap_{i \in I} A_i \right). \tag{8.3.1}
\]

We now give a characterization of belief functions analogue to the characterization of belief functions on finite spaces in terms of the basic belief assignment. On finite spaces, there is a one-to-one correspondence between the basic belief assignment \( m : 2^\Omega \rightarrow [0, 1] \) and \( P : 2^\Omega \) given by \( P(\{S\}) = m(S) \). Since probability distributions on general spaces are no longer determined by their values on singletons, there is no place for a basic belief assignment \( m \) in the general theory. For \( P \), we need an appropriate domain that in general will no longer be
8.3. BELIEF FUNCTIONS ON INFINITE SPACES

the entire space $2^\Omega$. Since we at least want to talk about the probability of the collection \{C : C \subseteq A\} for sets $A \in \mathcal{A}$ on which our belief function is defined, we introduce the following notation. If $\mathcal{A} \subseteq 2^\Omega$ is an algebra of sets, we write

$$\mathcal{A}^* := \{\{C : C \subseteq A\} : A \in \mathcal{A}\},$$

(8.3.2)

which is a collection of collections of sets. Note that $\mathcal{A}^*$ is not an algebra, since the complement of $\{C : C \subseteq A\}$ is

$$\{C : C \nsubseteq A\}$$

(8.3.3)

and can not be written as the collection of all subsets of some other element $A' \in \mathcal{A}$. The collection $\mathcal{A}^*$ is, however, closed under intersections since

$$\{C : C \subseteq A_1\} \cap \{C : C \subseteq A_2\} = \{C : C \subseteq A_1 \cap A_2\}$$

(8.3.4)

and $A_1 \cap A_2 \in \mathcal{A}$ if $A_1, A_2 \in \mathcal{A}$. This means that (finite additive) probability distributions $P$ defined on $a(\mathcal{A}^*)$, the algebra generated by $\mathcal{A}^*$, are uniquely determined by their values on $\mathcal{A}^*$. We now present the result, that gives a one-to-one correspondence between finitely additive probability distributions on $a(\mathcal{A}^*)$ and belief functions on $\mathcal{A}$.

**Theorem 8.3.2.** Let $\mathcal{A} \subseteq 2^\Omega$ be an algebra of sets. A function $\text{Bel} : \mathcal{A} \rightarrow [0, 1]$ is a belief function if and only if there is a finitely additive probability measure $P : a(\mathcal{A}^*) \rightarrow [0, 1]$ such that

$$\text{Bel}(A) = P(\{C : C \subseteq A\})$$

(8.3.5)

for every $A \in \mathcal{A}$.

**Proof.** The ‘if’ part of the follows from elementary properties of probability measures. For the proof of ‘only if’ part, we will construct an appropriate $P : a(\mathcal{A}^*) \rightarrow [0, 1]$. We start with the following definition. We say that $\mathcal{P} \subseteq \mathcal{A}$ is a $\mathcal{A}$-partition if $|\mathcal{P}| < \infty$, for $A, B \in \mathcal{P}$ we have $A \cap B = \emptyset$ and $\cup_{A \in \mathcal{P}} A = \Omega$. For an arbitrary $\mathcal{A}$-partition $\mathcal{P}$, we define $\text{Bel}_\mathcal{P} : 2^\mathcal{P} \rightarrow [0, 1]$ by

$$\text{Bel}_\mathcal{P}(S) := \text{Bel}\left(\bigcup_{S \in \mathcal{S}} S\right)$$

(8.3.6)

121
for any $S \subseteq P$. Now we observe that
\[
\text{Bel}_P \left( \bigcup_{i=1}^{N} S_i \right) = \text{Bel} \left( \bigcup_{S \in \bigcup_i S_i} S \right) \\
= \text{Bel} \left( \bigcup_{i=1}^{N} \bigcup_{S \in S_i} S \right) \\
\geq \sum_{I \subseteq \{1, \ldots, N\}, I \neq \emptyset} (-1)^{|I|+1} \text{Bel} \left( \bigcap_{i \in I} \bigcup_{S \in S_i} S \right) \\
= \sum_{I \subseteq \{1, \ldots, N\}, I \neq \emptyset} (-1)^{|I|+1} \text{Bel} \left( \bigcup_{S \in \bigcap_{i \in I} S} S \right) \\
= \sum_{I \subseteq \{1, \ldots, N\}, I \neq \emptyset} (-1)^{|I|+1} \text{Bel}_P \left( \bigcap_{i \in I} S_i \right)
\]

(8.3.7)

Hence Bel$_P$ is a belief function on the finite outcome space $P$. Write $m_P : 2^P \to [0, 1]$ for the corresponding basic belief assignment and define $\mu_P : 2^\Omega \to [0, 1]$ by
\[
\mu_P \left( \bigcup_{i=1}^{N} A_i \right) := m_P \left( \{ A_1, A_2, \ldots, A_N \} \right) \\
\text{if } A_1, A_2, \ldots, A_N \in P \text{ and set } \mu_P(A) = 0 \text{ for all other } A \subseteq \Omega. \text{ Note that we have }
\sum_{A \subseteq \Omega} \mu_P(A) = 1,
\]

(8.3.8)

which is well defined because $\mu_P$ is only positive on a finite number of subsets.

We want to see $\mu_P$ as an ‘approximation’ for the probability measure $P$ we want to construct and define $P$ as the limit of taking more and more refined partitions. We now make this precise. Write $\mathcal{Y}$ for the space of all $A$-partitions. We now make this precise. Write $\mathcal{Y}$ for the space of all $A$-partitions. We write $\mathcal{P}' \geq \mathcal{P}$ for $A$-partitions $\mathcal{P}'$ and $\mathcal{P}$ if $\mathcal{P}'$ is a refinement of $\mathcal{P}$, i.e. for every $A' \in \mathcal{P}'$ there is a $A \in \mathcal{P}$ such that $A' \subseteq A$. Now we show that
\[
\{ \{ \mathcal{P}' : \mathcal{P}' \geq \mathcal{P} \} : \mathcal{P} \in \mathcal{Y} \}
\]

is a filter base, by checking it is closed under intersections. Let $\mathcal{P}_1, \mathcal{P}_2 \in \mathcal{Y}$ and set
\[
\mathcal{P} := \{ A \cap B : A \in \mathcal{P}_1, B \in \mathcal{P}_2 \},
\]

(8.3.10)

which is clearly a $A$-partition such that
\[
\{ \mathcal{P}' : \mathcal{P}' \geq \mathcal{P}_1 \} \cap \{ \mathcal{P}' : \mathcal{P}' \geq \mathcal{P}_2 \} = \{ \mathcal{P}' : \mathcal{P}' \geq \mathcal{P} \}.
\]

(8.3.11)

Now set
\[
\mathcal{F} := \{ S \subseteq \mathcal{Y} : \exists \mathcal{P} \ S \supseteq \{ \mathcal{P}' : \mathcal{P}' \geq \mathcal{P} \} \},
\]

(8.3.12)
which is the filter generated by this filter base.

Let \( C \subseteq 2^\Omega \) be fixed and set \( \phi_C : \mathcal{Y} \to [0,1] \) by

\[
\phi_C(\mathcal{P}) := \sum_{C \in \mathcal{C}} \mu(\mathcal{P}), \quad (8.3.14)
\]

which gives, for every \( \mathcal{P} \in \mathcal{Y} \), our ‘approximate’ probability of \( C \) based on \( \mathcal{P} \).

We could now define \( P \) as the limit of \( \phi_C \) along the filter \( \mathcal{F} \). A limit along a filter does, however, not necessarily exist for every \( C \subseteq 2^\Omega \). Instead of worrying about whether or not this limit exists for every \( C \in a(\mathcal{A}^*) \), we can simply use the ultrafilter lemma to extend \( \mathcal{F} \) to an ultrafilter \( \mathcal{U} \). Of course this extension is not unique, but we only care about existence here. The limit along an ultrafilter always exists, so we can define \( P \) on \( 2^{2^\Omega} \), by setting

\[
P(\mathcal{C}) := \mathcal{U} - \lim(\phi_C), \quad (8.3.15)
\]

the limit of \( \phi_C \) along the ultrafilter \( \mathcal{U} \). We clearly have \( P(2^\Omega) = 1 \) and by definition of \( \phi_C \) and the fact that limits along filters are additive, we find that \( P \) is additive and hence a probability measure.

We now let \( A \in \mathcal{A} \), take \( \mathcal{P} \geq \{A, A^c\} \) and write \( \mathcal{P}_A := \{B \in \mathcal{P} : B \subseteq A\} \). Then

\[
\phi_{\{C : C \subseteq A\}}(\mathcal{P}) = \sum_{C \subseteq A} \mu(\mathcal{P}),
\]

\[
\sum_{S \subseteq \mathcal{P}_A} m(\mathcal{S}) = \text{Bel}(\mathcal{P}_A)
\]

\[
= \text{Bel}(A).
\]

So

\[
P(\{C \subseteq A\}) = \text{Bel}(A). \quad (8.3.17)
\]

Restricting \( P \) to \( a(\mathcal{A}^*) \) gives the desired result.

**8.3.1 Continuous belief functions?**

One might wonder why we did not include the constraint of \( \sigma \)-continuity in the definition of a belief function, i.e. that we would require that a belief function \( \text{Bel} : \mathcal{A} \to [0,1] \) has the property that for \( A_1, A_2, A_3, \ldots \in \mathcal{A} \), we have

\[
\text{Bel}\left(\bigcup_{i=1}^\infty A_i\right) = \sup_k \text{Bel}\left(\bigcup_{i=1}^k A_i\right) \quad (8.3.18)
\]

and

\[
\text{Bel}\left(\bigcap_{i=1}^\infty A_i\right) = \inf_k \text{Bel}\left(\bigcap_{i=1}^k A_i\right). \quad (8.3.19)
\]
Of course this requirement is debatable in the same way it is debatable for probability measures (see the first part of this thesis). But in the case of belief functions, another more practical reason to not require \( \sigma \)-continuity, is the following. One would hope that by considering \( \sigma \)-continuous belief functions, one could prove Theorem 8.3.2 again with finitely additive probability measures replaced by \( \sigma \)-additive probability measures. This, however, can not be done, as illustrated by the next counterexample.

Consider the space \( \mathbb{N} \) and, for \( 1 \leq j \leq 4^k \), we set

\[
A_{k,j} := \{ j, j + 4^k, j + 2 \cdot 4^k, j + 3 \cdot 4^k, \ldots \} \subseteq \mathbb{N}
\] (8.3.20)

and \( m_k : 2^\mathbb{N} \to [0, 1] \) by

\[
m_k(A_{k,j}) := \frac{1}{4^k}
\] (8.3.21)

and \( m_k(C) := 0 \) for all other \( C \). We note that for every \( C \subseteq \mathbb{N} \) that is not a singleton, there exists a \( K \) such that for all \( k > K \) and \( j \) we have \( C \not\subseteq A_{k,j} \). For \( n \in \mathbb{N} \), we have \( \{ n \} \subseteq A_{k,n} \) for all \( k \geq n \) and \( \{ n \} \not\subseteq A_{k,j} \) for all \( j \neq n \). This means that \( \text{Bel} : 2^\Omega \to [0, 1] \) given by

\[
\text{Bel}(C) := \lim_{k \to \infty} \sum_{B \subseteq C} m_k(B).
\] (8.3.22)

is well defined, because the limit exists for every \( C \subseteq \mathbb{N} \).

We claim that \( \text{Bel} \) is a belief function. Further, we have

\[
\text{Bel} \left( \bigcap_{i=1}^\infty C_i \right) = \lim_{k \to \infty} \sum_{B \subseteq \bigcap_{i=1}^\infty A_i} m_k(B)
\]

\[
= \lim_{k \to \infty} \lim_{n \to \infty} \sum_{B \subseteq \bigcap_{i=1}^n C_i} m_k(B)
\]

\[
= \lim_{n \to \infty} \lim_{k \to \infty} \sum_{B \subseteq \bigcap_{i=1}^n C_i} m_k(B)
\]

\[
= \lim_{n \to \infty} \text{Bel} \left( \bigcap_{i=1}^n C_i \right).
\] (8.3.23)

A similar argument works of course for continuity from below.

Now suppose that \( P : \sigma((2^\mathbb{N})^*) \to [0, 1] \) is a \( \sigma \)-additive probability measure such that \( P(\{ C \subseteq A \}) = \text{Bel}(A) \) for every \( A \subseteq \mathbb{N} \). For \( k \geq 1 \), we set

\[
B_k := \bigcup_{j=k}^{4^k} \{ C \subseteq A_{k,j} \}.
\] (8.3.24)

Then, on the one hand we have

\[
P \left( \bigcap_{k=1}^\infty B_k \right) = P(\{ \emptyset \}) = 0,
\] (8.3.25)
8.4. INFINITE EPISTEMIC REGRESS

and on the other hand we have

\[
P \left( \bigcap_{k=1}^{n} B_k \right) = 1 - P \left( \bigcup_{k=1}^{n} B_k^c \right) \\ \geq 1 - \sum_{k=1}^{n} P(B_k^c) \\ = 1 - \sum_{k=1}^{n} \frac{k - 1}{4^k} \geq \frac{8}{9}.
\]

(8.3.26)

This contradicts the fact that \( P \) is a \( \sigma \)-additive probability measure. So there is no \( \sigma \)-additive probability measure such that \( P(\{C \subseteq A\}) = \text{Bel}(A) \) for every \( A \subseteq \mathbb{N} \).

8.4 Infinite epistemic regress

8.4.1 Three examples of infinite epistemic regress

Now that we have introduced belief functions on infinite spaces, we finally turn our attention to situations of genuine infinite probabilistic regress. We describe three situations of infinite regress, all framed in the form of a dialogue, as before. All three examples are such that they satisfy (8.1.1) with \( a = 0 \) and \( b = 1/2 \), with a suitable interpretation of conditioning. However, we will see that the degree to which the agent believes \( E_0 \) is different each time, namely 1/3, 0, and 2/3 respectively.

The first example is a classical situation as described by Peijnenburg and Atkinson.

Example 8.4.1. After a good portion of thinking about all kinds of experiments, T. contacts R.

T: ‘Let \( t_0 = 1 \), \( t_1 = 1/2 \), \( t_2 = 1/3 \), \( t_3 = 1/4 \),… be moments in time. Now suppose that I have a lamp and keep myself to the following instruction. If the lamp is on at time \( t_k + 1 \), I switch the lamp off so that it is off at time \( t_k \). If the lamp is off at time \( t_k + 1 \), I flip a fair coin randomly, switching the lamp on at time \( t_k \) if it is heads and not do anything if it is tails.’

R: ‘This sounds like Thomson’s lamp reversed and with a probabilistic twist.’

T: ‘Exactly! So, what is your belief that the lamp is on at time \( t_0 \)?’

R: ‘Well, I think we are now in the situation of (8.1.1) with \( a = 0 \) and \( b = 1/2 \), so according to (8.1.2) the answer is 1/3. This is the degree to which I believe that the lamp is on at time \( t_0 \).’

T: ‘Right! And I assume that if I would tell you that the lamp is on at time \( k + 1 \), that your belief that the lamp is on at time \( k \) is zero?’

R: ‘Of course. Similarly, if you would tell me that the lamp is off at time \( k + 1 \), I would have a degree of belief of 1/2 that the lamp is on at time \( k \).’
Of course this dialogue is an example of an infinite probabilistic regress problem that can be described by a probability distribution, as the dialogue itself already suggests. We now turn to an example where this is no longer the case.

Example 8.4.2. R. contacts T. for a follow-up on the lamp.
R: ‘I am still thinking about the thought experiment with the lamp. Let’s suppose I go about the process completely differently. Suppose I throw a tetrahedron whose sides are numbered 1, 2, 3 and 4. If I roll 1, I switch the lamp on at \( t_0, t_2, t_4, \ldots \) and I switch it off at \( t_1, t_3, t_5, \ldots \). If I roll 2, I do the opposite: I switch the lamp off at \( t_0, t_2, t_4, \ldots \) and on at \( t_1, t_3, t_5, \ldots \). If I roll 3, I switch the lamp off at all times. Finally, if I roll 4 I switch the lamp on at all even times but I do not tell you what happens at the other moments in time. To what degree would you believe that the lamp is on at time \( t_0 \)?’
T: ‘Well, that is easy. Only if you roll 1 or 4, I know that the lamp is on at time \( t_0 \). The lamp is off when you roll 2 or 3, so the probability that the lamp is on at time \( t_0 \) is clearly 1/2, so that is the amount of money I would be willing to bet if the payout is 1.’
R: ‘That sounds completely reasonable to me. But, if I would tell you that the lamp is on at time \( t_{k+1} \), what do you believe about the lamp being on at time \( t_k \)? In other words, what is your conditional belief in the lamp being on at time \( t_k \) given that it is on at time \( t_{k+1} \)?’
T: ‘I think I must distinguish between two cases here. The first case is that \( k+1 \) is even, say \( k+1 = 4 \), just to be concrete. In that case, I know that you did not roll 2 or 3. I also know that if you rolled 1, the lamp is off at time \( t_3 \). If you rolled 4, I cannot conclude anything about the lamp being on at time \( t_3 \) since you gave me no information. For all I know, the lamp may be always off at time \( t_3 \) if you roll 4 and in the long run this would bankrupt me if I would be willing to pay anything at all for this. So no matter what, I have no belief in the lamp being on at time \( t_3 \) so my belief would be zero.’
R: ‘I do not understand the last bit. If I roll 4, I agree that you have no information, so wouldn’t that amount to fifty-fifty in that case?’
T: ‘No. We saw this before. Having no information is not the same as fifty-fifty. Again, for all I know it might always be the case that if you roll 4 the lamp is off at time \( t_3 \). I simply do not know. If we then would repeat this bet many times, I would lose a lot even if I would be willing to pay a small amount for this bet. So it really is rational not to pay anything for this bet given the information I have.’
R: ‘All right, fair enough, but what about the case when \( k+1 \) is odd, let’s say 3?’
T: ‘Ok, so now I suppose the lamp is on at time \( t_3 \). In that case, I know that you did not roll 1 or 3. And I know that if you rolled 2 the lamp is off at time \( t_2 \). Of course, if you would have rolled 4, the lamp would be on at time \( t_2 \), so it seems that I would be willing to pay 1/2. But hold on, this is not correct! If you roll 4, then the lamp may or may not be on at time \( t_3 \), both options are possible as far as I know. But at time \( t_2 \) it would always be on. Since my project is to say something intelligent about the relative frequency of the lamp being on at time \( t_2 \) relative to it being on at time \( t_3 \), I should completely ignore you throwing 4.’
R: ‘This I do not understand.’
T: ‘Well, the crucial point is that leaving out instances where the lamp is on at both time $t_2$ and $t_3$ will only decrease the relative frequency of the lamp being on at time $t_2$ relative to it being on at time $t_3$. So ignoring is safe for me, whereas betting any positive amount on it may prove disastrous.’
R: ‘All right, I understand now. You want to play safe, and you look for the maximum amount you want to bet to make sure that you play safe.’
T: ‘Indeed.’
R: ‘All right, so you have convinced me now that the conditional belief in the lamp being on at time $t_k$ given that it is on at time $t_{k+1}$ is zero. But what if I told you that the lamp is off at time $t_k$?’
T: ‘Well let’s see. Again, two cases have to be distinguished. The first case is that $k + 1$ is even. In that case, I know that you did not roll 1 or 4. Two cases remain. If you rolled a 2, the lamp is on at time $t_k$. If you rolled 3, it is not. This case is therefore much simpler than the previous one: out of the two remaining options, I know that one has the lamp on at time $k$, so my belief would be $1/2$. The second case is that $k + 1$ is odd. In that case, I know that you did not roll 2. I also know that if you rolled 1, the lamp is on at time $t_k$ and if you rolled 3 it is not. The possibility that you rolled 4 is special again, and the argument to ignore it is exactly the same as above. So out of the two options that I need to consider, only one gives that the lamp is on at time $t_2$, and therefore my belief would be $1/2’.

The most natural way to model the infinite case is to consider the outcome space $\Omega := \{0, 1\}^\infty$ and setting

$$E_i := \{(\omega_0, \omega_1, \omega_2, \ldots) \in \Omega : \omega_i = 1\}. \quad (8.4.1)$$

Write $\mathcal{F} := 2^\Omega$ for all subsets of $\Omega$. Consider a probability distribution $P : 2^\mathcal{F} \to [0, 1]$ such that there are a finite number of sets $A_1, \ldots, A_n$ such that

$$\sum_{i=1}^n P(\{A_i\}) = 1. \quad (8.4.2)$$

Then

$$\text{Bel}(A) = P(\{C \in \mathcal{F} : C \subseteq A\}) \quad (8.4.3)$$

is a belief function by Theorem 8.3.2. For these $P$, we again write

$$m(A) = P(\{A\}). \quad (8.4.4)$$

We now model Example 8.2.3 with a belief function. The corresponding basic belief assignment is

$$m(E_0 \cap E_0^c \cap E_2 \cap E_3^c \cap \cdots) = 1/4,$$
$$m(E_0^c \cap E_1 \cap E_2^c \cap E_3 \cap \cdots) = 1/4,$$
$$m(E_0^c \cap E_1^c \cap E_3^c \cap \cdots) = 1/4,$$
$$m(E_0 \cap E_2 \cap E_4 \cap \cdots) = 1/4. \quad (8.4.5)$$
It is easy to see that \( \text{Bel}(E_0) = 1/2 \) (coming from the first and fourth line). Next we compute the conditional beliefs. Taking \( k = 2 \) for concreteness, we find that
\[
\text{Bel}_{E_3}(E_2) = 0
\] (8.4.6)
since \( \text{Bel}(E_2 \cap E_3) = 0 \). Furthermore
\[
\text{Bel}_{E_3}(E_2) = \frac{\text{Bel}(E_2 \cap E_3^c)}{\text{Bel}(E_2 \cap E_3^c) + 1 - \text{Bel}(E_2 \cup E_3)} = \frac{1/4}{1/4 + 1 - 3/4} = 1/2,
\]
in complete agreement with the conclusion of T. in the dialogue.

We now give an example of infinite epistemic regress in which we need to consider necessary conditioning.

Example 8.4.3. R. contacts T. for another follow-up on the lamp.

R: ‘I am still thinking about the thought experiment with the lamp. Suppose I go about the process completely differently. I just throw one fair coin randomly. If it is heads, I assure you that if the lamp is off at time \( t_{k+1} \), I switch the lamp to be on at \( t_k \). That is all the information I give to you. I do not tell you what happens when the coin shows tails. What would your belief in the lamp being on at time \( t_0 \) be this time?’

T: ‘If you do it that way, I cannot say anything at all about the lamp being on at \( t_0 \). After all, you only told me something conditional, and I do not see how I can get any information about the real state of affairs from what you told me. So my degree of belief in the lamp being on at \( t_0 \) would be zero.’

R: ‘Fair enough. So that would be different from our first setup. But, if I would tell you that the lamp is necessarily on at time \( t_{k+1} \), what would you believe about the lamp being on at time \( t_k \)?’

T: ‘Again your information simply does not give any information which would be useful to answer this question. You did not tell me enough to conclude anything at all about the lamp being on at time \( t_k \), so my degree of belief regarding the lamp being on at \( t_k \) would still be zero.’

R: ‘And if I told you that the lamp is necessarily off at time \( t_{k+1} \)?’

T: ‘Well, that changes my position dramatically of course. In that case I know that if you flip heads, that the lamp will be on at time \( t_k \). And since you flip heads half of the time, my degree of belief regarding the lamp being on at \( t_k \) would be 1/2.’

We now show how to model Example 8.4.3 with belief functions. T. only knows that half of the time, the lamp is not off at two consecutive time moments. Hence this example can be modeled by
\[
m(\Omega) = m((E_1 \cup E_2) \cap (E_2 \cup E_3) \cap \cdots) = \frac{1}{2},
\] (8.4.7)
It is easy to see \( \text{Bel}(E_0) = 0 \). This is precisely the conclusion of T. Since R. supposes that the lamp is necessarily off or necessarily on at \( t_k \), we need to
condition in the necessary sense on respectively $E_{n+1}^c$ and $E_{n+1}$. If we do so, we find
\[ \text{Bel}_{E_{n+1}^c, \text{nec}}(E_n) = \frac{1}{2}, \]  
for all $n$, which is precisely the conclusion of T.

### 8.4.2 Bounds for the various forms of conditioning

Finally we present some results about bounds for the various notion of conditioning. Since the completion of the infinite probabilistic regress is no longer unique, it is a natural question as to how much freedom the new theory actually gives under the assumption that the conditional belief (in any of the three versions) of $E_n$ given $E_{n+1}$ and $E_{n+1}^c$ is equal to $a$ and $b$ respectively. The bounds below may not always be optimal. Finding better bounds is an interesting mathematical problem, but is beyond the scope of this thesis.

We start with necessary conditioning. Necessary conditioning on $H$ is only applicable if Bel($H^c$) = 0. So in order to be able to condition in the necessary sense on $E_n$ and $E_n^c$, we must require that
\[ \text{Bel}(E_n) = \text{Bel}(E_n^c) = 0 \]  
for all $n$. Applying (B2) with $N = 2$, $A_1 = E_1 \cup E_2$ and $A_2 = E_1 \cup E_2^c$, and using that $\text{Bel}_{E_{n+1}^c, \text{nec}}(E_n) = \text{Bel}(E_1 \cup E_{n+1}^c) = a$ and $\text{Bel}_{E_{n+1}^c, \text{nec}}(E_n^c) = \text{Bel}(E_1 \cup E_{n+1}) = b$ gives
\[ 0 = \text{Bel}(E_1) \geq \text{Bel}(E_1 \cup E_2) + \text{Bel}(E_1 \cup E_2^c) - 1 \]
\[ = a + b - 1. \]

It follows that for necessary conditioning we must require that $a + b \leq 1$.

**Theorem 8.4.4.** (Necessary conditioning) Suppose that $a + b \leq 1$ and that
\[ \text{Bel}_{E_{n+1}^c, \text{nec}}(E_n) = a \quad \text{and} \quad \text{Bel}_{E_{n+1}^c, \text{nec}}(E_n^c) = b \]  
for all $n$. Then we have
\[ 0 \leq \text{Bel}(E_0) \leq \min\{a, b\} \]  
and these bounds are attained.

Next we give bounds for contingent conditioning.

**Theorem 8.4.5.** (Contingent conditioning) Suppose that for all $n$ we have
\[ \text{Bel}_{E_{n+1}}(E_n) = a \quad \text{and} \quad \text{Bel}_{E_{n+1}^c}(E_n) = b. \]  
(a) We have
\[ \text{Bel}(E_0) \geq \min\{a, b\}. \]
Furthermore, if \(1 \neq a \leq b\) then for every \(\epsilon > 0\) there is a Bel satisfying (8.4.13) such that \(\text{Bel}(E_0) \leq \min\{a, b\} + \epsilon\).

(b) Let \(b > 0\) and \(a < 1\). For every \(\epsilon > 0\) there is a Bel satisfying (8.4.13) such that

\[
\text{Bel}(E_0) = \frac{1}{2 - a} - \epsilon. \tag{8.4.15}
\]

Finally, here are bounds for mixed conditioning.

**Theorem 8.4.6.** (Mixed conditioning) Suppose that for all \(n\) we have

\[
\text{Bel}_{E_{n+1}, \text{mix}}(E_n) = a \quad \text{and} \quad \text{Bel}_{E_{n+1}, \text{mix}}(E_n) = b. \tag{8.4.16}
\]

(a) If \(b > 0\) we have

\[
\text{Bel}(E_0) \geq \max \left\{ 0, \frac{a + b - 1}{b} \right\}. \tag{8.4.17}
\]

and this bound is attained.

(b) There is a Bel satisfying (8.4.16) such that

\[
\text{Bel}(E_0) = \max\{a, b\}. \tag{8.4.18}
\]

(c) Let \(a < 1\) or \(b < 1\). Then for all Bel satisfying (8.4.16) we have

\[
\text{Bel}(E_0) \leq \frac{a + b - 2ab}{1 - ab}. \tag{8.4.19}
\]

If \(b \leq 1/2\), this bound is attained.

### 8.4.3 Proofs of the bounds in Section 8.4.2

**Proof of Theorem 8.4.4.** To see that zero is attained, consider Bel given by

\[
\begin{align*}
m((E_0 \cup E_1^c) \cap (E_1 \cup E_2^c) \cap \cdots) &= a, \\
m((E_0 \cup E_1) \cap (E_1 \cup E_2) \cap \cdots) &= b, \\
m(\Omega) &= 1 - a - b. \tag{8.4.20}
\end{align*}
\]

For the upper bound, we observe that

\[
\text{Bel}(E_0) \leq \text{Bel}(E_0 \cup E_1^c) = a \tag{8.4.21}
\]

and

\[
\text{Bel}(E_0) \leq \text{Bel}(E_0 \cup E_1) = b. \tag{8.4.22}
\]

Hence \(\text{Bel}(E_0) \leq \min\{a, b\}\). To see that this bound is attained, first assume that \(a < b\) and consider

\[
\begin{align*}
m(E_0 \cap (E_1 \cup E_2^c) \cap \cdots) &= a, \\
m((E_0 \cup E_1) \cap (E_1 \cup E_2) \cap \cdots) &= b - a, \\
m(\Omega) &= 1 - b. \tag{8.4.23}
\end{align*}
\]
If \( b \leq a \), consider

\[
m((E_0 \cup E_1^c) \cap (E_1 \cup E_2^c) \cap \cdots) = a - b, \\
m(E_0 \cap (E_1 \cup E_2) \cap \cdots) = b, \\
m(\Omega) = 1 - a.
\]  

(8.4.24)

\[\Delta := \text{Bel}(E_0) - \text{Bel}(E_0 \cap E_1) - \text{Bel}(E_0 \cap E_1^c),
\]
\[P_1 := 1 - \text{Bel}(E_0 \cup E_1^c),
\]
\[P_2 := 1 - \text{Bel}(E_0 \cup E_1).
\]  

(8.4.25)

\[m((E_0 \cup E_1) \cap (E_1 \cup E_2) \cap \cdots) = C(1 - a),
\]
\[m(E_0 \cap E_1 \cap E_2 \cap \cdots) = a(1 - Z),
\]
\[m(E_0 \cap E_1^c \cap E_2 \cap E_3^c \cap \cdots) = Z,
\]
\[m(E_0^c \cap E_1 \cap E_2^c \cap E_3 \cap \cdots) = Z.
\]  

We can get \( \text{Bel}(E_0) \) arbitrarily close to \( a \) by choosing \( C \) close enough to 1.

(b) Let \( b > 0 \) and \( a < 1 \). Let \( 0 < \delta < 1 \) and consider Bel given by

\[
m(E_0 \cap E_1^c \cap E_2 \cap \cdots) = \delta,
\]
\[m(E_0^c \cap E_1 \cap E_2^c \cap \cdots) = \delta,
\]
\[m(E_0^c \cap E_1^c \cap E_2 \cap \cdots) = \frac{1 - b}{b} \delta,
\]
\[m(E_0 \cap E_1 \cap E_2 \cap \cdots) = \frac{a}{1 - a} \delta + \frac{a}{1 - a} Z,
\]
\[m(E_0 \cap E_2 \cap E_4 \cap \cdots) = Z,
\]
\[m(E_1 \cap E_3 \cap E_5 \cap \cdots) = Z,
\]
where
\[ Z = \frac{(1 - a)(1 - \frac{1-b}{b}\delta)}{2 - a} - \delta. \] (8.4.30)

We can get Bel\((E_0)\) arbitrarily close to \(\frac{1}{2-a}\) by choosing \(\delta\) close enough to 0.

**Proof of Theorem 8.4.6.** (a) First suppose that \(a + b \geq 1\). We have
\[
\begin{align*}
\text{Bel}(E_n) &\geq \text{Bel}(E_n \cup E_{n+1}) + \text{Bel}(E_n \cup E_n^c) - 1 \\
&= b + (1 - b)\text{Bel}(E_{n+1}) + a + (1 - a)\text{Bel}(E_{n+1}^c) - 1 \\
&\geq a + b - 1 + (1 - b)\text{Bel}(E_{n+1}).
\end{align*}
\] (8.4.31)

This gives an inequality between Bel\((E_n)\) and Bel\((E_{n+1})\) and re-iterating this inequality gives
\[
\text{Bel}(E_0) \geq \frac{a + b - 1}{b}. \] (8.4.32)

To see that this bound is attained, consider Bel\((n)\) given by
\[
\begin{align*}
m(E_0 \cap E_1 \cap E_2 \cap \cdots) &= \frac{a + b - 1}{b}, \\
m((E_0 \cup E_1) \cap (E_1 \cup E_2^c) \cap \cdots) &= a - \frac{a + b - 1}{b}, \quad (8.4.33) \\
m((E_0 \cup E_1) \cap (E_1 \cup E_2) \cap \cdots \cap (E_{n-1} \cup E_n)) &= 1 - a.
\end{align*}
\]

Now suppose that \(a + b < 1\). In this case, to see that zero is attained as a belief for \(E_0\), consider Bel given by
\[
\begin{align*}
m((E_0 \cup E_1^c) \cap (E_1 \cup E_2^c) \cap \cdots) &= a, \\
m((E_0 \cup E_1) \cap (E_1 \cup E_2) \cap \cdots) &= b, \quad (8.4.34) \\
m(\Omega) &= 1 - a - b.
\end{align*}
\]

(b) If \(a \geq b\), consider Bel given by
\[
\begin{align*}
m((E_0 \cap E_1 \cap E_2 \cap \cdots) &= a, \\
m((E_0 \cup E_1) \cap (E_1 \cup E_2) \cap \cdots) &= b(1 - a), \quad (8.4.35) \\
m(\Omega) &= 1 - a - b(1 - a).
\end{align*}
\]

If \(b \geq a\), consider Bel given by
\[
\begin{align*}
m((E_0 \cap E_1 \cap E_2 \cap \cdots) &= b, \\
m((E_0 \cup E_1^c) \cap (E_1 \cup E_2^c) \cap \cdots) &= a(1 - b), \quad (8.4.36) \\
m(\Omega) &= 1 - b - a(1 - b).
\end{align*}
\]

(c) We have
\[
\begin{align*}
\text{Bel}(E_0) &\leq \text{Bel}(E_0 \cup E_1) - \text{Bel}(E_1) + \text{Bel}(E_0 \cap E_1) \\
&= b - b\text{Bel}(E_1) + \text{Bel}(E_0 \cap E_1) \\
&\leq b + (1 - b)\text{Bel}(E_0 \cap E_1).
\end{align*}
\] (8.4.37)
In the same way we find
\[ \text{Bel}(E_0) \leq a + (1 - a)\text{Bel}(E_0 \cap E_0^c). \] (8.4.38)

If \( a = 0, b = 0 \) or \( \text{Bel}(E_0) = 0 \), then the inequality is clearly satisfied. So now suppose that \( a, b, \text{Bel}(E_0) > 0 \) and set
\[ u := \frac{\text{Bel}(E_0 \cap E_1)}{\text{Bel}(E_0)} \in [0, 1]. \] (8.4.39)

Then
\[ \text{Bel}(E_0) \leq b + (1 - b)u\text{Bel}(E_0) \] (8.4.40)
and
\[ \text{Bel}(E_0) \leq a + (1 - a)(1 - u)\text{Bel}(E_0). \] (8.4.41)

So
\[
\text{Bel}(E_0) \leq \min \left\{ \frac{b}{1 - (1 - b)u}, \frac{a}{1 - (1 - a)(1 - u)} \right\}
\leq \sup_{x \in [0, 1]} \min \left\{ \frac{b}{1 - (1 - b)x}, \frac{a}{1 - (1 - a)(1 - x)} \right\}
= \frac{a + b - 2ab}{1 - ab}. \] (8.4.42)

To show that this bound is attained, consider \( \text{Bel} \) given by
\[
m(E_0 \cap E_1 \cap E_2 \cap \cdots) = \frac{a(1 - b)}{1 - ab},
\]
\[
m(E_0^c \cap E_1 \cap E_2^c \cap E_3 \cap \cdots) = \frac{b(1 - a)}{1 - ab},
\]
\[
m((E_1 \cup E_2) \cap (E_2 \cup E_3) \cap \cdots) = \frac{b(1 - a)}{1 - ab},
\]
\[
m(\Omega) = \frac{(1 - a)(1 - 2b)}{1 - ab}. \] (8.4.43)
In the first part of this thesis, uniform probability measures on countable sets are discussed. These probability measures are finitely additive, but not countably additive. In Chapter 2, a notion of uniformity is introduced that is called ‘weak thinnability’, which is a generalization of the property of uniform probability measures on $[0, 1]$ or on a finite spaces that if we condition on any suitable subset, the resulting conditional probability measure is again uniform on that subset. It is showed that in the class of probability measures satisfying this notion, the probability of some subsets of natural numbers that do not have a natural density, is uniquely determined.

In Chapter 3, we mathematically model the supertask, introduced by Hansen [26, 27], in which an infinity of gods together select a random natural number by each randomly removing a finite number of balls from an urn, leaving one final ball. We showed that this supertask is highly underdetermined, i.e. there are many scenarios consistent with the supertask. In particular we show that the notion of uniformity for finitely additive probability measures on the natural numbers emerging from this supertask is unreasonably weak.

The second part of this thesis is dedicated to epistemic probability and more specifically to the theory of belief functions. In Chapter 5, the calculus of belief functions is discussed, based on a frequentist interpretation of belief functions. With respect to conditioning, instead following the distinction of Dubois and Prade [14] between ‘focusing’ and ‘updating’, a distinction between learning that something is contingently true (or has occurred contingently) and learning that something is necessarily true (or has necessarily occurred). Contingent conditioning coincides with what Dubois and Prade call ‘focusing’ and corresponds to a rule of conditioning that was already described by Dempster [9]. The rule that corresponds to necessary conditioning can be seen as the special case of Dempster-Shafer conditioning in which we have $\text{Bel}(H^c) = 0$, if $H$ is the set on which we condition. Having introduced these two notions of conditioning, it is explained how Dempster-Shafer conditioning, the rule of conditioning that corresponds to what Dubois and Prade call ‘updating’, as a ‘mix’ between contingent and necessary conditioning. After interpreting both Dempster-Shafer conditioning and strong independence, it is argued that there is no justification for Dempster’s rule of combination within the interpretation of Chapter 5.

In Chapter 6, the proposed calculus of Chapter 5 is justified from a betting
interpretation of belief functions. The first step is the justification of belief functions themselves from a betting interpretation, in an argument analogous to the Dutch Book argument. The theory of Peter Walley [61], which already provides betting schemes with a distinction between buying and selling prices, serves as a natural starting point here. Then it is argued that, on top of the constraint of coherence that Walley gives, reasonable agents should adhere to an additional constraint.

In Chapter 7, the theory developed in Chapter 5 is applied to forensic problems. This is done for two variants of the classical island problem and a case of parental identification. In these cases, a crucial problem of the classical approach is that one needs to assign prior probability to the criminal (in case of the island problem) and father (in case of parental identification). A uniform distribution is what is typically used, this does not really represent ignorance. With the theory of belief functions, we are able to take an actual uninformative prior. It is showed that doing so, leads to different answers than the classical answers and we compare these answers.

In Chapter 8, the theory is used to study an infinite regress problem. In this problem, we have an infinite sequence $E_0, E_1, E_2, \ldots$ of propositions. The central question is: if we know the conditional beliefs of $E_n$ given $E_{n+1}$ of some agent for every $n$, what can we infer about his or her (unconditional) beliefs about the individual propositions, and proposition $E_0$ in particular. For the analysis this problem, some basic theory of belief functions on infinite spaces is developed in this chapter. It is showed that, contrary to certain claims in the literature, this regress problem does in general not have a unique solution.


