Chapter 2
Mathematics for Reading Later Chapters

2.1 Introduction and Summary

This chapter covers a number of mathematical concepts that are used in the following chapters. The basic concepts that play important roles in this book are the *probability capacity*, which is a set function defined on the state space that is not necessarily additive over a disjoint family of subsets, the integral with respect to a probability capacity, called the *Choquet integral*, and an extension of a probability capacity to a Markovian stochastic environment, which we call the *capacitary kernel*. We define these concepts carefully and then present some of their important properties, which we repeatedly use in this book.

Some results are well known and their proofs are easily available in the literature unless otherwise stated. Regularly cited works include those of Dellacherie (1970), Shapley (1971), and Schmeidler (1986).

On the other hand, some results first appeared in the authors' own works. If such cases, we provide the proofs as fully as possible and the readers are referred to the Appendix.

2.2 Probability Charges and Probability Measures

2.2.1 Algebra, σ-Algebra, and Measurable Spaces

Let S be a set. We call a family of subsets $\mathcal{A}$ of a set S an *algebra* if it satisfies the three conditions: (1) $\phi \in \mathcal{A}$, (2)[1] $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, and (3) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$. Furthermore, consider the following condition that strengthens (3): (4) $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{A}$. Condition (4) is clearly strengthening of (3), because we can simply set $A_3 = A_4 = \cdots = \phi$ in (4). We call a family of subsets $\mathcal{A}$ of a set S a

[1]Here, A^c denotes the complement of A in S.

K.G. Nishimura and H. Ozaki, *Economics of Pessimism and Optimism*,
DOI 10.1007/978-4-431-55903-0_2

σ-algebra if it satisfies (1), (2), and (4). An element of an algebra and a σ-algebra is called an *event*. A pair of a set and an algebra or a σ-algebra defined on that set, $(S, \mathcal{A})$, is a *measurable space*. While it is often the case that whether we are talking about an algebra or a σ-algebra is clear from the context, we will be as specific as possible. The family of subsets of a given set S consisting of all its subsets is called the *power set* and is denoted by 2^S. Clearly, 2^S is a σ-algebra and $(S, 2^S)$ is a measurable space.

2.2.2 Probability Charge and Finite Additivity

Given a measurable space $(S, \mathcal{A})$, a set function $p : \mathcal{A} \to [0, +\infty]$ that satisfies the following two conditions is called a *finitely additive measure* or a *charge*:

$$p(\phi) = 0 \text{ and} \tag{2.1}$$

$$(\forall A, B \in \mathcal{A}) \quad A \cap B = \phi \;\Rightarrow\; p(A \cup B) = p(A) + p(B)\,. \tag{2.2}$$

Condition (2.2) is called a *finite additivity*. It immediately follows that a charge p is *monotonic* in the sense that

$$(\forall A, B \in \mathcal{A}) \quad A \subseteq B \;\Rightarrow\; p(A) \le p(B)\,. \tag{2.3}$$

To see this, note that for any A and B such that $A \subseteq B$, it holds that $B = A \cup (B \backslash A)$ and $A \cap (B \backslash A) = \phi$. This implies that $p(B) = p(A) + p(B \backslash A) \ge p(A)$, where the equality holds by the finite additivity of p and the inequality holds by the fact that p takes on only non-negative values.

A charge p that satisfies $p(S) < +\infty$ is a *finite charge* and a finite charge that satisfies $p(S) = 1$ is a *finitely additive probability measure* or a *probability charge*. We denote the set of all probability charges on a measurable space $(S, \mathcal{A})$ by $\mathcal{M}(S, \mathcal{A})$.

A probability charge p is *simple* if the set given by $\{\, s \in S \mid p(\{s\}) \ne 0 \,\}$ is a finite set. In particular, we write the simple probability charge such that $p(\{s\}) = 1$ for some $s \in S$ as δ_s. That is, δ_s is a point mass concentrated at s.

A probability charge p on a measurable space $(S, \mathcal{A})$ is said to be *convex-ranged* or *strongly nonatomic* if it satisfies the condition: $(\forall A \in \mathcal{A})(\forall r \in [0, p(A)])(\exists B \in \mathcal{A})\ B \subseteq A$ and $p(B) = r$. Roughly speaking, the convex-rangedness or strong nonatomicity requires that the measurable space together with the given charge should have a "rich" structure. A closely related concept is *nonatomicity*. A probability charge p on a measurable space $(S, \mathcal{A})$ is said to be *nonatomic* if it satisfies $(\forall A \in \mathcal{A})\ [\, p(A) > 0 \;\Rightarrow\; (\exists B \in \mathcal{A})\ B \subseteq A$ and $p(B) \in (0, p(A))\,]$. As the names suggest, if a probability charge is convex-ranged or strongly nonatomic, then it is nonatomic. However, the converse does not hold in general.[2]

[2] See, however, Proposition 2.2.1.

2.2.3 Dunford-Schwartz Integral with Respect to Charge

This subsection briefly explains the *Dunford-Schwartz integral* with respect to a probability charge. If a charge happens to be a measure, which will be defined in the next subsection, it coincides with the well known Lebesgue integral.

Let S be a set and let $\mathcal{A}$ be an algebra on it. We denote $B(S, \mathcal{A})$, or more simply B, as the set of all $\mathcal{A}$-measurable and bounded real-valued functions defined on a measurable space $(S, \mathcal{A})$. Here, a function $a : S \to \mathbb{R}$ is $\mathcal{A}$-*measurable* if for any Borel set E on $\mathbb{R}$, $a^{-1}(E) := \{ s \in S \mid a(s) \in E \} \in \mathcal{A}$. We denote by $B_0(S, \mathcal{A})$ or B_0 the subset of $B(S, \mathcal{A})$ consisting of functions, called *simple functions*, whose ranges are finite sets.

Given a probability charge on $(S, \mathcal{A})$ and $a \in B(S, \mathcal{A})$, the Dunford-Schwartz integral of a with respect to p is denoted by

$$\int_S a(s)\, dp(s)\,.$$

The Dunford-Schwartz integral is a functional defined on B or B_0. Instead of defining it formally, we choose to characterize it by some axioms. For its definition, see Dunford and Schwartz (1988) and Rao and Rao (1983).

To this end, we introduce some definitions about a functional. Let $I : B \to \mathbb{R}$ be a functional on a measurable space $B(S, \mathcal{A})$. It is *homogeneous* if for any $x \in \mathbb{R}$ and for any $a \in B$, $I(xa) = xI(a)$, and it is *additive* if $(\forall a, b \in B)\ I(a + b) = I(a) + I(b)$. A functional I is a *linear functional* if it is both homogeneous and additive. Also, a functional I is *monotonic* if $(\forall a, b \in B)\ a \geq b \;\Rightarrow\; I(a) \geq I(b)$. Finally, a functional I is *norm-continuous* if for any sequence $\langle a_n \rangle_{n=1}^{\infty} \subseteq B$ and for any element $a \in B$, $\|a - a_n\| \to 0 \;\Rightarrow\; |I(a) - I(a_n)| \to 0$, where $\|\cdot\|$ is the sup norm on B.

The Dunford-Schwartz integral is linear and norm-continuous. One of the most important results of the Dunford-Schwartz integral is that the converse holds.

Theorem 2.2.1 (Riesz Representation Theorem) *For any linear functional $I : B \to \mathbb{R}$ that is norm-continuous and satisfies $I(\chi_S) = 1$, it holds that*

$$(\forall a \in B) \quad I(a) = \int_S a(s)\, dp(s)\,. \tag{2.4}$$

Here, p is the probability charge on $(S, \mathcal{A})$ defined by $(\forall A \in \mathcal{A})\ p(A) = I(\chi_A)$.

In the theorem, χ denotes the *indicator function*. That is, for any $A \in \mathcal{A}$, χ_A is the measurable function on $(S, \mathcal{A})$ such that $\chi_A(s) = 1$ if $s \in A$ and $\chi_A(s) = 0$ if $s \notin A$. For the proof of the Riesz Representation Theorem, see Rao and Rao (1983, p.135, Theorem 4.7.4).

It is often not easy to verify the norm-continuity of a given functional. In that case, the next corollary is convenient. Let K be a convex set that satisfies $[-1, 1] \subseteq K \subseteq \mathbb{R}$ and denote the subset of B (resp. B_0) consisting of all the K-valued functions by $B(K)$ (resp. $B_0(K)$).

Corollary 2.2.1 *Let $I : B(K) \to \mathbb{R}$ be a functional. If I is additive, monotonic, and satisfying $I(\chi_S) = 1$, then (2.4) holds with B replaced by $B(K)$.*

2.2.4 Probability Measure and σ-Additivity

Let $\mathcal{A}$ be a σ-algebra and let p be a charge on $(S, \mathcal{A})$. A charge p that satisfies the following condition is called a *σ-additive measure* or simply a *measure*: $(\forall i, j \text{ such that } i \neq j)\ A_i \cap A_j = \phi$

$$\Rightarrow \quad p(A_1 \cup A_2 \cup \ldots) = p(A_1) + p(A_2) + \cdots . \tag{2.5}$$

Note that $\mathcal{A}$ needs to be a σ-algebra. Otherwise, the second line of (2.5) is not well defined. The condition (2.5) is called *countable additivity* or *σ-additivity*. Clearly, σ-additivity implies finite additivity because we may set A_i to be ϕ except for finite i values and because $p(\phi) = 0$ by (2.1).

A measure p on a measurable space $(S, \mathcal{A})$ that satisfies $p(S) = 1$ is called a *probability measure*. In other words, if a set function $p : \mathcal{A} \to [0, +\infty]$ satisfies (2.1), (2.5) and $p(S) = 1$, then p is called a probability measure. These conditions constitute the so-called *Kolmogorov's Axioms*.

In a very similar manner, we may define the convex-rangedness (or equivalently, the strong nonatomicity) and the nonatomicity for probability measures. Unlike probability charges, however, these two concepts coincide as the next proposition states[3]:

Proposition 2.2.1 *A probability measure is convex-ranged if and only if it is nonatomic.*

Note that a probability measure p is *continuous from below* in the sense that $(\forall \langle A_i \rangle_i \subseteq \mathcal{A})\ A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \Rightarrow p(\cup_i A_i) = \lim_{i\to\infty} p(A_i)$ and that it is *continuous from above* in the sense that $(\forall \langle A_i \rangle_i \subseteq \mathcal{A})\ A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \Rightarrow p(\cap_i A_i) = \lim_{i\to\infty} p(A_i)$. Therefore, a probability measure p is *continuous* in the sense that it is continuous from below and above. Note that when $\mathcal{A}$ is a σ-algebra, if a probability charge defined on $(S, \mathcal{A})$ is continuous, then it turns out to be a probability measure (that is, it is σ-additive).

---

[3] For the proof, see Rao and Rao (1983).

2.3 Probability Capacity

2.3.1 Basic Definitions

Let $(S, \mathcal{A})$ be a measurable space, where $\mathcal{A}$ may only be an algebra (instead of a σ-algebra). A set function $\theta : \mathcal{A} \to [0, +\infty]$ is a *nonadditive measure* or a *capacity* by definition if it only satisfies both (2.1) and the monotonicity condition, (2.3). As we have already seen, a charge is monotonic. However, the monotonicity never implies finite additivity.

A capacity θ such that $\theta(S) < +\infty$ is called a *finite capacity* or a *game*. A finite capacity that satisfies $\theta(S) = 1$ is a *probability capacity* by definition.[4]

A probability capacity θ is said to be *convex* if it holds that

$$(\forall A, B \in \mathcal{A}) \quad \theta(A \cup B) + \theta(A \cap B) \ \geq \ \theta(A) + \theta(B)\,. \tag{2.6}$$

If the converse inequality always holds in (2.6), θ is said to be *concave*. Note that if the inequality always holds with an equality, θ turns out to be a probability charge.

Given a probability capacity θ, we can define its *conjugate*, denoted by θ', by $(\forall A \in \mathcal{A}) \ \ \theta'(A) := 1 - \theta(A^c)$. It can be easily verified that if a probability capacity is convex, then its conjugate is concave, and vice versa. Because the convexity implies that $1 = \theta(S) \geq \theta(A) + \theta(A^c)$, $\theta(A) \leq 1 - \theta(A^c) = \theta'(A)$ holds if θ is convex.

Similar to charges and measures, a probability capacity θ is said to be *convex-ranged* if the next condition holds: $(\forall A \in \mathcal{A})(\forall r \in [0, \theta(A)])(\exists B \in \mathcal{A}) \ B \subseteq A$ and $\theta(B) = r$.

When $\mathcal{A}$ is a σ-algebra, a probability capacity θ is *continuous from below* if

$$(\forall \langle A_i \rangle_i \subseteq \mathcal{A}) \quad A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \ \Rightarrow \ \theta(\cup_i A_i) = \lim_{i \to \infty} \theta(A_i) \tag{2.7}$$

and it is *continuous from above* if

$$(\forall \langle A_i \rangle_i \subseteq \mathcal{A}) \quad A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \ \Rightarrow \ \theta(\cap_i A_i) = \lim_{i \to \infty} \theta(A_i)\,. \tag{2.8}$$

A probability capacity θ is *continuous* if it is continuous both from below and above. Note that any finite *measure* is continuous, and that continuity and finite additivity together imply countable additivity.

Remark 2.3.1 Assume that S is a topological space. Then, let $\mathcal{A}$ be the Borel σ-algebra on S, that is, the smallest σ-algebra that contains all open sets on S. Sometimes, a set function $\theta : \mathcal{A} \to [0, +\infty]$ is *defined* as a capacity if it satisfies (2.1), (2.3) as well as (2.8) and if (2.7) also holds when A_i is restricted to be open

[4] In the definitions that follow, unity can be replaced by $\theta(S)$ for a finite capacity, while we exclusively consider probability capacities in this book.

(see, for example, Huber and Strassen, 1973). However, in this book, we stick to (2.1) and (2.3) as defining properties of the capacity.

2.3.2 Decomposability

Given a probability capacity θ on $(S, \mathcal{A})$ and a nondecreasing function $g : [0, 1] \to [0, 1]$ such that $g(0) = 0$ and $g(1) = 1$, we define a mapping $g \circ \theta : \mathcal{A} \to [0, 1]$ by $(\forall A \in \mathcal{A})\ g \circ \theta(A) = g(\theta(A))$. Then $g \circ \theta$ is a probability capacity. Furthermore, $g \circ \theta$ is convex (resp. concave, continuous) when both g and θ are convex (resp. concave, continuous). (See A.1.2 in the Appendix.)

Given a probability charge p on $(S, \mathcal{A})$ and a nondecreasing function g on $[0, 1]$ that satisfies $g(0) = 0$ and $g(1) = 1$, we define a mapping $\theta = g \circ p : \mathcal{A} \to [0, 1]$ by

$$(\forall A \in \mathcal{A}) \quad \theta(A) = g \circ p(A) = g(p(A)) . \tag{2.9}$$

Then, θ is clearly a probability capacity. However, it is not true that any probability capacity can be decomposed in this way as the next example by Chateauneuf (1991, Example 4, p.364)

Example 2.3.1 Let $S := \{1, 2, 3, 4\}$ and define a mapping $m : 2^S \to [0, 1]$ by $m(\{1\}) = m(\{3\}) = \frac{1}{5}$; $m(\{2\}) = m(\{4\}) = m(\{2, 4\}) = \frac{1}{6}$; $m(S) = \frac{1}{10}$ and for any other $A \subseteq S$, $m(A) = 0$. Furthermore, if we define a mapping $\theta : 2^S \to [0, 1]$ by: $(\forall A)\ \theta(A) = \sum_{B \subseteq A} m(B)$, then it can be easily verified that θ thus defined is a convex probability capacity. Now suppose that (2.9) holds for some probability charge p and for some nondecreasing function g. Then, it holds that $\theta(\{1\}) = \theta(\{3\}) = \frac{1}{5} > \frac{1}{6} = \theta(\{2\}) = \theta(\{4\})$ and that $p(\{1\}) > p(\{2\})$ and $p(\{3\}) > p(\{4\})$ because g is nondecreasing, and hence, it follows that $p(\{1, 3\}) > p(\{2, 4\})$. However, because $\theta(\{1, 3\}) = \frac{2}{5} < \frac{1}{2} = \theta(\{2, 4\})$, we obtain a contradiction. ■

Some conditions for a probability capacity to be decomposable as in (2.9) are known. A probability capacity θ is *weakly additive* if it satisfies the next condition: $(\forall A, B, E, F \subseteq S)\ [E \subseteq A \cap B, F \subseteq (A \cup B)^c, \theta((A \backslash E) \cup F) > \theta((B \backslash E) \cup F) \Rightarrow \theta(A) > \theta(B)]$. Then the next theorem holds.

Theorem 2.3.1 (Scott) *Let $\mathcal{A}$ be 2^S. Then, for any convex-ranged probability capacity θ on $(S, \mathcal{A})$, it is weakly additive if and only if there exists a unique strictly increasing function $g : [0, 1] \to [0, 1]$ and a unique convex-ranged probability charge p such that $\theta = g \circ p$.*

Note that the function g in the theorem must be *strictly* increasing. Gilboa (1985) proved that it can be only nondecreasing if the weak additivity is replaced by "almost weak additivity" and "infinite decomposability." See Gilboa (1985) for these axioms and for more details.

2.3.3 The ε-Contamination

A very important example of a probability capacity is the *ε-contamination*, which is used repeatedly in this book. The ε-contamination is a sort of distortion of a given probability charge p obtained by "contaminating" p in the degree of ε. We present three versions of ε-contamination in this subsection.

Example 2.3.2 (**The ε-Contamination**) Let $p \in \mathcal{M}(S, \mathcal{A})$, let $\varepsilon \in [0, 1]$ and let θ be defined by

$$(\forall A \in \mathcal{A}) \quad \theta(A) = \begin{cases} (1-\varepsilon)p(A) & \text{if } A \neq S \\ 1 & \text{if } A = S\,. \end{cases}$$

The probability capacity thus defined is called the *ε-contamination of p*. It can be easily verified that θ is a convex probability capacity. Note that the ε-contamination is *not* in general[5] continuous from below even if the charge p is continuous. To see this, consider an increasing sequence of measurable sets such that each component is not equal to the whole state space and the limit (the union) is equal to it. ■

Example 2.3.3 (**The Naïve ε-Contamination**) Let $p \in \mathcal{M}(S, \mathcal{A})$, let $\varepsilon \in [0, 1]$ and let $\hat{\theta}$ be defined by

$$(\forall A \in \mathcal{A}) \quad \hat{\theta}(A) = \begin{cases} (1-\varepsilon)p(A) & \text{if } p(A) \neq 1 \\ 1 & \text{if } p(A) = 1\,. \end{cases}$$

The probability capacity thus defined is called the *naïve ε-contamination of p*. The naïve ε-contamination of p can be decomposed as $g \circ p$ where $g : [0, 1] \to [0, 1]$ is defined by $g(x) = (1-\varepsilon)x$ if $x < 1$ and $g(1) = 1$. By the observation made in the previous subsection, the naïve ε-contamination is convex because g is convex. However, the naïve ε-contamination is *not* in general continuous from below even if p is continuous because the mapping g defined above is not continuous. ■

While the difference between the naïve and ordinary ε-contamination is subtle, they *are* in fact different. To see this, let p be a probability charge on $(S, \mathcal{A})$. Note that the two concepts of the ε-contamination are distinct if there exists a set A such that $A \neq S$ and $p(A) = 1$ because $\hat{\theta}(A) = 1$ and $\theta(A) = 1 - \varepsilon$. However, such a situation is common. For instance, let $S = [0, 1]$, let $\mathcal{A}$ be the family of Lebesgue measurable subsets of $[0, 1]$, let p be the Lebesgue measure, and let $A = (0, 1)$.

Example 2.3.4 (**The δ-Approximation of the ε-Contamination**) Let $p \in \mathcal{M}(S, \mathcal{A})$, let $\varepsilon, \delta \in (0, 1]$ and let θ_δ be defined by

$$(\forall A \in \mathcal{A}) \quad \theta_\delta(A) = \begin{cases} (1-\varepsilon)p(A) & \text{if } p(A) \leq 1-\delta \\ (1-\varepsilon)p(A) + \varepsilon\left(\dfrac{p(A)-1}{\delta} + 1\right) & \text{if } p(A) > 1-\delta\,. \end{cases}$$

[5] We say "in general" because the correctness of this statement hinges upon the information structure as assumed.

The probability capacity thus defined is called the *δ-approximation of the ε-contamination of* p. It can be decomposed as $g \circ p$ where $g : [0, 1] \to [0, 1]$ is some convex function that can be easily figured out (see Sect. 9.3.2), and hence, it is a convex probability capacity. Importantly, the δ-approximation of the ε-contamination *is* continuous when p is continuous because the mapping g thus figured out *is* continuous. ■

The δ-approximation of the ε-contamination was introduced by Nishimura and Ozaki (2004). We discuss this probability capacity in more detail in Chap. 9.

2.3.4 The Core

We denote by core(θ) the *core* of a probability capacity θ and define it by

$$\text{core}(\theta) := \{\, p \in \mathcal{M}(S, \mathcal{A}) \mid (\forall A \in \mathcal{A}) \ \ p(A) \geq \theta(A) \,\} .$$

The inequalities must hold for all events, and hence, writing $p(A) \geq \theta(A)$ is equivalent to writing $\theta'(A) \geq p(A) \geq \theta(A)$. If a probability capacity θ turns out to be a probability charge, core(θ) consists only of θ itself. Furthermore, it can be shown that if $\mathcal{A}$ is a σ-algebra and if θ is continuous, any element of core(θ) is countably additive; that is, a probability measure.

Any element of the core of θ can be thought of as an allocating scheme that cannot be blocked by any coalition in the cooperative game characterized by θ.[6] It is well known that the core of a "convex game" is nonempty. (See Shapley 1971.)

Proposition 2.3.1 *When θ is convex,* core (θ) is nonempty.

Conveniently, the core of some probability capacity can be calculated explicitly.

Example 2.3.5 (**The ε-Contamination**) Let $p \in \mathcal{M}(S, \mathcal{A})$, let $\varepsilon \in [0, 1]$, and let θ be the ε-contamination of p. Then, the core of θ is given by the following simple form:

$$\text{core}\ (\theta) = \{\ (1 - \varepsilon)p + \varepsilon q \mid q \in \mathcal{M}(S, \mathcal{A})\ \} . \tag{2.10}$$

The set itself of probability charges defined by the right-hand side of (2.10) is often called ε-contamination of p. We sometimes denote it simply by $\{p\}^{\varepsilon}$. See Chaps. 12 and 14. ■

Here, we remark that (2.10) is the core of the ε-contamination of p, *not* that of the *naïve* ε-contamination of p. Actually, in general, the former is a *proper* superset of the latter. To see this, let $S = \{1, 2\}$, let p be the probability charge such that $p(\{1\}) = 0$ and $p(\{2\}) = 1$, and let θ and $\hat{\theta}$ be the ordinary and naïve ε-contamination of p,

[6] In this context, the requirement that a capacity (and a charge) of the whole space be unity is a mere normalization. See the discussion in Sect. 1.1.4.

respectively. Then, $\text{core}(\theta) = \{(x, 1-x) \mid x \in [0, \varepsilon]\}$ and $\text{core}(\hat{\theta}) = \{(0, 1)\}$, the former of which equals (2.10).

For the core of the δ-approximation of the ε-contamination, see Chap. 10.

2.3.5 Updating Probability Capacity

To conclude this section, we introduce the concept of *updating*. Let θ be a probability capacity on $(S, \mathcal{A})$, and let $A, B \in \mathcal{A}$. By writing $\theta_B(A)$, we mean the probability capacity of A when we know that the event B has already occurred. We always require that given $B \in \mathcal{A}$, $\theta_B(\cdot) : \mathcal{A} \to [0, 1]$ should be a probability capacity. If this is the case, $\theta_B(A)$ is a *conditional probability capacity* given B or an *update* of θ given B.

We discuss three updating rules, all of which are identical and coincide with Bayes' rule for a probability charge p; i.e., $p(A \cap B)/p(B)$, if θ is a probability charge (that is, it is additive).

The most simple updating rule for a probability capacity is the *naïve Bayes' rule*, or equivalently, the *generalized Bayesian updating rule*, which is a natural extension of Bayes' rule for probability charges. Thus, it is defined by $(\forall A)\ \theta_B^{\boldsymbol{B}}(A) := \theta(A \cap B)/\theta(B)$ as far as $\theta(B) \neq 0$. It is easy to see that $\theta_B^{\boldsymbol{B}}$ is convex as far as it is well defined if θ is convex.

The next updating rule is what Denneberg (1994) calls the *general updating rule*. This is studied by Denneberg (1994) and the authors cited there. According to this rule, the conditional probability capacity of $A \in \mathcal{A}$ given $B \in \mathcal{A}$, denoted by $\theta_B^{\boldsymbol{G}}(A)$, is defined by

$$\theta_B^{\boldsymbol{G}}(A) := \frac{\theta(A \cap B)}{\theta(A \cap B) + \theta'(A^c \cap B)}.$$

This updating rule is well defined as far as $\theta(B) > 0$ if θ is convex because $\theta(A \cap B) + \theta'(A^c \cap B) = \theta(A \cap B) + \theta(S) - \theta(A \cup B^c) \geq \theta(A \cap B) + [\theta(A \cup B^c) + \theta(B) - \theta(A \cap B)] - \theta(A \cup B^c) = \theta(B) > 0$ where the weak inequality holds by the convexity of θ and the fact that $(A \cup B^c) \cap B = A \cap B$ and $(A \cup B^c) \cup B = S$. Therefore, the general updating rule $\theta_B^{\boldsymbol{G}}$ is well defined if θ is convex and if $\theta(B) > 0$. It can be also shown that $\theta_B^{\boldsymbol{G}}$ is convex as far as $\theta(B) > 0$ if θ is convex (Denneberg 1994, Proposition 2.3 (ii), (iv), and Proposition 2.5). Furthermore, if θ is convex, it follows that

$$(\forall A \in \mathcal{A}) \quad \theta_B^{\boldsymbol{G}}(A) = \min\{\, P(A \cap B)/P(B) \mid P \in \text{core}(\theta)\,\} \tag{2.11}$$

(Denneberg 1994, Theorem 2.4). That is, the general updating rule coincides with the minimum of the updates of the probability charges in the core of θ.

The final updating rule we introduce in this subsection is the *Dempster-Shafer updating rule*. According to this rule, the probability capacity of $A \in \mathcal{A}$ updated given $B \in \mathcal{A}$, denoted by $\theta_B^{\boldsymbol{DS}}(A)$, is defined by

$$\theta_B^{\boldsymbol{DS}}(A) := \frac{\theta((A \cap B) \cup B^c) - \theta(B^c)}{\theta'(B)}.$$

This updating rule is well defined as far as $\theta(B) > 0$ if θ is convex because $\theta'(B) \geq \theta(B)$ when θ is convex. Also, it turns out that $\theta_B^{\boldsymbol{DS}}$ is convex as far as $\theta(B) > 0$ if θ is convex.[7]

Gilboa and Schmeidler found another important expression of the Dempster-Shafer updating rule. They showed that the following holds true: $(\forall A)$

$$\theta_B^{\boldsymbol{DS}}(A) = \min\left\{ \frac{P(A \cap B)}{P(B)} \,\middle|\, P \in \text{core}(\theta) \text{ and } P(B) = \theta'(B) \right\} \tag{2.12}$$

(Gilboa and Schmeidler 1993, Theorem 3.3, and Denneberg 1994, Theorem 3.4). Because $(\forall B)\ \theta'(B) = \max\{P(B) \mid P \in \text{ core } (\theta)\}$ by Corollary 2.4.1, (2.12) shows that the Dempster-Shafer updating rule is equal to the *maximum-likelihood updating rule*: it keeps the priors in the core of θ, which assign the maximum probability charge to the actually realized event and then minimizes the conditional probabilities of these remaining priors. It is obvious from (2.11) and (2.12) that $(\forall A)\ \theta_B^{\boldsymbol{G}}(A) \leq \theta_B^{\boldsymbol{DS}}(A)$, and hence, the general updating rule by Denneberg is more "cautious" than the Dempster-Shafer rule.[8]

We take up these updating rules again when we discuss conditional preferences and their representations by the updated probability capacities in Sect. 3.13, and when we develop an economic model where a learning process takes place under Knightian uncertainty in Chap. 14.

2.4 Choquet Integral

2.4.1 Definition

Recall from Sect. 2.2.3 that we denote by $B(S, \mathcal{A})$, or more simply B, the set of all $\mathcal{A}$-measurable and bounded real-valued functions defined on a measurable space $(S, \mathcal{A})$. We also denote by $B_0(S, \mathcal{A})$, or B_0, the subset of $B(S, \mathcal{A})$ consisting of simple functions.

---

[7] To see this, note that θ is convex $\Leftrightarrow \theta'$ is concave $\Rightarrow (\theta')^{\boldsymbol{B}}$ is concave $\Leftrightarrow \left((\theta')^{\boldsymbol{B}}\right)'$ is convex $\Leftrightarrow ((\theta')')^{\boldsymbol{DS}}$ is convex $\Leftrightarrow \theta^{\boldsymbol{DS}}$ is convex, where the second and fourth arrows are by Denneberg, 1994, Proposition 3.2 (vi) and (iv), respectively.

[8] Another important distinction between these two updating rules exists in the consequence of their iterated applications. Let $B, C \in \mathcal{A}$ be such that $\theta(B \cap C) > 0$. Then, $\left(\theta_B^{\boldsymbol{DS}}\right)_C^{\boldsymbol{DS}} = \theta_{B\cap C}^{\boldsymbol{DS}}$ while $\left(\theta_B^{\boldsymbol{G}}\right)_C^{\boldsymbol{G}} \geq \theta_{B\cap C}^{\boldsymbol{G}}$ (Gilboa and Schmeidler (1993), Theorem 3.3, and Denneberg (1994), Proposition 2.6 and Proposition 3.2 (viii)). The equality above for the Dempster-Shafer updating rule is referred to as the *commutativity* by Gilboa and Schmeidler (1993).

Given a probability capacity θ, we define a (nonlinear) functional $I : B \to \mathbb{R}$ by: $(\forall a \in B)$

$$\begin{aligned} I(a) &= \int a\, d\theta = \int_S a(s)\, d\theta(s) \\ &= \int_{-\infty}^{0} (\theta(a \geq y) - 1)\, dy + \int_0^{+\infty} \theta(a \geq y)\, dy \\ &= \int_{-\infty}^{0} (\theta(\{ s \mid a(s) \geq y \}) - 1)\, dy + \int_0^{+\infty} \theta(\{ s \mid a(s) \geq y \})\, dy\,. \end{aligned} \tag{2.13}$$

Here, the two integrals in the third line are Riemann integrals in a broad sense. To see that these integrals are well defined, first note that the integrands are nonincreasing functions because of the monotonicity of θ. Because a nonincreasing function has at most countably many discontinuous points, it is Riemann integrable. Second, note that a is a bounded function. Therefore, a Riemann integral in a broad sense here is a finite number. The functional I defined by (2.13) is called a *Choquet integral*.

One of immediate consequences of the definition of the Choquet integral is the following fact: $(\forall a \in B)$

$$\begin{aligned} \int (-a)\, d\theta &= \int_{-\infty}^{0} (\theta(-a \geq y) - 1)\, dy + \int_0^{+\infty} \theta(-a \geq y)\, dy \\ &= \int_{-\infty}^{0} (\theta(a < -y) - 1)\, dy + \int_0^{+\infty} \theta(a < -y)\, dy \\ &= \int_{-\infty}^{0} (1 - \theta'(a \geq -y) - 1)\, dy + \int_0^{+\infty} (1 - \theta'(a \geq -y))\, dy \\ &= \int_0^{+\infty} (1 - \theta'(a \geq y) - 1)\, dy + \int_{-\infty}^{0} (1 - \theta'(a \geq y))\, dy \\ &= -\int_{-\infty}^{0} (\theta'(a \geq y) - 1)\, dy - \int_0^{+\infty} \theta'(a \geq y)\, dy \\ &= -\int a\, d\theta'\,, \end{aligned}$$

where the first and last equalities are definitional; the second equality is trivial; the third equality holds by the definition of the conjugate; the fourth equality holds by the change of variable $(-y \to y)$; and the fifth equality holds by exchanging the first and the second terms. The combined equalities thus imply the next proposition.

Proposition 2.4.1 (Choquet Integral by the Conjugate)

$$(\forall a \in B) \qquad \int_S a\, d\theta' = -\int_S (-a)\, d\theta$$

where θ' is the conjugate of θ.

For a function $a \in B$, if we let $\underline{a} := \inf_s a(s)$, then $a - \underline{a} \geq 0$ and it holds that:

$$\int (a - \underline{a})\, d\theta = \int_0^{+\infty} \theta(a - \underline{a} \geq y)\, dy = \int a\, d\theta - \underline{a}\,. \tag{2.14}$$

By (2.14), we may apply only the definition of the Choquet integral for non-negative functions:

$$I(a) = \int_0^{+\infty} \theta(\{\, s \mid a(s) \geq y \,\})\, dy$$

when we calculate the Choquet integrals of bounded functions, which largely simplifies the story.

The expression of the Choquet integral is largely simplified when the integrand is a simple function. Given $a \in B_0$, we denote it by $a = \sum_{i=1}^{k} \alpha_i \chi_{E_i}$. Here, we let $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_k \geq 0$ and $\langle E_i \rangle_{i=1}^{k}$ is a partition of S such that $(\forall i)\ E_i = a^{-1}(\{\alpha_i\})$. Because a is $\mathcal{A}$-measurable, E_i is certainly an event for each i. (Such a partition is called an $\mathcal{A}$-*measurable partition*.) Then, the definition of a Choquet integral (2.13) implies

$$\begin{aligned}\int a\, d\theta = \int a(i)\, \theta(di) \ &= \sum_{i=1}^{k} (\alpha_i - \alpha_{i+1}) \theta \left(\cup_{j=1}^{i} E_j \right) \\ &= \alpha_1 \theta(E_1) + \sum_{i=2}^{k} \alpha_i \left(\theta \left(\cup_{j=1}^{i} E_j \right) - \theta \left(\cup_{j=1}^{i-1} E_j \right) \right)\end{aligned} \tag{2.15}$$

where $\alpha_{k+1} := 0$.

Finally, we define the Choquet integral of a function that is not necessarily bounded in a similar fashion. Let $L(S, \bar{\mathbb{R}})$ be the space of $\mathcal{A}$-measurable functions from S into $\bar{\mathbb{R}}$, where $\bar{\mathbb{R}}$ denotes the set of extended real numbers, $[-\infty, +\infty]$, and define the *Choquet integral* of $u \in L(S, \bar{\mathbb{R}})$ with respect to a capacity θ by (2.13) unless the expression is $(-\infty) + \infty$.

2.4.2 Properties of Choquet Integral

For the remainder of this chapter, θ is a probability capacity on $(S, \mathcal{A})$, where $\mathcal{A}$ is an algebra on S.

The following two results are immediate from the definition of the Choquet integral.

Proposition 2.4.2 (Monotonicity)

$$(\forall a, b \in B) \quad a \geq b \ \Rightarrow \int_S a\, d\theta \geq \int_S b\, d\theta\,.$$

Proposition 2.4.3 (Positive Homogeneity)

$$(\forall a \in B)(\forall \lambda \geq 0) \quad \int_S \lambda a \, d\theta = \lambda \int_S a \, d\theta \, .$$

The next result concerns the convexity and the concavity of a probability capacity.

Proposition 2.4.4 (Super- (Sub-)Additivity) *A probability capacity is convex (resp. concave) if and only if*

$$(\forall a, b \in B) \quad \int (a+b) \, d\theta \geq (\text{ resp. } \leq) \int a \, d\theta + \int b \, d\theta \, .$$

As this result indicates, the Choquet integral is not linear in general. However, there is an important case where it does become linear. For any pair of functions $a, b \in B$, they are said to be *co-monotonic* if it holds that $(\forall s, t \in S)$ $(a(s) - a(t))(b(s) - b(t)) \geq 0$. The co-monotonicity requires that two functions should move in the same direction when the state changes. Intuitively, one function does not work as a "hedge" of the other function.

The co-monotonicity has an important implication for the Choquet integral. It immediately follows that for any pair of functions $b, c \in B_0$, they are co-monotonic if and only if there exist a natural number k, an $\mathcal{A}$-measurable partition $\langle E_i \rangle_{i=1}^k$, and two k-dimensional vectors $(\beta_1, \beta_2, \ldots, \beta_k)$ and $(\gamma_1, \gamma_2, \ldots, \gamma_k)$ such that $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_k$ and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k$ and such that $b = \sum_{i=1}^k \beta_i \chi_{E_i}$ and $c = \sum_{i=1}^k \gamma_i \chi_{E_i}$. It also can be shown that for any function $a \in B$, there exists a sequence of pairs of co-monotonic simple functions, (a_n, b_n), which satisfies that $(\forall n)$ $a_n \leq a_{n+1} \leq \cdots \leq a \leq \cdots \leq b_{n+1} \leq b_n$ and $\lim_{n\to\infty} I(a_n) = I(a) = \lim_{n\to\infty} I(b_n)$. These two facts are used to show the next proposition.

Proposition 2.4.5 (Co-monotonic Additivity) *For any pair of functions $a, b \in B$, if a and b are co-monotonic, then it holds that*

$$\int_S (a+b) \, d\theta = \int_S a \, d\theta + \int_S b \, d\theta \, .$$

The next result is extremely important in the interpretation of the Choquet integral in the framework of economics. Therefore, we name this result the *fundamental theorem of the Choquet integral*.

Theorem 2.4.1 (Fundamental Theorem of Choquet Integral) *A probability capacity θ is convex if and only if*

$$(\forall a \in B) \quad \int_S a \, d\theta = \min \left\{ \int_S a \, dp \,\middle|\, p \in \text{core}(\theta) \right\} .$$

The integral in the right-hand side is the Dunford-Schwartz integral (Sect. 2.2.3). Note that because a is bounded and measurable and the core is compact in the weak * topology, the minimum is actually attained.[9]

Let θ be a convex probability capacity and let $a \in B$. Because the fundamental theorem of the Choquet integral applies to $-a$, we have

$$\begin{aligned} -\int_S (-a)\, d\theta &= -\min\left\{ \int_S (-a)\, dp \,\middle|\, p \in \operatorname{core}(\theta) \right\} \\ &= \max\left\{ \int_S a\, dp \,\middle|\, p \in \operatorname{core}(\theta) \right\}. \end{aligned}$$

Because, however, the first term is equal to the Choquet integral of a with respect to the conjugate of θ by Proposition 2.4.1, we have established the next corollary.

Corollary 2.4.1 *A probability capacity θ is convex if and only if*

$$(\forall a \in B) \quad \int_S a\, d\theta' = \max\left\{ \int_S a\, dp \,\middle|\, p \in \operatorname{core}(\theta) \right\},$$

where θ' is the concave *conjugate probability capacity of θ.*

Given any convex probability capacity θ and any bounded measurable function a, we define a set of probability charges, $\mathcal{P}(\theta, a)$, by

$$\mathcal{P}(\theta, a) := \arg\min\left\{ \int_S a\, dp \,\middle|\, p \in \operatorname{core}(\theta) \right\}. \tag{2.16}$$

This is the set of probability charges on S that are "equivalent" to θ with respect to the Choquet integral of a given the fundamental theorem of the Choquet integral. The same theorem guarantees that $\mathcal{P}(\theta, a)$ is nonempty. Rather, in general, it is not a singleton set unless θ happens to be a probability charge.

---

[9]To be more precise, let $p : \mathcal{A} \to \mathbb{R}$ be a *bounded charge* on $(S, \mathcal{A})$. That is, p satisfies a finite additivity (2.2) and is such that $(\exists M > 0)(\forall A \in \mathcal{A})\ |p(A)| < M$. The set of all bounded charges on $(S, \mathcal{A})$ is denoted by $ba(S, \mathcal{A})$ and it turns out that $ba(S, \mathcal{A}) = B^*(S, \mathcal{A})$, where B^* denotes the *dual* space of B, *i.e.*, the space of linear functionals on B that are continuous with respect to the sup norm topology on B (Dunford and Schwartz, 1988). Note that the linearity does not imply the continuity automatically for infinite-dimensional spaces. Also note that any linear functional on B is specified by the Dunford-Schwartz integral by some bounded charge, which is a version of Riesz Representation Theorem (Theorem 2.2.1). Therefore, B^* can be identified as a set of bounded charges. Furthermore, the unit ball in B^* is *weak * compact* by Banach-Alaoglu's theorem (Dunford and Schwartz, 1988, p.424). Here, the weak * topology, or equivalently, $\sigma(ba, B)$-topology, refers to the *weakest* topology on B^* with respect to which any element of B should be continuous, in which we identify any element of B as a linear functional on B^* in a natural way. As is well known, B^{**}, the dual of $B^* = ba$, is a *proper* superset of B. Hence, the weak * topology is strictly coarser than the weak topology on B^*. Finally, because $\operatorname{core}(\theta)$ is the weak * closed subset of the unit ball in ba, and hence, it is weak * compact (Munkres, 1975, p.165, Theorem 5.2) and because a is assumed to be an element of B, Weierstrass' theorem (Munkres, 1975, p.167, Theorem 5.5) proves the claim in the main text.

Given a pair of bounded and measurable functions on $\mathcal{A}$, (u, v), v is *u-measurable* if v is measurable with respect to the smallest algebra on S that makes u measurable. A function v is *u-ordered* if u and v are co-monotonic and if v is u-measurable. As for the u-measurability and u-orderedness, the following result is of importance.

Theorem 2.4.2 *Let θ be a convex probability capacity on $(S, \mathcal{A})$, let u be a bounded and measurable function on it, and let $P, Q \in \mathcal{P}(\theta, u)$. If v is u-measurable, then*

$$\int_S v\, dP = \int_S v\, dQ\,,$$

and if v is u-ordered, then

$$\int_S v\, dP = \int_S v\, dQ = \int_S v\, d\theta\,.$$

The proof of this theorem can be found in Ozaki (2000).

The next result is easy to prove, but essential in solving dynamic economic models that appear later in this book. (See Sect. 7.1.)

Theorem 2.4.3 *Let θ be a convex capacity. Then,*

$$(\forall u \in B(S, \mathbb{R})) \qquad \left|\int u\, d\theta - \int v\, d\theta\right| \leq \int |u - v|\, d\theta'\,.$$

We have already seen that the Choquet integral is monotonic and co-monotonically additive. Similar to the Riesz Representation Theorem, we now see that these properties are sufficient to characterize the Choquet integral. A functional I is *co-monotonically additive* if $(\forall a, b \in B)$ $I(a + b) = I(a) + I(b)$ whenever a and b are co-monotonic. If I is co-monotonically additive, it is homogeneous with respect to positive rational numbers. To see this, let $r = m/n$ $(m, n \in \mathbb{N})$. Then the co-monotonic additivity implies that $nI((m/n)a) = I(n(m/n)a) = I(ma) = mI(a)$. Also, in a similar manner to the proof of Corollary 2.2.1, if a functional $I : B \to \mathbb{R}$ is co-monotonically additive, monotonic, and satisfies $I(\chi_S) = 1$, then I satisfies the norm-continuity. Therefore, if I is co-monotonically additive, monotonic, and satisfies $I(\chi_S) = 1$, then it is homogeneous with respect to positive rational numbers and it is norm-continuous; hence, it follows that it is positively homogeneous.

The next theorem is a Choquet integral version of the Riesz Representation Theorem.

Theorem 2.4.4 (Schmeidler's Representation Theorem) *Suppose that $I : B \to \mathbb{R}$ is a functional satisfying $I(\chi_S) = 1$. Then I satisfies the co-monotonic additivity and monotonicity if and only if I can be represented by the Choquet integral with respect to the probability capacity θ defined by* $(\forall A \in \mathcal{A})$ $\theta(A) = I(\chi_A)$.

For proofs of this and the following theorems, see Schmeidler (1986). Note that there is a stark contrast between this theorem and Corollary 2.2.1. In this and the next theorems, B and $B(K)$ can be replaced by B_0 and $B_0(K)$, respectively.[10]

Theorem 2.4.5 (Schmeidler 1986) *Suppose that a functional* $I : B(K) \to \mathbb{R}$ *satisfies the following three conditions: (i)* $(\forall\lambda \in K)$ $I(\lambda\chi_S) = \lambda$*; (ii) For any triplet of functions,* (a, b, c)*, any two of which are co-monotonic, if* $I(a) > I(b)$*, then it holds that* $(\forall\alpha \in (0, 1))$ $I(\alpha a + (1-\alpha)c) > I(\alpha b + (1-\alpha)c)$*; and (iii)* $a \geq b \Rightarrow I(a) \geq I(b)$*. Then the function* I *can be represented by the Choquet integral with respect to the probability capacity* θ *defined by* $(\forall A \in \mathcal{A})$ $\theta(A) = I(\chi_A)$.

The remaining results of this section assume that a probability capacity is continuous.

Theorem 2.4.6 (Monotone Convergence Theorem) *Let* $(S, \mathcal{A})$ *be a measurable space where* $\mathcal{A}$ *is a* σ*-algebra and let* θ *be a probability capacity on it. (a) Suppose that* θ *is continuous from below and let* $\langle u_n\rangle_{n=0}^{\infty}$ *be a sequence of* $\mathcal{A}$*-measurable functions such that* $u_0 \leq u_1 \leq u_2 \leq u_3 \leq \cdots$ *and* $\int u_0\, d\theta > -\infty$*. Then,*

$$\lim_{n\to\infty} \int u_n\, d\theta = \int \lim_{n\to\infty} u_n\, d\theta\,.$$

(b) Suppose that θ *is continuous from above and let* $\langle u_n\rangle_{n=0}^{\infty}$ *be a sequence of* $\mathcal{A}$*-measurable functions such that* $u_0 \geq u_1 \geq u_2 \geq u_3 \geq \cdots$ *and* $\int u_0\, d\theta < +\infty$*. Then,*

$$\lim_{n\to\infty} \int u_n\, d\theta = \int \lim_{n\to\infty} u_n\, d\theta\,.$$

Note that by the monotone convergence theorem, all of the above properties of the Choquet integrals hold true for any continuous capacity θ and for any function $u \in L(S, \bar{\mathbb{R}})$ whenever the integral is well defined.

Theorem 2.4.7 (Fatou's Lemma) *Let* θ *be a probability capacity that satisfies: for any decreasing sequence of* $\mathcal{A}$*-measurable subsets of* S*,* $\langle A_n\rangle_{n=1}^{\infty}$*,* $\lim_{n\to\infty}\theta(A_n) \leq \theta\left(\lim_{n\to\infty} A_n\right)$*. Also, let* $\langle u_n\rangle_{n=1}^{\infty}$ *be a sequence of non-negative* $\mathcal{A}$*-measurable functions such that* $(\exists M \in \mathbb{R})(\forall n)$ $u_n \leq M$*. Then,*

$$\overline{\lim_{n\to\infty}} \int u_n(s)\,\theta(ds) \leq \int \overline{\lim_{n\to\infty}}\, u_n(s)\theta(ds)\,.$$

2.5 Capacitary Kernel

A mapping $\theta : S \times \mathcal{A} \to [0, 1]$ is a *capacitary kernel* (from S to S) if it satisfies $(\forall s \in S)$ θ_s is a probability capacity on $(S, \mathcal{A})$, and $(\forall B \in \mathcal{A})$ $\theta_\cdot(B)$ is $\mathcal{A}$-measurable.

---

[10] We only need to assume here that K is convex.

A capacitary kernel is *convex* (resp. *continuous*) if θ_s is convex (resp. continuous) for all s. In particular, if θ_s is a probability measure for all s, θ is called *stochastic kernel* (Stokey and Lucas, 1989, p.226).

As for a capacitary kernel, the next result is used repeatedly.

Theorem 2.5.1 (Fubini Property) *Let θ be a* continuous *capacitary kernel. Then for any $(\mathcal{A} \otimes \mathcal{A})$-measurable function u, the mapping*

$$s \mapsto \int u(s, s_+)\, \theta_s(ds_+) \tag{2.17}$$

is $\mathcal{A}$-measurable.

In this theorem, the continuity of θ cannot be dispensed with, as the following example shows.

Example 2.5.1 Let $(S, \mathcal{A})$ be a measurable space such that any singleton set is included in $\mathcal{A}$, and let a capacitary kernel θ be defined by

$$(\forall A \in \mathcal{A})(\forall s \in S) \quad \theta_s(A) = \begin{cases} 0 & \text{if } A \neq S \\ 1 & \text{if } A = S\,. \end{cases}$$

It is immediately apparent that θ thus defined is certainly a capacitary kernel that is convex. Furthermore, for any $(\mathcal{A} \otimes \mathcal{A})$-measurable function u, it turns out that

$$\int_S u(s, s_+)\, \theta_s(ds_+) = \inf_{s_+ \in S} u(s, s_+)\,. \tag{2.18}$$

To see this, fix $s \in S$. Then, there exists a sequence $\langle s_+^n \rangle_{n=1}^\infty \subseteq S$ such that $u(s, s_+^n) \to \inf_{s_+ \in S} u(s, s_+)$. For each n, we denote by δ^n the point mass concentrated at s_+^n. By the weak *-compactness of the core, there exists a subsequence, $\langle \delta^{n_j} \rangle_{j=1}^\infty$, of $\langle \delta^n \rangle_{n=0}^\infty$ that converges in the weak *-topology to some probability charge p^0 in the core of θ_s because each δ^n is obviously in the core. Thus, p^0 actually attains the infimum in the right-hand side of (2.18) because $u\left(s, s_+^{n_j}\right) \to \inf_{s_+ \in S} u(s, s_+)$ and

$$u\left(s, s_+^{n_j}\right) = \int_S u(s, s_+)\, \delta^{n_j}(ds_+) \to \int_S u(s, s_+)\, p^0(ds_+)$$

by the weak *-convergence of $\langle \delta^{n_j} \rangle$. (Note that p^0 may not be a point mass.)

Note that for any $a \in \mathbb{R}$, $\{s \mid \inf_{s_+ \in S} u(s, s_+) < a\} = \{s \mid (\exists s_+)\, u(s, s_+) < a\}$, the latter of which is the projection of the set $\{(s, s_+) \mid u(s, s_+) < a\}$ onto S. Unfortunately, the projection of a measurable set is not necessarily measurable (for instance, the projection of a Lebesgue measurable set in $\mathbb{R}^2$ onto $\mathbb{R}$ is not necessarily Lebesgue measurable). Therefore, the right-hand side of (2.18) is not necessarily $\mathcal{A}$-measurable as a function of s.

The set that is defined as the projection of the measurable set in the product measurable space is called *analytic* set. For the analytic set, see Dellacherie and Meyer (1988), Bertsekas and Shreve (1978), and Remark 2.5.1 right below. ■

The Fubini property is key for dynamic analyses conducted in the latter half of this book. An extension of the ε-contamination to a capacitary kernel inherits the same difficulty as Example 2.5.1. The δ-approximation of the ε-contamination (Example 2.3.4) is a mechanism to resolve this problem. See Chap. 9.

Remark 2.5.1 Note that the capacitary kernel in Example 2.5.1 *is* continuous from above but is *not* continuous from below. Furthermore, if we assume that S is a topological space and if we define the capacity by way of Remark 2.3.1, the "continuity from below" in the weak sense described there follows because the conjugate of θ_s is "continuous" with respect to a decreasing sequence of closed sets by the finite intersection property of compact sets. (Munkres 1975, p.170, Theorem 5.9) If we assume that θ_s is a capacity in the sense of Remark 2.3.1, the mapping defined by (2.17) is always $\mathcal{A}$-*analytic* instead of $\mathcal{A}$-measurable. (Dellacherie and Meyer 1988) The analyticity is a concept that is weaker than the measurability and Epstein and Wang (1995) use this concept to analyze asset pricing in the presence of Knightian uncertainty (see Chap. 8).

The rest of this chapter assumes that S is a nonempty Borel-measurable subset of a *Polish* space (*i.e.*, a Borel-measurable subset of a topological space that is a homeomorph of a *complete separable metric* space). Also, when S is a topological space, the algebra on it should be always understood to be the *Borel* σ-*algebra*, which is the smallest σ-algebra containing all open sets. We denote it by $\mathcal{B}_S$, and hence, $\mathcal{A} = \mathcal{B}_S$ in what follows.

A capacitary kernel θ is *strongly continuous* if $(\forall \langle s^n \rangle_{n=1}^{\infty} \to s^0)\ \sup_{E \in \mathcal{B}_S} |\theta_{s^n}(E) - \theta_{s^0}(E)| \to 0$.

The concept of upper quasi-continuity was introduced by Ozaki and Streufert (1996) for a more general class of operators that includes as a special case the Choquet integral with respect to a capacitary kernel. A capacitary kernel θ is *upper quasi-continuous beneath* a function $\bar{u} \in L(S, \bar{\mathbb{R}})$ if $(\forall \langle s^n \rangle_{n=1}^{\infty} \to s^0 \in S)(\forall \langle u_n \rangle_{n=1}^{\infty} \leq \bar{u})$

$$\overline{\lim_{n\to\infty}} \int_S u_n(s)\theta_{s^n}(ds) \leq \int_S \overline{\lim_{n\to\infty}}\, u_n(s)\theta_{s^0}(ds)\,.$$

The concept of upper semi-continuity was introduced by Ozaki (2002) for a more general class of operators that includes as a special case the Choquet integral with respect to a capacitary kernel. Assume that X is another Polish space. A capacitary kernel θ is *upper semi-continuous on* $S \times X$ *beneath* a function $\bar{u} \in L(S, \bar{\mathbb{R}})$ if $(\forall \langle (s^n, x^n) \rangle_{n=1}^{\infty} \to (s^0, x^0) \in S \times X)(\forall u \in L(S \times X, \bar{\mathbb{R}}))$

$$\begin{aligned} &u \ \text{ is upper semi-continuous and } \ (\forall n \geq 1)\ \ u(\cdot, x^n) \leq \bar{u} \\ &\qquad \Rightarrow \overline{\lim_{n\to\infty}} \int_S u(s, x^n)\, \theta_{s^n}(ds) \leq \int_S u(s, x^0)\, \theta_{s^0}(ds)\,. \end{aligned}$$

The next result is proved by Ozaki (2002), and it states that the upper quasi-continuity is a stronger assumption than the upper semi-continuity.

Theorem 2.5.2 *Suppose that a capacitary kernel θ is upper quasi-continuous beneath a function $\bar{u}$. Then for any Polish space X, θ is upper semi-continuous on $S \times X$ beneath $\bar{u}$.*

The next theorem provides a sufficient condition for θ to be upper semi-continuous.

Theorem 2.5.3 (Upper Semi-continuity) *Assume that a capacitary kernel θ is strongly continuous and that $(\forall s)$ $\theta_s(\cdot)$ is continuous from above. Then, for any Polish space X, θ is upper semi-continuous on $S \times X$ beneath any constant function.*

For the remainder of this chapter, we exploit the orderedness of the state space by setting $S = Z := [\underline{z}, \bar{z}]$ for some $\underline{z}$ and $\bar{z}$ such that $0 \le \underline{z} \le \bar{z} < +\infty$. We let $\mathcal{A} := \mathcal{B}_Z$, the Borel σ-algebra on Z.

A capacitary kernel θ is *stochastically nondecreasing* if for each nondecreasing function $h : Z \to \mathbb{R}$, the mapping defined by

$$z \mapsto \int_Z h(z')\, \theta_z(dz') \tag{2.19}$$

is nondecreasing. The definition of the stochastic nondecrease here extends that of (Topkis 1998, p.159) for a probability measure to a probability capacity. The concept of the stochastic nondecrease for the capacitary kernel first appeared in Ozaki and Streufert (2001), where the state space is assumed to be a finite set.

Topkis uses (2.20) below for additive θ to define this concept, which turns out to be equivalent to the definition that uses (2.19). This equivalence also holds for nonadditive θ as the next theorem shows.

Theorem 2.5.4 (Stochastic Nondecrease) *A continuous capacitary kernel θ is stochastically nondecreasing if and only if a mapping defined by*

$$z \mapsto \theta_z(\{z' \in Z | z' \ge t\}) \tag{2.20}$$

is nondecreasing for each $t \ge 0$.

Similarly, a capacitary kernel θ is *stochastically convex* if for each nondecreasing function $h : Z \to \mathbb{R}$, the mapping defined by (2.19) is convex. For this concept, we have a counterpart of Theorem 2.5.4.

Theorem 2.5.5 (Stochastic Convexity) *A continuous capacitary kernel θ is stochastically convex if and only if a mapping defined by (2.20) is convex for each $t \ge 0$.*

Some of the assumptions on capacitary kernels introduced so far remain to be satisfied even after some distortion. The next theorem shows this fact.

Theorem 2.5.6 *Assume that $f : [0, 1] \to [0, 1]$ is a convex and continuous function satisfying $f(0) = 0$ and $f(1) = 1$. Also, assume that θ is a convex and continuous capacitary kernel that is stochastically nondecreasing (resp. stochastically convex, upper semi-continuous). Then, a mapping $f \circ \theta : Z \times \mathcal{B}_Z \to [0, 1]$ defined by $(\forall z)(\forall A)\ \ (f \circ \theta)_z\,(A) = f(\theta_z(A))$ is a convex and continuous capacitary kernel that is stochastically nondecreasing (resp. stochastically convex, upper semi-continuous).*

In general, a stochastic kernel need not be stochastically nondecreasing nor upper semi-continuous. The next example provides a stochastic kernel P that is strongly continuous, stochastically nondecreasing, and stochastically convex. Because a stochastic kernel is automatically continuous from above, P in the example is also upper semi-continuous by Theorem 2.5.3.

Example 2.5.2 Let $Z = [0, 1]$ and let P be a stochastic kernel defined by

$$(\forall z, t \in Z) \quad F_z(t) = P_z([0, t]) := \int_0^t (2 - z)\, d\mu\,,$$

where F is the associated (conditional) distribution function and μ is the Lebesgue measure. That is, $P_z(\cdot)$ is the uniform distribution on $[0, 1/(2 - z)]$. Then, P is strongly continuous, stochastically nondecreasing, and stochastically convex. (See A.1.12 in the Appendix.)

Suppose that P is a stochastic kernel that is upper semi-continuous, stochastically nondecreasing, and stochastically convex. The existence of such a stochastic kernel is guaranteed by Example 2.5.2. Also, suppose that $f : [0, 1] \to [0, 1]$ is a convex and continuous function satisfying $f(0) = 0$ and $f(1) = 1$. Then, by Theorem 2.5.6, $\theta := f \circ P$ is convex and a continuous capacitary kernel that is upper semi-continuous and stochastically nondecreasing, as well as stochastically convex, and hence it satisfies all the assumptions for some results appearing in Chap. 11.

2.6 Remarks: Upper Quasi/Semi-Continuity of a Stochastic Kernel

Easily verifiable conditions for the upper quasi-continuity and the upper semi-continuity are still unknown. We only have the following conjecture.

Conjecture 2.6.1 (Upper Quasi-continuity) *A capacitary kernel θ is upper quasi-continuous beneath $\bar{u} \in L(S, \bar{\mathbb{R}})$ if $(\theta, \bar{u})$ satisfies*

M1. $\left(\forall \langle s^n \rangle_{n=1}^{\infty} \to s^0\right) \quad \sup\limits_{E \in \mathcal{B}_S} |\theta_{s^n}(E) - \theta_{s^0}(E)| \to 0$ *and*

M2. $\left(\forall \langle s^n \rangle_{n=1}^{\infty} \to s^0\right) \quad \lim\limits_{b \to +\infty} \overline{\lim_{n \to +\infty}} \int_{\{s' \mid \bar{u}(s') \geq b\}} \bar{u}(s')\, \theta_{s^n}(ds') = 0\,.$

The conjecture holds true if θ is a *stochastic kernel* (see Ozaki and Streufert (1996), p.424, Lemma C.1).

For the upper semi-continuity capacitary kernel, we have

Conjecture 2.6.2 (Upper Semi-continuity) *Let X be a Polish space. Then a capacitary kernel θ is upper semi-continuous on $S \times X$ beneath $\bar{u} \in L(S, \bar{\mathbb{R}})$ if $(\theta, \bar{u})$ satisfies **M2** and*

M1⁻. $\left(\forall \langle s^n \rangle_{n=1}^{\infty} \to s^0\right)$ *θ_{s^n} converges to θ_{s^0}* "weakly" *as $n \to +\infty$.*

We think that the strong continuity of a capacitary kernel assumed in Theorem 2.5.3 that gives sufficient conditions for the upper semi-continuity may be too strong. We think so because the above conjecture holds true if θ is a *stochastic kernel* (see Ozaki 2002, p.30, Theorem 1) where the sense in which θ weakly converges is clear. For a general capacitary kernel, there are several proposals about what a weak convergence is. See, for example, Narukawa et al. (2003).

References

Bertsekas, D.P., and S.E. Shreve. 1978. *Stochastic optimal control: the discrete time case*. Academic Press.

Chateauneuf, A. 1991. On the use of capacities in modeling uncertainty aversion and risk aversion. *Journal of Mathematical Economics* 20: 343–369.

Dellacherie, C. 1970. *Quelques commentaires sur les prolongements de capacités*, Seminaire Probabilités V, Strasbourg, Lecture Notes in Math., vol. 191, Springer-Verlag, Berlin and New York.

Dellacherie, C., and P.A. Meyer. 1988. *Probabilities and potential C*. New York: North-Holland.

Denneberg, D. 1994. Conditioning (updating) non-additive measures. *Annals of Operations Research* 52: 21–42.

Dunford, N., and J.T. Schwartz. 1988. *Linear operator part I: general theory*. Wiley Classics Library.

Epstein, L.G., and T. Wang. 1995. Uncertainty, risk-neutral measures and security price booms and crashes. *Journal of Economic Theory* 67: 40–82.

Gilboa, I. 1985. Subjective distortions of probabilities and non-additive probabilities, Working Paper, The Foerder Institute for Economic Research, Tel Aviv University.

Gilboa, I., and D. Schmeidler. 1993. Updating ambiguous beliefs. *Journal of Economic Theory* 59: 33–49.

Huber, P.J., and V. Strassen. 1973. Minimax tests and the Neyman-Pearson lemma for capacities. *The Annals of Statistics* 1: 251–263.

Munkres, J.R. 1975. *Topology: a first course*. New Jersey: Prentice-Hall.

Narukawa, Y., T. Murofushi, and M. Sugeno. 2003. Space of fuzzy measures and convergence. *Fussy Sets and Systems* 138: 497–506.

Nishimura, K.G., and H. Ozaki. 2004. Search and Knightian uncertainty. *Journal of Economic Theory* 119: 299–333.

Ozaki, H. 2000. Choquet capacity, Knightian uncertainty and indeterminacy of equilibrium prices, Tohoku University, Mimeo.

Ozaki, H. 2002. Dynamic programming with upper semi-continuous stochastic aggregator. *Advances in Mathematical Economics* 4: 25–39.

Ozaki, H., and P.A. Streufert. 1996. Dynamic programming for non-additive stochastic objectives. *Journal of Mathematical Economics* 25: 391–442.

Ozaki, H., and P.A. Streufert. 2001. Solutions for some dynamic problems with uncertainty aversion. *The Japanese Economic Review* 52: 251–283.

Rao, K.P.S.B., and M.B. Rao. 1983. *Theory of charges*. Academic Press.

Schmeidler, D. 1986. Integral representation without additivity. *Proceedings of the American Mathematical Society* 97: 255–261.

Shapley, L.S. 1971. Cores of convex games. *International Journal of Game Theory* 1: 11–26.

Stokey, N.L., and R.E. Lucas Jr. 1989. *Recursive methods in economic dynamics*. Cambridge: Harvard University Press.

Topkis, D.M. 1998. *Supermodularity and complementarity*. Princeton, New Jersey: Princeton University Press.

Springer

http://www.springer.com/978-4-431-55901-6

Economics of Pessimism and Optimism
Theory of Knightian Uncertainty and Its Applications
Nishimura, K.G.; Ozaki, H.
2017, XX, 326 p. 12 illus. in color., Hardcover
ISBN: 978-4-431-55901-6