Uncertainty and Capacities in Finance

Monotone Set Functions and the Choquet Integral

Alexander von Felbert*

Munich, May 2019

Abstract

This review article provides an introduction to monotone set functions and Choquet integrals. A monotone set function, also known as capacity, extends the classical measure theory by restricting to monotonicity instead of requiring (σ-) additivity. The generalization of classical measures is motivated by the uncertainty and human behavior present in finance. Important concepts such as modularity, generalized distribution and survival functions as well as distorted probabilities are explained and put into context. After studying capacities, we introduce Choquet integrals with respect to capacities. Many (new) examples throughout the article have been added to illustrate the theory.

Keywords – uncertainty, human behavior, capacities, monotone set functions, non-additive measures, Choquet integral, Choquet functionals, general distribution function, general survival function

*University of Mannheim, Lehrstuhl fuer Wirtschaftsmathematik I, thought@deep-mind.org

1 Introduction

There are some things that you know to be true, and others that you know to be false; yet, despite this extensive knowledge that you have, there remain many things whose truth or falsity is not known to you. We say that you are uncertain about them. You are uncertain, to varying degrees, about everything in the future; much of the past is hidden from you; and there is a lot of the present about which you do not have full information. Uncertainty is everywhere and you cannot escape from it. – DENNIS LINDLEY. *Understanding Uncertainty (2006).*

Banks, insurance companies and other financial institutions are accustomed to taking certain financial risks and generating profit from it. These risks are (in the best case) quantifiable and can therefore be limited and steered based on objective information. Financial institutions are, however, also exposed to risks, that cannot be reasonable measured or guessed. In such situations only subjective perception remains. Typical examples are so-called non-financial risks such as political, compliance, conduct or legal risk. Apparently, the term 'risk' is loosely used in financial practice since quantifiable and non-quantifiable risks can entail far-reaching and crucial differences. F. H. KNIGHT first distinguished in 1921 between 'risk' and 'uncertainty' in his seminal book [Kni09]. On the one hand, *Knightian* or *complete uncertainty* may be characterized as the complete absence of information or knowledge about a situation or an outcome of an event. *Knightian risk*, on the other hand, may be characterized as a situation, where the true information on the probability distribution is available. In practice, no parameter or distribution can be known for sure a priori. Hence, we find it preferable to think about degrees of uncertainty.[1] Both terms, risk and uncertainty, are obviously not rigorously[2] defined, however, it serves as an orientation, motivation and starting point for the survey below.

Quantifiable risks can be modeled using a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where Ω is the sample space, $\mathcal{A}$ a non-empty collection of all possible events[3], and $\mathbb{P}$ a probability measure. The frequentist's interpretation of the probability $\mathbb{P}(A)$ of an event $A \in \mathcal{A}$ is indicated by the limit[4] of its relative frequency in a large number of trials. For Bayesians, the concept of probability is interpreted as a degree of belief or confidence, representing a state of knowledge or a quantification of a personal belief, about a statement. Both approaches assume additivity. Please also refer to [Sha81], for instance, for more details on the additivity in the Bayesian approach.

Most individuals prefer decision making, where more information is available to decisions with less available information. More information actually means a lower degree of uncertainty, while less information implies a higher degree of uncertainty. The following paradox suggests that the so-called *uncertainty* or *ambiguity aversion* cannot be expressed through an additive model.[5] That is, an additive (probability) measure might not be suitable to represent situations where high uncertainty and human behavior is involved.

[1] Also refer to, for instance,[Tal15] and [Fin17]

[2] For instance, it is not clear a priori what 'objective' actually means. In addition, please note that even empirical information or knowledge is subject to uncertainty.

[3] This collection might be a σ-algebra or the power set depending on Ω

[4] The term 'limit' here is not a mathematical limit as outlined in [Lin08]. The axiomatic approach avoids the difficulties, and the empirical observation will not be used to define probability, but only to suggest the axioms.

[5] Refer to section 5.3.3 in [Gra16]

1.1 Example (Ellsberg Paradox, see [Ell61] and [FS04]): A person is shown two urns, A and B. Each of them containing 10 balls of red or black color. Urn A contains 5 black and 5 red balls, while there is no additional information about urn B. That is, all balls in urn B could be black or red or any combination in between. One ball is drawn at random from each urn. The person is offered to make a bet on color of the ball chosen from either urn. Winning a bet the person receives, say, 1000 EUR. Possible bets are, for example, 'the ball drawn from urn A is black' denoted by Ab, 'the ball drawn from urn A is red' denoted by Ar, and, similarly Bb and Br. Let us assume that the person can pick one option for each of the following four bets:

(i) Bet on Ar, Ab or indifferent;

(ii) Bet on Br, Bb or indifferent;

(iii) Bet on Ar, Br or indifferent;

(iv) Bet on Ab, Bb or indifferent.

We denote the corresponding probabilities of an occurrence of Ar and Ab by p_{Ar} and p_{Ab}, respectively. A similar notation is used for the occurrence of Br and Bb. The probabilities of p_{Ar} and p_{Ab} as well as p_{Br} and p_{Bb} both each add up to one given the additivity of the model.

It has been observed empirically that most subjects prefer any bet on urn A to a bet on urn B. The decision that most subjects make with respect to the four bets is usually indifferent for (i) and (ii), Ar for (iii) and Ab for (iv).

From the empirically suggested choice for bet (iii) of drawing a red ball in urn B, we infer $p_{Br} < p_{Ar}$. Similarly, bet (iv) implies $p_{Bb} < p_{Ab} \Leftrightarrow 1 - p_{Br} < 1 - p_{Ar} \Leftrightarrow p_{Br} > p_{Ar}$ and thus a contradiction. Please also refer to Example 2.75 in Foellmer & Schied [FS04].

Going beyond the classical measure theory might be required in a human-centered model when the degree of uncertainty increases towards complete uncertainty as shown in Ellsberg's Paradox. For instance, when probabilities cannot be assigned to all events in $\mathcal{A}$, particularly, when only little information or scarce data is available. But even if a sufficient data basis is available, there are many sources of uncertainty such as experimental, parameter, algorithmic, interpolation or model uncertainty. Compensating the lack of objective empirical information by, for instance, subjective expert judgments can be fruitful and is quite common in finance. However, then it might become inevitable to extend the classical measure theory.

A *capacity* extends a classical measure by requiring monotonicity of a set function instead of additivity. It can therefore capture a higher degree of uncertainty, which can make the contradiction of the Ellsberg paradox disappear.[6]

Given the diversity of applications it is not surprising that there are many fields of research in different independent domains such as measure theory, theory of aggregation functions, decision theory, artificial intelligence and discrete mathematics (e.g. game theory and combinatorial optimization). [FS04] provides an introduction to financial stochastic models in discrete time. In addition, the interconnections to different types of risk measures (coherent, convex, etc.) are derived. [Gra16] focuses more on game-theoretic aspects but provides a good introduction to

[6] Refer to Example 2.77 in [FS04] or section 5.3.4 in [Gra16]

capacities and Choquet integrals. [WK09] provides a measure-theoretic introduction to different types of capacities.

In this note, we provide an introduction to *capacities* and *Choquet integrals*. In the next section 2, we provide some useful notation and basic facts. Capacities, modularity of capacities as well as generalized distribution functions are introduced and studied in section 3. Section 4 contains a short outline of the Choquet integral.

2 Prerequisites

Before we actually start with defining what a capacity is, it is convenient to introduce and extend certain notations as we need to consider higher-dimensional generalized distribution functions, for instance.

Let $\mathbb{N}$ denote the *natural numbers*, $\mathbb{R}$ the *real numbers*, $\mathbb{R}_+$ the set of *non-negative real numbers*, and $\mathbb{R}_-$ the set of *non-positive real numbers*. If A is a subset of Ω, then $A^C := \Omega \setminus A$ is the *complement set*. Further, in this note let $n \in \mathbb{N}$ be a positive integer.

2.1 Definition: 1) Ω denotes the non-empty *sample space* or *basic set*, whereby Ω might be finite, countable or an uncountable set;

2) $\mathcal{A}$ is a non-empty family of subsets of Ω. If $\mathcal{A}$ is equipped with some mathematical structure, such as ring, algebra, σ-ring, σ-algebra and the like, it is explicitly mentioned;

3) The pair $(\Omega, \mathcal{A})$ is called, in general, *space*. If $\mathcal{A}$ is a σ-algebra, then $(\Omega, \mathcal{A})$ is called *measurable space*. In general, probability measures are denoted by $\mathbb{P}$;

4) 2^Ω is the *power set* of Ω, i.e., $2^\Omega := \{A | A \subseteq \Omega\}$;

5) A *n-interval* is the Cartesian product of n real intervals and a *n-box* is a closed n-interval;

6) The *unit n-cube* $[0,1]^n$ is the n-box denoted by $[0,1]^n = [(0,\ldots,0),(1,\ldots,1)]$;

7) Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\} = [-\infty, \infty]$ and $\overline{\mathbb{R}}^n := \underbrace{\overline{\mathbb{R}} \times \ldots \times \overline{\mathbb{R}}}_{n \text{ times}}$. We extend the natural order of the reals through $-\infty < x < \infty$ for all $x \in \mathbb{R}$ using the following conventions. For all $x \in \mathbb{R}$ we define

$$x + (\pm\infty) = (\pm\infty) + x = (\pm\infty) + (\pm\infty) = \pm\infty.$$

Moreover,

$$x \cdot (\pm\infty) = (\pm\infty) \cdot x = \begin{cases} \pm\infty & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ \mp\infty & \text{for } x < 0 \end{cases}$$

and $\frac{x}{\pm\infty} = 0$. Last, we set $(\pm\infty) \cdot (\pm\infty) = +\infty$, $(\pm\infty) \cdot (\mp\infty) = -\infty$, $\frac{1}{0} := +\infty$. Thereby we have extended the total order to $\overline{\mathbb{R}}$. However, the operations

$$(\pm\infty) - (\pm\infty)$$

are undefined since they do not make any sense in this context.

The extended addition and multiplication on $\overline{\mathbb{R}}$ are commutative and associative, but the structure of $\overline{\mathbb{R}}$ together with the extended operations is not an algebraic field anymore. In addition, we set $\inf(\varnothing) = \infty = \sup(\mathbb{R})$, $\inf(\mathbb{R}) = -\infty = \sup(\varnothing)$. Those conventions will become important when dealing with inverse functions.

2.2 Definition: Let $A_1, \ldots, A_n$ be non-empty sets of $\mathbb{R}$, $F : A_1 \times \ldots \times A_n \to \overline{\mathbb{R}}$, and let $B = [x, y]$ be a n-box, where $x = (x_1, \ldots, x_n) < y = (y_1, \ldots, y_n)$. Let further $c = (c_1, \ldots, c_n) \in B$.

1) A *vertex* of a n-box B is a point $c = (c_1, \ldots, c_n) \in B$, where c_k equals either x_k or y_k $\forall k \in \{1, \ldots, n\}$. That is, a vertex is one of the 'corner points' of the n-box;

2) If the vertices of B are all distinct (which is equivalent to saying $x < y$), then

$$\operatorname{sgn}_B(c) := \begin{cases} 1 & \text{if } c_k = x_k \text{ for an even number of } k\text{'s} \\ -1 & \text{if } c_k = x_k \text{ for an odd number of } k\text{'s} \end{cases}$$

If the vertices of B are not all distinct, then $\operatorname{sgn}_B(c) = 0$. The function sgn is called the *sign-function* of B;

3) If B is a n-box, whose elements are all in the domain of F, then the *F-volume* of B is defined as the sum

$$V_F(B) := \sum_{c \text{ a vertex of } B} [\operatorname{sgn}_B(c) \cdot F(c)],$$

where the summation is taken over all vertices $c = (c_1, \ldots, c_n)$ of B;

4) The function F is called *n-increasing* if $V_F(B) \geq 0$ for all n-boxes, whose elements are in the domain of F.

Definitions 1) to 4) are still well-defined when extended to the more general n-intervals. For instance, in [DS16], the same concepts are applied on left-half-open n-intervals.

For the 2-dimensional case, we obtain

$$V_F([(x_1, x_2), (y_1, y_2)]) = F(y_1, y_2) - F(x_1, y_2) - F(y_1, x_2) + F(x_1, x_2).$$

Please note that $c = (y_1, x_2)$, for instance, contains only the element x_2, and, thus an odd number of elements of $\{x_1, x_2\}$. If $n = 1$, so that the domain of F is a subset of $\mathbb{R}$, then F is 1-increasing if and only if F is non-decreasing. Distribution functions of probability spaces are n-increasing.

The concept of n-increasing functions in a n-dimensional space is analogous to non-decreasing functions in one dimension. A n-dimensional function is n-increasing, if the F-volume of any n-box B in the domain of F is non-negative, i.e., their cumulative measure (or probability) should be always grater or equal than zero.

2.3 Definition: A set function on Ω is a mapping $\xi : \mathcal{A} \to \mathbb{R}$, assigning a real number to any subset in $\mathcal{A}$. A set function can be

1) *Additive* if $\xi(A \dot{\cup} B) = \xi(A) + \xi(B)$ for every disjoint $A, B \in \mathcal{A}$;

2) *Sub-additive* if $\xi(A \dot{\cup} B) \leq \xi(A) + \xi(B)$ for every disjoint $A, B \in \mathcal{A}$;

3) *Super-additive* if $\xi(A \dot{\cup} B) \geq \xi(A) + \xi(B)$ for every disjoint $A, B \in \mathcal{A}$;

4) *Monotone* if $\xi(A) \leq \xi(B)$ whenever $A \subseteq B$;

5) *Grounded* if $\xi(\emptyset) = 0$;

6) *Normalized* if $\xi(\Omega) = 1$;

7) *Positive homogeneity* if $\xi(x \cdot A) = x\xi(A)$;

8) *Bounded* if $\sup_{A \in \mathcal{A}} |\nu(A)| < +\infty$.

Furthermore, let $\vee$ = sup denote the supremum and $\wedge$ = inf denote the infimum. A poset $\mathcal{L} := \langle \Omega; \subseteq \rangle$ is called *lattice* if there exists an infimum and a supremum for every pair $A, B \subseteq \Omega$. Because of the required existence of infimum and supremum and their uniqueness, the maps

$$\begin{aligned} \wedge : \Omega \times \Omega \to \Omega \qquad\qquad & (A, B) \mapsto A \wedge B := \inf\{A, B\} \\ \vee : \Omega \times \Omega \to \Omega \qquad\qquad & (A, B) \mapsto A \vee B := \sup\{A, B\} \end{aligned}$$

are well-defined.

The Lebesgue measure on Ω, that is, a measure generated by $\mathbb{L}([a, b]) = b - a$ for $[a, b] \subseteq \Omega$, is denoted by $\mathbb{L}$. The *sign function* $\operatorname{sgn} : \mathbb{R} \to \{0, \pm 1\}$ of a real number extracts the sign of its argument. That is,

$$x \mapsto \operatorname{sgn}(x) := \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}.$$

A weight vector w with weights $(w_1, \ldots, w_n)$ is a vector $w \in \mathbb{R}^n$ satisfying $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0$ for all $i = 1, 2, \ldots, n$.

For a more detailed treatment of probability theory, we refer to standard literature such as [Bil95] and [Lin08].

3 Capacities

3.1 Basic Definitions

Aside from too high uncertainty in human-centered systems, there are also other real-world situations that cannot be adequately represented by an additive measure. Usually, the price of a set of goods equals the sum of the prices of its elements. The case where a collection is worth more than the single pieces is discussed in a) of the next example.

3.1 Example: a) Suppose there are two volumes x_1 and x_2 of a rare book.[7] Let $\Omega := \{x_1, x_2\}$ stand for the entire collection and suppose that there is a secondhand bookseller who buys the single volumes at the prices $\nu(\{x_1\})$ and $\nu(\{x_2\})$. If $\nu(\Omega)$ symbolizes the price of both volumes, then $\nu(\Omega) > \nu(\{x_1\}) + \nu(\{x_2\})$ as the entire collection is worth more than sum of the single volumes. This type of capacity is called super-additive and it is capable of expressing a beneficial relation between sets in terms of the measured property, which is in this example the price(s) of the books.

b) Consider a group $A \subseteq \Omega$ of firms, usually called a coalition because it is supposed that the individuals in A cooperate in some sense to achieve a common goal.[8] The annual profit $\mu(A)$ might be used to determine to which extend the group A has been able to achieve its common objective. Correspondingly, a group $\varnothing$ with no members should yield no profit at all, i.e. $\nu(\varnothing) = 0$. However, in a real-world situation monotonicity may be violated if the collaboration is not beneficial overall. If one firm k is close to bankruptcy one could have $\nu(A \setminus \{k\}) > \nu(A)$.

Apparently, real-world situations require the use of set functions that, like (probability) measures, are monotone with respect to set inclusion, but, unlike (probability) measures, are not additive, not even finitely additive. A capacity generalizes the classical concept of measure theory by dropping additivity and requiring monotonicity instead. The basic heuristic is as follows: the larger the set we observe, the more confident we can -usually- be that an event of interest is going to occur.

3.2 Definition: A set function $\nu : \mathcal{A} \to \mathbb{R}_+$ on $(\Omega, \mathcal{A})$ is called a *capacity*, if it satisfies the following two properties:

i) it is *grounded*, i.e. $\nu(\varnothing) = 0$;

ii) it is *monotone*, i.e. $A, B \in \mathcal{A}$, $A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$.

ν is called *finite* or *infinite* if $\nu(\Omega)$ is finite or infinite, respectively. A capacity is called *normalized* if $\nu(\Omega) = 1$.

Example 3.1 b) tells us that even the fairly general capacities do have their limitation to real-world problems. However, capacities are capable of modeling a) of Example 3.1 appropriately.

The present Definition 3.2 of a capacity deviates from CHOQUET's original since we strive to keep the note as general as possible. Choquet provided several definitions of different types of

[7] Example taken from [Gra16] and [Nar13]

[8] Example taken from [Gra16]

capacities, which are all too restrictive for most of our purposes. That is, Choquet requires certain types of continuity, see [Cho54], which we will only require when needed.

Generalizing the Bayesian approach to derive degrees of belief (priors) for statements leads us to the so-called belief functions. This special class of normalized capacities is quite popular and can be obtained thorough, for instance, DEMPSTER's upper & lower probabilities ([Dem67], [Dem68]), MATHERON's & KENDALL's random sets ([Ken74], [Mat74]) as well as SHAFER's evidence theory ([Sha76]). Whereas probability functions assume that mass is assigned to all countable singletons $x \in \Omega$ and aggregated via the additivity property, belief functions allow basic mass numbers to be assigned to subsets $A \subseteq \Omega$, called focal element, *without* further subdivision. That is, the belief (prior) allocated to focal elements $A \in \mathcal{A}$ is not further divided into proper subsets $B \in \mathcal{A}$, $B \subsetneq A$. All non-focal elements $B \in \mathcal{A}$ will be assigned with zero belief. The theory is well-explained in the seminal book [Sha76].

Please note that a capacity is defined on a family of sets $\mathcal{A}$ without having any specific mathematical structure. This changes, for instance, in section 3.4, where we define the distribution function associated with a capacity as continuity comes into the game.

An immediate consequence of the given definition is that a capacity ν is *positive*, i.e. $\nu(A) \geq 0$, $\forall A \in \mathcal{A}$ on a suitable (measurable) space $(\Omega, \mathcal{A})$. Let $\varnothing \in \mathcal{A}$, then the assertion $0 = \nu(\varnothing) \leq \nu(A)$ follows since $\varnothing \subset A$ and ν is monotone.

3.3 Example: Let $\Omega = \{1,2\}$ represent the set of the two volumes as outlined in a) of Example 3.1. Further, let $\mathcal{A} = 2^\Omega$ and $\nu' : 2^\Omega \to [0,\infty)$ be defined by

$$\begin{aligned} \varnothing &\mapsto 0 & \{1\} &\mapsto 500 \\ \{2\} &\mapsto 750 & \Omega = \{1,2\} &\mapsto 1500. \end{aligned}$$

This simple set function ν' is a capacity since it is grounded and monotone as can be double-checked manually. Be aware that ν' is not additive as $\nu'(\{1,2\}) > \nu'(\{1\}) + \nu'(\{2\})$, for instance. Let us now consider the corresponding 'normalized' version $\nu : 2^\Omega \to [0,1]$ on $(\{1,2\}$ defined by

$$\nu(A) := \nu'(A) \cdot \frac{1}{1500} = \begin{cases} 1 & \text{if } A = \{1,2\} \\ \frac{1}{3} & \text{if } A = \{1\} \\ \frac{1}{2} & \text{if } A = \{2\} \\ 0 & \text{if } A = \varnothing \end{cases}.$$

The set function ν is also non-additive, grounded, normalized and monotone as can be double-checked manually. Hence, ν is a normalized capacity.

Some technical examples will help to understand capacities better.

3.4 Example: a) Let $\mathcal{P}$ be a class of probability measures on the measurable space $(\Omega, \mathcal{A})$. ν_+, ν_- is defined as follows

$$\nu_+(A) := \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(A), \qquad \nu_-(A) := \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(A).$$

Then ν_+ and ν_- are normalized capacities[9], which follows from the definition of the probability measures $\mathbb{P} \in \mathcal{P}$. $P(\varnothing) = 0$ for all $\mathbb{P} \in \mathcal{P}$ and thus $\nu_+(\varnothing) = \nu_-(\varnothing) = 0$. In a similar manner, we can derive $\nu_+(\Omega) = \nu_-(\Omega) = 1$. Given that all probability measures are monotone, we can furthermore infer that ν_+ and ν_- are.

b) If we extend the finite counting measure via $|A| := \infty$ for all infinite $A \in 2^{\mathbb{N}}$, we receive $|\cdot| : \overline{\mathcal{A}} \to \overline{\mathbb{N}}$ that is still finitely additive but also σ-additive.[10] To see the σ-additivity property consider a sequence $(A_k) \in \overline{\mathcal{A}}$ with $k \in \mathbb{N}$ of pairwise disjoint sets, where each A_k can either be finite or infinite. If (at least) one of the A_k is infinite, say A_i, then $\bigcup_{k\in\mathbb{N}} A_k \supset A_i$ is also infinite, and

$$\sum_{k\in\mathbb{N}} |A_k| = \sum_{k<i} |A_i| + |A_i| + \sum_{k>i} |A_k| = \infty = |\bigcup_{k\in\mathbb{N}} A_k|.$$

If infinitely many of the disjoint A_k are non-empty, the union $\bigcup_{k\in\mathbb{N}} A_k$ is also infinite and

$$\sum_{k\in\mathbb{N}} |A_k| = \sum_{k\in\mathbb{N}, A_k\neq\varnothing} |A_k| \geq \sum_{k\in\mathbb{N}} 1 = \infty = |\bigcup_{k\in\mathbb{N}} A_k|.$$

If only finitely many A_k are non-empty and all A_k are finite, the assertion follows from the finite additivity. Please refer also to the similar but different Example 3.13.

c) Let ν be a capacity on the Borel measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then, the *k-th projection* ν_k for a Borel set $A \in \mathcal{B}(\mathbb{R}^n)$ with $A = (A_1, \dots, A_k, \dots, A_n)$ defined by

$$\nu_k(A) := \nu(\mathbb{R} \times \dots \mathbb{R} \times A_k \times \mathbb{R} \times \dots \times \mathbb{R})$$

is a capacity on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let A be a set with $A_k = \varnothing$, then $\nu_k(A) = \nu(\varnothing) = 0$ since the Cartesian product of any set with the empty set equals the empty set. If A is a set with $A_k = \mathbb{R}$, then $\nu_k(A) = \nu(\mathbb{R}^n) = 1$. The monotonicity of ν_k follows directly from the monotonicity of ν and its definition.

A (σ-)additive measure $\mu : \mathcal{A} \to \mathbb{R}$ does have the property that the range of the set function μ can be fully derived from the values $\mu(x)$, $x \in \Omega$ provided that Ω is at most countable.[11] That is, requiring additivity for such a μ actually means to only model the single elements of the basic set and deriving the measures of all other subsets by using additivity. This is, in general, not possible for capacities. How to effectively assign the real numbers $\nu(A)$ to all events (also called random sets[12]) $A \in \mathcal{A}$ of a capacity ν is a general problem since additivity cannot be used anymore.

The need and the chance to assign real numbers not only to isolated elements but to all subsets of Ω meet real-world requirements and offers great flexibility. Note that the number of subsets increase exponentially as there are 2^n subsets for a sample or outcome set Ω of size $|\Omega| = n$. The problem is even more complicated if we deal with infinite sample sets. A couple of solutions have been proposed to avoid combinatorial explosion or to create capacities from already existing ones. A subjective selection can be found in the following list of definitions.

[9] Example taken from [Sca96]
[10] Please also refer to Example 2.2 in [Den94]
[11] Note that the Lebesgue measure of a single $x \in \Omega$ is zero
[12] Refer to [Mol12] and [Ngu06], for instance

3.5 Definition: Let ν be a capacity on a space $(\Omega, \mathcal{A})$.

1) Let $\Omega = \{x_1, \ldots, x_n\}$ be finite, $\lambda > -1$ a fixed real number and ν a capacity. We call ν a *Sugeno λ-measure* if it complies with the so-called λ-rule

$$\nu(A \cup B) = \nu(A) + \nu(B) + \lambda \cdot \nu(A)\nu(B) \tag{3.1}$$

for all disjoint $A, B \subseteq \Omega$. The parameter λ can be derived by assigning real values to all single outcomes $\{x_1\}, \ldots, \{x_n\}$ and solving the following equation[13]

$$1 + \lambda = \prod_{i=1}^{n} (1 + \lambda\nu(x_i)) \Leftrightarrow 0 = \frac{\prod_{i=1}^{n} (1 + \lambda\nu(x_i)) - 1}{\lambda} - 1. \tag{3.2}$$

2) Let $\mathbb{P}$ be a probability measure on the measurable space $(\Omega, \mathcal{A})$ and $h : [0,1] \to [0,1]$ be an increasing function with $h(0) = 0$ and $h(1) = 1$. Then $\nu = h \circ \mathbb{P}$ is a capacity.[14] The capacity $\nu = h \circ \mathbb{P}$ is also called *distorted probability* and h the corresponding *distortion function*.[15]

3) Let $\Omega = \{x_1, \ldots, x_n\}$ be finite and $\mathrm{Bel} : 2^{\Omega} \to [0,1]$ be a grounded and normalized set function. If

$$\mathrm{Bel}(A_1 \cup \ldots \cup A_n) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \mathrm{Bel}\left(\bigcap_{i \in I} A_i\right) \tag{3.3}$$

for every collection of subsets $A_1, \ldots, A_n \subseteq \Omega$, then Bel is a *belief function* on Ω.[16]

4) Let $\mathcal{A}_1 \subset \mathcal{A}_2$ be two classes of subsets of a nonempty basic set Ω, and let ν_1, ν_2 be the corresponding set functions defined on $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. The set function ν_2 is called an *extension* of ν_1 from $\mathcal{A}_1$ to $\mathcal{A}_2$ if $\nu_1(A) = \nu_2(A)$ for every $A \in \mathcal{A}_1$.

5) Given a capacity ν, the set function $\widetilde{\nu}$ defined by $\widetilde{\nu}(A) := \nu(\Omega) - \nu(A^C)$ is called the *dual capacity* of ν. If $A, B \in \mathcal{A}$ with $A \subseteq B$, then $A^C \supseteq B^C$, and, thus $\widetilde{\nu}(A) \leq \widetilde{\nu}(B)$. The feature $\widetilde{\nu}(\emptyset) = 0$ can be inferred directly from the definition.

Regarding 1) of Definition 3.5, the commonly admitted interpretation of the Sugeno measure is as follows. If enough statistical evidence is available on the realization of A, λ would be set to zero and thus ν would be a probability measure. That is, $\nu(A) = \mathbb{P}(A)$ for all $A \in \mathcal{A}$. If not, it means that our knowledge on the experiment or statement is incomplete, and we lack evidence on the realization of A. Hence, the amount of certainty (accumulated evidence) on

[13] Refer to Theorem 4.2 of [WK09] where it is shown–in an even more general way–that the solution of equation (3.2) provides a suitable parameter λ

[14] To prove the assertion that a distortion is a capacity, let $A, B \in \mathcal{A}$ with $A \subseteq B$, then $\nu(A) = h(P(A)) \leq h(P(B)) = \nu(B)$ because h is increasing and P a probability measure. The set function is grounded and normalized as $\nu(\emptyset) = h(0) = 0$ and $\nu(\Omega) = h(1) = 1$, respectively

[15] In the discrete case, i.e. $|\Omega| \in \mathbb{N}$, the transformation $h : [0,1] \to [0,1]$ can also be seen as an importance weighting vector.

[16] Belief functions can also be generated with so-called *basic probability assignments* $m : \mathcal{A} \to [0,1]$, where m fulfills $\sum_{A \subseteq \Omega} m(A) = 1$. Then $\mathrm{Bel}_m(A) = \sum_{B \subseteq A} m(B)$ for all $A, B \in \mathcal{A}$ is a belief measure. Note that the basic probability assignment m is a set and not a point function. We refer to chapter 2 of [Sha76] for further details.

the realization of $A = A_1 \cup \ldots A_n$ (with $A_i \cap A_j = \varnothing$), quantified by $\nu(A)$, should be less than the probability $\mathbb{P}(A) = \mathbb{P}(A_1 \cup \ldots \cup A_n)$ of the event A.

Given the capacity $\nu : \mathcal{A} \to \mathbb{R}$, we can extend ν from $\mathcal{A}$ to $\mathcal{A}'$ via

$$B \mapsto \nu'(A') := \sup_{A \subseteq A'}\{\nu(A)\} \qquad \forall A' \in \mathcal{A}' \text{ and } A \in \mathcal{A}. \tag{3.4}$$

The *extension* ν' via the supremum is also a capacity since monotonicity is fulfilled by definition and all other requirements are inherited. Note that ν' is one of many possible extensions. It is therefore reasonable to restrict the study to continuous extensions or extensions that fulfill other interesting additional structural requirements. For a more detailed treatment of extensions of capacities, we refer to chapter 5 of [WK09]. Also refer to c) of the next Example 3.6.

Dual capacities may be important when dealing with lattices of capacities with respect to the set inclusion.[17] Please note that we are going to use distorted probability measures for the calculation of continuous Choquet integrals in section 4.

The next example applies Sugeno's λ-measure on a very small sample space and we also touch upon the role of the corresponding lattice with respect to the set inclusion. In addition, we study a quite simple extension using equation (3.4).

3.6 Example: Let $\Omega := \{x_1, x_2, x_3\}$ and $\mathcal{A} := 2^{\Omega}$, such that $(\Omega, \mathcal{A})$ is a measurable space. Let further $\nu : \mathcal{A} \to [0,1]$ be a capacity. The values $\nu(x_1) = 0.2, \nu(x_2) = 0.2$ and $\nu(x_3) = 0.3$ have been estimated by experts. The values of the other lattice elements, sketched in Figure 1, are either derived by using the Sugeno λ-measure or by using the extension formula (3.4).

a) First, we need to solve the equation $1+\lambda = (1+0.2\cdot\lambda)(1+0.2\cdot\lambda)(1+0.3\cdot\lambda)$ considering that $\lambda > -1$ according to equation (3.2). Second, the unique solution $\lambda = \frac{5}{3}$ may be employed to derive the values of the other subsets of Ω.

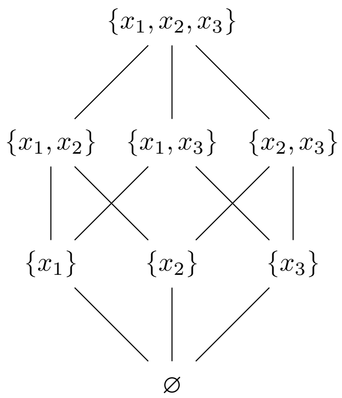

Figure 1: *Boolean lattice with eight elements*

To this end, we apply formula (3.1) to calculate $\nu(\{x_1, x_2\})$ via

[17] Please refer, for instance, to [DP08]

$$\begin{aligned}\nu(\{x_1\} \cup \{x_2\}) &= \nu(\{x_1\}) + \nu(\{x_2\}) + \frac{5}{3} \cdot \nu(\{x_1\}) \cdot \nu(\{x_2\}) \\ &= 0.2 + 0.2 + \left(\frac{5}{3} \cdot 0.2 \cdot 0.2\right) = \frac{7}{15}.\end{aligned}$$

Combining the disjoint sets $\{x_1, x_2\}$ and $\{x_3\}$ yields the desired normalization for Ω:

$$\begin{aligned}\nu(\{x_1, x_2\} \cup \{x_3\}) &= \nu(\{x_1, x_2\}) + \nu(\{x_3\}) + \frac{5}{3} \cdot \nu(\{x_1, x_2\}) \cdot \nu(\{x_3\}) \\ &= \frac{7}{15} + 0.3 + \left(\frac{5}{3} \cdot \frac{7}{15} \cdot 0.3\right) = 1.\end{aligned}$$

The set function values of $\{x_1, x_3\}$ and $\{x_2, x_3\}$ could be derived similarly. Both sets result in a value of $\frac{3}{5}$.

b) The dual capacity $\widetilde{\nu}(A) = \nu(\Omega) - \nu(A^C) = 1 - \nu(A^C)$ of a) can be determined by simple arithmetic and the results in the last example a). By definition, $\widetilde{\nu}(\varnothing) = 1 - \nu(\varnothing^C) = 1 - \nu(\Omega) = 0$. Accordingly, we get $\widetilde{\nu}(\{x_1\}) = 1 - \nu(\{x_2, x_3\}) = 1 - \frac{3}{5} = \frac{2}{5}$, $\widetilde{\nu}(\{x_3\}) = 1 - \nu(\{x_1, x_2\}) = 1 - \frac{7}{15} = \frac{8}{15}$ and $\widetilde{\nu}(\{x_1, x_2\}) = 1 - \nu(\{x_3\}) = 1 - 0.3 = 0.7$, for example.

c) The extension of ν from $\{\ \{x_1\}, \{x_2\}, \{x_3\}\ \}$ to $\mathcal{A} = 2^\Omega$ is quite straightforward using equation (3.4), i.e. the extension via the supremum function. For instance, $\nu(\{x_1, x_2\}) = \sup_{A \subseteq \{x_1, x_2\}}\{\nu(A)\}$, that is, $\nu(\{x_1, x_2\}) = \sup\{\nu(x_1), \nu(x_2), \nu(\varnothing)\} = 0.2$. Analogously, $\nu(\{x_1, x_3\}) = \nu(\{x_2, x_3\}) = \nu(\Omega) = 0.3$. Apparently, this type of filling up the missing values is more a technical exercise than a practical way to determine all values of a capacity.

3.2 Modularity

The next definitions are important in the context of convex/concave sets and functions.

3.7 Definition: A capacity ν on $(\Omega, \mathcal{A})$ is called

1) *submodular* if $\nu(A \cup B) + \nu(A \cap B) \leq \nu(A) + \nu(B)$ for all $A, B \in \mathcal{A}$;
2) *supermodular* if $\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B)$ for all $A, B \in \mathcal{A}$;
3) *modular* if the capacity is sub- and supermodular, i.e. if $\nu(A \cup B) + \nu(A \cap B) = \nu(A) + \nu(B)$ for all $A, B \in \mathcal{A}$.

Let $(\Omega, \mathcal{A})$ be a suitable non-empty space and $\mu : \mathcal{A} \to \mathbb{R}$ a set function. Modularity can be achieved by setting

$$\mu(A) := \mu(\varnothing) + \sum_{x \in A} \mu(\{x\}). \tag{3.5}$$

If $\mu(\varnothing) = 0$ and $\mu(\{x\})$ for all $x \in \Omega$ are prescribed arbitrarily, then the set function defined by (3.5) is additive and thus modular by design.

The following examples are going to illustrate some of the definitions made.

3.8 Example: a) Any probability measure $(\Omega, \mathcal{A}, \mathbb{P})$ yields an example of a modular capacity. The modularity of ν is obvious if the considered sets are disjoint since ν is additive. If $A_1, A_2 \in \mathcal{A}$ intersect in a non-empty set $A_1 \cap A_2 \neq \emptyset$, as illustrated in Figure 3.8, modularity can be seen by applying additivity to the disjoint sets $A_1 \setminus A_2$, $A_2 \setminus A_1$ and $(A_1 \cap A_2)$.

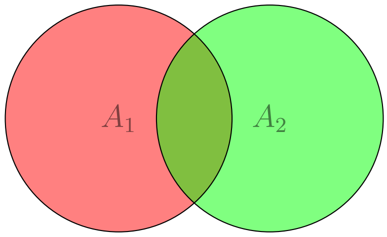

Figure 2: *Venn diagram of two intersecting sets A_1 and A_2*

b) Let $\Omega := \{v_1, v_2, v_3, v_4\}$ be the matrix

$$A := \begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 & v_4 \end{matrix} \\ \begin{pmatrix} 1 & 2 & 4 & 0 \\ 1 & 3 & 5 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

on the field $\mathbb{R}$ of real numbers and $\mathcal{A} := 2^{\Omega}$. The rank function $r : \mathcal{A} \to \mathbb{N} \subseteq \mathbb{R}$ of the sub-matrices defines a finite and submodular capacity. If $A_1, A_2 \in \mathcal{A}$ with $A_1 \subseteq A_2$, monotonicity $r(A_1) \leq r(A_2)$ follows since A_2 comprises more column vectors as A_1 and has therefore at least the rank of A_1. Finally, the sub-modularity of the rank function follows directly from the well-known rank-nullity theorem. A more general proof that the matroid rank function is submodular can be found in section 1.3 of [Oxl11].

c) It is common practice to reduce prices by granting a discount when many similar objects are bought together. If $\Omega := \{x_1, \ldots, x_n\}$ represents a finite set of $n \in \mathbb{N}$ similar objects for sale, and ν represents the price set function on $(\Omega, 2^{\Omega})$, then ν is a (sub-)additive measure. That is, $\nu(A \dot\cup B) \leq \nu(A) + \nu(B)$ for every disjoint $A, B \in 2^{\Omega}$. Let us further assume that each single object $x \in \Omega$ costs without any discount one unit. The granted discount depends on the relative number of objects which a counterparty agrees to buy.

Let us consider the linear function $\nu_0(A = \{x_1, \ldots, x_m\}) = |A| - \frac{|A|}{|\Omega|} = m - \frac{m}{n}$ for $A \in 2^{\Omega}$, where $x_i \neq x_j$ for all pairwise distinct $x_i, x_j \in A$. The measure ν_0 is additive and thus modular as can be double-checked using the identity $|A \dot\cup B| = |A| + |B|$.

If we define $\nu_1(A = \{x_1, \ldots, x_m\}) = |A| - \left(\frac{|A|}{|\Omega|}\right)^2$, which is a non-linear set function, ν_1 is a sub-additive measure. That is, $\nu_1(A \dot\cup B) < \nu_1(A) + \nu_1(B)$ for all $A, B \in \mathcal{A}$.

Submodular (supermodular) capacities are also sub-additive (super-additive), but the converse does not necessarily hold. We also refer to Examples 3.1 and 3.3, where examples of super-additive capacities are given.

3.9 Definition: A capacity ν on $(\mathcal{A}, \Omega)$ is called

1) *n-monotone* if for every $k \in \{2, \dots, n\}$ and $A_1, \dots, A_k \in \mathcal{A}$

$$\nu\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\varnothing \neq I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} \nu\left(\bigcap_{i \in I} A_i\right) \tag{3.6}$$

where $|I|$ is the cardinality of $I \in 2^{\{1,\dots,k\}} \setminus \varnothing$. We say that ν is 1-monotone if it is monotone.

2) *totally monotone* if it is k-monotone for every integer k.

Be aware that for $n = 2$ and $A_1, A_2 \in \mathcal{A}$ the inequality (3.6) becomes $\nu(A_1 \cup A_2) \geq [\nu(A_1) + \nu(A_2)] - \nu(A_1 \cap A_2)$. The n-monotone property is a generalization of the supermodularity, which is on the other hand related to convex functions. Note the interconnection between n-monotone capacities and belief functions, defined in 3) of Definition 3.5. That is, every belief function is totally monotone by definition.

The next lemma is going to be useful when continuous supermodular or submodular capacity examples are required. We are going to consider two of those example after the next lemma.

3.10 Lemma: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability measure and $h : [0,1] \to [0,1]$ be an increasing injective function with $h(0) = 0$ and $h(1) = 1$. Then, the function $A \mapsto \nu(A) := h(\mathbb{P}(A))$ is a distorted probability function of $\mathbb{P}$. If the distortion function h is convex (concave), the capacity ν is supermodular (submodular).

Proof. According to Definition 3.5 the composition $\nu = h \circ \mathbb{P}$ is a capacity. To prove the assertion that the capacity is supermodular if h is convex, we depart from modularity of the probability measure

$$\mathbb{P}(A \cap B) + \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

We have to show that $\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B)$. We pick $A, B \in \mathcal{A}$ such that $a := \mathbb{P}(A) \leq \mathbb{P}(B) =: b$. Then, $\mathbb{P}(A \cap B) =: u \leq a \leq b \leq v := \mathbb{P}(A \cup B)$. The intervals $[u, v] \supseteq [a, b]$ have common centers $\frac{1}{2}u + \frac{1}{2}v = \frac{1}{2}a + \frac{1}{2}b$ as $\mathbb{P}$ is modular. Applying the convexity of h infers $\frac{1}{2}h(u) + \frac{1}{2}h(v) \geq \frac{1}{2}h(a) + \frac{1}{2}h(b)$, which can be easily transformed to $h(u) + h(v) \geq h(a) + h(b)$. The set function ν is also 2-monotone as ν is supermodular. Please also compare with Example 2.1 in [Den94]. □

The following two examples show how the lemma can be used to generate sub- and supermodular examples.

3.11 Example: a) Let $\mathbb{U} : [1,4] \to [0,1]$ be the uniform distribution on the corresponding Borel sets and let $h(x) = x^2$ be a convex function with $h(0) = 0$ and $h(1) = 1$. The function $A \mapsto \nu(A) := h(\mathbb{U}(A))$ with $A \subseteq [1,4]$ is a supermodular capacity. The probability measure $A \mapsto \nu(A)$ is defined by $\nu(A) = \mathbb{U}(A)^2 = (\frac{1}{3}\mathbb{L}(A))^2$, e.g. $\nu([3,4]) = \mathbb{U}([3,4])^2 = (\frac{1}{3})^2 = \frac{1}{9}$. Obviously, ν is not additive since $\nu([1,2] \cup [2,3]) = \nu([1,3]) = (\frac{2}{3})^2 = \frac{4}{9}$, while $\nu([1,2]) + \nu([2,3]) = (\frac{1}{3})^2 + (\frac{1}{3})^2 = \frac{2}{9}$.

b) Let now $\mathbb{U} : [1,2] \to [1, \sqrt{2}]$ be the uniform distribution on the corresponding Borel sets, $h(x) = \sqrt{x}$ be a concave function with $h(0) = 0$ and $h(1) = 1$. The function

$A \mapsto \nu(A) := h(\mathbb{U}(A))$ with $A \subseteq [1,2]$ is a submodular capacity and its probability measure $A \mapsto \nu(A)$ is defined by $\nu(A) = \sqrt{\mathbb{U}(A)} = \sqrt{\mathbb{L}(A)}$. Obviously, ν is not additive since $\nu([1,1.5] \cup [1.5,2]) = \nu([1,2]) = \sqrt{1} = 1$, while $\nu([1,1.5]) + \nu([1.5,2]) = \sqrt{\frac{1}{2}} + \sqrt{\frac{1}{2}} > 1$.

Some authors directly call a set function convex (concave) if it is supermodular (submodular).

3.3 Continuity

As already mentioned in section 3.1, we strive to keep the note as general as possible. Hence, we only require continuity from above or below where needed.

3.12 Definition: A capacity ν on $(\mathcal{A}, \Omega)$ is called

1) *continuous from below* if $A_k \in \mathcal{A}$, $A_k \subseteq A_{k+1} \forall k \in \mathbb{N}$, implies $\lim_{k\to\infty} \nu(A_k) = \nu\left(\bigcup_{k=1}^{\infty} A_k\right)$;

2a) *continuous from above* if $A_k \in \mathcal{A}$, $A_k \supseteq A_{k+1} \forall k \in \mathbb{N}$, and $\nu(A_k) < \infty$ for at least one integer k, implies $\lim_{k\to\infty} \nu(A_k) = \nu\left(\bigcap_{k=1}^{\infty} A_k\right)$;

2b) *continuous from above at* $\varnothing$ if $A_k \in \mathcal{A}$, $A_k \supseteq A_{k+1} \forall k \in \mathbb{N}$, and $\nu(A_k) < \infty$ for at least one integer k, implies $\lim_{k\to\infty} \nu(A_k) = \nu(\varnothing) = 0$;

3) *continuous* if ν is continuous from above and below.

In plain English, property 1) of the last definition means that the limit of the increasing measures $\mu(A_1) \leq \ldots \leq \mu(A_k) \leq \ldots$ converges towards the measure $\mu(\bigcup_{k\in\mathbb{N}} A_k)$ for any increasing monotone sequence of sets $(A_k)_{k\in\mathbb{N}} \in \mathcal{A}$.

Regarding the properties 2a) and 2b), we note, that 2b) is a special case of the more general 2a). Further, we have to assume that at least one set A_k with finite measure exists in the case of decreasing sequence, to avoid counter-examples such as the following:
Let $|\cdot|$ denote the counting measure, as defined in Example 3.4 b). Consider the decreasing family of sets $\{A_k\}_{k\in\mathbb{N}}$, defined by $A_k := \{k, k+1, \ldots\}$. Apparently, $A_k \subseteq A_{k+1} \subseteq \ldots \downarrow \varnothing$ as $k \to \infty$. However, each set A_k still comprises infinitely many natural numbers. Hence, $|A_k| = \infty$ for all $k \in \mathbb{N}$, but $|\bigcap_{k\in\mathbb{N}} A_k| = |\varnothing| = 0$.

Let us illustrate continuity in this context with two simple examples.

3.13 Example: Let the discrete set function $\nu : \mathcal{A} \to \{0,1\}$ be defined on $(\Omega := \mathbb{N}, \mathcal{A} := 2^{\mathbb{N}})$ by

$$\nu_\alpha(A) := \begin{cases} \alpha & \text{if } |A| = \infty \\ 0 & \text{otherwise} \end{cases}$$

with $\alpha \in (0, \infty]$. The discrete set function ν_α is a capacity since it is grounded and monotone. For $k \in \mathbb{N}$ set $A_k := \{1, \ldots, k\}$.

a) The set function $\nu_{\alpha=1}$ is a normalized, submodular capacity. First, we show that ν is not supermodular. Consider the sets $\mathbb{N}_O := \{1,3,5,\ldots\}$ and $\mathbb{N}_E := \{2,4,6,\ldots\}$ of all odd and even natural numbers, respectively. Apparently, $\mathbb{N}_O \cap \mathbb{N}_E = \varnothing$, which is why we get $\nu_{\alpha=1}(\mathbb{N}_O \cup \mathbb{N}_E) + \nu_{\alpha=1}(\mathbb{N}_O \cap \mathbb{N}_E) = \nu_{\alpha=1}(\mathbb{N}) + \nu_{\alpha=1}(\varnothing) = 1 \ngeq \nu_{\alpha=1}(\mathbb{N}_O) + \nu_{\alpha=1}(\mathbb{N}_E) = 2$.

Arbitrary subsets $A, B \subseteq \mathbb{N}$ can either be finite or infinite. In all four possible combinations, the validity of the inequality $\nu_{\alpha=1}(A \cup B) + \nu_{\alpha=1}(A \cap B) \leq \nu_{\alpha=1}(A) + \nu_{\alpha=1}(B)$ is ensured. Hence, $\nu_{\alpha=1}$ is a submodular capacity. For all finite subsets A_k, we have $\nu_{\alpha=1}(A_k) = 0$. The infinite union of finite strictly increasing sets is infinite, that is, $\nu_{\alpha=1}(\bigcup_{k\in\mathbb{N}} A_k) = 1$ and $\nu_{\alpha=1}$ is therefore not continuous from below.

b) Now consider the capacity $\nu_{\alpha=\infty}$, which is even a finitely additive measure since ultimately $\infty + \infty = \infty$. According to the definition, we have $\nu(A_k) = 0$, $A_k \uparrow \mathbb{N}$. On the other hand, $\nu_{\alpha=\infty}\left(\bigcup_{k=1}^{\infty} A_k\right) = \nu_{\alpha=\infty}(\mathbb{N}) = \infty \neq \lim_{k\to\infty} \nu_{\alpha=\infty}(A_k) = 0 = \sum_{k\in\mathbb{N}} \nu_{\alpha=\infty}(\{k\})$. The capacity $\nu_{\alpha=\infty}$ is therefore not continuous from below and it is not σ-additive.

A finitely additive measure μ on a ring is continuous from below if and only if μ is σ-additive. In addition, continuity from below implies continuity from above at $\varnothing$ under certain conditions as pointed out in the following theorem.

3.14 Theorem: Let μ be a positive and additive set function on a ring $\mathcal{R}$ above Ω. Then the following holds:

$$\mu \text{ is } \sigma\text{-additive } \Leftrightarrow \mu \text{ is continuous from below } \Rightarrow \mu \text{ is continuous from above.}$$

Proof. Let μ be σ-additive and $A_n \in \mathcal{R}$ with $A_n \uparrow A \in \mathcal{R}$. Any countable union can be written as a countable union of disjoint sets. Let $A_1, A_2, \ldots \in \mathcal{R}$ and define $D_1 := A_1$, $D_2 := A_2 \setminus A_1$, $D_3 := A_3 \setminus (A_1 \cup A_2), \ldots$ with $(D_i) \in \mathcal{R}$. Then, (D_i) is a collection of disjoint sets with $A = \bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}} D_i = \sum_{i\in\mathbb{N}} D_i$. Due to the σ-additivity as well as the relation between A_i and D_i, we can derive

$$\mu(A) = \mu(\bigcup_{i\in\mathbb{N}} D_i) = \mu(\sum_{i\in\mathbb{N}} D_i) = \sum_{i\in\mathbb{N}} \mu(D_i) = \lim_{k\to\infty} \sum_{i=1}^{k} \mu(D_i) = \lim_{k\to\infty} \mu(\sum_{i=1}^{k} D_i) = \lim_{k\to\infty} \mu(A_k),$$

which proves that μ is continuous from below.
Now, let μ be continuous from below and $(D_i) \in \mathcal{R}$ be a series of *disjoint* sets with $\sum_{i\in\mathbb{N}} D_n \in \mathcal{R}$. Due to the fact that $\sum_{i=1}^{k} D_i \uparrow \sum_{n\in\mathbb{N}} D_n$ if $k \to \infty$, we infer $\lim_{k\to\infty} \sum_{i=1}^{k} \mu(D_i) = \mu(\lim_{k\to\infty} \sum_{i=1}^{k} D_i) = \mu(\sum_{i\in\mathbb{N}} D_i) = \mu(\bigcup_{i\in\mathbb{N}} D_i)$ by applying the (finite) additivity of μ as well as the continuity from below. Apparently, μ is σ-additive.
Finally, let μ be continuous from below with $A_1, A_2, \ldots \in \mathcal{R}$ and $A_i \downarrow A \in \mathcal{R}$. If $\mu(A_1) < \infty$, then all subsequent sets $A_2, A_3, \ldots$ will also result in a real number (smaller than ∞). Using the same argument, we can also infer that $\mu(A) < \infty$. Set $B_i := A_1 \setminus A_i \in \mathcal{R}$ for every $i \in \mathbb{N}$ and $B_i = A_1 \setminus A_i \uparrow A_1$. That is, we have transformed a decreasing set function into an increasing one. Hence,

$$\lim_{i\to\infty} \mu(B_i) = \lim_{i\to\infty} \mu(A_1 \setminus A_i) = \mu(\lim_{i\to\infty} B_i) = \mu(\bigcup_{i\in\mathbb{N}} B_i) = \mu(A_1).$$

Continuity from above at $\varnothing$ is just a special case of the already proven direction. □

According to the last theorem and Proposition 3.30 in [PS16], probability measures are continuous by design. Capacities, on the contrary, can be non-continuous as outlined in the last examples.

The following example, taken from [WK09], exhibits the root-cause why a continuous extension, as defined in section 3.1, is not in all cases possible.

3.15 Example: Let $(\Omega := \mathbb{N}, \mathcal{A} := 2^{\Omega})$ be a space, and, $\mathcal{F} \subseteq \mathcal{A}$ be the class of all finite subsets of Ω. Define the set function $\nu : \mathcal{F} \to \{0,1\}$ by

$$F \mapsto \nu(F) := \begin{cases} 0 & \text{if } F = \varnothing \\ 1 & \text{otherwise} \end{cases} . \tag{3.7}$$

Then, ν is a finite continuous capacity on $\mathcal{F}$. Note that a monotone decreasing sequence can only converge towards $\varnothing$ if $\varnothing$ is one of its elements. If a capacity ν' is an extension of ν from $\mathcal{F}$ to $\mathcal{A}$, then $\varnothing \neq A \mapsto \nu'(A) \geq 1$ since ν' is monotone. There exist at least one infinite A_k, for instance, $A_k = \{k, k+1, \ldots\}$ and a corresponding family of sets (F_i), for instance, $F_i := \{k, k+1, \ldots, k+i-1\}$ with $i \in \mathbb{N}$, such that $F_i \uparrow A_k$.
There are now two possible cases for the extension ν'. First, we assume that there exist some infinite A_k with $\nu'(A_k) = \infty$. Since F_i is finite and ν' an extension with $F_i \uparrow A_k$, we have $\nu'(F_i) = \nu(F_i) = 1$ for all $i \in \mathbb{N}$. This shows that an arbitrary extension cannot be continuous from below. Second, if $\nu'(A_k) < \infty$ for any infinite set A_k with $k \in \mathbb{N}$. From $A_k \downarrow \varnothing$ and $\nu'(A_k) \geq 1$ for every $k \in \mathbb{N}$, we know that ν' is not continuous from above at $\varnothing$. Consequently, the extension ν' cannot be a continuous capacity on $\mathcal{A}$.

3.4 Generalized Distribution Functions

Based on [Sca96], we define a distribution function of a capacity as follows.

3.16 Definition ([Sca96]): Let ν be a capacity on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The function $F_\nu : \overline{\mathbb{R}^n} \to \mathbb{R}$ defined by

$$F_\nu(x_1, \ldots, x_n) \ := \ \nu\left([-\infty, x_1] \times \ldots \times [-\infty, x_n]\right) \tag{3.8}$$

is called a *(generalized joint) distribution function* associated with the capacity ν. Correspondingly, the distribution function associated with the k-th projection ν_k is the function

$$F_{\nu_k}(x_1, \ldots, x_k, \ldots, x_n) \ := \ \nu\left(\overline{\mathbb{R}^n} \times \ldots \times [-\infty, x_k] \times \ldots \overline{\mathbb{R}^n}\right). \tag{3.9}$$

Note that the k-th projection of a capacity on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is a capacity according to c) of Example 3.4.
The function $G_\nu : \overline{\mathbb{R}^n} \to \mathbb{R}$ defined by

$$G_\nu(x_1, \ldots, x_n) \ := \ \nu\left([x_1, \infty] \times \ldots \times [x_n, \infty]\right) \tag{3.10}$$

is called *(generalized joint) survival function* associated with the capacity ν.

A specific type of survival function will be more suitable for the definition of the so-called Choquet integral in the next section. In this section, however, we will focus on the distribution function with respect to a capacity. The distribution or survival function of a discrete capacity[18] is not covered in Definition 3.16.

[18] If we are interested in a capacity on a finite or countable set $\Omega \subseteq \mathbb{R}^n$, we would have to adapt the definition accordingly

The function F_ν can have countable many jump discontinuities in one dimension.[19] No matter how small a jump $a \in \mathbb{R}$ would be, we could still find a $n \in \mathbb{N}$, such that $\frac{1}{n} < a$. A maximum of n such jumps would fit into the range $[0,1]$ of F_ν. Thus, the set of all jumps $\bigcup_{n \in \mathbb{N}} \{\text{jumps of size } \frac{1}{n} \text{ or more}\}$ contains countably many discontinuities at most.

Before we actually study the connection between the properties and the necessary assumptions, let us consider a simple example of a generalized distribution function.

3.17 Example: Recall the supermodular capacity $A \mapsto \nu(A) := h(\mathbb{U}(A))$, $A \subseteq [1,4]$ on the corresponding Borel sets as outlined in Example 3.11. ν is defined by a convex distortion $h(x) = x^2$ of the uniform distribution $\mathbb{U}[1,4]$. Latter one can be represented by its probability distribution function $F_{\mathbb{U}}(t) = \frac{1}{3}(t-1)$.

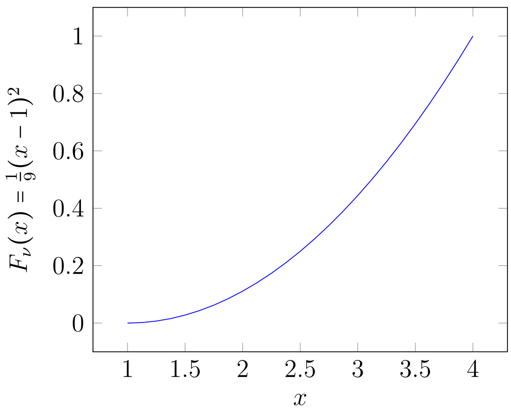

Figure 3: *Continuous generalized distribution function of a supermodular capacity*

According to equation (3.8), the continuous distribution function associated with ν equals $F_\nu(x) = h(F_{\mathbb{U}}(x)) = \frac{1}{9}(x-1)^2$ for $x \in [1,4]$ as sketched in Figure 3.17.

The last example exhibits a continuous generalized distribution function, however, in general capacities are not continuous from below but they do comply with (df1) - (df3).

3.18 Proposition: Any multivariate distribution function F_ν associated with a capacity $\nu : \Omega := \overline{\mathbb{R}^n} \to \mathbb{R}$ satisfies:

(df1) $F_\nu(x_1, \ldots, x_n)$ is increasing[20] in each argument;

(df2) $F_\nu(x_1, \ldots, x_n) = 0$ if $\min\{x_1, \ldots, x_n\} \to -\infty$ and ν is continuous from above;

(df3) $F_\nu(x_1, \ldots, x_n) = \nu(\Omega) = 1$ if $\lim x_k = \infty$ for *all* $k \in \{1, \ldots, n\}$ and ν is continuous from below;

Proof. Let (x_{ki}), $k \in \{1, \ldots, n\}$ and $i \in \mathbb{N}$ be a set of n arbitrary sequences of real numbers.

[19] Refer to Example 3.19 for the multi-dimensional case

[20] Let $a, b \in \mathbb{R}$ with $a < b$, then F is called increasing in the k-th element if $F(x_1, \ldots, x_k = a, \ldots, x_n) \leq F(x_1, \ldots, x'_k = b, \ldots, x_n)$

(df1) Consider the n-boxes $A := [-\infty, x_1] \times \ldots \times [-\infty, x_k] \times \ldots \times [-\infty, x_n]$ and $B := [-\infty, x_1] \times \ldots \times [-\infty, x'_k] \times \ldots \times [-\infty, x_n]$ with $k \in \{1, \ldots, n\}$. Given that $A \subseteq B$, we have $\nu(A) \leq \nu(B)$ and thus $F_\nu(x_1, \ldots, x_k, \ldots, x_n) \leq F_\nu(x_1, \ldots, x'_k, \ldots, x_n)$ because of the definition of F_ν.

(df2) Let $x_{\min,i} := \min\{x_{1i}, \ldots, x_{ni}\}$ be the minimum of the set of sequences (x_{ki}) that converges towards $-\infty$ for $i \to \infty$. According to the definition of F_ν, we receive a corresponding sequence $([-\infty, x_{1i}] \times \ldots \times [-\infty, x_{\min i}] \times \ldots \times [-\infty, x_{ni}])$ of n-boxes, where at least the interval with index min converges towards $\varnothing$ if $i \to \infty$. The Cartesian product of any set comprising the empty set equals the empty set. The assertion therefore follows by using the definition of F_ν and the continuity from above:

$$\begin{aligned}\lim_{i\to\infty} F(x_{1i}, \ldots, x_{\min i}, \ldots, x_{ni}) &= \lim_{i\to\infty} \nu([-\infty, x_{1i}] \times \ldots \times [-\infty,_{ni}]) \\ &= \nu\left(\bigcap_{i=1}^{\infty} ([-\infty,_{1i}] \times \ldots \times [-\infty,_{ni}])\right) \\ &= \nu(\varnothing) = 0.\end{aligned}$$

(df3) Let again $([-\infty, x_{1i}] \times \ldots \times [-\infty, x_{ni}])$ be a sequence of n-boxes, where *all* x_{ki}, $k \in \{1, \ldots, n\}$ converges towards ∞ if $i \to \infty$, denoted by $(x_{ki}) \overset{i\to\infty}{\to} \Omega = \overline{\mathbb{R}^n}$. Interpreting this geometrically means that the corresponding union of those sequences of n-intervals converges increasingly towards $\Omega = \overline{\mathbb{R}^n}$. Given that ν is continuous from below, we infer

$$\begin{aligned}\lim_{x \overset{i\to\infty}{\to} \infty} F_\nu(x_{1i}, \ldots, x_{ki}, \ldots, x_{ni}) &= \lim_{(x_{ki}) \overset{i\to\infty}{\to} \infty} \nu([-\infty, x_{1i}] \times \ldots \times [-\infty, x_{ni}]) \\ &= \nu\left(\bigcup_{i=1}^{\infty} [-\infty, x_{1i}] \times \ldots \times [-\infty, x_{ni}]\right) \\ &= \nu(\Omega) = 1.\end{aligned}$$

□

A generalized distribution function may not be n-increasing, i.e., it may not be a probability distribution function as the following example shows.

3.19 Example: Let $F : \overline{\mathbb{R}^2} \to [0, 1]$ be defined by

$$F(x, y) = \begin{cases} 0 & \text{if } y + x < 0 \\ 1 & \text{if } y + x \geq 0 \end{cases}.$$

The function F takes the value 0 for the points below the line $y = -x$ in $\mathbb{R}^2$, whereby it takes the value 1 for the points above and on that line. The blue-dashed area including the line in Figure 4 represents the part of the domain that is mapped to 1 via F, while the area below the line is mapped to 0.

F is increasing in each argument, i.e., it complies with (df1), as can be seen by keeping either x or y constant and varying the other variable from $-\infty$ to ∞. Apparently, F also converges to 0 and 1 as required in (df2) and (df3), even though an underlying capacity is not continuous from below.[21]

[21] To see that the underlying capacity is not continuous from below, let $(A_k)_{k\in\mathbb{N}}$ be a family of sets defined by $A_k := [-\infty, -\frac{1}{k}] \times [-\infty, -\frac{1}{k}]$. Apparently, $A_k \uparrow A := [-\infty, 0] \times [-\infty, 0]$ and $\nu(A) = 1$. However, $\nu(A_k) = 0$ for all $k \in \mathbb{N}$.

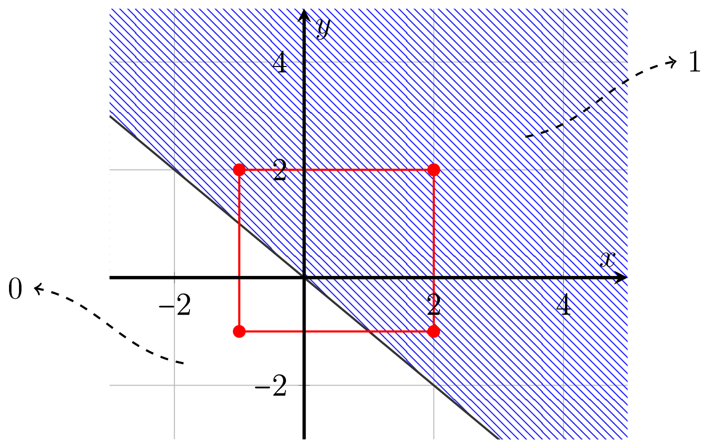

Figure 4: *Graphical segmentation of F's domain $\overline{\mathbb{R}^2}$ as well as an indication for the corresponding function values*

F is, however, not 2-increasing. To see this, pick the left-lower corner of a 2-box below the line $y = -x$. The other three points need to be above the line $y = -x$. For instance, set $x_1 = 2$, $x_2 = -1$, $y_1 = 2$ and $y_2 = -1$ as sketched in Figure 4, then we get $+F(2,2) - F(2,-1) + F(-1,-1) - F(-1,2) = -1$. Hence, F is not 2-increasing and it is therefore not a probability distribution function.

Capacities are in general not n-increasing as outlined in Example 3.19. The next lemma shows what property a capacity needs to overcome this shortfall.

3.20 Lemma: ([Sca96]) If ν on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is n-monotone, then F_ν is k-increasing for every positive integer $k \le n$.

Proof. Refer to Lemma 6 of [Sca96]. □

In particular for the case $n = 2$, the distribution function of any supermodular capacity is increasing and 2-increasing.

3.21 Corollary: ([Sca96]) If ν is a n-monotone capacity on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, then there exists a finitely additive probability measure μ on M, such that $F_\mu = F_\nu$.

Proof. By Lemma 3.20, if ν is n-monotone, then its distribution function F_ν is k-increasing for every positive $k \le n$. According to [MDCQ08] this coincides with the distribution function of a finitely additive probability measure. □

A function F that complies with (df1) - (df3) might be used to construct a capacity whose distribution function is exactly F.

3.22 Example: We continue with Example 3.19 and show that a corresponding capacity can be derived from F. We define $\nu([-\infty, x] \times [-\infty, y]) := F(x, y)$ for all 2-boxes $A := [-\infty, x] \times [-\infty, y]$ in $\overline{\mathbb{R}^2}$. Consider the extension of ν from the set of all 2-boxes to the Borel sets, defined by

$$B \mapsto \nu'(B) := \sup_{A \subset B} \nu(A) \qquad \text{for all } B \in \mathcal{B}. \tag{3.11}$$

The supremum of all known measures $\nu(A)$ with $\mathcal{A} \ni A \subset B \in \mathcal{B}$ is used to come up with an extended measure on the Borel sets $B \in \mathcal{B}$. Since the measure is defined via the supremum, it is monotone by definition. The remaining properties of a capacity are also fulfilled. Please refer to Example 3.6 c) for an example with finite basic set.

The next example illustrates that a distribution function associated with a capacity does not uniquely characterize a capacity. That is, each function F satisfying properties (df1) to (df3) may be used to construct a capacity ν, whose distribution function is F. Contrary to the case of probability measures, the derived capacity ν is not uniquely determined by F.

3.23 Example: Consider the probability distribution function $F_{\mathbb{L}} : [0,1]^2 \to [0,1]$ defined by $F_{\mathbb{L}}(x_1, x_2) = (1 - x_1)(1 - x_2)$ on the Borel space $(\overline{\mathbb{R}^2}, \mathcal{B}(\overline{\mathbb{R}^2}), \mathbb{L})$.[22] Applying the probability

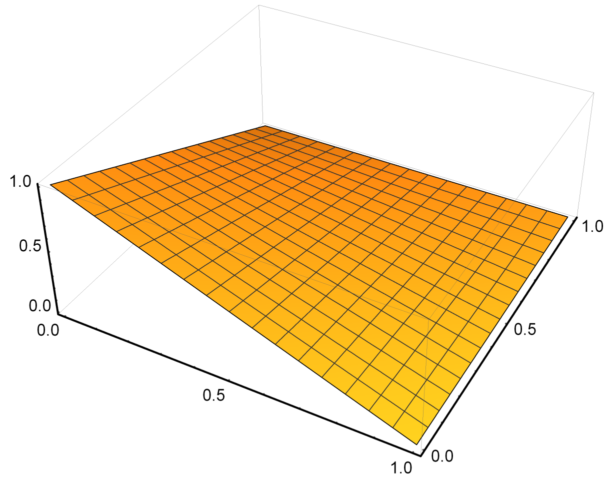

Figure 5: *Graph of probability distribution function $F_{\mathbb{L}}$ on* $[0,1]^2$

measures on the 2-box $A = [\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{4}, \frac{3}{4}]$, we obtain the F-volume $\mathbb{L}(A) = \frac{1}{4}$. If we instead use the extension defined in (3.4) to determine the capacity ν (and take monotonicity into account), we get

$$\begin{aligned}
\nu(A) &= \sup_{(x_1,x_2)\in A} F(x_1, x_2) \\
&= \sup \left\{ F\left(\frac{1}{4}, \frac{1}{4}\right), F\left(\frac{1}{4}, \frac{3}{4}\right), F\left(\frac{3}{4}, \frac{1}{4}\right), F\left(\frac{3}{4}, \frac{3}{4}\right) \right\} \\
&= \sup \left\{ \frac{3}{16}, \frac{3}{16}, \frac{1}{16}, \frac{9}{16} \right\} = \frac{9}{16}.
\end{aligned}$$

Note, that the Lebesgue measure $\mathbb{L}$ as well as ν are both capacities with $F_{\mathbb{L}}$ as its distribution function.

As in the additive case, the multivariate distribution function of a supermodular capacity satisfies the Frechet bounds.

3.24 Theorem (Frechet bounds for supermodular capacities, [Sca96]): If the capacity ν on

[22] A similar example can be found in [DS10]

$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is supermodular, then

$$\max\left(\sum_{k=1}^{n} F_{\nu_k} - n + 1, 0\right) \leq F_\nu(x_1, \ldots, x_n) \leq \min(F_{\nu_1}(x_1), \ldots, F_{\nu_n}(x_n)), \tag{3.12}$$

where F_ν is the distribution function associated with ν and F_{ν_k} the corresponding marginal distribution functions.

Proof. The second inequality $F_\nu(x_1, \ldots, x_n) \leq \min(F_{\nu_1}(x_1), \ldots, F_{\nu_n}(x_n))$ is immediate by monotonicity of ν and the definition of F_{ν_k} for $k \in \{1, \ldots, n\}$. To proof the first inequality, let $A_i := (\overline{\mathbb{R}} \times \ldots \times \overline{\mathbb{R}} \times [-\infty, x_i] \times \overline{\mathbb{R}} \times \ldots \times \overline{\mathbb{R}})$. By supermodularity of ν applied on the two sets A_1 and $(A_2 \cap \ldots \cap A_n)$ we get

$$\nu(A_1 \cap \ldots \cap A_n) \geq \nu(A_1) + \nu(A_2 \cap \ldots \cap A_n) - \nu(A_1 \cup (A_2 \cap \ldots \cap A_n)).$$

Iterating this principle $(n-1)$ times on $(A_2 \cap \ldots \cap A_n)$, i.e., applying the supermodularity of ν on the two sets A_2 and $(A_3 \cap \ldots \cap A_n)$ and considering that $-\nu(A_1 \cup (A_2 \cap \ldots \cap A_n) \geq -1$ we get

$$\begin{aligned} \nu(A_1 \cap \ldots \cap A_n) &\geq \nu(A_1) + \nu(A_2) + \nu(A_3 \cap \ldots \cap A_n) - \nu(A_2 \cup (A_3 \cap \ldots \cap A_n)) - 1 \\ &\vdots \\ &\geq \nu(A_1) + \ldots + \nu(A_n) - n + 1, \end{aligned}$$

which proves the assertion. □

4 Choquet Integral

The Lebesgue integral has a lot of useful properties for both theory and application. However, it may also be that the Lebesgue integral is too restrictive for certain situations. The root-cause is again the σ-additivity, just as in the case of probability measures. A natural question is therefore, how we could meaningfully define the integral $\int f \, d\nu$ of a function f with respect to a capacity ν?

The answer is not unique as there are several approaches to define such a generalized integral. We will focus on the so-called Choquet integral. Choquet defined[23] an integration operation with respect to the non-necessarily additive set function ν in the 1950s. He has shown that it is possible to develop a rich theory of integration in a non-additive setting. Schmeidler [Sch86] rediscovered and extended the Choquet integral in the 1980s. This section is based on [MM03], [Gra16] and [Sch86], that take also Schmeidler's ideas into account. The Choquet integral will be introduced in several steps, beginning with non-negative and then extending the approach to general functions. For the sake of understanding, it is helpful to know how the Lebesgue measure and integral is defined and derived since there are several similarities between Choquet's and Lebesgue's approach.

4.1 Basic Definitions

Before we can actually define the Choquet integral of a function $f : \Omega \to \mathbb{R}$, we first need to study so-called upper level sets, which divide $\mathrm{ran}(f) = \mathbb{R}$ into a non-increasing family of sets. For the sake of completeness and understanding, we will also define so-called lower level sets. Please also refer to, for instance, [Den94].

[23] Refer to page 265 ff in [Cho54]

4.1 Definition: Let $(\Omega, \mathcal{A})$ be a measurable space, $f : \Omega \to \mathbb{R}$ be a $\mathcal{A}$-measurable function and $t \in \mathbb{R}$, then the set

$$\{f \geq t\} := \{x \in \Omega |\ f(x) \leq t\ \} \subseteq \Omega$$

is called a *upper level set of f for the level t.* The system of all upper level sets of f is called the *family of upper level sets* of f. A set $\{f > t\} = \{x \in \Omega |\ f(x) > t\ \}$ is called *strict upper level set of f for the level t.* If we replace '≤' ('<') by '≥' ('>') in the definition of an upper level set we would receive (*strict*) *lower level sets.*

In this note, however, we focus on upper level sets since these are more suitable to define the Choquet integral. The following examples illustrate the definition of lower level sets using simple functions.

4.2 Example: a) Let us consider the linear function $f(x) = -2x + 1$ defined on $f : \Omega := [0, 2] \to [-3, 1]$ as depicted in Figure 6. For $t \in \{-3, -2, -1, 0, 1\} \subset \text{ran}(f)$, we get the

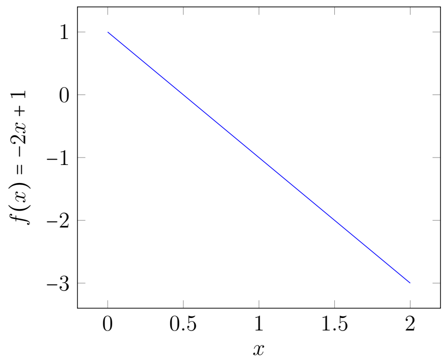

Figure 6: *Graph of function* $f(x) = -2x + 1$ *on domain* $[0, 2]$

following upper level sets by solving the defining inequalities:

$$\{f \geq -3\} = \{x \in [0,2] \mid -2x + 1 \geq -3\ \} = [0,2] \qquad \{f \geq -2\} = \{x \in [0,2] \mid -2x + 1 \geq -2\ \} = [0, 1.5]$$
$$\{f \geq -1\} = \{x \in [0,2] \mid -2x + 1 \geq -1\ \} = [0,1] \qquad \{f \geq 0\} = \{x \in [0,2] \mid -2x + 1 \geq 0\ \} = [0, 0.5]$$
$$\{f \geq 1\} = \{x \in [0,2] \mid -2x + 1 \geq +1\ \} = \{0\}.$$

If $t \in [-3, 1)$, the solution is given by $[0, \frac{1-t}{2}]$. Apparently, for $t = 1$ we get $\{0\}$ as solution. The upper level sets are non-increasing for increasing t's, i.e, $\{f \geq -3\} \supset \{f \geq -2\} \supset \{f \geq -1\} \supset \{f \geq 0\} \supset \{f \geq 1\}$.

b) Let us now consider the non-injective function $f : [-2, 2] \to \mathbb{R}$ defined by $f(x) = x^2$ with $\Omega := [-2, 2]$. The corresponding upper level sets for $t \in \{0, 1, 2, 3, 4\} \subset \text{ran}(f)$ are as follows:

$$\{f \geq 0\} = \{x \in [-2,2] \mid x^2 \geq 0\ \} = [-2, 2] \qquad \{f \geq 1\} = \{x \in [-2,2] \mid x^2 \geq 1\ \} = \Omega \setminus (-1, 1)$$
$$\{f \geq 2\} = \{x \in [-2,2] \mid x^2 \geq 2\ \} = \Omega \setminus (-\sqrt{2}, \sqrt{2}) \qquad \{f \geq 3\} = \{x \in [-2,2] \mid x^2 \geq 3\ \} = \Omega \setminus (-\sqrt{3}, \sqrt{3})$$
$$\{f \geq 4\} = \{x \in [-2,2] \mid x^2 \geq +4\ \} = \{-2, +2\}.$$

Again, the upper level sets are non-increasing for increasing t's, i.e, $\{f \geq 0\} \supset \{f \geq 1\} \supset \{f \geq 2\} \supset \{f \geq 3\} \supset \{f \geq 4\}$. Apparently, $\Omega \setminus (-\sqrt{t}, \sqrt{t})$ provides the upper level sets $\{x^2 \geq t\}$ of f for levels $t \in [0,4] = \mathrm{ran}(f)$.

c) Let us now consider the exponential function $f : [0,1] \to \mathbb{R}$ defined by $f(x) = \exp(x) = e^x$ with $\Omega := [0,1]$. The corresponding upper level sets for $t \in \{1,2,e\} \subset \mathrm{ran}(f) = [1,e]$ are as follows:

$\{f \geq 1\} = \{x \in [0,1] \mid e^x \geq 1\} = [\ln(1), 1] = [0,1]$ $\quad$ $\{f \geq 2\} = \{x \in [0,1] \mid e^x \geq 2\} = [\ln(2), 1]$
$\{f \geq e\} = \{x \in [0,1] \mid e^x \geq e\} = [\ln(e), 1] = \{1\}$

The upper level sets are again non-increasing for increasing t's, i.e, $\{f \geq 1\} \supset \{f \geq 2\} \supset \{f \geq e\}$. Apparently, $[\ln(t), 1]$ provides the upper level sets $\{e^x \geq t\}$ for levels $t \in [1,e] = \mathrm{ran}(f)$.

The following lemma is going to formalize the observation made in the examples above.

4.3 Lemma: Let $(\Omega, \mathcal{A})$ be a measurable space and $f : \Omega \to \mathbb{R}$ be a $\mathcal{A}$-measurable function. The upper level sets $\{f \geq t\}$ are non-increasing with respect to increasing $t \in [0, \infty)$ and the set inclusion.

Proof. Let $t_1, t_2 \in \mathbb{R}$ two arbitrary real numbers with $t_1 < t_2$. We need to prove that $\{f \geq t_1\} \supseteq \{f \geq t_2\}$. A real value $f(x)$ is assigned to every $x \in \Omega$ since f is a function. Thus $\{f \geq t_1\} \supseteq \{f \geq t_2\}$ due to the definition of upper level sets and $t_1 < t_2$. □

Another important concept, that needs to be reassessed in the context of general measures, is that of null sets. A set N, that cannot be seen by a classical additive measure μ in the sense that it's measure $\mu(N)$ is zero, is called null set[24] in classical measure theory. The definition of a Lebesgue null set is, however, not suitable for capacities as it is possible that there exist measurable sets $A, B \in \mathcal{A}$ (and a super-additive measure ν), such that $\nu(A) = \nu(B) = 0$ but $\nu(A \cup B) > 0$. For capacities this notion can be extended as follows.

4.4 Definition: ([MS91] and [Gra16]) Let ν be a capacity on the measurable space $(\Omega, \mathcal{A})$. A set $N \in \mathcal{A}$ is called a *null set* with respect to ν, if

$$\nu(A \cup N) = \nu(A) \qquad \textit{for all } A \in \mathcal{A}.$$

A null set with respect to a capacity might be interpreted as a set which is invisible and/or unimportant with respect to ν when adding it with *all* measurable sets $A \in \mathcal{A}$. The main properties of null sets with respect to a capacity are given in the following proposition.

4.5 Proposition ([Gra16]): Let ν be a capacity on the measurable space $(\Omega, \mathcal{A})$, then the following holds:

[24] A subset $N \subseteq \mathbb{R}$ is said to have Lebesgue measure 0 or is called a *Lebesgue null set* if, for every $\delta > 0$, N can be covered by a countable number of open intervals the sum of whose length is less than δ. For instance, every countable set is a Lebesgue null set.

(i) The empty set is a null set;

(ii) If N is a null set, then $\nu(N) = 0$;

(iii) If N is a null set, then every $M \in \mathcal{A}$, $M \subseteq N$ is a null set;

(iv) If $\mathcal{A}$ is closed under finite unions, the finite union of null sets is a null set;

(v) If $\mathcal{A}$ is closed under countable unions and if ν is continuous from below, the countable unions of null sets is a null set;

(vi) If ν is sub-additive, every measurable set of capacity zero is a null set.

Proof. (i) By definition $\nu(A \cup \varnothing) = \nu(A)$ $\forall A \in \mathcal{A}$ as well as $\nu(\varnothing) = 0$.

(ii) Let $N \in \mathcal{A}$ be a null set, then $\nu(A \cup N) = \nu(A)$ for all $A \in \mathcal{A}$. Due to the fact that ν is monotone and $A \subseteq A \cup N$ is valid, we get $\nu(A) \leq \nu(A \cup N) \leq \nu(A)$ which implies $\nu(N) = 0$.

(iii) The assertion follows directly from the proof of (ii) since $M \subseteq N$ and hence $\nu(M) \leq \nu(N) = 0$.

(iv) Let $N_1, \ldots, N_k \in \mathcal{A}$ a finite collection of null sets and $N := N_1 \cup \ldots \cup N_k$ the finite union of these null sets. Since N_i with $i \in \{1, \ldots, k\}$ is a null set, $\nu(A \cup N_i) = \nu(A)$ for all $A \in \mathcal{A}$. Note that $A' := N \setminus N_i$ also belongs to $\mathcal{A}$ as $N_1, \ldots, N_k$ are measurable sets and $\mathcal{A}$ is closed under countable unions. Hence, $\nu(A \cup N) = \nu(A \cup N_1 \cup \ldots \cup N_i \cup \ldots \cup N_k) = \nu(A \cup A' \cup N_i) = \nu(A)$ for all $A \in \mathcal{A}$ (incl. $A \cup A'$) and all $i \in \{1, \ldots, k\}$.

(v) The assertion follows directly from the proof of (iv) and the continuity assumption.

(vi) First, let us assume that ν is sub-additive and $\nu(N) = 0$ for $N \in \mathcal{A}$. We have to show that $\nu(A \cup N) = \nu(A)$ for all $A \in \mathcal{A}$. Without loss of generality we assume that $A \cap N = \varnothing$. Due to the sub-additivity of ν, we get $\nu(A \cup N) \leq \nu(A) + \nu(N) = \nu(A)$. Since ν is monotone and $A \subseteq A \cup N$, we get $\nu(A) \leq \nu(A \cup N)$. Hence, $\nu(A \cup N) = \nu(A)$ for an arbitrary $A \in \mathcal{A}$, which proves the assertion.

□

The '*almost everywhere*' concept is strongly related to null sets and might be defined in the same way as in classical measure theory.

4.6 Definition: Let ν be a capacity on a space $(\Omega, \mathcal{A})$, $\mathcal{P}$ be a property that is declared for each $x \in \Omega$. The property $\mathcal{P}$ holds *almost everywhere* (a.e.) if the set for which the property holds takes up nearly all relevant possibilities. More specifically, a property $\mathcal{P}$ holds a.e., if the set of elements for which the property does not hold is a null set N.

4.7 Lemma: Let ν be a capacity on a measurable space $(\Omega, \mathcal{A})$. A measurable set N is a null set if and only if $\nu(A \cap N^C) = \nu(A)$ for all $A \in \mathcal{A}$.

Proof. If N is a null set, then we have $\nu(A \cup N) = \nu((A \cap N^C) \cup N) = \nu((A \cup N) \cap (N^C \cup N)) = \nu(A \cup N) = \nu(A)$ for every $A \in \mathcal{A}$. □

Let us now turn to specific spaces of functions that are important for the definition of the Choquet integral.

4.8 Definition: Let ν be a capacity on the measurable space $(\Omega, \mathcal{A})$. The set of all *bounded $\mathcal{A}$-measurable functions* $f : \Omega \to \mathbb{R}$ is denoted by $B(\mathcal{A})$. Accordingly, the set of all *bounded $\mathcal{A}$-measurable non-negative functions* is denoted by $B^+(\mathcal{A})$.

Both $B(\mathcal{A})$ and $B^+(\mathcal{A})$ show a familiar structure with respect to the standard order on functions as stated in the next proposition.

4.9 Proposition: The set $B(\mathcal{A})$ endowed with the usual order on functions is a lattice. That is, if $f, g \in B(\mathcal{A})$, then $f \wedge g, f \vee g \in B(\mathcal{A})$. If, in addition, $\mathcal{A}$ is a σ-algebra, then $B(\mathcal{A})$ is a vector space.

Proof. Let $f, g \in B(\mathcal{A})$. We only prove that $(f \wedge g)^{-1}(a, b) \in \mathcal{A}$ for any open (possibly unbounded) interval $(a, b) \subseteq \mathbb{R}$. The other cases being similar. For eacht $t \in \mathbb{R}$, the following holds:

$$\begin{aligned}(f \wedge g > t) &= (f > t) \cup (g > t)\\(f \wedge g < t) &= (f < t) \cup (g < t).\end{aligned}$$

Hence,

$$\begin{aligned}(f \wedge g)^{-1}(a, b) &= (f \wedge g > a) \cup (f \wedge g < b)\\&= ((f > a) \cup (g > a)) \cap ((f < b) \cap (g < b)) \in \mathcal{A},\end{aligned}$$

as desired. Finally, the fact that $B(\mathcal{A})$ is a vector space when $\mathcal{A}$ is a σ-algebra is a standard result in measure theory, see, for instance, section 19 of [Bil95]. □

The same statement holds true for the set $B^+(\mathcal{A})$.

Let f be a bounded measurable real-valued function. We can decompose f into its so-called *positive* and *negative parts*, that is,

$$f = f^+ - f^- \qquad \text{with } f^+ = 0 \vee f \text{ and } f^- = (-f)^+.$$

Note that f^+ and f^- are non-negative, bounded and measurable functions, that is, $f^+, f^- \in B^+(\mathcal{A})$. The decomposition of f in positive and negative parts as well as the vector space structure become useful when extending the Choquet integral of a non-negative function, as set out in the next section, to general functions.

4.2 Non-Negative Functions

Now we have all ingredients for the definition of the Choquet integral of a non-negative measurable function. That is, families of upper level sets are used in combination with a Riemann integral to define the Choquet integral of a non-negative function $f : \Omega \to \mathbb{R}$.

4.10 Definition: (Choquet Integral of Non-Negative Function) Let $(\Omega, \mathcal{A})$ be a measurable space, ν a capacity and $f : \Omega \to \mathbb{R}$ a *non-negative* $\mathcal{A}$-measurable function. The *Choquet integral* of a *non-negative function* f with respect to ν is given by

$$(C) \int_\Omega f d\nu := \int_0^\infty \nu(\{f \geq t\})\, dt = \int_0^\infty \nu(\{x \in \Omega | f(x) \geq t\})\, dt, \tag{4.1}$$

where the integral on the right of the equation is a well-defined Riemann integral.

To see why the Riemann integral of the non-negative function f is well defined, first observe that

$$f^{-1}([t,\infty)) = \{f \geq t\} = \{x \in \Omega | f(x) \geq t\} \in \mathcal{A} \qquad \text{for each } t \in \mathbb{R} \tag{4.2}$$

holds true. That is, the generalized inverse of an interval $[t,\infty) \in \mathcal{B}$ can be traced back to a set $A \in \mathcal{A}$ since f is $\mathcal{A}$-$\mathcal{B}$-measurable function. By requiring that $\{f \geq t\}$, the range $\mathbb{R}$ of the measurable function f is divided into sets of non-increasing intervals and thus elements of $\mathcal{B}$. A set $A \in \mathcal{A}$ is assigned to each of the interval sets via (4.2). Note that an interval set can also consist of an union of distinct intervals as shown in b) of Example 4.2. The Borel σ-algebra, on the other hand, is an extension of all (half-open, open or closed) intervals as set out in § 15, [Hal78]. The Borel sets $\mathcal{B}$ can therefore be generated using the family of all (half-open, open or closed) intervals and the extension theorems as explained in chapter III of [Hal78].

A relative of the distribution function, the so-called *survival function* $G_{\nu,f} : \mathbb{R} \to \mathbb{R}$ *of* $f : \Omega \to \mathbb{R}$ with respect to ν, is defined by $G_{\nu,f}(t) := \nu(\{f \geq t\})$ for each $t \in \mathbb{R}$. Note that the definition of $G_{\nu,f}$ differs from the joint survival function (3.10) since f does not need to be defined on the Borel σ-algebra. Using the survival function of the non-negative measurable function f, we can write (4.7) as

$$(C)\int_\Omega f\ d\nu = \int_0^\infty \nu(\{f \geq t\})\ dt = \int_0^\infty G_{\nu,f}(t)\ dt.$$

The familiy $\{f \geq t\}$ with $t \in \mathbb{R}$ is a decreasing series of sets according to Lemma 4.3, i.e., it is a chain[25] with respect to the set inclusion. Since ν is both non-negative and bounded, the function $G_{\nu,f}$ is non-negative, decreasing and with compact support.[26] By standard results on Riemann integration, see Theorem 2.26 in [DSK11], for instance, we conclude that the Riemann integral $\int_0^\infty G_{\nu,f}(t)\ dt$ exists, and so the Choquet integral (4.7) is well defined.

One reason for introducing survival functions $G_{\nu,f}(t)$ is that they make it possible to replace integrals over Ω by integrals over $t \in [0,\infty)$, i.e.,

$$(C)\int_\Omega f d\nu = \int_0^\infty \nu(\{f \geq t\})\ dt = \int_0^\infty \nu(\{x \in \Omega | f(x) \geq t\})\ dt = \int_0^\infty G_{\nu,f}(t)\ dt.$$

This replacement of the integration variable is ultimately based on the Fubini integral theorem as pointed out at page 172 in [Rud87]. A consequence of the same principle is, that the Choquet integral $\int f\ d\nu$ reduces to the standard additive integral when the capacity is additive.

4.11 Proposition: ([MM03]) Let $f \in B^+(\mathcal{A})$ and μ be an additive capacity (i.e. a measure) on a measurable space $(\Omega, \mathcal{A})$. Then

$$(C)\int_\Omega f\ d\mu = \int_\Omega f\ d\mu = \int_0^\infty \mu(\{f \geq t\})\ dt$$

holds true, where $\int_\Omega f\ d\mu$ is the standard integral with respect to an additive capacity.

[25] A collection $\mathcal{C}$ in $\mathcal{A}$ is a chain with respect to the set inclusion if for each $A, B \in \mathcal{C}$ it holds either $A \subseteq B$ or $B \subseteq A$. Throughout we assume that $\varnothing, \Omega \in \mathcal{C}$.

[26] Note that the corresponding distribution function does not have a compact support and is therefore not that suitable for the definition of the Choquet integral.

Proof. Let us fix one arbitrary $x \in \Omega$. We have

$$\int_0^\infty \mathbb{1}_{\{f \geq t\}}(x)\ dt = \int_0^\infty \mathbb{1}_{[0,f(x)]}(t)\ dt = \int_0^{f(x)}\ dt = f(x).$$

Equivalently, $f(x) = \int_0^\infty \mathbb{1}_{\{f \geq t\}}(x)\ d\mathbb{L}$, where $\mathbb{L}$ is the Lebesgue measure on $\mathbb{R}$. For the standard integral $\int_\Omega f\ d\mu$, we can write

$$\begin{aligned}\int_\Omega f\ d\mu &= \int_\Omega \left(\int_0^\infty \mathbb{1}_{\{f \geq t\}}(x)\ d\mathbb{L} \right)\ d\mu \\ &= \int_0^\infty \left(\int_\Omega \mathbb{1}_{\{f \geq t\}}(x)\ d\mu \right)\ d\mathbb{L} \\ &= \int_0^\infty \mu(\{f \geq t\})\ d\mathbb{L} = \int_0^\infty \mu(\{f \geq t\})\ dt\end{aligned}$$

by using the Fubini theorem. This proves the assertion. □

The Choquet integral could have also been defined using strict upper level sets $\{f > t\}$. As stated before, lower level sets $\{f \leq t\}$ and the corresponding distribution functions would lead to a non-compact support, which is inconvenient for the definition of the Choquet integral.

4.12 Proposition: ([MM03]) Let ν be a capacity and f a non-negative function in $B(\mathcal{A})$. Then,

$$\int_0^\infty \nu(\{f \geq t\})\ dt = \int_0^\infty \nu(\{f > t\})\ dt$$

Proof. Set $G_{\nu,f}(t) := \nu(\{f \geq t\})$ for each $t \in \mathbb{R}$. Moreover, set $G'_{\nu,f} := \nu(\{f > t\})$ for each $t \in \mathbb{R}$. The order $\{x \in \Omega | f(x) \geq t + \frac{1}{n}\} \subseteq \{x \in \Omega | f(x) > t\} \subseteq \{x \in \Omega | f(x) \geq t\}$ holds for each $t \in \mathbb{R}$ and for every $n \in \mathbb{N}$ according to Lemma 4.3. Hence, $G_{\nu,f}(t + \frac{1}{n}) \geq G'_{\nu,f}(t) \geq G_{\nu,f}(t)$ for each $t \in \mathbb{R}, n \in \mathbb{N}$. The Riemann-integrable function $G_{\nu,f}$ is continuous except on an at most countable set $D \subset \mathbb{R}$. If $G_{\nu,f}$ is continuous at t, we have $G_{\nu,f}(t) = \lim_{n\to\infty} G_{\nu,f}(t + \frac{1}{n}) \geq G'_{\nu,f}(t) \geq G_{\nu,f}(t)$, so that $G'_{\nu,f}(t) = G_{\nu,f}(t)$ for each $t \in \mathbb{R} \setminus D$. Standard Riemann integration results imply that their integrals are equal. □

Note that a by-product of the last proposition is that the Riemann-integrable survival function $G_{\nu,f}$ is continuous except on an at most countable set $D \subset \mathbb{R}$. Typical examples are 'smoothly continuous' functions or step functions.

In the following examples, we use distorted probability functions $\nu = h \circ \mathbb{L}$, where $\mathbb{L}$ is the Lebesgue measure, to determine the Choquet integral of a (non-)additive measure. Be aware that there are capacities that cannot be represented via distorted probabilities.[27]

4.13 Example: a) Let $\Omega := [0,1]$, $f(x) = x$ for all $x \in \Omega$ and $\mathcal{B} = \mathcal{B}([0,1])$ the class of all Borel sets in $[0,1]$. Let furthermore $\nu : \mathcal{B} \to [0,1]$ be a (distorted) probability measure defined by $\nu([a,b]) = \mathbb{L}([a,b]) = (b - a)$. Here, the Lebesgue measure $\mathbb{L}$ is 'distorted' by the function $h : [0,1] \to [0,1]$ with $h(t) = t$. According to (4.7) the Choquet integral of

[27] For instance, refer to [NT05] for the discrete case

the non-negative function f with respect to ν is

$$
\begin{aligned}
(C)\int_\Omega f d\nu = \int_0^\infty G_\nu(t)\ dt = \int_0^\infty \nu(\{x\in\Omega | f(x)\ge t\})\ dt = \int_0^\infty \nu(\{x\in[0,1] | x\ge t\})\ dt \\
= \int_0^1 \mathbb{L}([t,1])\ dt \ = \int_0^1 (1-t)\ dt \\
= \int_0^1 (1-t)\ dt = (t-\frac{t^2}{2})\ |_0^1 = \frac{1}{2}.
\end{aligned}
$$

Apparently, this Choquet integral is simply the Riemann integral of the function f.

b) Let $\Omega := [0,1]$, $f(x) = x$ for all $x\in\Omega$ and $\mathcal{B} = \mathcal{B}([0,1])$ the class of all Borel sets in $[0,1]$.[28] Let furthermore $\nu : \mathcal{B} \to [0,1]$ be a distorted probability measure defined by $\nu([a,b]) = [\mathbb{L}([a,b])]^2 = (b-a)^2$. Here, the Lebesgue measure $\mathbb{L}$ is distorted by the function $h:[0,1]\to[0,1]$ with $h(t) = t^2$. According to (4.7) the Choquet integral of the non-negative function f with respect to ν is

$$
\begin{aligned}
(C)\int_\Omega f d\nu = \int_0^\infty G_\nu(t)\ dt = \int_0^\infty \nu(\{x\in\Omega | f(x)\ge t\})\ dt = \int_0^\infty \nu(\{x\in[0,1] | x\ge t\})\ dt \\
= \int_0^1 \mathbb{L}([t,1])^2\ dt \ = \int_0^1 (1-t)^2\ dt \\
= \int_0^1 1-2t+t^2\ dt = (t-t^2+\frac{1}{3}t^3)\ |_0^1 = \frac{1}{3}.
\end{aligned}
$$

c) Let $\Omega := [0,1]$, $f(x) = x^2$ for all $x\in\Omega$ and $\mathcal{B} = \mathcal{B}([0,1])$. Let furthermore $\nu : \mathcal{B}\to[0,1]$ be a distorted probability measure defined by $\nu([a,b]) = [\mathbb{L}([a,b])]^3 = (b-a)^3$. Here, the Lebesgue measure $\mathbb{L}$ is distorted by the function $h:[0,1]\to[0,1]$ with $h(t) = t^3$. According to (4.7) and Example 4.2 b), the Choquet integral of the non-negative function f with respect to ν is

$$
\begin{aligned}
(C)\int_\Omega f d\nu = \int_0^\infty \nu(\{f\ge t\})\ dt = \int_0^\infty \nu(\{x\in[0,1] | x^2\ge t\})\ dt \\
= \int_0^1 \mathbb{L}([0,1]\setminus[0,\sqrt{t}))^3\ dt \ = \int_0^1 (1-\sqrt{t})^3\ dt = \frac{1}{10}.
\end{aligned}
$$

We now establish further basic properties of the survival function G_ν. Before that, we introduce the notions of essential supremum and essential infimum.

4.14 Definition: (Essential Supremum and Infimum) For any $f\in B^+(\mathcal{A})$ and any capacity ν on the measurable space $(\Omega,\mathcal{A})$, the *essential supremum* and *essential infimum* of f with respect to ν are defined by

$$
\begin{aligned}
\text{ess sup}_\nu f &:= \inf\{\ t\in\mathbb{R} : \{x\in\Omega |\ f(x) > t\} \text{ is null w.r.t. } \nu\ \} \\
\text{ess inf}_\nu f &:= \sup\{\ t\in\mathbb{R} : \{x\in\Omega |\ f(x) < t\} \text{ is null w.r.t. } \nu\ \}
\end{aligned}
$$

respectively.

[28]Example taken form chapter 11 of [WK09]

4.15 Example: Let us consider the function $f(x) = \begin{cases} x^2 & x \in [1,2) \cup (2,3] \\ 10 & x = 2 \end{cases}$ as sketched in Figure 4.15. Apparently, the supremum of the function f on the support $[1,3]$ is 10, however, the essential supremum is only 9 since $x = \{2\}$ is a null set. The difference between the

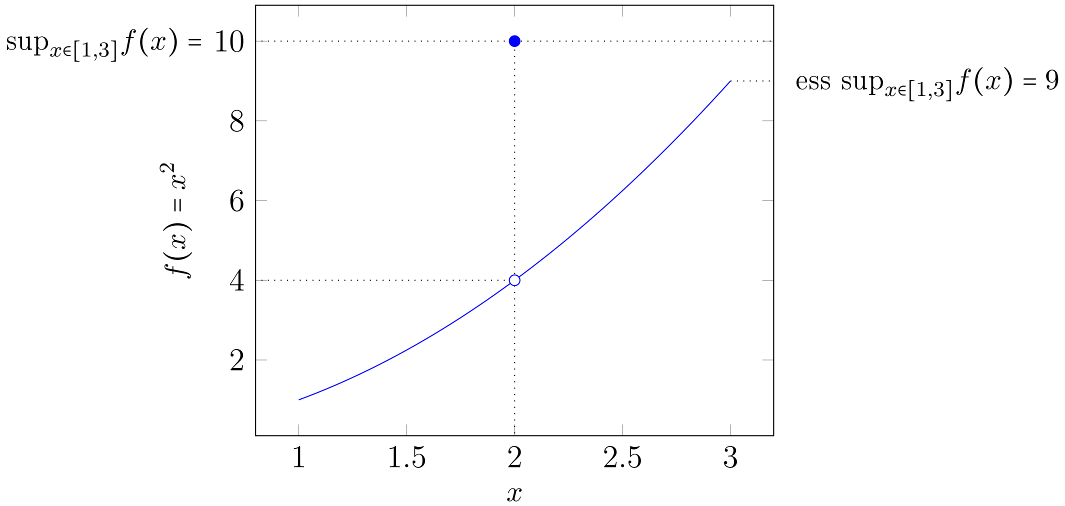

Figure 7: *Supremum and essential supremum of f*

infimum and essential infimum is similar. Please also think about how this would affect the corresponding survival function.

For a more detailed outline of essential supremum and infimum, we refer to section 4.2 in [Gra16]. We can now use both new terms to describe the range of f and the support of the corresponding survival function $G_{\nu,f}$ of a non-necessarily additive set function ν.

4.16 Lemma: ([Gra16]) Let $f \in B^+(\mathcal{A})$ and any capacity ν on the measurable space $(\Omega, \mathcal{A})$. Then, $G_{\nu,f} : \mathbb{R} \to \mathbb{R}$

(i) is a non-negative and non-increasing function with $G_{\nu,f}(0) = \nu(\Omega)$;

(ii) $G_{\nu,f}(t) = \nu(\Omega)$ on the interval $[0, \text{ess inf}_\nu f]$;

(iii) has a compact support, namely $[0, \text{ess sup}_\nu f]$ with $G_{\nu,f}(t) = 0$ for $t > \text{ess sup}_\nu f$.

Proof. (i) Obvious by monotonicity of ν and the fact that $t < t'$ implies $\{f \geq t\} \supseteq \{f \geq t'\}$.

(ii) By definition, $N := \{f < \text{ess inf}_\nu f\}$ is a null set, hence $G_{\nu,f}(\text{ess inf}_\nu f) = \nu(\Omega \setminus N) = \nu(\Omega)$ by (ii) of Lemma 4.5.

(iii) Since f is bounded, so is its essential supremum. Now, by definition $\{x \in \Omega | f(x) > \text{ess sup}_\nu f\}$ is a null set, therefore $G_{\nu,f}(t) = 0$ if $t > \text{ess sup}_\nu f$. □

4.3 General Functions

In the last section 4.2, we have defined the Choquet integral for non-negative $\mathcal{A}$-measurable functions $f \in B^+(\mathcal{A})$. The extension of the Choquet integral to general $\mathcal{A}$-measurable functions

$f \in B(\mathcal{A})$ needs several additional steps. First, we need to identify each element of $\mathcal{A}$ with a specific functional[29]. Second, each non-negative functional in return is identified with its corresponding Choquet integral. Finally, each general function is identified with a functional (i.e. a generalized integral) that fulfills a specific property. Methods from functional analysis will be handy for this purpose.

Let us consider a Choquet integral type, which is by definition additive with respect to its positive and negative part. That is,

$$(C)\int f\,d\nu := (C)\int f^+\,d\nu - (C)\int f^-\,d\nu, \tag{4.3}$$

where f^+ is the positive and f^- the negative part. Note that both $f^+, f^- \in B^+(\mathcal{A})$. Further, for $f \in B(\mathcal{A})$ the symmetry property

$$(C)\int(-f)\,d\nu = (-1)\cdot(C)\int f\,d\nu$$

holds true since $(-f)^+ = \sup\{0,-f\} = f^-$ and $(-f)^- = \sup\{0,f\} = f^+$. In addition, the system of upper level sets $\{f \geq t\}$ lead to the same Riemann integral as the system of lower level sets $\{f \leq t\} = \{(-f) \geq t\}$ (over Ω). The integral, as defined in (4.3), is therefore called *symmetric Choquet integral*.

Let us define two important terms.

4.17 Definition: 1) We say that a functional on $B(\mathcal{A})$ is *translation invariant* (t.i.) if for every $f \in B(\mathcal{A})$ and every $\alpha \in \mathbb{R}$ the equation $\nu_c(f+\alpha\mathbb{1}_\Omega) = \nu_c(f) + \alpha\nu_c(\mathbb{1}_\Omega)$ holds true;

2) Let $(\Omega,\mathcal{A},\nu)$ be a measurable space and $f \in B(\mathcal{A})$. A translation invariant functional $\nu_c : B(\mathcal{A}) \to \mathbb{R}$ defined by $\nu_c(f) := (C)\int_\Omega f\,d\nu$ is called *Choquet functional*.

The following example shows that the symmetric Choquet integral is not translation invariant.

4.18 Example: Let ν be a capacity over the measurable space $(\Omega = \mathbb{R},\mathcal{A})$.[30] Let $f = \operatorname{sgn}(x)$ be the signum function. According to (4.3), the symmetric Choquet integral equals

$$\begin{aligned}(C)\int \operatorname{sgn}\,d\nu =& (C)\int \mathbb{1}_{\mathbb{R}_+}\,d\nu - (C)\int \mathbb{1}_{\mathbb{R}_-}\,d\nu \\ =& \nu(\mathbb{R}_+) - \nu(\mathbb{R}_-),\end{aligned}$$

whereby we have used Lemma 4.19 for the last equation. If we translate (i.e. shift) the function f to get $g := \mathbb{1}_\mathbb{R} + \operatorname{sgn}$ for all $x \in \mathbb{R}$, we have

$$(C)\int \mathbb{1}_\mathbb{R} + \operatorname{sgn}\,d\nu = \int \mathbb{1}_\mathbb{R} + \operatorname{sgn}\,d\nu = 2\nu(\mathbb{R}_+).$$

Note that $\mathbb{1}_\mathbb{R} + \operatorname{sgn} = 0$ for $x \in (-\infty, 0]$.

[29] A *functional* is a mapping of a function $f : \Omega \to \mathbb{R}$ to the value of the function at a point $x \in \Omega$, i.e. $f \mapsto f(x)$ with $x \in \Omega$.

[30] Example taken from section 4.3.1 in [Gra16]

Let us take a step back to consider the overall picture. Each capacity ν on $(\Omega, \mathcal{A})$ induces a functional $\nu_c : B^+(\mathcal{A}) \to \mathbb{R}$ given by $\nu_c(f) = (C)\int f \, d\nu$ for each $f \in B^+(\mathcal{A})$. A special case of this observation, where f is a characteristic function $\mathbb{1}_A$ with $A \in \mathcal{A}$, leads to the following lemma.

4.19 Lemma: Let $\mathbb{1}_A$ be a characteristic function with $A \in \mathcal{A}$. Then, for every capacity ν on $(\Omega, \mathcal{A})$ we have

$$(C)\int_\Omega \mathbb{1}_A \, d\nu = \nu(A).$$

Proof. The functional $\mathbb{1}_A$ is non-negative for any set of its domain and thus belongs to $B^+(\mathcal{A})$. According to the definition of the characteristic function $\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{else} \end{cases}$ and the Choquet integral, we derive

$$(C)\int_\Omega \mathbb{1}_A \, d\nu = \int_0^\infty \nu(\{x \in \Omega | \mathbb{1}_A(x) \geq t\}) \, dt = \int_0^1 \nu(A) \, dt = \nu(A)$$

for all $A \in \mathcal{A}$. Note that for $t = 0$, we get $\{\mathbb{1}_A \geq 0\} = \Omega$ since $\operatorname{ran}(\mathbb{1}_A) = \{0, 1\}$. For $t \in (0, 1]$, we get $\{\mathbb{1}_A \geq t\} = A$. All other disjoint measurable subset of $[0, \infty)$ is (by definition of $\mathbb{1}_A$) mapped to zero. □

The lemma suggests that we can identify each element $A \in \mathcal{A}$ with the functional $\mathbb{1}_A$.

Choquet functionals can be viewed as an extension of the capacity ν from $\mathcal{A}$ to $B^+(\mathcal{A})$. The problem of extending the Choquet integral from $B^+(\mathcal{A})$ to $B(\mathcal{A})$ can be viewed as the problem of how to extend the Choquet functionals from $B^+(\mathcal{A})$ to the entire space $B(\mathcal{A})$. Possible extensions depend on the conditions that we want it to satisfy. A natural property to require is that the extended functional $\nu_c : B(\mathcal{A}) \to \mathbb{R}$ is translation invariant.

$$A \xrightarrow[\mathbb{1}_A]{A \in \mathcal{A}} \mathcal{A} \xrightarrow[(C)\int_\Omega f \, d\nu]{f \in B^+(\mathcal{A})} B^+(\mathcal{A}) \xrightarrow[\text{t.i. } \nu_c(f)]{f \in B(\mathcal{A})} B(\mathcal{A})$$

The symmetric Choquet integral is not translation invariant as pointed out in Example 4.18. However, there exists a well-defined, translation-invariant extension of the Choquet integral to $B(\mathcal{A})$ given by

$$(C)\int f \, d\nu \;=\; (C)\int f^+ \, d\nu - (C)\int f^- \, d\widetilde{\nu}, \tag{4.4}$$

where $\widetilde{\nu}$ is the dual capacity. The integral (4.4) is called asymmetric Choquet integral since in general $(C)\int (-f) \, d\nu \neq (-1) \cdot (C)\int f \, d\nu$ for $f \in B(\mathcal{A})$. A more explicit expression can be received by using the definitions of the applied objects as well as Proposition 4.12 to derive

$$\begin{aligned}
-(C)\int_\Omega f^- \, d\widetilde{\nu} &= -(C)\int_0^\infty \widetilde{\nu}(\{-f \geq s\}) \, ds = -(C)\int_0^\infty \nu(\Omega) - \nu(\{-f \geq s\}^C) \, ds \\
&= (C)\int_0^\infty \nu(\{-f \geq s\}^C) - \nu(\Omega) \, ds = (C)\int_0^\infty \nu(\{-f < s\}) - \nu(\Omega) \, ds \\
&= (C)\int_0^\infty \nu(\{f > -s\}) - \nu(\Omega) \, ds = (C)\int_{-\infty}^0 \nu(\{f > t\}) - \nu(\Omega) \, dt \\
&= (C)\int_{-\infty}^0 \nu(\{f \geq t\}) - \nu(\Omega) \, dt
\end{aligned}$$

with $t := -s$.

The next proposition provides an explicit expression of the desired translation-invariant extension.

4.20 Proposition: Let $(\Omega, \mathcal{A}, \nu)$ be a measurable space and $f \in B(\mathcal{A})$. A Choquet functional $\nu_c : B^+(\mathcal{A}) \to \mathbb{R}$ admits an unique translation invariant extension, given by

$$\nu_c(f) = \int_0^\infty \nu(\{f \geq t\})\ dt + \int_{-\infty}^0 \nu(\{f \geq t\} - \nu(\Omega))\ dt \tag{4.5}$$

for each $f \in B(\mathcal{A})$, where on the right we have two Riemann integrals.

Proof. Let $\nu_c : B(\mathcal{A}) \to \mathbb{R}$ be a functional that is t.i. and coincides with the Choquet integral on $B^+(\mathcal{A})$. It suffices to show that ν_c has the form

$$\widehat{\nu_c}(f) = \int_0^\infty \nu(\{f \geq t\})\ dt + \int_{-\infty}^0 \nu(\{f \geq t\} - \nu(\Omega))\ dt. \tag{4.6}$$

Take $f \in B(\mathcal{A})$ and suppose that $\inf f =: \gamma < 0$. By translation invariance, $\nu_c(f - \gamma \mathbb{1}_\Omega) = \nu_c(f) - \gamma \nu_c(\mathbb{1}_\Omega)$. As $f - \gamma \mathbb{1}_\Omega$ belongs to $B^+(\mathcal{A})$, we can write

$$\begin{aligned}
\nu_c(f) &= \nu_c(f - \gamma) + \gamma \nu_c(\mathbb{1}_\Omega) \\
&= \int_0^\infty \nu(\{f - \gamma \mathbb{1}_\Omega \geq t\})\ dt + \gamma \nu(\Omega) \\
&= \int_\gamma^\infty \nu(\{f \geq \tau\})\ d\tau + \gamma \nu(\Omega) \\
&= \int_\gamma^0 \nu(\{f \geq \tau\})\ d\tau + \int_0^\infty \nu(\{f \geq \tau\})\ d\tau - \int_\gamma^0 \nu(\Omega)\ d\tau
\end{aligned}$$

where the penultimate equality is due to the change of variable $\tau := t + \gamma$. As $[\nu(\{f \geq \tau\}) - \nu(\Omega)] = 0$ for all $\tau \leq \gamma$, the following holds:

$$\nu_c(f) = \int_0^\infty \nu(\{f \geq \tau\})\ d\tau + \int_{-\infty}^0 \nu(\{f \geq \tau\}) - \nu(\Omega)\ d\tau.$$

□

Finally, we can define the general (asymmetric) Choquet integral.

4.21 Definition: Let $(\Omega, \mathcal{A})$ be a measurable space, ν a capacity and $f : \Omega \to \mathbb{R}$ a $\mathcal{A}$-measurable function. The (asymmetric) *Choquet integral* of f with respect to ν is given by

$$(C) \int_\Omega f d\nu := \int_0^\infty \nu(\{f \geq t\})\ dt + \int_{-\infty}^0 \nu(\{f \geq t\} - \nu(\Omega))\ dt, \tag{4.7}$$

where the integrals on the right of the equations are well-defined Riemann integrals.

4.4 Comonotonic Functionals

In general, a Choquet functional $\nu_c = (C) \int_\Omega f\ d\nu$ on a measurable space $(\Omega, \mathcal{A}, \nu)$ is –in general– not additive. That is, the equation

$$\nu_c(f + g) = \nu_c(f) + \nu_c(g)$$

is *in general not* true. However, additivity in a restricted sense holds true if there is a certain relation between f and g. This relation is called comonotonicity and is defined as follows.

4.22 Definition: (Comonotonic functions) Two real functions $f, g : \Omega \to \mathbb{R}$ are called *comonotonic* (short for 'commonly monotonic') if

$$[f(x_1) - f(x_2)][g(x_1) - g(x_2)] \geq 0 \qquad \text{for any } x_1, x_2 \in \Omega.$$

A *class of functions* $\mathcal{C}$ is said to be *comonotonic* if for every pair $(f, g) \in \mathcal{C}^2$ is comonotonic.

A pair of functions f, g is comonotonic, if either

- $f(x_1) \leq f(x_2)$ and $g(x_1) \leq g(x_2)$, or
- $f(x_1) \geq f(x_2)$ and $g(x_1) \geq g(x_2)$

holds for all $x_1, x_2 \in \Omega$. It is definitely *not* comonotonic if there are $x_1, x_2 \in \Omega$, such that $f(x_1) < f(x_2)$ and $g(x_1) > g(x_2)$. However, the negation of this characterization is also necessary and sufficient to characterize comonotonicity as stated in [Gra16]. That is, a pair of function f, g is comonotonic if and only if *there are no* $x_1, x_2 \in \Omega$, such that $f(x_1) < f(x_2)$ and $g(x_1) > g(x_2)$. Please also refer to [MS91].

Grabisch [Gra16] interprets the actual definition of comonotonicity as follows:

> *Roughly speaking, two comonotonic functions have a similar pattern of variation, however one should be careful that comonotonicity is in fact more demanding than simply to be increasing and decreasing on the same domains.*

The desired feature of comonotonic functions is that both functions have similar patterns with respect to the set inclusion. Let us illustrate the definition of comonotonic functions with some examples.

4.23 Example: a) Let $f, g : \mathbb{R} \to \mathbb{R}$ be defined as noted in Figure 8. The pair of function is not comonotonic on $\mathbb{R}$ as for $x_1 = -4$ and $x_2 = 5$, for example, the inequality of the definition does not hold true, even though both functions are increasing and decreasing on the same domain. Note that the function value f is at $x_1 = -4$ smaller than at $x_2 = 5$. However, g is at $x_1 = -4$ greater than at $x_2 = 5$. This is only possible if f and g are not comonotonic.

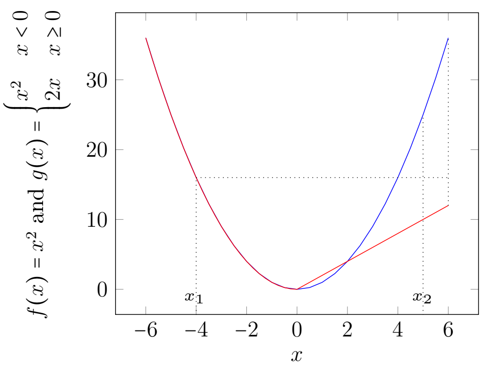

Figure 8: *Graph of non-comonotonic pair of functions f (blue) and g (red)*

b) Consider the real functions $f(x) = |x|$ and $g(x) = x^2$ with $x_1, x_2 \in \mathbb{R}$ as illustrated in Figure 9. If $x_1 \leq x_2$, then $|x_1| \leq |x_2|$ and $x_1^2 \leq x_2^2$. Furthermore, $f(x), g(x) \geq 0$ for all

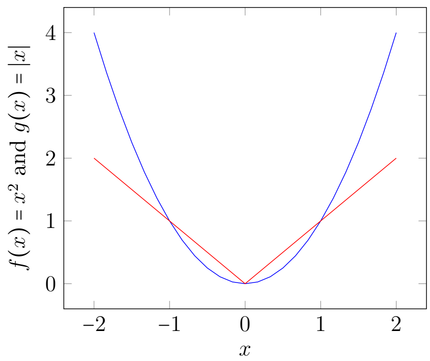

Figure 9: *Graph of comonotonic pair of functions f (blue) and g (red)*

$x \in \mathbb{R}$. Hence, $[f(x_1) - f(x_2)] \geq 0$ as well as $[g(x_1) - g(x_2)] \geq 0$, which proves that f and g are comonotonic.

c) Now, let us consider the real functions $f(x) = x^2$ and $g(x) = 1$ $\forall x \in \mathbb{R}$ as sketched in Figure 10. This pair of function is also comonotonic since $g(x_1) - g(x_2) = 0$ for all $x_1, x_2 \in \mathbb{R}$. Evidently, $[f(x_1) - f(x_2)][g(x_1) - g(x_2)] \geq 0$ holds therefore for all real $x_1, x_2 \in \mathbb{R}$. In general, a constant function is comonotonic with any other function. A

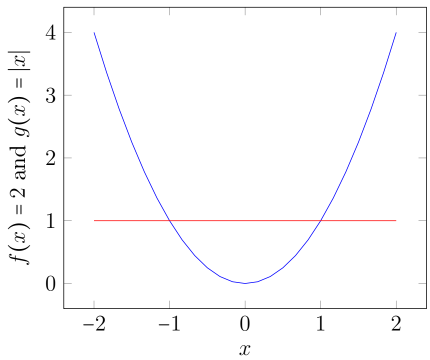

Figure 10: *Graph of comonotonic pair of functions f (blue) and g (red)*

direct consequence of this generalized example is, that the binary relation 'is comonotonic with' is not transitive.

The following lemma is going to characterize comonotonicity by connecting this concept to upper level sets and to the existence of non-decreasing functions.

4.24 Lemma: Let ν be a capacity on a measurable space $(\Omega, \mathcal{A})$. If $f, g : \Omega \to \mathbb{R}$, then the following statements are equivalent.

(i) f and g are comonotonic;

(ii) There exist non-decreasing functions $u, v : \mathbb{R} \to \mathbb{R}$ and a function $h : \Omega \to \mathbb{R}$ such that $f = u \circ h$ and $g = v \circ h$.

(iii) The collection $\{f \geq t\}_{t\in\mathbb{R}} \cup \{g \geq t\}_{t\in\mathbb{R}}$ is a chain.

Moreover, suppose now that Ω is finite, i.e. $|\Omega| = n \in \mathbb{N}$. Then,

(iv) $f = (f_1, \ldots, f_n)$ and $g = (g_1, \ldots, g_n)$ are comonotonic if and only if there exists a permutation σ on X such that $f_{\sigma(1)} \leq \ldots \leq f_{\sigma(n)}$ and $g_{\sigma(1)} \leq \ldots \leq g_{\sigma(n)}$.

Proof. (i) $\Rightarrow$ (ii): If either f, g are constant the result holds trivially. We can therefore assume that f is not constant. Choose any increasing function u on $\mathbb{R}$ and define $h := u^{-1} \circ f$. Then, $f = u \circ (u^{-1} \circ f) = \mathrm{id} \circ f$. Now, define $v : \mathrm{ran}(h) \to \mathbb{R}$ by $v(h(x)) = g(x)$ for every $x \in X$. We still have to show that v is non-decreasing on $\mathrm{ran}(h)$. To this end, take $x_1, x_2 \in \Omega$ such that $h(x_1) > h(x_2)$, which is possible because f is not constant and u is increasing. This is equivalent to $u^{-1}(f(x_1)) > u^{-1}(f(x_2))$, which is in turn equivalent to $f(x_1) > f(x_2)$. Since f, g are comonotonic, it follows that $g(x_1) \geq g(x_2)$.

(ii) $\Rightarrow$ (iii): Take $t_1, t_2 \in \mathbb{R}$ and consider the upper level sets $\{f \geq t_1\}$ and $\{f \geq t_2\}$. We have

$$\{f \geq t_1\} = \{u \circ h \geq t_1\} = \{h \geq s_1\}$$

with $s := \inf\{u^{-1}(t_1)\}$. Similarly, $\{g \geq t_2\} = \{h \geq s_2\}$ with $s_2 := \inf\{v^{-1}(t_2)\}$. These two level sets are then in inclusion relation, which proves the assertion.

(iii) $\Rightarrow$ (i): Suppose there exist $x_1, x_2 \in \Omega$ such that $f(x_1) < f(x_2)$ and $g(x_1) > g(x_2)$. Consider the upper level sets $A := \{f \geq f(x_2)\}$ and $B := \{g \geq g(x_1)\}$. Then, $x_1 \in B \setminus A$ and $x_2 \in A \setminus B$, however, it cannot be that $A \subseteq B$ or $B \subseteq A$.

(iv): This is clear from the definition. $\square$

The following theorem shows that the integral of a pair of comonotonic functions is additive in a restricted sense.

4.25 Theorem: Let $f, g \in B(\mathcal{A})$ be a comonotonic pair of functions such that $f + g \in B(\mathcal{A})$. Then, for any corresponding capacity ν the Choquet integral is comonotonically additive, that is,

$$(C)\int f + g \, d\nu = (C)\int f \, d\nu + (C)\int g \, d\nu. \tag{4.8}$$

Proof. Refer to [Den94] or [MM03], for instance. $\square$

References

[Bil95] Patrick Billingsley. *Probability and Measure*. JOHN WILEY & SONS INC, 1995. 608 pp. ISBN: 0471007102.

[Cho54] Gustave Choquet. 'Theory of capacities'. In: *Annales de l'institute Fourier* 5 (1954), pp. 131–295.

[Dem67] A. P. Dempster. 'Upper and Lower Probabilities induced by a Multivalued Mapping'. In: *The Annals of Mathematical Statistics* 38.2 (1967), pp. 325–339.

[Dem68] A. P. Dempster. 'Upper and Lower Probabilities Generated by a Random Closed Interval'. In: *The Annals of Mathematical Statistics* 39.3 (1968), pp. 957–966.

[Den94] Dieter Denneberg. *Non-Additive Meausres and Integral*. Springer, 31st May 1994. 178 pp. ISBN: 9048144043.

[DP08] Fabrizio Durante and Pierluigi Papini. 'The lattice-theoretic structure of the sets of triangular norms and semi-copulas'. In: *Nonlinear Analysis* (25th Apr. 2008).

[DS10] Fabrizio Durante and Fabio Spizzichino. 'Semi-copulas, capacities and families of level sets'. In: *Fuzzy Sets and Systems* 161 (2010), pp. 269–276.

[DS16] Fabrizio Durante and Carlo Sempi. *Principles of Copula Theory*. Taylor & Francis Inc, 2016. 332 pp. ISBN: 1439884420.

[DSK11] Charles W. Swartz Douglas S. Kurtz. *Theories of Integration*. World Scientific Publishing Company, 2011. 314 pp. ISBN: 9814368997.

[Ell61] Daniel Ellsberg. 'Risk, Ambiguity, and the Savage Axioms'. In: *The Quarterly Journal of Economics* 75.4 (1961), pp. 643–669.

[Fin17] Bruno de Finetti. *Theory of Probability: A critical introductory treatment (Wiley Series in Probability and Statistics)*. Wiley, 2017. ISBN: 1119286379.

[FS04] Hans Follmer and Alexander Schied. *Stochastic Finance: An Introduction In Discrete Time 2 (Degruyter Studies in Mathematics)*. De Gruyter, 2004. ISBN: 9783110183467.

[Gra16] Michel Grabisch. *Set Functions, Games and Capacities in Decision Making*. Springer-Verlag GmbH, 21st June 2016. ISBN: 331930688X.

[Hal78] Paul R. Halmos. *Measure Theory*. Springer New York, 1978. 324 pp. ISBN: 0-387-90088-8.

[Ken74] D. G. Kendall. 'Foundations of a Theory of Random Sets'. In: *Stochastic Geometry*. Wiley (New York), 1974, pp. 322–376.

[Kni09] Frank H. Knight. *Risk, Uncertainty, and Profit*. Signalman Publishing, 2009. ISBN: 0984061428.

[Lin08] Dennis V. Lindley. *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 1, Probability.* CAMBRIDGE UNIV PR, 2008. 272 pp. ISBN: 0521298679.

[Mat74] G Matheron. *Random sets and integral geometry (Wiley series in probability and mathematical statistics).* Wiley, 1974. ISBN: 0471576212.

[MDCQ08] Enrique Miranda, Gert De Cooman and Erik Quaeghebeur. 'Finitely Additive Extensions of Distribution Functions and Moment Sequences: The Coherent Lower Presivision Approach'. In: *International Journal of Approximate Reasoning* 48.1 (Apr. 2008), pp. 132–155.

[MM03] Massimo Marinacci and Luigi Montrucchio. 'Introduction to the Mathematics of Ambiguity'. In: *APPLIED MATHEMATICS WORKING PAPER SERIES* 34/2003 (2003).

[Mol12] Ilya Molchanov. *Theory of Random Sets.* Springer London, 14th Mar. 2012. 504 pp. ISBN: 1849969493.

[MS91] Toshiaki Murofushi and Michio Sugeno. 'A Theory of Fuzzy Measures: Representations, the Choquet Integral, and Null Sets'. In: *Journal of Mathematical Analysis and Applications* 159 (1991), pp. 532–549.

[Nar13] Yasuo Narukawa. 'Integral with Respect to a Non Additive Measure: An Overview'. In: *Non-Additive Measures - Theory and Applications.* Springer, 2013. ISBN: 978-3-319-03154-5.

[Ngu06] Hung T. Nguyen. *An Introduction to Random Sets.* Chapman and Hall/CRC, 2006. ISBN: 158488519X.

[NT05] Yasuo Narukawa and Vicenç Torra. 'Fuzzy measure and probability distributions: distorted probabilities'. In: *IEEE Transactions on Fuzzy Systems* 13.5 (2005), pp. 617–629.

[Oxl11] James G. Oxley. *Matroid Theory.* Oxford University Press, 2011. 704 pp. ISBN: 0199603391.

[PS16] Michael A. Proschan and Pamela A. Shaw. *Essentials of Probability Theory for Statisticians.* Chapman & Hall/Crc Texts in Statistical Science, 2016. ISBN: 978-1498704199.

[Rud87] Walter Rudin. *REAL & COMPLEX ANALYSIS 3E (5P) (Int'l Ed).* McGraw-Hill Education - Europe, 1987. 483 pp. ISBN: 9780071002769.

[Sca96] Marco Scarsini. 'Copulae of Capacities on Product Spaces'. In: *Distributions with Fixed Marginals and Related Topics. IMS Lecture Notes - Monograph Series* 28 (1996).

[Sch86] David Schmeidler. 'Integral Representation without Additivity'. In: *Proceedings of the American Mathematical Society* 97.2 (1986).

[Sha76] Glenn Shafer. *A Mathematical Theory of Evidence.* PRINCETON UNIV PR, 1976. 314 pp. ISBN: 069110042X.

[Sha81] Glenn Shafer. 'Constructive Probability'. In: *Synthese* 48.1 (1981), pp. 1–60.

[Tal15] Nassim Nicholas Taleb. *Silent Risk. TECHNICAL INCERTO: LECTURES NOTES ON PROBABI L I TY, VOL 1.* DesCartes Publishing, 2015.

[WK09] Zhenyuan Wang and George J. Klir. *Generalized Measure Theory.* Springer US, 2009. DOI: 10.1007/978-0-387-76852-6.