Ludwig-Maximilians-University Munich

Master thesis

G-Framework in statistics

Nonlinear expectations for decision-making under uncertainty

Author:
Yusuf Sale

Supervisors:
Prof. Dr. Thomas Augustin,
Prof. Dr. Georg Schollmeyer

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

Faculty of Mathematics, Computer Science and Statistics
Department of Statistics

May 4, 2022

Abstract

In order to achieve reliable results via statistical methodology, one important goal is to account for potential uncertainty. Shige Peng introduced an uncertainty counterpart of Kolmogorov's probabilistic setting: the G-Framework. While this framework is well-known in mathematical finance, work within the G-Framework in statistics is limited. This thesis motivates nonlinear expectations for decision-making under uncertainty in dynamic and non-dynamic situations. Switching the viewpoint from probability spaces to expectation spaces, we discuss the theoretical foundations of the G-Framework, emphasizing comprehensibility. We motivate nonlinear expectations for subsequent application in statistics by notions that emerged in various academic communities and are likewise concerned with decision-making under uncertainty: Choquet expectations express probabilistic uncertainty from the viewpoint of non-additive measures and g-expectations, which represent a nonlinear class of expectations based on backward stochastic differential equations (BSDE). For explicit understanding, we provide the required foundations of stochastic calculus in a self-contained form. The applicability of the G-Framework in statistics is particularly evident from the respective Law of Large Numbers and Central Limit Theorem. To emphasize the applicability, this thesis motivates a notion of sublinear regression.

Contents

Frequently used notation **3**

1 Introduction **5**
1.1 Risk, uncertainty and probability 5
1.2 Towards a statistical theory under uncertainty 12

2 Choquet expectation **13**

3 Backward SDE and related g-expectation **20**
3.1 (Backward) stochastic differential equations 20
3.2 g-expectation 42
3.3 Applications 46

4 Peng's G-Framework **47**
4.1 Nonlinear expectation spaces 47
4.2 (Viscosity) solutions of parabolic PDE 57
4.3 CLT and LLN under sublinear expectation 68
4.4 Remarks on related work 71

5 Sublinear expectation regression **72**

6 Conclusion & Outlook **75**

A Probability Theory **78**

B Functional Analysis **84**

Frequently used notation

Here we provide a list of some frequently used mathematical notation. However, rather specific notation will, in general, not be listed here.

Abbreviations

$a.a.$	Almost all
$a.s.$	Almost sure
$a.e.$	Almost everywhere

Number sets

$\mathbb{C}$	Complex numbers
$\mathbb{N}$	Natural numbers
$\mathbb{R}, \mathbb{R}_{\geq 0}$	Real numbers, non-negative real numbers

Spaces of functions

$\mathcal{C}[0,T]$	Space of continuous, real-valued functions on $[0,T]$
$\mathcal{C}^k([0,T])$	Space of k-times differentiable, continuous, real-valued functions on $[0,T]$
$\mathcal{L}^p_{\mathcal{F}}$	$\mathcal{L}^p(\Omega, \mathcal{F}, P)$
$L^p_{\mathcal{F}}$	$L^p(\Omega, \mathcal{F}, P)$
$L^2(\lambda \otimes P)$	$L^2\left([0,T] \times \Omega, \mathfrak{B}([0,T]) \otimes \mathcal{F}, \lambda \otimes P\right)$
$L^2_{\mathsf{ad}}([0,T])$	Space of measurable, $\mathfrak{F}$-adapted and square integrable processes
$L^2_{0,\mathsf{ad}}([0,T])$	Space of measurable, $\mathfrak{F}$-adapted and square integrable step processes
$C_{l.lip}(\mathbb{R}^n)$	Space of functions ψ satisyfing $$\lvert\psi(x) - \psi(y)\rvert \leq C(1 + \lvert x\rvert^m + \lvert y\rvert^m)\lvert x - y\rvert$$ with $x, y \in \mathbb{R}^n, m \in \mathbb{N}$ and $C > 0$.

Miscellaneous

$\langle \cdot, \cdot \rangle$	Inner product
$\overset{d}{\sim}$	Identically distributed under sublinear expectation

$\|\cdot\|$	Norm
λ	Lebesgue measure
$\lambda \otimes P$	Product measure
$\mathfrak{B}(\mathbb{R}^n)$	Borel σ-algebra on $\mathbb{R}^n$
$\mathfrak{N}$	$\{A \subseteq \Omega : \exists B \in \mathcal{F}, B \subset A, P(B) = 0\}$, collection of P-null sets
$\mathfrak{F}$	Filtration, usually augmented Brownian filtration
$\mathcal{T}$	Stopping time
$\mathbb{E}$	Linear expectation
$\mathcal{E}$	Nonlinear expectation
$\mathcal{E}_c$	Choquet expectation
$\mathcal{E}_g$	g-expectation
X^+	$\max\{0, X\}$, positive part of X
X^-	$\max\{0, -X\}$, negative part of X
$\triangle(\Omega)$	Set of all probability measures on the measurable space $(\Omega, \mathcal{F})$

1 Introduction

1.1 Risk, uncertainty and probability

> *O False and treacherous Probability,*
> *Enemy of truth, and friend to wickednesse;*
> *With those bleare eyes Opinion learns to see*
> *Truth's feeble party here, and barennesse.*
>
> *Keynes (1921, p. 166)*

This first introductory chapter, intentionally named similarly to Frank H. Knight's 1921 work (Knight, 1921) and ironically starting with a poetic quote by John M. Keynes, should sensibilize the reader on the (proper) meaning of uncertainty, especially when it comes to statistics. For this purpose, we will revisit and discuss Knight's most famous work in a pretty philosophical fashion and contrast it with statistical elements, to enhance the importance of dealing with uncertainty in statistical applications. Furthermore, since this thesis is highly concerned with a formal decision-theoretic build-up, the introduction should also allow the familiarization with fundamental decision-theoretic notions, by briefly reviewing the respective development in decision-making under risk and particularly under uncertainty. While Knight's work is mainly engaged with economic theory and how profit can emerge under conditions of market competition, our primary goal is to clarify the notion of uncertainty and derive a possible taxonomy for concepts of uncertainty within the broader frame of statistics. For this quite difficult objective, it is also essential to consider, albeit only partially, the development of probability theory, since the notion of *risk*, *uncertainty*, and *probability* have been entangled in different ways throughout history.

Pierre-Simon Laplace' remarks on probability theory (and uncertainty):

> *On voit par cet Essai, que la théorie des probabilités n'est au fond, que le bon sens réduit au calcul : elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte. Elle ne laisse rien d'arbitraire dans le choix des opinions et des partis à prendre, toutes les fois que l'on peut, à son moyen, déterminer le choix le plus avantageux. Par là, elle devient le supplément le plus heureux à l'ignorance et à la faiblesse de l'esprit humain*[1] (Laplace, 1829, p. 273 ff.).

Laplace's viewpoint is a striclty classical one: "probability theory is only common sense

[1] It is seen in this essay that the theory of probabilities is at bottom only common sense reduced to calculus; it makes us appreciate with exactitude that which exact minds feel by a sort of instinct without being able ofttimes to give a reason for it. It leaves no arbitrariness in the choice of opinions and sides to be taken; and by its use can always be determined the most advantageous choice. Thereby it supllements most happily the ignorance and the weakness of the human mind (Laplace, 2007, p. 196).

reduced to calculus". He points out that one could always determine the most advantageous choice by its means, i.e., the calculations. This nearly 200-year-old viewpoint will lead to a curious contrast with the later development of probability theory (resp. decision-theory). However, at this point, one could ask whether it is *always* possible to determine the most advantageous choice solely through probabilistic "calculations".

Laplace continues:

> *[. . .] si l'on observe ensuite que dans les choses mêmes qui ne peuvent être soumises au calcul, elle donne les aperçus les plus sûrs qui puissent nous guider dans nos jugemens, et qu'elle apprend à se garantir des illusions qui souvent nous égarent, on verra qu'il n'est point de science plus digne de nos méditations, et qu'il soit plus utile de faire entrer dans le système de l'instruction publique*[2] (Laplace, 1829, p. 274).

Although concepts and theories of uncertainty were far from Laplace's lifetime, he gave a crucial hint by remarking that probability theory (in the sense of Laplace) "[. . .] teaches us to avoid the illusions which offtimes confuse us". From a modern standpoint of thinking about uncertainties, we would reformulate and ask where these confusing "illusions" come from and how to deal with them. This should also demonstrate that it is not possible to think in terms of uncertainty without considering probability. A few decades earlier, Daniel Bernoulli had already dealt with determining the "most advantageous choice". In his *Specimen Theoriae novae de Mensura Sortis*[3] 1738, the Swiss mathematician formulates the so-called St. Petersburg Paradox and a proposed solution that is of particular importance for the further development of modern decision-theory. In the St. Petersburg Paradox, a fair coin is tossed until tails fall for the first time. If tails fall on the kth toss, you receive 2^{k-1} euros. Now, the important question arises as to what entry fee one is willing to pay. Since the expected value is infinite, one would expect to be willing to pay fairly high entry fees. However, this is opposed to reality. Daniel Bernoulli suggested a possible way out of this contradictory situation: Instead of striving to maximize the amount of profit, one looks at the benefit an individual achieves through the respective utility. In this context, this is also referred to as the *expected utility model*. Axiomatization did not occur until 1944 with John von Neumann and Oskar Morgenstern's fundamental work (von Neumann and Morgenstern, 1947) on game theory. Since von Neumann and Morgenstern's axiomatization relies on known, objective probabilities, it cannot deal with situations in which probabilities are

[2] If we consider again that, even in the things which cannot be submitted to calculus, it gives the surest hints which can guide us in our judgments, and that it teaches us to avoid the illusions which ofttimes confuse us, then we shall see that there is no science more worthy of our meditations, and that no more useful one could be incorporated in the system of public instruction (Laplace, 2007, p. 196).

[3] (Bernoulli, 1896) is a version translated from Latin. It should also be remarked that Nicolaus Bernoulli already mentioned the paradox in 1713 (de Montmort, 1708, p. 402).

not known (exactly). So as to deal with these situations, Leonard J. Savage axiomatized a subjective counterpart in 1954 (Savage, 1972). Savage's work builds heavily on the von Neumann-Morgenstern expected utility model yet it no longer regards probabilities as given. Instead, the so-called *subjective expected utility model* uses a subjective notion of probability that Bruno de Finetti already treated in 1937 (de Finetti, 1937). Perhaps de Finetti's point of view on probability is best signified in his own words:

> *My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simple this:*
>
> *PROBABILITY DOES NOT EXIST.*
>
> *The abandonment of superstitious beliefs about the existence of Phlogiston, the Cosmic Ether, Absolute Space and Time, . . . , or Fairies and Witches, was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilisitc beliefs* (de Finetti, 1990, p. x).

It is important to note that, while Laplace speaks of the avoidance of illusions through probability theory, de Finetti sees the objective conception of probabilities as the illusion itself. Thus, while objective, unknown probabilities can be replaced with subjective ones in a subjective approach, the concepts of uncertainty and risk per se must be distinguished strictly. Knight was the first, motivated by economic observations, to prominently introduce this distinction:

> *But Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated. [. . .] The essential fact is that "risk" means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomenon depending on which of the two is really present and operating.* (Knight, 1921, p. 19 f.)

Knight postulates to distinguish between uncertainty in the proper sense and risk, that is at bottom explainable with means of probability theory (also compare Laplace's viewpoint) and statistics, such that is no uncertainty at all:

> *It will appear that a measurable uncertainty, or "risk" proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all.* (Knight, 1921, p. 20)

Later on, we will go a step further, especially in terms of statistical applications, and define a taxonomy of uncertainty. Concerning the theory of profit in an economic setting, Knight continues with the following important observation:

> *It is this "true" uncertainty, and not risk, as has been argued, which forms the basis of a valid theory of profit and accounts for the divergence between actual and theoretical competition.* (Knight, 1921, p. 20)

As statisticians, we could look at the above quote differently: Is not the actual profit of statistics hidden in the uncertain, rather than in the certain? Knight's uncertainty-risk distinction is also well-explained by another economic pioneer, John M. Keynes:

> *By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only possible. The game of roulette is not subject, in this sense, to uncertainty; nor is the prospect of a Victory bond being drawn. Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the absolescence of a new invention, or the position of private wealth-owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.* (Keynes, 1937, p. 213 ff.)

In a decision-theoretic approach, we tend to refer to these situations of proper uncertainty as *ambiguity*. So far, not a word has been said about *statistics* itself. Before doing so and deriving in detail a taxonomy of uncertainty concerning statistical applications, we will briefly introduce the so-called *Ellsberg experiment* (and the respective paradox) as a famous example of a situation under ambiguity. Thereafter, we will discuss probable solutions.

Ellsberg experiment:

There are 90 balls in an urn. Of these, 30 are red and the remaining 60 can be black or yellow (the proportions are unknown!). Now, consider the following two bets:

1. A ball is drawn at random. You can bet on {red} or {black} and win 100€ if you are right.
2. A ball is drawn at random. You can bet on {red, yellow} or {black, yellow} and win 100€ if you are right.

Formalized:

Action	$\{r\}$	$\{y\}$	$\{b\}$
f_1 : Bet on $\{r\}$	100	0	0
f_2 : Bet on $\{b\}$	0	0	100
f_3 : Bet on $\{r, y\}$	100	100	0
f_4 : Bet on $\{b, y\}$	0	100	100

Table 1: Setting of the Ellsberg experiment.

As a result of this experiment (Ellsberg, 1961), Daniel Ellsberg observes that, very often, f_1 is preferred over f_2 and f_4 is preferred over f_3, i.e., $f_1 \succ f_2$ and $f_4 \succ f_3$ (at this point, we only consider "$\succ$" as a formalization of the linguistic "is preferred over" and do not take into consideration the proper mathematical structure of the preference relation). We will demonstrate why these preferences lead to a contradiction (= the so-called Ellsberg paradox).

In the sense of expected utility theory, we define, for reasons of simplicity, the following *utility function*:

$$u : \{0; 100\} \to \mathbb{R} : \; x \mapsto u(x) = \begin{cases} 1, & \text{if} \quad x = 100 \\ 0, & \text{if} \quad x = 0 \end{cases} \tag{1.1.1}$$

In order to be consistent with expected utility theory, we expect from above (experimentally) observed preferences

$$\begin{aligned} f_1 \succ f_2 &\Leftrightarrow \mathbb{E}[u \circ f_1] > \mathbb{E}[u \circ f_2] \\ &\Leftrightarrow P(\{r\}) > P(\{b\}) \end{aligned} \tag{1.1.2}$$

$$\begin{aligned} f_4 \succ f_3 &\Leftrightarrow \mathbb{E}[u \circ f_4] > \mathbb{E}[u \circ f_3] \\ &\Leftrightarrow P(\{b\}) > P(\{r\}), \end{aligned} \tag{1.1.3}$$

where P is a (subjective) probability measure on the state space $\mathcal{S} = \{r, b, y\}$. Clearly (1.1.2) and (1.1.3) are contradictory. Hence, it is not possible to find any (subjective) probability measure P to represent the observed preferences. According to Machina and Schmeidler (1992), one speaks of *probabilistic sophistication* if the uncertainty of events can always be described by (classical) probability measures. The outcome of the Ellsberg experiment is that the decision-maker (DM) is not always probabilistically sophisticated. The ambiguity experienced by the DM is due to the unknown proportion $c \in [0, 2/3]$ of black balls in the urn. David Schmeidler, among others, famously dealt with this problem and proposed an expected utility model that does not entail the use of classical probabilities, and considers expected utility in terms of non-additive measures (Schmeidler, 1989). In a later section, *Choquet expectations* will be discussed in more detail. In addition, we show how the above problem can be solved using this new notion of expected utility.

Based on Lo and Mueller (2010), a detailed taxonomy of uncertainty is presented and the degree of uncertainty which this work is mainly concerned is highlighted.

The following table summarizes the different levels of uncertainty, which will now be described in detail and contrasted with previous observations. For an intuitive example from physics, consider Lo and Mueller (2010, p. 14 ff.).

Level	Description	Theoretical framework
1	Complete certainty	Physics: Newton's laws
2	Risk without uncertainty	Kolmogorov's axiomatic setting
3	Fully reducible uncertainty	Statistical inference
4	Partially reducible uncertainty	Partially: Bayesian statistics, ?
5	Irreducible uncertainty	?

Table 2: Taxonomy of uncertainty.

Level 1: Complete certainty

Complete certainty refers to situations in which everything is deterministic, and there is nothing to worry about. This level of uncertainty is suited best for the setting of classical physics with Newton's idealized laws of motion (Lo and Mueller, 2010, p. 10). Generally, we do not encounter such situations of complete certainty in statistics.

Level 2: Risk without uncertainty

We are now in Kolmogorov's axiomatic setting (Kolmogorov, 1933), where uncertainty can be described by way of the respective probabilistic foundations. This level corresponds precisley to Knights' definition of risk. Now, let $X_t : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ be a random variable for all $t \in I$, where I can be interpreted as a time-index. At this level, we would expect the random variables X_t to be identically and independently distributed (abbreviated as i.i.d), e.g.,

$$X_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$$

where σ^2 is assumed to be known.

Level 3: Fully reducible uncertainty

At this level, the methodology of statistical inference comes into play. With sufficiently large amounts of data, one can reduce this type of uncertainty arbitrarily close to Level 2 uncertainty; hence, one speaks of *fully reducible uncertainty*. This would mean that we now have

$$X_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2).$$

Here, we do not know σ^2 exactly, such that we now have to estimate this parameter with tools of statistical inference. It is also crucial that this level requires unknown parameters to be stationary, e.g., it is not possible that σ^2 changes over time.

Level 4: Partially reducible uncertainty

Partially reducible uncertainty refers to situations in which "uncertainty regarding the underlying structures generating the data cannot be reduced to Level 2 uncertainty, even with an infinite amount of data" (Lo and Mueller, 2010, p. 12). Some examples relevant to statistics are (Lo and Mueller, 2010, p. 11):

i) Stochastic or time-varying parameters vary too frequently to be estimated accurately.

ii) Nonlinearities are too complex to be captured by existing models, techniques, and datasets.

iii) Non-stationarities and non-ergocities that render useless the *Law of Large Numbers*, *Central Limit Theorem*, and other methods of statistical inference and approximation.

iv) Dependence on relevant but unknown and unknowable conditioning information.

In light of the previous case, this translates to

$$X_t \sim \mathcal{T} \in \{\text{set of probability measures on } (\Omega', \mathcal{A}')\}$$

namely, X_t could be distributed arbitarly for $t \in I$ within a set of probability measures on $(\Omega', \mathcal{A}')$. This applies also in the sense that there could be more than one model generating the data.

Level 5: Irreducible uncertainty

This is the final level of uncertainty, where we "simply do not know". No meaningful probabilistic reasoning or statistical inference techniques could deal with this kind of uncertainty. "This type of uncertainty is the domain of philosophers and religious leaders, who focus on not only the unknown but the unknowable" (Lo and Mueller, 2010, p. 13).

Our work will focus mainly on Level 4 uncertainty and discuss a more suitable framework to cope with uncertainties regarding underlying structures generating the data. For this framework, one considers nonlinear expectation spaces instead of linear ones.

1.2 Towards a statistical theory under uncertainty

The realm of statistics is to "teach us to avoid illusions", to reduce uncertainty, and discover the unseen. It seems inappropriate to believe that uncertainty, as it appears in different degrees, is simply to be ignored. For this reason, we will discuss methodology allowing us to deal with respective uncertainty. In the case of the Ellsberg experiment, we already saw that by a nonlinear notion of expectation, we were able to model the occurring uncertainty appropriately. In particular, the nonlinearity of the Choquet expectation is induced by a non-additive measure. Choquet expected utility (Schmeidler, 1986, 1989) and the related notion of multiple prior utility (Gilboa and Schmeidler, 1989) are mainly concerned with non-dynamic situations. However, to deal with uncertainty occurring in dynamic situations, recursive multiple priors utilities, representing an intertemporal counterpart to multiple prior utility, were considered in Chen and Epstein (2002). This approach to dynamic modeling is based on a particular class of nonlinear expectations, so-called g-expectations (Peng, 1997). Since g-expectations are defined via particular backward stochastic differential equations (BSDE), before discussing g-expectations in more detail, we first elaborate on the underlying stochastic calculus. In dynamic as well as non-dynamic situations under uncertainty the aforementioned nonlinear expectations motivate a generalization of Kolmogorov's classical framework $(\Omega, \mathcal{F}, P)$ to nonlinear expectation spaces $(\Omega, \mathcal{H}, \mathcal{E})$ providing an unified way of dealing with uncertainty. We will mainly be concerned with sublinear expectation spaces. This framework, based on Peng (2004, 2007a, 2008) amongst further work by the same author, is called G-Framework. Albeit not always in the classical sense, we can state elementary notions such as independence and distribution of a random variable in this nonlinear framework. Since the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) play a crucial role in statistics, it is essential that analogous theorems can be stated in this new framework. Peng (2007b) and Peng (2008) proved versions of the LLN and CLT within the G-Framework. This work aims to illustrate fundamental concepts in this nonlinear framework and to motivate theoretical foundations in more detail. Possible applications in the area of regression models are discussed based on regressions models considered in Lin et al. (2013). Before introducing the basics of stochastic calculus for a more comprehensive understanding of g-expectations, we first consider how Choquet expectations can be used to model the observed preferences in the Ellsberg experiment. These considerations also motivate how concepts from imprecise probabilities may relate here.

2 Choquet expectation

Based on the results of the Ellsberg experiment, the previous section gave way to the observation that classical probability measures are not sufficient to model the underlying uncertainty in some situations. Therefore, this section will motivate non-additive measures, i.e., capacities[4], and revisit *Choquet expectation* using a notion of integration with respect to non-additive measures, first discussed by Choquet (1954). This nonlinear expectation can be seen as a generalization of the classical linear case. In particular, Choquet expectations will be quite interesting for later comparisons in a broader setting, once we derive so-called g-expectations, which are in turn highly motivated through BSDE.

Definition 2.1 (Capacity)
Let $(\Omega, \mathcal{F})$ be a measurable space. The set-function

$$\mu : \mathcal{F} \to [0, \infty[$$

is called capacity, if

C1) $\mu(\emptyset) = 0,\ \mu(\Omega) = 1$

C2) $\forall A, B \in \mathcal{F} :\ A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

Moreover, the set function

$$\mu_d : \mathcal{F} \to [0, \infty[:\ A \mapsto 1 - \mu(A^c)$$

is called μ-dual capacity. Then $(\Omega, \mathcal{F}, \mu)$ is referred to as capacity space and respectively $(\Omega, \mathcal{F}, \mu_d)$ is called μ-dual capacity space.

Now we call a capacity μ

(i) 2-alternating, if for all $A, B \in \mathcal{F}$ we have

$$\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B).$$

(ii) 2-monotone, if for all $A, B \in \mathcal{F}$ we have

$$\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B).$$

Integration with respect to capacities requires a new notion of integration, since the usual integral in the Lebesgue sense is not suitable due to the lack of additivity. The

[4] The notion *capacity* dates back to Choquet (1954). In addition, be warned, as the terms capacity and non-additive measure are not always used consistently. Nevertheless, these two terms should be treated as synonyms for our purposes.

standard reference for integration with respect to non-additive measures is Denneberg (2013). At this point, we will only formulate the results necessary to motivate Choquet expectations. For some earlier work on Choquet Integration and the role in expected utility theory, consider Sale (2021) and the references therein.

Definition 2.2 (Choquet expectation)
Let $(\Omega, \mathcal{F}, \mu)$ be a capacity space and $X : \Omega \to \mathbb{R}$ a bounded, measurable function. Then, the Choquet expectation of X with respect to the capacity μ is defined by

$$\mathcal{E}_c[X] \coloneqq \int_0^\infty X^+ d\mu - \int X^- d\mu_d \tag{2.0.1}$$

$$= \int_0^\infty \mu(X \geq x)\, dx + \int_{-\infty}^0 (\mu(X \geq x) - 1)\, dx, \tag{2.0.2}$$

where we have $X^+ = \max\{0, X\}$ and $X^- = \max\{0, -X\}$. The integrals involved in (2.0.1) *are called Choquet Integrals.*

In particular, the form of (2.0.1) illustrates the relationship between the notions of expectation and capacity. This relationship also holds true for classical probability measures and linear expectations:

Proposition 2.1
Let X be a random variable on the probability space $(\Omega, \mathcal{F}, P)$. Thus, the (linear) expectation can be represented by

$$\mathbb{E}[X] = \int_0^\infty P(X \geq x)\, dx + \int_{-\infty}^0 (P(X \geq x) - 1)\, dx. \tag{2.0.3}$$

Proof. First, we note that any non-negative random variable can be written in integral form (also referred to as layer-cake representation[5]) by

$$X = \int_0^\infty \mathbf{1}_{\{X > x\}} dx. \tag{2.0.4}$$

Taking the expectation and changing the order of integration (by applying Fubini-Tonelli) yields

$$\begin{aligned}
\mathbb{E}[X] &= \int_\Omega \int_0^\infty \mathbf{1}_{\{X>x\}}\, dx\, dP \\
&= \int_0^\infty \int_\Omega \mathbf{1}_{\{X>x\}}\, dP\, dx \\
&= \int_0^\infty P(X \geq x) dx.
\end{aligned}$$

[5] See Lieb and Loss (2001) for the layer-cake representation.

Since we can write any random variable as $X = X^+ - X^-$ we get

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X^+] - \mathbb{E}[X^-] \\ &= \int_0^\infty P(X^+ \geq x)dx - \int_0^\infty P(X^- > x)dx \\ &= \int_0^\infty P(X \geq x)dx - \int_{-\infty}^0 P(X < x)dx \\ &= \int_0^\infty P(X \geq x)dx + \int_{-\infty}^0 \left(P(X \geq x) - 1\right) dx.\end{aligned}$$

This completes the proof. □

Theorem 2.1 (Properties of Choquet expectation)
Let X, Y be two bounded, measurable functions and $(\Omega, \mathcal{F}, \mu)$ a capacity space. Then, the Choquet expectation has the following properties:

1.) $\forall A \in \mathcal{F} : \mathcal{E}_c[\mathbb{1}_A] = \mu(A)$

2.) $\forall \alpha > 0 : \mathcal{E}_c[\alpha X] = \alpha\, \mathcal{E}_c[X]$

3.) $X \leq Y \Rightarrow \mathcal{E}_c[X] \leq \mathcal{E}_c[Y]$

4.) If μ is a 2-alternating capacity then the respective Choquet expectation is subadditive, i.e.

$$\mathcal{E}_c[X + Y] \leq \mathcal{E}_c[X] + \mathcal{E}_c[Y].$$

5.) If μ is a 2-monotone capacity then the respective Choquet expectation is superadditive, i.e.

$$\mathcal{E}_c[X + Y] \geq \mathcal{E}_c[X] + \mathcal{E}_c[Y].$$

The proof and further properties of the Choquet expectation can be found in Denneberg (2013, p. 64 ff.). Given the properties, it is easy to see that the Choquet expectation is a nonlinear functional. In particular, we will consider functionals satisfying positive homegeneity and subadditvity in later sections. The following theorem shows how subadditivity resp. superadditivity depends on the properties of the underlying capacity.

Theorem 2.2
Let X, Y be two random variables on the a capacity space $(\Omega, 2^\Omega, \mu)$. Then we have:

i.) μ is 2-alternating, if and only if

$$\mathcal{E}_c[X + Y] \leq \mathcal{E}_c[X] + \mathcal{E}_c[Y].$$

ii.) μ *is 2-monotone, if and only if*

$$\mathcal{E}_c[X+Y] \geq \mathcal{E}_c[X] + \mathcal{E}_c[Y].$$

See Denneberg (2013, p. 71) for proof of the above, so-called *Subadditivity-* resp. *Superadditivity theorem*. Note that this theorem was proven many times under different assumptions.

Remark 2.1
From (2.0.2)*, it should be clear that the Choquet expectation can not be additive until the underlying capacity is additive (in this case the Choquet expectation simply reduces to the classical expectation w.r.t. a probability measure), e.g., let* $(\Omega, \mathcal{F}, \mu)$ *be a capacity space, thus for all* $A, B \subseteq \Omega$ *with* $A \cap B = \emptyset$ *we have*

$$\mathcal{E}_c[\mathbf{1}_A + \mathbf{1}_B] = \mathcal{E}_c[\mathbf{1}_{A\cup B}] = \mu(A \cup B) \neq \mu(A) + \mu(B) = \mathcal{E}_c[\mathbf{1}_A] + \mathcal{E}_c[\mathbf{1}_B].$$

Furthermore, it is noteworthy that the Choquet expectation is comonoton additive, which is an important feature for application in mathematical finance. We will not elaborate on this point, hence the reader may refer to the above references.

Ellsberg experiment revisited:
We now revist the Ellsberg experiment and consider the following capacity, instead of a classical probability measure to model the observed preferences:

Event E	$\{r\}$	$\{y\}$	$\{b\}$	$\{r,y\}$	$\{r,b\}$	$\{y,b\}$	$\{r,b,y\}$
$\mu(E)$	$\frac{1}{3}$	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	1

Table 3: Capacity μ

Calculating the resp. Choquet expectations yields

$$\begin{aligned} f_1 \succ f_2 &\Leftrightarrow \mathcal{E}_c[u \circ f_1] > \mathcal{E}_c[u \circ f_2] \\ &\Leftrightarrow \mu(\{r\}) > \mu(\{b\}) \\ &\Leftrightarrow \frac{1}{3} > 0 \end{aligned}$$

$$\begin{aligned} f_4 \succ f_3 &\Leftrightarrow \mathcal{E}_c[u \circ f_4] > \mathcal{E}_c[u \circ f_3] \\ &\Leftrightarrow \mu(\{y,b\}) > \mu(\{r,y\}) \\ &\Leftrightarrow \frac{2}{3} > \frac{1}{3} \end{aligned}$$

Thus, with a suitable capacity specification, it is possible to model the observed preferences in the Ellsberg experiment in terms of expected utility. The so-called *Choquet-expected utility* was axiomatized by Schmeidler (1989), Gilboa (1987) and Sarin and Wakker (1992). A helpful overview can be found in Gilboa (2009) and references therein. Having established Choquet expectations as a valuable tool for decision-making under uncertainty, we will now present a related approach, *maximin-expected utlity*, also called *multiple prior utility* (Gilboa and Schmeidler, 1989). To do so, we revisit the setting of the Ellsberg experiment and consider a different approach to circumvent the observed contradiction by considering the following set of probability measures

$$\mathfrak{C}_\mu \coloneqq \{P \in \triangle(\mathcal{S}) : \, P(\{r\}) = 1/3\}\,, \tag{2.0.5}$$

where $\triangle(\mathcal{S})$ denotes the set of all probability measures on the state space $\mathcal{S} = \{r, b, y\}$. The set of probability measures (2.0.5) is also called *Credal-set*. Instead of evaluating acts w.r.t. the classical expectation or Choquet expectation, this approach makes possible the use of *lower expectation* to model uncertainty appropriately.

Definition 2.3 (Lower and upper expectation)
Let $\mathcal{M} \subseteq \triangle(\mathcal{S})$ be a closed, and convex credal set on the state space $\mathcal{S}$. Thus,

$$\underline{\mathbb{E}}_P[f] = \min_{P \in \mathcal{M}} \int f \, dP \tag{2.0.6}$$

$$\bar{\mathbb{E}}_P[f] = \max_{P \in \mathcal{M}} \int f \, dP \tag{2.0.7}$$

is called lower resp. upper expectation of (suitable) f.

Assuming the "worst case" (acting ambiguity averse), we now consider the following calculations:

$$\underline{\mathbb{E}}[u(f_1)] = \min_{P \in \mathfrak{C}_\mu} \int (u \circ f_1) dP = \frac{1}{3} \tag{2.0.8}$$

$$\underline{\mathbb{E}}[u(f_2)] = \min_{P \in \mathfrak{C}_\mu} \int (u \circ f_2) dP = \min_{c \in [0,2/3]} c = 0 \tag{2.0.9}$$

$$\underline{\mathbb{E}}[u(f_3)] = \min_{P \in \mathfrak{C}_\mu} \int (u \circ f_3) dP = \min_{c \in [0,2/3]} (1 - c) = \frac{1}{3} \tag{2.0.10}$$

$$\underline{\mathbb{E}}[u(f_4)] = \min_{P \in \mathfrak{C}_\mu} \int (u \circ f_4) dP = \frac{2}{3} \tag{2.0.11}$$

Hence, this allows for appropriate modeling of the observed preferences in the Ellsberg experiment. In particular, note that both approaches coincide here, although this is generally not true. See also the remark at the end of this section. The following theorem shows the conditions under which the two approaches coincide.

Theorem 2.3
Let $(\Omega, \mathcal{F}, \mu)$ be a capacity space and

$$\mathcal{C}_\mu = \{P \in \triangle(\Omega) : \ P(E) \geq \mu(E) \quad \textit{for all} \quad E \in \mathcal{F}\}$$

the anti-core[6] *of the capacity μ. Hence, for all μ- and μ_d-integrable functions f, the following statements are equivalent:*

i) μ is 2-monoton and continuous from below.

ii) $\mathcal{E}_c[f] = \min_{P \in \mathcal{C}_\mu} \int f \, dP$.

This theorem was proven under different conditions in Huber and Strassen (1973) or Schmeidler (1986). Further details and a complete proof can be found in Dyckerhoff (1994, p. 55 ff.).

Remark 2.2
Note that we have $\mu(E) = \inf_{P \in \mathfrak{C}_\mu} P(E)$ for all events E, where μ is the capacity defined in Table 3. Since μ is 2-monoton, the preceding theorem tells us that we have the equality

$$\mathcal{E}_c[u \circ f] = \min_{P \in \mathfrak{C}_\mu} \int u \circ f \, dP$$

for all acts f and utility functions u. Hence, Choquet expectation and the lower expectation coincide in this case.

In order not to cause a deceptive impression, the next example shows that, in general the *lower envelope* of a credal set is not 2-monotone.

Example 2.1
Let $\Omega = \{1, 2, 3, 4\}$. Then $P = (0.5, 0.2, 0.2, 0.1)$ and $Q = (0.6, 0.1, 0.1, 0.2)$ are two probability measures on $(\Omega, 2^\Omega)$. Now, we define $\mathfrak{C}$ to be the convex hull of P_0 and P_1. Hence, by taking the lower envelope of this credal set, we define a capacity μ, i.e.,

$$\forall A \in 2^\Omega : \ \mu(A) = \inf_{P \in \mathfrak{C}} P(A). \tag{2.0.12}$$

With $A = \{1, 2\}$ and $B = \{1, 3\}$ we get

$$\mu(A \cup B) + \mu(A \cap B) = 0.8 + 0.5 < 0.7 + 0.7 = \mu(A) + \mu(B).$$

Thus, μ is not 2-monotone and therefore Theorem 2.3 no longer holds true.

Furthermore, even if the capacity μ is 2-montone, it may happen that the underlying credal set is not the anticore of the capacity. Consider the following example:

[6] (Anti-)core is a game-theoretic term, but appears in the literature under many different names.

Example 2.2
Let $\Omega = \{1, 2, 3\}$. *Then* $P = (0.5, 0.5, 0)$ *and* $Q = \left(\frac{4}{6}, \frac{1}{6}, \frac{1}{6}\right)$ *are two probability measures on* $(\Omega, 2^{\Omega})$. *Now, we define* $\mathfrak{C}$ *to be the convex hull of* P_0 *and* P_1, *i.e.,*

$$\mathfrak{C} = \left\{ \left(\frac{3+\lambda}{6}, \frac{3-2\lambda}{6}, \frac{\lambda}{6} \right) \Big| \lambda \in [0,1] \right\}.$$

If we now define a capacity analogous to (2.0.12), *it is easy to see that in this case* μ *is 2-monotone. However, we see that only* $\mathfrak{C} \subsetneq \mathfrak{C}_\mu$ *applies, where*

$$\mathfrak{C}_\mu = \left\{ \left(\frac{3+t}{6}, \frac{3-s-t}{6}, \frac{s}{6} \right) \Big| s, t \in [0,1] \right\}.$$

denotes the anti-core of the capacity μ.

Example 2.1 and Example 2.2 can be found in Huber and Strassen (1973, p. 254) or Dyckerhoff (1994, p. 56).

3 Backward SDE and related g-expectation

One first step to the broader framework of the G-Framework is to revisit Peng's so-called g-expectation, first discussed in Peng (1997). In particular, we note that this nonlinear, filter-consistent expectation is a solution to backward stochastic differential equations (BSDE). For this purpose, we will introduce the corresponding basics of stochastic differential equations (SDE) and the respective Itô's calculus before discussing g-expectations in detail. For a more detailed introduction into stochastic calculus, the reader may refer to Oksendal (2013), Steele (2001) or Karatzas and Shreve (2012). Since this procedure requires some prior knowledge on probability theory and stochastic processes, the reader not familiar with these foundations may refer to the respective Appendix.

3.1 (Backward) stochastic differential equations

To familiarize with the notion and conceptual idea of SDE, we will define ordinary differential equations (ODE) first. An ordinary differential equation relates a (real- or complex-valued) function and its derivatives.

Definition 3.1 (Differential equation)
For $U \subset \mathbb{R} \times \mathbb{R}^{d(k+1)}$ and $F : U \to \mathbb{R}^d$ the equation

$$F\left(t, x, x^{(1)}, ..., x^{(k)}\right) = 0 \tag{3.1.1}$$

with $d, k \in \mathbb{N}$ is called real-valued, d-dimensional implicit differential equation of order k.

A function $\lambda : I \to \mathbb{R}^d$ is called solution of the differential equation (3.1.1)*, if the following holds:*

1. *$I \subset \mathbb{R}$ has non-empty interior.*
2. *$\lambda \in \mathcal{C}^k(I)$, where $\mathcal{C}^k(I)$ denotes the space of k-times differentiable, continuous, real-valued functions on I.*
3. *$\forall t \in I$:*
   - *i.) $\left(t, \lambda(t), \lambda^{(1)}(t), \ldots, \lambda^{(k)}(t)\right) \in U$*
   - *ii.) $F\left(t, \lambda(t), \lambda^{(1)}(t), \ldots, \lambda^{(k)}(t)\right) = 0$*

Furthermore, for $V \subset \mathbb{R} \times \mathbb{R}^{dk}$ and $F : U \to \mathbb{R}^d$ the equation

$$x^{(k)} = f\left(t, x, x^{(1)}, ..., x^{(k-1)}\right) \tag{3.1.2}$$

with $d, k \in \mathbb{N}$ is called real-valued, d-dimensional explicit differential equation of order k.

Notation:

$$x^{(1)}(t) = \frac{dx}{dt}(t), \ldots, x^{(d)}(t) = \frac{d^d x}{dt^d}(t)$$

A differential equation is called

(i) scalar, if $d = 1$.

(ii) autonom, if $F\left(t, x, x^{(1)}, ..., x^{(k)}\right) = F\left(x, x^{(1)}, ..., x^{(k)}\right)$.

(iii) linear, if the mapping

$$\mathbb{R}^{d(k+1)} \to \mathbb{R}^d : X \mapsto F(t, X)$$

is linear.

Example 3.1 (Population growth model)
Consider the population growth model

$$\frac{dN_t}{dt} = \alpha N_t \tag{3.1.3}$$

with $N_t \coloneqq N(t)$ and $N_0 \coloneqq N(0) =$ constant, where N_t denotes the population size at time t, and α a constant growth rate. It is easy to verify that the function

$$N_t = N_0 e^{\alpha t}$$

solves the linear differential equation (3.1.3).

Given the definition of a differential equation, we introduce a stochastic version by allowing randomness for some of the coefficients of the differential equation. With the aid of the population growth model, we will now motivate stochastic differential equations by formulating a stochastic counterpart of (3.1.3). Given previous considerations in the introductory chapter, we now add a stochastic component to the constant growth rate, i.e., we define

$$\alpha_t \coloneqq r_t + \alpha V_t,$$

where V_t denotes, some noise term, yet to be specified. Hence, we call

$$\frac{dN_t}{dt} = r_t N_t + \alpha N_t V_t \tag{3.1.4}$$

stochastic differential equation. Since the way in which to derive a solution for such SDEs is not apparent, the rest of this introductory is concerned mainly with answering the following important questions:

1.) How can the noise term be specified appropriately?

2.) How can we solve such SDEs, or more, generally speaking, equations of the form

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)V_t \tag{3.1.5}$$

adequately?

To answer the first question, it would be intuitive to find a stochastic process $\{V_t\}_{t\in I}$ with some parameter set I yet to be specified, representing the noise term in (3.1.5) with some suitable properties:

(i) $\forall t_i, t_j \in I,\ i \neq j : V_{t_i}$ and V_{t_j} are independent.

(ii) $\{V_t\}_{t\in I}$ is stationary.

(iii) $\forall t \in I : \mathbb{E}(V_t) = 0$.

For I at most countable, any sequence of independent and identically distributed, centered random variables will satisfy the properties (i) - (iii). In case where I is uncountable, usually, we assume in the following, $I = \mathbb{R}_{\geq 0}$ there does not exist a stochastic process satisfying the above properties, since such a stochastic process can not have continuous paths:

Theorem 3.1
Let $(\Omega, \mathcal{F}, P, \{V_t\}_{t\in I})$ be a non-trivial stochastic process[7] *with state space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ satisyfing the above properties, where $\mathfrak{B}(\mathbb{R})$ denotes the Borel σ-algebra on $\mathbb{R}$. Consequently, $\{V_t\}_{t\in I}$ does not have continuous paths.*

Proof. Consider a non-trivial stochastic process $(\Omega, \mathcal{F}, P, \{V_t\}_{t\in I})$ with state space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ satisyfing the above properties. Assume that $\{V_t\}$ has continuous paths, i.e.,

$$\forall \omega \in \Omega : t \mapsto V_t(\omega)$$

is a continuous mapping. For all $N \in \mathbb{N}$ we now define $V_t^{(N)} := \max\{-N, \min\{N, V_t\}\}$. From the assumed continuity we know $\lim_{s\to t}\left(V_t^{(N)} - V_s^{(N)}\right)^2 = 0$ for fixed $t \in I$. Since, by definition, we also have $\forall t \in I : \left(V_t^{(N)} - V_s^{(N)}\right)^2 \leq 4N^2$ for fixed $N \in \mathbb{N}$, applying the dominated convegence theorem yields

$$\lim_{s\to t}\mathbb{E}\left[\left(V_t^{(N)} - V_s^{(N)}\right)^2\right] = \lim_{s\to t}\int\left(V_t^{(N)} - V_s^{(N)}\right)^2 dP$$

[7] In the following we will denote a stochastic process by $\{V\}_{t\in I}$ and omit the quadruple $(\Omega, \mathcal{F}, P, \{V_t\}_{t\in I})$ notation. If there is no possibility of confusion, we also drop the parameter set I in our notation.

$$= \int \lim_{s \to t} \left(V_t^{(N)} - V_s^{(N)} \right)^2 dP$$
$$= 0.$$

On the other hand we derive by independence and stationarity

$$\begin{aligned}
\mathbb{E}\left[\left(V_t^{(N)} - V_s^{(N)} \right)^2 \right] &= \mathbb{E}\left[{V_t^{(N)}}^2 \right] - 2\mathbb{E}\left[V_t^{(N)} \right] \cdot \mathbb{E}\left[V_s^{(N)} \right] + \mathbb{E}\left[{V_s^{(N)}}^2 \right] \\
&= 2\mathbb{E}\left[{V_t^{(N)}}^2 \right] \\
&= 2 \operatorname{Var}\left[V_t^{(N)} \right].
\end{aligned}$$

Since $\operatorname{Var}\left[V_t^{(N)} \right] = 0$ if, and only if $\mathbb{E}\left[V_t^{(N)} \right] = V_t^{(N)}$ $a.s.$, we conclude $V_t^{(N)}$ is degenerate, i.e., V_t is degenerate. By assumption (iii) we get $V_t \equiv 0$. This contradicts the non-triviality assumption. □

Actually, a much stronger statement holds: If we assume mutual independence and $\mathbb{E}\left[V_t^2\right] = 1$, there is no measurable stochastic process that satisfies the above properties. For the precise statement and proof, we refer to Kallianpur (2013, p. 10 ff.). Nevertheless, it is possible to construct a stochastic process satisfying the above properties, called "white noise process", by introducing *generalized functions* in terms of Distribution Theory. However, since this approach is not within the scope of this thesis, we will primarily deal with formally replacing the noise term with a suitable stochastic process, as suggested in Oksendal (2013).

To answer the second question, we will now introduce the basics of the related Itô's calculus[8]. Here, we primarily follow Oksendal (2013) and Steele (2001). So as to avoid notational difficulties, we further define $X_k \coloneqq X_{t_k}$ and $\Delta t_k = t_{k+1} - t_k$, where we set $0 = t_0 < t_1 < t_2 < \cdots < t_m = t$. If we now consider a discrete version of (3.1.5), we get

$$X_k - X_{k+1} = b(t_k, X_k) \cdot \Delta t_k + \sigma(t_k, X_k) V_k \cdot \Delta t_k. \tag{3.1.6}$$

From $X_k = X_0 + \sum_{j=0}^{k-1} (X_{k+1} - X_k)$ and (3.1.6), we obtain

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j) \cdot \Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j) \cdot \Delta B_j, \tag{3.1.7}$$

where, based on previous considerations, we set $V_j \cdot \Delta t_j = \Delta B_j$ with B_j denoting standard Brownian motion.

[8]Note: Since access to general stochastic integrals via Semimartingales is technically quite challenging, only Itô's approach is discussed in this thesis. For a mathematically rigorous approach via Semimartingales consider Karatzas and Shreve (2012).

Remark 3.1
Since the Brownian motion has almost surely nowhere differentiable paths, the above replacement would not make much sense. Therefore, the expression $\frac{dB}{dt}$ should be understood as the distributional derivative of the Brownian motion in the aforementioned sense of Distribution Theory.

Thus (3.1.7) suggests for $\Delta t_j \to 0$:

$$\boxed{X_k = X_0 + \int_0^t b(s, X_s)ds + \underbrace{\int_0^t \sigma(s, X_s)dB_s}_{(1)}} \tag{3.1.8}$$

Since the unbounded variation of Brownian motion makes it unfeasible to treat integrals of the form (1) as ordinary Stieltjes Integrals, we will now focus on the existence and calculation of stochastic integrals. However, before defining a suitable class of integrands to define the Itô Integral, we fix some notation for later convenience:

Let $(\Omega, \mathcal{F}, P)$ be a measure space, then we define for $1 \leq p < \infty$:

- $\mathcal{L}^p_{\mathcal{F}} \coloneqq \mathcal{L}^p(\Omega, \mathcal{F}, P) = \{f : \Omega \to \mathbb{R} \quad \text{is measurable, } \int_\Omega |f|^p dP < \infty\}$ and

  $$\|\cdot\|_{\mathcal{L}^p} : \mathcal{L}^p_{\mathcal{F}} \to \mathbb{R} : \ f \mapsto \|f\|_{\mathcal{L}^p} \coloneqq \left(\int_\Omega |f|^p dP\right)^{1/p},$$

  where $\|f\|_{\mathcal{L}^p}$ is a seminorm on $\mathcal{L}^p_{\mathcal{F}}$.
- $L^p_{\mathcal{F}} \coloneqq L^p(\Omega, \mathcal{F}, P) = \mathcal{L}^p/\mathcal{N}$ with $\mathcal{N} \coloneqq \{f : \ f = 0\, P - a.e.\}$, where $\|f\|_{L^p} \coloneqq \|f\|_{\mathcal{L}^p}$ is a norm on $L^p_{\mathcal{F}}$. Henceforth, $f \in L^p_{\mathcal{F}}$ shall refer to the respective representant in $\mathcal{L}^p_{\mathcal{F}}$ than the whole equivalence class.
- $L^2_{\mathcal{F}}$ equipped with the inner product

  $$\langle f, g\rangle \coloneqq \int_\Omega f \cdot g \, dP$$

  is a Hilbert space.

Definition 3.2
Let $\{B_t\}_{t\geq 0}$ be a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathfrak{F} \coloneqq \{\mathcal{F}_t\}_{t\in[0,T]}$ the respective augmented filtration with $\mathcal{F}_t \coloneqq \sigma\{B_s \cup \mathfrak{N} : \ 0 \leq s \leq t\}$, where $\mathfrak{N} \coloneqq \{A \subseteq \Omega |\ \exists B \in \mathcal{F}, B \subset A, P(B) = 0\}$ denotes the collection of P-null sets. Define $L^2_{ad}([0,T])$ to be the class of functions

$$f(t, \omega) : \ [0, \infty) \times \Omega \to \mathbb{R}$$

such that

(i) $(t,\omega) \to f(t,\omega)$ *is* $\mathfrak{B}([0,\infty)) \times \mathcal{F}$*-measurable.*

(ii) $f(t,\omega)$ *is* $\mathfrak{F}$*-adapted, i.e.,* $\forall t \in [0,T] : f(t,\cdot)$ *is* $\mathcal{F}_t$*-measurable.*

(iii) $\mathbb{E}\left[\int_0^T |f(t,\omega)|^2 dt\right] < \infty$.

Further denote by $L^2_{0,ad}([0,T])$ *the subset of* $L^2_{ad}([0,T])$ *consisting of all functions of the form*

$$f(t,\omega) = \sum_{j=0}^{n-1} e_j(\omega)\mathbf{1}_{[t_j,t_{j+1}]}(t). \tag{3.1.9}$$

Functions of the form (3.1.9) *are called step processes.*

Remark 3.2
Conditions (i) and (iii) can also be expressed by

$$f \in L^2\left([0,T]\times\Omega, \mathfrak{B}([0,T])\otimes\mathcal{F}, \lambda\otimes P\right) =: L^2(\lambda\otimes P).$$

Hence

$$\|f\|_{L^2(\lambda\otimes P)} = \left(\int_\Omega \int_0^T |f(t,\omega)|^2 dt\, dP\right)^{\frac{1}{2}}$$

defines a norm on $L^2(\lambda\otimes P)$ *induced by the respective inner product. This remark also emphasizes that* $L^2_{0,ad}([0,T])$ *is a closed linear subspace of* $L^2(\lambda\otimes P)$.

With these prerequisites in place, we can define the Itô Integral for step processes:

Definition 3.3 (Itô Integral for $f \in L^2_{0,\mathsf{ad}}([0,T])$)
Let $f \in L^2_{0,ad}([0,T])$ *and* $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$*, then the functional*

$$\mathcal{I} : L^2_{0,ad}([0,T]) \to L^2_{\mathcal{F}} : \mathcal{I}(f)(\omega) = \sum_{j=0}^{n-1} e_j(\omega)(B_{t_{j+1}} - B_{t_j}) \tag{3.1.10}$$

is called Itô Integral. Note that this functional is a random variable itself.

The next step is to extend the functional $\mathcal{I}$ from $L^2_{0,\mathsf{ad}}([0,T])$ to $L^2_{\mathsf{ad}}([0,T])$. In order to do so, we state the following two important lemmata:

Lemma 3.1 (Itô Isometry on $L^2_{0,\mathsf{ad}}([0,T])$)
For $f \in L^2_{0,ad}([0,T])$ *we have*

$$\|\mathcal{I}(f)\|_{L^2} = \|f\|_{L^2(\lambda\otimes P)}. \tag{3.1.11}$$

Proof. Let $f \in L^2_{0,\mathsf{ad}}[0,T]$, then we have

$$\begin{aligned}
\|\mathcal{I}(f)\|^2_{L^2} &= \mathbb{E}\left[\sum_{j=0}^{n-1} e_j^2(\omega)(B_{t_{j+1}} - B_{t_j})^2\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}\left[e_j^2 \cdot (B_{t_{j+1}} - B_{t_j})^2\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}\left[e_j^2\right] \mathbb{E}\left[(B_{t_{j+1}} - B_{t_j})^2\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}[e_j^2] \cdot (t_{j+1} - t_j)
\end{aligned}$$

and

$$\begin{aligned}
\|f\|^2_{L^2(\lambda\otimes P)} &= \mathbb{E}\left[\int_0^T f^2(t,\omega)dt\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{n-1} e_j^2(\omega)\mathbf{1}_{[t_j,t_{j+1}]}(t)\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}[e_j^2] \cdot (t_{j+1} - t_j)
\end{aligned}$$

This yields $\|\mathcal{I}(f)\|_{L^2} = \|f\|_{L^2(\lambda\otimes P)}$. □

Lemma 3.2 ($L^2_{0,\mathsf{ad}}([0,T])$ is dense in $L^2_{\mathsf{ad}}([0,T])$)
For any $f \in L^2_{ad}([0,T])$ there exists a sequence $\{f_n\}$ with $f_n \in L^2_{0,ad}([0,T])$ such that

$$\|f_n - f\|_{L^2(\lambda\otimes P)} \to 0 \quad \textit{for} \quad n \to \infty.$$

See Steele (2001, p. 90 ff.) for the proof of a more general statement. With these prerequisites, it becomes easy to extend the functional $\mathcal{I}$ from $L^2_{0,\mathsf{ad}}[0,T]$ to $L^2_{\mathsf{ad}}([0,T])$ by defining

$$\mathcal{I}(f) \coloneqq \lim_{n\to\infty} \mathcal{I}(f_n) \tag{3.1.12}$$

for all $f \in L^2_{\mathsf{ad}}([0,T])$. It remains to show that the expression (3.1.12) is actually well-defined. This follows immediately since:

(i) Let $\{f_n\} \subset L^2_{\text{ad}}([0,T])$ such that $f_n \to f$ in $L^2(\lambda \otimes P)$ for $n \to \infty$. Since the space $L^2(\lambda \otimes P)$ is complete, this implies that $\{f_n\}$ is a Cauchy-sequence. By Lemma 3.1, we know that $\{\mathcal{I}(f_n)\}$ is also a Cauchy-sequence in L^2, hence converges to some $\mathcal{I}(f)$.

(ii) Let $\{f'_n\} \subset L^2_{\text{ad}}([0,T])$ be another sequence, such that $f'_n \to f$ in $L^2(\lambda \otimes P)$ for $n \to \infty$. The triangle-inequality

$$\|f_n - f_n\|_{L^2(\lambda\otimes P)} \leq \|f - f_n\|_{L^2(\lambda\otimes P)} + \|f - f_n\|_{L^2(\lambda\otimes P)},$$

and the application of Lemma 3.1 yields $\|\mathcal{I}(f_n) - \mathcal{I}(f'_n)\|_{L^2} \to 0$.

Since the Itô Isometry is a technically important instrument, we show that the statement also holds true for functions in $L^2_{\text{ad}}([0,T])$.

Lemma 3.3 (Itô Isometry on $L^2_{\text{ad}}([0,T])$)
For $f \in L^2_{ad}([0,T])$ we have

$$\|\mathcal{I}(f)\|_{L^2} = \|f\|_{L^2(\lambda\otimes P)}. \tag{3.1.13}$$

Proof. Let $f \in L^2_{\text{ad}}([0,T])$. By Lemma 3.2 we can choose $\{f_n\} \subset L^2_{0,\text{ad}}([0,T])$ such that $\|f_n - f\|_{L^2(\lambda\otimes P)} \to 0$ for $n \to \infty$. Applying the reverse triangle-inequality yields

$$0 \leq \Big|\, \|f_n\|_{L^2(\lambda\otimes P)} - \|f\|_{L^2(\lambda\otimes P)} \,\Big| \leq \|f_n - f\|_{L^2(\lambda\otimes P)} \overset{n\to\infty}{\to} 0.$$

Hence, $\|f_n\|_{L^2(\lambda\otimes P)} \to \|f\|_{L^2(\lambda\otimes P)}$ for $n \to \infty$. With similar reasoning, we also get $\|\mathcal{I}(f_n)\|_{L^2} \to \|\mathcal{I}(f)\|_{L^2}$. The Itô Isometry on $L^2_{0,\text{ad}}([0,T])$ tells us now

$$\|\mathcal{I}(f_n)\|_{L^2} = \|f_n\|_{L^2(\lambda\otimes P)}.$$

Taking limits completes the proof. □

Since it is essential to understand such integrals as stochastic processes, the following theorem strongly connects Itô Integrals and martingales.

Theorem 3.2
Let $f \in L^2_{ad}([0,T])$. Then there exists a t-continuous stochastic process $\{M_t\}$ on $(\Omega, \mathcal{F}, P)$ that is a martingale with respect to the filtration of the Brownian motion $\mathfrak{F}$ such that

$$\forall t \in [0,T] : \; P\left[M_t = \mathcal{I}(m_t f)\right] = 1,$$

where

$$m_t(\omega, s) \coloneqq \mathbf{1}_{[0,1]}(\omega, s) = \begin{cases} 1, & \text{for } s \in [0,t] \\ 0, & \text{else.} \end{cases}$$

Proof. Let $f \in L^2_{\mathsf{ad}}([0,T])$. Thus, according to Lemma 3.2, there exists a sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq L^2_{0,\mathsf{ad}}([0,T])$ such that $\|f - f_n\|_{L^2(\lambda\otimes P)} \to 0$ for $n \to \infty$. Since we also have $\{m_t f_n\}_{n\in\mathbb{N}} \subseteq L^2_{0,\mathsf{ad}}([0,T])$, we can define

$$\mathcal{I}_n(t,\omega) \coloneqq \mathcal{I}(m_t f_n). \tag{3.1.14}$$

From the explicit formulation (3.1.10) of the Itô Integral for step processes, it follows immediately that $\mathcal{I}_n(\cdot,\omega)$ is t-continuous for all $n \in \mathbb{N}$. Now, we show that $\mathcal{I}_n(t,\omega)$ is a martingale with respect to $\mathfrak{F}$ for all $n \in \mathbb{N}$, i.e., for all $t, s \in [0,T]$ with $s \le t$ we have $\mathbb{E}[\mathcal{I}_n(t,\omega)|\mathcal{F}_s] = \mathcal{I}_n(s,\omega)$:

Let $t, s \in [0,T]$ with $s \le t$, then we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{I}_n(t,\omega)\middle|\,\mathcal{F}_s\right] &= \mathbb{E}\left[\left(\mathcal{I}_n(s,\omega) + \mathcal{I}_n(t,\omega) - \mathcal{I}_n(s,\omega)\right)\middle|\,\mathcal{F}_s\right] \\
&= \mathbb{E}\left[\mathcal{I}_n(s,\omega)\middle|\,\mathcal{F}_s\right] + \mathbb{E}\left[\mathcal{I}_n(t,\omega) - \mathcal{I}_n(s,\omega)\middle|\,\mathcal{F}_s\right] \\
&= \mathcal{I}_n(s,\omega) + \mathbb{E}\left[\sum_{s\le t_j^{(n)}\le t_{j+1}^{(n)}\le t} e_j^{(n)}(B_{j+1} - B_j)\middle|\,\mathcal{F}_s\right] \\
&= \mathcal{I}_n(s,\omega) + \sum_{s\le t_j^{(n)}\le t_{j+1}^{(n)}\le t} \mathbb{E}\left[\mathbb{E}\left[e_j^{(n)}(B_{j+1} - B_j)\middle|\,\mathcal{F}_{t_j^{(n)}}\right]\middle|\,\mathcal{F}_s\right] \\
&= \mathcal{I}_n(s,\omega) + \sum_{s\le t_j^{(n)}\le t_{j+1}^{(n)}\le t} \mathbb{E}\left[e_j^{(n)}\,\mathbb{E}\left[(B_{j+1} - B_j)\middle|\,\mathcal{F}_{t_j^{(n)}}\right]\middle|\,\mathcal{F}_s\right] \\
&= \mathcal{I}_n(s,\omega).
\end{aligned}$$

Hence, $\mathcal{I}_n$ is a martingale for all $n \in \mathbb{N}$. Note that the last equality holds, since $(B_{j+1} - B_j)$ is independent of $\mathcal{F}_{t_j^{(n)}}$ and $\mathbb{E}\left[(B_{j+1} - B_j)\right] = 0$. By applying Doob's maximal inequality on the submartingale $\mathcal{I}_n - \mathcal{I}_m$ for $n \ge m$ we get

$$\begin{aligned}
P\left(\sup_{0\le t\le T} \left|\mathcal{I}_n(t,\omega) - \mathcal{I}_m(t,\omega)\right| > \epsilon\right) &\le \frac{1}{\epsilon^2}\cdot\mathbb{E}\left[\left|\mathcal{I}_n(T,\omega) - \mathcal{I}_m(T,\omega)\right|^2\right] \\
&\le \frac{1}{\epsilon^2}\|f_n - f_m\|^2_{L^2(\lambda\otimes P)}.
\end{aligned}$$

We know that $\|f - f_n\|_{L^2(\lambda\otimes P)} \to 0$ for $n \to \infty$, yielding an increasing subsequence $\{f_{n_k}\}$ such that

$$\max_{n\ge n_k} \|f_n - f_m\|^2_{L^2(\lambda\otimes P)} \le 2^{-3k}.$$

By fixing $\epsilon = 2^{-k}$ we get

$$P\left(\sup_{0\le t\le T} \left|\mathcal{I}_{n_{k+1}}(t,\omega) - \mathcal{I}_{n_k}(t,\omega)\right| > 2^{-k}\right) \le 2^{-k}.$$

Finally the Borel-Cantelli lemma provides us with

$$P\left(\sup_{0\leq t\leq T}|\mathcal{I}_{n_{k+1}}(t,\omega)-\mathcal{I}_{n_k}(t,\omega)|>2^{-k}\quad\text{for infinitely many } k\right)=0.$$

Thus, for almost all $\omega\in\Omega$ there is a $\widetilde{k}(\omega)$ such that

$$\forall k\geq\widetilde{k}(\omega):\ \sup_{0\leq t\leq T}|\mathcal{I}_{n_{k+1}}(t,\omega)-\mathcal{I}_{n_k}(t,\omega)|\leq 2^{-k}.$$

Consequently, $\{\mathcal{I}_{n_k}\}$ is a Cauchy sequence in $\mathcal{C}([0,T])$, and since this space is complete, this sequence has, in particular, a limit, which we denote by M_t. To sum up, M_t is a t-continuous martingale. Since we already know that $\mathcal{I}_{n_k}$ converges to $\mathcal{I}_t\coloneqq\mathcal{I}(m_tf)$ in L^2 we immediately deduce by unqiueness of limits

$$\forall t\in[0,T]:\ \mathcal{I}_t=M_t\quad a.s.$$

This completes the proof. □

Remark 3.3
Let $f\in L^2_{ad}([0,T])$. If the requirements of Theorem 3.2 are fulfilled, we henceforth also write

$$M_t(\omega)=\int_0^t f(s,\omega)\,dB_s\quad\textit{for all}\quad 0\leq t\leq T.$$

The final step in deriving the Itô Integral is to relax the integrability constraints in Definition 3.2 sufficiently enough. For this purpose, efforts will now be made to achieve what is known as *localization* in Stochastics. Although the formalization is quite abstract, one will quickly realize that this new space naturally fits as a proper domain of the Itô Integral. In this case, the reader may also refer to Steele (2001, p. 96 ff.). We now consider the following class of functions:

Definition 3.4
Let $\{B_t\}_{t\geq 0}$ be a Brownian motion on the probability space $(\Omega,\mathcal{F},P)$ and $\mathfrak{F}\coloneqq\{\mathcal{F}_t\}_{t\in[0,T]}$ the respective augmented filtration. Define $L^2_{loc}([0,T])$ to be the class of functions

$$f(t,\omega):\ [0,\infty)\times\Omega\to\mathbb{R}$$

such that

(i) $(t,\omega)\to f(t,\omega)$ *is $\mathfrak{B}([0,\infty))\times\mathcal{F}$-measurable.*

(ii) $f(t,\omega)$ *is $\mathfrak{F}$-adapted, i.e., $\forall t\in[0,T]:f(t,\cdot)$ is $\mathcal{F}_t$-measurable.*

(iii) $P\left(\int_0^T|f(t,\omega)|^2dt<\infty\right)=1.$

Note that in contrast to Definition 3.2, we weakened condition (iii) here.

Definition 3.5 (Stopping time)
Let $(\Omega, \mathcal{F})$ be a measurable space endowed with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Then we call the mapping

$$\mathcal{T} : \Omega \to \mathbb{R} \cup \{\infty\} \tag{3.1.15}$$

stopping time w.r.t. $\{\mathcal{F}_t\}_{t\geq 0}$, if we have $\{\mathcal{T} \leq t\} = \{\omega \in \Omega : \mathcal{T}(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Example 3.2 (Hitting time)
Let $\{B_t\}_{t\geq 0}$ be a Brownian motion. Then the so-called hitting-time

$$\mathcal{T}_a(\omega) \coloneqq \inf\{t \in \mathbb{R}_{\geq 0} \mid B_t(\omega) = a\}, \quad a \in \mathbb{R}$$

is a stopping time. For an exemplary visualization, consider Figure 1.

Before we begin the localization procedure, we briefly state a theorem concerned with the persistence of identity under integration. This theorem will be helpful for later proofs.

Theorem 3.3 (Persistence of identity)
Let $f, g \in L^2_{ad}([0,T])$ and $\mathcal{T}$ a stopping time, such that $f(s,\omega) = g(s,\omega)$ for almost all $\omega \in \{\omega :\ s \leq \mathcal{T}(\omega)\}$. Then the integrals

$$X_t(\omega) = \int_0^t f(s,\omega)dB_s \quad \textit{and} \quad Y_t(\omega) = \int_0^t g(s,\omega)dB_s$$

are for $a.a.\ \omega \in \{\omega :\ s \leq \mathcal{T}(\omega)\}$ equal.

See Steele (2001, p. 89) for the proof of Theorem 3.3.

Definition 3.6 ($L^2_{\text{ad}}([0,T])$ localizing sequence)
Let $f \in L^2_{loc}([0,T])$. An increasing sequence of stopping times $\{\mathcal{T}_n\} \subset L^2_{ad}([0,T])$ is called a $L^2_{ad}([0,T])$-localizing sequence if we have

$$\forall n \in \mathbb{N} :\ f_n(\omega,t) = f(\omega,t)\mathbf{1}_{\{t\leq \mathcal{T}_n\}} \in L^2_{ad}([0,T])$$

and

$$P\left(\bigcup_{n=1}^{\infty} \{\omega \in \Omega : \mathcal{T}_n(\omega) = T\}\right) = 1.$$

For simplicity, we will henceforth speak of localizing sequences rather than $L^2_{ad}([0,T])$ localizing sequence, since there is no possibility of confusion.

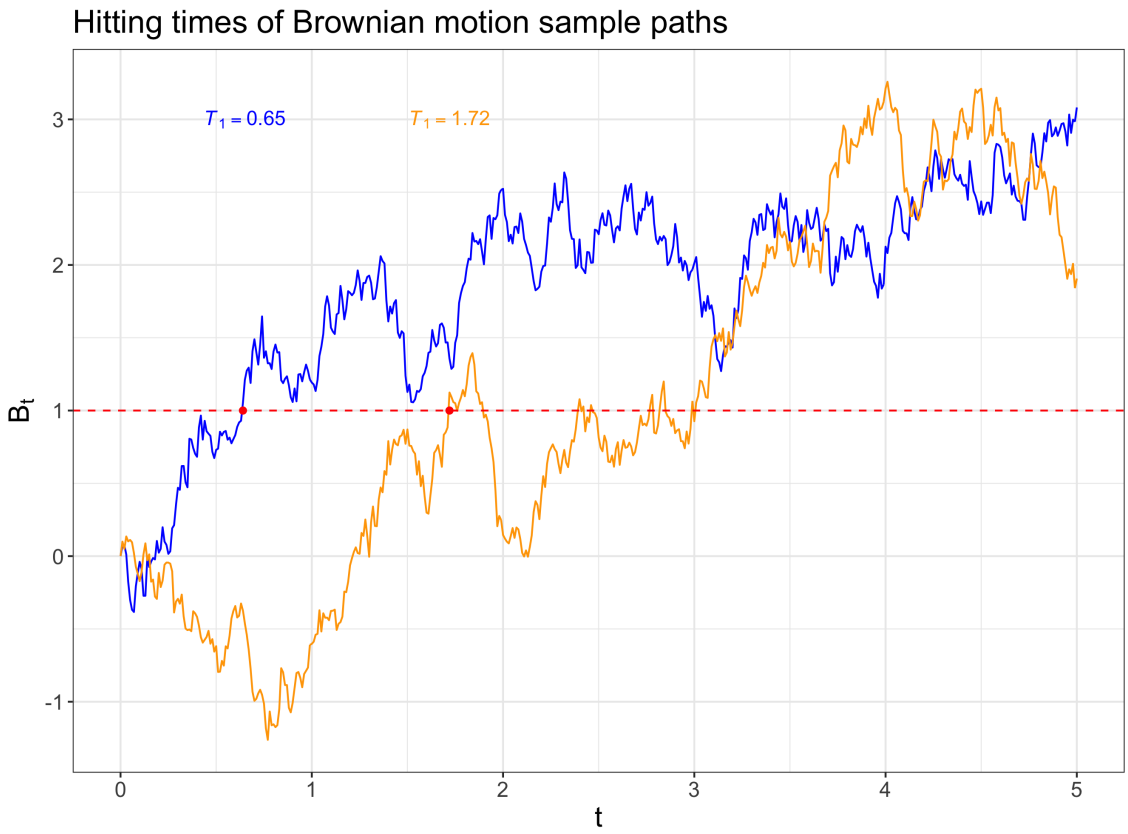

Figure 1: Hitting time T_1 of Brownian motion sample paths.

Proposition 3.1
Let $f \in L^2_{loc}([0,T])$. Consequently, the sequence defined by

$$\mathcal{T}_n = \inf\left\{ s \in [0,T] : \int_0^s |f(\omega,t)|^2 dt \geq n \quad \textit{or} \quad s \geq T \right\}$$

is a localizing sequence.

Proof. We have

$$\bigcup_{n=1}^{\infty} \{\omega : \mathcal{T}_n = T\} = \left\{ \omega : \int_0^T |f(\omega,t)|^2 dt < \infty \right\}$$

for $f \in L^2_{\text{loc}}([0,T])$, where the second set has by Definition 3.4 probability one. Since we also have

$$\|f_n\|^2_{L^2(\lambda \otimes P)} = \mathbb{E}\left[\int_0^T f_n^2(\omega,t) \mathbf{1}_{\{t \leq \mathcal{T}_n\}} dt \right] = \mathbb{E}\left[\int_0^{\mathcal{T}_n} f_n^2(\omega,t) \mathbf{1}_{\{t \leq \mathcal{T}_n\}} dt \right] \leq \mathbb{E}[n] = n,$$

hence, $f_n \in L^2_{\text{ad}}([0,T])$. Therefore, we conclude that $\{\mathcal{T}_n\}$ is a localizing sequence. □

With these prerequisites in place, we can now start extending the Itô Integral to functions defined on $L^2_{\text{loc}}([0,T])$. We will proceed as follows:

1.) Proposition 3.1 tells us that, for any $f \in L^2_{\text{loc}}([0,T])$ there is a localizing sequence $\{\mathcal{T}_n\}$ such that $f_n(\omega,s) = f(\omega,s)\mathbf{1}_{\{s\leq\mathcal{T}_n\}} \in L^2_{\text{ad}}([0,T])$. Hence, we know from previous results[9] that there exists a unique t-continuous stochastic process $\{M_{t,n}\}$ such that $M_{t,n} = \mathcal{I}(m_t f_n)$.

2.) According to the preceding point, we define the Itô Integral for $f \in L^2_{\text{loc}}([0,T])$ as the limit of $\{M_{t,n}\}$ for $n \to \infty$. Thus, we have to show that $\{M_{t,n}\}$ converges to some unique t-continuous stochastic process $\{M_t\}$ such that

$$\forall t \in [0,T] : \; P\left(M_t = \lim_{n\to\infty} M_{t,n}\right) = 1.$$

3.) The final step consists of checking that the above construction is independent of the specific choice of the localizing sequence $\mathcal{T}_n$.

Proposition 3.2
Let $f \in L^2_{loc}([0,T])$ and $\{\mathcal{T}_n\}$ any localizing sequence. If $\{M_{t,n}\}$ is the continuous martingale version of the Itô Integral $\mathcal{I}\left(m_t f(\omega,s)\mathbf{1}_{\{s\leq\mathcal{T}_n\}}\right)$, then we have

$$\forall t \in [0,T] \;\; \forall n \geq m : \; M_{t,n} = M_{t,m} \quad \textit{for } a.a. \quad \{\omega : t \leq \mathcal{T}_m\}.$$

Proof. By definition of localizing sequences, we have $\mathcal{T}_m \leq \mathcal{T}_n$ for $m \leq n$, and hence the functions

$$f_m(\omega,t) = f(\omega,t)\mathbf{1}_{\{t\leq\mathcal{T}_m\}} \quad \text{and} \quad f_n(\omega,t) = f(\omega,t)\mathbf{1}_{\{t\leq\mathcal{T}_n\}}$$

are equal on the set $\{\omega : t \leq \mathcal{T}_n\}$. The application of Theorem 3.3 completes the proof. □

According to point 2.) from our preliminary considerations, we show that such $\{M_{t,n}\}$ has a unique, t-continuous limit, denoted by $\{M_t\}$. The proof of the following statement clarifies that, with Proposition 3.2, the desired convergence follows immediately.

Proposition 3.3
There exists a t-continuous process $\{M_t\}$ such that

$$\forall t \in [0,T] : \; P\left(\lim_{n\to\infty} M_{t,n} = M_t\right) = 1.$$

[9] Specifically, here we refer to Theorem 3.2.

Proof. First, define $N \coloneqq \min\{n \in \mathbb{N} : \mathcal{T}_n = T\}$. By way of Definition 3.6, we already know that $P(\{\omega \in \Omega : N(\omega) < \infty\}) = 1$. In addition, let $\widetilde{\Omega}$ denote the set with probability one such that all functions $t \mapsto X_t(\omega)$ are t-continuous. For any $\omega \in \widetilde{\Omega} \cap \{\omega \in \Omega : N(\omega) < \infty\}$, we now fix $M_t(\omega) = M_{t,N}(\omega)$. Thus, $\{M_t\}$ is a t-continuous process. By means of the previous proposition, we have

$$\forall t \in [0,T] : \; P\left(\lim_{n\to\infty} M_{t,n} = M_{t,N}\right) = 1,$$

since the equality $M_{t,n} = M_{t,N}$ holds for all $t \in [0,T]$ and $n \geq N$ on the sets $\{\omega \in \Omega : t \leq \mathcal{T}_N(\omega)\}$. □

To complete our derivation of the Itô Integral for $L^2_{\text{loc}}([0,T])$ functions, we finally check that the construction is independent of the choice of the localizing sequence $\{\mathcal{T}_n\}$, namely:

Proposition 3.4
Let $\mathcal{T}_n$ and $\widetilde{T}_n$ be two localizing sequences for $f \in L^2_{\text{loc}}([0,T])$. Then the respective t-continuous martingale versions $M_{t,n}$ and $\widetilde{M}_{t,n}$ satisfy

$$\lim_{n\to\infty} M_{t,n} = \lim_{n\to\infty} \widetilde{M}_{t,n}$$

with probability one for each $t \in [0,T]$.

Proof. Let $f \in L^2_{\text{loc}}([0,T])$ and $\mathcal{T}_n$ and $\widetilde{\mathcal{T}}_n$ be two localizing sequences. Now, define $\tau_n \coloneqq \min\{\mathcal{T}_n, \widetilde{\mathcal{T}}_n\}$. By Theorem 3.3, it holds for all $n \geq m$ that

$$\widetilde{M}_{t,n} = M_{t,n} \quad a.e. \text{ on } \quad \{t \leq \tau_m\}. \tag{3.1.16}$$

Since Proposition 3.3 yields convergence of $M_{t,n}$ and $\widetilde{M}_{t,n}$ to a unique limit $a.s.$, we have with (3.1.16)

$$\lim_{n\to\infty} \widetilde{M}_{t,n} = \lim_{n\to\infty} M_{t,n} \quad a.e. \text{ on } \quad \{t \leq \tau_m\}.$$

By definition of a localizing sequence, the proof is complete. □

This completes the derivation of the Itô Integral. Although we have a well-defined stochastic integral notion for a sufficiently large class of functions, it seems pretty cumbersome to calculate integrals of the form

$$\int_0^t \sigma(s, X_s) dB_s$$

strictly via the above-derived definition. Bear in mind that the elegance of the "usual" integral calculus bears witness to its direct link to differential calculus. In particular, the

fundamental theorem of calculus contributes to the remarkable simplicity of calculating integrals. The aim now is to formulate a suitable formula to calculate Itô Integrals, namely Itô's formula. We will only develop the most general version of Itô's formula. For other versions, see the references above.

Definition 3.7 (Itô process)
An Itô process is a stochastic process $\{X_t\}$ of the form

$$X_t = X_0 + \int_0^t f(\omega, s)ds + \int_0^t g(\omega, s)dB_s \quad \textit{for} \quad 0 \leq t \leq T, \tag{3.1.17}$$

where we assume $g(\cdot,\cdot) \in L^2_{loc}([0,T])$ and $f(\cdot,\cdot) \in L^1_{loc}([0,T])$, where $L^1_{loc}([0,T])$ denotes the space of all measurable, $\mathfrak{F}$-adapted processes satisying

$$P\left(\int_0^T |f(t,\omega)|dt < \infty\right) = 1.$$

Remark 3.4
We note that the above assumptions on $g(\cdot,\cdot)$ and $f(\cdot,\cdot)$ guarantee that all integrals involved in (3.1.17) *are well-defined. Furthermore, we often use the following "shorthand notation"*

$$dX_t = fdt + gdB_s$$

for (3.1.17).

Theorem 3.4
Let $\{X_t\}$ be an Itô process given by

$$dX_t = fdt + gdB_s,$$

and $g(t,x) \in \mathcal{C}^2(\mathbb{R}_{\geq 0} \times \mathbb{R})$. Then the process denoted by $Y_t = g(t, X_t)$ is again an Itô process and we have

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2. \tag{3.1.18}$$

For the formal "symbol multiplication", we refer to the following table:

Remark 3.5
Some peculiarities explained in more detail:

1.) *In Table 4, the formal multiplication $(dB_t)^2 = dt$ is justified with the fact that the quadratic variation of a Brownian motion is equal to t.*

$\cdot$	dt	dB_t
dt	0	0
dB_t	0	dt

Table 4: Formal symbol multiplication.

2.) We can express (3.1.18) *equivalently by*

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2.$$

The substitution $dX_t = f dt + g dB_t$ *and Table 4 justify the above equivalence.*

See Steele (2001, Chap. 8) for the proof of Itô's formula. Before we revisit the population growth model and find a suitable solution to the stochastic differential equation (3.1.4), we first consider a classic example to illustrate Itô's formula.

Example 3.3
Let $\{B_t\}_{t\geq 0}$ *be a Brownian motion on the probability space* $(\Omega, \mathcal{F}, P)$*. We will demonstrate Itô's formula by solving the integral*

$$\int_0^t B_s \, dB_s.$$

With $X_t = B_t$ *and* $g(t, x) = \frac{1}{2}x^2$ *we get* $Y_t = g(t, B_t) = \frac{1}{2}B_t^2$ *and hence, applying Itô's formula*

$$dY_t = \frac{\partial g}{\partial t}(t, B_t)dt + \frac{\partial g}{\partial x}(t, B_t)dB_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, B_t)(dB_t)^2$$

yields

$$d\left(\frac{1}{2}B_t^2\right) = B_t dB_t + \frac{1}{2}dt.$$

Rearranging results in

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t.$$

Now, consider the SDE

$$\frac{dN_t}{dt} = r_t N_t + \alpha N_t V_t,$$

that we derived earlier as a stochastic version of the population growth model (3.1.3), or equivalently stated

$$\frac{dN_t}{N_t} = r_t\, d_t + \alpha\, dB_t, \tag{3.1.19}$$

where we set $dB_t = V_t d_t$. For the sake of simplicity, we also assume $r_t = r =$ constant. It seems natural to chose $g(t,x) = \ln(x)$, which is clearly a twice differentiable function such that applying Itô's formula yields

$$\begin{aligned} d\left(\ln N_t\right) &= \frac{1}{N_t} dN_t + \frac{1}{2}\left(-\frac{1}{N_t^2}(dN_t)^2\right) \\ &= \frac{dN_t}{N_t} - \frac{1}{2}\alpha^2 dt \\ \Leftrightarrow \qquad \frac{dN_t}{N_t} &= d(\ln N_t) + \frac{1}{2}\alpha^2 dt. \end{aligned}$$

By (3.1.19), we finally get

$$N_t = N_0 \exp\left((r - \frac{1}{2}\alpha^2)t + \alpha B_t\right). \tag{3.1.20}$$

In Figure 2, we visualized the solution of the population growth model with random noise for different α values and $r = 1$ and an initial population size $N_0 = 50$. This simple illustration shows that different parameter values strongly influence the population dynamics. Now, before we present the conceptual specificity of so-called Backward stochastic differential equations (BSDE), we consider an essential theorem that makes a statement about the existence and uniqueness of SDE. So far, we had silently assumed existence and uniqueness.

Definition 3.8 (Strong solution of SDE)
Let $b, \sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ be measurable functions. Then, the t-continuous, $\mathfrak{F}$-adapted process $\{Y_t\}$ is called strong solution of the SDE

$$\begin{cases} dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dB_t, \quad 0 \le t \le T, \\ Y_0 = X, \end{cases} \tag{3.1.21}$$

if for all $t \in [0,T]$ the equation

$$Y_t = X + \int_0^t b(s, Y_s)ds + \int_0^t \sigma(s, Y_s)dB_s \tag{3.1.22}$$

is $a.s.$ true.

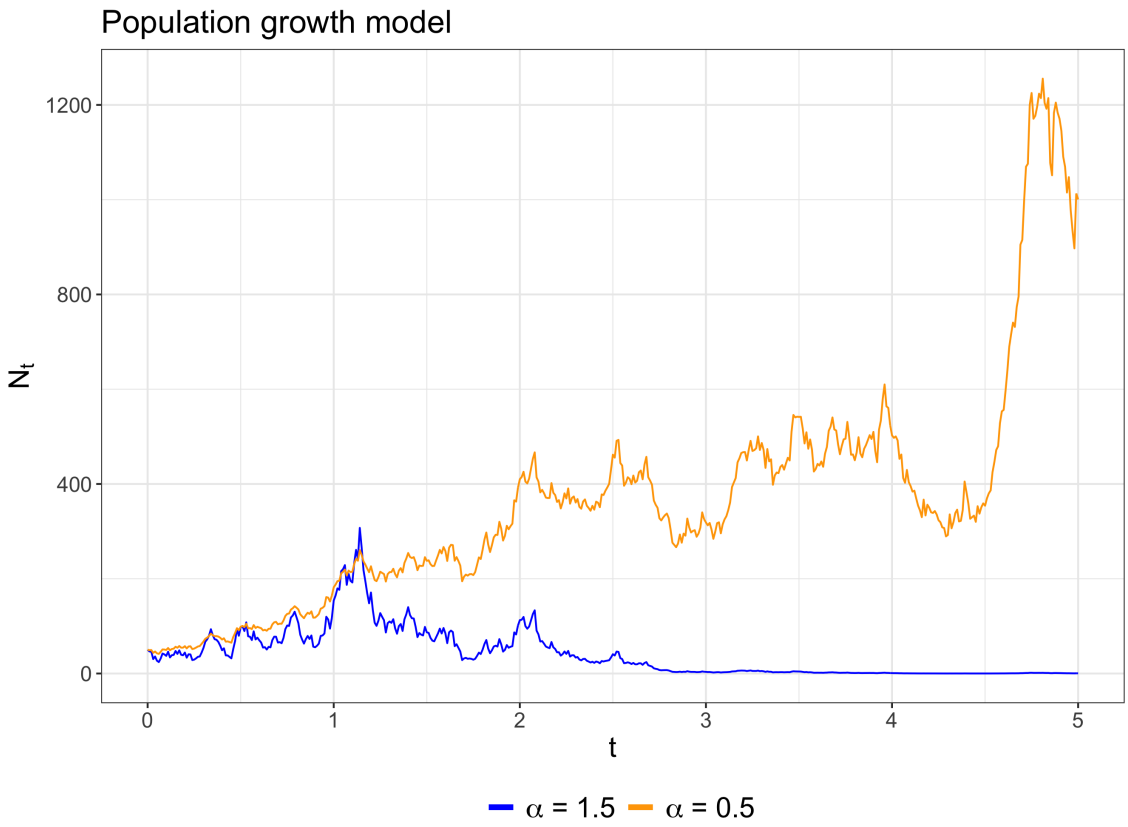

Figure 2: Visualization of (3.1.20) with different α values, $r = 1$ and $N_0 = 50$.

Note that, in the context of SDEs, one distinguishes between strong and weak solutions, in which case, weak solutions are characterized by distributions, roughly speaking. For this, consider the explanations in Oksendal (2013, p. 72 ff.).

Theorem 3.5 (Existence and uniqueness of SDE)
Let $b, \sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ be measurable functions satisyfing the Lipschitz condition

$$\forall x, y \in \mathbb{R}\ \forall t \in [0,T] : \ |\ b(t,x) - b(t,y)| \ + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y| \quad (3.1.23)$$

for some constant D and additionally the growth condition

$$\forall x \in \mathbb{R}\ \forall t \in [0,T] : \ |b(t,x)| + |\sigma(t,x)| \leq C(1+|x|) \quad (3.1.24)$$

for some constant C. If $X \in L^2_{\mathcal{F}_T}$ is $\mathcal{F}_0$-measurable, then the SDE (3.1.21) has a unique t-continuous, $\mathfrak{F}$-adapted solution.

At this point, we will not prove the existence and uniqueness theorem, since this requires some tedious preparation. For the respective proof, we refer to Oksendal (2013, p. 69 ff.).

Remark 3.6
Some technical remarks:

1.) *In Definition 3.8, we implicity assume* $b(\cdot,\cdot) \in L^1_{loc}([0,T])$ *and* $\sigma(\cdot,\cdot) \in L^2_{loc}([0,T])$. *Thus, all integrals involved in* (3.1.22) *are well-defined.*

2.) *The growth condition* (3.1.24) *ensures that the solution of* (3.1.21) *does not explode, i.e.,* $|Y_t(\omega)|$ *does not tend to infinity in finite time, whereas the Lipschitz condition guarantess the uniqueness of the SDE. For intuitive (counter)-examples, the reader may refer to Oksendal (2013, p. 68 ff.).*

3.) *Theorem 3.5 requires* $X \in L^2_{\mathcal{F}_T}$ *to be* $\mathcal{F}_0$*-measurable, where we denoted by* $\mathfrak{F} \coloneqq \{\mathcal{F}_t\}_{t\in[0,T]}$ *the augmented Brownian filtration. Albeit slightly generalized, one could consider* X *to be independent of* $\{\mathcal{F}_t\}_{t\in[0,T]}$ *and say that* $\{Y_t\}$ *is adapted to the filtration* $\mathcal{F}^X_t \coloneqq \sigma(X, \mathcal{F}_s : s < t)$. *This would allow for adapted solutions without necessarily* $\mathcal{F}_0$*-measurable initial condition* X.

Finally, we formulate the counterpart of Theorem 3.2. Theorem 3.2 states that the Itô Integral of any function $f \in L^2_{\text{ad}}([0,T])$ is a martingale with respect to the augmented filtration of the Brownian motion. Now, we will see that the converse direction is also true, i.e., any martingale adapted with respect to the augmented Brownian filtration can be represented as an Itô Integral. This theorem is considered particularly important.

Theorem 3.6 (Martingale representation)
Let $\{M_t\}$ *be a* L^2*-martingale with respect to the augmented Brownian filtration* $\mathfrak{F}$. *Then there exists a unique stochastic process* $g(s,\omega)$ *such that* $g(\cdot,\cdot) \in L^2_{ad}([0,T])$ *for all* $t \geq 0$ *and*

$$\forall t \geq 0 : \; M_t(\omega) = X + \int_0^t g(s,\omega) dB_s \quad a.s. \tag{3.1.25}$$

where we have $X \coloneqq \mathbb{E}[M_0]$.

The Martingale representation theorem is proven in Oksendal (2013, p. 53 f.). From (3.1.21), it should be clear that, in the context of SDEs, we are working with initial conditions, i.e., $Y_0 = X$. If we impose a terminal condition $X \coloneqq Y_T$ to the stochastic differential equation instead of an initial condition, we work *backward* in time, hence *backward* stochastic differential equations. While the theory of BSDE dates back in the linear case to Bismut (1973), we will consider the more general formulation of Pardoux and Peng (1990). For a brief introduction of BSDEs, we mainly consider El Karoui and Mazliak (1997) and Zhang (2017).

First, we adjust some further notation. As we did before, we denote by $\mathcal{F}_t \coloneqq \sigma\{B_s \cup \mathfrak{N} : 0 \leq s \leq t\}$ the augmented Brownian filtration. Furthermore, we will mainly be engaged with the following spaces of functions for $t \in [0,T]$:

- $L^2_{\mathcal{F}_t} \coloneqq L^2(\Omega, \mathcal{F}_t, P)$, space of all real-valued, $\mathcal{F}_t$-measurable and square integrable random variables equipped with the L^2-norm.
- $L^2_{\mathcal{F}}([0,T];\mathbb{R}^n)$ space of all $\mathbb{R}^n$-valued, $\{\mathcal{F}_t\}_{t\in[0,T]}$-adapted processes such that

$$\mathbb{E}\left[\int_0^T |\phi(s,\omega)|^2 ds\right] < \infty.$$

- $L^2_{\mathcal{F}}([0,T], \mathbb{R}\times\mathbb{R}^n) \coloneqq L^2_{\mathcal{F}}([0,T],\mathbb{R}) \times L^2_{\mathcal{F}}([0,T],\mathbb{R}^n)$.

Before considering the general case, we will first motivate the idea behind BSDE. As in the case of SDE, but now given a terminal condition, we would hope for an adapted solution of the respective BSDE. An adapted solution is generally not accessible when working backward in time in the case of SDEs, e.g., consider the following SDE

$$dY_t = \sigma(t, Y_t) dB_t, \quad 0 \leq t \leq T \tag{3.1.26}$$

with terminal condition $Y_T = X$ for some measurable function $\sigma : [0,T]\times\mathbb{R} \to \mathbb{R}$. Consider $\sigma \equiv 0$, then it is obvious that one solution of (3.1.26) is given by $X = Y_t$ for all $t \in [0,T]$. It should become clear that this solution is not adapted (can not be measurable with respect to smaller σ-algebras). Hence, to hope for an adapted solution for a given terminal condition, we must expect Y_t to be a martingale. Let $X \in L^2_{\mathcal{F}_T}$, then $Y_t \coloneqq \mathbb{E}[X|\,\mathcal{F}_t]$ is a martingale induced by X. By Theorem 3.6, there exists a unique stochastic process $Z(\cdot,\cdot) \in \mathcal{H}^2[0,T]$ such that

$$Y_t = Y_0 + \int_0^t Z_s dB_s. \tag{3.1.27}$$

In particular, we have

$$Y_T = Y_0 + \int_0^T Z_s dB_s \tag{3.1.28}$$

$$\Leftrightarrow \quad Y_T = Y_t + \int_t^T Z_s dB_s \tag{3.1.29}$$

$$\Leftrightarrow \quad Y_t = Y_T - \int_t^T Z_s dB_s. \tag{3.1.30}$$

We will now extend this to a more general setting, in which a nonlinear term is involved. We consider the following general BSDE first introduced by Pardoux and Peng (1990)

$$\begin{cases} -dY_t = g(t, Y_t, Z_t)dt - Z_t dB_t, \quad 0 \leq t \leq T, \\ \quad Y_T = X, \end{cases} \tag{3.1.31}$$

where the (nonlinear) function

$$g : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} : \ (\omega, t, y, z) \mapsto g(\omega, t, y, z)$$

is called the *generator* of the BSDE (3.1.31). This function is of particular interest in the context of decision making under uncertainty. As in the case of SDE, the BSDE (3.1.31) can be written equivalently in the integral form

$$Y_t = X + \int_t^T g(s, Y_s, Z_s)ds - \int_t^T Z_s dB_s.$$

Definition 3.9 (Strong solution of BSDE)
A pair $(Y_t, Z_t)_{t\in[0,T]} \in L^2_{\mathcal{F}}([0, T], \mathbb{R} \times \mathbb{R}^n)$ *solving the BSDE* (3.1.31) *is called strong solution. Given that BSDEs generally just admit strong solutions, we can omit the respective prefix here.*

Note that the solution of a BSDE consists of a pair $(Y_t, Z_t)_{t\in[0,T]}$. As stated in the motivation of BSDE, this is why the process is adapted. As before, in the case of SDEs, we can now answer when unique solutions of BSDEs of the form (3.1.31) exist. We will impose appropriate constraints on the function g.

Theorem 3.7 (Existence and uniqueness of BSDE)
If we assume that the function g satisfies the following conditions

(i) $g(\cdot, y, z) \in L^2_{\mathcal{F}_t}$ *for each* $y \in \mathbb{R}, z \in \mathbb{R}^n$,

(ii) $\exists C > 0$ *such that for all* $y_1, y_2 \in \mathbb{R}$ *and* $z_1, z_2 \in \mathbb{R}^n$, *we have*

$$|g(t, y_1, z_1) - g(t, y_2, z_2)| \leq C(|y_1 - y_2| + |z_1 - z_2|),$$

then there exists a unique pair of processes $(Y_t, Z_t)_{t\in[0,T]} \in L^2_{\mathcal{F}}([0, T], \mathbb{R} \times \mathbb{R}^n)$ *solving the BSDE* (3.1.31).

For the detailed proof, we refer to Pardoux and Peng (1990). Before introducing g-expectations, we will first state two fundamental theorems in BSDE theory without proof.

Theorem 3.8 (Continuous dependency)
Let $X_1, X_2 \in L^2_{\mathcal{F}}$ and $\phi^1_t, \phi^2_t \in L^2_{\mathcal{F}}([0,T];\mathbb{R})$. Now, let $(Y^i_t, Z^i_t)_{t\in[0,T]} \in L^2_{\mathcal{F}}([0,T], \mathbb{R}\times\mathbb{R}^n)$ for $i \in \{1,2\}$ be solutions of the BSDE

$$\begin{cases} -dY^i_t = f(t, Y^i_t, Z^i_t)dt - Z^i_t dB_t, \quad 0 \le t \le T, \\ Y^i_T = X_i. \end{cases}$$

Additionally, we assume that the conditions of Theorem 3.7 hold true for the function g. Then we have the following continuous dependence property

$$\sup_{0\le t\le T} \mathbb{E}\left[\left|Y^1_t - Y^2_t\right|^2\right] + \mathbb{E}\left[\int_0^T \left|Z^1_s - Z^2_s\right|^2 ds\right] \le C\,\mathbb{E}\left[\left|X_1 - X_2\right|^2\right] + C\,\mathbb{E}\left[\int_0^T |\phi^1_s - \phi^2_s ds|\right]$$

The next theorem allows to compare solutions of BSDEs in terms of the terminal conditions and the respective generators.

Theorem 3.9 (Comparison theorem)
Let assume that the assumptions of the previous theorem hold. If we have

$$Y^1_T \ge Y^2_T \; a.s. \quad \phi^1_t \ge \phi^2_t \; a.s., a.e.$$

then for $t \in [0,T]$

$$Y^1_t \ge Y^2_t \, a.s.$$

Moreover, we have

$$Y^1_t = Y^2_t \; a.s. \quad \textit{iff} \quad Y^1_T = Y^2_T \; a.s. \quad \phi^1_t = \phi^2_t \; a.s., a.e.$$

Now that the most necessary results are formulated, we can frame g-expectations. The theory subject to BSDE offers a theoretically very interesting, although mathematically deep framework. Thus, we refer to corresponding works without going into technical subtleties. For a good overview, consider El Karoui and Mazliak (1997) and references therein, in particular, Peng (1991), Pardoux and Peng (1992), Peng et al. (1992), Peng (1992), and Peng (1993).

3.2 g-expectation

With these preparations, g-expectations can now be defined immediately. Furthermore, we will also consider the basic properties of this dynamic nonlinear expectation. The reader may also refer to Coquet et al. (2002).

Definition 3.10 (g-expectation)
Let $(Y_t, Z_t)_{t\in[0,T]} \in L^2_{\mathcal{F}}([0,T], \mathbb{R}\times\mathbb{R}^n)$ be the unique solution of the BSDE

$$\begin{cases} -dY_t = g(t, Y_t, Z_t)dt - Z_t dB_t, & 0 \le t \le T, \\ Y_T = X, \end{cases}$$

where the function

$$g : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} : \ (\omega, t, y, z) \mapsto g(\omega, t, y, z)$$

satisfies the conditions in Theorem 3.7 as well as

$$\forall y \in \mathbb{R} : \ g(\cdot, y, 0) \equiv 0. \tag{3.2.1}$$

Then the functional $\mathcal{E}_g : L^2_{\mathcal{F}} \to \mathbb{R} : \ X \mapsto \mathcal{E}_g[X] = Y_0$ is called g-expectation of X. In the following, we often write $g(t,y,z)$ instead of $g(\omega,t,y,z)$.

Theorem 3.10 (Properties of g-expectation)
Let $\mathcal{E}_g$ be the g-expectation functional and $X, Y \in L^2_{\mathcal{F}_T}$, then we have:

(i) $\forall c \in \mathbb{R} : \mathcal{E}_g[c] = c$

(ii) $\forall X \le Y \ a.s. : \mathcal{E}_g[X] \le \mathcal{E}_g[Y]$ *and* $\mathcal{E}_g[X] = \mathcal{E}_g[Y]$ *if and only if* $X = Y \ a.s.$

(iii) For all $T > 0$, there exists a constant C_T, such that, for all X_1 and X_2 in $L^2_{\mathcal{F}_T}$

$$|\mathcal{E}_g[X_1] - \mathcal{E}_g[X_2]|^2 \le C_T \, \mathbb{E}[|X_1 - X_2|^2].$$

Proof. Since $c \in \mathbb{R}$ is deterministic we have $Z = 0$ and by assumption (3.2.1) the solution of the respective BSDE is $(c, 0)$, hence $\mathcal{E}_g[c] = c$. This proves (i). Properties (ii) and (iii) are immediate corollaries of Theorem 3.9 and Theorem 3.8. □

Analogous to the classical notion of conditional expectation, we can also define a conditional g-expectation that satisfies the classical conditions. At this point, the reader may also refer to the corresponding section in the appendix regarding classical results on conditional expectations.

Proposition 3.5
Let $X \in L^2_{\mathcal{F}_T}$. Then there exists a $a.s.$ unique $\eta \in L^2_{\mathcal{F}_T}$ satisfying

(i) η *is* $\mathcal{F}_t$*-measurable*

(ii) $\forall A \in \mathcal{F}_t : \ \mathcal{E}_g[X\mathbf{1}_A] = \mathcal{E}_g[\eta\mathbf{1}_A]$

Proof. We will now show existence and uniqueness.

(i) *Existence:*

Let $t \geq 0$ and $X \in L^2_{\mathcal{F}_T}$, as well as $(Y_t, Z_t)_{t\in[0,T]} \in L^2_{\mathcal{F}}([0,T], \mathbb{R} \times \mathbb{R}^n)$ be the solution of the BSDE (3.1.31), then we have

$$Y_u = X + \int_u^T g(s, Y_s, Z_s)ds - \int_u^T Z_s dB_s, \quad 0 \leq u \leq T.$$

Because of the equality

$$g(\cdot, \mathbf{1}_A y, \mathbf{1}_A z) \equiv g(\cdot, y, \mathbf{1}_A z) \equiv \mathbf{1}_A g(\cdot, y, z)$$

we have for all $A \in \mathcal{F}_t$

$$\mathbf{1}_A Y_u = \mathbf{1}_A X + \int_u^T g(s, \mathbf{1}_A Y_s, \mathbf{1}_A Z_s)ds - \int_u^T \mathbf{1}_A Z_s dB_s, \quad t \leq u \leq T.$$

Thus, it follows from Definition 3.10

$$\mathcal{E}_g[\mathbf{1}_A X] = \mathcal{E}_g[\mathbf{1}_A Y_t].$$

Since Y_t is $\mathcal{F}_t$-measurable this completes the proof of existence.

(ii) *Uniqueness:*

Let $\eta_1, \eta_2 \in L^2_{\mathcal{F}_T}$ satisfying (i) and (ii). Then we have for all $A \in \mathcal{F}_t$

$$\mathcal{E}_g[\eta_1\mathbf{1}_A] = \mathcal{E}_g[\eta_2\mathbf{1}_A].$$

Define $A_1 \coloneqq \{\eta_1 \geq \eta_2\}$ and $A_2 \coloneqq \{\eta_1 \leq \eta_2\}$. We know that in particular $A_1, A_2 \in \mathcal{F}_t$ and hence

$$\mathcal{E}_g[\eta_1\mathbf{1}_{\{\eta_1\geq\eta_2\}}] = \mathcal{E}_g[\eta_1\mathbf{1}_{\{\eta_2\geq\eta_2\}}]$$
$$\mathcal{E}_g[\eta_1\mathbf{1}_{\{\eta_1\leq\eta_2\}}] = \mathcal{E}_g[\eta_1\mathbf{1}_{\{\eta_2\leq\eta_2\}}].$$

Further we have

$$\mathbf{1}_{\{\eta_1\geq\eta_2\}}\eta_1 \geq \mathbf{1}_{\{\eta_1\geq\eta_2\}}\eta_2$$
$$\mathbf{1}_{\{\eta_1\leq\eta_2\}}\eta_1 \geq \mathbf{1}_{\{\eta_1\leq\eta_2\}}\eta_2.$$

From these observations and with property (ii) in Theorem 3.10 we get

$$\mathbf{1}_{\{\eta_1 \geq \eta_2\}}\eta_1 = \mathbf{1}_{\{\eta_1 \geq \eta_2\}}\eta_2 \quad a.s.$$

$$\mathbf{1}_{\{\eta_1 \leq \eta_2\}}\eta_1 = \mathbf{1}_{\{\eta_1 \leq \eta_2\}}\eta_2 \quad a.s.$$

and hence $\eta_1 = \eta_2$ $a.s.$.

This completes the proof. □

Definition 3.11 (Conditional g-expectation)
A random variable η satisfying the conditions of Proposition 3.5 is called conditional g-expectations and is often denoted by $\mathcal{E}_g[X|\,\mathcal{F}_t]$.

Lemma 3.4
Let $0 \leq t \leq T$. The conditional g-expectation has following properties:

(i) If X is $\mathcal{F}_t$-measurable, then $\mathcal{E}_g[X|\,\mathcal{F}_t] \equiv X$.

(ii) Let $t > s$, then
$$\mathcal{E}_g[\mathcal{E}_g[X|\ \mathcal{F}_t]|\ \mathcal{F}_s] = \mathcal{E}_g[X|\,\mathcal{F}_s].$$

(iii) Let $X_1 \leq X_2$, then
$$\mathcal{E}_g[X_1|\,\mathcal{F}_t] \leq \mathcal{E}_g[X_2|\,\mathcal{F}_2].$$

(iv) For all $A \in \mathcal{F}_t$ we have $\mathcal{E}_g[X\mathbf{1}_A|\,\mathcal{F}_t] = \mathbf{1}_A\,\mathcal{E}_g[X|\,\mathcal{F}_t]$.

Proof. Property (i) is an immediate consequence of Definition 3.11.

(ii) Let $t > s$, then we know $\mathcal{F}_t \supseteq \mathcal{F}_s$. Thus, this yields

$$\mathcal{E}_g[\mathcal{E}_g[\mathcal{E}_g[X|\,\mathcal{F}_t]|\,\mathcal{F}_s]\mathbf{1}_A] = \mathcal{E}_g[X|\,\mathcal{F}_s].$$

Hence, by definition

$$\mathcal{E}_g[\mathcal{E}_g[X|\,\mathcal{F}_t]|\,\mathcal{F}_s] = \mathcal{E}_g[X|\,\mathcal{F}_s].$$

(iii) Define $\eta_1 \coloneqq \mathcal{E}_g[X|\,\mathcal{F}_t]$ and $\eta_2 \coloneqq \mathcal{E}_g[X_2|\,\mathcal{F}_t]$. Let $A \in \mathcal{F}_t$, then we have for $X_1 \geq X_2$ with Theorem 3.10 (ii) and Definition 3.11

$$\mathcal{E}_g[\eta_1\mathbf{1}_A] = \mathcal{E}_g[X\mathbf{1}_A]] \leq \mathcal{E}_g[Y\mathbf{1}_A] = \mathcal{E}_g[\eta_2\mathbf{1}_A].$$

If we define $A \coloneqq \{\eta_1 > \eta_2\}$ and assume $P(A) > 0$, then we get again by monotonicity

$$\mathcal{E}_g[\eta\mathbf{1}_A] > \mathcal{E}_g[\eta_2\mathbf{1}_A].$$

These observations yield $P(A) = 0$. Thus, we have $\mathcal{E}_g[X_1|\,\mathcal{F}_t] \leq \mathcal{E}_g[X_2|\,\mathcal{F}_2]$.

(iv) Let $B \in \mathcal{F}_t$, then we have by definition

$$\begin{aligned}\mathcal{E}_g[\mathcal{E}_g[X\mathbf{1}_A|\,\mathcal{F}_t]\mathbf{1}_B] &= \mathcal{E}_g[X\mathbf{1}_A\mathbf{1}_B] \\ &= \mathcal{E}_g[\mathcal{E}_g[X|\,\mathcal{F}_t]\mathbf{1}_{A\cap B}] \\ &= \mathcal{E}_g[[\mathcal{E}_g[X|\,\mathcal{F}_t]\mathbf{1}_A]\mathbf{1}_B].\end{aligned}$$

Hence, $\mathcal{E}_g[X\mathbf{1}_A|\,\mathcal{F}_t] = \mathbf{1}_A\,\mathcal{E}_g[X|\,\mathcal{F}_t]$.

□

Before concluding this section, we would like to clarify how the classical expectation and the g-expectation are related. The reader may also consider the remarks in Chen et al. (2005, p. 5).

Proposition 3.6
Let $X \in L^2_{\mathcal{F}_T}$ and $\mathcal{E}_g[X]$ the respective g-expectation according to Definition 3.10. If $g \equiv 0$, then the classical expectation and the g-expectation of the random variable X coincide, i.e., $\mathcal{E}_g[X] = \mathbb{E}[X]$. Hence, the classical expectation $\mathbb{E}[\cdot]$ is a special case of the g-expectation $\mathcal{E}_g[\cdot]$.

Proof. Let $X \in L^2_{\mathcal{F}_T}$ and $\mathcal{E}_g[X]$ the respective g-expectation, where we set $g \equiv 0$. This yields the following BSDE

$$\begin{cases} dY_t = Z_t dB_t, \quad 0 \le t \le T, \\ Y_T = X. \end{cases} \tag{3.2.2}$$

The we can equivalently state (3.2.2) by

$$Y_t = \underbrace{Y_T}_{=X} - \int_t^T Z_s dB_s. \tag{3.2.3}$$

Now, taking the conditional expectation on both sides of (3.2.3) implies

$$Y_t = \mathcal{E}[X|\,\mathcal{F}_t] = \mathbb{E}[X|\,\mathcal{F}_t].$$

If we fix $t = 0$, and recall that $\mathcal{F}_0 = \{\Omega, \emptyset\}$, then we conclude $Y_0 = \mathcal{E}_g[X] = \mathbb{E}[X]$, where we trivally have $\mathbb{E}[X|\,\mathcal{F}_0] = \mathbb{E}[X]$. □

The most important terms regarding g-expectations being introduced, we immediately see the relation to classical expectation. It is easy to see that, based on previous observations, g-expectations are filtration consistent, which is a particularly desirable property in expected-utility theory in dynamic situations. Such expectations are usually

referred to as $\mathcal{F}$-expectations. Before considering some applications of g-expectations, we explore the aforementioned relationship between Choquet and g-expectations, where we rely on results worked out in Chen et al. (2005). To avoid getting lost in the technical details, we only explain the main result and refer to Chen et al. (2005) for the proof and further technical details.

Theorem 3.11
Assume g satisfies the usual conditions (i.e., conditions stated in Definition 3.10). Then there exists a Choquet expectation $\mathcal{E}_c$ restricted to $L^2_{\mathcal{F}_T}$ equal to a g-expectation $\mathcal{E}_g$ if and only if g does not depend on y and is linear in z, i.e., there exists a continuous function $v(t)$ such that

$$g(y, z, t) = v(t)z.$$

Explicitly, this theorem states: Both expectations only coincide in the class of classical linear expectation.

3.3 Applications

Finally, we consider possible applications of g-expectations. Especially in mathematical finance, g-expectations enjoy high popularity. For applications in finance consider El Karoui et al. (1997). Further, Gianin (2006) defines risk measures via g-expectations. The link between risk measures and robust methods in statistics and machine learning leads to an interesting research question. Chouzenoux et al. (2019) considers risk measures developed in mathematical finance for robust machine learning. This approach illustrates that translating ideas from different academic communities can be of particular interest for future research. From a decision-theoretic point of view, based on g-expectations, an intertemporal counterpart to the multiple prior model can be formulated, referred to as *recursive multiple priors utility* (Chen and Epstein, 2002). Recursive multiple priors utilities represent a generalization of *stochastic differential utility* (Duffie and Epstein, 1992) to dynamic situations under uncertainty. Moreover, we note that g-expectations can be used for dynamic modeling in areas apart from mathematical finance. For example, Shamarova et al. (2017) models stochastic gene expression and protein level dynamics based on BSDEs. In this sense, employment in biostatistics or epidemiology is also conceivable.

4 Peng's G-Framework

4.1 Nonlinear expectation spaces

So far, we have seen that nonlinear expectations, such as the Choquet expectation or the g-expectation, are particularly suitable for modeling uncertainty. This section presents basic notions within the so-called G-Framework, which can be seen as an uncertainty counterpart of Kolmogorov's axiomatic setting. In particular, this new framework is based on Peng (2004, 2007a,b, 2008). We will rely on previously established results to clarify the connection between different parts of this thesis. We first fix some notations and elaborate on fundamental concepts. These are usually taken for granted. Nevertheless, since we are mainly concerned with making this framework understandable to a broad audience, some additional preparations are inevitable.

Definition 4.1 (Vector lattice)
Let Ω be an arbitrary set. A vector space $\mathcal{H}$ of real-valued functions $X : \Omega \to \mathbb{R}$ satisfying the conditions

(i) $\forall X, Y \in \mathcal{H} : X \vee Y \in \mathcal{H}$

(ii) $\forall X, Y \in \mathcal{H} : X \wedge Y \in \mathcal{H}$

is called vector lattice, where $X \vee Y = \max\{X, Y\}$ and resp. $X \wedge Y = \min\{X, Y\}$. Since we can equivalently write

$$X \wedge Y = X + Y - (X \vee Y)$$
$$X \vee Y = Y + (X - Y) \vee 0$$

for all $X, Y \in \mathcal{H}$, a vector space of real-valued functions is a vector lattice, if for all $f \in \mathcal{H}$ we also have $X^+ = X \vee Y \in \mathcal{H}$ (this was defined previously to be the positive part of a real-valued function). Now, let $\mathcal{H}$ be the vector space of real-valued functions such that $X \in \mathcal{H} \Rightarrow |\mathcal{H}| \in \mathcal{H}$. Then $\mathcal{H}$ is a vector lattice, since we can write

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

Thus in the above definition, it is sufficient to impose $X \in \mathcal{H} \Rightarrow |X| \in \mathcal{H}$ for all $X \in \mathcal{H}$.

Throughout this section, $\mathcal{H}$ denotes the vector space of real-valued functions satisfying the conditions

(i) $\forall c \in \mathbb{R} : c \in \mathcal{H}$

(ii) $X \in \mathcal{H} \Rightarrow |X| \in \mathcal{H}$

where condition (i) guarantees constant preserving. Furthermore $C_{l.lip}(\mathbb{R}^n)$ denotes the space of functions ψ satisyfing

$$|\psi(x) - \psi(y)| \leq C(1 + |x|^m + |y|^m)|x - y|$$

with $x, y \in \mathbb{R}^n, m \in \mathbb{N}$ and $C > 0$. We further assume that if $X_1, \ldots, X_n \in \mathcal{H}$ then we also have $\psi(X_1, \ldots, X_n) \in \mathcal{H}$ for all $\psi \in C_{l.lip}(\mathbb{R}^n)$. Technially $C_{l.lip}(\mathbb{R}^n)$ can be replaced by other suitable spaces, compare Peng (2010, p. 5).

Definition 4.2 (Sublinear expectation)
A functional $\mathcal{E} : \mathcal{H} \to \mathbb{R}$ satisfying

(i) Monotonicity:
$$\forall X \leq Y : \mathcal{E}[X] \leq \mathcal{E}[Y]$$

(ii) Constant preserving:
$$\forall c \in \mathbb{R} : \mathcal{E}[c] = c$$

(iii) Subadditivity:
$$\forall X, Y \in \mathcal{H} : \ \mathcal{E}[X + Y] \leq \mathcal{E}[X] + \mathcal{E}[Y]$$

(iv) Positive homogeneity:
$$\forall \lambda \geq 0 : \ \mathcal{E}[\lambda X] = \lambda \mathcal{E}[X]$$

is called sublinear expectation. A functional satisfying only (i) + (ii) is called nonlinear expectation. We will refer to $(\Omega, \mathcal{H}, \mathcal{E})$ as sublinear (resp. nonlinear) expectation space. Property (iii) is also known as the self-dominance property, since

$$\begin{aligned} & \mathcal{E}[X] = \mathcal{E}[(X - Y) + Y] \leq \mathcal{E}[X - Y] + \mathcal{E}[Y] \\ \Leftrightarrow \quad & \mathcal{E}[X] - \mathcal{E}[Y] \leq \mathcal{E}[X - Y]. \end{aligned}$$

For instance, the Choquet expectation (with respect to a 2-alternating capacity) or the g-expectation discussed in Section 3 is a nonlinear expectation (see also the discussion in Section 2). From a purely probabilistic perspective, it may seem bizarre why we are now devoting ourselves to expectation spaces instead of probability spaces. In the classic linear case, we know that there is a one-to-one correspondence between expectation and probability measure:

Theorem 4.1 (Daniell-Stone)
Let $\mathbb{E} : \mathcal{H} \to \mathbb{R}$ be linear functional on $\mathcal{H}$ satisyfing following conditions

i.) $\forall \lambda, c \in \mathbb{R} : \mathbb{E}[\lambda X + c] = \lambda \mathbb{E}[X] + c.$

ii.) $\forall X, Y \in \mathcal{H}$ *with* $X \geq Y : \mathbb{E}[X] \geq \mathbb{E}[Y]$.

iii.) *Let* $\{X_n\}_{n\in\mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ *with* $X_n \downarrow 0 \Rightarrow \mathbb{E}[X_n] \to 0$ *for* $n \to \infty$.

Then there exists a unique probability measure P *on* $(\Omega, \mathcal{F})$ *with* $\mathcal{F} \coloneqq \sigma(\mathcal{H})$ *such that* $P(A) = \mathbb{E}[1_A]$ *for all* $A \in \mathcal{F}$.

The proof and other related notions regarding *Daniell-Integration* can be found in Ash (2014). The reader may also find some helpful insights in Royden and Fitzpatrick (1988). However, in the nonlinear case, such a one-to-one correspondence is no more valid. Hence it is convenient to work with expectation spaces rather than probability spaces. The following example taken from Coquet et al. (2002, p. 2) illustrates such a failure of one-to-one correspondence.

Example 4.1
Let $(\Omega, \mathcal{F}, P)$ *be a probability space and* f *a strictly increasing function on* $\mathbb{R}$ *such that* $f(x) = x$ *for* $0 \leq x \leq 1$.
Now, define

$$\mathcal{E}^f[X] \coloneqq f^{-1}\left(\mathbb{E}[f(X)]\right). \tag{4.1.1}$$

Then (4.1.1) *is a nonlinear expectation, unless* f *is linear. Any expectation of the above form induces the same probability measure, i.e.,* $P(A) = \mathbb{E}[\mathbf{1}_A] = \mathcal{E}^f[\mathbf{1}_A]$ *for any* $A \subseteq \Omega$. *Hence, the one-to-one correspondence fails to hold.*

As in the case of Choquet expectation, the question arises whether and under which conditions sublinear expectations can be represented by linear expectations.

Theorem 4.2 (Representation of sublinear expectations)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ *be a sublinear expectation space. Then there exists a family of linear expectations* $\{\mathbb{E}_\theta\}_{\theta\in\Theta}$ *with* $\mathbb{E}_\theta : \mathcal{H} \to \mathbb{R}$, *such that*

$$\mathcal{E}[X] = \max_{\theta\in\Theta} \mathbb{E}_\theta[X] \quad \textit{for} \quad X \in \mathcal{H}. \tag{4.1.2}$$

Furthermore, for each $X \in \mathcal{H}$ *there exists* $\theta_X \in \Theta$ *such that* $\mathcal{E}[X] = \mathbb{E}_{\theta_X}[X]$.

Proof. Let $X \in \mathcal{H}$ and define $L \coloneqq \{\alpha X : \alpha \in \mathbb{R}\}$. Note, that L is clearly a subspace of $\mathcal{H}$, since $0 \cdot X = 0 \in L$, for all $\alpha \in \mathbb{R}$ by definition $\alpha X \in L$ and $\alpha X + \beta X = (\alpha + \beta)X \in L$. Then $\ell : L \to \mathbb{R} : \alpha X \mapsto \alpha\mathcal{E}[X]$ defines a linear functional on L such that

$$\forall x \in L : \ell(x) \leq \mathcal{E}(x).$$

Since $\mathcal{E}[\cdot]$ is in particular subadditive and positive homogeneous we know by applying the Hahn-Banach Theorem that there exists a linear functional $\mathbb{E}[\cdot]$ on $\mathcal{H}$ extending $\ell[\cdot]$ such that

(i) $\mathbb{E}_{|Y} = \ell$

(ii) $\forall X \in \mathcal{H} : \mathbb{E}[X] \leq \mathcal{E}[X]$

Hence, this linear functional $\mathbb{E}$ is dominated by $\mathcal{E}$. Thus we have $\mathbb{E}[X] = \mathcal{E}[X]$. Since $\mathcal{E}$ is per definition monotone, we have for $X \geq 0$

$$\mathbb{E}[X] = -\mathbb{E}[-X] \geq -\mathcal{E}[-X] \geq 0.$$

And by the constant preserving property of $\mathcal{E}$ it follows that

$$-\mathbb{E}[c] = \mathbb{E}[-c] \leq \mathcal{E}[-c] \quad \text{and} \quad \mathbb{E}[c] \leq \mathcal{E}[c] = c.$$

Hence, $\{\mathbb{E}_\theta\}_{\theta\in\Theta}$ is a family of linear expectations. □

This result was first proved for finite Ω in Huber (1981). Since then, many authors have established the theorem under different conditions.

Remark 4.1
The linear expectation in (4.1.2) *is presumably with respect to a finitely additive probability measure. To guarantee a σ-additive probability measure, we have to assume $\mathbb{E}[X_n] \to 0$ for $n \to \infty$ for $\{X_n\}_{n\in\mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ with $X_n \downarrow 0$. For such representation w.r.t. to σ-additive probability measure refer to Peng (2010, p. 7). See also later remarks on uncertain probability measures associated with sublinear expectations.*

Now that expectation spaces have been sufficiently motivated, elemental notions such as distribution of random variables under sublinear expectations and independence must be defined. Note that these terms are fundamentally different from the classical ones.

Definition 4.3 (Distribution under sublinear expectation)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space, and $X \in \mathcal{H}$. Then we call the functional

$$\mathbb{F}_X : C_{l.lip}(\mathbb{R}^n) \to \mathbb{R} : \phi \mapsto F_X[\psi] := \mathcal{E}[\psi(X)]$$

distribution of X under the sublinear expectation $\mathcal{E}$. Particularly $(\mathbb{R}, C_{l.lip}(\mathbb{R}), \mathbb{F}_X)$ is also a sublinear expectation space.
Let $(\Omega_1, \mathcal{H}_1, \mathcal{E}_1)$ and $(\Omega_2, \mathcal{H}_2, \mathcal{E}_2)$ be two sublinear expectation spaces. Two random variables $X, Y \in \mathcal{H}$ are identically distributed, if

$$\forall \psi \in C_{l.lip}(\mathbb{R}) : \mathcal{E}_1[\psi(X_1)] = \mathcal{E}_2[\psi(X_2)],$$

denoted by $X_1 \overset{d}{\sim} X_2$. More generally, we can replace sublinear with nonlinear in the above definition. However, since we were mainly concerned with sublinear expectations, this is always evident from the context.

The distribution of a random variable $X \in \mathcal{H}$ under a sublinear expectation $\mathcal{E}$ is typically characterized by the mean parameter $\mu := -\mathcal{E}[-X]$, $\bar{\mu} := \mathcal{E}[X]$ and the variance parameter $\underline{\sigma}^2 := -\mathcal{E}[-X^2]$, $\bar{\sigma}^2 := \mathcal{E}[X^2]$. In fact, $\underline{\mu} \leq \bar{\mu}$ holds, since:

$$\begin{aligned} & -\mathcal{E}[X + (-X)] = -\mathcal{E}[0] = 0 \geq -\mathcal{E}[-X] - \mathcal{E}[X] \\ \Leftrightarrow \quad & \mathcal{E}[X] \geq -\mathcal{E}[-X] \\ \Leftrightarrow \quad & \bar{\mu} \geq \underline{\mu}. \end{aligned}$$

In the first inequality, we used the fact that $-\mathcal{E}[-(X+Y)] \geq -\mathcal{E}[-X] - \mathcal{E}[-Y]$, which follows directly from the subadditivity property of sublinear expectations. Since $(\mathbb{R}, C_{l.lip}(\mathbb{R}), \mathbb{F}_X)$ is again a sublinear expectation space, the interesting question arises whether probability measures can also represent the distribution of a random variable under a sublinear expectation. The following two lemmas give an affirmative answer.

Lemma 4.1
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $X \in \mathcal{H}$. For any sequence $\{\psi_n\}_{n\in\mathbb{N}} \subset C_{l.lip}(\mathbb{R})$ satisyfing $\psi_n \downarrow 0$ for $n \to \infty$ we have $\mathcal{E}[\psi_n(X)] \downarrow 0$.

Proof. Let $\{\psi_n\}_{n\in\mathbb{N}} \subset C_{l.lip}(\mathbb{R})$ be a sequence sucht that $\psi_n \downarrow 0$ for $n \to \infty$. Define $k_{n,N} := \max_{|x|\leq N}$ for fixed $N > 0$. Then we have for all $x \in \mathbb{R}$

$$\psi_n(x) \leq k_{n,N} + \psi_1(x)\mathbf{1}_{|x|>N} \leq k_{n,N} + \frac{\psi_1(x)|x|}{N}.$$

This yields

$$\begin{aligned} & \mathcal{E}[\psi_n(X)] \leq k_{n,N} + \frac{1}{N}\mathcal{E}[\psi_1(X)|X|] \\ \Rightarrow \quad & \lim_{n\to\infty} \mathcal{E}[\psi_n(X)] \leq \frac{1}{N}\mathcal{E}[\psi_1(X)|X|], \end{aligned}$$

where the implication follows from $\psi_n \downarrow 0 \Rightarrow k_{n,N} \downarrow 0$. For N large enough we get the desired result. Thus, the proof is complete. □

Lemma 4.2
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $\mathbb{F}_X[\psi]$ the distribution of $X \in \mathcal{H}$. Then there exists a family of probability measures $\{P_{X,\theta}\}_{\theta\in\Theta}$ defined on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ such that

$$\mathbb{F}_X[\psi] = \sup_{\theta\in\Theta} \int_{\mathbb{R}} \psi(x) dP_{X,\theta}(x), \quad \psi \in C_{l.lip}(\mathbb{R}). \tag{4.1.3}$$

Proof. From Lemma 4.1 we know that a sequence $\{\psi_n\}_{n\in\mathbb{N}} \subset C_{l.lip}(\mathbb{R})$ satisyfing $\psi_n \downarrow 0$ for $n \to \infty$ yields $\mathbb{F}_X[\psi_n] \downarrow 0$. Hence, applying Theorem 4.1 completes the proof. □

Since sublinear distributions play a crucial role in this framework, we will elaborate on the meaning of such distributions and relate this to associated uncertainty measures. These considerations are also highly related to some considerations in the section on Choquet expectations.

Remark 4.2
For simplicity of techniques and illustration purposes, we restrict ourselves to finite Ω and a convex, closed set of prior distributions $\mathcal{P} \subseteq \mathcal{W}$, where $\mathcal{W}$ denotes the family of all probability measures on the measurable space $(\Omega, \mathcal{F})$. Now, define the functional $\mathcal{E}_1[X] = \sup_{P\in\mathcal{P}} \int X\, dP$. Then, by results obtained in Huber (1981), we know that such a set of prior distributions can be represented by

$$\mathcal{P} = \{P \in \mathcal{M} : \ \mathbb{E}_P[X] \leq \mathcal{E}_1[X] \quad \textit{for all } X\}\,, \tag{4.1.4}$$

where $\mathbb{E}_P[\cdot]$ denotes the classical expectation with respect to the probability measure P. Hence, $\mathcal{P}$ can be interpreted as the uncertainty sets of distributions associated with the sublinear expectation $\mathcal{E}_1[\cdot]$. Now, we consider a second prior distribution set $\mathcal{Q} \subseteq \mathcal{W}$ and define similarly a sublinear expectation $\mathcal{E}_2[X] = \sup_{Q\in\mathcal{Q}} \int X\, dQ$. Analogous argumentation yields the representation

$$\mathcal{Q} = \{Q \in \mathcal{M} : \ \mathbb{E}_Q[X] \leq \mathcal{E}_2[X] \quad \textit{for all } X\}\,. \tag{4.1.5}$$

Hence, $\mathcal{Q}$ denotes the uncertainty sets associated with the sublinear expectation $\mathcal{E}_2[\cdot]$. Thus, clearly $\mathcal{E}_1[X] \leq \mathcal{E}_2[X]$ implies $\mathcal{P} \subseteq \mathcal{Q}$. A stronger sublinear expectation (hence distribution) accounts for a larger associated uncertainty set in sublinear expectation spaces, i.e., for $\psi \in C_{l.lip}(\mathbb{R})$ we have

$$\mathbb{F}_{X_1}[\psi] \ \leq \mathbb{F}_{X_2}[\psi] \Rightarrow \{P_{X_1,\theta}\}_{\theta\in\Theta_1} \subseteq \{P_{X_2,\theta}\}_{\theta\in\Theta_2}.$$

From Definition 4.3, a notion of convergence can be derived directly, which should be understood analogously to the classical notion of convergence in distribution.

Definition 4.4 (Convergence in distribution)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $\{X_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$ a sequence of random variables. Then $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to $X \in \mathcal{H}$ if

$$\forall \psi \in C_{l.lip}(\mathbb{R}) : \lim_{n\to\infty} \mathbb{F}_{X_n}[\psi] = \mathbb{F}_X[\psi].$$

Note that $\mathbb{F}_X[\psi]$ is a sublinear expectation on $(\mathbb{R}, C_{l.lip}(\mathbb{R}))$.

Now we turn to another important aspect, the independence of random variables. We will motivate the notion of independent random variables under sublinear expectation by starting with a result regarding the usual notion of independence.

Proposition 4.1
Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ two random variables. If X, Y are independent and $\mathbb{E}[|\psi(X,Y)|] < \infty$ we have

$$\mathbb{E}[\psi(X,Y)|X] = \mathbb{E}[\psi(x,Y)]_{x=X}, \tag{4.1.6}$$

or equivalently stated

$$\mathbb{E}[\psi(X,Y)] = \mathbb{E}[\mathbb{E}[\psi(x,Y)]_{x=X}]. \tag{4.1.7}$$

The last equations follows from the tower property of conditional expectations (see respective Appendix).

Proof. Let $h(x) = \mathbb{E}[\psi(x,Y)]$ and $\sigma(X) = \{X^{-1}(B) : B \in \mathfrak{B}(\mathbb{R})\}$, where $\sigma(X)$ is the σ-algebra generated by X. Since $h(X)$ is $\sigma(h(X))$-measurable by definition, and $\sigma(h(X)) \subset \sigma(X)$, we know that $h(X)$ is also $\sigma(X)$-measurable. Further, we know that for any $A \in \sigma(X)$ there exists $B \in \mathfrak{B}(\mathbb{R})$ such that $A = X^{-1}(B)$. Thus, this yields

$$\begin{aligned}
\int \psi(X,Y)\mathbf{1}_A dP &= \int \psi(x,y)\mathbf{1}_B(x) dP_{X,Y}(x,y) \\
&= \int \left(\int \psi(x,y) dP_Y(y) \right) \mathbf{1}_B(x) dP_X(x) \\
&= \int \mathbb{E}[\psi(X,Y)]\mathbf{1}_A dP.
\end{aligned}$$

Hence, we conclude that $\mathbb{E}[\psi(x,Y)]_{x=X}$ is the conditional expectation of $\psi(X,Y)$ given X. In the second equality, we used the independence assumption. □

This observation will serve us as the definition of independence in the sublinear expectation framework. However, in the sublinear case, one needs to pay attention since X independent of Y does no more imply that Y is also independent of X.

Definition 4.5 (Independence)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $X, Y \in \mathcal{H}$ two random variables. Then Y is independent of X (under the sublinear expectation $\mathcal{E}$) if we have

$$\forall \psi \in C_{l.lip}(\mathbb{R}^2) : \ \mathbb{E}[\psi(X,Y)] = \mathbb{E}[\mathbb{E}[\psi(x,Y)]_{x=X}]. \tag{4.1.8}$$

Two random variables $X, Y \in \mathcal{H}$ are called identical and independent (i.i.d) if $X \overset{d}{\sim} Y$ and Y is independent of X.

For further perspective on the independence in the G-Framework, the reader may also refer to Hu and Li (2014). The following example based on Peng (2010, p. 11 ff.) illustrates the aforementioned specificity of the notion of independence in this framework

Example 4.2
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $X, Y \in \mathcal{H}$ two identically distributed random variables with $\mathcal{E}[X] = -\mathcal{E}[-X] = 0$ and $\bar{\sigma}^2 := \mathcal{E}[X^2] \geq \underline{\sigma}^2 := -\mathcal{E}[-X^2]$. Now, we assume $\mathcal{E}[|X|] = \mathcal{E}[X^+ + X^-] > 0$, where we used $|X| = X^+ + X^-$. This implies $\mathcal{E}[X^+] = \frac{1}{2}\mathcal{E}[[X| + X] = \frac{1}{2}\mathcal{E}[|X|] > 0$, where we used $X^+ = \frac{|X|+X}{2}$ in the first equation. Define $\psi(x, y) = xy^2$.

Now, consider the following cases:

1.) Y is independent of X:

$$\begin{aligned}\mathcal{E}[\psi(X,Y)] &= \mathcal{E}[\mathcal{E}[\psi(x,Y)]_{x=X}] \\ &= \mathcal{E}[\mathcal{E}[xY^2]_{x=X}] \\ &= \mathcal{E}[X^+\bar{\sigma}^2 - X^-\underline{\sigma}^2] \\ &= \mathcal{E}[X^+](\bar{\sigma}^2 - \underline{\sigma}^2) \\ &> 0\end{aligned}$$

2.) X is independent of Y:

$$\begin{aligned}\mathcal{E}[\psi(X,Y)] &= \mathcal{E}[\mathcal{E}[\psi(X,y)]_{y=Y}] \\ &= \mathcal{E}[\mathcal{E}[Xy^2]_{y=Y}] \\ &= \mathcal{E}[Y^2\,\mathcal{E}[X]] \\ &= 0\end{aligned}$$

Hence Y independent of X is not equivalent to the converse statement.

Now that the notion of independence has also been established, it is natural to consider product spaces.

Definition 4.6 (Product space)
Let $(\Omega_1, \mathcal{H}_1, \mathcal{E}_1)$ and $(\Omega_2, \mathcal{H}_2, \mathcal{E}_2)$ be two sublinear expectation spaces. Then we call the sublinear expectation space $(\Omega_1 \times \Omega_2, \mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ product space, where we define

$$\mathcal{H}_1 \otimes \mathcal{H}_2 := \{Z(\omega_1, \omega_2) = \psi(X(\omega_1), Y(\omega_2)) : (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2,$$

$$X \in \mathcal{H}_1, Y \in \mathcal{H}_2, \psi \in C_{l.lip}(\mathbb{R}^2)\}$$

and for $Z \in \mathcal{H}_1 \otimes \mathcal{H}_2$

$$(\mathcal{E}_1 \otimes \mathcal{E}_2)[Z] \coloneqq \mathcal{E}_1[\mathcal{E}_2[\phi(X,Y)]].$$

Analogously we define

$$\left(\prod_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{H}_i, \bigotimes_{i=1}^{n} \mathcal{E}_i\right), \quad n \in \mathbb{N}.$$

If we have $(\Omega_i, \mathcal{H}_i, \mathcal{E}_i) = (\Omega_1, \mathcal{H}_1, \mathcal{E}_1)$ *for* $i \in \{1, \dots, n\}$*, then we write*

$$\left(\Omega_1^n, \mathcal{H}_1^{\otimes n}, \mathcal{E}_1^{\otimes n}\right).$$

Now, let $(\Omega_1, \mathcal{H}_1, \mathcal{E}_1)$ be a sublinear expectation space, and $X, Y \in \mathcal{H}$ two identically distributed random variables, i.e. $X \overset{d}{\sim} Y$.
Then

$$(X,Y) : \Omega \to \mathbb{R}^2 : \omega \mapsto (X(\omega), Y(\omega))$$

is a random vecotor, and $(\mathbb{R}^2, C_{l.lip}(\mathbb{R}^2), \mathbb{F}_{X,Y})$ a sublinear expectation space, where $\mathbb{F}_{X,Y}$ denotes the joint distribution of X and Y. If Y is independent of X, then we have in terms of Definition 4.6 for $\psi \in C_{l.lip}(\mathbb{R}^2)$

$$\begin{aligned}
\mathbb{F}_{X,Y}[\psi] &= \mathcal{E}_1^{\otimes 2}[\psi(X,Y)] \\
&= \mathcal{E}_1[\mathcal{E}_1[\psi(X,Y)]] \\
&= \sup_{\theta_1 \in \Theta} \int_{\mathbb{R}} \left[\sup_{\theta_2 \in \Theta} \int_{\mathbb{R}} \psi(x,y) dP_{\theta_2}(y) \right] dP_{\theta_1}(x).
\end{aligned}$$

Where the last equality is a consequence of Lemma 4.2 and the assumption that X and Y are identically distributed. The next proposition illustrates how to construct independent random variables defined on sublinear expectation spaces.

Proposition 4.2
Let X_i *be random variables on sublinear expectation spaces* $(\Omega_i, \mathcal{H}_i, \mathcal{E}_i)$ *for* $i \in \{1, \dots n\}$*. Now, define projections by*

$$Y_i : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R} : (\omega_1, \dots, \omega_n) \mapsto X_i(\omega_i).$$

Then $Y_i \overset{d}{\sim} X_i$ *and* Y_{i+1} *is independent of* $(Y_1, \dots, Y_i)$ *for all* $i \in \{1, \dots, n-1\}$*. Further,* Y_i *is for all* $i \in \{1, \dots, n\}$ *a random vector on the product space* $(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{H}_i, \bigotimes_{i=1}^n \mathcal{E}_i)$.

Moreover, the proposition is easy to interpret since an analogous construction of independent random variables on a product space is possible in the classical case. We conclude this first section on the G-Framework with an example that illustrates previous concepts. This example is based on Peng (2010, p. 15 ff.).

Example 4.3
There are 100 balls in an urn, which are either black b_1 or white w_1. Now, define the following random variable

$$X_1 : \{b_1, w_1\} \to \{1, -1\} : \omega \mapsto X_1(\omega) = \mathbf{1}_{\{b_1\}}(\omega) - \mathbf{1}_{\{w_1\}}(\omega). \tag{4.1.9}$$

Let $c \in [\underline{\mu}, \bar{\mu}]$ with $0 \leq \underline{\mu} < \bar{\mu} \leq 1$ denote the unknown proportion of black balls in the urn, then the we have the following law

$$P_X = \begin{bmatrix} X & 1 & -1 \\ P & 1-c & c \end{bmatrix}.$$

This game is repeated, where c can be changed in each round within the above range $[\underline{\mu}, \bar{\mu}]$. In this way, we now construct a sequence of random variables $\{X_i\}_{i \in \mathbb{N}}$. We consider now the sublinear expectation

$$\mathcal{E}[\psi(X_i)] = \max_{c \in [\underline{\mu}, \bar{\mu}]} \left[c\psi(1) + (1-c)\psi(-1) \right], \tag{4.1.10}$$

where ψ can be thought of as a suitable loss function. Since for all $i, j = 1, 2, \ldots$ we have $\mathcal{E}[\psi(X_i)] = \mathcal{E}[\psi(X_j)]$, the sequence of random variables is identically distributed. Further, X_{i+1} is independent of $(X_1, \ldots, X_i)$, since by definition (4.1.10) *we know that*

$$\mathbb{E}[\psi(X_1, \ldots, X_i, X_{i+1})] = \mathbb{E}[\mathbb{E}[\psi(x_1, \ldots, x_i, X_{i+1})]_{x_j = X_j, 1 \leq j \leq i}].$$

The above sequence is a Bernoulli sequence under distributional uncertainty.

4.2 (Viscosity) solutions of parabolic PDE

This section deals with the (viscosity) solution of a particular partial differential equation (PDE), the G-heat equation. PDEs rarely appear in classical literature on probability theory. However, for our purposes, it is particularly worthwhile to look at the classical heat equation since we will see later that in the nonlinear case, there is a close connection with the so-called G-normal distribution and G-heat equation, which is a core utensil in the G-Framework. Although this section goes an extra step, understanding this connection is worth the effort. For this purpose, it is inevitable to make a few preparations regarding PDEs. Our presentation is mainly based on Evans (2010, 2012). At the beginning of this thesis, we stated that ordinary differential equations relate a real- or complex-valued function with its derivatives. Partial differential equations link an unknown function of two or more variables with their partial derivatives.

Definition 4.7 (Partial differential equation)
For $U \subset \mathbb{R}^d$ open and given function

$$F : \mathbb{R}^{n^k} \times \mathbb{R}^{n^{k-1}} \times \cdots \times \mathbb{R}^n \times \mathbb{R} \times U \to \mathbb{R}$$

the equation

$$F\left(D^k u(x), D^{k-1}u(x), \ldots, Du(x), u(x), x\right) = 0, \quad x \in U \tag{4.2.1}$$

with $u : U \to \mathbb{R}$ unknown is called k-th order partial differential equation. A solution of the PDE (4.2.1) *consists of all functions u satisfying this equation.*

We proceed now as follows: First, we will formulate the so-called heat equation and derive the *fundamental solution*[10]. Then we will consider how we can study the heat equation and its respective solution from a probabilistic point of view. The (homogeneous) heat equation (also called the diffusion equation)

$$\boxed{u_t = \alpha u_{xx}, \quad \alpha > 0 \quad \text{(Diffiusion constant)}} \tag{4.2.2}$$

is of great interest in physics and many related areas of applied mathematics. The heat equation describes the distribution of some quantity (like heat) in a given body. In (4.2.2) we apply following notation

$$u_t = \frac{\partial u}{\partial t}, \quad u_{xx} = \frac{\partial^2 u}{\partial x^2},$$

where $u : U \times \mathbb{R}_{\geq 0} : (x,t) \mapsto u(x,t)$ denotes the unknown function. If we now assume that a function $u(x,t)$ satisfies the heat equation (4.2.2), then necessarily $u(\sqrt{\lambda}x, \lambda t)$

[10] We will be concerned for simplicity reasons with one spatial dimension. See Evans (2010, p. 44 ff.) for a more general solution.

must be a solution to this PDE. Hence, we would expect solutions of the form $u(\frac{x}{\sqrt{t}})$. In particular, we are now looking for a function of the form

$$u(x,t) = \frac{1}{t^\beta} v\left(\frac{x}{\sqrt{t}}\right), \quad x \in \mathbb{R},\, t > 0. \tag{4.2.3}$$

Rougly speaking, the extra term in (4.2.3) reduces the PDE to an ODE in later steps. Now, we plug (4.2.3) in the heat equation (4.2.2) and get

$$\frac{\partial}{\partial t}\left(t^{-\beta} v\left(xt^{-1/2}\right)\right) = \alpha \frac{\partial^2}{\partial x^2}\left(t^{-\beta} v\left(xt^{-1/2}\right)\right).$$

Applying the product rule and simplifying yields

$$\beta v\left(\frac{x}{\sqrt{t}}\right) + \frac{x/\sqrt{t}}{2} v^{(1)}\left(\frac{x}{\sqrt{t}}\right) + \alpha v^{(2)}\left(\frac{x}{\sqrt{t}}\right) = 0, \tag{4.2.4}$$

where we denote with $v^{(1)}$ (resp. $v^{(2)}$) the derivative with respect to $\frac{x}{\sqrt{t}}$. Hence by substituting $y = \frac{x}{\sqrt{t}}$ we get from (4.2.4)

$$\beta v(y) + \frac{y}{2} v^{(1)}(y) + \alpha v^{(2)}(y) = 0. \tag{4.2.5}$$

Thus, we have transformed the original PDE into an ODE. Now, fix $\beta = \frac{1}{2}$. Then we get

$$\begin{aligned}
&\frac{1}{2}\left(v(y) + y v^{(1)}(y)\right) + \alpha v^{(2)}(y) = 0 \\
\Rightarrow \quad &\alpha v^{(2)}(y) + \frac{1}{2}\left(y v(y)\right)^{(1)} = 0 \\
\Rightarrow \quad &\left(\alpha v^{(1)}(y) + \frac{1}{2} y v(y)\right)^{(1)} = 0 \\
\Rightarrow \quad &\alpha v^{(1)}(y) + \frac{1}{2} y v(y) = C,
\end{aligned}$$

where $C =$ constant. Hence the last equation is now a first order ODE. We fix now $C = 0$ and get

$$\begin{aligned}
&\frac{v^{(1)}(y)}{v(y)} = -\frac{y}{2\alpha} \\
\Rightarrow \quad &\ln\left(|v(y)|\right)^{(1)} = -\frac{y}{2\alpha}
\end{aligned}$$

$$\Rightarrow \qquad \ln\left(|v(y)|\right) = -\frac{y^2}{4\alpha} + \tilde{C},$$

where $\tilde{C} =$ constant. Solving for $v(y)$ yields

$$v(y) = \hat{C} \exp\left(-\frac{y^2}{4\alpha}\right), \tag{4.2.6}$$

where $\hat{C} =$ constant. Now, fixing $\hat{C} = \frac{1}{\sqrt{4\pi\alpha}}$ and plugging (4.2.6) in (4.2.3) yields

$$\boxed{u(x,t) = \frac{1}{\sqrt{4\pi\alpha t}} \exp\left\{\frac{-x^2}{4\alpha t}\right\}.} \tag{4.2.7}$$

Cleary, $u(x,t)$ is for $t > 0$ and $\alpha = \frac{D}{2}$ the density of the $\mathcal{N}(0, Dt)$ distribution.

Definition 4.8 (Fundamental solution)
The function

$$f(x,t) \coloneqq \begin{cases} \frac{1}{\sqrt{4\pi\alpha t}} \exp\left\{\frac{-x^2}{4\alpha t}\right\}, & \text{for } x \in \mathbb{R},\ t > 0 \\ 0, & \text{for } x \in \mathbb{R},\ t < 0 \end{cases}$$

is called the fundamental solution of the heat equation (4.2.2)*. Compare also Figure 3.*

Before turning to the probabilistic interpretation, we clarify what is meant by an *initial-value problem* or *Cauchy problem* in terms of the heat equation.

Intial-value problem:
Let f be the fundamental solution of the heat eqaution (4.2.2). Then we call

$$\begin{cases} u_t = \alpha u_{xx}, & \text{on } x \in \mathbb{R} \times (0, \infty) \\ u = g, & \text{on } x \in \mathbb{R} \times \{t = 0\}. \end{cases} \tag{4.2.8}$$

the initial-value problem of the heat equation. Then, without proof, we claim that the *convolution*

$$u(x,t) = f \star g(x) \coloneqq \int_{\mathbb{R}} f(x-y,t) g(y)\, dy$$

solves the initial-value problem (4.2.8), where $f(x-y,t)$ is defined according to Definition 4.8. For a rigorous proof the reader may refer to Evans (2010, p. 47 ff.). Now we turn to the probabilistic interpretation of the heat equation and try to motivate a probabilistic derivation. This approach is closely related to Einstein et al. (1905) and Bachelier (1900). Now, let the real line be divided in intervals of the length $\triangle x$

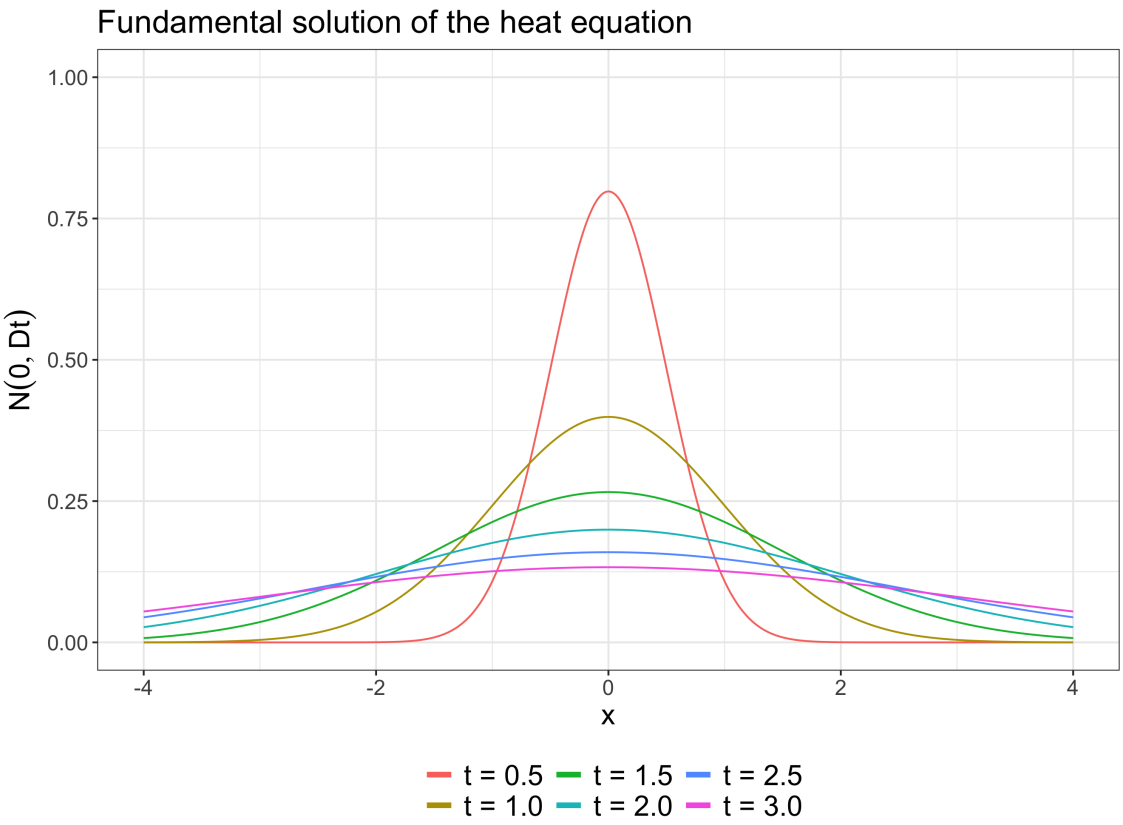

Figure 3: Visualization of the fundamental solution for fixed $D = 2$ and different t.

and consider time increments $\triangle t$. Hence, this setting can be thought of as a two-dimensional lattice with a space and time component (see Figure 4 for a visualization of this idea). We start at the initial point $(x_0, t_0) = (0, 0)$ and in each time step we do a fair Bernoulli experiment, such that with probability $\frac{1}{2}$ we move $\triangle x$ to the left and with the same probabilty an amount $\triangle x$ to the right. Now let $\{X_i\}_{i \in \mathbb{N}}$ be a family of random variables such that

$$P(X_i = 0) = P(X_i = 1) = \frac{1}{2}.$$

Based on this family of random variables, we define the number of total moves to right in $\triangle t = n \triangle t$ time increments, i.e.,

$$S_n \coloneqq \sum_{i=1}^{n} X_i.$$

Hence, S_n is a Binomial distributed random variable. By denoting $X(t)$ the posiiton

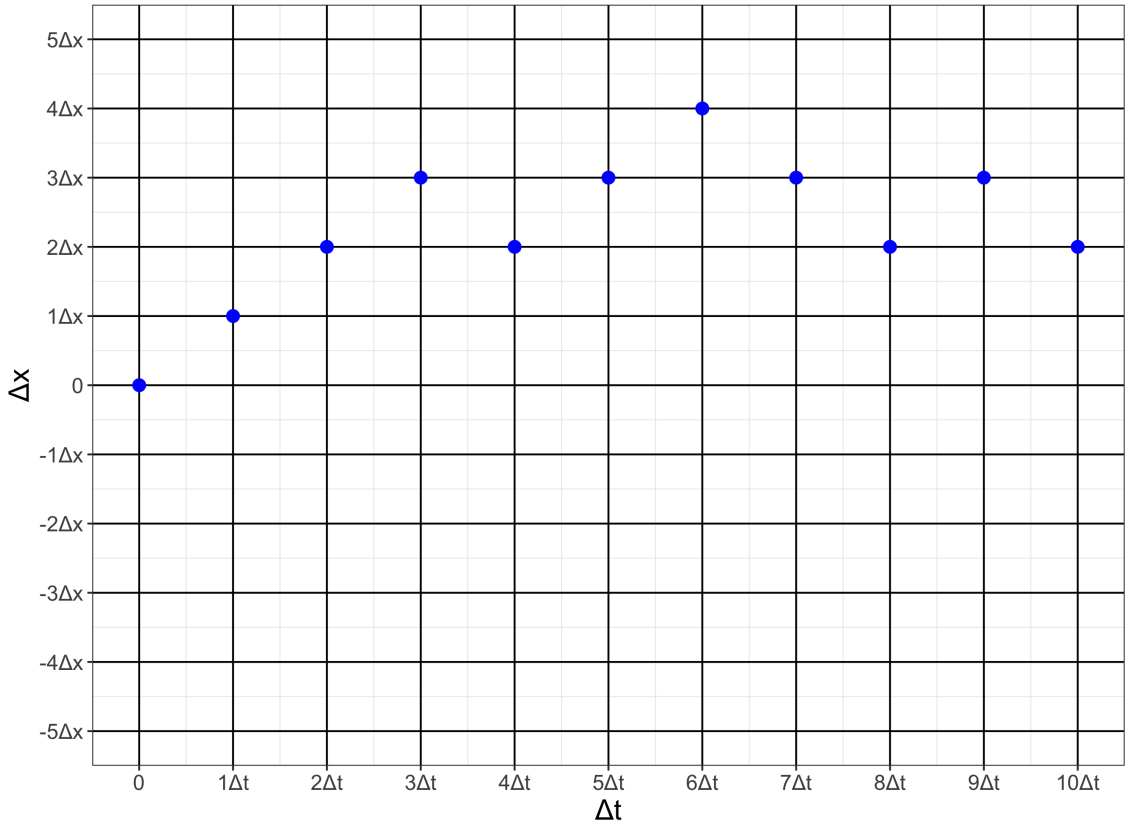

Figure 4: Visualization of time-space lattice and an examplary realization of random walk with initial point $(0, 0)$.

at time $n\triangle t$, we get

$$\begin{aligned} X(t) = & \underbrace{S_n \triangle x}_{\text{number of moves to the right times } \triangle x} + \underbrace{(n - S_n)(-\triangle x)}_{\text{number of moves to the left times } -\triangle x} \\ & = (2S_n - n)\triangle x. \end{aligned} \tag{4.2.9}$$

By assuming the time-space scaling $D = \frac{(\triangle x)^2}{\triangle t}$ and plugging in (4.2.9) we get

$$X(t) = (2S_n - n)\triangle x = \left(\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \right) \sqrt{tD}.$$

Finally, applying the Central-Limit Theorem yields

$$\lim_{n\to\infty} \; P(a \leq X(t) \leq b) = \lim_{n\to\infty} \left(\frac{a}{\sqrt{tD}} \leq \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \leq \frac{b}{\sqrt{tD}} \right)$$

$$= \frac{1}{\sqrt{2\pi Dt}} \int_a^b \exp\left\{-\frac{x^2}{2Dt}\right\}.$$

This again gives us a $\mathcal{N}(0, Dt)$ distribution, whose density again solves the heat equation (4.2.2), where we set $\alpha = \frac{D}{2}$.

Remark 4.3
At this point, we make some technical remarks clarifying any ambiguities about the terms involved so far in this section:

(i) A linear, second-order PDE with constant coefficients is of the form

$$au_{xx} + 2bu_{xy} + cu_{yy} + du_x + eu_y + f = 0.$$

If $b^2 - 4ac = 0$, we call the respective equation parabolic. Thus, the heat equation (4.2.2) is a parabolic PDE.

(ii) In the context of nonlinear PDEs we will use the notion of viscosity solution. However, to avoid getting lost in technical details, these should be understood as classical solutions.

Now that the relationship between the heat equation and the Normal distribution has been clarified, we formulate a nonlinear version of the heat equation. However, first, we formulate a classical result concerning Normal distributed random variables to clarify the latter definition of so-called G-normal distributed random variables.

Proposition 4.3
Let $(\Omega, \mathcal{F}, P)$ be a probability space, and X, Y two i.i.d (in the classical sense) real-valued random variables. Then, $X \sim \mathcal{N}(0, \sigma^2)$ if and only if $aX + bY$ and $\sqrt{a^2 + b^2}X$ are identically distributed, for $a, b \geq 0$.

See also Mathai and Pederzoli (1977) for related charaterizations of the Normal distirbution. Analogously, this proposition serves as a definition of G-normal distributed random variables in the sublinear case, hence:

Definition 4.9 (G-normal distribution)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $X \in \mathcal{H}$ a random variable with

$$\bar{\sigma}^2 = \mathcal{E}[X^2], \quad \underline{\sigma}^2 = -\mathcal{E}[-X^2].$$

Then, X is G-normal distributed, denoted by $X \overset{d}{\sim} \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, if for all $Y \in \mathcal{H}$ independent of X and $Y \overset{d}{\sim} X$ we have

$$\forall a, b \geq 0 : aX + bY \overset{d}{\sim} \sqrt{a^2 + b^2}X. \tag{4.2.10}$$

The random variable X has no mean-uncertainty since from (4.2.10) we know that

(i) $\sqrt{2}\,\mathcal{E}[X] = \mathcal{E}[X+Y] = 2\,\mathcal{E}[X] \Rightarrow \mathcal{E}[X] = 0.$

(ii) $\sqrt{2}\,\mathcal{E}[-X] = \mathcal{E}[-X-Y] = 2\,\mathcal{E}[-X] \Rightarrow \mathcal{E}[-X] = 0.$

Hence, X has no mean uncertainty. However, in the following section, when we look at the Law of Large Numbers and the Central Limit Theorem, we will also consider distributions with mean uncertainty.

Consider the sublinear expectation space $(\Omega, \mathcal{H}, \mathcal{E})$ and a random variable $X \in \mathcal{H}$. Then we call

$$\begin{cases} u_t = G(u_{xx}), & \text{on } x \in \mathbb{R} \times (0, \infty) \\ u = \psi, & \text{on } x \in \mathbb{R} \times \{t = 0\} \end{cases} \tag{4.2.11}$$

initial-value problem of the G-heat equation, where we refer to the nonlinear function

$$G(\alpha) := \frac{1}{2}\,\mathcal{E}[\alpha X^2] = \frac{1}{2}(\bar{\sigma}^2\alpha^+ - \underline{\sigma}^2\alpha^-),\ \alpha \in \mathbb{R} \tag{4.2.12}$$

as the *generating-function* of the nonlinear heat-equation, e.g. see Figure 5. This PDE generates the *G-normal distribution* $\mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. "Generating" can be understood as already in the classical case of the heat equation: The solution of the nonlinear heat equation is the G-normal distribution. We will now show that this statement is indeed true. This is one of the few results we actually prove in this section since previous concepts are directly involved in the proof.

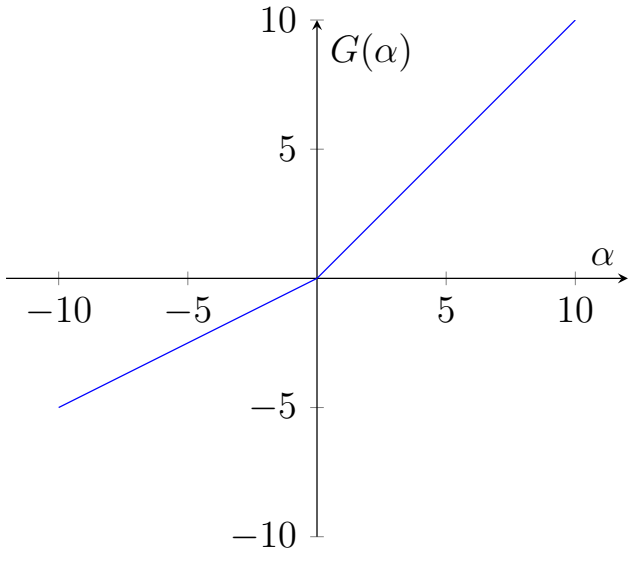

Figure 5: Example of a generating function G.

Definition 4.10 (Sub- and supersolution)
Consider a real-valued function $u \in \mathcal{C}([0,T] \times \mathbb{R})$.

(i) If for all $\psi \in \mathcal{C}_b^{1,3}((0,\infty) \times \mathbb{R})$ and all minima $(t,x) \in (0,\infty) \times \mathbb{R}$ of $\psi - u$ we have

$$\psi_t - G(\psi_{xx}) \leq 0$$

then we call u subsolution of the G-heat equation (4.2.11).

(ii) If for all $\psi \in \mathcal{C}_b^{1,3}((0,\infty) \times \mathbb{R})$ and all maxima $(t,x) \in (0,\infty) \times \mathbb{R}$ of $\psi - u$ we have

$$\psi_t - G(\psi_{xx}) \geq 0$$

then we call u supersolution of the G-heat equation (4.2.11).

(iii) If u is a sub- and supersolution, then we call u a viscosity solution of the G-heat equation.

Theorem 4.3
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and X a G-normal distributed random variable. For all $\psi \in C_{l.lip}(\mathbb{R})$ we define

$$u : [0,\infty) \times \mathbb{R} \to \mathbb{R} : \ (t,x) \mapsto u(t,x) \coloneqq \mathcal{E}[\psi(x, \sqrt{t}X)].$$

Then we have:

(i) $\forall s \geq 0 : \ u(t+s,x) = \mathcal{E}[u(t, x+\sqrt{s}X)]$.

(ii) For all $T > 0$, there exist constants $C, k > 0$ such that, for all $t, s \in [0,T]$ and $x, y \in \mathbb{R}$ we have

$$|u(t,x) - u(t,y)| \ \leq C(1 + |x|^k + |y|^k)|x - y| \tag{4.2.13}$$

$$|u(t,x) - u(t+s,x)| \ \leq C(1 + |x|^k)|s|^{1/2} \tag{4.2.14}$$

(iii) u is the unique viscosity solution of the G-heat equation (4.2.11).

Proof. For $T > 0$ and $s, t \in [0,T]$ we have

$$\begin{aligned} u(t,x) - u(t,y) &= \mathcal{E}[\psi(x + \sqrt{t}X)] - \mathcal{E}[\psi(x + \sqrt{t}X)] \\ &\leq \mathcal{E}[\psi(x + \sqrt{t}X) - \psi(y + \sqrt{t}X)] \\ &\leq \mathcal{E}[\tilde{C}(1 + |X|^k + |x|^k + |y|^k)|x - y|] \end{aligned}$$

$$\leq C(1+|x|^k+|y|^k)|x-y|,$$

where the third inequality follows by the assumption $\psi \in C_{l.lip}(\mathbb{R})$. Thus, (4.2.13) follows. Let Y be independent of X, such that we have

$$\begin{aligned} u(t+s,x) &= \mathcal{E}[\psi(x+\sqrt{t+s}X)] \\ &= \mathcal{E}[\psi(x+\sqrt{s}X+\sqrt{t}Y)] \\ &= \mathcal{E}[\mathcal{E}[\psi(x+\sqrt{s}z+\sqrt{t}Y)]_{z=X}] \\ &= \mathcal{E}[u(t,x+\sqrt{s}X)], \end{aligned}$$

where the second equality follows since X is G-Normal distributed, see Definition 4.9. Hence, (i) follows. With (i) we get

$$\begin{aligned} u(t+s,x)-u(t,x) &= \mathcal{E}[u(t,x+\sqrt{s}X)-u(t,x)] \\ &\leq \mathcal{E}[\tilde{C}(1+|x|^k+|X|^k)|s|^{1/2}|X|] \\ &\leq C(1+|x|^k)|s|^{1/2}. \end{aligned}$$

This yields (4.2.14). We now prove that u is a subsolution of the G-heat equation (4.2.11). Now, let $\psi \in C_b^{1,3}([0,\infty)\times\mathbb{R})$ such that $\psi \geq u$ and $u(t,x)=\psi(t,x)$ for fixed minima $(t,x)\in(0,\infty)\times\mathbb{R}$. Then, we have for $\delta\in(0,t)$

$$\begin{aligned} \psi(t,x)=\psi((t-\delta)+\delta,x) &= u((t-\delta)+\delta,x) \\ &\overset{\text{(i)}}{=} \mathcal{E}[u(t-\delta,\sqrt{\delta}X)] \\ &\leq \mathcal{E}[\psi(t-\delta,\sqrt{\delta}X)]. \end{aligned}$$

Thus, rearranging yields

$$\begin{aligned} & 0 \leq \mathcal{E}[\psi(t-\delta,\sqrt{\delta}X)]-\psi(t,x) \\ \Rightarrow \quad & 0 \leq \mathcal{E}[\psi(t-\delta,\sqrt{\delta}X)-\psi(t,x)]. \end{aligned}$$

By Taylor expansion we get

$$\begin{aligned} 0 &\leq \mathcal{E}[\psi(t-\delta,\sqrt{\delta}X)-\psi(t,x)] \\ &\overset{(1)}{\leq} -\psi_t(t,x)\delta+\mathcal{E}[\psi_x(t,x)\sqrt{\delta}X+\frac{1}{2}\psi_{xx}(t,x)\delta X^2]+\tilde{C}\delta^{3/2} \\ &= -\psi_t(t,x)\delta+\mathcal{E}[\frac{1}{2}\psi_{xx}(t,x)\delta X^2]+\tilde{C}\delta^{3/2} \end{aligned}$$

$$= -\psi_t(t,x)\delta + \delta G(\psi_{xx}(t,x)) + \tilde{C}\delta^{3/2}.$$

Thus, we conclude

$$\psi_t - G(\psi_{xx}) \leq 0.$$

Similarly one shows, that u is a supersolution of the G-heat equation (4.2.11). Hence, u is a viscosity solution. This completes the proof. □

Remark 4.4
We see that the G-normal distribution $\mathcal{N}(0,[\underline{\sigma}^2,\bar{\sigma}^2])$ is a sublinear distribution defined on $(\mathbb{R}, C_{l.lip}(\mathbb{R}))$. Let $X \overset{d}{\sim} \mathcal{N}(0,[\underline{\sigma}^2,\bar{\sigma}^2])$, then by Lemma 4.2 we know that there exists a family of probability measures $\{P_{X,\theta}\}_{\theta\in\Theta}$ defined on $(\mathbb{R},\mathfrak{B}(\mathbb{R}))$ such that

$$\mathbb{F}_X[\psi] = \sup_{\theta\in\Theta}\int_{\mathbb{R}} \psi(x) dP_{X,\theta}(x). \tag{4.2.15}$$

Now we could carelessly conclude that the uncertainty set associated with the G-normal distribution corresponds to $\{\mathcal{N}(0,\sigma^2)|\,\sigma^2 \in [\underline{\sigma}^2,\bar{\sigma}^2]\}$. However, this is not true since the uncertainty set of the G-normal distribution is much larger.

The following proposition shows that the classical normal distribution is a special case of the G-normal distribution.

Proposition 4.4
If $\underline{\sigma}^2 = \bar{\sigma}^2$, then the G-normal distribution $\mathcal{N}(0,[\underline{\sigma}^2,\bar{\sigma}^2])$ reduces to the classical normal distribution $\mathcal{N}(0,\bar{\sigma}^2)$.

Proof. Let $\underline{\sigma}^2 = \bar{\sigma}^2$, then the G-heat equation (4.2.11) becomes

$$\begin{aligned} u_t = G(u_{xx}) &\Leftrightarrow u_t = \frac{1}{2}(\bar{\sigma}^2 u_{xx}^+ - \underline{\sigma}^2 u_{xx}^-) \\ &\Leftrightarrow u_t = \frac{1}{2}(\bar{\sigma}^2(u_{xx}^+ - u_{xx}^-)) \\ &\Leftrightarrow u_t = \frac{\bar{\sigma}^2}{2} u_{xx} \end{aligned}$$

With $\alpha = \frac{\bar{\sigma}^2}{2}$ we get the classical heat equation (4.2.2). Hence, $u(t,x)$ is again the Normal distribution. More precisely, we get

$$\mathcal{E}[\psi(X)] = u(1,x) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}}\int_{-\infty}^{\infty} \psi(y)\exp\left\{-\frac{y^2}{2\bar{\sigma}^2}\right\} dy.$$

Clearly, $X \sim \mathcal{N}(0,\bar{\sigma}^2)$. □

Finally, we consider two special cases of the G-normal distribution that are very interesting, especially for possible applications. Now, let X be a G-normal distributed random variable, then we have:

(i) If $\psi \in C_{l.lip}(\mathbb{R})$ is convex, then

$$\mathcal{E}[\psi(X)] = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}} \int_{-\infty}^{\infty} \psi(y) \exp\left\{-\frac{y^2}{2\bar{\sigma}^2}\right\} dy. \tag{4.2.16}$$

(ii) If $\psi \in C_{l.lip}(\mathbb{R})$ is concave, then

$$\mathcal{E}[\psi(X)] = \frac{1}{\sqrt{2\pi\underline{\sigma}^2}} \int_{-\infty}^{\infty} \psi(y) \exp\left\{-\frac{y^2}{2\underline{\sigma}^2}\right\} dy. \tag{4.2.17}$$

Since we can prove both cases analogous, we now consider a convex $\psi \in C_{l.lip}(\mathbb{R})$. Then, we have for $\lambda \in [0,1]$

$$\begin{aligned} u(t, \lambda x + (1-\lambda)y) &= \mathcal{E}[\psi(\lambda x + (1-\lambda)y + \sqrt{t}X)] \\ &= \lambda\, \mathcal{E}[\psi(x + \sqrt{t}X)] + (1-\lambda)\, \mathcal{E}[\psi(x + \sqrt{t}X)] \\ &= \lambda u(t,x) + (1-\lambda) u(t,y), \end{aligned}$$

where the convexity of ψ yields the second equality. Hence, $u(t,x)$ is a convex function. Since the second derivative of a convex function is always non-negative we have $(u_{xx})^- = \max\{0, -u_{xx}\} = 0$ and thus the G-heat equation (4.2.11) reduces to

$$\begin{aligned} u_t = G(u_{xx}) &\Leftrightarrow u_t = \frac{1}{2}(\bar{\sigma}^2 u_{xx}^+ - \underline{\sigma}^2 u_{xx}^-) \\ &\Leftrightarrow u_t = \frac{\bar{\sigma}^2}{2} u_{xx}. \end{aligned}$$

This completes the proof. Since we know that concave functions always yield negative second derivatives, we can prove (ii) by analogous reasoning. These results might be important for later investigations, especially in the convex optimization context of Machine Learning.

4.3 CLT and LLN under sublinear expectation

Finally, we present a version of the Central Limit Theorem (CLT) and the Law of Large Numbers (LLN) in this sublinear expectation framework. These results originate in Peng (2007b) and can also be found in further work of the same author under different conditions. Before framing these theorems, we consider another crucial distribution in the sublinear framework, the so-called Maximal distribution.

Definition 4.11 (Maximal distribution)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $X \in \mathcal{H}$ a random variable with

$$\underline{\mu} := \mathcal{E}[X] < -\mathcal{E}[-X] =: \bar{\mu}.$$

Then, X is Maximal distributed, if for all $Y \in \mathcal{H}$ independent of X and $Y \overset{d}{\sim} X$, we have

$$\forall a, b \geq 0 : aX + bY \overset{d}{\sim} (a+b)X. \tag{4.3.1}$$

The Maximal distribution can also be characterized equivalently by: If X is Maximal distributed, we have for $\psi \in C_{l.lip}(\mathbb{R})$

$$\mathbb{F}_X[\psi] = \mathcal{E}[\psi(X)] = \sup_{\underline{\mu} \leq y \leq \bar{\mu}} \psi(y).$$

Thus, the Maximal distribution can be interpreted as "worst-case distribution". We denote a Maximal distributed random variable by $X \overset{d}{\sim} \mathcal{N}([\underline{\mu}, \bar{\mu}] \times \{0\})$ or by $\mathcal{M}_{[\underline{\mu}, \bar{\mu}]}$.

Theorem 4.4 (Law of Large Numbers)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $\{X_i\}_{i \in \mathbb{N}}$ an i.i.d[11] sequence of random variables. Further, assume that the uniform integrability condition

$$\lim_{\lambda \to \infty} \mathcal{E}[(|X_1| - \lambda)^+] = 0. \tag{4.3.2}$$

is satisfied. Now, define

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then S_n converges in distribution[12] to a Maximal distributed random variable X, i.e.

$$\lim_{n \to \infty} \mathcal{E}[\psi(\bar{X}_n)] = \mathcal{E}[\psi(X)], \tag{4.3.3}$$

where $\mathcal{E}[\psi(X)] = \sup_{\underline{\mu} \leq y \leq \bar{\mu}} \psi(y)$ with $\underline{\mu} = -\mathcal{E}[-X] < \mathcal{E}[X] = \bar{\mu}$, for all $\psi \in \mathcal{C}(\mathbb{R})$ satisyfing the linear growth condition $|\psi(x)| \leq C(1 + |x|)$.

[11] Recall: $\{X_i\}_{i \in \mathbb{N}}$ i.i.d means that $X_{i+1} \overset{d}{\sim} X_i$ and X_{i+1} independent of $\{X_1, \dots, X_i\}$ for all $i \in \{1, 2, \dots\}$. See also Proposition 4.2.
[12] See Definition 4.4.

Theorem 4.5 (Central Limit Theorem)
Let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $\{X_i\}_{i\in\mathbb{N}}$ an i.i.d sequence of random variables with mean-certainty, i.e., $\mathcal{E}[X_1] = -\mathcal{E}[-X_1] = 0$. Further, assume that the condition

$$\lim_{\lambda\to\infty} \mathcal{E}[(|X_1|^2 - \lambda)^+] = 0.$$

is satisfied. Now, define

$$\bar{X}_n \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

Then S_n converges in distribution to a G-normal distributed random variable X, i.e.

$$\lim_{n\to\infty} \mathcal{E}[\psi(\bar{X}_n)] = \mathcal{E}[\psi(X)], \tag{4.3.4}$$

for all $\psi \in \mathcal{C}(\mathbb{R})$ satisyfing the linear growth condition $|\psi(x)| \leq C(1 + |x|)$.

Remark 4.5
Some remarks on the assumptions in Theorem 4.4 and Theorem 4.5: The uniform integrability condition (4.3.2) *can be equivalently written by*

$$\lim_{n\to\infty} \mathcal{E}[|X_1|\mathbf{1}_{|X_1|>n}] = 0. \tag{4.3.5}$$

The LLN and CLT was proved under different conditions, we refer to the notes in Peng (2010, p. 44) and references therein.

Before we conclude this chapter with some references regarding related research, we present an unbiased estimator for the parameters of the Maximal distribution (Jin and Peng, 2016). This is especially important from a statistical perspective since just a few statistical tools have been provided so far in the sublinear framework. For the rest of this section, let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and $\{X_i\}_{i\in\mathbb{N}}$ an i.i.d sequence of Maximal distributed random variables with respective parameters $\underline{\mu} \leq \bar{\mu}$.

Definition 4.12 (Unbiased estimator)
Let $\psi_n \in \mathcal{C}(\mathbb{R}^n)$ and $-\infty < \underline{\mu} \leq \bar{\mu} < \infty$, then the statistic $T_n = \psi_n(X_1, \ldots, X_n)$ is called an unbiased estimator of

(i) $\bar{\mu}$, if $\mathcal{E}[\psi(X_1, \ldots, X_n)] = \bar{\mu}$.

(ii) $\underline{\mu}$, if $\mathcal{E}[\psi(X_1, \ldots, X_n)] = \underline{\mu}$.

Lemma 4.3
Let $\psi_n \in C_{l.lip}(\mathbb{R}^n)$. If the estimator $T_n = \psi_n(X_1, \ldots, X_n)$ is unbiased for $\bar{\mu}$, then for all $\underline{\mu} \leq \bar{\mu}$ we have

$$\max_{(x_1,\ldots,x_n)\in[\underline{\mu},\bar{\mu}]^n} \psi_n(x_1, \ldots, x_n) = \bar{\mu}, \tag{4.3.6}$$

$$\min_{(x_1,\dots,x_n)\in[\underline{\mu},\bar{\mu}]^n} \psi_n(x_1,\dots,x_n) = \underline{\mu}. \tag{4.3.7}$$

Consequently (4.3.6) *and* (4.3.7) *imply*

$$f_n(x_1,\dots,x_n) \leq \bar{\mu}, \tag{4.3.8}$$

$$f_n(x_1,\dots,x_n) \geq \underline{\mu}. \tag{4.3.9}$$

Proof. For reasons of clarity we just prove the case $n = 2$. Thus, we have for $\psi_2 \in C_{lip}(\mathbb{R}^2)$

$$\begin{aligned}
\mathcal{E}[\psi_2(X_1,X_2)] &\overset{i.i.d}{=} \mathcal{E}[\mathcal{E}[\psi_2(x_1,X_2)]_{x_1=X_2}] \\
&= \mathcal{E}[\mathcal{E}[\max_{\underline{\mu}\leq x_2\leq\bar{\mu}} \psi_2(x_1,x_2)]_{x_1=X_1}] \\
&= \mathcal{E}[\max_{\underline{\mu}\leq x_2\leq\bar{\mu}} \psi_2(X_1,x_2)]] \\
&= \max_{\underline{\mu}\leq x_1\leq\bar{\mu}} \max_{\underline{\mu}\leq x_2\leq\bar{\mu}} \psi_2(x_1,x_2) \\
&= \max_{(x_1,x_2)\in[\underline{\mu},\bar{\mu}]^2} \psi_2(x_1,x_2).
\end{aligned}$$

Since $\psi_2(X_1,X_2)$ is unbiased, this completes the proof. Clearly, this implies (4.3.8). Analogously one shows (4.3.7) and (4.3.9). Note that, in the second equation, we used the fact that the random variables are Maximal distributed. A more general proof can be found in Jin and Peng (2016, p. 9 f.). □

Theorem 4.6
Let $X_1,\dots,X_n$ be i.i.d. distributed random variables with $X_i \overset{d}{\sim} \mathcal{M}_{[\underline{\mu},\bar{\mu}]}$, where $\underline{\mu} \leq \bar{\mu}$ are unknown parameters of the Maximal distribution. Now, we define the estimator for the upper mean $\hat{\bar{\mu}} := \max\{X_1(\omega),\dots,X_n(\omega)\}$ and resp. $\hat{\underline{\mu}} := \min\{X_1(\omega),\dots,X_n(\omega)\}$. Then, the following statements hold true:

(i) $\underline{\mu} \leq \hat{\underline{\mu}} \leq \hat{\bar{\mu}} \leq \bar{\mu}$.

(ii) $\hat{\bar{\mu}}$ *is the largest unbiased estimator for* $\bar{\mu}$.

(iii) $\hat{\underline{\mu}}$ *is the smallest unbiased estimator for* $\underline{\mu}$.

Proof. The first statement (i) is by construction obvious. Again, we just consider the case $n = 2$. It is clear, that $\hat{\underline{\mu}}$ and $\hat{\bar{\mu}}$ are unbiased, since X_1 and X_2 are i.i.d and

Maximal distributed. Let $T_2 = \psi_n(X_1, X_2)$ be a given unbiased estimator of $\bar{\mu}$ and for $x_1, x_2 \in \mathbb{R}$ set

$$\bar{\tilde{\mu}} = \max\{x_1, x_2\}. \tag{4.3.10}$$

According to Lemma 4.3 this yields

$$\psi_2(x_1, x_2) \leq \bar{\tilde{\mu}} = \max\{x_1, x_2\}. \tag{4.3.11}$$

Since $y_1, y_2 \in \mathbb{R}$ were chosen arbitrarly, (ii) follows. Analogously one proves (iii). This completes the proof. □

For a more general estimator, asymptotic studies and further results see Jin and Peng (2016). We conclude this section with some remarks on related work.

4.4 Remarks on related work

We complete this chapter by pointing out further work within the G-Framework. Since our presentation was mainly focused on clarity, we only considered the most basic structures in more detail. In particular, so-called G-*Brownian motions* can be defined based on the G-normal distribution. This allows, especially, to formulate many concepts presented in the first section of this thesis in a sublinear context, i.e., stochastic calculus under sublinear expectations. We mainly refer to Peng (2010). Albeit our explanations do not consider multidimensional cases, all concepts can also be stated for the multidimensional case. From a statistical point of view, we point out Peng and Zhou (2020) since the authors provide a hypothesis-testing perspective within the G-Framework. Closely related to the aforementioned work, we also refer here to Sun and Ji (2016), which is concerned with a generalized Neyman-Pearson lemma for sublinear expectations. A generalized Neyman-Pearson type lemma was previously studied also in the context of g-probabilities by Ji and Zhou (2010) relying on concepts discussed in previous sections of this thesis. In particular, to better understand G-normal distributions and also provide a means for simulations, Li et al. (2021) considers a bridging link between normal distributions and G-normal distributions, called the Semi-G-normal distribution. A Stein type characterization is proposed in Hu et al. (2017). Although not in direct reference to the G-Framework, Cohen (2017) also considers nonlinear expectations to express statistical uncertainty. Moreover, Nendel (2021) seeks to link work among the communities of imprecise probabilities and mathematical finance. Finally, we note that these references are not exhaustive, and there is more existing work related to the G-Framework, albeit the "classical" statistical work in this area is pretty limited.

5 Sublinear expectation regression

In this section, we motivate a possible application of sublinear expectations in the context of regression models. The notion of *sublinear expectation regression* is based on Lin et al. (2013). This section will show that the sublinear regression model is in fact identifiable. At the end of section, we will state some further research questions to be addressed in future work. First we recall the "classical" linear regression model, and record where problems may arise in the respective specifications. We examine only the simple linear regression model in this section for the sake of clarity, although a much larger class of regression models can be considered under sublinear expectations.

(Simple) linear regression:
Let $(\Omega, \mathcal{F}, P)$ be a probability space and X, Y two real-valued random variables, where we refer to Y as response variable and to X as covariate. Let $\mathbb{E}[Y|X]$ be the conditional expectation of Y given X, then we consider

$$Y = \mathbb{E}[Y|X] + \varepsilon, \tag{5.0.1}$$

where ε defines an error term, yet to be specified. In the case of *linear regression* we specify

$$\mathbb{E}[Y|X] = f(X), \quad f \text{ linear function of } X. \tag{5.0.2}$$

We usually have $f(X) = \beta_0 + \beta_1 X$, where β_0 is called intercept and $\boldsymbol{\beta}^t = (\beta_0, \beta_1)$ is the respective regression coefficient vector. Hence, we also write $f(X) = \boldsymbol{X}^t\boldsymbol{\beta}$, where we define $\boldsymbol{X}^t = (1, X)$. Since we are just interested in simple linear regression, we do not state the case with multiple covariates here, and the reader may refer e.g. to Fahrmeir et al. (2013). Additionally to the previous assumptions, we further consider the *conditional error variance*

$$\sigma^2(X) = \mathrm{Var}(\varepsilon|X) = \mathbb{E}[\epsilon^2|X]. \tag{5.0.3}$$

By assuming homoskedasticity, we simply have $\mathbb{E}[\varepsilon^2|X] = \mathbb{E}[\varepsilon^2] =: \sigma^2$. In the following, we will stay with homoskedastic errors. We also know that $\mathbb{E}[\varepsilon^2] = 0$, since

$$\begin{aligned}\mathbb{E}[\epsilon|X] \overset{(5.0.1)}{=} \mathbb{E}[(Y - [\mathbb{E}[Y|X]])|X] &= \mathbb{E}[Y|X] - \mathbb{E}[\mathbb{E}[Y|X]|X] \\ &= 0.\end{aligned}$$

Thus, $\mathbb{E}[\varepsilon] = \mathbb{E}[\mathbb{E}[\varepsilon|X]] = 0$. It is also common to assume Normal distributed errors, i.e., $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. In the case where this assumption is violated, different methods have evolved in statistics, but as noted in Lin et al. (2013, p. 3), these methods are entirely unrelated to nonlinear expectations. This lack of attention is because the G-

Framework and its theoretical foundations are widely unknown in statistics. We will now discuss the sublinear counterpart of simple linear regression.

(Simple) sublinear expectation regression:
Relying on results from previous sections, let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and X, Y two real-valued random variables. Again X will denote our covariate, Y the response variable. If we assume $\mathcal{E}[Y|X]$ to be a sublinear expectation, then we know from Lemma 4.2 that there exists a family of probability measures $\{P_{Y|X,\theta}\}_{\theta\in\Theta}$ defined on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ such that

$$\mathcal{E}[Y|X] = \sup_{\theta\in\Theta} \mathbb{E}_{P_{Y|X,\theta}}[Y|X].$$

Guided by our considerations regarding uncertainty in the introductory section, we state the following simple regression model

$$Y = \mathcal{E}[Y|X] + \varepsilon, \tag{5.0.4}$$

where we assume $\varepsilon \overset{d}{\sim} \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, i.e., the error ε is G-normal distributed (hence, ε is mean certain, but variance uncertain).

Proposition 5.1
Let $Y = \mathcal{E}[Y|X] + \epsilon$, where $\epsilon \overset{d}{\sim} \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ and $\mathcal{E}[Y|X] = X\beta$. Thus, we omit the intercept term here. If $\mathbb{E}[X^2] < \infty$, then β is identifiable by

$$\beta = (\mathbb{E}[X^2])^{-1}\, \mathbb{E}[X\, \mathcal{E}[Y|X]].$$

Proof. $\mathcal{E}[Y|X] = X\beta$ yields

$$\begin{aligned}
& X\, \mathcal{E}[Y|X] = X^2\beta \\
\Leftrightarrow \quad & \mathbb{E}[X\, \mathcal{E}[Y|X]] = \mathbb{E}[X^2\beta] \\
\Leftrightarrow \quad & \mathbb{E}[X\, \mathcal{E}[Y|X]] = \mathbb{E}[X^2]\beta \\
\Leftrightarrow \quad & (\mathbb{E}[X^2])^{-1}\, \mathbb{E}[X\, \mathcal{E}[Y|X]] = \beta,
\end{aligned}$$

where $\mathbb{E}[\cdot]$ denotes the classical expectation. This completes the proof. □

This generally allows the formulation of regression models in a sublinear context. For further details and considerations, see Lin et al. (2013). For some recent work on regressions modeling under distribution uncertainty, the reader may refer to Yang and Yao (2021). Finally, we state some further research questions in terms of regression modeling under sublinear expectations. The author intends to deal with these explicitly in further work. It seems that the sublinear expectation framework is particularly suited to formulate *imprecise regression models*. In this sense, we make following remark:

Remark 5.1
Is it possible to express "imprecise covariates" through sublinear distributions? Precisely, let $(\Omega, \mathcal{H}, \mathcal{E})$ be a sublinear expectation space and X, Y two random variables. Now, consider the following regression model

$$Y = X\beta + \varepsilon, \quad X \overset{d}{\sim} \mathcal{M}_{[\underline{\mu},\bar{\mu}]}, \quad \varepsilon | X \sim \mathcal{N}(0, \sigma^2). \tag{5.0.5}$$

Hence, the covariate X is assumed to be Maximal distributed with $0 < \underline{\mu} < \bar{\mu}$. We interpret this assumption as follows: Instead of X being a deteriministic covariate, X can be seen as an imprecise covariate with lower mean $-\mathcal{E}[-X] = \underline{\mu}$ and $\mathcal{E}[X] = \bar{\mu}$, respectively upper mean. Hence, by (5.0.5) *we expect*

$$\begin{aligned}\mathcal{E}[Y] = \mathcal{E}[X\beta + \epsilon] &\leq \mathcal{E}[X\beta] + \mathcal{E}[\varepsilon] \\ &= \beta^+ \mathcal{E}[X] + \beta^- \mathcal{E}[-X] \\ &= \beta^+ \bar{\mu} - \beta^- \underline{\mu}.\end{aligned}$$

By assuming a multivariate Maximal distribution, this idea can be generalized to the multiple covariate case. Since multivariate sublinear random variables were not considered in this thesis, we do not explicitly address this point here.

6 Conclusion & Outlook

This thesis considered nonlinear expectations from different perspectives and their suitability for decision-making under uncertainty in dynamic and non-dynamic situations. For a compact summary of the notions presented in this thesis, the reader may refer to Figure 6. A primary goal of this work was to invoke a nonlinear (probabilistic) way of thinking and emphasize the potential applicability of nonlinear expectations in statistics. Therefore, this illustration demands a translation of ideas from different academic communities. First, we clarified the fundamental difference between risk and uncertainty by entangling the notions of risk, probability, and uncertainty under a historical perspective. Then, taking the Ellsberg experiment as an example, we took a closer look at a typical decision situation under uncertainty. Next, we noted why the expected utility theory is not suitable to explain the observed preferences in the Ellsberg experiment. Finally, motivated by the fact that classical probability measures are insufficient to model the underlying uncertainty in some situations, we discussed integrals with respect to non-additive measures or capacities. However, these integrals are nonlinear due to the lack of additivity of the underlying probability measure. We refer to the nonlinear functional

$$\mathcal{E}_c : L^2_{\mathcal{F}} \to \mathbb{R} : \; X \mapsto \mathcal{E}_c[X] = \int_0^\infty X^+ d\mu - \int X^- d\mu_d \tag{I}$$

as *Choquet expectation*. This nonlinear functional was used to define the so-called Choquet expected utility and thus appropriately model the observed preferences in the Ellsberg case. In addition, we noticed that Choquet expectations could be defined in terms of linear expectations if certain conditions are satisfied. Further, we revisited a nonlinear, filter-consistent expectation in the framework of so-called backward stochastic differential equations. For this purpose, stochastic differential equations and thus elementary concepts of stochastic calculus were first presented in more detail. If the pair of processes $(Y_t, Z_t)_{t\in[0,T]} \in L^2_{\mathcal{F}}([0,T], \mathbb{R}\times\mathbb{R}^n)$ is the unique solution of the BSDE

$$\begin{cases} -dY_t = g(t, Y_t, Z_t)dt - Z_t dB_t, \quad 0 \le t \le T, \\ \quad Y_T = X, \end{cases}$$

we call the functional

$$\mathcal{E}_g : L^2_{\mathcal{F}_T} \to \mathbb{R} : \; X \mapsto \mathcal{E}_g[X] = Y_0 \tag{II}$$

g-expectation of the random variable X. It turns out that g-expectations and corresponding conditional g-expectations satisfy many properties of the classical expectation. So far, g-expectations are mainly applied in mathematical finance. However, g-expectations are also used to define recursive multiple priors utility, an intertemporal counterpart to the multiple prior model. Based on (I) and (II), we record that in both

non-dynamic and dynamic situations, nonlinear expectations are helpful for modeling underlying uncertainty. This observation motivates a framework under nonlinear expectations, the G-Framework. It is worth mentioning that, especially in mathematical finance, these results under nonlinear expectations are already well known but have only been made available to a small audience among (applied) statisticians due to underlying theoretical requirements. This thesis considered a straightforward illustration of the fundamental concepts in this nonlinear framework and tried to motivate theoretical foundations in more detail. We shifted our focus from the usual probabilistic setting $(\Omega, \mathcal{F}, P)$ to nonlinear (mostly sublinear) expectation spaces $(\Omega, \mathcal{H}, \mathcal{E})$ and presented crucial notions such as independence and distribution of random variables under sublinear expectations. Since the so-called G-normal distribution is characterized via the G-heat equation

$$\begin{cases} u_t = G(u_{xx}), & \text{on } x \in \mathbb{R} \times (0, \infty) \\ u = \psi, & \text{on } x \in \mathbb{R} \times \{t = 0\}, \end{cases}$$

we explicitly exploited the relationship between the classical normal distribution and the heat equation in a probabilistic and non-probabilistic setting. We note that this relationship also holds in the nonlinear case. Of particular interest for later applications in statistics are the versions of the Law of Large Numbers and the Central Limit Theorem under sublinear expectations. Especially with the estimation of the mean parameters of the Maximal distribution, we illustrated that nonlinear versions of classical estimation and test theory could, or should, evolve in future work to increase applicability in statistics. Furthermore, the identifiability of the sublinear expectation regression served as another example of the potential applicability of this framework.

As increasing attention is being paid to potential uncertainty in many areas of statistics and its applications, we believe that the G-Framework may have a notable role to play in future applications. Future research will require statistical tools in the G-Framework to guarantee applicability to real-life questions. Furthermore, nonlinear (resp. sublinear) expectations have to be motivated and made available to a broad (applied) audience by employing real-life applicability. Moreover, suitable numerical methods in this context are also inevitable. In particular, it is worthwhile to examine the sublinear expectation regression motivated in Section 5 and reflect whether a possible imprecise generalization in this setting is possible, as proposed in the respective remark of the section. This approach connects directly to ideas emerging from the imprecise probability community, making a transfer of different concepts conceivable.

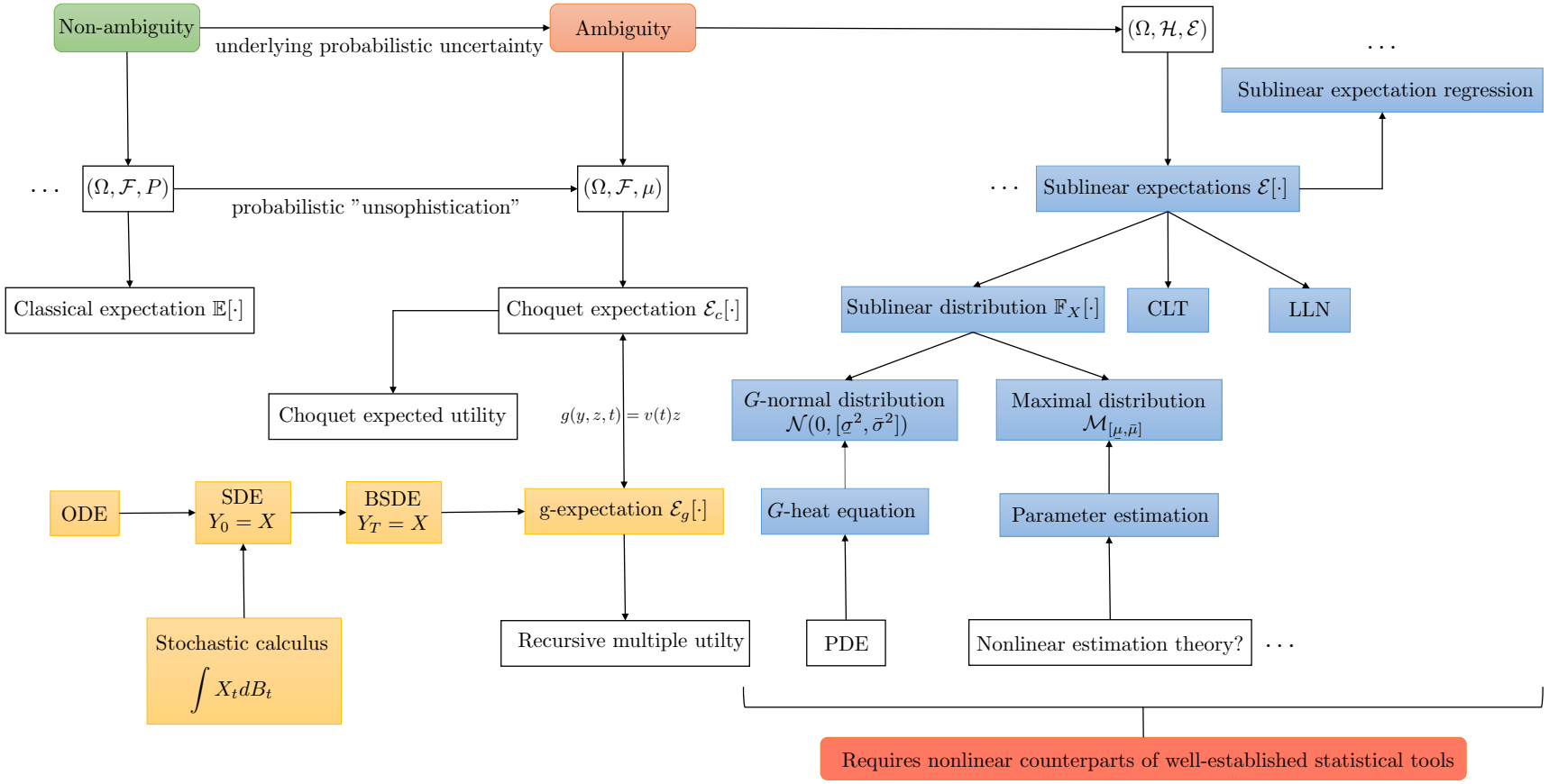

Figure 6: Visuzalisation of key elements in this thesis.

A Probability Theory

The material in this section can be found in any introductory book on Measure and Probability Theory, e.g. consider Bauer (2011).

Definition A.1 (σ-algebra)
Let Ω be a given set and $\mathcal{A} \subseteq 2^\Omega$ a family of subsets. Then $\mathcal{A}$ is called a σ-algebra, if

(i) $\emptyset, \Omega \in \mathcal{A}$

(ii) $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$

(iii) I *countable*, $(A_i)_{i\in I} \in \mathcal{A}^I \Longrightarrow \bigcup\limits_{i\in I} A_i \in \mathcal{A}$

The pair $(\Omega, \mathcal{A})$ is called a measurable space.

Given any family $\mathcal{U}$ of subsets of Ω there is always a smallest σ-Algebra containing $\mathcal{U}$, namely the σ-algebra generated by $\mathcal{U}$, i.e.

$$\sigma(\mathcal{U}) \coloneqq \bigcap \{\mathcal{H} : \mathcal{H} \text{ is } \sigma\text{-algebra on } \Omega, \mathcal{U} \subset \mathcal{H}\}.$$

Definition A.2 (Measure)
A set-function $\mu : \mathcal{A} \longrightarrow [0, \infty]$ is called measure, if $\mathcal{A}$ is a σ-algebra and

(i) $\forall A \in \mathcal{A} : \mu(A) \geq 0$

(ii) $\mu(\emptyset) = 0$

(iii) For $(A_k)_{k\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ pairwise disjunct, we have:

$$\mu\left(\dot{\bigcup_{k\in\mathbb{N}}} A_k\right) = \sum_{k\in\mathbb{N}} \mu(A_k)$$

The tripel $(\Omega, \mathcal{A}, \mu)$ is called measure space.

Example A.1

1. *$(\mathbb{R}^d, \mathcal{A})$ with $\mathcal{A} = \mathfrak{B}(\mathbb{R}^d)$ Borel-σ-algebra, is the smallest σ-algebra containing all open (resp. closed) sets.*

   $$\exists! \text{ measure } \mu : \mathcal{A} \to [0,\infty] : \begin{cases} \mu([0,1]^d) = 1 \\ \forall A \in \mathcal{A}\ \forall x \in \mathbb{R}^d :\ \mu(x + A) = \mu(A) \end{cases}$$

   This measure μ is called Borel-measure.

2. $(\mathbb{R}^d, \mathcal{A})$ *with* $\mathcal{A} = \mathfrak{L}(\mathbb{R}^d)$ *Lebesgue-*σ*-algebra, is the smallest* σ*-algebra containing all open (resp. closed) sets and all nullsets.*

$$\exists! \text{ measure } \mu : \mathcal{A} \to [0,\infty] : \begin{cases} \mu([0,1]^d) = 1 \\ \forall A \in \mathcal{A}\ \forall x \in \mathbb{R}^d :\ \mu(x + A) = \mu(A) \end{cases}$$

This measure μ *is called Lebesgue-measure.*

Definition A.3 (σ-finite measure)
Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space.*

(i) μ *is finite, if* $\mu(\Omega) < \infty$.

(ii) μ *is* σ*-finite, if there is a sequence* $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ *such that* $\bigcup_{n\in\mathbb{N}} A_n = \Omega$ *and* $\mu(A_n) < \infty$ *for all* $n \in \mathbb{N}$.

Definition A.4 (μ-null set)
Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space. Sets* $A \in \mathcal{F}$ *with* $\mu(A) = 0$ *are called* μ*-null-sets or just null-sets.*

Definition A.5 (Complete measure space)
A measure space $(\Omega, \mathcal{A}, \mu)$ *is called complete, if every subset of a nullset* $A \in \mathcal{A}$ *also belongs to* $\mathcal{A}$ *(and is therefore itself a nullset). In this case we also say that the measure* μ *is complete.*

Definition A.6 (Measurable mapping)
Let $(\Omega, \mathcal{F})$ *and* $(\Omega', \mathcal{F}')$ *be two measurable spaces. A mapping* $X : \Omega \longrightarrow \Omega'$ *is called* $\mathcal{A} - \mathcal{A}'$*-measurable, if every* $\mathcal{A}'$*-measurable set* B *has a* $\mathcal{A}$*-measurable preimage* $X^{-1}(B)$*, i.e.*

$$\forall B \in \mathcal{A}' : X^{-1}(B) \in \mathcal{A}\,.$$

A measurable mapping is often referred to as a random variable and denoted by capital letters, e.g. X, Y.

Definition A.7 (Conditional expectation)
Let $(\Omega, \mathcal{F}, P)$ *be a probability space and* X *a real-valued random variable with* $\mathbb{E}[|X|] < \infty$ *and* $\Sigma \subseteq \mathcal{F}$ *a sub-*σ*-algebra. Then* $\mathbb{E}[X|\Sigma] : \Omega \to \mathbb{R}$ *is called conditional expectation of* X *given* Σ *if*

(i) $\mathbb{E}[X|\Sigma]$ *is* Σ*-measurable.*

(ii) $\forall A \in \Sigma :\ \mathbb{E}[\mathbb{E}[X|\Sigma]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$.

The conditional expectation is almost surely unique, hence all statements regarding the condtional expectations should be understood up to a null-set.

Theorem A.1 (Properties of conditional expectation)
Let X, Y be two random variables on the probabilty space $(\Omega, \mathcal{F}, P)$ and $a, b \in \mathbb{R}$. Then the conditional expectation has following properties:

(i) *If X is Σ-measurable, then $\mathbb{E}[X|\Sigma] = X$.*

(ii) $\mathbb{E}[aX + bY|\Sigma] = a\,\mathbb{E}[X|\Sigma] + b\,\mathbb{E}[X|\Sigma]$.

(iii) $\mathbb{E}[\mathbb{E}[X|\Sigma]] = X$.

(iv) $\mathbb{E}[X|\Sigma] = \mathbb{E}[X]$, *if X is independent of Σ.*

(v) $\mathbb{E}[Y \cdot X|\Sigma] = Y \cdot \mathbb{E}[X|\Sigma]$, *if Y is Σ-measurable.*

(vi) *For $\Sigma_1 \subseteq \Sigma_2 \subseteq \mathcal{F}$ we have*

$$\mathbb{E}[\mathbb{E}[X|\Sigma_2]|\Sigma_1] = \mathbb{E}[X|\Sigma_1].$$

Theorem A.2 (Borel-Cantelli Lemma)
Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F}$ a sequence of events. Then we have

$$\sum_{i=0}^{\infty} P(A_i) < \infty \Rightarrow P\left(\sum_{i=0}^{\infty} \mathbf{1}_{A_i} < \infty\right) = 1.$$

If $\{B_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F}$ is now a sequence of independent events, then

$$\sum_{i=0}^{\infty} P(B_i) = \infty \Rightarrow P\left(\sum_{i=0}^{\infty} \mathbf{1}_{B_i} = \infty\right) = 1.$$

Theorem A.3 (Fubini-Tonelli)
Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two σ-finite measure spaces. Further, let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathfrak{B}(\mathbb{R})$-measurable function. If f is non-negative or $\mu_1 \otimes \mu_2$ integrable, then there is a μ_2-nullset $N \subseteq \Omega_2$, such that for $\omega_2 \in \Omega_2 \backslash N$ the function

$$f(\cdot, \omega_2) : \Omega_1 \to \mathbb{R} : \ \omega_1 \mapsto f(\omega_1, \omega_2)$$

is integrable with respect to μ_1.
Further, with

$$f_{\Omega_1} : \Omega_2 \backslash N \to \mathbb{R} : \ \omega_2 \mapsto \int f(\cdot, \omega_2) d\mu_1$$

we have

$$\int_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)d(\mu_1\otimes\mu_2)(\omega_1,\omega_2) = \int_{\Omega_2\setminus N}\left(\int_{\Omega_1} f(\omega_1,\omega_2)d\mu_1(\omega_1)\right)d\mu_{2_{|\Omega_2\setminus N}}(\omega_2). \tag{A.0.1}$$

Theorem A.4 (Dominated convergence theorem)
Let $(\Omega,\mathcal{F},P)$ be a probability space and $\{X_n\}_{n\in\mathbb{N}}$ a sequence of random variables such that

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

where X is a random variable. If there exists a random variable Y such that

$$\int Y dP < \infty \quad \textit{and} \quad |X_n(\omega)| \leq Y(\omega)$$

for every $n\in\mathbb{N}$ and almost all $\omega\in\Omega$, then

$$\lim_{n\to\infty}\int X_n dP = \int X dP.$$

Definition A.8 (Stochastic process)
The quadruple $(\Omega,\mathcal{F},P,\{X_t\}_{t\in I})$, where $(\Omega,\mathcal{F},P)$ is a probability space and $\{X_t\}_{t\in I}$ a family of random variables with values in a measurable space $(E,\mathcal{E})$, is called a stochastic process. For each $\omega\in\Omega$ the mapping of I into E defined by $t\mapsto X_t(\omega)$ is called a path of the process. Instead of the quadruple notation we denote $\{X_t\}_{t\in I}$ as a stochastic process. Usually $(E,\mathcal{E}) = (\mathbb{R},\mathfrak{B}(\mathbb{R}))$ and $I=\mathbb{R}_{\geq 0}$.

Definition A.9 (Stationary process)
A stochastic process $\{X_t\}_{t\in I}$ is called stationary if the respective joint distributions of $X_{t_1+s},\ldots,X_{t_n+s}$ are independent of s for all $t_1,\ldots,t_n$ and $n\geq 1$.

Definition A.10 (Brownian motion)
A stochastic process $\{B_t\}_{t\in I}$ with $I=\mathbb{R}_{\geq 0}$ is called (standard) Brownian motion, if it satisfies:

(i) $B_0 = 0 \quad a.s.$

(ii) For all $t>s\geq 0$, we have

$$B_t - B_s \sim \mathcal{N}(0,t-s).$$

(iii) For all $n\geq 2$ and $t_n > t_{n-1} > \cdots > t_0 \geq 0$, the random variables

$$B_{t_i} - B_{t_{i-1}}, \quad i=1,\ldots,n$$

are independent.

(iv) $\{B_t\}_{t\geq 0}$ *has* $a.s.$ *continuous samples paths.*

Definition A.11 (Filtration)
Let $(\Omega, \mathcal{F})$ *be a measurable space. A family of* σ*-algebras* $\mathcal{F}_t \subseteq \mathcal{F}$ *is called filtration if*

$$\forall s, t : s \leq t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t \,.$$

If P *defines a probability measure on* $(\Omega, \mathcal{F})$*, then we call* $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t\in I})$ *a filtered probability space.*

Definition A.12 (Brownian filtration)
Let $\{B_t\}_{t\in I}$ *be the Brownian motion. Then we call the filtration* $\mathfrak{F} \coloneqq \{\mathcal{F}_t\}_{t\in I}$ *Brownian filtration (or filtration generated by the Brownian motion), if*

$$\mathcal{F}_t = \sigma(B_s : s \leq t).$$

Definition A.13 (Augmented filtration)
Let $\{X_t\}_{t\in I}$ *be a stochastic process on the probability space* $(\Omega, \mathcal{F}, P)$*. Then we call* $\mathfrak{F} \coloneqq \{\mathcal{F}_t\}_{t\in I}$ *the respective augmented filtration with* $\mathcal{F}_t \coloneqq \sigma\{X_s \cup \mathfrak{N} : \, 0 \leq s \leq t\}$*, where* $\mathfrak{N} \coloneqq \{A \subseteq \Omega |\, \exists B \in \mathcal{F}, B \subset A, P(B) = 0\}$ *denotes the collection of respective* P*-null sets.*

Definition A.14 (Martingale)
Let $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\{\mathcal{F}_t\}_{t\in I}$ *a filtration. A stochastic process* $\{M_t\}_{t\in I}$ *on* $(\Omega, \mathcal{F}, P)$ *is called martingale w.r.t. to the filtration* $\{\mathcal{F}_t\}_{t\in I}$ *if*

(i) $\forall t \in I : \; M_t \in L^1(\Omega, \mathcal{F}, P)$.

(ii) M_t *is* $\mathcal{F}_t$*-measurable for all* $t \in I$.

(iii) For all $s, t \in I$ *with* $s \leq t$ *we have*

$$\mathbb{E}[M_t |\, \mathcal{F}_s] = M_s \quad a.s.$$

Definition A.15 (Sub- and supermartingale)
Let $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t\in I})$ *be a filtered probability space. An adapted stochastic process* $\{X_t\}_{t\in I}$ *is called*

(i) Submartingal, if

$$\forall s, t \in I : s \leq t \Rightarrow X_s \leq \mathbb{E}[X_t |\, \mathcal{F}_s] \quad a.s.$$

(ii) Supermartingal, if

$$\forall s,t \in I : s \leq t \Rightarrow X_s \geq \mathbb{E}[X_t | \mathcal{F}_s] \quad a.s.$$

Theorem A.5 (Doob's maximal inequality)
Let $\{M_n\}$ be a non-negative submartingale and define $\bar{M}_n := \sup_{0\leq m\leq n} M_m$. For $\lambda > 0$, we have

$$\lambda P(\bar{M}_n \geq \lambda) \leq \mathbb{E}[M_n \mathbf{1}_{\bar{M}_n \geq \lambda}] \leq \mathbb{E}[M_n].$$

Theorem A.6 (Law of Large Numbers)
Let $(\Omega, \mathcal{F}, P)$ a probability space and $\{X_n\}_{n\in\mathbb{N}}$ a family of pairwise independent and identicaly distributed random variables with $\mathbb{E}[|X_1|] < \infty$. Then, we have

$$S_n := \frac{1}{n}\sum_{i=1}^{n} X_n \xrightarrow{a.s.} \mathbb{E}[X_1] \quad \text{for } n \to \infty.$$

Theorem A.7 (Central Limit Theorem)
Let $(\Omega, \mathcal{F}, P)$ a probability space and $\{X_n\}_{n\in\mathbb{N}}$ a family of i.i.d random variables with $\mathbb{E}[|X_1|] = \mu < \infty$ and $\mathrm{Var}[X_1] = \sigma^2 < \infty$.
Define

$$S_n := \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}.$$

Then, S_n converges in distribution to a standard-normal distributed random variable.

B Functional Analysis

All results stated here can be found e.g. in Brezis (2011) or in similar introductory books on functional analysis.

Definition B.1 (L^p spaces)
Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $1 \leq p < \infty$. Then, we call

$$L^p_{\mathcal{F}} = \left\{ f : \Omega \to \mathbb{R} \text{ measurable}, \int |f|^p \, d\mu < \infty \right\}$$

L^p space equipped with the L^p norm

$$\|f\|_{L^p} = \Big(\int_\Omega |f|^p \, d\mu \Big)^{\frac{1}{p}}.$$

Further, for $p = \infty$ we set

$$L^\infty_{\mathcal{F}} = \left\{ f : \Omega \to \mathbb{R} \;\middle|\; \begin{array}{l} f \text{ is measurable and exists a constant } C \\ \text{such that } |f(x)| \leq C \; a.e. \text{ on } \Omega \end{array} \right\}$$

with

$$\|f\|_{L^\infty} = \inf\{C : |f(x)| \leq C \;\; a.e. \text{ on } \Omega\}.$$

Theorem B.1 (Hahn-Banach)
Let X be a $\mathbb{R}$-vector space and p be a real-valued function on X with the following properties:

(i) Positive homogeneity,

$$p(\lambda x) = \lambda p(x) \quad \text{for all } \lambda > 0$$

for all $x \in X$.

(ii) Subadditivity,

$$p(x + y) \leq p(x) + p(y)$$

for all $x, y \in X$.

Let Y now denote a subspace of X on which a linear functional l dominated by p is defined, i.e.

$$\forall y \in Y : l(y) \leq p(y)$$

Then, there exists a linear extension $L : X \to \mathbb{R}$ such that:

(i) $L_{|Y} = l$

(ii) $\forall x \in X : L(x) \leq p(x)$

References

Ash, R. B. (2014). *Real analysis and probability: probability and mathematical statistics: a series of monographs and textbooks*, Academic press.

Bachelier, L. (1900). Théorie de la spéculation, *Annales scientifiques de l'École normale supérieure*, Vol. 17, pp. 21–86.

Bauer, H. (2011). *Probability theory*, Vol. 23, Walter de Gruyter.

Bernoulli, D. (1896). *Die Grundlage der modernen Wertlehre: Versuch einer neuen Theorie der Wertbestimmung von Glücksfällen (Specimen Theoriae novae de Mensura Sortis).*, Duncker & Humblot.

Bismut, J.-M. (1973). Conjugate convex functions in optimal stochastic control, *Journal of Mathematical Analysis and Applications* **44**(2): 384–404.

Brezis, H. (2011). *Functional analysis, Sobolev spaces and partial differential equations*, Vol. 2, Springer.

Chen, Z., Chen, T. and Davison, M. (2005). Choquet expectation and Peng's g-expectation, *The Annals of Probability* **33**(3): 1179–1199.

Chen, Z. and Epstein, L. (2002). Ambiguity, risk, and asset returns in continuous time, *Econometrica* **70**(4): 1403–1443.

Choquet, G. (1954). Theory of capacities, *Annales de l'institut Fourier*, Vol. 5, pp. 131–295.

Chouzenoux, E., Gérard, H. and Pesquet, J.-C. (2019). General risk measures for robust machine learning, *arXiv preprint arXiv:1904.11707* .

Cohen, S. N. (2017). Data-driven nonlinear expectations for statistical uncertainty in decisions, *Electronic Journal of Statistics* **11**(1): 1858–1889.

Coquet, F., Hu, Y., Mémin, J. and Peng, S. (2002). Filtration-consistent nonlinear expectations and related g-expectations, *Probability Theory and Related Fields* **123**(1): 1–27.

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré* **17**: 1–68.

de Finetti, B. (1990). *Theory of probability: A critical introductory treatment*, Vol. 1, John Wiley & Sons.

de Montmort, P. R. (1708). *Essay d'analyse sur les jeux de hazard.*, Chez Jacque Quillau, Imprimeur-Juré-Libraire de l'Université, rue Galande.

Denneberg, D. (2013). *Non-additive measure and integral*, Vol. 27, Springer Science & Business Media.

Duffie, D. and Epstein, L. G. (1992). Stochastic differential utility, *Econometrica: Journal of the Econometric Society* pp. 353–394.

Dyckerhoff, R. (1994). *Choquet-Erwartungsnutzen und antizipierter Nutzen*, PhD thesis, Ph. D. thesis, University of the Federal Armed Forces Hamburg.

Einstein, A. et al. (1905). On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat, *Annalen der physik* **17**(549-560): 208.

El Karoui, N. and Mazliak, L. (1997). Backward Stochastic Differential Equations. Pitman Research Notes in Mathematics Series 364, *El-Karoui and S. Mazliak edts* .

El Karoui, N., Peng, S. and Quenez, M. C. (1997). Backward stochastic differential equations in finance, *Mathematical finance* **7**(1): 1–71.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms, *The Quarterly Journal of Economics* **75**(4): 643 – 669.

Evans, L. C. (2010). *Partial differential equations*, Vol. 19, American Mathematical Soc.

Evans, L. C. (2012). *An introduction to stochastic differential equations*, Vol. 82, American Mathematical Soc.

Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013). Regression models, *Regression*, Springer, pp. 21–72.

Gianin, E. R. (2006). Risk measures via g-expectations, *Insurance: Mathematics and Economics* **39**(1): 19–34.

Gilboa, I. (1987). Expected utility with purely subjective non-additive probabilities, *Journal of mathematical Economics* **16**(1): 65–88.

Gilboa, I. (2009). *Theory of decision under uncertainty*, Vol. 45, Cambridge university press.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior, *Journal of mathematical economics* **18**(2): 141–153.

Hu, M. and Li, X. (2014). Independence Under the G-Expectation Framework, *Journal of Theoretical Probability* **27**(3): 1011–1020.

Hu, M., Peng, S. and Song, Y. (2017). Stein type characterization for G-normal distributions, *Electronic Communications in Probability* **22**: 1–12.

Huber, P. J. (1981). Robust statistics, *Wiley Series in Probability and Mathematical Statistics* .

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

Ji, S. and Zhou, X. Y. (2010). A generalized Neyman–Pearson lemma for g-probabilities, *Probability theory and related fields* **148**(3): 645–669.

Jin, H. and Peng, S. (2016). Optimal unbiased estimation for maximal distribution, *arXiv preprint arXiv:1611.07994* .

Kallianpur, G. (2013). *Stochastic filtering theory*, Vol. 13, Springer Science & Business Media.

Karatzas, I. and Shreve, S. (2012). *Brownian motion and stochastic calculus*, Vol. 113, Springer Science & Business Media.

Keynes, J. M. (1921). *A treatise on probability*, Macmillan and Company, limited.

Keynes, J. M. (1937). The general theory of employment, *The quarterly journal of economics* **51**(2): 209–223.

Knight, F. H. (1921). *Risk, uncertainty and profit*, Vol. 31, Houghton Mifflin.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, 1933*, Springer, Berlin.

Laplace, P.-S. (1829). *Essai philosophique sur les probabilités*, H. Remy.

Laplace, P.-S. (2007). *A Philosophical Essay on Probabilities*, Cosimo, Inc.

Li, Y., Kulperger, R. and Yu, H. (2021). Semi-G-normal: a Hybrid between Normal and G-normal, *arXiv preprint arXiv:2104.04910* .

Lieb, E. H. and Loss, M. (2001). *Analysis*, Vol. 14, American Mathematical Soc.

Lin, L., Shi, Y., Wang, X. and Yang, S. (2013). Sublinear expectation linear regression, *arXiv preprint arXiv:1304.3559* .

Lo, A. W. and Mueller, M. T. (2010). Warning: physics envy may be hazardous to your wealth!, *arXiv preprint arXiv:1003.2688* .

Machina, M. J. and Schmeidler, D. (1992). A more robust definition of subjective probability, *Econometrica: Journal of the Econometric Society* pp. 745–780.

Mathai, A. M. and Pederzoli, G. (1977). *Characterizations of the normal probability law*, John Wiley & Sons.

Nendel, M. (2021). On nonlinear expectations and Markov chains under model uncertainty, *International Journal of Approximate Reasoning* **130**: 226–245.

Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications*, Springer Science & Business Media.

Pardoux, E. and Peng, S. (1990). Adapted solution of a backward stochastic differential equation, *Systems & Control Letters* **14**(1): 55–61.

Pardoux, E. and Peng, S. (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations, *Stochastic partial differential equations and their applications*, Springer, pp. 200–217.

Peng, S. (1991). Probabilistic interpretation for systems of quasilinear parabolic partial differential equations, *Stochastics and stochastics reports (Print)* **37**(1-2): 61–74.

Peng, S. (1992). A generalized dynamic programming principle and Hamilton-Jacobi-Bellman equation, *Stochastics: An International Journal of Probability and Stochastic Processes* **38**(2): 119–134.

Peng, S. (1993). Backward stochastic differential equations and applications to optimal control, *Applied Mathematics and Optimization* **27**(2): 125–144.

Peng, S. (1997). Backward SDE and related g-expectation, *Pitman research notes in mathematics series* pp. 141–160.

Peng, S. (2004). Nonlinear expectations, nonlinear evaluations and risk measures, *Stochastic methods in finance*, Springer, pp. 165–253.

Peng, S. (2007a). G-expectation, G-Brownian motion and related stochastic calculus of Itô type, *Stochastic analysis and applications*, Springer, pp. 541–567.

Peng, S. (2007b). Law of large numbers and central limit theorem under nonlinear expectations, *arXiv preprint math/0702358* .

Peng, S. (2008). A new central limit theorem under sublinear expectations, *arXiv preprint arXiv:0803.2656* .

Peng, S. (2010). *Nonlinear expectations and stochastic calculus under uncertainty*, Springer.

Peng, S. and Zhou, Q. (2020). A hypothesis-testing perspective on the G-normal distribution theory, *Statistics & Probability Letters* **156**: 108623.

Peng, S. et al. (1992). A nonlinear Feynman-Kac formula and applications, *Proceedings of Symposium of System Sciences and Control Theory*, World Scientific, pp. 173–184.

Royden, H. L. and Fitzpatrick, P. (1988). *Real analysis*, Vol. 32, Macmillan New York.

Sale, Y. (2021). Choquet-Erwartungsnutzen: Zur Choquet Integrationstheorie und ihrer Anwendung in der Entscheidungstheorie, Seminarpaper.

Sarin, R. and Wakker, P. (1992). A simple axiomatization of nonadditive expected utility, *Econometrica: Journal of the Econometric Society* pp. 1255–1272.

Savage, L. J. (1972). *The foundations of statistics*, Courier Corporation.

Schmeidler, D. (1986). Integral representation without additivity, *Proceedings of the American Mathematical Society* **97**(2): 255–261.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity, *Econometrica* **57**(3): 571–587.

Shamarova, E., Chertovskih, R., Ramos, A. F. and Aguiar, P. (2017). Backward-stochastic-differential-equation approach to modeling of gene expression, *Physical Review E* **95**(3): 032418.

Steele, J. M. (2001). *Stochastic calculus and financial applications*, Vol. 1, Springer.

Sun, C. and Ji, S. (2016). A generalized Neyman-Pearson lemma for sublinear expectations, *arXiv preprint arXiv:1605.05094* .

von Neumann, J. and Morgenstern, O. (1947). *Theory of games and economic behavior*, Princeton University Press.

Yang, S. and Yao, J. (2021). Linear regression under model uncertainty, *arXiv preprint arXiv:2108.02140* .

Zhang, J. (2017). *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*, Springer New York.

List of Figures

1 Hitting time T_1 of Brownian motion sample paths. 31
2 Visualization of (3.1.20) with different α values, $r = 1$ and $N_0 = 50$. . 37
3 Visualization of the fundamental solution for fixed $D = 2$ and different t. 60
4 Visualization of time-space lattice and an examplary realization of random walk with initial point $(0, 0)$. 61
5 Example of a generating function G. 63
6 Visuzalisation of key elements in this thesis. 77

List of Tables

1 Setting of the Ellsberg experiment. 8
2 Taxonomy of uncertainty. 10
3 Capacity μ 16
4 Formal symbol multiplication. 35

Declaration:

I declare that I completed this Master thesis independently and used only these materials that are listed. All materials used, from published as well as unpublished sources, whether directly quoted or paraphrased, are duly reported. Furthermore I declare that the Master thesis, or any abridgment of it, was not used for any other degree seeking purpose.

Place, Date

Signature (Yusuf Sale)