STATISTICAL CONSISTENCY FOR RISK MEASURES WITH THE LEBESGUE PROPERTY

by

Saumitra Mazumder
Bachelor of Science, Ryerson University, 2019

A thesis
presented to Ryerson University
in the partial fulfillment of the
requirements for the degree of
Master of Science (M.Sc.)
in the program of
Applied Mathematics

Toronto, Ontario, Canada, 2022

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public

Statistical consistency for risk measures with the Lebesgue property

Master of Science (M.Sc.), 2022

Saumitra Mazumder

Applied Mathematics

Ryerson University

Abstract

When estimating the risk $\rho(X)$ of a random variable X from historical data or Monte Carlo simulation, the asymptotic behaviour of the plug in estimator $\widehat{\rho}_n$ is of utmost importance. In their celebrated article [19], the Kraätschmer et al. showed that any finite-valued law-invariant convex risk measure ρ defined on an Orlicz heart $\mathcal{H}^\Phi$ is statistically consistent. That is, the plug-in estimator $\widehat{\rho}_n$ converges in the almost sure sense to $\rho(X)$. This result is very general, yet it does not cover the case where ρ is non-convex nor the case where ρ is defined on the entire Orlicz space $\mathcal{L}^\Phi$. The aim of this thesis is to fill this gap. In particular, we prove that any law-invariant risk measure with the Lebesgue property is statistically consistent on the entire Orlicz space. The Lebesgue property is a continuity condition that is automatically satisfied by all convex and finite-valued risk measures on Orlicz hearts. Thus our result can be viewed as a generalization of Theorem 2.6 in [19].

Acknowledgements

I would like to thank my supervisor, Professor Foivos Xanthos. His patience and his generosity in his time and understanding were a kindness during a difficult period. His mathematical prowess and almost encyclopedic knowledge were simultaneously humbling and inspiring. Finally, thank you for allowing me to ask silly questions and make silly mistakes.

Dedication

To my cats, meow.

To my Kate, I couldn't have done it without you. Sorry for all the late nights.

Table of Contents

Chapter 1

Introduction and Motivation

In very general terms, the concept of a risk measure is simply a function that maps a financial position with a real number such that it "measures" the riskiness of holding that financial position. Of course, the context of riskiness will depend on the risk measures' use. For example, a risk measure may be used to satisfy capital buffer requirements set by the regulator in order to buffer against unexpected loss. Regardless, only a certain subset of this mapping will be acceptable to a risk manager (see Section 3.2.1). A particular example of a risk measure that is widely used, perhaps to the detriment of a risk manager is *VaR*. VaR estimates how much a risky position might lose (within a given probability), given normal market conditions. In industry, it is used to gauge the amount of easily liquid assets needed to cover possible losses. This measure initially saw many benefits when popularized by J.P. Morgan and its wide use in industry makes it well studied and modelled. It has a prominent use in the Basel regulatory requirements as the boundary between expected loss and unexpected loss. Furthermore VaR quantifies levels of risk that are easily explainable to non-quants.

The overuse of VaR is thought to be one of the myriad of causes of the global financial crisis of 2007-2008 [13]. Indeed, it is known that VaR lacks a diversification property of holding more than one risky positions [27].

In lieu of this and other criticisms, *Coherent Risk Measures* were introducted by Artzner et. al in their seminal work, "Thinking Coherently" [2]. They proposed four axioms that risk measures must satisfy in order to be "coherent" (see Section 3.2.1). Their work was generalized to convex risk measures, first by Folmer and Schied and then by Delbaen [15, 11]. Work by Cherdito and Li and by Kraätschmer, Schied, et. al. extended convex risk measures to the Orlicz heart [9, 19].

Together, these results are very general, yet they do not cover the case where the risk measure is non-convex nor the case where it is defined on the entire Orlicz space. We aim to fill this gap. In particular, we prove that any law-invariant risk measure with the Lebesgue property is statistically consistent on the entire Orlicz space. Thus our result can be viewed as a generalization of Theorem 2.6 in [19].

This thesis contains the following in order: First we review some preliminary material required for our understanding of probability theory and functional analysis on convex risk measures, then we introduce the Orlicz spaces that we will be working on. Finally, we concluded with the final result of this thesis. An appendix section collects various intermediate results that were useful in the final proof. As an aside, the author hopes that it is an approachable read for those interested in convex risk measures.

1.1 Estimating Risk, The Strong Law of Large Numbers and Glivenko-Cantelli's Theorem

A layman unfamiliar with the particulars of probability may assume that risk measurement is much like a casino where the probabilistic structures of the games are generally known. However, a risk manager is faced with concurrent dilemmas: (1) The risk manager does not know the outcome of an event until is has occured, and even worse, (2) they cannot make any assumptions of the probabilistic structure of an economic situation.

To have any hope of being able to manage risk, the risk manager must rely on historical information to make inferences on the true probabilistic structure of the world through the use of the *empirical distribution.* Via the *strong law of large numbers* a risk manager can be assured (under the assumption that each realization of an event is independent of others) that the empirical distribution created by many realizations of the random variable of interest will differ from the true distribution by a probability of 0. Indeed, this method is exploited in industry to estimate the distribution of losses under a sequence of realizations.

Lets make more precise the law of large numbers and the empirical distribution.

Definition 1.1.1 (Empirical Distribution Function)**.** *Let X_n be a sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ having common distribution function F. The empirical distribution function for $(X_1, \cdots, X_n)$ is defined by*

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

We make the note that the distribution function is right-continuous, that is for any $x_n \downarrow a$, we have $F(x_n) \to F(a)$. This right-sided continuity is a matter of historical preference and one can similarly construct a left-continuous version.

Definition 1.1.2 (Strong Law of Large Numbers (SLLN))**.** *Let* $(X_1, \cdots, X_n)$ *be a sequence of i.i.d random variables such that* $\mathbb{E}[X_{n_0}] < \infty$*. Denote* $S_n = \sum_{i=1}^{n} X_n$*. Then as* $n \to \infty$*, we have*

$$\frac{S_n}{n} \overset{a.s.}{\to} \mathbb{E}[X_{n_0}].$$

Moreover, when $X_{n_0} \in \mathcal{L}^1$*, we have,*

$$\mathbb{E}\left[\left|\frac{S_n}{n} - \mathbb{E}[X_{n_0}]\right|\right] \to 0.$$

We make note that in the above definition of the empirical distribution that each $\mathbb{E}[\mathbb{1}_{\{X_i \leq x\}}] < 1$ and hence by the *strong law of large numbers*, the empirical distribution function converges to the true distribution as $n \to \infty$ almost surely. That is,

$$F_n(x) \overset{\text{a.s.}}{\to} F(x).$$

The risk manager may use the *Kolmogorov-Smirnov statistic*, to test the goodness of fit between the empirical distribution created by sampling and the assumed distribution that the samples have. Such a statistic is an application of the following theorem,

Theorem 1.1.3 (Glivenko–Cantelli Theorem)**.** *Suppose* $F_n(x)$ *is a sequence of random variables as in Definition 1.1.1. We know by SLLN for that every fixed* x*, we have that* $F_n(x)$ *converges almost surely to* $F(x)$*. Glivenko and Cantelli tells us that,*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \overset{a.s.}{\to} 0.$$

The above theorem is also known as the *Fundamental Theorem of Statistics* and strengthens the asymptotic properties of the empirical distribution function.

Let us take a moment to dive a bit deeper into Glivenko-Cantelli's theorem. In particular, it strengthens the usual *SLLN* convergence to *uniform* convergence of F_n to F. The reader may note that $F_n(x) \overset{\text{a.s.}}{\to} F(x)$ already implies uniform convergence when F is continuous, since distributions are bounded and monotone. However, the conclusion of Glivenko-Cantelli's theorem strengthens uniform convergence to intervals that contain discontinuities. That is, F_n is a reasonable estimate of F independent of x chosen [16, 29].

For an introductory treatment of the above process in view of finance, we direct the reader to [37]. For a comprehensive treatment of risk measure procedures, the reader can see [10].

An Example of SLLN and Glivenko-Cantelli's Theorem

Suppose we take repeated realizations of $X \sim \mathcal{N}(0,1)$ and create the empirical distribution as in Definition 1.1.1. As n gets very large, we can see that $F_n(x)$ becomes a better approximation of $F(x)$.

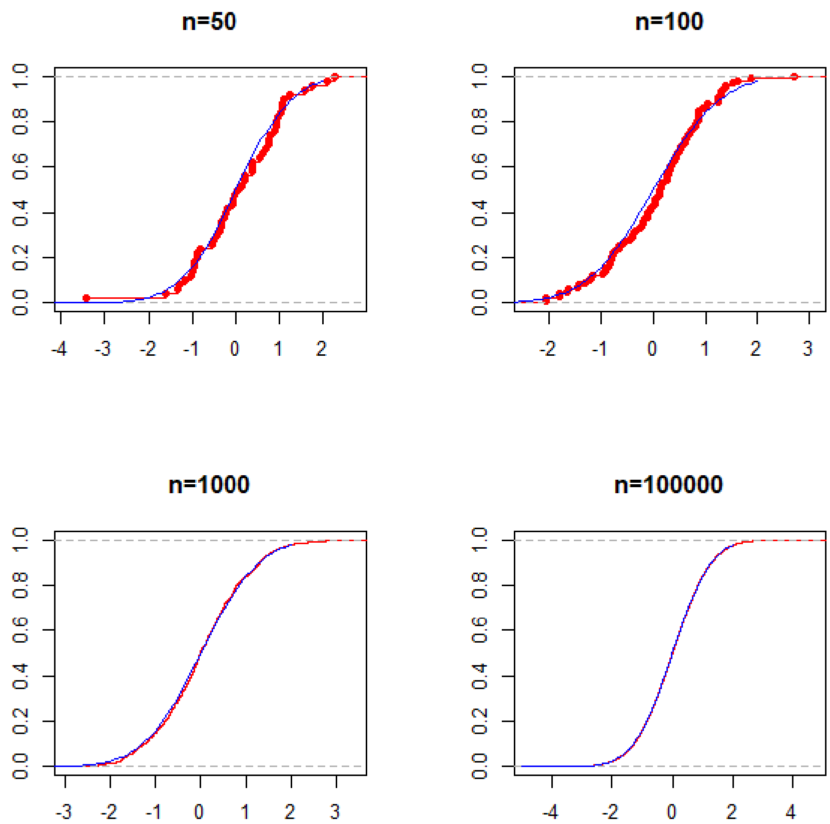

Figure 1.1: Empirical distributions with the normal distribution overlayed for sample sizes $n = 50, 100, 1000, 100000$.

Here we see the SLLN in action as n becomes large. That is, for a fixed $x \in \mathbb{R}$ we find that $F_n(x)$ is a good enough approximation of $F(x)$. Furthermore, by Theorem 1.1.3, we can find an arbitrary uniform bound between F_n and F regardless of x chosen.

Chapter 2

Preliminaries

This preliminaries section contains what the author believes is required for consistency of the lebesgue risk measures. It contains has two parts, some measure theory the author reviewed while attempting to understand the material in the [9, 19], and a lot of basic functional analysis that the author felt was useful in the formation of $\mathcal{L}^p$ spaces. Although this section was the most valuable to the author, it perhaps has the least valuable to the reader. This section can clearly be skipped for readers who have knowledge in either areas expanded upon in this section.

Throughout this thesis, we consider an non-atomic measure space, denoted $(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, we take our random variables from the $\mathcal{L}^0 := \mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ space and we do not make a distinction between random variables that are equivalent $\mathbb{P}$-almost surely.

2.1 Some Measure Theory

We now recall some basic definitions and results from Measure Theory. We will then move to after which we will move towards probability theory. Probability theory is the ground on which this thesis is built upon. A comprehensive treatment of what is forthcoming can be found in [14, 35, 16].

We first construct sets that are relevant to us. These sets are constructed so that we can measure and thus compare them.

Definition 2.1.1 (Measurable Space)**.** *A **Measure Space** is an ordered pair* $(\Omega, \mathcal{F})$ *where* Ω *is a set and* $\mathcal{F}$ *is a collection of subsets of* Ω*, that satisfy the following axioms:*

1. *The empty set* $\emptyset$ *and* Ω *itself belong to* $\mathcal{F}$*.*

2. *For any member, say* E*, in* $\mathcal{F}$*, the complement,* $\Omega \setminus E$ *is also a member of* $\mathcal{F}$*.*

3. *Countable union of members of* $\mathcal{F}$ *is also a member of* $\mathcal{F}$*.*

If $\mathcal{F}$ *satisfies the above conditions, then it is a* ***σ-algebra****.*

Once we have the sets that we can measure, we can move on to the properties of the measure that are useful to us.

Definition 2.1.2 (Measure)**.** *A* ***measure*** *is a function* $\mu : \mathcal{F} \mapsto \overline{\mathbb{R}}_+$ *such that*

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$$

for any pairwise-disjoint *members* E_n *of* $\mathcal{F}$*.*

1. μ *is bounded if* $\mu(\Omega) < \infty$*.*

2. μ *is* σ*-finite if there are countable pairwise disjoint members* E_n *of* $\mathcal{F}$ *such that* $\mu(E_n) < \infty$ *and* $\bigcup_{n\in\mathbb{N}} E_n = \Omega$*.*

Clearly if μ is bounded, then it is also σ-finite, since we know that $\emptyset \cap \Omega = \emptyset$ and $\mu(\emptyset \cup \Omega) = \mu(\Omega) < \infty$. However, the converse is not necessarily true.

A measurable space endowed is a measure is them a *measure space* denoted by the triple $(\Omega, \mathcal{F}, \mu)$. An immediate consequence of the previous definition is as follows:

Theorem 2.1.3 (Continuous Measure)**.** *For* $E \in \mathcal{F}$*,* μ *is* ***continuous from below*** *at* E *if for any* $(E_n)_{n\geq 1} \in \mathcal{F}$ *such that* $E_n \uparrow E$*, then we have,*

$$\mu(E_n) \to \mu(E).$$

Similarly, for $E \in \mathcal{F}$*,* μ *is* ***continuous from above*** *at* E *if for any* $(E_n)_{n\geq 1} \in \mathcal{F}$ *such that* $E_n \downarrow E$ *and there is some* n_0 *where* $\mu(E_{n_0}) < \infty$*, then we have,*

$$\mu(E_n) \to \mu(E).$$

Once we have our measurable sets and a method for measuring them, we need to consider what type of functions interest us. In particular, we only see relevance in functions wherein its inverse is one of our measurable sets.

Definition 2.1.4 (Measurable Functions). *A function* $f : ((\Omega, \mathcal{F}) \mapsto (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ *is **$\mathcal{F}$-measurable** or simply **measurable** if for any* $B \in \overline{\mathcal{B}}$ *we have*

$$f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F}.$$

Using tests for measurability (Lemma A.1.3), the following useful facts emerge,

1. For $f : \Omega \to \overline{\mathbb{R}}_+$, there exists a sequence of nonnegative simple functions $(f_n)_{n\in\mathbb{N}}$ such that $f_n \uparrow f$.

2. For measurable functions $f, g : \Omega \to \overline{\mathbb{R}}$ and $\alpha \in \mathbb{R}$, αf, $f + g$ and fg, if they exist, are also measurable.

3. For $f : \Omega \to \overline{\mathbb{R}}$, $f^+ := \sup(f, 0)$, $f^- := \sup(-f, 0)$ and $|f| := f^+ + f^-$, if they exist, are measurable.

4. For a measurable sequence of functions $f_n : \Omega \to \overline{\mathbb{R}}$, $\sup f_n$, $\inf f_n$, $\limsup f_n$, $\liminf f_n$, if they exist, are also measurable.

5. For a measurable sequence of functions $f_n : \Omega \to \overline{\mathbb{R}}$, if $f_n \to f$ pointwise, then f is measurable.

An immediate consequence of Definitions A.1.2 and 2.1.4 is that we can define a sequence of real-valued measurable functions on Ω, denoted $(f_n)_{n\in\mathbb{N}}$, which leads to the following theorems which are of great importance in measure theory:

Lemma 2.1.5 (Fatou's Lemma). *Fix a measure space* $(\Omega, \mathcal{F}, \mu)$ *and put* $\mathcal{L}^+$ *as the space of all measurable functions from* Ω *to* $[0, \infty]$.

If (f_n) *is a sequence in* $\mathcal{L}^+$ *then,*

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

A trivial corollary to Fatou's Lemma in [14] is as follows,

Corollary 2.1.6. *Fix a measure space* $(\Omega, \mathcal{F}, \mu)$ *and put* $\mathcal{L}^+$ *as the space of all measurable functions from* Ω *to* $[0, \infty]$.

If (f_n) *is a sequence in* $\mathcal{L}^+$ *and* f *is in* $\mathcal{L}^+$ *such that* $f_n \xrightarrow{\mu} f$, *then*

$$\int_\Omega f d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

Theorem 2.1.7 (Dominated Convergence Theorem)**.** *Fix a measure space* $(\Omega, \mathcal{F}, \mu)$*. For* $(f_n) \in \mathcal{L}^0$ *a function* $g \in \mathcal{L}^1$ *such that* $|f_n| \leq g$ *a.e., we have*

$$-\infty < \int (\liminf_{n\to\infty} f_n) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu \leq \limsup_{n\to\infty} \int f_n d\mu \leq \int (\limsup_{n\to\infty} f_n) d\mu < \infty.$$

As implied in the title, we are interested in the *statistical properties of a risk measure.* This requires us to work on a space less rich than a general measure space. In fact, we will run into trouble if we were to consider a general measure. Thus we restrict ourselves to a particular *finite* measure with certain properties as will be stated. The discussion of measure theory above applies wholly to probability theory.

When dealing with probability theory, we refer to a Measure Space as a *Probability Space.* That is, a probability space is a measure space endowed with the complete measure $\mathbb{P}$ such that, $\mathbb{P} : (\Omega, \mathcal{F}) \to ([0,1], \mathcal{B}_{[0,1]})$, and $\mathbb{P}(\Omega) = 1$. A measurable set in the σ-algebra $\mathcal{F}$ is an *event.* Finally, a measurable function X is a *random variable* which is a mapping $X : (\Omega, \mathcal{F}) \to (\overline{R}, \mathcal{B}(\overline{\mathbb{R}}))$.

Definition 2.1.8 (Probability Law)**.** *The* ***Probabilty Law*** *of a random variable* X *is defined by*

$$\mathbb{P}_X(B) = \mathbb{P}[X^{-1}(B)]$$

for any B *in the Borel* σ*-algebra* $\mathcal{B}(\overline{\mathbb{R}})$*.*

With the probability law defined, we can speak of types of convergence in our probability space. We go from the strongest to the weakest.

The strongest form of convergence of interest is,

Definition 2.1.9 (Almost Sure Convergence)**.** *A sequence of random variables* $(X_n)_{n\in\mathbb{N}}$ ***converges almost surely*** *to a random variable* X*, denoted* $X_n \overset{a.s}{\to} X$ *as* $n \to \infty$ *if there is a member of* $\mathcal{F}$*, say* E_0 *such that* $\mathbb{P}(E_0) = 1$ *and* $X_n(\omega) \to X(\omega)$ *as* $n \to \infty$*, for every* $\omega \in E_0$*.*

The above almost sure convergence implies the weaker *convergence in probability*,

Definition 2.1.10 (Convergence in Probability)**.** *A sequence of random variables* $(X_n)_{n\in\mathbb{N}}$ ***converges in probability*** *to a random variable* X*, denoted* $X_n \overset{\mathbb{P}}{\to} X$ *as* $n \to \infty$ *if for any* $\varepsilon > 0$ *we have* $\mathbb{P}(|X_n - X| \geq \varepsilon) \to 0$*.*

Convergence in probability has the following necessary and sufficient condition that is quite useful [14].

$X_n \xrightarrow{\mathbb{P}} X$ if and only if for all subsequences X_{n_k} of X_n there exists a further subsequence $X_{n'_k} \xrightarrow{\text{a.s.}} X$.

When in a probability space, our integrals are denoted as follows,

Definition 2.1.11 (Expectation)**.** *When X is integrable, it's integral is denoted $\mathbb{E}[X]$ and is called the **expectation** of random variable X:*

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} x d\mathbb{P}_X(x)$$

Using our definition of expectation, we have convergence in $\mathcal{L}^p$ norm,

Definition 2.1.12 (Convergence in $\mathcal{L}^p$ norm)**.** *Fix some $p \geq 1$, a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ **converges in $\mathcal{L}^p$ norm** to a random variable X, if $\mathbb{E}[|X_n|^p]$ and $\mathbb{E}[|X|^p]$ are finite and*

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

Such a convergence is denoted $X_n \xrightarrow{\mathcal{L}^p} X$ or $X_n \xrightarrow{\|\cdot\|_{\mathcal{L}^p}} X$. Convergence in $\mathcal{L}^p$ norm implies convergence in probability by *Markov's Inequality.* Furthermore, by (3) from Property B.4.3, we have if $X_n \xrightarrow{\mathcal{L}^p} X$ then $\lim_{n\to\infty} \mathbb{E}[|X_n|^p] \to \mathbb{E}[|X|^p]$. We note that the expectation operator $\|\cdot\|_{\mathcal{L}^p} := \mathbb{E}[|\cdot|]$ is a norm in the functional analytic sense when we do not make a distinction between random variables equal almost surely (see Appendix B).

Before we continue to the weakest convergence of interest for us, we introduce distribution functions and some of their properties.

Definition 2.1.13. *A real-valued function F defined on $\mathbb{R}$ is a **distribution function** for $\mathbb{R}$ if it is increasing and right-continuous such that*

$$\lim_{x\to-\infty} F(x) = 0 \quad \textit{and} \quad \lim_{x\to\infty} F(x) = 1$$

Proposition 2.1.14. *Let $\mathbb{Q}$ be a probability measure on $(\mathbb{R}, \mathcal{B})$, then $F(x) = \mathbb{Q}(-\infty, x]$ is a distribution function.*

Proof. By monotonicity of $\mathbb{Q}$, we have that F is increasing. Furthermore, since $\mathbb{Q}$ is complete we have for $(\infty, x_n] \downarrow (\infty, x]$ that $\lim_{n\to\infty} \mathbb{Q}\left(\cap_{i=1}^n (\infty, x_n]\right) = \mathbb{Q}((\infty, x])$ for $x_n \downarrow x$.

Similarly, by continuity of $\mathbb{Q}$, and $x_n \downarrow -\infty$ and $x_n \uparrow \infty$, we have $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$ respectively. ■

We also observe a fact about the distribution function with respect to a non-atomic probability space.

Observation 2.1.15. *The distribution function F_X is continuous if and only if $\mathbb{P}(x) = 0$ for any $x \in \mathbb{R}$*

Returning to convergence, the weakest notion of convergence of our interest is *convergence in distribution.*

Definition 2.1.16 (Convergence in Distribution)**.** *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is said to* ***converge in distribution*** *to a random variable X, denoted $X_n \xrightarrow{d} X$ as $n \to \infty$ if*

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x) \quad \text{for any } x \text{ such that } \mathbb{P}(X = x) = 0.$$

In view of Proposition 2.1.14, we have equivalently that $X_n \xrightarrow{d} X$ if $\lim_{n\to\infty} F_n(x) = F(x)$ for every $x \in \mathbb{R}$ at which F is continuous. Here, F_n and F are distribution functions for X_n and X respectively.

Perhaps surprisingly, even if a sequence of random variables $X_n \xrightarrow{\text{a.s.}} X$, it may be the case that neither $X_n \xrightarrow{\mathcal{L}^1} X$ nor $\mathbb{E}[X_n] \to \mathbb{E}[X]$ may hold. For a sufficient condition when that occurs, we turn to *Lebesgue's Dominated Convergence Theorem.*

Theorem 2.1.17 (Lebesgue's Dominated Convergence Theorem)**.** *If $X_n \to X$ almost surely, and there is an random variable $Y \in \mathcal{L}^1$ such that $|X_n| \leq Y$, then $X \in \mathcal{L}^1$, $\mathbb{E}[X_n] \to \mathbb{E}[X]$ and $\mathbb{E}[|X_n - X|] \to 0$.*

We will show in Section 2.3 that uniform integrability provides a more general sufficient condition giving rise to an *improved* dominated convergence theorem. In particular, we will consider the case of $\mathbb{E}[|\Phi(k_0 |X_n|) - \Phi(k_0 |X|)|] \to 0$ for a particular convex function Φ and a constant $k_0 > 0$.

2.2 Some Functional Analysis and Scheffe's Lemma

Functional Analysis deals with spaces that have a distance (norm) between two objects in the space. The Riesz-Fisher Theorem shows not only is the familiar $\mathcal{L}^p$ space with the expectation operator acting as our distance function a functional space, but also it is a complete linear space (Theorem B.4.1).

Before we get to that stage, lets precisely state our definition for complete linear spaces.

Definition 2.2.1 (Banach Space)**.** *A **Banach Space** is a* normed linear space *which is also* complete*. That is,*

*If (x_n) is a Cauchy sequence in a normed linear space $\mathcal{X}$, then there is at least one x in $\mathcal{X}$ such that $x_n \to x$. Banach spaces are also called **complete normed spaces**.*

In general, a normed linear space is not a Banach Space. That is a normed linear space may be "missing" some element x such that $x_n \to x$. Because Cauchy sequences are the sequences whose terms grow arbitrarily closer together, the normed linear space where *all* Cauchy sequences converge are the spaces that are not "missing" any numbers. Furthermore, any divergent sequence is "truly" divergent, that is, there is no bigger normed linear space which makes it convergent.

Since in a normed linear space each convergent sequence is Cauchy, we have the following Cauchy convergence criterion:

Theorem 2.2.2 (Cauchy Convergence Criterion)**.** *In a Banach space, a sequence is convergent if and only it is Cauchy.*

Corollary 2.2.3. *If $X_n \stackrel{\mathcal{L}^p}{\to} X$ for $1 \leq p < \infty$, then there is a subsequence (X_{n_k}) such that $X_{n_k} \stackrel{a.s.}{\to} X$.*

In view of the Cauchy convergence criterion and Riesz-Fisher's theorem, the proof of the above is clear, since every convergent sequence is a Cauchy sequence and by Theorem B.4.1, every Cauchy sequence contains an almost everywhere convergence subsequence with limit equal to the norm limit of the Cauchy sequence.

Theorem 2.2.4 (Scheffe's Lemma)**.** *For the normed space $\mathcal{L}^1$, suppose there is a sequence (X_n) and (X) both in $\mathcal{L}^1$ such that $X_n \stackrel{a.s}{\to} X$. Then $\|X_n\| \to \|X\|$, if and only if $X_n \stackrel{\|\cdot\|_{\mathcal{L}^1}}{\to} X$.*

Proof. The "only if" part is straight forward. Suppose, $\lim_{n\to\infty} \|X_n - X\| = 0$. we have,

$$\begin{aligned}
\big| \|X_n\| - \|X\| \big| &= \Big| \mathbb{E}[|X_n|] - \mathbb{E}[|X|] \Big| = \Big| \mathbb{E}\left[\big|X_n\big| - |X| \right] \Big| \\
&\leq \mathbb{E}\left[\Big| \big|X_n\big| - \big|X\big| \Big| \right] \text{ by (3) in Property B.4.3} \\
&\leq \mathbb{E}\left[|X_n - X| \right] \text{ by reverse triangle inequality} \\
&= \|X_n - X\| \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Hence, $\|X_n\| \to \|X\|$.

For "if", suppose $\|X_n\| \to \|X\|$. Fix n and partition $\Omega = \{\omega : X_n(\omega)X(\omega) \geq 0\} \cap \{\omega : X_n(\omega)X(\omega) < 0\}$. Then,

$$\mathbb{E}[|X_n - X|] = \mathbb{E}[|X_n - X| \mathbb{1}_{\{X_n X \geq 0\}}] + \mathbb{E}[|X_n - X| \mathbb{1}_{\{X_n X < 0\}}].$$

On $\{X_n X \geq 0\}$, either X_n and X are both nonpositive or are both nonnegative. In this case, $\mathbb{E}[|X_n - X| \mathbb{1}_{\{X_n X \geq 0\}}] = \mathbb{E}\left[\big||X_n| - |X|\big| \mathbb{1}_{\{X_n X \geq 0\}}\right]$. On $\{X_n X < 0\}$, X_n and X have opposite signs, and hence $\mathbb{E}[|X_n - X| \mathbb{1}_{\{X_n X < 0\}}] = \mathbb{E}\left[\big||X_n| + |X|\big| \mathbb{1}_{\{X_n X < 0\}}\right]$.

Thus,

$$\begin{aligned}
\mathbb{E}[|X_n - X|] &= \mathbb{E}\left[\big||X_n| - |X|\big| \mathbb{1}_{\{X_n X \geq 0\}}\right] + \mathbb{E}\left[\big||X_n| + |X|\big| \mathbb{1}_{\{X_n X < 0\}}\right] \\
&\leq \mathbb{E}\left[\big||X_n| - |X|\big| \mathbb{1}_{\{X_n X \geq 0\}}\right] + \mathbb{E}\left[(\big||X_n| - |X|\big| + 2|X|) \mathbb{1}_{\{X_n X < 0\}}\right] \\
&= \mathbb{E}\left[\big||X_n| - |X|\big|\right] + 2\mathbb{E}[|X| \mathbb{1}_{\{X_n X < 0\}}]
\end{aligned}$$

Now, applying the fact that $|x| = x^+ + x^- = x^+ - x^- + 2x^-$ and hence,

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n| - |X|] + 2\mathbb{E}[(|X_n| - |X|)^-] + 2\mathbb{E}[|X| \mathbb{1}_{\{X_n X < 0\}}].$$

By sending $n \to \infty$ and assumption, we have $\mathbb{E}[|X_n| - |X|] = 0$. By $X_n \overset{\text{a.s.}}{\to} X$ and noticing that (1) $(|X_n| - |X|)^- \leq |X|$ regardless of whether $|X_n| \geq |X|$ or $|X_n| < |X|$, (2) $X \in \mathcal{L}^1$ and (3) $\{X_n X < 0\} \downarrow \varnothing$, and applying *dominating convergence theorem* as in Theorem 2.1.17, we finally have,

$$\mathbb{E}[|X_n - X|] \to 0.$$

■

2.3 Uniform Integrability

At last in this section we reach the refinement of dominating convergence theorem that we stated earlier. First, we define *Uniform Integrability*.

Definition 2.3.1 (Uniform Integrability). *A non-empty family $\mathcal{X} \subset \mathcal{L}^0$ of random variables is said to be* ***uniformly integrable (UI)*** *if*

$$\lim_{n \to \infty} \left(\sup_{X \in \mathcal{X}} \mathbb{E}\left[|X| \mathbb{1}_{\{|X| \geq n\}}\right] \right) = 0.$$

That is, for integrable random variables, the far tails contribute very little to the expectation.

Proposition 2.3.2.

$$\lim_{n\to\infty}\left(\mathbb{E}\left[|X|\,\mathbb{1}_{\{|X|\geq n\}}\right]\right)=0 \textit{ if and only if } X\in\mathcal{L}^1$$

Proof. Suppose $X\in\mathcal{L}^1$, that is, $\mathbb{E}[|X|]<\infty$.

Now,

$$n\mathbb{P}\{|X|\geq n\}=\mathbb{E}[n\mathbb{1}_{\{|X|\geq n\}}]\leq\mathbb{E}[|X|\,\mathbb{1}_{\{|X|\geq n\}}]\leq\mathbb{E}[|X|] \text{ for every } n\in\mathbb{N}.$$

Hence,

$$\mathbb{P}\{|X|=\infty\}=\lim_{n\to\infty}\mathbb{P}\{|X|\geq n\}\leq\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[|X|]=0.$$

Since we are taking an integral over a null set, $\mathbb{P}\{|X|=\infty\}=0$, we have,

$$\lim_{n\to\infty}\mathbb{E}[\mathbb{1}_{\{|X|\geq n\}}\,|X|]=\lim_{n\to\infty}\int_{\{|X|\geq n\}}|X|\,d\mathbb{P}=0.$$

To show the other direction, toward a contradiction, suppose $\lim_{n\to\infty}\left(\mathbb{E}\left[|X|\,\mathbb{1}_{\{|X|\geq n\}}\right]\right)=0$ and $X\notin\mathcal{L}^1$.

Notice that,

$$\begin{aligned}\mathbb{E}[|X|]&=\mathbb{E}[|X|\,\mathbb{1}_{\{|X|\geq n\}}]+\mathbb{E}[|X|\,\mathbb{1}_{\{|X|<n\}}] \text{ for every } n\\&\leq\mathbb{E}[|X|\,\mathbb{1}_{\{|X|\geq n\}}]+n\mathbb{P}\{|X|\leq n\}\\&\leq\mathbb{E}[|X|\,\mathbb{1}_{\{|X|\geq n\}}]+n_0 \text{ for some } n_0\in\mathbb{R}.\end{aligned}$$

Taking limits over n, we have $+\infty=\mathbb{E}[|X|]\leq n_0$, which is a contradiction. Thus, we have $\lim_{n\to\infty}\left(\mathbb{E}\left[|X|\,\mathbb{1}_{\{|X|\geq n\}}\right]\right)=0$ if and only if $X\in\mathcal{L}^1$.

■

Some immediate observations from Definition 2.3.1 on the way to a *Criterion for Uniform Integrability*:

Observation 2.3.3 (Uniform Integrability)**.**

1. *The second condition of* Lebesgue convergence *from Theorem 2.1.17 implies uniform integrability of* (X_n).

   Proof. *Notice,* $\mathbb{E}\left[\sup_{X\in\mathcal{X}}|X|\,\mathbb{1}_{\{|X|\geq k\}}\right]\leq\mathbb{E}\left[Y\mathbb{1}_{\{|Y|\geq k\}}\right]$ *by the monotonicity of expectation. By Proposition 2.3.2, the proof follows.* ■

2. *Since $\mathbb{P}(\Omega) = 1$ and when $\mathbb{P}$ a complete measure, any $X \in \mathcal{L}^1$ is trivially uniformly integrable.*

   Proof. *Recall that for $X \in \mathcal{L}^1$, we have that $|X|$ is finite a.e. Set $X_n := \mathbb{1}_{\{|X| \geq n\}} |X|$, then we clearly have $X_n \leq |X| \leq M_0$ a.e., and the proof follows from the first observation above.* ■

3. *$\mathcal{X}$ is a class of uniformly integrable random variables, if and only if*

   (i) $\sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < +\infty$, *and*

   (ii) for every $\varepsilon > 0$ there exists $\delta > 0$ such that, for every $A \in \mathcal{F}$ such that $\mathbb{P}(A) < \delta$ and every $X \in \mathcal{X}$, we have $\mathbb{E}\left[\mathbb{1}_A |X|\right] < \varepsilon$.

   Proof. *Fix $X \in \mathcal{X}$, such that $\mathcal{X}$ is a class of uniformly integrable random variables. Indeed, $\mathbb{E}[|X|] = \mathbb{E}\left[|X| \mathbb{1}_{\{|X| \geq k\}}\right] + \mathbb{E}\left[|X| \mathbb{1}_{\{|X| < k\}}\right] \leq \mathbb{E}\left[|X| \mathbb{1}_{\{|X| \geq k\}}\right] + k_0$. And again by Proposition 2.3.2, it follows that $\sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < +\infty$.*

   Now,

$$\begin{aligned}\mathbb{E}[|X| \mathbb{1}_A] &= \mathbb{E}[|X| \mathbb{1}_{A \cap \{|X| \geq k\}}] + \mathbb{E}[|X| \mathbb{1}_{A \cap \{|X| < k\}}] \\ &\leq \mathbb{E}[|X| \mathbb{1}_{A \cap \{|X| \geq k\}}] + k_0 \mathbb{P}(A).\end{aligned}$$

   For sufficiently large k, $\mathbb{E}[|X_n| \mathbb{1}_{A \cap \{|X_n| \geq k\}}] \leq \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| \geq k\}}] < \frac{\varepsilon}{2}$. By putting $\mathbb{P}(A) < \delta = \frac{\varepsilon}{2k_0}$, we have the desired inequality, $\mathbb{E}[|X_n| \mathbb{1}_A] \leq \varepsilon$.

   For the other direction, suppose $\sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < +\infty$ and $\lim_{\mathbb{P}(A) \to 0} \mathbb{E}[|X| \mathbb{1}_A] = 0$ holds for some collection of random variables $\mathcal{X}$ and put $M := \sup_{X \in \mathcal{X}} \mathbb{E}[|X|]$. By Chebshev, we have:

$$\mathbb{P}(\{|X| \geq n\}) \leq \frac{1}{n} \mathbb{E}[|X|] \leq \frac{M}{n}$$

   For large enough n from (ii), we can put $\frac{M}{n} < \delta = \varepsilon$, then $\sup_{X \in \mathcal{X}} \mathbb{E}[|X| \mathbb{1}_{\{|X| \geq n\}}] \leq \sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < \varepsilon$ as desired.

   ■

Theorem 2.3.4 (Uniform Integrability Criterion). *For a sequence $(X_n) \in \mathcal{L}_1$ such that $X_n \overset{a.s.}{\to} X$, the following are equivalent:*

1. *Each X_n is uniformly integrable,*

2. *$X \in \mathcal{L}^1$ and $X_n \overset{\|\cdot\|_{\mathcal{L}_1}}{\to} X$,*

3. $\|X_n\| \to \|X\|$.

Any of the above conditions imply $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

The reader can find the proof of the above in [16]. We can weaken the above as follows:

Theorem 2.3.5 (Vitali's Convergence Theorem)**.** *For a class of uniformly integrable random variables,*$(X_n) \in \mathcal{X}$*,* $X_n \xrightarrow{\mathbb{P}} X$ *if and only if* $X \in L^1$ *and* $X_n \overset{\|\cdot\|_{\mathcal{L}^1}}{\to} X$*.*

Indeed, together Theorems 2.3.4 and 2.3.5 are necessary and sufficient conditions for $\mathcal{L}^1$ convergence. The above is a *Generalization of Lebesgue's Dominated Convergence Theorem.*

Proof. The necessary condition holds in general by Chebychev that for a fixed $p \geq 1$, we have $\mathbb{P}(\{|X_n - X|^p\} > \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}[|X_n - X|^p]$. For the sufficient condition, consider that if $X_n \xrightarrow{\mathbb{P}} X$, there exists a subsequence $X_{n_k} \xrightarrow{\text{a.s.}} X$. By an application of Fatou's Lemma as in Corollary 2.1.6, we have

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf_{k\to\infty} |X_{n_k}|] \leq \liminf_{k\to\infty} \mathbb{E}[|X_{n_k}|] \leq \sup_{X_n \in \mathcal{X}} \mathbb{E}[|X_n|] < \infty,$$

since each $\mathcal{X}$ is a uniformly integrable set and thus bounded in $\mathcal{L}^1$. Hence $X \in \mathcal{L}^1$ and thus $\{|X_n - X|\}$ is also a uniformly integrable set by Observation 2.3.3.

Now to show convergence in norm, notice that since there exists a subsequence $X_{k_n} \xrightarrow{\text{a.s.}} X$, we have a further subsequence $X_{n'_k} \xrightarrow{\mathbb{P}} X$. Hence there exists another further subsequence $X_{n''_k} \xrightarrow{\text{a.s.}} X$. By the first part of the sufficient condition, we have $X_{n''_k} \overset{\|\cdot\|_{\mathcal{L}^1}}{\to} X$. But notice that the limit X does not depend on the subsequence. Hence by reductio ad absurdum, we have that X_n itself converges to X in norm.

■

2.4 Orders and Lattices

We shall use this section to refine some of the notions of convergence introduced in the previous sections. As before, we first need to get some preliminaries out of the way.

Definition 2.4.1 (Lattice)**.** *A partially ordered set* $(\mathcal{X}, \leq)$ *is a lattice if each pair of elements* $x, y \in \mathcal{X}$ *has a supremum (or least upper bound) and an infimum (or greatest lower bound).*

For a lattice $\mathcal{X}$, if $X \in \mathcal{X}$, then $|X| \in \mathcal{X}$, since $|X| = \sup\{X, 0\} + \sup\{-X, 0\}$. An ordered vector space that is also a lattice is called a Riesz space or a vector lattice. A norm $\|\cdot\|$ is called a *lattice norm* if:

$$|x| \leq |y| \text{ then } \|x\| \leq \|y\| \text{ for each } x, y \in \mathcal{X}.$$

In particular, a Banach lattice is a Banach space endowed with a partial ordering such that $(\mathcal{X}, \leq)$ is a vector lattice and the the norm $\|\cdot\|$ is a lattice norm.

Definition 2.4.2 (Order Bounded Sets). *A subset $\mathcal{A}$ of $\mathcal{L}^0$ is* ***order bounded from above*** *if there is an element $u \in \mathcal{A}$ that dominates each element of $\mathcal{A}$. Similary, sets* ***order bounded from below*** *if there is an element $u \in \mathcal{A}$ that is dominated by every element of $\mathcal{A}$. Notice that a subset $\mathcal{A}$ is order bounded from above (below) if and only if $-\mathcal{A}$ is order bounded from below (above). A subset $\mathcal{A}$ is* ***order bounded*** *if $\mathcal{A}$ is both order bounded from above and from below.*

Definition 2.4.3 (Order Convergence). *A sequence of random variables $(X_n) \in \mathcal{A} \subset \mathcal{L}^0$ is* ***order convergent*** *to X, denoted $X_n \stackrel{o}{\to} X$, if there is a sequence (Z_n) decreasing almost everywhere pointwise to 0, i.e., $Z_n \downarrow 0$, such that $|X_n - X| \leq Z_n$ for every $n \in \mathbb{N}$.*

Lemma 2.4.4 (Lemma 8.17 from [1]). *An order bounded sequence $(X_n) \in \mathcal{L}^1$ satisfies $X_n \stackrel{o}{\to} X$ if and only if $X_n \stackrel{a.s.}{\to} X$.*

Proof. Since that (X_n) is order bounded, then there is some Y that dominates each X_n. Hence, $|X_n - X| \leq Y + |X|$.

Assume $X_n \stackrel{\text{a.s.}}{\to} X$, Then, $|X_n - X| \leq Y + |X| \leq 2Y$ almost surely and by dominated convergence, we have $X \in \mathcal{L}^1$, and hence $\sup_{m \geq n} |X_n - X| \in \mathcal{L}^1$ for every $n \in \mathbb{N}$. Putting $Z_n := \sup_{m \geq n} |X_n - X|$ we have $|X_n - X| \leq Z_n \downarrow 0$.

The other direction is rather trivial. Suppose $X_n \stackrel{\text{o}}{\to} X$, so there exists $(Z_n) \downarrow 0$ on the same directed set as (X_n) such that $|X_n - X| \leq Z_n$. Thus $|X_n(\omega) - X(\omega)| \leq Z_n(\omega)$ for every n and $\omega \in E$ such that $\mathbb{P}(E) = 1$. That is $X_n \stackrel{\text{a.s.}}{\to} X$. ∎

The reader may be interested to know that the following version of lower semicontinuity is equivalent to the *Fatou property* of risk measures [8, 5, 17, 18].

Definition 2.4.5 (Order Continuity). *A function $\rho : \mathcal{L}^0 \to \overline{R}$ is* ***order continuous*** *if $X_n \stackrel{o}{\to} X$ implies $\rho(X_n) \to \rho(X)$. Similarly, ρ is* ***order lower semicontinuous*** *if $X_n \stackrel{o}{\to} X$ implies $\rho(X) \leq \liminf \rho(X_n)$.*

2.5 Skorohod's Representation Theorem

In general, the risk manager does not know the probability structure of their random variable to interest. Hence, they must first construct their empirical distribution as be assured that such a distribution is a good enough approximation of the true distribution.

Once the risk manager has a method for approximating the true distribution, they can create their random variables of interest thusly,

Proposition 2.5.1. *Let F be a distribution function. Then there exists a unique probability measure $\mathbb{Q}$ on $(\mathbb{R}, \mathcal{B})$ such that $\mathbb{Q}((-\infty, x]) = F(x)$. Moreover, a random variable X with distribution function F can be constructed as follows: Let $\Omega = (0, 1)$ and let $\mathbb{P}$ be the uniform distribution on Ω and define*

$$X(\omega) = \inf\{x : F(x) \geq \omega\}, \quad 0 < \omega < 1.$$

The inverse in the above proposition is the *quantile function* and is used in the calculation of *Value at Risk (VaR)* and *Expected Shortfall (ES)*, wherein the former is the loss in the top α-th level of the empirical distribution and the latter is the expectation given losses are above α-th level from the empirical distribution.

Before we continue the proof of the above proposition, we make some observations and state a theorem.

Observation 2.5.2. *Some things to note from our generalized inverse $X(\omega) = \inf\{x : F(x) \geq \omega\}, 0 < \omega < 1$, if F is continuous, then the following are true:*

1. *$X(\omega)$ is strictly increasing,*

   Proof. *Toward contradiction, suppose that $X(\omega)$ is not strictly increasing. Then, for some $\alpha < \beta$, we have $X(\alpha) = X(\beta) = x \in R$. By definition of X, we have $F(x - \frac{1}{n}) \leq \alpha < \beta \leq F(x - \frac{1}{n})$ for any $n \in \mathbb{N}$. By taking limits over n, we have $F(x) < F(x)$, a contradiction.* ■

2. *$F(X(\omega)) \geq \omega$ for every $\omega \in [0, 1]$.*

   Proof. *Assume $X(\omega) \in \mathbb{R}$. We can then find some sequence (x_n) such that $x_n \geq F$ and $x_n \to X(\omega)$. By taking limits and using the right continuity of F we get that $F(X(\omega)) \geq \omega$.* ■

We also note the following well known interplay between $F, X(\omega)$ and $\mathcal{U}(0, 1)$ [37].

Theorem 2.5.3. *Let F be a cumulative distribution and $\mathcal{U}(0,1)$ the standard uniform distribution. Then,*

1. $\mathbb{P}(X(F(X)) = X) = 1$ *if X has distribution F,*
2. *If U has distribution $\mathcal{U}(0,1)$, then $X(U)$ has distribution F,*
3. *If F is continuous and X has distribution F, then $F(X)$ has distribution $\mathcal{U}(0,1)$.*

We now return to the initial Proposition 2.5.1.

Proof. We state the proof directly from [16]. From Lemma A.1.3, notice that X is a measurable function. Let $\mathbb{Q}$ be the probability function of X. We want to show $F(y) = \mathbb{P}(\{\omega : X(\omega) \leq y\})$ is the distribution function of $\mathbb{Q}$. Further notice that X is an increasing function on $(0,1)$ and the event $\{\omega : X(\omega) \leq y\}$ is bounded below by 0 and bounded above by $\sup\{\omega : X(\omega) \leq y\}$.

By the definition of X and right continuity of F, we have $F(X(\omega)) \geq \omega$. Hence for $\omega \in \{\omega : X(\omega) \leq y\}$ we have $F(y) \geq F(X(\omega)) \geq \omega$. Hence, $F(y)$ is an upper bound for $\{\omega : X(\omega) \leq y\}$.

On the other hand, $F(y) \in \{\omega : X(\omega) \leq y\}$, since $X(F(y)) \leq y$. Hence, $F(y) = \sup\{\omega : X(\omega) \leq y\}$. ■

From this empirical distribution, the risk manager can create a sequence of random variables and their respective probability measures from their empirical distribution. A risk manager would surely want to know the asymptotic properties these random variables. In particular, for large enough sample size, the risk manager wants to know if the random variable induced by the empirical distribution (X_n) is a good enough estimator for the random variable of interest (X). The *Skorokhod's Representation* shows that such a procedure produces a consistent estimator. Thus, we give a version of Skorokhod's representation,

Proposition 2.5.4 (Skorokhod's Representation)**.** *Let $\mathbb{Q}$ and a sequence $(\mathbb{Q}_n)$ be probability measures on $\mathbb{R}$ and suppose that $\mathbb{Q}_n \to \mathbb{Q}$ as $n \to \infty$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{R}$-valued random variables X and a sequence (X_n) defined on Ω, such that the distributions of X and X_n are $\mathbb{Q}$ and $\mathbb{Q}_n$, respectively, and $X_n \to X$ almost everywhere as $n \to \infty$.*

Proof. We state the proof for the above directly from [16]. Let F and F_n be the distribution functions corresponding to $\mathbb{Q}$ and $\mathbb{Q}_n$, respectively. Let $\Omega = (0,1)$, $\mathcal{F}$ its Borel field and $\mathbb{P}$ as the Lebesgue measure on $(\Omega, \mathcal{F})$. For $\omega \in \Omega$, set

$$X(\omega) = \inf \{x : F(x) \geq \omega\},$$
$$X_n(\omega) = \inf \{x : F_n(x) \geq \omega\}$$

Fix ω and $\delta > 0$ such that $X(\omega) - \delta$ is a point at which F is continuous. Then,

$$F(X(\omega) - \delta) < \omega \text{ and hence, } F_n(X(\omega) - \delta) < \omega,$$

for large enough n. Furthermore, for such n, we have $X_n(\omega) \geq X(\omega) - \delta$. Thus,

$$\liminf_{n\to\infty} X_n(\omega) \geq X(\omega) - \delta.$$

Now, let $\delta \downarrow 0$ to conclude that

$$\liminf_{n\to\infty} X_n(\omega) \geq X(\omega)$$

Notice that X is monotone and has countably many points of discontinuity. Thus for almost every ω, X is continuous. Fix ω such that $X(\omega)$ is continuous, fix $\epsilon > 0$ and fix $\delta > 0$ such that $X(\omega + \epsilon) + \delta$ is a point at which F is continuous. Then,

$$F(X(\omega + \epsilon) + \delta) \geq \omega + \epsilon.$$

For every large enough n, we have

$$F_n(X(\omega + \epsilon) + \delta) \geq \omega,$$

hence,

$$X(\omega) \leq X(\omega + \epsilon) + \delta.$$

Thus,

$$\limsup_{n\to\infty} X_n(\omega) \leq X(\omega + \epsilon) + \delta.$$

Now, let $\delta \downarrow 0$ and $\epsilon \downarrow 0$ to conclude that that

$$\limsup_{n\to\infty} X_n(\omega) \leq X(\omega).$$

That is, $X_n \to X$ almost surely. ■

Before we continue, let us re-evaluate what the above theorem allows us to do. It allows one to replace a sequence of random variables by a new sequence of random variables defined by the left continuous inverse of the empirical distributions from realizations the previous random variables. Hence, when using expectation as a risk measure, we can approximate the true risk statistic by taking many realizations of the risk measure.

However, our focus is on convex risk. We are left with the question, can we use the same approach when our risk measure is a convex functional? To get to this answer, we must work on generalization of the usual $\mathcal{L}^p$ spaces and see if we can extend this representation to that space.

Chapter 3

Risk measures on Orlicz spaces

3.1 Orlicz Spaces and the Orlicz Heart

The Orlicz space generalizes the notion of the $\mathcal{L}^p$ space. In particular, we consider an Orlicz function which generalizes $|x|^p$ as follows,

Definition 3.1.1 (Young's Function)**.** *A **Young's function** is a function*

$$\Phi : [0, \infty) \to [0, \infty]$$

satisfying:

1. $\Phi(0) = 0$.
2. Φ *is* left-continuous*:* $\lim_{u\uparrow x} \Phi(u) = \Phi(x)$.
3. Φ *is* increasing*: if* $u_1 \leq u_2$ *then* $\Phi(u_1) \leq \Phi(u_2)$.
4. Φ *is* convex*:* $\Phi(au_1 + (1-a)u_2) \leq a\Phi(u_1) + \Phi((1-a)u_2)$ *for* $0 \leq a \leq 1$.
5. Φ *is nontrivial:* $\Phi(u) > 0$ *for some* $u > 0$ *and* $\Phi(u) < \infty$ *for some* $u > 0$.

We remark that $\Phi(x)$ has only one discontinuity (the point at which it jumps to ∞) and if X is a random variable (i.e., a measurable function $X : \mathbb{R} \to \overline{\mathbb{R}}$), then $\Phi(X) : \mathbb{R} \to \overline{\mathbb{R}}^+$ is measurable for $(\Omega, \mathcal{F})$ by Observations A.2.4 and A.3.2.

We some useful facts about Young functions are as follows,

Fact 3.1.2.

1. *Φ achieves its minimum at* 0*,*

2. *Since Φ is increasing, it is differential almost everywhere.*

3. *Φ is* finite *if its effective domain, dom $(\Phi) := \{x \in \mathbb{R}^+$ such that $\Phi(x) < \infty\}$, is $\mathbb{R}^+$.*

4. *The Young's function can equivalently be written as*

$$\Phi(x) = \int_0^{|x|} \phi(x)dx$$

where ϕ is a positive right continuous nondecreasing function such that $\phi(0) = 0$ and $\lim_{x\to\infty} \phi(x) = \infty$*.*

Example 3.1.3 (Young's Functions)**.** *Some useful Young's functions on $[0, \infty]$ are:*

1. $\exp(x)$*,*

2. *$|x|^p$ for $p \geq 1$, and $-|x|^p$ for $0 \leq p < 1$,*

3. $e^{|x|} - |x| - 1$

4. $-\ln x$*,*

5. *$e^{|x|^\delta} - 1$ for $\delta > 1$. In particular, the derivative of this function has property (4) in Fact 3.1.2 [31] .*

Definition 3.1.4 (Orlicz Modular)**.** *On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the functional $\rho : \mathcal{L}^0 \to [0, \infty]$ is the **Orlicz Modular** for Young function Φ defined by*

$$M_\Phi(X) = \mathbb{E}\left[\Phi(|X|)\right] := \int_\Omega \Phi(X(\omega))\mathbb{P}(d\omega)$$

The *Young's class* of Φ is the set $\mathcal{L}^\Phi_M := \mathcal{L}^\Phi_M(\Omega, \mathcal{F}, \mathbb{P})$ consisting of all random variables such that $M_\Phi(X)$ is finite [21, 12].

Proposition 3.1.5. *The Orlicz modular is convex.*

Proof. Indeed for $0 \le \alpha \le 1$ we have,

$$\begin{aligned} M_\Phi(\alpha X + (1-\alpha)Y) &= \mathbb{E}[\Phi(|\alpha X + (1-\alpha)Y|)] \le \mathbb{E}[\Phi(\alpha |X| + (1-\alpha)|Y|)] \\ &\le \alpha \mathbb{E}[\Phi(|X|)] + (1-\alpha)\mathbb{E}[\Phi(|Y|)] = \alpha M_\Phi(X) + (1-\alpha) M_\Phi(Y) \end{aligned}$$

■

Definition 3.1.6 (Orlicz Spaces, The Orlicz Heart and the Luxemburg Norm)**.** *For $\lambda > 0$, define $\mathcal{L}^\Phi_\lambda(\Omega, \mathcal{F}, \mu) := \left\{X : M_\Phi\left(\frac{X}{\lambda}\right) < \infty\right\}$ for random variable X.*

1. ***Orlicz Spaces*** *are given by*

$$\mathcal{L}^\Phi = \mathcal{L}^\Phi(\Omega, \mathcal{F}, \mu) := \bigcup_{\lambda > 0} L^\Phi_\lambda(\Omega, \mathcal{F}, \mu)$$

2. *The* ***Orlicz Heart*** *is given by*

$$\mathcal{H}^\Phi = \mathcal{H}^\Phi(\Omega, \mathcal{F}, \mu) := \bigcap_{\lambda > 0} L^\Phi_\lambda(\Omega, \mathcal{F}, \mu)$$

   That is, $\mathcal{H}^\Phi$ is the set of finite elements of $\mathcal{L}^\Phi$.

3. *The* ***Luxembourg Norm*** *for a random variable X is $\|X\|_\Phi := \inf\left\{\lambda > 0 : M_\Phi\left(\left|\frac{X}{\lambda}\right|\right) \le 1\right\}$.*

4. *Both spaces, $(\mathcal{L}^\Phi, \|\cdot\|_\Phi)$ and $(\mathcal{H}^\Phi, \|\cdot\|_\Phi)$ are Banach Spaces.*

That is, *Orlicz spaces* are the space of random variables such that $M_\phi\left(\frac{X}{\lambda}\right)$ is finite for *some* $\lambda > 0$, and the *heart* of the Orlicz space is the space of random variables such that $M_\phi\left(\frac{X}{\lambda}\right)$ is finite for *every* $\lambda > 0$. The obvious inclusions are $\mathcal{H}^\Phi \subset \mathcal{L}^\Phi_M \subset \mathcal{L}^\Phi$.

We have the further inclusion $\mathcal{L}^\Phi \subset \mathcal{L}^1$. Indeed, for any convex function Φ, we have the property $\Phi(x) \ge \Phi(0) + \Phi'(0)x$. That is, $\Phi(x)$ sits above any of its tangent lines. Then for any $X \in \mathcal{L}^\Phi$ and the linearity and monotonicity of the expectation, we have,

$$\Phi(0) + \Phi'(0)\mathbb{E}[|X|] \le \mathbb{E}[\Phi(|X|)] < \infty.$$

We further note that $\mathcal{L}^\Phi = \mathcal{H}^\Phi$ if and only if Φ holds the Δ_2 condition:

there exist $C, x_0 > 0$ such that $\Phi(2x) \le C\Phi(x)$ for all $x \ge x_0$.

Indeed, the Δ_2 condition is satisfied for $\Phi(x) = |x|^p$ for $p \in [1, \infty)$, which corresponds to the familiar $\mathcal{L}^p$ space. Although $p = \infty$ is not technically an Orlicz space, one can think of $\mathcal{L}^\infty$ as the limiting case of $|x|^p$.

Returning to our discussion of Young's functions, with each function Φ as in Definition 3.1.1, there is another associated convex function Ψ, called the *complementary function*, having similar properties which is defined by

$$\Psi(y) := \sup\{x\,|y| - \Phi(x) : x \geq 0\}, \quad y \in \mathbb{R}.$$

Indeed, from the above, we have

1. $\Psi(0) = 0$,
2. Ψ is left continuous,
3. Ψ is increasing,
4. Ψ is convex,

Furthermore, the pair (Φ, Ψ) satisfies *Young's Inequality*: $xy \leq \Phi(x) + \Psi(y)$ for any $x, y \in \mathbb{R}$.

Proposition 3.1.7. *Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For complementary pair of Young's functions (Φ, Ψ) and their corresponding Orlicz spaces $(\mathcal{L}^\Phi, \mathcal{L}^\Psi)$, if X is a random variable such that $XY \in L^1$ for every $Y \in \mathcal{L}^\Psi$, then $X \in \mathcal{L}^\Phi$.*

We end our discussion of the Orlicz space and Orlicz heart with the following useful proposition from [12]:

Proposition 3.1.8 (2.1.10 Stopping Times and Directed Processes)**.**

1. $\|X_n - X\|_\Phi \to 0$ *if and only if* $M_\Phi\left(k(X - X_n)\right) \to 0$ *any* $k > 0$
2. $\|X_n - X\|_\Phi \to 0$*, then* $X_n \xrightarrow{\mathbb{P}} X$*.*

The proof Proposition 3.1.8 can be found in [12]. We will freely use these properties.

Finally, we note the extension of Lemma 2.4.4 to the Orlicz space as found in [28]:

Theorem 3.1.9. *For a sequence $(X_n) \in \mathcal{L}^\Phi$, $X_n \xrightarrow{o} X$ if and only if X_n is order bounded and $X_n \xrightarrow{a.s.} X$.*

3.2 Convex Risk Measures, The Lebesgue Property and Statistically Consistency

3.2.1 The Basics

In the following $\mathcal{X}$ denotes a fixed subspace of $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ unless otherwise stated.

Definition 3.2.1 (Risk Measures)**.** *The mapping $\rho : \mathcal{X} \mapsto (-\infty, \infty]$ a monetary risk measure on $\mathcal{X}$ if it has the following three properties:*

***(F) Finiteness at** 0: $\rho(0) \in \mathbb{R}$*

***(M) Monotonicity**: $\rho(X) \geq \rho(Y)$ for all $X, Y \in \mathcal{X}$ such that $X \leq Y$*

***(T) Translation property**: $\rho(X + m) = \rho(X) - m$ for all $X \in \mathcal{X}$ and $m \in \mathbb{R}$*

We call a monetary risk measure convex if it also satisfies

***(C) Convexity**: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for all $X, Y \in \mathcal{X}$ and $\lambda \in (0, 1)$*

The description for each property is as follows:

(F) This is clear, as the risk of holding nothing is nothing.

(M) The value $\rho(X)$ is understood to be the capital requirement of holding risky position X. In particular, we have that the capital requirement for X should be greater than Y if it is clear that the payoff of position X is smaller than risky position Y almost surely.

(T) The capital requirement for the convex combination of two risky positions is bounded above the the convex combination of the separate capital requirements.

(C) Adding an amount of money to a risky position reduces the capital requirement by that amount of money.

A convex monetary risk measure is called coherent if it further fulfills:

***(P) Positive homogeneity**: $\rho(\lambda X) = \lambda\rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \in \mathbb{R}_+$*

If (P) holds, then the capital requirements scale linearly when net risky positions are multiplied with non-negative constants and (C) is equivalent to

(S) Subadditivity: $\rho(X+Y) \leq \rho(X) + \rho(Y)$ *for all* $X, Y \in \mathcal{X}$.

Definition 3.2.2. *The acceptance set of a risk measure with properties (F), (M) and (T)* $\rho : \mathcal{X} \to (-\infty, \infty]$ *is*

$$\mathcal{C} := \{X \in \mathcal{X} : \rho(X) \leq 0\}$$

Furthermore, we can represent the previously defined risk measure as

$$\rho(X) = \inf\{m \in \mathbb{R} : \rho(X+m) \leq 0\}$$

That is, ρ *is the minimum capital requirement to "move" a risky position into the acceptance set.*

Definition 3.2.3. *We say that a map* $\rho : \mathcal{X} \mapsto (-\infty, \infty]$ *is law invariant if* $\rho(X) = \rho(Y)$ *whenever* $X, Y \in \mathcal{X}$ *have the same probability law. That is* $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ *for any* $B \in \mathcal{B}(\overline{\mathbb{R}})$.

We now define the *Lebesgue Property* for a functional ρ and discuss some of its properties.

Definition 3.2.4 (Lebesgue Property)**.** *A functional* $\rho : \mathcal{X} \to (-\infty, \infty]$ *has the* ***Lebesgue property*** *if*

$$\rho(X_n) \to \rho(X)$$

Whenever $(X_n) \in \mathcal{X}$ *such that* $X_n \xrightarrow{a.s} X$ *and* $|X_n| \leq Z$ *for some* $Z \in \mathcal{X}$.

In proving the next important result, we shall rely on the following version of a lemma found in [1] and stated in its entirety in A.3.6.

Lemma 3.2.5. *Let* $\rho : \mathcal{X} \to \mathbb{R}$ *be convex and positively homogeneous, then if* ρ *is continuous at* 0 *it is continuous everywhere.*

Proof. Let $X_n \overset{\|\cdot\|}{\to} X$, then $X_n - X \overset{\|\cdot\|}{\to} 0 \Rightarrow \rho(X_n - X) \to 0$. We want to show that $\rho(X_n) \to \rho(X)$. By homogeneity and convexity of ρ (notice that a homogeneous function is convex if and only if it is subadditive) we have,

$$\rho(X_n) = \rho(X_n - X + X) \leq \rho(X_n - X) + \rho(X)$$

and similarly,

$$-\rho(X) = -\rho(X - X_n + X_n) \geq \rho(X_n - X) - \rho(X_n).$$

And hence, $|\rho(X_n) - \rho(X)| \leq \rho(X_n - X)$. Taking $n \to \infty$, we have,

$$\lim_{n\to\infty} \rho(X_n) = \rho(X).$$

■

Theorem 3.2.6. *Any convex monotone increasing functional $\rho : \mathcal{X} \to \mathbb{R}$ on a Banach lattice $\mathcal{X}$ is continuous.*

Proof. In view of Theorem A.3.6 and Lemma 3.2.5, it suffices to show that ρ is continuous at 0.

Pick a sequence in $\mathcal{X}$, say (X_n) such that $X_n \overset{\|\cdot\|}{\to} 0$ and suppose $\rho(0) = 0$ (otherwise, take a linear transform to set $\tilde{\rho} := \rho - \rho(0)$). Toward a contradiction, take a subsequence $(X_{n_k})_{k\in\mathbb{N}}$ such that $\rho(X_{n_k})$ that is not convergent to 0, that is $\rho(X_{n_k}) \geq \varepsilon$ for any $k \in \mathbb{N}$.

Now, by supposition, we have $\|X_{n_k}\| \leq \frac{1}{k2^k}$ for any $k \in \mathbb{N}$.

Putting $Y_m := \sum_{i=1}^m i\,|X_{n_i}|$, we that that,

$$Y_m \uparrow Y := \sum_{i=1}^{\infty} i\,|X_{n_i}| \in \mathcal{X},$$

since $\mathcal{X}$ is a Banach space and $\sum_{i=1}^m \|i\,|X_{n_i}|\| \leq \sum_{i=1}^m \frac{1}{2^i} \in \mathbb{R}$.

Finally, by monotonicity and convexity of ρ, we have

$$\begin{aligned}
|\rho(X_{n_k})| &\leq \rho(|X_{n_k}|) \\
&\leq \rho\left(\frac{1}{k}k\,|X_{n_k}| + \left(1 - \frac{1}{k}\right)0\right) \\
&\leq \frac{1}{k}\rho(k\,|X_{n_k}|) + \left(1 - \frac{1}{k}\right)\rho(0) \leq \frac{1}{k}\rho(Y_m) \leq \frac{1}{k}\rho(Y) \in \mathbb{R}
\end{aligned}$$

Taking limits over $k \to \infty$, we have $\rho(X_{n_k}) \to 0$, a contradiction. ■

Corollary 3.2.7. *Any convex monotone increasing functional $\rho : \mathcal{H}^\Phi \to \mathbb{R}$ has the Lebesgue property.*

Proof. Let $(X_n) \subseteq \mathcal{H}^\Phi$ and $X \in \mathcal{X}$ such that $X_n \xrightarrow{o} X$. Then there exists $Y \in \mathcal{H}^\Phi$ such that $|X_n| \leq Y$. Fix some $k > 0$, then $M_\Phi(kY) < \infty$ and hence by the dominated convergence theorem it follows that,

$$M_\Phi(2kX_n) \to M_\Phi(2kX).$$

Since Φ is convex and nondecreasing, we have,

$$0 \leq M_\Phi(k(X_n - X)) \leq \frac{1}{2}\Big(M_\Phi(2kX_n) + M_\Phi(2kX)\Big).$$

Thus is easy to see that $\Phi(k(X_n - X))$ is uniformly integrable. By the continuity of Φ we have $\Phi(k(X_n - X)) \xrightarrow{a.s.} 0$. Therefore Vitali's convergnce theorem ensures that $M_\Phi(k(X_n - X)) \to 0$. By Proposition 3.1.8 it follows that $X_n \xrightarrow{\|\cdot\|} X$. In view of Theorem 3.2.6 we get that $\rho(X_n) \to \rho(X)$ and thus ρ has the Lebesgue property. ■

3.2.2 Statistically Consistency

In this section we closely follow the notations from [19].

In the following $\rho : \mathcal{L}^\Phi \to (-\infty, \infty]$ will denote a law-invariant risk measure. Let us denote by $\mathcal{M}(\mathcal{L}^\Phi) := \{\mathbb{P} \circ X^{-1} \ : \ X \in \mathcal{L}^\Phi\}$ the class of Borel probability measures on $\mathbb{R}$ that arise as the distribution of some $X \in \mathcal{L}^\Phi$. The risk functional $\mathcal{R}_\rho : \mathcal{M}(\mathcal{L}^\Phi) \to (-\infty, \infty]$ associated with ρ is defined via the following formula.

$$\mathcal{R}_\rho(\mathbb{P} \circ X^{-1}) = \rho(X), \ X \in \mathcal{L}^\Phi.$$

Let (X_n) be a sequence of i.i.d. We denote by

$$\widehat{F}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

the empirical distribution of $X_1, ..., X_n$.

Definition 3.2.8. *We say that ρ is statistical consistent if for any sequence (X_n) of i.i.d. with $X_n \sim X \in \mathcal{L}^\Phi$ we have that*

$$\mathcal{R}_\rho(\widehat{F}_n) \xrightarrow{a.s.} \mathcal{R}_\rho(\mathbb{P} \circ X^{-1})$$

Before we prove our main result, we need the following intermediate steps.

Lemma 3.2.9 (A Skorokhod Representation on Orlicz Spaces)**.** *Let (μ_n) be a sequence in $\mathcal{M}(L^\Phi)$ such that $\mu_n \to \mu \in \mathcal{M}(L^\Phi)$ and $\int \Phi(k|x|)d\mu_n \to \int \Phi(k|x|)d\mu < \infty$ for some $k > 0$. Then there exists a sequence (X_n) and X in $\mathcal{L}^\Phi$ such that $X_n \sim \mu_n$, $X \sim \mu$ and*

$$X_n \xrightarrow{a.s.} X; \Phi(k|X_n|) \xrightarrow{\|\cdot\|_{\mathcal{L}^1}} \Phi(k|X|).$$

Proof. By the standard Skorohod representation theorem we may find random variables X_n and X in $\mathcal{L}^\Phi$ such that $X_n \sim \mu_n, X \sim \mu$ and $X_n \xrightarrow{a.s.} X$. In view of the continuity of Φ we get that $\Phi(k|X_n|) \xrightarrow{a.s.} \Phi(k|X|)$. We also have

$$\mathbb{E}[\Phi(k|X_n|)] = \int \Phi(k|x|)d\mu_n \to \int \Phi(k|x|)d\mu = \mathbb{E}[\Phi(k|X|)].$$

Thus by Theorem 2.2.4, we have $\Phi(k|X_n|) \xrightarrow{\|\cdot\|_{\mathcal{L}^1}} \Phi(k|X|)$. ■

Lemma 3.2.10. *Let (X_n) be a sequence $\mathcal{L}^1$ such that $X_n \xrightarrow{||\cdot||_1} X$, then the exists a subsequence (X_{k_n}) that is order bounded in $\mathcal{L}^1$.*

Proof. Recall, for $X_n \in \mathcal{L}^1$ we have $|X_n|$ is finite a.s. Now since $||X_n - X|| \to 0$, by Theorem B.4.1, there exists a subseqence $X_{n_k} \xrightarrow{a.s.} X$. Putting $Z_k := \sup_{m \geq k} |X_{n_k} - X|$, we have $|X_{n_k} - X| \leq Z_k \downarrow 0$, a.s., the proof follows. ■

Now we are at the crux of the matter. We answer the question posed at the end of Section 2.5. To be explicit, we show that a risk manager can use realizations of a convex risk and be assured that the mean of sure a risk will be consistent to the true value of the convex risk.

Theorem 3.2.11. *Any $\rho : \mathcal{L}^\Phi \to (-\infty, \infty]$ with the* Lebesgue property *is statistically consistent.*

Proof. Let (X_n) a sequence of i.i.d. with $X_n \sim X \in \mathcal{L}^\Phi$, $\widehat{F}_n$ be the corresponding empirical distribution and $k_0 > 0$ such that $\mathbb{E}[\Phi(k_0|X|) < \infty$. By the Law of Large numbers and the Glivenko-Cantelli theorem there exists a set Ω_0 with $\mathbb{P}(\Omega_0) = 1$ such that

$$\int \Phi(k_0\,|x|)\widehat{F}_n^\omega(dx) \to \int \Phi(k_0\,|x|)(\mathbb{P} \circ X^{-1})(dx) < \infty \tag{3.1}$$

and

$$\widehat{F}_n^{\omega} \to \mathbb{P} \circ X^{-1} \tag{3.2}$$

for all $\omega \in \Omega_0$. For the following we fix $\omega_0 \in \Omega_0$. We claim that $\mathcal{R}_\rho(\widehat{F}_n^{\omega_0}) \to \mathcal{R}_\rho(\mathbb{P}\circ X^{-1})$. Indeed, suppose that the claim is false, then we can find $\epsilon > 0$ and a subsequence of $\widehat{F}_{k_n}^{\omega_0}$ of $\widehat{F}_{k_n}^{\omega_0}$ such that

$$|\mathcal{R}_\rho(\widehat{F}_{k_n}^{\omega_0}) - \mathcal{R}_\rho(\mathbb{P} \circ X^{-1})| > \epsilon \ \forall\ n \in \mathbb{N} \tag{3.3}$$

By Eq. (3.1), Eq. (3.2) and Lemma Lemma 3.2.9 there exists a sequence X_n in $\mathcal{L}^\Phi$ and $X_0 \in \mathcal{L}^\Phi$ such that

$X_n \sim \widehat{F}_{k_n}^{\omega_0}$, $X_0 \sim \mathbb{P} \circ X^{-1}$ and

$$X_n \xrightarrow{a.s.} X; \Phi(k_0|X_n|) \xrightarrow{\|\cdot\|_{\mathcal{L}^1}} \Phi(k_0|X_0|).$$

In view of Lemma 3.2.10 and by passing to a subsequence of (X_n) we have

$$\mathbb{E}[\sup_n\{\Phi(k_0\,|X_n|)\}] < \infty.$$

By monotonicity and almost sure continuity of Φ, we get that

$$\mathbb{E}[\Phi(\sup_n\{k_0\,|X_n|\})] = \mathbb{E}[\sup_n\{\Phi(k_0\,|X_n|)\}] < \infty.$$

That is $\sup\{X_n\} \in \mathcal{L}^\Phi$. Since $\sup_n\{X_n\} \in \mathcal{L}^\Phi$ and $X_n \xrightarrow{a.s.} X$ by the Lebesgue property of ρ we get that

$$\rho(X_n) = \mathcal{R}_\rho(\widehat{F}_{k_n}^{\omega_0}) \to \rho(X) = \mathcal{R}_\rho(\mathbb{P} \circ X^{-1}).$$

The above contradicts Eq. (3.3) and thus ρ is statistically consistent.

■

Appendix

Appendix A

Measures and Measurability

A.1 Some Basics

Definition A.1.1 (Almost Everywhere/Almost Sure)**.** *Fix* $(\Omega, \mathcal{F}, \mu)$ *and consider* $\mathfrak{P}$ *as some property. Property* $\mathfrak{P}$ *is said to hold almost everywhere (a.e) / almost surely (a.s.) if*

$$\exists E \in \mathcal{F} \text{ such that } \mu(E^c) = 0 \text{ and } \mathfrak{P} \text{ holds for every } \omega \in E.$$

Definition A.1.2 (Borel σ-algebra)**.** *The Borel* σ*-algebra of* $\mathbb{R}$*, denoted* $\mathcal{B}(\mathbb{R})$*, is the* σ*-algebra generated by the open sets in* $\mathbb{R}$*. That is, if* $\mathcal{O}$ *denotes the collection of all open subsets of* $\mathbb{R}$*, then* $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O})$*.*

Lemma A.1.3 (Tests for Measurability)**.** *For* $(\Omega, \mathcal{F}, \mu)$ *and* $f : \Omega \to \overline{\mathbb{R}}$*,* f *is measurable if and only if any one of the following hold:*

(i) $f^{-1}((-\infty, x]) \in \mathcal{F}$ *for any* $x \in \mathbb{R}$*,*

(ii) $f^{-1}((-\infty, x)) \in \mathcal{F}$ *for any* $x \in \mathbb{R}$*,*

(iii) $f^{-1}([x, \infty)) \in \mathcal{F}$ *for any* $x \in \mathbb{R}$*,*

(iv) $f^{-1}((x, \infty)) \in \mathcal{F}$ *for any* $x \in \mathbb{R}$*,*

A.2 Complete Measure and Extended σ-algebras

Definition A.2.1 (Complete Measure Space)**.** *For a σ-algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$ and a σ-additive measure $\mu : \mathcal{F} \to \overline{\mathbb{R}}^{+}$, the triplet $(\Omega, \mathcal{F}, \mu)$ is complete if,*

For $A \in \mathcal{F}$ such that $\mu(A) = 0$ and for any $E \subset A$ we have, $E \in \mathcal{F}$. By monotonicity of measure μ we also have $\mu(E) = 0$.

If a measure space $(\Omega, \mathcal{F}, \mu)$ is not complete, we can find a unique measure $\overline{\mu}$ and a σ-algebra $\overline{\mathcal{F}} \supset \mathcal{F}$ by setting $\overline{\mathcal{F}} = \{A \cup N : A \in \mathcal{F}, N \subset E \in \mathcal{F}, \text{ where } \mu(E) = 0\}$ and by setting $\overline{\mu} = \mu$ for any $A \in \mathcal{F}$.

A complete measure space is useful in the following:

Lemma A.2.2. *Fix $(\Omega, \mathcal{L}, \lambda)$, where $\mathcal{L}$ and λ are the borel σ-algebra, Lebesgue σ-algebra and Lebesgue measure, respectively. Suppose $f : \mathbb{R} \to \overline{\mathbb{R}}$ is $\mathcal{L}$-measurable and $g : \mathbb{R} \to \overline{\mathbb{R}}$ such that $g = f$ a.e. Then g is also $\mathcal{L}$-measurable.*

Proof. Fix $A \in \mathcal{B}$. Since $g = f$ a.e., there exists $E \in \mathcal{L}$ such that $\lambda(E^c) = 0$ and $f(\omega) = g(\omega)$ for every $\omega \in E$. Now,

$$\begin{aligned}\{\{\omega \in \Omega : g(\omega) \in A\} \cap E\} &= \{\{\omega \in \Omega : g(\omega) \in A\} \cap E\} \cup \{\{\omega \in \Omega : g(\omega) \in A\} \cap E^c\} \\ &= \underbrace{\{\omega \in \Omega : f(\omega) \in A\}}_{\in \mathcal{L}} \cup \underbrace{\{\{\omega \in \Omega : g(\omega) \in A\} \cap E^c\}}_{\subset E^c \text{ and } \in \mathcal{L} \text{ since } \lambda \text{ is complete}}\end{aligned}$$

Hence, g is measurable.

■

Definition A.2.3. *Fix $(\Omega, \mathcal{F}, \mu)$ and consider $f : \mathcal{F} \to \overline{\mathbb{R}}$ and the extended borel σ-algebra $\overline{\mathcal{B}} := \{A \cup B : A \in \mathcal{B} \text{ and } B \subset \{-\infty, +\infty\}\}$. The function f is measurable if $f^{-1}(A) \in \mathcal{F}$ for any $A \in \overline{\mathcal{B}}$.*

Observation A.2.4. *For $(\Omega, \mathcal{F}, \mu)$ and $f : \Omega \to \overline{R}$ and $g : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ such that f is $\mathcal{F}$-measurable and g is $\overline{\mathcal{B}}$-measurable. denote $g \circ f : \Omega \to \mathbb{R}$ as the composition of functions, $g(f(x))$ for any $x \in \overline{\mathbb{R}}$. Then $g \circ f$ is measurable.*

Proof. Fix $A \in \overline{\mathcal{B}}$, and notice $g \circ f^{-1}(A) = f^{-1}(g^{-1}(A))$. Since $g^{-1}(A) \in \overline{\mathcal{B}}$ and $f^{-1}(B) \in \mathcal{F}$, we have $f^{-1}(g^{-1}(A)) \in \mathcal{F}$. ■

A.3 Functions

Definition A.3.1 (Continuous function). *A function $f : X \to Y$ between topological spaces is* ***continuous*** *if for every $x \in X$ and every neighbourhood N of $f(x)$ there is a neighbourhood M of x such that $f(M) \subset N$*

Observation A.3.2 (Almost Everywhere Continuous Functions are Measurable). *Suppose Ω is a topological space such that $\mathcal{F} \supset \mathcal{O}$, where $\mathcal{F}$ is a σ-algebra and $\mathcal{O}$ is a topology Ω. For a continuous function $f : \Omega \to \overline{\mathbb{R}}$, we have that if f is almost everywhere continuous, then f is measurable.*

Definition A.3.3 (Inverse Functions). *For every function $f : X \to Y$, we can define an inverse function $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(x)$ by*

$$f^{-1}(A) = \{x \in X : f(x) \in A \text{ and } A \subset Y\}$$

Some properties of this definition are as follows

1. *For every $A \subset Y$, $(f^{-1}(A))^c = f^{-1}(A^c)$;*
2. *If $A, B \subset Y$ are disjoint, so are $f^{-1}(A), f^{-1}(B) \subset X$*
3. *$f^{-1}(Y) = X$,*
4. *If $A_n \subset Y$ is a sequence of subsets, then $f^{-1}(\cup_{n\in N}) = \cup_{n\in N} f^{-1}(A_n)$*

Fact A.3.4. *A function f over two normed linear spaces $\mathcal{X}$ and $\mathcal{Y}$ is continuous if either of the following hold:*

1. *if for each $(u_n)_{n\in\mathbb{N}} \in \mathcal{X}$,*

$$\text{If } \lim_{n\to\infty} u_n = u \text{ then } \lim_{n\to\infty} f(u_n) = f(u),$$

2. *or, if for each $u \in \mathcal{X}$ and every $\epsilon > 0$, there exists a number $\delta(\epsilon, u) > 0$ such that*

$$\text{If } \|v - u\| < \delta(\epsilon, u) \text{ and } v \in \mathcal{X} \text{ then } \|f(v) - f(u)\| < \epsilon.$$

Definition A.3.5 (Homogeneous, Linear, Affine and Convex Functions). *For function $f : \mathcal{X} \to \mathcal{Y}$ between two spaces:*

*f is **homogeneous** if*

$$f(\lambda x) = \lambda f(x) \text{ for every } \lambda \neq 0$$

1. *f is **Linear** if*

$$f(\lambda x + y) = \lambda f(x) + f(y) \text{ for every } \lambda \in \mathbb{R} \text{ and } x, y \in \mathcal{X}.$$

2. *f is **affine** if*

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y) \text{ for every } \lambda \in \mathbb{R} \text{ and } x, y \in \mathcal{X}.$$

3. *f is **convex** if*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for any $\lambda \in [0,1]$ and $x, y \in \mathcal{X}$.

Finally, we end this section with *Theorem 5.43* from [1] in its entirety:

Theorem A.3.6 (Global continuity of convex functions)**.** *For a convex function $f : \mathcal{C} \to \mathbb{R}$ on an open convex subset of a topological vector space, the following statements are equivalent:*

1. *f is continuous on $\mathcal{C}$,*

2. *f is upper semicontinuous on $\mathcal{C}$,*

3. *f is bounded above on a neighborhood of* each *point in $\mathcal{C}$,*

4. *f is bounded above on a neighborhood of* some *point in $\mathcal{C}$,*

5. *f is continuous at some point in $\mathcal{C}$*

Appendix B

Complete Linear Spaces and $\mathcal{L}^p$-Spaces

B.1 Basics

Definition B.1.1 (Linear Space)**.** *$\mathcal{X}$ is **linear space** over a field $\mathbb{K}$, if $\mathcal{X}$ is a set equipped with two operations,* addition(+) *and* scalar multiplication($\cdot$) *defined as follows:*

The operation + has the following properties:

(i) + is commutative,

(ii) + is associative,

(iii) and $\mathcal{X}$ is a group.

The operation $\cdot$ has the following properties:

(i) $\cdot$ is associative,

(ii) $\cdot$ is distributive,

(iii) and there exists a neutral element represented by 1 in the field $\mathbb{K}$

For convenience, we suppress $\cdot$ and simply write αx for $x \in \mathcal{X}$ and $\alpha \in \mathbb{K}$.

Unless stated otherwise, the field $\mathbb{K}$ is assumed to be real numbers $\mathbb{R}$.

Definition B.1.2 (Linear Subspace)**.** *For a linear space $\mathcal{X}$ over a field $\mathbb{K}$, $\mathcal{Y} \subseteq \mathcal{X}$ is a* ***linear subspace*** *(also called a* ***sub-linear space****) if x and y are in $\mathcal{Y}$ and α in $\mathbb{K}$, then $\alpha x + y$ is also in $\mathcal{Y}$.*

Definition B.1.3 (Normed Linear Space)**.** *Consider a linear space represented by $\mathcal{X}$ and a field $\mathbb{K}$ which will be either $\mathbb{R}$ or $\mathbb{C}$, we say that a* ***norm*** *represented by N is a function from $X \mapsto \mathbb{R}_+$ with the following properties:*

1. *$N(x)$ is* nonnegative *and $N(x) = 0 \iff x = 0$ for any x in X (definiteness),*
2. *for all α in K and for any x in X, $N(\alpha x) = |\alpha| N(x)$ (homogeneity),*
3. *for any two elements in X, say x and y, we have $N(x + y) \leq N(x) + N(y)$ (Sub-additivity).*

If $N(x)$ follows all of the above conditions, except perhaps the latter part of (1), then it is a ***semi-norm****.*

A ***normed linear space*** *is a linear space equipped with a norm as defined above. For convenience, we represent $N(x)$ as $\|x\|$.*

Definition B.1.4 (Cauchy Sequence)**.** *For a linear space $\mathcal{X}$ equipped with a norm $\|\cdot\|$, the sequence (x_n) is a* ***Cauchy Sequence****,*

if for any $\epsilon > 0$, there is at least one integer k_0, such that $\|x_j - x_k\| \leq \epsilon$ for any $k, j \geq k_0$.

Definition B.1.5 (Complete Space)**.** *$\mathcal{X}$ is a* ***complete space*** *if* every *Cauchy sequence converges in $\mathcal{X}$.*

Notice that a Cauchy sequence may not converge in a given space. E.g., take sequence $x_n = \frac{1}{n}$ in the interval $(0, 1)$. Then it is clear that x_n is Cauchy, but x_n converges to 0 which is *not* in the given space.

Proposition B.1.6. *In a normed space, each convergent sequence is Cauchy.*

Proof. For a convergent sequence x_n in a normed linear space $\mathcal{X}$, we have $\|x_n - x\| \leq \frac{\epsilon}{2}$ for some $n \geq k_0$.

Then, $\|x_n - x_m\| = \Big\|(x_n - x) + (x - x_m)\Big\| \leq \Big\|x_n - x\Big\| + \Big\|x - x_m\Big\| \leq \epsilon$ for $n, m \geq k_0$. ∎

Theorem B.1.7. *In a normed linear space $\mathcal{X}$, every Cauchy sequence (x_n) is convergent if and only if (x_n) has a convergent subsequence.*

B.2 Equivalence Classes

Definition B.2.1 (Equivalent modulo)**.** *For a linear subspace $\mathcal{Y}$ of linear space $\mathcal{X}$ over a field $\mathbb{K}$, we introduce a relation, $(\sim)$, between two elements of $\mathcal{X}$. Two arbitrary elements in $\mathcal{X}$, say x, y, are* ***equivalent modulo****, denoted $x \sim y$, if $x - y$ is in subspace $\mathcal{Y}$.*

Lemma B.2.2. *The relation $(\sim)$ is an equivalence relation:*

$$x \sim x \quad \textit{(Reflexive)}. \tag{B.1}$$

$$\textit{If } x \sim y, \textit{then } y \sim x \quad \textit{(Symmetric)}. \tag{B.2}$$

$$\textit{If } x \sim y \textit{ and } y \sim z, \textit{then } x \sim z \quad \textit{(Transitive)}. \tag{B.3}$$

Notice we can partition the linear space $\mathcal{X}$ into distinct equivalence classes modulo linear subspace $\mathcal{Y}$. We denote this set of equivalence classes as $\mathcal{X}\mid_{\mathcal{Y}}$ stated "$\mathcal{X}$ mod $\mathcal{Y}$".

Definition B.2.3 (Quotient Space)**.** *For x, y in a linear space $\mathcal{X}$, and a sublinear space $\mathcal{Y}$, we define the following:*

1. *$[x] = \{y \in \mathcal{X} : x \sim y\} := \{y \in \mathcal{X} : (x - y) \in \mathcal{Y}\}$. That is for $x \in \mathcal{X}$, $[x]$ represents all the elements of $y \in \mathcal{X}$ such that the linear combination $(x - y) \in \mathcal{Y}$.*
2. *We represent the* ***Quotient Space*** *of $\mathcal{X}$ modulo $\mathcal{Y}$ as $\mathcal{X}\mid_{\mathcal{Y}} = \{[x] : x \in \mathcal{X}\}$. That is $\mathcal{X}\mid_{\mathcal{Y}}$ is the set of all equivalence classes of elements of $\mathcal{X}$.*

Remark B.2.4. *From Definitions B.2.1 and B.2.3 it is clear that any representative from an equivalence class can be used to denote that equivalence class. That is, for any element $z \in [x]$, $[z] = [x]$. Furthermore, the operations:*

1. *$[x] + [y] = [x + y]$*
2. *$\alpha[x] = [\alpha x]$*

are well defined.

B.3 Using $\mathcal{L}^p$ and Dominated Convergence Theorem

Theorem B.3.1 (Convergence in Measure and Almost Everywhere Convergence)**.** *In a finite measure space $(\Omega, \mathcal{F}, \mu)$, and for a sequence of measurable functions (f_n), $f_n \to f$ in measure if and only if* every *subsequence $(f_{n_k}) \to f$ almost everywhere.*

Proof. Fix $(\Omega, \mathcal{F}, \mu)$ and suppose $(f_n) \to f$ a.e. By definition, we have $\mu(|f_n - f| > \varepsilon) \to 0$ as $n \to \infty$ for any $\varepsilon > 0$. So, we can fix $\varepsilon = \frac{1}{2^k}$ such that $\mu(|f_n - f| > \frac{1}{2^k}) \leq \frac{1}{2^k}$ for $n \geq n_k$.

Now, define $A_n := \left\{x \in \Omega : |f_{n_k}(x) - f(x)| > \frac{1}{2^k}\right\}$. Since $\mu(A_k) \leq \sum_{k=1}^{\infty} \frac{1}{2^k}$ is a convergent series, and Borel-Cantelli tells us that $\mu(\limsup_{k\to\infty} A_k) = 0$, we have $\lim_{k\to\infty} f_{n_k} = f$ almost everywhere. ∎

In fact, one can strengthen the above to almost uniform convergence as in *Egoroff's Theorem.*

Definition B.3.2 ($\mathcal{L}^p$ spaces)**.** *Fix a measure space $(\Omega, \mathcal{F}, \mu)$, for a measurable function f on Ω, and $1 \leq p < \infty$, define $\|f\|_p := \left(\int_\Omega |f|^p \, d\mu\right)^{\frac{1}{p}}$. We define the L^p* ***spaces*** *as follows:*

$$\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathcal{F}, \mu) = \left\{ f : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ and } \|f\|_p < \infty \right\}$$

A more general characterization of Theorem 2.1.17 is as follows,

Theorem B.3.3 ($\mathcal{L}^p$ Dominated Convergence Theorem)**.** *Fix a measure space $(\Omega, \mathcal{F}, \mu)$, put $\mathcal{L}^{p^+}$ as the space of all measurable functions from Ω to $[0, \infty]$ and put $1 \leq p < +\infty$.*

If (f_n) is a sequence in $\mathcal{L}^{p^+}$ such that $f_n \to f$ almost everywhere, and $f_n \leq g$ almost everywhere, for $\int g d\mu < +\infty$ then

$$\int f d\mu < \infty \text{ and } \int f_n d\mu \to \int f d\mu$$

B.4 Riesz-Fisher Theorem and Topologies in $\mathcal{L}^p$

Theorem B.4.1 (Riesz–Fischer Theorem)**.** *Under the $f \sim g$ a.s. equivalence class, $(\mathcal{L}^p, \|\cdot\|_p)$ is a Banach space.*

Theorem B.4.2 (Cauchy Sequences in $\mathcal{L}^p$)**.** *Suppose* $(f_n) \in \mathcal{L}^p$ *is a Cauchy sequence, then there exists* $f \in \mathcal{L}^p$ *such that* $f_n \overset{\|\cdot\|}{\to} f$.

Property B.4.3 (Properties of Integrals)**.** *For* $(\Omega, \mathcal{F}, \mu)$ *and* $f, g : \Omega \to \overline{\mathbb{R}}$ *such that* f, g *are measurable and integrable and* $A, B \in \mathcal{F}$ *such that* $A \cap B = \varnothing$

(1) $|f| < +\infty$ *a.e.*,

(2) *For measurable* $h : \Omega \to \overline{\mathbb{R}}$ *such that* $|h| \leq f$, h *is integrable.*

(3) $\left|\int f d\mu\right| \leq \int |f| \, d\mu$,

(4) *If* $f \geq 0$, *then* $\int f d\mu \geq 0$,

(5) *If* $f = g$ *a.e., then* $\int f d\mu = \int g d\mu$,

(6) *If* $f \geq 0$ *and* $\int f d\mu = 0$, *then* $f = 0$ *a.e.*,

(7) $f \pm g$ *and* cf *are integrable for any* $c \in \mathbb{R}$,

(8) $\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu$,

(9) $\mathbb{1}_A f$ *is integrable and* $\int_A f d\mu = \int \mathbb{1}_A f d\mu$ *for every* $A \in \mathcal{F}$,

(10) *If* $|f| \leq c$ *on* $E \in \mathcal{F}$, $|f| = 0$ *on* E^c *and* $\mu(E) < \infty$, *then* f *is integrable,*

References

[1] C.D. Aliprantis. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2006.

[2] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Thinking coherently - generalised scenatios rather than var should be used when calculating regulatory capital. *Risk : managing risk in the world's financial markets*, 10(11):68–72, 1997.

[3] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[4] G. Bachman and L. Narici. *Functional Analysis*. Dover Publications, Inc., 2000.

[5] S. Biagini and M. Frittelli. On the extension of the namioka-klee theorem and on the fatou property for risk measures. In *Optimality and risk-modern trends in mathematical finance*, pages 1–28. Springer, 2009.

[6] P. Billingsley. *Probability and Measure*. Wiley series in probability and mathematical statistics. Wiley India Pvt. Limited, 2008.

[7] L. Breiman. *Probability*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1968.

[8] S. Chen, N. Gao, and F. Xanthos. The strong fatou property of risk measures. *Dependence Modeling*, 6(1):183–196, 2018.

[9] P. Cheridito and T. Li. Risk measures on orlicz hearts. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(2):189–214, 2009.

[10] R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quantitative finance*, 10(6):593–606, 2010.

[11] F. Delbaen. *Coherent Risk Measures on General Probability Spaces*, pages 1–37. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[12] G.A. Edgar, L. Sucheston, P.E.G. Edgar, G.C. Rota, B. Doran, P. Flajolet, M. Ismail, T.Y. Lam, and E. Lutwak. *Stopping Times and Directed Processes*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1992.

[13] D. Einhorn. Private profits and socialized risk. *GARP Risk Review*, 2008.

[14] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications*. A Wiley-Interscience publication. Wiley, 1999.

[15] H. Föllmer and A. Schied. Convex risk measures. *Encyclopedia of Quantitative Finance*, 2010.

[16] B.E. Fristedt and L.F. Gray. *A modern approach to probability theory*. Springer Science & Business Media, 2013.

[17] N. Gao, D. Leung, C. Munari, and F. Xanthos. Fatou property, representations, and extensions of law-invariant risk measures on general orlicz spaces. *Finance and Stochastics*, 22(2):395–415, 2018.

[18] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the fatou property. In *Advances in mathematical economics*, pages 49–71. Springer, 2006.

[19] V. Kraätschmer, A Schied, , and H. Zähle. Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, 2014.

[20] S. Kusuoka. *On law invariant coherent risk measures*, pages 83–95. Springer Japan, Tokyo, 2001.

[21] C. Labuschagne, H. Ouerdiane, and I. Salhi. Risk measures on orlicz heart spaces. *Communications on Stochastic Analysis*, 9(2):2, 2015.

[22] C. Landim. Doctorate program: Functional analysis (2019). Online Lecture, 2019. Accessed 15/07/2021.

[23] P. Lax. *Functional Analysis*. John Wiley & Sons, 2002.

[24] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. A Series of Modern Surveys in Mathematics Series. Springer, 1991.

[25] D. Li and H. Queffélec. *Introduction to Banach Spaces: Analysis and Probability*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2017.

[26] S. Mazumder. Notes on functional analysis lectures by claudio landim. unpublished, 2021.

[27] A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[28] J. Orihuela and M. Ruiz Galán. Lebesgue property for convex risk measures on orlicz spaces. *Mathematics and Financial Economics*, 6(1):15–35, 2012.

[29] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 2005.

[30] M. M. Rao and Z. D. Ren. *Applications of Orlicz spaces*. CRC Press, 2002.

[31] M. M. Rao and Z. D. Ren. *Theory of Orlicz spaces*. CRC Press, 2002.

[32] T. Tao. *An Epsilon of Room, I: Real Analysis*. An Epsilon of Room. American Mathematical Society, 2010.

[33] Wikipedia contributors. Atom (measure theory) — Wikipedia, the free encyclopedia, 2021. [Online; accessed 5-August-2021].

[34] Wikipedia contributors. Skorokhod's representation theorem — Wikipedia, the free encyclopedia, 2021. [Online; accessed 5-August-2021].

[35] F. Xanthos. AM8001: Analysis and probability. University Lecture, 2019.

[36] F. Xanthos. AM8205: Applied statistical methods. University Lecture, 2019.

[37] F. Xanthos. MTH800: Financial mathematics II. University Lecture, 2019.

[38] E. Zeidler. *Applied Functional Analysis: Applications to Mathematical Physics*. Springer Publishing Company, Incorporated, 2012.