

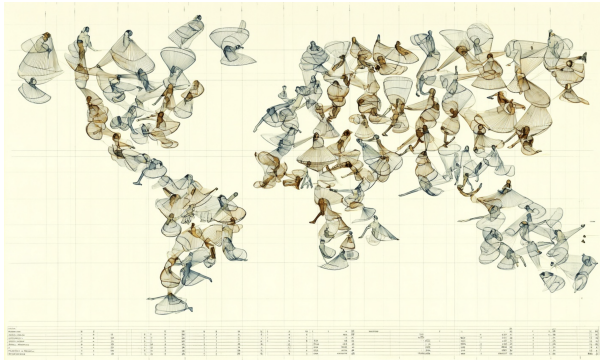
# An Atlas of Monetary Risk Measures

Stephen J. Mildenhall

Wednesday Apr 8, 2026 (20:41:30.000GMT+1)

## Abstract

This article offers a survey of monetary risk measures on  $L^\infty$  guided by a two-dimensional Atlas. The horizontal dimension tracks increasingly flexible portfolio additivity, from linear to comonotonic additive, coherent, convex, and monetary functionals, and the vertical tracks increasingly restrictive continuity assumptions, from the absence of a Fatou assumption to its presence, to law invariance. The Atlas provides a compact framework for locating familiar examples and for understanding how algebraic structure and topological regularity interact through convex duality. The article adopts the actuarial loss sign convention and a pricing-oriented perspective. Alongside the classical theory, it explains how finitely additive and other non-Fatou examples describe what lies beyond the standard countably additive setting and why the usual regularity assumptions matter for risk management.



## Introduction

A famous paper puts “order in risk measures”, that order emerging from characterizing axioms where each axiom scaffolds a tower of mathematical implications [Frittelli and Rosazza Gianin \(2002\)](#). The order is powerful because these implications resonate well with both intuitive notions of risk and the practice of risk management. The scaffold is at once beautiful and useful, but intimidating too as it clads advanced mathematics. Much of that mathematics is taken for granted in academic papers, making the literature hard for the novice. This paper aims to provide just enough mathematical background to make sense of the theory, saving the reader from thousand-page, multi-volume analysis texts.

The presentation is arranged around an Atlas of risk measures, [Figure 1](#). Its horizontal dimension tracks increasingly flexible algebraic notions of portfolio additivity, and the vertical increasingly restrictive notions of continuity. This two-dimensional framework allows us to locate any risk measure at the intersection of its economic behavior (diversification) and its mathematical regularity (limit behavior).

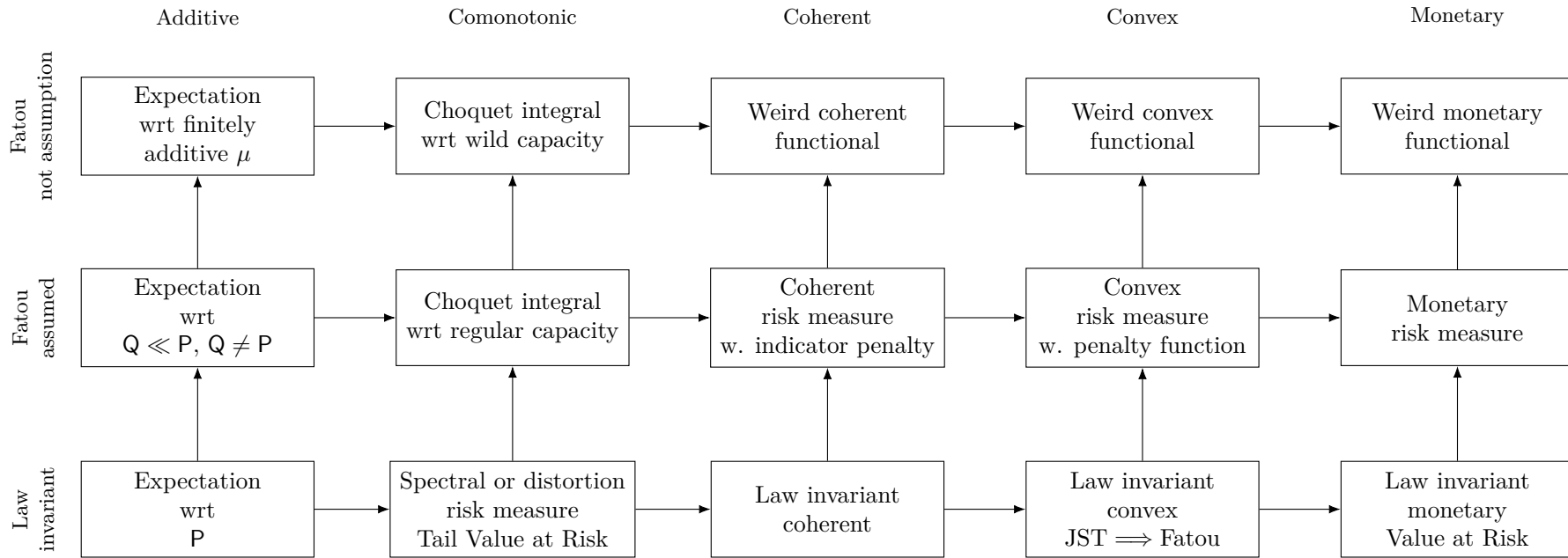


Figure 1: Atlas of risk measures. Continuity increases down each column and diversification additivity becomes more flexible across each row.

All measures in the Atlas are monetary: they share the foundational properties of monotonicity—if option  $B$  is worse than option  $A$  in every outcome, its risk must be higher—and cash invariance, ensuring that risk scales linearly with certain payments. Moving from left to right, we systematically relax the requirement of additivity over portfolios. An additive (linear) measure is the most rigid, where risk is strictly additive and no diversification benefit exists—not a good model for insurance, although additive rules are often used to model prices in efficient financial markets. Comonotonic additive measures relax additivity to hold only for risks that move together. Coherent measures introduce the *sine qua non* of risk theory: sub-additivity, which formalizes the benefit of diversification and requires that the risk of a pool is no greater than the sum of its parts, independent of position size. Convex measures, a further relaxation, acknowledge that diversification benefits may diminish as position sizes grow, often due to liquidity constraints or concentrations of risk. Finally, monetary measures are the most flexible, where only monotonicity and cash invariance remain, the two basic axioms for all risk functionals we consider.

While the columns define how a risk measure treats portfolio diversification, moving down the rows represents an increase in the measure’s mathematical niceness, or continuity. The top row represents measures in their most general, finitely additive form, operating with an absence of an explicit Fatou continuity assumption—defined below. In the middle row, we impose the Fatou property, a continuity assumption that ensures the risk measure behaves predictably under limits and allows for a dual representation using standard countably additive densities. Lastly, the bottom row imposes law invariance, requiring that the risk measure depends only on the distribution of a random variable, not on the underlying state space. In the convex setting, law invariance has powerful consequences, including the Fatou property. It represents the highest degree of regularity in the Atlas and forces the risk measure to respect several intuitive notions of risk.

An important mathematical idea unifying the Atlas is that of a convex function. Convexity is a mathematical way to express the benefits of diversification and pooling. In mathematics, convex functions are regarded as highly tractable, almost on par with continuous and differentiable functions. They often possess a dual representation based on affine (shifted linear) functions, making them easy to understand, interpret, and use in theory and practice. A central objective of the theory is describing exactly when this dual representation holds. We shall see that there is a subtle interplay between the measure’s algebraic flexibility and its topological continuity that emerges precisely through this dual representation.

Once you set up an axiomatic framework, mathematics tends to push back—“Problems worthy of attack prove their worth by hitting back,” as Piet Hein says. Things you wish true turn out false, while unexpected examples exhibit inconvenient truths. Mathematics pushes back particularly strongly in risk theory, so strongly that we can formulate a spoof rule: “No universal statement about risk is true.” The spoof rule certainly feels true, even though it is logically inconsistent. As we explore the Atlas we will uncover many unwanted implications. One is an interplay between the axioms. The monetary assumption already enforces a lot of continuity. Thus, in a sense convex duality is already lurking in the left four columns—the right-most lacks convexity. But continuity depends on topology, a notion of nearness, and that can be defined in many ways. The difference between the top two rows comes down to different topologies, and it is quite subtle. It leads to weird examples that are usually assumed away. We include them here to show *why* they are assumed away. Once you meet them, you will know.

This article has a different aim from comprehensive treatments such as the excellent book Föllmer and Schied (2016). The Atlas is intended as a tour guide rather than a comprehensive treatment. We include more background theory, and cover only a subset of the material in Föllmer and Schied Chapter 4, “Monetary measures of risk.” We use the actuarial loss sign convention throughout, and take a pricing-oriented viewpoint rather than organizing the subject around acceptance sets or capital requirements. We also pay more attention to the edge of the map: not only the classical convex, coherent, and law-invariant cases, but also the stranger objects that lie beyond them, and the assumptions that rule them out. Finally, we make no explicit mention of utility and preference theory (their Chapter 2), instead keeping the focus on risk functionals useful in insurance and the relationships among them.

We are now a generation removed from the coherent revolution initiated by Artzner et al. (1999). The excitement of the early papers, and the rapid development of convex duality methods for risk measures in the early 2000s, are still palpable when reading that literature today. Yet the subject has now become part of the foundation of the field, and every new researcher must come to grips with it. That is not an easy task. This survey article aims to help the next generation. It focuses on common points of difficulty or confusion (at least in the author’s experience), and is offered in the hope that it will be useful to students and researchers as they learn the theory.

With that introduction and motivation, we turn to the contents. There are seven sections. First, we define the axioms and terms that we use to signpost the Atlas, Section 1. Then, Section 2 lays out some basic results that follow easily from them. Section 3 explains the fundamental concept of convex duality. We then state and sketch, in Section 4, four deeper results in the theory: Schmeidler’s characterization of comonotonic additive functionals (second column), two implications of law invariance (bottom row), that it implies the Fatou property and that it respects second-order stochastic dominance, and finally Kusuoka’s representation of law-invariant convex risk measures, which explains why the law-invariant convex region is generated from Tail Value at Risk and clarifies the special role of distortion measures. Section 5 lays out examples in each region of the Atlas, Section 6 offers thoughts on three miscellaneous topics: risk vs. value measures, acceptance sets, and pricing vs. regulation. Section 7 provides with a very brief literature review. Section 8 covers deeper background material and proofs of two technical results. Finally, **@sec-atlas-redux** reprises everything we have covered.

## 1 Definitions

We work on  $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ , the space of essentially bounded random variables modulo almost sure equality. We use the **loss sign convention**: larger values of  $X$  are worse.

A **risk functional** is a map

$$\rho : L^\infty \rightarrow \mathbb{R}.$$

For  $X, Y \in L^\infty$  and  $m \in \mathbb{R}$ , we make the following definitions.

- **Monotone**:  $X \leq Y$  a.s. implies  $\rho(X) \leq \rho(Y)$ .
- **Cash invariant**:  $\rho(X + m) = \rho(X) + m$ .
- **Monetary**: monotone and cash invariant.
- **Normalized**:  $\rho(0) = 0$ .
- **Convex**:  $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$ , for  $0 \leq \lambda \leq 1$ .
- **Positively homogeneous**:  $\rho(\lambda X) = \lambda\rho(X)$ , for  $\lambda \geq 0$ . Positively homogeneous implies normalized since  $\rho(0) = \rho(0 \cdot 0) = 0\rho(0) = 0$ .
- **Subadditive**:  $\rho(X + Y) \leq \rho(X) + \rho(Y)$ .
- **Convex risk measure**: monetary and convex.
- **Coherent risk measure**: monetary, convex, and positively homogeneous; equivalently, monetary, subadditive, and positively homogeneous by the 2/3 lemma below.
- **Comonotone**:  $X$  and  $Y$  are comonotone if

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0$$

for all  $\omega, \omega'$  after choosing representatives. Equivalently,  $X = f(Z)$  and  $Y = g(Z)$  a.s. for some random variable  $Z$  and some increasing functions  $f, g$ , Proposition 2.2. In fact, we can take  $Z = X + Y$ .

- **Comonotonic additive**: if  $X, Y$  are comonotone then

$$\rho(X + Y) = \rho(X) + \rho(Y).$$

- **Law invariant**: if  $X$  and  $Y$  have the same distribution ( $X \stackrel{d}{=} Y$ ) then  $\rho(X) = \rho(Y)$ .
- **Fatou property**: if  $X_n \in L^\infty$ ,  $\sup_n \|X_n\|_\infty < \infty$ , and  $X_n \rightarrow X$  a.s., then

$$\rho(X) \leq \liminf_{n \rightarrow \infty} \rho(X_n).$$

The Fatou property guarantees that the limit position of a sequence of acceptable positions remains acceptable because risk does not jump up unexpectedly.

*Remark 1.1.* Cash invariance is one of the most useful, and most revealing, axioms in risk theory. It looks innocent and seems reasonable, but it has major consequences. It says that risk is measured on a linear cash scale: a certain dollar is a certain dollar, no matter what other wealth is present. The axiom therefore separates risk from wealth. That separation is mathematically powerful and is one of the reasons the theory of monetary risk measures is so tractable. The existence of risk-free cash is also implicit.

The cash invariance axiom marks a genuine fork in the road: it is not the way expected utility behaves. Under diminishing marginal utility, a dollar matters more when wealth is low than when it is high. Cash invariance rules that out. For a commercial enterprise, or a trading desk, one often wants to measure the risk of a position itself, not the welfare of the ultimate owners after combining the position with all their outside wealth, and we should separate risk from wealth in that setting.

## 2 Basic Results

*Remark 2.1.* Under the loss sign convention, positions  $X$  with  $\rho(X) \leq 0$  are desirable or **acceptable** in risk management terminology. Cash invariance implies  $\rho(X - \rho(X)) = \rho(X) - \rho(X) = 0$  and so any risk  $X$  becomes acceptable when combined with a payment of  $\rho(X)$ ; under the loss sign convention, receiving premium reduces loss and therefore enters with a minus sign. This algebra is the rationale for interpreting  $\rho(X)$  as the premium needed to write  $X$ .

*Remark 2.2.* Cash invariance implies

$$\rho(m) = \rho(0 + m) = \rho(0) + m, \quad \forall m \in \mathbb{R},$$

and therefore if  $\rho$  is normalized

$$\rho(m) = m \quad \text{for all } m \in \mathbb{R},$$

consistent with the idea that cash is riskless. For a normalized risk measure, cash invariance is equivalent to cash additivity:  $\rho(X + m) = \rho(X) + \rho(m)$ .

For most purposes in the monetary setting, it is no loss of generality to assume  $\rho$  is normalized. In the few cases it is needed, we assume it explicitly.

The next proposition shows that monetary risk functionals are automatically  $L^\infty$ -Lipschitz continuous.

**Proposition 2.1.** *If  $\rho$  is monetary, then for all  $X, Y \in L^\infty$ ,*

$$|\rho(X) - \rho(Y)| \leq \|X - Y\|_\infty.$$

*Proof.* Let

$$\delta := \|X - Y\|_\infty.$$

Then

$$Y - \delta \leq X \leq Y + \delta \quad \text{a.s.}$$

By monotonicity,

$$\rho(Y - \delta) \leq \rho(X) \leq \rho(Y + \delta).$$

By cash invariance,

$$\rho(Y - \delta) = \rho(Y) - \delta, \quad \rho(Y + \delta) = \rho(Y) + \delta.$$

Hence

$$\rho(Y) - \delta \leq \rho(X) \leq \rho(Y) + \delta,$$

which is equivalent to

$$|\rho(X) - \rho(Y)| \leq \delta = \|X - Y\|_\infty.$$

□

**Lemma 2.1** (2/3 lemma). *Let  $\rho$  be monetary and normalized. Any two of the following three properties imply the third:*

1. convexity,
2. positive homogeneity,

3. *subadditivity.*

*Proof.* Convexity and positive homogeneity implies subadditivity. For any  $X, Y$ ,

$$\rho(X + Y) = 2\rho\left(\frac{X + Y}{2}\right) \leq 2\left(\frac{\rho(X) + \rho(Y)}{2}\right) = \rho(X) + \rho(Y).$$

Subadditivity and positive homogeneity implies convexity. For  $0 \leq \lambda \leq 1$ ,

$$\rho(\lambda X + (1 - \lambda)Y) \leq \rho(\lambda X) + \rho((1 - \lambda)Y) = \lambda\rho(X) + (1 - \lambda)\rho(Y).$$

Convexity and subadditivity implies positive homogeneity. First, convexity with  $Y = 0$  and normalization gives, for  $0 \leq \lambda \leq 1$ ,

$$\rho(\lambda X) = \rho(\lambda X + (1 - \lambda)0) \leq \lambda\rho(X).$$

Next, for  $\lambda \geq 1$ , write

$$X = \frac{1}{\lambda}(\lambda X) + \left(1 - \frac{1}{\lambda}\right)0.$$

Convexity gives

$$\rho(X) \leq \frac{1}{\lambda}\rho(\lambda X),$$

so

$$\rho(\lambda X) \geq \lambda\rho(X).$$

On the other hand, subadditivity gives for integers  $n \geq 1$ ,

$$\rho(nX) \leq n\rho(X).$$

Hence for rational  $\lambda = m/n \geq 1$ ,

$$n\rho\left(\frac{m}{n}X\right) \geq \rho(mX) \quad \text{by subadditivity,}$$

and therefore

$$\rho\left(\frac{m}{n}X\right) \leq \frac{m}{n}\rho(X).$$

Combined with the reverse inequality from convexity, this yields

$$\rho(\lambda X) = \lambda\rho(X) \quad \text{for rational } \lambda \geq 1.$$

The same argument for  $0 \leq \lambda \leq 1$  gives equality there too. Since a monetary functional is Lipschitz,  $\lambda \mapsto \rho(\lambda X)$  is continuous, so the rational identity extends to all  $\lambda \geq 0$ .  $\square$

**Lemma 2.2.** *Let  $\rho$  be convex, normalized, and monetary. Then for every  $X \in L^\infty$ :*

1. *If  $0 \leq \lambda \leq 1$ , then*

$$\rho(\lambda X) \leq \lambda\rho(X);$$

2. *If  $\lambda \geq 1$ , then*

$$\rho(\lambda X) \geq \lambda\rho(X).$$

*Proof.* For  $0 \leq \lambda \leq 1$ , convexity and normalization give

$$\rho(\lambda X) = \rho(\lambda X + (1 - \lambda)0) \leq \lambda\rho(X) + (1 - \lambda)\rho(0) = \lambda\rho(X).$$

Now let  $\lambda \geq 1$ . Then

$$X = \frac{1}{\lambda}(\lambda X) + \left(1 - \frac{1}{\lambda}\right)0.$$

Applying convexity again,

$$\rho(X) \leq \frac{1}{\lambda}\rho(\lambda X) + \left(1 - \frac{1}{\lambda}\right)\rho(0) = \frac{1}{\lambda}\rho(\lambda X).$$

Multiplying by  $\lambda$  gives

$$\lambda\rho(X) \leq \rho(\lambda X).$$

$\square$

*Remark 2.3.* Lemma 2.2 shows the scaling signature of convexity. Small positions are sub-linearly risky, so splitting or reducing a position weakly lowers the capital requirement per unit. Large positions are super-linearly risky, so scaling up is discouraged. In liquidity language, convex risk measures penalize concentration. Small trades are weakly cheaper per unit than large trades, reflecting the idea that liquidity worsens as position size increases.

Next, we present some basic results about comonotonicity.

*Remark 2.4.* Every random variable is comonotone with every constant. Indeed, if  $Y \equiv c$  a.s., then for all  $\omega, \omega'$ ,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) = (X(\omega) - X(\omega')) \cdot 0 = 0 \geq 0,$$

and hence  $X$  and  $c$  are comonotone.

Comonotonic additivity gives a translation formula.

**Lemma 2.3.** *Assume  $\rho$  is comonotonic additive. Then for every constant  $m$ ,*

$$\rho(X + m) = \rho(X) + \rho(m).$$

*Proof.* By Remark 2.4,  $X$  and  $m$  are comonotone. □

*Remark 2.5.* To get cash invariance from Lemma 2.3 we still need

$$\rho(m) = m \quad \text{for constants } m,$$

which follows if  $\rho$  is normalized, but it can fail without.

The next lemma shows that monetary, comonotonic additive functionals are positive homogeneous.

**Lemma 2.4.** *Assume  $\rho$  is monetary and comonotonic additive. Then  $\rho$  is positively homogeneous.*

*Proof.* The first step proves integer homogeneity holds. Any random variable is comonotone with positive multiples of itself, thus for  $n \in \mathbb{N}$

$$\rho(nX) = \rho(X + \dots + X) = n\rho(X).$$

The second step extends to rational homogeneity. Let  $\lambda = m/n \in \mathbb{Q}_+$  with  $m, n \in \mathbb{N}$ . Then

$$n\rho\left(\frac{m}{n}X\right) = \rho\left(n\frac{m}{n}X\right) = \rho(mX) = m\rho(X),$$

so

$$\rho(\lambda X) = \lambda\rho(X).$$

The last step extends from rationals to all  $\lambda \geq 0$ . Because  $\rho$  is monetary, it is  $L^\infty$ -Lipschitz. Hence

$$|\rho(\lambda X) - \rho(\mu X)| \leq \|(\lambda - \mu)X\|_\infty = |\lambda - \mu| \|X\|_\infty,$$

showing that  $\lambda \mapsto \rho(\lambda X)$  is continuous on  $[0, \infty)$ . Since homogeneity holds for all nonnegative rationals, it holds for all  $\lambda \geq 0$  because  $\mathbb{Q}$  is dense in  $\mathbb{R}$ . □

**Corollary 2.1.** *Every monetary comonotonic additive convex risk measure is coherent.*

*Proof.* Lemma 2.4 gives positive homogeneity. Since constants are comonotone,  $\rho(0 + m) = \rho(0) + \rho(m)$  and so  $\rho(0) = 0$  and  $\rho$  is normalized. Thus we can apply the 2/3 lemma to turn convexity into subadditivity. □

**Proposition 2.2.** *Let  $X, Y \in L^\infty$ . The following are equivalent.*

1.  $X$  and  $Y$  are comonotone.
2. There exist a random variable  $Z$  and increasing functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$X = f(Z), \quad Y = g(Z) \quad \text{a.s.}$$

*Proof.* The implication (2)  $\Rightarrow$  (1) is immediate: if  $X = f(Z)$  and  $Y = g(Z)$  with  $f, g$  increasing, then for all  $\omega, \omega'$ ,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) = (f(Z(\omega)) - f(Z(\omega')))(g(Z(\omega)) - g(Z(\omega'))) \geq 0,$$

because the two factors always have the same sign.

For (1)  $\Rightarrow$  (2), we can take

$$Z := X + Y,$$

and show that  $X$  and  $Y$  are increasing functions of  $Z$ . Assume  $X$  and  $Y$  are comonotone, and suppose

$$Z(\omega) \leq Z(\omega').$$

We claim this implies

$$X(\omega) \leq X(\omega') \quad \text{and} \quad Y(\omega) \leq Y(\omega').$$

For if, say,  $X(\omega) > X(\omega')$ , then comonotonicity forces  $Y(\omega) \geq Y(\omega')$ . But, then

$$Z(\omega) = X(\omega) + Y(\omega) > X(\omega') + Y(\omega') = Z(\omega'),$$

a contradiction. Hence  $X(\omega) \leq X(\omega')$ . The same argument gives  $Y(\omega) \leq Y(\omega')$ . Thus both  $X$  and  $Y$  are order-preserving functions of  $Z$ : whenever  $Z(\omega) \leq Z(\omega')$ , one has  $X(\omega) \leq X(\omega')$  and  $Y(\omega) \leq Y(\omega')$ . Therefore one may define increasing functions  $f$  and  $g$  on the range of  $Z$  by

$$f(Z(\omega)) := X(\omega), \quad g(Z(\omega)) := Y(\omega),$$

and extend them monotonically to all of  $\mathbb{R}$ . Then

$$X = f(Z), \quad Y = g(Z) \quad \text{a.s.}$$

□

**Lemma 2.5** (Denneberg (1994)). *Let  $X_1, \dots, X_n \in L^\infty$  be pairwise comonotone. Then there exist a random variable  $Z$  and increasing functions  $f_1, \dots, f_n$  such that*

$$X_i = f_i(Z) \quad \text{a.s. for } i = 1, \dots, n.$$

Consequently, if  $\rho$  is comonotonic additive, then

$$\rho\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \rho(X_i).$$

*Proof.* The first statement is the finite-family extension of Proposition 2.2: if  $X_1, \dots, X_n$  are pairwise comonotone, then with

$$Z := X_1 + \dots + X_n$$

each  $X_i$  is an increasing function of  $Z$ .

For the second statement, the representation

$$X_i = f_i(Z), \quad i = 1, \dots, n,$$

shows that every partial sum

$$S_k := X_1 + \dots + X_k$$

is also an increasing function of  $Z$ . Hence  $S_{k-1}$  and  $X_k$  are comonotone for each  $k$ , and comonotonic additivity gives

$$\rho(S_k) = \rho(S_{k-1}) + \rho(X_k).$$

Induction on  $k$  yields

$$\rho\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \rho(X_i).$$

□

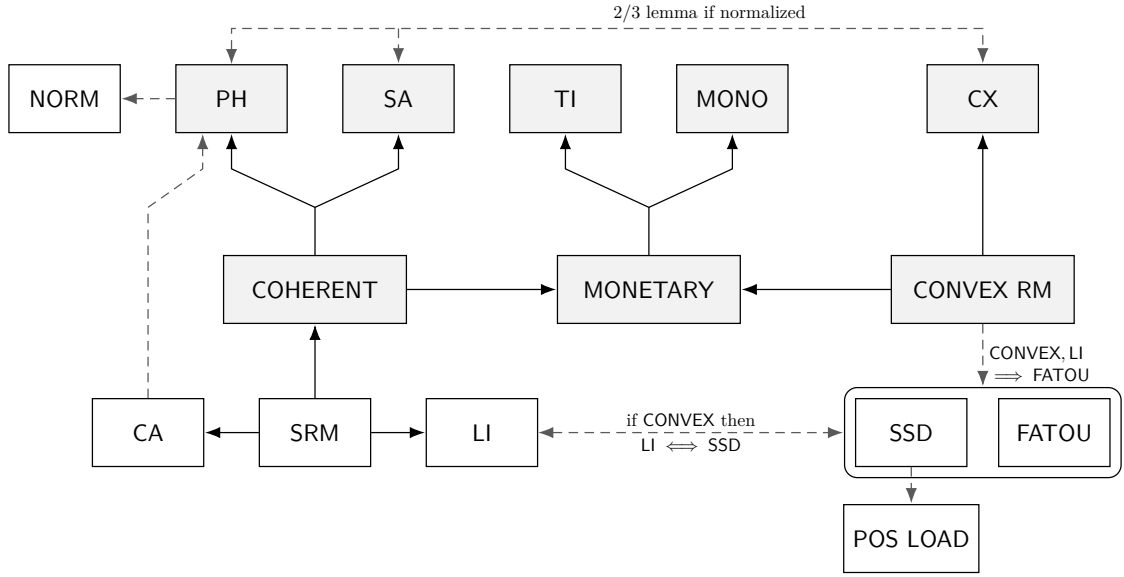


Figure 2: Relationships between properties of risk measures. Adapted from Mildenhall and Major (2022).

*Remark 2.6.* Denneberg’s lemma is useful because it reduces a comonotone family to a one-factor picture: all variables move in step with the same underlying state variable  $Z$ . It explains why Choquet functionals are additive on comonotone sums. For allocations  $X = \xi_0 + \xi_1$ , it shows there is an equivalence between allocations that increase with  $X$  and allocations where  $\xi_i$  are comonotone.

Figure 2 summarizes the results presented in this section. In the figure, properties defining coherent and convex risk measures are in grey. A solid line indicates a definition, for example, a monetary risk measure is monotone and translation invariant. A cell is defined by all solid lines emerging from it, reading the arrows as implies. Dashed lines indicate derived results, for example, positive homogeneous implies normalized. Respecting second order stochastic dominance implies a positive loading (lower right).

### 3 Convex Duality

We now summarize the facts about convex duality needed for the Atlas. The material is standard, but the risk measure specialization is so central that it is worth showing in some detail. The theory exhibits the beauty of mathematics: abstract convex duality determines a representation of risk measures as the worst outcome over a set of scenarios, with each suitably scenario handicapped, a representation that comports with an intuitive rational, and a regulator’s or risk manager’s approach to risk quantification. We start with the general theory, then specialize to  $L^\infty$  and the coherent case, and end with a geometric discussion.

#### 3.1 General Theory

To keep the discussion reasonably concrete, let  $X$  be a Banach space and let  $X'$  denote a subspace of its algebraic dual that separates points. The topology on  $X$  is taken to be the weakest topology making all  $x' \in X'$  continuous, denoted

$$\sigma(X, X').$$

Then, the topological dual of  $(X, \sigma(X, X'))$  is precisely  $X'$ . The weak topology is generally much weaker than the norm topology: viewing  $X$  inside the product space  $\mathbb{R}^{X'}$ , the weak topology uses coordinate-wise convergence, whereas norm topology demands uniform control over the dual unit ball. The pairing between  $X$  and  $X'$  is written

$$\langle x, \mu \rangle = \mu(x).$$

See Section 8.3 for a more about dual pairs.

**Definition 3.1.** Let

$$f : X \rightarrow \mathbb{R} \cup \{+\infty\}.$$

The function  $f$  is **proper** if  $f(x) < \infty$  for some  $x$ . The **convex conjugate** of a proper function is

$$f^*(\mu) := \sup_{x \in X} \mu(x) - f(x), \quad \mu \in X',$$

and the **biconjugate** is

$$f^{**}(x) := \sup_{\mu \in X'} \mu(x) - f^*(\mu), \quad x \in X.$$

**Definition 3.2.** Let  $X$  be a topological space, and let  $f : X \rightarrow (-\infty, \infty]$ . Then  $f$  is **lower semicontinuous**, abbreviated lsc, if for every  $c \in \mathbb{R}$  the sub-level set

$$\{x \in X : f(x) \leq c\}$$

is closed. Equivalently,  $f$  is lower semicontinuous if for every  $c \in \mathbb{R}$  the strict super-level set  $\{x \in X : f(x) > c\}$  is open.

A lower semicontinuous function may jump downward, but it may not jump upward. In risk management, that is exactly the direction one wants: if a sequence of positions converges to a limit, we do not want the limiting position to have mysteriously greater risk than the approximating positions indicate. The definition depends on the topology because it is expressed in terms of closed sets. Convergence of positions is relative to the same topology.

*Remark 3.1.* For any proper  $f$ , the function  $f^*$  is convex (it is the sup of affine functions and affine functions are convex) and  $\sigma(X', X)$ -lower semicontinuous (affine functions are lsc and sup of lsc functions is lsc), and  $f^{**}$  is convex and  $\sigma(X, X')$ -lower semicontinuous.

It is always true that  $f^{**} \leq f$ . Lower semicontinuity is the extra ingredient to ensure equality. The precise statement is:

**Theorem 3.1** (Fenchel–Moreau convex duality). *Let  $(X, X')$  be a dual pair, and equip  $X$  with the topology  $\sigma(X, X')$ . Let*

$$f : X \rightarrow (-\infty, \infty]$$

*be proper. Then the biconjugate  $f^{**}$  is always convex and  $\sigma(X, X')$ -lower semicontinuous, and*

$$f^{**} \leq f.$$

*If, in addition,  $f$  is convex and  $\sigma(X, X')$ -lower semicontinuous, then*

$$f^{**} = f.$$

*Proof.* For each fixed  $y \in X'$ , the function

$$x \mapsto \langle x, y \rangle - f^*(y)$$

is affine and  $\sigma(X, X')$ -continuous, hence convex and lower semicontinuous. Since  $f^{**}$  is the supremum of these affine functions, it follows that  $f^{**}$  is convex and  $\sigma(X, X')$ -lower semicontinuous.

Next we show  $f^{**} \leq f$ . For any  $x \in X$  and  $y \in X'$ ,

$$f^*(y) = \sup_{u \in X} \langle u, y \rangle - f(u) \geq \langle x, y \rangle - f(x),$$

so

$$\langle x, y \rangle - f^*(y) \leq f(x).$$

Taking the supremum over  $y \in X'$  gives

$$f^{**}(x) \leq f(x).$$

Now assume  $f$  is convex and  $\sigma(X, X')$ -lower semicontinuous, and fix  $x_0 \in X$ . To prove the reverse inequality  $f(x_0) \leq f^{**}(x_0)$ , let  $r < f(x_0)$ . Then

$$(x_0, r) \notin \text{epi}(f), \quad \text{epi}(f) := \{(x, t) \in X \times \mathbb{R} : t \geq f(x)\}.$$

Since  $f$  is convex,  $\text{epi}(f)$  is convex; since  $f$  is  $\sigma(X, X')$ -lower semicontinuous,  $\text{epi}(f)$  is closed in  $X \times \mathbb{R}$  for the product topology  $\sigma(X, X') \times$  the usual topology on  $\mathbb{R}$ . Hence, by the separating hyperplane theorem, there exist  $(a, b) \in (X \times \mathbb{R})' = X' \times \mathbb{R}$ , with  $(a, b) \neq (0, 0)$ , and  $\alpha \in \mathbb{R}$  such that

$$\langle a, x \rangle + bt \geq \alpha \quad \text{for all } (x, t) \in \text{epi}(f),$$

while

$$\langle a, x_0 \rangle + br < \alpha.$$

One must have  $b > 0$ :  $b < 0$  is impossible because the epigraph is upward closed in the  $t$ -direction, and  $b = 0$  contradicts the fact that  $(x_0, f(x_0)) \in \text{epi}(f)$ .

Set

$$y = -a/b, \quad c = \alpha/b.$$

Then the separation inequality becomes

$$f(x) \geq \langle x, y \rangle + c \quad \text{for all } x \in X,$$

and the strict inequality at  $(x_0, r)$  becomes

$$r < \langle x_0, y \rangle + c.$$

The first display implies

$$\langle x, y \rangle - f(x) \leq -c \quad \text{for all } x \in X,$$

hence

$$f^*(y) \leq -c.$$

Therefore

$$\langle x_0, y \rangle - f^*(y) \geq \langle x_0, y \rangle + c > r,$$

so by definition of  $f^{**}$ ,

$$f^{**}(x_0) \geq r.$$

Since this holds for every  $r < f(x_0)$ , we obtain

$$f^{**}(x_0) \geq f(x_0).$$

Combined with  $f^{**}(x_0) \leq f(x_0)$ , this gives

$$f^{**}(x_0) = f(x_0).$$

As  $x_0$  is arbitrary,  $f^{**} = f$ . □

Convexity is used in the proof to ensure  $\text{epi}(f)$  is convex and lsc to ensure it is closed; together these allow application of the hyperplane separation theorem.

*Remark 3.2.* Lower semicontinuity needed for the biconjugate formula is exactly lower semicontinuity with respect to the topology induced by the chosen dual pair. In the risk-measure application below, the relevant dual pair is

$$(L^\infty, L^1),$$

so one needs  $\sigma(L^\infty, L^1)$ -lower semicontinuity.

**Definition 3.3.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be proper and convex. The **subdifferential** of  $f$  at  $x \in X$  is

$$\partial f(x) := \{ \mu \in X' : f(y) \geq f(x) + \mu(y - x) \text{ for all } y \in X \}.$$

By definition,

$$\mu \in \partial f(x) \iff f(y) \geq \mu(y) - (\mu(x) - f(x)) \text{ for all } y,$$

and therefore  $\mu \in \partial f(x)$  means that  $\mu$  defines a supporting affine functional to the epigraph of  $f$  at  $x$ .

**Proposition 3.1** (Fenchel-Young inequality). *For every  $x \in X$  and  $\mu \in X'$ ,*

$$f(x) + f^*(\mu) \geq \mu(x).$$

*Proof.* By its definition as a sup,  $f^*(\mu) \geq \mu(x) - f(x)$ , which rearranges to give the result.  $\square$

**Proposition 3.2.** *Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be proper and convex. For  $x \in X$  and  $\mu \in X'$  the following are equivalent:*

1.  $\mu \in \partial f(x)$ ;
2.  $x \in \partial f^*(\mu)$ ;
3. Fenchel–Young’s inequality is an equality:  $f(x) + f^*(\mu) = \mu(x)$ .

*Proof.* We show (1)  $\iff$  (3)  $\iff$  (2).

Assume  $\mu \in \partial f(x)$ . Then for all  $y \in X$ ,

$$f(y) \geq f(x) + \mu(y - x),$$

so

$$\mu(y) - f(y) \leq \mu(x) - f(x).$$

Taking the supremum over  $y$  gives

$$f^*(\mu) \leq \mu(x) - f(x).$$

Fenchel–Young gives the reverse inequality and so we obtain equality  $f(x) + f^*(\mu) = \mu(x)$ . Thus (1)  $\implies$  (3).

Conversely, if

$$f(x) + f^*(\mu) = \mu(x),$$

then for every  $y \in X$ ,

$$\mu(y) - f(y) \leq f^*(\mu) = \mu(x) - f(x),$$

hence

$$f(y) \geq f(x) + \mu(y - x),$$

and so  $\mu \in \partial f(x)$ . Thus (3)  $\implies$  (1).

Now apply to  $f^*$  in place of  $f$ . Since

$$f(x) + f^*(\mu) = \mu(x)$$

is symmetric in  $f$  and  $f^*$ , condition (3) is equivalent to

$$x \in \partial f^*(\mu).$$

Hence (2)  $\iff$  (3).  $\square$

*Remark 3.3.* The equivalences in Proposition 3.2,

$$\mu \in \partial f(x) \iff x \in \partial f^*(\mu) \iff f(x) + f^*(\mu) = \mu(x),$$

are a simultaneous optimality relation:  $x$  is optimal for the supremum defining  $f^*(\mu)$ , and  $\mu$  is optimal for the supremum defining  $f^{**}(x)$ .

### 3.2 Specialization to risk measures on $L^\infty$

We now return to a convex monetary risk measure  $\rho : L^\infty \rightarrow \mathbb{R}$ . We use the dual pair  $(L^\infty, L^1)$ , with pairing  $\langle X, Z \rangle := \mathbb{P}(XZ)$ . This is the countably additive dual pair used for Fatou-type representations; it is not the full Banach dual of  $L^\infty$ .

At first sight the dual variable in the representation of a convex risk measure appears as an arbitrary element  $Z \in L^1$ . Later, [?@prp-cashinv-dual](#) shows that whenever  $\rho^*(Z) < \infty$ , the variable  $Z$  is in fact the Radon–Nikodym derivative of a countably additive probability measure  $\mathbb{Q} \ll \mathbb{P}$ , so that  $Z = d\mathbb{Q}/d\mathbb{P}$ . Thus the pairing can be written equivalently as

$$\mathbb{P}(XZ) = \mathbb{Q}(X) = \int X d\mathbb{Q}$$

We therefore move freely between the density notation  $Z$  and the measure notation  $Q$ , using whichever is more intuitive in context.

Once the dual variables are read as probability measures  $Q \ll P$ , the dual representation acquires a natural risk-management interpretation. A risk measure evaluates a position by examining its expected loss under a family of alternative probability measures, and then taking the worst penalized value. Each such measure  $Q$  is a **scenario**: not necessarily a single deterministic event, but a way the baseline model  $P$  is tilted toward more adverse outcomes. In that sense, the dual representation expresses risk measurement as systematic stress testing across a family of scenarios.

**Definition 3.4.** The convex conjugate of  $\rho$  is

$$\rho^*(Z) := \sup_{X \in L^\infty} P(XZ) - \rho(X), \quad Z \in L^1.$$

If  $\rho$  is  $\sigma(L^\infty, L^1)$ -lower semicontinuous, then Fenchel–Moreau gives

$$\rho(X) = \sup_{Z \in L^1} P(XZ) - \rho^*(Z).$$

**?@prp-cashinv-dual** links properties of  $\rho$  with those of the penalty function  $\alpha$  used to define it. It starts with an arbitrary  $\alpha$  on the dual space and defines  $\rho = \alpha^*$  on the primal space. Then

$$\rho^* = \alpha^{**},$$

and therefore  $\rho^* = \alpha$  exactly when  $\alpha$  is convex and lower semicontinuous in the relevant dual topology.

**Proposition 3.3.** Let  $\alpha : ba(\Omega, \mathcal{F}) \rightarrow (-\infty, \infty]$  be any function, and define

$$\rho(X) := \sup_{\nu \in ba} \langle X, \nu \rangle - \alpha(\nu), \quad X \in L^\infty(\Omega, \mathcal{F}, P),$$

where  $\langle X, \nu \rangle = \int X d\nu$ . Assume  $\rho(X) < \infty$  for all  $X \in L^\infty$ .

Then the following hold.

1.  $\rho$  is cash invariant if and only if

$$\text{dom } \alpha \subseteq \{ \nu \in ba : \nu(\Omega) = 1 \}.$$

2.  $\rho$  is monotone if and only if

$$\text{dom } \alpha \subseteq \{ \nu \in ba : \nu \text{ is positive} \}.$$

3.  $\rho$  is normalized ( $\rho(0) = 0$ ) if and only if

$$\inf_{\nu \in ba} \alpha(\nu) = 0.$$

Equivalently,

$$\inf_{\nu \in \text{dom } \alpha} \alpha(\nu) = 0.$$

4. If  $\rho$  is well defined on  $L^\infty(\Omega, \mathcal{F}, P)$ , equivalently it depends only on  $P$ -a.s. equivalence classes, then

$$\text{dom } \alpha \subseteq \{ \nu \in ba : \nu \ll P \}.$$

*Proof.* For (1), let  $m \in \mathbb{R}$ . Then

$$\begin{aligned} \rho(X + m) &= \sup_{\nu \in ba} \langle X + m, \nu \rangle - \alpha(\nu) \\ &= \sup_{\nu \in ba} \langle X, \nu \rangle + m\nu(\Omega) - \alpha(\nu). \end{aligned}$$

If every  $\nu$  in  $\text{dom } \alpha$  satisfies  $\nu(\Omega) = 1$ , then

$$\rho(X + m) = \sup_{\nu \in ba} \langle X, \nu \rangle + m - \alpha(\nu) = \rho(X) + m,$$

so  $\rho$  is cash invariant.

Conversely, assume  $\rho$  is cash invariant, and let  $\nu \in \text{dom } \alpha$ . For every  $X \in L^\infty$  and every  $m \in \mathbb{R}$ ,

$$\rho(X + m) \geq \langle X, \nu \rangle + m\nu(\Omega) - \alpha(\nu).$$

Since  $\rho(X + m) = \rho(X) + m$ , we get for all  $m$

$$\rho(X) \geq \langle X, \nu \rangle + m(\nu(\Omega) - 1) - \alpha(\nu).$$

If  $\nu(\Omega) \neq 1$ , choosing  $m \rightarrow \pm\infty$  with the appropriate sign gives a contradiction. Hence  $\nu(\Omega) = 1$ .

For (2), assume first that every  $\nu$  in  $\text{dom } \alpha$  is positive. If  $X \leq Y$  almost surely, then

$$\langle X, \nu \rangle \leq \langle Y, \nu \rangle$$

for every such  $\nu$ , hence

$$\rho(X) \leq \rho(Y).$$

So  $\rho$  is monotone.

Conversely, assume  $\rho$  is monotone, and let  $\nu \in \text{dom } \alpha$ . Suppose  $\nu$  is not positive. Then there exists  $A \in \mathcal{F}$  with  $\nu(A) < 0$ . Let

$$X := -nA, \quad Y := 0.$$

Then  $X \leq Y$ , so monotonicity gives

$$\rho(X) \leq \rho(0).$$

But

$$\rho(X) \geq \langle X, \nu \rangle - \alpha(\nu) = -n\nu(A) - \alpha(\nu) \rightarrow +\infty,$$

since  $\nu(A) < 0$ , a contradiction. Thus  $\nu$  is positive.

For (3),

$$\rho(0) = \sup_{\nu \in \text{ba}} -\alpha(\nu) = - \inf_{\nu \in \text{ba}} \alpha(\nu).$$

Therefore  $\rho(0) = 0$  if and only if  $\inf \alpha = 0$ .

Finally, for (4), let  $\nu \in \text{dom } \alpha$ , and suppose  $\nu \not\ll P$ . Then there exists  $A \in \mathcal{F}$  with

$$P(A) = 0, \quad \nu(A) \neq 0.$$

In  $L^\infty$  we have  $A = 0$ , so for every  $X \in L^\infty$  and every  $n \in \mathbb{N}$ ,

$$X + n1_A = X \quad \text{in } L^\infty.$$

Hence  $\rho(X + nA) = \rho(X)$ . But, then the representation gives

$$\rho(X) = \rho(X + nA) \geq \langle X, \nu \rangle + n\nu(A) - \alpha(\nu).$$

Replacing  $A$  by a subset if necessary, we may assume  $\nu(A) > 0$ , and then the right-hand side tends to  $+\infty$  as  $n \rightarrow \infty$ . This contradicts finiteness of  $\rho(X)$  for some  $X$ . Therefore every  $\nu$  in  $\text{dom } \alpha$  must satisfy  $\nu \ll P$ .  $\square$

*Remark.* There is obviously some chicken-and-egg going on here: do we define  $\rho$  first or the penalty? There is another route to the penalty via the acceptance set (see also Section 6.2). If  $\rho$  is a monetary risk measure, its **acceptance set** is defined as

$$\mathcal{A}_\rho := \{X \in L^\infty : \rho(X) \leq 0\}.$$

When  $\rho$  is convex,  $\mathcal{A}_\rho$  is convex, and one may define the associated penalty function by

$$\alpha_{\mathcal{A}}(Q) := \sup_{X \in \mathcal{A}_\rho} Q(X).$$

This is a canonical, or minimal, penalty: if

$$\rho(X) = \sup_Q \{Q(X) - \alpha(Q)\},$$

then necessarily

$$\alpha_{\mathcal{A}}(Q) \leq \alpha(Q)$$

for all  $Q$ . Under the usual convex lower semicontinuity assumptions, this minimal penalty coincides with the convex conjugate  $\rho^*$ . See

**Proposition 3.4.** *Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be convex and monetary. Then  $\rho$  has the Fatou property if and only if  $\rho$  is  $\sigma(L^\infty, L^1)$ -lower semicontinuous.*

*Remark 3.4.* Proposition 3.4 is a pivotal result because it links two continuity notions formulated in strikingly different languages. The Fatou property is probabilistic: it concerns a sequence  $(X_n)$  that converges almost surely to  $X$  and is uniformly bounded in  $L^\infty$ , and asks that  $\rho(X)$  not exceed  $\liminf_n \rho(X_n)$ . That is the natural notion from the perspective of a risk manager or pricing actuary, because it describes what it means for a sequence of loss positions to settle down state by state, except on a null set, without blowing up in size. Lower semicontinuity, by contrast, is inherently topological: it is defined in terms of closed sub-level sets, or equivalently by behavior under convergence in a specified topology. The two notions do not obviously have anything to do with one another. In particular, almost sure convergence is not itself topological, and in fact cannot be generated by any topology on  $L^\infty$ .

The proposition shows that for convex monetary functionals on  $L^\infty$  the natural probability-theory closure condition used in practice is exactly equivalent to lower semicontinuity for the specific dual-pair topology  $\sigma(L^\infty, L^1)$ . The uniform  $L^\infty$  bound is crucial in the bridge from one formulation to the other: if  $|X_n| \leq M$  and  $X_n \rightarrow X$  almost surely, then for every  $Z \in L^1$  we have  $|X_n Z| \leq M|Z|$  with  $M|Z| \in L^1$ , so Lebesgue's dominated convergence theorem gives  $\mathbb{P}(X_n Z) \rightarrow \mathbb{P}(XZ)$ . In other words, bounded almost sure convergence implies  $\sigma(L^\infty, L^1)$ -convergence. That implication is the key mechanism behind the equivalence, and it is what allows us to pass from a statement about converging risks to the convex-analytic machinery of Fenchel–Moreau duality.

The proof relies on the following technical lemma, which we prove in Section 8.8.

**Lemma 3.1.** *Let  $C \subseteq L^\infty$  be convex. Assume that  $C$  is closed under almost sure convergence of uniformly bounded sequences, in the sense that whenever  $(X_n) \subseteq C$  satisfies*

$$\sup_n \|X_n\|_\infty < \infty \quad \text{and} \quad X_n \rightarrow X \text{ a.s.},$$

*we have  $X \in C$ . Then  $C$  is  $\sigma(L^\infty, L^1)$ -closed.*

The lemma isolates the only nontrivial functional-analytic step in the proof of Proposition 3.4. The Fatou property shows that each sub-level set

$$C_m = \{X \in L^\infty : \rho(X) \leq m\}$$

is closed under almost sure convergence of uniformly bounded sequences. Since  $C_m$  is also convex, Lemma 3.1 implies that each  $C_m$  is  $\sigma(L^\infty, L^1)$ -closed, which is exactly the lower semicontinuity of  $\rho$ .

*Proof.* Suppose  $\rho$  has the Fatou property and let

$$C_m := \{X \in L^\infty : \rho(X) \leq m\}.$$

Because  $\rho$  is convex, each  $C_m$  is convex. To prove  $\sigma(L^\infty, L^1)$ -lower semicontinuity of  $\rho$ , it suffices to show each  $C_m$  is  $\sigma(L^\infty, L^1)$ -closed. To do that, let  $(X_\alpha)$  be a net in  $C_m$  with

$$X_\alpha \rightarrow X \quad \text{in } \sigma(L^\infty, L^1).$$

A standard closure argument on convex sets in  $L^\infty$  reduces this to showing that  $C_m$  is closed under bounded a.s. convergence. But that is exactly what the Fatou property gives: if  $\sup_n \|X_n\|_\infty < \infty$  and  $X_n \rightarrow X$  a.s. with each  $X_n \in C_m$ , then

$$\rho(X) \leq \liminf_n \rho(X_n) \leq m,$$

so  $X \in C_m$ . Hence each  $C_m$  is  $\sigma(L^\infty, L^1)$ -closed, and therefore  $\rho$  is  $\sigma(L^\infty, L^1)$ -lower semicontinuous.

For the converse, assume  $\rho$  is  $\sigma(L^\infty, L^1)$ -lower semicontinuous. Let  $(X_n)$  satisfy

$$\sup_n \|X_n\|_\infty < \infty \quad \text{and} \quad X_n \rightarrow X \text{ a.s.}$$

We need show that  $\rho(X) \leq \liminf_n \rho(X_n)$ . Because the sequence is uniformly bounded, say

$$|X_n| \leq M \quad \text{a.s. for all } n,$$

and  $X_n \rightarrow X$  a.s., dominated convergence implies that for every  $Z \in L^1$ ,

$$\mathbb{P}(X_n Z) \rightarrow \mathbb{P}(X Z).$$

That is exactly the statement that

$$X_n \rightarrow X \quad \text{in } \sigma(L^\infty, L^1).$$

Now lower semicontinuity of  $\rho$  for this topology gives

$$\rho(X) \leq \liminf_n \rho(X_n).$$

Hence  $\rho$  has the Fatou property. □

*Remark 3.5.* Combined with Fenchel–Moreau, Proposition 3.4 yields the dual representation

$$\rho(X) = \sup_{Z \in L^1} \mathbb{P}(XZ) - \rho^*(Z)$$

for every convex monetary risk measure with the Fatou property.

*Remark 3.6.* The  $L^1$  representation for a Fatou-continuous convex risk measure is generally a supremum, not a maximum:

$$\rho(X) = \sup_{Z \in L^1} \mathbb{P}(XZ) - \rho^*(Z).$$

If  $\rho(X) = \text{ess sup } X$ , the supremum is not attained in  $L^1$  when  $\{X = \text{ess sup } X\}$  has measure zero.

At the level of the full Banach dual,  $(L^\infty)^* = ba$ , one works instead with finitely additive dual variables and obtains the broader representation

$$\rho(X) = \max_{\mu \in ba} \mu(X) - \rho^*(\mu).$$

Some lower semicontinuity hypothesis is always needed for exact biconjugacy. For the dual pair  $(L^\infty, L^1)$ , the needed hypothesis is  $\sigma(L^\infty, L^1)$ -lower semicontinuity, which for convex monetary risk measures is equivalent to the Fatou property, as we have seen. For the full Banach dual pair  $(L^\infty, ba)$ , Fenchel–Moreau requires  $\sigma(L^\infty, ba)$ -lower semicontinuity. For convex functions on a Banach space, however, weak lower semicontinuity with respect to the full dual is equivalent to norm lower semicontinuity. Since a monetary risk measure is Lipschitz in the  $L^\infty$  norm, it is norm continuous, hence norm lower semicontinuous, and so exact biconjugacy over  $ba$  is automatic in the convex monetary setting.

*Remark 3.7.* For an arbitrary function, weak-\* lower semicontinuity implies norm lower semicontinuity, since the norm topology is stronger. The converse is false in general. For convex functions, however, the converse holds: a convex function on a Banach space is norm lower semicontinuous if and only if it is lower semicontinuous for the weak topology induced by the full dual. Equivalently, its epigraph is norm closed if and only if it is weakly closed. (For convex subsets of a Banach space, norm-closedness is equivalent to closedness for the weak topology induced by the full dual.) Thus for a convex monetary risk measure on  $L^\infty$ , Lipschitz continuity gives norm lower semicontinuity, and convexity upgrades this to  $\sigma(L^\infty, ba)$ -lower semicontinuity.

*Remark 3.8.* The scenario interpretation connects the convex-duality theory to several familiar strands of insurance and risk-management practice. In internal models and regulatory work, one often studies named or highly specified stresses: catastrophe events, market dislocations, reserve deterioration, operational failures, or multi-factor combinations of these. Lloyd’s Realistic Disaster Scenarios provide a good example of such concrete scenario design. A dual measure  $Q$  plays a similar mathematical role, though usually at a more abstract level: it specifies a re-weighting of the reference model toward adverse states, and the dual representation asks how costly the position looks under that stress.

The same interpretation also links risk measures to the robust and ambiguity-sensitive literature. From that viewpoint, the reference measure  $\mathbf{P}$  is not trusted completely. Instead, we contemplate a family of plausible alternatives  $\mathbf{Q}$ , and assess risk by guarding against unfavorable members of that family, either equally in the coherent case or with penalties in the convex case. The penalty function records how far one is willing to move away from the baseline model, or how implausible, expensive, or ambiguous a given scenario is judged to be. Thus the dual representation can be read as a mathematical formalization of model uncertainty: risk is not evaluated under one probabilistic view of the world, but under a controlled family of stressed views.

Law invariance changes the flavor of the scenarios. In the general case,  $\mathbf{Q}$  may encode very specific statewise stresses. In the law-invariant case, the relevant stresses are distributional rather than narrative: they are tied to ranks, quantiles, tail layers, or return periods, rather than to named states of the world. Distortion and spectral risk measures make that especially clear. Their scenarios are not “earthquake in region A plus equity shock in sector B,” but rather stresses attached to adverse percentiles of the loss distribution. That distinction mirrors an important divide in practice between narrative scenarios and return-period or percentile-based capital views.

Taking the sup of a set of convex risk measures is a standard recipe to create a new one; it moves us from left to right in the Atlas. It is justified by Proposition 3.5.

**Proposition 3.5.** *Let  $(\rho_i)_{i \in I}$  be a family of convex monetary risk measures on  $L^\infty$ , and assume that each  $\rho_i$  admits the dual representation*

$$\rho_i(X) = \sup_{Z \in \mathcal{D}} \mathbf{P}(XZ) - \rho_i^*(Z),$$

for some common dual domain  $\mathcal{D} \subseteq L^1$ . Assume also that

$$\sup_{i \in I} \rho_i(0) < \infty.$$

Define

$$\rho(X) := \sup_{i \in I} \rho_i(X).$$

Then  $\rho$  is finite-valued, convex, and monetary, and it admits the dual representation

$$\rho(X) = \sup_{Z \in \mathcal{D}} \mathbf{P}(XZ) - \rho^*(Z), \quad \rho^*(Z) = \inf_{i \in I} \rho_i^*(Z).$$

*Proof.* Since each  $\rho_i$  is monetary it is monotone. For every  $X \in L^\infty$ ,

$$\rho_i(X) \leq \rho_i(0) + \|X\|_\infty,$$

and therefore

$$\rho(X) = \sup_{i \in I} \rho_i(X) \leq \sup_{i \in I} \rho_i(0) + \|X\|_\infty < \infty.$$

Thus  $\rho$  is finite-valued. Since a supremum of convex functions is convex, and a supremum of monetary functionals is monetary,  $\rho$  is convex and monetary.

Now compute:

$$\begin{aligned} \rho(X) &= \sup_{i \in I} \rho_i(X) \\ &= \sup_{i \in I} \sup_{Z \in \mathcal{D}} \mathbf{P}(XZ) - \rho_i^*(Z) \\ &= \sup_{Z \in \mathcal{D}} \sup_{i \in I} \mathbf{P}(XZ) - \rho_i^*(Z) \\ &= \sup_{Z \in \mathcal{D}} \mathbf{P}(XZ) - \inf_{i \in I} \rho_i^*(Z). \end{aligned}$$

Hence

$$\rho^*(Z) = \inf_{i \in I} \rho_i^*(Z),$$

as claimed. Note the derivation is elementary; no minimax theorem is involved.  $\square$

Theorem 3.2 identifies the exact continuity requirement for  $\rho$  to ensure a representation on countably additive measures and to ensure the sup in the dual representation is achieved.

**Theorem 3.2.** *Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be a convex monetary risk measure with the Fatou property, and let*

$$\alpha_{\min}(\nu) := \sup_{X \in \mathcal{A}_\rho} \nu(X), \quad \mathcal{A}_\rho := \{X \in L^\infty : \rho(X) \leq 0\},$$

denote its minimal penalty on ba. Then the following are equivalent.

1.  $\rho$  is continuous from above: if  $X_n \downarrow X$  almost surely, then

$$\rho(X_n) \downarrow \rho(X).$$

2. The minimal penalty takes finite values only on countably additive probabilities:

$$\alpha_{\min}(\nu) < \infty \implies \nu \in L^1, \nu \geq 0, \mathbf{P}\nu = 1.$$

3. In the countably additive dual representation

$$\rho(X) = \sup_{Z \in L^1} \{P(XZ) - \rho^*(Z)\},$$

the supremum is attained for every  $X \in L^\infty$ .

In the payoff sign convention used by Follmer and Schied, condition (1) is written as continuity from below.

See Section 8.9 for a sketch of the proof.

*Remark.* The theorem explains the role of continuity from above in the Atlas. The Fatou property is enough to obtain an  $L^1$  representation, but not enough to force attainment:  $\text{ess sup}$  is the standard counterexample. Continuity from above is the extra condition that eliminates purely finitely additive dual variables from the minimal penalty and thereby upgrades the countably additive dual supremum to a maximum.

Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be a convex monetary risk measure with the Fatou property. If  $\rho$  is continuous from above, then it is continuous from below.

*Proof.* By Theorem 3.2, continuity from above implies that for every  $X \in L^\infty$  the supremum in the countably additive dual representation

$$\rho(X) = \sup_{Z \in L^1} P(XZ) - \rho^*(Z)$$

is attained.

Now let  $X_n \uparrow X$  almost surely. By monotonicity,

$$\rho(X_n) \leq \rho(X) \quad \text{for all } n,$$

so  $(\rho(X_n))$  is increasing and bounded above by  $\rho(X)$ .

Choose  $Z \in L^1$  attaining the dual maximum for  $X$ , so

$$\rho(X) = P(XZ) - \rho^*(Z).$$

Then for every  $n$ ,

$$\rho(X_n) \geq P(X_n Z) - \rho^*(Z).$$

Since  $X_n \uparrow X$  almost surely and the sequence is uniformly bounded in  $L^\infty$ , dominated convergence gives

$$P(X_n Z) \rightarrow P(XZ).$$

Therefore

$$\liminf_{n \rightarrow \infty} \rho(X_n) \geq \mathbf{P}(XZ) - \rho^*(Z) = \rho(X).$$

Since  $\rho(X_n) \leq \rho(X)$  for all  $n$ , we have

$$\limsup_{n \rightarrow \infty} \rho(X_n) \leq \rho(X) \leq \liminf_{n \rightarrow \infty} \rho(X_n) \leq \limsup_{n \rightarrow \infty} \rho(X_n).$$

Hence

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

□

### 3.3 The Coherent Case

Now assume  $\rho$  is coherent, that is, monetary, subadditive, and positively homogeneous.

**Proposition 3.6.** *Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be coherent with the Fatou property. Then  $\rho^*$  is the convex indicator function of a  $\sigma(L^1, L^\infty)$ -closed convex set*

$$M \subseteq \{ Z \in L^1 : Z \geq 0, \mathbf{P}Z = 1 \}$$

that is,

$$\rho^*(Z) = \begin{cases} 0, & Z \in M, \\ +\infty, & Z \notin M. \end{cases}$$

Consequently,

$$\rho(X) = \sup_{Z \in M} \mathbf{P}(XZ).$$

*Proof.* By Proposition 3.4  $\rho$  is  $\sigma(L^\infty, L^1)$ -lower semicontinuous. Hence Fenchel–Moreau Theorem 3.1 applies and yields

$$\rho(X) = \sup_{Z \in L^1} \{ \mathbf{P}(XZ) - \rho^*(Z) \},$$

where

$$\rho^*(Z) = \sup_{X \in L^\infty} \{ \mathbf{P}(XZ) - \rho(X) \}.$$

We now show that  $\rho^*$  is the indicator of a closed convex set. Fix  $Z \in L^1$ . If there exists  $X_0$  such that  $\mathbf{P}(X_0Z) - \rho(X_0) > 0$ , then for every  $\lambda > 0$ ,

$$\mathbf{P}((\lambda X_0)Z) - \rho(\lambda X_0) = \lambda(\mathbf{P}(X_0Z) - \rho(X_0)) \rightarrow \infty.$$

Hence  $\rho^*(Z) = +\infty$ . On the other hand, if  $\mathbf{P}(XZ) - \rho(X) \leq 0$  for all  $X \in L^\infty$ , then taking  $X = 0$  gives  $\rho^*(Z) \geq 0$ , while the definition of  $\rho^*$  implies  $\rho^*(Z) \leq 0$ , and therefore  $\rho^*(Z) = 0$ . Together, these arguments show  $\rho^*(Z) \in \{0, +\infty\}$ , and

$$\begin{aligned} \mathcal{M} &:= \{ Z \in L^1 : \rho^*(Z) = 0 \} \\ &= \{ Z \in L^1 : \mathbf{P}(XZ) \leq \rho(X) \ \forall X \in L^\infty \}. \end{aligned}$$

Since  $\rho^*$  is convex and  $\sigma(L^1, L^\infty)$ -lower semicontinuous, the set  $\mathcal{M}$  is convex and  $\sigma(L^1, L^\infty)$ -closed. Thus  $\rho^*$  is the convex indicator of  $\mathcal{M}$ .

Substituting into the biconjugate formula gives

$$\rho(X) = \sup_{Z \in L^1} \mathbf{P}(XZ) - \rho^*(Z) = \sup_{Z \in \mathcal{M}} \mathbf{P}(XZ).$$

Finally, by [?@prp-cashinv-dual](#) we know  $\mathcal{M}$  is a closed convex set of non-negative densities. □

*Remark 3.9.* A coherent risk measure is exactly the support function of a closed convex set of dual elements. The convex case allows a nontrivial penalty  $\rho^*$ ; positive homogeneity collapses that penalty to an indicator.

The set  $\mathcal{M}$  can be identified as  $\partial\rho(0)$  when  $\rho$  is normalized.

**Corollary 3.1.** *Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be coherent, normalized, and have the Fatou property. Then*

$$\partial\rho(0) = \{ Z \in L^1 : \rho^*(Z) = 0 \}.$$

*Proof.* For any proper convex function with  $\rho(0) = 0$ ,

$$Z \in \partial\rho(0) \iff \rho(Y) \geq \rho(0) + \mathbb{P}(ZY) \text{ for all } Y \iff \mathbb{P}(ZY) - \rho(Y) \leq 0 \text{ for all } Y.$$

Taking the supremum over  $Y$  gives

$$Z \in \partial\rho(0) \iff \rho^*(Z) \leq 0.$$

But always  $\rho^*(Z) \geq -\rho(0) = 0$ , by testing at  $Y = 0$ . Hence

$$Z \in \partial\rho(0). \iff \rho^*(Z) = 0.$$

□

### 3.4 Dual Geometry, Extreme Points, and Law-Invariant Orbits

Assume that  $\rho$  is a convex monetary risk measure on  $L^\infty$  with the Fatou property. Then  $\rho$  admits the countably additive dual representation

$$\rho(X) = \sup_{Z \in L^1} \mathbb{P}(XZ) - \rho^*(Z),$$

where

$$\rho^*(Z) = \sup_{X \in L^\infty} \mathbb{P}(XZ) - \rho(X).$$

By [?@prp-cashinv-dual](#), the effective domain

$$D_1 := \text{dom } \rho^* = \{ Z \in L^1 : Z \geq 0, \mathbb{P}Z = 1, \rho^*(Z) < \infty \}$$

consists of probability densities. In this section we relate properties of  $\rho$  to the geometry of its dual objects.

In the coherent case,  $\rho^*$  takes only the values 0 and  $\infty$ , so it is the indicator of a closed convex set

$$D \subseteq \{ Z \in L^1 : Z \geq 0, \mathbb{P}Z = 1 \},$$

and the representation becomes

$$\rho(X) = \sup_{Z \in D} \mathbb{P}(ZX).$$

Thus the natural dual object for a coherent risk measure is a closed convex set of probability densities. Convexity of  $D$  follows because  $D = \{ Z : \rho^*(Z) = 0 \}$  and  $\rho^*$  is convex.

In the convex, non-coherent case, the domain alone is no longer enough: we must retain the penalty values as well. The natural dual object is then the epigraph

$$D := \text{epi}(\rho^*) = \{ (Z, t) \in L^1 \times \mathbb{R} : \rho^*(Z) \leq t \},$$

and the dual representation may be written

$$\rho(X) = \sup_{(Z, t) \in D} \mathbb{P}(ZX) - t.$$

Coherence is therefore described by a set of scenarios, whereas convexity is described by a set of scenarios together with a cost attached to each one.

Suppose now that the relevant dual object is compact in the appropriate topology. In the coherent case, this means compactness of  $D$  in  $\sigma(L^1, L^\infty)$ . Then Krein–Milman gives

$$D = \overline{\text{co}}(\text{ext}(D)),$$

the closed convex hull of  $D$ . For fixed  $X$ , the map

$$Z \mapsto \mathbf{P}(ZX)$$

is affine and  $\sigma(L^1, L^\infty)$ -continuous. Hence, if the supremum is attained, Bauer's maximum principle shows that it is attained at an extreme point. In that case

$$\rho(X) = \max_{Z \in \text{ext}(D)} \mathbf{P}(ZX).$$

Thus the extreme points are the active dual scenarios. See Aliprantis and Border (2006) for Bauer and Krein-Milman.

The same idea extends to the convex case. For fixed  $X$ , the map

$$(Z, t) \mapsto \mathbf{P}(ZX) - t$$

is affine and continuous on  $L^1 \times \mathbb{R}$  for the product topology  $\sigma(L^1, L^\infty) \times$  the usual topology on  $\mathbb{R}$ . Hence, whenever the relevant maximizing slice of  $\text{epi}(\rho^*)$  is compact and the supremum is attained, Bauer's principle again shows that the maximum occurs at an extreme point of the epigraph. So the convex case has the same geometric flavor, except that the primitive object is now the epigraph of the penalty rather than the dual domain alone.

Compactness does not come for free. For example, if  $\rho(X) = \text{ess sup } X$ , then

$$\rho(X) = \sup_{Z \in D} \mathbf{P}(ZX), \quad D = \{ Z \in L^1 : Z \geq 0, \mathbf{P}Z = 1 \},$$

but  $D$  is not  $\sigma(L^1, L^\infty)$ -compact. Correspondingly, the supremum need not be attained unless  $X$  reaches its essential supremum on a set of positive probability.

A second organizing principle concerns law invariance and the symmetry it imposes on the dual geometry. On an atomless space, law invariance says that  $\rho$  depends only on the distribution of  $X$ , not on the labels of states. In particular, law invariance implies

$$\rho(X \circ T) = \rho(X)$$

for every measure-preserving transformation  $T : \Omega \rightarrow \Omega$ .

The corresponding symmetry on the dual side is most naturally expressed in terms of measures rather than densities. Let  $\mathbf{Q} \ll \mathbf{P}$  be a probability measure with Radon–Nikodym derivative

$$Z = \frac{d\mathbf{Q}}{d\mathbf{P}}.$$

For a measure-preserving transformation  $T$ , define the push-forward measure in the usual way:

$$T_{\#}\mathbf{Q}(A) := \mathbf{Q}(T^{-1}(A)).$$

Then

$$\langle X \circ T, \mathbf{Q} \rangle = \langle X, T_{\#}\mathbf{Q} \rangle.$$

If  $T$  is invertible and measure preserving, then  $T_{\#}\mathbf{Q}$  has density  $Z \circ T^{-1}$  with respect to  $\mathbf{P}$ . Thus the density picture may be viewed as a rearrangement of  $Z$ , but the measure formulation is more general.

The next proposition makes the link between primal and dual symmetry explicit.

**Proposition 3.7.** *Assume the underlying probability space is atomless, and let  $\rho : L^\infty \rightarrow \mathbb{R}$  be a convex monetary risk measure with the Fatou property. Then  $\rho$  is law invariant if and only if  $\rho^*$  is law invariant, in the sense that*

$$\frac{d\mathbf{Q}_1}{d\mathbf{P}} \stackrel{d}{=} \frac{d\mathbf{Q}_2}{d\mathbf{P}} \implies \rho^*(\mathbf{Q}_1) = \rho^*(\mathbf{Q}_2).$$

*Proof.* We identify  $\mathbf{Q} \ll \mathbf{P}$  with its density  $Z = d\mathbf{Q}/d\mathbf{P}$ .

Assume first that  $\rho$  is law invariant, and let  $Z, Z' \in L^1$  satisfy  $Z \stackrel{d}{=} Z'$ . To show  $\rho^*(Z) = \rho^*(Z')$ , fix  $X \in L^\infty$ . Since the space is atomless, any coupling of  $X$  and  $Z$  may be realized with  $Z'$  in place of  $Z$ : there exists  $\hat{X} \in L^\infty$  such that

$$(\hat{X}, Z') \stackrel{d}{=} (X, Z).$$

Therefore

$$\mathbb{P}(\hat{X}Z') = \mathbb{P}(XZ).$$

Also  $\hat{X} \stackrel{d}{=} X$ , so law invariance gives  $\rho(\hat{X}) = \rho(X)$ . Hence

$$\mathbb{P}(XZ) - \rho(X) = \mathbb{P}(\hat{X}Z') - \rho(\hat{X}) \leq \rho^*(Z').$$

Taking the supremum over  $X$  yields  $\rho^*(Z) \leq \rho^*(Z')$ , and symmetry gives equality.

Conversely, assume  $\rho^*$  is law invariant. Let  $X, X' \in L^\infty$  satisfy  $X \stackrel{d}{=} X'$ . Fix  $Z \in L^1$ . Since the space is atomless and  $X \stackrel{d}{=} X'$ , there exists  $Z' \in L^1$  such that

$$(X', Z') \stackrel{d}{=} (X, Z).$$

Then

$$\mathbb{P}(X'Z') = \mathbb{P}(XZ),$$

and  $Z' \stackrel{d}{=} Z$ , so by law invariance of  $\rho^*$ ,

$$\rho^*(Z') = \rho^*(Z).$$

Therefore

$$\mathbb{P}(XZ) - \rho^*(Z) = \mathbb{P}(X'Z') - \rho^*(Z') \leq \rho(X').$$

Taking the supremum over  $Z$  gives  $\rho(X) \leq \rho(X')$ . By symmetry,  $\rho(X') \leq \rho(X)$ , and therefore  $\rho(X) = \rho(X')$ .  $\square$

This proposition lets us describe law-invariant dual geometry in terms of orbits under measure-preserving rearrangements. For a dual probability measure  $\mathbb{Q} \ll \mathbb{P}$ , define its orbit by

$$[\mathbb{Q}] := \{ T_{\#} \mathbb{Q} : T \text{ measure preserving} \}.$$

At the density level, these correspond to rearrangements of  $Z = d\mathbb{Q}/d\mathbb{P}$ . A law-invariant coherent dual set is built from the closed convex hull of such orbits; in the convex case, the same applies to the epigraph, with the penalty constant along each orbit.

This viewpoint leads to a useful informal picture. A single dual measure gives a linear expectation. One orbit, together with its law-invariant closed convex hull, gives the spectral/comonotonic class. Several distinct orbits are needed for law-invariant coherent examples that are not comonotonic additive. Finally, adding a nontrivial penalty on top of the dual geometry produces convex, non-coherent examples. This is the geometric intuition behind Kusuoka-type representations and their refinements. These ideas are developed in Shapiro (2012).

Figure 3 provides a schematic of the geometry for the dual representations of monetary risk measures. It highlights the interplay between their algebraic structure and the shape of their penalty function epigraphs. The horizontal axis represents decreasing restrictions on the measure's algebraic properties: moving from left to right, the structure is systematically relaxed from purely linear (additive), to comonotonic additive, to sub-additive (coherent), and finally to general convex. This algebraic relaxation dictates a corresponding decrease in restrictions (increase in flexibility) on the dual set, which expands from a single fixed point (linear), to a structured capacity core (comonotonic), to an arbitrary flat base domain (coherent), and ultimately into a volumetric epigraph permitting arbitrary penalty heights (convex). The vertical axis illustrates the additional symmetry imposed by law invariance. The top row depicts generic, asymmetric dual sets, and the bottom demonstrates how law invariance forces these spaces into symmetric configurations with respect to the reference measure  $\mathbb{P}$ . By Ryff-type rearrangement results, this symmetrization is represented by passing from individual dual elements to the closed convex hull of their law-invariant orbits under measure-preserving transformations. Notably, this visualizes how symmetrizing the rigid polytope of a general comonotonic capacity restricts it to a distortion capacity, generating the characteristic symmetric orbit of a spectral risk measure.

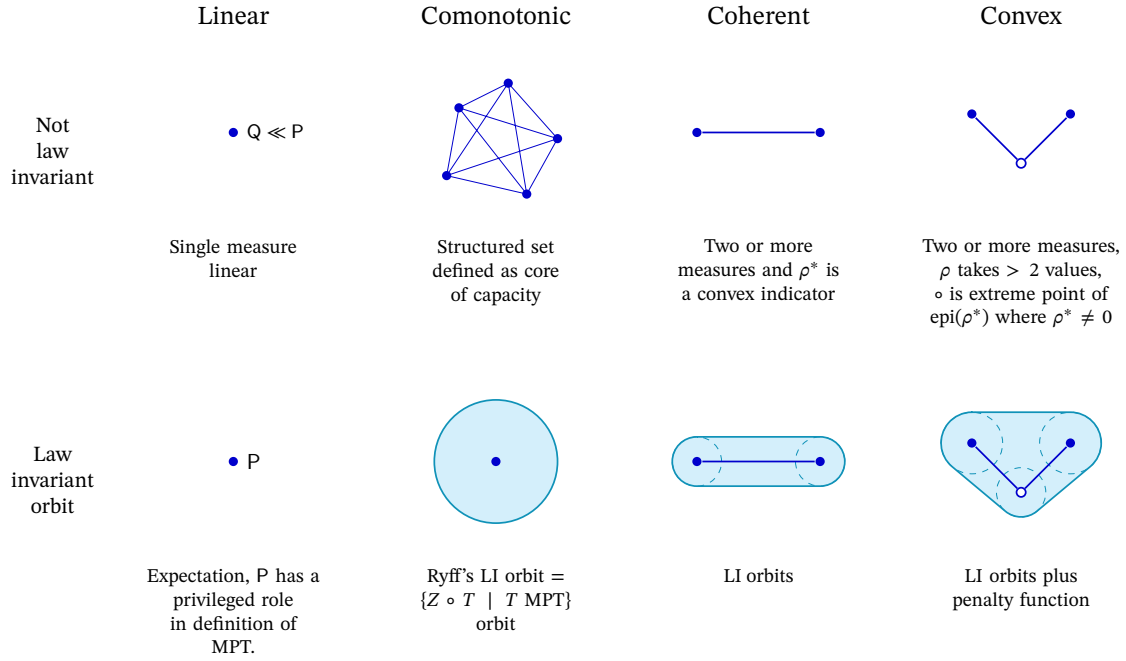


Figure 3: Geometric representations of penalty function epigraphs,  $\text{epi}(\rho^*)$ , across risk measure classes.

Finally, Figure 4 is a schematic of the topological structure of  $ba = (L^\infty)^*$ , the Banach dual of  $L^\infty$ , consisting of bounded finitely additive measures. The outer grey boundary represents all of  $ba$ . Nestled within it is the white region,  $ba_{1,+}$ , representing the convex, weak\* compact subset of normalized, positive finitely additive probability measures which is the generalized dual domain for coherent and convex risk measures defined on  $L^\infty$ .

The defining visual feature of the spaces is the stylized  $H$ -tree mesh—cyan in the general vector space ( $L^1$ ) and green in the positive probability cone ( $L^1_{1,+}$ ). This mesh represents the subspace of countably additive measures, which can be identified with standard probability densities via the Radon-Nikodym theorem. The skeletal nature is deliberate: it visualizes Goldstine's Theorem. By Goldstine, the unit ball of  $L^1$  is weak\* dense in the unit ball of  $ba$ ; therefore, the mesh must reach into every neighborhood of the larger space. However, because  $L^1$  is a proper subspace with an empty interior in the norm topology of  $ba$ , it cannot be drawn as a solid volume. It is everywhere in the weak\* sense, yet volumetrically nowhere in the norm sense, leaving infinite porous gaps (the white and grey voids) where non-countably additive measures to reside.

The magnified call out to the right illustrates the mechanics of the Yosida-Hewitt decomposition and the counter-intuitive reality of sequences in the weak\* topology. The thick green branches represent the local fibers of the  $L^1_{1,+}$  mesh. A sequence of perfectly well-behaved, countably additive density functions ( $Z_n$ ) exists strictly on these fibers. The sequence converges in the weak\* topology—meaning  $\int X Z_n dP$  converges for every bounded payoff  $X \in L^\infty$ . However, because  $L^1$  is not weak\* closed, the limit of this sequence is not guaranteed to remain a standard density. The red dashed path tracks the sequence as its limit jumps the tracks off the countably additive fibers and into the white void, converging to the red point  $\mu_{pfa}$ , a purely finitely additive measure. These objects carry no standard density and assign zero probability to any countably generated event, instead concentrating their mass on sequences of sets whose standard probabilities vanish. In the context of risk measurement, these generalized measures act analogously to Dirac deltas at the essential supremum of a random variable. The diagram highlights why restricting dual representations to standard  $L^1$  densities is insufficient: sequences of worst-case scenarios (e.g., for ess sup) inevitably push the defining risk weights off the  $L^1$  mesh and into  $ba$ .

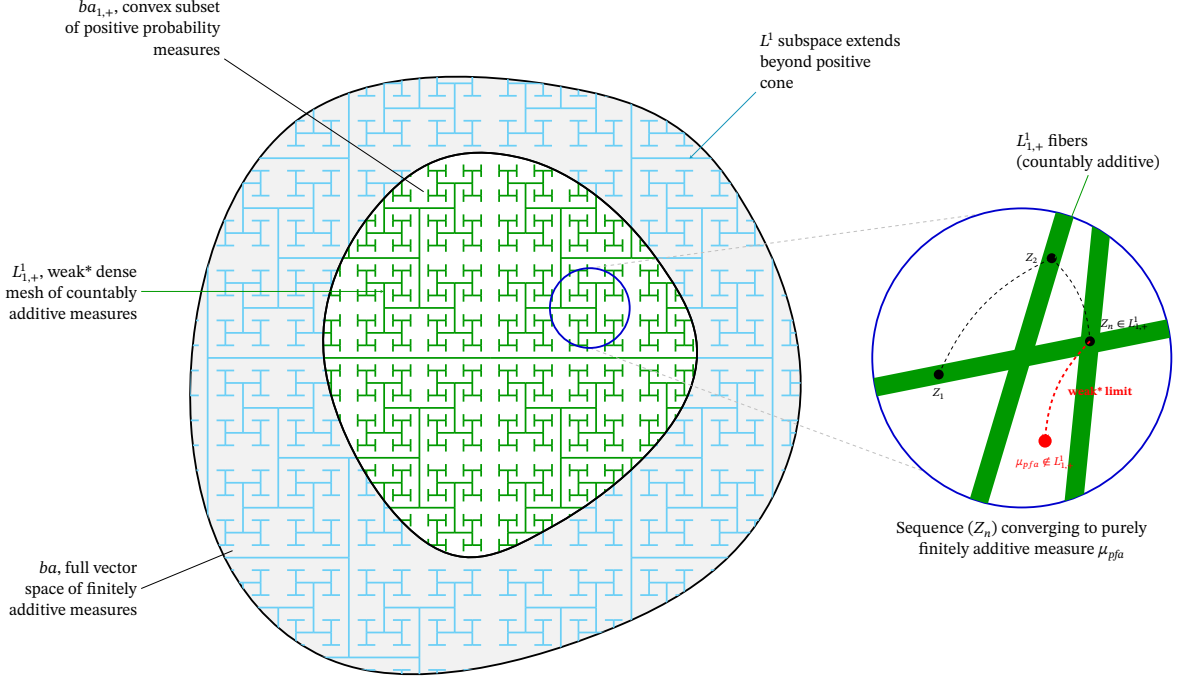


Figure 4: Topology of the dual space  $ba = (L^\infty)^*$ . The green sub-region shows the probability measures (positive, total measure 1) within  $ba$ .

## 4 Four Deeper Results

### 4.1 Law Invariant Risk Measures Preserve Second Order Stochastic Dominance

**Definition 4.1.** For  $X, Y \in L^\infty$ :

1.  $X \preceq_{cx} Y$  means that  $X$  is dominated by  $Y$  in **convex order**, that is,

$$P(\phi(X)) \leq P(\phi(Y)) \quad \text{for every convex } \phi : \mathbb{R} \rightarrow \mathbb{R}$$

whenever the expectations exist. On  $L^\infty$  they always exist for bounded convex  $\phi$ , and this is equivalent to the usual formulation with all convex  $\phi$ .

2.  $X \preceq_{ssd} Y$  means that  $X$  is dominated by  $Y$  in **second-order stochastic dominance**, equivalently increasing convex order for losses, if

$$P(\phi(X)) \leq P(\phi(Y)) \quad \text{for every increasing convex } \phi : \mathbb{R} \rightarrow \mathbb{R}.$$

Thus  $Y$  is the riskier loss.

*Remark 4.1.* In the definition of convex order we can take  $\phi(x) = x$  and  $\phi(x) = -x$  to deduce that  $X \preceq_{cx} Y$  implies  $X$  and  $Y$  have the same means. As a result, if  $PX = PY$ , then

$$X \preceq_{ssd} Y \iff X \preceq_{cx} Y.$$

On an atomless probability space, convex order admits two standard characterizations. Ryff's theorem says that  $X \preceq_{cx} Y$  iff  $X$  belongs to the closed convex hull of the equidistribution class of  $Y$  Ryff (1967). Strassen's theorem says that  $X \preceq_{ssd} Y$  iff one can couple versions  $(\tilde{X}, \tilde{Y})$  so that  $\tilde{X} \leq P(\tilde{Y} | \tilde{X})$  a.s. In the equal-mean case this condition becomes the martingale characterization of convex order,  $\tilde{X} = P(\tilde{Y} | \tilde{X})$  a.s. Strassen (1965).

**Theorem 4.1.** Assume the underlying probability space is standard and atomless. Let

$$\rho : L^\infty \rightarrow \mathbb{R}$$

be a monetary convex risk measure. Then  $\rho$  is law invariant if and only if it respects second-order stochastic dominance:

$$X \preceq_{ssd} Y \implies \rho(X) \leq \rho(Y).$$

This result is the risk-measure version of Svindland's characterization of convex lower-semicontinuous functionals by law invariance and convex-order monotonicity [Svindland \(2014\)](#). Since a monetary risk measure on  $L^\infty$  is Lipschitz, lower semicontinuity is automatic, so his theorem applies directly.

*Remark 4.2.* It is a consequence of [Theorem 4.1](#) that law invariant risk measures have the positive loading property:  $\rho(X) \geq \mathbb{P}(X)$  for all  $X$ , because expectations reduce risk and so  $\mathbb{P}(X) \preceq_{ssd} X$ . This remark justifies the implication to a positive loading in the lower left corner of [Figure 2](#).

*Proof.* Because  $\rho$  is monetary, it is  $L^\infty$ -Lipschitz and hence norm continuous.

First assume  $\rho$  is law invariant. We show it respects SSD. Suppose first that  $X \preceq_{cx} Y$ . By Ryff's theorem,  $X$  lies in the closed convex hull of the orbit

$$M(Y) := \left\{ Z \in L^\infty : Z \stackrel{d}{=} Y \right\}.$$

Since  $\rho$  is law invariant, it is constant on  $M(Y)$ :  $\rho(Z) = \rho(Y)$  for all  $Z \in M(Y)$ . By convexity, the same upper bound holds on the convex hull:

$$\rho\left(\sum_{i=1}^n \lambda_i Z_i\right) \leq \sum_{i=1}^n \lambda_i \rho(Z_i) = \rho(Y).$$

By continuity, the inequality extends to the closure. Hence

$$X \preceq_{cx} Y \implies \rho(X) \leq \rho(Y).$$

Now suppose  $X \preceq_{ssd} Y$ . By Strassen's theorem, there exist copies  $(\tilde{X}, \tilde{Y})$  on a common probability space such that  $\tilde{X} \leq \mathbb{P}(\tilde{Y} | \tilde{X})$  a.s. By monotonicity  $\rho(X) = \rho(\tilde{X}) \leq \rho(\mathbb{P}(\tilde{Y} | \tilde{X}))$ .

We next claim that  $\mathbb{P}(\tilde{Y} | \tilde{X}) \preceq_{cx} \tilde{Y}$ . To see this, let  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  be any convex function. Jensen's inequality gives  $\phi(\mathbb{P}(\tilde{Y} | \tilde{X})) \leq \mathbb{P}(\phi(\tilde{Y}) | \tilde{X})$  a.s. Taking expectations,  $\mathbb{P}\phi(\mathbb{P}(\tilde{Y} | \tilde{X})) \leq \mathbb{P}\phi(\tilde{Y})$ . Since this holds for every convex  $\phi$ , we obtain  $\mathbb{P}(\tilde{Y} | \tilde{X}) \preceq_{cx} \tilde{Y}$ .

By the first step of the proof, convex order dominance implies  $\rho(\mathbb{P}(\tilde{Y} | \tilde{X})) \leq \rho(\tilde{Y}) = \rho(Y)$ . Combining the two inequalities yields  $\rho(X) \leq \rho(Y)$ . Therefore  $\rho$  respects second-order stochastic dominance.

Conversely, if  $\rho$  respects SSD and  $X \stackrel{d}{=} Y$ , then both  $X \preceq_{ssd} Y$  and  $Y \preceq_{ssd} X$  hold trivially, since the two variables have the same law. Hence  $\rho(X) \leq \rho(Y)$  and  $\rho(Y) \leq \rho(X)$ , so  $\rho(X) = \rho(Y)$ . Thus  $\rho$  is law invariant.  $\square$

## 4.2 Schmeidler's Characterization of Comonotonic Additive Functionals

This section outlines the theory developed in [Schmeidler \(1986\)](#). The corresponding economic ideas are expounded in [Schmeidler \(1989\)](#).

**Definition 4.2.** A **capacity** on  $(\Omega, \mathcal{F})$  is a set function

$$\nu : \mathcal{F} \rightarrow [0, 1]$$

such that

$$\nu(\emptyset) = 0, \quad \nu(\Omega) = 1,$$

and

$$A \subseteq B \implies \nu(A) \leq \nu(B).$$

Thus a capacity is a normalized monotone set function; additivity is not assumed. If

$$\nu(A) = \mathbb{P}(A)$$

for all  $A \in \mathcal{F}$ , then  $\nu$  is an ordinary probability measure. More generally, if

$$\nu(A) = g(\mathbb{P}(A))$$

for some distortion  $g : [0, 1] \rightarrow [0, 1]$ , then  $\nu$  is a distorted probability capacity.

*Remark 4.3.* The terminology capacity, non-additive probability, and fuzzy measure are often used for closely related notions. Schmeidler's theorem relies on monotonicity and normalization, not additivity.

**Definition 4.3.** Let  $\nu$  be a capacity, and let  $X \in L^\infty$  be nonnegative. The **Choquet integral** of  $X$  with respect to  $\nu$  is

$$\int X d\nu := \int_0^\infty \nu \{X \geq t\} dt.$$

Since  $X \in L^\infty$ , the integral is finite, and one may equally write

$$\int X d\nu = \int_0^{\|X\|_\infty} \nu \{X \geq t\} dt.$$

For a general bounded random variable  $X \in L^\infty$ , the Choquet integral is defined by

$$\int X d\nu := \int_0^\infty \nu \{X \geq t\} dt + \int_{-\infty}^0 (\nu \{X \geq t\} - 1) dt.$$

Equivalently, writing  $M := \|X\|_\infty$ ,

$$\int X d\nu = \int_0^M \nu \{X \geq t\} dt + \int_{-M}^0 (\nu \{X \geq t\} - 1) dt.$$

This is the usual bounded-variable form of the Choquet integral.

*Remark 4.4.* Given a capacity  $\nu$ , define its dual capacity by

$$\check{\nu}(A) := 1 - \nu(A^c).$$

It is easy to check  $\check{\nu}$  is a capacity, and that the notation is consistent with the dual of a distortion when  $\nu(A) = g(\mathbb{P}(A))$  for a distortion function  $g$ . Using the dual capacity, for bounded  $X$ , we can write the Choquet integral in a pleasantly symmetric form:

$$\int X d\nu = - \int_{-\infty}^0 \check{\nu} \{X \leq x\} dx + \int_0^\infty \nu \{X > x\} dx.$$

Equivalently, if  $X = X^+ - X^-$  is the difference of two positive functions,

$$\int X d\nu = - \int X^- d\check{\nu} + \int X^+ d\nu.$$

Thus the negative part of the Choquet integral is naturally governed by the dual capacity, exactly paralleling the use of the dual distortion in distortion risk theory.

*Remark 4.5.* If  $\nu$  is sigma additive, that is, a probability measure, then the Choquet integral reduces to the ordinary Lebesgue integral:

$$\int X d\nu = \nu(X).$$

Thus the Choquet integral extends expectation from additive to non-additive set functions.

**Theorem 4.2** (Schmeidler's representation theorem). *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and*

$$\rho : L^\infty \rightarrow \mathbb{R}$$

*be a monetary, normalized, and comonotonic additive functional. Then there exists a unique capacity*

$$\nu : \mathcal{F} \rightarrow [0, 1]$$

such that

$$\rho(X) = \int X d\nu \quad \text{for all } X \in L^\infty.$$

Conversely, for every capacity  $\nu$ , the Choquet integral

$$X \mapsto \int X d\nu$$

is monetary, normalized, comonotonic additive, and sup-norm continuous.

*Remark 4.6.* In the monetary setting, the earlier basic results show that comonotonic additivity plus monetary already force positive homogeneity, so the theorem identifies the entire comonotonic branch with Choquet integration against capacities. The law-invariant sub-branch corresponds to capacities of the special form

$$\nu(A) = g(\mathbb{P}(A)).$$

*Remark 4.7.* The Choquet integral is continuous in the sup norm by Proposition 2.1.

*Proof.* The first step is to recover a candidate capacity from indicators. To that end, define

$$\nu(A) := \rho(1_A) = \rho(A), \quad A \in \mathcal{F},$$

where here and going-forward we identify a set with its indicator function. Monetary (monotone and cash invariant) and normalized imply  $\nu(\emptyset) = 0$ ,  $\nu(\Omega) = \rho(1)$ , and  $A \subseteq B \implies \nu(A) \leq \nu(B)$ . After normalizing constants, this gives a capacity candidate. Uniqueness is immediate, because any representing capacity must satisfy

$$\nu(A) = \int A d\nu = \rho(A).$$

The second step is to prove the formula for nested simple functions. Consider a simple function written in decreasing chain form

$$X = \sum_{k=1}^n a_k A_k, \quad a_k \geq 0, \quad A_1 \supseteq A_2 \supseteq \dots \supseteq A_n.$$

The summands are pairwise comonotone because the sets are nested (the pairwise values of indicators on nested sets are  $(0, 0)$ ,  $(1, 0)$  and  $(1, 1)$ ). Hence comonotonic additivity gives

$$\rho(X) = \sum_{k=1}^n a_k \rho(A_k) = \sum_{k=1}^n a_k \nu(A_k).$$

But this is exactly the Choquet integral of such a nested simple function. Equivalently, if one writes

$$X = \sum_{k=1}^n (x_k - x_{k-1}) \{ X \geq x_k \}$$

with

$$0 = x_0 \leq x_1 \leq \dots \leq x_n,$$

then

$$\rho(X) = \sum_{k=1}^n (x_k - x_{k-1}) \nu \{ X \geq x_k \} = \int X d\nu.$$

The third step extends from nonnegative simple functions to bounded nonnegative  $X$ . Approximate  $X \geq 0$  uniformly by nested simple functions  $X_n$  built from its level sets. Since  $\rho$  is sup-norm continuous,

$$\rho(X_n) \rightarrow \rho(X).$$

The Choquet integral is also sup-norm continuous on bounded functions, and for each  $n$  we already know

$$\rho(X_n) = \int X_n d\nu.$$

Passing to the limit yields

$$\rho(X) = \int X d\nu \quad \text{for all } X \geq 0.$$

Finally, extend to general bounded  $X$ , use the standard signed Choquet formula. This gives the representation for all bounded  $X$  and proves existence and uniqueness.

The converse direction is easier. For any capacity  $\nu$ : Monotonicity of  $X \mapsto \int X d\nu$  follows from monotonicity of the level sets. Comonotonic additivity is immediate on nested simple functions and then extends by uniform approximation. And sup-norm continuity follows because changing  $X$  by at most  $\varepsilon$  shifts all level sets by at most  $\varepsilon$ .  $\square$

*Remark 4.8.* The theorem involves several key assumptions and ideas working together. Comonotonicity lets us decompose a function along its own nested level sets. Nested indicators play the role that disjoint indicators play for ordinary integration. The values of the functional on indicators define the representing set function. Comonotonic additivity then forces the nested-simple-function formula, which is exactly the Choquet integral. The **characterization of SRMs** uses the same flow.

### 4.3 Law Invariance Implies Fatou

**Theorem 4.3** (Jouini, Schachermayer, and Touzi (2006), Svindland (2010)). *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a nonatomic probability space, and let*

$$\rho : L^\infty \rightarrow \mathbb{R}$$

*be a finite-valued, law invariant, convex, monetary risk measure. Then  $\rho$  has the Fatou property.*

*Remark 4.9.* Recall that by Proposition 3.4, the Fatou property is equivalent to lower semicontinuity for the  $\sigma(L^\infty, L^1)$  topology for convex monetary risk measures on  $L^\infty$ .

*Remark 4.10.* The functional  $\rho$  is **quasiconvex** if

$$\rho(\lambda x + (1 - \lambda)y) \leq \max\{\rho(x), \rho(y)\}, \quad 0 \leq \lambda \leq 1,$$

or, equivalently, if every sublevel set

$$\{x : \rho(x) \leq m\}$$

is convex. Every convex functional is quasiconvex, but not conversely. For example, any monotone function is quasiconvex because its sub-level sets are intervals or rays but most are not convex. Quasiconvexity expresses a weak preference for diversification: mixing two positions never produces risk larger than the worse of the two endpoints.

Svindland's argument is really about convex law-invariant sublevel sets, so quasiconvexity is enough. He shows that every law-invariant norm-lower-semicontinuous quasiconvex function on  $L^\infty$  is lower semicontinuous for every  $\sigma(L^\infty, L^q)$  topology,  $1 \leq q < \infty$ .

A set  $C \subseteq L^\infty$  is law invariant if  $X \in C$  and  $Y \stackrel{d}{=} X$ , then  $Y \in C$ . The proof of the theorem relies on the next proposition, which in turn relies on the lemma.

**Proposition 4.1.** *Let  $C \subseteq L^\infty$  be convex, law invariant, and closed for the norm topology of  $L^\infty$ . Then  $C$  is closed for every  $\sigma(L^\infty, L^q)$  topology,  $1 \leq q < \infty$ .*

**Lemma 4.1.** *If  $C \subseteq L^\infty$  is convex, law invariant, and norm closed, then for every sub- $\sigma$ -algebra  $\mathcal{A} \subseteq \mathcal{F}$  and every  $X \in C$ , we have  $\mathbb{P}(X \mid \mathcal{A}) \in C$ .*

*Proof.* Here is a sketch of Svindland's proof.

Step 1: constants. First prove that

$$\mathbb{P}X \in C$$

for every  $X \in C$ . Take  $\varepsilon > 0$ . Write  $q_X$  for the quantile function of  $X$ . Partition  $(0, 1)$  into  $n$  equal intervals, with  $n$  large enough that the oscillation of  $q_X$  on each interval is at most  $\varepsilon$ . Partition  $\Omega$  into sets  $B_1, \dots, B_n$  of probability  $1/n$ . On each  $B_k$ , build a random variable uniformly distributed on one of the quantile intervals; then permute these quantile pieces over the blocks  $B_k$ . Each such rearrangement has the same law as  $X$ , hence lies in  $C$  by law invariance. Averaging over all permutations gives a random

variable  $X_n \in C$  by convexity, and by construction  $X_n$  is uniformly within  $\varepsilon$  of the constant  $PX$ . Since  $C$  is norm closed,  $PX \in C$ .

Step 2: finite conditional expectations. Now let  $\mathcal{A} = \sigma(D_1, \dots, D_r)$  for a finite partition. Restrict to each cell  $D_i$ , normalize the probability there, and apply Step 1 on that smaller nonatomic probability space. This gives approximants inside the restricted version of  $C$  converging uniformly to the conditional mean on each cell. Pasting the pieces back together yields

$$P(X | \mathcal{A}) \in C.$$

Step 3: general  $\mathcal{A}$ . Approximate  $\mathcal{A}$  by an increasing sequence of finite sub- $\sigma$ -algebras  $\mathcal{A}_n$  such that

$$\|P(X | \mathcal{A}_n) - P(X | \mathcal{A})\|_\infty \rightarrow 0.$$

Since each  $P(X | \mathcal{A}_n) \in C$  and  $C$  is norm closed, one gets

$$P(X | \mathcal{A}) \in C.$$

□

*Remark 4.11.* A useful general fact is that closed convex sets do not depend on the choice of compatible locally convex topology. In particular, a norm-closed convex set in a Banach space is closed for every compatible locally convex topology; in particular weakly or weak-\* closed, as appropriate. This fact follows using a standard separation argument. Let  $C$  be convex and norm closed, and let  $x \notin C$ . By the Hahn-Banach separation theorem, there exists a continuous linear functional  $\mu$  and a real number  $a$  such that

$$\mu(x) > a \geq \mu(y) \quad \text{for all } y \in C.$$

Now  $\mu$  is not only norm continuous; it is also weakly continuous. Therefore

$$U := \{z : \mu(z) > a\}$$

is weakly open. It contains  $x$  and is disjoint from  $C$ . So every point outside  $C$  has a weakly open neighborhood disjoint from  $C$ , which means that the complement of  $C$  is weakly open. Hence  $C$  is weakly closed.

We can now prove Proposition 4.1.

*Proof.* Take a net  $(X_i)$  in  $C$  converging to  $X$  in  $\sigma(L^\infty, L^q)$ . Fix a finite sub- $\sigma$ -algebra  $\mathcal{G}$ . Then conditional expectation onto  $\mathcal{G}$  turns every continuous linear functional on the finite-dimensional space  $L^\infty(\mathcal{G})$  into an  $L^\infty$  density. Hence

$$P(X_i | \mathcal{G}) \rightarrow P(X | \mathcal{G})$$

in the weak-\* topology  $\sigma(L^\infty, L^{\infty*})$ . By the previous lemma,

$$P(X_i | \mathcal{G}) \in C \quad \text{for all } i.$$

Because a norm-closed convex set is also weak-\* closed (Remark 4.11), it follows that

$$P(X | \mathcal{G}) \in C.$$

Now choose finite  $\mathcal{G}_n$  with

$$\|P(X | \mathcal{G}_n) - X\|_\infty \rightarrow 0.$$

Each conditional expectation belongs to  $C$ , and  $C$  is norm closed, so  $X \in C$ . Thus  $C$  is  $\sigma(L^\infty, L^q)$ -closed. □

Finally, we can prove Theorem 4.3.

*Proof.* Apply Proposition 4.1 to the sublevel sets

$$C_m := \{X \in L^\infty : \rho(X) \leq m\}.$$

They are convex by convexity of  $\rho$  and law invariant by law invariance of  $\rho$ . Since  $\rho$  is monetary, it is  $L^\infty$ -Lipschitz, hence norm continuous, and so each  $C_m$  is norm closed. Therefore, each  $C_m$  is  $\sigma(L^\infty, L^1)$ -closed, so  $\rho$  is  $\sigma(L^\infty, L^1)$ -lower semicontinuous and hence has the Fatou property. □

*Remark 4.12.* The original proof that law invariance implies Fatou was Jouini, Schachermayer, and Touzi (2006). They work on standard nonatomic spaces and use measure-preserving transformations. Svindland (2010) shows that standardness is unnecessary; on a general nonatomic space one may replace the transformation argument by a more hands-on quantile-based rearrangement construction.

*Remark 4.13.* The same circle of ideas also connects law invariance with dilatation monotonicity. A functional  $\rho$  is called **dilatation monotone** if

$$\rho(\mathbb{P}(X \mid \mathcal{G})) \leq \rho(X) \quad \text{for every sub-}\sigma\text{-field } \mathcal{G} \subseteq \mathcal{F}.$$

This condition says that averaging out information cannot increase risk.

On an atomless probability space, a convex risk measure is law invariant if and only if it is dilatation monotone. One direction is immediate from Theorem 4.1: if  $\rho$  is law invariant and  $Y = \mathbb{P}(X \mid \mathcal{G})$ , then  $Y$  is less risky than  $X$  in the sense of second-order stochastic dominance, so  $\rho(Y) \leq \rho(X)$ . Thus law invariance implies SSD-monotonicity implies dilatation monotonicity.

The converse is deeper and is due to Cherny and Grigoriev (2007). Their argument shows that on an atomless space, dilatation monotonicity already forces law invariance for convex risk measures. The key construction uses conditional expectations along suitable finite partitions, and then passes through a tower of conditional expectations that approximates an equidistributed rearrangement. In this way, one compares random variables with the same law by repeatedly averaging over finer and finer  $\sigma$ -fields.

#### 4.4 Kusuoka's Theorem: Law-Invariant Convex Risk Measures are Built from TVaR

Kusuoka's theorem is the representation result for the lower convex-right corner of the Atlas. It says that, on an atomless space, every law-invariant convex risk measure with suitable continuity can be built from Tail Value at Risk. Thus Schmeidler explains the comonotonic column, while Kusuoka explains much of the law-invariant convex and coherent region.  $\mathcal{M}_1([0, 1])$  denotes the set of probability measures on  $[0, 1]$ . The omitted endpoint corresponds to *ess sup*.

**Theorem 4.4** (Kusuoka's representation theorem). *Assume  $(\Omega, \mathcal{F}, P)$  is atomless, and let  $\rho : L^\infty \rightarrow \mathbb{R}$  be a law-invariant convex monetary risk measure with the Fatou property. Then there exists a proper convex lower semicontinuous penalty*

$$\beta : \mathcal{M}_1([0, 1]) \rightarrow (-\infty, \infty]$$

such that

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1([0, 1])} \left\{ \int_{[0, 1]} \text{TVaR}_\alpha(X) \mu(d\alpha) - \beta(\mu) \right\}.$$

Equivalently, every such  $\rho$  is the supremum of penalized spectral risk measures.

*Remark.* The theorem says that the building blocks for law-invariant convex risk measures are not arbitrary scenarios on the underlying state space, but tail averages indexed by confidence level. This is a much sharper statement than the general  $L^1$  dual representation. Law invariance collapses the dual description from arbitrary countably additive measures  $Q \ll P$  to a one-dimensional family of tail functionals, mixed by probability measures  $\mu$  on  $(0, 1]$  and penalized by  $\beta$ .

Thus the law-invariant convex part of the Atlas is generated from TVaR in the same way that the general convex part is generated from affine functionals. The representation is especially natural in insurance language: a law-invariant convex risk measure is assembled from return-period views of the loss distribution, rather than from named statewise scenarios.

*Remark.* For each probability measure  $\mu$  on  $(0, 1]$ , the functional

$$\rho_\mu(X) := \int_{(0, 1]} \text{TVaR}_\alpha(X) \mu(d\alpha)$$

is a spectral risk measure. Spectral measures are coherent, law invariant, and continuous from above. They are also distortion risk measures: there is a concave distortion function  $g_\mu$  such that

$$\rho_\mu(X) = \int_0^1 q_X(u) d\check{g}_\mu(u),$$

where  $\check{g}_\mu$  is the dual distortion. Hence Kusuoka's theorem can be read as saying that law-invariant convex risk measures are obtained by taking penalized suprema of distortion-type building blocks.

This makes precise the relationship between the convex and comonotonic columns. A single distortion produces a comonotonic additive risk measure, as Schmeidler's theorem predicts. A supremum over many distortions usually destroys comonotonic additivity while retaining convexity and law invariance.

*Remark.* In the coherent case the penalty collapses to an indicator function. Then there is a closed convex set  $\mathcal{K} \subseteq \mathcal{M}_1((0, 1])$  such that

$$\rho(X) = \sup_{\mu \in \mathcal{K}} \int_{(0,1]} \text{TVaR}_\alpha(X) \mu(d\alpha).$$

Thus every law-invariant coherent risk measure with the Fatou property is the supremum of a family of spectral risk measures.

The comonotonic additive case is the extreme special case where only one distortion is needed. More precisely, a law-invariant coherent risk measure is comonotonic additive if and only if its Kusuoka representation can be taken with a single measure  $\mu$ , equivalently with a single spectral, or distortion, risk measure. This is exactly the point where the coherent column meets the comonotonic column in the bottom row of the Atlas.

*Proof.* (Sketch.) The proof is deeper than the general dual representation because law invariance must be turned into a statement about quantiles rather than states. The key steps are as follows.

First, by the law-invariance implies Fatou result already discussed,  $\rho$  has an  $L^1$  dual representation. Because  $\rho$  depends only on the distribution of  $X$ , one can average dual variables over measure-preserving transformations and pass from arbitrary dual densities to rearrangement-invariant ones. That reduces the dual problem to objects determined by the quantile structure of  $X$ .

Second, one identifies the extreme law-invariant coherent building blocks as spectral risk measures, equivalently mixtures of TVaR. In other words, TVaR plays for law-invariant coherent measures the role that linear functionals play in ordinary convex duality.

Third, a general law-invariant convex risk measure is recovered by allowing a penalty over the spectral building blocks. That yields the displayed supremum over  $\mu$  with penalty  $\beta$ .

The coherent case is the same argument with the penalty replaced by an indicator of an admissible set of measures  $\mathcal{K}$ . Finally, comonotonic additivity forces the representation to collapse from a supremum over many distortions to one distortion, because a supremum of distinct comonotonic additive functionals is generally only convex. Thus the single-distortion case is exactly the comonotonic additive one.  $\square$

*Remark.* Kusuoka's theorem is one reason TVaR occupies such a central place in risk theory. Value at Risk is the simplest law-invariant monetary functional, and distortion measures describe the comonotonic additive law-invariant coherent class, but TVaR is the true atomic building block for the whole law-invariant convex world. The theorem explains why one sees TVaR, weighted TVaR, spectral measures, and distortions repeatedly across the literature: they are not merely examples, but the structural components of the entire region.

## 5 Examples

It is best to read this section with the Atlas, Figure 1, at hand. The examples place familiar functionals, show that each box is distinct, and how to construct a measure in each box. We start by describing four archetypes: mean, Value at Risk, Tail Value at Risk, and the ess sup. After that, we continue exploring the Atlas from the bottom up.

### 5.1 Four Archetypes: P, VaR, TVaR, and ess sup

**Expectation** with respect to the underlying probability  $\mathbb{P}$  lies at the lower left corner of the Atlas. Expected loss is certainly an important, albeit very simple, attribute of a risk. Expectation is linear over portfolios, meaning it provides no diversification benefit at the level of expected loss. It has exemplary

continuity and limit behavior, and provides benchmark continuity behavior against which other measures can be compared.

**Value at Risk**, the grandparent risk functional, lives at the bottom right corner. VaR is risk-management speak for the lower quantile, defined for  $p \in (0, 1)$  as

$$\text{VaR}_p(X) = \inf \{ t \in \mathbb{R} : \mathbb{P}(X \leq t) \geq p \} =: q_X(p).$$

VaR is law invariant, monotone, cash invariant, and positive homogeneous. Famously, it is not subadditive and therefore not convex, by the 2/3 lemma. Since VaR is not convex, it does not admit a standard dual representation, but it still shares some structural similarities with convex measures. Because VaR is comonotonic additive, it can be represented as a Choquet integral with respect to a capacity, and since it is law invariant, that capacity is a distorted probability. The distortion function is the indicator step function

$$g(u) = \begin{cases} 1 & u \geq 1 - p \\ 0 & \text{otherwise.} \end{cases}$$

The failure of concavity in  $g$  corresponds to the failure of subadditivity.

We can ask for the smallest coherent risk measure that dominates  $\text{VaR}_p$ , i.e.,

$$\inf \{ \rho(X) : \rho \text{ coherent, } \rho \geq \text{VaR}_p \}.$$

That majorant is called **Tail Value at Risk**, and is denoted  $\text{TVaR}_p$ . Thus  $\text{TVaR}_p$  is the minimal coherent correction of  $\text{VaR}_p$ . It is defined for  $p \in [0, 1)$  as

$$\text{TVaR}_p(X) := \frac{1}{1-p} \int_p^1 q_X(u) du.$$

Obviously,  $\text{TVaR}_0(X) = \mathbb{P}X$ . TVaR is law invariant, coherent, comonotonic additive, and hence spectral. It respects SSD and has the Fatou property by the results in Section 4. Kusuoka's theorem shows that a law-invariant coherent risk measure is a supremum of mixtures of TVaRs; the special case of a single mixture is the spectral, comonotonic additive class. Thus, TVaR can be regarded as the building block of many risk measures in the Atlas.

The final archetype is the **essential supremum**, the smallest real number  $a$  such that  $X \leq a$  almost surely:

$$\text{ess sup } X := \inf \{ a \in \mathbb{R} : \mathbb{P}(X \leq a) = 1 \}.$$

(We have to use the essential supremum rather than the supremum because elements of  $L^\infty$  are equivalence classes of functions modulo differences on sets of measure zero.) By convention, we interpret  $\text{TVaR}_1(X) = \text{ess sup } X$ . Like TVaR, it is law invariant, coherent, comonotonic additive, and has the Fatou property. It is continuous from below but not from above: the decreasing sequence  $X_n = [0, 1/n]$  of indicator functions converges to  $X = 0$  a.s., but  $\text{ess sup } X_n = 1$  does not converge to  $\text{ess sup } X = 0$ . Therefore it also fails to be Lebesgue continuous. The ess sup has a dual representation over  $L^1$ , but it is a supremum, not a maximum: if  $\mathbb{P}(X = \text{ess sup } X) = 0$ , there is no density  $Z \geq 0$  with  $\mathbb{P}Z = 1$  such that  $\mathbb{P}(XZ) = \text{ess sup } X$ , although there is a sequence of such densities  $Z_n \in L^1$  with  $\mathbb{P}(XZ_n) \rightarrow \text{ess sup } X$ .

It is helpful to remember that TVaR interpolates between the mean and the maximum:

$$\mathbb{P}X = \text{TVaR}_0(X) \leq \text{TVaR}_p(X) \leq \text{ess sup } X,$$

with  $p \uparrow 1$  pushing mass toward the upper tail.

*Remark.* Although  $\text{VaR}_p$  is not convex, we can still try to analyze it in the duality framework. The biconjugate  $\text{VaR}_p^{**}$  gives the lower semicontinuous convex envelope of  $\text{VaR}_p$  relative to the chosen dual pairing and topology. For the dual pair  $(L^\infty, L^1)$  on an atomless space, however, that envelope turns out to be completely trivial: for every  $p \in (0, 1)$ , one has  $\text{VaR}_p^{**} \equiv -\infty$ . In particular, the biconjugate is not proper. This collapse shows, in a particularly vivid way, that convex duality does not meaningfully extend to non-convex risk functionals merely by taking conjugates. Here are the details.

Let  $f(X) = \text{VaR}_p(X)$ , and consider its conjugate  $f^*(Z) = \sup_X \{ \mathbb{P}(XZ) - f(X) \}$  for  $Z \in L^1$ . As usual, we need consider only  $Z$  satisfying  $Z \geq 0$  and  $\mathbb{P}Z = 1$ . Since the space is atomless, we can choose a set  $A$  with  $\mathbb{P}(A) < 1 - p$  and  $\mathbb{P}(Z1_A) > 0$ . Now take  $X_n = n1_A$ . Because  $X_n = 0$  on  $A^c$ , and  $\mathbb{P}(A^c) > p$ , the  $p$ -quantile of  $X_n$  is 0, so  $\text{VaR}_p(X_n) = 0$  for all  $n$ . On the other hand,  $\mathbb{P}(X_n Z) = n\mathbb{P}(Z1_A) \rightarrow \infty$ . Hence  $f^*(Z) = +\infty$  for every  $Z \in L^1$ . Therefore  $f^{**}(X) = \sup_Z \{ \mathbb{P}(XZ) - f^*(Z) \} = -\infty$  for every  $X$ .

## 5.2 The Bottom Row: Law Invariant Measures

### 5.2.1 Spectral Risk Measures

The spectral risk measure box is one of the strictest in the Atlas in the sense that its members must satisfy so many axioms. Nevertheless, we agree with Cherny and Orlov (2011) who say “This [spectral risk measure] class is very wide and, in our opinion, is sufficient for any practical application of coherent risks.” By Kusuoka’s theorem, every spectral measure is a weighted average of TVaRs. Equivalently, it has the form

$$\rho_\phi(X) = \int_0^1 q_X(u) \phi(u) du,$$

where  $\phi \geq 0$  is increasing and

$$\phi(0+) + \int_0^1 \phi(u) du = 1 \quad \phi(0+) := \lim_{x \downarrow 0} \phi(x).$$

The function  $\phi$  governs how the measure reacts to risk across the spectrum determined by percentile level  $p$ . In distortion form we write

$$\rho_g(X) = \int X d(g \circ \mathbb{P}),$$

with  $g$  a concave distortion function. The two parameterizations are related by

$$\phi(u) = g'(1 - u) = \check{g}'(u)$$

when  $g$  is differentiable.

For TVaR, the subdifferential has a simple tail-density form. Let  $q_X(p)$  denote an  $p$ -quantile. Then

$$Z \in \partial \text{TVaR}_p(X)$$

if and only if

$$0 \leq Z \leq \frac{1}{1-p}, \quad \mathbb{P}Z = 1,$$

and, more specifically,

$$Z = 0 \text{ on } \{X < q_X(p)\}, \quad Z = \frac{1}{1-p} \text{ on } \{X > q_X(p)\},$$

with any quantile plateau  $\{X = q_X(p)\}$  used only to interpolate so that  $\mathbb{P}Z = 1$ . This is an intuitive, prudent risk measurement principle: put all probability evenly on the worst  $(1-p)$ -tail. Notice that the probability is spread evenly over this tail, so TVaR is risk neutral in the tail, [Jouini, Schachermayer, and Touzi \(2008\)](#).

Results about SRMs can often be proved by building up from TVaR. For example, if

$$\rho(X) = \int_0^1 \text{TVaR}_p(X) \mu(dp)$$

is a general spectral risk measure, then

$$\partial \rho(X) = \left\{ \int_0^1 Z_p \mu(dp) : Z_p \in \partial \text{TVaR}_p(X) \text{ } \mu\text{-a.e.} \right\}.$$

Thus, spectral subgradients are weighted averages of tail selectors.

### 5.2.2 Coherent but Not Comonotonic Additive

The next box right is obtained by taking a supremum of two or more spectral functionals, [Kusuoka \(2001\)](#). To obtain a measure that is not comonotonic the chosen functional cannot be strictly ordered, so for example,  $\rho(X) = \max\{\text{TVaR}_{0.95}(X), \text{TVaR}_{0.99}(X)\} = \text{TVaR}_{0.99}(X)$  is still comonotonic. A good general recipe is to take the maximum of spectral measures associated with different distortions, such as the proportional hazard, Wang or dual, [Mildenhall and Major \(2022, sec. 11.3\)](#). Comonotonic additivity

fails because the maximizing measure can switch between components, and that destroys additivity even for comonotone pairs of risks. See Mildenhall and Major (2022) Example 229 for a specific example. More generally, any nontrivial supremum

$$\rho(X) = \sup_{i \in I} \rho_i(X),$$

with each  $\rho_i$  spectral, lands in the bottom-row coherent box but usually outside the spectral box.

### 5.2.3 Convex but Not Coherent

Next right are convex risk measures. These handicap component risk measures with a non-trivial function. Thus, generically,

$$\rho(X) := \max \{ \rho_1(X) - c_1, \rho_1(X) - c_2 \},$$

with  $c_1 \neq c_2$  and both finite is convex but not coherent. The values of  $c$  can be chosen to reflect assumptions about the corresponding scenarios.

An archetypal bottom-row convex-but-not-coherent example is the **entropic risk measure**

$$\rho_\gamma(X) := \frac{1}{\gamma} \log \mathbb{P}[e^{\gamma X}], \quad \gamma > 0.$$

It is law invariant, normalized, monetary, and convex, but not positive homogeneous and hence not coherent. It has dual representation

$$\rho_\gamma(X) = \sup_{Z \geq 0, \mathbb{P}Z=1} \{ \mathbb{P}(XZ) - \rho_\gamma^*(X) \} \quad \rho_\gamma^*(X) := \frac{1}{\gamma} \mathbb{P}(Z \log Z).$$

*Remark 5.1.* The penalty  $\mathbb{P}(Z \log Z)$  is called relative entropy. If

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}},$$

then it is sometimes written  $D(\mathbb{Q} \parallel \mathbb{P})$ , and is called the Kullback–Leibler divergence of  $\mathbb{Q}$  from  $\mathbb{P}$ . Hence the entropic risk measure admits the robust representation

$$\rho_\gamma(X) = \sup_{\mathbb{Q} \ll \mathbb{P}} \mathbb{Q}X - \frac{1}{\gamma} D(\mathbb{Q} \parallel \mathbb{P}).$$

The penalty charges more for moving further from the reference model  $\mathbb{P}$  as measured by  $D$ .

## 5.3 Middle Row: Fatou Assumption but Not Law Invariant

In the middle row we drop the strict symmetry required by law invariance, but explicitly add the Fatou condition that came for free with law invariance. Law invariance fails when risk is evaluated against a fixed state variable or some distinguished dual element. The market or the state of the economy is a common state variable used in finance.

*Remark 5.2.* Dropping law invariance has a dramatic effect on the dual representation. Suppose

$$\rho(X) = \sup_{\substack{Z \geq 0 \\ \mathbb{P}Z=1}} \{ \mathbb{P}(XZ) - \rho^*(Z) \}$$

is a convex dual representation on  $L^\infty$ , and  $\rho$  is law invariant. Then the penalty may be chosen to be law invariant:

$$Z \stackrel{d}{=} \tilde{Z} \implies \rho^*(Z) = \rho^*(\tilde{Z}).$$

Consequently, its effective domain

$$\text{dom } \rho^* := \{ Z : \rho^*(Z) < \infty \}$$

is saturated under equidistribution:

$$Z \in \text{dom } \rho^*, \tilde{Z} \stackrel{d}{=} Z \implies \tilde{Z} \in \text{dom } \rho^*.$$

Thus every admissible test density brings with it its entire equidistribution class. In an atomless space, that means that law-invariant dual sets are necessarily huge: once one density is admitted, all of its measure-preserving rearrangements are admitted as well. In the middle row, by contrast, one may choose a few distinguished scenario measures  $Z_1, Z_2, \dots$  and build a risk measure from them. In the bottom row, law invariance forbids such preferences for particular states. Any admissible density must be accompanied by all of its rearrangements, and the penalty must be constant on that whole orbit.

The linear case, risk-adjusted expected value is discussed in Section 5.4.

### 5.3.1 Comonotonic: Choquet Integral With a Regular Non-Distortion Capacity

Fix a nontrivial measurable set  $B$  with  $0 < P(B) < 1$ , and define

$$\nu(A) := \frac{1}{2} P(A | B) + \frac{1}{2} \sqrt{P(A | B^c)}.$$

Then  $\nu$  is a normalized monotone set function, continuous from below and above on  $\mathcal{F}$ , hence a regular capacity. It is not of distortion form

$$\nu(A) \neq g(P(A))$$

in general, because it depends on how  $A$  sits relative to the distinguished set  $B$ , not only on its probability. The associated Choquet integral is coherent and comonotonic additive, but not law invariant. Any capacity obtained by distorting different regions differently gives the same phenomenon.

### 5.3.2 Coherent Not Comonotonic or Law Invariant

Just as in the bottom row, the max of two or more comonotonic measures is coherent. Take two fixed densities  $Z_1, Z_2 \in L^1$  with

$$Z_i \geq 0, \quad PZ_i = 1,$$

and suppose they are not a.s. equal. Define

$$\rho(X) := \max \{ P(XZ_1), P(XZ_2) \}.$$

Then  $\rho$  is coherent: its dual is the convex indicator function of the convex set generated by the densities

$$\mathcal{M} = \text{co} \{ Z_1, Z_2 \}.$$

It is not comonotonic additive in general for the same reasons as in the law invariant row: the active branch can change with  $X$ . It is not law invariant unless the two densities collapse to a law invariant family. The dependence on the particular states encoded by  $Z_1$  and  $Z_2$  is exactly what law invariance rules out.

### 5.3.3 Convex but Not Coherent or Law Invariant

To move from coherent to convex, we simply add a nontrivial penalty. For example,

$$\rho(X) := \sup_{i=1,2} \{ P(XZ_i) - c_i \}$$

is convex if the constants  $c_i$  are not all 0, or

$$\rho(X) := \sup_{Z \geq 0, PZ=1} \{ P(XZ) - \rho^*(Z) \},$$

where  $\rho^*$  is a nontrivial convex penalty whose effective domain depends on a fixed sub- $\sigma$ -field or on a preferred state variable. For example, we could take a weighted average of two entropic measures with different reference probabilities:

$$\rho(X) = \log \left( \lambda e^{P(XZ_1)} + (1 - \lambda) e^{P(XZ_2)} \right), \quad 0 < \lambda < 1.$$

These examples are normalized, monotone, convex and cash invariant, but not positively homogeneous and not law invariant.

## 5.4 Left-Hand Side

The left hand side of the Atlas fractures more than we have indicated, and the coherent or convex regions can strengthen in two different directions, with full additivity at the extreme right:

$$\text{coherent} \dashrightarrow \begin{cases} \text{comonotonic additive} \\ \text{independent additive} \end{cases} \dashrightarrow \text{additive/linear.}$$

Independent additivity is an intermediate linearity condition. It requires

$$\rho(X + Y) = \rho(X) + \rho(Y)$$

for independent  $X$  and  $Y$ , while allowing non-linearity for dependent sums. It rules out any diversification benefit across independent sources of risk.

**Example 5.1.** The entropic risk measure is independent additive. If  $X$  and  $Y$  are independent then

$$\begin{aligned} \rho_\gamma(X + Y) &= \frac{1}{\gamma} \log \mathbf{P}(e^{\gamma(X+Y)}) \\ &= \frac{1}{\gamma} \log(\mathbf{P}(e^{\gamma X})\mathbf{P}(e^{\gamma Y})) \\ &= \rho_\gamma(X) + \rho_\gamma(Y). \end{aligned}$$

Notice that  $\rho_\gamma$  is essentially the cumulant generating function of  $X$ . Borch (1962) points out that any independent additive functional must be a function of cumulants.

Once a functional is fully additive, the distinctions between comonotonic and independent additivity cease to matter: all diversification effects disappear, and the value of a sum is just the sum of values. In that sense the left hand column is the world of linear pricing and the functionals of efficient market theory. A traded claim is priced by expectation under a single measure,

$$\pi(X) = \mathbf{P}(X) \quad \text{or} \quad \pi(X) = \mathbf{Q}(X) = \mathbf{P}(XZ), \quad Z = \frac{d\mathbf{Q}}{d\mathbf{P}},$$

depending on whether we work with a physical, objective measure, or a risk-neutral pricing kernel measure. There is no room for ambiguity, no probability distortions, no scenario penalty, and no nonlinear premium effects for concentration or tail dependence.

The top row shows, however, that linearity alone does not force countable additivity. Additive functionals may also be expectations with respect to finitely additive measures in  $ba \setminus L^1$ . Thus the left edge contains both the familiar efficient-market case, based on a single countably additive measure, and the much weirder top-row linear examples, based on a single finitely additive one.

*Remark 5.3 (Ambiguity and Spreads).* Additivity is natural in a frictionless market with a single probabilistic model: the price of a sum is the sum of the prices. Convexity, and in particular subadditivity, can emerge once pricing reflects ambiguity or model uncertainty. If a decision maker considers a family of plausible probability measures and evaluates a loss by a worst-case or penalized worst-case expectation, then the resulting functional is the supremum of linear prices, and hence convex. In that interpretation, subadditivity reflects not only diversification but also ambiguity aversion. Whereas diversification is usually beneficial and at worst neutral, combining positions can reduce or exacerbate exposure to ambiguity, for example through model error.

The same robust perspective also suggests an alternative origin for bid-ask spreads than transaction costs in an intermediated market. In a single probability model, with a linear pricing rule, the law of one price holds. In a model with ambiguity, the seller may evaluate a liability using a worst-case upper price while the buyer uses the corresponding lower price. The gap between the two is then a nonlinear spread generated by uncertainty, rather than by transaction costs. Castagnoli, Maccheroni, and Marinacci (2004) show that for spectral functionals the existence of one frictionless risk forces the functional to be additive. Said another way, all spectral risk measures have a non-zero bid ask spread on any non-trivial risk.

## 5.5 Top Row: Non-Fatou Examples

The top row is easy to forget and is not a focus in the literature: it is usually “assumed away”. The weird examples here allow a sequence of random variables to pass a risk test  $\rho(X_n) \leq c$  individually, but to converge a.s. to a variable  $X$  that jumps unexpectedly  $\rho(X) > c$ . Dropping Fatou introduces finitely additive dual objects which make the space of possibilities vastly larger than in the countably additive world. Together, these undesirable behaviors explain why the Fatou assumption is near ubiquitous. If you are happy with that explanation and do not want to see the details you can skip the rest of this section.

*Remark 5.4.* Monetary risk measures are automatically Lipschitz for the  $L^\infty$  norm, and hence norm continuous. In particular, a convex monetary risk measure on  $L^\infty$  is norm lower semicontinuous. For convex functions on a Banach space, norm lower semicontinuity is equivalent to lower semicontinuity for the weak topology induced by the full dual. Thus every convex monetary risk measure on  $L^\infty$  is automatically  $\sigma(L^\infty, (L^\infty)^*)$ -lower semicontinuous, and therefore admits a convex dual representation with respect to the full dual  $(L^\infty)^* = ba$ .

The Fatou property is different. It is not about norm continuity or full-dual lower semicontinuity, but about lower semicontinuity for the much weaker topology  $\sigma(L^\infty, L^1)$ . Since weaker topologies have fewer open and closed sets and more convergent nets, lower semicontinuity for them is a stronger requirement. That stronger requirement does not come for free from monetary assumption. Within the convex part of the Atlas, the top row consists precisely of functionals that are norm continuous, and hence well behaved from the Banach-space point of view, but fail the additional  $\sigma(L^\infty, L^1)$  lower semicontinuity and therefore fail countably additive  $L^1$  duality.

*Remark 5.5.* There are two distinct ways in which  $L^1$  duality can fail. The first is a mild failure: there is a valid  $L^1$  representation, but no maximizer in  $L^1$ . The canonical example is

$$\rho(X) = \text{ess sup } X,$$

which has the Fatou property. As a result, we still have the countably additive dual formula

$$\text{ess sup } X = \sup_{Z \geq 0, \mathbb{P}Z=1} \mathbb{P}(XZ),$$

and the  $L^1$  representation remains correct. What fails is attainment: unless the essential supremum is attained on a set of positive probability, there is no density  $Z \in L^1$  that places all mass on the top points of  $X$ . The maximizing mass wants to concentrate where  $X$  is largest, and this may require a finitely additive dual object. Thus  $\text{ess sup}$  already points toward  $ba$ , but only at the level of  $\max$  versus  $\sup$  in the dual representation.

The second failure is more fundamental: there is no  $L^1$  representation at all. We can have

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \mu(X), \quad \mathcal{M} \subset ba,$$

with purely finitely additive  $\mu$  essential to the representation. In such cases there may be no analogous  $L^1$  formula. Here the countably additive dual does not merely fail to attain the optimum; it fails to describe the functional at all.

The functional  $\text{ess sup}$  is therefore a boundary case. It is still Fatou, and it still admits a valid  $L^1$  supremum representation, but exact dual attainment already pushes one toward finitely additive measures. Top-row non-Fatou examples go further: they require finitely additive dual elements not just for attainment, but for validity.

We now explain what these finitely additive objects in  $ba \setminus L^1$  actually look like, and how they populate the top row of the Atlas.

### 5.5.1 Ultrafilters and $\{0, 1\}$ -valued finitely additive probabilities

Top-row examples are built from ultrafilters (see Section 8.6 for an overview of filters). Let  $(\Omega, \mathcal{F})$  be a measurable space. If  $\mathcal{U} \subseteq \mathcal{F}$  is an ultrafilter, define

$$\mu_{\mathcal{U}}(A) := 1_{\mathcal{U}}(A) = \begin{cases} 1 & A \in \mathcal{U} \\ 0 & A \notin \mathcal{U} \end{cases}.$$

Then  $\mu_{\mathcal{U}}$  is a finitely additive probability measure taking only the values 0 and 1.

Conversely, if  $\mu : \mathcal{F} \rightarrow \{0, 1\}$  is a finitely additive probability, then

$$\mathcal{U}_{\mu} := \{ A \in \mathcal{F} : \mu(A) = 1 \}$$

is an ultrafilter. Thus ultrafilters and  $\{0, 1\}$ -valued finitely additive probabilities are equivalent descriptions of the same object.

If the ultrafilter is principal, say generated by a point  $\omega \in \Omega$ , then

$$\mu_{\mathcal{U}}(A) = 1_{\{\omega \in A\}},$$

so  $\mu_{\mathcal{U}}$  is just the Dirac mass at  $\omega$ , and is countably additive.

For the non-principal case, the conclusion depends on the measurable space. On the spaces used in probability theory it is easy to prove that non-principal  $\{0, 1\}$ -valued measures are purely finitely additive.

**Proposition 5.1.** *Let  $(\Omega, \mathcal{F})$  be a countably generated measurable space whose measurable sets separate points. If  $\mu : \mathcal{F} \rightarrow \{0, 1\}$  is a finitely additive probability and is not principal, then  $\mu$  is purely finitely additive.*

*Proof.* Assume for contradiction that  $\mu$  dominates a nonzero countably additive measure  $\nu$ . After normalization, we may suppose  $\nu$  is a countably additive probability.

Let  $\{E_n\}_{n \geq 1}$  generate  $\mathcal{F}$  and separate points. For each  $n$ , exactly one of  $E_n$  and  $E_n^c$  has  $\mu$ -mass 1; call that set  $B_n$ . Since  $\nu \leq \mu$ , every  $\mu$ -null set is also  $\nu$ -null, and therefore  $\nu(B_n) = 1$  for all  $n$ .

Since each  $B_n^c$  is  $\nu$ -null, countable additivity implies

$$\nu \left( \bigcup_{n \geq 1} B_n^c \right) = 0,$$

and therefore

$$\nu \left( \bigcap_{n \geq 1} B_n \right) = 1.$$

Because the  $E_n$  separate points, the intersection contains at most one point. Hence it must be a singleton  $\{\omega\}$ , and so  $\nu = \delta_{\omega}$ . But then

$$1 = \nu \{ \omega \} \leq \mu \{ \omega \},$$

forcing  $\mu \{ \omega \} = 1$ , contrary to non-principality.  $\square$

This proposition applies to all the familiar settings in analysis, including  $\mathbb{N}$  with its power set, and  $[0, 1]$  with either the Borel or Lebesgue  $\sigma$ -field.

*Remark 5.6.* Proposition 5.1 is intentionally restrictive on  $\Omega$  and  $\mathcal{F}$ . In full generality the statement fails if one allows measurable-cardinal-strength set theory: a non-principal countably complete ultrafilter gives a non-principal countably additive  $\{0, 1\}$ -valued measure. The absence of a familiar counterexample is therefore not an accident. In full generality, a counterexample requires set-theoretic strength beyond ordinary analysis, namely a measurable cardinal, and so is not provable in ZFC alone.

### 5.5.2 The Prototype on $\mathbb{N}$ : Ultrafilter Expectations and Failure of Fatou

The simplest top-row example lives on  $\mathbb{N}$ . Let  $\mathcal{U}$  be a non-principal ultrafilter on  $\mathbb{N}$ , and define

$$\mu(A) := 1_{\mathcal{U}}(A), \quad A \subseteq \mathbb{N}.$$

By the preceding subsection,  $\mu$  is a  $\{0, 1\}$ -valued finitely additive probability. Because  $\mathcal{U}$  is non-principal and the power set  $\mathcal{P}(\mathbb{N})$  is countably generated and separates points,  $\mu$  is purely finitely additive.

Every non-principal ultrafilter on  $\mathbb{N}$  contains every cofinite set. Hence every finite set has  $\mu$ -mass 0, while every cofinite set has  $\mu$ -mass 1. This already shows that  $\mu$  cannot be countably additive:

$$\mathbb{N} = \bigcup_{n \geq 1} \{n\}, \quad \mu \{n\} = 0 \text{ for all } n, \quad \mu(\mathbb{N}) = 1.$$

A useful way to think about  $\mu$  is as a generalized limit. If  $X = (x_n)_{n \geq 1} \in \ell^\infty$ , define

$$\rho(X) := \int X d\mu.$$

Then  $\rho$  is linear, monotone, and normalized by  $\rho(1) = 1$ . Since linearity implies comonotonic additivity,  $\rho$  lies in the top-left corner of the Atlas: it is coherent, comonotonic additive, and admits the trivial dual representation

$$\rho(X) = \max_{\nu \in ba_+, \nu(\Omega)=1} \left\{ \int X d\nu - \rho^*(\nu) \right\}, \quad \rho^*(\nu) = \begin{cases} 0, & \nu = \mu, \\ \infty, & \nu \neq \mu. \end{cases}$$

The issue is not the absence of a dual representation, but that the representation necessarily uses dual variables from  $ba \setminus L^1$ .

**Example 5.2.** Define

$$X_n = (\underbrace{1, \dots, 1}_{n \text{ terms}}, 0, 0, \dots) = 1_{\{1, \dots, n\}}, \quad X = (1, 1, 1, \dots).$$

Then  $X_n \uparrow X$  pointwise, but

$$\rho(X_n) = \mu\{1, \dots, n\} = 0 \quad \text{for all } n,$$

whereas

$$\rho(X) = \mu(\mathbb{N}) = 1.$$

Therefore  $\rho$  jumps-up unexpectedly at the limit

$$\rho(X) = 1 > \liminf_{n \rightarrow \infty} \rho(X_n) = 0,$$

and so  $\rho$  fails the Fatou property. This is very bad behavior for a risk measure: if we have controlled the risk of each  $X_n$  we expect to have controlled the limit.

Example 5.2 already captures the essential top-row weirdness. The functional is linear and perfectly well behaved from the Banach-space point of view, but it is not lower semicontinuous for the natural probability topology  $\sigma(L^\infty, L^1)$ . The defect is not algebraic. It is entirely topological, and it comes from allowing dual variables in  $ba$  but not in  $L^1$ .

It is also worth emphasizing how alien the underlying expectation is. For an ordinary countably additive probability on  $\mathbb{N}$ , the expectation of a bounded sequence averages its values using weights at individual points. The ultrafilter expectation ignores every finite initial segment and records only what happens on a set in the ultrafilter. In that sense it behaves like evaluation at a ghost point at infinity. The next example shows that the same phenomenon occurs on the familiar atomless space  $[0, 1]$ . There the weirdness can be visualized as concentration at a point infinitesimally to the right of 0.

### 5.5.3 The Atomless Prototype on $[0, 1]$ : Concentration Near 0

The discrete example on  $\mathbb{N}$  is reasonably explicit, but it may still feel slightly artificial because the underlying space is atomic. The same weirdness appears on the familiar atomless space  $[0, 1]$ . Let

$$I_n := (0, 1/n), \quad n \geq 1,$$

and let  $\mathcal{U}$  be an ultrafilter on the measurable subsets of  $[0, 1]$  extending the filter generated by the sets  $I_n$ . Define

$$\mu(A) := 1_{\mathcal{U}}(A), \quad A \in \mathcal{F},$$

where  $\mathcal{F}$  is either the Borel or the Lebesgue  $\sigma$ -field. Then  $\mu$  is a  $\{0, 1\}$ -valued finitely additive probability. It is not principal, because no singleton belongs to  $\mathcal{U}$ : for each  $x > 0$ , choose  $n$  with  $1/n < x$ , so  $I_n \subseteq [0, 1] \setminus \{x\}$ ; and  $\{0\}$  is disjoint from every  $I_n$ . Hence, by Proposition 5.1,  $\mu$  is purely finitely additive.

This measure is easiest to picture as concentrated at a point infinitesimally to the right of 0. Every punctured neighborhood  $(0, 1/n)$  has mass 1, every set bounded away from 0 has mass 0, and even the

singleton  $\{0\}$  has mass 0. Thus  $\mu$  behaves like a Dirac mass at an ideal point just to the right of 0, not at any genuine point of  $[0, 1]$ .

As in the discrete case, define

$$\rho(X) := \int X d\mu.$$

Then  $\rho$  is linear, monotone, coherent, and comonotonic additive. It therefore lies in the same top-left corner of the Atlas as the ultrafilter expectation on  $\mathbb{N}$ . Again the issue is not the lack of a dual representation, but the fact that the representing measure is purely finitely additive.

**Example 5.3.** Let

$$X_n = 1_{[1/n, 1]}, \quad X = 1_{(0, 1]}.$$

Then  $X_n \uparrow X$  pointwise on  $[0, 1]$ . However,

$$\rho(X_n) = \mu([1/n, 1]) = 0 \quad \text{for every } n,$$

because  $[1/n, 1]$  misses the infinitesimal neighborhood selected by the ultrafilter. On the other hand,

$$\rho(X) = \mu((0, 1]) = 1,$$

since every  $I_n$  is contained in  $(0, 1]$ . Therefore

$$\rho(X) = 1 > \liminf_{n \rightarrow \infty} \rho(X_n) = 0.$$

So  $\rho$  fails the Fatou property on the atomless space  $[0, 1]$  in exactly the same way as on  $\mathbb{N}$ .

Example 5.3 is worth dwelling on, because it removes several possible sources of confusion. The failure of Fatou has nothing to do with atoms, with discrete spaces, or with awkward combinatorics on  $\mathbb{N}$ . It persists on the standard interval model of an atomless probability space. What matters is only that the dual variable  $\mu$  is purely finitely additive.

**Example 5.4.** The construction of Example 5.3 gives a striking interpretation of oscillatory functions such as

$$X(x) = \sin(1/x), \quad x \in (0, 1].$$

An ultrafilter extending the neighborhoods  $(0, 1/n)$  chooses a generalized limiting value of  $X$  as  $x \downarrow 0$ . In particular, by choosing the ultrafilter appropriately one can arrange

$$\int X d\mu = 1, \quad \int X d\mu = -1,$$

or indeed any prescribed value in  $[-1, 1]$ .

To see this, fix  $c \in [-1, 1]$  and define

$$A_n(c) := \{x \in (0, 1] : |\sin(1/x) - c| < 1/n\}.$$

Each  $A_n(c)$  accumulates at 0, and every finite intersection

$$I_m \cap A_{n_1}(c) \cap \cdots \cap A_{n_k}(c)$$

is nonempty. Hence these sets generate a filter, which can be extended to an ultrafilter  $\mathcal{U}_c$ . For the associated measure  $\mu_c$  one has

$$\int \sin(1/x) d\mu_c = c.$$

Thus the ultrafilter integral records not an ordinary limit, since none exists, but any chosen cluster value.

*Remark 5.7.* There is a conceptual connection with Robinson's nonstandard analysis [Goldblatt \(1998\)](#). The ultrafilter extending the sets  $(0, 1/n)$  behaves as though it were evaluating functions at an infinitesimal  $\varepsilon > 0$  satisfying  $\varepsilon < 1/n$  for every standard  $n$ . In that heuristic picture,

$$\int X d\mu$$

acts like the value of  $X$  at an ideal point infinitesimally to the right of 0. We do not need nonstandard analysis here, but its language provides a vivid mental picture for what the ultrafilter is doing.

Together with the example on  $\mathbb{N}$ , this interval construction isolates the essential defect of the top row. Once purely finitely additive expectations are admitted, the Fatou property can fail in the most direct possible way, even for linear functionals on familiar spaces. The remaining top-row examples merely decorate this basic defect: take maximums and add a non-trivial penalty function.

Another extension is to pass from additive measures to capacities and use the associated Choquet integral. Starting from a purely finitely additive probability  $\mu$ , one can combine it with an ordinary probability  $P$  to form a capacity such as

$$v(A) = \lambda \mu(A) + (1 - \lambda) g(P(A)), \quad 0 < \lambda < 1,$$

where  $g$  is a distortion function. This capacity is generally nonadditive, and it is non-law-invariant as soon as the  $\mu$ -term is non-law-invariant. The associated Choquet integral is comonotonic additive but no longer linear, and the underlying weirdness is unchanged: the defect comes from the purely finitely additive component.

*Remark 5.8.* The top row is not merely a weird fringe; it is a vastly larger and less rigid world. On a standard Borel probability space, the countably additive dual objects in the middle and bottom rows are the  $L^1$  densities

$$Z \geq 0, \quad PZ = 1,$$

and there are only continuum many of these. By contrast, once one allows finitely additive probabilities, the supply of dual objects becomes enormous. For example, the set of free ultrafilters on  $\mathbb{N}$  has cardinality

$$2^{2^{\aleph_0}},$$

and each such ultrafilter gives a 0-1 finitely additive probability. Thus the top row is not just larger in some vague sense: once countable additivity is dropped, finitely additive extensions proliferate.

## 5.6 Right of Right

### 5.6.1 Non-Monotone Examples

Monotonicity is not forced by convexity or duality alone. The classical counterexample is mean-variance

$$\rho(X) := PX + \frac{\lambda}{2} \text{var}(X)$$

which is cash invariant and convex and has a perfectly good convex dual, but it is not monotone. It lies outside the Atlas, to the right of the right-hand column.

Maccheroni et al. (2009) construct the monotone mean-variance functional as the largest monotone modification of the mean-variance criterion, and show that it has the variational representation

$$V_\theta(f) = \min_Q \left\{ E_Q[f] + \frac{1}{2\theta} C(Q\|P) \right\},$$

where  $C(Q\|P)$  is an explicit convex divergence penalty. Their point is exactly that ordinary mean-variance has the wrong order structure, and that monotonicity can be restored by passing to the closest “economically meaningful” (their phrase) monotone envelope. Acciaio (2007) fits naturally beside this: one can start from a non-monotone monetary functional and then ask for its best monotone approximation, which one way to think about the non-monotone border of the Atlas.

### 5.6.2 Non-Cash Invariant Examples

Cash invariance is one of the defining axioms of a monetary risk measure. It says that adding a sure loss of size  $m$  increases the risk measure by exactly  $m$ . This is natural when the pricing or risk functional is calibrated in a fixed numeraire and certain deterministic shifts are treated linearly. But outside that setting the axiom can be too strong. If the translation from today’s unit of account to tomorrow’s loss is itself uncertain, or if deterministic shifts are not priced one-for-one, then exact cash invariance need not hold.

One possible weakening is cash subadditivity, introduced by El Karoui and Ravanelli (2009) in the presence of interest-rate ambiguity and related numeraire uncertainty. In the loss convention, the defining inequality is

$$\rho(X + m) \leq \rho(X) + m, \quad m \geq 0.$$

Here  $\rho$  evaluates a future loss position in today's units of account. Thus adding a sure loss of size  $m$  in the future can increase the assessed risk by at most  $m$  today, but not necessarily by exactly  $m$ . A deterministic shift in the future loss need not map exactly into present values when discounting is ambiguous or the numeraire is itself risky. A cash-subadditive functional allows such a deterministic increment to contribute less than its face value in today's units, because of discounting ambiguity, default risk, or other numeraire effects. El Karoui and Ravanelli (2009) show that these functionals admit robust representations, but the dual objects are modified accordingly. They work with subprobability-type dual variables or with an enlarged space on which the functional becomes cash additive. More broadly, as emphasized by Drapeau and Kupper (2013), once cash invariance is dropped one can still retain monotonicity and diversification-type structure and seek robust representations of the resulting risk preferences.

## 5.7 Dual Representations

It is easy to deduce the behavior of a risk measure from properties of its dual function. It is less obvious how to determine the dual from a given measure. This section catalogs some examples.

We start by describing the **optimized certainty equivalent** (OCE), which provides a systematic way to generate convex risk measures and their conjugates. OCE includes several well-known examples. Let  $\ell : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a closed, proper, convex function, with

$$\ell(0) = 0 \quad \text{and} \quad \ell(x) \geq x \quad \text{for all } x \in \mathbb{R}.$$

The associated OCE risk measure is

$$\rho(X) = \inf_{t \in \mathbb{R}} t + \mathbb{P}(\ell(X - t)).$$

If, in addition,  $\ell$  is nondecreasing, then  $\rho$  is monotone; equivalently,  $\text{dom } \ell^* \subset [0, \infty)$ .

**Proposition 5.2.** *Let*

$$\rho(X) = \inf_{t \in \mathbb{R}} t + \mathbb{P}(\ell(X - t))$$

*be an OCE risk measure on  $L^\infty$ . Then its convex conjugate is*

$$\rho^*(Z) = \begin{cases} \mathbb{P}(\ell^*(Z)), & \text{if } Z \in L^1, \mathbb{P}(Z) = 1, \\ \infty, & \text{otherwise,} \end{cases}$$

*where*

$$\ell^*(z) = \sup_{x \in \mathbb{R}} zx - \ell(x)$$

*is the convex conjugate of  $\ell$ .*

*Proof.* By definition,

$$\rho^*(Z) = \sup_{X \in L^\infty} \left( \mathbb{P}(ZX) - \inf_{t \in \mathbb{R}} t + \mathbb{P}(\ell(X - t)) \right).$$

Replacing the infimum by a supremum and setting  $Y = X - t$  gives

$$\begin{aligned} \rho^*(Z) &= \sup_{Y \in L^\infty} \sup_{t \in \mathbb{R}} \mathbb{P}(Z(Y + t)) - t - \mathbb{P}(\ell(Y)) \\ &= \sup_{Y \in L^\infty} \sup_{t \in \mathbb{R}} \mathbb{P}(ZY) - \mathbb{P}(\ell(Y)) + t(\mathbb{P}(Z) - 1). \end{aligned}$$

If  $\mathbb{P}(Z) \neq 1$ , the supremum over  $t$  equals  $\infty$ . Hence  $\rho^*(Z) = \infty$  unless  $\mathbb{P}(Z) = 1$ .

Assume now that  $\mathbb{P}(Z) = 1$ . Then

$$\rho^*(Z) = \sup_{Y \in L^\infty} \mathbb{P}(ZY - \ell(Y)).$$

By the usual interchange argument for decomposable spaces and normal integrands,

$$\rho^*(Z) = \mathbb{P} \left( \sup_{y \in \mathbb{R}} Zy - \ell(y) \right) = \mathbb{P}(\ell^*(Z)),$$

proving the formula. □

We may equivalently start from the convex conjugate  $\ell^*$  rather than from  $\ell$ . In the **divergence risk measure** approach, we write  $g = \ell^*$  and  $Z = dQ/dP$ , and rewrite  $\rho^*$  as

$$P(g(Z)) = D_g(Q | P) = \begin{cases} P g\left(\frac{dQ}{dP}\right), & Q \ll P \\ \infty, & \text{otherwise} \end{cases}.$$

Under the usual assumptions on  $g$ , including convexity, lower semicontinuity,  $g(1) = 0$ , and suitable super-linear growth, the corresponding convex risk measure is then just

$$\rho(X) = \sup_Q Q(X) - D_g(Q | P).$$

Thus OCEs and divergence risk measures are convex-dual descriptions of the same basic construction.

Table 1 gives several useful examples of  $\rho, \rho^*$  pairs. The first four rows arise naturally from OCE/divergence constructions; the remaining rows are deviation-based examples included for comparison.

Table 1: Example risk measure and convex dual pairs.

Risk measure	Measure $\rho(X)$	Convex dual $\rho^*(Z)$	Monotone
Entropic	$\theta^{-1} \log \mathbb{P}(e^{\theta X})$	$\begin{cases} \theta^{-1} \mathbb{P}(Z \log Z - Z + 1) & Z \geq 0, \mathbb{P}Z = 1 \\ \infty & \text{otherwise} \end{cases}$	Yes
Optimized certainty equivalent	$\inf_t (t + \mathbb{P}(\ell(X - t)))$	$\begin{cases} \mathbb{P}(\ell^*(Z)) & \mathbb{P}Z = 1 \\ \infty & \text{otherwise} \end{cases}$	iff $\ell$ is non-decreasing
TVaR	$\inf_t \left( t + \frac{1}{1-p} \mathbb{P}((X - t)_+) \right)$	$\begin{cases} 0 & 0 \leq Z \leq (1-p)^{-1}, \mathbb{P}Z = 1 \\ \infty & \text{otherwise} \end{cases}$	Yes
Mean-variance (raw)	$\mathbb{P}(X) + c \text{var}(X)$	$\begin{cases} (4c)^{-1} \mathbb{P}((Z - 1)^2) & \mathbb{P}Z = 1 \\ \infty & \text{otherwise} \end{cases}$	No
Monotone mean-variance	$\inf_{Y \geq X} (\mathbb{P}(Y) + c \text{var}(Y))$	$\begin{cases} (4c)^{-1} \mathbb{P}((Z - 1)^2) & Z \geq 0, \mathbb{P}Z = 1 \\ \infty & \text{otherwise} \end{cases}$	Yes
Mean- $L^p$ -deviation	$\mathbb{P}(X) + c \ X - \mathbb{P}(X)\ _p$	$\begin{cases} 0 & \ Z - 1\ _q \leq c, \mathbb{P}Z = 1 \\ \infty & \text{otherwise} \end{cases}$	No
Mean- $p$ -semi-deviation	$\mathbb{P}(X) + c \ (X - \mathbb{P}(X))^+\ _p$	$\begin{cases} 0 & Z = 1 + W - \mathbb{P}(W), W \geq 0, \ W\ _q \leq c \\ \infty & \text{otherwise} \end{cases}$	Yes for $c \leq 1$
Mean-Gini	$\mathbb{P}(X) + c \mathbb{P} X_1 - X_2 $	$\begin{cases} 0 & Z \prec_{cx} 1 + c(2U - 1), U \sim U(0, 1) \\ \infty & \text{otherwise} \end{cases}$	Yes for $c \leq 1/2$

## Notes

Proposition 5.2 shows that the OCE class is parameterized by the loss function  $\ell$ , or equivalently by the divergence function  $g = \ell^*$ . If  $\ell$  decreases on some interval, then  $\ell^*(z)$  is finite for some  $z < 0$ , so signed dual densities are admissible and the resulting risk measure is not monotone. If  $\ell$  is nondecreasing, then  $\text{dom } \ell^* \subset [0, \infty)$  and the OCE is monotone.

- Entropic risk measure corresponds to the OCE and divergence pair

$$\ell(x) = \frac{1}{\theta}(e^{\theta x} - 1), \quad g(z) = \ell^*(z) = \frac{1}{\theta}(z \log z - z + 1).$$

Since,  $\mathbb{P}(d\mathbb{Q}/d\mathbb{P}) = 1$ , we get

$$D_g(\mathbb{Q} \mid \mathbb{P}) = \frac{1}{\theta} \mathbb{P} \left[ \frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}} \right],$$

which is the scaled Kullback-Leibler divergence, or relative entropy.

- Tail Value at Risk is the OCE obtained from

$$\ell(x) = \frac{1}{1-p} x_+.$$

- Mean-variance corresponds to

$$\ell(x) = x + cx^2.$$

Here,  $\ell$  decreases on  $(-\infty, -1/(2c))$ , so the resulting OCE is not monotone. Completing the square yields the quadratic penalty.

- Monotone mean-variance. To restore monotonicity, truncate  $\ell$  at its minimum:

$$\ell(x) = \begin{cases} x + cx^2, & x \geq -\frac{1}{2c}, \\ -\frac{1}{4c}, & x < -\frac{1}{2c}. \end{cases}$$

Then  $\ell$  is nondecreasing, and  $\ell^*(z) = \infty$  for  $z < 0$ , so the dual problem is restricted to nonnegative densities

$$\rho(X) = \inf_{Y \geq X} \mathbb{P}(Y) + c \text{var}(Y).$$

This functional is the monotone envelope studied by Maccheroni et al. (2009). They show that the monotone envelope of mean-variance also admits a variational representation with quadratic divergence penalty

$$C(\mathbb{Q} \parallel \mathbb{P}) = \mathbb{P} \left[ \left( \frac{d\mathbb{Q}}{d\mathbb{P}} \right)^2 \right] - 1,$$

called the relative Gini concentration index, or  $\chi^2$ -distance. The Gini index is a concentration index, like Shannon entropy. The quantity  $C(\mathbb{Q} \parallel \mathbb{P}) = \mathbb{P}(Z^2) - 1$ , with  $Z = d\mathbb{Q}/d\mathbb{P}$ , is the  $\chi^2$ -divergence; since  $\mathbb{P}Z = 1$ , it equals  $\text{var}(Z)$ .

- Mean- $L^p$ -deviation. Here  $\rho^*$  is the convex indicator on the  $L^q$  ball around 1 of radius  $c$ . Because the norm constraint is symmetric around 1, negative values of  $Z$  are allowed, so the functional is generally not monotone. By Hölder's inequality, the dual density's  $L^q$  distance from the neutral measure  $Z = 1$  is strictly bounded by  $c$ .
- Mean- $p$ -semi-deviation. The one-sided deviation produces the representation on  $Z$  expressed as

$$Z = 1 + W - \mathbb{P}(W), \quad W \geq 0, \quad \|W\|_q \leq c.$$

The non-negativity of  $W$  reflects the fact that only the right tail is penalized. The resulting risk measure is coherent for  $c \leq 1$ .

- Mean-Gini is a deviation-based functional defined as

$$\rho(X) = \mathbb{P}(X) + c \mathbb{P}|X_1 - X_2|.$$

The dual set is described by a convex-order constraint

$$Z \prec_{cx} 1 + c(2U - 1), \quad U \sim U(0, 1),$$

so it is naturally understood as a law-invariant coherent risk measure with a structured dual set. It is monotone and coherent for  $c \in [0, 1/2]$ .

- References: Ben-Tal and Teboulle (2007), Föllmer and Weber (2015), Shapiro (2012), Pflug and Römisch (2007).

## 6 Three Miscellaneous Topics

### 6.1 Measures of Risk vs. Measures of Value

In the literature, it is common to encounter two related viewpoints: risk vs. value. The risk view, the one used in this article, assess the riskiness or cost of a position, using the loss sign convention: a larger value of  $X$  is worse, because it means a larger loss. Alternatively, the utility or valuation view assesses the value of a position. Value measures usually use the payoff sign convention: a larger value of  $X$  is better, because it means more wealth or more payoff. The mathematics is almost the same, but the interpretation runs in opposite directions.

The connection between the two viewpoints is simple. Risk is just negative value. If  $V$  is a valuation functional defined on payoffs, then the corresponding risk functional on loss positions is obtained by applying  $V$  to the opposite position and then changing sign:

$$\rho(X) = -V(-X).$$

The inner minus sign converts a loss into the corresponding payoff, and the outer minus sign converts value back into risk. This identity is elementary, but it is worth keeping in mind because many formulas in the literature are written in the value convention, whereas here we use the risk convention throughout.

The sign switch explains a great deal of the apparent notational friction in the subject. Statements that look different are often the same statement seen from opposite sides of the ledger. A cash-additive valuation becomes a cash-invariant risk measure after the sign change. Monotonicity reverses its verbal interpretation: for value, more is better; for risk, more is worse. Even dual representations are unchanged in substance, though the minus signs move around. Once this translation is kept in mind, it becomes much easier to read across the utility, pricing, and risk-measure literatures without confusion.

### 6.2 Acceptance Sets

Acceptance sets appear throughout the literature as another way to define a risk measure. Here is the idea. Given a risk functional  $\rho$ , define its acceptance set to be

$$\mathcal{A} = \{ X : \rho(X) \leq 0 \}.$$

This is the set of positions regarded as acceptable without adding extra capital. Conversely, once  $\mathcal{A}$  is given, the risk measure can be recovered from it by

$$\rho(X) = \inf \{ m \in \mathbb{R} : X - m \in \mathcal{A} \}.$$

Thus the risk measure and its acceptance set contain the same information. One describes acceptability directly; the other describes how much cash must be added to make a position acceptable.

This viewpoint is often very useful because the structural properties of  $\rho$  become geometric properties of  $\mathcal{A}$ . Monotonicity means that if  $X \in \mathcal{A}$  and  $Y \leq X$ , then  $Y \in \mathcal{A}$ : if a position is acceptable, any uniformly better one is too. Cash invariance means that acceptability is stable under deterministic cash shifts, and allows recovery of  $\rho$  from the boundary of  $\mathcal{A}$ . Convexity of  $\rho$  corresponds to convexity of  $\mathcal{A}$ , so diversification of acceptable positions remains acceptable. Positive homogeneity corresponds to

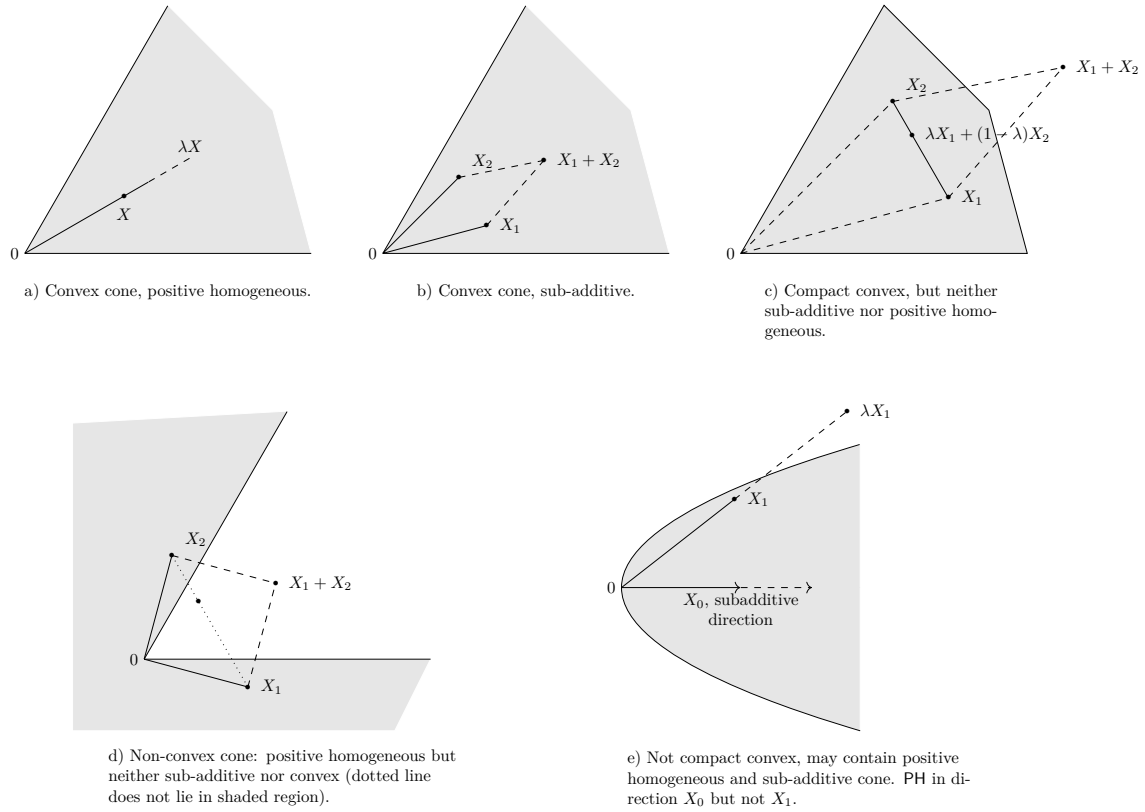


Figure 5: Acceptance sets and corresponding risk measure properties. The shaded region represents the acceptance set  $\mathcal{A}$ . Black lines represent the boundary; unbounded region have no boundary line. Only (c) is bounded. The origin is marked with 0. (a) A convex cone is positive homogeneous, the line  $\lambda X$  is contained in the shaded region  $\mathcal{A}$  and (b) is also sub-additive, the sum of two points in  $\mathcal{A}$  is again in  $\mathcal{A}$ . (c) A compact convex  $\mathcal{A}$  is neither positive homogeneous nor sub-additive, but the line joining two points in  $\mathcal{A}$  is in  $\mathcal{A}$ . (d) A non-convex cone is positive homogeneous but not sub-additive. (e) A non-compact convex  $\mathcal{A}$  can be sub-additive and positive homogeneous in some directions but not others. Source: Mildenhall and Major (2022).

$\mathcal{A}$  being a cone, so in the coherent case the acceptance set is a convex cone. For a merely convex risk measure,  $\mathcal{A}$  is convex but not generally conic. These ideas are illustrated in #fig-acceptance-sets.

Topology enters through whether or not  $\mathcal{A}$  is closed. If  $\rho$  has the Fatou property, then  $\mathcal{A}$  is closed under bounded almost sure convergence, and for convex monetary risk measures this is equivalent to being  $\sigma(L^\infty, L^1)$  closed. Being closed is exactly what allows an exact dual representation on countably additive densities. Without it, one is pushed into the larger finitely additive dual world. So acceptance sets provide a clean geometric way to see why convexity and closedness matter so much: convexity gives supporting hyperplanes, and closedness ensures that those supporting hyperplanes recover the set exactly.

For law-invariant and spectral examples, the acceptance set has extra symmetry. Law invariance means membership of  $\mathcal{A}$  depends only on the distribution of  $X$ , not on the labels of the underlying states. For spectral and other comonotonic additive measures, the acceptance set reflects the same distribution-based structure that appears in the distortion or Choquet representation. In that sense acceptance sets are another Atlas of the theory: linear measures correspond to half-spaces, coherent measures to closed convex cones, convex measures to closed convex sets, and law invariance adds symmetry under rearrangement.

These ideas mirror our discussion of the topology of the domain of the dual function, #fig-atlas-topol.

### 6.3 Pricing vs. Regulation

We can formulate the theory of risk measures equally well in terms of pricing or in terms of regulatory capital. In the standard presentation, we ask how much certain cash must be added to a position to make it acceptable “to the regulator” or “supervising agency”. In a pricing presentation, we ask how much premium or initial consideration must be received to make taking the position acceptable to the insurer or issuing entity. These are mathematically equivalent ways of looking at the same object. If  $X$  is a loss, then  $\rho(X)$  can be read either as the capital requirement for holding  $X$ , or as the premium required for accepting  $X$ . The basic structure does not change: monotonicity still says that a worse loss requires a higher premium; convexity still captures diversification and concentration effects; positive homogeneity still gives scale invariance; law invariance still means the price depends only on the distribution of the loss under the chosen objective measure.

From an insurance point of view, the pricing formulation is often the more natural one. Insurance is a business of premiums, not of attaching a labeled packet of capital to each policy. The premium is real, contractually specified, paid at inception, booked in accounts, and tied to observable cash flows. The phrase “adding premium to make a position acceptable” therefore corresponds to an actual commercial transaction. By contrast, “adding capital to make a position acceptable” is a hypothetical supervisory or accounting construction. Capital is fungible across the firm, and allocated capital at the unit level is typically a fictional internal attribution, not a distinct flow of money attached to the position itself.

The pricing viewpoint is older than the modern risk-measure literature. The actuarial premium principle tradition already studies monotone and diversification-sensitive pricing rules, and work such as Goovaerts, De Vylder, and Haezendonck (1984) place premium calculation at center stage. In that language, many familiar risk measure properties are exactly the properties one wants from a premium principle. A linear price corresponds to expectation under a pricing measure. A coherent price principle rewards diversification through subadditivity. A convex price principle allows for liquidity effects, friction, or increasing marginal cost of risk. A law-invariant price principle depends only on the loss distribution, not on the labels of the underlying states. The mathematics is the same; only the interpretation changes.

The acceptance-set formulation also translates directly. Instead of saying that  $X$  is acceptable when enough capital is added, one can say that  $X$  is acceptable at inception when the insurer receives enough premium. Then

$$\rho(X) = \inf \{ m \in \mathbb{R} : X - m \in \mathcal{A} \}$$

can be read as the least premium offset that moves the loss position into the acceptable set. The dual representation has the same interpretation too. The supporting measures or densities are stress scenarios or pricing scenarios, and the penalty function is the extra cost attached to using each scenario. Nothing essential is lost by speaking in terms of premium rather than capital; indeed, in insurance, something important is arguably gained, because the interpretation is closer to underwriting reality.

Cynically, we may suspect the dominance of the capital-requirement language is not just a matter of mathematical convenience. It may also reflect academic incentives. A paper framed as helping regulators, solvency regimes, or capital adequacy fits comfortably into the modern finance-and-regulation ecosystem. A paper framed as a theory of premiums is immediately exposed to harder questions: does it match market prices, underwriting practice, competition, and data? What are the empirical tests? Capital allocation is much harder to observe and much easier to leave as an internal or regulatory fiction. Pricing is concrete, commercial, and testable. Regulation is vaguer, more abstract, and often harder to falsify. That may be one reason the literature speaks so confidently about making positions acceptable by adding capital, while saying much less about the more natural and economically meaningful question of what premium actually gets paid.

Pricing actuaries can benefit from the modern analytic framework developed for risk measures—and they outnumber those working on regulation or capital allocation. Read as pricing functionals, risk measures provide a disciplined way to think about premium principles that go beyond expected loss or proportional loads by adding an explicit risk margin, with monotonicity, diversification, convexity, dual representations, and law invariance all available as useful organizing ideas. That perspective is especially valuable in lines such as property catastrophe insurance, where the risk margin is often not a small adjustment but the dominant part of the premium, commonly two to four times the expected loss. In such cases the abstract theory is not peripheral. It speaks directly to the central commercial problem.

## 7 Related Literature

This section provides a brief overview of some important papers in the development of convex risk measures. See Föllmer and Schied (2016) for a more thorough review.

*Coherent Measures of Risk*, Artzner et al. (1999) introduces the axiomatic approach, makes subadditivity and positive homogeneity central, and frames risk as a capital requirement rather than a utility. Much that follows is either an extension, a relaxation, or a reinterpretation of their original setup. It currently counts 14,194 Google Scholar citations.

*Coherent Risk Measures on General Probability Spaces*, Delbaen (2002) and the Pisa notes Delbaen (2000) present the coherent theory in full functional-analytic form. Delbaen takes the finite-scenario setup from Artzner et al. (1999) to general probability spaces, links coherent risk measures to closed convex sets of probability measures, discusses extension to larger spaces of random variables, and draws connections to distorted probabilities and cooperative games.

*Coherent risk measures and good-deal bounds*, Jaschke and Küchler (2001) provides an economic interpretation by showing that coherent risk measures are essentially equivalent to generalized good-deal bounds, connecting the new axioms to pricing bounds and portfolio problems. This is one of the first papers that explains why the dual set is not just mathematically convenient, but economically meaningful.

*On law invariant coherent risk measures*, Kusuoka (2001) classifies law-invariant coherent measures. In the later literature this becomes the structural theorem that coherent, law invariant, and Fatou means mixtures or suprema of TVaRs. Kusuoka generalizes for convex measures and remains the model for a law-invariant representation theorem. Frittelli and Gianin (2005) explicitly present their paper as a generalization of Kusuoka’s result, giving a convex law-invariant representation theorem. Shapiro (2012) discusses uniqueness of the dual representation.

*Convex measures of risk and trading constraints*, Föllmer and Schied (2002) is one of two foundational convex papers. They introduce convex risk measures as the right relaxation of coherence, prove the corresponding dual representation with a penalty term, and relate to super-hedging under convex constraints and to utility-based shortfall. Their paper is the source of the now-standard “worst penalized expected loss” picture.

*Putting order in risk measures*, Frittelli and Rosazza Gianin (2002) is the second foundational convex paper. It has a somewhat different flavor. It introduces axioms for convex risk measures and uses duality to obtain a representation theorem and a link with pricing rules. It recasts the subject in order, acceptance, and pricing language, closer to financial valuation, rather than via the trading-constraints route of Föllmer-Schied.

*Robust Preferences and Convex Measures of Risk*, Föllmer and Schied (2002) introduces the idea of model-uncertainty. It explicitly connects convex risk measures to robust preferences in the Gilboa-Schmeidler, where the penalty term becomes an ambiguity penalty rather than a purely formal conjugate, Gilboa and Schmeidler (1993). It extends Föllmer and Schied (2002) from the duality theorem to a decision-theoretic meaning of the dual variables and the penalty.

A number of papers consider dynamic generalizations of static measures, starting with Frittelli and Gianin (2004) and continued in Cheridito, Delbaen, and Kupper (2004). These show that risk measurement is not just about terminal payoff uncertainty, but about cash-flow streams and timing. Föllmer and Penner (2006) turns the penalty function into a dynamic object and shows how the dual variables evolve over time. Detlefsen and Scandolo (2005) treats the conditional case.

Beyond the basic  $L^\infty$  setting, several works deepen the relation between continuity, duality, and model space. Biagini and Frittelli (2007) connect Fatou-type lower semicontinuity with simplified dual representations on more general lattices. Cheridito and Li (2008) explore how duality reflects structural properties of monetary risk measures on Orlicz hearts.

## 8 Deep Background

### 8.1 Continuity From Above and Below and Lebesgue Continuity

Föllmer and Schied (2016) define the following terminology. A monetary risk measure  $\rho$  is

- **Continuous from above:** if  $(X_n) \subset L^\infty$  satisfies  $X_n \downarrow X$  almost surely, then

$$\rho(X_n) \downarrow \rho(X);$$

- **Continuous from below:** if  $(X_n) \subset L^\infty$  satisfies  $X_n \uparrow X$  almost surely, then

$$\rho(X_n) \uparrow \rho(X);$$

- and has the **Lebesgue property:** if  $(X_n) \subset L^\infty$  satisfies  $\sup_n \|X_n\|_\infty < \infty$  and  $X_n \rightarrow X$  almost surely, then

$$\rho(X_n) \rightarrow \rho(X).$$

Föllmer and Schied (2016) use the payout sign convention whereas we use the loss. That leads to confusion: “Up is down and good is bad.” In the loss sign convention, a large positive is bad: it is an amount owed; in the payout it is good: an amount received. In the loss convention, we are concerned with the right tail,  $p \uparrow 1$ , and ess sup. In the payout, it is the left tail,  $p \downarrow 0$ , and ess inf. As already noted, ess sup is continuous from below but not above:  $X_n(0, 1/n) \downarrow 0$ , but each ess sup  $X_n = 1$ ; conversely, ess inf is continuous from above but not below as evidenced by  $X_n(1/n, 1)$ .

The functional  $\rho$  is Lebesgue iff it is continuous from above and below. Implies is easy because monotone sequences are uniformly bounded. For the converse, let  $X_n \rightarrow X$  almost surely with  $\sup_n \|X_n\|_\infty \leq M$ . Define

$$Y_n := \inf_{k \geq n} X_k, \quad Z_n := \sup_{k \geq n} X_k.$$

Then  $Y_n \uparrow X$  a.s. and  $Z_n \downarrow X$  a.s., and

$$Y_n \leq X_n \leq Z_n.$$

By monotonicity,

$$\rho(Y_n) \leq \rho(X_n) \leq \rho(Z_n).$$

By continuity from below and above,

$$\rho(Y_n) \uparrow \rho(X), \quad \rho(Z_n) \downarrow \rho(X).$$

Hence the squeeze theorem gives

$$\rho(X_n) \rightarrow \rho(X).$$

The equivalence uses only monotonicity; cash invariance is not needed.

**Lemma 8.1.** *Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be monotone. Then the following are equivalent.*

1.  $\rho$  has the Fatou property.
2.  $\rho$  is continuous from below: if  $X_n \uparrow X$  almost surely, then

$$\rho(X_n) \uparrow \rho(X).$$

*Proof.* First assume  $\rho$  has the Fatou property, and let  $X_n \uparrow X$  almost surely. By monotonicity,

$$\rho(X_n) \leq \rho(X) \quad \text{for all } n.$$

Since  $(X_n)$  is uniformly bounded in  $L^\infty$  by  $\|X\|_\infty$ , Fatou gives

$$\rho(X) \leq \liminf_{n \rightarrow \infty} \rho(X_n).$$

Combining the two inequalities yields

$$\rho(X_n) \uparrow \rho(X).$$

Conversely, assume  $\rho$  is continuous from below, and let  $X_n \rightarrow X$  almost surely with

$$\sup_n \|X_n\|_\infty < \infty.$$

Define

$$Y_n := \inf_{k \geq n} X_k.$$

Then  $Y_n \uparrow X$  almost surely. Moreover  $Y_n \leq X_k$  for every  $k \geq n$ , so by monotonicity

$$\rho(Y_n) \leq \rho(X_k) \quad \text{for all } k \geq n.$$

Hence

$$\rho(Y_n) \leq \inf_{k \geq n} \rho(X_k).$$

By continuity from below,

$$\rho(X) = \lim_{n \rightarrow \infty} \rho(Y_n).$$

Therefore

$$\rho(X) \leq \sup_n \inf_{k \geq n} \rho(X_k) = \liminf_{n \rightarrow \infty} \rho(X_n).$$

Thus  $\rho$  has the Fatou property.  $\square$

**Proposition 8.1.** *Let  $\rho : L^\infty \rightarrow \mathbb{R}$  be a convex monetary risk measure with the Fatou property. Then the following are equivalent.*

1.  $\rho$  has the Lebesgue property.
2. For every  $X \in L^\infty$ , there exists  $Z_X \in L^1$  such that

$$\rho(X) = \mathbb{P}(XZ_X) - \rho^*(Z_X).$$

*Proof.* (Sketch.) Because  $\rho$  is convex, monetary, and Fatou, it admits the dual representation

$$\rho(X) = \sup_{Z \in \mathcal{D}} \{\mathbb{P}(XZ) - \rho^*(Z)\},$$

where

$$\mathcal{D} := \{Z \in L^1 : Z \geq 0, \mathbb{P}Z = 1, \rho^*(Z) < \infty\}.$$

To prove (1)  $\Rightarrow$  (2), one shows first that the sub-level sets

$$\{Z \in L^1 : \rho^*(Z) \leq c\}$$

are uniformly integrable. By Dunford–Pettis, they are relatively weakly compact in  $\sigma(L^1, L^\infty)$ . Since

$$Z \mapsto \mathbb{P}(XZ) - \rho^*(Z)$$

is  $\sigma(L^1, L^\infty)$ -upper semicontinuous, the supremum over a weakly compact set is attained. Hence there exists an optimizer  $Z_X$ .

To prove (2)  $\Rightarrow$  (1), let  $X_n \rightarrow X$  almost surely with  $\sup_n \|X_n\|_\infty < \infty$ . Choose an optimizer  $Z$  for  $X$ , so

$$\rho(X) = \mathbb{P}(XZ) - \rho^*(Z).$$

Then

$$\rho(X_n) \geq \mathbb{P}(X_n Z) - \rho^*(Z)$$

for every  $n$ . Since  $|X_n| \leq M$  and  $X_n \rightarrow X$  a.s., dominated convergence gives

$$\mathbb{P}(X_n Z) \rightarrow \mathbb{P}(XZ).$$

Therefore

$$\liminf_n \rho(X_n) \geq \rho(X).$$

That is the Fatou inequality.

To upgrade from Fatou to full convergence, use monotonicity and the already-known equivalence

$$\text{Fatou} \iff \text{CFB}.$$

Thus if  $\rho$  also has continuity from above, then it has the Lebesgue property. Conversely, the Lebesgue property obviously implies continuity from above. Hence, under Fatou,

$$\text{Lebesgue} \iff \text{CFA}.$$

Combining with (1)  $\iff$  (2) yields the desired equivalence between CFA and attainment.  $\square$

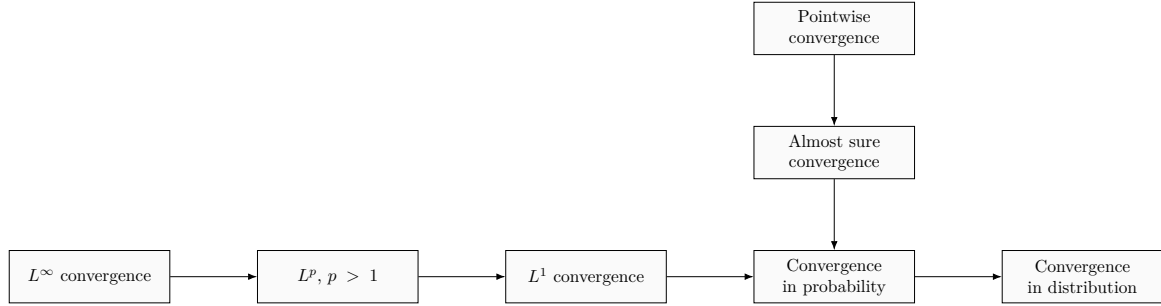


Figure 6: Different modes of convergence for random variables.

## 8.2 Modes of Convergence for Random Variables

Random variables can converge in several different ways. Here is a brief introduction, highlighting examples of the different behaviors that can occur. Many of the examples below are taken from Stoyanov (2013).

There are at least six different notions of convergence for random variables.

1.  $X_n$  converges to  $X$  **pointwise** if  $X_n(\omega) \rightarrow X(\omega)$  for all  $\omega$ .
2.  $X_n$  converges to  $X$  **almost surely** or **almost everywhere** if  $X_n(\omega) \rightarrow X(\omega)$  for almost all  $\omega$ , i.e., for all  $\omega$  in a set of probability 1. Thus, almost sure convergence requires  $\Pr(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1$ .
3.  $X_n$  converges to  $X$  in **probability** if for any  $\epsilon > 0$  we have  $\Pr(|X_n(\omega) - X(\omega)| > \epsilon) \rightarrow 0$  as  $n \rightarrow \infty$ .
4.  $X_n$  converges to  $X$  in **distribution** or **law** or **weakly** or in the **weak\* topology** if  $F_n(x) \rightarrow F(x)$  for all  $x$  for which  $F(x)$  is continuous, where  $F_n, F$  are the distribution functions of  $X_n, X$ .
5.  $X_n$  converges to  $X$  in the  **$L^1$ -norm** if  $\int |X_n(\omega) - X(\omega)| \Pr(d\omega) \rightarrow 0$  as  $n \rightarrow \infty$ . More generally, for  $p \geq 1$  there is convergence for the  **$L^p$ -norm** if  $\int |X_n(\omega) - X(\omega)|^p \Pr(d\omega) \rightarrow 0$ .
6.  $X_n$  converges to  $X$  in the **sup-norm** if  $\sup_{\Omega} |X_n - X| \rightarrow 0$ .

Pointwise convergence is the strongest notion, and it obviously implies almost sure convergence. Almost sure convergence implies convergence in probability, which implies convergence in distribution.  $L^1$  convergence implies convergence in probability. The figure lays out the relationships schematically.  $L^p$  convergence implies  $L^r$  convergence for  $p \geq r \geq 1$ . sup-norm convergence can be regarded as a special case of  $L^p$  as  $p \rightarrow \infty$ . Notice that since probability spaces have total probability (measure) 1, we are concerned about large values of  $X$  only. Random variables never fail to be integrable because of small values of  $X$ . (On  $[1, \infty)$  the variable  $X(x) = 1/x$  is divergent, but  $[1, \infty)$  does not have finite measure.)

Convergence in distribution is special to probability theory. It is equivalent to a number of other conditions, spelled out in the *Portmanteau theorem*, Billingsley (2012). In particular, on a standard probability space, convergence in distribution is equivalent to  $\Pr(X_n \in A) \rightarrow \Pr(X \in A)$  for all events  $A$  whose boundary has probability zero and to  $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(x)]$  for all bounded, continuous functions  $g$ . The last condition partially explains the condition for convergence in distribution using Fourier transforms (moment generating functions), since  $g(x) = e^{2\pi i x \theta}$  is bounded for fixed  $\theta$ .

The modes along the main horizontal line are all metrizable. For the  $L^p$  spaces, the topology is generated by the metric induced by the respective norm,  $d(X, Y) = \|X - Y\|_p$ . Convergence in probability is generated by the capping metric

$$d(X, Y) = \mathbb{P} \left[ \frac{|X - Y|}{1 + |X - Y|} \right]$$

on the space of random variables  $L^0$ . Convergence in distribution corresponds to the weak convergence of probability measures. For random variables taking values in a separable metric space (like  $\mathbb{R}$ ), this topology is metrizable using the Lévy-Prokhorov metric or the bounded Lipschitz metric. However, almost sure convergence is not metrizable or even topologizable. This follows from an analysis of the Urysohn property.

The relationships between the different modes of convergence are best understood by considering examples.

1. Convergence in probability but not almost surely.
  - $X_n$  independent with  $P(X_n = 1) = 1/n$  and  $P(X_n = 0) = 1 - 1/n$ ; let  $X = 0$ . Almost sure convergence  $X_n \rightarrow X$  would mean that for almost every  $\omega$ , there exists  $N$  such that  $X_n(\omega) = 0$  for all  $n \geq N$ . But for any fixed  $N$ ,  $P(X_n = 0 \text{ for all } n \geq N) = \prod_{n \geq N} (1 - \frac{1}{n}) = 0$  (take logs and use  $\log(1 - 1/n) < -1/n$  and the fact  $\sum_n 1/n$  diverges). Hence  $P(X_n(\omega) \rightarrow 0) = 0$ ,  $X_n$  does not converge to  $X$  almost surely.
  - (Typewriter sequence.) Each integer  $n \geq 1$  can be written uniquely as  $n = 2^m + k$  for  $0 \leq k < 2^m$ . Let  $X_n(\omega) = 1$  if  $\omega \in [k2^{-m}, (k+1)2^{-m}]$  and 0 otherwise, and  $X = 0$ . Then  $X_n$  converges to  $X$  in probability but not almost surely (for given  $\omega$ ,  $X_n(\omega) = 1$  for one  $k$  for each  $m$  and is zero otherwise, hence it takes the values 0 and 1 infinitely many times and so  $X_n(\omega)$  does not converge for any  $\omega$ ).
2. Convergence in distribution but not in probability.
  - Let  $X_n = X$  be Bernoulli and  $Y = 1 - X$ . Then  $X_n$  tends to  $X$  and  $Y$  in distribution (they have the same distribution) but not in probability because  $\Pr(X_n = Y) = \Pr(X = Y) = 0$ . Just as law invariant risk measures do not *see* the actual events, convergence in distribution does not consider explicit events.
  - The same example works if  $X$  is any non-trivial, symmetric random variable, and  $Y = -X$ .
  - Let  $X_n$  be uniform on  $k/n$  for  $k = 0, 1, \dots, n-1$  and  $X$  be uniform on  $[0, 1]$ . Then  $X_n$  converges to  $X$  in distribution (the distribution of  $X_n$  is a finer and finer stair-step function converging to the distribution of  $X$ ) but not probability (the distribution of  $X_n$  is supported on rational numbers, which have probability zero.)
3.  $L^1$  convergence or almost sure but not both.
  - $X_n(\omega) = n$  if  $\omega < 1/n$  and 0 otherwise converges to  $X = 0$  almost surely but not in  $L^1$ , since  $\int X_n = 1$  for all  $n$  but  $\int X = 0$ . Note  $X_n$  is unbounded; if  $X_n$  is dominated by an integrable function then Lebesgue's dominated convergence theorem ensures  $L^1$  convergence.
  - The typewriter sequence has  $L^1$  convergence but not almost sure convergence, since  $\int X_n \rightarrow 0$ . In fact, it converges in  $L^p$  for all  $p < \infty$ . It does not converge in  $L^\infty$  since  $\sup X_n = 1 \neq \sup X = 0$ .
4. Equivalent formulations for convergence in distribution.
  - The test function  $g$  must be continuous. Let  $X_n = 1/n$  with probability  $1 - 1/n$  and 1 otherwise.  $X_n$  converges to 0 in probability (for all  $\epsilon > 0$ ,  $\Pr(X_n > \epsilon) \rightarrow 0$  as  $n \rightarrow \infty$ ). Let  $g(x) = 0$  for  $x \leq 0$  and  $g(x) = 1$  for  $x > 0$ . For all  $n$ ,  $g(X_n) = 1$ , but  $g(0) = 0$ .
  - Test sets  $A$  must have a boundary of probability zero. Apply the third (uniform) example from (2) to  $A = \mathbb{Q} \cap [0, 1]$ , the rationals in  $[0, 1]$ .  $\Pr(X_n \in A) = 1$  for all  $n$ , but  $\Pr(X \in A) = 0$ . In this case the boundary of  $A$  is the set of all irrational numbers, which has probability 1 (the rationals have probability zero: they can be covered by an open set of arbitrarily small probability by putting an open interval of width  $\epsilon/2^{n+1}$  around the  $n$ th rational).
5. The strong law of large numbers is a statement that the sample mean converges to the true mean almost surely. For an iid sequence it holds iff  $E[|X_1|] < \infty$ .
6. The weak law of large numbers is a statement that the sample mean converges in probability, which is true under weaker conditions that do not require the mean exists, see Feller (1971).
7. The central limit theorem is a statement about the convergence of the distribution of the mean of a sample as the sample size increases.

### 8.3 Dual Pairs

Let  $X$  be a real vector space. Its algebraic dual, often written  $X^*$ , is the collection of all linear maps from  $X$  to  $\mathbb{R}$ . At this stage no topology is present. Now choose a subspace  $X' \subseteq X^*$  and regard the elements of  $X'$  as the linear probes with which points of  $X$  will be tested. There is then a canonical topology on  $X$ , denoted  $\sigma(X, X')$ , defined to be the weakest topology making every  $x' \in X'$  continuous. This topology is locally convex. If  $X'$  separates points of  $X$ , meaning that for every nonzero  $x \in X$  there exists  $x' \in X'$  with  $x'(x) \neq 0$ , then  $\sigma(X, X')$  is Hausdorff. In practice, we assume separation and call  $(X, X')$  a **dual pair**.

It is an important fact that once the topology  $\sigma(X, X')$  is imposed on  $X$ , its continuous dual is exactly the space  $X'$  we started with:

$$(X, \sigma(X, X'))' = X'.$$

This is the basic dual-pair construction. It is worth emphasizing because it explains a source of circularity

in the notation. Sometimes  $X'$  means the continuous dual of a previously given topology on  $X$ ; at other times one starts with a chosen  $X' \subseteq X^*$  and uses it to define the topology. In the second setup, the resulting continuous dual is precisely the chosen space  $X'$ . If  $X$  is a normed space, then its topological dual with respect to the norm topology is what is usually denoted by  $X'$  in functional analysis. Thus the symbol  $X'$  only acquires meaning after a topology on  $X$  has been specified.

Let  $X$  be a Banach space with its norm topology, and let  $X'$  denote its full continuous dual with respect to that norm. Then  $\sigma(X, X')$  is the usual weak topology on  $X$ . A natural first guess is that because we are using all norm-continuous linear functionals, the resulting weak topology ought to recover the norm topology. In infinite dimensions this is false. The weak topology is generally strictly weaker than the norm topology. This feels surprising only until we notice that a topology is not determined merely by the set of continuous linear functionals, but by how neighborhoods are built out of them. A basic weak neighborhood of the origin has the form

$$\{x : |x'_1(x)| < \varepsilon, \dots, |x'_n(x)| < \varepsilon\},$$

with only finitely many functionals at a time. By contrast, norm control is uniform control over the entire dual unit ball, as expressed by the Hahn-Banach formula

$$\|x\| = \sup_{\|x'\| \leq 1} |x'(x)|.$$

Thus weak convergence means that each fixed functional sees convergence, whereas norm convergence requires simultaneous uniform control over all norm-bounded functionals.

These topological distinctions are clarified through an embedding into a product space. Given a separating pair  $(X, X')$ , define the evaluation map  $J : X \rightarrow \mathbb{R}^{X'}$  by  $J(x) = (x'(x))_{x' \in X'}$ . Then  $\sigma(X, X')$  is exactly the subspace topology induced on  $J(X)$  from the product topology on  $\mathbb{R}^{X'}$ . Weak convergence is therefore nothing more than coordinatewise convergence under this embedding: a net  $x_\alpha$  converges to  $x$  in  $\sigma(X, X')$  if and only if  $x'(x_\alpha) \rightarrow x'(x)$  for every  $x' \in X'$ . This makes the comparison with norm topology very intuitive. The product topology controls only finitely many coordinates at once, whereas the norm behaves like a supremum over an entire bounded family of coordinates. Weak convergence is therefore analogous to pointwise convergence, while norm convergence is analogous to uniform convergence.

The standard examples on  $L^\infty$  and  $L^1$  illustrate the distinction. On  $L^\infty[0, 1]$ , paired with  $L^1[0, 1]$  via  $\langle X, Z \rangle = \mathbb{P}(XZ)$ , consider the Rademacher functions  $r_n(t) = \text{sign}(\sin 2^n \pi t)$ . Each  $r_n$  takes only the values  $\pm 1$ , so its  $L^\infty$  norm is always one. Hence the sequence cannot converge to zero in norm. On the other hand, for every  $g \in L^1[0, 1]$  we have  $\int_0^1 r_n(t)g(t) dt \rightarrow 0$ , and so  $r_n \rightarrow 0$  in the weak topology  $\sigma(L^\infty, L^1)$ . Thus the weak topology is strictly weaker than the norm topology on  $L^\infty$ . This phenomenon is closely related in spirit to the Riemann-Lebesgue theorem: highly oscillatory functions may fail to become small in norm, but they wash out against each fixed integrable test function.

A related example appears on  $L^1[0, 1]$ , paired with  $L^\infty[0, 1]$ . The exponential functions  $f_n(t) = e^{2\pi i n t}$  all have  $L^1$  norm equal to one, so again there is no convergence to zero in norm. But for every bounded measurable  $h$  one has  $\int_0^1 f_n(t)h(t) dt \rightarrow 0$ , and hence  $f_n \rightarrow 0$  in the weak topology  $\sigma(L^1, L^\infty)$ . This is perhaps even closer in flavor to the Riemann-Lebesgue theorem. The message of both examples is that weak convergence records how a sequence behaves when tested against each fixed dual element, while norm convergence records its absolute size.

The idea of a dual pair is especially important for  $L^\infty$ , because several different duals appear. First there is the algebraic dual, consisting of all linear maps from  $L^\infty$  to  $\mathbb{R}$ ; this space is enormous and usually too large to handle concretely. Second there is the Banach dual, meaning the full continuous dual of  $L^\infty$  with respect to the norm topology. This is the space  $ba$  of bounded finitely additive signed measures absolutely continuous with respect to the reference probability. Third, one often works with the smaller space  $L^1$ , viewed as a subspace of the Banach dual via integration. These three spaces should not be confused. In particular,  $ba$  is not the algebraic dual of  $L^\infty$ ; it is the full norm-continuous dual.

Since  $ba$  is the full continuous dual of normed  $L^\infty$ , the topology  $\sigma(L^\infty, ba)$  is the usual weak topology on  $L^\infty$ . It is still strictly weaker than the norm topology, because  $L^\infty$  is infinite-dimensional. The topology  $\sigma(L^\infty, L^1)$  is weaker again, because  $L^1$  is a proper subspace of  $ba$ . Thus one has the strict chain

$$\sigma(L^\infty, L^1) \subsetneq \sigma(L^\infty, ba) \subsetneq \text{norm topology}.$$

The first inclusion reflects the fact that countably additive functionals are only part of the full Banach dual; the second reflects the general distinction between weak and norm topologies on infinite-dimensional normed spaces.

This hierarchy is exactly what lies behind the distinction between countably additive and finitely additive dual representations in risk-measure theory. If one works with the dual pair  $(L^\infty, L^1)$ , then lower semicontinuity means  $\sigma(L^\infty, L^1)$ -lower semicontinuity, and this is the topology relevant for Fatou-type results and exact dual representations over countably additive densities. If one instead works with the full dual pair  $(L^\infty, ba)$ , then one enters the larger finitely additive world. The passage from  $ba$  down to  $L^1$  is therefore not a mere technicality: it changes the topology, and hence changes which convex functionals are lower semicontinuous and which dual representations are exact.

All of the subtlety here comes from the fact that  $X$  is infinite-dimensional. In finite dimensions, the distinctions largely collapse: every linear functional is automatically continuous for the Euclidean norm, or any equivalent locally convex topology; the weak and norm topologies on  $X$  coincide; and likewise the weak-\* and norm topologies on  $X'$  coincide. Equivalently, boundedness, continuity, and convergence all become much simpler because there is no distinction between coordinatewise and uniform control once only finitely many directions are present. Thus the entire phenomenon of weak convergence without norm convergence, and the resulting distinction between different dual topologies, is a genuinely infinite-dimensional feature.

## 8.4 Lower Semicontinuity

**Definition 8.1.** Let  $X$  be a topological space, and let  $f : X \rightarrow (-\infty, \infty]$ . Then  $f$  is **lower semicontinuous**, abbreviated lsc, if for every  $c \in \mathbb{R}$  the sub-level set

$$\{x \in X : f(x) \leq c\}$$

is closed. Equivalently,  $f$  is lower semicontinuous if for every  $c \in \mathbb{R}$  the strict super-level set  $\{x \in X : f(x) > c\}$  is open.

If  $X$  is described by a convergence structure, the equivalent net formulation is:  $f$  is lower semicontinuous if whenever  $x_\alpha \rightarrow x$  in the given topology,

$$f(x) \leq \liminf_\alpha f(x_\alpha).$$

In metric or first-countable spaces, sequences suffice in place of nets.

*Remark.* Here is a proof of the equivalence in the sequence case. Assume first that sub-level sets are closed, and let  $x_n \rightarrow x$ . Write

$$\ell := \liminf_{n \rightarrow \infty} f(x_n).$$

If  $\ell = +\infty$  there is nothing to prove, so assume  $\ell < \infty$ . Suppose for contradiction that  $f(x) > \ell$ . Choose  $c$  with

$$\ell < c < f(x).$$

By the definition of  $\liminf$ , infinitely many  $n$  satisfy  $f(x_n) \leq c$ . Hence there is a subsequence  $(x_{n_k})$  such that

$$f(x_{n_k}) \leq c \quad \text{for all } k.$$

Thus every  $x_{n_k}$  lies in the closed sublevel set  $F_c := \{y \in X : f(y) \leq c\}$ . Since  $x_{n_k} \rightarrow x$  and  $F_c$  is closed, we must have  $x \in F_c$ , so  $f(x) \leq c$ , contradicting  $c < f(x)$ . Therefore  $f(x) \leq \liminf_{n \rightarrow \infty} f(x_n)$ .

Next, assume the  $\liminf$  condition, and suppose  $F_c := \{x : f(x) \leq c\}$  is not closed for some  $c$ . In a sequential space this means there exists a sequence  $(x_n)$  in  $F_c$  with  $x_n \rightarrow x$  and  $x \notin F_c$ . Since each  $x_n \in F_c$ ,  $f(x_n) \leq c$  for all  $n$ , and therefore  $\liminf_{n \rightarrow \infty} f(x_n) \leq c$ . But the  $\liminf$  condition gives  $f(x) \leq \liminf_{n \rightarrow \infty} f(x_n) \leq c$ , which says  $x \in F_c$ , a contradiction. Hence  $F_c$  is closed.

Convex functions have nice lower semicontinuity behavior. The reason is geometric. The natural geometric object associated with a function  $f$  is its epigraph,

$$\text{epi}(f) = \{(x, t) \in X \times \mathbb{R} : t \geq f(x)\}.$$

This set records the region on or above the graph of the function. A function is convex if and only if its epigraph is a convex set, and a function is lower semicontinuous if and only if its epigraph is closed in the product topology on  $X \times \mathbb{R}$ . Thus, for convex functions, lower semicontinuity is exactly the statement that its epigraph is convex and closed.

Closed convex sets in a locally convex space can be recovered from their supporting half-spaces. More precisely, if a point lies outside a closed convex set, then the Hahn-Banach separation theorem produces a continuous affine functional that separates the point from the set. In the case of an epigraph, such a separating affine functional is exactly a supporting affine minorant of the function. The function can then be reconstructed as the supremum of all its supporting affine minorants. That is the conceptual content of the Fenchel-Moreau theorem. Convexity provides the right geometry; lower semicontinuity ensures that no boundary points are missing; and separation by continuous linear functionals turns geometry into a dual representation.

Suppose  $X$  is a Banach space with its norm topology, and let  $X'$  be its full continuous dual. Then the weak topology  $\sigma(X, X')$  is strictly weaker than the norm topology whenever  $X$  is infinite-dimensional. For arbitrary functions, norm-lower-semicontinuity and weak-lower-semicontinuity are different. For convex functions, by contrast, they coincide. The reason is again geometric: a convex set is norm closed if and only if it is weakly closed. Applying this to the epigraph shows that a convex function on a Banach space is norm-lower-semicontinuous if and only if it is weakly lower semicontinuous with respect to the full dual. This is one of the central simplifications of convex analysis. In the presence of convexity, the weak topology induced by the full continuous dual is already strong enough to detect all closed convex sets.

## 8.5 Convergence, Compactness, and the Geometry of Banach Spaces

In functional analysis, the behavior of a space is dictated by the tests linear functionals that define its topology. Whether we are working in  $L^p$  spaces or general Banach spaces, the hierarchy of these topologies can be the difference between a space where the unit ball is a solid compact object and one where it is an empty non-compact shell.

In a **dual pair**  $(X, Y)$ ,  $X$  is a vector space and  $Y$  is a subspace of the algebraic dual. (The algebraic dual includes all linear functionals. The topological or continuous dual is a subspace of the algebraic dual consisting only of those linear functionals that are continuous with respect to the topology on  $X$ . The algebraic dual of  $X$  is often  $X^*$  and the topological dual  $X'$ .) The topology  $\sigma(X, Y)$  is the coarsest (weakest) topology such that every  $y \in Y$  is continuous.

Consider two sets of test functions  $Y_1 \subset Y_2 \subset X'$ . When there are more tests there are more open sets. As we move from  $Y_1$  to  $Y_2$ , we are demanding that more functions be continuous. To satisfy this, we must add more pre-images to our collection of open sets. Thus,  $\sigma(X, Y_2)$  has more open sets than  $\sigma(X, Y_1)$ . We call  $\sigma(X, Y_2)$  a stronger (or finer) topology and  $\sigma(X, Y_1)$  weaker (or coarser) one. It also becomes harder to for a sequence to converge: a sequence  $x_n \rightarrow x$  in  $\sigma(X, Y_2)$  must pass every test in  $Y_2$ . Since  $Y_2$  is larger, it is harder to converge. Conversely, it is easier for a set to be compact in the weaker topology  $\sigma(X, Y_1)$  because there are fewer open covers.

The fundamental trade-off in infinite-dimensional spaces is between the resolution of the topology and the compactness of the unit ball. In the norm topology (the strongest standard topology), the unit ball is never compact if the space is infinite-dimensional. (Rademacher functions have norm 1 but  $L^1$  integral that  $\rightarrow 0$ .)

**Theorem 8.1** (Banach-Alaoglu). *Let  $X$  be a normed vector space. The closed unit ball  $B_{X'} = \{f \in X' : \|f\| \leq 1\}$  in the dual space  $X'$  is compact in the weak-\* topology  $\sigma(X', X)$ .*

By weakening the topology of the dual space until only the evaluations at points in  $X$  are continuous, we collapse the space enough to regain compactness.

**Example 8.1.** In the Hilbert space  $\ell^2$ , consider the standard orthonormal basis  $\{e_n\}_{n=1}^\infty$ . Under the strong topology the distance between any two distinct basis vectors is  $\|e_n - e_m\| = \sqrt{2}$ . Thus, the set  $E = \{e_n : n \in \mathbb{N}\}$  is closed and has no limit points in the norm topology. Specifically,  $e_n \not\rightarrow 0$  strongly. However, under the weak topology, for any  $y \in \ell^2$ , the action  $y(e_n) = \langle e_n, y \rangle$  is simply the  $n$ -th coordinate of  $y$ . By Bessel's inequality,  $y(e_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Therefore,  $e_n \rightarrow 0$  weakly. This shows that 0 is in

the weak closure of  $E$ , even though it is not in the strong closure. This smearing occurs because  $E$  is not a convex set.

**Theorem 8.2** (Mazur). *Let  $C$  be a convex subset of a locally convex space  $X$ . Then  $C$  is weakly closed if and only if it is strongly (norm) closed.*

Let  $X$  be a topological space and let  $A \subseteq X$ .  $A$  is **relatively compact** if its closure  $\bar{A}$  is compact. If  $X$  carries the weak topology, then  $A$  is relatively weakly compact if it is relatively compact for the weak topology; equivalently, the weak closure of  $A$  is weakly compact.

A family  $\mathcal{K} \subseteq L^1$  is **uniformly integrable** if

$$\lim_{M \rightarrow \infty} \sup_{X \in \mathcal{K}} \mathbb{P}(|X|1_{\{|X| > M\}}) = 0.$$

Equivalently, for every  $\varepsilon > 0$  there exists  $M < \infty$  such that

$$\sup_{X \in \mathcal{K}} \mathbb{P}(|X|1_{\{|X| > M\}}) < \varepsilon.$$

Thus uniform integrability means that the tails of the family can be made uniformly negligible: no mass can escape to infinity in a way that depends on the chosen member of the family.

**Theorem 8.3** (Dunford–Pettis). *A subset  $\mathcal{K} \subseteq L^1$  is relatively weakly compact, for the topology  $\sigma(L^1, L^\infty)$ , if and only if it is uniformly integrable.*

*Remark.* In practice, Dunford–Pettis is the compactness theorem that turns integrability control into weak compactness on the dual side. In the theory of convex risk measures, it is used to show that suitable sublevel sets of the penalty function  $\rho^*$  are weakly compact in  $L^1$ ; once that is known, the map

$$Z \mapsto \mathbb{P}(XZ) - \rho^*(Z)$$

attains its supremum on those sets, yielding existence of a dual optimizer.

## 8.6 Filters and Ultrafilters

A **filter**  $\mathcal{F}$  on  $\Omega$  is a collection of subsets of  $\Omega$  such that

1.  $\emptyset \notin \mathcal{F}$ ,
2. if  $A, B \in \mathcal{F}$  then  $A \cap B \in \mathcal{F}$ ,
3. if  $A \in \mathcal{F}$  and  $A \subseteq B \subseteq \Omega$ , then  $B \in \mathcal{F}$ .

A filter is a family of sets closed upward under inclusion and closed under finite intersections. If a set is large, then any larger set is also large. If two sets are large, then their overlap is still large.

The simplest example is a principal filter. Fix a point  $\omega_0 \in \Omega$ . Then

$$\mathcal{F}_{\omega_0} = \{ A \subseteq \Omega : \omega_0 \in A \}$$

is a filter. Here “large” simply means “contains  $\omega_0$ ”. More generally, if  $E \subseteq \Omega$  is nonempty, then

$$\mathcal{F}_E = \{ A \subseteq \Omega : E \subseteq A \}$$

is the filter generated by  $E$ . A filter is called **principal** if it is of this form for some nonempty  $E$ .

In a probability space, sets of measure 1 form a filter. The dual is the ideal of null sets.

A basic non-principal example is the Fréchet filter on  $\mathbb{N}$ :

$$\mathcal{F}^{\text{cof}} = \{ A \subseteq \mathbb{N} : A^c \text{ is finite} \}$$

containing the cofinite sets. A set is large if it contains all but finitely many integers. This filter is not principal. No finite set of indices determines it, and indeed no single point plays a distinguished role. This is the first hint that non-principal filters capture an asymptotic notion of largeness.

If  $\mathcal{A}$  is a family of subsets of  $\Omega$  with the finite intersection property, meaning that every finite intersection of members of  $\mathcal{A}$  is nonempty, then  $\mathcal{A}$  **generates** a filter:

$$\langle \mathcal{A} \rangle = \{ B \subseteq \Omega : \exists A_1, \dots, A_m \in \mathcal{A} \text{ with } A_1 \cap \dots \cap A_m \subseteq B \}.$$

In words, a set is large if it contains some finite intersection of the prescribed large sets. If the family  $\mathcal{A}$  is nested decreasing this simplifies: the generated filter consists of all sets that contain some  $A_n$ .

An **ultrafilter**  $\mathcal{U}$  on  $\Omega$  is a filter that is maximal among proper filters. Equivalently,  $\mathcal{U}$  is an ultrafilter if for every subset  $A \subseteq \Omega$ ,

$$A \in \mathcal{U} \quad \text{or} \quad A^c \in \mathcal{U},$$

and exactly one of these holds. This dichotomy is what makes ultrafilters powerful. Every set is either large or its complement is large; there is no middle ground.

A filter leaves many sets undecided. An ultrafilter is what you get by adjoining as many large sets as possible without ever forcing the empty set to become large.

A **principal ultrafilter** is one of the form

$$\mathcal{U}_{\omega_0} = \{ A \subseteq \Omega : \omega_0 \in A \}.$$

On an infinite set there are also non-principal ultrafilters, but their existence requires a choice principle. A filter on an infinite set is called **free** if it is not principal.

A standard theorem says:

**Theorem 8.4.** *Every proper filter on a set  $\Omega$  is contained in an ultrafilter.*

Equivalently, every family of sets with the finite intersection property (the intersection of any finite subcollection is non-empty) extends to an ultrafilter. The ultrafilter theorem is weaker than the full axiom of choice, but it is not provable in ordinary ZF alone. For our purposes, we use it as a standard existence result.

If  $\mathcal{F}$  is a filter, the ultrafilter theorem says there exists at least one ultrafilter  $\mathcal{U}$  with

$$\mathcal{F} \subseteq \mathcal{U}.$$

But in general the extension is very far from unique. An ultrafilter extension is a way of making a yes/no decision about every set not already decided by  $\mathcal{F}$ , and there are usually many such decisions available.

For the cofinite filter on  $\mathbb{N}$ , this is especially dramatic. The cofinite filter only declares sets with finite complement to be large. But there are infinitely many other sets—evens, odds, primes, unions of blocks, and so on—for which neither the set nor its complement is cofinite. An ultrafilter extending the cofinite filter must choose exactly one of each such pair

$$A, \quad A^c,$$

and must do so consistently with finite intersections. There are therefore many different extensions. Indeed, on  $\mathbb{N}$  there are not just many but enormously many non-principal ultrafilters extending the cofinite filter: in fact

$$2^{2^{\aleph_0}}$$

of them. So the cofinite filter is a very small amount of asymptotic information, and an ultrafilter extension is a huge refinement of it. That is why the notation

$$\lim_{\mathcal{U}} x_n$$

depends strongly on the choice of  $\mathcal{U}$ : different ultrafilters extending the same cofinite filter can produce different generalized limits for the same bounded sequence.

## 8.7 Alternative Primals: Orlicz Spaces and Orlicz Hearts

The theory we have developed so far lives in  $L^\infty$ . Boundedness is a simplification, but many losses of practical and theoretical interest are unbounded, including the normal and lognormal. We can extend to  $L^p$  spaces for  $1 \leq p < \infty$ , where we gain polynomial control over tails. For example,  $L^1$  corresponds to variables with a mean,  $L^2$  with a variance, and so forth. Orlicz spaces and hearts are more differentiating. They can distinguish, for example, between exponential, stretched-exponential, lognormal, and power-law tails, and so provide a more natural setting when the finiteness of a risk measure or the form of its dual representation depends on more than the existence of a few moments. Orlicz spaces also provide

a natural domain for convex duality beyond  $L^\infty$ , when we wish to accommodate unbounded losses. An Orlicz space is built from a convex growth function, called a Young function, and they come in two versions: the full Orlicz space, and the smaller Orlicz heart. See [my blog post](#), the books [Krasnosel'skii and Rutickii \(1961\)](#) and [Rao and Ren \(1991\)](#), and the papers [Cheridito and Li \(2008\)](#), [Cheridito and Li \(2009\)](#), [Arai \(2009\)](#), [Delbaen \(2010\)](#), [Gao et al. \(2018\)](#) and [Gao, Leung, and Xanthos \(2019\)](#) for more details.

## 8.8 Proof of Lemma 3.1

The proof splits into two steps. The first step is probabilistic and elementary: bounded almost sure closure implies bounded closure in probability. The second step is functional-analytic: for convex sets in  $L^\infty$ , closure in probability on every norm-bounded ball implies  $\sigma(L^\infty, L^1)$ -closedness.

We first show that the hypothesis in Lemma 3.1 already implies the following stronger sequential closure property: whenever  $(X_n) \subseteq C$  is uniformly bounded in  $L^\infty$  and converges in probability to  $X$ , then  $X \in C$ .

Indeed, assume

$$(X_n) \subseteq C, \quad \sup_n \|X_n\|_\infty \leq M, \quad X_n \rightarrow X \text{ in probability.}$$

*Remark 8.1.* If  $X_n \rightarrow X$  in probability, then every subsequence  $(X_{n_k})$  has a further subsequence  $(X_{n_{k_j}})$  such that

$$X_{n_{k_j}} \rightarrow X \quad \text{almost surely.}$$

Let  $(X_{n_k})$  be any subsequence. Since  $X_{n_k} \rightarrow X$  in probability, for each  $j \geq 1$  we can choose  $k_j$  increasing so that

$$\mathbf{P}(|X_{n_{k_j}} - X| > 2^{-j}) < 2^{-j}.$$

Then

$$\sum_{j=1}^{\infty} \mathbf{P}(|X_{n_{k_j}} - X| > 2^{-j}) < \infty.$$

By the Borel–Cantelli lemma, with probability 1 only finitely many of the events

$$\{|X_{n_{k_j}} - X| > 2^{-j}\}$$

occur. Hence for almost every  $\omega$  there exists  $J(\omega)$  such that for all  $j \geq J(\omega)$ ,

$$|X_{n_{k_j}}(\omega) - X(\omega)| \leq 2^{-j}.$$

Since  $2^{-j} \rightarrow 0$ , we obtain

$$X_{n_{k_j}}(\omega) \rightarrow X(\omega) \quad \text{for almost every } \omega.$$

Thus  $X_{n_{k_j}} \rightarrow X$  almost surely.

Using Remark 8.1 we obtain a further subsequence converging almost surely to  $X$ . For completeness, choose inductively integers  $n_k$  so that

$$\mathbf{P}(|X_{n_k} - X| > 2^{-k}) < 2^{-k}.$$

Then

$$\sum_{k=1}^{\infty} \mathbf{P}(|X_{n_k} - X| > 2^{-k}) < \infty.$$

By the Borel–Cantelli lemma, for almost every  $\omega$  only finitely many of the events

$$\{|X_{n_k} - X| > 2^{-k}\}$$

occur. Hence  $X_{n_k}(\omega) \rightarrow X(\omega)$  almost surely. Since the subsequence remains uniformly bounded by  $M$ , the assumed closure property of  $C$  gives  $X \in C$ .

Thus  $C$  is closed under convergence in probability on each norm-bounded subset of  $L^\infty$ .

For the second step, fix  $M < \infty$  and consider the truncated set

$$C_M := C \cap \{X \in L^\infty : \|X\|_\infty \leq M\}.$$

Since  $C$  is convex,  $C_M$  is convex. Step 1 shows that  $C_M$  is closed in probability.

We claim that  $C_M$  is  $\sigma(L^\infty, L^1)$ -closed. Here one uses a standard fact about bounded subsets of  $L^\infty$ : on a norm-bounded set, convex closure in probability agrees with closure for the weak topology  $\sigma(L^\infty, L^1)$ . Intuitively, probability convergence is strong enough on bounded sets to control all linear functionals of the form

$$X \mapsto \mathbb{P}(XZ), \quad Z \in L^1,$$

because boundedness lets us use dominated convergence after passing to almost surely convergent subsequences. Convexity then upgrades the sequential probabilistic closure to topological closure for the dual-pair topology.

A convenient way to package the final step is through the Krein–Smulian theorem. In the present setting, the relevant message is the following: to prove that a convex set in the dual space  $L^\infty = (L^1)^*$  is closed for the weak-star topology  $\sigma(L^\infty, L^1)$ , it is enough to check weak-star closedness after intersecting with each closed norm ball. In other words, one does not need to control the whole set at once; one only needs to control

$$C \cap \{X : \|X\|_\infty \leq M\}$$

for each fixed  $M$ .

Applying that principle here, it suffices to show that each  $C_M$  is  $\sigma(L^\infty, L^1)$ -closed. But, as noted above, on the bounded ball  $\{X : \|X\|_\infty \leq M\}$  the convex probability-closed sets are exactly the convex  $\sigma(L^\infty, L^1)$ -closed sets. Therefore each  $C_M$  is  $\sigma(L^\infty, L^1)$ -closed. Krein–Smulian then yields that  $C$  itself is  $\sigma(L^\infty, L^1)$ -closed.

That proves Lemma 3.1.

## 8.9 Sketch Proof of Theorem 3.2

We sketch the main ideas and refer to Föllmer and Schied (2016) the technical details. In the payoff sign convention used by Föllmer and Schied, condition (1) is written as continuity from below.

The starting point is the  $ba$  representation. Since  $\rho$  is convex and monetary, it is Lipschitz in the  $L^\infty$  norm, hence norm continuous. Fenchel–Moreau therefore gives an exact dual representation over  $ba$ , and the minimal penalty  $\alpha_{\min}$  agrees with the convex conjugate relative to the dual pair  $(L^\infty, ba)$ :

$$\rho(X) = \max_{\nu \in ba} \nu(X) - \alpha_{\min}(\nu).$$

Thus every  $X$  admits at least one maximizing finitely additive dual variable.

To prove (1)  $\Rightarrow$  (2), decompose  $\nu \in ba_+$  into its Yosida–Hewitt decomposition

$$\nu = \nu^r + \nu^s,$$

where  $\nu^r$  is countably additive and  $\nu^s$  is purely finitely additive. One shows that if  $\nu^s \neq 0$ , then continuity from above forces

$$\alpha_{\min}(\nu) = \infty.$$

The key idea is to exploit the defining property of a purely finitely additive measure: there exists a decreasing sequence of sets  $A_n \downarrow \emptyset$  almost surely such that  $\nu^s(A_n)$  does not tend to 0 (see ?). Since  $A_n \downarrow 0$  in  $L^\infty$ , continuity from above gives

$$\rho(-mA_n) \downarrow \rho(0) = 0 \quad \text{for each fixed } m > 0,$$

whereas the singular mass of  $\nu^s$  on the sets  $A_n$  makes

$$\nu(-mA_n) - \alpha_{\min}(\nu)$$

arbitrarily large unless  $\alpha_{\min}(\nu) = \infty$ . Hence finite penalty excludes the singular part, and only countably additive probabilities remain in the effective domain.

To prove (2)  $\Rightarrow$  (3), start from the exact *ba* representation above. For each  $X$ , choose a maximizer  $\nu_X \in \text{ba}$ . By (2), finite penalty implies that  $\nu_X$  is actually countably additive, so  $\nu_X \in L^1$ . Therefore the same maximizer attains the supremum in the  $L^1$  dual representation:

$$\rho(X) = \max_{Z \in L^1} \mathbb{P}(XZ) - \rho^*(Z).$$

To prove (3)  $\Rightarrow$  (1), let  $X_n \downarrow X$  almost surely. By monotonicity, the sequence  $\rho(X_n)$  is decreasing, so it has a limit  $\ell \geq \rho(X)$ . Choose  $Z \in L^1$  attaining the dual maximum for  $X$ :

$$\rho(X) = \mathbb{P}(XZ) - \rho^*(Z).$$

Then for every  $n$ ,

$$\rho(X_n) \geq \mathbb{P}(X_n Z) - \rho^*(Z).$$

Since  $X_n \downarrow X$  and  $Z \in L^1$ , dominated convergence gives

$$\mathbb{P}(X_n Z) \rightarrow \mathbb{P}(XZ).$$

Hence

$$\liminf_n \rho(X_n) \geq \rho(X).$$

Because  $\rho(X_n) \downarrow \ell$  and always  $\rho(X) \leq \ell$  by monotonicity, we conclude  $\ell = \rho(X)$ . Thus  $\rho$  is continuous from above.

::: {remark ?} Recall that nonprincipal ultrafilter measures can be used to create purely finitely additive measures. Countably additive measures satisfy continuity from above, but purely finitely additive measure can violate that property. For example, consider a nonprincipal ultrafilter  $U$  on  $\mathbb{N}$ . The associated  $\{0, 1\}$ -valued finitely additive probability  $\nu_U$  assigns mass 1 to every cofinite set, and hence to each tail set

$$A_n = \{n, n + 1, n + 2, \dots\}.$$

Then  $A_n$  is a sequence  $A_n \downarrow \emptyset$  while  $\nu_U(A_n) = 1$  for all  $n$ . Not every purely finitely additive measure is literally an ultrafilter measure, but the ultrafilter example is a good guide. It captures the idea that purely finitely additive mass can hide in an “infinitesimal tail” that survives every finite truncation even though the sets themselves shrink to nothing. In the duality theory of risk measures, continuity from above rules out exactly that kind of hidden singular mass. :::

## 9 Atlas Redux

This section contains a summary table of the results we have discussed, and three schematics: the construction triangle, showing how we can build a risk measure directly, or from its dual, or from its acceptance set; the correspondence between primal axioms to dual behavior; and the continuity hierarchy. The last two are extracted from the Table 2.

### 9.1 Summary Table

In Table 2 we reach the end of our all exploring and summarize the main results we have covered in the Atlas. The table includes internal references. The functional  $\rho$  is assumed to be monetary and convex, and to have the form

$$\rho(X) = \sup \nu(X) - \alpha(X)$$

for a penalty function  $\alpha$ , with a specified dual space. Thus  $\rho = \alpha^*$ , and  $\alpha = \rho^*$  if  $\alpha$  is convex and lsc. We call a (possibly finitely additive) measure a probability if it is positive and its total measure equals 1.

Table 2: Major results and implications. The functional  $\rho$  is a dual; unnamed functions assume MRM but not the dual form. The statement, “[A:] B  $\implies$  C” is read as “If MRM [and A], then B implies C.”

Ref	Statement	Proof
01.	$\rho$ is convex, lsc	Sup of affine functions
02.	ess sup is CFB; and ess inf is CFA	Indicators on $(0, 1/n)$ and $(1/n, 1)$ rule out CFA/B
03.	MRM: Lipschitz and hence LSC wrt norm $\ \cdot\ _\infty$ topology	Proposition 2.1
04.	$\rho$ is cash invariant $\iff \text{dom } \alpha \subseteq \{Z : \mathbb{P}Z = 1\}$ (total variation 1)	Proposition 3.3
05.	$\rho$ is monotone $\iff \text{dom } \alpha \subseteq \{Z : Z \geq 0\}$	Proposition 3.3
06.	$\rho$ is proper and finite at some $X \iff \text{dom } \alpha \subseteq \{Q : Q \ll \mathbb{P}\}$	Proposition 3.3
07.	$\rho$ is normalized $\iff \inf \rho^*(Z) = 0$	Proposition 3.3
08.	$\sup_i \rho_i$ is CX MRM and $(\sup_i \rho_i)^* = \inf_i \rho_i^*$	Proposition 3.5
09.	MRM + CX: exact <i>ba</i> dual representation over probabilities (Fenchel-Moreau duality)	$L' = ba$ , Lipschitz, Theorem 3.1
10.	CX: COH $\iff \alpha = \text{convex indicator}$	Proposition 3.6
11.	$\rho$ : LEB $\iff$ CFA and CFB	$Y_n := \inf_{k \geq n} X_k, Z_n := \sup_{k \geq n} X_k$ trick
12.	$\rho$ : CFB $\iff$ Fatou	Lemma 8.1 w $Y_n$
13.	$\rho$ : LEB $\implies$ Fatou	LEB $\implies$ CFB
14.	CX: Fatou $\iff$ LSC wrt $\sigma(L^\infty, L^1)$	Proposition 3.4
15.	CX + Fatou: $L^1$ duality (no fa measures)	Proposition 3.4, general theory
16.	CX + Fatou: CFA $\iff$ LEB $\iff$ sup is achieved in $L^1$	Theorem 3.2, Proposition 8.1
17.	COMON + CX $\implies$ COH	Corollary 2.1
18.	NORM: COMON $\implies$ Choquet representation (Schmeidler’s theorem)	Theorem 4.2
19.	Atomless, CX, Fatou: $\rho$ LI $\iff \rho^*$ LI	Proposition 3.7
20.	Atomless, CX: LI $\iff$ respects SSD	Theorem 4.1
21.	Atomless: LI + CX $\implies \sigma(L^\infty, L^1)$ LSC (JST’s theorem)	Theorem 4.3
22.	Atomless, LI + COH + Fatou $\implies$ Kusuoka representation	Theorem 4.4
23.	CX: closed in $L^1$ for a.s. unif convg $\implies \sigma(L^\infty, L^1)$ -closed	Lemma 3.1
24.	CX, LI, closed $ \cdot _\infty \implies \sigma(L^\infty, L^q)$ -closed $\forall 1 \leq q < \infty$	Proposition 4.1
25.	CX, LI, closed $ \cdot _\infty \implies$ closed under conditional exptns	Lemma 4.1

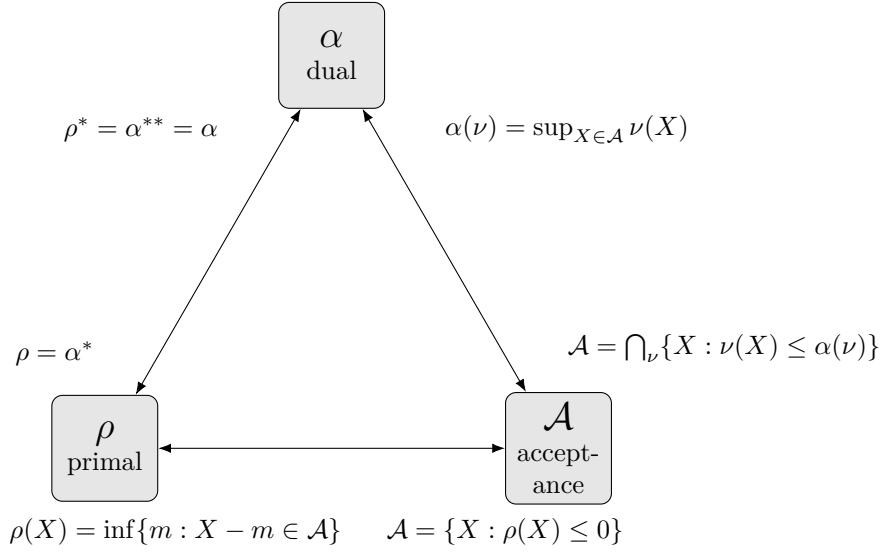


Figure 7: Starting from a convex, lsc  $\alpha$ .

*Remark.* Abbreviations: COMON(otonic) additive, CX convex, C(ontinuous) F(rom) A(bove)/B(elow), LEB(esgue), NORM(alized), L(aw) I(nvariant), MRM monetary risk measure, SSD second-order stochastic dominance.

## 9.2 Construction Triangle

Figure 7 shows the triangle of equivalent constructions starting from a convex, lsc monetary risk measure  $\rho$  on  $L^\infty$ , or  $\alpha$  on a dual space  $L'$ , or with the (closed) convex acceptance set of  $\mathcal{A}$  of  $\rho$ . The minimal penalty function is discussed in Föllmer and Schied (2016) Theorem 4.16

If we start with a convex, lsc penalty function  $\alpha$  on a chosen dual space  $L' \subset ba$  that separates points, and define

$$\rho(X) := \sup_{\nu \in L'} \nu(X) - \alpha(\nu),$$

then:

1.  $\rho$  is convex and  $\sigma(L^\infty, L')$ -lsc, because it is a supremum of affine functions of  $X$ .
2.  $\rho = \alpha^*$  and so  $\rho^* = \alpha^{**} = \alpha$  by duality for  $\alpha$ .
3. If  $\text{dom } \alpha \subseteq \{\nu : \nu \geq 0, \nu(\Omega) = 1\}$ , then  $\rho$  is monotone and cash invariant, and hence monetary.

If we start with a functional  $\rho$  that is monotone and cash invariant. Let  $\mathcal{A} = \{X \in L^\infty : \rho(X) \leq 0\}$  be the acceptance set of  $\rho$ . Define a function called the minimal penalty function (name to be justified)

$$\alpha_{\min}(\nu) := \sup_{X \in \mathcal{A}} \nu(X),$$

then:

1.  $\rho^*$ , the dual of  $\rho$  is supported in positive probability measures,  $\text{dom } \rho^* \subset \{\nu : \nu \geq 0, \nu(\Omega) = 1\}$
2.  $\alpha_{\min}$  is convex since sup is convex.
3.  $\alpha_{\min}$  is lsc since  $\{\nu : \alpha_{\min}(\nu) \leq c\} = \bigcap_{X \in \mathcal{A}} \{\nu : \nu(X) \leq c\}$ .
4. If  $\alpha$  represents  $\rho$  and is convex and  $\sigma(L', L^\infty)$ -lsc, then by 2 and 3 and biconjugacy,  $\rho^* = \alpha^{**} = \alpha$ .
5. It is easy to see  $\alpha_{\min}$  is minimal among penalties, meaning  $\alpha(\nu) \geq \alpha_{\min}(\nu)$  for all penalty functions  $\alpha$  representing  $\rho$ . This inequality follows because, for all  $X \in \mathcal{A}$  and all  $\nu \in L$ , we have

$$0 \geq \rho(X) \geq \nu(X) - \alpha(\nu).$$

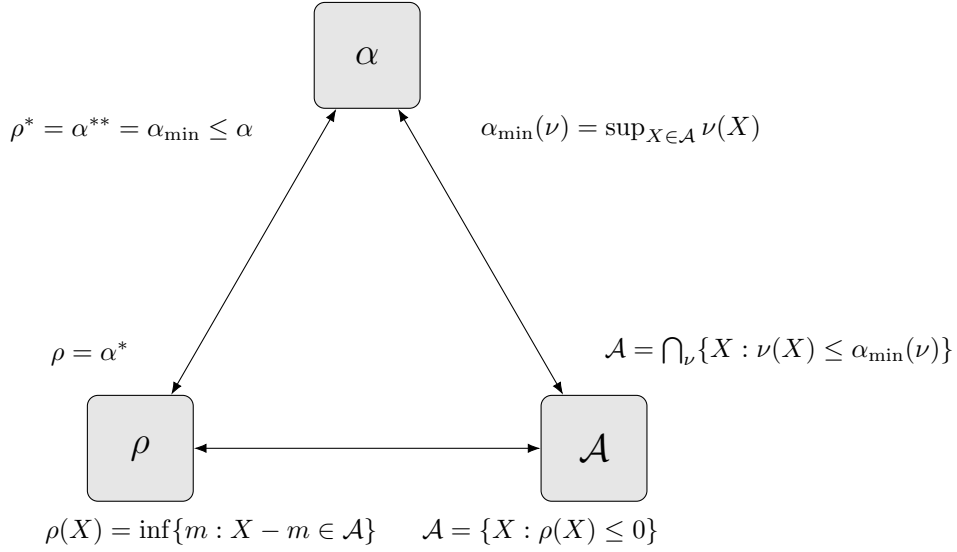


Figure 8: Starting from an arbitrary  $\alpha$ .

6. We can recover  $\rho$  from  $\mathcal{A}$  via

$$\begin{aligned} \inf \{ m \in \mathbb{R} : X - m \in \mathcal{A} \} &= \inf \{ m \in \mathbb{R} : \rho(X - m) \leq 0 \} \\ &= \inf \{ m \in \mathbb{R} : \rho(X) \leq m \} \\ &= \rho(X). \end{aligned}$$

The amount  $m$  is received and hence it is subtracted. If  $m > 0$ ,  $-m$  is the premium received.

7. To close the logical circle, it remains to show that  $\alpha_{\min}$  is a penalty, which is not entirely obvious. By cash invariance  $X - \rho(X) \in \mathcal{A}$ . Then  $\alpha_{\min}(\nu) \geq \nu(X - \rho(X))$  and so  $\rho(X) \geq \nu(X) - \alpha_{\min}(\nu)$  (assuming  $\nu(\Omega) = 1$ , which is cash invariance on the dual side) and therefore  $\rho(X) \geq \sup_{\nu} \nu(X) - \alpha_{\min}(\nu)$ . For the opposite inequality we need do real work. If  $m < \rho(X)$ , then  $X - m \notin \mathcal{A}_{\rho}$ . Assuming  $\mathcal{A}_{\rho}$  is convex and closed in the relevant topology, the separating hyperplane theorem yields a continuous linear functional  $\nu$  such that

$$\nu(X - m) > \sup_{Y \in \mathcal{A}_{\rho}} \nu(Y) = \alpha_{\min}(\nu).$$

After normalizing by cash invariance so that  $\nu(1) = 1$ , this becomes

$$m < \nu(X) - \alpha_{\min}(\nu).$$

Since the inequality holds for every  $m < \rho(X)$ , it follows that

$$\rho(X) \leq \sup_{\nu} \nu(X) - \alpha_{\min}(\nu),$$

showing  $\alpha_{\min}$  is a penalty function.

The case starting with the acceptance set is left as an exercise, but is largely covered in Section 6.2.

Lastly, Figure 8 shows the case where we start from an arbitrary  $\alpha$  that is not necessarily convex and lsc. We can still follow the definitions round and create  $\alpha_{\min}$ , which is the dual of  $\rho$ .

### 9.3 Primal Axioms: Dual Restrictions

Figure 9 shows the connections between axiomatic properties of the primal function and the restrictions each places on the duals.

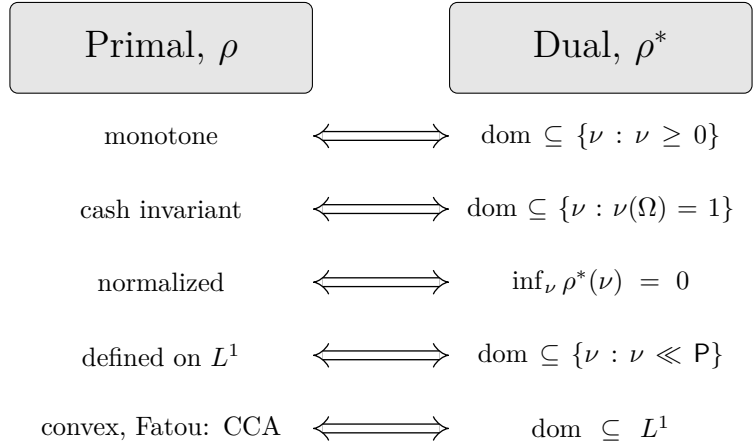


Figure 9: Primal axioms and corresponding restrictions on the dual function.

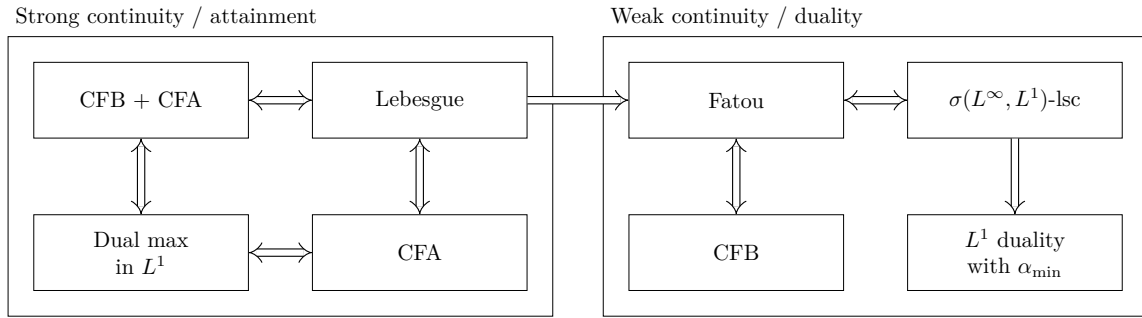


Figure 10: Relationships between continuity concepts for a convex monetary risk measure with a dual representation.

## 9.4 Continuity Hierarchy

Figure 10 shows the relationships between the different continuity concepts we have introduced. These results are contained in Föllmer and Schied (2016). The right-hand box is Theorem 4.22 and 4.33 and the left-hand in Corollary 4.35 and 4.35. Above and below swap because of the different sign conventions.

## References

- Acciaio, Beatrice, 2007, [Optimal risk sharing with non-monotone monetary functionals](#), *Finance and Stochastics* 11, 267–289.
- Aliprantis, Charalambos D., and Kim C. Border, 2006, *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Third. (Springer Verlag).
- Arai, Takuji, 2009, [Convex risk measures on Orlicz spaces: Inf-convolution and shortfall](#), *Mathematics and Financial Economics* 3, 73–88.
- Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David C. Heath, 1999, [Coherent measures of risk](#), *Mathematical finance* 9, 203–228.
- Ben-Tal, Aharon, and Marc Teboulle, 2007, [An old-new concept of convex risk measures: The optimized certainty equivalent](#), *Mathematical Finance* 17, 449–476.
- Biagini, Sara, and Marco Frittelli, 2007, [On the extension of the Namioka-Klee theorem and on the Fatou property for Risk Measures](#), *Optimality and risk-modern trends in mathematical ...*, 1–30.
- Billingsley, Patrick, 2012, *Probability and Measure*. Anniversary. (John Wiley & Sons).
- Borch, Karl, 1962, Equilibrium in a Reinsurance Market, *Econometrica* 30, 424–444.
- Castagnoli, Erio, Fabio Maccheroni, and Massimo Marinacci, 2004, [Choquet insurance pricing: A caveat](#), *Mathematical Finance* 14, 481–485.
- Cheridito, Patrick, Freddy Delbaen, and Michael Kupper, 2004, [Coherent and convex monetary risk measures for bounded càdlàg processes](#), *Stochastic Processes and their Applications* 112, 1–22.
- Cheridito, Patrick, and Tianhui Li, 2008, [Dual characterization of properties of risk measures on Orlicz hearts](#), *Mathematics and Financial Economics* 2, 29–55.
- Cheridito, Patrick, and Tianhui Li, 2009, [Risk Measures on Orlicz Hearts](#), *Mathematical Finance* 19, 189–214.
- Cherny, Alexander S., and Pavel G. Grigoriev, 2007, [Dilatation monotone risk measures are law invariant](#), *Finance and Stochastics* 11, 291–298.
- Cherny, Alexander S., and Dmitri Orlov, 2011, [On two approaches to coherent risk contribution](#), *Mathematical Finance* 21, 557–571.
- Delbaen, Freddy, 2000, [Coherent risk measures \(Pisa Notes\)](#), *Pisa Notes* 24, 733–739.
- Delbaen, Freddy, 2002, [Coherent risk measures on general probability spaces](#), *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, 1–37.
- Delbaen, Freddy, 2010, [Convex functions on dual Orlicz spaces](#), 222–223.
- Denneberg, Dieter, 1994, *Non-additive measure and integral* (Kluwer Academic, Dordrecht).
- Detlefsen, Kai, and Giacomo Scandolo, 2005, [Conditional and dynamic convex risk measures](#), *Finance and Stochastics* 9, 539–561.
- Drapeau, Samuel, and Michael Kupper, 2013, [Risk Preferences and Their Robust Representation](#), *Mathematics of Operations Research* 38, 28–62.
- El Karoui, Nicole, and Claudia Ravanelli, 2009, [Cash subadditive risk measures and interest rate ambi-](#)

guity, *Mathematical Finance* 19, 561–590.

Feller, William, 1971, *An Introduction to Probability Theory and its Applications, Volume 2*. Second. (J. Wiley; Sons).

Föllmer, Hans, and Irina Penner, 2006, [Convex risk measures and the dynamics of their penalty functions](#), *Statistics & Risk Modeling* 24, 61–96.

Föllmer, Hans, and Alexander Schied, 2002, [Robust Preferences and Convex Measures of Risk](#), *Advances in Finance and Stochastics*.

Föllmer, Hans, and Alexander Schied, 2002, [Convex measures of risk and trading constraints](#), *Finance and stochastics*.

Föllmer, Hans, and Alexander Schied, 2016, *Stochastic Finance: An Introduction in Discrete Time*. Fourth. (Walter de Gruyter, Berlin, Boston).

Föllmer, Hans, and Stefan Weber, 2015, [The Axiomatic Approach to Risk Measures for Capital Determination](#), *Annual Review of Financial Economics* 7, 301–337.

Frittelli, Marco, and Emanuela Rosazza Gianin, 2005, [Law invariant convex risk measures](#), *Advances in Mathematical Economics* 7, 33–46.

Frittelli, Marco, and Emanuela Rosazza Gianin, 2002, [Putting order in risk measures](#), *Journal of Banking and Finance* 26, 1473–1486.

Frittelli, M., and E. Rosazza Gianin, 2004, [Dynamic convex risk measures](#), *Lectures* 8, 227–248.

Gao, Niushan, Denny H. Leung, Cosimo Munari, and Foivos Xanthos, 2018, [Fatou property, representations, and extensions of law-invariant risk measures on general Orlicz spaces](#), *Finance and Stochastics* 22, 395–415.

Gao, Niushan, Denny H. Leung, and Foivos Xanthos, 2019, [Closedness of convex sets in Orlicz spaces with applications to dual representation of risk measures](#), *Studia Mathematica* 249, 329–347.

Gilboa, Itzhak, and David Schmeidler, 1993, [Updating Ambiguous Beliefs](#), *Journal of Economic Theory* 59, 33–49.

Goldblatt, Robert, 1998, *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis*. 188 (Springer New York).

Goovaerts, Marc J., F. E. De Vylder, and J. Haezendonck, 1984, *Insurance premiums* (North-Holland).

Jaschke, Stefan, and Uwe Küchler, 2001, [Coherent risk measures and good-deal bounds](#), *Finance and Stochastics* 5, 181–200.

Jouini, Elyès, Walter Schachermayer, and Nizar Touzi, 2006, [Law invariant risk measures have the Fatou property](#), in Shigeo Kusuoka and in Akira Yamazaki ed.: *Advances in Mathematical Economics* (Springer Japan).

Jouini, E., W. Schachermayer, and N. Touzi, 2008, [Optimal risk sharing for law invariant monetary utility functions](#), *Mathematical Finance* 18, 269–292.

Krasnosel'skii, M. A., and Yz B. Rutickii, 1961, [Convex functions and Orlicz spaces](#), *Noordhoff, Gröningen*.

Kusuoka, Shigeo, 2001, [On law invariant coherent risk measures](#), *Advances in Mathematical Economics* 3, 83–95.

- Maccheroni, Fabio, Massimo Marinacci, Aldo Rustichini, and Marco Taboga, 2009, Portfolio Selection with Monotone Mean-Variance Preferences, *Mathematical Finance* 19, 487–521.
- Mildenhall, Stephen J., and John A. Major, 2022, *Pricing Insurance Risk: Theory and Practice* (John Wiley & Sons, Inc.).
- Pflug, Georg Ch, and Werner Römisch, 2007, *Modeling, Measuring and Managing Risk* (World Scientific).
- Rao, M. M., and Z. D. Ren, 1991, *Theory of Orlicz Spaces*. 146 (Marcel Dekker).
- Ryff, John V., 1967, On Muirhead’s Theorem, *Pacific Journal of Mathematics* 21.
- Ryff, John V., 1967, [Extreme Points of Some Convex Subsets of  \$L^1\(0, 1\)\$](#) , *Proc American Mathematical Society* 18, 1026—1034.
- Schmeidler, David, 1986, [Integral representation without additivity](#), *Proceedings of the American Mathematical Society* 97, 255–255.
- Schmeidler, David, 1989, Subjective Probability and Expected Utility without Additivity, *Econometrica* 57, 571–587.
- Shapiro, Alexander, 2012, [On Kusuoka Representation of Law Invariant Risk Measures](#), *Mathematics of Operations Research* 38, 142–152.
- Stoyanov, Jordan M., 2013, *Counterexamples in Probability*. Third. (Dover Publications).
- Strassen, Volker, 1965, The existence of probability measures with given marginals, *The Annals of Mathematical Statistics* 36, 423–439.
- Svindland, Gregor, 2010, [Continuity properties of law-invariant \(quasi-\)convex risk functions on  \$L^\infty\$](#) , *Mathematics and Financial Economics* 3, 39–43.
- Svindland, Gregor, 2014, [Dilatation monotonicity and convex order](#), *Mathematics and Financial Economics* 8, 241–247.