Monetary Risk Measures

Dr. Robert Philipowski

Recommended book:


Contents

1 Monetary risk measures and their acceptance sets 2
2 Simple examples 5
3 Coherent and convex risk measures 6
4 Robust representation of convex risk measures 10
  4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
  4.2 Existence of a robust representation . . . . . . . . . . . . . . . . . . . . . 12
5 Convex risk measures on $L^\infty(\Omega, \mathcal{F})$ 14
6 Convex risk measures on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ 17
7 Average value at risk 21
8 Law-invariant risk measures and Kusuoka’s representation theorem 25

*Notes of the second part of a course on mathematical finance given at the University of Bonn during the winter semester 2015/2016*
1 Monetary risk measures and their acceptance sets

Given a risky financial position we want to quantify its risk in the sense of determining the minimal amount of money that has to be added in order to make the position “acceptable”.

**Definition 1.1.** Let $\Omega$ be a non-empty set (interpreted as the set of possible scenarios). A financial position is a function $X : \Omega \to \mathbb{R}$. Here $X(\omega)$ is interpreted as the (discounted) net worth of the position in scenario $\omega$ at the end of the period under consideration.

In the sequel we fix a set $G$ of “admissible” financial positions. (From the economic point of view $G$ should be as large as possible, ideally the space of all functions $X : \Omega \to \mathbb{R}$, but for mathematical reasons it can be convenient to impose some restrictions, e.g. measurability if $\Omega$ is equipped with a $\sigma$-field, or boundedness.)

**Assumption 1.2.** $G$ is a linear space containing the constants.

Given a subset $A$ of $G$ (interpreted as the set of acceptable positions or acceptance set) we will define the risk of a position $X \in G$ by

$$\rho_A(X) := \inf \{ m \in \mathbb{R} : X + m \in A \}$$

(with the convention $\inf \emptyset = \infty$). Conversely, given a “risk measure”, i.e. a function $\rho : G \to \mathbb{R} := \mathbb{R} \cup \{\pm \infty\}$, we will define the acceptance set of $\rho$ by

$$A_\rho := \{ X \in G : \rho(X) \leq 0 \}.$$

Of course not all subsets $A$ of $G$ are reasonable acceptance sets. We therefore define:

**Definition 1.3.** A subset $A$ of $G$ is called an acceptance set if:

1. $A \cap \{\text{constant functions}\} \neq \emptyset$. That is, there exists an amount of money $m$ such that having $m$ for sure is acceptable.

2. For all $X \in G$ there exists $m \in \mathbb{R}$ such that $X + m \notin A$. That is, no financial position is “infinitely good”. (This assumption is not evident; if $\mathbb{E}(X) = \infty$, risk-neutral people would pay arbitrary amounts of money to get the position $X$.)

3. $A$ is monotone in the following sense: For all $X \in A$ and all $Y \in G$ satisfying $Y \geq X$ (i.e. $Y(\omega) \geq X(\omega)$ for all $\omega \in \Omega$) we have $Y \in A$. That is, if some position is acceptable then all better positions are acceptable as well.

**Definition 1.4.** An acceptance set $A$ is normalized if $\inf \{ m \in \mathbb{R} \mid m \in A \} = 0$.

Hence if $A$ is normalized positive amounts of money are accepted, and negative amounts are not.

**Remark 1.5.** Suppose that $A$ is not normalized, and let $x_0 := \inf \{ m \in \mathbb{R} \mid m \in A \}$ and $A' := A - x_0$. Then $A'$ is normalized, and $X \in A$ if and only if $X - x_0 \in A'$. Hence one can restrict one’s attention to normalized acceptance sets without any loss of generality. Nevertheless it is sometimes convenient not to insist on normalization.

**Proposition 1.6.** Let $A \subset G$ be an acceptance set and define the function $\rho_A : G \to \mathbb{R}$ by

$$\rho_A(X) := \inf \{ m \in \mathbb{R} \mid X + m \in A \}$$

(that is, $\rho_A(X)$ is the minimal amount of money which, when added to $X$, makes $X$ acceptable). Then $\rho_A$ has the following properties:
1. \( \rho_A \) does not take the value \(-\infty\).

2. For each constant \( m \in \mathbb{R} \), \( \rho_A(m) \) is finite.

3. \( \rho_A \) is monotone, i.e.
   \[ Y \geq X \Rightarrow \rho_A(Y) \leq \rho_A(X) \]

4. \( \rho_A \) is cash-invariant, i.e. for all positions \( X \) and all constants \( m \) we have
   \[ \rho_A(X + m) = \rho_A(X) - m. \]

5. \( A \) is normalized if and only if \( \rho_A(0) = 0 \).

Proof. These properties are immediate consequences of the definition of an acceptance set.

Definition 1.7. A function \( \rho : \mathcal{G} \to \overline{\mathbb{R}} \) with the first four of the above properties is called a monetary risk measure. It is called normalized if \( \rho(0) = 0 \).

Remark 1.8. Again we can always pass to a normalized monetary risk measure by replacing \( \rho \) with \( \rho - \rho(0) \).

Lemma 1.9. Let \( \rho : \mathcal{G} \to \overline{\mathbb{R}} \) be a monetary risk measure. Then
   \[ \rho(X) \leq \rho(Y) + \|X - Y\|_{\infty} \]
for all \( X, Y \in \mathcal{G} \).

Proof. We have
   \[ Y \leq X + \|X - Y\|_{\infty} \]
so that, using monotonicity and cash invariance,
   \[ \rho(Y) \geq \rho(X + \|X - Y\|_{\infty}) = \rho(X) - \|X - Y\|_{\infty}. \]

Corollary 1.10. Let \( \rho : \mathcal{G} \to \overline{\mathbb{R}} \) be a monetary risk measure.

1. \( \rho \) is continuous with respect to the supremum norm on \( \mathcal{G} \).

2. It is even 1-Lipschitz in the following sense: Whenever \( X, Y \in \mathcal{G} \) are such that \( \rho(X) \) or \( \rho(Y) \) is finite, we have
   \[ |\rho(X) - \rho(Y)| \leq \|X - Y\|_{\infty}. \]

Proof. To prove the first statement let \( (X_n)_{n \in \mathbb{N}} \) be a sequence in \( \mathcal{G} \) such that \( \|X_n - X\|_{\infty} \to 0 \) for some \( X \in \mathcal{G} \). Then on the one hand
   \[ \rho(X_n) \leq \rho(X) + \|X_n - X\|_{\infty}, \]
so that
   \[ \limsup_{n \to \infty} \rho(X_n) \leq \rho(X), \]
and on the other hand
   \[ \rho(X) \leq \rho(X_n) + \|X_n - X\|_{\infty}, \]
so that
   \[ \liminf_{n \to \infty} \rho(X_n) \geq \rho(X), \]
Hence
   \[ \lim_{n \to \infty} \rho(X_n) = \rho(X). \]
The second statement follows immediately from the lemma.
Proposition 1.11. Let $\rho : \mathcal{G} \to \bar{\mathbb{R}}$ be a monetary risk measure. Then the set

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

is an acceptance set (called the acceptance set of $\rho$). Moreover, $A_\rho$ is closed with respect to the supremum norm on $\mathcal{G}$.

Proof. The first statement follows immediately from the definition of a monetary risk measure. The second statement follows from the fact that $A_\rho$ is the pre-image of the closed set $\mathbb{R}_-$ under the continuous map $\rho$. $\square$

Proposition 1.12.

1. For each monetary risk measure $\rho : \mathcal{G} \to \bar{\mathbb{R}}$ we have

$$\rho_{A_\rho} = \rho.$$

2. For each acceptance set $A \subset \mathcal{G}$ we have

$$A_{\rho_A} \supseteq A,$$

with equality if and only if $A$ is closed with respect to $\| \cdot \|_\infty$.

Proof.

1. We have

$$\rho_{A_\rho}(X) = \inf \{ m \in \mathbb{R} \mid X + m \in A_\rho \} = \inf \{ m \in \mathbb{R} \mid \rho(X + m) \leq 0 \} = \inf \{ m \in \mathbb{R} \mid \rho(X) \leq m \} = \rho(X).$$

2. a) To prove that $A_{\rho_A} \supseteq A$ let $X \in \mathcal{G} \setminus A_{\rho_A}$. Then $\rho_A(X) > 0$, i.e.

$$\inf \{ m \in \mathbb{R} \mid X + m \in A \} > 0.$$

Consequently $X \notin A$.

b) If $A$ is not closed with respect to $\| \cdot \|_\infty$, equality cannot hold because $A_{\rho_A}$ is closed.

c) Suppose that $A$ is closed, and let $X \in A_{\rho_A}$. Then $\rho_A(X) \leq 0$, i.e.

$$\inf \{ m \in \mathbb{R} \mid X + m \in A \} \leq 0.$$

Consequently $X + m \in A$ for each $m > 0$. Since $A$ is closed, it follows that $X \in A$. $\square$

Conclusion. There is a one-to-one correspondence between monetary risk measures and closed acceptance sets:

$$\{ \text{Monetary risk measures} \} \leftrightarrow \{ \text{Closed acceptance sets} \}$$

$$\rho \mapsto A_\rho = \{ X \in \mathcal{G} : \rho(X) \leq 0 \}$$

$$\rho_A(X) = \inf \{ m \in \mathbb{R} \mid X + m \in A \} \mapsto A$$
2 Simple examples

Which monetary risk measure (or, equivalently, which closed acceptance set) should one use in practice?

Example 2.1 (Worst case risk measure). Extremely pessimistic people would choose the worst case risk measure defined by

$$\rho(X) := - \inf X.$$ 

This is indeed a normalized monetary risk measure. Its acceptance set is

$$\mathcal{A}_\rho = \{ X \in \mathcal{G} \mid X(\omega) \geq 0 \text{ for all } \omega \in \Omega \}.$$ 

For practical purposes, however, the worst case risk measure is too pessimistic.

Example 2.2 (Value at risk). Idea: Very unlikely events should be neglected. To make this idea precise we assume that $\Omega$ is equipped with a $\sigma$-field $\mathcal{F}$ and a probability measure $\mathbb{P}$ and that all elements of $\mathcal{G}$ are measurable with respect to $\mathcal{F}$. Given a number $\alpha \in [0, 1)$ (e.g. $\alpha = 0.01$) we regard a financial position as acceptable if the probability to obtain a negative value is $\leq \alpha$. That is, we choose

$$\mathcal{A} := \{ X \in \mathcal{G} \mid \mathbb{P}[X < 0] \leq \alpha \}.$$ 

$\mathcal{A}$ is indeed a normalized acceptance set. (Exercise: Is $\mathcal{A}$ closed?) The corresponding monetary risk measure is called value at risk at level $\alpha$ and denoted $\text{VaR}_\alpha$. We have

$$\text{VaR}_\alpha(X) = \inf\{ m \in \mathbb{R} : X + m \in \mathcal{A} \} = \inf\{ m \in \mathbb{R} : \mathbb{P}[X < -m] \leq \alpha \}.$$ 

Question. Is value at risk a good risk measure?

- It is easy to understand.
- It is easy to calculate as soon as $\mathbb{P}$ is specified.

For these reasons value at risk has been the standard so far. However, it can penalize diversification: Consider two risky financial positions $X_1$ and $X_2$. Diversification means taking a convex combination of them, i.e.

$$X = \lambda X_1 + (1 - \lambda)X_2$$

for some $\lambda \in [0, 1]$. Diversification should not increase the risk, i.e. the risk of $X$ should be at most as high as the maximum of the risks of $X_1$ and $X_2$. Hence a reasonable monetary risk measure $\rho$ should satisfy

$$\rho(\lambda X_1 + (1 - \lambda)X_2) \leq \max(\rho(X_1), \rho(X_2))$$

for all $X_1, X_2 \in \mathcal{G}$ and all $\lambda \in [0, 1]$. In terms of acceptability, if two positions $X_1$ and $X_2$ are acceptable, then each convex combination of them should be acceptable as well. In other words, the acceptance set should be convex.

Does value at risk have this property?
Example 2.3. Let $\alpha = 0.01$, and let $X_1$ and $X_2$ be two independent random variables with
\[ P[X_1 = 1,000] = 0.992, \quad P[X_1 = -10,000] = 0.008. \]
(Think of $X_i$ as investing 10,000 € in a bond with interest rate 10% and default probability 0.8%.) We have $E(X_i) = 992 - 80 = 912$, so that the expected return of each position equals 9.12%. Moreover, the probability of a loss is less than $\alpha$ for each position, so that both of them are “acceptable”. (We obtain $\text{Var}_{0.01}(X_i) = -1,000$ for both positions.)

Instead of choosing $X_1$ or $X_2$ it should be even less risky to diversify, i.e. to choose $X := \frac{1}{2}X_1 + \frac{1}{2}X_2$. However, the probability that at least one of the two variables is negative equals $2 \cdot 0.008 - 0.008^2 \approx 0.016 > \alpha$. Hence $X$ is “not acceptable”. (We obtain $\text{Var}_{0.01}(X) = 4,500$.)

Conclusion. Value at risk can penalize diversification.

What should we do?

a) Try to modify value at risk in such a way that this problem does not occur any more.

b) Systematic approach initiated by Artzner, Delbaen, Eber and Heath (“Coherent measures of risk”, 1999): Formulate a set of axioms that a good risk measure should satisfy, and then investigate the structure of these risk measures.

3 Coherent and convex risk measures

Following Artzner et al. (1999) we define:

Definition 3.1 (Coherent risk measure). A monetary risk measure $\rho : \mathcal{G} \to \bar{\mathbb{R}}$ is coherent if it is

1. positively homogeneous, i.e.
\[ \rho(\lambda X) = \lambda \rho(X) \]
for all $X \in \mathcal{G}$ and all $\lambda \geq 0$ (with the convention $0 \cdot \infty := 0$),

2. subadditive, i.e.
\[ \rho(X + Y) \leq \rho(X) + \rho(Y) \]
for all $X, Y \in \mathcal{G}$.

Remark 3.2.

- Positive homogeneity seems very reasonable: if we double a financial position, then the risk should also double. Note that value at risk is positively homogeneous and that each positively homogeneous risk measure is normalized.

- Subadditivity allows to decentralize the task of managing the risk arising from a collection of different positions: if separate risk limits are given to different “desks”, then the risk of the aggregate position is bounded by the sum of the individual risk limits. Note that positive homogeneity and subadditivity together imply that
\[ \rho(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda \rho(X_1) + (1 - \lambda)\rho(X_2) \leq \max(\rho(X_1), \rho(X_2)), \]
so that a coherent risk measure does not penalize diversification. In particular, value at risk is not coherent (unless $\alpha = 0$).
Remark 3.3. A monetary risk measure $\rho$ is positively homogeneous if and only if its acceptance set $A_\rho$ is a cone, i.e. if and only if for all $X \in A_\rho$ and all $\lambda \geq 0$ we have $\lambda X \in A_\rho$. Consequently, if $X$ is acceptable, positive homogeneity implies that for each $\lambda \geq 0$ the position $\lambda X$ is acceptable as well. Föllmer and Schied ("Convex measures of risk and trading constraints", 2002), however, argued that this property is questionable and that consequently the axioms of positive homogeneity and subadditivity are too strong. Instead, one should only make sure that diversification is not penalized. This property can be formulated in several equivalent ways:

Lemma 3.4. For a monetary risk measure $\rho : G \to \bar{\mathbb{R}}$ the following properties are equivalent:

1. $\rho$ is quasi-convex, i.e.
   \[ \rho(\lambda X_1 + (1 - \lambda)X_2) \leq \max(\rho(X_1), \rho(X_2)) \]
   for all $X_1, X_2 \in G$ and all $\lambda \in [0, 1]$.

2. The acceptance set of $\rho$ is convex.

3. $\rho$ is convex, i.e.
   \[ \rho(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda \rho(X_1) + (1 - \lambda) \rho(X_2) \]
   for all $X_1, X_2 \in G$ and all $\lambda \in [0, 1]$.

Proof. 1. $\Rightarrow$ 2. trivial.
3. $\Rightarrow$ 1. trivial.
2. $\Rightarrow$ 3. Suppose that $A_\rho$ is convex and that $\rho(X_1)$ and $\rho(X_2)$ are both finite. Then by cash invariance
   \[
   \rho\left(\lambda X_1 + (1 - \lambda)X_2\right) - \lambda \rho(X_1) - (1 - \lambda) \rho(X_2) \\
   = \rho\left(\lambda X_1 + (1 - \lambda)X_2 + \lambda \rho(X_1) + (1 - \lambda) \rho(X_2)\right) \\
   = \rho\left(\left(\lambda X_1 + \rho(X_1)\right) + (1 - \lambda) \left(X_2 + \rho(X_2)\right)\right) \\
   \in_{\mathcal{A}_\rho} \in_{\mathcal{A}_\rho} \text{ because } \mathcal{A}_\rho \text{ is convex} \\
   \leq 0. \quad \square
   
Definition 3.5. A monetary risk measure with these properties is called convex.

Remark 3.6.

1. Every coherent risk measure is convex.

2. A convex risk measure is coherent if and only if it is positively homogeneous.

3. Let $\rho$ be a normalized convex risk measure. Then for each $X \in G$ and each $\lambda \geq 0$,
   \[
   \rho(\lambda X) \begin{cases} 
   \leq \lambda \rho(X) & \text{if } \lambda \in [0, 1], \\
   \geq \lambda \rho(X) & \text{if } \lambda \geq 1.
   \end{cases}
   
7
Proof. The first two statements are trivial. If \( \lambda \in [0, 1] \) we obtain

\[
\rho(\lambda X) = \rho(\lambda X + (1 - \lambda) \cdot 0) \\
\leq \lambda \rho(X) + (1 - \lambda) \rho(0) \\
= \lambda \rho(X).
\]

If \( \lambda \geq 1 \), then \( 1/\lambda \leq 1 \), so that

\[
\rho(X) = \rho \left( \frac{1}{\lambda} \lambda X \right) \leq \frac{1}{\lambda} \rho(\lambda X). \tag*{\square}
\]

Example 3.7 (Acceptance in terms of expected utility). Let \( \Omega \) be equipped with a \( \sigma \)-field \( \mathcal{F} \) and a probability measure \( P \), and assume that all elements of \( \mathcal{G} \) are measurable with respect to \( \mathcal{F} \) and bounded. Let moreover \( u : \mathbb{R} \to \mathbb{R} \) be strictly increasing and concave (“the utility function of a risk-averse or risk-neutral decision maker”). Fix \( x_0 \in \mathbb{R} \), and let

\[
\mathcal{A} := \{ X \in \mathcal{G} \mid \mathbb{E}[u(X)] \geq u(x_0) \}.
\]

That is, \( X \) is accepted if and only if the certainty equivalent of \( X \) – i.e. \( u^{-1}(\mathbb{E}[u(X)]) \) – is greater than or equal to \( x_0 \). \( \mathcal{A} \) is clearly an acceptance set, which is normalized if and only if \( x_0 = 0 \). Moreover, concavity of \( u \) implies that for \( X, Y \in \mathcal{A} \) and \( \lambda \in [0, 1] \),

\[
\mathbb{E}[u(\lambda X + (1 - \lambda)Y)] \geq \mathbb{E}[\lambda u(X) + (1 - \lambda)u(Y)] \\
= \lambda \mathbb{E}[u(X)] + (1 - \lambda)\mathbb{E}[u(Y)] \\
\geq u(x_0),
\]

so that \( \mathcal{A} \) is convex. The corresponding convex risk measure is given by

\[
\rho_{\mathcal{A}}(X) = \inf\{ m \in \mathbb{R} \mid \mathbb{E}[u(X + m)] \geq u(x_0) \}.
\]

Let us consider two important special cases:

a) Let

\[
u(x) = x.
\]

Then

\[
\rho_{\mathcal{A}}(x) = \inf\{ m \in \mathbb{R} \mid \mathbb{E}(X) + m \geq x_0 \} \\
= x_0 - \mathbb{E}(X) \\
= \mathbb{E}(-X) + x_0.
\]

b) Let

\[
u(x) = 1 - e^{-\beta x} \quad (\beta > 0)
\]

and \( x_0 = 0 \). We have

\[
u'(x) = \beta e^{-\beta x} > 0, \\
u''(x) = -\beta^2 e^{-\beta x} < 0,
\]

so that \( u \) is indeed strictly increasing and concave. The condition \( \mathbb{E}[u(X + m)] \geq u(x_0) \) now reads

\[
1 - e^{-\beta m} \mathbb{E} \left[ e^{-\beta X} \right] \geq 0
\]
or, equivalently,
\[ m \geq \frac{1}{\beta} \log \mathbb{E} \left[ e^{-\beta X} \right]. \]

Hence
\[ \rho_A(X) = \frac{1}{\beta} \log \mathbb{E} \left[ e^{-\beta X} \right]. \]

(“entropic risk measure”).

**Proposition 3.8.** The entropic risk measure can be represented as follows:
\[ \frac{1}{\beta} \log \mathbb{E} \left[ e^{-\beta X} \right] = \sup_{Q \in \mathcal{M}_1(\Omega, \mathcal{F})} \left( \mathbb{E}_Q(-X) - \frac{1}{\beta} H(Q \| P) \right), \]

Here \( H(Q \| P) \) is the relative entropy of \( Q \) with respect to \( P \), defined by
\[ H(Q \| P) := \begin{cases} +\infty & \text{if } Q \text{ is not absolutely continuous with respect to } P \\ \mathbb{E} \left( \frac{dQ}{dP} \log \frac{dQ}{dP} \right) = \mathbb{E}_Q \left( \log \frac{dQ}{dP} \right) & \text{if } Q \text{ is absolutely continuous with respect to } P. \end{cases} \]

**Lemma 3.9.** For all probability measures \( P \) and \( Q \),
\[ H(Q \| P) \geq 0, \]
with equality if and only if \( Q = P \).

**Proof.** We may assume that \( Q \) is absolutely continuous with respect to \( P \); we denote the density by \( \rho \). Since the function \( f(x) := x \log x \) is strictly convex, Jensen’s inequality implies that
\[ H(Q \| P) = \int_{\Omega} f(\rho) \, dP \geq f \left( \int_{\Omega} \rho \, dP \right) = f(1) = 0, \]
with equality if and only if \( \rho \) is constant \( P \)-a.s, i.e. if and only if \( Q = P \).

Proposition 3.8 follows immediately from the following lemma:

**Lemma 3.10.** For any probability measure \( Q \) on \( (\Omega, \mathcal{F}) \),
\[ H(Q \| P) \geq \beta \mathbb{E}_Q(-X) - \log \mathbb{E} \left[ e^{-\beta X} \right], \]
with equality if and only if
\[ Q = P^X := \frac{e^{-\beta X}}{\mathbb{E}[e^{-\beta X}]} P. \]

**Proof.** We have
\[ \frac{dQ}{dP} = \frac{dQ}{dP^X} \frac{dP^X}{dP} = \frac{dQ}{dP^X} \frac{e^{-\beta X}}{\mathbb{E}[e^{-\beta X}]} \]
and consequently
\[ \log \frac{dQ}{dP} = \log \frac{dQ}{dP^X} - \beta X - \log \mathbb{E}[e^{-\beta X}]. \]
Integration with respect to $Q$ then yields

$$H(Q \mid \mathbb{P}) = E_Q \left( \log \frac{dQ}{d\mathbb{P}} \right)$$

$$= E_Q \left( \log \frac{dQ}{d\mathbb{P}X} \right) - \beta E_Q(X) - \log \mathbb{E}[e^{-\beta X}]$$

$$\geq \beta E_Q(-X) - \log \mathbb{E}[e^{-\beta X}],$$

with equality if and only if $Q = \mathbb{P}^X$.

Example 3.11. We work in the setting of the last example, but instead of just one probability measure $\mathbb{P}$ we are given a whole non-empty set $S$ of probability measures. Moreover, for each $Q \in S$ we are given a number $x_0(Q) \in \mathbb{R}$. Now let

$$A := \bigcap_{Q \in S} \{ X \in \mathcal{G} \mid E_Q[u(X)] \geq u(x_0(Q)) \}. $$

That is, $X$ is accepted if and only if for each probability measure $Q \in S$ the certainty equivalent of $X$ with respect to $Q$ is greater than or equal to $x_0(Q)$. $A$ is not necessarily an acceptance set because it can be empty. However, if the function $Q \mapsto x_0(Q)$ is bounded, $A$ is a convex acceptance set.

In the special case $u(x) = x$ we get

$$\rho_A(X) = \inf \{ m \in \mathbb{R} \mid X + m \in A \}$$

$$= \inf \{ m \in \mathbb{R} \mid \forall Q \in S, E_Q(X + m) \geq x_0(Q) \}$$

$$= \inf \{ m \in \mathbb{R} \mid \inf_{Q \in S} \left( E_Q(X) + m - x_0(Q) \right) \geq 0 \}$$

$$= - \inf_{Q \in S} \left( E_Q(X) - x_0(Q) \right)$$

$$= \sup_{Q \in S} \left( E_Q(-X) + x_0(Q) \right).$$

Note that this expression looks quite similar to the formula that we obtained for the entropic risk measure. Remarkably, “many” convex risk measures can be represented in this way.

4 Robust representation of convex risk measures

4.1 Introduction

In this section we suppose that $\Omega$ is equipped with a $\sigma$-field $\mathcal{F}$ and that all elements of $\mathcal{G}$ are $\mathcal{F}$-measurable. Our goal is to show that if $\mathcal{G}$ is “nice”, then all “nice” convex risk measures $\rho : \mathcal{G} \to \bar{\mathbb{R}}$ can be represented in the form

$$\rho(X) = \sup_{Q \in S} \left( E_Q(-X) - \alpha(Q) \right)$$

where

10
• $S$ is a non-empty set of probability measures on $(\Omega, \mathcal{F})$ such that all elements of $\mathcal{G}$ are integrable with respect to all $Q \in S$,

• the function $\alpha : S \to \mathbb{R} \cup \{+\infty\}$ is bounded from below and not identically $+\infty$.

**Proposition 4.1.** A function $\rho : \mathcal{G} \to \bar{\mathbb{R}}$ of that form is a convex risk measure.

*Proof.* Since $\alpha$ is not identically $+\infty$, $\rho$ does not take the value $-\infty$. Since $\alpha$ is bounded from below the constants have finite risk. Monotonicity, cash-invariance and convexity are clear. \(\square\)

**Remark 4.2.** A representation of that form is called a robust representation. The idea is as follows:

1. For various probabilistic models $Q$ one computes the expected loss with respect to $Q$.
2. These models are taken more or less seriously, as specified by the penalty function $\alpha$.
3. Finally one computes $\rho(X)$ as the worst case of the penalized expected loss over all probabilistic models in the class $S$.

**Remark 4.3.** Probability measures $Q \in S$ for which $\alpha(Q) = +\infty$ do not contribute to the supremum so that the corresponding terms can be dropped. On the other hand, one can artificially add probability measures $Q$ to the supremum by choosing $\alpha(Q) := +\infty$. One may therefore assume that $S$ is the set of all probability measures on $(\Omega, \mathcal{F})$ with respect to which all elements of $\mathcal{G}$ are integrable.

**Proposition 4.4.**

1. For each $Q \in S$ we have
   \[
   \alpha(Q) \geq \alpha_{\text{min}}(Q) := \sup_{X \in \mathcal{G}} \left( E_Q(-X) - \rho(X) \right)
   = \sup_{X \in \mathcal{A}_\rho} E_Q(-X).
   \]

2. The robust representation also holds with $\alpha_{\text{min}}$ instead of $\alpha$.

*Proof.* 1. The robust representation immediately implies that for all $Q \in S$ and all $X \in \mathcal{G}$ we have
   \[
   \alpha(Q) \geq E_Q(-X) - \rho(X)
   \]
   and consequently
   \[
   \alpha(Q) \geq \alpha_{\text{min}}(Q).
   \]
   Moreover,
   \[
   \sup_{X \in \mathcal{G}} \left( E_Q(-X) - \rho(X) \right) = \sup_{X \in \mathcal{G}} E_Q(-[X + \rho(X)])
   \leq \sup_{X' \in \mathcal{A}_\rho} E_Q(-X')
   \]
   and
   \[
   \sup_{X \in \mathcal{A}_\rho} E_Q(-X) \leq \sup_{X \in \mathcal{A}_\rho} E_Q(-[X + \rho(X)])
   \leq \sup_{X \in \mathcal{G}} E_Q(-[X + \rho(X)]).\]
2. On the one hand, since \( \alpha_{\text{min}}(Q) \leq \alpha(Q) \) the robust representation immediately implies that
\[
\rho(X) \leq \sup_{Q \in \mathcal{S}} \left( \mathbb{E}_Q(-X) - \alpha_{\text{min}}(Q) \right).
\]
On the other hand, for each \( X \in \mathcal{G} \) we have \( \alpha_{\text{min}}(Q) \geq \mathbb{E}_Q(-X) - \rho(X) \) and consequently
\[
\sup_{Q \in \mathcal{S}} \left( \mathbb{E}_Q(-X) - \alpha_{\text{min}}(Q) \right) \leq \sup_{Q \in \mathcal{S}} \left( \mathbb{E}_Q(-X) - \mathbb{E}_Q(X) + \rho(X) \right) = \rho(X). \tag*{□}
\]

**Proposition 4.5.**  For a convex risk measure \( \rho \) admitting a robust representation the following statements are equivalent:

1. \( \rho \) is coherent.
2. There exists a robust representation whose penalty function \( \alpha \) only takes the values 0 and \( \infty \).
3. \( \alpha_{\text{min}} \) only takes the values 0 and \( \infty \).

**Proof.**  3. \( \Rightarrow \) 2. Trivial.
2. \( \Rightarrow \) 1. If \( \alpha \) only takes the values and \( \infty \), then \( \rho \) is clearly positively homogeneous and hence coherent.
1. \( \Rightarrow \) 3. Let \( Q \in \mathcal{S} \) and \( \lambda > 0 \). Recall (Remark 3.3) that the acceptance set of a coherent risk measure is a cone, so that \( X \in \mathcal{A}_\rho \) if and only if \( \lambda X \in \mathcal{A}_\rho \). Consequently
\[
\alpha_{\text{min}}(Q) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}_Q(-X) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}_Q(-\lambda X) = \lambda \alpha_{\text{min}}(Q). \tag*{□}
\]

### 4.2 Existence of a robust representation

We now suppose that \( \mathcal{G} \) is equipped with a norm \( \| \cdot \| \) such that \( (\mathcal{G}, \| \cdot \|) \) is a Banach space. Our main example will be \( L^\infty(\Omega, \mathcal{F}) \), the space of all bounded \( \mathcal{F} \)-measurable functions on \( \Omega \) equipped with the supremum norm.

The idea is to consider the Legendre transform of \( \rho \), defined by
\[
\rho^*(\ell) := \sup_{X \in \mathcal{G}} \left( \ell(X) - \rho(X) \right), \quad \ell \in \mathcal{G}' := \text{dual space of } \mathcal{G}.
\]

Applying the Legendre transform a second time (together with the canonical embedding of \( \mathcal{G} \) into its double dual \( \mathcal{G}'' \)) we get
\[
\rho^{**}(X) := \sup_{\ell \in \mathcal{G}'} \left( \ell(X) - \rho^*(\ell) \right), \quad X \in \mathcal{G}.
\]

If \( \rho^{**} = \rho \), we are on the way to obtain a robust representation.

**Remark 4.6.** For each \( \ell \in \mathcal{G}' \) the function \( X \mapsto \ell(X) - \rho^*(\ell) \) is affine and continuous. Consequently, \( \rho^{**} \) is convex and lower semicontinuous. Hence \( \rho^{**} = \rho \) can only hold if \( \rho \) itself is convex and lower semicontinuous.

**Theorem 4.7** (Fenchel-Moreau). A function \( \rho : \mathcal{G} \to \overline{\mathbb{R}} \) satisfies \( \rho^{**} = \rho \) if and only if \( \rho \) is convex and lower semicontinuous.

**Corollary 4.8.** Each convex and lower semicontinuous risk measure \( \rho : \mathcal{G} \to \overline{\mathbb{R}} \) has the representation
\[
\rho(X) = \sup_{\ell \in \mathcal{G}'} \left( \ell(X) - \rho^*(\ell) \right).
\]
**Example 4.9.** Every monetary risk measure is even Lipschitz continuous with respect to the supremum norm. Hence on $L^\infty(\Omega, \mathcal{F})$ every convex risk measure has such a representation.

**Proposition 4.10.** Whenever $\rho^*(\ell) < \infty$, we have

1. $\ell \leq 0$, i.e. $\ell(X) \leq 0$ for all $X \geq 0$,

2. $\ell(1) = -1$.

Hence for each $\ell \in \mathcal{G}'$ which contributes to the supremum, $-\ell$ behaves as integration with respect to a probability measure.

**Proof of Proposition 4.10.** Let $\ell \in \mathcal{G}'$ be such that $\rho^*(\ell) < \infty$.

1. Take $X \geq 0$. Then

$$
+\infty > \rho^*(\ell) = \sup_{Y \in \mathcal{G}} \left( \ell(Y) - \rho(Y) \right) 
\geq \sup_{\lambda \geq 0} \left( \ell(\lambda X) - \rho(\lambda X) \right) 
\geq \sup_{\lambda \geq 0} \lambda \ell(X) - \rho(0).
$$

This supremum can only be less than $+\infty$ if $\ell(X) \leq 0$.

2. Similarly,

$$
+\infty > \sup_{Y \in \mathcal{G}} \left( \ell(Y) - \rho(Y) \right) 
\geq \sup_{\lambda \in \mathbb{R}} \left( \ell(\lambda) - \rho(\lambda) \right) 
= \sup_{\lambda \in \mathbb{R}} \left( \lambda \ell(1) - \rho(0) + \lambda \right) 
= \sup_{\lambda \in \mathbb{R}} \lambda \left( \ell(1) + 1 \right) - \rho(0).
$$

This supremum can only be less than $+\infty$ if $\ell(1) = -1$. \qed

**Corollary 4.11.** Every convex and lower semicontinuous risk measure $\rho : \mathcal{G} \to \bar{\mathbb{R}}$ has the representation

$$
\rho(X) = \sup_{\ell \in \mathcal{G}'}, \ell \leq 0, \ell(1) = -1 \left( \ell(X) - \rho^*(\ell) \right).
$$

Under appropriate additional conditions each $\ell \in \mathcal{G}'$ with $\ell \leq 0$ and $\ell(1) = -1$ is of the form $\ell(X) = -\mathbb{E}_Q(X)$ for some probability measure $Q$ on $(\Omega, \mathcal{F})$. Defining $\alpha(\mathcal{Q}) := \rho^*(\ell)$ then gives the representation

$$
\rho(X) = \sup_Q \left( \mathbb{E}_Q(-X) - \alpha(Q) \right)
$$

where the supremum is over a certain class of probability measures.
5 Convex risk measures on $\mathcal{L}^\infty(\Omega, \mathcal{F})$

We now suppose that $\mathcal{G} = \mathcal{L}^\infty(\Omega, \mathcal{F})$. Assuming boundedness of all financial positions $X$ is not very restrictive from the economic point of view, but quite convenient because it implies that

1. $\rho(X)$ is finite for every monetary risk measure $\rho$,
2. $\mathbb{E}_Q(-X)$ is well-defined and finite for every probability measure $Q$ on $(\Omega, \mathcal{F})$.

Let $\rho$ be a convex risk measure on $\mathcal{L}^\infty(\Omega, \mathcal{F})$. Since $\rho$ is even Lipschitz continuous with respect to $\| \cdot \|$ we have

$$\rho(X) = \sup \left\{ \ell(X) - \rho^*(\ell) \mid \ell \in \mathcal{L}^\infty(\Omega, \mathcal{F})', \ell \leq 0, \ell(1) = -1 \right\}.$$  

How do elements of $\mathcal{L}^\infty(\Omega, \mathcal{F})'$ look like?

**Definition 5.1.**

1. A function $\mu : \mathcal{F} \to \mathbb{R}$ is called a **finitely additive set function** if
   
   (a) $\mu(\emptyset) = 0$,
   
   (b) $\mu(A \cup B) = \mu(A) + \mu(B)$ for all disjoint $A, B \in \mathcal{F}$.

2. The **total variation** of a finitely additive set function $\mu$ is

$$\|\mu\|_{\text{var}} := \sup \left\{ \sum_{i=1}^n |\mu(A_i)| \mid n \in \mathbb{N}, A_1, \ldots, A_n \in \mathcal{F} \text{ are pairwise disjoint} \right\}.$$  

3. We denote the set of all finitely additive set functions of finite total variation by $\mathcal{M}_f(\Omega, \mathcal{F})$.

**Definition 5.2.** The integral of a **bounded** measurable function $f : \Omega \to \mathbb{R}$ with respect to a finitely additive set function with finite total variation is defined as follows:

1. Let $\mathcal{G}_0$ be the space of **elementary functions** of the form

$$X = \sum_{i=1}^n \alpha_i 1_{A_i}$$

with $\alpha_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$. Note that this space is dense in $\mathcal{L}^\infty(\Omega, \mathcal{F})$ with respect to the supremum norm. For such a function $X \in \mathcal{G}_0$ we set

$$I_\mu(X) := \sum_{i=1}^n \alpha_i \mu(A_i).$$

One has to check that this definition is independent of the choice of the representation of $X$, but this can be done in the same way as when $\mu$ is a measure.

2. Choose a representation of $X$ with pairwise disjoint sets $A_i$. Then

$$|I_\mu(X)| \leq \sum_{i=1}^n |\alpha_i| |\mu(A_i)| \leq \|X\|_\infty \|\mu\|_{\text{var}}.$$  

Hence there is a unique extension of $I_\mu$ to a continuous linear functional on $\mathcal{L}^\infty(\Omega, \mathcal{F})$. We write

$$\int \Omega X d\mu$$

for this extension applied to a function $X \in \mathcal{L}^\infty(\Omega, \mathcal{F})$.  

14
Theorem 5.3. For each continuous linear functional $\ell$ on $G$ there exists a finitely additive set function $\mu \in \mathcal{M}_f(\Omega, \mathcal{F})$ such that

$$\ell(X) = \int_{\Omega} X d\mu$$

for each $X \in \mathcal{L}^\infty(\Omega, \mathcal{F})$.

Proof. We define

$$\mu(A) := \ell(1_A).$$

Finite additivity of $\mu$ is clear. To prove that $\mu$ is of finite variation let $A_1, \ldots, A_n$ be pairwise disjoint. Let $A_+ = \bigcup \{A_i : \mu(A_i) \geq 0\}$ and $A_- = \bigcup \{A_i : \mu(A_i) < 0\}$. Then

$$\sum_{i=1}^n |\mu(A_i)| = \mu(A_+) - \mu(A_-) = \ell(1_{A_+} - 1_{A_-}) \leq ||\ell||\|1_{A_+} - 1_{A_-}\|_\infty \leq ||\ell||.$$

Hence $||\mu||_{\text{var}} \leq ||\ell||$. \qed

Corollary 5.4. Every convex risk measure $\rho : \mathcal{L}^\infty(\Omega, \mathcal{F}) \to \mathbb{R}$ can be represented in the form

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \left( E_Q(-X) - \alpha(Q) \right),$$

where $\mathcal{M}_{1,f}$ denotes the space of all finitely additive probability measures on $(\Omega, \mathcal{F})$, i.e. the set of all non-negative finitely additive set functions $Q$ satisfying $Q(\Omega) = 1$.

Question. Under which conditions is there a representation in terms of true (i.e. $\sigma$-additive) probability measures? In other words, which convex risk measures $\rho$ can be represented in the form

$$\rho(X) = \sup_{Q \in \mathcal{M}_1} \left( E_Q(-X) - \alpha(Q) \right),$$

where $\mathcal{M}_1$ denotes the space of all true (i.e. $\sigma$-additive) probability measures on $(\Omega, \mathcal{F})$?

We start with a necessary condition:

Proposition 5.5. Let $\rho : \mathcal{L}^\infty(\Omega, \mathcal{F}) \to \mathbb{R}$ be a convex risk measure that can be represented in terms of true probability measures. Then it has the following “Fatou property”: For each bounded and pointwise convergent sequence $(X_n)_{n \in \mathbb{N}}$ we have

$$\rho(\lim_{n \to \infty} X_n) \leq \liminf_{n \to \infty} \rho(X_n).$$

Proof. Using Lebesgue’s dominated convergence theorem (which holds for true probability measures, but not for finitely additive ones) we obtain

$$\rho(\lim_{n \to \infty} X_n) = \sup_{Q \in \mathcal{M}_1} \left( E_Q(-\lim_{n \to \infty} X_n) - \alpha(Q) \right)$$

$$= \sup_{Q \in \mathcal{M}_1} \lim_{n \to \infty} \left( E_Q(-X_n) - \alpha(Q) \right)$$

$$\leq \liminf_{n \to \infty} \sup_{Q \in \mathcal{M}_1} \left( E_Q(-X_n) - \alpha(Q) \right)$$

$$= \liminf_{n \to \infty} \rho(X_n). \qed$$
Proposition 5.6. For a monetary risk measure $\rho : \mathcal{L}^\infty(\Omega, \mathcal{F}) \to \mathbb{R}$ the Fatou property is equivalent to continuity from above, i.e.

$$X_n \searrow X \implies \rho(X_n) \nearrow X.$$ 

Proof. Suppose first that $\rho$ has the Fatou property, and let $X_n \searrow X$. Note that this implies that the sequence $(X_n)_{n \in \mathbb{N}}$ is bounded (from below by $\inf X$ and from above by $\sup X_1$). Hence the Fatou property implies that

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

Moreover, monotonicity of $\rho$ implies that the sequence $(\rho(X_n))_{n \in \mathbb{N}}$ is convergent, and

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

Hence $\rho(X_n) \nearrow \rho(X)$.

Conversely, assume that $\rho$ is continuous from above, and let $(X_n)_{n \in \mathbb{N}}$ be a bounded and pointwise convergent sequence in $\mathcal{L}^\infty(\Omega, \mathcal{F})$. Let $X := \lim_{n \to \infty} X_n$ and $Y_n := \sup_{m \geq n} X_m$. Then $Y_n \searrow X$, so that by continuity from above

$$\rho(Y_n) \nearrow \rho(X).$$

Moreover, $X_n \leq Y_n$, so that

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

A sufficient condition for the existence of a robust representation in terms of true probability measures is continuity from below:

Proposition 5.7. Let $\rho$ be a convex risk measure on $\mathcal{L}^\infty(\Omega, \mathcal{F}) \to \mathbb{R}$ that is continuous from below, i.e.

$$X_n \nearrow X \implies \rho(X_n) \searrow X,$$

and has the representation

$$\rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \left( \mathbb{E}_Q(-X) - \alpha(Q) \right).$$

Then $\alpha(Q) = \infty$ for all $Q \in \mathcal{M}_{1,f}$ which are not $\sigma$-additive, and consequently

$$\rho(X) = \sup_{Q \in \mathcal{M}_1} \left( \mathbb{E}_Q(-X) - \alpha(Q) \right).$$

Proof. Let $Q \in \mathcal{M}_{1,f}$ be such that $\alpha(Q) < \infty$. In order to show that $Q$ is $\sigma$-additive we show that for every sequence $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{F}$ that increases to $\Omega$ we have $Q(A_n) \nearrow 1$. To this end let $X = \lambda 1_{A_n}$ for some $\lambda > 0$. We have

$$\rho(\lambda 1_{A_n}) \geq \mathbb{E}_Q(-\lambda 1_{A_n}) - \alpha(Q)$$

and consequently

$$\lambda Q(A_n) \geq -\rho(\lambda 1_{A_n}) - \alpha(Q).$$

As $n \to \infty$, $\lambda 1_{A_n} \nearrow \lambda$, hence continuity from below implies that

$$\rho(\lambda 1_{A_n}) \searrow \rho(\lambda) = \rho(0) - \lambda.$$
It follows that
\[ \lambda \liminf_{n \to \infty} Q(A_n) \geq \lambda - \rho(0) - \alpha(Q) \]
and consequently
\[ \liminf_{n \to \infty} Q(A_n) \geq 1 - \frac{\rho(0) + \alpha(Q)}{\lambda}. \]
Letting \( \lambda \to \infty \) gives the result. \( \Box \)

**Summary.** For a convex risk measure \( \rho \) on \( \mathcal{L}^\infty(\Omega, \mathcal{F}) \) we have the following implications:

- \( \rho \) continuous from below
  - \( \Rightarrow \rho \) has a robust representation in terms of true probability measures
  - \( \Rightarrow \rho \) continuous from above
  - \( \Leftrightarrow \rho \) has the Fatou property

### 6 Convex risk measures on \( \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \)

We now fix a probability measure \( \mathbb{P} \) on \( (\Omega, \mathcal{F}) \) and consider risk measures \( \rho \) on \( \mathcal{L}^\infty(\Omega, \mathcal{F}) \) with the property
\[ X = Y \mathbb{P}\text{-a.s.} \Rightarrow \rho(X) = \rho(Y). \]
Such risk measures can be regarded as functionals on the space \( \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) of equivalence classes (w.r.t. to \( \mathbb{P}\)-a.s. equality) of elements of \( \mathcal{L}^\infty(\Omega, \mathcal{F}) \). We therefore call them *risk measures on \( \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \)*.

**Proposition 6.1.** Let \( \rho \) be a convex risk measure on \( \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) that has the representation
\[ \rho(X) = \sup_{Q \in \mathcal{M}_{1,f}} \left( \mathbb{E}_Q(-X) - \alpha(Q) \right). \]
Then \( \alpha(Q) = \infty \) for all \( Q \in \mathcal{M}_{1,f} \) that are not absolutely continuous with respect to \( \mathbb{P} \). Consequently, \( \rho \) has the representation
\[ \rho(X) = \sup_{Q \in \mathcal{M}_{1,f}(\mathbb{P})} \left( \mathbb{E}_Q(-X) - \alpha(Q) \right), \]
where \( \mathcal{M}_{1,f}(\mathbb{P}) \) denotes the set of all finitely additive probability measures on \( (\Omega, \mathcal{F}) \) that are absolutely continuous with respect to \( \mathbb{P} \).

**Proof.** Let \( Q \in \mathcal{M}_{1,f} \) be not absolutely continuous with respect to \( \mathbb{P} \). Then there exists \( A \in \mathcal{F} \) with \( \mathbb{P}(A) = 0 \) and \( Q(A) > 0 \). Let \( X_n := -n1_A \). Then \( X_n = 0 \) \( \mathbb{P}\)-a.s, and consequently
\[ \alpha(Q) \geq \mathbb{E}_Q(n1_A) - \rho(-n1_A) = nQ(A) - \rho(0) \to \infty \quad (n \to \infty). \quad \Box \]

For convex risk measures on \( \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) existence of a representation in terms of true probability measures can be characterized in a number of ways. To formulate and prove the next result we have to recall some definitions and results from functional analysis:
1. The space $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ can be identified with the dual space of $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Namely, every continuous linear functional $\ell$ on $L^1(\Omega, \mathcal{F}, \mathbb{P})$ is of the form

$$\ell(X) = \int_\Omega XYd\mathbb{P}$$

for some $Y \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

2. Given a Banach space $G$, the \textit{weak* topology} on $G'$ is the coarsest topology for which all functionals of the form

$$\ell \mapsto \ell(X), \quad X \in G,$$

are continuous. For example, the weak* topology on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ (regarded as dual space of $L^1(\Omega, \mathcal{F}, \mathbb{P})$) is the coarsest topology for which all functionals of the form

$$Y \mapsto \int_\Omega XYd\mathbb{P}, \quad X \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

are continuous. One can show that every linear functional on $G'$ which is continuous with respect to the weak* topology is of the form

$$\ell \mapsto \ell(X)$$

for some $X \in G$.

3. \textit{Krein-Shmulyan theorem:} A convex subset $K$ of $G'$ is weak* closed if and only if for each $r > 0$ the set

$$K_r := K \cap \{x \in G' | \|x\|_{G'} \leq r\}$$

is weak* closed.

\textbf{Theorem 6.2.} For a convex risk measure $\rho$ on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ the following properties are equivalent:

1. $\rho$ can be represented in the form

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(P)} \left( E_Q(-X) - \alpha(Q) \right).$$

2. $\rho$ can be represented in the form

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(P)} \left( E_Q(-X) - \alpha(Q) \right),$$

where $\mathcal{M}_1(P)$ denotes the set of all true probability measures on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to $\mathbb{P}$.

3. $\rho$ is continuous from above, i.e.

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

4. $\rho$ has the Fatou property, i.e. for every bounded and pointwise convergent sequence $(X_n)_{n \in \mathbb{N}},$

$$\rho(\lim_{n \to \infty} X_n) \leq \liminf_{n \to \infty} \rho(X_n).$$
5. \( \rho \) is lower semicontinuous with respect to the weak* topology on \( L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) (regarded as dual pace of \( L^1(\Omega, \mathcal{F}, \mathbb{P}) \)).

6. The acceptance set \( A_\rho \) is closed in the weak* topology of \( L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \).

**Remark 6.3.**

1. In 3. and 4. one can replace pointwise convergence with \( \mathbb{P} \)-almost sure convergence.

2. Every convex risk measure on \( L^\infty(\Omega, \mathcal{F}) \) is Lipschitz continuous and consequently lower semicontinuous with respect to \( \| \cdot \|_\infty \). However, the weak* topology is coarser than the norm topology, so that lower semicontinuity with respect to the weak* topology is a stronger property than with respect to the norm topology. Note however that for a convex function on a Banach space \( G \) lower semicontinuity with respect to the weak topology is equivalent to lower semicontinuity with respect to the norm topology. (Suppose that \( f : G \to \bar{\mathbb{R}} \) is lower semicontinuous with respect to \( \| \cdot \|_\infty \). Then for each \( c \in \mathbb{R} \) the set \( \{ f \leq c \} \) is strongly closed. Since moreover it is convex it is weakly closed. The last fact (every strongly closed and convex subset \( K \) of a Banach space \( G \) is weakly closed) can be proved as follows: Let \( x \in G \setminus K \). Then by the Hahn-Banach separation theorem there exists a continuous linear functional \( \ell \) such that \( \ell|_K \leq 0 \) and \( \ell(x) > 0 \). Hence the set \( \{ \ell > 0 \} \) is a weakly open neighborhood of \( x \) contained in the complement of \( K \). Since \( x \in G \setminus K \) is arbitrary it follows that \( G \setminus K \) is weakly open.)

**Proof of Theorem 6.2.**

1. The equivalence 1. \( \iff \) 2. follows from the last proposition, and the implications 1. \( \Rightarrow \) 3. \( \iff \) 4. have already been proved (Propositions 5.5 and 5.6). Since \( A_\rho = \{ X \in L^\infty | \rho(X) \leq 0 \} \), the implication 5. \( \Rightarrow \) 6. is trivial. In the sequel we will prove 4. \( \Rightarrow \) 5. and 6. \( \Rightarrow \) 2.

2. We now prove 4. \( \Rightarrow \) 5. We have to show that for each \( c \in \mathbb{R} \) the set

\[
K := \{ X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) | \rho(X) \leq c \}
\]

is weak* closed. Convexity of \( \rho \) implies that \( K \) is convex, so that by the Krein-Shmulyan theorem it suffices to prove that for each \( r > 0 \) the set

\[
K_r = \{ X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) | \| X \|_\infty \leq r \text{ and } \rho(X) \leq c \}
\]

is weak* closed.

(a) We first show that \( K_r \) is closed with respect to the \( L^1 \)-norm (note that \( K_r \subseteq L^\infty \subseteq L^1 \)). To this end let \( (X_n)_{n \in \mathbb{N}} \) be a sequence in \( K_r \) converging in \( L^1 \) to some \( X \in L^1 \). Then there exists a subsequence \( (X_{n_k})_{k \in \mathbb{N}} \) converging \( \mathbb{P} \)-almost surely to \( X \), so that

\[
\| X \|_\infty \leq r.
\]

Moreover, by the Fatou property

\[
\rho(X) \leq \liminf_{k \to \infty} \rho(X_{n_k}) \leq c.
\]

Hence \( X \in K_r \), and \( K_r \) is closed with respect to the \( L^1 \)-norm.

(b) Since \( K_r \) is convex it follows that it is also weakly closed in \( L^1 \).
(c) Weak* convergence in $L^\infty$ implies weak convergence in $L^1$ because the space of test functions is larger ($L^1$ instead of $L^\infty$). Consequently weak closedness in $L^1$ implies weak* closedness in $L^\infty$. Hence $K_r$ is weak* closed in $L^\infty$.

3. We now prove 6. $\Rightarrow$ 2. To this end we will show that for all $X \in L^\infty$,

$$
\rho(X) = \sup_{Q \in M_1(\mathbb{P})} \left( E_Q(-X) - \alpha_{\min}(Q) \right),
$$

where we recall that

$$
\alpha_{\min}(Q) = \sup_{X \in L^\infty} \left( E_Q(-X) - \rho(X) \right) = \sup_{X \in \mathcal{A}_\rho} E_Q(-X).
$$

Let

$$
m := \sup_{Q \in M_1(\mathbb{P})} \left( E_Q(-X) - \alpha_{\min}(Q) \right).
$$

Clearly $m \leq \rho(X)$, so that it suffices to show that $m \geq \rho(X)$, i.e.

$$
X + m \in \mathcal{A}_\rho.
$$

Suppose instead that $X + m \notin \mathcal{A}_\rho$. Under this assumption we will construct a probability measure $\mathbb{Q}_0$ on $(\Omega, \mathcal{F})$ such that

$$
E_{\mathbb{Q}_0}(-X) - \alpha_{\min}(\mathbb{Q}_0) > m,
$$

in contradiction to the definition of $m$.

(a) We apply the Hahn-Banach separation theorem to the convex and weak*-closed set $\mathcal{A}_\rho$ and the point $\{X + m\}$, and obtain a linear functional $\ell$ on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ that is continuous with respect to the weak* topology and separates $\mathcal{A}_\rho$ and $X + m$ in the sense that

$$
\inf_{Y \in \mathcal{A}_\rho} \ell(Y) > \ell(X + m).
$$

(b) Since $\ell$ is continuous with respect to the weak* topology of $L^\infty$, it is of the form

$$
\ell(Y) = \int_\Omega YZd\mathbb{P}
$$

for some $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

(c) Now we show that $Z \geq 0$: For each $Y \geq 0$ and each $\lambda > 0$ we have $\rho(\lambda Y) \leq \rho(0)$, so that

$$
\lambda Y + \rho(0) \in \mathcal{A}_\rho
$$

and consequently

$$
\lambda \ell(Y) + \ell(\rho(0)) = \ell(\lambda Y + \rho(0)) > \ell(X + m).
$$

It follows that

$$
\ell(Y) > \frac{\ell(X + m - \rho(0))}{\lambda}.
$$

By letting $\lambda \to \infty$ we see that $\ell(Y) \geq 0$ and consequently $Z \geq 0$. Moreover, $\ell \not\equiv 0$ and consequently $Z \not\equiv 0$. 

20
(d) It follows that

\[ Q_0 := \frac{Z}{\mathbb{E}(Z)} \mathbb{P} \]

defines a probability measure on \((\Omega, \mathcal{F})\). Moreover,

\[
\mathbb{E}_{Q_0}(X) + m = \frac{\mathbb{E}(XZ)}{\mathbb{E}(Z)} + m = \frac{\mathbb{E}((X + m)Z)}{\mathbb{E}(Z)} = \frac{\ell(X + m)}{\mathbb{E}(Z)} < \frac{\inf_{Y \in \mathcal{A}_\mu} \ell(Y)}{\mathbb{E}(Z)} = \inf_{Y \in \mathcal{A}_\mu} \mathbb{E}(YZ) = \inf_{Y \in \mathcal{A}_\mu} \mathbb{E}_{Q_0}(Y) = -\sup_{Y \in \mathcal{A}_\mu} \mathbb{E}_{Q_0}(-Y) = -\alpha_{\min}(Q_0). \]

7 Average value at risk

In this section we suppose that \(\Omega\) is equipped with a \(\sigma\)-field \(\mathcal{F}\) and a probability measure \(\mathbb{P}\) and that all elements of \(\mathcal{G}\) are measurable with respect to \(\mathcal{F}\). For \(\lambda \in (0, 1]\) we define the average value at risk (or expected shortfall or conditional value at risk) of \(X\) at level \(\lambda\) by

\[
\text{AVaR}_\lambda(X) := \frac{1}{\lambda} \int_0^\lambda \text{VaR}_\alpha(X) d\alpha.
\]

Is this integral well-defined? We have to recall some facts about quantiles.

**Definition 7.1.** Let \(\mu\) be a probability measure on \(\mathbb{R}\) and \(\alpha \in (0, 1)\).

1. A number \(q \in \mathbb{R}\) is called an \(\alpha\)-quantile of \(\mu\) if

\[
\mu((\infty, q]) \geq \alpha \quad \text{and} \quad \mu([q, \infty)) \geq 1 - \alpha.
\]

2. A function \(q_\mu : (0, 1) \rightarrow \mathbb{R}\) is called a quantile function of \(\mu\) if for each \(\alpha \in (0, 1)\), \(q_\mu(\alpha)\) is an \(\alpha\)-quantile of \(\mu\).

**Remark 7.2.**

1. For each \(\alpha \in (0, 1)\) the set of \(\alpha\)-quantiles of \(\mu\) is a non-empty bounded closed interval. We denote its endpoints by \(q_\mu^- (\alpha)\) and \(q_\mu^+ (\alpha)\).

2. Pictures (nice case, discontinuity, interval of constancy)!!!

3. Numbers \(\alpha\) for which \(q_\mu^- (\alpha) < q_\mu^+ (\alpha)\) correspond to intervals of constancy of the cumulative distribution function of \(\mu\). It follows that the set \(\{\alpha \in (0, 1) | q_\mu^- (\alpha) < q_\mu^+ (\alpha)\}\) is countable.
4. Let $X$ be a random variable with distribution $\mu$. Then

$$-\text{VaR}_\alpha(X) = \sup\{c \in \mathbb{R} \mid \mu((\neg \infty, c]) \leq \alpha\}.$$ 

It follows that $-\text{VaR}_\alpha(X)$ is the largest $\alpha$-quantile of $\mu$,

$$-\text{VaR}_\alpha(X) = q^+_\mu(\alpha).$$

5. Let $U$ be a standard uniform random variable (i.e. a random variable which is uniformly distributed on $(0, 1)$), and let $q_\mu$ be any quantile function of $\mu$. Then $q_\mu(U)$ has distribution $\mu$. (Standard method to simulate random variables with a given distribution $\mu$)

**Proposition 7.3.** The integral

$$\int_0^1 \text{VaR}_\alpha(X) d\alpha$$

exists in $\mathbb{R}$ if and only if $X$ is quasi-integrable (i.e. has a well-defined expectation). In this case

$$\int_0^1 \text{VaR}_\alpha(X) d\alpha = E(-X).$$

**Corollary 7.4.** If $X$ is integrable, $\text{AVaR}_\lambda(X)$ exists and is finite for all $\lambda \in (0, 1]$.

**Proof of Proposition 7.3.** Let $\mu$ be the distribution of $X$, and let $U$ be a standard uniform random variable. Supposing that all integrals exist we obtain

$$\int_0^1 \text{VaR}_\alpha(X) d\alpha = -\int_0^1 q^+_\mu(\alpha) d\alpha$$

$$= -E[q^+_\mu(U)]$$

$$= E(-X).$$

Moreover, this reasoning shows that $\int_0^1 \text{VaR}_\alpha(X) d\alpha$ exists if and only if $X$ is quasi-integrable. $\blacksquare$

From now on we consider only integrable random variables, i.e. we assume that $\mathcal{G} \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

**Proposition 7.5.** Let $q_X$ be any quantile function of the distribution of $X$. Then

$$\text{AVaR}_\lambda(X) = \frac{1}{\lambda} E[(q_X(\lambda) - X)_+] - q_X(\lambda).$$

**Proof.** Let $U$ be a standard uniform random variable. Then $X$ has the same distribution as $q_X(U)$, and consequently

$$\frac{1}{\lambda} E[(q_X(\lambda) - X)_+] - q_X(\lambda) = \frac{1}{\lambda} E[(q_X(\lambda) - q_X(U)_+)] - q_X(\lambda)$$

$$= \frac{1}{\lambda} \int_0^1 (q_X(\lambda) - q_X(\alpha)_+) d\alpha - q_X(\lambda)$$

$$= \frac{1}{\lambda} \int_0^\lambda (q_X(\lambda) - q_X(\alpha)) d\alpha - q_X(\lambda)$$

$$= -\frac{1}{\lambda} \int_0^\lambda q_X(\alpha) d\alpha$$

$$= \frac{1}{\lambda} \int_0^\lambda \text{VaR}_\alpha(X) d\alpha$$

$$= \text{AVaR}_\lambda(X). \blacksquare$$
Proposition 7.6. We have

\[ \text{AVaR}_\lambda(X) = \sup \left\{ \mathbb{E}_Q(-X) \mid Q \ll P, \frac{dQ}{dP} \leq \frac{1}{\lambda} \right\}. \]

Corollary 7.7. Average value at risk is a coherent risk measure.

Proof of Proposition 7.6. The supremum on the right-hand side is equal to

\[ \sup \left\{ \mathbb{E}(\varphi X) \mid \varphi \in L^1(\Omega, \mathcal{F}, P), \mathbb{E}(\varphi) = 1, 0 \leq \varphi \leq \frac{1}{\lambda} \right\}. \]

\( \mathbb{E}(\varphi X) \) is large if \( \varphi \) takes large values at points where \( X \) takes small values and vice versa. Hence the supremum is attained for

\[ \varphi := \begin{cases} \frac{1}{\lambda} & \text{on } \{X < q_X(\lambda)\} \\ 0 & \text{on } \{X > q_X(\lambda)\} \\ \kappa & \text{on } \{X = q_X(\lambda)\}, \end{cases} \]

where \( q_X \) is any quantile function of the distribution of \( X \) and \( \kappa \) is such that \( \mathbb{E}(\varphi) = 1 \), i.e.

\[ \frac{1}{\lambda} \mathbb{P}(X < q_X(\lambda)) + \kappa \mathbb{P}(X = q_X(\lambda)) = 1. \]

(Note that \( \kappa \) necessarily satisfies \( \kappa \in [0, 1/\lambda] \) because \( \mathbb{P}(X < q_X(\lambda)) \leq \lambda \leq \mathbb{P}(X \leq q_X(\lambda)) \).

It follows that

\[
\sup \left\{ \mathbb{E}(\varphi X) \mid \varphi \in L^1(\Omega, \mathcal{F}, P), \mathbb{E}(\varphi) = 1, 0 \leq \varphi \leq \frac{1}{\lambda} \right\} \\
= -\frac{1}{\lambda} \mathbb{E}(X \cdot 1_{\{X < q_X(\lambda)\}} + \kappa \mathbb{E}(X \cdot 1_{\{X = q_X(\lambda)\}})) \\
= -\frac{1}{\lambda} \mathbb{E}(X \cdot 1_{\{X < q_X(\lambda)\}}) - \kappa q_X(\lambda) \mathbb{P}(X = q_X(\lambda)) \\
= -\frac{1}{\lambda} \mathbb{E}(X \cdot 1_{\{X < q_X(\lambda)\}}) - q_X(\lambda) + \frac{1}{\lambda} q_X(\lambda) \mathbb{P}(X < q_X(\lambda)) \\
= \frac{1}{\lambda} \mathbb{E}((q_X(\lambda) - X) \cdot 1_{\{X < q_X(\lambda)\}}) - q_X(\lambda) \\
= \frac{1}{\lambda} \mathbb{E}((-X + 1_{\{X < q_X(\lambda)\}}) - q_X(\lambda) \\
= \text{AVaR}_\lambda(X). \Box

Proposition 7.8. For \( \lambda \in (0, 1) \) let

\[ \text{WCE}_\lambda(X) := \sup \{ \mathbb{E}(-X | A) \mid A \in \mathcal{F}, \mathbb{P}(A) > \lambda \} \]

be the worst conditional expectation at level \( \lambda \) and

\[ \text{TCE}_\lambda(X) := \mathbb{E}(-X \mid -X \geq \text{VaR}_\lambda(X)) \]

be the tail conditional expectation at level \( \lambda \). Then

\[ \text{AVaR}_\lambda(X) \geq \text{WCE}_\lambda(X) \geq \text{TCE}_\lambda(X) \geq \text{VaR}_\lambda(X). \]

Moreover, the first two inequalities are equalities if \( \mathbb{P}(-X \geq \text{VaR}_\lambda(X)) = \lambda \). (By the definition of \( -\text{VaR}_\lambda(X) \) as a \( \lambda \)-quantile of \( X \) this probability is always greater than or equal to \( \lambda \), and it is equal to \( \lambda \) if \( X \) has a continuous distribution.)
Proof. The third inequality is trivial.

To prove the first inequality let \( A \in \mathcal{F} \) with \( \mathbb{P}(A) \geq \lambda \) and define \( Q := \mathbb{P}(. \, | \, A) \). Then \( Q \ll \mathbb{P} \) with \( dQ/d\mathbb{P} \leq 1/\lambda \), so that

\[
\mathbb{E}(-X \, | \, A) \leq \text{AVaR}_\lambda(X).
\]

To prove the second inequality note that for every \( \varepsilon > 0 \),

\[
\mathbb{P}\{-X \geq \text{VaR}_\lambda(X) - \varepsilon\} > \lambda.
\]

(This follows immediately from the fact that \( -\text{VaR}_\lambda(X) = \sup\{c \in \mathbb{R} \mid \mathbb{P}[X \leq c] \leq \lambda\} \). Consequently, for every \( \varepsilon > 0 \),

\[
\text{WCE}_\lambda(X) \geq \mathbb{E}(-X \mid -X \geq \text{VaR}_\lambda(X) - \varepsilon).
\]

Letting \( \varepsilon \to 0 \) gives the second inequality (using Lebesgue’s dominated convergence theorem).

Finally, suppose that \( \mathbb{P}\{-X \geq \text{VaR}_\lambda(X)\} = \lambda \). Then on the one hand

\[
\text{TCE}_\lambda(X) = \frac{\mathbb{E}(-X \cdot 1_{\{-X \geq \text{VaR}_\lambda(X)\}})}{\mathbb{P}\{-X \geq \text{VaR}_\lambda(X)\}} = \frac{1}{\lambda} \mathbb{E}(-X \cdot 1_{\{-X \geq \text{VaR}_\lambda(X)\}}),
\]

and on the other hand

\[
\text{AVaR}_\lambda(X) = \frac{1}{\lambda} \mathbb{E}((-\text{VaR}_\lambda(X) - X)_+) + \text{VaR}_\lambda(X) = \frac{1}{\lambda} \mathbb{E}((-\text{VaR}_\lambda(X) - X) \cdot 1_{\{-X \geq \text{VaR}_\lambda(X)\}}) + \text{VaR}_\lambda(X) = \frac{1}{\lambda} \mathbb{E}(-X \cdot 1_{\{-X \geq \text{VaR}_\lambda(X)\}})
\]
as well. \( \square \)

Remark 7.9.

1. The inequalities in the proposition can be strict.

2. \( \text{WCE}_\lambda \) is coherent.

3. \( \text{TCE}_\lambda \) is not even convex.

4. In contrast to the other risk measures appearing in the proposition, \( \text{WCE}_\lambda(X) \) depends not only on the distribution of \( X \) under \( \mathbb{P} \), but also on the structure of the underlying probability space.

Example 7.10. Let \( \Omega = \{\omega_1, \omega_2, \omega_3\} \), \( \mathcal{F} = \mathcal{P}(\Omega) \), \( \mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = 0.1 \) and \( \mathbb{P}(\{\omega_3\}) = 0.8 \). Define \( X_1, X_2 : \Omega \to \mathbb{R} \) by \( X_1(\omega_1) = -100, X_1(\omega_2) = X_1(\omega_3) = 0, X_2(\omega_2) = -100 \) and \( X_2(\omega_1) = X_2(\omega_3) = 0 \). We obtain

\[
\begin{align*}
\text{VaR}_{0.15}(X_1) &= 0, \\
\text{AVaR}_{0.15}(X_1) &= \frac{1}{0.15} \cdot 0.1 \cdot 100 = \frac{200}{3} \approx 67, \\
\text{WCE}_\lambda(X_1) &= 50, \\
\text{TCE}_\lambda(X_1) &= 10.
\end{align*}
\]
Moreover, letting \( Y := \frac{X_1 + X_2}{2} \) we have \( Y(\omega_1) = Y(\omega_2) = -50 \) and \( Y(\omega_3) = 0 \), and consequently
\[
\text{VaR}_{0.15}(Y) = 50, \\
\text{TCE}_{0.15}(Y) = 50,
\]
which shows that TCE is not convex.

## 8 Law-invariant risk measures and Kusuoka’s representation theorem

We still suppose that \( \Omega \) is equipped with a \( \sigma \)-field \( \mathcal{F} \) and a probability measure \( \mathbb{P} \) and that all elements of \( \mathcal{G} \) are measurable with respect to \( \mathcal{F} \). If we regard \( \mathbb{P} \) as the “true probability measure”, it is natural to concentrate on law-invariant risk measures.

**Definition 8.1.** A monetary risk measure \( \rho : \mathcal{G} \rightarrow \bar{\mathbb{R}} \) is called law-invariant or distribution-based if \( \rho(X) = \rho(Y) \) whenever \( X \) and \( Y \) have the same distribution under \( \mathbb{P} \).

**Example 8.2.**
- Value at risk (not convex),
- Average value at risk (convex).

Law-invariance clearly implies that \( \rho(X) = \rho(Y) \) whenever \( X = Y \) \( \mathbb{P} \)-a.s. Hence every law-invariant convex risk measure on \( \mathcal{L}^\infty(\Omega, \mathcal{F}) \) that is continuous from above has the representation
\[
\rho(X) = \sup_{Q \in M_1(\Omega, \mathcal{F}, \mathbb{P})} \left( \mathbb{E}_Q(-X) - \alpha_{\min}(Q) \right)
\]
with
\[
\alpha_{\min}(Q) = \sup_{X \in \mathcal{L}^\infty(\Omega, \mathcal{F})} \left( \mathbb{E}_Q(-X) - \rho(X) \right) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}_Q(-X).
\]
If \( \rho \) is law-invariant it will turn out that the only relevant aspect of the probability measures \( Q \) is the distribution of the density \( \frac{dQ}{d\mathbb{P}} \) under \( \mathbb{P} \). Consequently, the idea of Kusuoka’s representation theorem is to write \( \rho(X) \) as a supremum over distributions of probability densities with respect to \( \mathbb{P} \).

**Assumption 8.3.** The probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is atomless, i.e. for every \( A \in \mathcal{F} \) with \( \mathbb{P}(A) > 0 \) there exists \( A' \in \mathcal{F} \) with \( A' \subset A \) and \( 0 < \mathbb{P}(A') < \mathbb{P}(A) \).

**Lemma 8.4.** For a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) the following properties are equivalent:

1. \( (\Omega, \mathcal{F}, \mathbb{P}) \) is atomless.
2. There exists a standard uniform random variable \( U \) on \( (\Omega, \mathcal{F}, \mathbb{P}) \).
3. For every real-valued random variable \( X \) on \( (\Omega, \mathcal{F}, \mathbb{P}) \) there exists a standard uniform random variable \( U \) on \( (\Omega, \mathcal{F}, \mathbb{P}) \) such that
\[
X = q_X(U) \ \mathbb{P}\text{-a.s.,}
\]
where \( q_X \) is any quantile function of the distribution of \( X \).
Proof. The implications 3. $\Rightarrow$ 2. $\Rightarrow$ 1. are trivial. The implication 1. $\Rightarrow$ 3. can be found in the book by Föllmer and Schied.

We will encounter maximization problems of the following form: Given two random variables $X$ and $Y$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$,

$$\text{maximize } \mathbb{E}(X \tilde{Y})$$

over all random variables $\tilde{Y}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with the same distribution as $Y$. On an atomless probability space this problem can be easily solved:

**Lemma 8.5.** Let $X$ and $Y$ be two random variables on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and assume that one of them is integrable and the other one is bounded (this last assumption is made to ensure that $\mathbb{E}(X \tilde{Y})$ exists and is finite). Let $U$ be a standard uniform random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X = q_X(U) \mathbb{P}$-a.s. ($U$ exists by the previous lemma.) Then the maximum is attained by $\tilde{Y} = q_Y(U)$, and consequently

$$\sup_{\tilde{Y} \sim Y} \mathbb{E}(X \tilde{Y}) = \mathbb{E}(q_X(U)q_Y(U)) = \int_0^1 q_X(t)q_Y(t) dt.$$ 

Proof (heuristic argument). $\mathbb{E}(X \tilde{Y})$ becomes large if $\tilde{Y}$ takes large values when $X$ does so, and small values when $X$ does so. Since quantile functions are non-decreasing, this requirement is achieved by the choice $\tilde{Y} = q_Y(U)$. 

**Proposition 8.6.** Let $\rho$ be a law-invariant convex risk measure $\rho$ on $L^\infty(\Omega, \mathcal{F})$ that is continuous from above (and hence admits a robust representation in terms of true probability measures). Then its minimal penalty function is given by

$$\alpha_{\text{min}}(Q) = \sup_{X \in \mathbb{A}_\rho} \int_0^1 \text{VaR}_t(X)q_{dQ/d\mathbb{P}}(1-t) dt.$$ 

**Corollary 8.7.** $\alpha_{\text{min}}(Q)$ depends only on the distribution of $dQ/d\mathbb{P}$ under $\mathbb{P}$.

Proof of Proposition 8.6. Using the law-invariance of $\rho$ and the last lemma we obtain

$$\alpha_{\text{min}}(Q) = \sup_{X \in \mathbb{A}_\rho} \mathbb{E}_Q(-X)$$

$$= \sup_{X \in \mathbb{A}_\rho} \sup_{\tilde{X} \sim X} \mathbb{E}_Q(-\tilde{X})$$

$$= \sup_{X \in \mathbb{A}_\rho} \sup_{\tilde{X} \sim X} \mathbb{E}(\tilde{X}q_{dQ/d\mathbb{P}})$$

$$= \sup_{X \in \mathbb{A}_\rho} \int_0^1 q_{-X}(t)q_{dQ/d\mathbb{P}}(t) dt.$$ 

Moreover,

$$q_{-X}(t) = -q_X(1-t) = \text{VaR}_{1-t}(X).$$ 

for all but countably many $t$. Hence the claim follows.

The last proposition suggests that

$$\sup_{Q \in \mathbb{M}_1(\mathbb{P})} \left( \mathbb{E}_Q(-X) - \alpha_{\text{min}}(Q) \right)$$
might be formulated as a supremum over quantile functions $q$ of probability densities with respect to $P$. Such functions are non-decreasing and non-negative and satisfy $\int_0^1 q(t) dt = E(q(U)) = 1$. Conversely, if $(\Omega, F, P)$ is atomless, every non-decreasing function $q : (0, 1) \to \mathbb{R}_+$ with $\int_0^1 q(t) dt = 1$ is a quantile function of a probability density, namely of $q(U)$ for some standard uniform random variable $U$.

**Proposition 8.8** (Preliminary version of Kusuoka’s representation theorem). Let $\rho$ be a law-invariant convex risk measure on $L^\infty(\Omega, F)$ that is continuous from above. Then

$$\rho(X) = \sup_q \left( \int_0^1 \text{VaR}_t(X) q(1-t) dt - \beta(q) \right),$$

where the supremum is over all non-decreasing functions $q : (0, 1) \to \mathbb{R}_+$ with $\int_0^1 q(t) dt = 1$, and

$$\beta(q) = \sup_{X \in A_P} \int_0^1 \text{VaR}_t(X) q(1-t) dt.$$

**Proof.** We have

\[
\rho(X) = \sup_{Q \in M_1(\mathbb{P})} \left( E_Q(-X) - \alpha_{\min}(Q) \right)
\]

\[
= \sup_{Q \in M_1(\mathbb{P})} \left( E \left( -X \frac{dQ}{d\mathbb{P}} \right) - \alpha_{\min}(Q) \right)
\]

\[
= \sup_{Q \in M_1(\mathbb{P})} \sup_{\tilde{Q} \sim Q} \left( E \left( -X \frac{d\tilde{Q}}{d\mathbb{P}} \right) - \alpha_{\min}(\tilde{Q}) \right)
\]

\[
= \sup_{Q \in M_1(\mathbb{P})} \left( \sup_{\tilde{Q} \sim Q} \left( E \left( -X \frac{d\tilde{Q}}{d\mathbb{P}} \right) - \alpha_{\min}(\tilde{Q}) \right) \right)
\]

\[
= \sup_{Q \in M_1(\mathbb{P})} \left( \int_0^1 q_{-X(t)} q_{dQ/d\mathbb{P}}(t) dt - \alpha_{\min}(Q) \right)
\]

\[
= \sup_{Q \in M_1(\mathbb{P})} \left( \int_0^1 \text{VaR}_t(X) q_{dQ/d\mathbb{P}}(1-t) dt - \beta(q_{dQ/d\mathbb{P}}) \right).
\]

Writing this as a supremum over all quantile functions of probability densities (with respect to $\mathbb{P}$) we get the result. \qed

**Lemma 8.9.** There is a one-to-one correspondence between left-continuous non-decreasing functions $q : (0, 1) \to \mathbb{R}_+$ with $\int_0^1 q(t) dt = 1$, and probability measures $\mu$ on $(0, 1]$, which is given as follows: With a probability measure $\mu$ on $(0, 1]$ we associate the function $q$ defined by

$$q(t) = \int_{(1-t, 1]} \frac{1}{s} \mu(ds).$$

**Proof.**

1. For a given probability measure $\mu$ on $(0, 1]$ define the measure $\nu$ on $(0, 1]$ by

$$\nu(ds) = \frac{1}{s} \mu(ds).$$

Then

$$q(t) = \nu((1 - t, 1]).$$
from which it is clear that \( q \) is non-negative, non-decreasing and left-continuous. Moreover,

\[
\int_0^1 q(t)\,dt = \int_0^1 q(1-t)\,dt = \int_0^1 \nu((t, 1])\,dt = \int_0^1 \int_{(0,1]} 1_{t\in(t,1]} \nu(ds)\,dt = \int_{(0,1]} 1_{t\in[0,s]}\,dt \, \nu(ds) = \int_{(0,1]} sv(ds) = \mu((0,1]) = 1.
\]

2. Conversely, let \( q : (0, 1) \to \mathbb{R}_+ \) be non-negative, non-decreasing and left-continuous and such that \( \int_0^1 q(t)\,dt = 1 \). We extend \( q \) to a function on \([0, 1)\) by setting \( q(0) := 0 \), and define \( f : (0, 1] \to \mathbb{R}_+ \) by

\[ f(t) = -q(1-t). \]

\( f \) is non-decreasing and right-continuous and therefore induces a measure \( \nu \) (the Stieltjes measure) on \((0, 1]\) by

\[ \nu((a, b]) := f(b) - f(a) = q(1-a) - q(1-b). \]

In particular, for \( b = 1 \) we get

\[ \nu((a, 1]) = q(1-a) \]

(because \( q(0) = 0 \)). Let now

\[ \mu(ds) := s\nu(ds). \]

By essentially the same computation as above we obtain that \( \mu \) is a probability measure:

\[
\mu((0,1]) = \int_{(0,1]} sv(ds) = \int_{(0,1]} \int_0^1 1_{t\in(0,s]}\,dt \, \nu(ds) = \int_0^1 \int_{(0,1]} 1_{s\in(t,1]} \nu(ds)\,dt = \int_0^1 \nu((t, 1])\,dt = \int_0^1 q(1-t)\,dt = \int_0^1 q(t)\,dt = 1.
\]
Finally, the two procedures \((\mu \to \nu \to q \text{ and } q \to \nu \to \mu)\) are clearly inverse to each other. 

**Theorem 8.10** (Kusuoka’s representation theorem). Let \(\rho\) be a law-invariant convex risk measure \(\rho\) on \(L^\infty(\Omega, \mathcal{F})\) that is continuous from above. Then

\[
\rho(X) = \sup_{\mu \in \mathcal{M}_1((0,1])} \left( \int_{(0,1]} \text{AVaR}_\lambda(X) \mu(d\lambda) - \gamma(\mu) \right),
\]

where

\[
\gamma(\mu) = \sup_{X \in \mathcal{A}_r} \int_{(0,1]} \text{AVaR}_\lambda(X) \mu(d\lambda).
\]

**Proof.** By the preliminary version of Kusuoka’s representation theorem,

\[
\rho(X) = \sup_q \left( \int_0^1 \text{VaR}_t(X) q(1-t) dt - \beta(q) \right),
\]

where the supremum is over all non-decreasing functions \(q : (0, 1) \to \mathbb{R}_+\) with \(\int_0^1 q(t) dt = 1\), and

\[
\beta(q) = \sup_{X \in \mathcal{A}_r} \int_0^1 \text{VaR}_t(X) q(1-t) dt.
\]

For

\[
q(t) = \int_{[1-t,1]} \frac{1}{s} \mu(ds)
\]

we obtain

\[
\int_0^1 \text{VaR}_t(X) q(1-t) dt = \int_0^1 \text{VaR}_t(X) \int_{(t,1]} \frac{1}{s} \mu(ds) dt
\]

\[
= \int_0^1 \int_{(0,1]} \text{VaR}_t(X) 1_{s \in (t,1]} \frac{1}{s} \mu(ds) dt
\]

\[
= \int_{(0,1]} \int_0^1 \text{VaR}_t(X) 1_{t \in (0,s]} \frac{1}{s} dt \mu(ds)
\]

\[
= \int_{(0,1]} \frac{1}{s} \int_0^s \text{VaR}_t(X) dt \mu(ds)
\]

\[
= \int_{(0,1]} \text{AVaR}_s(X) \mu(ds).
\]

In view of the one-to-one correspondence between functions \(q\) and probability measures \(\mu\) this proves the result. 

\[\square\]