Distribution-Invariant Risk Measures, Entropy, and Large Deviations

Stefan Weber *

Cornell University

June 24, 2004; this version December 4, 2006

Abstract

The simulation of distributions of financial positions is an important issue for financial institutions. If risk measures are evaluated for a simulated distribution instead of the model-implied distribution, errors of risk measurements needs to be analyzed. For distribution-invariant risk measures which are continuous on compacts we employ the theory of large deviations to study the probability of large errors. If the approximate risk measurements are based on the empirical distribution of independent samples, the rate function equals the minimal relative entropy under a risk measure constraint. For shortfall risk and average value at risk (AVaR) we solve this minimization problem explicitly.

Key words: Risk measures, average value at risk, shortfall risk, Monte Carlo, large deviation principle, Sanov's theorem, relative entropy

JEL Classification: G11; G12; G13

Mathematics Subject Classification (2000): 91B30; 49Q20; 62B10; 62D05; 91B28

1 Introduction

The portfolios of banks consist of financial assets such as stocks, bonds, credits and options. The quantification of the risk associated with these positions is of crucial importance, since banks need to manage their risks and are obliged to respect regulatory constraints. This requires both suitable models of portfolio holdings and appropriate numerical measures of risk. In practice, financial positions are frequently modeled as real-valued random variables on some underlying probability space. In such a setting, the modeling assumptions determine in particular the distributions of the financial positions. A standard approach to measure risk is to use certain functionals of these distributions, namely static distribution-invariant risk measures.

A theory of such risk measures is already well developed. Nevertheless, the implementation of risk measurements requires further analysis. Model distributions are often not directly tractable, but

*School of Operations Research and Industrial Engineering, Cornell University, 279 Rhodes Hall, Ithaca, NY 14853, USA, email sweber@orie.cornell.edu. I would like to thank Boris Buchmann, Hans Föllmer and Ulrich Horst for helpful discussions. I am very grateful for useful comments of an anonymous referees. I acknowledge financial support by Deutsche Forschungsgemeinschaft via Graduiertenkolleg 251 'Stochastische Prozesse und Probabilistische Analysis' and SFB 373 'Quantifikation und Simulation ökonomischer Prozesse'.

can only be simulated by Monte Carlo methods. If risk measures are evaluated for the simulated instead of the model-implied distributions, actual risk measurements deviate from the model-implied risk and the errors of these measurements need to be analyzed.

In the current article we employ the theory of large deviations to study these errors for various risk measures. We investigate large deviation bounds for a broad class of static risk measures. Specific examples include shortfall risk, and average value at risk. For an axiomatic analysis of coherent risk measures we refer to Artzner, Delbaen, Eber & Heath (1999); for extensions to general probability spaces and convex risk measures see Delbaen (2002), Föllmer & Schied (2002), Frittelli & Rosazza (2002), and Föllmer & Schied (2004).

The paper is outlined as follows. In a first step, we describe how the error of the risk measurements and large deviations are related. In Section 2 we investigate conditions under which a large deviation principle (LDP) holds for risk measurements. A LDP can be derived from a contraction principle, if the risk measure satisfies a certain regularity property, i.e., is continuous on compacts. This notion is introduced in Section 2, and a contraction principle for the corresponding class of risk measures is formulated. Section 3 analyzes the notion of continuity on compacts. In particular, we characterize coherent distribution-invariant risk measures which are continuous on compacts. Examples include average value at risk (AVaR) and shortfall risk. Further properties of these risk measures are discussed in Föllmer & Schied (2004), Weber (2006), Giesecke, Schmidt & Weber (2005), and Dunkel & Weber (2005).

If risk measurements are based on empirical distributions, the rate function of the LDP can be characterized more explicitly. For independent samples the rate function of the large deviations of the risk measurements is given as the minimal relative entropy under a risk measure constraint. Based on general methods of Csiszar (1975), we calculate the minimal relative entropy explicitly for both shortfall risk and average value at risk in Sections 4 and 5.

For a shortfall risk constraint, the calculation of the minimal relative entropy only involves a linear constraint. A solution to the problem is obtained in Section 4. In the case of AVaR which we consider in the Section 5 the analysis is more complicated. Our solution is based on a particular representation of AVaR as the expected loss under the worst case measure which can be computed by means of the Neyman-Pearson lemma, see equation (10). The constraint set of the minimization problem is in general not convex, and the calculation is quite involved. Necessary and sufficient conditions for the existence of a solution are formulated in terms of the parameters of the problem. For AVaR, we solve the original problem in two steps. The first step consists of minimizing the relative entropy under a linear constraint. Minimizing densities and minimal relative entropies are explicitly calculated. In a second step, a minimization problem with three varying parameters has to be solved.

2 Large deviation bounds

We always assume that $(\Omega, \mathcal{F}, P)$ is a rich probability space, i.e., a probability space on which a random variable with continuous distribution exists. We recall the following definition.

Definition 2.1. *A mapping $\rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is called a distribution-invariant risk measure, if it satisfies the following conditions for all $X, Y \in L^\infty$:*

- Monotonicity: *If $X \le Y$, then $\rho(X) \ge \rho(Y)$.*
- Translation-property: *If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$.*

- Distribution-invariance: *If* $P \circ X^{-1} = P \circ Y^{-1}$, *then* $\rho(X) = \rho(Y)$.

We denote by $\mathcal{M}_{1,c} = \mathcal{M}_{1,c}(\mathbb{R})$ the space of Borel probability measures on $\mathbb{R}$ with compact support. We endow $\mathcal{M}_{1,c}$ with the weak topology. A distribution-invariant risk measure ρ defines a functional $\rho' : \mathcal{M}_{1,c} \to \mathbb{R}$ by $\rho'(\mu) = \rho(X)$ for some $X \in L^\infty$ with distribution $\mathcal{L}(X) := P \circ X^{-1} = \mu$. For more details see Weber (2006) and Weber (2004).

We consider the following situation. Assume that we are interested in the risk of a financial position $X \in L^\infty$ with distribution $\mu = \mathcal{L}(X)$. Suppose that the distribution μ is not directly tractable, but that samples of μ can be generated.

For example, let (X_i) be a sequence of independent random variables on the probability space $(\Omega, \mathcal{F}, P)$ with identical distribution μ. The empirical distribution of the first n samples $X_1, X_2, \ldots, X_n$ is then given by the random measure

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}. \tag{1}$$

Here, δ_x denotes the Dirac measure placing all mass on $x \in \mathbb{R}$. Then (μ_n) converges P-almost surely to μ in the weak topology.

A naive Monte Carlo procedure for simulating $\rho(X)$ is to calculate $\rho'(\mu_n)$, $n \in \mathbb{N}$. A possible measure of the quality of the nth approximation is the probability that the error of the simulated risk deviates from the true risk of X by more than a given bound $\epsilon > 0$, i.e.,

$$P\left(\left|\rho'(\mu_n) - \rho(X)\right| > \epsilon\right).$$

If ρ is regular enough, asymptotic upper and lower exponential bounds for these error probabilities can be obtained from the theory of large deviations.

Definition 2.2. *A distribution-invariant risk measure* $\rho : L^\infty \to \mathbb{R}$ *is called* continuous on compacts, *if for all compact sets* $K \subseteq \mathbb{R}$ *the restriction of* ρ' *to* $\mathcal{M}_1(K)$ *is continuous. Here,* $\mathcal{M}_1(K)$ *denotes the space of probability measures supported in* K.

The notion of continuity on compacts is weaker than continuity of risk measures with respect to the weak topology. For risk measures which are continuous on compacts a LDP is an immediate consequence of the contraction principle.

Proposition 2.3. *Let* ρ *be a distribution-invariant risk measure that is continuous on compacts. Assume that* $(\mu_n) \subseteq \mathcal{M}_{1,c}$ *is a sequence of random measures that satisfies a LDP with rate* (γ_n) *and rate function* I. *Additionally, assume that there exists a compact set* $K \subseteq \mathbb{R}$ *such that* supp $\mu_n \subseteq K$ *for all* n. *Then* $(\rho'(\mu_n))_n$ *satisfies a LDP with rate* (γ_n) *and rate function*

$$J(x) := \inf\left\{I(\nu) : \nu \in \mathcal{M}_{1,c},\ x = \rho'(\nu)\right\}.$$

Proof. The proposition is an immediate consequence of the contraction principle for Hausdorff spaces (see Dembo & Zeitouni (1998), Theorem 4.2.1, p. 126). □

Remark 2.4. *Value at risk is not continuous on compacts. Nevertheless, it is possible to derive upper large deviation bounds by direct calculations, see e.g. Fu, Jin & Xiong (2003) and Weber (2004).*

We specialize now to the case of empirical measures (μ_n) as defined in equation (1). For independent samples, the sequence $(\mu_n)_n$ converges P-almost surely to μ. Thus, a strong law of large numbers holds for the risk measurements. At the same time, we have the following LDP for the risk measures:

Corollary 2.5. *Let ρ be continuous on compacts. Then $(\rho'(\mu_n))_n$ satisfies a LDP with rate n and rate function*

$$J(x) = \inf\{H(\nu|\mu) : \ \nu \in \mathcal{M}_{1,c}, \quad x = \rho'(\nu)\}. \tag{2}$$

Here, $H(\nu|\mu)$ denotes the relative entropy of the probability measure ν with respect to μ defined by

$$H(\nu|\mu) := \begin{cases} \int f \log f d\mu & \text{if } f := \frac{d\nu}{d\mu} \text{ exists} \\ \infty & \text{otherwise.} \end{cases} \tag{3}$$

Proof. The proof is a corollary of Sanov's Theorem (see e.g. Dembo & Zeitouni (1998), Theorem 6.2.10) and the contraction principle for empirical measures stated in Corollary 2.3. □

3 Continuity on compacts

In this section we characterize risk measures which are continuous on compacts and thus satisfy the contraction principle of the preceding section.

The following theorem is elementary, but allows us to identify examples of risk measures which are continuous on compacts. We recall that a risk measure ρ is called *continuous from above*, if $X_n \searrow X$ P-almost surely implies $\rho(X_n) \nearrow \rho(X)$. Analogously, ρ is called *continuous from below*, if $X_n \nearrow X$ P-almost surely implies $\rho(X_n) \searrow \rho(X)$.

Theorem 3.1. *Let ρ be a distribution-invariant risk measure. The following conditions are equivalent:*

(1) ρ is continuous on compacts.

(2) ρ is both continuous from above and from below.

(3) ρ is continuous for bounded sequences, i.e., for every bounded sequence (X_n) converging P-almost surely to some X it holds that $\lim_{n\to\infty} \rho(X_n) = \rho(X)$.

Proof. (2)⇒(3): Let (X_n) be bounded and converging P-almost surely to X. Then $\left(\sup_{m\geq n} X_m\right)_n$ and $(\inf_{m\geq n} X_m)_n$ converge to X from above and below, respectively. Thus,

$$\rho(X) = \lim_n \rho\left(\sup_{m\geq n} X_m\right) \leq \liminf_{n\to\infty} \rho(X_n) \leq \limsup_{n\to\infty} \rho(X_n) \leq \lim_n \rho\left(\inf_{m\geq n} X_n\right) = \rho(X).$$

(3)⇒(2): If (X_n) converges to X from below or from above, then (X_n) is bounded. This implies the claim.

(1)⇒(3): Let (X_n) be a bounded sequence converging P-almost surely to some X. Then there exists a compact $K \subseteq \mathbb{R}$ such that P-almost surely $X_n, X \in K$ $(n \in \mathbb{N})$. Clearly, $\mathcal{L}(X_n), \mathcal{L}(X) \in \mathcal{M}_1(K)$ and $\mathcal{L}(X_n) \to \mathcal{L}(X)$. Thus, $\rho(X_n) = \rho'(\mathcal{L}(X_n)) \to \rho'(\mathcal{L}(X)) = \rho(X)$.

(3)⇒(1): Let $K \subseteq \mathbb{R}$ be a compact set and assume that $\mu_n \Rightarrow \mu$ for $\mu_n, \mu \in \mathcal{M}_1(K)$. Denote by F_n, F the distribution functions of μ_n, μ, respectively. Since $(\Omega, \mathcal{F}, P)$ is rich, there exists a random variable Z with $\mathcal{L}(Z) = \text{unif}\,[0,1]$. Define $X_n := F_n^{-1}(Z)$, $X := F^{-1}(Z)$, where F_n^{-1} and F^{-1} are the right-continuous inverses of F_n and F, respectively. Observe that $X_n \to X$ P-a.s. as $n \to \infty$. Moreover, $X_n, X \in K$ P-a.s. Hence,

$$\rho'(\mu_n) = \rho(X_n) \to \rho(X) = \rho'(\mu). \tag{4}$$

□

We provide examples for risk measures which are continuous on compacts. The current industry standard *value at risk* is not continuous on compacts. In contrast, the coherent risk measure *average value at risk* is continuous on compacts. We recall that a risk measure is coherent, if it satisfies for $X, Y \in L^\infty$ both

- convexity: $\rho(\alpha X + (1-\alpha)Y) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$, $\alpha \in [0,1]$, and
- positive homogeneity: $\rho(\alpha X) = \alpha\rho(X)$, $\alpha \geq 0$.

Example 3.2. *Value at risk at level $\lambda \in (0,1)$ is defined as*

$$VaR_\lambda(X) = \inf\{m \in \mathbb{R} : \ P[m + X < 0] \leq \lambda\}.$$

In order to see that value at risk is not continuous on compacts, we consider the following example. Let the probability space $(\Omega, \mathcal{F}, P)$ given by the unit interval $[0,1]$ with Lebesgue measure. For $\frac{1}{n} < 1-\lambda$ define $X_n = \mathbf{1}_{[\lambda+1/n,1]}$. Then $X_n \nearrow X := \mathbf{1}_{[\lambda,1]}$ as $n \to \infty$. But, $VaR_\lambda(X_n) = 0$ does not converge to $VaR_\lambda(X) = -1$ as $n \to \infty$.

Example 3.3. Average value at risk *at level $\lambda \in (0,1]$ is defined as*

$$AVaR_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda VaR_\gamma(X)d\gamma.$$

Average value at risk is continuous on compacts.

Proof. According to Theorem 4.47 in Föllmer & Schied (2004), $AVaR_\lambda$ is continuous from below. By Theorem 4.31, $AVaR_\lambda$ is continuous from above. Thus, $AVaR$ is continuous on compacts by Theorem 3.1. □

Average value at risk is an important building block for coherent distribution-invariant risk measures. We quote the following theorem of Kusuoka (2001).

Theorem 3.4. *On a rich probability space, a coherent distribution-invariant risk measure ρ is continuous from above, if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \int_{(0,1]} AVaR_\lambda(X)\mu(d\lambda) \tag{5}$$

for some set $\mathcal{M} \subseteq \mathcal{M}_1((0,1])$. Here, $\mathcal{M}_1((0,1])$ denotes the space of Borel probability measures on $(0,1]$. The assumption of continuity from above can be dropped, if the underlying probability space is standard, see Jouini, Schachermayer & Touzi (2006).

This representation of coherent distribution-invariant risk measures provides a further perspective on risk measures which are continuous on compacts, see Theorem 3.5 and Proposition 3.9 below.

We denote by $AVaR_0(X) := \|X^-\|$ the essential infimum of X. Motivated by Kusuoka's Theorem, we introduce the following notation. If a measure $\mu \in \mathcal{M}_1([0,1])$, i.e., μ is a Borel probability measure on $[0,1]$, then we write

$$\rho_\mu(X) := \int_{[0,1]} AVaR_\lambda(X)\mu(d\lambda) \qquad (X \in L^\infty).$$

For coherent risk measures we can now state necessary and sufficient conditions for continuity on compacts.

Theorem 3.5. *For a distribution-invariant coherent risk measure ρ the following properties are equivalent:*

(1) ρ is continuous from below.

(2) ρ is continuous on compacts.

(3) ρ is continuous in the Mackey topology $\tau(L^\infty, L^1)$.

(4) There exists a law-invariant set $\mathcal{D} \subseteq L^1$ of densities representing ρ such that the supremum is attained:

$$\rho(X) = \max_{h\in\mathcal{D}} E[-hX] \quad \text{for all } X \in L^\infty.$$

(5) There exists a law-invariant set $\mathcal{D} \subseteq L^1$ of densities representing ρ which is $\sigma(L^1, L^\infty)$-compact:

$$\rho(X) = \sup_{h\in\mathcal{D}} E[-hX] \quad \text{for all } X \in L^\infty.$$

(6) The maximal representing set

$$\mathcal{D}_{\max} =: \left\{ h \in L^1 : \|h\|_{L^1} = 1,\ h \geq 0,\ \sup_{X\in L^\infty} [E[-hX] - \rho(X)] = 0 \right\}$$

is $\sigma(L^1, L^\infty)$-compact.

(7) There exists a law-invariant set $\mathcal{D} \subseteq L^1$ of densities representing ρ such that the supremum is attained:

$$\rho(X) = \max_{h\in\mathcal{D}} \int_0^1 q_h(t)q_{-X}(t)dt \quad \text{for all } X \in L^\infty.$$

Here, q_Y denotes the quantile function of a random variable Y.

(8) There exists a set $\mathcal{M} \subseteq \mathcal{M}((0,1])$ of probability measures representing ρ such that the supremum is attained:

$$\rho(X) = \max_{\mu\in\mathcal{M}\subseteq\mathcal{M}_1((0,1])} \int_{(0,1]} AVaR_\gamma(X)\mu(d\gamma) \quad \text{for all } X \in L^\infty.$$

In (4), (5) and (7) the set $\mathcal{D}$ may be chosen convex. The set $\mathcal{D}_{\max}$ in (6) is always convex.

Proof. (1) $\Leftrightarrow$ (2): This follows from Corollary 4.35 and Theorem 4.31 in Föllmer & Schied (2004) and Theorem 3.1.
(5) $\Rightarrow$ (1) $\Leftrightarrow$ (4): Except for the law-invariance of $\mathcal{D}$ in (4), this is an immediate consequence of Corollary 4.35 in Föllmer & Schied (2004). We show that $\mathcal{D}$ can be chosen law-invariant. By

Corollary 4.34 in Föllmer & Schied (2004) we may choose $\mathcal{D}$ as the maximal representing set of densities

$$\mathcal{D}_{\max} =: \left\{ h \in L^1 : \ \|h\|_{L^1} = 1, \ h \geq 0, \ \alpha_{\min}(h) = 0 \right\}$$

which is defined in terms of the minimal penalty function

$$\alpha_{\min}(h) = \sup_{X \in L^\infty} \left[E[-hX] - \rho(X) \right].$$

By Theorem 4.54 in Föllmer & Schied (2004) $\alpha_{\min}$ depends only on the law of its argument. Thus, $\mathcal{D}_{\max}$ is law-invariant.

(4) $\Rightarrow$ (5): This is a consequence of James' theorem, see Theorem A.66 in Föllmer & Schied (2004). The proof parallels the argument in Corollary 4.35 in Föllmer & Schied (2004). $\mathcal{D}$ is law-invariant by assumption.

(4) $\Rightarrow$ (6): As we pointed out in the proof of (5) $\Rightarrow$ (1) $\Leftrightarrow$ (4), the set $\mathcal{D}$ in (4) can be assumed to be $\mathcal{D}_{\max}$. The weak compactness of $\mathcal{D}_{\max}$ follows again by James' theorem.

(6) $\Rightarrow$ (4): This is a special case of the implication (5) $\Rightarrow$ (4).

(4) $\Rightarrow$ (3): This is an immediate consequence of Theorem 5.102 in Aliprantis & Border (1999).

(3) $\Rightarrow$ (4): By Theorem 5.102 in Aliprantis & Border (1999) we have $\rho(X) = \max_{h \in \mathcal{D}} E[-hX]$ where

$$\mathcal{D} = \left\{ h \in L^1 : E[-hX] \leq \rho(X) \ \forall X \in L^\infty \right\}$$

is $\sigma(L^1, L^\infty)$-compact.

If $h \in \mathcal{D}$, then $h \geq 0$ P-almost surely. Suppose not. With $X = \mathbf{1}_{\{h<0\}}$ we obtain for $n \in \mathbb{N}$ that

$$E[-nh\mathbf{1}_{\{h<0\}}] \leq \rho(nX) \leq \rho(X).$$

The left-hand side of this inequality converges to ∞ as $n \to \infty$, contradicting $\rho(X) \in \mathbb{R}$.

Finally, we obtain with $X = 1$ that $E[-h] \leq \rho(1) = -1$, thus $E[h] \geq 1$, and with $X = -1$ that $E[h] \leq \rho(-1) = 1$. This implies that $\mathcal{D}$ is a set of densities. Since $(\Omega, \mathcal{F}, P)$ is a rich probability space and since ρ is distribution-invariant, we can use Theorem 6.10 in Kallenberg (1997) to show that $\mathcal{D}$ is law-invariant.

(7) $\Leftrightarrow$ (8): For a probability density $h \in L^1$ we define

$$[h] = \left\{ \hat{h} \in L^1 : \ \hat{h} \overset{P}{\sim} h \right\}.$$

We show that the equivalence classes

$$\{[h] : \ h \in L^1, h \geq 0, \|h\|_{L^1} = 1\}$$

and the probability measures $\mathcal{M}((0,1])$ are in 1-to-1 correspondence.

Let μ be given. By Lemma 4.63 in Föllmer & Schied (2004) μ corresponds uniquely to an increasing and concave function $\psi : [0,1] \to [0,1]$ with $\psi(0) = 0$ and $\psi(1) = 1$. Its right-continuous right-hand side derivative ψ'_+ is therefore positive and

$$\int_0^1 \psi'_+(t)dt = \psi(1) - \psi(0) = 1.$$

ψ'_+ defines hence a probability density on $((0,1), \mathcal{B}, \lambda)$ where λ denotes Lebuesgue measure. Since $(\Omega, \mathcal{F}, P)$ is a rich probability space, there exists $h \in L^1$ with $\mathcal{L}(h; P) = \mathcal{L}(\psi'_+; \lambda)$ where the notation $\mathcal{L}(\cdot;\cdot)$ signifies the respective law. We define $M(\mu) := [h]$.

Conversely, let $h \in L^1$ be a probability density with equivalence class $[h]$. Then $\psi(t) = \int_0^t q_h(1-u)du$ depends on the distribution of h only. We have $\psi(0) = 0$ and $\psi(1) = E_P(h) = 1$, since h is a probability density. ψ is increasing, since $q_h \geq 0$. Moreover, $t \mapsto q_h(1-t)$ is decreasing which implies that ψ is concave. By Lemma 4.63 in Föllmer & Schied (2004) ψ corresponds to a unique probability measure $\mu \in \mathcal{M}_1((0,1])$. From the construction it is obvious that $[h] = M(\mu)$.

For fixed $\mu \in \mathcal{M}((0,1])$ we set $\rho_\mu(X) = \int_{(0,1]} AVaR_\gamma(X)\mu(d\gamma)$. Let ψ be according to Lemma 4.63 in Föllmer & Schied (2004) the function corresponding to μ. For $h \in M(\mu)$ we have by construction that $q_h(t) = \psi'_+(1-t)$ for $t \in [0,1]$. By Lemma 4.63 and Theorem 4.64 in Föllmer & Schied (2004)

$$\rho_\mu(X) = \int_0^1 q_{-X}(t)\psi'_+(1-t)dt = \int_0^1 q_{-X}(t)q_h(t)dt.$$

Assume now that (7) holds. Then

$$\rho(X) = \max_{h\in\mathcal{D}} \int_0^1 q_h(t)q_{-X}(t)dt = \max_{h=M^{-1}([h]),\, h\in\mathcal{D}} \rho_\mu(X) = \max_{\mu\in\mathcal{M}} \rho_\mu(X),$$

where $\mathcal{M} = \left\{M^{-1}([h]) : \ h \in \mathcal{D}\right\}$.

Conversely, if (8) holds, we obtain

$$\rho(X) = \max_{\mu\in\mathcal{M}} \rho_\mu(X) = \max_{h\in\mathcal{D}} \int_0^1 q_h(t)q_{-X}(t)dt,$$

where $\mathcal{D} = \bigcup_{\mu\in\mathcal{M}} M(\mu)$.

(4) $\Rightarrow$ (7): For $h \in \mathcal{D}$ we have by the Hardy-Littlewood inequality, see Theorem A.24 in Föllmer & Schied (2004),

$$E[-hX] \leq \int_0^1 q_h(t)q_{-X}(t)dt.$$

Since $(\Omega, \mathcal{F}, P)$ is rich, there exists a pair of random variables $(\tilde{h}, Y)$ such that $\mathcal{L}((\tilde{h}, -Y); P) = \mathcal{L}((q_h, q_{-X}); \lambda)$. Note that, of course, $\tilde{h} \overset{P}{\sim} h$ and $Y \overset{P}{\sim} X$, but $\tilde{h}$ and Y are anticomonotonic. This implies that

$$\int_0^1 q_h(t)q_{-X}(t)dt = E(-\tilde{h}Y) \leq \rho(Y) = \rho(X).$$

Taking suprema leads to $\rho(X) = \max_{h\in\mathcal{D}} E[-hX] = \sup_{h\in\mathcal{D}} \int_0^1 q_h(t)q_{-X}(t)dt$. Letting $h^* = \text{argmax}_{h\in\mathcal{D}} E[-hX]$, we have

$$\rho(X) = E[-h^*X] \leq \int_0^1 q_{h^*}(t)q_{-X}(t)dt \leq \rho(X).$$

Hence, the supremum in (7) is attained.

(8) $\Rightarrow$ (1): For fixed $\mu \in \mathcal{M}((0,1])$ we set $\rho_\mu(X) = \int_{(0,1]} AVaR_\gamma(X)\mu(d\gamma)$. By Corollary 4.74 in Föllmer & Schied (2004) there exists a set of densities $\mathcal{D}_\mu$ such that $\rho_\mu(X) = \max_{h\in\mathcal{D}_\mu} E[-hX]$. Letting $\mathcal{D} = \bigcup_{\mu\in\mathcal{M}} \mathcal{D}_\mu$ we obtain that $\rho(X) = \max_{h\in\mathcal{D}} E[-hX]$. Corollary 4.35 in Föllmer & Schied (2004) implies that ρ is continuous from below. □

Remark 3.6. *If $(\Omega, \mathcal{F}, P)$ is a standard probability space, then every distribution-invariant convex risk measure is continuous from above. This has recently been pointed out in Jouini et al. (2006). Their theorem provides another justification for the fact that continuity on compacts is implied by continuity from below.*

Remark 3.7. *If the maximizing density in Corollary 3.5(3) can explicitly be calculated, this can be used in order to find explicitly the minimal relative entropy under the risk measure constraint. This provides a way to obtain the rate function in Corollary 2.5. We will consider the case of* Average value at risk *in Section 5.*

Corollary 3.8. *For a risk measure $\rho : L^\infty \to \mathbb{R}$ the following properties are equivalent:*

(1) ρ is distribution-invariant, coherent and continuous on compacts.

(2) The acceptance set $\mathcal{A}$ of ρ is a law-invariant convex cone, and for increasing sequences $(X_n)_n \subseteq L^\infty$ with $X = \lim_{n\to\infty} X_n$ P-a.s., $X \in \mathcal{A}$ and any $\epsilon > 0$ we have $X_n + \epsilon \in \mathcal{A}$ for large enough n.

Proof. (1) $\Rightarrow$ (2): Obvious. (2) $\Rightarrow$ (1): Law-invariance and coherence of ρ are immediate. By Theorem 3.5 we need to verify continuity from below. Letting $Y_n \nearrow Y$ P-a.s., we have $X = Y + \rho(Y) \in \mathcal{A}$, $X_n = Y_n + \rho(Y) \nearrow X$ P-a.s. For any $\epsilon > 0$ we have $\rho(X_n + \epsilon) \leq 0$ for large enough n, thus $\rho(Y_n) \leq \rho(Y) + \epsilon$. This implies that $\rho(Y_n)$ converges to $\rho(Y)$ as $n \to \infty$. □

Corollary 3.9. *Suppose a coherent distribution-invariant risk measure ρ admits a representation (5) for some weakly compact set $\mathcal{M} \subseteq \mathcal{M}_1((0,1])$. Then the supremum in (5) is actually a maximum, and ρ is continuous on compacts.*

Proof. Since $\lambda \mapsto AVaR_\lambda(X)$ is continuous and bounded in $[-\|X\|, \|X\|]$, $\mu \mapsto \int_{(0,1]} AVaR_\gamma(X)\mu(d\gamma)$ defines a weakly continuous functional on $\mathcal{M}((0,1])$. Since $\mathcal{M}$ is weakly compact, the supremum in (5) is attained. By Theorem 3.5 we obtain that ρ is continuous on compacts. □

The condition of Proposition 3.9 is, of course, satisfied, if the set $\mathcal{M}$ is a singleton. In this case, ρ is simply a mixture of average value at risk at different levels. By a theorem of Schmeidler (1986) the class of such risk measures is closely related to the family of distribution-invariant risk measures that are comonotonic, see Section 4.7 in Föllmer & Schied (2004).

Theorem 3.10. *On a rich probability space, the class of risk measures*

$$\rho_\mu(X) = \int AVaR_\lambda(X)\mu(d\lambda), \qquad \mu \in \mathcal{M}_1([0,1])$$

is precisely the class of all distribution-invariant convex risk measures on L^∞ that are comonotonic. In particular, these risk measures are also coherent. ρ_μ is continuous on compacts, if and only if $\mu(\{0\}) = 0$.

Proof. For the first part of the theorem see Theorem 4.87 in Föllmer & Schied (2004). The second part follows from Corollary 4.74 in Föllmer & Schied (2004) and Theorem 3.1. □

Remark 3.11. *ρ_μ can also be represented as a Choquet integral with respect to a continuous concave distortion of the underlying probability measure, see Corollary 4.71 in Föllmer & Schied (2004)*

Finally, we discuss an important class of distribution-invariant convex risk measures which are continuous on compacts, namely shortfall risk.

Definition 3.12. *Let $\ell : \mathbb{R} \to \mathbb{R}$ be a convex loss function, i.e. an increasing, non constant and convex function. Assume that z is an interior point of the range of ℓ. We define the acceptance set*

$$\mathcal{A} = \left\{ X \in L^\infty : \int \ell(-X)dP \leq z \right\}.$$

The shortfall risk is defined by

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

Shortfall risk is a distribution-invariant risk measure which is continuous from above and below, cf. Proposition 4.104 & Theorem 4.31 in Föllmer & Schied (2004). Thus, by Theorem 3.1 shortfall risk is continuous on compacts. Like average value at risk, shortfall risk has many desirable properties. In contrast to value at risk, it encourages diversification, since it is convex, and does not neglect the size of losses. For a detailed analysis of this risk measure, including applications to dynamic risk measurement and Monte Carlo simulations, we refer to Weber (2006), Giesecke et al. (2005), and Dunkel & Weber (2005).

4 Entropy minimization under a shortfall risk constraint

As we have seen in Corollary 2.5, for independently generated samples the rate function of the large deviations of risk measures is determined by the minimal relative entropy under a risk measure constraint. In the current section, we consider a first example of the entropy minimization problem under a risk measure constraint: we discuss a shortfall risk constraint. Shortfall risk has many appealing properties. It is distribution-invariant, coherent and sensitive to the size of losses. In contrast to average value at risk, shortfall risk can be used for the weakly consistent dynamic evaluation of financial positions, cf. Weber (2006).

The minimization problem Let $\ell : \mathbb{R} \to \mathbb{R}$ be a convex loss function, $z \in \mathbb{R}$ be a point in the interior of the range of ℓ, and ρ be shortfall risk associated with ℓ and z. Fix a reference probability $\mu \in \mathcal{M}_{1,c}$ with compact support and a constant $y \in \mathbb{R}$. We are interested in the problem of minimizing $H(\nu|\mu)$ where $\nu \in \mathcal{M}_{1,c}$ and $\rho'(\nu) = y$. We set $a := \inf\{x \in \mathbb{R} : x \in \text{supp } \mu\}$, and $b := \sup\{x \in \mathbb{R} : x \in \text{supp } \mu\}$. Then supp $\mu \subseteq [a, b]$. Observe that supp $\nu \not\subseteq$ supp μ implies $\nu \not\ll \mu$, thus $H(\nu|\mu) = \infty$. Thus, we may restrict our attention to the constraint set

$$\mathcal{C} := \{\nu \in \mathcal{M}_1([a, b]) : \rho'(\nu) = y\}.$$

4.1 Existence of solutions

A necessary and sufficient criterion for the existence of solutions can be formulated in terms of the parameters a, b and y. We need the following general result that will also be used in Section 5. Since shortfall risk is continuous on compacts, the entropy minimization problem under a shortfall risk-constraint represents a special case of the next lemma.

Let $\mu \in \mathcal{M}_{1,c}$, and $a, b, y \in \mathbb{R}$ be given as above. For any distribution-invariant risk measure $\rho : L^\infty \to \mathbb{R}$ which is continuous on compacts we define

$$\mathcal{C}_\rho := \{\nu \in \mathcal{M}_1([a, b]) : \rho'(\nu) = y\}.$$

We consider the minimization problem of $H(\cdot|\mu)$ on $\mathcal{C}_\rho$.

Lemma 4.1. *Suppose that $\mathcal{C}_\rho \neq \emptyset$. There exists a solution to the entropy minimization problem with constraint set $\mathcal{C}_\rho$. If there exists a $\nu \in \mathcal{C}_\rho$ such that $H(\nu|\mu) < \infty$, then the minimizer has finite relative entropy.*

Proof. If $H(\nu|\mu) = \infty$ for all $\nu \in \mathcal{C}_\rho$, then any $\nu \in \mathcal{C}_\rho$ minimizes the relative entropy. Otherwise observe that $\mathcal{M}_1([a,b])$ is weakly compact, since $[a,b]$ is compact. Since ρ is continuous on compacts, $\mathcal{C}_\rho$ is a weakly compact set. Since $H(\cdot|\mu)$ is lower semicontinuous, it achieves its minimum on $\mathcal{C}_\rho$. □

In the case of shortfall risk, the following characterization theorem is essential. A proof is contained in the proof of Proposition 4.104 in Föllmer & Schied (2004).

Proposition 4.2. *Let $X \in L^\infty$. Then $\rho(X) = y$, if and only if $\int \ell(-X-y)dP = z$.*

The following proposition characterizes the existence of solutions.

Proposition 4.3. *The following conditions are equivalent.*

(1) There exists $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$.

(2) The minimal value of the relative entropy on $\mathcal{C}$ is finite and attained for some element of $\mathcal{C}$.

(3) a is an atom of μ and $\ell(-a-y) = z$, or b is an atom of μ and $\ell(-b-y) = z$, or

$$\ell(-b-y) < z < \ell(-a-y). \tag{6}$$

If (6) holds, then there exists $\nu \approx \mu$, $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$.

Proof. See appendix. □

4.2 Structure of the solution

Since shortfall risk imposes a linear constraint, the solution to the minimization problem can be characterized. Its density with respect to μ is of exponential form. The exponent is a linear combination of the constraint functions. We quote a theorem of Csiszar (1975).

Theorem 4.4. *For $i = 1, 2, \ldots, I$ let $f_i : \mathbb{R} \to \mathbb{R}$ be measurable functions and $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$, and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \;\; i = 1, 2, \ldots, I \right\}.$$

Assume there exists $\nu \in \hat{\mathcal{C}}$ with $\nu \approx \mu$ and $H(\nu|\mu) < \infty$. Then there exists a unique minimizer on $\hat{\mathcal{C}}$ with finite relative entropy. ν is the minimizer, if and only if its μ-density is of the following form

$$\frac{d\nu}{d\mu} = c \cdot \exp\left(\sum_{i=1}^{I} h_i f_i \right),$$

with normalizing constant $c > 0$ and $h_i \in \mathbb{R}$ $(i = 1, 2, \ldots, I)$.

Corollary 4.5. *Assume that (6) holds. Then ν is the unique minimizer of the relative entropy on $\mathcal{C}$, if and only if its μ-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left(h \cdot \ell(-x-y)\right). \tag{7}$$

Here, $c > 0$ is a normalizing constant, $h \in \mathbb{R}$, and the following conditions need to be satisfied:

$$z = c \cdot \int \ell(-x-y) \exp\left(h \cdot \ell(-x-y)\right) \mu(dx), \tag{8}$$

$$1 = c \cdot \int \exp\left(h \cdot \ell(-x-y)\right) \mu(dx). \tag{9}$$

Proof. By Proposition 4.3 there exists $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$ and $\nu \approx \mu$. It follows from Theorem 4.4 that there exists a unique minimizer of the relative entropy. By Proposition 4.2, a measure $\nu \in \mathcal{M}_1([a,b])$ is an element of $\mathcal{C}$, if and only if $\int \ell(-x-y)\nu(dx) = z$. Thus, the minimizer ν has μ-density (7) by Theorem 4.4. (8) is required by the constraint, (9) is a normalization. □

If ν is the minimizing density characterized in Corollary 4.5, then the minimal relative entropy is given by the expression

$$H(\nu|\mu) = \log c + h \cdot z.$$

Remark 4.6. *If a is an atom of μ and $\ell(-a-y) = z$, or if b is an atom of μ and $\ell(-b-y) = z$, then $\mathcal{C} = \{\delta_x\}$ with $x = a$ or $x = b$, respectively. Here, δ_x denotes the Dirac measure on $x \in \mathbb{R}$. Then the minimizer ν of the relative entropy is trivially unique, and $H(\nu|\mu) = -\log \mu\{x\}$ with $x = a$ or $x = b$, respectively.*

5 Entropy minimization under $AVaR$-constraints

In the current section, we will discuss the minimization problem for a second risk measure: *average value at risk*. $AVaR$ is a risk measures with appealing properties. It is distribution-invariant and coherent. In the event of a large loss, $AVaR$ takes its size into account. The last fact follows, for example, from the following representation of average value at risk at level λ:

$$AVaR_\lambda(X) = \frac{1}{\lambda} E((q-X)^+) - q,$$

where q is some λ-quantile of the random variable X.

The minimization problem Fix a reference probability measure $\mu \in \mathcal{M}_{1,c}$ with compact support, let $\lambda \in (0,1)$ be a level and $y \in \mathbb{R}$ a constant. We are interested in the problem of minimizing $H(\nu|\mu)$ where $\nu \in \mathcal{M}_{1,c}$ and $AVaR_\lambda(\nu) = y$.

We set $a := \inf\{x \in \mathbb{R} : \ x \in \text{supp } \mu\}$, and $b := \sup\{x \in \mathbb{R} : \ x \in \text{supp } \mu\}$. As in the case of shortfall risk, we may restrict our attention to the constraint set

$$\mathcal{C} := \{\nu \in \mathcal{M}_1([a,b]) : \ AVaR_\lambda(\nu) = y\}.$$

5.1 Existence of solutions

A necessary and sufficient criterion for the existence of solutions can be formulated in terms of the parameters a, b and y. The derivation is based on Lemma 4.1.

In the case of AVaR, the existence of a minimizer with finite relative entropy can be rephrased in terms of the parameters of the problem. For this purpose, it is useful to recall a particular representation of $AVaR_\lambda$, cf. Föllmer & Schied (2004).

Proposition 5.1. *Let* $\lambda \in (0,1)$*, and* $\nu \in \mathcal{M}_{1,c}$*. Then*

$$AVaR_\lambda(\nu) = -\int x f_\nu(x)\nu(dx), \tag{10}$$

where f_ν *is the following density of a probability measure with respect to* ν*:*

$$f_\nu(x) = \frac{1}{\lambda}\left(\mathbf{1}_{(-\infty,q)} + \kappa \mathbf{1}_{\{q\}}\right) \tag{11}$$

Here, q *is a* λ*-quantile of* ν*, i.e.*

$$\int \mathbf{1}_{(-\infty,q)} d\nu \quad \leq \quad \lambda \tag{12}$$

$$\int \mathbf{1}_{(-\infty,q]} d\nu \quad \geq \quad \lambda \tag{13}$$

The parameter κ *is defined as follows:*

$$\kappa = \begin{cases} 0 & \text{if } \nu\{q\} = 0 \\ \frac{\lambda - \nu(-\infty,q)}{\nu\{q\}} & \text{if } \nu\{q\} \neq 0 \end{cases} \tag{14}$$

The following proposition characterizes the existence of solutions.

Proposition 5.2. *The following conditions are equivalent:*

(1) There exists $\nu \in \mathcal{C}$ *such that* $H(\nu|\mu) < \infty$*.*

(2) The minimal value of the relative entropy on $\mathcal{C}$ *is finite and attained for some element of* $\mathcal{C}$*.*

(3) $a < -y < b$*, or* $-y$ *is an atom of* μ*.*

Proof. See appendix. □

5.2 Structure of the solutions

Classical results of Csiszar (1975) determine the general structure of the minimizer. We compute the solution explicitly. In order to avoid trivial cases, we will always assume that one and thus all of the equivalent conditions of Proposition 5.2 is satisfied. We distinguish two cases of different complexity: (A) μ does not have any atoms; (B) μ possibly has atoms.

5.2.1 A two-step procedure I

First we focus on case (A). We can restrict our attention to probability measures which are absolutely continuous with respect to μ. A minimizer ν will thus not have any atoms. The formulas characterizing the density f_ν in (10) simplify to $f_\nu = \frac{1}{\lambda}\mathbf{1}_{(-\infty,q)}$, $\int \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \lambda$.

The original problem can be reduced to a family of relative entropy minimization problems under linear constraints and a one-dimensional minimization problem.

Step 1 Fix some quantile level $q \in \mathbb{R}$. Minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures $\nu \ll \mu$ which satisfy the constraint

$$-\frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) \quad = \quad y, \tag{15}$$

$$\int \mathbf{1}_{(-\infty,q)}(x)\nu(dx) \quad = \quad \lambda. \tag{16}$$

We will provide conditions under which this problem has a solution. Then the solution is unique and can be represented by an exponential density.

Step 2 As we will see, if for $q \in \mathbb{R}$ the minimization problem in step 1 has a solution with finite relative entropy, the minimizer will be unique. We denote this minimizer by ν^q. Otherwise, we set $\nu^q = \dagger$ with the convention $H(\dagger|\mu) = \infty$. With this notation, the solution of the original problem is given by the set $\operatorname{argmin}_{\nu\in\mathcal{D}} H(\nu|\mu)$ with $\mathcal{D} = \{\nu^q : \ q \in \mathbb{R}\}$.

5.2.2 Entropy minimization under linear constraints I

We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ without atoms and $q \in \mathbb{R}$. In this section we consider the minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraints (15) and (16).

Proposition 5.3. *The following conditions are equivalent:*

(1) There exists a probability measure ν with $H(\nu|\mu) < \infty$ that satisfies the constraint.

(2) There exists a probability measure ν equivalent to μ with $H(\nu|\mu) < \infty$ that satisfies the constraint.

(3) Under the constraint there exists a unique minimizer of the relative entropy.

(4) $a < -y < q < b$, $\mu(-y, q) > 0$.

Proof. See appendix. □

If one and thus all of the equivalent conditions of Proposition 5.3 are satisfied, the unique minimizer can be characterized. The proof follows directly from Theorem 4.4.

Corollary 5.4. *Assume that one of the equivalent conditions of Proposition 5.3 holds. ν is the unique minimizer of the relative entropy under the constraints (15) and (16), if and only if its μ-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left((h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\right). \tag{17}$$

Here, $c > 0$ is a normalizing constant, $h_1, h_2 \in \mathbb{R}$, and the following conditions need to be satisfied:

$$-\lambda y \quad = \quad c \cdot \int x \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\mu(dx), \tag{18}$$

$$\lambda \quad = \quad c \cdot \int \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\mu(dx), \tag{19}$$

$$1 \quad = \quad c \cdot \int \exp\left\{(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x)\right\}\mu(dx). \tag{20}$$

If ν is the minimizing density characterized in Corollary 5.4, the minimal relative entropy is given by the expression $H(\nu|\mu) = \log c + \lambda(h_1 - h_2 y)$.

5.2.3 The original problem I

Assuming that one and thus all of the equivalent conditions of Proposition 5.3 are satisfied, we denote the unique solution of the minimization problem under the linear constraints (15) and (16) by ν^q. The proof of the following proposition is immediate.

Proposition 5.5. *There exists a solution to the minimization problem of the relative entropy on* $\mathcal{C}$, *if and only if* $a < -y < b$. *The collection of all solutions is given by* $\{\nu^q : \; q \in Q^*\}$ *where* $Q^* =$ *argmin* $\{H(\nu^q|\mu) : \; a < -y < q < b, \; \mu(-y, q) > 0\}$.

5.2.4 A two-step procedure II

We now consider case (B), i.e., μ may have atoms. As in case (A) the problem can be decomposed into two subproblems, but atoms make the problem more complicated.

Step 1 Fix some quantile level $q \in \mathbb{R}$. We distinguish two cases:

a) q is not an atom of the reference measure μ.

b) q is an atom of μ.

Case a) is only slightly more complicated than the situation which we considered in Proposition 5.3. Case b) involves two additional parameters. Let $d \in [0, \lambda]$ and $u \in [0, 1-\lambda]$. If $d = u = 0$, we set $d \cdot (d+u)^{-1} = 0$. In step 1 we need to minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures $\nu \ll \mu$ which satisfy the constraint

$$\lambda - d \quad = \quad \int \mathbf{1}_{(-\infty,q)}(x)\nu(dx), \tag{21}$$

$$\lambda + u \quad = \quad \int \mathbf{1}_{(-\infty,q]}(x)\nu(dx), \tag{22}$$

$$y \quad = \quad -\frac{1}{\lambda}\int x\left(\mathbf{1}_{(-\infty,q)}(x) + \frac{d}{d+u}\cdot \mathbf{1}_{\{q\}}(x)\right)\nu(dx). \tag{23}$$

We will provide conditions when the problems have a solution. Then the solution is unique and can again be represented by a density which is of exponential form outside the set where it vanishes. The solution will not always be equivalent to the reference measure μ.

Step 2 As we will show, if the minimization problems a) and b) in step 1 have a solution with finite relative entropy for fixed parameters q, u, d, then the minimizer will be unique. Let $q \in \mathbb{R}$, $d \in [0, \lambda]$ and $u \in [0, 1-\lambda]$. If q is not an atom of μ, we are in the situation of case a). If the minimizer exists and if $u = d = 0$, we denote it by $\nu^{q,d,u} = \nu^{q,0,0}$. If q is an atom, we consider case b). If the minimizer exists, we denote it by $\nu^{q,d,u}$. In all other cases, we set $\nu^{q,u,d} = \dagger$ with the convention $H(\dagger|\mu) = \infty$. With this notation, the solution of the original problem is given by the set of minimizers

$$\text{argmin}_{\nu\in\mathcal{D}}\, H(\nu|\mu), \quad \mathcal{D} = \{\nu^{q,d,u} : \; q \in \mathbb{R}, \; d \in [0,\lambda], \; u \in [0, 1-\lambda]\}. \tag{24}$$

5.2.5 Entropy minimization under linear constraints II

Case a) We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and $q \in \mathbb{R}$. We assume that q is *not* an atom of μ. Nevertheless, the measure μ may have atoms. In this section we consider the minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraints (15) and (16).

Proposition 5.6. *The following conditions are equivalent:*

(1) There exists a probability measure ν with $H(\nu|\mu) < \infty$ that satisfies the constraint.

(2) Under the constraint there exists a unique minimizer of the relative entropy.

(3) One of the following conditions holds:

(a) $a < -y < q < b$, $\mu(-y, q) > 0$.

(b) $a \leq -y < q < b$, $-y$ is an atom of μ.

Moreover, if condition (3)(a) holds, then there exists a probability measure ν equivalent to μ with $H(\nu|\mu) < \infty$ that satisfies the constraint.

Proof. See appendix. □

Definition 5.7. *For $i = 1, 2, \ldots, I$ let $f_i : \mathbb{R} \to \mathbb{R}$ be measurable functions and $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$, and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \quad i = 1, 2, \ldots, I \right\}.$$

By $\bar{\mathcal{C}}$ we denote the subset of elements $\nu \in \hat{\mathcal{C}}$ with $H(\nu|\mu) < \infty$. A measurable set $N \subseteq \mathbb{R}$ is called maximal common nullset *if*

(1) $\forall\, \nu \in \bar{\mathcal{C}}$: $\nu(N) = 0$,

(2) $\exists\, \nu \in \bar{\mathcal{C}}$: $\mathbf{1}_N + \frac{d\nu}{d\mu} > 0$ μ-almost surely.

Remark 5.8. *Maximal common nullsets always exist, if $\bar{\mathcal{C}} \neq \emptyset$. This can e.g. be demonstrated as an application of Zorn's lemma.*

Remark 5.9. *A maximal common nullset is indeed a maximal set in the following sense: Let $M \subseteq \mathbb{R}$ be a measurable set that satisfies condition (1) of Definition 5.7, and assume that $N \subseteq M$ is a maximal common nullset. Then M is also a maximal common nullset, and $\mu(M \setminus N) = 0$. For a proof see Weber (2004).*

Remark 5.10. *μ-equivalent elements of $\bar{\mathcal{C}}$ exist under the following equivalent conditions. For a proof see Weber (2004). (i) There exists $\nu \in \bar{\mathcal{C}}$ with $\nu \approx \mu$. (ii) Some maximal common nullset is a μ-nullset. (iii) Any maximal common nullset is a μ-nullset. (iv) Any μ-nullset is a maximal common nullset. (v) Maximal common nullset and μ-nullsets coincide. (vi) The empty set is a maximal common nullset.*

In the context of the minimization problem of the current section maximal common nullsets can be characterized in terms of the parameters of the problem. If condition (3)(a) of Proposition 5.6 holds, maximal common nullset are μ-nullsets. The next proposition investigates maximal common nullsets, if condition (3)(a) is *not* satisfied, but condition (3)(b) holds.

Proposition 5.11. *Assume that condition (3)(b) of Proposition 5.6 holds.*

(1) If $a = -y$ and $\mu(-y, q) = 0$, then any maximal common nullset is a μ-nullset. I.e., there exists a μ-equivalent probability measure ν with $H(\nu|\mu) < \infty$ that satisfies the constraint.

(2) If $a = -y$ and $\mu(-y, q) > 0$, then (a, q) is a maximal common nullset.

(3) If $a < -y$ and $\mu(-y, q) = 0$, then $[a, -y)$ is a maximal common nullset.

Proof. See appendix. □

The minimizers are characterized by the following theorem of Csiszar (1975).

Theorem 5.12. *For $i = 1, 2, \dots, I$ let $f_i : \mathbb{R} \to \mathbb{R}$ be measurable functions and $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$, and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \quad i = 1, 2, \dots, I \right\}.$$

Assume there exists $\nu \in \hat{\mathcal{C}}$ with $H(\nu|\mu) < \infty$. Let N be a maximal common nullset. Then there exists a unique minimizer on $\hat{\mathcal{C}}$ with finite relative entropy. ν is the minimizer, if and only if its μ-density is of the following form

$$\frac{d\nu}{d\mu} = c \cdot \exp\left(\sum_{i=1}^{I} h_i f_i \right) \cdot \mathbf{1}_{N^c},$$

with normalizing constant $c > 0$ and $h_i \in \mathbb{R}$ $(i = 1, 2, \dots, I)$.

Corollary 5.13. *Assume that one and thus all of the equivalent conditions of Proposition 5.6 hold. Let N be a maximal common nullset, cf. Propositions 5.6 & 5.11. ν is the unique minimizer of the relative entropy under the constraints (15) and (16), if and only if its μ-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\big((h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \big) \cdot \mathbf{1}_{N^c}(x). \tag{25}$$

Here, $c > 0$ is a normalizing constant, $h_1, h_2 \in \mathbb{R}$, and the following conditions need to be satisfied:

$$-\lambda y = c \cdot \int x \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \cdot \mathbf{1}_{N^c}(x)\mu(dx), \tag{26}$$

$$\lambda = c \cdot \int \exp(h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \cdot \mathbf{1}_{N^c}(x)\mu(dx), \tag{27}$$

$$1 = c \cdot \int \exp\big\{ (h_1 + h_2 x)\mathbf{1}_{(-\infty,q)}(x) \big\} \cdot \mathbf{1}_{N^c}(x)\mu(dx). \tag{28}$$

Proof. The proof follows directly from Theorem 5.12. (28) is a normalization, (26) and (27) are required by the constraint. □

If ν is the minimizing density characterized in Corollary 5.13, the minimal relative entropy is given by the expression $H(\nu|\mu) = \log c + \lambda(h_1 - h_2 y)$.

Case b) We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and parameters $q \in \mathbb{R}$, $d \in [0, \lambda]$ and $u \in [0, 1-\lambda]$. Now we assume that q *is* an atom of μ. In this section we consider the minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraint (21), (22) and (23).

Proposition 5.14. *The following conditions are equivalent:*

(1) There exists a probability measure ν with $H(\nu|\mu) < \infty$ that satisfies the constraint.

(2) Under the constraint there exists a unique minimizer of the relative entropy.

(3) $a \leq -y \leq q \leq b$, and one of the following conditions holds:

(a) $d = 0$, $-y < q$, and $-y$ is an atom of μ

(b) $d = 0$, $a < -y < q$, $\mu(-y, q) > 0$

(c) $d > 0$, $a < -y$ and with $\bar{a} := \sup\{x \in \operatorname{supp} \mu : \; x < q\}$

$$-y \quad > \quad \frac{\lambda - d}{\lambda} \cdot a \; + \; \frac{d}{\lambda} \cdot q$$

$$-y \quad < \quad \frac{\lambda - d}{\lambda} \cdot \bar{a} \; + \; \frac{d}{\lambda} \cdot q$$

(d) $d > 0$, and for some atom $r \in \mathbb{R}$ of μ,

$$-y \quad = \quad \frac{\lambda - d}{\lambda} \cdot r \; + \; \frac{d}{\lambda} \cdot q$$

Moreover, if conditions (3)(b) holds, or if condition (3)(c) holds, then there exists a probability measure ν equivalent to μ with $H(\nu|\mu) < \infty$ that satisfies the constraint.

Proof. See appendix. □

The next proposition investigates maximal common nullsets, if the conditions in part (3) of Proposition 5.14 are satisfied. This characterization together with Theorem 5.12 allows us to specify the solution of the minimization problem under the constraints (21), (22) and (23).

Proposition 5.15. *Assume that $a \leq -y \leq q \leq b$ holds.*

(1) Suppose that $d = 0$ and that $-y = a < q$ is an atom of μ. Then $(-y, q)$ is a maximal common nullset.

(2) Suppose $d = 0$ and $a < -y < q$. If $\mu(-y, q) = 0$ and $-y$ is an atom of μ, then $[a, -y)$ is a maximal common nullset. If $\mu(-y, q) > 0$, then the empty set is a maximal common nullset.

(3) If condition (3)(c) of Proposition 5.14 holds, the empty set is a maximal common nullset.

(4) Suppose that condition (3)(d) of Proposition 5.14 holds.

(a) If $\lambda = d$, then $(-\infty, q)$ is a maximal common nullset.

(b) If $\lambda \neq d$ and $r = a$, then (a, q) is a maximal common nullset.

(c) If $\lambda \neq d$, $r > a$ and $\mu(r, q) = 0$, then $[a, r)$ is a maximal common nullset.

Proof. See appendix. □

Remark 5.16. *Proposition 5.15 covers all cases which are considered in Proposition 5.14. The proof is left to the reader.*

As a corollary of Proposition 5.15 and Theorem 5.12 we finally obtain a characterization of the solution.

Corollary 5.17. *Assume that one of the equivalent conditions of Proposition 5.14 holds. Let N be a maximal common nullset, cf. Propositions 5.15. ν is the unique minimizer of the relative entropy under the constraints (21), (22) and (23), if and only if its μ-density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \cdot \exp\left\{(h_1 + h_2 + h_3 x)\mathbf{1}_{(-\infty,q)}(x) + \left(h_2 + h_3 \frac{d}{d+u} x\right) \mathbf{1}_{\{q\}}(x)\right\} \cdot \mathbf{1}_{N^c}(x). \tag{29}$$

Here, $c > 0$ is a normalizing constant, $h_1, h_2, h_3 \in \mathbb{R}$, and conditions (21), (22) and (23) need to be satisfied.

In particular, the minimizer ν is equivalent to μ, if and only if the empty set is a maximal common nullset.

Proof. The proof follows directly from Theorem 5.12. □

If ν is the minimizing density characterized in Corollary 5.17, the minimal relative entropy is given by the expression $H(\nu|\mu) = \log c + h_1(\lambda - d) + h_2(\lambda + u) - h_3 \lambda y$.

5.2.6 The original problem II: the general case

The solution of the entropy minimization problem on the constraint set $\mathcal{C}$ can be obtained by minimizing over the solutions under linear constraints. The proof is now immediate.

Proposition 5.18. *There exists a solution to the minimization problem of the relative entropy on $\mathcal{C}$, if and only if $a < -y < b$, or $-y$ is an atom of μ. The collection of all solutions is given by (24).*

A Proofs for Sections 4 and 5

Proof of Proposition 4.3. (1) and (2) are clearly equivalent by Lemma 4.1. Assume now that (1) holds. Suppose that neither a nor b are atoms of μ. Then neither a nor b are atoms of ν. Since $\nu \in \mathcal{C}$, we have supp $\nu \subseteq [a, b]$. If $\ell(-a - y) < z$, then $\ell(-x - y) < z$ for $x \in (a, b]$. If $\ell(-a - y) = z$, then $\ell(-x - y) < z$ for $x \in (a, b]$, since $\ell^{-1}\{z\}$ is a singleton and ℓ is increasing. Thus, if $\ell(-a - y) \le z$, then $\int \ell(-x-y)\nu(dx) < z$, since a is not an atom of ν. By Proposition 4.2 $y \ne \rho'(\nu)$, a contradiction. Thus, $\ell(-a - y) > z$. Analogously, one can show that $\ell(-b - y) < z$. If a or b are atoms, then additionally $\ell(-a - y) = z$ or $\ell(-b - y) = z$ is possible. This proves that (1) implies (3).

Conversely, assume that (3) holds. If a is an atom of μ and $\ell(-a - y) = z$, define $\nu \ll \mu$ by $\frac{d\nu}{d\mu} = \frac{1}{\mu\{a\}}\mathbf{1}_{\{a\}}$. Then, $H(\nu|\mu) < \infty$ and $\int \ell(-x - y)\nu(dx) = \ell(-a - y) = z$, thus $\rho'(\nu) = y$ by Proposition 4.2. Analogously, if b is an atom of μ and $\ell(-b - y) = z$, then $\frac{d\nu}{d\mu} = \frac{1}{\mu\{b\}}\mathbf{1}_{\{b\}}$ defines $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$.

Finally, assume that $\ell(-b - y) < z < \ell(-a - y)$. Since $\ell^{-1}\{z\}$ is a singleton, we obtain $\ell(-x - y) < z$ for $x > q$, $\ell(-x - y) > z$ for $x < q$, where $q := -\ell^{-1}(z) - y \in (a, b)$. Thus,

$$u_+ \;:=\; \int \mathbf{1}_{[a,q]}(x)\ell(-x - y)\mu(dx) \;>\; z, \qquad u_- \;:=\; \int \mathbf{1}_{(q,b]}(x)\ell(-x - y)\mu(dx) \;<\; z.$$

Choose $\alpha \in (0,1)$ such that $\alpha u_+ + (1-\alpha)u_- = z$. Then, we define $\nu \ll \mu$ by

$$\frac{d\nu}{d\mu} = \alpha \mathbf{1}_{[a,q]} + (1-\alpha)\mathbf{1}_{(q,b]}.$$

Clearly, $H(\nu|\mu) < \infty$ and $\nu \in \mathcal{C}$ by Proposition 4.2. □

Proof of Proposition 5.2. (1) and (2) are clearly equivalent by Lemma 4.1.

We will now show that (1) and (3) are equivalent. First suppose that $-y$ is an atom of μ. Then we set

$$\frac{d\nu}{d\mu} = \frac{1}{\mu\{-y\}} \cdot \mathbf{1}_{\{-y\}}.$$

In this case, $-y$ is a λ-quantile of ν, and $AVaR_\lambda(\nu) = y$.

Next suppose that $-y$ is not an atom of μ and that (1) holds. Then either $-y \le a$, $a < -y < b$, or $-y \ge b$. Let $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$. In particular, $\nu \ll \mu$. Since supp $\nu \subseteq [a,b]$, it follows that $-y = -AVaR_\lambda(\nu) \in [a,b]$. If $a = -y = -AVaR_\lambda(\nu)$, then a must be an atom of ν. Since $\nu \ll \mu$, a is then also an atom of μ, a contradiction. Analogously, it can be shown that $b \neq -y$. We obtain therefore $a < -y < b$.

Finally, we have to show that for $a < -y < b$ there exists always $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$. We consider two cases:

(a) There exists $q \in \mathbb{R}$ with $a < -y < q < b$ such that $\mu(-y,q) > 0$. Since μ has at most countably many atoms, we may and will assume that q is not an atom of μ.

(b) There exists no such $q \in \mathbb{R}$. This implies that $\mu(-y,b) = 0$. Then b must be an atom of μ, since $b = \sup\{x \in \mathbb{R} : \; x \in \text{supp } \mu\}$.

We consider first case (a). Since $a < -y < q$, there exists $\alpha' \in (0,1)$ such that

$$-y = \alpha' \frac{1}{\mu[a,-y)} \int x \mathbf{1}_{[a,-y)}(x)\mu(dx) + (1-\alpha')\frac{1}{\mu[-y,q)} \int x \mathbf{1}_{[-y,q)}(x)\mu(dx). \tag{30}$$

Define the weights

$$\alpha = \frac{\lambda\alpha'}{\mu[a,-y)}, \quad \beta = \frac{\lambda(1-\alpha')}{\mu[-y,q)}, \quad \gamma = \frac{1-\lambda}{\mu[q,b]}. \tag{31}$$

We define a probability measure by

$$\frac{d\nu}{d\mu} = \alpha \mathbf{1}_{[a,-y)} + \beta \mathbf{1}_{[-y,q)} + \gamma \mathbf{1}_{[q,b]}. \tag{32}$$

Then $H(\nu|\mu) < \infty$. We show that $\nu \in \mathcal{C}$. First, by calculation we obtain that $\nu(-\infty,q) = \nu(-\infty,q] = \lambda$. Thus, q is a λ-quantile of ν. Second,

$$\begin{aligned} AVaR_\lambda(\nu) &= -\frac{1}{\lambda}\int x \mathbf{1}_{[a,q)}(x)\nu(dx) \\ &= -\left(\frac{\alpha'}{\mu[a,-y)}\int x \mathbf{1}_{[a,-y)}(x)\mu(dx) + \frac{1-\alpha'}{\mu[-y,q)}\int x \mathbf{1}_{[-y,q)}(x)\mu(dx)\right) \\ &= y. \end{aligned}$$

Next we consider case (b). In this case we set

$$\bar{a} := \frac{1}{\mu(-\infty,-y]}\int x \mathbf{1}_{(-\infty,-y]}\mu(dx) \; < \; -y.$$

Let $\gamma := \lambda(b+y)/(b-\bar{a}) \in (0,\lambda)$. We define a probability measure ν via its density

$$\frac{d\nu}{d\mu} = \frac{\gamma}{\mu[a,-y]}\mathbf{1}_{[a,b)} + \frac{1-\gamma}{\mu\{b\}}\mathbf{1}_{\{b\}}.$$

Then $H(\nu|\mu) < \infty$. We now verify that $\nu \in \mathcal{C}$. Observe that

$$\int \mathbf{1}_{(-\infty,b)}(x)\nu(dx) = \frac{\gamma}{\mu[a,-y]}\int \mathbf{1}_{(-\infty,-y]}(x)\mu(dx) = \gamma < \lambda.$$

This implies that b is a λ-quantile for ν. We can now calculate $AVaR_\lambda$ using (11). Here, $\kappa = (\lambda-\gamma)/\nu\{b\}$. Thus,

$$\begin{aligned} AVaR_\lambda(\nu) &= \frac{-\gamma}{\lambda\cdot\mu[a,-y]}\int x\mathbf{1}_{(-\infty,b)}(x)\mu(dx) - \frac{(\lambda-\gamma)b}{\lambda} \\ &= \frac{-\bar{a}\gamma}{\lambda} - \frac{(\lambda-\gamma)b}{\lambda} = \frac{(b-\bar{a})\gamma}{\lambda} - b = y. \end{aligned}$$

□

Proof of Proposition 5.3. (3) trivially implies (1). In order to show that (1) implies (3) observe that the constraint set defined by (15) and (16) is variation-closed and convex. Theorem 2.1. of Csiszar (1975) implies that the minimization problem has a solution with finite relative entropy. The uniqueness of the minimizer follows, since the constraint set is convex and $H(\cdot|\mu)$ is strictly convex on its essential domain. Altogether, we have shown that (1) and (3) are equivalent.

Next, we show that (2) $\Rightarrow$ (1) $\Rightarrow$ (4) $\Rightarrow$ (2). The first implication is clear. Assume that (1) holds. We show that this implies (4). By assumption, $\nu \ll \mu$, and ν does not have any atoms. Since $\operatorname{supp}\nu \subseteq [a,b]$, we obtain that $q \in (a,b)$. Thus,

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu[a,q)}\int x\mathbf{1}_{[a,q)}(x)\nu(dx) \in [a,q). \tag{33}$$

$-y$ is not an atom of μ. Then (33) implies that $a < -y$. Suppose moreover $\mu(-y,q) = 0$, thus $\nu[-y,q) = 0$. Then,

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu(a,-y)}\int x\mathbf{1}_{(a,-y)}(x)\nu(dx) < -y,$$

a contradiction.

Finally, we show that (4) implies (2). Define the density of ν with respect to μ by (32) with coefficients given by (30) and (31). This defines a measure ν which is equivalent to μ and satisfies the constraints (15) and (16). □

Proof of Proposition 5.6. Proving that (1) and (2) are equivalent, is analogous to the proof of the equivalence of (1) and (3) in Proposition 5.3.

Next, we show that (1) and (3) are equivalent. Assume that (1) holds. By assumption, $\nu \ll \mu$, and ν does not have any atom at q. Since $\operatorname{supp}\nu \subseteq [a,b]$, we obtain that $q \in (a,b)$. Thus,

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu[a,q)}\int x\mathbf{1}_{[a,q)}(x)\nu(dx) \in [a,q). \tag{34}$$

Suppose that $-y$ is not an atom of μ. Then (34) implies that $a < -y$. Suppose moreover $\mu(-y,q) = 0$, thus $\nu[-y,q) = 0$. Then,

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu[a,-y)}\int x\mathbf{1}_{[a,-y)}(x)\nu(dx) < -y,$$

a contradiction.

Finally, we show that (3) implies (1). If $-y$ is an atom for μ, we set

$$\frac{d\nu}{d\mu} \;=\; \frac{\lambda}{\mu\{-y\}}\cdot \mathbf{1}_{\{-y\}} \;+\; \frac{1-\lambda}{\mu(q,b]}\cdot \mathbf{1}_{(q,b]}. \tag{35}$$

The measure ν satisfies the constraints (15) and (16) and has finite relative entropy. Nevertheless, it might not be equivalent to μ.

Otherwise, $a < -y$ and $\mu(-y,q) > 0$. Define the density of ν with respect to μ by (32) with coefficients given by (30) and (31). This defines a measure ν which is equivalent to μ and satisfies the constraints (15) and (16). □

Proof of Proposition 5.11. In case (1) equation (35) defines a density of a μ-equivalent probability measure ν with $H(\nu|\mu) < \infty$ that satisfies the constraint. In order to verify (2), set $N := (a,q)$. Let ν be a measure with $H(\nu|\mu) < \infty$ that satisfies the constraint. If $\nu(N) > 0$, then

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}\nu(dx) > -y, \tag{36}$$

a contradiction. Next, define a measure ν by density (35). As shown in the proof of Proposition 5.6, ν satisfies the constraints (15) and (16), and $H(\nu|\mu) < \infty$. Moreover, μ-almost surely,

$$\mathbf{1}_N \;+\; \frac{d\nu}{d\mu} \;>\; 0.$$

The proof of (3) is completely analogous to the proof of (2). We simply have to set $N := [a,-y)$ and to reverse the inequality in (36). □

Proof of Proposition 5.14. Proving that (1) and (2) are equivalent, is analogous to the proof of the equivalence of (1) and (3) in Proposition 5.3.

We now show that (1) implies (3). First consider the case $d = 0$. Then

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu(-\infty,q)}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) < q. \tag{37}$$

If $-y$ is an atom of μ, then (a) holds. Next, suppose that $-y$ is not an atom of μ, thus neither of ν. From (37) follows that $-y > a$. Suppose $\mu(-y,q) = 0$. Then $\nu(-\infty,-y) = \nu(-\infty,q) = \lambda$. Hence, $-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu(-\infty,-y)}\int x\mathbf{1}_{(-\infty,-y)}(x)\nu(dx) < -y$, a contradiction. This implies (b).

Next consider the case $d > 0$. Then q is an atom for ν. If $\lambda = d$, then (d) holds with $r = q$. Otherwise, $\nu(-\infty,q) > 0$ and $a < q$. We obtain from (23),

$$-y \;=\; \frac{\lambda-d}{\lambda}\;\underbrace{\frac{1}{\nu(-\infty,q)}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx)}_{=:r} \;+\; \frac{d}{\lambda}\cdot q.$$

Then $r \in [a,q)$ and $a < -y$. If r is not an atom of μ, it is not an atom of ν and $a < r < \bar{a}$. This implies (c). If r is an atom of μ, then (d) holds.

Finally, we have to show that (3) implies (1). In all cases we will specify a density with respect to μ such that the resulting measure ν satisfies $H(\nu|\mu) < \infty$ and the constraint. In case (a) choose

$$\frac{d\nu}{d\mu} = \frac{\lambda}{\mu\{-y\}}\cdot \mathbf{1}_{\{-y\}} + \frac{1-\lambda}{\nu[q,\infty)}\cdot \mathbf{1}_{[q,\infty)}. \tag{38}$$

Next, consider case (b). Then clearly $\frac{1}{\mu[-y,q)}\int x\mathbf{1}_{[-y,q)}(x)\mu(dx) > -y$. Hence, there exists $\alpha' \in (0,1)$ which satisfies (30). We choose a density according to (31) and (32). As in the proof of Proposition 5.2 simple calculations show that the constraints are satisfied. Observe that the measure ν specified by (32) is equivalent to μ.

Assume that (3)(c) is satisfied. Then there exists $r \in (a, \bar{a})$ such that $-y = \frac{\lambda-d}{\lambda}\cdot r + \frac{d}{\lambda}\cdot q$. By the definition of a and $\bar{a}$ it holds that $\mu[a,r) > 0$ and $\mu[r,q) > 0$. Moreover,

$$r \; > \; \frac{1}{\mu[a,r)}\int x\mathbf{1}_{[a,r)}(x)\mu(dx) \; =: \; g_-, \qquad r \; < \; \frac{1}{\mu[r,q)}\int x\mathbf{1}_{[r,q)}(x)\mu(dx) \; =: \; g_+.$$

Thus, there exists $\alpha \in (0,1)$ such that $r = \alpha g_- + (1-\alpha)g_+$. Define

$$\frac{d\nu}{d\mu} = (\lambda - d)\left(\frac{\alpha}{\mu[a,r)}\cdot\mathbf{1}_{[a,r)} + \frac{1-\alpha}{\mu[r,q)}\cdot\mathbf{1}_{[r,q)}\right) + \frac{d+u}{\mu\{q\}}\cdot\mathbf{1}_{\{q\}} + \frac{1-\lambda-u}{\mu(q,\infty)}\cdot\mathbf{1}_{(q,\infty)}.$$

Then ν satisfies the constraints (21), (22) and (23). Observe that ν is equivalent to μ.

Finally, if (3)(d) holds, then e.g. the following density defines an appropriate measure μ:

$$\frac{d\nu}{d\mu} \quad = \quad \frac{\lambda-d}{\mu\{r\}}\cdot\mathbf{1}_{\{r\}} \; + \; \frac{d+u}{\mu\{q\}}\cdot 1_{\{q\}} \; + \; \frac{1-\lambda-u}{\mu(q,\infty)}\cdot\mathbf{1}_{(q,\infty)}. \tag{39}$$

□

Proof of Proposition 5.15. Denote by $\bar{\mathcal{C}}$ the set of measures ν with $H(\nu|\mu) < \infty$ that satisfies the constraints (21), (22) and(23). It follows from Proposition 5.14 that $\bar{\mathcal{C}}$ is the never empty for the cases considered in the current proposition.
ad (1): Let $\nu \in \bar{\mathcal{C}}$. If $\nu(-y,q) > 0$, then (23) implies that

$$-y = \frac{\nu\{-y\}}{\nu[-y,q)}\cdot(-y) + \frac{\nu(-y,q)}{\nu[-y,q)}\cdot\frac{1}{\nu(-y,q)}\int x\mathbf{1}_{(-y,q)}(x)\nu(dx) > -y,$$

a contradiction. Thus, $\nu(-y,q) = 0$.

Next, define $\nu \in \bar{\mathcal{C}}$ by density (38). Then, μ-almost surely $\mathbf{1}_{(-y,q)} + \frac{d\nu}{d\mu} > 0$.
ad (2): We consider first the case $\mu(-y,q) = 0$. Let $\nu \in \bar{\mathcal{C}}$. Assume that $\nu[a,-y) > 0$. Then (23) implies that $-y = \frac{\nu\{-y\}}{\nu[a,-y]}\cdot(-y) + \frac{\nu[a,-y)}{\nu[a,-y]}\cdot\frac{1}{\nu[a,-y)}\int x\mathbf{1}_{[a,-y)}(x)\nu(dx) < -y$, a contradiction. Thus, $\nu[a,-y) = 0$.

If $\nu \in \bar{\mathcal{C}}$ is specified via density (38), then μ-almost surely $\mathbf{1}_{[a,-y)} + \frac{d\nu}{d\mu} > 0$.

Secondly, we consider the case $\mu(-y,q) > 0$. By Proposition 5.14 there exists a μ equivalent $\nu \in \bar{\mathcal{C}}$. The claim follows from Remark 5.10.
ad (3): This follows from Proposition 5.14 and Remark 5.10.
ad (4): If $\lambda = d$, then $\nu(-\infty,q) = 0$ for $\nu \in \bar{\mathcal{C}}$ by (21).

In case (b) assume for $\nu \in \bar{\mathcal{C}}$ that $\nu(a,q) > 0$. Then we obtain from (23),

$$-y = \frac{1}{\lambda}\int x\mathbf{1}_{(-\infty,q)}(x)\nu(dx) + \frac{d}{\lambda}\cdot q = \frac{\lambda-d}{\lambda}\cdot\frac{1}{\lambda-d}\int x\mathbf{1}_{[r,q)}(x)\nu(dx) + \frac{d}{\lambda}\cdot q > -y,$$

a contradiction. Thus, $\nu(a,q) = 0$.

In case (c) assume for $\nu \in \bar{\mathcal{C}}$ that $\nu[a,r) > 0$. Arguing analogously, we obtain that $\nu[a,r) = 0$.

Finally, set $N := (-\infty,q)$ in case (a), $N := (a,q)$ in case (b), and $N := [a,r)$ in case (c). Define $\nu \in \bar{\mathcal{C}}$ by density (39). Then, μ-almost surely $\mathbf{1}_N + \frac{d\nu}{d\mu} > 0$. □

References

Aliprantis, Charalambos D. & Kim C. Border (1999), *Infinite Dimensional Analysis : A Hitchhiker's Guide*, Springer, Berlin.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber & David Heath (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

Csiszar, Imre (1975), 'I-divergence geometry of probability distributions and minimization problems', *Annals of Probability* **3**(1).

Delbaen, Freddy (2002), Coherent risk measures on general probability spaces, *in* K. Sandmann & P. J. Schönbucher, eds, 'Advances in Finance and Stochastics', Springer-Verlag Berlin, pp. 1–38.

Dembo, Amir & Ofer Zeitouni (1998), *Large Deviations Techniques and Applications*, Springer, New York.

Dunkel, Jörn & Stefan Weber (2005), Efficient Monte Carlo methods for risk measures. Working Paper.

Föllmer, Hans & Alexander Schied (2002), Robust preferences and convex meaures of risk, *in* K. Sandmann & P. J. Schönbucher, eds, 'Advances in Finance and Stochastics', Springer-Verlag Berlin, pp. 39–56.

Föllmer, Hans & Alexander Schied (2004), *Stochastic Finance - An Introduction in Discrete Time (2nd edition)*, Walter de Gruyter, Berlin.

Frittelli, Marco & Gianin E. Rosazza (2002), 'Putting order in risk measures', *Journal of Banking and Finance* **26**(7), 1473–1486.

Fu, Michael C., Xing Jin & Xiaoping Xiong (2003), 'Probabilistic error bounds for simulation quantile estimators', *Management Science* **49**(2).

Giesecke, Kay, Thorsten Schmidt & Stefan Weber (2005), Measuring the risk of extreme events, *in* Marco Avellaneda, ed., 'Event Risk', Risk Books, London.

Jouini, Elvyès, Walter Schachermayer & Nizar Touzi (2006), Law invariant risk measures have the Fatou property. Working Paper.

Kallenberg, Olav (1997), *Foundations of Modern Probability*, Springer, New York.

Kusuoka, Shigeo (2001), 'On law invariant coherent risk measures', *Advances in Mathematical Economics* **3**, 83–95.

Schmeidler, David (1986), 'Integral representation without additivity', *Proceedings American Mathematical Society* **97**(2), 255–261.

Weber, Stefan (2004), Measures and models of financial risk. Ph.D. Thesis, Humboldt-Universität zu Berlin.

Weber, Stefan (2006), 'Distribution-invariant risk measures, information, and dynamic consistency', *Mathematical Finance* **16**(2), 419–442.