Maximum Entropy Principle with General Deviation Measures

Bogdan Grechuk
Stevens Institute of Technology, Hoboken, NJ
email: bgrechuk@stevens.edu

Anton Molyboha
Stevens Institute of Technology, Hoboken, NJ
email: amolyboh@stevens.edu

Michael Zabarankin
Stevens Institute of Technology, Hoboken, NJ
email: mzabaran@stevens.edu `http://personal.stevens.edu/~mzabaran`

An approach to the Shannon and Rényi entropy maximization problems with constraints on the mean and law invariant deviation measure for a random variable has been developed. The approach is based on the representation of law invariant deviation measures through corresponding convex compact sets of nonnegative concave functions. A solution to the problem has been shown to have an alpha-concave distribution (log-concave for Shannon entropy), for which in the case of comonotone deviation measures, an explicit formula has been obtained. As an illustration, the problem has been solved for several deviation measures, including mean absolute deviation (MAD), conditional value-at-risk (CVaR) deviation, and mixed CVaR-deviation. Also, it has been shown that the maximum entropy principle establishes a one-to-one correspondence between the class of alpha-concave distributions and the class of comonotone deviation measures. This fact has been used to solve the inverse problem of finding a corresponding comonotone deviation measure for a given alpha-concave distribution.

Key words: Deviation measures; Maximum entropy principle; Shannon entropy; Rényi entropy

MSC2000 Subject Classification: Primary: 94A17; Secondary: 46N10, 91B44

OR/MS subject classification: Primary: Probability: Entropy; Secondary: Desicion Analysis: Risk

---

1. Introduction. The principle of entropy maximization is widely used in a variety of applications ranging from statistical thermodynamics and quantum mechanics to decision theory and financial engineering. The principle was first introduced by Jaynes [12, 13] and is intended for finding the least-informative probability distribution[1] given any available information about the distribution. A classical example of an application for this principle is to find the maximally noncommittal probability distribution of a random variable (r.v.) given its first two moments, or, equivalently, mean and standard deviation. It is well known that if the Shannon differential entropy is chosen as a measure of uncertainty and is maximized subject to these constraints, then the distribution in question is the normal one, with the mean and standard deviation given by the corresponding constraints.

In application to decision and finance theories, the principle is extensively used for stock and option pricing through estimating corresponding probability distributions as well as for investigating agents' risk preferences. For example, Cozzolino and Zahner [7] derived the maximum-entropy distribution for the future market price of a stock under the assumption that the expectation and variance[2] of the price are known, while Thomas [29] considered the maximum-entropy principle in application to decision making under uncertainty for the oil spill abatement planning problem with discrete distributions and linear constraints. Also, the principle with the Rényi entropy, which is a generalization of the Shannon entropy, was applied to option pricing [5] and was investigated under constraints on covariance [14]. For option pricing with the maximum-entropy principle, see also Stutzer [28] and Buchen and Kelly [4]. For the application of generalized relative entropy to statistical learning of risk preferences, the reader may refer to [11].

A recently emerged theory of general deviation measures, developed by Rockafellar et al. [20, 21], generalizes the notion of standard deviation and provides an alternative way to measure "nonconstancy" in an r.v. In general, these measures are no longer symmetric, i.e., in contrast to standard deviation, they do not penalize the ups and downs of an r.v. equally, which is a desirable property in applications such as portfolio optimization, actuarial science, etc. Examples of deviation measures include standard devi-

[1] Jaynes [13] calls it least biased or maximally noncommittal probability distribution with regard to missing information.
[2] Variance is derived under the assumption that the price is a stochastic process with stationary and independent increments.

ation, *lower and upper semideviations*, *mean absolute deviation*, *median absolute deviation*, *Conditional Value-at-Risk (CVaR) deviation*, *mixed CVaR-deviation*, and *worst-case mixed-CVaR deviation*; see [21] for other examples. The aforementioned measures are law invariant, i.e., depend only on distributions of r.v.s. Rockafellar et al. [21] investigated the relationship between deviation and risk measures and showed that deviation measures provide significant customization in expressing agent's risk preferences. In application to portfolio optimization, Rockafellar et al. generalized a number of results originally stated for standard deviation, including Markowitz's portfolio selection problem [22, 23], in which standard deviation was replaced by an arbitrary deviation measure, the One Fund theorem [22, 23], the Capital Asset Pricing Model (CAPM) [22] and conditions on market equilibrium with investors using a variety of deviation measures [24]. Also, the role of deviation measures in a generalized linear regression was investigated in [25]. These developments raise a question: If a particular deviation measure reflects agent's risk preferences, what is the least informative distribution for which only the mean and corresponding deviation are available? Answering this question is the main motivation for this paper.

In this work, we consider the class of law invariant deviation measures and show that they can be represented in a quantile-based form on an atomless probability space over a corresponding set of nonnegative concave functions. We show that for an arbitrary law invariant deviation measure, this set is convex and compact, and in the case of comonotone deviation measures consists of a single element. Using these results, we determine which class of r.v.s maximizes the Shannon differential entropy with given constraints on the mean and a law invariant deviation measure and show that for every law invariant deviation measure the solution to the corresponding entropy maximization problem has a log-concave distribution. Moreover, for the case of comonotone deviation measure, an explicit formula for the optimal distribution is obtained.

Having the whole class of deviation measures at our disposal, we also address the inverse problem: Given a distribution for an r.v., determine a deviation measure that corresponds to this distribution through the entropy maximization principle. We show that the corresponding deviation measure can be constructed if and only if the given distribution is log-concave. For the case of log-concave distribution, we prove that several law invariant deviation measures may correspond to the same distribution through the maximum-entropy principle, however, only one of them is comonotone. In addition, an explicit formula for generating this comonotone deviation measure for a given log-concave distribution is provided. Solving the inverse problem paves the way for "restoring" agent's risk preferences encoded in a corresponding deviation through estimating appropriate probability distributions from historical data.

All the obtained results are extended for the entropy maximization problem with the Rényi entropy. Although the Shannon entropy remains a preferred choice as a measure of information, the Rényi entropy and particularly Rényi quadratic entropy have recently gained in popularity due to their computational efficiency in applications. For the last years, there have been extensive studies of Rényi entropy maximization with constraints on variance, covariance, and p-moment [6, 14, 17]. Distributions that maximize the Rényi entropy subject to constraints on a deviation measure are not necessarily log-concave and as a result, are more appropriate for estimating future returns of financial assets with heavy tails, e.g., returns of stokes, indices, etc.

The paper is organized into five sections. Section 2 reviews the main properties of deviation measures and represents law invariant deviation measures in a quantile-based form over a corresponding set of nonnegative concave functions. Section 3 proves that the set is compact and establishes several corollaries from this fact. Section 4 considers Shannon and Rényi differential entropy maximization problems subject to constraints on the mean and a law invariant deviation measure and solves the direct and inverse problems. Section 5 concludes this work.

2. Law invariant deviation measures.

This section reviews main properties of deviation measures and derives several representations for law invariant deviation measures.

Let $(\Omega, \mathcal{M}, P)$ be a probability space, where Ω denotes the designated space of future states ω, $\mathcal{M}$ is a field of sets in Ω, and P is a probability measure on $(\Omega, \mathcal{M})$. A random variable (r.v.) is any measurable function from Ω to $\mathbb{R}$. In this paper, we restrict our attention to r.v.s from spaces $\mathcal{L}^p(\Omega) = \mathcal{L}^p(\Omega, \mathcal{M}, P)$, $p = [1, \infty]$, with norms $||X||_p = (E[|X|^p])^{1/p}$, $p < \infty$, and $||X||_\infty = \operatorname{ess\,sup} |X|$. We also introduce the space $\mathcal{L}^F(\Omega) \subset \mathcal{L}^\infty(\Omega)$ of r.v.s that assume only a finite number of values. For an r.v. X, we denote $F_X(x)$, $f_X(x)$ and $q_X(\alpha) = \inf\{x | F_X(x) > \alpha\}$ its cumulative distribution function, probability density function (PDF), and quantile function, respectively. Throughout the paper, we assume that the probability space Ω is *atomless*, i.e., there exists an r.v. with continuous cumulative distribution function. This assumption implies existence of r.v.s on Ω with all possible[3] distribution functions (see, e.g., [10]).

General deviation measures, introduced by Rockafellar et al. in [20, 21], are defined as follows.

DEFINITION 2.1 *(deviation measures). A* deviation measure[4] *is any functional* $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$ *satisfying*

(D1) $\mathcal{D}(X) = 0$ *for constant* X*, but* $\mathcal{D}(X) > 0$ *otherwise (*nonnegativity*),*

(D2) $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ *for all* X *and all* $\lambda > 0$ *(*positive homogeneity*),*

(D3) $\mathcal{D}(X + Y) \le \mathcal{D}(X) + \mathcal{D}(Y)$ *for all* X *and* Y *(*subadditivity*),*

(D4) set $\{X \in \mathcal{L}^p(\Omega) \big| \mathcal{D}(X) \le c\}$ *is closed for all* $c < \infty$ *(*lower semicontinuity*).*

Axioms D2 and D3 imply convexity, and axioms D1–D3 have the consequence, shown in [21], that

$$\mathcal{D}(X + C) = \mathcal{D}(X) \text{ for all constants } C \text{ (\textit{insensitivity to constant shift})}. \tag{1}$$

A deviation measure is called *symmetric* if axiom D2 extends also to $\lambda < 0$ as $\mathcal{D}(\lambda X) = |\lambda| \mathcal{D}(X), \lambda \in \mathbb{R}$. A deviation measure $\mathcal{D}$ is called *law invariant* if $\mathcal{D}(X_1) = \mathcal{D}(X_2)$ for any two r.v.s X_1 and X_2 yielding the same distribution function on $\mathbb{R}$, and $\mathcal{D}$ is called *proper* if $\mathcal{D}(X) < \infty$ for some nonconstant X.

The most well known examples of deviation measures are:

(i) standard deviation $\sigma(X) = \sqrt{E[X - EX]^2}$;

(ii) *lower* and *upper semideviations* $\sigma_-(X) = \sqrt{E[X - EX]_-^2}$ and $\sigma_+(X) = \sqrt{E[X - EX]_+^2}$, respectively, where $[X]_- = \max\{0, -X\}$ and $[X]_+ = \max\{0, X\}$.

(iii) *mean absolute deviation* $\mathrm{MAD}(X) = E|X - EX|$;

(iv) *conditional value-at-risk (CVaR) deviation*, defined for any $\alpha \in (0, 1)$ by

$$\mathrm{CVaR}_\alpha^\Delta(X) \equiv EX - \frac{1}{\alpha} \int_0^\alpha q_X(\beta) d\beta. \tag{2}$$

All these deviation measures are law invariant. Whereas standard deviation, semideviations, and mean absolute deviation are well known measures of deviation, CVaR-deviation was introduced by Rockafellar et al. [21] as a "deviation analog" of conditional value-at-risk widely used in financial applications as a coherent measure of risk [1]. For the detailed discussion and other examples, see [21].

An r.v. X is said to dominate Y with respect to concave ordering, or $X \succcurlyeq_c Y$, if $EX = EY$ and $\int_{-\infty}^x F_X(t)dt \le \int_{-\infty}^x F_Y(t)dt$ for all $x \in \mathbb{R}$; see [9]. The result of Dana [9, Theorem 4.1.] implies that on an atomless probability space a deviation measure is law invariant if and only if it is consistent with concave ordering, i.e., $X \succcurlyeq_c Y$ implies $\mathcal{D}(X) \le \mathcal{D}(Y)$.

The conjugate function of a deviation measure $\mathcal{D}$ is associated with *risk envelope*, see [20, 21].

DEFINITION 2.2 *(risk envelope). A* risk envelope *is a subset* $\mathcal{Q}$ *of* $\mathcal{L}^q(\Omega)$*, where* $1/p + 1/q = 1$*, which satisfies the following axioms:*

(Q1) $\mathcal{Q}$ *is a convex, closed set containing 1 (constant r.v.),*

(Q2) $EQ = 1$ *for every* $Q \in \mathcal{Q}$*,*

(Q3) for every nonconstant $X \in \mathcal{L}^p(\Omega)$ *there is a* $Q \in \mathcal{Q}$ *such that* $E[XQ] < EX$*.*

---

[3]For any nondecreasing right-continuous function $F : \mathbb{R} \to [0, 1]$ with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, there exists an r.v. $X : \Omega \to \mathbb{R}$ such that its cumulative distribution function is $F(x)$.

[4]The axioms are those in [25]. In [20, 21], deviation measures are defined on $\mathcal{L}^2(\Omega)$, and originally axiom D4 was not included in the definition. Deviation measures satisfying D4 were called *lower-semicontinuous* deviation measures.

As shown in [20, 21], in the case of $p = q = 2$, there is a one-to-one correspondence between deviation measures and risk envelopes

$$\begin{aligned} \mathcal{Q} &= \left\{ Q \in \mathcal{L}^2(\Omega) \;\middle|\; E[X(1-Q)] \le \mathcal{D}(X) \;\; \forall X \right\}, \\ \mathcal{D}(X) &= \sup_{1-Q\in\mathcal{Q}} E[XQ]. \end{aligned} \tag{3}$$

In particular, standard deviation σ corresponds to $\mathcal{Q} = \{Q \big| \sigma(1-Q) \le 1,\, EQ = 1\}$, see [21].

In fact, the representation (3) holds true for all deviation measures for $p \in [1, \infty)$, and for deviation measures satisfying the Fatou property[5] for $p = \infty$ (see, e.g., [26]). Jouini et al. [15, Theorem 2.2] proved that on an atomless probability space every law invariant functional, satisfying D2–D4, has the Fatou property, which extends the representation (3) to all $p \in [1, \infty]$.

For law invariant deviation measures, the relationship (3) implies the following representation.

Proposition 2.1 *Every law invariant deviation measure $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$ can be represented in the form*

$$\mathcal{D}(X) = \sup_{1-Q\in\mathcal{Q}} \int_0^1 q_Q(\alpha)\, q_X(\alpha) d\alpha, \tag{4}$$

where $\mathcal{Q}$ is the corresponding risk envelope.

Proof. We write $X \sim Y$ if X and Y have the same distribution function. We have

$$\begin{aligned} \mathcal{D}(X) &= \sup_{Y:Y\sim X} \mathcal{D}(Y) = \sup_{Y:\, Y\sim X} \left[\sup_{Q:\, 1-Q\in\mathcal{Q}} E[QY] \right] \\ &= \sup_{Q:\, 1-Q\in\mathcal{Q}} \left[\sup_{Y:\, Y\sim X} E[QY] \right] = \sup_{1-Q\in\mathcal{Q}} \int_0^1 q_X(\alpha) q_Q(\alpha) d\alpha, \end{aligned}$$

where the first equality holds thanks to law invariance of $\mathcal{D}$, the second equality follows from (3), and the last one follows from [10, Lemma 4.55].[6] □

Next, we derive several representations for law invariant deviation measures, which will be used for solving the entropy maximization problem with constraints on the mean and a law invariant deviation measure.

Proposition 2.2 *For a functional $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$, the following are equivalent*

(a) $\mathcal{D}$ *is a law invariant deviation measure*

(b)

$$\mathcal{D}(X) = \sup_{\phi(\alpha)\in\Lambda} \int_0^1 \phi(\alpha)\, q_X(\alpha)\, d\alpha, \tag{5}$$

where Λ is a collection of nondecreasing functions $\phi(\alpha) \in \mathcal{L}^q(0, 1)$, $1/p + 1/q = 1$, such that $\int_0^1 \phi(\alpha)\, d\alpha = 0$, and containing at least one nonzero element;

(c)

$$\mathcal{D}(X) = \sup_{\phi(\alpha)\in\Lambda} \int_0^1 \mathrm{CVaR}_\alpha^\Delta(X)\, d(\psi(\alpha)), \qquad d(\psi(\alpha)) = \alpha d(\phi(\alpha)), \tag{6}$$

(d)

$$\mathcal{D}(X) = \sup_{g(\alpha)\in G} \int_0^1 g(\alpha)\, d(q_X(\alpha)), \tag{7}$$

where G is a collection of positive concave functions $g : (0, 1) \to \mathbb{R}$.

[5] Functional $\mathcal{D}$ satisfies the Fatou property, if $\mathcal{D}(X) \le \liminf_{n\to\infty} \mathcal{D}(X_n)$ for any bounded sequence X_n with $X_n \to X$ a.s.

[6] Lemma 4.55 in [10] is proved for the case $X \in \mathcal{L}^\infty(\Omega)$, $Q \in \mathcal{L}^1(\Omega)$. However, it can be readily extended to the general case $X \in \mathcal{L}^p(\Omega)$, $Q \in \mathcal{L}^q(\Omega)$ with $1/p + 1/q = 1$.

Proof. Proposition 2.1 implies that every law invariant deviation measure can be represented in the form (5) with $\Lambda = \left\{\phi(\alpha) \middle| \phi(\alpha) = q_Q(\alpha),\ 1-Q \in \mathcal{Q}\right\}$. Thus, (a) implies (b).

Now let $\mathcal{D}(X)$ be given by (6). For every nondecreasing $\phi(\alpha)$, we have $d(\psi(\alpha)) = \alpha d(\phi(\alpha)) \geq 0$. Thus, the properties D1–D4 for $\mathcal{D}$ and law invariance of $\mathcal{D}$ follow from the corresponding properties and law invariance of $\text{CVaR}_\alpha^\Delta$. Consequently, (c) implies (a).

The proofs of $(b) \to (c)$ and $(b) \leftrightarrow (d)$ reduce to integrating (5) by parts and are presented in Appendix A. □

Definition 2.3 *A collection G of nonnegative concave functions $g : (0, 1) \to \mathbb{R}$, for which* (7) *holds, will be called* g-envelope *of a law invariant deviation measure $\mathcal{D}$.*

It follows from the proof of Proposition 2.2 that if $\mathcal{D}$ can be represented in the form (5) with some Λ, it can be represented in the form (7) with g-envelope $G = \{g | g(\alpha) = -\int_0^\alpha \phi(u)du,\ \phi \in \Lambda\}$.

Remark 2.1 *The representation* (6) *is similar to the well known Kusuoka's representation [16] for coherent risk measures.*[7] *However, in contrast to the latter, it is insensitive to constant shift* (1).

Remark 2.2 *The representation* (7) *is equivalent to*

$$\mathcal{D}(X) = \sup_{g(\alpha)\in G} \int_{-\infty}^{\infty} g(F_X(x))dx, \tag{8}$$

under the assumption that $g(0) = g(1) = 0$, which extends the integration interval from $(\operatorname{ess\,inf} X, \operatorname{ess\,sup} X)$ *to $\mathbb{R}$. The reader can verify* (8) *by substituting $\alpha = F_X(x)$ into* (7).

Next, we prove that deviation measures having single-element g-envelope G are *comonotone* (comonotonically additive).

Definition 2.4 *Two r.v.s $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are said to be* comonotone *if there exists a set $A \subseteq \Omega$ such that $P[A] = 1$ and*

$$(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0 \qquad \forall\, \omega_1, \omega_2 \in A.$$

Definition 2.5 *A deviation measure $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$ is called* comonotone *if for any two comonotone r.v.s $X \in \mathcal{L}^p(\Omega)$ and $Y \in \mathcal{L}^p(\Omega)$, we have*

$$\mathcal{D}(X + Y) = \mathcal{D}(X) + \mathcal{D}(Y). \tag{9}$$

To derive a representation for comonotone law invariant deviation measures, we need to show that a law invariant deviation measure is uniquely defined by its restriction to the space of r.v.s that assume only a finite number of values.

Proposition 2.3 *Let $\mathcal{D}_1(X)$ and $\mathcal{D}_2(X)$ be two law invariant deviation measures such that $\mathcal{D}_1(X) = \mathcal{D}_2(X)$ for every $X \in \mathcal{L}^F(\Omega)$. Then $\mathcal{D}_1(X) = \mathcal{D}_2(X)$ for every $X \in \mathcal{L}^p(\Omega)$.*

Proof. We prove by contradiction. Suppose there exists an r.v. $X \in \mathcal{L}^p(\Omega)$ such that $\mathcal{D}_1(X) > \mathcal{D}_2(X)$. Fix $n \in \mathbb{N}$. For $i = -2^{2n}, \ldots, -1, 0, 1, \ldots, 2^{2n} - 1$ denote $\Omega_i = \{\omega \in \Omega | \frac{i}{2^n} \leq X(\omega) < \frac{i+1}{2^n}\}$, and let $\Omega_+ = \{\omega \in \Omega |\ X(\omega) \geq 2^n\}$, $\Omega_- = \{\omega \in \Omega |\ X(\omega) < -2^n\}$. Then $\mathcal{F}_n = \{\Omega_i | -2^{2n} \leq i \leq 2^{2n} - 1\} \cup \{\Omega_+\} \cup \{\Omega_-\}$ is a partition of Ω, and for $X_n = E[X|\mathcal{F}_n]$, we have $X_n \succcurlyeq_c X$, whence $\mathcal{D}_2(X_n) \leq \mathcal{D}_2(X)$. Because $X_n \in \mathcal{L}^F(\Omega)$, we have $\mathcal{D}_1(X_n) = \mathcal{D}_2(X_n) \leq \mathcal{D}_2(X)$ for every $n \in \mathbb{N}$.

Let $\Omega_* = \Omega_+ \cup \Omega_-$. Then $\Omega_* = \emptyset$ for $p = \infty$ and $n > \log_2 ||X||_\infty$, and thus, $\operatorname{ess\,sup} |X - X_n| \leq 2^{-n}$. For $p < \infty$, we have

$$\int_\Omega |X - X_n|^p d\omega = \int_{\Omega_*} |X - X_n|^p d\omega + \int_{\Omega/\Omega_*} |X - X_n|^p d\omega \leq \int_{\Omega_*} |2X|^p d\omega + 2^{-pn}.$$

[7] A coherent risk measure [1] is a functional $\mathcal{R} : \mathcal{L}^p(\Omega) \to [0, \infty]$ satisfying D2, D3 together with $\mathcal{R}(X + C) = \mathcal{R}(X) - C$ for all X and constant C, and $\mathcal{R}(X) \leq 0$ for all $X \geq 0$.

Consequently, $X_n \to X$ in $\mathcal{L}^p(\Omega)$ as $n \to \infty$, and lower semicontinuity of $\mathcal{D}_1(X)$ implies $\mathcal{D}_1(X) \le \mathcal{D}_2(X)$ which contradicts our initial assumption. By similar reasoning, we exclude the case $\mathcal{D}_1(X) < \mathcal{D}_2(X)$. □

Observe that, in general, in an infinite-dimensional space, two lower-semicontinuous convex positive homogeneous functionals, assuming same values on a dense subset, may not be equal (see [3]).

For comonotone law invariant deviation measures the following representation holds.

PROPOSITION 2.4 *A functional $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$ is a proper* comonotone *law invariant deviation measure if and only if it can be represented in the form*

$$\mathcal{D}(X) = \int_0^1 g(\alpha)\, d(q_X(\alpha)), \tag{10}$$

for some positive concave function $g : (0,1) \to \mathbb{R}$.

PROOF. By Proposition 2.2, the functional (10) is a law invariant deviation measure, which is also comonotone, because $q_{X+Y}(\alpha) = q_X(\alpha) + q_Y(\alpha)$ for comonotone r.v.s X and Y (see, e.g., [10, Lemma 4.84]).

To prove necessity, we show that every proper comonotone law invariant deviation measure $\mathcal{D}$ can be represented in the form (10) with $g(\alpha) = \mathcal{D}(X_\alpha)$, $\alpha \in (0,1)$, where X_α is a collection of comonotone r.v.s given by[8]

$$X_\alpha = \begin{cases} -1 & \text{with probability} \quad p = \alpha, \\ 0 & \text{with probability} \quad p = 1 - \alpha. \end{cases} \tag{11}$$

First, we prove that $g(\alpha)$ is a concave function on $(0,1)$. Indeed, for every $0 < \alpha_1 < \alpha_2 < \alpha_3 < 1$, and $\lambda \in (0,1)$ such that $\alpha_2 = \lambda\alpha_1 + (1-\lambda)\alpha_3$, we have $\lambda X_{\alpha_1} + (1-\lambda)X_{\alpha_3} \not\succeq_c X_{\alpha_2}$, whence $\mathcal{D}(X_{\alpha_2}) \ge \mathcal{D}(\lambda X_{\alpha_1} + (1-\lambda)X_{\alpha_3})$. Because $\mathcal{D}$ is comonotone, $\mathcal{D}(\lambda X_{\alpha_1} + (1-\lambda)X_{\alpha_3}) = \lambda\mathcal{D}(X_{\alpha_1}) + (1-\lambda)\mathcal{D}(X_{\alpha_3})$. This implies that $g(\alpha_2) \ge \lambda g(\alpha_1) + (1-\lambda)g(\alpha_3)$, and consequently, g is concave.

Thus, through (10), the function $g(\alpha)$ defines some comonotone law invariant deviation measure $\mathcal{D}'$. Next, we establish that $\mathcal{D}'(X) = \mathcal{D}(X)$ for every $X \in \mathcal{L}^p(\Omega)$. In fact, the equality holds for every X_α. Further, if an r.v. $X \in \mathcal{L}^F(\Omega)$ takes values $a_1 < a_2 < \ldots < a_n$ with probabilities $p_1, p_2, \ldots, p_n$, respectively, then X has the same distribution as $a_n + (a_n - a_{n-1})X_{q_{n-1}} + \ldots + (a_2 - a_1)X_{q_1}$, where $q_i = \sum_{j=1}^i p_j$, $i = 1, \ldots, n-1$. Because all r.v.s X_{q_i}, $i = 1, \ldots, n-1$ are comonotone, the comonotonicity and law invariance of $\mathcal{D}'$ and $\mathcal{D}$ imply that $\mathcal{D}'(X) = \mathcal{D}(X)$ for all $X \in \mathcal{L}^F(\Omega)$. It is left to apply Proposition 2.3. □

EXAMPLE 2.1 CVaR-deviation (2) *is a comonotone deviation measure and can be represented in the form* (10) *with*

$$g(\beta) = \begin{cases} (1/\alpha - 1)\beta & \beta \le \alpha, \\ 1 - \beta & \beta \ge \alpha. \end{cases} \tag{12}$$

Detail. Rockafellar et al. [21] showed that $\mathrm{CVaR}_\alpha^\Delta$ is a law invariant deviation measure, and comonotonicity of $\mathrm{CVaR}_\alpha^\Delta$ follows from the comonotonicity of the quantile function. By virtue of Proposition 2.4, $\mathrm{CVaR}_\alpha^\Delta$ can be represented in the form (10) with $g(\beta) = \mathrm{CVaR}_\alpha^\Delta(X_\beta)$, where X_β is given by (11). □

REMARK 2.3 *The deviation measure $\mathcal{D}(X) = \operatorname{ess\,sup} X - \operatorname{ess\,inf} X$ is comonotone and can be represented in the form* (7)*, with G containing single element $g(\alpha) \equiv 1$. However, it cannot be represented in the form* (5) *with some single-element collection Λ.*

[8] To construct such a collection, one can choose some r.v. U, uniformly distributed on [0,1], and take $X_\alpha = -1$ whenever $U < \alpha$, and $X_\alpha = 0$ whenever $U \ge \alpha$.

3. Maximal g-envelope. The representation (7) is a surjective mapping of g-envelopes $G \subset \mathcal{G}$ onto the set of deviation measures,[9] where $\mathcal{G}$ is the set of all nonnegative concave functions $g : (0,1) \to [0,\infty)$. A one-to-one correspondence can be established between law invariant deviation measures and maximal g-envelopes, defined below.

Definition 3.1 *Let $\mathcal{D}$ be a law invariant deviation measure. Its g-envelope G_M is called* maximal g-envelope*, if there is no g-envelope G of $\mathcal{D}$ such that $G_M \subset G$.*

The next proposition characterizes the maximal g-envelope G_M of an arbitrary law invariant deviation measure.

Proposition 3.1 *Let $\mathcal{D}$ be a law invariant deviation measure. Then it has exactly one maximal g-envelope G_M, which is a convex set given by*

$$G_M = \left\{ g(\alpha) \in \mathcal{G} \,\middle|\, \int_0^1 g(\alpha)\, d(q_X(\alpha)) \le \mathcal{D}(X) \qquad \forall X \in \mathcal{L}^F(\Omega) \right\}. \tag{13}$$

For an arbitrary g-envelope $G \subset \mathcal{G}$ of $\mathcal{D}$, we have $G \subseteq G_M$.

Proof. We first establish that the set

$$G'_M = \left\{ g(\alpha) \in \mathcal{G} \,\middle|\, \int_0^1 g(\alpha)\, d(q_X(\alpha)) \le \mathcal{D}(X) \qquad \forall X \in \mathcal{L}^p(\Omega) \right\} \tag{14}$$

is a maximal g-envelope of $\mathcal{D}$.

Let G be an arbitrary g-envelope of the deviation measure $\mathcal{D}$. Then for every $g(\alpha) \in G$, we have

$$\int_0^1 g(\alpha)\, d(q_X(\alpha)) \le \mathcal{D}(X) \qquad \forall X \in \mathcal{L}^p(\Omega).$$

Consequently, $G \subseteq G'_M$. This implies

$$\mathcal{D}(X) = \sup_{g(\alpha)\in G} \int_0^1 g(\alpha)\, d(q_X(\alpha)) \le \sup_{g(\alpha)\in G'_M} \int_0^1 g(\alpha)\, d(q_X(\alpha)) \le \mathcal{D}(X),$$

and thus the equality holds, which means that G'_M is a g-envelope of $\mathcal{D}$. The convexity of G'_M follows from (14), and the maximality of G'_M is guaranteed by the inclusion $G \subseteq G'_M$ for an arbitrary g-envelope G of $\mathcal{D}$.

It is left to prove that $G'_M = G_M$. Obviously, $G'_M \subseteq G_M$. According to Proposition 2.2, the set G_M is a g-envelope for some law invariant deviation measure $\overline{\mathcal{D}}$. Then for every $X \in \mathcal{L}^F(\Omega)$, the inequality $\overline{\mathcal{D}}(X) \le \mathcal{D}(X)$ follows from (13), and $\overline{\mathcal{D}}(X) \ge \mathcal{D}(X)$ follows from the fact that $G'_M \subseteq G_M$. Thus, $\overline{\mathcal{D}}(X) = \mathcal{D}(X)$ for every $X \in \mathcal{L}^F(\Omega)$, and by Proposition 2.3, we have $\overline{\mathcal{D}}(X) = \mathcal{D}(X)$ for every $X \in \mathcal{L}^p(\Omega)$. Consequently, G_M is a g-envelope of $\mathcal{D}$, whence $G_M \subseteq G'_M$, and the proof is finished. □

The relation (13) along with (7) introduces a one-to-one correspondence between law invariant deviation measures and their maximal g-envelopes.

Example 3.1 Mean absolute deviation $\mathrm{MAD}(X) = E|X - EX|$ *can be represented in the form* (7) *with the maximal g-envelope given by*

$$G_M(\mathrm{MAD}) = \left\{ g(\alpha) \in \mathcal{G} \,\middle|\, g(0+) = g(1-) = 0,\ g'(0+) - g'(1-) \le 2 \right\}. \tag{15}$$

Detail. As shown in [21], MAD is a law invariant deviation measure. Its maximal g-envelope $G_M(\mathrm{MAD})$ is given by (13). We need to prove that $G_M(\mathrm{MAD}) = A$, where $A \subset \mathcal{G}$ is the right-hand side in (15). For a sequence of r.v.s $X_n, n \ge 2$, given by

$$X_n = \begin{cases} -n & \text{with probability} \quad p = 1/n, \\ 0 & \text{with probability} \quad p = 1 - 2/n, \\ n & \text{with probability} \quad p = 1/n, \end{cases}$$

[9] Different sets G in (7) can produce the same deviation measure.

we have $\mathrm{MAD}(X_n) = 2$, and $\lim_{n\to\infty} \int_0^1 g(\alpha)\, d(q_{X_n}(\alpha)) = g'(0+) - g'(1-) + \lim_{n\to\infty} n(g(0+) + g(1-))$. Thus, (13) implies that $g(0+) = g(1-) = 0$ and $g'(0+) - g'(1-) \le 2$ for every $g \in G_M(\mathrm{MAD})$, i.e. $G_M(\mathrm{MAD}) \subseteq A$.

To prove the reverse inclusion, we show that

$$\int_0^1 g(\alpha)\, d(q_X(\alpha)) \le \mathrm{MAD}(X) \quad \forall X \in \mathcal{L}^F(\Omega), \quad \forall g(\alpha) \in A. \tag{16}$$

For every $g(\alpha) \in A$, concavity implies $g(\alpha) \le g'(0+)\alpha$ and $g(\alpha) \le -g'(1-)(1-\alpha)$. Thus, for every $y \in [-g'(1-), 2 - g'(0+)]$, $g(\alpha)$ is pointwise less than

$$g_y(\alpha) = \begin{cases} (2-y)\alpha & \alpha \le y/2, \\ y(1-\alpha) & \alpha \ge y/2. \end{cases}$$

Consequently, we need to prove (16) only for $g(.) = g_y(.)$, $y \in (0,2)$. Clearly, we can assume that $EX = 0$, and then integrating by parts, obtain $\int_0^1 g_y(\alpha)\, d(q_X(\alpha)) = -2\int_0^{y/2} q_X(\alpha) d\alpha \le E|X| = \mathrm{MAD}(X)$, whence $G_M(\mathrm{MAD}) \supseteq A$. □

The following proposition establishes some properties of G_M.

Proposition 3.2 *Let $\mathcal{D}$ be a proper law invariant deviation measure and let G_M be its maximal g-envelope. Then,*

(a) functions $g \in G_M$ are uniformly bounded.

(b) G_M is a compact set in the topology induced by the pointwise convergence.

Proof. We begin with proving (a). Because $\mathcal{D}$ is a proper deviation measure, there exists a nonconstant r.v. $X \in \mathcal{L}^p(\Omega)$ such that $\mathcal{D}(X) < \infty$. For nonconstant X, there exist a and b such that $0 < a < b < 1$ and $q_X(a) < q_X(b)$. Then, for every $g(\alpha) \in G_M$, we have

$$\mathcal{D}(X) \ge \int_0^1 g(\alpha)\, d(q_X(\alpha)) \ge \int_a^b g(\alpha)\, d(q_X(\alpha)) \ge \min_{\alpha\in[a,b]} g(\alpha)(q_X(b) - q_X(a)).$$

Consequently, $\min_{\alpha\in[a,b]} g(\alpha)$ is bounded from above by some constant M independent of g, and as a result, $g(\alpha_0) \le M$ for some $\alpha_0 \in [a,b]$. Concavity of g implies $g(\alpha) \le g(\alpha_0)\frac{\alpha}{\alpha_0} \le M\frac{\alpha}{\alpha_0}$ for $\alpha \ge \alpha_0$, and $g(\alpha) \le g(\alpha_0)\frac{1-\alpha}{1-\alpha_0} \le M\frac{1-\alpha}{1-\alpha_0}$ for $\alpha \le \alpha_0$. Finally, $g(\alpha) \le \max(M/\alpha_0, M/(1-\alpha_0)) \le \max(M/a, M/(1-b))$ for all α.

To prove (b), we first establish that G_M is a closed set with respect to pointwise convergence. Let a sequence $g_n(\alpha) \in G_M$ converge pointwise to some limit $g(\alpha)$. Then, obviously, the limit function $g(\alpha)$ is nonnegative and concave, and thus, $g(\alpha) \in \mathcal{G}$. The fact that for any $X \in \mathcal{L}^F(\Omega)$,

$$\int_0^1 g_n(\alpha)\, d(q_X(\alpha)) \to \int_0^1 g(\alpha)\, d(q_X(\alpha)) \quad \textit{as} \quad n \to \infty \tag{17}$$

follows from the dominated convergence theorem and statement (a). This, along with Proposition 3.1, implies that $g(\alpha) \in G_M$, and consequently, G_M is a closed set with respect to pointwise convergence.

Now compactness of G_M follows from Tychonoff's product theorem (see, e.g., [30, Theorem 17.8]), stating that the product of any collection of compact topological spaces is compact in the product topology (topology induced by pointwise convergence). Indeed, for any $C > 0$, the set of all functions from $[0,1]$ to $[0,C]$ is the product of a continuum of closed intervals $[0,C]$, which are compact, and therefore, the set is compact with respect to the product topology. By virtue of (a), G_M is a subset of this set for some C, and because it is closed, it is compact. □

The compactness of G_M is critical for establishing the existence of solution to optimization problems over G_M. In particular, we state the following result.

Proposition 3.3 *Let $\mathcal{D}$ be a proper law invariant deviation measure and let G_M be its maximal g-envelope. Then for every* bounded *r.v. X_0, there exists $g_0(\alpha) \in G_M$ (g-identifier of X_0) such that*

$$\mathcal{D}(X_0) = \int_0^1 g_0(\alpha)\, d(q_{X_0}(\alpha)).$$

Proof. Because G_M is compact, we only need to show that for every bounded r.v. X_0, the functional

$$S(g) = \int_0^1 g(\alpha)\, d(q_{X_0}(\alpha))$$

is continuous on G_M with respect to the pointwise convergence. Then we can apply Weierstrass's theorem (see, e.g., [2]) to conclude that the maximum in (7) is attained.

Let $\{g_n(\alpha)\} \subseteq G_M$ be a sequence converging pointwise to some limit $g^\star(\alpha)$. By Proposition 3.2(a), $g_n(\alpha)$ are uniformly bounded by some constant C. Because $\int_0^1 C\, d(q_{X_0}(\alpha)) < \infty$ for $X_0 \in \mathcal{L}^\infty(\Omega)$, we can apply the dominated convergence theorem, which states that $\lim\limits_{n\to\infty} S(g_n) = S(g^\star)$, and the proof is finished. □

We say that $g_1 \in \mathcal{G}$ dominates $g_2 \in \mathcal{G}$ and write $g_1 \succcurlyeq g_2$ if $g_1(\alpha) \geq g_2(\alpha)$ for all α. Set $G \subset \mathcal{G}$ is called dominance closed if $g_1 \in G$ implies $g_2 \in G$ whenever $g_1 \succcurlyeq g_2$. The following proposition provides another characterization for maximal g-envelope.

Proposition 3.4 *Set $G \subset \mathcal{G}$, containing at least one nonzero element, is a maximal g-envelope of some deviation measure if and only if it is convex, dominance closed, and closed with respect to pointwise convergence.*

Proof. Let $\mathcal{D}$ be a deviation measure with a g-envelope G. If G is the maximal g-envelope of $\mathcal{D}$, then (13) implies that it is convex and dominance closed, and its closedness with respect to pointwise convergence follows from Proposition 3.2. Let us prove the converse − i.e., if G is convex, closed, and dominance closed, then $G = G_M$, where G_M is the maximal g-envelope of $\mathcal{D}$.

Obviously, $G \subseteq G_M$, and we only need to prove that $g^* \in G$ for every $g^* \in G_M$. Let $k \in N$, $n = 2^k - 1$, $a_i = \frac{i}{2^k}$, $i = 1, \ldots, n$. Because G is convex and closed, the set $B = \{b = (b_1, \ldots, b_n) | \exists g \in G : b_i \leq g(a_i), i = 1, \ldots, n\}$ is also convex and closed in $\mathbb{R}^n$. Thus, if $(g^*(a_1), \ldots, g^*(a_n)) \notin B$, then by separation principle, $\sum_{i=1}^n \lambda_i g^*(a_i) \geq \epsilon + \sum_{i=1}^n \lambda_i b_i$ for $\epsilon > 0$, some $\lambda = (\lambda_1, \ldots, \lambda_n)$, and every $b \in B$. Because $(0, \ldots, 0, b_i = C, 0, \ldots, 0) \in B$ for every $i = 1, \ldots, n$ and $C < 0$, $\sum_{i=1}^n \lambda_i g^*(a_i) \geq \epsilon + \lambda_i C$, and consequently, $\lambda_i \geq 0$. Let X be an r.v. assuming values 0, λ_1, $\lambda_1 + \lambda_2$, …, $\sum_{i=1}^n \lambda_i$ with equal probabilities. Then $\int_0^1 g^*(\alpha) d(q_X(\alpha)) = \sum_{i=1}^n \lambda_i g^*(a_i) \geq \epsilon + \sum_{i=1}^n \lambda_i g(a_i) = \epsilon + \int_0^1 g(\alpha) d(q_X(\alpha))$ for every $g \in G$, which implies $\int_0^1 g^*(\alpha) d(q_X(\alpha)) \geq \epsilon + \mathcal{D}(X)$. However, this contradicts $g^* \in G_M$. Thus, $(g^*(a_1), \ldots, g^*(a_n)) \in B$.

Let $g_k(\alpha)$ be a piecewise-linear function with $n + 1 = 2^k$ pieces and with "vertexes" $g_n(a_i) = g^*(a_i)$ $i = 1, \ldots, n$, and $g_n(0+) = g_n(1-) = 0$. Because G is dominance closed, $(g^*(a_1), \ldots, g^*(a_n)) \in B$ implies $g_k(\alpha) \in G$. Then $\{g_k(\alpha)\}_{k \in \mathbb{N}}$ is a monotonically increasing sequence of nonnegative functions, and we have $\lim\limits_{k\to\infty} g_k(\alpha) = g^*(\alpha)$ pointwise. Thus, $g^*(\alpha) \in G$, whence $G = G_M$, and the proof is finished. □

As a corollary of Proposition 3.4, the relationships (13) and (7) introduce a one-to-one correspondence between law invariant deviation measures and convex, closed, dominance closed sets of nonnegative concave functions $g : (0, 1) \to [0, \infty)$.

The following example presents the maximal g-envelope of a comonotone deviation measure.

Example 3.2 *Let $\mathcal{D}$ be a proper comonotone law invariant deviation measure. Then its maximal g-envelope has the form $G_M = \{h \in \mathcal{G} | g \succcurlyeq h\}$, where the function g is given by* (10).

Detail. By Proposition 2.4, the set $\{h \in \mathcal{G} | g \succcurlyeq h\}$ is a g-envelope of $\mathcal{D}$. Because it is convex, closed, and dominance closed, by Proposition 3.4, it is a maximal g-envelope. □

4. Deviation measures and entropy maximization.

This section investigates Shannon and Rényi entropy maximization problems with constraints on the mean and a law invariant deviation measure.

4.1 Problem formulation. Let $\mathcal{X} \subset \mathcal{L}^1(\Omega)$ be the set of all r.v.s having continuous PDFs[10]. Then for an arbitrary $X \in \mathcal{X}$, the Shannon differential entropy $S(X)$ (see [27]) is defined by

$$S(X) = -\int_{-\infty}^{+\infty} f_X(x) \ln f_X(x) dx = -\int_{-\infty}^{+\infty} F'_X(x) \ln F'_X(x) dx, \tag{18}$$

where $f_X(x)$ is the PDF of X, whereas the Rényi differential entropy $H_\beta(X)$, being a generalization of the Shannon differential entropy, is introduced by[11]

$$H_\beta(X) = \frac{1}{1-\beta} \ln \int_{-\infty}^{+\infty} (f_X(x))^\beta dx, \qquad \beta > \frac{1}{2}, \quad \beta \neq 1, \tag{19}$$

see [19]. When $\beta \to 1$, $H_\beta(X)$ converges to $S(X)$. By convention, let

$$H_1(X) = S(X). \tag{20}$$

The entropy maximization problem with constraints on the mean and a proper law invariant deviation measure is formulated as

$$\max_{X \in \mathcal{X}} S(X) \qquad \text{s.t.} \quad EX = \mu, \quad \mathcal{D}(X) \le d, \tag{21}$$

$$\max_{X \in \mathcal{X}} H_\beta(X) \qquad \text{s.t.} \quad EX = \mu, \quad \mathcal{D}(X) \le d, \tag{22}$$

where $\beta > \frac{1}{2}$, and the constants μ and $d > 0$ are given. Because $H_\beta(kX) = H_\beta(X) + \ln k$, $k > 0$, the constraint $\mathcal{D}(X) \le d$ in (22) is always active.

In Shannon entropy maximization (21), Boltzmann's theorem [8, Theorem 11.1.1] plays a central role.

Proposition 4.1 (Boltzmann's theorem) *Let $V \subseteq \mathbb{R}$ be a closed subset and let $h_1, \ldots, h_n$ be measurable functions. Also, let $\mathcal{B}$ be the set of all continuous r.v.s X with the support V (i.e., those whose PDFs are zero outside of V) and satisfying the conditions*

$$E(h_j(X)) = a_j \qquad j = 1, \ldots, n, \tag{23}$$

where $a_1, \ldots, a_n$ are given. If there is an r.v. in $\mathcal{B}$ whose PDF is positive everywhere in V, and if there exists a Shannon maximum-entropy distribution in $\mathcal{B}$, then its PDF $f_X(x)$ is determined by

$$f_X(x) = c \exp\left(\sum_{j=1}^{n} \lambda_j h_j(x)\right) \qquad \forall x \in V, \tag{24}$$

where the constants c and λ_j are determined from (23) and the condition that the integral of $f_X(x)$ over V is 1.

If both constraints in (21) can be expressed in the form (23), then a solution to (21) is given by (24).

Example 4.1 (Shannon Entropy Maximization with Standard Deviation) *For standard deviation $\mathcal{D}(X) = \sigma(X) = \sqrt{E[X - EX]^2}$, the constraints $EX = \mu$ and $\sigma(X) = d$ in (21) can be represented in the form (23) with $V = (-\infty, \infty)$, $h_1(X) = X$, $a_1 = \mu$, $h_2(X) = (X - \mu)^2$, and $a_2 = d^2$. In this case, a solution to (21) is the normal distribution $\mathcal{N}(\mu, d^2)$.*

Example 4.2 (Shannon Entropy Maximization with Lower Semideviation) *For lower semideviation $\mathcal{D}(X) = \sigma_-(X) = \sqrt{E\big([X - EX]_-\big)^2}$, the constraints $EX = \mu$ and $\sigma_-(X) = d$ in (21) correspond to (23) with $V = (-\infty, \infty)$, $h_1(X) = X$, $a_1 = \mu$, $h_2(X) = [X - \mu]_-^2$, and $a_2 = d^2$. In this case, a solution to (21) is determined by*

$$f_X(x) = c \exp\left(\lambda_1 x + \lambda_2 [x - \mu]_-^2\right) \qquad \forall x \in \mathbb{R}. \tag{25}$$

In particular, if $\mu = 0$ and $d = 1$ then $c \approx 0.260713$, $\lambda_1 \approx -0.638833$ and $\lambda_2 = -0.5$.

[10] The entropy maximization problem is formulated on the space $\mathcal{L}^1(\Omega)$, which includes $\mathcal{L}^p(\Omega)$ for all $p \in [1, \infty]$.

[11] The Rényi differential entropy is defined for $\beta > 0$, $\beta \neq 1$. However, we restrict our attention to the case $\beta > 1/2$ to guarantee that the maximum-entropy distribution belongs to $\mathcal{L}^1(\Omega)$.

Detail. The formula (25) follows from Boltzmann's theorem, and the constants c, λ_1, and λ_2 are found from the conditions $\int_{-\infty}^{\infty} f_X(x)dx = 1$, $\int_{-\infty}^{\infty} x f_X(x)dx = \mu$, and $\int_{-\infty}^{\mu}(x-\mu)^2 f_X(x)dx = d^2$. Because the first and third integrals cannot be expressed in terms of elementary functions, the constants are found numerically. □

Example 4.3 (Shannon Entropy Maximization with Lower Range Deviation) *For lower range deviation $\mathcal{D}(X) = EX - \operatorname{ess\,inf} X$ (see [21]), the constraints $EX = \mu$ and $EX - \operatorname{ess\,inf} X = d$ in* (21) *are equivalent to* (23) *with $V = (\mu - d, \infty)$, $h_1(X) = X$, and $a_1 = \mu$. In this case, a solution to* (21) *is the shifted exponential distribution with $f_X(x) = \frac{1}{d}\exp(\frac{-x+\mu-d}{d})$, $x \geq \mu - d$.*

The solution of the Rényi entropy maximization problem (22) with standard deviation $\mathcal{D}(X) = \sigma(X)$ is also well known (see, e.g., [6, 14]).

Example 4.4 (Rényi Entropy Maximization with Standard Deviation) *For standard deviation $\mathcal{D}(X) = \sigma(X) = \sqrt{E[X - EX]^2}$, $\beta \neq 1$, and for $\mu = 0$, $d = 1$, a solution to* (22) *has the PDF*[12]

$$f_X(x) = A\left[1 - \tfrac{\beta-1}{3\beta-1}\,x^2\right]_+^{\frac{1}{\beta-1}} \qquad \forall x \in \mathbb{R}, \tag{26}$$

where $[x]_+ = \max\{x, 0\}$, and the constant A is defined by

$$A = \begin{cases} \left(\Gamma(\frac{1}{1-\beta})\sqrt{\frac{1-\beta}{3\beta-1}}\right) / \left(\Gamma(\frac{1}{1-\beta} - \frac{1}{2})\sqrt{\pi}\right) & \beta < 1, \\ \left(\Gamma(\frac{\beta}{\beta-1} + \frac{1}{2})\sqrt{\frac{\beta-1}{3\beta-1}}\right) / \left(\Gamma(\frac{\beta}{\beta-1})\sqrt{\pi}\right) & \beta > 1. \end{cases} \tag{27}$$

Detail. This result is a particular case of [14, Proposition 1.3] with $n = 1$. □

If the constraint $\mathcal{D}(X) = d$ can be expressed in the form (23), a distribution maximizing the Rényi entropy in (22) for $\beta \neq 1$ can be represented in the form similar to (24) in Boltzmann's theorem (see Appendix B).

Example 4.5 (Rényi Entropy Maximization with Lower Semideviation) *For lower semideviation $\mathcal{D}(X) = \sigma_-(X) = \sqrt{E[X - EX]_-^2}$, and $\beta \neq 1$, a solution to* (22) *has the PDF*

$$f_X(x) = c\left[1 + \lambda_1 x + \lambda_2 [x-\mu]_-^2\right]_+^{\frac{1}{\beta-1}} \qquad \forall x \in \mathbb{R}, \tag{28}$$

where c, λ_1 and λ_2 are found numerically from the conditions

$$\int_{-\infty}^{\infty} f_X(x)dx = 1, \qquad \int_{-\infty}^{\infty} x f_X(x)dx = \mu, \qquad \int_{-\infty}^{\mu} (x-\mu)^2 f_X(x)dx = d^2.$$

In particular, for $\mu = 0$, $d = 1$ these coefficients are shown in Figure 1.

Figure 1: Coefficients c, λ_1 and λ_2 in (28) as functions of β.

However, not for every deviation measure, the constraint $\mathcal{D}(X) = d$ in (22) can be represented in the form (23). A simple necessary condition can be formulated in terms of *mixtures*.

Definition 4.1 *Given r.v.s X and Y, and a number $\lambda \in [0,1]$, an r.v. Z with the cumulative distribution function $F_Z(z) = \lambda F_X(z) + (1-\lambda)F_Y(z)$ is called* λ-mixture *of X and Y. We write $Z = \lambda X \oplus (1-\lambda)Y$.*

Proposition 4.2 *Let $\mathcal{D}$ be a deviation measure such that the set $P_{\mathcal{D}} = \{X : EX = 0,\ \mathcal{D}(X) = 1\}$ can be expressed in the form* (23). *Then $\lambda X \oplus (1-\lambda)Y \in P_{\mathcal{D}}$ for $X, Y \in P_{\mathcal{D}}$ and any $\lambda \in [0,1]$.*

[12]In fact, (26) remains correct for more general case $\beta > \frac{1}{3}$. For $\beta \leq \frac{1}{3}$, the solution does not exist; see [6].

Proof. Let $X, Y \in P_{\mathcal{D}}$ and let $P_{\mathcal{D}}$ be represented in the form (23). Then $E(h_j(X)) = a_j$ and $E(h_j(Y)) = a_j$, $j = 1, \ldots, n$, and for the λ-mixture $Z = \lambda X \oplus (1-\lambda)Y$, we have

$$E(h_j(Z)) = \int_{-\infty}^{+\infty} h_j(z)\, d(\lambda F_X(z) + (1-\lambda)F_Y(z)) = \lambda \int_{-\infty}^{+\infty} h_j(z) d(F_X(z)) +$$
$$+(1-\lambda)\int_{-\infty}^{+\infty} h_j(z) d(F_Y(z)) = \lambda E(h_j(X)) + (1-\lambda)E(h_j(Y)) = a_j,$$

which implies $Z \in P_{\mathcal{D}}$. □

Example 4.6 *For* CVaR-deviation (2)*, the necessary condition, established in Proposition 4.2, does not hold, and consequently, the set* $\{X : EX = 0,\ \mathrm{CVaR}_\alpha^\Delta(X) = 1\}$ *cannot be expressed in the form* (23)*.*

Detail. Let two r.v.s X and Y be defined by $\Pr\{X = -3\} = 1/4$ and $\Pr\{X = 1\} = 3/4$, and $\Pr\{Y = -1\} = 3/4$ and $\Pr\{Y = 3\} = 1/4$, respectively. Then their 1/2-mixture Z is determined by $\Pr\{Z = -3\} = 1/8$, $\Pr\{Z = -1\} = 3/8$, $\Pr\{Z = 1\} = 3/8$, and $\Pr\{Z = 3\} = 1/8$. We have $EX = EY = 0$, $\mathrm{CVaR}_{1/2}^\Delta(X) = \mathrm{CVaR}_{1/2}^\Delta(Y) = 1$, but $\mathrm{CVaR}_{1/2}^\Delta(Z) = 3/2 \neq 1$. Counterexamples for arbitrary α can be constructed similarly. □

As in the case of CVaR-deviation, the necessary condition in Proposition 4.2 is not satisfied for the *mixed CVaR-deviation* (see [21]), which is a comonotone deviation measure, defined by

$$\mathcal{D}(X) = \int_0^1 \mathrm{CVaR}_\alpha^\Delta(X) d\mathfrak{m}(\alpha), \tag{29}$$

where $\mathfrak{m}$ is a weighting measure on $(0, 1)$ (nonnegative with total measure 1). Thus, in general, the problem (22) cannot be solved by direct application of Boltzmann's theorem, even in the case of Shannon entropy maximization ($\beta = 1$).

Lutwak et al. [17] used relative entropy to find Rényi maximum-entropy distribution for an r.v. with given p-th moment. The next section couples this approach with the representation (7) to characterize maximum-entropy distribution in (22) for an arbitrary law invariant deviation measure.

4.2 Characterization of Maximum-entropy Distribution. This section investigates the problem (22) for an arbitrary law invariant deviation measure $\mathcal{D}$.

The Rényi entropy (19) can be equivalently represented through the quantile function $q_X(\alpha)$. For $X \in \mathcal{X}$, the quantile $q_X(\alpha)$ is the inverse function of $F_X(x)$ and is differentiable almost everywhere. Substituting $\alpha = F_X(x)$ into (19), we have $d\alpha = f_X(x)dx$ and $x = q_X(\alpha)$. By the Inverse Function Theorem, $f_X(x) = \frac{1}{q_X'(F_X(x))}$, and the Rényi differential entropy takes the form

$$H_\beta(X) = \frac{1}{1-\beta} \ln \int_0^1 (q_X'(\alpha))^{1-\beta} d\alpha, \qquad \beta > \frac{1}{2}, \quad \beta \neq 1, \tag{30}$$

and similarly

$$H_1(X) = S(X) = \int_0^1 \ln q_X'(\alpha) d\alpha. \tag{31}$$

Because $H_\beta(X + C) = H_\beta(X)$ for any r.v. X and constant C, and $H_\beta(kX) = H_\beta(X) + \ln k$ for any X and $k > 0$, an r.v. X^* solves (22) if and only if $(X^* - \mu)/d$ solves the problem

$$\max_{X \in \mathcal{X}} H_\beta(X) \qquad \text{s.t.} \quad EX = 0, \quad \mathcal{D}(X) = 1. \tag{32}$$

We use an approach similar to that of Lutwak et al. [17] and introduce

$$N_\beta[X, Y] = \frac{\left(\int_0^1 q_X'(\alpha)^{-\beta} q_Y'(\alpha) d\alpha\right)^{\frac{1}{\beta}} \left(\int_0^1 q_X'(\alpha)^{1-\beta} d\alpha\right)^{\frac{1}{1-\beta}}}{\left(\int_0^1 q_Y'(\alpha)^{1-\beta} d\alpha\right)^{\frac{1}{\beta(1-\beta)}}}, \qquad \beta > \frac{1}{2}, \quad \beta \neq 1, \tag{33}$$

and

$$N_1[X,Y] = \left(\int_0^1 q'_X(\alpha)^{-1} q'_Y(\alpha) d\alpha\right) e^{H_1(X)-H_1(Y)} \tag{34}$$

provided that all the integrals exist. $\ln N_\beta[X,Y]$ is a version of relative β-Rényi entropy, which in contrast to the one of Lutwak [17], uses quantiles instead of probability density functions.

PROPOSITION 4.3 *$N_\beta[X,Y] \geq 1$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{X}$.*

PROOF. Let $\beta = 1$, then

$$\ln N_\beta[X,Y] = \ln \int_0^1 \frac{q'_Y(\alpha)}{q'_X(\alpha)} d\alpha - \int_0^1 \ln \frac{q'_Y(\alpha)}{q'_X(\alpha)} d\alpha \geq 0,$$

where the equality follows from (34) and (31), and the inequality is Jensen's one. For $\beta \neq 1$, the proof follows from Hölder's inequality. Let $\beta < 1$, $f = q'_Y(\alpha)$, and $g = q'_X(\alpha)$, then

$$\int_0^1 f^{1-\beta} d\alpha = \int_0^1 (g^{-\beta} f)^{1-\beta} g^{(1-\beta)\beta} d\alpha \leq \left(\int_0^1 g^{-\beta} f d\alpha\right)^{1-\beta} \left(\int_0^1 g^{1-\beta} d\alpha\right)^{\beta},$$

which implies $N_\beta[X,Y]^{\beta(1-\beta)} \geq 1$. Finally, let $\beta > 1$, $f = q'_Y(\alpha)^{-1}$, and $g = q'_X(\alpha)^{-1}$, then

$$\int_0^1 g^{\beta-1} d\alpha = \int_0^1 (g^\beta f^{-1})^{\frac{\beta-1}{\beta}} (f^{\beta-1})^{\frac{1}{\beta}} d\alpha \leq \left(\int_0^1 g^\beta f^{-1} d\alpha\right)^{\frac{\beta-1}{\beta}} \left(\int_0^1 f^{\beta-1} d\alpha\right)^{\frac{1}{\beta}},$$

which implies $N_\beta[X,Y]^{1-\beta} \leq 1$. Consequently, in both cases, $N_\beta[X,Y] \geq 1$. □

The next result is auxiliary and addresses existence of the integrals in (33).

PROPOSITION 4.4 *Let $g : (0,1) \to \mathbb{R}$ be a positive concave function. Then*

$$\int_0^1 g(\alpha)^u d\alpha < \infty \quad \text{for any constant } u > -1, \tag{35}$$

and the indefinite integral

$$h(\alpha) = \int g(\alpha)^u d\alpha \in \mathcal{L}^1(0,1) \quad \text{for any constant } u > -2. \tag{36}$$

PROOF. For $u \geq 0$, $g(\alpha)^u$ is a bounded continuous function of α on $(0,1)$, and thus, both (35) and (36) hold. Now let $u < 0$. The concavity and positiveness of $g(\alpha)$ imply $g(\alpha) \geq (1-2\alpha)g(0+) + 2\alpha g(1/2) \geq 2\alpha g(1/2)$ for $\alpha \in (0,1/2]$, and similarly, $g(\alpha) \geq 2(1-\alpha)g(1/2) + (1-2(1-\alpha))g(1-) \geq 2(1-\alpha)g(1/2)$ for $\alpha \in [1/2,1)$. Consequently, $g(\alpha) \geq g_0(\alpha) = 2\min\{\alpha,(1-\alpha)\}g(1/2)$ for all α, and thus,

$$\int_0^1 g(\alpha)^u d\alpha \leq 2\int_0^{1/2} (2\alpha g(1/2))^u d\alpha < \infty \qquad 0 > u > -1,$$

which proves (35) and (36) for $u > -1$. For $u \in (-2,-1]$, we obtain

$$|h(\beta)| \leq |C| + \left|\int_{\frac{1}{2}}^{\beta} g(\alpha)^u d\alpha\right| \leq |C| + \underbrace{\left|\int_{\frac{1}{2}}^{\beta} (g_0(\alpha))^u d\alpha\right|}_{I_u(\beta)},$$

where C is a constant. Observe that $I_u(\beta) \sim O\left(\beta^{u+1}\right)$ as $\beta \to 0$ and $I_u(\beta) \sim O\left((1-\beta)^{u+1}\right)$ as $\beta \to 1$ for $u \in (-2,-1)$. Also $I_u(\beta)$ grows as a logarithm as $\beta \to 0$ or $\beta \to 1$, for $u = -1$. Consequently, in both cases, $I_u(\beta) \in \mathcal{L}^1(0,1)$ and (36) follows. □

Next, Proposition 4.3 is applied to characterize a solution to problem (32).

PROPOSITION 4.5 *Let $\mathcal{D}$ be a law invariant deviation measure with maximal g-envelope G_M. Let $g \in G_M$, and let X be an r.v. such that $q'_X(\alpha) = C_g\, g(\alpha)^{-\frac{1}{\beta}}$, where $C_g = 1\Big/\int_0^1 g(u)^{1-\frac{1}{\beta}} du$. Then $e^{\mathcal{H}_\beta(Y)} \leq e^{\mathcal{H}_\beta(X)}\, \mathcal{D}(Y)$ for every r.v. $Y \in \mathcal{X}$.*

Proof. Because (35) holds, C_g is finite and positive for $\beta > 1/2$. Expressing $g(\alpha)$ through $q'_X(\alpha)$, we obtain $g(\alpha) = C(q'_X(\alpha))^{-\beta}$ with $C = C_g^\beta = 1 \Big/ \int_0^1 q'_X(\alpha)^{1-\beta} d\alpha$, and from $g \in G_M$, we have

$$\mathcal{D}(Y) \geq \int_0^1 g(\alpha) q'_Y(\alpha) d\alpha = \left(\int_0^1 q'_X(\alpha)^{1-\beta} d\alpha \right)^{-1} \int_0^1 q'_X(\alpha)^{-\beta} q'_Y(\alpha) d\alpha$$
$$= e^{\mathcal{H}_\beta(Y) - \mathcal{H}_\beta(X)} N_\beta[X,Y]^\beta \geq e^{\mathcal{H}_\beta(Y) - \mathcal{H}_\beta(X)},$$

where the second equality follows from (33) and (30), and the last inequality follows from Proposition 4.3. □

Proposition 4.5 implies that for the specified r.v. X, $\mathcal{H}_\beta(X) \geq \mathcal{H}_\beta(Y)$ whenever $\mathcal{D}(Y) \leq 1$. Thus, if $\mathcal{D}(X) \leq 1$ for some $g \in G_M$, then X solves (32). For a comonotone deviation measure $\mathcal{D}$, this leads to the following result.

Proposition 4.6 *Let $\mathcal{D}$ be a proper* comonotone *law invariant deviation measure. Then an r.v. X solves problem* (32) *if and only if $EX = 0$ and*

$$q'_X(\alpha) = C_g\, g(\alpha)^{-\frac{1}{\beta}}, \qquad C_g = 1 \Bigg/ \int_0^1 g(u)^{1-\frac{1}{\beta}} du\,, \tag{37}$$

where the function g is given by (10).

Proof. With the representation (10), $\mathcal{D}(X) = \int_0^1 g(\alpha) q'_X(\alpha) d\alpha = C_g \int_0^1 g(\alpha)^{1-\frac{1}{\beta}} d\alpha = 1$. By Proposition 4.5, $\mathcal{H}_\beta(X) \geq \mathcal{H}_\beta(Y)$ whenever $\mathcal{D}(Y) \leq 1$, and thus, X is a solution to (32). Because the Rényi differential entropy is a strictly concave functional of q'_X, X is unique. □

The quantile function $q_X(\alpha)$ of an r.v. X that solves (32) with an arbitrary comonotone deviation measure $\mathcal{D}$ is found from (37):

$$q_X(\alpha) = C_g \int g(\alpha)^{-\frac{1}{\beta}} d\alpha + C, \tag{38}$$

where C_g is determined in (37) and the integration constant C is chosen to satisfy the constraint $EX = 0$. For $\beta > 1/2$, we have $-\frac{1}{\beta} > -2$, and in view of (36), we obtain $q_X(\alpha) \in \mathcal{L}^1(0,1)$. This is the reason why (30) is defined for $\beta > \frac{1}{2}$.

Next we characterize the class of distributions that solve (32) with a comonotone deviation measure.

Definition 4.2 *A PDF $f_X(x)$ of an r.v. X is called α-concave, if*

- $\frac{f_X(x)^\alpha}{\alpha}$ *is a concave function on the support of $f_X(x)$ for $\alpha \neq 0$,*
- $\ln f_X(x)$ *is a concave function on the support of $f_X(x)$ for $\alpha = 0$ (such functions are also called* log-concave*).*

Proposition 4.7

(a) For a proper comonotone *law invariant deviation measure $\mathcal{D}$, the unique PDF $f_X(x)$ that solves* (32) *is $(\beta - 1)$-concave.*

(b) For a given r.v. X with a $(\beta-1)$-concave PDF $f_X(x)$, there exists a unique comonotone deviation measure $\mathcal{D}$ such that $f_X(x)$ solves (32) *with $\mathcal{D}$. This deviation measure is given by*

$$\mathcal{D}(Y) = \int_0^1 g(\alpha) dq_Y(\alpha), \tag{39}$$

where $g(\alpha) = C(q'_X(\alpha))^{-\beta}$ and $C = 1 \Big/ \int_0^1 q'_X(\alpha)^{1-\beta} d\alpha$.

Proof. The existence and uniqueness in (a) follow from Proposition 4.6. Let us prove that the solution has a $(\beta - 1)$-concave PDF. Because $g(\alpha)$ is concave, (37) implies that $h(\alpha) = q'_X(\alpha)^{-\beta}$ is a concave function. This holds if and only if the derivative $h'(\alpha) = -\beta(q'_X(\alpha))^{-\beta-1} q''_X(\alpha)$ exists almost everywhere and is decreasing. Differentiating the equality $F_X(q_X(\alpha)) = \alpha$, we obtain $q'_X(\alpha) = \frac{1}{f_X(q_X(\alpha))}$

and $q''_X(\alpha) = -\frac{q'_X(\alpha) f'_X(q_X(\alpha))}{f_X^2(q_X(\alpha))} = -\frac{f'_X(q_X(\alpha))}{f_X^3(q_X(\alpha))}$. Thus, $h'(\alpha)$ is decreasing if and only if $(f_X(x))^{\beta+1}\frac{f'_X(x)}{f_X^3(x)}$ is decreasing. The last expression is the derivative of $\frac{f_X(x)^{\beta-1}}{\beta-1}$ on the support of $f_X(x)$ for $\beta \neq 1$, and is the derivative of $\ln f_X(x)$ for $\beta = 1$. This proves that $f_X(x)$ is $(\beta-1)$-concave.

The proof of (b) is straightforward. For any X with a $(\beta-1)$-concave PDF, $g(\alpha) = C(q'_X(\alpha))^{-\beta}$ is concave and positive, and by Proposition 2.4, $\mathcal{D}(Y) = \int_0^1 g(\alpha) dq_Y(\alpha)$ is a comonotone deviation measure. □

Now the problem (32) is investigated for an arbitrary (not necessarily comonotone) deviation measure $\mathcal{D}$. We prove that there exists $g(\alpha) \in G_M$ such that $\mathcal{D}(X) \leq 1$ for the r.v. X defined in Proposition 4.5. This $g(\alpha)$ solves the optimization problem from the next proposition.

PROPOSITION 4.8 *Let $\mathcal{D}$ be a proper law invariant deviation measure, and let G_M be its maximal g-envelope. For $g \in G_M$, let*

$$W_\gamma(g) = \begin{cases} \frac{1}{1-\gamma} \ln \int_0^1 g(\alpha)^{1-\gamma} d\alpha, & \gamma \in (0,2), \quad \gamma \neq 1, \\ \int_0^1 \ln g(\alpha) d\alpha, & \gamma = 1, \end{cases} \tag{40}$$

then the optimization problem

$$\max_{g \in G_M} W_\gamma(g) \qquad \gamma \in (0,2) \tag{41}$$

has a unique solution $g^\star(\alpha) \in G_M$.

PROOF. The functional $W_\gamma(g)$ is finite by Proposition 4.4 (see the formula (35)). It is upper semi-continuous (with respect to pointwise convergence) for $\gamma \geq 1$ by Fatou's lemma and is continuous for $\gamma < 1$ by Proposition 3.2(a) and the dominated convergence theorem. The existence of solution follows from the compactness of G_M (see Proposition 3.2(b)) and Weierstrass's theorem, and the uniqueness of solution is the result of the convexity of G_M and strict concavity of $W(g)$. □

The function $g^\star(\alpha)$, described in Proposition 4.8, has the following properties.

PROPOSITION 4.9 *Let $\mathcal{D}$ be a proper law invariant deviation measure and let G_M be its maximal g-envelope. Also, let $\gamma \in (0,2)$, and let $g^\star(\alpha) \in G_M$ solve (41). Then for any nonzero $g(\alpha) \in G_M$, we have*

$$\int_0^1 \frac{g^\star(\alpha)}{g(\alpha)^\gamma} d\alpha \geq \int_0^1 \frac{g(\alpha)}{g(\alpha)^\gamma} d\alpha \tag{42}$$

and

$$\int_0^1 \frac{g^\star(\alpha)}{g^\star(\alpha)^\gamma} d\alpha \geq \int_0^1 \frac{g(\alpha)}{g^\star(\alpha)^\gamma} d\alpha. \tag{43}$$

PROOF. We begin with proving (42). Let $\gamma = 1$ then

$$0 \leq \int_0^1 \ln g^\star(\alpha) d\alpha - \int_0^1 \ln g(\alpha) d\alpha = \int_0^1 \ln \frac{g^\star(\alpha)}{g(\alpha)} d\alpha \leq \ln \int_0^1 \frac{g^\star(\alpha)}{g(\alpha)} d\alpha \qquad \forall g(\alpha) \in G_M,$$

where the last inequality is Jensen's one, and (42) follows. For $\gamma \neq 1$, by definition of $g^\star(\alpha)$, we have

$$\frac{1}{1-\gamma} \ln \int_0^1 g(\alpha)^{1-\gamma} d\alpha \leq \frac{1}{1-\gamma} \ln \int_0^1 g^\star(\alpha)^{1-\gamma} d\alpha \qquad \forall\, g(\alpha) \in G_M. \tag{44}$$

Let $d\nu(\alpha) = g(\alpha)^{1-\gamma} d\alpha$ be a nonnegative measure on $(0,1)$, then $\int_0^1 d\nu(\alpha)$ is finite in view of (35) (see Proposition 4.4), and therefore, $d\mathfrak{m}(\alpha) = \frac{d\nu(\alpha)}{\int_0^1 d\nu(\alpha)}$ is a probability measure on $(0,1)$. Then (42) is equivalent to

$$\int_0^1 \frac{g^\star(\alpha)}{g(\alpha)} d\mathfrak{m}(\alpha) \geq 1. \tag{45}$$

Using Jensen's inequality, we obtain

$$\left(\int_0^1 \frac{g^\star(\alpha)}{g(\alpha)} d\mathfrak{m}(\alpha) \right)^{1-\gamma} \leq \int_0^1 \left(\frac{g^\star(\alpha)}{g(\alpha)} \right)^{1-\gamma} d\mathfrak{m}(\alpha) = \frac{\int_0^1 g^\star(\alpha)^{1-\gamma} d\alpha}{\int_0^1 d\nu(\alpha)} \leq 1 \qquad \text{for } \gamma > 1,$$

and

$$\left(\int_0^1 \frac{g^\star(\alpha)}{g(\alpha)} d\mathfrak{m}(\alpha)\right)^{1-\gamma} \geq \int_0^1 \left(\frac{g^\star(\alpha)}{g(\alpha)}\right)^{1-\gamma} d\mathfrak{m}(\alpha) = \frac{\int_0^1 g^\star(\alpha)^{1-\gamma} d\alpha}{\int_0^1 d\nu(\alpha)} \geq 1 \qquad \text{for } 0 < \gamma < 1,$$

which result in (45) (in the two lines above, the last inequality follows from (44)).

Now we show (43). The integral $J = \int_0^1 g^\star(\alpha)^{1-\gamma} d\alpha$ is finite in view of (35), and for each $\lambda \in (0,1)$, we have

$$\int_0^1 \frac{g^\star(\alpha)}{(\lambda g(\alpha) + (1-\lambda) g^\star(\alpha))^\gamma} d\alpha \leq \int_0^1 \frac{g^\star(\alpha)}{((1-\lambda) g^\star(\alpha))^\gamma} d\alpha = (1-\lambda)^{-\gamma} J. \tag{46}$$

Because G_M is convex, $g_\lambda(\alpha) = \lambda g(\alpha) + (1-\lambda) g^\star(\alpha) \in G_M$, and it follows from (42) that

$$\int_0^1 \frac{g^\star(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha \geq \int_0^1 \frac{\lambda g(\alpha) + (1-\lambda) g^\star(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha.$$

For $\lambda \neq 0$, this implies

$$\int_0^1 \frac{g^\star(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha \geq \int_0^1 \frac{g(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha. \tag{47}$$

Combining (47) and (46), we obtain

$$(1-\lambda)^{-\gamma} J \geq \int_0^1 \frac{g(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha.$$

Because $g(\alpha) g_\lambda(\alpha)^{-\gamma} \to g(\alpha) g^\star(\alpha)^{-\gamma}$ pointwise as $\lambda \to 0$, using Fatou's lemma, we have

$$J = \lim_{\lambda \to 0} \left((1-\lambda)^{-\gamma} J\right) \geq \liminf_{\lambda \to 0} \int_0^1 \frac{g(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha \geq \int_0^1 \lim_{\lambda \to 0} \frac{g(\alpha)}{g_\lambda(\alpha)^\gamma} d\alpha = \int_0^1 \frac{g(\alpha)}{g^\star(\alpha)^\gamma} d\alpha,$$

which proves (43). □

The next proposition characterizes solutions to the problem (32) with an arbitrary law invariant deviation measure.

PROPOSITION 4.10 *Let $\mathcal{D}$ be a proper law invariant deviation measure and let G_M be its maximal g-envelope. Also, let $g^\star(\alpha) \in G_M$ solve* (41) *for $\gamma = 1/\beta$. Then an r.v. X solves problem* (32) *if and only if $EX = 0$ and*

$$q'_X(\alpha) = C\, g^\star(\alpha)^{-\frac{1}{\beta}}, \qquad C = 1 \Big/ \int_0^1 g^\star(u)^{1-\frac{1}{\beta}} du. \tag{48}$$

PROOF. For X satisfying (48) and $\gamma = \frac{1}{\beta}$, we have

$$\mathcal{D}(X) = \sup_{g(\alpha) \in G_M} \int_0^1 g(\alpha)\, d(q_X(\alpha)) = \sup_{g(\alpha) \in G_M} \int_0^1 \frac{C\, g(\alpha)}{g^\star(\alpha)^\gamma} d\alpha = C \int_0^1 g^\star(\alpha)^{1-\gamma} d\alpha = 1,$$

where the third equality follows from (43). Thus, the constraints in (32) are satisfied, and Proposition 4.5 guarantees that X is a solution to (32). The uniqueness of X follows from strict concavity of $\mathcal{H}_\beta$. □

Consequently, to solve (32) with a law invariant deviation measure, for which the constraint $\mathcal{D}(X) = 1$ cannot be expressed in the form (23), we suggest the following approach:

(i) Given law invariant $\mathcal{D}$, find a solution $g^\star(\alpha)$ to the problem (41). In the case when $\mathcal{D}$ is comonotone, $g^\star(\alpha)$ coincides with $g(\alpha)$ in (10).

(ii) Find the quantile function of X as the antiderivative of (48) such that $EX = 0$.

The solution to the initial problem (22) is then $X^\star = \mu + d \cdot X$.

REMARK 4.1 *Compared to solving* (22) *directly, this approach has the obvious advantage for comonotone deviation measures.*

Proposition 4.10 is central to characterizing the class of maximum-entropy distributions with given mean and deviation $\mathcal{D}$.

Proposition 4.11 *Let $\mathcal{D}$ be a proper law invariant deviation measure. Then the problem* (32) *with $\mathcal{D}$ has a unique solution $f_X(x)$, which is $(\beta-1)$-concave (log-concave for Shannon entropy). Conversely, for an arbitrary r.v. X with $EX = 0$ and a $(\beta-1)$-concave PDF $f_X(x)$, there exist infinitely many deviation measures such that the solution to the corresponding problem* (32) *is $f_X(x)$. Exactly one of these deviation measures is comonotone and is given by* (39).

Proof. The proof follows from Propositions 4.7 and 4.10. □

4.3 Examples. The results obtained in the preceding section are illustrated for entropy maximization with the full-range deviation, CVaR-deviation, mixed CVaR-deviation, and mean absolute deviation (MAD).

First, maximum-entropy distributions for CVaR-deviation and mixed CVaR-deviation are derived for the Shannon entropy maximization problem (21) with $\mu = 0$ and $d = 1$:

$$\max_{X\in\mathcal{X}} S(X) \qquad \text{s.t.} \quad EX = 0, \quad \mathcal{D}(X) = 1. \tag{49}$$

Example 4.7 (Shannon Entropy Maximization with CVaR-Deviation) *A solution to* (49) *with CVaR-deviation* (2)*, which is comonotone, has the PDF*

$$f_X(x) = \begin{cases} (1-\alpha)\exp\left(\frac{1-\alpha}{\alpha}\left(x - \frac{2\alpha-1}{1-\alpha}\right)\right) & x \le \frac{2\alpha-1}{1-\alpha}, \\ (1-\alpha)\exp\left(-\left(x - \frac{2\alpha-1}{1-\alpha}\right)\right) & x \ge \frac{2\alpha-1}{1-\alpha}. \end{cases} \tag{50}$$

Figure 2 shows $f_X(x)$ for $\alpha = 0.01$, 0.3, 0.5, 0.7, 0.8, and 0.9.

For this $f_X(x)$, a deviation measure, restored by (39) *with $\beta = 1$, is* $\mathrm{CVaR}^{\Delta}_{\alpha}(X)$.

Figure 2: The PDF $f_X(x)$ (see (50) that solves the entropy maximization problem (49) with CVaR-deviation for $\alpha = 0.01$, 0.3, 0.5, 0.7, 0.8, and 0.9.

Detail. $\mathrm{CVaR}^{\Delta}_{\alpha}$ can be represented in the form (10) with $g(\beta)$ given by (12); see Example 2.1. If X solves (49) with CVaR-deviation, then by Proposition 4.6, we have

$$q'_X(\beta) = \frac{1}{g(\beta)} = \begin{cases} \frac{\alpha}{1-\alpha}\frac{1}{\beta} & \beta \le \alpha, \\ \frac{1}{1-\beta} & \beta \ge \alpha, \end{cases}$$

and for the quantile function of X, we obtain

$$q_X(\beta) = \int \frac{1}{g(\beta)} d\beta = q_0 + \begin{cases} \frac{\alpha}{1-\alpha}\ln\frac{\beta}{\alpha} & \beta \le \alpha, \\ -\ln\frac{1-\beta}{1-\alpha} & \beta \ge \alpha, \end{cases}$$

where the integration constant $q_0 = \frac{2\alpha-1}{1-\alpha}$ is found from the condition $EX = \int_0^1 q_X(\beta)d\beta = 0$, and consequently,

$$q_X(\beta) = \frac{2\alpha-1}{1-\alpha} + \begin{cases} \frac{\alpha}{1-\alpha}\ln\frac{\beta}{\alpha} & \beta \le \alpha, \\ -\ln\frac{1-\beta}{1-\alpha} & \beta \ge \alpha. \end{cases}$$

Finally, (50) is found as the derivative of the inverse function for $q_X(\beta)$. □

Example 4.8 (Shannon Entropy Maximization with Mixed CVaR-deviation) *An r.v. X solves* (49) *with* mixed CVaR-deviation (29)*, which is a comonotone deviation measure, if and only if $EX = 0$ and*

$$q'_X(\beta) = \left(\mathfrak{m}(0,\beta] + \beta\int_{\beta+}^{1-}\frac{d\mathfrak{m}(\alpha)}{\alpha} - \beta\right)^{-1}. \tag{51}$$

For this distribution, a deviation measure, restored by (39) *with $\beta = 1$, is mixed CVaR-deviation.*

Detail. According to Proposition 4.6, X solves (49) if and only if $EX = 0$ and $q'_X(\beta) = \frac{1}{g(\beta)}$, where $g(\beta)$ is defined by (10) and can be calculated as $g(\beta) = \mathcal{D}(X_\beta)$ with X_β (11). Substituting (10) with (12) (see Example 2.1) into (29), we obtain

$$g(\beta) = \int_0^1 \mathrm{CVaR}_\alpha^\Delta(X_\beta) d\mathfrak{m}(\alpha) = \int_0^\beta (1-\beta) d\mathfrak{m}(\alpha) + \int_{\beta+}^1 \left(\frac{1}{\alpha} - 1\right) \beta d\mathfrak{m}(\alpha) =$$

$$= \mathfrak{m}(0,\beta] - \beta\big(\mathfrak{m}(0,\beta] + \mathfrak{m}(\beta,1)\big) + \beta \int_{\beta+}^{1-} \frac{d\mathfrak{m}(\alpha)}{\alpha},$$

which proves (51). □

The next example presents a solution to the Rényi entropy maximization problem (32) with CVaR-deviation.

Example 4.9 (Rényi Entropy Maximization with CVaR-Deviation) *A solution to (32) with CVaR-deviation (2), which is comonotone, for $\beta \neq 1$ has the PDF*

$$f_X(x) = \begin{cases} \frac{\beta}{2\beta-1}\frac{1-\alpha}{\alpha^{1/\beta}}\left(\frac{\beta-1}{2\beta-1}\frac{1-\alpha}{\alpha^{1/\beta}}x + \frac{\alpha+\beta-1}{2\beta-1}\frac{1}{\alpha^{1/\beta}}\right)^{\frac{1}{\beta-1}} & x \le \frac{2\alpha-1}{1-\alpha}, \quad x \in V, \\ \frac{\beta}{2\beta-1}\frac{1-\alpha}{(1-\alpha)^{1/\beta}}\left(-\frac{\beta-1}{2\beta-1}\frac{1-\alpha}{(1-\alpha)^{1/\beta}}x + \frac{\beta-\alpha}{2\beta-1}\frac{1}{(1-\alpha)^{1/\beta}}\right)^{\frac{1}{\beta-1}} & x \ge \frac{2\alpha-1}{1-\alpha}, \quad x \in V, \\ 0 & x \notin V, \end{cases} \tag{52}$$

where $V = (-\infty, \infty)$ for $\beta < 1$ and $V = \left[-\frac{\alpha+\beta-1}{(\beta-1)(1-\alpha)}, \frac{\beta-\alpha}{(\beta-1)(1-\alpha)}\right]$ for $\beta > 1$. Figure 3 shows the function $f_X(x)$ for various α and β.

For this PDF, a deviation measure, restored by (39) with $\beta \neq 1$, is $\mathrm{CVaR}_\alpha^\Delta(X)$.

Detail. $\mathrm{CVaR}_\alpha^\Delta$ can be represented in the form (10) with g given by (12) (see Example 2.1). If X solves (32) with $\mathcal{D}(X) = \mathrm{CVaR}_\alpha^\Delta(X)$, then by Proposition 4.6, $q'_X(u) = C_g\, g(u)^{-\gamma}$, where $\gamma = \frac{1}{\beta}$ and $C_g = 1 \Big/ \int_0^1 g(u)^{1-\gamma} du = \frac{2-\gamma}{(1-\alpha)^{1-\gamma}}$. Thus, we have

$$q'_X(u) = \frac{C_g}{g(u)^\gamma} = \begin{cases} \frac{2-\gamma}{1-\alpha}\left(\frac{\alpha}{u}\right)^\gamma & u \le \alpha, \\ \frac{2-\gamma}{(1-\alpha)^{1-\gamma}}\frac{1}{(1-u)^\gamma} & u \ge \alpha. \end{cases}$$

Then the quantile function of X takes the form

$$q_X(u) = C + \int_\alpha^u q'_X(\lambda) d\lambda = C + \begin{cases} \frac{2-\gamma}{1-\gamma}\frac{\alpha^\gamma}{1-\alpha}u^{1-\gamma} - \frac{2-\gamma}{1-\gamma}\frac{\alpha}{1-\alpha} & u \le \alpha, \\ -\frac{2-\gamma}{1-\gamma}\frac{1}{(1-\alpha)^{1-\gamma}}(1-u)^{1-\gamma} + \frac{2-\gamma}{1-\gamma} & u \ge \alpha, \end{cases}$$

where the integration constant $C = \frac{2\alpha-1}{1-\alpha}$ is found from $EX = \int_0^1 q_X(u) du = 0$. Finally, (52) is obtained as the derivative of the inverse function for $q_X(u)$. □

(a) $\beta = 0.6$ (b) $\beta = 1.5$

(c) $\beta = 2$ (d) $\beta = 3$

Figure 3: The PDF $f_X(x)$ (see (52) that solves the Rényi entropy maximization problem (32) with CVaR-deviation for $\alpha = 0.01$, 0.3, 0.5, 0.7, 0.8, and 0.9 in four cases: (a) $\beta = 0.6$, (b) $\beta = 1.5$, (c) $\beta = 2$, and (d) $\beta = 3$.

The next example presents the maximum-entropy distribution in (32) for the full-range deviation.

Example 4.10 (Rényi Entropy Maximization with Full-Range Deviation) *A solution to* (32) *with the full-range deviation* $\mathcal{D}(X) = \operatorname{ess\,sup} X - \operatorname{ess\,inf} X$ *has the uniform distribution on* $(-\frac{1}{2}, \frac{1}{2})$. *If given this distribution, a deviation measure is restored by* (39), *then it is the full-range deviation.*

Detail. $\mathcal{D}(X) = \operatorname{ess\,sup} X - \operatorname{ess\,inf} X$ is comonotone and can be represented in the form (10) with $g(\alpha) \equiv 1$. Thus, for the solution $X^\star$ to (32), Proposition 4.6 implies that $q'_{X^\star}(\alpha) \equiv 1$, or $q_{X^\star}(\alpha) = \alpha + C$. The condition $EX^\star = 0$ yields $C = -\frac{1}{2}$, and consequently, $X^\star$ has the uniform distribution on $(-\frac{1}{2}, \frac{1}{2})$. □

Next, two examples present solutions to the Shannon and Rényi entropy maximization problems with MAD.

Example 4.11 (Shannon Entropy Maximization with Mean Absolute Deviation) *A solution to* (49) *with* $\mathrm{MAD}(X) = E|X - EX|$ *has the PDF*

$$f_X(x) = \begin{cases} 1/2\, e^{x} & x \le 0, \\ 1/2\, e^{-x} & x \ge 0. \end{cases} \tag{53}$$

Detail. The formula (53) follows from Boltzmann's theorem, because the constraints $EX = 0$ and $\mathrm{MAD}(X) = 1$ can be represented in the form (23) with $h_1(x) = x$, $a_1 = 0$, $h_2(x) = |x|$, and $a_2 = 1$.

As an illustration for the developed approach, we also prove (53) using Proposition 4.10. Because the logarithm is a monotonic function, a solution to (41) with $\gamma = 1$ and G_M given by (15) can be represented in the form

$$g_x(\beta) = \begin{cases} (2-x)\beta & \beta \le x/2, \\ x(1-\beta) & \beta \ge x/2, \end{cases}$$

for some $x \in (0, 2)$. For $\int_0^1 \ln g_x(\beta) d\beta = \ln x + \ln(1 - x/2) - 1$, the maximum is attained at $x = 1$, and thus,

$$g^*(\beta) = \begin{cases} \beta & \beta \le 1/2, \\ 1-\beta & \beta \ge 1/2. \end{cases}$$

A solution to (49) is given by $q'_X(\beta) = \frac{1}{g^\star(\beta)}$ (see Proposition 4.10), and the PDF (53) is the derivative of the inverse function for $q_X(\beta)$. □

Example 4.12 (Rényi Entropy Maximization with Mean Absolute Deviation) *A solution to* (32) *with* $\mathrm{MAD}(X) = E|X - EX|$ *for* $\beta \ne 1$ *has the PDF*

$$f_X(x) = \begin{cases} \frac{\beta}{4\beta-2}\left(\frac{\beta-1}{2\beta-1}x + 1\right)^{\frac{1}{\beta-1}} & x \le 0, \quad x \in V \\ \frac{\beta}{4\beta-2}\left(-\frac{\beta-1}{2\beta-1}x + 1\right)^{\frac{1}{\beta-1}} & x \ge 0, \quad x \in V, \\ 0 & x \notin V, \end{cases} \tag{54}$$

where $V = (-\infty, \infty)$ *for* $\beta < 1$ *and* $V = \left[-\frac{2\beta-1}{\beta-1}, \frac{2\beta-1}{\beta-1}\right]$ *for* $\beta > 1$.

Detail. First, we solve (41) with G_M given by (15). Because $h(y) = \frac{y^{1-\gamma}}{1-\gamma}$ is a monotonic function, (41) attains its maximum at one of the functions

$$g_x(u) = \begin{cases} (2-x)u & u \le x/2, \\ x(1-u) & u \ge x/2, \end{cases}$$

for some $x \in (0, 2)$. Because $\int_0^1 \frac{g_x(u)^{1-\gamma}}{1-\gamma} du = \frac{1}{(2-\gamma)2^{2-\gamma}} \frac{(2-x)^{1-\gamma} + x^{1-\gamma}}{1-\gamma}$, the maximum will be attained at $x = 1$, and we have

$$g^*(u) = \begin{cases} u & u \le 1/2, \\ 1-u & u \ge 1/2. \end{cases}$$

By Proposition 4.10, a solution X to (32) is such that $EX=0$ and $q'_X(u)=C_g\, g^*(u)^{-\gamma}$, where $\gamma=\frac{1}{\beta}$ and $C_g=1\Big/\int_0^1 g^*(u)^{1-\gamma}du\,=(2-\gamma)2^{1-\gamma}$. Then the quantile function of X takes the form

$$q_X(u)=C+\int_{\frac{1}{2}}^{u} q'_X(\lambda)d\lambda=C+\begin{cases}\frac{2-\gamma}{1-\gamma}(2u)^{1-\gamma}-\frac{2-\gamma}{1-\gamma} & u\le\frac{1}{2},\\ \frac{2-\gamma}{1-\gamma}[2(1-u)]^{1-\gamma}+\frac{2-\gamma}{1-\gamma} & u\ge\frac{1}{2},\end{cases}$$

where the integration constant $C=0$ is found from the condition $EX=\int_0^1 q_X(u)du=0$. Finally, (54) is obtained as the derivative of the inverse function for $q_X(u)$. □

Remark 4.2 *Applying* (39) *to the PDF* (54), *we obtain* median absolute deviation $\mathcal{D}(X)=\mathrm{CVaR}_{1/2}(X)$. *This illustrates the fact that different deviation measures may lead to the same optimal PDF in* (32). *Among all these measures, the formula* (39) *provides only the comonotone one.*

In particular, this fact suggests a surjective mapping of all deviation measures to the class of comonotone deviation measures. Example 4.1 shows that the solution to the Shannon entropy maximization problem (49) with standard deviation is the standard normal distribution $\mathcal{N}(0,1)$. We consider the inverse problem: What comonotone deviation measure corresponds to $\mathcal{N}(0,1)$ through the maximum-entropy principle?

Example 4.13 (Maximum-Entropy Inverse Problem for $\mathcal{N}(0,1)$) *A comonotone deviation measure $\mathcal{D}$ that produces $\mathcal{N}(0,1)$ as the solution to the maximum-entropy problem* (49) *is given by* (10) *with*

$$g(\alpha)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(\Phi^{-1}(\alpha))^2\right), \tag{55}$$

where $\Phi^{-1}(\alpha)$ is the inverse to the cumulative distribution function of $\mathcal{N}(0,1)$.

Detail. For an r.v. $X\sim\mathcal{N}(0,1)$, we have $q_X(\alpha)=\Phi^{-1}(\alpha)$ and $q'_X(\alpha)=\frac{1}{\Phi'(\Phi^{-1}(\alpha))}=\sqrt{2\pi}\cdot\exp\left(\frac{1}{2}(\Phi^{-1}(\alpha))^2\right)$. Substituting this $q'_X(\alpha)$ into (39) with $\beta=1$, we obtain (55). □

Thus, we conclude that standard deviation corresponds to the comonotone deviation measure $\mathcal{D}$ in Example 4.13 through the maximum-entropy principle.

Similarly, we address the inverse problem with the Rényi entropy for $\mathcal{N}(0,1)$.

Example 4.14 (Inverse Problem with the Rényi Entropy for $\mathcal{N}(0,1)$) *A comonotone deviation measure producing $\mathcal{N}(0,1)$ as a solution to the Rényi entropy maximization problem* (32) *with $\beta<1$ is given by* (10) *with*

$$g(\alpha)=\sqrt{\frac{\beta}{2\pi}}\exp\left(-\frac{\beta}{2}(\Phi^{-1}(\alpha))^2\right), \tag{56}$$

where $\Phi^{-1}(\alpha)$ is the inverse to the cumulative distribution function of $\mathcal{N}(0,1)$.

Detail. Because $\mathcal{N}(0,1)$ is log-concave, it is $(\beta-1)$-concave for $\beta<1$. For an r.v. $X\sim\mathcal{N}(0,1)$, we have $q'_X(\alpha)=\sqrt{2\pi}\exp\left(\frac{1}{2}(\Phi^{-1}(\alpha))^2\right)$, see Example 4.13. Thus, according to Proposition 4.7(b), a comonotone deviation measure that produces $\mathcal{N}(0,1)$ as an outcome of (32) is given by (10) with $g(\alpha)=C\big(q'_X(\alpha)\big)^{-\beta}=C(2\pi)^{-\beta/2}\exp\left(-\frac{\beta}{2}(\Phi^{-1}(\alpha))^2\right)$ where $C=1\Big/\int_0^1 q'_X(\alpha)^{1-\beta}d\alpha$. With $\alpha=\Phi(x)$, we obtain $C=\sqrt{\beta}(2\pi)^{(\beta-1)/2}$, and consequently, (56) follows. □

The next example highlights practical aspects of the maximum-entropy principle with general deviation measures. It solves the inverse problem: Given historical data for stock's rate of return, estimate the probability distribution for the rate of return and find a deviation measure that produces that distribution through the maximum-entropy principle. We rely on the belief that risk preferences of the investors are fully reflected in stock's expected rate of return and some deviation measure. In fact, this belief is the extension of Markowitz's mean-variance approach [18], according to which all investors are concerned only with the mean and variance (or equivalently, standard deviation) of stocks' rates of return. Solving the

(a) empirical distribution and its approximation (b) the function $g(\alpha)$ in (39), $\beta = 1$

Figure 4: (a) the empirical distribution of the monthly historical rates of return for the Bank of America Corporation stock and its log-concave approximation; (b) the function $g(\alpha)$ for the deviation measure (39) ($\beta = 1$) restored through the maximum-entropy principle for the log-concave distribution in (a).

inverse problem is based on the fact that the Shannon maximum-entropy principle establishes one-to-one correspondence between the class of log-concave PDFs and the class of comonotone deviation measures (see Proposition 4.7, item (b), case $\beta = 1$).

Example 4.15 (Restored Deviation Measure) *Using monthly historical rates of return for the Bank of America Corporation stock for the last eight years, we approximate the empirical distribution of the rate of return by a log-concave distribution*[13] *and then restore the comonotone deviation measure using* (39) *with* $\beta = 1$ *for the approximating distribution. The deviation measure is given by* (10)*, where* $g(\alpha)$ *is calculated numerically and is shown on Figure 4.*

Concluding the section, we reexamine Example 4.15 with the Rényi entropy. As in Example 4.15, solving the inverse problem is based on the fact that for any fixed $\beta > 1/2$, the Rényi entropy maximization problem (22) establishes one-to-one correspondence between the class of $(\beta - 1)$-concave distributions and the class of comonotone deviation measures. If for a given β we denote this class of $(\beta - 1)$-concave distributions by $\mathcal{C}_\beta$, then $\mathcal{C}_{\beta_2} \subseteq \mathcal{C}_{\beta_1}$ for any $\beta_1 < \beta_2$. Because we restrict β to be $\beta > 1/2$, the set $\mathcal{C}_{1/2}$ is the largest among those with $\beta \geq 1/2$.

(a) empirical distribution and its approximation (b) the function $g(\alpha)$ in (39)

Figure 5: (a) the empirical distribution of the monthly historical rates of return for the Bank of America Corporation stock and its approximation; (b) the function $g(\alpha)$ for the deviation measure (39) restored through the maximum-entropy principle with the Rényi entropy ($\beta \to 1/2$) and for the approximating distribution in (a).

Example 4.16 *Using the same historical data for the rate of return for the Bank of America Corporation stock as in Example 4.15, we approximate the empirical distribution of the rate of return by the distribution with the PDF* $f_X(x)$ *such that* $\frac{(f_X(x))^{\beta-1}}{\beta-1}$ *is a concave function for* $\beta \to 1/2$[14] *and then restore a comonotone deviation measure using* (39) *with the approximating distribution. The deviation measure is given by* (10)*, where* $g(\alpha)$ *is calculated numerically for* $\beta \to 1/2$ *and is shown on Figure 5.*

Comparing Figures 4 and 5, we conclude that although the approximating distributions are sufficiently close, the corresponding functions $g(\alpha)$ differ significantly. This observation suggests that the choice of β in the Rényi entropy has a strong impact on a restored deviation and, consequently, on agent's risk preferences associated with that deviation measure.

5. Conclusions.

This work has formulated the problem of Shannon and Rényi entropy maximization with a constraint on a general deviation measure introduced by Rockafellar et al. and has generalized the recent results on Rényi entropy maximization with constraints on standard deviation and pth moment. It has also developed a new representation for deviation measures (Proposition 2.2(d)) that played a pivotal role in adapting existing entropy-maximization approaches to solving the formulated problem. The chain of intermediate propositions and auxiliary results has culminated in Proposition 4.10. As an illustration, new maximum-entropy distributions for the Shannon and Rényi entropies, in particular with conditional value-at-risk deviation, have been obtained. As another major contribution, this work has

[13] The logarithm of the empirical distribution is convexified, and the approximating distribution is determined as the normalized exponential function of the convexified distribution.

[14] In this example, we choose $\beta = 0.5 + 10^{-9}$.

solved the inverse entropy-maximization problem: Finding a deviation measure that corresponds to a given probability distribution function through the maximum-entropy principle. This problem finds its application in financial engineering and risk analysis. In particular, it could be used for restoring risk preferences of an agent from historical rates of return of agent's financial instruments.

Appendix A. Proof of Proposition 2.2. Let us show that $(b) \to (c)$. It follows from (2) that

$$-\alpha\,\mathrm{CVaR}_\alpha^\Delta(X) = \int_0^\alpha (q_X(t) - EX)dt, \tag{57}$$

which, along with the property $\int_0^1 \phi(\alpha)\,d\alpha = 0$, reduces the integral in (5) to

$$I = \int_0^1 \phi(\alpha) q_X(\alpha) d\alpha = \int_0^1 \phi(\alpha)(q_X(\alpha) - EX) d\alpha = \int_0^1 \phi(\alpha) d(-\alpha\,\mathrm{CVaR}_\alpha^\Delta(X)).$$

Integrating the last integral by parts, we obtain

$$I = \left(-\alpha\,\mathrm{CVaR}_\alpha^\Delta(X)\,\phi(\alpha)\right)\Big|_0^1 + \int_0^1 \alpha\,\mathrm{CVaR}_\alpha^\Delta(X) d(\phi(\alpha)). \tag{58}$$

It is left to prove that the first term in (58) vanishes. First, we show that

$$\lim_{\alpha\to 0}\left(-\alpha\,\mathrm{CVaR}_\alpha^\Delta(X)\,\phi(\alpha)\right) = 0. \tag{59}$$

With (57) and the fact that for sufficiently small α, the function $|\phi|$ monotonously decreases on $(0,\alpha)$, we obtain

$$\left|\alpha\,\mathrm{CVaR}_\alpha^\Delta(X)\phi(\alpha)\right| = \left|\phi(\alpha)\int_0^\alpha (q_X(t) - EX)\,dt\right| \le \int_0^\alpha \left|\phi(\alpha)(q_X(t)-EX)\right|dt \le \int_0^\alpha \left|\phi(t)(q_X(t)-EX)\right|dt.$$

Because $q_X(t) \in \mathcal{L}^p(0,1)$, and $\phi(t) \in \mathcal{L}^q(0,1)$, the integral $\int_0^1 |\phi(t)||q_X(t) - EX|dt$ is finite. Consequently, the fact $\int_0^\alpha \left|\phi(t)(q_X(t) - EX)\right|dt \to 0$ as $\alpha \to 0$ can be shown by Lebesgue's dominated convergence theorem. This proves (59).

Similarly, we can show that $\lim_{\alpha\to 1}\left(-\alpha\,\mathrm{CVaR}_\alpha^\Delta(X)\,\phi(\alpha)\right) = 0$. Consequently, the representations (5) and (6) are equivalent.

Now, we show that $(b) \to (d)$. For every nonzero $\phi(\alpha) \in \Lambda$, the function $g(\alpha) = -\int_0^\alpha \phi(t)dt$ is positive and concave and satisfies $g(0) = g(1) = 0$. Integrating the original integral I by parts, we obtain

$$I = \int_0^1 \phi(\alpha) q_X(\alpha) d\alpha = (-g(\alpha)\,q_X(\alpha))\Big|_0^1 + \int_0^1 g(\alpha)\,d(q_X(\alpha)). \tag{60}$$

It is left to prove that the first term in (60) vanishes. Indeed, if $q_X(\alpha) \to -\infty$ as $\alpha \to 0$ then for sufficiently small α we have $|g(\alpha)\,q_X(\alpha)| \le |q_X(\alpha)| \int_0^\alpha |\phi(t)|dt \le \int_0^\alpha |q_X(t)||\phi(t)|dt \to 0$ as $\alpha \to 0$ (the last integral vanishes because $q_X(t) \in \mathcal{L}^p(0,1)$, and $\phi(t) \in \mathcal{L}^q(0,1)$). Similarly, $|g(\alpha)\,q_X(\alpha)| \to 0$ as $\alpha \to 1$. This proves that $(b) \to (d)$.

Finally, we show that $(d) \to (a)$, i.e. for every collection G of positive concave functions $g : (0,1) \to \mathbb{R}$, the functional (7) is a law invariant deviation measure. Because the axioms D1–D4 are preserved under the supremum operation, it suffices to establish $(d) \to (a)$ for $\mathcal{D}_g(X) = \int_0^1 g(\alpha)\,d(q_X(\alpha))$ with a positive concave function $g(\alpha)$.

First, we assume that $g(\alpha)$ is a piecewise-linear concave function with finite number of linear pieces and such that $g(\alpha) > 0$ for $\alpha \in (0,1)$ and $g(0+) = g(1-) = 0$. Denote $a = g'(0+)$ and $b = g'(1-)$ and $\phi(\alpha) = -g'(\alpha)$, where the derivative exists. Then integrating $\int_0^1 g(\alpha)\,d(q_X(\alpha))$ by parts, we obtain

$$\mathcal{D}_g(X) = \int_0^1 g(\alpha)\,d(q_X(\alpha)) = b\,\lim_{\alpha\to 1}((1-\alpha)\,q_X(\alpha)) - a\,\lim_{\alpha\to 0}(\alpha\,q_X(\alpha)) + \int_0^1 \phi(\alpha)\,q_X(\alpha)d\alpha. \tag{61}$$

Because $q_X(\alpha) \in \mathcal{L}^p(0,1) \subset \mathcal{L}^1(0,1)$, both limits in (61) vanish (as those in (60)). Because in addition $\phi(\alpha)$ is a nondecreasing nonzero function, $\phi(\alpha) \in \mathcal{L}^\infty(0,1) \subset \mathcal{L}^q(0,1)$, and $\int_0^1 \phi(\alpha)d\alpha = 0$, it follows from the part $(b) \to (a)$ that the functional $\mathcal{D}_g(X)$ is a law invariant deviation measure.

Now let $g \in G$ be an arbitrary nonzero function and let $g_n(\alpha)$ for every $n \in \mathbb{N}$ be a piecewise-linear function with 2^n pieces such that $g_n(0+) = g_n(1-) = 0$ and $g_n(i/2^n) = g(i/2^n)$ for $i = 1, \ldots, 2^n - 1$. Then $\{g_n(\alpha)\}_{n\in\mathbb{N}}$ is a monotonically increasing sequence of nonnegative functions, and we have $\lim_{n\to\infty} g_n(\alpha) = g(\alpha)$ pointwise. It follows from the monotone convergence theorem that

$$\int_0^1 g(\alpha)\, d(q_X(\alpha)) = \lim_{n\to\infty} \int_0^1 g_n(\alpha)\, d(q_X(\alpha)) = \sup_{n\in\mathbb{N}} \int_0^1 g_n(\alpha)\, d(q_X(\alpha)). \tag{62}$$

Because the axioms D1–D4 are preserved under the supremum operation, the functional (62) is a law invariant deviation measure and, consequently, so is (7).

Appendix B. Version of Boltzmann's Theorem for the Rényi Entropy. If the constraint $\mathcal{D}(X) = d$ can be expressed in the form (23), a distribution maximizing the Rényi entropy in (22) for $\beta \neq 1$ can be represented in the form similar to (24) in Boltzmann's theorem.

Let $V \subseteq \mathbb{R}$ be a closed subset and let $h_1, \ldots, h_n$ be measurable functions. Also, let $\mathcal{B}$ be the set of all continuous r.v.s X with the support V (i.e., those whose PDFs are zero outside of V) and satisfying the conditions

$$E(h_j(X)) = a_j \qquad j = 1, \ldots, n, \tag{63}$$

where $a_1, \ldots, a_n$ are given.

A general formulation of the Rényi entropy maximization problem subject to (63) is given by

$$\begin{aligned} &\max \int_V \big(f_X(x)\big)^\beta dx \quad \text{if } \beta < 1 \qquad \text{or} \qquad \min \int_V \big(f_X(x)\big)^\beta dx \quad \text{if } \beta > 1 \\ &\text{s.t.} \int_V h_j(x) f_X(x) dx = a_j \qquad j = 0, \ldots, n, \\ &f_X(x) \geq 0 \end{aligned}$$

where $h_0(x) \equiv 1$ and $a_0 = 1$.

With Lagrange multipliers $\lambda_0, \ldots, \lambda_n$ and $\mu(x)$, the Lagrangian for this problem takes the form

$$\mathcal{L} = \int_V \Big[\big(f_X(x)\big)^\beta + \sum\nolimits_{j=0}^n \lambda_j h_j(x) f_X(x) + \mu(x) f_X(x)\Big]\, dx - \sum\nolimits_{j=0}^n \lambda_j a_j$$

and the necessary optimality conditions are determined by

$$\beta\big(f_X(x)\big)^{\beta-1} + \sum\nolimits_{j=0}^n \lambda_j h_j(x) + \mu(x) = 0$$

with the complementarity conditions $\mu(x) f_X(x) = 0$ and $\mu(x) \geq 0$ for $\beta < 1$ ($\mu(x) \leq 0$ for $\beta > 1$), whence

$$f_X(x) = \begin{cases} \Big(-\frac{1}{\beta} \sum_{j=0}^n \lambda_j h_j(x)\Big)^{\frac{1}{\beta-1}}, & \text{if } \sum_{j=0}^n \lambda_j h_j(x) \leq 0 \text{ or } \beta < 1, \\ 0, & \text{if } \sum_{j=0}^n \lambda_j h_j(x) > 0 \text{ and } \beta > 1, \end{cases}$$

or equivalently,

$$f_X(x) = \left[-\frac{1}{\beta} \sum\nolimits_{j=0}^n \lambda_j h_j(x)\right]_+^{\frac{1}{\beta-1}}.$$

In particular, for standard lower semideviation σ_-, the constraints $EX = \mu$ and $\sigma_-(X) = d$ correspond to (63) with $V = (-\infty, \infty)$, $h_1(X) = X$, $a_1 = \mu$, $h_2(X) = [X - \mu]_-^2$, and $a_2 = d^2$. In this case, a solution to (22) is determined by (28).

Acknowledgments. The authors are grateful to the anonymous referees for their valuable comments and suggestions, which helped to improve the quality of the paper.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, *Coherent Measures of Risk*, Mathematical Finance **9** (1999), 203–227.

[2] J. Bialas, Y. Nakamura, *The Theorem of Weierstrass*, Formalized Mathematics **5**(3) (1996), 353–359.

[3] J. Benoist, A. Daniilidis, *Coincidence Theorems for Convex Functions*, Journal of Convex Analysis **9**(1) (2002), 259–268.

[4] P. Buchen, M. Kelly, *The Maximum Entropy Distribution of an Asset Inferred from Option Prices*, Journal of Financial and Quantitative Analysis **31**(1) (1996), 143–159.

[5] D. C. Brody, R. C. Buckley Ian, I. C. Constantinou, *Option Price Calibration from Rényi Entropy*, Physics Letters A **366** (2007), 298–307.

[6] J. Costa, A. Hero, C. Vignat, *On solutions to multivariate maximum-entropy problems*, Lecture Notes in Computer Science (A. Rangarajan, M. Figueiredo, J. Zerubia, eds.), Springer-Verlag, Berlin, **2683**, 2003, pp. 211–228.

[7] J. M. Cozzolino, M. J. Zahner, *The Maximum-Entropy Distribution of the Future Market Price of a Stock*, Operations Research **21**(6) (1973), 1200–1211.

[8] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 1st ed., Wiley, New York, 1991.

[9] R.-A. Dana, *A representation result for concave Schur-concave functions*, Mathematical Finance **15**(4) (2005), 613–634.

[10] H. Follmer, A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, 2nd ed. de Gruyter, Berlin, 2004.

[11] C. Friedman, J. Huang, and S. Sandow, *A Utility-Based Approach to Some Information Measures*, Entropy **9** (2007), 1–6.

[12] E. T. Jaynes, *Prior Probabilities*, IEEE Transactions on Systems Science and Cybernetics **4** (1968), 227–251.

[13] E. T. Jaynes, *Information Theory and Statistical Mechanics*, Physical Review **106**(4) (1957), 620–630.

[14] O. Johnson, C. Vignat, *Some results concerning maximum Rényi entropy distributions*, Annales de l'Institute Henri Poincare (B) Probability and Statistics, **43**(3) (2007), 339–351.

[15] E. Jouini, W. Schachermayer, N. Touzi, *Law invariant risk measures have the Fatou Property*, Advances in Mathematical Economics **9** (2006), 49–71.

[16] S. Kusuoka, *On law invariant coherent risk measures*, Advances in Mathematical Economics **3** (2001), 83–95.

[17] E. Lutwak, D. Yang, G. Zhang, *Cramer-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information*, IEEE Transactions on Information Theory **51** (2005), 473–478.

[18] H. M. Markowitz, *Portfolio selection*, The Journal of Finance **7**(1) (1952), 77–91.

[19] A. Rényi, *On measures of information and entropy*, Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, University of California Press, Berkeley, CA (1961), 547–561.

[20] R. T. Rockafellar, S. Uryasev, M. Zabarankin, *Deviation Measures in Risk Analysis and Optimization*, Research Report 2002-7, Dept. of Industrial and Systems Engineering, University of Florida, Gainesville, 2002.

[21] R. T. Rockafellar, S. Uryasev, M. Zabarankin, *Generalized deviations in risk analysis*, Finance and Stochastics **10**(1) (2006), 51–74.

[22] R. T. Rockafellar, S. Uryasev, M. Zabarankin, *Optimality Conditions in Portfolio Analysis with General Deviation Measures*, Mathematical Programming **108**(2–3) (2006), 515–540.

[23] R.T. Rockafellar, S. Uryasev, M. Zabarankin, *Master Funds in Portfolio Analysis with General Deviation Measures*, The Journal of Banking and Finance **30**(2) (2006), 743–777.

[24] R. T. Rockafellar, S. Uryasev, M. Zabarankin, *Equilibrium with Investors Using a Diversity of Deviation Measures*, The Journal of Banking and Finance **31**(11) (2007), 3251–3268.

[25] R. T. Rockafellar, S. Uryasev, M. Zabarankin, *Risk Tuning with Generalized Linear Regression*, Mathematics of Operations Research **33**(3) (2008), 712–729.

[26] B. Rudloff, *Hedging in incomplete markets and testing compound hypotheses via convex duality*, Dissertation, Martin-Luther University, Halle-Wittenberg, Germany, 2006.

[27] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), 379–423, 623–656.

[28] M. Stutzer, *A Simple Nonparametric Approach to Derivative Security Valuation*, Journal of Finance **51**(5) (1996), 1633–1652.

[29] M. U. Thomas, *A Generalized Maximum Entropy Principle*, Operations Research **27**(6) (1979), 1188–1196.

[30] S. Willard, *General Topology*, Courier Dover Publications, Mineola, NY, 2004.