Information Geometry in Mathematical Finance: Model Risk, Worst and Almost Worst Scenarios

Thomas Breuer
PPE Research Centre
FH Vorarlberg
A-6850 Dornbirn, Austria
Email: tb@fhv.at

Imre Csiszár
A. Rényi Institute of Mathematics
Hungarian Academy of Sciences
H-1364 Budapest, P.O.Box 127, Hungary
Email: csiszar@renyi.hu

***Abstract*—The mathematical problem addressed is minimising the expectation of a random variable over a set of feasible distributions $\mathbb{P} \in \Gamma$, given as a level set of a convex integral functional. As special cases, Γ may be an I-divergence or f-divergence ball or a Bregman ball around a default distribution. Our approach is motivated by geometric intuition and relies upon the theory of minimising convex integral functionals subject to moment constraints. One main result is that all "almost minimisers" $\mathbb{P} \in \Gamma$ belong to a small Bregman ball around a specified distribution or defective distribution $\overline{\mathbb{P}}$, equal to the strict minimiser if that exists but well defined also otherwise.**

I. Introduction

Assume a random variable X describes the utility of some portfolio/action/choice, depending on the realisation of some risk factors $r \in \Omega$. When the risk factor distribution is not known exactly but may be assumed to belong to some set Γ of distributions on Ω, the worst case expected utility

$$\inf_{\mathbb{P}\in\Gamma} E_{\mathbb{P}}(X) \tag{1}$$

may be taken as (the negative of a) measure of model risk caused by the lack of knowledge about $\mathbb{P}$. Other kinds of risk are also considered in the literature but it is a basic result that any risk measure satisfying some natural postulates ("coherence") can be represented by the negative of (1) for some convex set Γ of distributions on Ω. The same expression enters the theory of *preferences* of ambiguity averse decision makers. Such decision makers are averse to the uncertainty about the distribution $\mathbb{P}$ and prefer acts X with higher expected utilities in the worst case over Γ. (1) is their decision criterion. See [8], [11], [9] for details.

The axiomatic theory does not specify the set Γ but if a "best guess" $\mathbb{P}_0$ of the unknown risk factor distribution is available, it is natural to take Γ as those distributions that are "close" to $\mathbb{P}_0$, for example in I-divergence (relative entropy), f-divergence or Bregman distance from $\mathbb{P}_0$. To our knowlegde, I-divergence has been used for this purpose first by [10], other f-divergences (including "weighted" ones) by [12], while we have not met Bregman distances yet in this context.

This work has been supported by the Hungarian National Foundation for Scientific Research under Grant K76088 and by the Austrian Forschungsförderungsgesellschaft in its Josef Ressel Center for Optimisation under Uncertainty. The second author has benefited from discussions with František Matúš.

With this choice of Γ, Problem (1) is related to the information geometric problem of minimising a divergence as above of $\mathbb{P}_0$ from the manifold of distributions with $E_{\mathbb{P}}(X) = b$. In this paper we employ the information theoretical perspective to solve Problem (1), relying upon the results of [7]. We address the problem (1) for Γ equal to the level set of some convex integral functional H, which includes as special cases all of the above divergences, see Examples 1 and 2 in the next section. We also address a familiar relaxation of Problem (1),

$$W \triangleq \inf_{p:\int p d\mu=1} \left[\int X p d\mu + \lambda H(p)\right], \quad \lambda > 0. \tag{2}$$

Note that $-W$ is a convex risk measure, non-coherent in general. Decision makers with so-called *divergence preferences* rank alternatives X according to this criterion.

The main new result of this paper is a characterisation of "almost worst case" distributions, which almost achieve the infimum in (1), whether or not the infimum is attained. From a practical point of view the localisation of almost worst case distributions determines the efficiency of a hedge against the worst case distribution. To our knowledge this question has not been addressed in the literature. Additionally, we generalise previous results [1] about the evaluation of (1) and (2).

II. Preliminaries

A. Geneneral Framework

Let Ω be any set equipped with a (finite or σ-finite) measure μ on a σ-algebra not mentioned in the sequel. Let H be a convex integral functional defined for (measurable) non-negative functions p on Ω by

$$H(p) = H_\beta(p) \triangleq \int_\Omega \beta(r, p(r))\mu(dr), \tag{3}$$

determined by a function $\beta(r, s)$ of $r \in \Omega, s \in \mathbb{R}$, measurable in r for each $s \in \mathbb{R}$, strictly convex and differentiable in s on $(0, +\infty)$ for each $r \in \Omega$, and satisfying

$$\beta(r, 0) = \lim_{s\downarrow 0} \beta(r, s), \quad \beta(r, s) \triangleq +\infty \text{ if } s < 0. \tag{4}$$

Then β is a convex normal integrand [14], which ensures the measurability of $\beta(r, p(r))$ in (3) and of similar functions later on. A necessary condition for $H(p) < \infty$ is $p \geq 0$ a.e..

Let X be a real-valued measurable function and $\mathbb{P}_0$ a default distribution on Ω with $\mathbb{P}_0 \ll \mu$, $d\mathbb{P}_0/d\mu = p_0$, such that the expectation $E_{\mathbb{P}_0}(X) = \int_\Omega X(r)p_0(r)d\mu(r) \triangleq b_0$ exists. Let m and M denote the μ-ess inf and μ-ess sup of X, and adopt as standing assumptions

$$-\infty \leq m \quad < \quad b_0 < M \leq +\infty \tag{5}$$

$$H(p) \quad \geq \quad H(p_0) = 0 \qquad \text{whenever} \int p d\mu = 1. \tag{6}$$

Problem (1) will be considered for

$$\Gamma \triangleq \{\mathbb{P} : d\mathbb{P} = pd\mu, H(p) \leq k\}. \tag{7}$$

Example 1. Take $\mu = \mathbb{P}_0$, thus $p_0 = 1$, and let $\beta(r, s) = f(s)$ be an autonomous convex integrand, with $f(s) \geq f(1) = 0$. Then $H(p)$ in (3) is the *f-divergence* $D_f(\mathbb{P}\,||\,\mathbb{P}_0)$ [6]. If f is cofinite, i.e. if $\lim_{s\to+\infty} f(s)/s = +\infty$, then $\mathbb{P} \ll \mathbb{P}_0$ is a necessary condition for $D_f(\mathbb{P}\,||\,\mathbb{P}_0) < +\infty$, hence in that case Γ equals the f-divergence ball $\{\mathbb{P} : D_f(\mathbb{P}\,||\,\mathbb{P}_0) \leq k\}$.

Example 2. Let f be a strictly convex differentiable function on $(0, +\infty)$, and for $s \geq 0$ let $\beta(r, s) = \Delta_f(s, p_0(r))$ where

$$\Delta_f(s, t) \triangleq f(s) - f(t) - f'(t)(s - t); \tag{8}$$

in case $f'(0) = -\infty$ assume that $p_0 > 0$ μ-a.e. Then $H(p)$ in (7) equals the *Bregman distance* [3]

$$B_{f,\mu}(p, p_0) \triangleq \int_\Omega \Delta_f(p(r), p_0(r)\mu(dr), \tag{9}$$

and Γ is a Bregman ball of radius k around $\mathbb{P}_0$.

In the special case $f(s) = s\log s - s + 1$, both examples give the I-divergence ball $\Gamma = \{\mathbb{P} : D(\mathbb{P}\,||\,\mathbb{P}_0) \leq k\}$.

B. Basic approach

For Γ in (7) the infimum in (1) equals

$$V(k) \triangleq \inf_{p:\int pd\mu=1, H(p)\leq k} \int Xpd\mu. \tag{10}$$

To address the minimisation problem (10) consider the related "moment problem"

$$F(b) \triangleq \inf_{p:\int pd\mu=1, \int Xpd\mu=b} H(p). \tag{11}$$

Problems (10) and (11) are closely related. What is the objective function in one problem is the constraint in the other, and vice versa (see Fig. 1). The following lemma makes this relation explicit.

Lemma 1. *If* $0 < k < k_{\max} \triangleq \lim_{b\downarrow m} F(b)$, *there exists a unique* b *satisfying*

$$F(b) = k, \qquad m < b < b_0 \tag{12}$$

and then $V(k) = b$. *The minimum in* (10) *is attained if and only if that in* (11) *is attained (for the* b *in* (12)*), and then the same* p *attains both minima.*

Proof: The convex function F attains its minimum 0 at b_0, see (6). This easily implies the assertion, using a technical detail given in Corollary 3. For details see [4]. ∎

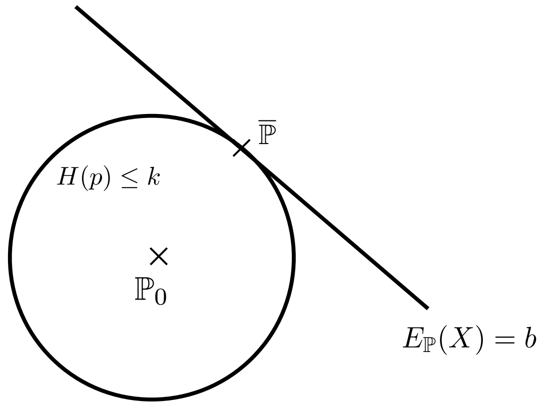

Fig. 1. **Geometric view of Lemma 1.**

Remark 1. The condition $0 < k < k_{\max}$ is not restrictive, for if $k = 0$ or $k \geq k_{\max} > 0$ then trivially $V(k) = b_0$ resp. $V(k) = m$. If $k_{\max} = 0$ then the functional H is unsuitable to define a non-trivial measure of risk.

Lemma 1 admits to treat Problem (10) using known results about minimising convex integral functionals under moment constraints, specifically with moment mapping defined by $\phi(r) \triangleq (1, X(r))$. We will rely upon results[1] in [7], specified for this moment mapping. Then the value function in [7] becomes

$$J(a, b) \triangleq \inf_{p:\int pd\mu=a, \int Xpd\mu=b} H(p), \tag{13}$$

thus $F(b) = J(1, b)$.

C. Basic concepts and facts

The convex conjugate $J^*(\theta_1, \theta_2) \triangleq \sup_{a,b}[\theta_1 a + \theta_2 b - J(a, b)]$ of J, by [7, Theorem 1.1], equals

$$K(\theta_1, \theta_2) \triangleq \int \beta^*(r, \theta_1 + \theta_2 X(r))\mu(dr), \tag{14}$$

where β^* is the convex conjugate of β,

$$\beta^*(r, \tau) \triangleq \sup_{s\in\mathbb{R}} (s\tau - \beta(r, s)). \tag{15}$$

(Both the conjugate and the derivative of β are by the second variable.) For fixed $r \in \Omega$ the function β^* equals $-\beta(r, 0)$ for $\tau \leq \beta'(r, 0)$, it is strictly convex in the interval $(\beta'(r, 0), \beta'(r, +\infty))$, and equals $+\infty$ if $\beta'(r, +\infty)$ is finite and $\tau > \beta'(r, +\infty)$. For $\tau < \beta'(r, +\infty)$ the derivative $(\beta^*)'(r, \tau)$ exists, it is positive if $\tau > \beta'(r, 0)$, and increases to $+\infty$ as $\tau \uparrow \beta'(r, +\infty)$.

The following family of functions plays a similar role as exponential families do for the I-divergence minimisation:

$$p_{\theta_1,\theta_2}(r) \triangleq (\beta^*)'(r, \theta_1 + \theta_2 X(r)), \quad (\theta_1, \theta_2) \in \Theta \tag{16}$$

[1]Many of these results have been known earlier, though typically under less general conditions.

where, with $\operatorname{dom} K \triangleq \{(\theta_1,\theta_2) : K(\theta_1,\theta_2) < +\infty\}$,

$$\Theta \triangleq \{(\theta_1,\theta_2) \in \operatorname{dom} K : \theta_1 + \theta_2 X(r) < \beta'(r,+\infty)\ \mu\text{-a.e.}\}\,.$$

As K in (14) is the convex conjugate of J, it is a lower semicontinuous proper convex function, and

$$J(a,b) = K^*(a,b) = \sup_{(\theta_1,\theta_2)\in\mathbb{R}^2} \left(\theta_1 a + \theta_2 b - K(\theta_1,\theta_2)\right), \tag{17}$$

if (a,b) is in the interior of $\operatorname{dom} J$, or equivalently [7, Lemma 6.6], if $am < b < aM$. By [7, Corollary 3.8], K is differentiable at each $\theta = (\theta_1,\theta_2) \in \operatorname{int\,dom}\ K$, with

$$\nabla K(\theta_1,\theta_2) = \left(\int p_{\theta_1,\theta_2} d\mu, \int X p_{\theta_1,\theta_2} d\mu\right). \tag{18}$$

III. Main results

A. Worst Scenarios: Determining $V(k)$

Theorem 1. *Assuming* (5), (6), *and* $0 < k < k_{\max}$, *if*

$$\overline{\theta}_2 < 0, \qquad \int p_{\overline{\theta}_1,\overline{\theta}_2}\, d\mu = 1, \tag{19}$$

$$\overline{\theta}_1 + \overline{\theta}_2 \int X p_{\overline{\theta}_1,\overline{\theta}_2}\, d\mu - K(\overline{\theta}_1,\overline{\theta}_2) = k \tag{20}$$

for some $(\overline{\theta}_1,\overline{\theta}_2) \in \Theta$ *then the value of the inf in* (10) *is*

$$V(k) = \int X p_{\overline{\theta}_1,\overline{\theta}_2}\, d\mu. \tag{21}$$

Essential smoothness[2] *of K is a sufficient condition for the existence of such $(\overline{\theta}_1,\overline{\theta}_2)$. Further, a necessary and sufficient condition for p to attain the minimum in* (10) *is $p = p_{\overline{\theta}_1,\overline{\theta}_2}$ for some $(\overline{\theta}_1,\overline{\theta}_2) \in \Theta$ satisfying* (19) *and* (20).

Corollary 1. *If the equations*

$$\frac{\partial}{\partial\theta_1} K(\theta_1,\theta_2) = 1, \tag{22}$$

$$\theta_1 + \theta_2 \frac{\partial}{\partial\theta_2} K(\theta_1,\theta_2) - K(\theta_1,\theta_2) = k \tag{23}$$

have a solution $(\overline{\theta}_1,\overline{\theta}_2) \in$ int dom K with $\overline{\theta}_2 < 0$ then $\overline{\theta}_1,\overline{\theta}_2$ satisfy (19) *and* (20), *and the solution to Problem* (10) *equals*

$$V(k) = \left.\frac{\partial K(\theta_1,\theta_2)}{\partial\theta_2}\right|_{(\theta_1,\theta_2)=(\overline{\theta}_1,\overline{\theta}_2)}. \tag{24}$$

The Corollary follows from the Theorem by (18). However, if K is not essentially smooth, $(\overline{\theta}_1,\overline{\theta}_2) \in$ int dom K is not a necessary condition for (19), (20).

Proof: We sketch the proof of Theorem 1, referring to [4] for details. By [7, Lemmata 4.4, 4.10], if $(\overline{\theta}_1,\overline{\theta}_2) \in \Theta$ satisfies

$$\int p_{\theta_1,\theta_2} d\mu = 1, \int X p_{\theta_1,\theta_2} d\mu = b \tag{25}$$

then $\overline{\theta}_1,\overline{\theta}_2$ attains the supremum in the definition (17) of $K^*(a,b)$. Moreover, the existence of $(\overline{\theta}_1,\overline{\theta}_2) \in \Theta$ satisfying (25) is necessary and sufficient for the attainment of the minimum in (13) if (a,b) is in the interior of $\operatorname{dom} J$, and then $p = p_{\overline{\theta}_1,\overline{\theta}_2}$ is the unique minimiser. Apply this to $a = 1$, $b = \int X p_{\overline{\theta}_1,\overline{\theta}_2} d\mu$ with $\overline{\theta}_1,\overline{\theta}_2$ satisfying (19) and (20); it follows using (17) that $F(b) = k$, and Lemma 1 gives the first and last assertions of Theorem 1. For this, one checks that $m < b < b_0$. When K is essentially smooth, the gradient vectors $\nabla K(\theta_1,\theta_2)$, $(\theta_1,\theta_2) \in$ int dom K, cover int dom $K^* =$ int dom J, see [13, Corollary 26.4.1], hence to b in (12) there exists a $(\overline{\theta}_1,\overline{\theta}_2) \in$ int dom J with $\nabla K(\overline{\theta}_1,\overline{\theta}_2) = (1,b)$. By (18), then $(\overline{\theta}_1,\overline{\theta}_2)$ satisfies (19). ■

Remark 2. If a $\overline{\theta}_2 > 0$ satisfies (20), $p_{\overline{\theta}_1,\overline{\theta}_2}$ has $b = \int X p d\mu > b_0$ rather than (12), and attains $\sup_{\mathbb{P}\in\Gamma} E_{\mathbb{P}}(X)$ rather than (1).

In Lemma 1 and Theorem 1 the condition $k_{\max} > 0$ has been essential. We now give a necessary and sufficient condition for this to hold.

Theorem 2. *A necessary and sufficient condition for $k_{\max} > 0$ is the existence of a $(\theta_1,\theta_2) \in \operatorname{dom} K$ with $\theta_2 < 0$. Sufficient conditions are $m > -\infty$ or essential smoothness of K.*

Proof: Suppose $F(b) > 0$ for some $m < b < b_0$. Then $(1,b) \in$ int dom J, which implies that there exists $(\overline{\theta}_1,\overline{\theta}_2)$ in $\operatorname{dom} K$ such that

$$F(b) = J(1,b) = \overline{\theta}_1 + \overline{\theta}_2 b - K(\overline{\theta}_1,\overline{\theta}_2). \tag{26}$$

Indeed, if $(\overline{\theta}_1,\overline{\theta}_2)$ belongs to the subdifferential of the convex function J at $(1,b)$, which is nonempty by [13, Theorem 23.4], then (17) implies by [13, Theorem 23.5] that $(\overline{\theta}_1,\overline{\theta}_2)$ satisfies (26). It follows from (17) and (26) that for each $t \in (m,M)$

$$F(t) = K^*(1,t) \geq \overline{\theta}_1 + \overline{\theta}_2 t - K(\overline{\theta}_1,\overline{\theta}_2) = F(b) + \overline{\theta}_2(t-b).$$

Substituting $t = b_0$, this proves $\overline{\theta}_2 < 0$, thus the necessity part of the Theorem. For the sufficiency part, see [4]. ■

Corollary 2. *In the setting of Example 1, if f is not cofinite then the necessary and sufficient condition for $k_{\max} > 0$ is $m > -\infty$*

Proof: Clearly, $f^*(\theta_1+\theta_2 X) < +\infty$ μ-a.e. is a necessary condition for $(\theta_1,\theta_2) \in \operatorname{dom} K$. In case $\theta_2 < 0$ this can hold only if $\theta_1 + \theta_2 m \leq \lim_{s\to+\infty} f(s)/s$. It follows that if f is not cofinite then the sufficient condition $m > -\infty$ for $k_{\max} > 0$ is necessary as well. ■

Corollary 3. *If $k_{\max} > 0$ then $\int p d\mu = 1$, $H(p) < +\infty$ imply that $\int X p d\mu$ exists and does not equal $-\infty$.*

Proof: Substitute in the Fenchel inequality $xs \leq \beta(r,s) + \beta^*(r,x)$ (a consequence of (15)) $x := \theta_1 + \theta_2 X(r)$, $s := p(r)$ and integrate. It follows that if $(\theta_1,\theta_2) \in \operatorname{dom} K$ and p satisfies the hypotheses then

$$\theta_1 + \theta_2 \int X p d\mu \leq H(p) + K(\theta_1,\theta_2) < +\infty.$$

Taking (θ_1,θ_2) with $\theta_2 < 0$, the assertion follows. ■

[2]A lower semicontinuous proper convex function is essentially smooth if its effective domain has nonempty interior, the function is differentiable there, and at non-interior points of the effective domain the directional derivatives in directions towards the interior are $-\infty$. The latter trivially holds if the effective domain is open.

B. Almost worst scenarios

For a specified $\epsilon > 0$ define an almost worst scenario to be a density p satisfying $H(p) \leq k$ and achieving $\int Xpd\mu < V(k) + \epsilon$. Our third main result needs a concept of Bregman distance slightly more general than (9). Define $\Delta_{\beta(r,\cdot)}(s,t)$ as in (8), with the convex function $\beta(r,\cdot) : s \mapsto \beta(r,s)$ playing the role of f. The mapping $(r,s,t) \mapsto \Delta_{\beta(r,\cdot)}(s,t)$ is a normal integrand [7, Lemma 2.10], hence if p and q are non-negative measurable functions on Ω then so is also $\Delta_{\beta(r,\cdot)}(p(r),q(r))$, denoted briefly by $\Delta_\beta(p,q)$. The required extended concept of Bregman distance is defined by

$$B_{\beta,\mu}(p,q) \triangleq \int \Delta_\beta(p,q)d\mu. \tag{27}$$

Like its special case in (9), it is non-negative and equals 0 if and only if $p = q$, μ-a.e. Moreover, if $B_{\beta,\mu}(p_n,q) \to 0$ for a sequence of functions p_n then this sequence converges to q locally in measure[3] [7, Corollary 2.14]. This implication holds for any integrand β meeting the assumptions in the Preliminaries, but for certain choices of β more is true, e.g., $D(p_n||q) \to 0$ implies $p_n \to q$ in $L_1(\mu)$.

Theorem 3. *Supposing* $0 < k < k_{\max}$, *there exists* $(\overline{\theta}_1,\overline{\theta}_2) \in \Theta$ *with* $\overline{\theta}_2 < 0$ *such that for each* p *with* $\int pd\mu = 1$

$$B_{\beta,\mu}(p,p_{\overline{\theta}_1,\overline{\theta}_2}) \leq H(p) - k - \overline{\theta}_2\left[\int Xpd\mu - V(k)\right]. \tag{28}$$

Corollary 4. *If* $\{p_n\}$ *is any sequence of functions satisfying the constraints in* (10) *such that* $\int Xp_nd\mu \to V(k)$ *then* p_n *converges to* $p_{\overline{\theta}_1,\overline{\theta}_2}$ *locally in measure.*

Note that the function $p_{\overline{\theta}_1,\overline{\theta}_2}$ here does not necessarily determine a probability distribution. For some choices of the integrand β, the integral of this $p_{\overline{\theta}_1,\overline{\theta}_2}$ may be less than 1, thus it determines a *defective distribution*.

Proof: The assertion is an easy consequence of the inequality

$$H(p) \geq \theta_1 + \theta_2 t - K(\theta_1,\theta_2) + B_{\beta,\mu}(p,p_{\theta_1,\theta_2}) \tag{29}$$

valid for each p with $\int pd\mu = 1$, $\int Xpd\mu = t$ and any $(\theta_1,\theta_2) \in \Theta$. (29) follows from a general version of the Pythagorean inequality [7, Lemma 4.15], applied to $(a_1,a_2) = (1,t)$. (There, actually, an equality is given, with a specified error term added to the right hand side, which vanishes in some important cases.) Indeed, consider $b \in (m,b_0)$ satisfying $F(b) = k$ as in Lemma 1, and take $(\overline{\theta}_1,\overline{\theta}_2) \in \Theta$ with $\overline{\theta}_2 < 0$ such that

$$k = F(b) = \overline{\theta}_1 + \overline{\theta}_2 b - K(\overline{\theta}_1,\overline{\theta}_2); \tag{30}$$

for its existence, see the proof of Theorem 2. Since $b = V(k)$ by Lemma 1, combining (29) and (30) gives rise to (28). ■

Theorem 3 describes a clustering property of almost worst scenarios in the Bregman neighbourhood of $p_{\overline{\theta}_1,\overline{\theta}_2}$. This clustering property holds both in the case where a worst case density exists (in which case it is equal to $p_{\overline{\theta}_1,\overline{\theta}_2}$), and also in case the infimum in (10) is not attained. Corollary 4 describes the same clustering in neighbourhoods "in measure". Theorem 3 also describes clustering of distributions which slightly violate the constraint $H(p) \leq k$.

[3]This means that for each $C \subset \Omega$ with $\mu(C)$ finite, and any $\epsilon > 0$, $\mu(\{r \in C : |p_n(r) - q(r)| > \epsilon\}) \to 0$. If μ is a finite measure, this is equivalent to standard (global) convergence in measure.

Example 3. Let us briefly check how the unified framework leads, in the special case of sufficiently small I-divergence balls, to the results of [5, Theorem 1]. Set $\mu = \mathbb{P}_0$, $\beta(r,s) = s\log s - s + 1$, then $H(p) = D(p||p_0)$, $B(p,q) = D(p||q)$,

$$K(\theta_1,\theta_2) = \exp[\theta_1+\Lambda(\theta_2)]-1, \quad \Lambda(\theta) \triangleq \log\int\exp(\theta X)d\mu.$$

The functions p_{θ_1,θ_2} of (16) are of form $\exp(\theta_1+\theta_2X(r))$, and integrate to 1 if and only if $\theta_1 = -\Lambda(\theta_2)$. With this constraint the two variables θ_1,θ_2 can be replaced by the single variable θ_2, for which we write θ. The p_{θ_1,θ_2} which integrate to 1 form the exponential family $\exp[\theta X - \Lambda(\theta)]$. From (17),

$$\begin{aligned} F(b) = K^*(1,b) &= \sup\nolimits_{\theta_1,\theta_2}[\theta_1 + \theta_2 b - \exp(\theta_1 + \Lambda(\theta_2)) + 1] \\ &= \sup\nolimits_\theta[-\Lambda(\theta) + \theta b] = \Lambda^*(b), \end{aligned}$$

if $m < b < M$. By Lemma 1, for $0 < k < k_{\max}$, $V(k)$ equals the unique b in (m,b_0) satisfying $F(b) = k$. For this b, the above maximum is attained by $\overline{\theta} < 0$ satisfying $\Lambda'(\overline{\theta}) = b$, i.e., by the negative solution of the equation

$$-\Lambda(\theta) + \theta\Lambda'(\theta) = k, \tag{31}$$

if it exists. With this $\overline{\theta}$ we recover [5, Theorem 1] that $V(k) = \Lambda'(\overline{\theta})$ and the distribution $\overline{\mathbb{P}}$ with density $\exp[\overline{\theta}X - \Lambda(\overline{\theta})]$ is the worst case distribution. If no $\theta < 0$ satisfies (31), which happens if $\operatorname{dom}\Lambda$ contains its left endpoint $\theta_{\min}$ and $k > -\Lambda(\theta_{\min})+\theta_{\min}\Lambda'(\theta_{\min})$, then $\overline{\theta} = \theta_{\min}$ attains the maximum in the representation of $\Lambda^*(b)$. In this case, no worst case distribution exists, i.e., in (10) the minimum is not attained. In either case, all ϵ-worst distributions, i.e. those in the I-divergence ball of radius k around $\mathbb{P}_0$ with $E_{\mathbb{P}}(X) \leq V(k)+\epsilon$, cluster in the I-divergence neighborhood

$$\{\mathbb{P} : D(\mathbb{P}||\overline{\mathbb{P}}) \leq -\overline{\theta}\epsilon\}$$

of $\overline{\mathbb{P}}$, by Theorem 3.

IV. Divergence preferences

Finally, we briefly address the problem (2), related to (10) but mathematically simpler. Note that $-W$, though not a coherent risk measure in general, belongs to the larger class of convex risk measures. Below, we consider (2) with any convex integral functional (3) in the role of H. Typically, one takes the I-divergence or an f-divergence of $\mathbb{P}$ from $\mathbb{P}_0$, while [12] proposed weighted f-divergence, corresponding to the choice $\beta(r,s) = w(r)f(s)$ in (3), where w is a normalised, non-negative weight function.

Define a new convex integrand and integral functional by

$$\tilde{\beta}(r,s) \triangleq X(r)s + \lambda\beta(r,s), \quad \tilde{H}(p) \triangleq \int\tilde{\beta}(r,p(r))d\mu(r),$$

(where $\lambda > 0$ is fixed), then

$$W = \inf_{p: \int p d\mu = 1} \left[\int X p d\mu + \lambda H(p) \right] = \inf_{p: \int p d\mu = 1} \tilde{H}(p). \tag{32}$$

Thus, the problem is to minimize the functional $\tilde{H}(p)$ under the single constraint $\int p d\mu = 1$.

In analogy to (13), consider

$$\tilde{J}(a) := \inf_{p: \int p d\mu = a} \tilde{H}(p), \quad a \in \mathbb{R}.$$

Note that $\tilde{\beta}$ meets the basic assumptions on β (though (6) does not hold for $\tilde{H}$), and that

$$\begin{aligned} (\tilde{\beta})^*(r, x) &= \sup_s [xs - X(r)s - \lambda\beta(r, s)] \\ &= \lambda\beta^* \left(r, \frac{x - X(r)}{\lambda} \right). \end{aligned}$$

It follows by [7, Theorem 1.1] that the convex conjugate of $\tilde{J}$ equals

$$\tilde{K}(\theta) := \int (\tilde{\beta})^*(r, \theta) d\mu(r) = \lambda \int \beta^* \left(r, \frac{\theta - X(r)}{\lambda} \right) d\mu(r),$$

or, with the notation (14),

$$\tilde{J}^*(\theta) = \tilde{K}(\theta) = \lambda K(\frac{\theta}{\lambda}, -\frac{1}{\lambda}), \quad \theta \in \mathbb{R}.$$

As the interior of $\mathrm{dom}\, \tilde{J}$ is $(0, +\infty)$, it follows that $\tilde{J}(a) = \tilde{K}^*(a)$ for each $a > 0$. In particular,

$$\begin{aligned} W = \tilde{J}(1) &= \tilde{K}^*(1) = \sup_{\theta \in \mathbb{R}} (\theta - \tilde{K}(\theta)) \\ &= \sup_{\theta \in \mathbb{R}} \left[\theta - \lambda K \left(\frac{\theta}{\lambda}, -\frac{1}{\lambda} \right) \right] \\ &= \lambda \sup_{\theta_1 \in \mathbb{R}} \left[\theta_1 - K \left(\theta_1, -\frac{1}{\lambda} \right) \right]. \end{aligned}$$

Proposition 1. *The necessary and sufficient condition for $W > -\infty$ in* (32) *is the existence of $\theta_1 \in \mathbb{R}$ with*

$$(\theta_1, -1/\lambda) \in \mathrm{dom}\, K, \tag{33}$$

and then

$$W = \lambda \sup_{\theta_1} [\theta_1 - K(\theta_1, -1/\lambda)]. \tag{34}$$

If to $\theta_2 = -1/\lambda$ there exists θ_1 with $(\theta_1, \theta_2) \in \Theta$ such that the function p_{θ_1,θ_2} in (16) *has integral equal to one, then θ_1 attains the maximum in* (34)*, and $p = p_{\theta_1,\theta_2}$ attains the minimum in* (32)*. Otherwise, among the numbers θ_1 satisfying* (33) *there exists a largest one $\theta_{1\,\max}$, and $p_{\theta_{1\,\max},-1/\lambda}$ has integral less than one; then $\theta_1 = \theta_{1\,\max}$ attains the maximum in* (34)*, while the infimum in* (32) *is not attained.*

Proof: Clearly, $W = \tilde{J}(1) > -\infty$ if and only if $\tilde{J}$ never equals $-\infty$, thus its conjugate $\tilde{K}$ is not identically $+\infty$; by the formula for $\tilde{K}$, this proves the first assertion. The second assertion follows from (33). As the supremum in (34) is the same as the supremum defining $\tilde{K}^*(1)$ in (33) (with θ/λ substituted by θ_1), the next assertion follows from a simple instance of [7, Lemma 4.10](note that the function $(\beta^*)'(r, \theta)$ in that Lemma, replacing β by $\tilde{\beta}$ and θ by $\theta_1\lambda$, gives the function p_θ in the Proposition). For the last assertion, recall that the maximum in the definition of $\tilde{K}^*(1)$, and therefore in (34), is always attained, because $a = 1$ is in the interior of $\mathrm{dom}\, \tilde{K}^*$. Then the (left) derivative by θ_1 of $K(\theta_1, -1/\lambda)$ at the maximiser, say θ_1^*, has to be ≤ 1, and the strict inequality can hold only if $\theta^* = \theta_{1\,\max}$. As the mentioned derivative equals the integral of p_{θ^*} with $\theta^* = (\theta_1^*, -1/\lambda)$, this completes the proof. ■

Remark 3. Proposition 1 extends [2, Theorem 4.2] about Problem (2) with an f-divergence penalty term, i.e., in our terminology, when β is an autonomous integrand. As shown in [1], that result of [2] admits to represent also the solution of Problem (10) by a convex optimization formula involving two real variables (or one in the I-divergence case). In a similar manner, Proposition 1 can be used to extend this result of [1] to non-autonomous integral functionals. As a technical point, note that in the mentioned references X has been assumed bounded, which assumption is not needed here.

REFERENCES

[1] Amir Ahmadi-Javid. Entropic Value at Risk: a new coherent risk measure. *Journal of Optimizaton Theory and Applications*, 155(3):1105–1123, 2011.

[2] Aharanov Ben-Tal and Marc Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17:449–476, 2007.

[3] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[4] Thomas Breuer and Imre Csiszár. Measuring Model Risk. *ArXiv e-prints*, January 2013. http://adsabs.harvard.edu/abs/2013arXiv1301.4832B.

[5] Thomas Breuer and Imre Csiszár. Systematic stress tests with entropic plausibility constraints. *Journal of Banking and Finance*, 37:1552–1559, 2013.

[6] Imre Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 8:85–108, 1963.

[7] Imre Csiszár and František Matúš. Generalized minimizers of convex integral functionals, Bregman distance, Pythagorean identities. *Kybernetika*, 48:637–689, 2012.

[8] Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter, 2nd edition, 2004.

[9] Itzhak Gilboa. *Theory of Decision under Uncertainty*, volume 45. Cambridge University Press, 2009.

[10] Lars Peter Hansen and Thomas Sargent. Robust control and model uncertainty. *American Economic Review*, 91:60–66, 2001.

[11] Lars Peter Hansen and Thomas Sargent. *Robustness*. Princeton University Press, 2008.

[12] Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74:1447–1498, 2006.

[13] Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[14] Ralph Tyrell Rockafellar and Roger J-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 1997.