Information Geometry in Multiple Prior Models Worst Case and Almost Worst Case Distributions

Imre Csiszár[1] Thomas Breuer[2]

[1]Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Hungary

[2]PPE Research Centre, FH Vorarlberg, Austria

Bachelier Finance Society, World Congress
New York, 16 July 2016
Dedicated to Shin-ichi Amari for his 80th birthday

Summary

Problem arising in model risk measurement stress testing:

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

where set Γ of prior distributions is defined in terms of some generalised entropy functional.

Questions addressed here:

- (Risk measurement: What is the inf?)
- Construct hedges: Where is the worst case distribution, if it exists? What if the inf is not achieved?
- Efficiency of hedges: How good are hedges if not worst case happens but almost worst case?

How to chose the plausible set Γ

Intuition: the plausible distributions are those not deviating much from a default distribution $\mathbb{P}_0$, resulting e.g. by estimation from historical data.
Of many measures of deviation of distributions, I-divergence (relative entropy) appears most versatile. Familiar larger classes of measures of deviation are f-divergences and Bregman distances.

We consider plausible sets consisting of distributions dominated by a given (σ-finite) measure μ, of form

$$\Gamma = \{\mathbb{P} : d\mathbb{P} = p d\mu, \ H(p) \leq k\} \tag{2}$$

where H is an entropy functional (convex integaral functional).
I-divergence balls, and general f-divergence or Bregman distance balls around a default distribution arise by specific choices of H.

Formal definitions

σ-finite measure space $(\Omega, \mathcal{F}, \mu)$, nonnegative measurable functions denoted by p or q, equality $p = q$ in μ-a.e. sense.

Denote by $\mathbb{B}$ the class of functions $\beta(\omega, s)$ on $\Omega \times \mathbb{R}$ that are

- for each $s \in \mathbb{R}$, measurable in ω
- for each $\omega \in \Omega$, strictly convex, differentiable in s on $(0, +\infty)$, equal to $+\infty$ if $s < 0$, and $\beta(\omega, 0) = \lim_{s \downarrow 0} \beta(\omega, s)$.

For $\beta \in \mathbb{B}$ define the entropy functional, for functions $p \geq 0$, by

$$H(p) = H_\beta(p) \triangleq \int_\Omega \beta(\omega, p(\omega))\mu(d\omega). \tag{3}$$

The functions $\beta \in \mathbb{B}$ are convex normal integrands, hence $\beta(\omega, p(\omega))$ and similar functions later on are measurable.

History sketch

Multiple prior models, risk measures, theory of preferences: Főllmer and Schied 2004, Hansen and Sargent 2008, Gilboa 2009.
Plausible sets Γ: I-divergence balls Hansen and Sargent 2001, Ahmadi-Javid 2011; f-divergence balls Maccheroni, Marinacci, Rustichini 2006, Ben Tal and Teboulle 2007.

Axiomatic approach leading to specific divergences: In inference context, leading to I-divergence, with f-divergences and Bregman distances as alternatives: Csiszár 1991. In mathematical finance, distinguishing I-divergence: Strzalecki 2011.

Moment problem: Geometric view goes back to Chentsov 1972, clustering of approximate solutions to Topsoe 1979, Csiszár 1984, convex duality approach to Borwein and Lewis 1991, 1993. General results relied upon in this talk: Csiszár and Matús 2012.

Basic framework used in this talk: Breuer and Csiszár 2013.

Special cases

- Let μ equal the default distribution $\mathbb{P}_0$, let $\beta(\omega, s) = f(s)$ be an autonomous convex integrand with $f(1) = 0$. Then $H(p)$ in (3) with $p = d\mathbb{P}/d\mu$ is the *f-divergence* $D_f(\mathbb{P}\,||\,\mathbb{P}_0) = \int f(\frac{d\mathbb{P}}{d\mathbb{P}_0})d\mathbb{P}_0$.
- Let μ and the default distribution $\mathbb{P}_0 \ll \mu$ be arbitrary, f a strictly convex differentiable function on $(0, +\infty)$, and for $s \geq 0$ let $\beta(\omega, s) = \Delta_f(s, p_0(\omega))$ where

$$\Delta_f(s,t) \triangleq f(s) - f(t) - f'(t)(s-t); \qquad (4)$$

if f is steep ($f'(0) = -\infty$), assume that $p_0 > 0$ μ-a.e. Then $H(p)$ equals the Bregman distance $B_f(p, p_0)$.
- In the special case $f(s) = s \log s$, both examples above give for $H(p)$ in (3) with $p = d\mathbb{P}/d\mu$ the *I*-divergence $D(\mathbb{P}||\mathbb{P}_0) = \int p \log \frac{p}{p_0} d\mu$, thus for the plausible set Γ in (2) the *I*-divergence ball $\{\mathbb{P} : D(\mathbb{P}||\mathbb{P}_0) \leq k\}$.

Goal, tools

Goal: Determine

$$V(k) \triangleq \inf_{p: \int p d\mu = 1, H(p) \leq k} \int X p d\mu. \quad (5)$$

Main contribution: Almost Worst Case localisation theorem: Whether or not a worst case distribution minimising (5) exists, the almost worst case distributions cluster around an explicitly specified, perhaps incomplete distribution, which we call Almost Worst Case localiser.

Moment problem (General Maximum Entropy Problem):

$$J(a, b) \triangleq \inf_{p: \int p d\mu = a, \int X p d\mu = b} H(p). \quad (6)$$

Special choice: $F(b) \triangleq J(1, b)$.

We will use results available on the moment problem, taken from Csiszár and Matúš 2012, and convex duality tools.

Standing Assumptions

- X is a real valued measurable function and $\mathbb{P}_0 \ll \mu$ a default distribution on Ω, with density p_0
  Thermodynamic counterpart of X: energy
- $-\infty \leq m < b_0 < M \leq +\infty$ where
$$m \triangleq \mu\text{-ess inf X},$$
$$M \triangleq \mu\text{-ess sup X},$$
$$b_0 \triangleq E_{\mathbb{P}_0}(X) = \int X(\omega)p_0(\omega)\mu(d\omega),$$
- $H(p) \geq H(p_0) = 0 \quad$ whenever $\int p d\mu = 1.$
- $0 < k < k_{\max} \triangleq \lim_{b \downarrow m} F(b).$

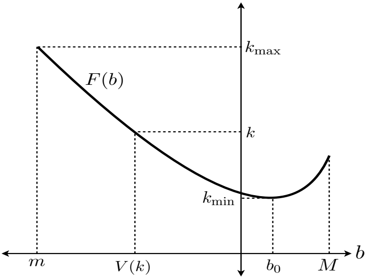
The Duality Lemma

Lemma (Duality Lemma)

Under these standing assumptions, there exists a unique b with

$$F(b) = k, \qquad m < b < b_0 \tag{7}$$

and then $V(k) = b$. *A density p attains the minimum in* (5)*, the definition of* $V(k)$*, if and only if it attains that in* (6) *for b in* (7)*.*



Invoking moment problem results

Instances of CsM 2012, Theorem 1.1 and Lemma 6.6 :

- Convex conjugate $J^*(\theta_1, \theta_2) \triangleq \sup_{a,b}[\theta_1 a + \theta_2 b - J(a, b)]$ of the function (6) equals

$$J^*(\theta_1, \theta_2) = K(\theta_1, \theta_2) \triangleq \int \beta^*(\omega, \theta_1 + \theta_2 X(\omega))\mu(d\omega). \quad (8)$$

  Convex conjugate and (later on) derivative of β are by its second variable.

- The interior of $\mathrm{dom} J \triangleq \{(a, b) : J(a, b) < +\infty\}$ is

$$\text{int dom } J = \{(a, b) : am < b < aM\}$$

Convex conjugate of F

Since $F(b) = J(1, b)$, standard convex duality results give for $(1, b)$ not on the boundary of $\mathrm{dom} J$, i.e., for $b \neq m,\ b \neq M$, that

$$F(b) = J(1, b) = J^{**}(1, b) = \sup_{\theta_1, \theta_2}[\theta_1 + \theta_2 b - K(\theta_1, \theta_2)] \quad (9)$$

$$= \sup_{\theta_2}[\theta_2 b - G(\theta_2)] = G^*(b) \quad (10)$$

where

$$G(\theta_2) \triangleq \inf_{\theta_1}[K(\theta_1, \theta_2) - \theta_1]. \quad (11)$$

Thermodynamic counterpart of G: log of partition function Z.
Thermodynamic counterpart of $-\theta_2$: absolute temperature β.

Lemma

$$G = F^* \quad (12)$$

In particular, $G(0) = -k_{\min}$.

Note: The equality $\theta_2 b - G(\theta_2) = G^*(b)$ holds if and only if a supporting line to the graph of G at θ_2 has slope b.

Evaluation of $V(k)$

Fix $k \in (0, k_{\max})$.
By Main Lemma, $k = F(b)$ where $b = V(k) \in (m, b_0)$.
For this b the maximizer θ_2 in (10) is negative, thus

$$k = \max_{\theta_2 < 0} [\, \theta_2 V(k) - G(\theta_2) \,]. \tag{13}$$

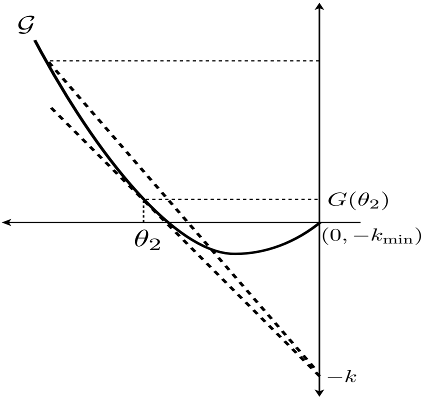
Consequence (extension of Ahmadi-Javid 2011, Theorem 5.1):

Theorem

$$V(k) = \max_{\theta_2 < 0} \frac{k + G(\theta_2)}{\theta_2}. \tag{14}$$

Of course, the maximizers in (13) and (14) are the same.

Evaluation of $V(k)$



Generalized exponential family

Key concept for the moment problem, see CsM 2012. In our case of $\phi(\omega) = (1, X(\omega))$, it consists of the nonnegative functions

$$p_{\theta_1,\theta_2}(\omega) \triangleq (\beta^*)'(\omega, \theta_1 + \theta_2 X(\omega)), \quad (\theta_1, \theta_2) \in \Theta, \qquad (15)$$

Thermodynamic counterpart of p_{θ_2}: state in canonical ensemble.

$$\Theta \triangleq \left\{ (\theta_1, \theta_2) \in \mathrm{dom} K : \theta_1 + \theta_2 X(\omega) < \beta'(\omega, +\infty) \ \ \mu\text{-a.e.} \right\} \quad (16)$$

Instance of CsM 2012, Lemma 4.10: For $b \in (m, M)$, in the definition (6) of $F(b)$ the minimum is attained if and only if some p_{θ_1,θ_2} in (15) is a density with $\int X p_{\theta_1,\theta_2} d\mu = b$. Then

$$H(p_{\theta_1,\theta_2}) = F(b) = \theta_1 + \theta_2 b - K(\theta_1, \theta_2).$$

Via Main Lemma it follows: For $k \in (0, k_{\max})$, in the definition (5) of $V(k)$ the minimum is attained if and only if some p_{θ_1,θ_2} in (15) with $\theta_2 < 0$ is a density with $H(p_{\theta_1,\theta_2}) = k$. Then θ_2 is a minimizer in (14), and the minimizer in (5) (worst case density) is $p = p_{\theta_1,\theta_2}$.

Additional auxiliary results

Clearly, if $K(\theta_1, \theta_2) < +\infty$ and $\overline{\theta}_1 < \theta_1$ then $(\overline{\theta}_1, \theta_2) \in \Theta$, hence $\mathrm{dom}\, K$ and Θ have the same projection to the θ_2 axis. Denote this projection by Θ_2 and its infimum by $\theta_{\min}$.

- $p_{\theta_1,\theta_2}(\omega) = 0$ if and only if $\theta_1 + \theta_2 X(\omega) \leq \beta'(\omega, 0)$. If $\beta'(\omega, 0) = -\infty$ μ-a.e then each p_{θ_1,θ_2} is positive μ-a.e.
- The generalized exponential family (15) contains the default density: $p_0 = p_{\theta_1,0}$ for some θ_1. In particular, $0 \in \Theta_2$.
- The standing assumption $k_{\max} > 0$ is equivalent to $\theta_{\min} < 0$, and implies that $F(b) > 0$ for each $b < b_0$.
- To each $\theta_2 \in \Theta_2$ there exists a unique $\theta_1 = \theta_1(\theta_2)$ with $G(\theta_2) = K(\theta_1, \theta_2) - \theta_1$. If $\int p_{\theta_1,\theta_2} d\mu = 1$ for some θ_1 then this θ_1 is unique and equals $\theta_1(\theta_2)$. Otherwise $\theta_1(\theta_2)$ is the largest θ_1 with $(\theta_1, \theta_2) \in \Theta$, and for this $\int p_{\theta_1,\theta_2} d\mu < 1$.

Main Result

Bregman distance corresponding to any $\beta \in \mathbb{B}$:

$$B(p,q) = B_\beta(p,q) \triangleq \int \Delta_{\beta(\omega,.)}(p(\omega), q(\omega))\mu(d\omega), \tag{17}$$

with $\Delta_\beta(\omega,.)$ defined as in (4), taking $f(s) = \beta(\omega, s)$ there.

Theorem (Worst case localiser)

For $k \in (0, k_{\max})$), let $\theta_2 < 0$ attain the maximum in Theorem 1, let $\theta_1 = \theta_1(\theta_2)$. Then each density p with $H(p) < +\infty$ satisfies

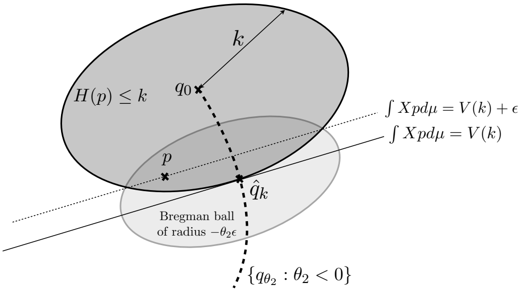
$$B(p, p_{\theta_1,\theta_2}) \leq H(p) - k - \theta_2 \left(\int X p d\mu \; - \; V(k) \right). \tag{18}$$

Consequently, each sequence of densities p_n with

$$H(p_n) \to k, \quad \int X p_n d\mu \to V(k) \tag{19}$$

converges to p_{θ_1,θ_2} in the Theorem, locally in measure.

Almost Worst Case Localisation



All ϵ-AWC densities
(i.e. which satisfy $H(p) \leq k$ and $\mathbb{E}_{\mathbb{P}}(X) \leq V(k) + \epsilon$)
are contained in a Bregman neighbourhood of q_θ of radius proportional to ϵ, with proportionality factor $-\theta_2$.

Comments

- The almost worst case localizer p_{θ_1,θ_2} is uniquely determined by $k \in (0, k_{\max})$, denoted by q_k.
- The parameters θ_1, θ_2 of the worst case localizer $q_k = p_{\theta_1,\theta_2}$ need not be unique, but they are if $q_k > 0$ μ-a.e.
- If a worst case density (attaining the minimum in (5)) exists, it equals the worst case localizer. However, the clustering property also holds when no worst case density exists.
- A sufficient condition for $q_k = p_{\theta_1,\theta_2}$ to be the worst case density is $(\theta_1, \theta_2) \in \text{int dom } K$.
- q_k fails to be a density if and only if θ_2 in Theorem 2 is such that $\int p_{\theta_1,\theta_2} d\mu < 1$ for each θ_1 with $(\theta_1, \theta_2) \in \Theta$.