Maximum Entropy Solution to Ill-posed Inverse Problems with Approximately Known Operator

Jean-Michel LOUBES and Bruno PELLETIER*

Institut de Mathématiques et de Modélisation de Montpellier
UMR CNRS 5149, Equipe de Probabilités et Statistique
Université Montpellier II, CC 051
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

loubes,pelletier@math.univ-montp2.fr

Abstract

We consider the linear inverse problem of reconstructing an unknown finite measure μ from a noisy observation of a generalized moment of μ defined as the integral of a continuous and bounded operator Φ with respect to μ. Motivated by various applications, we focus on the case where the operator Φ is unknown; instead, only an approximation Φ_m to it is available. An approximate maximum entropy solution to the inverse problem is introduced in the form of a minimizer of a convex functional subject to a sequence of convex constraints. Under several assumptions on the convex functional, the convergence of the approximate solution is established.
Index Terms — Maximum entropy, Inverse problems, Convex functionals.

1 Introduction

A number of inverse problems may be stated in the form of reconstructing an unknown measure μ from observations of generalized moments of μ, i.e., moments y of the form

$$y = \int_{\mathcal{X}} \Phi(x) d\mu(x),$$

*Corresponding author.

where $\Phi : \mathcal{X} \to \mathbb{R}^k$ is a given map. Such problems are encountered in various fields of sciences, like medical imaging, time-series analysis, speech processing, image restoration from a blurred version of the image, spectroscopy, geophysical sciences, crytallography, and tomography; see for example Decarreau et al (1992), Gzyl (2002), Hermann and Noll (2000), and Skilling (1988). Recovering the unknown measure μ is generally an ill-posed problem, which turns out to be difficult to solve in the presence of noise, i.e., one observes y^{obs} given by

$$y^{obs} = \int_{\mathcal{X}} \Phi(x) d\mu(x) + \varepsilon. \tag{1.1}$$

For inverse problems with known operator Φ, regularization techniques allow the solution to be stabilized by giving favor to those solutions which minimize a regularizing functional J, i.e., one minimizes $J(\mu)$ over μ subject to the constraint that $\int_{\mathcal{X}} \Phi(x)d\mu(x) = y$ when y is observed, or $\int_{\mathcal{X}} \Phi(x)d\mu(x) \in K_Y$ in the presence of noise, for some convex set K_Y containing y^{obs}. Several types of regularizing functionals have been introduced in the literature. For instance, the squared norm of the density of μ, when this latter is absolutely continuous with respect to a given reference measure, leads to the well-knwon Tikhonov functional. For additional references on regularization techniques, we refer to Engl, Hanke and Neubauer (1996).

Alternatively, one may opt for a regularization functional with grounding in information theory, generally expressed as a negative entropy, which leads to *maximum entropy* solutions to the inverse problem. Maximum entropy solutions have been studied in a deterministic context in Borwein and Lewis (1993, 1996). They may be given a Bayesian interpretation (Gamboa, 1999; Gamboa and Gassiat, 1999) and have proved useful in seismic tomography (Fermin, Loubes and Ludena, 2006) and in image analysis (Gzyl and Zeev, 2003; Skilling and Gull, 2001).

In many actual situations, however, the map Φ is not exactly known. Instead, only an approximation to it is available, say Φ_m, where m represents a degree of accuracy of the approximation or the order of a model. As an exemple, in remote sensing of aerosol vertical profiles, one wishes to recover the concentration of aerosol particles as a function of the altitude, from noisy observations of the radiance field at several wavelengths, i.e., from measurements of a radiometric quantity (see e.g. Gabella et al, 1997; Gabella, Kisselev and Perona, 1999). Under several physical assumptions, the radiance may be related to the aerosol vertical profile by a Fredholm integral equation of the first kind, the kernel of which is approximately known (i.e., known up to a given amount of uncertainty). Actually, the kernel expression results from several modelings at the micro-physical scale, is fairly complex to handle analytically, and so essentially comes in the form of a computer code.

The study of statistical inverse problems with unknown or approximately known operator has started out recently in the case of a linear operator. Efromovich and Koltchinskii (2001) and Cavalier and Hengartner (2005) derive consistency results for an inverse problem with a noisy linear operator and an additive random noise on the observation (see also Carasco, Florens and Renault, 2004). In this paper, based on an approximation Φ_m to Φ, and following lines devised in Gamboa (1999) and Gamboa and Gassiat (1999), we introduce an approximate maximum entropy on the mean (AMEM) estimate $\hat{\mu}_{m,n}$ of the measure μ_X to be reconstructed. This estimate is expressed in the form of a discrete measure concentrated on n points of $\mathcal{X}$. In our main result, we prove that $\hat{\mu}_{m,n}$ converges to the solution of the initial inverse problem (i.e., with the exact Φ) as $m \to \infty$ and $n \to \infty$. Besides, we derive a characterization of $\hat{\mu}_{m,n}$ allowing its construction in a practical setting.

The paper is organized as follows. Section 2 introduces some notation and the definition of the AMEM estimate. In Section 3, we state our main result (Theorem 3.1). Section 4 develops several applications, in particular application to the remote sensing problem. Section 5 is devoted to the proofs of our results. Finally, the Appendix, at the end of the paper, gathers some results on entropic projections and technical Lemmas.

2 Notation and definitions

2.1 Problem position

Let Φ be a continuous and bounded map defined on a subset $\mathcal{X}$ of $\mathbb{R}^d$ and taking values in $\mathbb{R}^k$. The set of finite measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ will be denoted by $\mathcal{M}(\mathcal{X})$, where $\mathcal{B}(\mathcal{X})$ denotes the Borel σ-field of $\mathcal{X}$. Let $\mu_X \in \mathcal{M}(\mathcal{X})$ be an unknown finite measure on $\mathcal{X}$ and consider the following equation:

$$y = \int_{\mathcal{X}} \Phi(x) d\mu_X(x). \tag{2.1}$$

Suppose that we observe a perturbed version y^{obs} of the response y:

$$y^{obs} = \int_{\mathcal{X}} \Phi(x) d\mu_X(x) + \varepsilon,$$

where ε is an error term supposed bounded in norm from above by some positive constant η, representing the maximal noise level. Based on the data y^{obs}, we aim at reconstructing the measure μ_X with a maximum entropy procedure. As explained in the Introduction, the true map Φ is *unknown* and we assume knowledge of an *approximating sequence* Φ_m to the map Φ.

To this aim, let us first introduce some notation. For all probability measure ν on $\mathbb{R}^n$, we shall denote by $\mathcal{L}_\nu$, Λ_ν, and Λ_ν^* the Laplace, log-Laplace, and Cramer transforms of ν, respectively defined by:

$$\begin{aligned}
\mathcal{L}_\nu(s) &= \int_{\mathbb{R}^n} \exp\langle s, x\rangle d\nu(x), \\
\Lambda_\nu(s) &= \log \mathcal{L}_\nu(s), \\
\Lambda_\nu^*(s) &= \sup_{u \in \mathbb{R}^n} \{\langle s, u\rangle - \Lambda_\nu(u)\},
\end{aligned}$$

for all $s \in \mathbb{R}^n$.
Define the set

$$K_Y = \{y \in \mathbb{R}^k : \|y - y^{obs}\| \leq \eta\},$$

i.e., K_Y is the closed ball centered at the observation y^{obs} and of radius η.

Now let ν_Z be a probability measure on $\mathbb{R}_+$. Let P_X be a probability measure on $\mathcal{X}$ having full support, and define the convex functional $I_{\nu_Z}(\mu|P_X)$ by:

$$I_{\nu_Z}(\mu|P_X) = \begin{cases} \int_{\mathcal{X}} \Lambda_{\nu_Z}^*\left(\frac{d\mu}{dP_X}\right) dP_X & \text{if } \mu << P_X \\ +\infty & \text{otherwise.} \end{cases}$$

Then, we consider as a solution of the inverse problem (2.1) a minimizer of the functional $I_{\nu_Z}(\mu|P_X)$ subject to the constraint

$$\mu \in S(K_Y) = \{\mu \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} \Phi(x) d\mu(x) \in K_Y\}.$$

2.2 The AMEM estimate

We introduce the approximate maximum entropy on the mean (AMEM) estimate as a sequence $\hat{\mu}_{m,n}$ of discrete measures on $\mathcal{X}$. In all of the following, the integer m indexes the approximating sequence Φ_m to Φ, while the integer n indexes a random discretization of the space $\mathcal{X}$. For the construction of the AMEM estimate, we proceed as follows.

Let $(X_1, \dots, X_n)$ be a random sample drawn from P_X. Thus the empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges weakly to P_X.

Let L_n be the discrete measure with random weights defined by

$$L_n = \frac{1}{n}\sum_{i=1}^n Z_i \delta_{X_i},$$

where $(Z_i)_i$ is a sequence of i.i.d. random variables on $\mathbb{R}$.

For $\mathcal{S}$ a set we denote by $\operatorname{co}\mathcal{S}$ its convex hull. Let $\Omega_{m,n}$ be the probability event defined by

$$\Omega_{m,n} = [K_Y \cap \operatorname{co} Supp\, F_* \nu_Z^{\otimes n} \neq \emptyset] \tag{2.2}$$

where $F : \mathbb{R}^n \to \mathbb{R}^k$ is the linear operator associated with the matrix

$$\mathbf{A}_{m,n} = \frac{1}{n}\begin{pmatrix} \Phi_m^1(X_1) & \dots & \Phi_m^1(X_n) \\ & \dots & \\ \Phi_m^k(X_1) & \dots & \Phi_m^k(X_n) \end{pmatrix},$$

and where $F_*\nu_Z^{\otimes n}$ denotes the image measure of $\nu_Z^{\otimes n}$ by F. For ease of notation, the dependence of F on m and n will not be explicitely written throughout.

Denote by $\mathcal{P}(\mathbb{R}^n)$ the set of probability measures on $\mathbb{R}^n$. For any map $\Psi : \mathcal{X} \to \mathbb{R}^k$ define the set

$$\Pi_n(\Psi, K_Y) = \left\{ \nu \in \mathcal{P}(\mathbb{R}^n) : \mathbb{E}_\nu \left[\int_{\mathcal{X}} \Psi(x) dL_n(x) \right] \in K_Y \right\}.$$

Let $\nu_{m,n}^\star$ be the I-projection of $\nu_Z^{\otimes n}$ on $\Pi_n(\Phi_m, K_Y)$ (see Appendix 2 for definitions and materials related to I-projections).

Then, on the event $\Omega_{m,n}$, we define the AMEM estimate $\hat{\mu}_{m,n}$ by

$$\hat{\mu}_{m,n} = \mathbb{E}_{\nu_{m,n}^\star}[L_n], \tag{2.3}$$

and we extend the definition of $\hat{\mu}_{m,n}$ to the whole probability space by setting it to the null measure on the complement $\Omega_{m,n}^c$ of $\Omega_{m,n}$. In other words, letting $(z_1, ..., z_n)$ be the expectation of the measure $\nu_{m,n}^\star$, the AMEM estimate may be rewritten more conveniently as

$$\hat{\mu}_{m,n} = \frac{1}{n}\sum_{i=1}^n z_i \delta_{X_i} \tag{2.4}$$

with $z_i = \mathbb{E}_{\nu_{m,n}^\star}(Z_i)$ on $\Omega_{m,n}$, and as $\hat{\mu}_{m,n} \equiv 0$ on $\Omega_{m,n}^c$.

Remark 2.1 *It is shown in Lemma A.1 that $\mathbb{P}(\Omega_{m,n}) \to 1$ as $m \to \infty$ and $n \to \infty$. Hence for m and n large enough, the AMEM estimate $\hat{\mu}_{m,n}$ may be expressed as in (2.4) with high probability, and asymptotically with probability 1.*

3 Convergence of the AMEM estimate

Assumption 1 The minimization problem admits at least one solution, i.e., there exists a continuous function $g_0 : \mathcal{X} \to \operatorname{co} \operatorname{Supp} \nu_Z$ such that

$$\int_{\mathcal{X}} \Phi(x) g_0(x) dP_X \in K_Y.$$

Assumption 2

(i) $\operatorname{dom} \Lambda_{\nu_Z} := \{s : |\Lambda_{\nu_Z}(s)| < \infty\} = \mathbb{R}$;

(ii) Λ'_{ν_Z} and Λ''_{ν_Z} are bounded.

Assumption 3 The approximating sequence Φ_m converges to Φ in $L^\infty(\mathcal{X}, P_X)$.

We are now in a position to state our main result.

Theorem 3.1 (Convergence of the AMEM estimate) *Suppose that Assumption 1, Assumption 2, and Assumption 3 hold. Let $\hat{\mu}$ be the minimizer of the functional*

$$I_{\nu_Z}(\mu|P_X) = \int_{\mathcal{X}} \Lambda^*_{\nu_Z} \left(\frac{d\mu}{dP_X} \right) dP_X$$

subject to the constraint

$$\mu \in S(K_Y) = \{\mu \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} \Phi(x) d\mu(x) \in K_Y\}.$$

Then the AMEM estimate $\hat{\mu}_{m,n}$ converges weakly to $\hat{\mu}$ as $m \to \infty$ and $n \to \infty$. Furthermore $\hat{\mu}$ may be written as

$$\hat{\mu} = \Lambda'_{\nu_Z}(\langle v^\star, \Phi(x)\rangle) P_X,$$

where $v^\star$ is the unique minimizer of

$$H(\Phi, v) = \int_{\mathcal{X}} \Lambda_{\nu_Z} \left(\langle \Phi(x), v \rangle \right) dP_X(x) - \inf_{y \in K_Y} \langle v, y \rangle.$$

Remark 3.1 *Assumption 2-(i) ensures that the function $H(\Phi, v)$ in Theorem 3.1 attains its minimum at a unique point $v^\star$ belonging to the interior of its domain. If this assumption is not met, Borwein and Lewis (1993) and Gamboa and Gassiat (1999) have shown that the minimizers of $I_{\nu_Z}(\mu|P_X)$ over $S(K_Y)$ may have a singular part with respect to P_X.*

4 Application

In remote sensing of aerosol vertical profiles, one wishes to recover the concentration of aerosol particules from noisy observations of the radiance field (i.e., a radiometric quantity), in several spectral bands (see e.g. Gabella et al, 1997; Gabella, Kisselev and Perona, 1999). More specifically, at a given level of modeling, the noisy observation y^{obs} may be expressed as

$$y^{obs} = \int_{\mathcal{X}} \Phi(x; t^{obs}) d\mu_X(x) + \varepsilon, \tag{4.1}$$

where $\Phi : \mathcal{X} \times \mathcal{T} \to \mathbb{R}^k$ is a given operator, and where t^{obs} is a vector of angular parameters observed simultaneously with y^{obs}. The aerosol vertical profile is a function of the altitude x and is associated with the measure μ_X to be recovered, i.e., the aerosol vertical profile is the Radon-Nykodim derivative of μ_X with respect to a given reference measure (e.g., the Lebesgue measure on $\mathbb{R}$). The analytical expression of Φ is fairly complex as it sums up several models at the microphysical scale, so that basically Φ is available in the form of a computer code. So this problem motivates the introduction of an efficient numerical procedure for recovering the unknwon μ_X from y^{obs} and arbitrary t^{obs}.

More generally, the remote sensing of the aerosol vertical profile is in the form of an inverse problem where some of the inputs (namely t^{obs}) are observed simultaneously with the noisy output y^{obs}. Suppose that random points $X_1, \ldots, X_n$ of $\mathcal{X}$ have been generated. Then, applying the maximum entropy approach would require the evaluations of $\Phi(X_i, t^{obs})$ each time t^{obs} is observed. If one wishes to process a large number of observations, say (y_i^{obs}, t_i^{obs}), for different values t_i^{obs}, the computational cost may become prohibitive. So we propose to replace Φ by an approximation Φ_m, the evaluation of which is faster in execution. To this aim, suppose first that $\mathcal{T}$ is a subset of $\mathbb{R}^p$. Let $T_1, ..., T_m$ be random points of $\mathcal{T}$, independent of $X_1, \ldots, X_n$, and drawn from some probability measure μ_T on $\mathcal{T}$ admitting a density f_T with respect to the Lebesgue measure on $\mathbb{R}^p$ such that $f_T(t) > 0$ for all $t \in \mathcal{T}$. Next, consider the operator

$$\Phi_m(x, t) = \frac{1}{f_T(t)} \frac{1}{m} \sum_{i=1}^{m} K_{h_m}(t - T_i) \Phi(x, T_i),$$

where $K_{h_m}(.)$ is a symetric kernel on $\mathcal{T}$ of smoothing sequence h_n. It is a classical exercise to prove that Φ_m converges to Φ provided h_m tends to 0 at a suitable rate. Since the T_i's are independent from the X_i, one may see that Theorem 3.1 applies, and so the solution to the approximate inverse problem

$$y^{obs} = \int_{\mathcal{X}} \Phi_m(x; t^{obs}) d\mu_X(x) + \varepsilon,$$

will converge to the solution to the original inverse problem in Eq. 4.1. In terms of computational complexity, the advantage of this approach is that the construction of the AMEM estimate requires, for each new observation (y^{obs}, t^{obs}), the evaluation of the m kernels at t^{obs}, i.e., $K_{h_m}(t^{obs} - T_i)$, the $m \times n$ ouputs $\Phi(X_i, T_j)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ having evaluated once and for all.

5 Proofs

The proof of our main result falls into three parts. First, we characterize the I-projection of $\nu_Z^{\otimes n}$ on the convex set $\Pi_n(\Phi_m, K_Y)$ (Theorem 5.1), from which we derive the characterization of the AMEM estimate (Corrolary 5.1). Next, we prove an equivalence result (Theorem 5.2) stating that the AMEM estimate is the solution of a discrete inverse problem close to the initial one. Finally, we show that the AMEM estimate sequence converges and we charaterize its accumulation point (Theorem 5.3), from which we deduce our main result (Theorem 3.1).

5.1 Characterization of the I-projection

The following theorem characterizes the I-projection $\nu^\star_{m,n}$ of $\nu_Z^{\otimes n}$ on $\Pi_n(\Phi_m, K_Y)$.

Theorem 5.1 *On the event $\Omega_{m,n}$, the measure $\nu_Z^{\otimes n}$ admits an I-projection $\nu^\star_{m,n}$ on $\Pi_n(\Phi_m, K_Y)$. Furthermore,*

$$d\nu^\star_{m,n} = \frac{\exp\langle \omega^\star_{m,n}, . \rangle}{\mathcal{L}_{\nu_Z^{\otimes n}}(\omega^\star_{m,n})} d\nu_Z^{\otimes n},$$

where $\omega^\star_{m,n} = (\omega^{\star,1}_{m,n}, ..., \omega^{\star,n}_{m,n}) \in \mathbb{R}^n$ has i^{th} component given by

$$\omega^{\star,i}_{m,n} = \langle v^\star_{m,n}, \Phi_m(X_i) \rangle,$$

where $v^\star_{m,n} \in \mathbb{R}^k$ minimizes over $\mathbb{R}^k$ the function

$$H_n(\Phi_m, v) = \frac{1}{n} \sum_{i=1}^{n} \Lambda_{\nu_Z} \left(\langle v, \Phi_m(X_i) \rangle \right) - \inf_{y \in K_Y} \langle v, y \rangle.$$

Proof For all $\nu \in \mathcal{P}(\mathbb{R}^n)$, we have

$$\mathbb{E}_\nu \int_{\mathcal{X}} \Phi_m(x) dL_n(x) = \mathbf{A}_{m,n} \mathbb{E}_\nu[Z],$$

where we recall that $\mathbf{A}_{m,n}$ is the $k \times n$ matrix defined by

$$\mathbf{A}_{m,n} = \frac{1}{n} \begin{pmatrix} \Phi_m^1(X_1) & \dots & \Phi_m^1(X_n) \\ & \dots & \\ \Phi_m^k(X_1) & \dots & \Phi_m^k(X_n) \end{pmatrix}.$$

Consequently,

$$\Pi_n(\Phi_m, K_Y) = \{P \in \mathcal{P}(\mathbb{R}^n) : \mathbf{A}_{m,n} \mathbb{E}_P[Z] \in K_Y\}.$$

Let $F : \mathbb{R}^n \to \mathbb{R}^k$ be the linear operator defined by $z \mapsto \mathbf{A}_{m,n} z$, for $z \in \mathbb{R}^n$. Set

$$\tilde{\Pi} = \{Q \in \mathcal{P}(\mathbb{R}^k) : \int_{\mathbb{R}^k} y dQ(y) \in K_Y\}.$$

Clearly we also have,

$$\Pi_n(\Phi_m, K_Y) = \{P \in \mathcal{P}(\mathbb{R}^n) : F_* P \in \tilde{\Pi}\}.$$

Since on the event $\Omega_{m,n}$ we have

$$K_Y \cap \operatorname{co} Supp\, F_* \nu_Z^{\otimes n} \neq \emptyset$$

then by Theorem B.1 and Theorem B.2, it follows that:

$$\begin{aligned} \frac{d\nu^\star_{m,n}}{d\nu_Z^{\otimes n}}(z) &= \frac{d(F_*\nu_Z^{\otimes n})^\star}{d(F_*\nu_Z^{\otimes n})}(F(z)) \\ &= \frac{\exp\langle u^\star_{m,n}, F(z)\rangle}{\int_{\mathbb{R}^k} \exp\langle u^\star_{m,n}, s\rangle d(F_*\nu_Z^{\otimes n})(s)}, \end{aligned}$$

where $u^\star_{m,n} \in \mathbb{R}^k$ minimizes over u the functional

$$G_n(\Phi_m, u) = \log \int_{\mathbb{R}^k} \exp\langle s, u\rangle d(F_*\nu_Z^{\otimes n})(s) - \inf_{y \in K_Y} \langle y, u\rangle,$$

and where the dual of $\mathbb{R}^k$ has been canonically identified with $\mathbb{R}^k$. But

$$\int_{\mathbb{R}^k} \exp\langle u, s\rangle d(F_*\nu_Z^{\otimes n})(s) = \int_{\mathbb{R}^n} \exp\langle u, F(x)\rangle d\nu_Z^{\otimes n}(x)$$

and

$$\langle u, F(z)\rangle = \langle \mathbf{A}_{m,n}^t u, z\rangle,$$

where $\mathbf{A}_{m,n}^t$ denotes the transpose of $\mathbf{A}_{m,n}$. Consequently,

$$\begin{aligned}\frac{d\nu_{m,n}^\star}{d\nu_Z^{\otimes n}}(z) &= \frac{\exp\langle \mathbf{A}_{m,n}^t u_{m,n}^\star, z\rangle}{\int_{\mathbb{R}^n} \exp\langle \mathbf{A}_{m,n}^t u_{m,n}^\star, x\rangle d\nu_Z^{\otimes n}(x)} \\ &= \frac{\exp\langle \mathbf{A}_{m,n}^t u_{m,n}^\star, z\rangle}{\mathcal{L}_{\nu_Z^{\otimes n}}\left(\mathbf{A}_{m,n}^t u_{m,n}^\star\right)},\end{aligned}$$

and the functional G_n may be rewritten as

$$G_n(\Phi_m, u) = \Lambda_{\nu_Z^{\otimes n}}\left(\mathbf{A}_{m,n}^t u\right) - \inf_{y\in K_Y}\langle y, u\rangle.$$

Now observe that, for all $s = (s_1, ..., s_n)$ in the domain of $\Lambda_{\nu_Z^{\otimes n}}$, we have

$$\Lambda_{\nu_Z^{\otimes n}}(s) = \sum_{i=1}^n \Lambda_{\nu_Z}(s_i),$$

and that

$$\mathbf{A}_{m,n}^t u = \frac{1}{n}\begin{pmatrix} \langle \Phi_m(X_1), u\rangle \\ \vdots \\ \langle \Phi_m(X_n), u\rangle \end{pmatrix}.$$

Thus we arrive at

$$\begin{aligned} G_n(\Phi_m, u) &= n\left[\frac{1}{n}\sum_{i=1}^n \Lambda_{\nu_Z}\left(\langle \Phi_m(X_i), \frac{u}{n}\rangle\right) - \inf_{y\in\mathbb{R}^k}\langle y, \frac{u}{n}\rangle\right] \\ &:= nH_n(\Phi_m, \frac{u}{n}). \end{aligned}$$

Clearly, $u_{m,n}^\star$ minimizes G_n if and only if $\frac{u_{m,n}^\star}{n}$ minimizes H_n. Setting $v_{m,n}^\star = \frac{u_{m,n}^\star}{n}$ and $\omega_{m,n}^\star$ the vector with i^{th} component $\omega_{m,n}^{\star,i} = \langle \Phi_m(X_i), v_{m,n}^\star\rangle$ leads to the desired result. □

Corollary 5.1 *Using the notation of Theorem 5.1, on the event $\Omega_{m,n}$, the AMEM estimate is given by*

$$\hat{\mu}_{m,n} = \frac{1}{n}\sum_{i=1}^n \Lambda_{\nu_Z}'\left(\langle v_{m,n}^\star, \Phi_m(X_i)\rangle\right)\delta_{X_i}.$$

Proof We have

$$\mathbb{E}_{\nu^\star_{m,n}}[L_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\nu^\star_{m,n}}[Z_i]\delta_{X_i}.$$

Now observe that

$$\mathcal{L}_{\nu_Z^{\otimes n}}(z_1, ..., z_n) = \prod_{i=1}^{n}\mathcal{L}_{\nu_Z}(z_i).$$

Thus letting $z^{(i)} = (z_1, ..., z_{i-1}, z_{i+1}, ..., z_n)$, we have

$$\begin{aligned}
\mathbb{E}_{\nu^\star_{m,n}}[Z_i] &= \int_{\mathbb{R}} z_i \exp(\omega^{\star,i}_{m,n} z_i) d\nu_Z(z_i) \frac{1}{\mathcal{L}_{\nu_Z^{\otimes n}}(\omega^\star_{m,n})}\int_{\mathbb{R}^{n-1}}\exp(\sum_{j\neq i}\omega^{\star,j}_{m,n} z_j) d\nu_Z^{\otimes(n-1)}(z^{(i)}) \\
&= \frac{\mathcal{L}'_{\nu_Z}(\omega^{\star,i}_{m,n})}{\mathcal{L}_{\nu_Z}(\omega^{\star,i}_{m,n})} \\
&= \Lambda'_{\nu_Z}(\omega^{\star,i}_{m,n}).
\end{aligned}$$

Consequently

$$\mathbb{E}_{\nu^\star_{m,n}}[L_n] = \frac{1}{n}\sum_{i=1}^{n}\Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi_m(X_i)\rangle\right)\delta_{X_i}.$$

□

5.2 Equivalence Theorem

Theorem 5.2 *Let* $P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$. *Let*

$$S_m(K_Y) = \{\mu \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}}\Phi_m(x)d\mu(x) \in K_Y\}.$$

On the event $\Omega_{m,n}$ *defined by 2.2, the following assertions are equivalent:*

(i) The convex functional $I_{\nu_Z}(.|P_n)$ *defined by*

$$I_{\nu_Z}(\mu|P_n) = \int_{\mathcal{X}}\Lambda^*_{\nu_Z}\left(\frac{d\mu}{dP_n}\right)dP_n$$

attains its minimum on $S_m(K_Y)$ *at* $\mu^\star_{m,n} := \mathbb{E}_{\nu^\star_{m,n}}[L_n]$.

(ii) The measure $\nu_Z^{\otimes n}$ *admits an I-projection* $\nu^\star_{m,n}$ *on* $\Pi_n(\Phi_m, K_Y)$.

Proof $(ii) => (i)$
First, observe that if $I_{\nu_Z}(P|P_n) < \infty$, then there exists $z = (z_1, ..., z_n) \in \mathbb{R}^n$ such that $P = \frac{1}{n}\sum_{i=1}^n z_i\delta_{X_i}$. Such a measure associated with z will be denoted by $P_n(z)$. Then we have

$$\begin{aligned} I_{\nu_Z}(L_n(z)|P_n(z)) &= \frac{1}{n}\Lambda^*_{\nu_Z}(z_i) \\ &= \frac{1}{n}\Lambda^*_{\nu_Z^{\otimes n}}(z). \end{aligned}$$

But

$$\Lambda^*_{\nu_Z^{\otimes n}}(z) = \inf\{H(\nu|\nu_Z^{\otimes n}), \nu \in \mathcal{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} x d\nu(x) = z\}.$$

Denote by ν^z the measure at which the infimum is attained. Thus we have shown that, for all $z \in \mathbb{R}^n$, there exists a measure ν^z such that

$$\begin{cases} \int_{\mathbb{R}^n} x d\nu^z(x) = z \\ \text{and} \\ I_{\nu_Z}(L_n(z)|P_n) = \frac{1}{n}H(\nu^z|\nu_Z^{\otimes n}). \end{cases}$$

In particular, for any measure $\mu \in S_m(K_Y)$ with $I_{\nu_Z}(\mu|P_n) < \infty$, there exists $z \in \mathbb{R}^n$ and a measure $\nu^z \in \mathcal{P}(\mathbb{R}^n)$ such that $\mu = L_n(z) = \mathbb{E}_{\nu^z}[L_n]$. Consequently, $\mathbb{E}_{\nu^z}[L_n] \in S_m(K_Y)$ and so $\nu^z \in \Pi_n(\Phi_m, K_Y)$. Then we deduce that

$$\inf\{I_{\nu_Z}(\mu|P_n) : \mu \in S_m(K_Y)\} \geq \inf_{\nu \in \Pi_n(\Phi_m, K_Y)} \frac{1}{n}H(\nu|\nu_Z^{\otimes n}). \tag{5.1}$$

Now let $\nu^\star_{m,n}$ be the projection of $\nu_Z^{\otimes n}$ on $\Pi_n(\Phi_m, K_Y)$. Since

$$\mathbb{E}_{\nu^\star_{m,n}}[L_n] = \frac{1}{n}\sum_{i=1}^n \alpha_i\delta_{X_i},$$

where

$$\alpha_i = \Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi_m(X_i)\rangle\right),$$

it follows that

$$\begin{aligned} I_{\nu_Z}(\mathbb{E}_{\nu^\star_{m,n}}[L_n]|P_n) &= \tfrac{1}{n}\Lambda^*_{\nu_Z^{\otimes n}}(\alpha_1, ..., \alpha_n) \\ &= \tfrac{1}{n}\inf\left\{H(\nu|\nu_Z^{\otimes n}), \nu \in \mathcal{P}(\mathbb{R}^n) : \textstyle\int_{\mathbb{R}^n} x d\nu(x) = \int_{\mathbb{R}^n} x d\nu^\star_{m,n}(x)\right\} \\ &\leq \tfrac{1}{n}H(\nu^\star_{m,n}|\nu_Z^{\otimes n}) \\ &= \inf_{\nu \in \Pi_n(\Phi_m, K_Y)} \tfrac{1}{n}H(\nu|\nu_Z^{\otimes n}). \end{aligned} \tag{5.2}$$

From (5.1) and (5.2) we deduce that

$$\inf\{I_{\nu_Z}(\mu|P_n) : \mu \in S_m(K_Y)\} = I_{\nu_Z}(\mathbb{E}_{\nu^\star_{m,n}}[L_n]|P_n).$$

$(i) => (ii)$
Let $z^\star \in \mathbb{R}^n$ be such that

$$I_{\nu_Z}(L_n(z^\star)|P_n) = \inf\{I_{\nu_Z}(\mu|P_n) : \mu \in S_m(K_Y)\}$$

Then

$$I_{\nu_Z}(L_n(z^\star)|P_n) = \frac{1}{n} H(\nu^{z^\star}|\nu_Z^{\otimes}),$$

where $\nu^{z^\star}$ satisfies $\int_{\mathbb{R}^n} x d\nu^{z^\star}(x) = z^\star$.
Now for all $\nu \in \Pi_n(\Phi_m, K_Y)$, we have $\mathbb{E}_\nu[L_n] \in S_m(K_Y)$, so

$$\begin{aligned}
\inf\{I_{\nu_Z}(\mu|P_n) : \mu \in \tilde{S}(K_Y)\} &\leq I_{\nu_Z}(\mathbb{E}_\nu[L_n]|P_n) \\
&= \frac{1}{n}\Lambda^*_{\nu_Z^{\otimes n}}\left(\mathbb{E}_\nu[Z]\right).
\end{aligned}$$

Consequently

$$\begin{aligned}
\frac{1}{n} H(\nu^{z^\star}|\nu_Z^{\otimes}) &= \inf\{I_{\nu_Z}(\mu|P_n) : \mu \in \tilde{S}(K_Y)\} \\
&\leq \inf_{\nu \in \Pi_n(\Phi_m, K_Y)} I_{\nu_Z}(\mathbb{E}_\nu[L_n]|P_n) \\
&= \inf_{\nu \in \Pi_n(\Phi_m, K_Y)} \frac{1}{n}\Lambda^*_{\nu_Z^{\otimes n}}\left(\mathbb{E}_\nu[Z]\right) \\
&= \inf_{\nu \in \Pi_n(\Phi_m, K_Y)} \inf_{\lambda \in \mathcal{P}(\mathbb{R}^n) : \mathbb{E}_\lambda[Z] = \mathbb{E}_\nu[Z]} \frac{1}{n} H(\lambda|\nu_Z^{\otimes n}) \\
&\leq \inf_{\nu \in \Pi_n(\Phi_m, K_Y)} \frac{1}{n} H(\nu|\nu_Z^{\otimes n}).
\end{aligned}$$

So $\nu^{z^\star}$ is the entropic projection of $\nu_Z^{\otimes n}$ on $\Pi_n(\Phi_m, K_Y)$. □

5.3 Convergence of the AMEM estimate

Theorem 5.3 *Suppose that Assumption 1, Assumption 2, and Assumption 3 hold. Then the sequence $\hat{\mu}_{m,n}$ converges weakly to the measure $\mu^\star$ given by*

$$\mu^\star = \Lambda'_{\nu_Z}\left(\langle v^\star, \Phi(x)\rangle\right) P_X,$$

where $v^\star$ is the unique minimum of

$$H(\Phi, v) = \int_{\mathcal{X}} \Lambda_{\nu_Z}\left(\langle \Phi(x), v\rangle\right) dP_X(x) - \inf_{y \in K_Y} \langle v, y\rangle.$$

Proof We first prove the convergence of the sequence $v^\star_{m,n}$ to $v^\star$. To this aim, we start by showing that

$$H_n(\Phi_m, v) = \frac{1}{n}\sum_{i=1}^{n} \Lambda_{\nu_Z}\left(\langle v, \Phi_m(X_i)\rangle\right) - \inf_{y\in K_Y}\langle v, y\rangle$$

converges pointwise in probability to $H(\Phi, v)$.

First, we have

$$\mathbb{E}\frac{1}{n}\sum_{i=1}^{n} \Lambda_{\nu_Z}\left(\langle v, \Phi_m(X_i)\rangle\right) = \mathbb{E}\Lambda_{\nu_Z}\left(\langle v, \Phi_m(X)\rangle\right).$$

and

$$H(\Phi_m, v) = H(\Phi, v) + \int_{\mathcal{X}} \Lambda'_{\nu_Z}(\xi)\langle \Phi_m(x) - \Phi(x), v\rangle dP_X(x)$$

for some $\xi = \xi(m) \in \mathbb{R}$. Then we deduce that

$$\{\mathbb{E}H_n(\Phi_m, v) - H(\Phi, v)\}^2 \leq \left(\sup|\Lambda'_{\nu_Z}|\right)^2 \|\Phi_m - \Phi\|^2_\infty \|v\|^2.$$

Second, we have

$$\begin{aligned}
&Var\left(\frac{1}{n}\sum_{i=1}^{n} \Lambda_{\nu_Z}\left(\langle v, \Phi_m(X_i)\rangle\right)\right)\\
&= \frac{1}{n}Var\left(\Lambda_{\nu_Z}\left(\langle v, \Phi_m(X)\rangle\right)\right)\\
&\leq \frac{1}{n}\mathbb{E}\Lambda^2_{\nu_Z}\left(\langle v, \Phi_m(X)\rangle\right)\\
&= \frac{1}{n}\int_{\mathcal{X}} \Lambda^2_{\nu_Z}\left(\langle v, \Phi_m(x)\rangle\right) dP_X(x)\\
&= \frac{1}{n}\int_{\mathcal{X}} \left(\Lambda_{\nu_Z}\left(\langle v, \Phi(x)\rangle\right) + \Lambda'_{\nu_Z}(\xi)\langle \Phi_m(x) - \Phi_x, v\rangle\right)^2 dP_X(x)\\
&\leq \frac{1}{n}\left[2 + \int_{\mathcal{X}} \left(\Lambda^2_{\nu_Z}\left(\langle v, \Phi(x)\rangle\right)\right)\right]
\end{aligned}$$

for m large enough. Hence we have shown that, for all v, there exist constants C and C' depending only on v such that

$$\mathbb{E}\left[H_n(\Phi_m, v) - H(\Phi, v)\right]^2 \leq C(v)\|\Phi_m - \Phi\|^2_\infty + \frac{C'(v)}{n}$$

from which it follows that $H_n(\Phi_m, v)$ converges pointwise to $H(\Phi, v)$ in probability as $m \to \infty$ and $n \to \infty$.

Thus, from Theorem 10.8 in Rockafellar (1997), the convergence is uniform on each compact subset, which implies that the sequence of minima $v^\star_{m,n}$ converges to $v^\star$.

Now, on the event $\Omega_{m,n}$, for all continuous and bounded function g on $\mathcal{X}$, we have

$$\int_{\mathcal{X}} g(x)d\hat{\mu}_{m,n}(x) = \frac{1}{n}\sum_{i=1}^{n} \Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi_m(X_i)\rangle\right) g(X_i).$$

We may write

$$\begin{aligned}
&\left|\int_{\mathcal{X}} g(x)d\hat{\mu}_{m,n}(x) - \int_{\mathcal{X}} g(x)d\mu^\star(x)\right| \\
&\quad \leq \left|\int_{\mathcal{X}} g(x)d\hat{\mu}_{m,n}(x) - \frac{1}{n}\sum_{i=1}^{n} \Lambda'_{\nu_Z}\left(\langle v^\star, \Phi(X_i)\rangle\right) g(X_i^n)\right| \\
&\qquad + \left|\frac{1}{n}\sum_{i=1}^{n} \Lambda'_{\nu_Z}\left(\langle v^\star, \Phi(X_i)\rangle\right) g(X_i) - \int_{\mathcal{X}} g(x)d\mu^\star(x)\right| \\
&\quad := I_1 + I_2.
\end{aligned}$$

For the first term I_1, we have

$$\begin{aligned}
I_1 &= \left|\frac{1}{n}\sum_{i=1}^{n} \Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi_m(X_i)\rangle\right) - \Lambda'_{\nu_Z}\left(\langle v^\star, \Phi(X_i)\rangle\right) g(X_i)\right| \\
&\leq \left|\frac{1}{n}\sum_{i=1}^{n} \Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi_m(X_i)\rangle\right) - \Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi(X_i)\rangle\right) g(X_i)\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n} \Lambda'_{\nu_Z}\left(\langle v^\star_{m,n}, \Phi(X_i)\rangle\right) - \Lambda'_{\nu_Z}\left(\langle v^\star, \Phi(X_i)\rangle\right) g(X_i)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \left|\Lambda''_{\nu_Z}(\xi)\langle v^\star_{m,n}, \Phi_m(X_i) - \Phi(X_i)\rangle\right| |g(X_i)| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n} \Lambda''_{\nu_Z}(\xi')\langle v^\star_{m,n} - v^\star, \Phi(X_i)\rangle g(X_i)\right| \\
&\leq C\|\Phi_m - \Phi\|_\infty + C'\|\frac{1}{n}\sum_{i=1}^{n} \Phi(X_i)g(X_i)\|\|v^\star_{m,n} - v^\star\|
\end{aligned}$$

where C and C' are positive constants. But $\frac{1}{n}\sum_{i=1}^{n} \Phi(X_i)g(X_i)$ converges in probability to $\int_{\mathcal{X}} \Phi(x)g(x)dP_X(x)$, and since $v^\star_{m,n} \to v^\star$, we conclude that

$I_1 \to 0$.
For the second term I_2, repeating the first step of the proof of the convergence of H_n to H with Λ'_{ν_Z} in place of Λ_{ν_Z}, we conclude that $I_2 \to 0$, since Λ''_{ν_Z} is bounded under Assumption 2.
To conclude the proof, we may write

$$\begin{aligned} \mathbb{P}\left(|\hat{\mu}_{m,n}g - \mu^\star g| > \varepsilon\right) &\leq \mathbb{P}\left(|\hat{\mu}_{m,n}g - \mu^\star g| > \varepsilon \cap \Omega_{m,n}\right) + \mathbb{P}(\Omega^c_{m,n}) \\ &\to 0. \end{aligned}$$

□

5.4 Proof of Theorem 3.1

Let $\hat{\mu}_{\infty,n} = \lim_{m\to\infty} \hat{\mu}_{m,n}$. Since $S_m(K_Y) \to S(K_Y)$ as $m \to \infty$, we have by Theorem 5.2 that

$$I_{\nu_Z}\left(\hat{\mu}_{\infty,n}|P_n\right) = \inf_{\mu\in S(K_Y)} I_{\nu_Z}\left(\mu|P_n\right). \tag{5.3}$$

Let T_n be the operator defined for all $\mu << P_X$ by

$$T_n(\mu) = \frac{1}{n}\sum_{i=1}^{n} \frac{d\mu}{dP_X}(X_i)\delta_{X_i}.$$

Since for all $\mu \in \mathcal{M}(\mathcal{X})$, if $I_{\nu_Z}(\mu|P_n)$ is finite then μ takes the form $\frac{1}{n}\sum_{i=1}^{n} z_i\delta_{X_i}$, it follows that

$$\inf_{\mu\in S(K_Y)} I_{\nu_Z}(\mu|P_n) \leq \inf_{\mu\in S(K_Y) \text{and}\, \mu<<P_X} I_{\nu_Z}(T_n(\mu)|P_n). \tag{5.4}$$

Inspecting the proof Theorem 5.3, it may be seen that $I_{\nu_Z}\left(\hat{\mu}_{\infty,n}|P_n\right)$ converges in probability to $I_{\nu_Z}(\mu^\star|P_X)$ as $n \to \infty$, where $\mu^\star$ is given by

$$\mu^\star = \Lambda'_{\nu_Z}(\langle v^\star, \Phi(x)\rangle)P_X.$$

Furthermore, for all $\mu << P_X$, $I_{\nu_Z}(T_n(\mu)|P_n)$ converges in probability to $I_{\nu_Z}(\mu|P)$. Observing that

$$\inf_{\mu\in S(K_Y)} I_{\nu_Z}(\mu|P_X) = \inf_{\mu\in S(K_Y) \text{and}\, \mu<<P_X} I_{\nu_Z}(\mu|P_X)$$

yields, together with Eq. (5.3) and Eq. (5.4), that

$$I_{\nu_Z}(\mu^\star|P_X) \leq \inf_{\mu\in S(K_Y)} I_{\nu_Z}(\mu|P_X).$$

Finally, observing that $S(K_Y)$ is compact for the weak topology on $\mathcal{M}(\mathcal{X})$ since Φ is continuous and bounded on $\mathcal{X}$, we also have that $\mu^\star \in S(K_Y)$. □

A Technical Lemma

Lemma A.1 *Suppose that Assumption 1 holds. Let $F : \mathbb{R}^n \to \mathbb{R}^k$ be the linear operator associated with the matrix*

$$\mathbf{A}_{m,n} = \frac{1}{n} \begin{pmatrix} \Phi_m^1(X_1) & \dots & \Phi_m^1(X_n) \\ & \dots & \\ \Phi_m^k(X_1) & \dots & \Phi_m^k(X_n) \end{pmatrix}.$$

Then

$$\mathbb{P}\left(K_Y \cap \mathrm{co}\, Supp\, F_* \nu_Z^{\otimes n} \neq \emptyset\right) \to 1$$

as $m \to \infty$ and $n \to \infty$.

Proof First observe that

$$F\left(Supp \nu_Z^{\otimes n}\right) \subset Supp F_* \nu_Z^{\otimes n},$$

and that since F is a linear operator, it follows that

$$F\left(\mathrm{co}\, Supp \nu_Z^{\otimes n}\right) \subset \mathrm{co}\, Supp F_* \nu_Z^{\otimes n}.$$

Furthermore

$$\begin{aligned} F\left(F^{-1}(K_Y) \cap \mathrm{co}\, Supp \nu_Z^{\otimes n}\right) &\subset K_Y \cap F\left(\mathrm{co}\, Supp \nu_Z^{\otimes n}\right) \\ &\subset K_Y \cap \mathrm{co}\, Supp F_* \nu_Z^{\otimes n}. \end{aligned}$$

Consequently, if $F^{-1}(K_Y) \cap \mathrm{co}\, Supp \nu_Z^{\otimes n}$ is nonempty, then so is $K_Y \cap \mathrm{co}\, Supp F_* \nu_Z^{\otimes n}$.

Now we proceed to show that $F^{-1}(K_Y) \cap \mathrm{co}\, Supp \nu_Z^{\otimes n}$ is nonempty for n large enough. First note that

$$\mathrm{co}\, Supp \nu_Z^{\otimes n} = \left(\mathrm{co}\, Supp \nu_Z\right)^n.$$

Under Assumption 1, there exists g_0 such that

$$\int_{\mathcal{X}} \Phi(x) g_0(x) dP_X \in K_Y.$$

Set

$$z_n = (g(X_1^n), ..., g(X_n^n)) \in \left(\mathrm{co}\, Supp \nu_Z\right)^n.$$

Now the result follows from the fact that i) $F(z_n)$ converges to $\int_{\mathcal{X}} \Phi_m(x) g_0(x) dP_X$ as $n \to \infty$ in probability as $n \to \infty$ and ii) $\int_{\mathcal{X}} \Phi_m(x) g_0(x) dP_X$ converges to $\int_{\mathcal{X}} \Phi(x) g_0(x) dP_X \in K_Y$ as $m \to \infty$. $\square$

B Entropic projection

Let $\mathcal{X}$ be a set, and let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. For $\nu, \mu \in \mathcal{P}(\mathcal{X})$, the relative entropy of ν with respect to μ is defined by

$$H(\nu|\mu) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{if} \nu << \mu \\ +\infty & \text{otherwise.} \end{cases}$$

Given a set $\mathcal{C} \in \mathcal{P}(\mathcal{X})$ and a probability measure $\mu \in \mathcal{P}(\mathcal{X})$, an element μ^* of $\mathcal{C}$ is called an *I-projection* of μ on $\mathcal{C}$ if

$$H(\mu^*|\mu) = \inf_{\nu \in \mathcal{C}} H(\nu|\mu).$$

Now we let $\mathcal{X}$ be a locally convex topological vector space of finite dimension. The dual of $\mathcal{X}$ will be denoted by $\mathcal{X}'$. The following two Theorems, due to Csiszar (1984), characterize the entropic projection of a given probability measure on a convex set. For their proofs, see Theorem 3 and Lemma 3.3 in Csiszar (1984), respectively.

Theorem B.1 *Let μ be a probability measure on $\mathcal{X}$. Let $\mathcal{C}$ be a convex subset of $\mathcal{X}$ whose interior has a non-empty intersection with the convex hull of the support of μ. Let*

$$\Pi\left(\mathcal{X}\right) = \{P \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} x dP(x) \in \mathcal{C}\}.$$

Then the I-projection μ^* of μ on $\Pi(\mathcal{C})$ is given by the relation

$$d\mu^*(x) = \frac{\exp \lambda^*(x)}{\int_{\mathcal{X}} \exp \lambda^*(u) d\mu(u)} d\mu(x),$$

where $\lambda^* \in \mathcal{X}'$ is given by

$$\lambda^* = \arg\max_{\lambda \in \mathcal{X}'} \left[\inf_{x \in \mathcal{C}} \lambda(x) - \log \int_{\mathcal{X}} \exp \lambda(x) d\mu(x) \right].$$

Given $F : \mathcal{X} \to \mathcal{Y}$ a measurable mapping between measurable spaces $\mathcal{X}$ and $\mathcal{Y}$, and a probability measure P on $\mathcal{X}$, the image measure of P under F will be denoted F_*P. With this notation, we have the following theorem.

Theorem B.2 *Let $F : \mathcal{X} \to \mathcal{Y}$ be a measurable mapping between two measurable spaces $\mathcal{X}$ and $\mathcal{Y}$. Let Π_Y be a convex set of probability measures on*

$\mathcal{Y}$ and let Π_X be the set of probability measures on $\mathcal{X}$ whose image under F belong to Π_Y, i.e.,

$$\Pi_X = \{P \in \mathcal{P}(\mathcal{X}) : F_* P \in \Pi_Y\}.$$

Then for any $\mu_X \in \mathcal{P}(\mathcal{X})$, if $H(\Pi_X|\mu_X) < \infty$, the I-projections μ_X^* of μ_X on Π_X and $(F_*\mu_X)^*$ of $F_*\mu_X$ on Π_Y are related by:

$$\frac{d\mu_X^*}{d\mu_X}(x) = \frac{d(F_*\mu_X)^*}{d(F_*\mu_X)}(F(x)) \; [\mu_X].$$

References

[1] J.M. Borwein and A.S. Lewis. Partially-finite programming in l^1 and the existence of maximum entropy estimates. *SIAM J. Optim.*, 3:248–267, 1993.

[2] J.M. Borwein, A.S. Lewis, and D. Noll. Maximum entropy reconstruction using derivative information i: Fisher information and convex duality. *Math. Oper. Res.*, 21:442–468, 1996.

[3] M. Carasco, J-P. Florens, and E. Renault. Linear inverse problems in structural econometrics. *Handbook of Econometrics*, 2004.

[4] L. Cavalier and N.W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361, 2005.

[5] I. Csiszár. Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.*, 12(3):768–793, 1984.

[6] A. Decarreau, D. Hilhorst, C. Lemaréchal, and J. Navaza. Dual methods in entropy maximization. Application to some problems in crystallography. *SIAM J. Optim.*, 2(2):173–197, 1992.

[7] S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Trans. Inform. Theory*, 47(7):2876–2894, 2001.

[8] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[9] A-K. Fermin, J-M. Loubes, and C. Ludeña. Bayesian methods in seismic tomography. *International Journal of Tomography and Statistics*, 4(6):1–19, 2006.

[10] M. Gabella, R. Guzzi, V. Kisselev, and G. Perona Retrieval of aerosol profile variations in the visible and near infrared: theory and application of the single-scattering approach. *Applied Optics*, 36(6):1328-1336, 1997.

[11] M. Gabella, V. Kisselev, and G. Perona Retrieval of aerosol profile variations from reflected radiation in the oxygen absorption A band. *Applied Optics*, 38(15):3190-3195, 1999.

[12] F. Gamboa. New Bayesian methods for ill posed problems. *Stat. Decis.*, 17(4):315–337, 1999.

[13] F. Gamboa and E. Gassiat. Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Statist.*, 25(1):328–350, 1999.

[14] H. Gzyl. Tomographic reconstruction by maximum entropy in the mean: unconstrained reconstructions. *Appl. Math. Comput.*, 129(2-3):157–169, 2002.

[15] H. Gzyl and N. Zeev. Probabilistic approach to an image reconstruction problem. *Methodol. Comput. Appl. Probab.*, 4(3):279–290, 2003.

[16] U. Hermann and D. Noll. Adaptive image reconstruction using information measures. *SIAM J. Control Optim.*, 38(4):1223–1240, 2000.

[17] R. T Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[18] J. Skilling. Maximum entropy spectroscopy—DIMES and MESA. In *Maximum-entropy and Bayesian methods in science and engineering, Vol. 2 (Laramie, WY, 1985 and Seattle, WA, 1986/1987)*, Fund. Theories Phys., pages 127–145. Kluwer Acad. Publ., Dordrecht, 1988.

[19] J. Skilling and S. Gull. Bayesian maximum entropy image reconstruction. In *Spatial statistics and imaging (Brunswick, ME, 1988)*, volume 20 of *IMS Lecture Notes Monogr. Ser.*, pages 341–367. Inst. Math. Statist., Hayward, CA, 1991.