Maximum Entropy on the Mean (MEM) method to solve inverse problems.

Eva Lawrence [1]

Workshop Potsdam-Toulouse on Stochastic Processes and Machine Learning - 13th March

[1] eva.lawrence@math.univ-toulouse.fr

Outline

Introduction

Context

- Probability space $(U, \mathcal{B}(U), P_U)$.

Introduction

Context

- Probability space $(U, \mathcal{B}(U), P_U)$.

- General method of **reconstruction** for a **multidimensional** function $f(x) = (f^1(x), \ldots, f^p(x))^T$ defined on U, from **partial knowledge**.
  - . Solution to an inverse problem.
  - . At designs points $\{x_l\}_{l=1,\ldots,N}$, we observe

$$\sum_{i=1}^{p} \lambda^i(x_l) f^i(x_l) = z_l \qquad 1 \leqslant l \leqslant N, \tag{1}$$

where $\lambda^i \in L^2(U, \mathbb{R})$ are known contribution functions.

Introduction

Context

- For generalization purpose, problem (1) is rewritten

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N, \tag{2}$$

where $\lambda^i \in L^2(U, \mathbb{R})$ are known contribution functions, Φ_l some positive measures.

Introduction
Context

- For generalization purpose, problem (1) is rewritten

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N, \tag{2}$$

where $\lambda^i \in L^2(U, \mathbb{R})$ are known contribution functions, Φ_l some positive measures.

- **Generalized moment and interpolation** problem features a mix of

$$\Phi_l(x) = \delta_{x_l}(x)$$
$$\Phi_l(x) = x^l P_U(x).$$

Introduction

MEM method in action

- Recall **Kullback-Leibler divergence** K of measure P with respect to Q

$$K(P,Q) = \begin{cases} \int \log\left(\frac{dP}{dQ}\right) dP & \text{if } P \ll Q \quad \text{and } \log\left(\frac{dP}{dQ}\right) \in L^1(P) \\ +\infty & \text{else.} \end{cases} \tag{3}$$

Introduction

MEM method in action

- Recall **Kullback-Leibler divergence** K of measure P with respect to Q

$$K(P,Q) = \begin{cases} \int \log\left(\frac{dP}{dQ}\right) dP & \text{if } P \ll Q \quad \text{and } \log\left(\frac{dP}{dQ}\right) \in L^1(P) \\ +\infty & \text{else.} \end{cases} \tag{3}$$

- **Problem**
  . Probability space $(V, \mathcal{B}(V), P_V)$.
  . Find probability measure $P \ll P_V$ on $(V, \mathcal{B}(V))$ such that

$$\int_V \varphi(t) dP(t) \in A, \;\; A \subset \mathbb{R}. \tag{4}$$

Introduction

MEM method in action

- Recall **Kullback-Leibler divergence** K of measure P with respect to Q

$$K(P,Q) = \begin{cases} \int \log\left(\frac{dP}{dQ}\right) dP & \text{if } P \ll Q \quad \text{and } \log\left(\frac{dP}{dQ}\right) \in L^1(P) \\ +\infty & \text{else.} \end{cases} \tag{3}$$

- **Problem**
  . Probability space $(V, \mathcal{B}(V), P_V)$.
  . Find probability measure $P \ll P_V$ on $(V, \mathcal{B}(V))$ such that

$$\int_V \varphi(t) dP(t) \in A, \ A \subset \mathbb{R}. \tag{4}$$

- **Maximum Entropy (ME) principle**: solution P^{ME} is solution of following problem with continuous φ

$$\begin{aligned} &\min K(P, P_V) \\ &s.t. \int_V \varphi(t) dP(t) \in A. \end{aligned} \tag{5}$$

Introduction

MEM method in action

- **Maximum Entropy on the Mean** (MEM) method
  - . Discretize set V with deterministic points $(t_i)_{i=1,\dots,n}$.
  - . Define random measure ν_n

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \delta_{t_i}, \qquad \text{with } \frac{1}{n}\sum_{i=1}^{n} \delta_{t_i} \to P_V \tag{6}$$

where δ_{t_i} are Dirac measures located at points t_i and Y_i are i.i.d. random amplitudes at point t_i, Q_n is a prior measure for the $(Y_1, \dots, Y_n)^T$.

Introduction

MEM method in action

- **Maximum Entropy on the Mean** (MEM) method
  - . Discretize set V with deterministic points $(t_i)_{i=1,\dots,n}$.
  - . Define random measure ν_n

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \delta_{t_i}, \qquad \text{with } \frac{1}{n}\sum_{i=1}^{n} \delta_{t_i} \to P_V \tag{6}$$

where δ_{t_i} are Dirac measures located at points t_i and Y_i are i.i.d. random amplitudes at point t_i, Q_n is a prior measure for the $(Y_1, \dots, Y_n)^T$.

Discretized moment constraint

$$\mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^{n}\varphi(t_i)Y_i\right] \in A.$$

MEM problem : solution Q_n^{MEM} is solution of following problem

$$\min K(Q, Q_n)$$
$$s.t.\ \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^{n}\varphi(t_i)Y_i\right] \in A. \tag{7}$$

Introduction

MEM method in action

- **Maximum Entropy on the Mean** (MEM) method
  - . Discretize set V with deterministic points $(t_i)_{i=1,\dots,n}$.
  - . Define random measure ν_n

$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \delta_{t_i}, \qquad \text{with } \frac{1}{n}\sum_{i=1}^{n} \delta_{t_i} \to P_V \tag{6}$$

where δ_{t_i} are Dirac measures located at points t_i and Y_i are i.i.d. random amplitudes at point t_i, Q_n is a prior measure for the $(Y_1, \dots, Y_n)^T$.

Discretized moment constraint

$$\mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^{n}\varphi(t_i)Y_i\right] \in A.$$

MEM problem : solution Q_n^{MEM} is solution of following problem

$$\min K(Q, Q_n)$$
$$s.t.\ \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^{n}\varphi(t_i)Y_i\right] \in A. \tag{7}$$

Define $P_n^{MEM} = \mathbb{E}_{Q_n^{MEM}}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\delta_{t_i}\right]$

We want $P_n^{MEM} \to P^{ME}$. [Gamboa, Gassiat, 1997]

To MEM problem for function reconstruction

From a convex analysis point-of-view

- In **convex analysis**, study of more general γ-divergence/γ-projection : [Borwein and Lewis, 1991], [Borwein and Lewis, 1993].

Alternative **objective function** to consider

. For a measure P

$$D_\gamma(P, P_U) = \int \gamma\left(\frac{dP^a}{dP_U}\right) dP_U \tag{8}$$

with

. γ is some convex function

. $P^a \ll P_U$.

To MEM problem for function reconstruction

- In **convex analysis**, study of more general γ-divergence/γ-projection : [Borwein and Lewis, 1991], [Borwein and Lewis, 1993].

Alternative **objective function** to consider

. For a measure P

$$D_\gamma(P, P_U) = \int \gamma\left(\frac{dP^a}{dP_U}\right) dP_U \qquad (8)$$

with

. γ is some convex function

. $P^a \ll P_U$.

. For a function p

$$I_\gamma(p) = \int \gamma\left(p\right) dP_U. \qquad (9)$$

To MEM problem for function reconstruction

From a convex analysis point-of-view

- In **convex analysis**, study of more general γ-divergence/γ-projection : [Borwein and Lewis, 1991], [Borwein and Lewis, 1993].

  Alternative **objective function** to consider

  . For a measure P

$$D_\gamma(P, P_U) = \int \gamma\left(\frac{dP^a}{dP_U}\right) dP_U \tag{8}$$

  with

  . γ is some convex function

  . $P^a \ll P_U$.

  . For a function p

$$I_\gamma(p) = \int \gamma(p)\, dP_U. \tag{9}$$

- Our problem is now

$$\begin{aligned} &\min \; I_\gamma(f) \\ &s.t. \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N. \end{aligned} \tag{10}$$

To MEM problem for function reconstruction

From a convex analysis point-of-view

- **Problem**

$$\min \; I_\gamma(f)$$

$$s.t. \; \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N.$$

To MEM problem for function reconstruction

From a convex analysis point-of-view

- **Problem**

$$\min \; I_\gamma(f)$$
$$s.t. \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N.$$

- **Assumptions**
  . γ is closed positive convex function differentiable on its domain,
  . γ is essentially strictly convex,
  . ψ convex conjugate of γ

$$\psi(z) = \sup_{y \in dom(\gamma)} \{y^T z - \gamma(y)\}$$

has full domain $\mathbb{R}^p$.

To MEM problem for function reconstruction

From a convex analysis point-of-view

- **Problem**

$$\begin{aligned} &\min \; I_\gamma(f) \\ &s.t. \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N. \end{aligned}$$

- **Assumptions**
  . γ is closed positive convex function differentiable on its domain,
  . γ is essentially strictly convex,
  . ψ convex conjugate of γ

$$\psi(z) = \sup_{y \in dom(\gamma)} \{y^T z - \gamma(y)\}$$

  has full domain $\mathbb{R}^p$.

- Recall

$$\begin{aligned} \gamma &: \mathbb{R}^p \to \mathbb{R} \\ \psi &: \mathbb{R}^p \to \mathbb{R} \end{aligned}$$

To MEM problem for function reconstruction

From a convex analysis point-of-view

Theorem

Suppose there exists a $\mathbb{R}^p$-valued function f which meets the constrains and is in the interior of γ domain for all $x \in U$, P_U-a.s.
Let L be the subspace of $\mathbb{R}^N$ such that for given $(\Phi_1, \ldots, \Phi_N)$ absolutely continuous with respect to P_U

$$L = \left\{ v \in \mathbb{R}^N : v_l = \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_l(x), l = 1, \ldots, N \right\}.$$

The minimum of the γ-projection under the constraints can be expressed by

$$I_\gamma(C) = \max_{v \in \mathbb{R}^N} \left\{ \inf_{c \in \mathcal{K} \cap L} \langle v, c \rangle - \int_U \psi \left(\tau^1(x, v), \ldots, \tau^p(x, v) \right) dP_U(x) \right\} \quad (11)$$

with $\tau^i(x, v) = \sum_{l=1}^{N} \lambda^i(x) v_l \phi_l(x)$.
Then, for $v^o \in \mathbb{R}^N$ optimum of (11), optimum function under the constraints

$$f^{i,o}(x) = \frac{\partial \psi}{\partial \tau_i} \left(\tau^1(x, v^o), \ldots, \tau^p(x, v^o) \right), \quad \forall i = 1, \ldots, p. \quad (12)$$

Proof relies on a Fenchel duality result.

To MEM problem for function reconstruction

From a convex analysis point-of-view

- When Φ_I is **not absolutely continuous** with respect to P_U? Recall Lebesgue decomposition of Φ_I

$$\Phi_I = \phi_I P_U + \Sigma$$

with ϕ_I the Radon-Nikodym derivative with respect to P_U.

To MEM problem for function reconstruction

From a convex analysis point-of-view

- When Φ_I is **not absolutely continuous** with respect to P_U? Recall Lebesgue decomposition of Φ_I

$$\Phi_I = \phi_I P_U + \Sigma$$

with ϕ_I the Radon-Nikodym derivative with respect to P_U.

- Problem constraints become

$$\int \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_I(x) \in \mathcal{K}_I$$

$$\Leftrightarrow \left\{ \int \sum_{i=1}^{p} \lambda^i(x) f^i(x) \phi_I(x) dP_U(x) + \int \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Sigma(x) \right\} \in \mathcal{K}_I$$

$$\Leftrightarrow \int \sum_{i=1}^{p} \lambda^i(x) f^i(x) \phi_I(x) dP_U(x) \in \tilde{\mathcal{K}}_I \neq \mathcal{K}_I.$$

To MEM problem for function reconstruction

From a convex analysis point-of-view

- When Φ_I is **not absolutely continuous** with respect to P_U? Recall Lebesgue decomposition of Φ_I

$$\Phi_I = \phi_I P_U + \Sigma$$

with ϕ_I the Radon-Nikodym derivative with respect to P_U.

- Problem constraints become

$$\int \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Phi_I(x) \in \mathcal{K}_I$$

$$\Leftrightarrow \left\{ \int \sum_{i=1}^{p} \lambda^i(x) f^i(x) \phi_I(x) dP_U(x) + \int \sum_{i=1}^{p} \lambda^i(x) f^i(x) d\Sigma(x) \right\} \in \mathcal{K}_I$$

$$\Leftrightarrow \int \sum_{i=1}^{p} \lambda^i(x) f^i(x) \phi_I(x) dP_U(x) \in \tilde{\mathcal{K}}_I \neq \mathcal{K}_I.$$

- Theorem is not well suited for the interpolation problem!

To MEM problem for function reconstruction

Via linear transfer

- Instead of reconstruction of **functions** vector f, we reconstruct a vector of **measures** F.

To MEM problem for function reconstruction

Via linear transfer

- Instead of reconstruction of **functions** vector f, we reconstruct a vector of **measures** F.

- That is, find measures $(F^1, \ldots, F^p)$ such that

$$\sum_{i=1}^{p} \int_V \varphi_l^i(t) dF^i(t) \in \mathcal{K}_l, \qquad l = 1, \ldots, N \tag{13}$$

To MEM problem for function reconstruction

Via linear transfer

- Instead of reconstruction of **functions** vector f, we reconstruct a vector of **measures** F.
- That is, find measures $(F^1, \ldots, F^p)$ such that

$$\sum_{i=1}^{p} \int_V \varphi_l^i(t) dF^i(t) \in \mathcal{K}_l, \qquad l = 1, \ldots, N \tag{13}$$

- **Links with the previous problem**.
  We choose $K^i(.,.)$ on $U \times V$ for each $i = 1, \ldots, p$ such that

$$f_K^i(x) = \int_V K^i(x,t) dF^i(t) \qquad \forall i = 1, \ldots, p \ ;$$

$$\varphi_l^i(t) = \int_U \lambda^i(x) K^i(x,t) d\Phi_l(x) \qquad \forall i = 1, \ldots, p \ , \ \forall l = 1, \ldots, N.$$

$$\sum_{i=1}^{p} \int_U \lambda^i(x) f_K^i(x) d\Phi_l(x) = \sum_{i=1}^{p} \int_V \varphi_l^i(t) dF^i(t). \tag{14}$$

To MEM problem for function reconstruction

Via linear transfer

- Minimization of γ-divergence under the constraints:

$$\begin{aligned} &\min \; D_\gamma(F, P_V) \\ &s.t. \; \sum_{i=1}^{p} \int_V \varphi_l^i(t) dF^i(t) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N. \end{aligned}$$

To MEM problem for function reconstruction

Via linear transfer

- Minimization of γ-divergence under the constraints:

$$\min \ D_\gamma(F, P_V)$$
$$s.t. \ \sum_{i=1}^{p} \int_V \varphi_l^i(t) dF^i(t) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N.$$

- Additional assumptions:
  . V is a compact metric space and P_V has full support,
  . φ_l^i are continuous,
  . $(\varphi_l^i)_l$ are linearly independent for each $i = 1, \ldots, p$.

To MEM problem for function reconstruction

Via linear transfer

- Minimization of γ-divergence under the constraints:

$$\min \; D_\gamma(F, P_V)$$
$$s.t. \sum_{i=1}^{p} \int_V \varphi_l^i(t) dF^i(t) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N.$$

- Additional assumptions:
  - . V is a compact metric space and P_V has full support,
  - . φ_l^i are continuous,
  - . $(\varphi_l^i)_l$ are linearly independent for each $i = 1, \ldots, p$.

- Problem is now similar to a ME problem introduced before.

To MEM problem for function reconstruction

For a discretized measure

- Discretize constraints of problem (13).
  - Discretize set V with deterministic points $(t_i)_{i=1,\dots,n}$.

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{t_i} \to P_V. \tag{15}$$

  - Define p random measures ν_n^i

$$\nu_n^i = \frac{1}{n}\sum_{j=1}^{n} Y_j^i \delta_{t_j}. \tag{16}$$

  where δ_{t_i} are Dirac measures located at points t_j and Y_j are random amplitudes with values in $\mathbb{R}^p$ at point t_i. Y_j are $\mathbb{R}^p$-valued i.i.d. samples of prior distribution Q_0.

To MEM problem for function reconstruction

For a discretized measure

- Discretize constraints of problem (13).
  - . Discretize set V with deterministic points $(t_i)_{i=1,\dots,n}$.

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{t_i} \to P_V. \tag{15}$$

  - . Define p random measures ν_n^i

$$\nu_n^i = \frac{1}{n}\sum_{j=1}^{n} Y_j^i \delta_{t_j}. \tag{16}$$

where δ_{t_j} are Dirac measures located at points t_j and Y_j are random amplitudes with values in $\mathbb{R}^p$ at point t_i. Y_j are $\mathbb{R}^p$-valued i.i.d. samples of prior distribution Q_0.

  - . Force ν_n^i to meet the constraints.

To MEM problem for function reconstruction

For a discretized measure

- Prior measure Q_0 for Y_i.

To MEM problem for function reconstruction

For a discretized measure

- Prior measure Q_0 for Y_i.
- Choice of ψ: log of Q_0 moment generating function

$$\psi(z) = \log\left(\int \exp(z^T y) dQ_0(y)\right). \tag{17}$$

To MEM problem for function reconstruction

For a discretized measure

- Prior measure Q_0 for Y_i.
- Choice of ψ: log of Q_0 moment generating function

$$\psi(z) = \log\left(\int \exp(z^T y) dQ_0(y)\right). \tag{17}$$

- Provides it exists $y_j = (y_j^1, \ldots, y_j^p)^T$ such that

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\varphi_l^i(t_j)y_j^i \in \mathcal{K}_l, \quad l = 1, \ldots, N, \tag{18}$$

By standard theory of the ME method, Q_n^{MEM} exists.
Q_n^{MEM} belongs to the exponential family through $Q_0^{\otimes n}$ spanned by the statistics (18) for $l = 1, \ldots, N$.

$$Q_n^{MEM} = \sum_{j=1}^{n} \exp\left(\tau_j^T y_j - \psi(\tau_j)\right) Q_0^{\otimes n} \tag{19}$$

with

. $\tau_j = (\tau^1(t_j, v_o), \ldots, \tau^p(t_j, v_o))$

. $\tau^i(t, v) = \sum_{l=1}^{N} v_l \varphi_l^i(t)$

To MEM problem for function reconstruction

For a discretized measure

$$Q_n^{MEM} = \sum_{j=1}^{n} \exp\left(\tau_j^T y_j - \psi(\tau_j)\right) Q_0^{\otimes n} \tag{20}$$

with

. $\tau_j = (\tau^1(t_j, v_o), \ldots, \tau^p(t_j, v_o))$

. $\tau^i(t, v) = \sum_{l=1}^{N} v_l \varphi_l^i(t)$

- v_o is maximizer of

$$H_n(v) = \inf_{c \in \mathcal{K}} \langle v, c \rangle - \frac{1}{n} \sum_{j=1}^{n} \psi\left(\tau^1(t_j, v), \ldots, \tau^p(t_j, v)\right). \tag{21}$$

To MEM problem for function reconstruction

For a discretized measure

$$Q_n^{MEM} = \sum_{j=1}^{n} \exp\left(\tau_j^T y_j - \psi(\tau_j)\right) Q_0^{\otimes n} \tag{20}$$

with

. $\tau_j = (\tau^1(t_j, v_o), \dots, \tau^p(t_j, v_o))$

. $\tau^i(t, v) = \sum_{l=1}^{N} v_l \varphi_l^i(t)$

- v_o is maximizer of

$$H_n(v) = \inf_{c \in \mathcal{K}} \langle v, c \rangle - \frac{1}{n} \sum_{j=1}^{n} \psi\left(\tau^1(t_j, v), \dots, \tau^p(t_j, v)\right). \tag{21}$$

- Function reconstruction is

$$f_{n,K}^i(x) = \frac{1}{n} \sum_{j=1}^{n} K^i(x, t_j) \frac{\partial \psi}{\partial \tau_i} \left(\tau^1(t_j, v_o), \dots, \tau^p(t_j, v_o)\right). \tag{22}$$

Perspectives

Directions for next work:

- Integration of prior information.
- So far, reconstruction in a simple case study.
Generalization to real problems in Thermodynamics.
- Study of **uncertainties propagation**.

Bibliography

- I. Csiszar, Fabrice Gamboa, and Elisabeth Gassiat. "MEM pixel correlated solutions for generalized moment and interpolation problems". In: *Proceedings. IEEE International Symposium on Information Theory.* 1998
- Fabrice Gamboa and Elisabeth Gassiat. "Maximum d'entropie et problème des moments: cas multidimensionnel". In: *PROBABILITY AND MATHEMATICAL STATISTICS, 12.1.6* (1991)
- Fabrice Gamboa. "Maximum d'entropie sur la moyenne.". PhD thesis. Orsay, 1989
- J. Borwein and A. Lewis. "Duality Relationships for Entropy-Like Minimization Problems". In: *SIAM Journal on Control and Optimization* (1991)
- J. Borwein and A. Lewis. "Partially-Finite Programming in L_1 and the Existence of Maximum Entropy Estimates". In: *SIAM Journal on Optimization* (1993)