

# Functional Calibration Estimation by the Maximum Entropy on the Mean Principle

*Statistics: A Journal of Theoretical and Applied Statistics*, 49, 989-1004. 2015.

Santiago Gallón  
(Joint work with F. Gamboa and J-M. Loubes)

Departamento de Matemáticas y Estadística  
Facultad de Ciencias Económicas – Universidad de Antioquia  
Institut de Mathématiques de Toulouse, Université Toulouse III – Paul Sabatier

Seminario Institucional – Escuela de Estadística  
Facultad de Ciencias – Universidad de Nacional de Colombia  
5 de septiembre, 2016

# Outline

The problem of calibration estimation

Functional calibration estimation

Maximum entropy on the mean –MEM– principle

Simulation study

References

# The problem of calibration estimation I

- Let  $U_N = \{1, \dots, N\}$  be a finite survey population.
- Assume that there is a survey variable  $Y_i, i \in U_N$ .

**The goal:** Estimate the finite population mean

$$\mu_Y = N^{-1} \sum_{i \in U_N} Y_i,$$

based on a sample  $a \subset U_N$  of  $n$  elements.

- The unbiased Horvitz-Thompson –HT– estimator:

$$\hat{\mu}_Y^{\text{HT}} = N^{-1} \sum_{i \in a} \pi_i^{-1} Y_i = N^{-1} \sum_{i \in a} d_i Y_i, \quad \pi_i = \mathbb{P}(i \in a).$$

- The HT estimator has **low precision** due to bad sample selection,
- The inference can be improved **modifying** the weights  $d_i = \pi_i^{-1}$ .

## The problem of calibration estimation II

- A method can be conducted by incorporating auxiliary information  $X_i \in \mathbb{R}^q$ ,  $i \in U_N$ ,  $q \geq 1$ , (Deville and Särndal (1992)).

Sources of auxiliary information:

- ✓ Census data
- ✓ Administrative registers
- ✓ Previous surveys

- Calibration estimation: modify the weights of  $\hat{\mu}_Y^{\text{HT}}$  by new weights  $w_i$  close enough to  $d_i$ 's according to some distance  $\mathcal{D}(w, d)$ , s.t.

$$N^{-1} \sum_{i \in a} \textcolor{red}{w}_i X_i = \mu_X = N^{-1} \sum_{i \in U_N} X_i \quad (\text{Calibration equation}).$$

- The weights  $w_i$  generally result in estimates with smaller variance.
- The calibration estimator of  $\mu_Y$  is given by

$$\hat{\mu}_Y = N^{-1} \sum_{i \in a} \textcolor{red}{w}_i Y_i.$$

# Functional calibration estimation

**Our aim:** Calibration estimation for  $\mu_Y$ , where  $Y(t) \in \mathcal{C}([0, T])$  using functional information  $X(t) \in \mathcal{C}([0, T]; \mathbb{R}^q)$ .

- The functional calibration constraint:

$$N^{-1} \sum_{i \in a} w_i(t) X_i(t) = \mu_X(t) = N^{-1} \sum_{i \in U_N} X_i(t), \quad \forall t \in [0, T].$$

- $w_i(t)$  are obtained by matching the calibration estimation with the maximum entropy on the mean principle.
- **Example** (Chaouch and Goga (2012)): Estimation of the mean electricity consumption (every 30 min.) during the 2nd week using the consumption in the 1st week as auxiliary information of 18902 meters sending electricity consumption for two weeks.

# Maximum entropy on the mean –MEM– principle I

- Proposed by Navaza (1985, 1986) to solve an inverse problem in crystallography.
- It is an alternative to the Tikhonov's regularization of ill-posed inverse problems.
- Maximum entropy solutions provide a simple and natural way to incorporate constraints on the solution.
- Let be the linear inverse problem

$$y = \mathcal{K}x,$$

where  $\mathcal{K}: \mathcal{X} \rightarrow \mathcal{Y}$  is a known bounded linear operator.

- The MEM considers  $x$  as the expectation of a r.v.  $X$ , such that

$$y = \mathcal{K}\mathbb{E}_\nu(X),$$

where  $\nu$  is an **unknown** probability measure.

## Maximum entropy on the mean –MEM– principle II

- The solution:  $\hat{x} = \mathbb{E}_{\nu^*}(X)$ , where  $\nu^*$  maximizes the entropy

$$S(\nu \parallel v) = -D(\nu \parallel v), \quad \text{subject to} \quad y = \mathcal{K}\mathbb{E}_v(X),$$

where  $D(\nu \parallel v)$  is the Kullback-Leibler divergence,

$$D(\nu \parallel v) = \begin{cases} \int \log\left(\frac{d\nu}{dv}\right) d\nu & \text{if } \nu \ll v \\ +\infty & \text{otherwise,} \end{cases}$$

- The MEM principle focuses on reconstructing a measure  $\nu^*$  that  $\max S(\nu \parallel v)$  between a feasible finite measure  $\nu$  w.r.t. a given measure  $v$  subject to a linear constraint.

## Reconstruction of $\nu^*$ I

- $\nu^*$  is rebuilt by the random measure method for infinite dimensional inverse problems (Gzyl and Velásquez (2011)).
- The calibration constraint is expressed as an infinite-dimensional linear inverse problem, writing  $w_i(t)$  as

$$w_i(t) = \int_0^1 K(s, t) \varpi_i(s) \, ds + d_i \quad \text{for each } i \in a,$$

where  $K(s, t)$  is a kernel function,  $\varpi_i = \mathbb{E}_\nu [\mathcal{W}_i(s)]$  and  $\mathcal{W}$  is a stochastic process.

## Reconstruction of $\nu^*$ II

- The inverse problem takes the form of a Fredholm integral equation of the first kind,

$$\begin{aligned}\mathbb{E}_\nu [\mathcal{K}\mathcal{W}] &= \mathbb{E}_\nu \left\{ \sum_{i \in a} \left[ \int_0^1 K(s, t) d\mathcal{W}_i(s) + d_i \right] \mathbf{X}_i(t) \right\} \\ &= \int_0^1 \sum_{i \in a} K(s, t) \mathbf{X}_i(t) \varpi_i(s) ds + \sum_{i \in a} d_i \mathbf{X}_i(t) \\ &= N \boldsymbol{\mu}_X(t)\end{aligned}$$

- The functions  $\varpi_i^*(s)$  that solve  $\mathbb{E}_\nu [\mathcal{K}\mathcal{W}] = N \boldsymbol{\mu}_X(t)$  are obtained by considering  $\varpi_i(s)$  as a density of a measure

$$d\mathcal{W}_i(s) = \varpi_i(s) ds, \quad i \in a.$$

# The main theorem

**Theorem (Csiszár (1984)).** Let  $v$  be a prior p.m.,  $\boldsymbol{\lambda} = \boldsymbol{\lambda}(t)$  a measure in the class  $\mathcal{M}(\mathcal{C}[0, 1]; \mathbb{R}^q)$ , and  $\mathcal{V} = \{\nu \ll v : Z_v(\boldsymbol{\lambda}) < +\infty\}$ , where  $Z_v(\boldsymbol{\lambda}) = \mathbb{E}_v [\exp \{\langle \boldsymbol{\lambda}, \mathcal{KW} \rangle\}]$ , with

$$\langle \boldsymbol{\lambda}, \mathcal{KW} \rangle = \int_0^1 \boldsymbol{\lambda}^\top(dt) \left( \int_0^1 \sum_{i \in a} K(s, t) \mathbf{X}_i(t) d\mathcal{W}_i(s) + \sum_{i \in a} d_i \mathbf{X}_i(t) \right).$$

Then, *there exists a unique* p.m.

$$\nu^* = \arg \max_{\nu \in \mathcal{V}} S(\nu \parallel v) \quad \text{subject to} \quad \mathbb{E}_\nu [\mathcal{KW}] = N \boldsymbol{\mu}_X(t),$$

which is achieved at

$$d\nu^*/dv = Z_v^{-1}(\boldsymbol{\lambda}^*) \exp \{\langle \boldsymbol{\lambda}^*, \mathcal{KW} \rangle\},$$

where  $\boldsymbol{\lambda}^*(t)$  minimizes the functional

$$H_v(\boldsymbol{\lambda}) = \log Z_v(\boldsymbol{\lambda}) - \langle \boldsymbol{\lambda}, N \boldsymbol{\mu}_X \rangle.$$

## Our results

**Lemma 1.** Let  $\nu$  be a prior centered stationary **Gaussian** measure on  $(\mathcal{C}([0, 1]), \mathcal{B}(\mathcal{C}([0, 1])))$ , and  $\boldsymbol{\lambda}(t) \in \mathcal{M}(C[0, 1]^q)$ . Then,

$$\hat{w}_i(t) = \int_0^1 K(s, t) \varpi^*(s) ds + d_i, \quad i \in a, \text{ where}$$

$$\varpi^*(s) = \sum_{i' \in a} \int_0^1 K(s, t') \mathbf{X}_{i'}^\top(t') \boldsymbol{\lambda}^*(dt').$$

**Lemma 2.** Let  $\mathcal{W}_i(s) = \sum_{k=1}^{N(s)} \xi_{ik}$  be a **compound Poisson** process, where  $N(s)$  is a Poisson process on  $[0, 1]$  with parameter  $\gamma > 0$ , and  $\xi_{ik}, k \geq 1$  are i.i.d. r.v.'s for each  $i \in a$  with distribution  $u$  independent of  $N(s)$ . Then,  $\hat{w}_i(t) = \int_0^1 K(s, t) \varpi_i^*(s) ds + d_i, \quad i \in a, \text{ where}$

$$\varpi_i^*(s) = \int_{\mathbb{R}} \xi_i \exp \left\{ \sum_{i \in a} \int_0^1 K(s, t) \xi_i \mathbf{X}_i^\top(t) \boldsymbol{\lambda}^*(dt) \right\} u(d\xi_i).$$

# Simulation study I

$$Y_i(t) = \alpha(t) + \mathbf{X}_i(t)^\top \boldsymbol{\beta}(t) + \varepsilon_i(t), \quad i \in U_N,$$

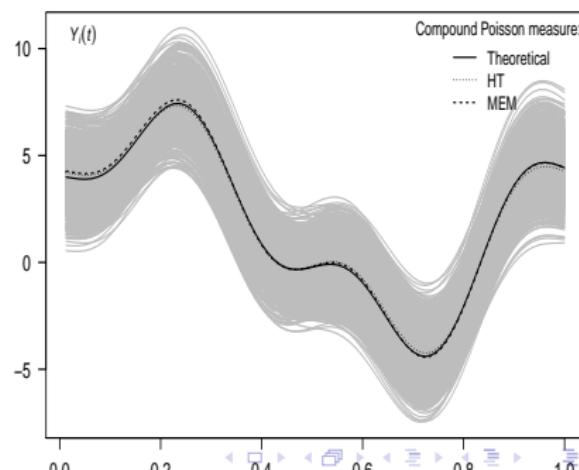
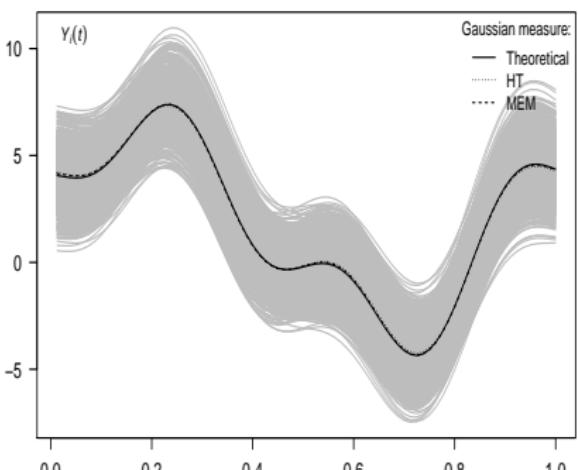
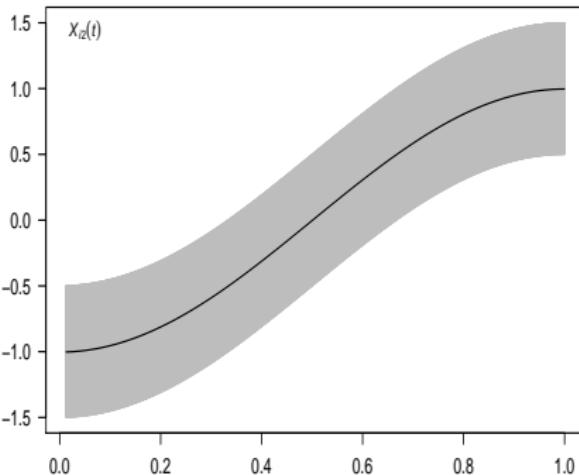
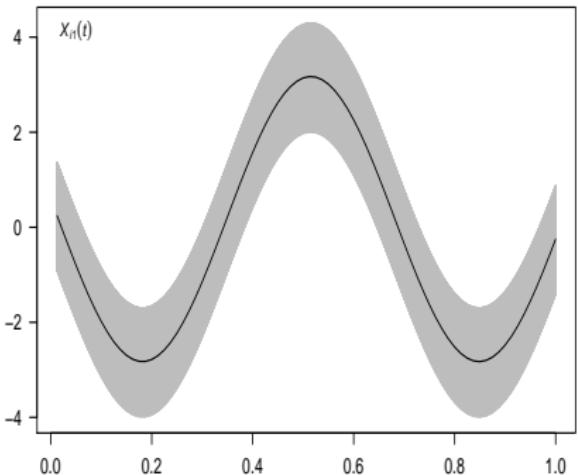
where

- $\alpha(t) = 1.2 + 2.3 \cos(2\pi t) + 4.2 \sin(2\pi t)$ ,
- $\boldsymbol{\beta}(t) = (\beta_1(t), \beta_2(t))^\top$   $\beta_1(t) = \cos(10t)$  and  $\beta_2(t) = t \sin(15t)$ ,
- $\varepsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2(1+t))$  with  $\sigma_\varepsilon^2 = 0.1$ , independent of  $\mathbf{X}_i(t)$ .
- $\mathbf{X}_i(t) = (X_{i1}(t), X_{i2}(t))^\top$ , where  $X_{i1}(t) = \mathcal{U}_{i1} + f_1(t)$  and  $X_{i2}(t) = \mathcal{U}_{i2} + f_2(t)$  with  $f_1(t) = 3 \sin(3\pi t + 3)$ ,  
 $f_2(t) = -\cos(\pi t)$ ,  $\mathcal{U}_{i1} \stackrel{\text{i.i.d.}}{\sim} U[-1, 1.3]$  and  $\mathcal{U}_{i2} \stackrel{\text{i.i.d.}}{\sim} U[-0.5, 0.5]$ .
- Uniform fixed-size sampling design sample  $a \in U_N$  of  $n = 0.12N$  elements without replacement, with  $N = 1000$ .
- $K(t, s) = \exp\left\{-|t-s|^2/2\sigma^2\right\}$  with  $\sigma^2 = 0.5$ .
- For the compound Poisson case  $\xi_i \stackrel{\text{i.i.d.}}{\sim} U[-1, 1]$ , and  $\gamma = 1$ .

## Simulation study II

Table : Bias-variance decomposition of MSE

Functional estimator	MSE	Bias <sup>2</sup>	Variance
Horvitz-Thompson	0.2391	0.0005	0.2386
MEM (Gaussian)	0.2001	0.0006	0.1995
MEM (Poisson)	0.2333	0.0084	0.2249



# References I

- M. Chaouch and C. Goga. Using complex surveys to estimate the  $l_1$ -median of a functional variable: application to electricity load curves. *Int. Stat. Rev.*, 80:40–59, 2012.
- I. Csiszár.  $I$ -divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3:146–158, 1975.
- I. Csiszár. Sanov property, generalized  $I$ -projection and a conditional limit theorem. *The Annals of Probability*, 12:768–793, 1984.
- J. C. Deville and C. E. Särndal. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382, 1992.
- H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht, 2000.
- W. A. Fuller. *Sampling Statistics*. Wiley Series in Survey Methodology. Wiley, 2009.
- F. Gamboa. *Méthode du maximum d'entropie sur la moyenne et applications*. PhD thesis, Université de Paris-Sud, Orsay, 1989.
- H. Gzyl and Y. Velásquez. *Linear Inverse Problems: the Maximum Entropy Connection*, volume 83 of *Series on Advances in Mathematics for Applied Sciences*. World Scientific, 2011.
- D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685, 1953.
- J. Navaza. On the maximum entropy estimate of electron density function. *Acta Crystallographica*, A41:232–244, 1985.
- J. Navaza. The use of non-local constraints in maximum-entropy electron density reconstruction. *Acta Crystallographica*, A42:212–223, 1986.

# Thanks!!!

$$Z_v(\boldsymbol{\lambda}) = \exp \left\{ \mathbb{E}_v [\langle \boldsymbol{\lambda}, \mathcal{K}\mathcal{W} \rangle] + \frac{1}{2} \text{Var}_v [\langle \boldsymbol{\lambda}, \mathcal{K}\mathcal{W} \rangle] \right\} =$$

$$\exp \left\{ \sum_{i \in a} d_i \int_0^1 \boldsymbol{\lambda}^\top (\mathrm{d}t) \mathbf{X}_i(t) + \frac{1}{2} \int_0^1 \left( \sum_{i \in a} \int_0^1 K(s, t) \boldsymbol{\lambda}^\top (\mathrm{d}t) \mathbf{X}_i(t) \right)^2 \mathrm{d}s \right\}$$

due to  $\mathbb{E}_v [d\mathcal{W}_i(s)] = 0$ , and  $\text{Var}_v [d\mathcal{W}_i(s)] = ds, i \in a$ .

$$\begin{aligned}
H_v(\boldsymbol{\lambda}) &= \log Z_v(\boldsymbol{\lambda}) - \langle \boldsymbol{\lambda}, N\boldsymbol{\mu}_X \rangle = \\
&\frac{1}{2} \int_0^1 \left( \sum_{i \in a} \int_0^1 K(s, t) \boldsymbol{\lambda}^\top(dt) \mathbf{X}_i(t) \right) \left( \sum_{i' \in a} \int_0^1 K(s, t') \boldsymbol{\lambda}^\top(dt') \mathbf{X}_{i'}(t') \right) ds \\
&+ \int_0^1 \boldsymbol{\lambda}^\top(dt) \left( \sum_{i \in a} d_i \mathbf{X}_i(t) - N\boldsymbol{\mu}_X(t) \right).
\end{aligned}$$

$\boldsymbol{\lambda}^*(dt)$  that minimizes  $H_v(\boldsymbol{\lambda})$  is given by

$$\sum_{i \in a} \sum_{i' \in a} \int_0^1 \int_0^1 K(s, t) K(s, t') \mathbf{X}_i(t) \mathbf{X}_{i'}^\top(t') \boldsymbol{\lambda}^*(dt') ds + \sum_{i \in a} d_i \mathbf{X}_i(t) = N\boldsymbol{\mu}_X(t)$$

which can be rewritten as

$$\sum_{i \in a} \left[ \int_0^1 K(s, t) \left( \sum_{i' \in a} \int_0^1 K(s, t') \mathbf{X}_{i'}^\top(t') \boldsymbol{\lambda}^*(dt') \right) ds + d_i \right] \mathbf{X}_i(t) = N\boldsymbol{\mu}_X(t)$$

$$m_i((a, b]) \triangleq \mathcal{W}_i(b) - \mathcal{W}_i(a) = \sum_{k=N(a)+1}^{N(b)} \xi_{ik}.$$

By the Lévy-Khintchine formula for Lévy processes, the m.g.f. of  $\mathcal{W}(s)$  is

$$\mathbb{E}_v [\exp \{\langle \boldsymbol{\alpha}, \mathcal{W}(s) \rangle\}] = \exp \left\{ s\gamma \int_{\mathbb{R}^n} \left( e^{\langle \boldsymbol{\alpha}, \boldsymbol{\xi}_k \rangle} - 1 \right) u(d\boldsymbol{\xi}_k) \right\}, \quad \boldsymbol{\alpha} \in \mathbb{R}^n,$$

$$\mathbb{E}_v [\exp \{\langle g(s), d\mathcal{W}_i \rangle\}] = \exp \left\{ \gamma \int_0^1 ds \int_{\mathbb{R}} (\exp \{g(s) \xi_i\} - 1) u(d\xi_i) \right\}.$$

$$\langle \boldsymbol{\lambda}, \mathcal{K}\mathcal{W} \rangle$$

$$= \int_0^1 \boldsymbol{\lambda}^\top(dt) \int_0^1 \sum_{i \in a} K(s, t) \mathbf{X}_i(t) d\mathcal{W}_i(s) + \int_0^1 \boldsymbol{\lambda}^\top(dt) \sum_{i \in a} d_i \mathbf{X}_i(t)$$

where  $g(s) = \int_0^1 \boldsymbol{\lambda}^\top(dt) \sum_{i \in a} K(s, t) \mathbf{X}_i(t)$ .

$$Z_v(\boldsymbol{\lambda})$$

$$= \exp \left\{ \gamma \int_0^1 \mathrm{d}s \int_{\mathbb{R}} (\exp \{g(s)\xi_i\} - 1) u(\mathrm{d}\xi_i) + \left\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \mathbf{X}_i(t) \right\rangle \right\}$$

$$H_v(\boldsymbol{\lambda})$$

$$\begin{aligned} &= \gamma \int_0^1 \mathrm{d}s \int_{\mathbb{R}} \left( \exp \left\{ \int_0^1 \boldsymbol{\lambda}^\top(\mathrm{d}t) \sum_{i \in a} K(s,t) \xi_i \mathbf{X}_i(t) \right\} - 1 \right) u(\mathrm{d}\xi_i) \\ &\quad + \int_0^1 \boldsymbol{\lambda}^\top(\mathrm{d}t) \left( \sum_{i \in a} d_i \mathbf{X}_i(t) - N \boldsymbol{\mu}_X(t) \right). \end{aligned}$$

$$\begin{aligned} &\sum_{i \in a} \left[ \int_0^1 K(s,t) \left( \int_{\mathbb{R}} \xi_i \exp \left\{ \sum_{i \in a} \int_0^1 K(s,t) \xi_i \mathbf{X}_i^\top(t) \boldsymbol{\lambda}^*(\mathrm{d}t) \right\} u(\mathrm{d}\xi_i) \right) \mathrm{d}s \right. \\ &\quad \left. + d_i \right] \mathbf{X}_i(t) = N \boldsymbol{\mu}_X(t) \end{aligned}$$