# Functional calibration estimation by the maximum entropy on the mean principle

Santiago Gallón[*][a,b], Jean-Michel Loubes[†][b], and Fabrice Gamboa[‡][b]

[a]*Departamento de Matemáticas y Estadística, Facultad de Ciencias Económicas, Universidad de Antioquia, Medellín, Colombia.*
[b]*Institut de Mathématiques de Toulouse, Université Toulouse III Paul Sabatier, Toulouse, France.*

January 10, 2014

## Abstract

We extend the problem of obtaining an estimator for the finite population mean parameter incorporating complete auxiliary information through calibration estimation in survey sampling but considering a functional data framework. The functional calibration sampling weights of the estimator are obtained by matching the calibration estimation problem with the maximum entropy on the mean principle. In particular, the calibration estimation is viewed as an infinite dimensional linear inverse problem following the structure of the maximum entropy on the mean approach. We give a precise theoretical setting and estimate the functional calibration weights assuming, as prior measures, the centered Gaussian and compound Poisson random measures. Additionally, through a simple simulation study, we show that our functional calibration estimator improves its accuracy compared with the Horvitz-Thompson estimator.

**Key words:** Auxiliary information; Functional calibration weights; Functional data; Infinite dimensional linear inverse problems; Survey sampling.

## 1 Introduction

In survey sampling, the well-known calibration estimation method proposed by Deville and Särndal [7] allows to construct an estimate for the finite population total or mean of a survey variable by incorporating complete auxiliary information on the study population in order to improve its efficiency. The main idea of the calibration method consists in modifying the standard sampling design weights $d_i$ of the unbiased Horvitz-Thompson estimator Horvitz and Thompson [17] by new weights $w_i$ close enough to $d_i$'s according to some distance function $\mathcal{D}(w, d)$, while satisfying a linear calibration equation in which the auxiliary information is taken into account. The estimator based

[*]santiagog@udea.edu.co
[†]jean-michel.loubes@math.univ-toulouse.fr
[‡]gamboa@math.univ-toulouse.fr

on these new calibration weights is asymptotically design unbiased and consistent with a variance smaller than the Horvitz-Thompson one.

The idea of calibration has been extended to estimate other finite population parameters, such as finite population variances, distribution functions and quantiles. See, for instance, Rao et al. [26], Kovaĉević [18], Théberge [31], Singh [30], Wu and Sitter [35], Wu [34], Harms and Duchesne [15], Rueda et al. [27], Särndal [28], and references therein. Recent developments have also been conducted toward, for example, the approach of (parametric and non-parametric) non-linear relationships between the survey variable and the set of auxiliary variables for the underlying assisting model, and a broad classes of conceivable calibration constraints functions (Breidt and Opsomer [1], Wu and Sitter [35], Wu [34], Montanari and Ranalli [21]).

One interesting extension emerges when both the survey and auxiliary variables are considered as infinite dimensional objects such as random functions. This generalization relies on the fact that, due to improvements in data collection technologies, large and complex databases are being registered frequently at very fine time scales, regarded these as functional datasets. This kind of data are collected in many scientific fields as molecular biology, astronomy, marketing, finance, economics, among many other. A depth overview on functional data analysis can be found in Ramsay and Silverman [24], Ramsay and Silverman [25] and Horváth and Kokoszka [16]. Functional versions of the Horvitz-Thompson estimator have been proposed recently by Cardot and Josserand [2] and Cardot et al. [3] for the cases of error free and noisy functional data, respectively.

The purpose of the present paper is to extend the problem of obtaining calibration sampling weights using functional data. This is conducted through the generalization of the work by Gamboa et al. [11], where the calibration estimation problem, which is considered as a linear inverse problem following Théberge [31], is matched with the maximum entropy on the mean approach under a finite dimensional setting. The maximum entropy on the mean principle applied to our goal focuses on reconstructing an unique posterior measure $\nu^*$ that maximizes the entropy $S(\nu \parallel \upsilon)$ between a feasible finite measure $\nu$ relative to a given prior measure $\upsilon$ subject to a linear constraint. Finally, the functional calibration sampling weights are defined as the mathematical expectation with respect to $\nu^*$ of a random variable with mean equal to the standard sampling design weights $d_i$. In this paper, we reconstruct $\nu^*$ adopting the random measure approach by Gzyl and Velásquez [14] under an infinite dimensional context.

The maximum entropy method on the mean was introduced by Navaza [22, 23] to solve an inverse problem in crystallography, and has been further investigated, from a mathematical point of view, by Gamboa [9], Dacunha-Castelle and Gamboa [6] and Gamboa and Gassiat [10]. Complementary references on the approach are Mohammad-Djafari [20], Maréchal [19], Gzyl [13], Gzyl and Velásquez [14] and Golan and Gzyl [12]. Maximum entropy solutions, as an alternative to the Tikhonov's regularization of ill-conditioned inverse problems, provide a very simple and natural way to incorporate constraints on the support and the range of the solution Gamboa and Gassiat [10], and its usefulness has been proven, e.g., in crystallography, seismic tomography and image reconstruction.

The paper is organized as follows. Sect. 2, presents the calibration estimation framework for the functional finite population mean. In Sect. 3, the connection between calibration and maximum entropy on the mean approaches is established, and the functional calibration sampling weights are obtained assuming two prior measures. In Sect. 4, the respective approximations of the functional maximum entropy on the mean estimators are derived. The performance of the estimator is studied through a simple simulation study in Sect. 5. Some concluding remarks are given in Sect. 6. Finally, the technical proofs of the technical results are gathered in the Appendix.

## 2 Calibration estimation for the functional finite population mean

Let $U_N = \{1, \ldots, N\}$ be a finite survey population from which a realized sample $a$ is drawn with fixed-size sampling design $p_N(a) = \mathbb{P}(A = a)$. Here $a \in \mathcal{A}$, where $\mathcal{A}$ is the collection of all subsets $A$ of $U_N$ that contains all possible samples of $n_N$ different elements randomly drawn from $U_N$ according to a given sampling selection scheme, and $\mathbb{P}$ a probability measure on $\mathcal{A}$. The first order inclusion probabilities, $\pi_{iN} = \mathbb{P}(i \in a) = \sum_{a \in A(i)} p_N(a)$, where $A(i)$ represents the set of samples that contain the $i$th element, are assumed to be strictly positive for all $i \in U_N$. See Särndal et al. [29] and Fuller [8] for details about survey sampling.

Associated with the $i$th element in $U_N$ there exists an unique functional random variable $Y_i(t)$ with values in the space of all continuous real-valued functions defined on $[0, T]$ with $T < +\infty$, $\mathcal{C}([0, T])$. However, only the sample functional data, $Y_i(t)$, $i \in a$ are observed. Additionally, an auxiliary $q$-dimensional functional vector is available for each $i \in U_N$, $\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{iq}(t))^\top \in \mathcal{C}([0, T]^q)$ with $q \geq 1$. The known functional finite population mean is denoted by $\boldsymbol{\mu}_X(t) = N^{-1} \sum_{i \in U_N} \boldsymbol{X}_i(t)$.

The main goal is to obtain a design consistent estimator for the unknown functional finite population mean, $\mu_Y(t) = N^{-1} \sum_{i \in U_N} Y_i(t)$, based on the calibration method. The idea consists in modify the basic sampling design weights, $d_i = \pi_i^{-1}$, of the unbiased functional Horvitz-Thompson estimator defined by $\hat{\mu}_Y^{HT}(t) = N^{-1} \sum_{i \in a} d_i Y_i(t)$, for new more efficient weights $w_i > 0$ incorporating the auxiliary information. These weights must to be sufficiently close to $d_i$'s according to some dissimilarity distance function $\mathcal{D}_a(w, d)$ on $\mathbb{R}^n_+$, and satisfying the set of calibration constraints

$$N^{-1} \sum_{i \in a} w_i \boldsymbol{X}_i(t) = \boldsymbol{\mu}_X(t).$$

The functional estimator for $\mu_Y(t)$ based on the calibration weights is expressed by the linear weighted estimator $\hat{\mu}_Y(t) = N^{-1} \sum_{i \in a} w_i Y_i(t)$. Different calibration estimators can be obtained depending on the chosen distance function Deville and Särndal [7]. However, it is well known that, in the finite dimensional setting, all of calibration estimators are asymptotically equivalent to the one obtained through the use of the popular chi-square distance function $\mathcal{D}_a(w, d) = \sum_{i \in a} (w_i - d_i)^2 / 2 d_i q_i$, where $q_i$ is an individual given positive weight uncorrelated with $d_i$.

Assuming a point-wise multiple linear regression model Ramsay and Silverman [25], $Y_i(t) = \boldsymbol{X}_i(t)^\top \boldsymbol{\beta}(t) + \varepsilon_i(t)$, where $\varepsilon_i(t)$ is the $i$th zero-mean measurement functional error independent of $\boldsymbol{X}_i(t)$ with variance structure given by a diagonal matrix with elements $1/q_i$ unrelated to $d_i$, then the estimator for $\mu_Y(t)$ from the restricted minimization problem can be expressed as

$$\hat{\mu}_Y(t) = \hat{\mu}_Y^{HT}(t) + \left\{ \boldsymbol{\mu}_X(t) - \hat{\boldsymbol{\mu}}_X^{HT}(t) \right\}^\top \widehat{\boldsymbol{\beta}}(t),$$

where $\hat{\boldsymbol{\mu}}_X^{HT}(t) = \sum_{i \in a} d_i \boldsymbol{X}_i(t)$ denotes the Horvitz-Thompson estimator for the functional vector $\boldsymbol{X}(t)$, and $\widehat{\boldsymbol{\beta}}(t) = \left\{ \sum_{i \in a} d_i q_i \boldsymbol{X}_i(t) \boldsymbol{X}_i(t)^\top \right\}^{-1} \sum_{i \in a} d_i q_i \boldsymbol{X}_i(t) Y_i(t)$ is the weighted estimator of the functional coefficient vector $\boldsymbol{\beta}(t)$, whose uniqueness relies on the existence of the inverse of the matrix $\sum_{i \in a} d_i q_i \boldsymbol{X}_i(t) \boldsymbol{X}_i(t)^\top$ for all $t$.

The calibration weights can be generalized allowing functional calibration weights $w_i(t)$ which can be obtained from the minimization of the generalized chi-square distance $\mathcal{D}_a^*(w, d)$, expressed below, subject to the functional calibration restriction

$$N^{-1} \sum_{i \in a} w_i(t) \boldsymbol{X}_i(t) = \boldsymbol{\mu}_X(t). \tag{1}$$

The existence of functional calibration weights is stated in the next theorem, which is a straightforward generalization of the finite dimensional results of Deville and Särndal [7].

**Theorem 1.** *Assume the existence of a functional vector $\boldsymbol{w}(t) = (w_1(t), \ldots, w_n(t))^\top$ such that (1) holds, and the inverse of the matrix $\sum_{i \in a} d_i q_i(t) \boldsymbol{X}_i(t) \boldsymbol{X}_i(t)^\top$. Then, for a fixed $t \in [0, T]$, $\hat{\boldsymbol{w}}(t)$ minimizes over $\mathcal{C}([0, T]^n)$ the generalized chi-square distance*

$$\mathcal{D}_a^*(w, d) = \sum_{i \in a} \frac{\left( w_i(t) - d_i \right)^2}{2 d_i q_i(t)}$$

*subject to (1), where the functional calibration weight $\hat{w}_i(t)$ for all $i \in a$ is given by*

$$\hat{w}_i(t) = d_i \left[ 1 + q_i(t) \left\{ \boldsymbol{\mu}_X(t) - \hat{\boldsymbol{\mu}}_X^{HT}(t) \right\}^\top \left\{ \sum_{i \in a} d_i q_i(t) \boldsymbol{X}_i(t) \boldsymbol{X}_i(t)^\top \right\}^{-1} \boldsymbol{X}_i(t) \right].$$

Note that, for this generalized setting, the functional calibration estimator for $\mu_Y(t)$ is expressed by

$$\hat{\mu}_Y(t) = N^{-1} \sum_{i \in a} \hat{w}_i(t) Y_i(t) = \hat{\mu}_Y^{HT}(t) + \left\{ \boldsymbol{\mu}_X(t) - \hat{\boldsymbol{\mu}}_X^{HT}(t) \right\}^\top \widehat{\boldsymbol{\beta}}(t),$$

where

$$\widehat{\boldsymbol{\beta}}(t) = \left\{ \sum_{i \in a} d_i q_i(t) \boldsymbol{X}_i(t) \boldsymbol{X}_i(t)^\top \right\}^{-1} \sum_{i \in a} d_i q_i(t) \boldsymbol{X}_i(t) Y_i(t),$$

provided the inverse of the matrix $\sum_{i \in a} d_i q_i(t) \boldsymbol{X}_i(t) \boldsymbol{X}_i(t)^\top$ exists for all $t$.

# 3 Maximum entropy on the mean for survey sampling

Let $(\widetilde{\mathcal{X}}, \mathcal{F})$ be an arbitrary measurable space over which we want to search for an $\sigma$-finite positive measure $\mu$. The maximum entropy on the mean principle provides an efficient way of getting an estimator for some linear functional $\mu_{\widetilde{Y}}(t) = \int_{\widetilde{\mathcal{X}}} \widetilde{Y}(t) d\mu$ satisfying a known $q$-vector of functionals $\int_{\widetilde{\mathcal{X}}} \widetilde{\boldsymbol{X}}(t) d\mu = \boldsymbol{\mu}_X(t)$, where $\widetilde{Y}(t) \colon \widetilde{\mathcal{X}} \to \mathcal{C}([0,T])$ and $\widetilde{\boldsymbol{X}}(t) \colon \widetilde{\mathcal{X}} \to \mathcal{C}([0,T]^q)$.

A natural unbiased and consistent estimator of $\mu_{\widetilde{Y}}(t)$ is the empirical functional mean $\hat{\mu}_{\widetilde{Y}}(t) = \int_{\mathcal{X}} \widetilde{Y}(t) d\mu_n = n^{-1} \sum_{i \in a} \widetilde{Y}_i(t)$, where $\mu_n = n^{-1} \sum_{i \in a} \delta_{T_i}$ is the corresponding empirical distribution with $T_1, \ldots, T_n$ an observed random sample from $\mu$. Despite properties of this estimator, it may not have the smallest variance in this kind of framework. Therefore, incorporating prior functional auxiliary information the variance of an asymptotically unbiased functional estimator can be reduced applying the maximum entropy on the mean principle.

The philosophy of the principle consists in to enhance $\hat{\mu}_{\widetilde{Y}}(t)$ considering the maximum entropy on the mean functional estimator

$$\hat{\mu}_{\widetilde{Y}}^{MEM}(t) = \int_{\mathcal{X}} \widetilde{Y}(t) d\hat{\mu}_n^{MEM} = n^{-1} \sum_{i \in a} \hat{p}_i(t) \widetilde{Y}_i(t), \quad \text{for all } t \in [0,T] \,,$$

where $\hat{\mu}_n^{MEM} = n^{-1} \sum_{i \in a} \hat{p}_i(t) \delta_{T_i}$ is a weighted version of the empirical distribution $\mu_n$, with $\hat{\boldsymbol{p}}(t) = (\hat{p}_1(t), \ldots, \hat{p}_n(t))^\top$ given by the expectation of the independent $n$-dimensional stochastic process $\boldsymbol{P}(t) = (P_1(t), \ldots, P_n(t))^\top$ drawn from a posterior finite distribution $\nu^*$, $\hat{\boldsymbol{p}}(t) = \mathbb{E}_{\nu^*}[\boldsymbol{P}(t)]$ for all $t \in [0,T]$, where $\nu^*$ must to be close to a prior distribution $\upsilon$, which transmits the information that $\hat{\mu}_n^{MEM}$ must to be sufficiently close to $\mu_n$.

Therefore, the maximum entropy on the mean principle focuses on reconstructing the posterior measure $\nu^*$ that maximizes the entropy, over the convex set of all probability measures, $S(\nu \parallel \upsilon) = -D(\nu \parallel \upsilon)$ subject to the linear functional constraint holds in mean,

$$\mathbb{E}_{\nu^*}\left[ n^{-1} \sum_{i \in a} P_i(t) \widetilde{\boldsymbol{X}}_i(t) \right] = \boldsymbol{\mu}_X(t), \qquad \forall t \in [0,T] \,.$$

We recall that $D(\nu \parallel \upsilon)$ is the $I$-divergence or relative divergence or Kullbach-Leibler information divergence between a feasible finite measure $\nu$ with respect to a given prior measure $\upsilon$ (see, for instance, Csiszár [4]) defined by

$$D(\nu \parallel \upsilon) = \begin{cases} \int_\Omega \log\left(\frac{d\nu}{d\upsilon}\right) d\nu - \nu(\Omega) + 1 & \text{if } \nu \ll \upsilon \\ +\infty & \text{otherwise.} \end{cases}$$

To establish the connection between calibration and maximum entropy on the mean approaches the following notation is adopted $\widetilde{Y}_i(t) = N^{-1} n d_i Y_i(t)$, $\widetilde{\boldsymbol{X}}_i(t) =$

$N^{-1}nd_i\boldsymbol{X}_i(t)$ and $p_i(t) = \pi_i w_i(t)$, such that the functional Horvitz-Thompson estimator of $\mu_Y(t)$ and the functional calibration constrain (1) can be, respectively, expressed as

$$\hat{\mu}_Y^{HT}(t) = N^{-1}\sum_{i\in a} d_i Y_i(t) = n^{-1}\sum_{i\in a} \widetilde{Y}_i(t)$$

and

$$n^{-1}\sum_{i\in a} p_i(t)\widetilde{\boldsymbol{X}}_i(t) = N^{-1}\sum_{i\in a} w_i(t)\boldsymbol{X}_i(t) = \boldsymbol{\mu}_X(t), \qquad \forall t \in [0,T]\,.$$

Hence, the functional calibration estimation problem follows the structure of the maximum entropy on the mean principle, where the corresponding estimator is defined by

$$\hat{\mu}_Y^{MEM}(t) = n^{-1}\sum_{i\in a} \hat{p}_i(t)\widetilde{Y}_i(t) = N^{-1}\sum_{i\in a} \hat{w}_i(t)Y_i(t).$$

The functional calibration weighting vector $\hat{\boldsymbol{p}}(t)$ with coordinates $\hat{p}_i(t) = \pi_i\hat{w}_i(t)$ for $i \in a$, is the expectation of the $n$-dimensional stochastic process $\boldsymbol{P}(t)$ with coordinates $P_i(t) = \pi_i W_i(t)$, drawn from $\nu^*$,

$$\hat{\boldsymbol{p}}(t) = \mathbb{E}_{\nu^*}\big[\boldsymbol{P}(t)\big], \qquad \forall t \in [0,T]\,,$$

where the posterior measure $\nu^* = \otimes_{i\in a}\nu_i^*$ (by the independence of $P_i$'s) maximizes the entropy $S(\cdot \parallel v)$ subject to the calibration constraint is fulfilled in mean,

$$\mathbb{E}_{\nu^*}\left[n^{-1}\sum_{i\in a} P_i(t)\widetilde{\boldsymbol{X}}_i(t)\right] = \mathbb{E}_{\nu^*}\left[N^{-1}\sum_{i\in a} W_i(t)\boldsymbol{X}_i(t)\right] = \boldsymbol{\mu}_X(t), \quad \forall t \in [0,T]\,.$$

Note that as $p_i(t) = \pi_i w_i(t)$ and $\hat{w}_i(t)$ must to be sufficiently close to $d_i$, then the $\hat{p}_i(t)$ must be close enough to 1 for each $i \in a$.

## 3.1 Reconstruction of the measure $\nu^*$

For simplicity and without loss generality we assume that $T = 1$. The posterior distribution $\nu^*$ can be reconstructed adopting the random measure approach for infinite dimensional inverse problems explained in detail by Gzyl and Velásquez [14]. To do this, we express the calibration constraint (1) 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 known continuous, real-valued and bounded kernel function and $\varpi_i = \mathbb{E}_\nu[\mathcal{W}_i(s)]$, where $\mathcal{W}$ is a stochastic process.

Hence, the infinite dimensional inverse problem, which takes the form of a Fredholm integral equation of the first kind, is

$$
\begin{aligned}
\mathbb{E}_\nu\left[\mathcal{KW}\right] &= \mathbb{E}_\nu\left\{\sum_{i\in a}\left[\int_0^1 K(s,t)\mathcal{W}_i(s)\,ds + d_i\right]\boldsymbol{X}_i(t)\right\} \\
&= \int_0^1\sum_{i\in a}K(s,t)\boldsymbol{X}_i(t)\varpi_i(s)\,ds + \sum_{i\in a}d_i\boldsymbol{X}_i(t) \\
&= N\boldsymbol{\mu}_X(t), \qquad t\in[0,1].
\end{aligned}
\tag{2}
$$

To obtain the functions $\varpi_i^*(s)$ that solve the integral equation $\mathbb{E}_\nu\left[\mathcal{KW}\right] = N\boldsymbol{\mu}_X(t)$, the random measure approach adopted considers $\varpi_i(s)$ as a density of a measure $\varpi_i(s)\,ds$, $i\in a$. Under this setting, we define the random measure $\mathcal{W}_i(a,b] = \mathcal{W}_i(b) - \mathcal{W}_i(a)$ for $(a,b]\subset[0,1]$ such that $dE_\nu\{\mathcal{W}_i(0,s]\} = \varpi_i(s)ds$ for each $i\in a$. The next theorem ensures the existence of the posterior distribution $\nu^*$ to obtain the functions $\varpi_i^*(s)$ depending on the assumed prior distribution $\upsilon$.

**Theorem 2.** *Let $\upsilon$ be a prior positive probability measure, $\boldsymbol{\lambda} = \boldsymbol{\lambda}(t)$ a measure in the class of continuous measures on $[0,1]^q$, $\mathcal{M}(C[0,1]^q)$, and $\mathcal{V} = \{\nu\ll\upsilon\colon Z_\upsilon(\boldsymbol{\lambda}) < +\infty\}$ a nonempty open class, where $Z_\upsilon(\boldsymbol{\lambda}) = \mathbb{E}_\upsilon[\exp\{\langle\boldsymbol{\lambda},\mathcal{KW}\rangle\}]$, with*

$$
\langle\boldsymbol{\lambda},\mathcal{KW}\rangle = \int_0^1\boldsymbol{\lambda}^\top(dt)\left(\int_0^1\sum_{i\in a}K(s,t)\boldsymbol{X}_i(t)d\mathcal{W}_i(s) + \sum_{i\in a}d_i\boldsymbol{X}_i(t)\right).
\tag{3}
$$

*Then there exists an unique probability measure*

$$
\nu^* = \arg\max_{\nu\in\mathcal{V}}S(\nu\parallel\upsilon),
$$

*subject to $E_\nu\left[\mathcal{KW}\right] = N\boldsymbol{\mu}_X(t)$, which is achieved at*

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

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

$$
H_\upsilon(\boldsymbol{\lambda}) = \log Z_\upsilon(\boldsymbol{\lambda}) - \langle\boldsymbol{\lambda}, N\boldsymbol{\mu}_X\rangle.
$$

Based on the Theorem 2, we will carry out the reconstruction of $\nu$, assuming the centered Gaussian and compound Poisson random measures as prior measures, in order to estimate the respective functional calibration weights $\hat{w}_i(t)$, $i\in a$. The estimates are given by the following two Lemmas.

**Lemma 1.** *Let $\upsilon$ be a centered stationary Gaussian measure on $\left(\mathcal{C}([0,1]), \mathcal{B}(\mathcal{C}([0,1]))\right)$, and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(t)\in\mathcal{M}(C[0,1]^q)$. Then, $\hat{w}_i(t) = \int_0^1 K(s,t)\varpi^*(s)ds + d_i$ $i\in a$, where*

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

**Lemma 2.** *Let* $\mathcal{W}_i(s) = \sum_{k=1}^{N(s)} \xi_{ik}$ *be a compound Poisson process, where $N(s)$ is a homogeneous Poisson process on $[0,1]$ with intensity parameter $\gamma > 0$, and $\xi_{ik}$, $k \geq 1$ are independent and identically distributed real-valued random variables for each $i \in a$ with distribution $u$ on $\mathbb{R}$ satisfying $u(\{0\}) = 0$, and independent of $N(s)$. Then, $\hat{w}_i(t) = \int_0^1 K(s,t) \varpi^*(s) ds + d_i$ $i \in a$, where*

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

# 4 Approximation of the maximum entropy on the mean functional estimator

To approximate the functional calibration weights and the functional maximum entropy on the mean estimator for the finite population mean of $Y(t)$ with the assumed prior measure, an Euler discretization scheme is used. Consider a partition of $(s,t) \in [0,1]^2$ in $J$ and $L$ equidistant fixed points, $(j-1)/J < s_j \leq j/J$, $j = 1, \ldots, J$, $(l-1)/L < t_l \leq l/L$, $l = 1, \ldots, L$, respectively. For the corresponding prior measures, the approximations for functions $Z_\upsilon(\boldsymbol{\lambda})$, $H_\upsilon(\boldsymbol{\lambda})$ and $\boldsymbol{\lambda}^*(t)$ are based on the respective results found in the Appendix.

## 4.1 Centered Gaussian measure

For a prior centered Gaussian random measure, the approximations of the linear moment calibration constraint (2) and the inner product $\langle \boldsymbol{\lambda}, \mathcal{K}\mathcal{W} \rangle$ are, respectively, given by

$$\mathbb{E}_\nu \left[ \sum_{j=1}^J \sum_{i \in a} K(s_j, t_l) \Delta \mathcal{W}_i(s_j) \boldsymbol{X}_i(t_l) + \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right] = N \boldsymbol{\mu}_X(t_l)$$

and

$$\frac{1}{L} \sum_{l=1}^L \boldsymbol{\lambda}^\top(t_l) \sum_{j=1}^J \sum_{i \in a} K(s_j, t_l) \Delta \mathcal{W}_i(s_j) \boldsymbol{X}_i(t_l) + \frac{1}{L} \sum_{l=1}^L \boldsymbol{\lambda}^\top(t_l) \sum_{i \in a} d_i \boldsymbol{X}_i(t_l)$$

$$= \frac{1}{L} \sum_{j=1}^J \sum_{i \in a} \sum_{l=1}^L K(s_j, t_l) \Delta \mathcal{W}_i(s_j) \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) + \frac{1}{L} \sum_{i \in a} d_i \sum_{l=1}^L \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l),$$

where $\Delta \mathcal{W}_i(s_j) = \mathcal{W}_i(s_j) - \mathcal{W}_i(s_{j-1})$ is the discrete version of $d\mathcal{W}_i(s)$ for $i \in a$.

Therefore, we have that $Z_\upsilon(\boldsymbol{\lambda})$ is approximated at the grid (see equation (6) of the

proof of Lemma 1 in the Appendix) by

$$
\mathbb{E}_v \left[ \exp \left\{ \frac{1}{L} \sum_{i \in a} d_i \sum_{l=1}^{L} \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) + \frac{1}{L} \sum_{j=1}^{J} \sum_{i \in a} \sum_{l=1}^{L} K(s_j, t_l) \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) \Delta \mathcal{W}_i(s_j) \right\} \right]
$$

$$
= \exp \left\{ \frac{1}{L} \sum_{i \in a} d_i \sum_{l=1}^{L} \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) + \sum_{j=1}^{J} \frac{1}{2J} \left( \frac{1}{L} \sum_{i \in a} \sum_{l=1}^{L} K(s_j, t_l) \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) \right)^2 \right\}
$$

$$
= \exp \left\{ \frac{1}{L} \sum_{i \in a} d_i \sum_{l=1}^{L} \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) \right\} \prod_{j=1}^{J} \exp \left\{ \frac{1}{2J} \sum_{i \in a} \sum_{i' \in a} h_i(s_j) h_{i'}(s_j) \right\}
$$

$$
= \exp \left\{ \frac{1}{L} \sum_{i \in a} d_i \sum_{l=1}^{L} \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l) \right\} \prod_{j=1}^{J} z_i \left( h_i(s_j) \right),
$$

where $h_i(s_j) = L^{-1} \sum_{l=1}^{L} K(s_j, t_l) \boldsymbol{\lambda}^\top(t_l) \boldsymbol{X}_i(t_l)$, $i \in a$, $j = 1, \ldots J$, and $l = 1, \ldots L$.

Now, the finite dimensional maxentropic solution for $\varpi_i(s_j)$ for each $i \in a$ is approximated by (see Gzyl and Velásquez [14])

$$
\varpi_i^*(s_j) = \frac{\mathrm{d} \log z_i \left( h_i(s_j) \right)}{\mathrm{d}(2J)^{-1} h_i(s_j)} \bigg|_{h_i(s_j) = \mathcal{K} \boldsymbol{\lambda}^*}
$$

$$
= \sum_{i' \in a} h_{i'}(s_j) \bigg|_{h_i(s_j) = \mathcal{K} \boldsymbol{\lambda}^*} \tag{4}
$$

$$
= \frac{1}{L} \sum_{l=1}^{L} \sum_{i' \in a} K(s_j, t_l') \boldsymbol{\lambda}^{*\top}(t_l') \boldsymbol{X}_{i'}(t_l'),
$$

where the finite dimensional version of $\boldsymbol{\lambda}^*(t_l')$, $(l-1)/L < t_l \leq l/L$, $l = 1, \ldots, L$, is the minimizer of $H_v(\boldsymbol{\lambda})$, whose approximation (see equation (7) of the proof of Lemma 1 in the Appendix) is

$$
\frac{1}{2} \sum_{l=1}^{L} \sum_{l=1}^{L} \boldsymbol{\lambda}^\top(t_l) \left( \frac{1}{JL^2} \sum_{j=1}^{J} K(s_j, t_l) K(s_j, t_l') \sum_{i \in a} \sum_{i' \in a} \boldsymbol{X}_i(t_l) \boldsymbol{X}_{i'}^\top(t_l') \right) \boldsymbol{\lambda}(t_l')
$$

$$
+ \frac{1}{L} \sum_{l=1}^{L} \left( \sum_{i \in a} d_i \boldsymbol{X}_i^\top(t_l) - N \boldsymbol{\mu}_X^\top(t_l) \right) \boldsymbol{\lambda}(t_l).
$$

The first order condition (see equation(8)) associated to this minimization problem is

$$
\frac{1}{JL^2} \sum_{j=1}^{J} \sum_{l=1}^{L} K(s_j, t_l) K(s_j, t_l') \sum_{i \in a} \sum_{i' \in a} \boldsymbol{X}_i(t_l) \boldsymbol{X}_{i'}^\top(t_l') \boldsymbol{\lambda}^*(t_l')
$$

$$
+ \frac{1}{L} \left( N \boldsymbol{\mu}_X(t_l) - \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right) = \mathbf{0},
$$

whose solution $\boldsymbol{\lambda}^*(t_l')$ is given by

$$\boldsymbol{\lambda}^*(t_l') = JL \left( \sum_{j=1}^{J} \sum_{l=1}^{L} K(s_j, t_l) K(s_j, t_l') \sum_{i \in a} \sum_{i' \in a} \boldsymbol{X}_i(t_l) \boldsymbol{X}_{i'}^{\top}(t_l') \right)^{-1}$$
$$\times \left( N \boldsymbol{\mu}_X(t_l) - \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right).$$

Finally, the approximation of the finite dimensional solution of $\hat{w}_i(t)$ is

$$\hat{w}_i(t_l) = \frac{1}{J} \sum_{j=1}^{J} K(s_j, t_l) \varpi_i^*(s_j) + d_i,$$

where $\varpi_i^*(s_j)$ es given by the equation (4).

## 4.2 Compound Poisson measure

Based on equations (9) and (10) of the proof of Lemma 2 in the Appendix, the approximation of $Z_v(\boldsymbol{\lambda})$ is given by

$$\mathbb{E}_v \left[ \exp \left\{ \langle g(s_j), d\mathcal{W}_i \rangle + \left\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right\rangle \right\} \right]$$

$$= \exp \left\{ \left\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right\rangle \right\} \mathbb{E}_v \left[ \exp \left\{ \langle g(s_j), d\mathcal{W}_i \rangle \right\} \right]$$

$$= \exp \left\{ \left\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right\rangle \right\}$$

$$\times \prod_{j=1}^{J} \exp \left\{ \frac{\gamma}{J} \int_{\mathbb{R}} \left( \exp \left\{ \frac{1}{L} \sum_{i \in a} \xi_i \sum_{l=1}^{L} K(s_j, t_l) \boldsymbol{\lambda}^{\top}(t_l) \boldsymbol{X}_i(t_l) \right\} - 1 \right) u\,(d\xi_i) \right\}$$

$$= \exp \left\{ \left\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right\rangle \right\} \prod_{j=1}^{J} \exp \left\{ \frac{\gamma}{J} \int_{\mathbb{R}} \left( \exp \left\{ \sum_{i \in a} \xi_i h_i(s_j) \right\} - 1 \right) u\,(d\xi_i) \right\}$$

$$= \exp \left\{ \left\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \right\rangle \right\} \prod_{j=1}^{J} z_i\,(h_i(s_j)), \qquad i \in a,$$

where $h_i(s_j) = L^{-1} \sum_{l=1}^{L} K(s_j, t_l) \boldsymbol{\lambda}^{\top}(t_l) \boldsymbol{X}_i(t)$, $i \in a$, $j = 1, \dots J$, and $\langle \boldsymbol{\lambda}, \sum_{i \in a} d_i \boldsymbol{X}_i(t_l) \rangle = L^{-1} \sum_{i \in a} d_i \sum_{l=1}^{L} \boldsymbol{\lambda}^{\top}(t_l) \boldsymbol{X}_i(t_l)$.

The approximated maxentropic solution for $\varpi_i(s_j)$ for each $i \in a$ is

$$
\begin{aligned}
\varpi_i^*(s_j) &= \left.\frac{d \log z_i\left(h_i(s_j)\right)}{d h_i(s_j)}\right|_{h_i(s_j)=\mathcal{K}\boldsymbol{\lambda}^*} \\
&= \left.\frac{\gamma}{J} \int_{\mathbb{R}} \xi_i \exp\left\{\sum_{i \in a} \xi_i h_i(s_j)\right\} u\left(d\xi_i\right)\right|_{h_i(s_j)=\mathcal{K}\boldsymbol{\lambda}^*} \\
&= \frac{\gamma}{J} \int_{\mathbb{R}} \xi_i \exp\left\{\frac{1}{L}\sum_{i \in a}\sum_{l=1}^{L} K(s_j,t_l)\xi_i \boldsymbol{X}_i^{\top}(t_l)\boldsymbol{\lambda}^*(t_l)\right\} u\left(d\xi_i\right),
\end{aligned}
\tag{5}
$$

where the finite dimensional version of $\boldsymbol{\lambda}^*(t_l)$, is the minimizer of $H_v(\boldsymbol{\lambda})$, whose approximation, by the equation (11) of the proof of Lemma 2 in the Appendix, is

$$
\begin{aligned}
H_v(\boldsymbol{\lambda}) = {}&\frac{\gamma}{J}\sum_{j=1}^{J}\int_{\mathbb{R}}\left(\exp\left\{\frac{1}{L}\sum_{i \in a}\sum_{l=1}^{L}K(s_j,t_l)\xi_i\boldsymbol{X}_i^{\top}(t_l)\boldsymbol{\lambda}(t_l)\right\}-1\right)u\left(d\xi_i\right) \\
&+\frac{1}{L}\sum_{l=1}^{L}\left(\sum_{i \in a}d_i\boldsymbol{X}_i^{\top}(t_l)-N\boldsymbol{\mu}_X^{\top}(t_l)\right)\boldsymbol{\lambda}(t_l)
\end{aligned}
$$

The corresponding equation for $\boldsymbol{\lambda}^*(t_l)$ that minimizes $H_v(\boldsymbol{\lambda})$ is given by the nonlinear system of equations (see equation (12) in the Appendix)

$$
\begin{aligned}
\sum_{i \in a}\left[\frac{1}{J}\sum_{j=1}^{J}K(s_j,t_l)\left(\gamma L\int_{\mathbb{R}}\xi_i\exp\left\{\frac{1}{L}\sum_{i \in a}\sum_{l=1}^{L}K(s_j,t_l)\xi_i\boldsymbol{X}_i^{\top}(t_l)\boldsymbol{\lambda}^*(t_l)\right\}u\left(d\xi_i\right)\right)+d_i\right] \\
\times \boldsymbol{X}_i(t_l) = N\boldsymbol{\mu}_X(t_l)
\end{aligned}
$$

Finally, as in the Gaussian measure case, the finite dimensional solution of $\hat{w}_i(t)$ is approximated by $\hat{w}_i(t_l) = J^{-1}\sum_{j=1}^{J}K(s_j,t_l)\varpi_i^*(s_j) + d_i$ with $\varpi_i^*(s_j)$ given by the equation (5).

## 5 Simulation study

We shall illustrate through a simple simulation study the performance of results obtained in the above section. Considering a finite population $U_N$ of size $N = 1000$, we generate a functional random variable $Y_i(t)$ by the point-wise multiple linear regression model

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

where $\alpha(t) = 1.2 + 2.3\cos(2\pi t) + 4.2\sin(2\pi t)$, $\boldsymbol{\beta}(t) = (\beta_1(t), \beta_2(t))^{\top}$ with $\beta_1(t) = \cos(10t)$ and $\beta_2(t) = t\sin(15t)$, $\boldsymbol{X}_i(t) = (X_{i1}(t), X_{i2}(t))^{\top}$, and $\varepsilon_i(t) \sim \mathcal{N}\left(0, \sigma_\varepsilon^2(1+t)\right)$ with $\sigma_\varepsilon^2 = 0.1$, and independent of $\boldsymbol{X}_i(t)$. The auxiliary functional covariates are defined by $X_{i1}(t) = \mathcal{U}_{i1} + f_1(t)$ with $f_1(t) = 3\sin(3\pi t + 3)$, and $X_{i2}(t) = \mathcal{U}_{i2} + f_2(t)$ with $f_2(t) = -\cos(\pi t)$, where $\mathcal{U}_{i1}$ and $\mathcal{U}_{i2}$ are independent and, respectively, i.i.d. uniform random variables on the intervals $[-1, 1.3]$ and $[-0.5, 0.5]$.

The design time points for $t \in [0,1]$ and $s \in [0,1]$ are $t_j = j/J$, $j = 1, \ldots, J$ and $s_l = l/L$, $l = 1, \ldots, L$, with $J = 50$ and $L = 80$ The Figures 1 and 2 show, respectively, the simulated finite population auxiliary functional covariates and functional responses for each $i \in U_N$, and the respective finite population functional means, $\boldsymbol{\mu}_X(t) = (\mu_{X_1}(t), \mu_{X_2}(t))^\top$ and $\mu_Y(t) = N^{-1} \sum_{i \in U_N} Y_i(t)$. Assuming a uniform fixed-size sampling design we drawn a sample $a \in U_N$ of $n = 0.12N$ elements without replacement. For the kernel function we assumed a Gaussian one, $K(t,s) = \exp\left\{-|t-s|^2/2\sigma^2\right\}$ with $\sigma^2 = 0.5$. The random variables $\xi_i$ for the compound Poisson case are assumed i.i.d. uniform on the interval $[-1,1]$, and $\gamma = 1$. To solve the nonlinear system of equations for $\boldsymbol{\lambda}^*(t_l)$ in the compound Poisson case, we used the R-package BB (see Varadhan [32] and Varadhan and Gilbert [33]).
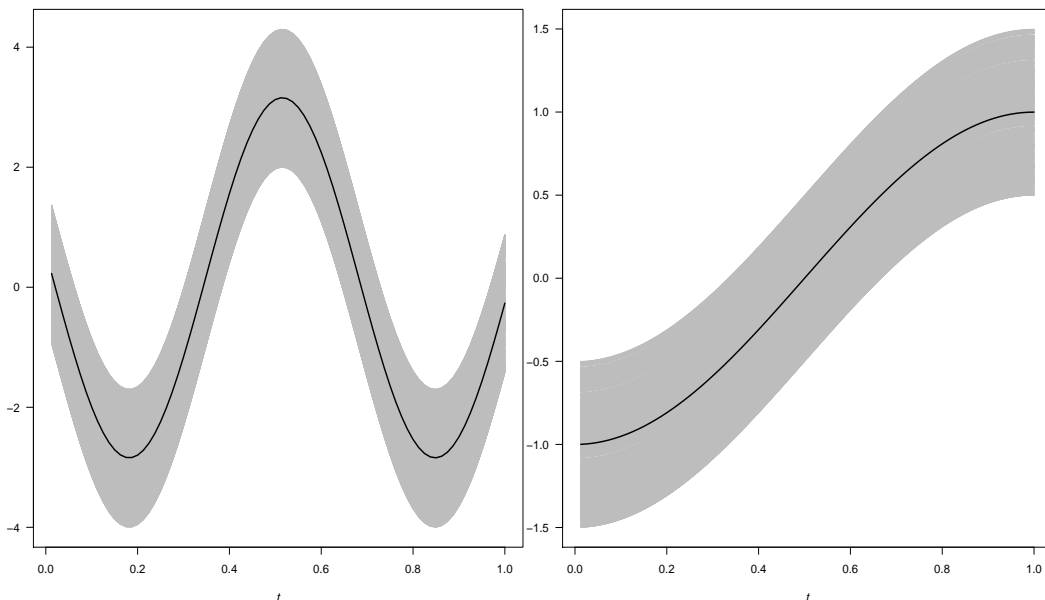


Figure 1: Population auxiliary functional variables (gray), $X_{i1}(t)$ (on the left) and $X_{i2}(t)$ (on the right). Functional finite population means, $\mu_{X_1}(t)$ and $\mu_{X_2}(t)$ (solid line)

The graphical comparisons of the estimators for a random selected repetition are illustrated in the Figure 2. The figure shows, in general, a good performance, specially for the estimator assuming the Gaussian measure. The principal differences with respect to the theoretical functional finite population mean are localized on the edges, particularly on the left edge. The Horvitz-Thompson estimator, in both cases, has a little departure localized around the deep valley. However our estimator has not this departure. A nice feature of the functional calibration method is that permits to check graphically how well the estimator satisfies the calibration constraints for each covariate, $N^{-1} \sum_{i \in a} \hat{w}_i(t) \boldsymbol{X}_i(t) = \boldsymbol{\mu}_X(t)$. This is illustrated in the Figure 3.

To evaluate the performance of the maximum entropic functional calibration estimator, $\hat{\mu}_Y^{MEM}(t)$, assuming the Gaussian and compound Poisson prior measures,
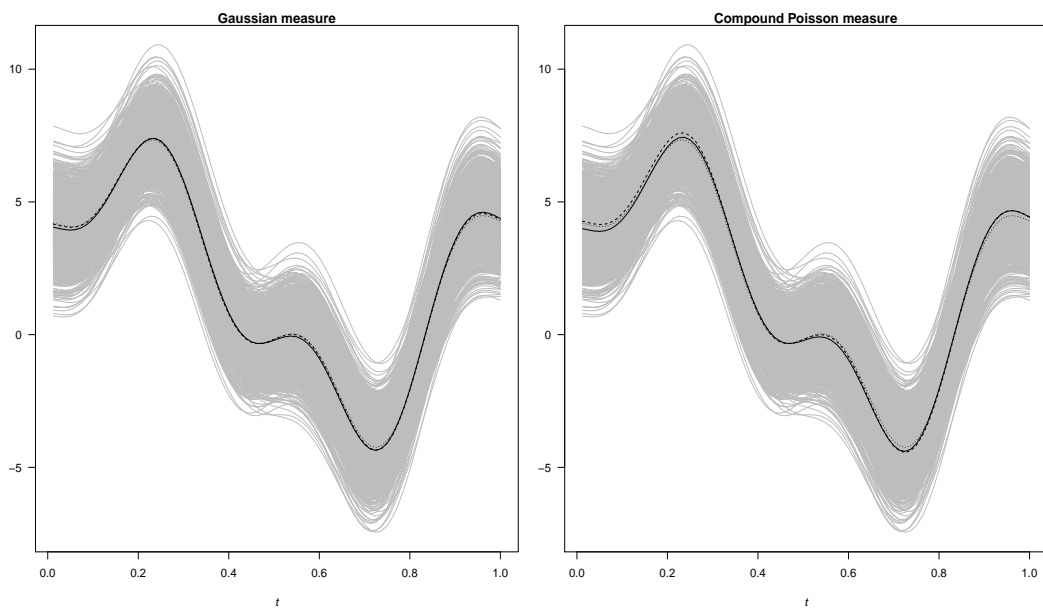
Figure 2: Population survey functions $Y_i(t)$ (in gray), functional finite population mean $\mu_Y(t)$ (solid line), and the Horvitz-Thompson (dotted line) and functional maximum entropy on the mean (dashed line) estimators
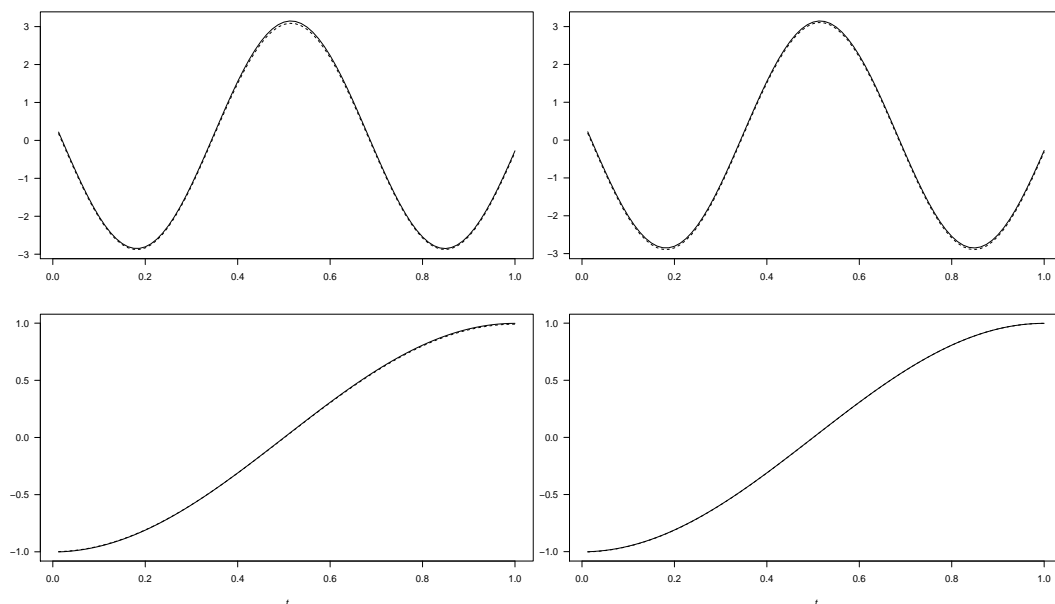


Figure 3: Functional calibration constraint (1) for Gaussian (on the left) and compound Poisson (on the right) measures. $\boldsymbol{\mu}_X(t)$ (solid line), $N^{-1}\sum_{i\in a}\hat{w}_i(t)\boldsymbol{X}_i(t)$ (dash)

Table 1: Bias-variance decomposition of MSE

| Functional estimator | MSE | Bias$^2$ | Variance |
|---|---|---|---|
| Horvitz-Thompson | 0.2391 | 0.0005 | 0.2386 |
| Maximum entropy on the mean (Gaussian) | 0.2001 | 0.0006 | 0.1995 |
| Maximum entropy on the mean (Poisson) | 0.2333 | 0.0084 | 0.2249 |

we calculated its empirical bias–variance decomposition of the mean square errors and compare it with the functional Horvitz-Thompson estimator $\hat{\mu}_Y^{HT}(t)$. The simulation study was conducted with 100 repetitions. In Table 1 we can see that, with respect to the Horvitz-Thompson estimator, the maximum entropic estimator has smaller variance and mean square error for both prior measures, particularly for the Gaussian prior. Although the Horvitz-Thompson estimator has smaller bias squared, the differences are not significant. Also, the small value for the bias confirm the unbiasedness of the functional maximum entropy on the mean and Horvitz-Thompson estimators.

## 6   Concluding remarks

In this paper we have proposed an extension to the problem of obtaining an estimator for the finite population mean of a survey variable incorporating complete auxiliary information under an infinite dimensional setting. Considering that both the survey and the set of auxiliary variables are functions, the respective functional calibration constraint is expressed as an infinite dimensional linear inverse problem, whose solution offers the functional survey weights of the calibration estimator. The solution of the problem is conducted by mean the maximum entropy on the mean principle, which is a powerful probabilistic-based regularization method to solve constrained linear inverse problems. Here we assume a centered Gaussian and compound Poisson random measures as prior measures to obtain the functional calibration weights. However, other random measures can be considered also.

The simulations study results show that the proposed functional calibration estimator improves its accuracy compared with the Horvitz-Thompson estimator. In the simulations, both the functional survey and auxiliary variables where assumed with amplitude variations (variation in the $y$-axis) only. More complex extensions allowing both amplitude and phase (variation in the $x$-axis) variations are possible.

Finally, a further interesting extension of the functional calibration estimation problem under the maximum entropy on the mean approach can be conducted following the idea of model-calibration proposed by Wu and Sitter [35], Wu [34] and Montanari and Ranalli [21]. This may be accomplished considering a nonparametric functional regression $Y_i(t) = \mu\left\{\boldsymbol{X}_i(t)\right\} + \varepsilon_i(t)$, $i \in U_N$, $t \in \big([0,T]$ to model the relation between the functional survey variable and the set of functional auxiliary covariates in order to allows a more effective use of the functional auxiliary information.

# Appendix

*Proof of Theorem 1.* The Lagrangian function associated to the restricted minimization problem is

$$L_a(\boldsymbol{w}, \boldsymbol{\lambda}) = \mathcal{D}_a^*(w, d) + \boldsymbol{\lambda}^\top(t)\left(\boldsymbol{\mu}_X(t) - N^{-1}\sum_{i \in a} w_i(t)\boldsymbol{X}_i(t)\right),$$

where $\boldsymbol{\lambda}(t)$ is the corresponding functional Lagrange multiplier vector. The first order conditions are

$$\frac{w_i(t) - d_i}{d_i q_i(t)} - \boldsymbol{\lambda}(t)^\top \boldsymbol{X}_i(t) = 0, \qquad i \in a$$

which can be expressed as

$$w_i(t) = d_i\left[1 + q_i(t)\boldsymbol{\lambda}(t)^\top \boldsymbol{X}_i(t)\right], \qquad i \in a$$

where, its uniqueness is guaranteed by the continuous differentiability of $\mathcal{D}_a^*(w, d)$ with respect to $w_i(t)$ for all $i \in a$, and by its strictly convexity.

From the functional calibration restriction (1) and by the existence assumption on the inverse of the matrix $\sum_{i \in a} d_i q_i(t)\boldsymbol{X}_i(t)\boldsymbol{X}_i(t)^\top$ for all $t$, the Lagrange functional multiplier vector is determined by

$$\hat{\boldsymbol{\lambda}}(t) = \left(\sum_{i \in a} d_i q_i(t)\boldsymbol{X}_i(t)\boldsymbol{X}_i(t)^\top\right)^{-1}\left(\boldsymbol{\mu}_X(t) - \hat{\boldsymbol{\mu}}_X^{HT}(t)\right).$$

Finally, replacing $\hat{\boldsymbol{\lambda}}(t)$ into the first order conditions, the calibration functional estimator $\hat{w}_i(t)$ of the Theorem is obtained. ∎

*Proof of Theorem 2.* Csiszár [5, Theorem 3, page 775]. ∎

*Proof of Lemma 1.* According to Theorem 2, the maximum of the entropy $S(\nu \parallel \upsilon)$ over the class $\mathcal{V} = \{\nu \ll \upsilon \colon Z_\upsilon(\boldsymbol{\lambda}) < \infty\}$ subject to the linear moment calibration constraint $\mathbb{E}_\upsilon[\mathcal{K}\mathcal{W}] = N\boldsymbol{\mu}_X(t)$ is attained at $d\nu^*/d\upsilon = Z_\upsilon^{-1}(\boldsymbol{\lambda}^*)\exp\{\langle \boldsymbol{\lambda}^*, \mathcal{K}\mathcal{W}\rangle\}$, where

$$Z_\upsilon(\boldsymbol{\lambda}) = \exp\left\{\mathbb{E}_\upsilon[\langle \boldsymbol{\lambda}, \mathcal{K}\mathcal{W}\rangle] + \frac{1}{2}\mathbb{V}_\upsilon[\langle \boldsymbol{\lambda}, \mathcal{K}\mathcal{W}\rangle]\right\}$$

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

$$\tag{6}$$

owing to that $\mathbb{E}_\upsilon[d\mathcal{W}_i(s)] = 0$, and $\mathbb{V}_\upsilon[d\mathcal{W}_i(s)] = ds$, $i \in a$.

Now we proceed with the problem of finding $\boldsymbol{\lambda}^*(dt) \in \mathcal{M}_b (C [0,1]^q)$, where $\mathcal{M}_b$ is the class of bounded continuous measures, such that minimizes

$$
\begin{aligned}
H_v(\boldsymbol{\lambda}) &= \frac{1}{2} \int_0^1 \left( \sum_{i \in a} \int_0^1 K(s,t) \boldsymbol{\lambda}^\top(dt) \boldsymbol{X}_i(t) \right) \left( \sum_{i' \in a} \int_0^1 K(s,t') \boldsymbol{\lambda}^\top(dt') \boldsymbol{X}_{i'}(t') \right) ds \\
&\quad + \int_0^1 \boldsymbol{\lambda}^\top(dt) \left( \sum_{i \in a} d_i \boldsymbol{X}_i(t) - N \boldsymbol{\mu}_X(t) \right) \\
&= \frac{1}{2} \sum_{i \in a} \sum_{i' \in a} \int_0^1 \int_0^1 \int_0^1 K(s,t) K(s,t') \boldsymbol{\lambda}^\top(dt) \boldsymbol{X}_i(t) \boldsymbol{X}_{i'}^\top(t') \boldsymbol{\lambda}(dt') ds \\
&\quad + \int_0^1 \boldsymbol{\lambda}^\top(dt) \left( \sum_{i \in a} d_i \boldsymbol{X}_i(t) - N \boldsymbol{\mu}_X(t) \right).
\end{aligned}
\tag{7}
$$

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

$$
\sum_{i \in a} \sum_{i' \in a} \int_0^1 \int_0^1 K(s,t) K(s,t') \boldsymbol{X}_i(t) \boldsymbol{X}_{i'}^\top(t') \boldsymbol{\lambda}^*(dt') ds + \sum_{i \in a} d_i \boldsymbol{X}_i(t) = N \boldsymbol{\mu}_X(t), \tag{8}
$$

which can be rewritten as

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

obtaining, by the moment calibration constraint (2), the Lemma's result. ∎

*Proof of Lemma 2.* For each $i \in a$, define a random variable $m_i ((a,b])$ for $(a,b] \subset [0,1]$,

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

By the Lévy-Khintchine formula for Lévy processes, the moment generating function of the $n$-dimensional compound Poisson process $\boldsymbol{\mathcal{W}}(s)$ is given by

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

where $\boldsymbol{\xi}_k = (\xi_{1k}, \dots, \xi_{nk})^\top$. This formula can be generalized for a continuous function $g(s)$ from $[0,1]$ to $\mathbb{R}$ and defining $\langle g(s), \mathcal{W}_i \rangle = \int_0^1 g(s) d\mathcal{W}_i(s)$ for each $i \in a$, which is approximated by $\sum_{j=1}^J g(s_{j-1}) m_i ((s_{j-1}, s_j])$, with $s_j = j/J$, $j = 1, \dots, J$. Thus, by

the independence of $m_i\left(\left(a,b\right]\right)$, we have that

$$
\begin{aligned}
\mathbb{E}_v\left[\exp\left\{\langle g(s), d\mathcal{W}_i\rangle\right\}\right] &= \lim_{J\to\infty}\prod_{j=1}^{J}\mathbb{E}_v\left[\exp\left\{g\left(s_{j-1}\right)m_i\left(\left(s_{j-1},s_j\right]\right)\right\}\right] \\
&= \lim_{J\to\infty}\prod_{j=1}^{J}\exp\left\{\mathbb{E}_v\left[\exp\left\{g\left(s_{j-1}\right)\xi_i\right\}\right]\right\} \\
&= \lim_{J\to\infty}\prod_{j=1}^{J}\exp\left\{\frac{\gamma}{J}\int_{\mathbb{R}}\left(\exp\left\{g\left(s_{j-1}\right)\xi_i\right\}-1\right)u\left(d\xi_i\right)\right\} \\
&= \exp\left\{\gamma\int_0^1 ds\int_{\mathbb{R}}\left(\exp\left\{g\left(s\right)\xi_i\right\}-1\right)u\left(d\xi_i\right)\right\}, \qquad i\in a.
\end{aligned}
\tag{9}
$$

Now, by the Theorem 2, the maximum of the entropy $S$ over the class $\mathcal{V}$ subject to $\mathbb{E}_v\left[\mathcal{KW}\right]=N\boldsymbol{\mu}_X(t)$ is achieved at $d\nu^*/d\upsilon = Z_v^{-1}(\boldsymbol{\lambda}^*)\exp\left\{\langle\boldsymbol{\lambda}^*,\mathcal{KW}\rangle\right\}$ with

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

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

Therefore,

$$
\begin{aligned}
Z_v(\boldsymbol{\lambda}) &= \exp\left\{\gamma\int_0^1 ds\int_{\mathbb{R}}\left(\exp\left\{g\left(s\right)\xi_i\right\}-1\right)u\left(d\xi_i\right)\right\}\exp\left\{\left\langle\boldsymbol{\lambda},\sum_{i\in a}d_i\boldsymbol{X}_i(t)\right\rangle\right\} \\
&= \exp\left\{\gamma\int_0^1 ds\int_{\mathbb{R}}\left(\exp\left\{g\left(s\right)\xi_i\right\}-1\right)u\left(d\xi_i\right)+\left\langle\boldsymbol{\lambda},\sum_{i\in a}d_i\boldsymbol{X}_i(t)\right\rangle\right\}
\end{aligned}
\tag{10}
$$

Finally, as in the proof of Lemma 1, the problem is concentrated to find $\boldsymbol{\lambda}^*(t)$ such that minimizes

$$
\begin{aligned}
H_v(\boldsymbol{\lambda}) &= \gamma\int_0^1 ds\int_{\mathbb{R}}\left(\exp\left\{\int_0^1\boldsymbol{\lambda}^\top(dt)\sum_{i\in a}K(s,t)\xi_i\boldsymbol{X}_i(t)\right\}-1\right)u\left(d\xi_i\right) \\
&\quad+\int_0^1\boldsymbol{\lambda}^\top(dt)\left(\sum_{i\in a}d_i\boldsymbol{X}_i(t)-N\boldsymbol{\mu}_X(t)\right).
\end{aligned}
\tag{11}
$$

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

$$
\sum_{i\in a}\left[\int_0^1 K(s,t)\left(\int_{\mathbb{R}}\xi_i\exp\left\{\sum_{i\in a}\int_0^1 K(s,t)\xi_i\boldsymbol{X}_i^\top(t)\boldsymbol{\lambda}^*(dt)\right\}u\left(d\xi_i\right)\right)ds+d_i\right] \\
\times\boldsymbol{X}_i(t)=N\boldsymbol{\mu}_X(t),
\tag{12}
$$

obtaining, by the moment calibration constraint (2), the Lemma's result. $\qquad\blacksquare$

# References

[1] J. Breidt, F and J. D. Opsomer. Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28:1026–1053, 2000.

[2] H. Cardot and E. Josserand. Horvitz-Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika*, 98:107–118, 2011.

[3] H. Cardot, D. Degras, and E. Josserand. Confidence bands for Horvitz-Thompson estimators using sampled noisy functional data. 2011. ArXiv:1105.2135v1. Available at http://arxiv.org/abs/1105.2135v1.

[4] I. Csiszár. *I*-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3:146–158, 1975.

[5] I. Csiszár. Sanov property, generalized *I*-projection and a conditional limit theorem. *The Annals of Probability*, 12:768–793, 1984.

[6] D. Dacunha-Castelle and F. Gamboa. Maximum d'entropie et problème des moments. *Annales de l'Institut Henri Poincaré. Series B, Probabilités et Statistiques*, 26:567–596, 1990.

[7] J. C. Deville and C. E. Särndal. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382, 1992.

[8] W. A. Fuller. *Sampling Statistics*. Wiley Series in Survey Methodology. Wiley, 2009.

[9] F. Gamboa. *Méthode du maximum d'entropie sur la moyenne et applications*. PhD thesis, Université de Paris-Sud, Orsay, 1989.

[10] F. Gamboa and E. Gassiat. Bayesian methods and maximum entropy for ill-posed inverse problems. *The Annals of Statistics*, 25:328–350, 1997.

[11] F. Gamboa, J.-M. Loubes, and P. Rochet. Maximum entropy estimation for survey sampling. *Journal of Statistical Planning and Inference*, 141:305–317, 2011.

[12] A. Golan and H. Gzyl. An entropic estimator for linear inverse problems. *Entropy*, 14:892–923, 2012.

[13] H. Gzyl. Maxentropic reconstruction of some probability distributions. *Studies in Applied Mathematics*, 105:235–243, 2000.

[14] 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.

[15] T. Harms and P. Duchesne. On calibration estimation for quantiles. *Survey Methodology*, 32:37–52, 2006.

[16] L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer, 2012.

[17] 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.

[18] M. Kovaĉević. Calibration estimation of cumulative distribution and quantile functions from survey data. *Proceedings of the Survey Methods Section, Statistical Society of Canada*, 47:139–144, 1997.

[19] P. Maréchal. On the principle of maximum entropy as a methodology for solving linear inverse problems. In B. Grigelionis, V. P. J. Kubilius, H. Pragarauskas, R. Ruszkis, and V. Statulevicius, editors, *Probability Theory and Mathematical Statistics*, Proceedings of the Seventh Vilnius Conference, pages 481–492, Vilnius, Lithuania, 1999. VPS/TEV.

[20] A. Mohammad-Djafari. A comparison of two approaches: Maximum entropy on the mean (MEM) and Bayesian estimation (BAYES) for inverse problems. In M. Sears, V. Nedeljkovic, N. E. Pendock, and S. Sibisi, editors, *Maximum Entropy and Bayesian Methods*, pages 77–91, Berg-en-Dal, South Africa, 1996. Kluwer Academic Publishers.

[21] G. E. Montanari and M. G. Ranalli. Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association*, 100:1429–1442, 2005.

[22] J. Navaza. On the maximum entropy estimate of electron density function. *Acta Crystallographica*, A41:232–244, 1985.

[23] J. Navaza. The use of non-local constraints in maximum-entropy electron density reconstruction. *Acta Crystallographica*, A42:212–223, 1986.

[24] J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies.* Springer-Verlag, 2002.

[25] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer Series in Statistics. Springer-Verlag, 2nd edition, 2005.

[26] J. N. K. Rao, J. G. Kovar, and H. J. Mantel. On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77:365–375, 1990.

[27] M. Rueda, S. Martínez, H. Martínez, and A. Arcos. Estimation of the distribution function with calibration methods. *Journal of Statistical Planning and Inference*, 137:435–448, 2007.

[28] C. E. Särndal. The calibration approach in survey theory and practice. *Survey Methodology*, 33:99–119, 2007.

[29] C. E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling.* Springer Series in Statistics. Springer, 1992.

[30] S. Singh. Generalized calibration approach for estimating variance in survey sampling. *Annals of the Institute of Statistical Mathematics*, 53:404–417, 2001.

[31] A. Théberge. Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94:635–644, 1999.

[32] R. Varadhan. R-package BB: Solving and optimizing large-scale nonlinear systems. 2012. Available at http://cran.r-project.org.

[33] R. Varadhan and P. Gilbert. BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *J. Stat. Softw.*, 32, 2009. Available at http://www.jstatsoft.org/v32/i04/.

[34] C. Wu. Optimal calibration estimators in survey sampling. *Biometrika*, 90:937–951, 2003.

[35] C. Wu and R. R. Sitter. A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*,

96:185–193, 2001.