

Francesca Bruno, Daniela Cocchi, Alessandro
Vaghegini

Finite population properties of predictors
based on spatial patterns

Quaderni di Dipartimento

Serie Ricerche 2011, n. 3
ISSN 1973-9346



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Abstract

When statistical inference is used for spatial prediction, the model-based framework known as kriging is commonly used. The predictor for an unsampled element of a population is a weighted combination of sampled values, in which weights are obtained by estimating the spatial covariance function. This solution can be affected by model misspecification and can be influenced by sampling design properties. In classical design-based finite population inference, these problems can be overcome; nevertheless, spatial solutions are still seldom used for this purpose. Through the efficient use of spatial information, a conceptual framework for design-based estimation has been developed in this study. We propose a standardized weighted predictor for unsampled spatial data, using the population information regarding spatial locations directly in the weighting system. Our procedure does not require model estimation of the spatial pattern because the spatial relationship is captured exclusively based on the Euclidean distances between locations (which are fixed and do not require assessment after sample selection). The individual predictor is a design-based ratio estimator, and we illustrate its properties for simple random sampling.

keywords: spatial sampling; ratio estimator; design-based inference; model-based inference; spatial information in finite population inference

1 Introduction

In spatial statistics, it is important to be able to predict the value of the variable under study in an unsampled location on the basis of sample data. The variable is assumed to be homogeneous in the spatial domain: points nearer to each other are more similar in value than points at greater distances. Tobler's law (Tobler, 1970), well known in geography, states this fact.

When no auxiliary information is available, each individual in the population is equipped with its own spatial coordinates. Such additional information usually accompanies the identification of population elements. Spatial statistics offers several means of exploiting information about spatial coordinates, all based on functions of spatial lags: i.e. distances between the points that constitute the population (Cressie, 1993; Gaetan and Guyon, 2010).

The prediction of unsampled values of a variable for which Tobler's law holds can be assisted by an appropriate system of weights. These weights are usually generated by using a function whose parameters must be estimated, as usually happens when kriging is performed (Diggle and Ribeiro Jr., 2000), or by using a deterministic function (Shepard, 1968; Webster and Oliver, 2007). When the proposed spatial lag function depends only on the spatial coordinates, it can be computed for each element of the population before

sampling and then can be used to support inferences. According to Tobler's law, several deterministic distance functions can be constructed as suitable pre-sampling systems for weighting data collected in a region (i.e. Barabesi (2010)). Heavier weights are assigned to sites that are nearer to the unsampled one and lighter weights to more distant sites.

Based on the above, it makes sense to explore the statistical potential of the finite population framework. A main goal in finite population inference is the estimation of global functions for the population values of the variable under study, such as means, totals or variance. The probabilistic tools involved may include a sampling design or a superpopulation model (Särndal et al., 1992). Each proposed estimator of these global functions can be expressed in predictive form, whatever the source of randomness (Bolfarine and Zacks, 1992). If no auxiliary variable is available, the same prediction is calculated for each unsampled element. One example is the Horvitz-Thompson estimator for population total in design-based inference, which can be expressed in predictive form as the sum of the sampled values plus a global estimator of the total for the unsampled portion calculated as a function of the whole sample. If, in contrast, auxiliary variables are available for population members, each individual unsampled value can be predicted based on either design-based randomness or a superpopulation model. The predicted population total is the sum of the sampled values plus a linear combination (i.e. the sum) of the individual, and in principle different, predicted values.

In finite population inference, labels associated with each individual play an important role. In many cases, labels may be considered non-informative; this means that labels do not provide any information about individuals. However, in what follows, we propose a way to exploit spatial information on population elements using the design-based approach in a way linked to label specificity.

Our proposal is different from the commonly used model-based geostatistical approach for spatial prediction (Cressie (1993), Schabenberger and Gotway (2005)), which requires the specification of a superpopulation model. This other approach bases the construction of weights on the sample estimate derived from the covariogram. Based on this approach, it is necessary to use the sampled values of the variable under study, taking into account its spatial structure. Instead of exploiting a model for the covariogram, we propose the intensive use of geographical information known for the whole population: we wish to explore the potential use of spatial information that is available before sampling and typically employed in non-stochastic spatial prediction. This proposal is innovative; non-stochastic spatial predictors are not usually equipped with measures of uncertainty. In Brus and de Gruijter (1997), Gregoire (1998) and Stevens (2006), the model-based and design-based approaches are compared.

Spatial prediction is based on the relationship between a specific unsam-

pled element and all the others: any element in an N -dimensional population can indeed be seen as potentially unsampled and can therefore to be predicted. The expression “all the others” refers to the whole population of $N - 1$ units excluding that specific element. Functions of spatial lags, like distance functions, can be computed for the whole population before sampling; they are different for the different non-sampled elements considered. That this information is available before sampling is the main reason why we are proposing the use of a design-based predictor. The predictor can be also compared with the k -nearest neighbors technique (Baffetta et al., 2009).

A predictor of the value of the variable under study in an unsampled location assigns appropriate weights to sampled individual values. Its statistical properties depend on the source of randomness chosen. On that basis, we propose a design-based predictor and investigate its properties. This seemingly simple estimator hides the complexity of ratio-type estimators in design-based finite population inference: bias and the need to calculate approximated properties.

The components of the weighting system are population constants that do not influence sampling and play the role of pure pre-sampling weights. Inference for an individual unobserved value will depend on its unique specific geographical relationship to the sample. Simple random sampling of sites is therefore appropriate. The construction of a system of distances between pairs of locations makes it possible to predict each potentially unsampled value using its own distances from the sampled sites.

Design-based inference for summary quantities for finite populations (e.g. totals) is centered on the inclusion probabilities (Herzel (1986), Fattorini (2006)) of the elements of the population in the sample. The equivalent concept used to predict unsampled values in a spatially distributed population is the association probability for every potentially unsampled element and every potentially sampled element of the population. As far as we know, the concept of association probability is introduced here for the first time.

This paper is organized as follows. The next section introduces linear spatial interpolation in the finite population framework. The following shows how a linear spatial interpolator can be converted into a random outcome from a population, in this case a ratio estimator in finite population design-based inference. Section 4 computes the suitable probabilities for such quantities. In that section, the concept of association probability is developed in detail. Section 5 derives the moments of the random quantities that are needed to calculate the properties of the linear predictor in the simple random sampling framework. Finally, Section 6 proposes the approximated properties of the ratio predictor introduced in Section 3. A series of appendices completes the paper.

2 Linear spatial interpolation in the finite population framework

Let us consider a spatial domain \mathcal{D} and a set of n locations, $\mathbf{u}_1, \dots, \mathbf{u}_n$, within which the variable under study z is known. The interpolation problem involves evaluating $z(\mathbf{u}_0)$ at an arbitrary location \mathbf{u}_0 not belonging to the set.

Let us define a n -dimensional weighting vector \mathbf{w}_0^* whose elements are a function of the Euclidean distance between \mathbf{u}_0 and each known location. This weighting system has to be constructed in a way such that the influence of values at $\mathbf{u}_1, \dots, \mathbf{u}_n$ decreases with the amount of distance from the arbitrary \mathbf{u}_0 , according to Tobler's law. Interpolation for the variable under study for an arbitrary point can thus be carried out using standardized weights (i.e. $\mathbf{1}'_n \mathbf{w}_0^* = 1$), to obtain the following member of the linear interpolator family:

$$\widehat{z}(\mathbf{u}_0) = \mathbf{z}' \mathbf{w}_0 = \frac{\mathbf{z}' \mathbf{w}_0^*}{\mathbf{1}'_n \mathbf{w}_0^*}, \quad (1)$$

where

$$\mathbf{w}_0 = \frac{\mathbf{w}_0^*}{\mathbf{1}'_n \mathbf{w}_0^*} \quad (2)$$

and $\mathbf{z} = [z(\mathbf{u}_1), \dots, z(\mathbf{u}_n)]'$ is the vector of the variable values at the known locations.

2.1 Sample-based linear spatial interpolation

In the language of finite populations, a variable of interest ζ is considered fixed but unknown in a population of N elements labeled as $\lambda = 1, \dots, N$. When a spatial domain is sampled, each element, here a point or a location, is denoted by its coordinates $\mathbf{u}_\lambda = (x_\lambda, y_\lambda)$.

Let us consider the value of the variable at each of the locations $\zeta(\mathbf{u}_\lambda)$ as a function of the spatial coordinates. When a sample s of n locations is drawn, the values $z(\mathbf{u}_i)$ ($i = 1, \dots, n$, i.e. $i \in s$), of the variable are observed.

The function which identifies the values of the variable under study is non-random, meaning that population values $\zeta(\mathbf{u}_\lambda)$ ($\lambda = 1, \dots, N$) and sampled values $z(\mathbf{u}_i)$ ($i = 1, \dots, n$) coincide. In the finite population framework, the identity

$$\zeta(\mathbf{u}_\lambda) = \zeta(\mathbf{u}_{l_i}) = z(\mathbf{u}_i),$$

holds for each sampled unit, with $\lambda \in s$, $l_i = 1, \dots, N$ and $i = 1, \dots, n$.

The set of known locations over the domain \mathcal{D} can be considered the realization of a probabilistic sampling design from a population of size N where the variable under study is observed at the n sampled locations $\mathbf{u}_1, \dots, \mathbf{u}_n$.

From this perspective, the linear standardized interpolator (1) is still valid even if the probabilistic drawing of the set of locations must be taken

into account. The information usually available at the population level needs to be defined.

2.2 Population geographical characteristics

The geographical relationships between pairs of elements, regarded as functions of the sampled locations up until now, are indeed a subset of characteristics that hold for the whole population before sampling. The distance functions for pairs of locations in the population can be organized in a symmetric $N \times N$ matrix Φ . Each column (or row) of this matrix contains the relationship between the λ -th location \mathbf{u}_λ ($\lambda = 1, \dots, N$) and any other $\mathbf{u}_{\lambda'}$ ($\lambda' = 1, \dots, N$). We consider the inverse of the squared Euclidean distance, which satisfies Tobler's law. Thus,

$$\Phi_{N \times N} = [\phi_{\lambda\lambda'}; \lambda, \lambda' = 1, \dots, N] = [\|\mathbf{u}_\lambda - \mathbf{u}_{\lambda'}\|^{-2}; \lambda, \lambda' = 1, \dots, N] , \quad (3)$$

where $\phi_{\lambda\lambda'} = 0$ when $\lambda = \lambda'$. Each distance in Φ makes it possible to compute a specific weight for any unsampled location. Next, to construct a predictor for any unsampled element of the population, one needs to use the column of Φ corresponding to that location.

If a specific location of the population $\mathbf{u}_{\bar{\lambda}}$ is potentially unsampled, the corresponding column in matrix Φ can be isolated as a column vector via post-multiplication by the N -dimensional vector of the canonical basis $\mathbf{e}_{\bar{\lambda}}$

$$\phi_{\bar{\lambda}} = [\phi_{\lambda\bar{\lambda}}; \lambda = 1, \dots, N] = \Phi \mathbf{e}_{\bar{\lambda}} . \quad (4)$$

For the distances between pairs, each constituted by a fixed unsampled location $\mathbf{u}_{\bar{\lambda}}$ and any sampled one, the following symmetric $N \times N$ matrix is defined for any $\bar{\lambda}$

$$\Phi_{\bar{\lambda}} = \phi_{\bar{\lambda}} \phi_{\bar{\lambda}}' = \Phi \mathbf{e}_{\bar{\lambda}} \mathbf{e}_{\bar{\lambda}}' \Phi = [\phi_{\lambda\bar{\lambda}} \phi_{\bar{\lambda}\lambda'}; \lambda, \lambda' = 1, \dots, N] , \quad (5)$$

that contains zero values in the $\bar{\lambda}$ -th column and the $\bar{\lambda}$ -th row.

Let us also arrange the values of the variable under study in a diagonal matrix

$$\mathbf{Z} = \text{diag} [\zeta(\mathbf{u}_\lambda); \lambda = 1, \dots, N] , \quad (6)$$

or in the vector

$$\boldsymbol{\zeta} = \mathbf{Z} \mathbf{1}_N . \quad (7)$$

Remark 1. Note that the $N \times N$ matrix $\mathbf{e}_{\bar{\lambda}} \mathbf{e}_{\bar{\lambda}}'$ contains all zero values except for a unit value at position $(\bar{\lambda}, \bar{\lambda})$. \square

For computational reasons, in the case of simple random sampling, the following definitions are needed.

Remark 2. The diagonal matrix

$$\mathbf{D}_{\bar{\lambda}} = \mathbf{I}_N - \mathbf{e}_{\bar{\lambda}} \mathbf{e}_{\bar{\lambda}}' , \quad (8)$$

has unit values except for a zero value at the $(\bar{\lambda}, \bar{\lambda})$ position.

From (8), the N -dimensional vector

$$\mathbf{d}_{\bar{\lambda}} = \mathbf{1}_N - \mathbf{e}_{\bar{\lambda}} = \mathbf{D}_{\bar{\lambda}} \mathbf{1}_N, \quad (9)$$

can be defined. \square

2.3 Relating geographical characteristics of sampled and unsampled locations

Conditionally on a generic unsampled element, the relationship between sampled locations and any of the unsampled locations in a population is determined using pre- and post-multiplication by the $N \times N$ symmetric diagonal matrices \mathbf{A}_s and $\mathbf{B}_{\bar{s}}$, which are the realizations of random matrices that will be defined later.

First, define matrix \mathbf{A}_r , which relates sampled and unsampled locations. It contains n non-null equal rows corresponding to the n sampled elements. These rows exhibit unit values for the unsampled elements and zero values for the sampled elements. The diagonal of \mathbf{A}_r contains zero values because a location cannot be included in and excluded from the sample at the same time. Matrix \mathbf{A}_r has $(N - n)$ non-null equal columns corresponding to each unsampled location $\mathbf{u}_{\bar{\lambda}} \notin s$. Here, the unit values correspond to the sampled locations and the zero values to the unsampled ones. In other words, the sampling information is conditioned on the event that an element has not been sampled. Matrix \mathbf{A}_r rank is 1.

Any non null column of matrix \mathbf{A}_r contains the same information: a unit value for sampled locations and a zero value for unsampled ones. Each column corresponding to unsampled elements $\mathbf{u}_{\bar{\lambda}} \notin s$ needs to be seen as a diagonal matrix $\mathbf{A}_{r(\bar{\lambda})}$. Its non-null rows are the vectors $\mathbf{e}'_{\lambda} = \mathbf{e}'_{l_i}$ of the N -dimensional canonical basis corresponding to each sampled label $\lambda \in s$, $l_i = 1, \dots, N$ and $i = 1, \dots, n$. The rank of $\mathbf{A}_{r(\bar{\lambda})}$ is n . Extracting a sample produces $(N - n)$ identical matrices $\mathbf{A}_{r(\bar{\lambda})}$ that can be denoted with the same symbol \mathbf{A}_s .

In matrix $\mathbf{B}_{\bar{s}}$, the $N - n$ non-null columns are the vectors $\mathbf{e}_{\bar{\lambda}} = \mathbf{e}_{l_j}$ of the N -dimensional canonical basis corresponding to each unsampled label $\bar{\lambda} \notin s$, $l_j = 1, \dots, N$ and $j = n + 1, \dots, N$, i.e. the matrix contains diagonal unit values at the places of the unsampled elements. Exclusion from the sample means that in the matrix ex-post describing the sample, the column that corresponds to the $\bar{\lambda}$ -th unselected element contains a unit value at the $\bar{\lambda}$ -th position. The rank of $\mathbf{B}_{\bar{s}}$ is $N - n$.

Matrices \mathbf{A}_s , and $\mathbf{B}_{\bar{s}}$ can be reordered as block matrices

$$\mathbf{A}_s^* = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix}$$

and

$$\mathbf{B}_s^* = \begin{bmatrix} \mathbf{0} & \bar{\mathbf{S}}' \end{bmatrix},$$

where $\mathbf{S}_{n \times N}$ is the block of non-zero rows of \mathbf{A}_s^* and $\bar{\mathbf{S}}_{(N-n) \times N}$ is the block of non-zero rows of \mathbf{B}_s^* . Note that blocks of zero values have dimensions such that \mathbf{A}_s^* and \mathbf{B}_s^* are $N \times N$ matrices.

Matrix $\mathbf{S}_{n \times N}$ contains n row vectors \mathbf{e}'_λ of the N -dimensional canonical basis with a unit value corresponding to each sampled location $\mathbf{u}_\lambda \in s$, whereas matrix $\bar{\mathbf{S}}_{(N-n) \times N}$ contains $N - n$ row vectors \mathbf{e}'_λ of the N -dimensional canonical basis with unit value corresponding to each unsampled location $\mathbf{u}_\lambda \notin s$. In other words, matrices $\mathbf{S}_{n \times N}$ ($\bar{\mathbf{S}}_{(N-n) \times N}$) are the selection (non-selection) matrices that isolate sampled (unsampled) elements in a population.

When matrix \mathbf{B}_s^* post-multiplies Φ , the columns of Φ corresponding to the unsampled locations are selected; columns with zero values correspond to the sampled locations $\mathbf{u}_\lambda \in s$. When matrix \mathbf{A}_s^* pre-multiplies $\Phi \mathbf{B}_s^*$, it selects the rows of $\Phi \mathbf{B}_s^*$ corresponding to the pairs of sampled and unsampled locations. When a sample is drawn, the product

$$\mathbf{A}_s^* \Phi \mathbf{B}_s^* = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \Phi \begin{bmatrix} \mathbf{0} & \bar{\mathbf{S}}' \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{S} \Phi \bar{\mathbf{S}}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (10)$$

isolates $\mathbf{S} \Phi \bar{\mathbf{S}}'$, i.e. a $n \times (N - n)$ block of matrix Φ , constituted by the distances between the n sampled locations and the $N - n$ unsampled locations. In other words, conditionally on each sample s (i.e. starting from the pre-multiplication by \mathbf{S}), the distances in Φ between sampled and unsampled elements are isolated via the post-multiplication by $\bar{\mathbf{S}}'$. Product (10) is computed because after sample selection, matrices $\mathbf{A}_{r(\bar{\lambda})}$ are all equal. Some characteristics of matrices \mathbf{S} and $\bar{\mathbf{S}}$ are summarized in Appendix A.

Example. In a population of five elements, $\lambda = 1, \dots, 5$, three elements are sampled such that $s = \{1, 3, 4\}$; the unsampled units $\bar{\lambda}$ are 2 and 5. Matrices \mathbf{A}_r , \mathbf{A}_s (for $\bar{\lambda} = 2, 5$) and \mathbf{B}_s^* are, respectively,

$$\mathbf{A}_r = [\mathbf{a}_{r1}, \dots, \mathbf{a}_{r\bar{\lambda}}, \dots, \mathbf{a}_{r5}] = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The position of the non-null equal columns corresponds to the unsampled elements 2 and 5. Any of the two non-null columns $\mathbf{a}_{r\bar{\lambda}}$ ($\bar{\lambda} = 2, 5$) of matrix

\mathbf{A}_r can be arranged in the diagonal matrix \mathbf{A}_s

$$\mathbf{A}_s = \mathbf{A}_{r2} = \mathbf{A}_{r5} = \text{diag} [\mathbf{a}_{r\bar{\lambda}}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

while matrix $\mathbf{B}_{\bar{s}}$ is

$$\mathbf{B}_{\bar{s}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

After reordering, matrices \mathbf{A}_s , and $\mathbf{B}_{\bar{s}}$ appear in blocks as

$$\mathbf{A}_s^* = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ & & \mathbf{0} & & \end{bmatrix}$$

and

$$\mathbf{B}_{\bar{s}}^* = \begin{bmatrix} & 0 & 0 \\ & 1 & 0 \\ \mathbf{0} & 0 & 0 \\ & 0 & 0 \\ & 0 & 1 \end{bmatrix},$$

and consequently

$$\mathbf{A}_s^* \Phi \mathbf{B}_{\bar{s}}^* = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} 0 & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} \\ \phi_{21} & 0 & \phi_{23} & \phi_{24} & \phi_{25} \\ \phi_{31} & \phi_{32} & 0 & \phi_{34} & \phi_{35} \\ \phi_{41} & \phi_{42} & \phi_{43} & 0 & \phi_{45} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{0} & \bar{\mathbf{S}}' \end{bmatrix} = \begin{bmatrix} & \phi_{12} & \phi_{15} \\ \mathbf{0} & \phi_{32} & \phi_{35} \\ & \phi_{42} & \phi_{45} \\ \mathbf{0} & & \mathbf{0} \end{bmatrix}.$$

2.4 Linear spatial interpolation conditional on sampling

Once a sample has been drawn, the interpolation is no longer managed by (1); instead, the product (10) is used. The linear interpolator can be explicitly rewritten with reference to sampling using blocks \mathbf{S} and $\bar{\mathbf{S}}$ of matrices \mathbf{A}_s^* and $\mathbf{B}_{\bar{s}}^*$. Interpolation is conceived for an unsampled element at a time, identified by the appropriate N -dimensional column of matrix $\bar{\mathbf{S}}'$; i.e. the vector $\mathbf{e}_{\bar{\lambda}}$ of the N -dimensional canonical basis. The $n \times (N - n)$ non-null block $\mathbf{S} \Phi \bar{\mathbf{S}}'$ of matrix (10) can be written according to its columns

$$\mathbf{S} \Phi \bar{\mathbf{S}}' = [\mathbf{S} \Phi \bar{\mathbf{s}}_1, \dots, \mathbf{S} \Phi \bar{\mathbf{s}}_{\bar{\lambda}}, \dots, \mathbf{S} \Phi \bar{\mathbf{s}}_{N-n}] . \quad (11)$$

Each column $\bar{\mathbf{s}}_\lambda$ must be used to predict the value of the variable under study $\zeta(\cdot)$ for the unsampled location \mathbf{u}_λ in the population of N locations.

When a sample has been drawn from the population, for any unsampled location $\mathbf{u}_\lambda \notin s$, each column of (11) is the individual non-standardized weight \mathbf{w}_λ^* . After it is normalized using the sampling constant $\mathbf{1}'_n \mathbf{S} \Phi \bar{\mathbf{s}}_\lambda$, the n -dimensional vector of weights (2) for an unsampled location is re-expressed as

$$\mathbf{w}_\lambda = \frac{\mathbf{S} \Phi \bar{\mathbf{s}}_\lambda}{\mathbf{1}'_n \mathbf{S} \Phi \bar{\mathbf{s}}_\lambda} .$$

Thus, under the framework highlighted in Section 2.1, the linear interpolator (1) is rewritten as

$$\hat{\zeta}(\mathbf{u}_\lambda) = \mathbf{z}' \mathbf{w}_\lambda = \frac{\mathbf{z}' \mathbf{S} \Phi \bar{\mathbf{s}}_\lambda}{\mathbf{1}'_n \mathbf{S} \Phi \bar{\mathbf{s}}_\lambda} , \quad (12)$$

where the weighting vector is different for each unsampled location. The linear predictor in (12), formally equal to (1), is now expressed as a function of any selected sample from a population.

3 Spatial prediction as random selection from a population

In Section 2, interpolator (12) is explicitly written as a function of the realized sample. In design-based finite population inference, randomization concerns population elements via their probabilities of entering in the sample. To evaluate design-based properties, sample quantities need to be converted into random population quantities so that the non-random interpolator (12) can be seen as a design-based spatial predictor.

3.1 Univariate random quantities

In Section 2.3 we defined, conditionally on sampling, the $N \times N$ matrices \mathbf{A}_r , \mathbf{A}_s and $\mathbf{B}_{\bar{s}}$. They are realizations of the $N \times N$ matrices of random indicators \mathbf{A} , \mathbf{A}_λ and \mathbf{B} constructed according to the chance that elements of whether the elements of the population are part of the sample.

For a generic location, the event corresponding to be included in the sample, $\mathbf{u}_{l_i} = \mathbf{u}_\lambda$ ($i \in s$), can be synthetically denoted as $\lambda \in s$; similarly not being sampled, $\mathbf{u}_{l_j} = \mathbf{u}_\lambda$ ($j \notin s$), can be denoted as $\bar{\lambda} \in \bar{s}$, (i.e. $\bar{\lambda} \notin s$).

The random matrix \mathbf{A} of the conditional indicators for inclusion in the

sample, given that a population element is unsampled, is

$$\mathbf{A} = \left[\mathcal{I}_{(\lambda \in s | \bar{\lambda} \notin s)}; \lambda, \bar{\lambda} = 1, \dots, N \right] = \begin{bmatrix} 0 & \mathcal{I}_{(1 \in s | 2 \notin s)} & \cdots & \mathcal{I}_{(1 \in s | N \notin s)} \\ \mathcal{I}_{(2 \in s | 1 \notin s)} & 0 & \cdots & \mathcal{I}_{(2 \in s | N \notin s)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{(N \in s | 1 \notin s)} & \mathcal{I}_{(N \in s | 2 \notin s)} & \cdots & 0 \end{bmatrix}. \quad (13)$$

Matrix \mathbf{A} is a collection of N -dimensional column vectors, each conditional on a fixed $\bar{\lambda} \notin s$. This means that if one of the N columns is related to a sampled element, the conditioning event is false, and the column is made of zero values: the number of null columns is n . These columns have a structural zero value at the position $\bar{\lambda} = \lambda$ because a location cannot be included in and excluded from the sample at the same time.

The realization of \mathbf{A} is matrix \mathbf{A}_r of Section 2.3. The non-null columns of \mathbf{A}_r are all equal because the structural zeros are not distinguishable from the zero realizations of indicator variables. In its non-null columns, each matrix \mathbf{A}_r replicates a unit value corresponding to the labels that constitute the sample. Thus, the number of potentially realized matrices \mathbf{A}_r coincides with the dimension of the universe of samples of size n from the N -dimensional population.

The marginal indicator variables of unsampled locations are collected in the diagonal matrix \mathbf{B}

$$\mathbf{B} = \text{diag} \left[\mathcal{I}_{(\bar{\lambda} \notin s)}; \bar{\lambda} = 1, \dots, N \right] = \begin{bmatrix} \mathcal{I}_{(1 \notin s)} & 0 & \cdots & 0 \\ 0 & \mathcal{I}_{(2 \notin s)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{I}_{(N \notin s)} \end{bmatrix}. \quad (14)$$

The realization of \mathbf{B} is matrix $\mathbf{B}_{\bar{s}}$ of Section 2.3.

Because the prediction problem is addressed to each location once at a time, they have to be identified separately. For this purpose, matrices \mathbf{A} and \mathbf{B} are arranged according to their column vectors

$$\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_{\bar{\lambda}}, \dots, \mathbf{a}_N], \quad (15)$$

and

$$\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_{\bar{\lambda}}, \dots, \mathbf{b}_N]. \quad (16)$$

So, for each potentially unsampled location, we define the diagonal matrix

$$\mathbf{A}_{\bar{\lambda}} = \text{diag} [\mathbf{a}_{\bar{\lambda}}] = \begin{bmatrix} \mathcal{I}_{(1 \in s | \bar{\lambda} \notin s)} & 0 & \cdots & 0 \\ 0 & \mathcal{I}_{(2 \in s | \bar{\lambda} \notin s)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{I}_{(N \in s | \bar{\lambda} \notin s)} \end{bmatrix}. \quad (17)$$

Matrices $\mathbf{A}_{\bar{\lambda}}$ are N . Their non-null realizations are matrices \mathbf{A}_s of Section 2.3.

In what follows, we consider the column vectors $\mathbf{b}_{\bar{\lambda}}$ of \mathbf{B}

$$\mathbf{b}_{\bar{\lambda}} = \begin{bmatrix} 0 \\ \vdots \\ \mathcal{I}_{(\bar{\lambda} \notin s)} \\ \vdots \\ 0 \end{bmatrix} = \mathcal{I}_{(\bar{\lambda} \notin s)} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \mathcal{I}_{(\bar{\lambda} \notin s)} \mathbf{e}_{\bar{\lambda}} . \quad (18)$$

The last equality of (18) shows that each $\mathbf{b}_{\bar{\lambda}}$ is the randomisation of the vector of the N -dimensional canonical basis $\mathbf{e}_{\bar{\lambda}}$ denoting exclusion from the sample.

3.2 The design-based predictor

The definitions of Section 3.1 make it possible to pass from the linear interpolator (12) to a function of the random variables that govern selection from population values as follows.

The relationship between the selected sample and any unsampled location is highlighted when matrix $\mathbf{A}_{\bar{\lambda}}$ and the corresponding vector $\mathbf{b}_{\bar{\lambda}}$ are considered jointly. The random selection of the distances contained in matrix (3) is managed by the N -dimensional random vector $\mathbf{A}_{\bar{\lambda}} \Phi \mathbf{b}_{\bar{\lambda}}$, that involves the product of the univariate random variables constructed in Section 3.1

$$\mathbf{A}_{\bar{\lambda}} \Phi \mathbf{b}_{\bar{\lambda}}; \quad \bar{\lambda} = 1, \dots, N . \quad (19)$$

The linear interpolator (12) has been redefined as a function of the sampled values \mathbf{z} and the proper weighting vector, function of the distances of each unsampled location from the sampled ones. These distances are contained in the n -dimensional non-random column $\mathbf{S} \Phi \bar{\mathbf{s}}_{\bar{\lambda}}$ of the non-null block of matrix (10). Note that once a sample is drawn, the non-null values of (19) correspond exactly to what is contained in the appropriate column vector of (11).

Remark. Product (10) can now be interpreted more clearly. It is the collection of column vectors, representing the realizations of the random event (19), each containing the distances between a sampled element and any unsampled one; they are different for each column. Also, matrix (11) should be seen as a collection of column vectors.

For the numerator and denominator of (12), the following equalities hold:

$$\mathbf{z}' \mathbf{S} \Phi \bar{\mathbf{s}}_{\bar{\lambda}} = \zeta' \mathbf{A}_{\bar{\lambda}} \Phi \mathbf{b}_{\bar{\lambda}} , \quad (20)$$

and

$$\mathbf{1}'_n \mathbf{S} \Phi \bar{\mathbf{s}}_{\bar{\lambda}} = \mathbf{1}'_N \mathbf{A}_{\bar{\lambda}} \Phi \mathbf{b}_{\bar{\lambda}} . \quad (21)$$

This is due to the joint role of random matrix $\mathbf{A}_{\bar{\lambda}}$ and random vector $\mathbf{b}_{\bar{\lambda}}$.

Thus we can start from interpolator (12) and write a predictor that is a function of the random variables managing the selection from population:

$$\hat{\zeta}(\mathbf{u}_{\bar{\lambda}}) = \frac{\mathbf{z}'\mathbf{S}\Phi\bar{\mathbf{s}}_{\bar{\lambda}}}{\mathbf{1}'_n\mathbf{S}\Phi\bar{\mathbf{s}}_{\bar{\lambda}}} = \frac{\boldsymbol{\zeta}'\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{b}_{\bar{\lambda}}}{\mathbf{1}'_N\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{b}_{\bar{\lambda}}}. \quad (22)$$

The expression above stresses that, in each unsampled location, the predictor is the ratio of random quantities whose randomness depends on the sampling design. The exact properties of the estimator, which is biased, cannot be derived analytically.

The necessary randomization is therefore bivariate. To compute the moments of the predictor (22), it is necessary to define the higher-order random quantities involved, their probabilities and their moments. This is done respectively in the following sub-section and in sections 4 and 5. In Section 6 the moments of predictor (22) are computed.

3.3 Bivariate random quantities for spatial design-based prediction

Starting from univariate events, the joint occurrence of an unsampled element and a selected sample will be employed for the properties of the design-based predictor (22). Based on (17) and the indicator random variable appearing in (18), for each $\bar{\lambda}$, the second-order random variable is the element of the diagonal matrix

$$\mathbf{K}_{\bar{\lambda}} = \mathcal{I}_{(\bar{\lambda} \notin s)} \mathbf{A}_{\bar{\lambda}} = \text{diag} \left[\mathcal{I}_{(\bar{\lambda} \notin s, \lambda \in s)}; \lambda = 1, \dots, N \right], \quad (23)$$

that has a zero value in the $(\bar{\lambda}, \bar{\lambda})$ position due to $\mathbf{A}_{\bar{\lambda}}$. This is because no location can be included in and excluded from the sample at the same time. Matrix (23) contains the joint random variables that for any unsampled unit, retrieve the entire sample in the population; it can also be represented in vector form:

$$\mathbf{k}_{\bar{\lambda}} = \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N = \left[\mathcal{I}_{(\bar{\lambda} \notin s, \lambda \in s)}; \lambda = 1, \dots, N \right]. \quad (24)$$

The definition (23) makes it possible to rewrite the random numerator (20) and denominator (21) using the bivariate random variables

$$\boldsymbol{\zeta}'\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{b}_{\bar{\lambda}} = \mathbf{1}'_N\mathbf{Z}'\mathbf{A}_{\bar{\lambda}}\Phi\mathcal{I}_{(\bar{\lambda} \notin s)}\mathbf{e}_{\bar{\lambda}} = \mathbf{1}'_N\mathbf{Z}'\mathcal{I}_{(\bar{\lambda} \notin s)}\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{e}_{\bar{\lambda}} = \mathbf{1}'_N\mathbf{Z}\mathbf{K}_{\bar{\lambda}}\phi_{\bar{\lambda}} = \boldsymbol{\zeta}'\mathbf{K}_{\bar{\lambda}}\phi_{\bar{\lambda}}, \quad (25)$$

and

$$\mathbf{1}'_N\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{b}_{\bar{\lambda}} = \mathbf{1}'_N\mathbf{A}_{\bar{\lambda}}\Phi\mathcal{I}_{(\bar{\lambda} \notin s)}\mathbf{e}_{\bar{\lambda}} = \mathbf{1}'_N\mathcal{I}_{(\bar{\lambda} \notin s)}\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{e}_{\bar{\lambda}} = \mathbf{1}'_N\mathbf{K}_{\bar{\lambda}}\phi_{\bar{\lambda}}. \quad (26)$$

In (25) and (26), the population vector (4) is retrieved, and (18) is split in a random and a non-random component. The linear predictor (22) can be finally written as

$$\widehat{\zeta}(\mathbf{u}_{\bar{\lambda}}) = \mathbf{z}'\mathbf{w}_{\bar{\lambda}} = \frac{\mathbf{z}'\mathbf{S}\Phi\bar{\mathbf{s}}_{\bar{\lambda}}}{\mathbf{1}'_n\mathbf{S}\Phi\bar{\mathbf{s}}_{\bar{\lambda}}} = \frac{\boldsymbol{\zeta}'\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{b}_{\bar{\lambda}}}{\mathbf{1}'_N\mathbf{A}_{\bar{\lambda}}\Phi\mathbf{b}_{\bar{\lambda}}} = \frac{\boldsymbol{\zeta}'\mathbf{K}_{\bar{\lambda}}\phi_{\bar{\lambda}}}{\mathbf{1}'_N\mathbf{K}_{\bar{\lambda}}\phi_{\bar{\lambda}}}, \quad (27)$$

where the randomness of the marginal and conditional univariate random variables flows into a bivariate random variable.

3.3.1 Squares of bivariate random quantities

To evaluate the variance of the proposed predictor, fourth-order random quantities need to be defined as special cases under simplifying hypotheses.

From (24), we define the random matrix

$$\mathbf{k}_{\bar{\lambda}}\mathbf{k}'_{\bar{\lambda}} = \begin{bmatrix} \mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)}^2 & \mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)} & \cdots & \mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, N\in s)} \\ \mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)} & \mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)}^2 & \cdots & \mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, N\in s)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{(\bar{\lambda}\notin s, N\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)} & \mathcal{I}_{(\bar{\lambda}\notin s, N\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)} & \cdots & \mathcal{I}_{(\bar{\lambda}\notin s, N\in s)}^2 \end{bmatrix} \\ = \begin{bmatrix} \mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)} & \mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)} & \cdots & \mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, N\in s)} \\ \mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)} & \mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)} & \cdots & \mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, N\in s)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{(\bar{\lambda}\notin s, N\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 1\in s)} & \mathcal{I}_{(\bar{\lambda}\notin s, N\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, 2\in s)} & \cdots & \mathcal{I}_{(\bar{\lambda}\notin s, N\in s)} \end{bmatrix} \quad (28)$$

where, in the diagonal, fourth-order randomness reduces to second-order due to the idempotence of indicator random variables,

$$\mathcal{I}_{(\bar{\lambda}\notin s, \lambda\in s)}^2 = \mathcal{I}_{(\bar{\lambda}\notin s, \lambda\in s)}.$$

Each out-of-diagonal product of two bivariate events can be rewritten as

$$\mathcal{I}_{(\bar{\lambda}\notin s, \lambda\in s)}\mathcal{I}_{(\bar{\lambda}\notin s, \lambda'\in s)} = \mathcal{I}_{(\lambda\in s|\bar{\lambda}\notin s)}\mathcal{I}_{(\bar{\lambda}\notin s)}\mathcal{I}_{(\lambda'\in s|\bar{\lambda}\notin s)}\mathcal{I}_{(\bar{\lambda}\notin s)} = \mathcal{I}_{(\bar{\lambda}\notin s)}\mathcal{I}_{(\lambda\in s|\bar{\lambda}\notin s)}\mathcal{I}_{(\lambda'\in s|\bar{\lambda}\notin s)}. \quad (29)$$

It is an idempotent third-order random indicator that assumes unit values when the three events $\lambda \in s$, $\lambda' \in s$ and $\bar{\lambda} \notin s$ occur jointly. In (29) the product $\mathcal{I}_{(\lambda\in s|\bar{\lambda}\notin s)}\mathcal{I}_{(\lambda'\in s|\bar{\lambda}\notin s)}$ has to be considered as

$$\mathcal{I}_{(\lambda, \lambda'\in s|\bar{\lambda}\notin s)} = \mathcal{I}_{(\lambda\in s|\bar{\lambda}\notin s)}\mathcal{I}_{(\lambda'\in s|\lambda\in s, \bar{\lambda}\notin s)}, \quad (30)$$

and thus, each third-order random indicator variable becomes

$$\mathcal{I}_{(\bar{\lambda}\notin s, \lambda, \lambda'\in s)} = \mathcal{I}_{(\bar{\lambda}\notin s)}\mathcal{I}_{(\lambda, \lambda'\in s|\bar{\lambda}\notin s)}.$$

4 Fundamental probabilities in finite population inference

To predict the value of the variable under study in an unsampled location for a spatially distributed population, randomness is managed not only by the individual probabilities of inclusion in the sample, but also by the association probabilities that indicate the relationship of each unsampled element to the whole sample. First-order association probabilities come from results regarding second-order probabilities in standard design-based inference. Similarly, second-order association probabilities are computed based on results for fourth-order probabilities in standard design-based inference. These probabilities are simpler to compute in the case of equal probability sampling that is the case of our study. A discussion of unequal probabilities is included in Appendix B.

A couple of elements will meet one the following conditions:

- a) both elements will be included in the sample;
- b) no element will be included in the sample;
- c) one element will be included in the sample, but the other will not.

4.1 Inclusion and exclusion probabilities in equal probability sampling

Cases a) and b) are commonly used in the theory of finite populations. Case a) involves second-order inclusion probabilities. In the case of equal probability sampling, the first-order (marginal and conditional) inclusion probabilities are

$$\pi_{\lambda} = \Pr(\mathcal{I}_{(\lambda \in s)}) = \frac{n}{N}; \quad \forall \lambda = 1, \dots, N,$$

and

$$\pi_{\lambda'|\lambda} = \Pr(\mathcal{I}_{(\lambda' \in s | \lambda \in s)}) = \frac{n-1}{N-1}; \quad \forall \lambda \neq \lambda' = 1, \dots, N.$$

Marginal first-order inclusion probabilities are arranged in a vector that is a special case of (84)

$$\boldsymbol{\pi} = \frac{n}{N} \mathbf{1}_N.$$

Second-order inclusion probabilities

$$\pi_{\lambda\lambda'} = \pi_{\lambda} \pi_{\lambda'|\lambda} = \begin{cases} \frac{n}{N} \frac{n-1}{N-1}; & \lambda \neq \lambda' = 1, \dots, N, \\ \frac{n}{N}; & \lambda = \lambda' = 1, \dots, N, \end{cases}$$

can be arranged (see Dol et al. (1996)) in a $N \times N$ matrix

$$\mathbf{\Pi}_I = \frac{n}{N} \frac{N-n}{N-1} \mathbf{I}_N + \frac{n}{N} \frac{n-1}{N-1} \mathbf{1}_N \mathbf{1}'_N, \quad (31)$$

that contains the probabilities for unit values of the bivariate indicators $\mathcal{I}_{(\lambda, \lambda' \in s)}$. This is the special case of (85) for simple random sampling.

Case b) is managed by second-order exclusion probabilities. In the case of equal probability sampling, the first-order (marginal and conditional) exclusion probabilities are

$$\pi_{\bar{\lambda}} = \Pr\left(\mathcal{I}_{(\bar{\lambda} \notin s)}\right) = \frac{N-n}{N}; \quad \forall \bar{\lambda} = 1, \dots, N, \quad (32)$$

and

$$\pi_{\bar{\lambda}'|\bar{\lambda}} = \Pr\left(\mathcal{I}_{(\bar{\lambda}' \notin s | \bar{\lambda} \notin s)}\right) = \frac{N-n-1}{N-1}; \quad \forall \bar{\lambda} \neq \bar{\lambda}' = 1, \dots, N.$$

The marginal first-order exclusion probabilities are arranged in a vector that is a special case of (88)

$$\bar{\boldsymbol{\pi}} = \frac{N-n}{N} \mathbf{1}_N. \quad (33)$$

The unit values that appear in the indicator random variable of expression (18) have probabilities (33).

The second-order exclusion probabilities

$$\pi_{\bar{\lambda}\bar{\lambda}'} = \pi_{\bar{\lambda}} \pi_{\bar{\lambda}'|\bar{\lambda}} = \begin{cases} \frac{N-n}{N} \frac{N-n-1}{N-1}; & \bar{\lambda} \neq \bar{\lambda}' = 1, \dots, N, \\ \frac{N-n}{N}; & \bar{\lambda} = \bar{\lambda}' = 1, \dots, N. \end{cases}$$

can be arranged, as in (31), in a $N \times N$ matrix

$$\mathbf{\Pi}_E = \frac{N-n}{N} \frac{n}{N-1} \mathbf{I}_N + \frac{N-n}{N} \frac{N-n-1}{N-1} \mathbf{1}_N \mathbf{1}'_N,$$

that contains the probabilities of unit values of the bivariate indicators $\mathcal{I}_{(\bar{\lambda}, \bar{\lambda}' \notin s)}$. This is a special case of (90) for simple random sampling.

4.2 First-order association probabilities in equal probability sampling

When we are interested in the relationship between a sampled location and an unsampled one, we can introduce first-order association probabilities (case c) defined as

$$\pi_{\lambda\bar{\lambda}} = \pi_{\bar{\lambda}\lambda} = \pi_{\bar{\lambda}} \pi_{\lambda|\bar{\lambda}}, \quad (34)$$

which is null when $\lambda = \bar{\lambda}$ because an element cannot jointly be included in and excluded from a sample. This type of association probability might come also from the product $\pi_\lambda \pi_{\bar{\lambda}|\lambda}$ that has no relevance in the problem we are dealing with.

Expression (34) is the probability for the unit value of the bivariate indicator variables in (23) and (24). In simple random sampling without replacement, the conditional probabilities of inclusion given an unsampled location are

$$\pi_{\lambda|\bar{\lambda}} = \Pr\left(\mathcal{I}_{(\lambda \in s | \bar{\lambda} \notin s)}\right) = \begin{cases} \frac{n}{N-1}; & \lambda = 1, \dots, N, \lambda \neq \bar{\lambda}, \\ 0; & \lambda = 1, \dots, N, \lambda = \bar{\lambda}, \end{cases} \quad (35)$$

i.e. the probabilities for unit values of the univariate conditional indicator variables that appear in (13). Thus, based on (32), the first-order association probabilities (34) are

$$\pi_{\lambda\bar{\lambda}} = \pi_{\bar{\lambda}} \pi_{\lambda|\bar{\lambda}} = \begin{cases} \frac{N-n}{N} \frac{n}{N-1}; & \lambda = 1, \dots, N, \lambda \neq \bar{\lambda}, \\ 0; & \lambda = 1, \dots, N, \lambda = \bar{\lambda}. \end{cases}$$

Defining the constant

$$c = \frac{N-n}{N} \frac{n}{N-1}, \quad (36)$$

the first-order association probabilities can be written as

$$\pi_{\lambda\bar{\lambda}} = \pi_{\bar{\lambda}} \pi_{\lambda|\bar{\lambda}} = \begin{cases} c; & \lambda = 1, \dots, N, \lambda \neq \bar{\lambda}, \\ 0; & \lambda = 1, \dots, N, \lambda = \bar{\lambda}. \end{cases} \quad (37)$$

The probabilities (34) can be arranged in column vectors

$$\boldsymbol{\pi}_{A\bar{\lambda}} = \frac{N-n}{N} \frac{n}{N-1} \mathbf{d}_{\bar{\lambda}} = c \mathbf{d}_{\bar{\lambda}}, \quad (38)$$

where (9) is used in order to have a zero value at the $\bar{\lambda}$ -th position.

The vectors (38) of association probabilities can be organized in a $N \times N$ matrix with zero values in the diagonal as follows :

$$\begin{aligned} \mathbf{\Pi}_A &= \frac{n-N}{N} \frac{n}{N-1} \mathbf{I}_N + \frac{N-n}{N} \frac{n}{N-1} \mathbf{1}_N \mathbf{1}'_N \\ &= c (\mathbf{1}_N \mathbf{1}'_N - \mathbf{I}_N) \\ &= c (\mathbf{1}_N \mathbf{1}'_N - \mathbf{e}_{\bar{\lambda}} \mathbf{e}'_{\bar{\lambda}} - \mathbf{D}_{\bar{\lambda}}). \end{aligned} \quad (39)$$

Note that in the last line of (39), matrix (8) appears. This is a special case of (92).

4.3 Second-order association probabilities in equal probability sampling

Second-order association probabilities are based on the third-order random indicators of (29). They can be defined, as is stated in (34) and using (30), as

$$\pi_{\bar{\lambda}\lambda\lambda'} = \pi_{\bar{\lambda}}\pi_{\lambda\lambda'|\bar{\lambda}}; \quad \forall \lambda, \lambda' = 1, \dots, N, \quad (40)$$

where the factorization into marginal and conditional probabilities follows the logic of the problem under study. Note that when $\lambda = \lambda'$, the probability (40) becomes the first-order association probability (34) because the joint event on the diagonal of matrix (28) is bivariate

$$\pi_{\bar{\lambda}\lambda\lambda} = \pi_{\bar{\lambda}}\pi_{\lambda|\bar{\lambda}} = \pi_{\bar{\lambda}\lambda}; \quad \forall \lambda = 1, \dots, N. \quad (41)$$

A summary of second-order association probabilities for varying probability sampling is presented in Appendix B.3.

Based on (32) and (35), the joint probabilities (40) and (41) are

$$\pi_{\bar{\lambda}\lambda\lambda'} = \begin{cases} \frac{N-n}{N} \frac{n}{N-1} \frac{n-1}{N-2}; & \lambda \neq \lambda' = 1, \dots, N, \quad \lambda \neq \lambda' \neq \bar{\lambda}, \\ \frac{N-n}{N} \frac{n}{N-1}; & \lambda = \lambda' = 1, \dots, N, \quad \lambda = \lambda' \neq \bar{\lambda}, \\ 0; & \lambda, \lambda'\bar{\lambda} = 1, \dots, N. \end{cases}$$

After the following constants have been defined, which enter in the subsequent results

$$g = \frac{n-1}{N-2}, \quad h = \frac{N-n-1}{N-2}, \quad m = g - c,$$

and using constant c , defined in (36), the joint probabilities (40) and (41) can be written as

$$\pi_{\bar{\lambda}\lambda\lambda'} = \begin{cases} cg; & \lambda \neq \lambda' = 1, \dots, N, \quad \lambda \neq \lambda' \neq \bar{\lambda}, \\ c; & \lambda = \lambda' = 1, \dots, N, \quad \lambda = \lambda' \neq \bar{\lambda}, \\ 0; & \lambda, \lambda'\bar{\lambda} = 1, \dots, N. \end{cases} \quad (42)$$

The second-order association probabilities (42) can be arranged, for each $\bar{\lambda}$, in a $N \times N$ matrix that generalizes vector (38)

$$\begin{aligned} \mathbf{\Pi}_{2A\bar{\lambda}} &= \frac{N-n}{N} \frac{n}{N-1} \left[\frac{N-n-1}{N-2} \mathbf{D}_{\bar{\lambda}} + \frac{n-1}{N-2} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right] \\ &= c \left[h \mathbf{D}_{\bar{\lambda}} + g \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right]. \end{aligned} \quad (43)$$

The last line of (43) is a function of (8) and (9). Such quantities make it possible to manage the $\lambda, \lambda' = \bar{\lambda}$ case, which corresponds to an impossible

event. As in the relation between vector (38) and matrix (39), matrix (43) is a block within the larger $N^2 \times N$ matrix for second-order association probabilities. Thus, (43) is the simple random sampling without replacement equivalent to one of the N blocks that constitute (94).

5 Moments of the random quantities in simple random sampling

The first two moments of the random quantities defined in Section 3 are derived in what follows, in the case of simple random sampling. Varying probability sampling is addressed in Appendix C.

5.1 Expectations of univariate random quantities

For each $\bar{\lambda}$, the expectations of the univariate random indicators for exclusion and inclusion conditional to exclusion are, respectively,

$$\mathbb{E} \left[\mathcal{I}_{(\bar{\lambda} \notin s)} \right] = \pi_{\bar{\lambda}}, \quad (44)$$

and

$$\mathbb{E} \left[\mathcal{I}_{(\lambda \in s | \bar{\lambda} \notin s)} \right] = \pi_{\lambda | \bar{\lambda}}; \quad \lambda = 1, \dots, N. \quad (45)$$

In simple random sampling, the expectation of random matrix (17) comes directly from (45) for the special case (35)

$$\mathbb{E} [\mathbf{A}_{\bar{\lambda}}] = \text{diag} \left[\mathbb{E} \left[\mathcal{I}_{(\lambda \in s | \bar{\lambda} \notin s)} \right]; \quad \lambda = 1, \dots, N \right] = \frac{n}{N-1} \mathbf{D}_{\bar{\lambda}},$$

where (8) is used.

Similarly, regarding the exclusion from sampling, for each $\bar{\lambda}$, the expectation of the random vector (18) from (44) for the special case (32) is

$$\mathbb{E} [\mathbf{b}_{\bar{\lambda}}] = \mathbb{E} \left[\mathcal{I}_{(\bar{\lambda} \notin s)} \right] \mathbf{e}_{\bar{\lambda}} = \frac{N-n}{N} \mathbf{e}_{\bar{\lambda}}.$$

Sampling with varying probability is presented in Appendix C.1.

5.2 Expectations of bivariate quantities

For each $\bar{\lambda}$, the expectation of the bivariate random indicator $\mathcal{I}_{(\bar{\lambda} \notin s, \lambda \in s)}$ is the first-order association probability defined in (34)

$$\mathbb{E} \left[\mathcal{I}_{(\bar{\lambda} \notin s, \lambda \in s)} \right] = \pi_{\lambda \bar{\lambda}}. \quad (46)$$

The expectation of (23) comes directly from (46) based on (37) and as before (8),

$$\mathbb{E} [\mathbf{K}_{\bar{\lambda}}] = \text{diag} \left[\mathbb{E} \left[\mathcal{I}_{(\bar{\lambda} \notin s, \lambda \in s)} \right]; \quad \lambda = 1, \dots, N \right] = c \mathbf{D}_{\bar{\lambda}}. \quad (47)$$

The expectation of (24), using (9), is

$$\mathbb{E} [\mathbf{k}_{\bar{\lambda}}] = c\mathbf{D}_{\bar{\lambda}}\mathbf{1}_N = c\mathbf{d}_{\bar{\lambda}} . \quad (48)$$

Finally, based on (47), the expectations of the numerator (25) and the denominator (26) are, respectively,

$$\mathbb{E} [\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}] = \zeta' \mathbb{E} [\mathbf{K}_{\bar{\lambda}}] \phi_{\bar{\lambda}} = c\zeta \mathbf{D}_{\bar{\lambda}} \phi_{\bar{\lambda}} = c\mathbf{1}'_N \mathbf{Z} \mathbf{D}_{\bar{\lambda}} \phi_{\bar{\lambda}} = c \sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda \bar{\lambda}} = cT_{1\bar{\lambda}} , \quad (49)$$

and

$$\mathbb{E} [\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}] = \mathbf{1}'_N \mathbb{E} [\mathbf{K}_{\bar{\lambda}}] \phi_{\bar{\lambda}} = c\mathbf{1}_N \mathbf{D}_{\bar{\lambda}} \phi_{\bar{\lambda}} = c\mathbf{1}'_N \mathbf{D}_{\bar{\lambda}} \phi_{\bar{\lambda}} = c \sum_{\lambda \neq \bar{\lambda}} \phi_{\lambda \bar{\lambda}} = cT_{2\bar{\lambda}} . \quad (50)$$

Sampling with varying probability is discussed in Appendix C.2.

5.3 Variance of the main bivariate quantity

The results here exposed for simple random sampling are summarized in Appendix C.3 for varying probability sampling.

Theorem 5.3.1

$$\mathbf{V}_{\bar{\lambda}} = c \left[h\mathbf{D}_{\bar{\lambda}} + m\mathbf{d}_{\bar{\lambda}}\mathbf{d}'_{\bar{\lambda}} \right] . \quad (51)$$

Proof

The variance of (24) is

$$\mathbf{V}_{\bar{\lambda}} = \mathbb{V} [\mathbf{k}_{\bar{\lambda}}] = \mathbb{E} [\mathbf{k}_{\bar{\lambda}}\mathbf{k}'_{\bar{\lambda}}] - \mathbb{E} [\mathbf{k}_{\bar{\lambda}}] \mathbb{E} [\mathbf{k}_{\bar{\lambda}}]' ,$$

where based on (42), the expected value of (28) is

$$\mathbb{E} [\mathbf{k}_{\bar{\lambda}}\mathbf{k}'_{\bar{\lambda}}] = \begin{cases} cg; & \lambda \neq \lambda' = 1, \dots, N , \lambda \neq \lambda' \neq \bar{\lambda} , \\ c; & \lambda = \lambda' = 1, \dots, N , \lambda = \lambda' \neq \bar{\lambda} , \\ 0; & \lambda, \lambda' \bar{\lambda} = 1, \dots, N . \end{cases}$$

Based on (43), this solution can be written as

$$\mathbb{E} [\mathbf{k}_{\bar{\lambda}}\mathbf{k}'_{\bar{\lambda}}] = c \left[h\mathbf{D}_{\bar{\lambda}} + g\mathbf{d}_{\bar{\lambda}}\mathbf{d}'_{\bar{\lambda}} \right] . \quad (52)$$

Finally, using (52) and (48), we obtain

$$\begin{aligned} \mathbf{V}_{\bar{\lambda}} &= c \left[h\mathbf{D}_{\bar{\lambda}} + g\mathbf{d}_{\bar{\lambda}}\mathbf{d}'_{\bar{\lambda}} \right] - c^2 \mathbf{d}_{\bar{\lambda}}\mathbf{d}'_{\bar{\lambda}} \\ &= c \left[h\mathbf{D}_{\bar{\lambda}} + m\mathbf{d}_{\bar{\lambda}}\mathbf{d}'_{\bar{\lambda}} \right] . \quad \square \end{aligned}$$

5.4 Expectations of squares of bivariate random quantities

The following expectations are computed via the fundamental theorem on quadratic forms (see Bolfarine and Zacks (1992)) and are necessary for computing the variance of the ratio-type predictor (27). In all derivations, some population properties described at the end of Section 2.2 are used.

Theorem 5.4.1

$$\mathbb{E} \left[(\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}) (\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}})' \right] = c [hT_{3\bar{\lambda}} + gT_{4\bar{\lambda}}] \quad (53)$$

Proof

Using the commutative property of diagonal matrices and definition (5), the square of the quantity $(\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}})$ can be expressed as

$$\begin{aligned} (\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}) (\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}})' &= \mathbf{1}'_N \mathbf{Z} \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \phi_{\bar{\lambda}}' \mathbf{K}_{\bar{\lambda}} \mathbf{Z} \mathbf{1}_N = \mathbf{1}'_N \mathbf{Z} \mathbf{K}_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{K}_{\bar{\lambda}} \mathbf{Z} \mathbf{1}_N \\ &= \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N = \mathbf{k}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{k}_{\bar{\lambda}}, \end{aligned} \quad (54)$$

where the population matrix (5) is retrieved for the first time. Based on (48), the expectation of the quadratic form (54) is

$$\mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{k}_{\bar{\lambda}} \right] = \text{tr} (\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{V}_{\bar{\lambda}}) + \mathbb{E} [\mathbf{k}_{\bar{\lambda}}]' \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbb{E} [\mathbf{k}_{\bar{\lambda}}] . \quad (55)$$

The first term of the previous equation needs expression (51) and the product

$$\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{V}_{\bar{\lambda}} = c \left[h \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{D}_{\bar{\lambda}} + m \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right] . \quad (56)$$

The traces of the two matrices appearing in (56) are computed separately as

$$\text{tr} (\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{D}_{\bar{\lambda}}) = \sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda}^2 \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda} = \sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda}^2 \phi_{\lambda \bar{\lambda}}^2 = \phi'_{\bar{\lambda}} \mathbf{Z}^2 \phi_{\bar{\lambda}} = T_{3\bar{\lambda}} ,$$

and

$$\text{tr} (\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}}) = \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \zeta_{\lambda} \zeta_{\lambda'} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda'} = \mathbf{d}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} = T_{4\bar{\lambda}} . \quad (57)$$

Finally the trace of matrix (56) is

$$\text{tr} (\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{V}_{\bar{\lambda}}) = c [hT_{3\bar{\lambda}} + mT_{4\bar{\lambda}}] .$$

Because the trace of any permutation of the product of compatible matrices remains the same (see Magnus and Neudecker (2007)), a permutation of $\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}}$ is the scalar $\mathbf{d}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}}$. Thus, trace (57) is equal to such scalar. Based on (48) the last term of (55) is

$$\mathbb{E} [\mathbf{k}_{\bar{\lambda}}]' \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbb{E} [\mathbf{k}_{\bar{\lambda}}] = c^2 \mathbf{d}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} = c^2 \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \zeta_{\lambda} \zeta_{\lambda'} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda'} = c^2 \mathbf{d}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} = c^2 T_{4\bar{\lambda}} . \quad (58)$$

Ultimately, expectation (55) becomes

$$\begin{aligned} \mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{k}_{\bar{\lambda}} \right] &= c [hT_{3\bar{\lambda}} + mT_{4\bar{\lambda}} + cT_{4\bar{\lambda}}] \\ &= c [hT_{3\bar{\lambda}} + gT_{4\bar{\lambda}}] . \quad \square \end{aligned}$$

Theorem 5.4.2

$$\mathbb{E} \left[(\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}) (\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}})' \right] = c [hT_{5\bar{\lambda}} + gT_{6\bar{\lambda}}] . \quad (59)$$

Proof

As in (54), the square of $(\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}})$ can be expressed

$$\begin{aligned} (\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}) (\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}})' &= \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \phi'_{\bar{\lambda}} \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N \\ &= \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N = \mathbf{k}'_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}} . \end{aligned} \quad (60)$$

Based on (48), the expectation of the quadratic form (60) is

$$\mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}} \right] = \text{tr} (\Phi_{\bar{\lambda}} \mathbf{V}_{\bar{\lambda}}) + \mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \right] \Phi_{\bar{\lambda}} \mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \right] . \quad (61)$$

As with (56), following (51), for the first term in equation (60), we obtain

$$\Phi_{\bar{\lambda}} \mathbf{V}_{\bar{\lambda}} = c \left[h \Phi_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}} + m \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right] . \quad (62)$$

The traces of the two matrices in (62) are computed separately as

$$\text{tr} (\Phi_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}}) = \sum_{\lambda \neq \bar{\lambda}} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda} = \sum_{\lambda \neq \bar{\lambda}} \phi_{\lambda \bar{\lambda}}^2 = \phi'_{\bar{\lambda}} \phi_{\bar{\lambda}} = T_{5\bar{\lambda}} ,$$

and

$$\text{tr} (\Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}}) = \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda'} = \mathbf{d}'_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} = T_{6\bar{\lambda}} .$$

Finally the trace of matrix (62) is obtained:

$$\text{tr} (\Phi_{\bar{\lambda}} \mathbf{V}_{\bar{\lambda}}) = c [hT_{5\bar{\lambda}} + mT_6] .$$

For the reasons indicated above regarding the equality of traces of any permutation of compatible matrices that induce (58), the last term of (61) is

$$\mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \right]' \Phi_{\bar{\lambda}} \mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \right] = c^2 \mathbf{d}'_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} = c^2 \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda'} = c^2 \mathbf{d}'_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} = c^2 T_{6\bar{\lambda}} . \quad (63)$$

Expectation (61) therefore becomes

$$\begin{aligned} \mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \Phi_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}} \right] &= c [hT_{5\bar{\lambda}} + mT_{6\bar{\lambda}} + cT_{6\bar{\lambda}}] \\ &= c [hT_{5\bar{\lambda}} + gT_{6\bar{\lambda}}] . \quad \square \end{aligned}$$

Theorem 5.4.3

$$\mathbb{E} \left[\left(\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \right) \left(\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \right)' \right] = c \left[hT_{7\bar{\lambda}} + gT_{8\bar{\lambda}} \right] \quad (64)$$

Proof

The product of (25) and (26), as (54) and (60), can be written as

$$\begin{aligned} \left(\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \right) \left(\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \right)' &= \mathbf{1}'_N \mathbf{Z} \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \phi_{\bar{\lambda}}' \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N = \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N \\ &= \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N = \mathbf{k}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}}, \end{aligned} \quad (65)$$

Because matrix $\mathbf{Z} \Phi_{\bar{\lambda}}$ is not symmetric we need to define (see Bao and Ullah (2010))

$$\mathbf{W}_{\bar{\lambda}} = \frac{\mathbf{Z} \Phi_{\bar{\lambda}} + \Phi_{\bar{\lambda}} \mathbf{Z}}{2},$$

to compute the expected value of (65)

$$\mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}} \right] = \text{tr} \left(\mathbf{W}_{\bar{\lambda}} \mathbf{V}_{\bar{\lambda}} \right) + \mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \right] \mathbf{W}_{\bar{\lambda}} \mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \right]. \quad (66)$$

As with (56), following (51), for the first term of equation (66) we have

$$\begin{aligned} \mathbf{W}_{\bar{\lambda}} \mathbf{V}_{\bar{\lambda}} &= \mathbf{W}_{\bar{\lambda}} \mathbf{V}_{\bar{\lambda}} \left[h \mathbf{D}_{\bar{\lambda}} + m \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right] \\ &= c \left[h \mathbf{W}_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}} + m \mathbf{W}_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right]. \end{aligned} \quad (67)$$

The traces of the two matrices in (67) are computed separately. Using the property of the trace of matrix permutations used to obtain (58), we have

$$\text{tr} \left(\mathbf{W}_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}} \right) = \frac{1}{2} \text{tr} \left(\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}} + \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{D}_{\bar{\lambda}} \right) = \text{tr} \left(\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}} \right) = \sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda \bar{\lambda}}^2 = T_{7\bar{\lambda}},$$

and

$$\begin{aligned} \text{tr} \left(\mathbf{W}_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right) &= \frac{1}{2} \text{tr} \left(\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} + \Phi_{\bar{\lambda}} \mathbf{Z} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right) = \text{tr} \left(\mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} \mathbf{d}'_{\bar{\lambda}} \right) \\ &= \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda'} = \mathbf{d}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} = T_{8\bar{\lambda}}. \end{aligned}$$

Finally the trace of matrix (67) is

$$\text{tr} \left(\mathbf{W}_{\bar{\lambda}} \mathbf{D}_{\bar{\lambda}} \right) = c \left[hT_{7\bar{\lambda}} + mT_{8\bar{\lambda}} \right].$$

For the reasons presented above about the equality of traces, that induce (58) and (63), the last term of (66) is

$$\mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \right] \mathbf{W}_{\bar{\lambda}} \mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \right] = c^2 \mathbf{d}'_{\bar{\lambda}} \mathbf{W}_{\bar{\lambda}} \mathbf{d}_{\bar{\lambda}} = c^2 \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda \bar{\lambda}} \phi_{\bar{\lambda} \lambda'} = c^2 T_{8\bar{\lambda}}.$$

Expectation (66) therefore becomes

$$\mathbb{E} \left[\mathbf{k}'_{\bar{\lambda}} \mathbf{Z} \Phi_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}} \right] = c \left[hT_{7\bar{\lambda}} + mT_{8\bar{\lambda}} + cT_{8\bar{\lambda}} \right] = c \left[hT_{7\bar{\lambda}} + gT_{8\bar{\lambda}} \right]. \quad \square$$

The relationships between traces are highlighted in Appendix D.

5.5 Variances of bivariate random quantities and covariances between bivariate random quantities

The variance of (25), based on (53) and (49) by means of (99), is

$$\begin{aligned} V[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}] &= E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \phi_{\bar{\lambda}}' \mathbf{K}_{\bar{\lambda}} \zeta] - E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^2 \\ &= c[hT_{3\bar{\lambda}} + gT_{4\bar{\lambda}}] - c^2 T_{1\bar{\lambda}}^2 = c[hT_{3\bar{\lambda}} + gT_{4\bar{\lambda}}] - c^2 T_{4\bar{\lambda}} \\ &= c[hT_{3\bar{\lambda}} + gT_{4\bar{\lambda}} - cT_{4\bar{\lambda}}] = c[hT_{3\bar{\lambda}} + mT_{4\bar{\lambda}}] . \end{aligned} \quad (68)$$

The variance of (26), based on (59) and (50) by means of (100), is

$$\begin{aligned} V[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}] &= E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \phi_{\bar{\lambda}}' \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N] - E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^2 \\ &= c[hT_{5\bar{\lambda}} + gT_{6\bar{\lambda}}] - c^2 T_{2\bar{\lambda}}^2 = c[hT_{5\bar{\lambda}} + gT_{6\bar{\lambda}}] - c^2 T_{6\bar{\lambda}} \\ &= c[hT_{5\bar{\lambda}} + gT_{6\bar{\lambda}} - cT_{6\bar{\lambda}}] = c[hT_{5\bar{\lambda}} + mT_{6\bar{\lambda}}] . \end{aligned} \quad (69)$$

Finally, the covariance between (25) and (26), based on (64), (49), (50), by means of (101), is

$$\begin{aligned} \text{Cov}(\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}, \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}) &= E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}} \phi_{\bar{\lambda}}' \mathbf{K}_{\bar{\lambda}} \mathbf{1}_N] - E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}] E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}] \\ &= c[hT_{7\bar{\lambda}} + gT_{8\bar{\lambda}}] - c^2 T_{1\bar{\lambda}} T_{2\bar{\lambda}} \\ &= c[hT_{7\bar{\lambda}} + gT_{8\bar{\lambda}} - cT_{8\bar{\lambda}}] = c[hT_{7\bar{\lambda}} + mT_{8\bar{\lambda}}] . \end{aligned} \quad (70)$$

6 Moments of the ratio-type predictor

The structure of predictor (22) is the ratio of random quadratic forms expressed by (27). Based on the theory of survey sampling, this predictor is known to be biased and to have approximated properties. We must assess the extent to which the expectation of a ratio can be approximated based on the ratio of the expectations in this special case and in this way evaluating bias. Similarly, we need to assess the approximation of the variance of the ratio.

6.1 Approximated expected value of the predictor

The approximated expected value of the ratio-type predictor (27) for each unsampled $\bar{\lambda}$ is

$$E[\widehat{\zeta}(\mathbf{u}_{\bar{\lambda}})] = E\left[\frac{\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}}{\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}}\right] \simeq \frac{E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]}{E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]} , \quad (71)$$

which, based on (49) and (50), is

$$E[\widehat{\zeta}(\mathbf{u}_{\bar{\lambda}})] \simeq \frac{c \mathbf{1}'_N \mathbf{Z} \mathbf{D}_{\bar{\lambda}} \phi_{\bar{\lambda}}}{c \mathbf{1}'_N \mathbf{D}_{\bar{\lambda}} \phi_{\bar{\lambda}}} = \frac{\sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda \bar{\lambda}}}{\sum_{\lambda \neq \bar{\lambda}} \phi_{\lambda \bar{\lambda}}} = \frac{T_{1\bar{\lambda}}}{T_{2\bar{\lambda}}} , \quad (72)$$

i.e. a weighted average of all population values, with weights that are functions of the distances of the unsampled value from all the others. The expected approximation does not depend on the sample size.

For any location $\bar{\lambda}$ where a prediction $\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})$ is computed via (12), several population quantities are important:

- a) the true value $\zeta(\mathbf{u}_{\bar{\lambda}})$ of the variable under study—that is, the target of prediction;
- b) the constant $T_{1\bar{\lambda}}/T_{2\bar{\lambda}}$ of (72) which is also the approximated expected value;
- c) the exact expected value of $\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})$ (i.e. $E[\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})]$), which can be computed only through the universe of samples and/or simulation.

For a given sample size, predictor $\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})$ of (12) is a biased estimator of (72) because the sample difference from the target is

$$d(\mathbf{u}_{\bar{\lambda}}) = \hat{\zeta}(\mathbf{u}_{\bar{\lambda}}) - \frac{T_{1\bar{\lambda}}}{T_{2\bar{\lambda}}}; \quad \forall \bar{\lambda} = 1, \dots, N,$$

with expectation

$$B[\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})] = E[d(\mathbf{u}_{\bar{\lambda}})] = E[\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})] - \frac{T_{1\bar{\lambda}}}{T_{2\bar{\lambda}}}; \quad \forall \bar{\lambda} = 1, \dots, N,$$

i.e. a bias that, given the properties of ratio-type estimators, decreases as sample size increases but is not directly related to the true value $\zeta(\mathbf{u}_{\bar{\lambda}})$. This is because no matter how the sample is extracted, each $\hat{\zeta}(\mathbf{u}_{\bar{\lambda}})$ is allowed to depend on any among the selected $\zeta(\mathbf{u}_{\lambda})$, but never on the $\bar{\lambda}$ -th.

So, for a given population, for any location $\mathbf{u}_{\bar{\lambda}}$, the “structural bias”

$$\delta(\mathbf{u}_{\bar{\lambda}}) = \zeta(\mathbf{u}_{\bar{\lambda}}) - \frac{T_{1\bar{\lambda}}}{T_{2\bar{\lambda}}}; \quad \forall \bar{\lambda} = 1, \dots, N, \quad (73)$$

is defined as the difference between the value to predict and the population ratio (72).

The choice of equal weights for all values simplifies (72) into

$$\frac{T_{1\bar{\lambda}}^*}{T_{2\bar{\lambda}}^*} = \frac{1}{N-1} \sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda}, \quad (74)$$

and consequently defines the following special case of structural bias

$$\delta^*(\mathbf{u}_{\bar{\lambda}}) = \zeta(\mathbf{u}_{\bar{\lambda}}) - \frac{T_{1\bar{\lambda}}^*}{T_{2\bar{\lambda}}^*}; \quad \forall \bar{\lambda} = 1, \dots, N. \quad (75)$$

The comments above show that geographical distances are useful for predicting unsampled values $\zeta(\mathbf{u}_{\bar{\lambda}})$ when the system of weights, different for each $\bar{\lambda}$, improves inference. If information on distances is available and is considered useful, the population quantity predicted through (12) is (72). When no information on locations is available, the unweighted population mean (74) which excludes the $\bar{\lambda}$ -th element, is the predicted quantity. In this special case, moreover, because $T_{1\bar{\lambda}}^*/T_{2\bar{\lambda}}^*$ does not use any spatial information, the only safeguard for acceptable inferences is that all the $\zeta(\mathbf{u}_{\bar{\lambda}})$ are similar (i.e. their variance in population is small).

Because, in the general case of (72), the distance between $\zeta(\mathbf{u}_{\bar{\lambda}})$ and $T_{1\bar{\lambda}}/T_{2\bar{\lambda}}$ depends on the distances between $\mathbf{u}_{\bar{\lambda}}$ and all the others, the question of when suitable weights make (73) smaller than (75) arises.

6.2 Approximated variance of the predictor

Theorem 6.2.1

$$V \left[\widehat{\zeta}(\mathbf{u}_{\bar{\lambda}}) \right] \simeq \frac{1}{cT_{6\bar{\lambda}}^2} \left[h(T_{3\bar{\lambda}}T_{6\bar{\lambda}} - 2T_{7\bar{\lambda}}T_{8\bar{\lambda}} + T_{4\bar{\lambda}}T_{5\bar{\lambda}}) + m(T_{4\bar{\lambda}}T_{6\bar{\lambda}} - 2T_{8\bar{\lambda}}^2 + T_{4\bar{\lambda}}T_{6\bar{\lambda}}) \right]. \quad (76)$$

Proof

Stuart and Ord (1987), suggest a three-term approximation for the variance of the ratio of random quantities

$$V \left[\frac{X}{Y} \right] \simeq \frac{V[X]}{E[Y]^2} - 2 \frac{\text{Cov}(X, Y) E[X]}{E[Y]^3} + \frac{E[X]^2 V[Y]}{E[Y]^4},$$

which, in this case, yields

$$V \left[\widehat{\zeta}(\mathbf{u}_{\bar{\lambda}}) \right] \simeq \frac{V[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]}{E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^2} - 2 \frac{\text{Cov}(\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}, \mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}) E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]}{E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^3} + \frac{E[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^2 V[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]}{E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^4}. \quad (77)$$

In fact, all denominators in (77) depend on (50).

The first term of (77) is, for (68) and (50),

$$\begin{aligned} \frac{V[\zeta' \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]}{E[\mathbf{1}'_N \mathbf{K}_{\bar{\lambda}} \phi_{\bar{\lambda}}]^2} &= \frac{c[hT_{3\bar{\lambda}} + mT_{4\bar{\lambda}}]}{c^2 T_{2\bar{\lambda}}^2} \\ &= \frac{[hT_{3\bar{\lambda}} + mT_{4\bar{\lambda}}]}{cT_{2\bar{\lambda}}^2}. \end{aligned} \quad (78)$$

Analogously, based on (70), (49) and (50), the second term is

$$\begin{aligned} \frac{\text{Cov}(\boldsymbol{\zeta}'\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}, \mathbf{1}'_N\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}) \text{E}[\boldsymbol{\zeta}'\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}]}{\text{E}[\mathbf{1}'_N\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}]^3} &= \frac{c^2 [hT_{7\bar{\lambda}} + mT_{8\bar{\lambda}}] T_{1\bar{\lambda}}}{c^3 T_{2\bar{\lambda}}^3} \\ &= \frac{[hT_{7\bar{\lambda}} + mT_{8\bar{\lambda}}] T_{1\bar{\lambda}}}{c T_{2\bar{\lambda}}^3}, \end{aligned} \quad (79)$$

whereas the third component, based on (69), (49) and (50), is

$$\begin{aligned} \frac{\text{E}[\boldsymbol{\zeta}'\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}]^2 \text{V}[\mathbf{1}'_N\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}]}{\text{E}[\mathbf{1}'_N\mathbf{K}_{\bar{\lambda}}\boldsymbol{\phi}_{\bar{\lambda}}]^4} &= \frac{c^3 [hT_{5\bar{\lambda}} + mT_{6\bar{\lambda}}] T_{1\bar{\lambda}}^2}{c^4 T_{2\bar{\lambda}}^4} \\ &= \frac{[hT_{5\bar{\lambda}} + mT_{6\bar{\lambda}}] T_{1\bar{\lambda}}^2}{c T_{2\bar{\lambda}}^4}, \end{aligned} \quad (80)$$

where the properties of traces (99), (100) and (101) are used. \square

7 A toy example

A synthetic simulation study was performed with a small population to verify the properties of the design-based spatial predictor presented in the previous sections. The weights that can be assigned to the sampled values of the variable under study depend only on the distances between the location of the value to be predicted and the sampled locations. The distances between the locations are structural characteristics of the elements of a spatially organized population.

Their construction and the successive definition of distance functions as in matrix Φ of (3) are usually independent of the values for the characteristics associated with the population elements. Using distance functions, a system of weights can be created before sampling for any variable and for any individual value to predict. For a population of N elements, a $N \times N$ matrix Φ can be constructed that contains distance functions for all population elements with zeros in the diagonal.

The aim of the simulation study is to assess:

- a) behavior of the estimator given the spatial structure, due both to the locations at play and the values for the characteristics under study for the particular locations;
- b) the effect of sample size on the reduction in the differences between the exact and approximated properties of the design-based spatial predictor;
- c) the consequences of different structural biases for prediction for each population element.

Figure 1 shows a labelled ($\lambda = 1, \dots, N$) small population with $N = 15$ locations. The values $\zeta(\mathbf{u}_\lambda)$ are assigned to each location. The increasing diameter of the dots associated with the labels corresponds to the four classes isolated by the quartiles of the population values (which follow Tobler's law because the more similar values are closer together).

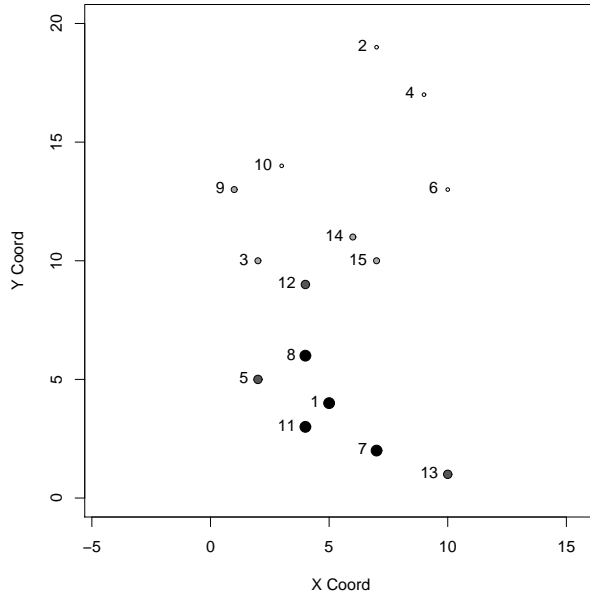


Figure 1: Site locations with the categorization delimited by quartiles of the population values

Table 1 reports the values of matrix Φ computed according to (3) for this population. Note that each of its column is different, since the location to predict varies.

Table 2 reports some important population values for each location. From now on, each location is denoted as a potentially unsampled $\bar{\lambda}$. The first and second columns contain the labels $\bar{\lambda}$ and the values to predict $\zeta(\mathbf{u}_{\bar{\lambda}})$ respectively. The third and fourth columns contain the population syntheses based on functions from matrix Φ of Table 1; namely, the population ratio (72) corresponding to the approximated expectation and the structural bias $\delta(\mathbf{u}_{\bar{\lambda}})$ of (73). The last two columns report the population syntheses that do not consider spatial information: $T_{1\bar{\lambda}}^*/T_{2\bar{\lambda}}^*$ of (74) and $\delta^*(\mathbf{u}_{\bar{\lambda}})$ of (75).

In Table 2, the fourth column highlights that different structural biases are associated to different locations when functions of distances are used as weights: the most favourable cases are those in which $\delta(\mathbf{u}_{\bar{\lambda}})$ is null or near to zero, as happens for labels 1 and 5, which are further studied in

Table 1: Matrix of distance functions

$$\mathbf{\Phi} = \begin{bmatrix} 0.000 & 0.066 & 0.149 & 0.074 & 0.316 & 0.097 & 0.354 & 0.447 & 0.102 & 0.098 \\ 0.066 & 0.000 & 0.097 & 0.354 & 0.067 & 0.149 & 0.059 & 0.075 & 0.118 & 0.156 \\ 0.149 & 0.097 & 0.000 & 0.101 & 0.200 & 0.117 & 0.106 & 0.224 & 0.316 & 0.243 \\ 0.074 & 0.354 & 0.101 & 0.000 & 0.072 & 0.243 & 0.066 & 0.083 & 0.112 & 0.149 \\ 0.316 & 0.067 & 0.200 & 0.072 & 0.000 & 0.088 & 0.171 & 0.447 & 0.124 & 0.110 \\ 0.097 & 0.149 & 0.117 & 0.243 & 0.088 & 0.000 & 0.088 & 0.108 & 0.111 & 0.141 \\ 0.354 & 0.059 & 0.106 & 0.066 & 0.171 & 0.088 & 0.000 & 0.200 & 0.080 & 0.079 \\ 0.447 & 0.075 & 0.224 & 0.083 & 0.447 & 0.108 & 0.200 & 0.000 & 0.131 & 0.124 \\ 0.102 & 0.118 & 0.316 & 0.112 & 0.124 & 0.111 & 0.080 & 0.131 & 0.000 & 0.447 \\ 0.098 & 0.156 & 0.243 & 0.149 & 0.110 & 0.141 & 0.079 & 0.124 & 0.447 & 0.000 \end{bmatrix}$$

what follows. The most unfortunate case is label 11: in this location the $\zeta(\mathbf{u}_{\bar{\lambda}})$ may be considered an outlier, and its prediction is problematic. If we compare columns 4 and 6, which contain the alternative form of structural bias (73) and (75), it emerges that predictor (12), which uses matrix $\mathbf{\Phi}$, is acceptable in almost all cases; because $\delta(\mathbf{u}_{\bar{\lambda}}) < \delta^*(\mathbf{u}_{\bar{\lambda}})$ except for labels 12 and 13. Because label 12 exhibits very similar structural biases for both weighting systems, our subsequent analysis focuses on the prediction of 13.

Table 2: Main population syntheses

$\bar{\lambda}$	$\zeta(\mathbf{u}_{\bar{\lambda}})$	$T_{1\bar{\lambda}}/T_{2\bar{\lambda}}$	$\delta(\mathbf{u}_{\bar{\lambda}})$	$T_{1\bar{\lambda}}^*/T_{2\bar{\lambda}}^*$	$\delta^*(\mathbf{u}_{\bar{\lambda}})$
1	5.81	5.81	0.00	3.57	2.23
2	1.07	2.95	-1.88	3.91	-2.84
3	2.42	3.66	-1.24	3.81	-1.39
4	1.20	2.92	-1.73	3.90	-2.71
5	5.14	5.07	0.07	3.62	1.52
6	1.58	3.26	-1.68	3.87	-2.29
7	5.75	4.86	0.89	3.58	2.17
8	7.27	4.48	2.78	3.47	3.80
9	1.65	3.26	-1.61	3.87	-2.22
10	1.59	3.13	-1.53	3.87	-2.28
11	11.49	4.48	7.01	3.16	8.33
12	3.19	3.81	-0.63	3.76	-0.57
13	2.79	4.76	-1.97	3.79	-1.00
14	2.30	3.34	-1.04	3.82	-1.52
15	2.57	3.43	-0.86	3.80	-1.23

In the following figures (Figures 2-4), the results of simulations to predict locations 1, 5 and 13 are reported. In all figures, the population values to be predicted, $\zeta(\mathbf{u}_{\bar{\lambda}})$, are indicated by bold dashed lines, expressions $T_{1\bar{\lambda}}/T_{2\bar{\lambda}}$

of (72) are indicated by thin dashed lines, and $T_{1\bar{\lambda}}^*/T_{2\bar{\lambda}}^*$ of (74) are indicated by dotted lines. In each figure, the structural biases (73) and (75) can easily be identified.

The left panel of each figure shows that as the sample size $n = 2, \dots, 14$ increases, the convergence of the biased predictor (12) to the approximated expectation (72) when matrix Φ of Table 1 is used for prediction. Each right panel shows the convergence of the (unweighted) sample mean to the unweighted population mean (74) for all values excluding the $\bar{\lambda}$ -th.

For increasing sample sizes, the distributions of predictor (12) in the sample space have been computed and summarized in the various panels. All panels show the reduction in variability as the sample size increases. The average of each distribution is the exact expectation of the predictor and is represented, in the left-hand panels, by the horizontal bar in each plot. In this case, the predictor is biased, with decreasing bias as the sample size increases and converges to (72). Conversely, when distances are not considered, the averages for the distributions in the sample space are represented in the right-hand panels. Because the predictor is unbiased, all horizontal bars have the same value which, moreover, coincides with the unweighted population average (74).

Based on the normal approximation for the exact distribution of the predictor in the sample space, the exact variances of the distributions of (12), for each n , are the starting point for computing the exact extremes $E[\hat{\zeta}_{\bar{\lambda}}] \pm 1.96sd(\hat{\zeta}_{\bar{\lambda}})$ of the 95% probability intervals that appear in each panel (shown by the darker boxes).

For increasing sample sizes, each panel of the figures shows the constants in (72), which are also the approximated expectations of predictor (12) and the approximated standard deviations based on (76). The approximated expectations, as previously noted, are constant regardless of the sample sizes because they do not depend on n .

When the normal approximation is also used to generate the approximated distribution of the predictor (12), the approximated variance (76) for each n is the starting point for computing the extremes of the 95% probability intervals, now centered in (72) in the corresponding panel. The extremes of the probability intervals of the approximated distributions are the bounds of the lighter boxes. Locations 1, 5 and 13 have higher approximated variances than the exact ones; for some individual predictions that are not reported in this example, this does not occur.

Figure 2 shows the results for prediction when $\bar{\lambda} = 1$. Here, the structural bias (73) is null: i.e. $\zeta(\mathbf{u}_{\bar{\lambda}}) = T_{1\bar{\lambda}}/T_{2\bar{\lambda}} = 5.8$. In this special case, the value to predict is equal to the population ratio in (72). The left-hand panel of Figure 2 shows the convergence of the biased predictor (12) to the approximated expectation (72) and therefore suitably captures the prediction of $\zeta(\mathbf{u}_{\bar{\lambda}})$. The variance shown in the right-hand panel is systematically

lower than the variance of the left-hand panel, but the unbiased estimator converges to a constant that is very different from the target because $\delta^*(\mathbf{u}_{\bar{\lambda}}) = 2.23$.

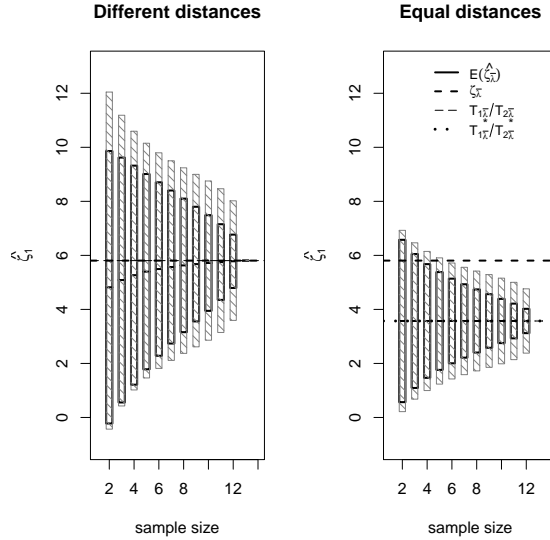


Figure 2: Prediction of $\bar{\lambda} = 1$ as n increases (left-hand panel uses distances for prediction, right-hand panel does not).

Figure 3 shows the results on prediction when $\bar{\lambda} = 5$. In this case, $\delta(\mathbf{u}_{\bar{\lambda}}) = 0.07$, i.e. it is very small even if not null. Given this consideration, the behaviour of the predictor is similar to the case illustrated in Figure 2.

Figure 4 shows the results on prediction when $\bar{\lambda} = 13$. In this case, both structural biases (73) and (75) are relevant, with the consequence that predictor (12) has a bad performance with or without weights based on distances. In the left-hand panel it converges to $T_{1\bar{\lambda}}/T_{2\bar{\lambda}} = 4.76$, while the target value $\zeta(\mathbf{u}_{\bar{\lambda}})$ is 2.79; in the right-hand panel it converges to $T_{1\bar{\lambda}}^*/T_{2\bar{\lambda}}^* = 3.79$, obtaining a slightly better result since $\delta^*(\mathbf{u}_{\bar{\lambda}}) < \delta(\mathbf{u}_{\bar{\lambda}})$.

8 Conclusions

This paper aims to demonstrate how a spatial linear interpolator can be seen under a random sampling framework. Through the use of selection matrices, it is possible to manage the relationships between sampled and unsampled locations, showing that a spatial linear interpolator can be seen as a design-based ratio-type predictor. In this case, we do not use a super-population model; we use only the population information available before sampling.

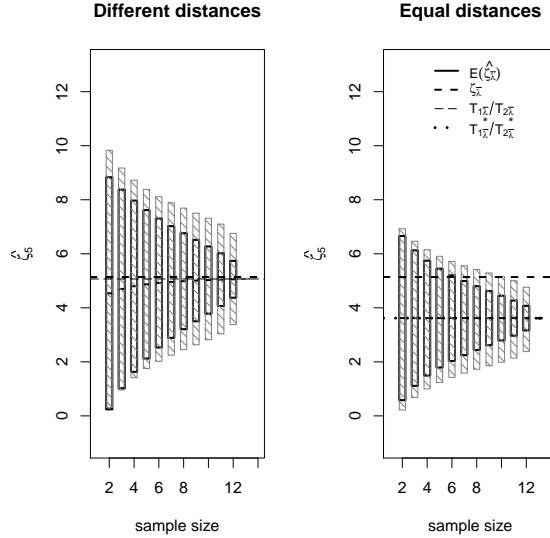


Figure 3: Prediction of $\bar{\lambda} = 5$ as n increases (left-hand panel uses distances for prediction, right-hand panel does not).

A fully design-based perspective has been explored in this paper. Individual prediction is the target of inference and is managed by the population association probabilities linking each potentially unsampled location with the others. Second- and fourth-order probabilities are therefore very important to spatial prediction: the concept of association probabilities is new and is crucial for predictions made under a design-based perspective. Indeed, these probabilities are the starting point for the evaluation of the statistical properties of predictors, which, because they must be seen as random ratios, can be analytically computed only as approximations. Our main achievement has been to reveal that it is possible to associate an uncertainty measure with interpolators: non-stochastic spatial predictors are not usually endowed by measures of uncertainty.

In this study, we proposed the simplest spatial extension of simple random sampling in finite population inference. It uses the distance functions contained in Φ as weights for the elements extracted from a population of $N - 1$ elements; they are different for each location. As a result of the selection procedure, from the design-based perspective, the predictor is a random ratio instead of a weighted average.

Our most unique findings are those that explore how a system of weights based only on geographical distances can improve inference without the use of a superpopulation model. Our main emphasis is on the difference, at the population level, between the individual value to predict and the function of population values to be estimated, expressed by (73). The proximity

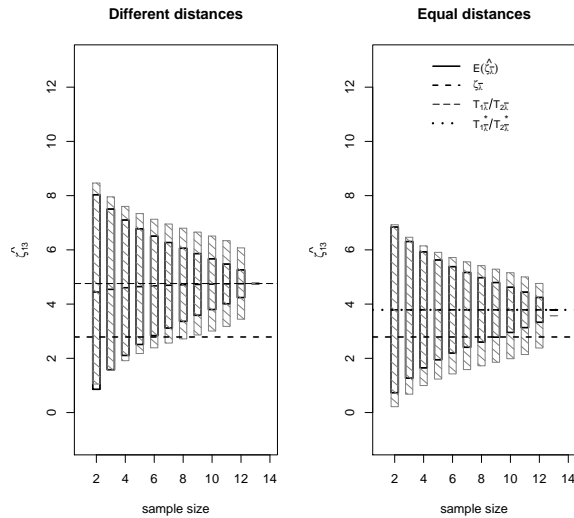


Figure 4: Prediction of $\bar{\lambda} = 13$ as n increases (left-hand panel uses distances for prediction, right-hand panel does not).

between the two population quantities may be improved prior to sampling using the distances of the location for which a value has to be predicted from all other locations in the population. In this way we exploit the spatial structure of the variable under study. In our work, we have not sought to optimize prediction; rather, we are determining to what degree geographical distance assists with inference with spatial data. In conducting spatial prediction using a design-based framework, we were assisted only by Tobler's first law of geography.

As the weighting system is the same for the entire population but different with respect to each element of that population, individual prediction may be more accurate for some locations and less so for others. This occurs irrespective of the sample size; it is due to the structural bias that characterizes the relationship between the value to be predicted and the population synthesis (72) predicted via the sample.

The results achieved constitute a first step into a dense field of foundational and analytical issues that, if explored, can help to create a new class of design-based spatial predictors that will not require the estimation of superpopulation parameters. The ease and rapidity of these computations ought to be compared via the kriging formulation, and their flexibility and performance in a variety of cases should also be considered.

Appendices

A Properties of matrices \mathbf{S} and $\bar{\mathbf{S}}$

Matrix $\mathbf{S}_{n \times N}$ ($\bar{\mathbf{S}}_{(N-n) \times N}$) is the selection (non-selection) matrix that isolates sampled (unsampled) units in a population. The symmetric permutation matrix $\mathbf{S}_{N \times N}^*$ for the sampled and unsampled elements of a population is

$$\mathbf{S}^* = \begin{bmatrix} \mathbf{S}' & \bar{\mathbf{S}}' \end{bmatrix}' = \begin{bmatrix} \mathbf{S} \\ \bar{\mathbf{S}} \end{bmatrix}$$

such that

$$\mathbf{S}^* \mathbf{S}^{*'} = \begin{bmatrix} \mathbf{S} \\ \bar{\mathbf{S}} \end{bmatrix} \begin{bmatrix} \mathbf{S}' & \bar{\mathbf{S}}' \end{bmatrix} = \begin{bmatrix} \mathbf{S}\mathbf{S}' & \mathbf{S}\bar{\mathbf{S}}' \\ \bar{\mathbf{S}}\mathbf{S}' & \bar{\mathbf{S}}\bar{\mathbf{S}}' \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & 0 \\ 0 & \mathbf{I}_{N-n} \end{bmatrix} = \mathbf{I}_N .$$

The relationship of matrices \mathbf{A}_s^* and $\mathbf{B}_{\bar{s}}^*$ with \mathbf{S}^* is

$$\mathbf{A}_s^* + (\mathbf{B}_{\bar{s}}^*)' = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{S}} \end{bmatrix} = \begin{bmatrix} \mathbf{S} \\ \bar{\mathbf{S}} \end{bmatrix} = \mathbf{S}^* .$$

Product (10)

$$\mathbf{A}_s^* \Phi \mathbf{B}_{\bar{s}}^* = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \Phi \begin{bmatrix} \mathbf{0} & \bar{\mathbf{S}}' \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{S}\Phi\bar{\mathbf{S}}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \Phi_{S\bar{S}} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} .$$

differs from the usual product in finite population prediction theory (see for instance Bolfarine and Zacks (1992)) which produces the following partition

$$\mathbf{S}^* \Phi \mathbf{S}^{*'} = \begin{bmatrix} \mathbf{S}' & \bar{\mathbf{S}}' \end{bmatrix}' \Phi \begin{bmatrix} \mathbf{S}' \\ \bar{\mathbf{S}}' \end{bmatrix} = \begin{bmatrix} \Phi_{SS} & \Phi_{S\bar{S}} \\ \Phi_{\bar{S}S} & \Phi_{\bar{S}\bar{S}} \end{bmatrix}$$

B Fundamental probabilities in varying probability sampling

As mentioned in Section 4, one of the following will be true about a pair of elements in a population:

- a) both elements will be included in the sample;
- b) neither element will be included in the sample;
- c) one element will be included in the sample and the other will not be.

B.1 Inclusion and exclusion probabilities in varying probability sampling

Cases a) and b) are commonly used in the theory of finite populations. Case a) is managed by second-order inclusion probabilities

$$\pi_{\lambda\lambda'} = \begin{cases} \pi_{\lambda}\pi_{\lambda'|\lambda}; & \lambda \neq \lambda' = 1, \dots, N, \\ \pi_{\lambda}; & \lambda = \lambda' = 1, \dots, N, \end{cases} \quad (81)$$

where the first-order (marginal and conditional) inclusion probabilities are

$$\pi_{\lambda} = \Pr(\mathcal{I}_{(\lambda \in s)}); \quad \forall \lambda = 1, \dots, N, \quad (82)$$

and

$$\pi_{\lambda'|\lambda} = \Pr(\mathcal{I}_{(\lambda' \in s | \lambda \in s)}); \quad \forall \lambda \neq \lambda' = 1, \dots, N. \quad (83)$$

The marginal first-order inclusion probabilities can be arranged in a vector

$$\pi = [\pi_{\lambda}; \quad \lambda = 1, \dots, N]. \quad (84)$$

The second-order inclusion probabilities (81) may be arranged in the $N \times N$ matrix

$$\mathbf{\Pi}_I \pi = [\pi_{\lambda\lambda'}; \lambda, \lambda' = 1, \dots, N], \quad (85)$$

which contains the probabilities of unit values of the bivariate indicators $\mathcal{I}_{(\lambda, \lambda' \in s)}$.

Case b) is managed by second-order exclusion probabilities. In varying probability sampling, first-order (marginal and conditional) exclusion probabilities are

$$\pi_{\bar{\lambda}} = \Pr(\mathcal{I}_{(\bar{\lambda} \notin s)}); \quad \forall \bar{\lambda} = 1, \dots, N, \quad (86)$$

and

$$\pi_{\bar{\lambda}'|\bar{\lambda}} = \Pr(\mathcal{I}_{(\bar{\lambda}' \notin s | \bar{\lambda} \notin s)}); \quad \forall \bar{\lambda} \neq \bar{\lambda}' = 1, \dots, N. \quad (87)$$

The marginal first-order exclusion probabilities can be arranged in a vector

$$\bar{\pi} = [\pi_{\bar{\lambda}}; \quad \bar{\lambda} = 1, \dots, N]. \quad (88)$$

They are the probabilities of the unit values of the indicator random variable in expression (18).

The second-order exclusion probabilities

$$\pi_{\bar{\lambda}\bar{\lambda}'} = \pi_{\bar{\lambda}}\pi_{\bar{\lambda}'|\bar{\lambda}} = \begin{cases} \pi_{\bar{\lambda}}\pi_{\bar{\lambda}'|\bar{\lambda}}; & \bar{\lambda} \neq \bar{\lambda}' = 1, \dots, N, \\ \pi_{\bar{\lambda}}; & \bar{\lambda} = \bar{\lambda}' = 1, \dots, N, \end{cases} \quad (89)$$

can be arranged in a $N \times N$ matrix

$$\mathbf{\Pi}_E \pi = [\pi_{\bar{\lambda}\bar{\lambda}'}; \quad \bar{\lambda}, \bar{\lambda}' = 1, \dots, N]. \quad (90)$$

which contains the probabilities of unit values of the bivariate indicators $\mathcal{I}_{(\bar{\lambda}, \bar{\lambda}' \notin s)}$.

B.2 First-order association probabilities in varying probability sampling

Case c) is managed by first-order association probabilities (34) that express the relationship between a sampled location and an unsampled one, defined as:

$$\pi_{\lambda\bar{\lambda}} = \pi_{\bar{\lambda}\lambda} = \pi_{\bar{\lambda}}\pi_{\lambda|\bar{\lambda}},$$

which is null when $\lambda = \bar{\lambda}$ because an element cannot both be part of and be excluded from a sample.

First-order association probabilities are the probability of the unit value for the bivariate indicator variables in (23) and (24). The probabilities of conditional inclusion given an unsampled location are

$$\pi_{\lambda|\bar{\lambda}} = \Pr\left(\mathcal{I}_{(\lambda \in s | \bar{\lambda} \notin s)}\right). \quad (91)$$

They are the probabilities of unit values of the indicator variables that appear in (13), whereas the probabilities of exclusion come from (86). The first-order association probabilities can be arranged in the $N \times N$ matrix

$$\mathbf{\Pi}_{A\pi} = [\pi_{\lambda\bar{\lambda}}; \lambda, \bar{\lambda} = 1, \dots, N] = [\pi_{A\bar{1}\pi} \dots \pi_{A\bar{\lambda}\pi} \dots \pi_{A\bar{N}\pi}] , \quad (92)$$

which contains zero values in the diagonal. In other words, each column vector $\pi_{A\bar{\lambda}\pi}$ contains a zero value at the $\bar{\lambda}$ -th position because of the definition of first-order association probabilities in (34).

B.3 Second-order association probabilities in varying probability sampling

The extension to varying probability sampling of the second-order association probabilities (40) under the independence hypothesis of (30) are

$$\pi_{\bar{\lambda}\lambda\lambda'} = \pi_{\bar{\lambda}}\pi_{\lambda\lambda'|\bar{\lambda}}; \quad \forall \lambda, \lambda' = 1, \dots, N .$$

When $\lambda = \lambda'$, it becomes the first-order association probability (34) because the joint event (41) in the diagonal of matrix (28) is bivariate

$$\pi_{\bar{\lambda}\lambda\lambda} = \pi_{\bar{\lambda}}\pi_{\lambda|\bar{\lambda}} = \pi_{\bar{\lambda}\lambda}; \quad \forall \lambda = 1, \dots, N .$$

In varying probability sampling, probabilities (40) become

$$\pi_{\bar{\lambda}\lambda\lambda'} = \begin{cases} \pi_{\bar{\lambda}}\pi_{\lambda\lambda'|\bar{\lambda}}; & \lambda, \lambda' = 1, \dots, N, \lambda \neq \lambda' \neq \bar{\lambda}, \\ \pi_{\bar{\lambda}}\pi_{\lambda|\bar{\lambda}}; & \lambda = \lambda' = 1, \dots, N, \lambda = \lambda' \neq \bar{\lambda}, \\ 0; & \lambda = \lambda' = \bar{\lambda} = 1, \dots, N . \end{cases} \quad (93)$$

The second-order association probabilities (93) can be arranged for each $\bar{\lambda}$ in a $N \times N$ matrix

$$\mathbf{\Pi}_{2A\bar{\lambda}\pi} = [\pi_{\bar{\lambda}\lambda\lambda'}; \lambda, \lambda' = 1, \dots, N] . \quad (94)$$

C Moments of the random quantities in varying probability sampling

C.1 Expectations of univariate random quantities

The expectation of (17) comes from (45) using (91)

$$\mathbb{E} [\mathbf{A}_{\bar{\lambda}}] = \text{diag} \left[\mathbb{E} \left[\mathcal{I}_{(\lambda \in s | \bar{\lambda} \notin s)} \right]; \lambda = 1, \dots, N \right] \quad (95)$$

Given the exclusion from sampling, for each $\bar{\lambda}$, based on (44) and using (86), the expectation of (18) is

$$\mathbb{E} [\mathbf{b}_{\bar{\lambda}}] = \mathbb{E} \left[\mathcal{I}_{(\bar{\lambda} \notin s)} \right] \mathbf{e}_{\bar{\lambda}} = \pi_{\bar{\lambda}} \mathbf{e}_{\bar{\lambda}} .$$

C.2 Expectations of bivariate random quantities

These expectations need to be derived using the association probabilities (34) for each potentially unsampled location $\bar{\lambda}$ and every potentially sampled one.

Given the expectation of (23), using the suitable column in (92), we obtain

$$\mathbb{E} [\mathbf{K}_{\bar{\lambda}}] = \text{diag} \left[\pi_{A\bar{\lambda}\pi} \right] ,$$

and

$$\mathbb{E} [\mathbf{k}_{\bar{\lambda}}] = \pi_{A\bar{\lambda}\pi} , \quad (96)$$

C.3 Variance of the main bivariate quantity

The variance of (24) is defined as

$$\mathbf{V}_{\bar{\lambda}} = \mathbb{V} [\mathbf{k}_{\bar{\lambda}}] = \mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}}' \right] - \mathbb{E} [\mathbf{k}_{\bar{\lambda}}] \mathbb{E} [\mathbf{k}_{\bar{\lambda}}]' .$$

The expected value of (28) is as follows based on (93)

$$\mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}}' \right] = \begin{cases} \pi_{\bar{\lambda}} \pi_{\lambda\lambda' | \bar{\lambda}}; & \lambda \neq \lambda' = 1, \dots, N, \lambda \neq \lambda' \neq \bar{\lambda}, \\ \pi_{\bar{\lambda}} \pi_{\lambda | \bar{\lambda}}; & \lambda = \lambda' = 1, \dots, N, \lambda = \lambda' \neq \bar{\lambda}, \\ 0; & \lambda = \lambda' = \bar{\lambda} = 1, \dots, N, \end{cases}$$

which, following (94), can be written as

$$\mathbb{E} \left[\mathbf{k}_{\bar{\lambda}} \mathbf{k}_{\bar{\lambda}}' \right] = \mathbf{\Pi}_{2A\bar{\lambda}\pi} . \quad (97)$$

Finally, using (97) and (96) we obtain

$$\mathbf{V}_{\bar{\lambda}} = \mathbf{\Pi}_{2A\bar{\lambda}\pi} - \pi_{A\bar{\lambda}\pi} (\pi_{A\bar{\lambda}\pi})' . \quad (98)$$

D Some useful relationships between traces

In sections 5.2 and 5.4, various constants were defined. Here, the relationships between them are stresses. Based on (49), we have

$$T_{1\bar{\lambda}}^2 = \left[\sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda\bar{\lambda}} \right]^2 = \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \zeta_{\lambda} \zeta_{\lambda'} \phi_{\lambda\bar{\lambda}} \phi_{\bar{\lambda}\lambda'} = T_{4\bar{\lambda}}. \quad (99)$$

Analogously, based on (50)

$$T_{2\bar{\lambda}}^2 = \left[\sum_{\lambda \neq \bar{\lambda}} \phi_{\lambda\bar{\lambda}} \right]^2 = \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \phi_{\lambda\bar{\lambda}} \phi_{\bar{\lambda}\lambda'} = T_{6\bar{\lambda}}. \quad (100)$$

Finally from (49) and (50)

$$T_{1\bar{\lambda}} T_{2\bar{\lambda}} = \left[\sum_{\lambda \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda\bar{\lambda}} \right] \left[\sum_{\lambda \neq \bar{\lambda}} \phi_{\lambda\bar{\lambda}} \right] = \sum_{\lambda \neq \bar{\lambda}} \sum_{\lambda' \neq \bar{\lambda}} \zeta_{\lambda} \phi_{\lambda\bar{\lambda}} \phi_{\bar{\lambda}\lambda'} = T_{8\bar{\lambda}}. \quad (101)$$

Acknowledgements

Work supported by the project PRIN 2008, New developments in sampling theory and practice, Project number 2008CEFF37, sector: Economics and Statistics, awarded by the Italian Government.

References

- Baffetta, F., L. Fattorini, S. Franceschi, and P. Corona (2009). Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. *Remote Sensing of Environment* 113, 463–475.
- Bao, Y. and A. Ullah (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference* 140, 1193–1205.
- Barabesi, L. (2010). mimeo. Facoltà di Economia “R. Goodwin”, Univeristà degli Studi di Siena.
- Bolfarine, H. and S. Zacks (1992). *Prediction theory for finite populations*. New York: Springer–Verlag.
- Brus, D. and J. de Gruijter (1997). Random sampling or geostatistical modelling? choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* 80, 1–44.

- Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley.
- Diggle, P. and P. Ribeiro Jr. (2000). *Model-based geostatistics*. São Paulo: Associação Brasileira de Estatística.
- Dol, W., T. Steerneman, and T. Wansbeek (1996). Matrix algebra and sampling theory: the case of the horvitz–thompson estimator. *Linear Algebra and its Applications* 237, 225–238.
- Fattorini, L. (2006). Applying the horvitz-thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities. *Biometrika* 93, 269–278.
- Gaetan, C. and X. Guyon (2010). *Spatial statistics and modeling. Modélisation et statistique spatiales*. New York: Springer.
- Gregoire, T. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research* 28, 1429–1447.
- Herzel, A. (1986). Sampling without replacement with unequal probabilities: sample designs with preassigned joint inclusion probabilities of any order. *Metron* 44, 49–68.
- Magnus, J. and H. Neudecker (2007). *Matrix differential calculus with application in statistics and econometrics*. New York: Wiley.
- Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model-assisted survey sampling*. New York: Springer-Verlag.
- Schabenberger, O. and C. Gotway (2005). *Statistical methods for spatial data analysis*. Boca Raton, Florida: Chapman & Hall.
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd Association for Computing Machinery national conference*, pp. 517–523.
- Stevens, D. (2006). Spatial properties of design-based versus model-based approaches to environmental sampling. *American Statistical Association; Section on Statistics & the Environment Newsletter* 10, 3–5.
- Stuart, A. and K. Ord (1987). *Kendall's advanced theory of statistics, distribution theory*, Volume 1. London: Arnold.
- Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography* 46, 234–240.
- Webster, R. and M. Oliver (2007). *Geostatistics for environmental scientists*. Chichester: John Wiley and Sons.