Ratio Mathematica Volume 38, 2020, pp. 261-285

Geometrical foundations of the sampling design with fixed sample size

Pierpaolo Angelini *

Abstract

We study the sampling design with fixed sample size from a geometric point of view. The first-order and second-order inclusion probabilities are chosen by the statistician. They are subjective probabilities. It is possible to study them inside of linear spaces provided with a quadratic and linear metric. We define particular random quantities whose logically possible values are all logically possible samples of a given size. In particular, we define random quantities which are complementary to the Horvitz-Thompson estimator. We identify a quadratic and linear metric with regard to two univariate random quantities representing deviations. We use the α-criterion of concordance introduced by Gini in order to identify it. We innovatively apply to probability this statistical criterion.
Keywords: tensor product; linear map; bilinear map; quadratic and linear metric; α-product; α-norm
2010 AMS subject classifications: 62D05. [1]

*Dipartimento di Scienze Statistiche, Università La Sapienza, Roma, Italia; pier.angelini@uniroma1.it

[1]Received on May 12th, 2020. Accepted on June 3rd, 2020. Published on June 30th, 2020. doi: 10.23755/rm.v38i0.511. ISSN: 1592-7415. eISSN: 2282-8214.

1 Introduction

Given a finite population having N elements, we are only interested in considering samples containing units of this population where no element of the population under consideration can be selected more than once in the same sample (Basu [1971]). We are not interested in considering ordered samples of a given size selected from a finite population (Basu [1958]). On the other hand, when we consider not ordered samples where repetitions are not allowed we have no loss of information about a given parameter of the population under consideration (Conti and Marella [2015]). All logically possible samples of a given size belong to a given set. We suppose that we are always able to number them. It is known that if the number of all logically possible samples of a given set is very large then it could be a very hard or impossible work to give to them a number (Godambe and Joshi [1965]). A sampling design is characterized by a pair of elements (Joshi [1971]). The first element of this pair represents the set of all logically possible samples selected from a finite population. The second element of this pair represents all probabilities assigned to the samples of the set of all logically possible samples of a given size. We consider a distribution of probability in this way (Hartley and Rao [1962]). Each element of the set of all logically possible samples of a given size can be viewed as a logically possible event of a finite partition of incompatible and exhaustive elementary events. It is then possible to assign a subjective probability to each logically possible event of this partition (Good [1962]). A probability subjectively assigned to each logically possible event of a finite partition of events must only be coherent. It is inadmissible when it is not coherent. A probability is subjectively assigned to each logically possible event of a finite partition of events even when it is an equal probability assigned to each of them. An equal probability assigned to each logically possible event of a finite partition of events is always a subjective judgment. We have to note a very important point: when we say that it is possible to assign a subjective and coherent probability to every logically possible event of a given set of events we mean that the choice of any value in the interval from 0 to 1 is allowed. Such an interval includes both endpoints. It would therefore be possible to assign to every logically possible event of a given set of events a probability equal to 0. This choice is absolutely coherent. We will however introduce a restriction that is concerned with this point. We have to note another very important point: we methodologically distinguish what it is logically possible from what it is subjectively probable. What it is logically possible at a given instant it is not either certainly true or certainly false. One and only one element of the elements belonging to the set containing all logically possible elements at a given instant will be true a posteriori. A subjective probability is then assigned to each element of the set containing all logically possible elements before knowing this thing.

2 Events as points in the space of random quantities

We consider a finite set of vectors denoted by $\mathcal{S}$ into the field $\mathbb{R}$ of real numbers. We enumerate them. We consequently write

$$s_1, \ldots, s_N, \tag{1}$$

where it turns out to be $s_i \in \mathcal{S}$, $i = 1, \ldots, N$. We consider a linear space over $\mathbb{R}$ of all linear combinations of elements of $\mathcal{S}$ expressed in the form

$$c_1 s_1 + \ldots + c_N s_N, \tag{2}$$

where every c_i, $i = 1, \ldots, N$, is a real number. We observe that (2) is completely determined by the real numbers $c_1, \ldots, c_N$. Each number c_i is associated with the element s_i of the set $\mathcal{S}$. It is known that an association is exactly a function. For each $s_i \in \mathcal{S}$ and $c \in \mathbb{R}$ we then consider

$$c s_i \tag{3}$$

to be the function that associates c with s_i and 0 with s_j, where we have $j \neq i$. Given $a \in \mathbb{R}$, we obtain

$$a(c s_i) = (ac) s_i. \tag{4}$$

Given $c' \in \mathbb{R}$, we obtain

$$(c + c') s_i = c s_i + c' s_i. \tag{5}$$

Thus, it is possible to consider a linear space over $\mathbb{R}$. It is the set of all functions of $\mathcal{S}$ into $\mathbb{R}$. These functions can be written in the form given by (2). The functions

$$1 s_1, \ldots, 1 s_N \tag{6}$$

are linearly independent so they represent a basis of the linear space under consideration. We have then to suppose that $c_1, \ldots, c_N$ are elements of $\mathbb{R}$ such that it is possible to obtain the zero function given by

$$c_1 s_1 + \ldots + c_N s_N = 0. \tag{7}$$

This means that we have $c_i = 0$ for every c_i, $i = 1, \ldots, N$. This thing consequently proves the linear independence under consideration. Moreover, it is always possible to write s_i instead of $1 s_i$. A sample belonging to the set of all logically possible samples of a given size is then expressed by the vector

$$\delta(s') = \begin{bmatrix} \delta(1; s') \\ \delta(2; s') \\ \vdots \\ \delta(N; s') \end{bmatrix} \tag{8}$$

having N components, where s' is a sample of the set of all logically possible samples denoted by $\mathcal{S}'$ (Godambe [1955]). We will always consider vectors viewed as ordered lists of real numbers within this context. A sample can be expressed by the real numbers of a linear combination of N-dimensional vectors by means of which another N-dimensional vector is obtained. If a sample is identified with an N-dimensional vector then its components express the real numbers of a linear combination of the elements of a basis of the linear space under consideration. This linear space is denoted by $\mathbb{R}^N$. Its basis is denoted by $\mathcal{S} = \{\mathbf{e}_j\}$, $j = 1, \ldots, N$. We always consider orthonormal bases within this context. We therefore write

$$\delta(1; s')\mathbf{e}_1 + \delta(2; s')\mathbf{e}_2 + \ldots + \delta(N; s')\mathbf{e}_N = \mathbf{y}, \tag{9}$$

where we have $\mathbf{y} \in \mathbb{R}^N$. We consider as many linear combinations of the elements of $\mathcal{S} = \{\mathbf{e}_j\}$, $j = 1, \ldots, N$, as logically possible samples there are into the set of all logically possible samples of a given size denoted by $\mathcal{S}'$. We note that the real numbers of every linear combination of the elements of $\mathcal{S} = \{\mathbf{e}_j\}$, $j = 1, \ldots, N$, represent one of the logically possible samples of $\mathcal{S}'$. We have evidently

$$\delta(i; s') = \begin{cases} 1 & \text{if } i \in s' \\ 0 & \text{if } i \notin s' \end{cases} \tag{10}$$

for every $i = 1, \ldots, N$, where the elements of the population under consideration are overall N. We consider all logically possible samples of $\mathcal{S}'$ having the same size denoted by n. Since the population has got N elements we observe that the number of n-combinations is equal to the binomial coefficient denoted by $\binom{N}{n}$. We observe that $\mathcal{S}'$ whose elements are elementary events is a subset of $\mathbb{R}^N$. We say that $\mathcal{S}'$ is embedded in $\mathbb{R}^N$.

3 Finite partitions of logically possible elementary events

Given N, all logically possible samples whose size is equal to n belong to the set denoted by $\mathcal{S}'$. We have

$$n = \sum_{i=1}^{N} \delta(i; s') \tag{11}$$

for every $s' \in \mathcal{S}'$. Every sample of the set of all logically possible samples corresponds to a vertex denoted by $\delta(s')$ of an N-dimensional unit hypercube denoted by $[0, 1]^N$. All logically possible samples of $\mathcal{S}'$ can be viewed as possible events of a finite partition of incompatible and exhaustive elementary events (de Finetti [1982b]). We are consequently able to define a univariate random quantity whose logically possible values are represented by all logically possible samples of $\mathcal{S}'$. The logically possible values of it are not real numbers but they are N-dimensional vectors of an N-dimensional linear space over $\mathbb{R}$. Every logically possible sample belonging to $\mathcal{S}'$ has a subjective probability of being selected (de Finetti [1989]. It represents the degree of belief in the selection of a logically possible sample assigned by a given individual (the statistician) at a certain instant with a given set of information. An evaluation of probability known over a set of possible and elementary events coinciding with all logically possible samples of $\mathcal{S}'$ is admissible when it is coherent. This means that it must be

$$\sum_{s' \in \mathcal{S}'} p(s') = 1. \tag{12}$$

It is essential to note a very important point: we have to introduce an unusual restriction with regard to the coherence because we exclude of choosing a subjective probability equal to 0 with respect to any possible and elementary event. This implies that any logically possible sample of $\mathcal{S}'$ has always a probability greater than zero of being selected. We have consequently

$$0 < p(s') \leq 1 \tag{13}$$

for every $s' \in \mathcal{S}'$ (Coletti et al. [2015]). Thus, conditions of coherence coincide with positivity of each probability of a random event and finite additivity of probabilities of incompatible and exhaustive events (Gilio and Sanfilippo [2014]). We will also consider bivariate random quantities whose components are two univariate random quantities (de Finetti [2011]). If the logically possible values of these univariate random quantities are the same vectors of the same N-dimensional linear space over $\mathbb{R}$ then these random quantities have the same marginal distributions of probability. They represent the same finite partition of incompatible and exhaustive elementary events. Putting them into a two-way table we observe that it is always a table having the same number of rows and columns.

4 First-order inclusion probabilities viewed as a coherent prevision of a univariate random quantity

We consider a univariate random quantity denoted by S whose logically possible values are vectors of $\mathbb{R}^N$. Given N and n, the number of the logically possible values of S coincides with the binomial coefficient expressed by

$$\binom{N}{n} = k. \tag{14}$$

The set of the logically possible values of S is then given by $I(S) = \{s'_1, \ldots, s'_k\}$, with $s'_i \in \mathcal{S}'$, $i = 1, \ldots, k$. A nonzero probability is assigned to each sample of the set of all logically possible samples. Let $p(s'_1), \ldots, p(s'_k)$ be these probabilities. It must therefore be

$$\sum_{i=1}^{k} p(s'_i) = 1, \tag{15}$$

with

$$0 < p(s'_i) \leq 1 \tag{16}$$

for every $i = 1, \ldots, k$. It is possible to obtain an N-dimensional vector after assigning a nonzero probability to each sample of $\mathcal{S}'$. We denote it with π. It represents the first-order inclusion probabilities of all units of the population under consideration. Thus, we write

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} = p(s'_1) \begin{bmatrix} \delta(1; s'_1) \\ \delta(2; s'_1) \\ \vdots \\ \delta(N; s'_1) \end{bmatrix} + \ldots + p(s'_k) \begin{bmatrix} \delta(1; s'_k) \\ \delta(2; s'_k) \\ \vdots \\ \delta(N; s'_k) \end{bmatrix}, \tag{17}$$

where we have $\pi_i > 0$ for every $i = 1, \ldots, N$. We have evidently written a convex combination of the vertices of the N-dimensional unit hypercube $[0, 1]^N$ corresponding to the samples of $\mathcal{S}'$. Each vertex is a sample having a nonzero weight representing a subjective probability. It is essential to note that π is a coherent prevision of S denoted by $\mathbf{P}(S)$. We therefore write

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} = \mathbf{P}(S) = \sum_{i=1}^{k} \delta(s'_i) p(s'_i). \tag{18}$$

We observe that the logically possible values of S are represented by vectors having N components so its coherent prevision must also be represented by a vector having N components. The logically possible values of S belong to the set denoted by $I(S)$. Each element of $I(S)$ contains first-order inclusion "a posteriori" probabilities. This implies that π must contain first-order inclusion "a priori" probabilities based on the degree of belief in the selection of all logically possible samples attributed by the statistician at a certain instant with a given set of information. An "a posteriori" probability of a unit of the population of being included in a given sample is always predetermined. If a unit of the population is contained "a posteriori" in the sample that has been selected then its probability is equal to 1. If a unit of the population does not belong "a posteriori" to the sample that has been selected then its probability is equal to 0. A convex combination coinciding with $\mathbf{P}(S)$ has conveniently been taken under consideration because the logically possible values of S are incompatible and exhaustive elementary events of a finite partition of random events. In general, if we consider an event divided into two or more than two incompatible events then we obtain that its coherent probability is the sum of two or more than two coherent probabilities. This sum is a linear combination of probabilities (de Finetti [1980]). We evidently consider a convex combination coinciding with $\mathbf{P}(S)$ within this context, where its weights or coefficients are "a priori" subjective probabilities connected with the samples of $\mathcal{S}'$ (de Finetti [1981]). This convex combination is characterized by k column vectors viewed as k matrices. Each row of every $N \times 1$ matrix is a first-order inclusion "a posteriori" probability. We therefore consider a linear combination of probabilities (de Finetti [1982a]).

5 First-order inclusion probabilities obtained by means of linear maps

We consider all logically possible samples belonging to the set $\mathcal{S}'$. Given N and n, let k be the number of all elements of $\mathcal{S}'$. We are consequently able to determine an $N \times k$ matrix in $\mathbb{R}$. We denote it by B. It is therefore possible to define a linear map expressed by

$$L_B \colon \mathbb{R}^k \to \mathbb{R}^N. \tag{19}$$

It depends on B. Moreover, it also depends on the choice of bases for $\mathbb{R}^k$ and $\mathbb{R}^N$. We choose standard bases for $\mathbb{R}^k$ and $\mathbb{R}^N$. We consider all probabilities assigned to the logically possible samples of $\mathcal{S}'$ whose size is equal to n. They can be viewed as a column vector. We denote it by Q. We have then

$$Q = \begin{bmatrix} p(s'_1) \\ p(s'_2) \\ \vdots \\ p(s'_k) \end{bmatrix}. \tag{20}$$

It therefore turns out to be

$$L_B(Q) = BQ = \pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix}. \tag{21}$$

We note that if $k = N$ then we are able to define a linear map expressed by

$$L_B \colon \mathbb{R}^N \to \mathbb{R}^N. \tag{22}$$

We observe that B is a square matrix. This linear map is an endomorphism. It is also an isomorphism. It is then an automorphism, so we write

$$B^{-1}\pi = \begin{bmatrix} p(s'_1) \\ p(s'_2) \\ \vdots \\ p(s'_k) \end{bmatrix}. \tag{23}$$

Given B, each row of Q can subjectively vary because an evaluation of probability known over a set of logically possible events must only be coherent. This means that the sum of all probabilities of the samples of $\mathcal{S}'$ must be equal to 1. We consequently observe that there are infinite ways of choosing all probabilities of the samples of $\mathcal{S}'$. They are conveniently caught by L_B. It is hence possible to obtain π as a multiplication of matrices according to a linear map depending on B and the standard bases of the linear spaces under consideration. Also, we always obtain

$$\sum_{i=1}^{N} \pi_i = n. \tag{24}$$

6 First-order and second-order inclusion probabilities obtained by means of tensor products

We consider a bivariate random quantity denoted by S_{12} whose components are two univariate random quantities denoted by ${}_1S$ and ${}_2S$. We therefore write $S_{12} = \{{}_1S, {}_2S\}$. Given N and n, the logically possible values of each univariate random quantity coincide with k samples belonging to the set $\mathcal{S}'$. They are all logically possible samples of $\mathcal{S}'$ whose size is equal to n. Each sample of $\mathcal{S}'$ is a vector of $\mathbb{R}^N$. We have to note a very important point: we suppose that the logically possible values of ${}_1S$ and ${}_2S$ are the same N-dimensional vectors of the same N-dimensional linear space over $\mathbb{R}$. These univariate random quantities have then the same marginal distributions of probability. Putting them into a two-way table we observe that it is always a square table. We observe that all probabilities of the joint distribution of probability outside of the main diagonal of this table are always equal to 0. The nonzero probabilities of the joint distribution of probability coincide with $p(s'_1), \ldots, p(s'_k)$. They are on the main diagonal of the table under consideration. A coherent prevision of S_{12} denoted by $\mathbf{P}(S_{12})$

is obtained by means of the sum of k square matrices. The number of rows and columns of every square matrix of this sum is equal to N. Each square matrix of this sum derives from a tensor product belonging to the same linear space denoted by $\mathbb{R}^N \otimes \mathbb{R}^N$. It is an N^2-dimensional linear space over $\mathbb{R}$. We always consider as many tensor products as joint probabilities are associated with the samples of $\mathcal{S}'$. We have then

$$p(s_i')\left(\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix},\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix}\right)\mapsto p(s_i')\left(\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix}\otimes\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix}\right) \tag{25}$$

for every $i = 1, \ldots, k$. We note that it turns out to be

$$p(s_i')\left(\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix}\otimes\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix}\right) = p(s_i')\begin{bmatrix}\delta(1;s_i')\\ \delta(2;s_i')\\ \vdots\\ \delta(N;s_i')\end{bmatrix}\begin{bmatrix}\delta(1;s_i') & \delta(2;s_i') & \ldots & \delta(N;s_i')\end{bmatrix}. \tag{26}$$

If we consider a coherent prevision of S_{12} then we deal with a bilinear map expressed by $\mathbb{R}^N \times \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$, where the linear space over $\mathbb{R}$ of the $N \times N$ matrices in $\mathbb{R}$ is denoted by $\mathcal{M}_{N,N}(\mathbb{R})$. This linear space is isomorphic to $\mathbb{R}^{N^2}$. The matrix product resulting from this bilinear map is factorized by means of the tensor product of vectors of $\mathbb{R}^N$. It is also factorized by means of a unique linear map whose domain coincides with $\mathbb{R}^N \otimes \mathbb{R}^N$. This is because we are able to know a basis of $\mathbb{R}^N \otimes \mathbb{R}^N$ as well as the value of the linear map under consideration on basis elements. We suppose that a basis of $\mathbb{R}^N \otimes \mathbb{R}^N$ results from the standard basis of $\mathbb{R}^N$, where $\mathbb{R}^N$ is evidently considered two times. It is therefore possible to say that there exists a unique linear map given by $\mathbb{R}^N \otimes \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$. It coincides with the product of a joint probability viewed as a scalar and a square matrix. We consider k products of a joint probability and a square matrix. We obtain k square matrices in this way. We consider the sum of these k square matrices in order to obtain a coherent prevision of S_{12}. We observe that $\mathbb{R}^N \times \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$ and $\mathbb{R}^N \otimes \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$ have the same codomain. A factorization of $\mathbb{R}^N \times \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$ is then realized by means of a bilinear map given by $\mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \otimes \mathbb{R}^N$ and a linear map given by $\mathbb{R}^N \otimes \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$. These two maps are connected, so we obtain a composition of functions identified with $\mathbb{R}^N \times \mathbb{R}^N \to \mathcal{M}_{N,N}(\mathbb{R})$. The following commutative diagram

$$\begin{tikzcd} \mathbb{R}^N \times \mathbb{R}^N \arrow[r] \arrow[dr] & \mathbb{R}^N \otimes \mathbb{R}^N \arrow[d] \\ & \mathcal{M}_{N,N}(\mathbb{R}) \end{tikzcd}$$

permits of visualizing what we have said. A coherent prevision of S_{12} is then bilinear and homogeneous. It is given by

$$\mathbf{P}(S_{12}) = \Pi = \begin{bmatrix} \pi_1 & \pi_{12} & \dots & \pi_{1N} \\ \pi_{21} & \pi_2 & \dots & \pi_{2N} \\ \dots & \dots & \dots & \dots \\ \pi_{N1} & \pi_{N2} & \dots & \pi_N \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_{12} & \dots & \pi_{1N} \\ \pi_{12} & \pi_2 & \dots & \pi_{2N} \\ \dots & \dots & \dots & \dots \\ \pi_{1N} & \pi_{2N} & \dots & \pi_N \end{bmatrix}. \tag{27}$$

It coincides with the symmetric matrix of the first-order and second-order inclusion probabilities. The trace of this matrix is evidently equal to n (Angelini [2020]).

7 The covariance of two univariate random quantities obtained by considering two bilinear maps

Given $S_{12} = \{{}_1S, {}_2S\}$, the covariance of ${}_1S$ and ${}_2S$ is expressed by

$$\mathrm{C}({}_1S, {}_2S) = \mathbf{P}(S_{12}) - \mathbf{P}({}_1S)\mathbf{P}({}_2S), \tag{28}$$

where $\mathbf{P}(S_{12})$ represents the prevision or mathematical expectation or expected value of S_{12}, while $\mathbf{P}({}_1S)$ and $\mathbf{P}({}_2S)$ represent the prevision or mathematical expectation or expected value of ${}_1S$ and ${}_2S$. We note that a coherent prevision of S_{12} derives from a bilinear map because we have

$$\mathbf{P}(S_{12}) = \begin{bmatrix} \pi_1 & \pi_{12} & \dots & \pi_{1N} \\ \pi_{21} & \pi_2 & \dots & \pi_{2N} \\ \dots & \dots & \dots & \dots \\ \pi_{N1} & \pi_{N2} & \dots & \pi_N \end{bmatrix}. \tag{29}$$

Moreover, since we have

$$\mathbf{P}({}_1S) = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} \tag{30}$$

as well as

$$\mathbf{P}({}_2S) = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix}, \tag{31}$$

we note that the product of these two linear maps is evidently bilinear. Such a product is expressed in the form

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_N \end{bmatrix} = \begin{bmatrix} \pi_1\pi_1 & \pi_1\pi_2 & \dots & \pi_1\pi_N \\ \pi_2\pi_1 & \pi_2\pi_2 & \dots & \pi_2\pi_N \\ \dots & \dots & \dots & \dots \\ \pi_N\pi_1 & \pi_N\pi_2 & \dots & \pi_N\pi_N \end{bmatrix}. \tag{32}$$

It is then evident that the covariance of ${}_1S$ and ${}_2S$ derives from two bilinear maps because we can write

$$\mathrm{C}({}_1S, {}_2S) = \begin{bmatrix} \pi_1 & \pi_{12} & \dots & \pi_{1N} \\ \pi_{21} & \pi_2 & \dots & \pi_{2N} \\ \dots & \dots & \dots & \dots \\ \pi_{N1} & \pi_{N2} & \dots & \pi_N \end{bmatrix} - \begin{bmatrix} \pi_1\pi_1 & \pi_1\pi_2 & \dots & \pi_1\pi_N \\ \pi_2\pi_1 & \pi_2\pi_2 & \dots & \pi_2\pi_N \\ \dots & \dots & \dots & \dots \\ \pi_N\pi_1 & \pi_N\pi_2 & \dots & \pi_N\pi_N \end{bmatrix}. \tag{33}$$

By writing

$$\mathrm{C}({}_1S, {}_2S) = \begin{bmatrix} (\pi_1 - \pi_1\pi_1) & (\pi_{12} - \pi_1\pi_2) & \dots & (\pi_{1N} - \pi_1\pi_N) \\ (\pi_{21} - \pi_2\pi_1) & (\pi_2 - \pi_2\pi_2) & \dots & (\pi_{2N} - \pi_2\pi_N) \\ \dots & \dots & \dots & \dots \\ (\pi_{N1} - \pi_N\pi_1) & (\pi_{N2} - \pi_N\pi_2) & \dots & (\pi_N - \pi_N\pi_N) \end{bmatrix} \tag{34}$$

we note that it is possible to consider as many random components as inclusion probabilities are studied. A unit of the population under consideration can be included, or not, in a given sample (Bondesson [2010]). This thing is uncertain until a given sample is selected (Hájek [1958]). Two different units of the population under consideration can be included, or not, in the same sample (Deville and Tillé [1998]). This thing is uncertain until a given sample is selected. A component associated with one or two different units of the population under consideration is evidently random for this reason (Connor [1966]). This means that each random component is characterized by a subjective probability. It is an "a priori" probability. It is also characterized by an "a posteriori" probability coinciding with one of the two logically possible values of a random event, 0 and 1. One and only one of these two logically possible values of a random event will be true "a posteriori". On the other hand, it is known that the notion of probability basically deals with an aspect that is included between two extreme aspects. The first extreme aspect deals with situations of non-knowledge or ignorance or uncertainty determining the set of all logically possible samples of a given size viewed as elementary events. They are evidently all logically possible alternatives that can be considered. The second extreme aspect deals with definitive certainty expressed in the form of what it is certainly true or certainly false. Thus, every logically possible sample of a given size definitively becomes true or false. Probability is subjectively distributed by the statistician as a mass over the domain of all logically possible samples of a given size before knowing which is the true sample to be selected "a posteriori". Having said that, the variance of every random component as well as the covariance of two random components are dealt with by means of the first-order and second-order inclusion probabilities. The variance of each random component is represented by every element on the main diagonal of the symmetric matrix given by (34). The covariance of two random components is represented by every element outside of the main diagonal of the square matrix given by (34).

8 A univariate random quantity representing deviations

We define a univariate random quantity representing deviations. We denote it by D. We firstly consider S whose values are all logically possible samples of a given size viewed as elementary events belonging to the set $\mathcal{S}'$. Given N and n, the number of the logically possible values of S is equal to $\binom{N}{n} = k$. The set of the logically possible values of S is then given by $I(S) = \{s'_1, \ldots, s'_k\}$, with $s'_i \in \mathcal{S}'$, $i = 1, \ldots, k$. A nonzero probability denoted by $p(s'_i)$, $i = 1, \ldots, k$, is assigned to each sample of $\mathcal{S}'$. We therefore obtain an N-dimensional vector denoted by π. It represents the first-order inclusion probabilities of all units of the population under consideration. They are all greater than zero. This vector is always independent of the origin of the coordinate system that we could consider. We note that the number of the logically possible values of D is equal to k. It is the same of the one of S. The set of the logically possible values of D is given by $I(D) = \{d'_1, \ldots, d'_k\}$, with

$$d'_i = \left(\begin{bmatrix} \delta(1; s'_i) \\ \delta(2; s'_i) \\ \vdots \\ \delta(N; s'_i) \end{bmatrix} - \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} \right), \tag{35}$$

where we have $i = 1, \ldots, k$. It follows that we have

$$p(s'_1)d'_1 + \ldots + p(s'_k)d'_k = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{36}$$

This means that $\mathbf{P}(S)$ is an N-dimensional vector such that all deviations from it that are multiplied by the corresponding probabilities represent N-dimensional vectors whose sum coincides with the zero vector of $\mathbb{R}^N$. We are now able to calculate the variance of S by using D. We refer to the α-criterion of concordance introduced by Gini. It is a statistical criterion that we innovatively apply to probability viewed as a mass. An absolute maximum of concordance is then realized when each d'_i, $i = 1, \ldots, k$, is multiplied by itself. If each d'_i, $i = 1, \ldots, k$, is multiplied by itself then we obtain k square matrices. Every multiplication that we consider is a tensor product of two vectors of $\mathbb{R}^N$. These two vectors represent two deviations which are the same. The components of these two vectors are then the same. Hence, the variance of S coincides with the sum of k traces of k square matrices. Each trace of the square matrix under consideration is an inner product viewed as an α-product. An α-product is a bilinear form. We consider each $p(s'_i)$, $i = 1, \ldots, k$, as a scalar. Each $p(s'_i)$, $i = 1, \ldots, k$, is firstly a subjective probability. Thus, it always characterizes a random quantity. It is nevertheless viewed as a scalar within this context. We can therefore multiply all components of d'_i by $p(s'_i)$, $i = 1, \ldots, k$. We note that the components of each d'_i, $i = 1, \ldots, k$, are always independent of the origin of the coordinate system that we could consider. We therefore write

$$\sigma_S^2 = \mathrm{tr}\left({d'_1}^T(p(s'_1)d'_1)\right) + \ldots + \mathrm{tr}\left({d'_k}^T(p(s'_k)d'_k)\right). \tag{37}$$

We have evidently introduced a quadratic and linear metric in this way. We therefore note that σ_S^2 is the sum of the squares of k α-norms. It is possible to verify that every trace of a square matrix is an α-product which is an α-commutative product, an α-associative product, an α-distributive product and an α-orthogonal product. We have to note a very important point: S and D are two different quantities from a geometric point of view because they are represented by different sets of N-dimensional vectors. They are nevertheless the same quantity from a randomness point of view. They are characterized by the same probabilities. We therefore observe the same events because we consider only a change of origin.

9 Intrinsic properties of a univariate random quantity representing deviations

Translations and rotations of vectors identifying a given univariate random quantity representing deviations are intrinsic properties of it. They do not depend on the choice of a basis of a given linear space. We say that all vectors of $\mathcal{S}'$ are subjected to the same translation when we consider k sums of two vectors. We consider k sums of two vectors because the number of the elements of $\mathcal{S}'$ is equal to k. The first vector of each sum of them is given by s_i', $i = 1, \ldots, k$. The second vector of each sum of them is given by an arbitrary N-dimensional vector which is always the same. We say that all vectors of $\mathcal{S}'$ are then subjected to the same change of origin. It follows that σ_S^2 is invariant with respect to a translation of all vectors of $\mathcal{S}'$. We say that a quadratic and linear metric is invariant with respect to a translation of all vectors of $\mathcal{S}'$. Concerning a rotation, let $A = (a_j^{i'})$ be an $N \times N$ orthogonal matrix. Each element of this matrix is denoted by two indices. We use contravariant and covariant indices without loss of generality. The contravariant indices represent the rows of the matrix. We have $i' = 1, \ldots, N$. The covariant indices represent the columns of the matrix. We have $j = 1, \ldots, N$. We observe that rotations of all vectors contained in $I(D) = \{d_1', \ldots, d_k'\}$ are characterized by A. We write

$$\mathcal{R}_A(d_i') \colon d_i' \Rightarrow A d_i' = (d_i')^*, \tag{38}$$

where we have $i = 1, \ldots, k$. We evidently denote by $(d_i')^*$ the result of the rotation of the vector d_i' obtained by means of the orthogonal matrix denoted by A. The vector $(d_i')^*$ is an N-dimensional vector. Its components are originated by N linear and homogeneous relationships. We have to note a very important point: $\mathbf{P}(S)$ is an N-dimensional vector such that all rotated deviations from it that are multiplied by the corresponding probabilities represent N-dimensional vectors whose sum coincides with the zero vector of $\mathbb{R}^N$. We have then

$$p(s_1')(d_1')^* + \ldots + p(s_k')(d_k')^* = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{39}$$

If we consider rotated deviations then we write

$$\sigma_{S^*}^2 = \operatorname{tr}\left((d_1')^{*T}(p(s_1')(d_1')^*)\right) + \ldots + \operatorname{tr}\left((d_k')^{*T}(p(s_k')(d_k')^*)\right), \tag{40}$$

where S^* represents a univariate random quantity connected with rotated deviations. Since it turns out to be

$$\sigma_S^2 = \sigma_{S^*}^2, \tag{41}$$

we say that the variance of S is invariant with respect to all rotated deviations obtained by means of the same orthogonal matrix denoted by A. We have therefore introduced a quadratic and linear metric which is invariant with respect to translations and rotations of vectors identifying a univariate random quantity representing deviations.

10 A univariate random quantity representing variations and its intrinsic properties

We define a univariate random quantity representing variations. We denote it by V. Given D, the set of the logically possible values of V is expressed by $I(V) = \{v'_1, \ldots, v'_k\}$, with

$$v'_i = d'_i \frac{1}{\sqrt{\sigma^2_S}}, \tag{42}$$

where we have $i = 1, \ldots, k$. We therefore note that S, D and V are different quantities from a geometric point of view. They are conversely the same quantity from a randomness point of view. It is possible to verify that it turns out to be

$$\sigma^2_V = 1. \tag{43}$$

This index is always equal to 1 independently of the components of d'_i, $i = 1, \ldots, k$. It is evident that these components identify σ^2_S, so we say that $\sigma^2_V = 1$ is also independent of σ^2_S. We observe that rotations of all vectors belonging to $I(V) = \{v'_1, \ldots, v'_k\}$ are always characterized by an $N \times N$ orthogonal matrix. We write

$$(v'_i)^* = (d'_i)^* \frac{1}{\sqrt{\sigma^2_S}}, \tag{44}$$

where we have $i = 1, \ldots, k$. If we consider translations and rotations of vectors identifying a univariate random quantity representing variations then we observe intrinsic properties that we have already considered. We note that V can be subjected to an affine transformation. If V is subjected to an affine transformation then we write

$$V \Rightarrow aV + b, \tag{45}$$

where we have $a \neq 0$. We therefore observe that each vector of $I(aV + b)$ is equal to the corresponding vector of $I(V)$. This means that the components of each vector of $I(aV + b)$ are the same of the ones of the corresponding vector of $I(V)$. Hence, we say that univariate random quantities representing variations are invariant with respect to an affine transformation. Given $S_{12} = \{_1S, _2S\}$, we note that we have ${}_1V = {}_2V = V$ if and only if it turns out to be ${}_1S = {}_2S = S$. It is possible to verify that the covariance of ${}_1V$ and ${}_2V$ is an α-product. It is always equal to 1. On the other hand, it coincides with the Bravais-Pearson correlation coefficient in the case of a perfect direct linear relationship between two quantities. It is possible to verify that the Bravais-Pearson correlation coefficient is invariant with respect to rotations of vectors belonging to $I(V)$. It is therefore invariant with respect to an affine transformation of V. We have to note a very important point: intrinsic properties that we have considered can be related to the random quantities themselves or to specific metric indices based on these quantities. Specific metric indices are evidently based on random quantities representing deviations or variations because we calculate them after taking such random quantities into account. We have to note another very important point: we are not interested in translating or rotating a geometric object in real terms but we are interested in studying its intrinsic properties because these properties are a fundamental consequence of its geometric representation.

11 Metric aspects of an estimate of the population mean

We want to wonder what happens from a metric point of view when we study one or more than one attribute with respect to each element of the population under consideration. We suppose of observing three different and independent characteristics of each element of the population under consideration. We admit this thing without loss of generality. We therefore consider three different and independent variables denoted by X, Y and Z. We note that X is the variable concerning the first attribute of each element of the population under consideration. The variable concerning the second attribute of each element of the population under consideration is denoted by Y. The variable concerning the third attribute of each element of the population under consideration is denoted by Z. If we study only one attribute of each element of the population under consideration then we estimate the population mean by using the univariate Horvitz-Thompson estimator. It is defined by

$$t_{HT}^{(x)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\pi_i} \delta(i; s') x_i, \tag{46}$$

where we have $s' \in \mathcal{S}'$. It is linear and homogeneous (Horvitz and Thompson [1952]). We note that s' is one of the logically possible samples of $\mathcal{S}'$. Also, the weight of the generic unit i of the population under consideration never depends on s'. It is obtained beginning from (17). We have conversely considered all logically possible samples of $\mathcal{S}'$ when we have defined S, D and V. We did not consider only one of them. These random quantities are complementary to the univariate Horvitz-Thompson estimator for this reason. Also, we have always taken $\mathbf{P}(S) = \pi$ into account when we have defined S, D and V. On the other hand, a coherent prevision of S is itself linear and homogeneous. The expected value of the univariate Horvitz-Thompson estimator is given by

$$\mathrm{E}[t_{HT}^{(x)}] = \mu_x. \tag{47}$$

It is equal to the population mean denoted by μ_x for any vector $(x_1\ x_2\ \ldots\ x_N)^T \in \mathbb{R}^N$. We have

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{48}$$

The variance of the univariate Horvitz-Thompson estimator is given by

$$\mathrm{V}(t_{HT}^{(x)}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{x_i}{\pi_i}\frac{x_j}{\pi_j}\Delta_{ij}, \tag{49}$$

where we have $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$, with $i, j = 1, \ldots, N$. We note that Δ_{ij}, $i, j = 1, \ldots, N$, is obtained by means of (34). Since we consider all logically possible samples whose size is equal to n we can also write

$$\mathrm{V}(t_{HT}^{(x)}) = -\frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^2 \Delta_{ij}, \tag{50}$$

where we have again $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$, with $i, j = 1, \ldots, N$ (Yates and Grundy [1953]). This variance is estimated by the univariate Yates-Grundy estimator given by

$$\hat{\mathrm{V}}_{YG}(t_{HT}^{(x)}) = \frac{1}{2N^2}\sum_{i\in s'}\sum_{j\in s'}\left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^2 \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}, \tag{51}$$

where we have $\pi_{ij} > 0$ because we assume that the sampling design is measurable and $\pi_{ij} \leq \pi_i\pi_j$, with $i, j = 1, \ldots, N$. The same thing goes when we consider Y and Z. We have to note a very important point: the variance of S denoted by σ_S^2 coincides with the variance of the univariate Horvitz-Thompson estimator given by (50) when the absolute values of each deviation of x_i from x_j, with $i \neq j = 1, \ldots, N$, are multiples of N. In addition to this thing, the variance of S coincides with the variance of the univariate Horvitz-Thompson estimator given by (50) when the entropy H of the sampling design with fixed sample size is maximum (Tillé and Wilhelm [2017]), where we have

$$H = -\sum_{s'\in\mathcal{S}'} p(s')\log p(s'). \tag{52}$$

We note that H is maximum when we have

$$p(s'_1) = p(s'_2) = \ldots = p(s'_k), \tag{53}$$

with $\sum_{i=1}^{k} p(s'_i) = 1$. It does not turn out to be $p(s') = 0$ within this context. However, if we observe $p(s') = 0$ with regard to (52) then it turns out to be $[0 \log 0] = 0$ by convention. We therefore say that the weights of the univariate Horvitz-Thompson estimator are based on a coherent prevision of S. We have obtained a linear and quadratic metric by considering two univariate random quantities representing deviations. We have obtained the variance of S by using this metric. The same thing goes when we consider Y and Z. We have to note another very important point: by studying three different and independent attributes of each element of the population under consideration we do not jointly consider three variables but we jointly consider two variables at a time. This is because it is not appropriate to use a trilinear form when we deal with metric relationships. If we jointly study two attributes of each element of the population under consideration then we estimate the bivariate population mean by using the bivariate Horvitz-Thompson estimator. We write

$$t_{HT}^{(xy)} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi_i} \delta(i; s') x_i \, \frac{1}{\pi_j} \delta(j; s') y_j \tag{54}$$

when we jointly consider X and Y, where all first-order inclusion probabilities are greater than zero. They are obtained by means of (17). We write

$$t_{HT}^{(xz)} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi_i} \delta(i; s') x_i \, \frac{1}{\pi_j} \delta(j; s') z_j \tag{55}$$

when we jointly consider X and Z, where all first-order inclusion probabilities are greater than zero. They are obtained by means of (17). We write

$$t_{HT}^{(yz)} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi_i} \delta(i; s') y_i \, \frac{1}{\pi_j} \delta(j; s') z_j \tag{56}$$

when we jointly consider Y and Z, where all first-order inclusion probabilities are greater than zero. They are obtained by means of (17). The bivariate Horvitz-Thompson estimator is obtained by multiplying two linear and homogeneous expressions. This means that what we have said concerning the weights of the univariate Horvitz-Thompson estimator does not change. The expected value of the bivariate Horvitz-Thompson estimator concerning X and Y is given by

$$\mathrm{E}[t_{HT}^{(xy)}] = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi_i} \mathrm{E}[\delta(i; s')] x_i \, \frac{1}{\pi_j} \mathrm{E}[\delta(j; s')] y_j. \tag{57}$$

We observe that it turns out to be $\mathrm{E}[\delta(i;s')] = \pi_i$ as well as $\mathrm{E}[\delta(j;s')] = \pi_j$ for every $s' \in \mathcal{S}'$, $i,j = 1,\ldots,N$. It is therefore evident that (57) is equal to the population mean denoted by $\mu_{(xy)}$ for any vector $(x_1 \ x_2 \ \ldots \ x_N)^T \in \mathbb{R}^N$ and $(y_1 \ y_2 \ \ldots \ y_N)^T \in \mathbb{R}^N$, where we have

$$\mu_{(xy)} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} x_i \, y_j. \tag{58}$$

The same thing goes when we consider the expected value of the bivariate Horvitz-Thompson estimator concerning X and Z as well as the expected value of the bivariate Horvitz-Thompson estimator concerning Y and Z. We consider an auxiliary variable denoted by X' related to X when the values of X given by x_i, $i = 1,\ldots,N$, are unknown. We consider an auxiliary variable denoted by Y' related to Y when the values of Y given by y_i, $i = 1,\ldots,N$, are unknown. We consider an auxiliary variable denoted by Z' related to Z when the values of Z given by z_i, $i = 1,\ldots,N$, are unknown. The known values of X' are given by x_i', $i = 1,\ldots,N$. We write

$$\mu_{x'} = \frac{1}{N}\sum_{i=1}^{N} x_i'. \tag{59}$$

If X and X' are approximately proportional then it turns out to be

$$\frac{x_i}{x_i'} \approx \text{constant}, \tag{60}$$

where we have $i = 1,\ldots,N$. The first-order inclusion probabilities chosen by the statistician are then given by

$$\pi_i = \frac{n x_i'}{N \mu_{x'}}, \tag{61}$$

where we have $i = 1,\ldots,N$. We note that such probabilities are used into (23) in order to obtain $p(s_i')$, $i = 1,\ldots,k$, when we have $k = N$. We observe that $p(s_i')$, $i = 1,\ldots,k$, are used in order to obtain a coherent prevision of S. If we have $k \neq N$ then we consider a system of N linear equations with k unknowns, where $\pi_1,\ldots,\pi_N$ are constant terms. We evidently refer to (21). We therefore observe that $\pi_1,\ldots,\pi_N$ represent a coherent prevision of S obtained beginning from $p(s_i')$, $i = 1,\ldots,k$. We observe that α-products and α-norms use $p(s_i')$, $i = 1,\ldots,k$, as scalars. Also the second-order inclusion probabilities characterize our metric structure. They are obtained by means of tensor products having $p(s_i')$, $i = 1,\ldots,k$, as scalars. They are chosen by the statistician because he subjectively chooses $p(s_i')$, $i = 1,\ldots,k$. He is consequently able to observe $\pi_{ij} > 0$, $i,j = 1,\ldots,N$. We have established them in (27). The same thing goes when we consider Y' and Z'.

12 A metric homoscedasticity of different variables identifying different and independent attributes of the units of the population

We have jointly to consider two variables at a time for a metric reason. When we jointly consider X and Y we have firstly to disaggregate $t_{HT}^{(xy)}$. Given

$$t_{HT}^{(x)} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\pi_i} \delta(i; s') x_i \tag{62}$$

and

$$t_{HT}^{(y)} = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\pi_j} \delta(j; s') y_j, \tag{63}$$

the covariance of these two univariate Horvitz-Thompson estimators is therefore expressed by

$$\mathrm{C}(t_{HT}^{(x)}, t_{HT}^{(y)}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{x_i}{\pi_i} \frac{y_j}{\pi_j} \Delta_{ij}, \tag{64}$$

where we have $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$, with $i, j = 1, \ldots, N$. We note that Δ_{ij}, $i, j = 1, \ldots, N$, is obtained by means of (34). The same thing goes when we jointly consider X and Z as well as Y and Z. We note that

$$\mathrm{C}(t_{HT}^{(x)}, t_{HT}^{(x)}) = \mathrm{V}(t_{HT}^{(x)}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{x_i}{\pi_i} \frac{x_j}{\pi_j} \Delta_{ij}, \tag{65}$$

where we have $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$, $i, j = 1, \ldots, N$. We observe that Δ_{ij}, $i, j = 1, \ldots, N$, is obtained by means of (34). We note that

$$\mathrm{C}(t_{HT}^{(y)}, t_{HT}^{(y)}) = \mathrm{V}(t_{HT}^{(y)}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \Delta_{ij}, \tag{66}$$

where we have $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$, with $i, j = 1, \ldots, N$. We observe that Δ_{ij}, $i, j = 1, \ldots, N$, is obtained by means of (34). It is also possible to write

$$\mathrm{C}(t_{HT}^{(z)}, t_{HT}^{(z)}) = \mathrm{V}(t_{HT}^{(z)}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{z_i}{\pi_i} \frac{z_j}{\pi_j} \Delta_{ij}, \tag{67}$$

where we have $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$, with $i, j = 1, \ldots, N$. We observe that Δ_{ij}, $i, j = 1, \ldots, N$, is obtained by means of (34). We are interested in knowing what happens from a metric point of view when we study three different and independent attributes with respect to each element of the population under consideration. We have defined S, D and V. In particular, we consider a bivariate random quantity representing deviations. It is expressed by $D_{12} = \{{}_1D, {}_2D\}$. Its components are two univariate random quantities, ${}_1D$ and ${}_2D$, identifying two sets of N-dimensional vectors. Each vector of a set of N-dimensional vectors is equal to the corresponding vector of the other set of N-dimensional vectors. We have consequently $I({}_1D) = I({}_2D) = \{d'_1, \ldots, d'_k\}$. Given $p(s'_i)$, $i = 1, \ldots, k$, we observe that ${}_1D$ is equal to ${}_2D$, so the covariance of ${}_1D$ and ${}_2D$ is equal to the variance of S denoted by σ^2_S. We observe this thing regardless of any pair of variables that we consider. We could indifferently consider X and Y or X and Z or Y and Z. On the other hand, if we take ${}_1V$ and ${}_2V$ into account then we note that their covariance is equal to 1. Since it turns out to be ${}_1V = {}_2V = V$ we say that the variance of V is equal to 1. We observe this thing regardless of any pair of variables that we consider. We could indifferently consider X and Y or X and Z or Y and Z. We therefore say that X, Y and Z are homoscedastic from a metric point of view. We say this thing after considering all logically possible samples having a given size belonging to $\mathcal{S}'$. We say this thing after defining S with respect to X, Y, Z. We say this thing because, given $p(s'_i)$, $i = 1, \ldots, k$, the variance of S is always the same. It is obtained by virtue of the metric structure that we have introduced.

13 What is all this for?

All the first-order inclusion probabilities derive from a coherent prevision of S. A coherent prevision of S always depends on $p(s'_i)$, $i = 1, \ldots, k$, where these probabilities are coherently chosen by the statistician. All the second-order inclusion probabilities derive from a coherent prevision of S_{12}. A coherent prevision of S_{12} always depends on $p(s'_i)$, $i = 1, \ldots, k$. A coherent prevision of S is linear and homogeneous. A coherent prevision of S_{12} is bilinear and homogeneous. The bivariate Horvitz-Thompson estimator is obtained by multiplying two linear and homogeneous expressions. This means that what we are going to say concerning the weights of the univariate Horvitz-Thompson estimator continues to be valid even when we make reference to the bivariate Horvitz-Thompson estimator. We therefore make reference to the first-order inclusion probabilities. If there exists a direct linear relationship between X' and X then the statistician chooses high inclusion probabilities denoted by π_i with respect to the units of the population under consideration having high attributes of X' denoted by x'_i, $i = 1, \ldots, N$. This is because they are likely associated with high attributes of X denoted by x_i, $i = 1, \ldots, N$. The same thing goes when we consider a direct linear relationship between Y' and Y as well as between Z' and Z. If X and X' are approximately proportional then the first-order inclusion probabilities chosen by the statistician are given by

$$\pi_i = \frac{nx'_i}{\sum_{j=1}^{N} x'_j}, \tag{68}$$

where we have $i = 1, \ldots, N$. If it turns out to be $\pi_i > 1$ for some unit of the population under consideration then we have $\pi_i = 1$ for all units of the population under consideration having i as a label and such that it turns out to be $nx'_i \geq \sum_{j=1}^{N} x'_j$ because x'_i is high. We consider $n > 1$ within this context. The statistician consequently chooses

$$\pi_i = (n - n_A)\frac{x'_i}{\sum_{\substack{j=1\\ j\notin A}}^{N} x'_j}, \tag{69}$$

where we have $i = 1, \ldots, N$, $i \notin A$, concerning the remaining units of the population under consideration. The set of the units of the population under consideration such that it turns out to be $nx'_i \geq \sum_{j=1}^{N} x'_j$ is denoted by A, while their number is denoted by n_A. The same thing goes when we consider Y' and Y as well as Z' and Z. Having said that, we evidently establish a linear relationship between $p(s'_i)$, $i = 1, \ldots, k$, and π_i, $i = 1, \ldots, N$. If the statistician chooses $p(s'_i)$, $i = 1, \ldots, k$, with $\sum_{i=1}^{k} p(s'_i) = 1$, then it is possible to get π_i, $i = 1, \ldots, N$, with $\sum_{i=1}^{N} \pi_i = n$. We write

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} = \sum_{i=1}^{k} \delta(s'_i) p(s'_i). \tag{70}$$

He is consequently able to obtain $\pi_i > 0$ for every $i = 1, \ldots, N$. Conversely, if the statistician chooses π_i, $i = 1, \ldots, N$, then it is possible to get $p(s'_i)$, $i = 1, \ldots, k$. We observe that α-products and α-norms use $p(s'_i)$, $i = 1, \ldots, k$, as scalars. We obtain different metric relationships by using α-norms whose scalars are $p(s'_i)$, $i = 1, \ldots, k$. We note that $\pi_1, \ldots, \pi_N$ are used into

$$B^{-1}\mathbf{P}(S) = \begin{bmatrix} p(s'_1) \\ p(s'_2) \\ \vdots \\ p(s'_k) \end{bmatrix} \tag{71}$$

in order to obtain $p(s'_i)$, $i = 1, \ldots, k$, when we have $k = N$. We note that B is a square matrix, while B^{-1} is its inverse. If we have $k \neq N$ then we consider a system of N linear equations with k unknowns, where $\pi_1, \ldots, \pi_N$ are constant terms. We evidently refer to

$$L_B(Q) = B \begin{bmatrix} p(s'_1) \\ p(s'_2) \\ \vdots \\ p(s'_k) \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} = \mathbf{P}(S). \tag{72}$$

It is known that if the statistician chooses appropriate inclusion probabilities then he is able to obtain a more efficient estimator of the population mean.

14 Conclusions

We have considered random quantities whose logically possible values are all logically possible samples of a given size belonging to a given set. Every logically possible sample belonging to a given set has a subjective probability of being selected. We have obtained the first-order inclusion probabilities by means of coherent previsions of univariate random quantities. We have defined bivariate random quantities whose components are two univariate random quantities having all logically possible samples of a given size as their logically possible values. All univariate random quantities which we have defined are complementary to the univariate Horvitz-Thompson estimator. It is linear and homogeneous like a coherent prevision of a univariate random quantity whose logically possible values are all logically possible samples of a given size belonging to a given set. A univariate random quantity representing deviations as well as a univariate random quantity representing variations are defined on the basis of a coherent prevision of a given univariate random quantity. These random quantities are the same quantity from a randomness point of view. We have identified a quadratic and linear metric with regard to two univariate random quantities representing deviations. We have used the α-criterion of concordance introduced by Gini in order to identify it.

References

Pierpaolo Angelini. A quadratic and linear metric characterizing the sampling design with fixed sample size considered from a geometric viewpoint. *European Scientific Journal*, 16(15):1–19, 2020.

D. Basu. On sampling with and without replacement. *Sankhya: The Indian Journal of Statistics*, 20(3-4):287–294, 1958.

D. Basu. An essay on the logical foundations of survey sampling, part one. In V. P. Godambe and D. A. Sprott, editors, *Foundations of Statistical Inference*. Holt, Rinehart & Winston, Toronto, 1971.

L. Bondesson. Recursion formulas for inclusion probabilities of all orders for conditional Poisson, Sampford, Pareto, and more general sampling designs. In M. Carlson, H. Nyquist, and M. Villani, editors, *Official statistics, methodology and applications in honour of Daniel Thorburn*. Brommatryck & Brolins AB, Stoccolma, 2010.

G. Coletti, R. Scozzafava, and B. Vantaggi. Possibilistic and probabilistic logic under coherence: default reasoning and system P. *Mathematica Slovaca*, 65(4): 863–890, 2015.

W. S. Connor. An exact formula for the probability that two specified sampling units will occur in a sample drawn with unequal probabilities and without replacement. *Journal of the American Statistical Association*, 61:384–390, 1966.

P. L. Conti and D. Marella. Inference for quantiles of a finite population: asymptotic versus resampling results. *Scandinavian Journal of Statistics*, 42:545–561, 2015.

B. de Finetti. Probability: beware of falsifications! In H. E. Kyburg jr. and H. E. Smokler, editors, *Studies in subjective probability*. R. E. Krieger Publishing Company, Huntington, New York, 1980.

B. de Finetti. The role of "dutch books" and of "proper scoring rules". *The British Journal of Psychology of Sciences*, 32:55–56, 1981.

B. de Finetti. Probability: the different views and terminologies in a critical analysis. In L. J. Cohen, J. Łoś, H. Pfeiffer, and K.-P. Podewski, editors, *Logic, Methodology and Philosophy of Science VI*, pages 391–394. North-Holland Publishing Company, Amsterdam, 1982a.

B. de Finetti. The proper approach to probability. In G. Koch and F. Spizzichino, editors, *Exchangeability in Probability and Statistics*. North-Holland Publishing Company, Amsterdam, 1982b.

B. de Finetti. Probabilism: A critical essay on the theory of probability and on the value of science. *Erkenntnis*, 31(2-3):169–223, 1989.

B. de Finetti. La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista. In B. de Finetti, editor, *Induzione e statistica*, pages 5–115. Springer, Heidelberg, 2011.

J.-C. Deville and Y. Tillé. Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85:89–101, 1998.

A. Gilio and G. Sanfilippo. Conditional random quantities and compounds of conditionals. *Studia logica*, 102(4):709–729, 2014.

V. P. Godambe. A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society*, B17(2):269–278, 1955.

V. P. Godambe and V. M. Joshi. Admissibility and bayes estimation in sampling finite populations. i. *The Annals of Mathematical Statistics*, 36(6):1707–1722, 1965.

I. J. Good. Subjective probability as the measure of a non-measurable set. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science*. Stanford University Press, Stanford, 1962.

J. Hájek. Some contributions to the theory of probability sampling. *Bulletin of the international Statistical Institute*, 36(3):127–134, 1958.

H. O. Hartley and J. N. K. Rao. Sampling with unequal probabilities and without replacement. *The Annals of Mathematical Statistics*, 33(2):350–374, 1962.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47 (260):663–685, 1952.

V. M. Joshi. A note on admissible sampling designs for a finite population. *The Annals of Mathematical Statistics*, 42(4):1425–1428, 1971.

Y. Tillé and M. Wilhelm. Probability sampling designs: principles for choice of design and balancing. *Statistical Science*, 32(2):176–189, 2017.

F. Yates and P. M. Grundy. Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B*, 15 (2):253–261, 1953.