A Quadratic and Linear Metric Characterizing the Sampling Design with Fixed Sample Size Considered From a Geometric Viewpoint

Pierpaolo Angelini,

Università ``La Sapienza'' di Roma, Dipartimento di Scienze Statistiche, Italy

Abstract

The first-order and second-order inclusion probabilities are chosen by the statistician. They are subjective probabilities. We innovatively define univariate and bivariate random quantities whose logically possible values are samples of a given size in order to obtain the first-order and second-order inclusion probabilities by means of their coherent previsions. We consider linear maps connected with univariate random quantities as well as bilinear maps connected with bivariate random quantities. The covariance of two univariate random quantities that are the components of a bivariate random quantity has been expressed by means of two bilinear maps. We show that a univariate random quantity denoted by S is complementary to the univariate Horvitz-Thompson estimator. We identify a quadratic and linear metric with regard to two univariate random quantities representing deviations that we innovatively define. We use the α-criterion of concordance introduced by Gini in order to identify it. It is a statistical criterion that we innovatively apply to probability.

Keywords: Tensor Product, Linear Map, Bilinear Map, Horvitz-Thompson Estimator, Quadratic And Linear Metric, Composition Of Functions

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], Hájek [1981], Kish [1965]). 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], Godambe and Joshi [1965]). All logically possible samples of a given size

belong to a given set (Islam et al. [2017]). We suppose that we are always able to number all logically possible samples of a given size belonging to a given set. 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. We simply disregard this thing. 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 (Gladys [2014]). We will always consider sets whose elements are all logically possible samples having a given size selected from a finite population (Oksuz [2015]).

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. A probability is then assigned to each element of this set and this means that it is possible to consider a distribution of probability (Brewer and Hanif [1983], 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 events (Kyburg jr. and Smokler [1964], Savage [1954]). It is then possible to assign a subjective probability to each logically possible event of this partition (Good [1962], Ramsey [1960]). A probability subjectively assigned to each logically possible event of a finite partition of events must be coherent. It is inadmissible only when it is not coherent (Koopman [1940], Jeffreys [1961]). 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. It must therefore be coherent. We have to note a very important point: when we say that it is possible to assign a 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. This implies that such an interval must include both endpoints. It is therefore 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.

1. The set of all logically possible samples of a given size embedded in a linear space provided with a metric on it

We consider a finite set of objects denoted by S in the field **R** of real numbers (Lang [1966]). We number these elements. We consequently write

$$s_1, \ldots, s_N$$

where it turns out to be $s_i \in S$, i = 1, …, N. Each element of S is nothing but a letter with a subscript.

We consider a linear space over **R** of all formal linear combinations of elements of S expressed in the form

$$c_1 s_1 + \ldots + c_N s_N \qquad (1),$$

where every c_i, i = 1, …, N, is a real coefficient.

For the moment, we do not describe the elements of this linear space because we are only interested in considering their addition given by (1). This thing is unusual but it can be done without problem. On the other hand, we speak about formal linear combinations for this reason. We observe that (1) is completely determined by the real coefficients c_1, …, c_N. Each coefficient c_i is associated with the element s_i of the set S. It is known that an association is exactly a function.

For each $s_i \in S$ and c $\in$ **R** we then consider cs_i to be the function that associates c to s_i and 0 to s_j, with $j \neq$ i.

Given a $\in$ **R**, we have a(cs_i) = (ac)s_i.

Given $c' \in$ **R**, we have $(c + c')\, s_i = cs_i + c's_i$.

Thus, it is possible to consider a linear space over **R**. It is the set of all functions of S in R. These functions can be written in the form given by (1).

The functions $1s_1$, …, $1s_N$ are linearly independent, so they represent a basis of the linear space under consideration (Handley [1961]).

We have then to suppose that c_1, …, c_N are elements of **R** such that it is possible to obtain the zero function given by

$$c_1 s_1 + \ldots + c_N s_N = 0.$$

This means that we have $c_i = 0$ for every c_i, i = 1, …, N. This thing consequently proves the linear independence under consideration. Moreover, it is always possible to write s_i instead of $1s_i$. Having said that, we observe that our objects denoted by letters having numbers as subscripts are not generic objects any more but they coincide with N-dimensional vectors. A sample belonging to the set of all logically possible samples of a given size is then expressed by the vector

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

having N components, where s' is a sample of the set of all logically possible samples denoted by S' (Cochran [1977], Särndal et al. [1992], Godambe [1955]).

We will always consider vectors viewed as ordered lists of real numbers within this context. A sample can evidently be denoted by a letter having a number as a subscript as well as it can be expressed by the real coefficients 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 coefficients of a linear combination of the elements of a basis of the linear space under consideration. This linear space is denoted by R^N. Its basis is denoted by $S = \{e_j\}$, j = 1, …, N.

We always consider orthonormal bases within this context. It follows that R^N is also a metric space.
We therefore write

$$\delta(1;\ s')\ e_1 + \delta(2;\ s')\ e_2 + \ldots + \delta(N;\ s')\ e_N = \text{y},$$

where we have y $\in R^N$.

We consider as many linear combinations of the elements of $S = \{e_j\}$, j = 1, …, N, as logically possible samples there are into the set of all logically possible samples of a given size denoted by S'.
We note that the real coefficients of every linear combination of the elements of $S = \{e_j\}$, j = 1, …, N, represent one of the logically possible samples of S'.

We have evidently $\delta(i;\ s') = \begin{cases} 1 \\ 0 \end{cases}$

for every i = 1, …, N, where the elements of the population under consideration are N. If i $\in s'$ then we have $\delta(i;\ s') = 1$, while we obtain 0 otherwise. We consider all logically possible samples of S' having the same size denoted by n. Since the population has N elements we observe that the number of n-combinations is equal to the binomial coefficient denoted by $\binom{N}{n}$. We observe that S' is a subset of R^N. We say that S' is embedded in R^N.

2. Finite partitions of logically possible events

Given N, all logically possible samples whose size is equal to n belong to the set denoted by S'. We have n $= \sum_{i=1}^{N} \delta(i;\ s')$ for every $s' \in S'$.

This means that 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 $[0, 1]^N$ (G. Coletti and D. Petturiti and B. Vantaggi [2016b]). All logically possible samples of S' can be viewed as possible and elementary events of a finite partition of incompatible and exhaustive events (de Finetti [1982b]).

4

We are consequently able to define a univariate random quantity whose logically possible values are represented by all logically possible samples of S' (de Finetti [2011]).

The logically possible values of this univariate random quantity are not real numbers but they are N-dimensional vectors of an N-dimensional linear space over $\mathbf{R}$ (Gilio and Sanfilippo [2014]).

We consider sampling designs specifying a subjective probability for every logically possible sample of S'. Every logically possible sample belonging to S' has a subjective probability of being selected.

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 (de Finetti [1975], de Finetti [1972]). An evaluation of probability known over a set of possible events coinciding with all logically possible samples of S' is admissible when it is coherent. Only coherence is really necessary (G. Coletti and R. Scozzafava [2002]).

This means that it must be $\sum_{s' \in S'} p(s') = 1$.

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 event. This implies that any logically possible sample of S' has always a probability greater than zero of being selected.

We have consequently $0 < p(s') \leq 1$ for every $s' \in S'$ (G. Coletti and D. Petturiti and B. Vantaggi [2016a]).

Thus, conditions of coherence coincide with positivity of every probability of a random event and finite additivity of probabilities of incompatible and exhaustive events (de Finetti [1989]). We will also consider bivariate random quantities whose components are two univariate random quantities. If the logically possible values of these univariate random quantities are the same vectors of the same N-dimensional linear space over $\mathbf{R}$ then these random quantities have the same marginal distributions of probability (G. Coletti and R. Scozzafava and B. Vantaggi [2015]).

They represent the same finite partition of incompatible and exhaustive events (G. Coletti and D. Petturiti and B. Vantaggi [2014]). Putting them into a two-way table we observe that it is always a table having the same number of rows and columns.

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

We innovatively define a univariate random quantity denoted by S whose logically possible values are vectors of $\mathbf{R}^N$. They are all logically possible samples of the set S'. Given $S = \{e_j\}$, j = 1, …, N, each sample of S' coincides with the real coefficients of a linear combination of the vectors of

$S = \{e_j\}$, j = 1, …, 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.

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

It must therefore be $\sum_{i=1}^{k} p(s'_i) = 1$, with $0 < p(s') \leq 1$ for every I = 1, …, k. It is possible to obtain an N-dimensional vector after assigning a nonzero probability to each sample of S'. We denote it with π. It represents the first-order inclusion probabilities of all units of the population under consideration. Thus, we write

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

where wehave $\pi_i > 0$ for every i = 1, …, 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 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 **P**(S). We therefore write

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

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 this set 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 **P**(S) has conveniently been taken under consideration because the logically possible values of S are incompatible and exhaustive 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], de Finetti [1981], de Finetti [1982a]). We evidently consider a convex combination coinciding with **P**(S) within this context, where its weights or coefficients are a priori subjective probabilities connected with the samples of S'. This convex combination is characterized by k column vectors viewed as k matrices. Each row of every N × 1 matrix is a first-order inclusion a posteriori probability. We therefore consider a linear combination of probabilities.

4. First-order inclusion probabilities obtained by means of linear maps

We consider all logically possible samples belonging to the set S'. Given N and n, let k be the number of all elements of S'. We are consequently able to determine an N × k matrix in **R**.

We denote it by B. It is therefore possible to define a linear map expressed by
$$L_B: \boldsymbol{R}^k \to \boldsymbol{R}^N.$$

This linear map depends on B. Moreover, it also depends on the choice of bases for $\boldsymbol{R}^k$ and $\boldsymbol{R}^N$.

We choose standard bases for $\boldsymbol{R}^k$ and $\boldsymbol{R}^N$. We consider all probabilities assigned to the logically possible samples of S' whose size is equal to n. They can be viewed as a column vector. We denote it by Q. We have then
$$Q = \begin{bmatrix} p(s'_1) \\ p(s'_2) \\ \vdots \\ p(s'_k) \end{bmatrix}.$$

Therefore, it turns out to be
$$L_B(Q) = BQ = \pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix}.$$

We note that if k = N then we are able to define a linear map expressed by
$$L_B: \boldsymbol{R}^N \to \boldsymbol{R}^N.$$

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

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

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 S' must be equal to 1. We consequently observe that there are infinite ways of choosing all probabilities of the samples of S'. They are conveniently caught by L_B. Hence, it is 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

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

1. **First-order and second-order inclusion probabilities obtained by means of tensor products**

We define 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 S'.

They are all logically possible samples of S' whose size is equal to n. Each sample of S' is a vector of $\boldsymbol{R}^N$. The logically possible values of $_1S$ and $_2S$ are the same N-dimensional vectors of the same N-dimensional linear space over **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 $p(s'_1)$, …, $p(s'_k)$. 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 results from a tensor product belonging to the same linear space denoted by $\boldsymbol{R}^N \otimes \boldsymbol{R}^N$. It is an N^2-dimensional linear space over **R**. We always consider as many tensor products as joint probabilities are associated with the samples of S'. We have then

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

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

$$p(s_i')\left(\begin{bmatrix}\delta(1;\ s_i')\\\delta(2;\ s_i')\\\vdots\\\delta(N;\ s_i')\end{bmatrix}\otimes\begin{bmatrix}\delta(1;\ s_i')\\\delta(2;\ s_i')\\\vdots\\\delta(N;\ s_i')\end{bmatrix}\right)=p(s_i')\begin{bmatrix}\delta(1;\ s_i')\\\delta(2;\ s_i')\\\vdots\\\delta(N;\ s_i')\end{bmatrix}[\delta(1;\ s_i')\ \ \delta(2;\ s_i')\ \dots$$
$$\delta(N;\ s_i')].$$

When we consider a coherent prevision of S_{12} we deal with a bilinear map expressed by $\boldsymbol{R}^N \times \boldsymbol{R}^N \to M_{N,N}$ (**R**), where the linear space over **R** of the N × N matrices in **R** is denoted by $M_{N,N}$ (**R**). This linear space is isomorphic to $\boldsymbol{R}^{N^2}$. The matrix product resulting from this bilinear map is factorized by means of the tensor product of vectors of $\boldsymbol{R}^N$. It is also factorized by means of a unique linear map whose domain coincides with $\boldsymbol{R}^N \otimes \boldsymbol{R}^N$. This is because we are able to know a basis of $\boldsymbol{R}^N \otimes \boldsymbol{R}^N$ as well as the value of the linear map under consideration on basis elements. We suppose that a basis of $\boldsymbol{R}^N \otimes \boldsymbol{R}^N$ results from the standard basis of $\boldsymbol{R}^N$, where $\boldsymbol{R}^N$ is evidently considered two times.

It is therefore possible to say that there exists a unique linear map given by $\boldsymbol{R}^N \otimes \boldsymbol{R}^N \to M_{N,N}$ (**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 $\boldsymbol{R}^N \times \boldsymbol{R}^N \to M_{N,N}$ (**R**) and $\boldsymbol{R}^N \otimes \boldsymbol{R}^N \to M_{N,N}$ (**R**) have the same codomain. A factorization of $\boldsymbol{R}^N \times \boldsymbol{R}^N \to M_{N,N}$ (**R**) is then realized by means of a bilinear map given by $\boldsymbol{R}^N \times \boldsymbol{R}^N \to \boldsymbol{R}^N \otimes \boldsymbol{R}^N$ and a linear map given by $\boldsymbol{R}^N \otimes \boldsymbol{R}^N \to M_{N,N}$ (**R**). These two maps are connected, so we obtain a composition of functions identified with $\boldsymbol{R}^N \times \boldsymbol{R}^N \to M_{N,N}$ (**R**). A coherent prevision of S_{12} is then bilinear and homogeneous. It is given by

$$P(S_{12})=\Pi=\begin{bmatrix}\pi_1 & \cdots & \pi_{1N}\\\vdots & \ddots & \vdots\\\pi_{N1} & \cdots & \pi_N\end{bmatrix}=\begin{bmatrix}\pi_1 & \cdots & \pi_{1N}\\\vdots & \ddots & \vdots\\\pi_{1N} & \cdots & \pi_N\end{bmatrix}.$$

It coincides with the symmetric matrix of the first-order and second-order inclusion probabilities. It is isomorphic to a vector of $\boldsymbol{R}^{N^2}$. The trace of this matrix is evidently equal to n.

6. 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 $C(_1S, {}_2S) = \mathbf{P}(S_{12}) - \mathbf{P}(_1S)\mathbf{P}(_2S)$,

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} results from a bilinear map because we have

$$\mathrm{P}(\,S_{12}) = \begin{bmatrix} \pi_1 & \cdots & \pi_{1N} \\ \vdots & \ddots & \vdots \\ \pi_{N1} & \cdots & \pi_N \end{bmatrix}$$

Moreover, since we have

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

as well as

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

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

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} [\pi_1\ \pi_2\ \dots\ \pi_N] = \begin{bmatrix} \pi_1\pi_1 & \cdots & \pi_1\pi_N \\ \vdots & \ddots & \vdots \\ \pi_N\pi_1 & \cdots & \pi_N\pi_N \end{bmatrix}.$$

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

$$C(_1S,\ _2S) = \begin{bmatrix} \pi_1 & \cdots & \pi_{1N} \\ \vdots & \ddots & \vdots \\ \pi_{N1} & \cdots & \pi_N \end{bmatrix} - \begin{bmatrix} \pi_1\pi_1 & \cdots & \pi_1\pi_N \\ \vdots & \ddots & \vdots \\ \pi_N\pi_1 & \cdots & \pi_N\pi_N \end{bmatrix}.$$

By writing

$$C(_1S,\ _2S) = \begin{bmatrix} (\pi_1 - \pi_1\pi_1) & \cdots & (\pi_{1N} - \pi_1\pi_N) \\ \vdots & \ddots & \vdots \\ (\pi_{N1} - \pi_N\pi_1) & \cdots & (\pi_N - \pi_N\pi_N) \end{bmatrix} \qquad (3)$$

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], Hájek [1958]). This thing is uncertain until a given sample is selected.

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 two logically possible values, 0 and 1. Only one of these two logically possible values 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. They must evidently be viewed as 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 (3). The covariance of two random components is represented by every element outside of the main diagonal of the square matrix given by (3).

7. Univariate and bivariate random quantities representing deviations

We define another univariate random quantity. We denote it by D. We note that D is based on S. Given N and n, the number of the logically possible values of S is equal to the binomial coefficient given by $\binom{N}{n} = k$. We have I(S) $= \{s'_1, \ldots, s'_k\}$, with $s'_i \in S'$, i = 1, …, k. A nonzero probability denoted by $p(s'_i)$, I = 1, …, k, is assigned to each sample of S'. We obtain π. 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

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

where we have i = 1, …, k.

It follows that we have

11

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

This means that **P**(S) is an N-dimensional vector such that all deviations from it multiplied by the corresponding probabilities represent N-dimensional vectors whose sum coincides with the zero vector of $\boldsymbol{R}^N$. We are able to calculate the variance of S by using D. We use $D_{12} = \{_1D, _2D\}$, where D_{12} is a bivariate random quantity representing deviations whose components are two univariate random quantities representing deviations which are the same. We denote them by $_1D$ and $_2D$.

We refer to the α-criterion of concordance introduced by Gini. It is a statistical criterion that we innovatively apply to probability. An absolute maximum of concordance is then realized when each d_i', i = 1, …, k, is multiplied by itself. If each d_i', i = 1, …, k, is multiplied by itself then we obtain k square matrices. Every multiplication that we consider is a tensor product of two vectors of $\boldsymbol{R}^N$. 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. 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, …, k, as a scalar. Each $p(s_i')$, i = 1, …, 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, …, k. We write $\sigma_S^2 = \mathrm{tr}\,(d_1'^T (p(s_1')\,d_1')) + \ldots + \mathrm{tr}\,(d_k'^T (p(s_k')\,d_k'))$.

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.

8.   Metric aspects of an estimate of the population mean
We wonder what happens from a metric viewpoint when we study one attribute with respect to every element of the population under consideration (Hassanein and Elmelegy [2014]).

Let X be the variable concerning this attribute. 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

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

where we have $s' \in S'$. It is linear and homogeneous (Horvitz and Thompson [1952]). We note that s' is one of the logically possible samples of S'. Also, the weight of the generic unit i of the population under consideration never depends on s'. It is obtained beginning from (2). We have conversely considered all logically possible samples of S' when we have defined S. We did not consider only one of them. We say that S is complementary to the univariate Horvitz-Thompson estimator for this reason. We have taken **P**(S) = π into account after defining S. We observe that a coherent prevision of S is itself linear and homogeneous. The expected value of the univariate Horvitz-Thompson estimator is given by

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

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

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

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

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

where we have $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$, with i, j = 1, …, N. We note that Δ_{ij}, i, j = 1, …, N, is obtained through (3). Since we consider all logically possible samples whose size is equal to n we can also write

$$V(t_{HT}^{(x)}) = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^2 \Delta_{ij} \qquad (4),$$

where we have again $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$, with i, j = 1, …, N (Yates and Grundy [1953]).

This variance is estimated by the univariate Yates-Grundy estimator given by

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

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, …, N. 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 (4) when the absolute values of each deviation of x_i from x_j, with i $\neq$ j = 1, …, 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 (4) when the entropy H of the sampling design with fixed sample size is maximum (Tillé and Wilhelm [2017]), where we have

$$H = - \sum_{s' \in S'} p(s') \log p(s') \qquad (5).$$

We note that H is maximum when we have

$$p(s_1') = p(s_2') = \dots = p(s_k'),$$

with $\sum_{i=i}^{k} p(s_i') = 1$.

It does not turn out to be $p(s') = 0$ within this context. However, if we observe $p(s') = 0$ with regard to (5) 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 a particular random quantity that we have innovatively defined. We have denoted it by S. On the other hand, we have obtained a linear and quadratic metric by considering two univariate random quantities denoted by $_1D$ and $_2D$. They are based on S. We have obtained the variance of S by using this metric.

9.   Why it is meaningful what we have shown

We consider an auxiliary variable denoted by X' related to X when the values of X given by x_i, i = 1, …, N, are unknown. The known values of X' are given by x_i', i = 1, …, N. We write

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

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

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

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

$$\pi_i = \frac{nx_i'}{N\mu_{x'}} \tag{6},$$

where we have i = 1, …, N. 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, …, N. This is because they are likely associated with high attributes of X denoted by x_i, i = 1, …, N. If X and X' are approximately proportional then the first-order inclusion probabilities chosen by the statistician are given by (6). It is also possible to write

$$\pi_i = \frac{nx_i'}{\sum_{j=1}^{N} x_j'} ,$$

where we have i = 1, …, 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

$$\pi_i = (n - n_A) \frac{x_i'}{\sum_{j=1}^{N} x_j'} ,$$

where we have j ∉ A, i = 1, …, N, i ∉ 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. Having said that, we evidently establish a linear relationship between $p(s'_i)$, i = 1, …, k, and π_i, i = 1, …, N.

If the statistician chooses $p(s'_i)$, i = 1, …, k, with $\sum_{i=i}^{k} p(s'_i) = 1$, then it is possible to get π_i, i = 1, …, N, with $\sum_{i=1}^{N} \pi_i = $ n. We write

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

He is consequently able to obtain $\pi_i > 0$ for every i = 1, …, N. He methodologically distinguishes what it is logically possible from what it is subjectively probable. All samples belonging to S' are logically possible because they are not either certainly true or certainly false. Conversely, if the statistician chooses π_i, i = 1, …, N, then it is possible to get $p(s'_i)$, i = 1, …, k. We observe that α-products and α-norms use $p(s'_i)$, i = 1, …, k, as scalars. We obtain different metric relationships by using α-norms whose scalars are $p(s'_i)$, i = 1, …, k.

We note that π_1, …, π_N are used into

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

in order to obtain $p(s'_i)$, i = 1, …, k, when we have k = N. We note that B is a square matrix while B^{-1} is its inverse. If we have k ≠ N then we consider a system of N linear equations with k unknowns, where π_1, …, π_N are constant terms. We evidently refer to

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

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.

Conclusion

   We have defined univariate and bivariate 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 a coherent prevision of a univariate random quantity denoted by S whose logically possible values are all logically possible samples of a given size belonging to a given set. We have defined a

bivariate random quantity denoted by S_{12} whose components are two univariate random quantities having all logically possible samples of a given size as their logically possible values. We have shown that S is complementary to the univariate Horvitz-Thompson estimator. This estimator is linear and homogeneous like a coherent prevision of S. We have identified a quadratic and linear metric with regard to two univariate random quantities representing deviations that we have innovatively defined. We have used the α-criterion of concordance introduced by Gini in order to identify it. It is a statistical criterion that we have innovatively applied to probability.

References:
1. D. Basu, On sampling with and without replacement, Sankhya 20, 287-294, 1958.
2. D. Basu, An essay on the logical foundations of survey sampling, part one. In V. P. Godambe and D. A. Sprott (Eds.), Foundations of Statistical Inference. Holt, Rinehart & Winston, Toronto, 1971.
3. 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 (Eds.), Official statistics, methodology and applications in honour of Daniel Thorburn. Brommatryck & Brolins AB, Stoccolma, 2010.
4. K. R. W. Brewer and M. Hanif, Sampling with unequal probabilities, Springer Verlag, New York, 1983.
5. W. G. Cochran, Sampling techniques, third edition, J. Wiley & Sons, New York, 1977.
6. 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.
7. 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.
8. B. de Finetti, Probability, Induction, and Statistics (The art of guessing), J. Wiley & Sons, London-New York-Sydney-Toronto, 1972.
9. B. de Finetti, Theory of probability, 2 vols, J. Wiley & Sons, London-New York-Sydney-Toronto, 1975.
10. B. de Finetti, Probability: beware of falsifications!. In H. E. Kyburg jr. and H. E. Smokler (Eds.), Studies in subjective probability. R. E. Krieger Publishing Company, Huntington, New York, 1980.
11. B. de Finetti, The role of "dutch books" and of "proper scoring rules", The British Journal of Psychology of Sciences 32, 55-56, 1981.

12. B. de Finetti, Probability: the different views and terminologies in a critical analysis, Logic, Methodology and Philosophy of Science VI: Proceedings of the Sixth International Congress of Logic, Methodology, and Philosophy of Science, Hannover 1979, 391-394, 1982a.

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

14. 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.

15. B. de Finetti, La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista. In B. de Finetti (Ed.), Induzione e statistica, 1-122, Springer, Heidelberg, 2011.

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

17. G. Coletti and D. Petturiti and B. Vantaggi, Bayesian inference: the role of coherence to deal with a prior belief function, Statistical Methods & Applications 23, 4, 519-545, 2014.

18. G. Coletti and D. Petturiti and B. Vantaggi, When upper conditional probabilities are conditional possibility measures, Fuzzy Sets and Systems 304, 45-64, 2016a.

19. G. Coletti and D. Petturiti and B. Vantaggi, Conditional belief functions as lower envelopes of conditional probabilities in a finite setting, Information Sciences 339, 64-84, 2016b.

20. G. Coletti and R. Scozzafava, Probabilistic logic in a coherent setting, Kluwer Academic Publishers, Dordrecht/Boston/London, 2002.

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

22. Gilio and G. Sanfilippo, Conditional random quantities and compounds of conditionals, Studia logica 102, 4, 709-729, 2014.

23. C.-O. Gladys, Impacts of mathematical skills on national security, European Scientific Journal 10, 33, 109-114, 2014.

24. V. P. Godambe, A unified theory of sampling from finite populations, Journal of the Royal Statistical Society B17, 2, 269-278, 1955.

25. 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.

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

27. J. Hájek, Some contributions to the theory of probability sampling, Bulletin of the International Statistical Institute 36, 3, 127-134, 1958

28. J. Hájek, Sampling from a finite population, M. Dekker, New York, 1981.

29. G. Handley, Linear algebra, Addison-Wesley, Reading MA, 1961.

30. 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.

31. W. A. Hassanein and A. A. Elmelegy, Clustering algorithms for categorical data using concepts of significance and dependence of attributes, European Scientific Journal 10, 3, 381-400, 2014.

32. 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.

33. M. R. Islam and I. B. Shafique and K. Rahman and A. Haque, A simple study on weight and height of students, European Scientific Journal 13, 6, 63-71, 2017.

34. H. Jeffreys, Theory of probability, 3rd edition, Clarendon Press, Oxford, 1961.

35. V. M. Joshi, A note on admissible sampling designs for a finite population, The Annals of Mathematical Statistics 42, 4, 1425-1428, 1971.

36. L. Kish, Survey sampling, J. Wiley & Sons, New York, 1965.

37. B. O. Koopman, The axioms and algebra of intuitive probability, Annals of Mathematics 41, 269-292, 1940.

38. H. E. Kyburg jr. and H. E. Smokler, Studies in subjective probability, J. Wiley & Sons, New York, London, Sydney, 1964.

39. S. Lang, Linear algebra, Addison-Wesley, Reading, Massachusetts, 1966.

40. C. Oksuz, Examining Primary school students' levels of mathematics motivation, European Scientific Journal 11, 28, 51-63, 2015.

41. F. P. Ramsey, The foundations of mathematics and other logical essays. Edited by R. B. Braithwaite. With a preface by G. E. Moore. Littlefield, Adams & Co, Paterson, N. J., 1960.

42. C. E. Särndal, B. Swensson, and J. Wretman, Model assisted survey sampling, Springer Verlag, New York, 1992.

43. L. J. Savage, The foundations of Statistics, J. Wiley & Sons, New York, 1954.

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

45. 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.