Resenhas IME-USP 1998, Vol. 3, No. 3, 273 - 290.

The Operational Baysean Approach for Finite Populations

Heleno Bolfarine, Loretta B. Gasco
and Pilar L. Iglesias

Abstract: In this paper we discuss invariant prediction in finite populations. It is assumed that the distribution of the observable quantities is invariant under an orthogonal group of transformations. The quantities of interest are introduced as operational parameters, which depend only on observable quantities. Interest centers on the population total and on the finite population regression coefficient although predictors for the finite population variance are also considered. An operational likelihood function is defined which is a function of the operational parameters. Bayes estimators for the operational parameters are obtained by using the operational likelihood under noninformative and informative prior distributions. As shown, the Pearson type II distributions plays an important role on deriving the main results.

key words: Operational Parameters; Bayesian approach; Pearson Type II distribution: Inference in finite populations.

1 Introduction

Inference in finite populations deals with the problem of gaining information about certain quantities that describe the behavior of one or a set of variables in a finite population of individuals or subjects, by using information on a subset (sample) of the population. In opinion pools, for example, one problem of great interest is to infer on the total of supporters of a certain candidate in a certain city or the total of smokers in a certain region. In others, the interest is on studying the relationship between two quantities of interest as, for example, expenditure and income. In situations like the ones described, there is a finite population of N identifiable units denoted by $\mathcal{P} = \{1, \ldots, N\}$ where N, the population size, is known. We consider that associated with unit k of $\mathcal{P}$ there is a p-dimensional vector $\mathbf{X}_k$, of known quantities and the unknown value of the characteristic of interest that we shall denote by Y_k, $k = 1, \ldots, N$.

In matrix notation, we consider

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} 1 & \mathbf{X}_1' \\ \vdots & \vdots \\ 1 & \mathbf{X}_N' \end{pmatrix},$$

of dimension $N \times m$ with $m = p + 1$, of rank m and where $\mathbf{Y}$ is the vector of unknown quantities associated with the finite population $\mathcal{P}$. For example, the vector $\mathbf{Y}$ may correspond to expenditure of the N families and the elements of $\mathbf{X}_j$ to salary, size of the family, and social indicators, among others. To obtain information on an unknown quantity $\theta(\mathbf{Y}, \mathbf{X})$, a sample s of size n of $\mathcal{P}$ is selected according to some specified sampling plan.

There are several approaches for treating this type of problem. One approach, known as the classical approach to sampling theory (see Cochran, 1973; Godambe, 1966; Basu, 1969, 1971) is based on the fact that the sampling plan used to select the sample s is the unique probability structure used for making inference on θ. Actually, it is a distribution free component of the general theory, as, for example, randomization tests and bootstrapping in infinite populations. Another approach is known as the superpopulation approach to the prediction problem in finite populations (Bolfarine and Zacks, 1992) where it is considered that the population is a random sample generated from an infinite population. Typically, the superpopulation model is represented in a parametric form, with the distribution of $\mathbf{Y}$ being partially or totally specified, for example, by considering that $\mathbf{Y}$ is normally distributed. According to this approach, the superpopulation model establishes the main relations between observed and unobserved units of $\mathbf{Y}$. Moreover, an intermediate step in making inference on the finite population quantity θ is to deal with the unknown parameters (typically termed as superparameters) of the superpopulation model. Making inference on θ becomes then a prediction problem. Within the ordinary Bayesian formulation, it is necessary to assign priors for the superparameters as can be seen in Bolfarine and Zacks (1992). An alternative formulation for the finite population problem is the operational (predictivistic) Bayesian approach which has recently been considered in Iglesias (1993), Barlow and Mendel (1993) and Mendel (1994), among others. The main idea is to define a family of distributions for the vector of unknown population quantities $\mathbf{Y}$ and then find an statistics $\theta(\mathbf{Y}, \mathbf{X})$, which is also a sufficient statistics for this family. This statistics is termed by Mendel (1994) (see also Barlow and Mendel, 1993) as the operational parameter for the family of distributions. Thus, once a prior distribution is specified for $\theta(\mathbf{Y}, \mathbf{X})$, inference for this quantity is based on the posterior distribution. The family of distributions for $\mathbf{Y}$ is usually specified in terms of invariance conditions on its arguments. In this paper, distributions which are invariant under subgroups of the group of orthogonal transformations are considered in connection with finite populations when auxiliary variables are present.

Quantities $\theta(\mathbf{Y}, \mathbf{X})$ like the population total $T = \sum_{i=1}^{N} Y_i$ or the finite population regression coefficient $\mathbf{B}_N = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ have been the subject of a great attention in the recent statistical literature. Comprehensive reviews can be found in Cassel et al. (1977) and Bolfarine and Zacks (1992). Most of the literature associated with the subject consider that the quantities Y_k and $\mathbf{X}_k$ are related through the linear relation

$$Y_k = \beta_0 + \mathbf{X}_k'\boldsymbol{\beta} + \epsilon_k, \tag{1.1}$$

$k = 1, \ldots, N$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, a p-dimensional vector of fixed and unknown parameters and $\epsilon = (\epsilon_1, \ldots, \epsilon_N)'$ is a vector of random errors, typically satisfying $E[\epsilon] = \mathbf{0}$ and $Var[\epsilon] = \sigma^2 \mathbf{I}_N$, with σ^2 unknown and $\mathbf{I}_N$ the identity matrix of dimension N. To obtain information on quantities $\theta(\mathbf{Y}, \mathbf{X})$ like T or $\mathbf{B}_N$, a sample $\mathbf{s}$ is selected from $\mathcal{P}$ according to some specified sampling plan. The unobserved part of $\mathcal{P}$ is denoted by $\mathbf{r} = \mathcal{P} - \mathbf{s}$. Given $\mathbf{s}$, we denote by $\mathbf{Y}_s = (Y_1, \ldots, Y_n)'$ and $\mathbf{Y}_r = (Y_{n+1}, \ldots, Y_N)'$, the observed and unobserved parts of $\mathbf{Y}$, respectively, with the corresponding partition

$$\mathbf{X}_s = \begin{pmatrix} 1 & \mathbf{X}_1' \\ \vdots & \vdots \\ 1 & \mathbf{X}_n' \end{pmatrix} \quad \text{and} \quad \mathbf{X}_r = \begin{pmatrix} 1 & \mathbf{X}_{n+1}' \\ \vdots & \vdots \\ 1 & \mathbf{X}_N' \end{pmatrix},$$

of the matrix $\mathbf{X}$. As pointed out above, the superpopulation (or model based) approach to the prediction problem consider the model (or superpopulation) parameters $(\beta_0, \boldsymbol{\beta}, \sigma^2)$ as the main connection between $\mathbf{s}$ and $\mathbf{r}$, no matter which sampling plan is used and inference should only be based on the superpopulation model (1.1). Thus, under the perspective of the superpopulation approach, according to the conditionality principle (Basu, 1975), the sampling plan is not relevant for inference. The design-model based approach utilizes model (1.1) only in the definition of estimators and to propose convenient sampling plans. However, the merits of an estimator is totally judged by its performance with respect to the sampling plan.

In this paper, we focus on the prediction of the population quantities $\theta(\mathbf{Y}, \mathbf{X})$ from a pure predictivistic approach. The main assumption is that the distribution of $\mathbf{Y}$ is $O_N(M)$ invariant, that is, $\mathbf{Y}$ and $\boldsymbol{\Gamma}\mathbf{Y}$ are identically distributed for all $\boldsymbol{\Gamma}$ in $O_N(M)$, where

$$O_N(M) = \{\boldsymbol{\Gamma} \in O_N; \boldsymbol{\Gamma}\mathbf{x} = \mathbf{x}, \mathbf{x} \in M\},$$

where O_N is the compact subgroup of orthogonal $N \times N$ matrices and M is the space generated by the columns of the matrix $\mathbf{X}$. A consequence of the invariance assumption is the Pearson type II representation of the marginal distributions of $\mathbf{Y}$ given in Corollary 2.1 and Example 2.2. One important consequence of the assumptions is that $\mathbf{Y}_s$ and $\mathbf{X}_s$ are related through the linear model

$$\mathbf{Y}_s = \mathbf{X}_s \mathbf{B}_N + \mathbf{e}_s, \tag{1.2}$$

with $\mathbf{B}_N$ as above and the distribution of $\mathbf{e}_s = (e_1, \ldots, e_n)'$ given $\mathbf{B}_N$ and $\mathbf{S} = \mathbf{Y}'\mathbf{Q}_M\mathbf{Y}$ is the Pearson type II distribution $MPII(\mathbf{0}, \mathbf{S}(\mathbf{I}_n - \mathbf{X}_s(\mathbf{X}_s'\mathbf{X}_s)^{-1}\mathbf{X}_s', \frac{N-m-n-2}{2})$. We call attention to the fact that the invariance condition determines all the elements of the Bayesian model, the parameters, however, being functions only of observable quantities. The marginal distribution of $\mathbf{Y}$ then becomes completely specified as soon as prior distributions are specified for the (finite population) model parameters.Another important aspect of the development is that it provides justification for the fact that the normal distribution is typically associated with the distribution of the error vector ϵ. Operational Bayesian approach is the nomenclature typically associated with such approach (Mendel, 1994) and the finite population parameters (quantities of interest) are termed operational parameters. This approach was considered by Mendel(1992) and Barlow and Mendel (1993) in reliability theory.

The paper is organized as follows. Section 2 discusses the construction of $O_N(M)$-invariant distributions in finite populations and presents also some representation theorems for the distribution of the observed part of the population, represented as projected measures of $O_N(M)$-invariant distributions. In Section 3 a systematic approach is considered for proposing operational parameters in finite populations. As a consequence, operational likelihood and operational prior are formulated allowing the development of an operational prediction Bayesian theory in finite populations. Finally, Bayesian inference solutions are presented for the operational parameters in Section 4.

2 $O_N(M)$-invariant distributions in finite populations

We consider the distribution of the population vector $\mathbf{Y}$ to be $O_N(M)$-invariant. Consequences of this assumption including characterizations in terms of maximal invariants and representation of such distributions as mixtures of uniform distributions defined on the orbits induced by the maximal invariants under the groups considered are studied. Moreover, following some results in Diaconis et al. (1992), stochastic representations of such distributions in terms of the Pearson type II distributions are considered. These results will be used in Section 3 to obtain operational likelihoods for the prediction problem in finite populations.

2.1 Construction of $O_N(M)$-invariant distributions

Let M denote an m-dimensional subspace of $\Re^N$. Moreover, let's denote by O_N the compact subgroup of all real orthogonal $N \times N$ matrices and by $O_N(M) = \{\mathbf{\Gamma} \in O_N; \mathbf{\Gamma x} = \mathbf{x}, \mathbf{x} \in M\}$. Let $\mathcal{L}_N$ be the set of all N-dimensional real vectors, and $\mathbf{Y}$ a N-dimensional random vector taking values in $\mathcal{L}_N$. The random vector $\mathbf{Y}$ is said to be $O_N(M)$-invariant if $\mathbf{Y}$ and $\mathbf{\Gamma Y}$ are identically distributed for all $\mathbf{\Gamma} \in O_N(M)$. The main interest is on representing such $O_N(M)$-invariant distributions (and marginals) as a mixture of appropriate uniform (or marginals of uniform) distributions.

Definition 2.1. *Let* $\mathbf{U}$ *be a random matrix with values in* $O_N(M)$. *Then* $\mathbf{U}$ *is uniformly distributed on* $O_N(M)$ *provided* $\mathbf{U}$ *is distributed according to the probability measure* ν, *which is the unique invariant probability measure on* $O_N(M)$.

Existence and uniqueness of the probability measure ν is guaranteed by the fact that $O_N(M)$ is a compact subgroup (Nachbin, 1965). Furthermore, using results in Diaconis et al. (1992), it can be shown that

$$E[\mathbf{U}] = \mathbf{P}_M \quad \text{and} \quad Var[\mathbf{U}] = \frac{1}{N-m}\mathbf{Q}_M \otimes \mathbf{Q}_M, \tag{2.1}$$

where $\mathbf{P}_M$ is the orthogonal projection matrix on M, $\mathbf{Q}_M = \mathbf{I}_N - \mathbf{P}_M$ is the projection matrix on $M^\perp$ and $\mathbf{A} \otimes \mathbf{B}$ denotes the Kroenecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$. The action of $O_N(M)$ on $\mathcal{L}_N$ yields a partition of $\mathcal{L}_N$ into orbits as follows. If $\mathbf{y} \in \mathcal{L}_N$ then the orbit of $\mathbf{y}$, O_y, is the set

$$O_y = \{\mathbf{z} \in \mathcal{L}_N; \mathbf{z} = \mathbf{\Gamma y}, \text{ for } \quad \text{some} \quad \mathbf{\Gamma} \in O_N(M)\}.$$

An alternative way of characterizing the orbits of $\mathcal{L}_N$ is by using maximal invariants corresponding to the action of the group $O_N(M)$. Denoting by $t(.)$ such maximal invariant, it follows that

$$O_y = \{\mathbf{z} \in \mathcal{L}_N; t(\mathbf{z}) = t(\mathbf{y})\}.$$

We call attention to the fact that the orbits of $\mathcal{L}_N$ are indexed by the maximal invariant statistics $t(.)$.

In the special case of the group $O_N(M)$, where $M = \text{span}(\mathbf{1}_N)$, namely, the space generated by the N-dimensional vector of ones, the orbits O_y corresponding to $\mathbf{y} \in \Re^N$ are

$$O_y = \{\mathbf{z} \in \mathcal{L}_N; \mathbf{P}_M\mathbf{z} = \mathbf{P}_M\mathbf{y}, \mathbf{z}'\mathbf{Q}_M\mathbf{z} = \mathbf{y}'\mathbf{Q}_M\mathbf{y}\}, \tag{2.2}$$

which geometrically represent spheres in $M^\perp$ centered in $\mathbf{P}_M\mathbf{y}$ and with radius $\mathbf{y}'\mathbf{Q}_M\mathbf{y}$.

Now, the main purpose is to build the uniform distribution on O_y, the orbit of $\mathbf{y}$ generated by the action of $O_N(M)$. Since O_y is compact in $\mathcal{L}_N$ (a locally compact Hausdorff space with an enumerable basis for the Euclidean topology in $\Re^N$) and $O_N(M)$ acts transitively on O_y it follows that there exists a unique $O_N(M)$-invariant probability measure, ν, defined on the orbit O_y (Nachbin, 1965). However, if $\mathbf{U}$ is a random matrix uniformly distributed over $O_N(M)$ and $\mathbf{y} \in \mathcal{L}_N$ then $\mathbf{Uy}$ is a random matrix with values in O_y and with a $O_N(M)$-invariant distribution generated by ν, which we denote by ν_y. Thus, uniqueness of this measure implies that it is the corresponding invariant probability measure on the orbit O_y. In this sense, it is said that the random matrix $\mathbf{Uy}$ has uniform distribution on O_y, which is the surface of an $(N-m)$-dimensional sphere as in (2.2). Note from (2.1) that $E[\mathbf{Uy}] = \mathbf{P}_M\mathbf{y}$ and $Var[\mathbf{Uy}] = {S_y}^2\mathbf{Q}_M$, with $S_y^2 = \mathbf{y}'\mathbf{Q}_M\mathbf{y}/(N-m)$. Consequently, if P denotes the probability measure of a N-dimensional random vector $\mathbf{Y}$ with

$$P = \int \nu_y P(dy), \tag{2.3}$$

then $\mathbf{Y}$ has a $O_N(M)$-invariant distribution. The converse also holds. A simple situation is illustrated in this context in the following examples.

Example 2.1. *$O_N(M)$-invariance.* As before, let M denotes a m-dimensional subspace of $\Re^N$ and $O_N(M)$, the compact set of $N \times N$ orthogonal matrices leaving invariant the elements of M and $\mathbf{P}_M$ the orthogonal projection matrix onto M. Moreover, let $\mathbf{Y}$ a N-dimensional random vector with a O_N-invariant distribution P. In this case, $t(\mathbf{Y}) = (\mathbf{P}_M\mathbf{Y}, ||\mathbf{Y} - \mathbf{P}_M\mathbf{Y}||^2)$ with $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, is a maximal invariant under the action of $O_N(M)$. As a consequence, the conditional distribution of $\mathbf{Y}$ given $\mathbf{P}_M\mathbf{Y} = \mathbf{c}$ and $||\mathbf{Y} - \mathbf{P}_M\mathbf{Y}||^2 = r^2$ is uniform over the set $S_N(\mathbf{c}, r) = \{\mathbf{Y} \in \Re^N; \mathbf{P}_M\mathbf{Y} = \mathbf{c}, ||\mathbf{Y} - \mathbf{P}_M\mathbf{Y}||^2 = r^2\}$, $\mathbf{c} \in M$ and $r > 0$. Thus, the distribution P can be represented as a center-radial mixture of these uniform distributions. The measure in the mixture is the P-law of $t(\mathbf{Y}) = (\mathbf{P}_M, ||\mathbf{Y} - \mathbf{P}_M\mathbf{Y}||^2)$. For example, if $\mathbf{Y} \sim N_N(\mathbf{m}, \mathbf{I}_N)$, then $\mathbf{P}_M\mathbf{Y}$ and $||\mathbf{Y} - \mathbf{P}_M\mathbf{Y}||^2$ are independent with $\mathbf{P}_M\mathbf{Y} \sim N_N(\mathbf{P}_M\mathbf{m}, \mathbf{P}_M)$ and $||\mathbf{Y} - \mathbf{P}_M\mathbf{Y}||^2 \sim \chi^2_{N-m}$. If M is the column space generated by $\mathbf{1}_N$, then the measure P can be represented as a mixture of uniform distributions over the set $S_N(a, r) = \{\mathbf{Y} \in \Re^N; \bar{Y} = a, \sum_{i=1}^N (Y_i - \bar{Y})^2 = r^2\}$, $a, r \in \Re$, $r > 0$, where $\bar{Y} = \sum_{i=1}^n Y_i/n$. Similarly, if M is the column space generated by a $N \times p$ matrix $\mathbf{X}$ with rank p ($p < N$), then P can be represented as a mixture of uniform distributions over the set

$$S_N(\mathbf{b}, r) = \{\mathbf{Y} \in \Re^N; \mathbf{B}_N = \mathbf{b}, ||\mathbf{Y} - \mathbf{X}\mathbf{B}_N||^2 = r^2\},$$

where $\mathbf{B}_N$ is the finite population regression coefficient and $\mathbf{b} \in \Re^p$.

2.2 Marginal distributions

In the section, we consider the problem characterizing the distribution of the observed part $\mathbf{Y}_s$ of the unobserved $\mathbf{Y}$ given that the distribution of $\mathbf{Y}$ is $O_N(M)$-invariant. Let $\mathbf{Y}$ be a $N \times p$-dimensional random matrix with a $O_N(M)$-invariant distribution P and $\mathbf{\Pi}$ denotes a $n \times N$ matrix such that $\mathbf{\Pi}\mathbf{\Pi}' = \mathbf{I}_n$. The matrix $\mathbf{\Pi}$ is a projection matrix and $\mathbf{\Pi}P$ is called the projected measure. For example, if $P^{(n)}$, $n < N$ is the marginal law of the $N \times p$ random matrix $\mathbf{Y}$, then $P^{(n)} = \mathbf{\Pi}P$, with $\mathbf{\Pi} = [\mathbf{I}_n \ \mathbf{0}]$, which is of dimension $n \times N$. Because P is a $O_N(M)$-invariant measure, it follows from representation (2.3) that

$$\mathbf{\Pi}P = \int \mathbf{\Pi}\nu_y P(d\mathbf{y}). \tag{2.4}$$

That is, the projected measure of a $O_N(M)$-invariant measure can be represented as a mixture of projected uniform measures on spheres generated by a maximal invariant associated with $O_N(M)$. A general result presented in Diaconis et al. (1992) can be used to obtain marginal distributions from uniform distributions over such spheres. It describes how to obtain the distribution of $\mathbf{V} = \mathbf{AUB}'$, where $\mathbf{U}$ is a $N \times N$ random matrix with uniform distribution over $O_N(M)$, and

$\mathbf{A}$ and $\mathbf{B}$ are arbitrary matrices of dimensions $r \times N$ and $s \times N$, respectively, with $\max\{r, s\} < N - m$ and $m = \dim(M)$, the dimension of M. Following Diaconis et al. (1992) we introduce the two non-negative definite matrices, $\mathbf{C} = \mathbf{A}\mathbf{Q}_M\mathbf{A}'$ and $\mathbf{S} = \mathbf{B}\mathbf{Q}_M\mathbf{B}'$, where $\mathbf{Q}_M = \mathbf{I}_N - \mathbf{P}_M$, as well as their unique non-negative definite square roots $\mathbf{C}^{1/2}$ and $\mathbf{D}^{1/2}$. Also, in this section as well as in the ones that follow, $\mathcal{L}(X)$ denotes the "law of X".

Proposition 2.1. (Diaconis et al., 1992) *Let* $\mathbf{U}$ *be the random matrix with uniform distribution on* $O_N(M)$ *and let* $\mathbf{Z}$ *be the* $r \times s$ *upper left corner block of* $\mathbf{U}^*$, *where* $\mathbf{U}^*$ *is a random matrix with uniform distribution on* O_{N-m} *Then, the law of* $\mathbf{V} = \mathbf{AUB}'$ *is such that* $\mathcal{L}(\mathbf{V}) = \mathcal{L}(\mathbf{C}^{1/2}\mathbf{Z}\mathbf{S}^{1/2} + \mathbf{A}\mathbf{P}_M\mathbf{B}')$. *Further, if* $r + s \leq N - m$ *and* $s \leq r$ *then* $\mathbf{Z}$ *has a density function (with respect to the Lebesgue measure) concentrated on the set where* $\mathbf{I}_s - \mathbf{Z}'\mathbf{Z}$ *and* $\mathbf{Z}'\mathbf{Z}$ *are positive definite matrices. Moreover, the density of* $\mathbf{Z}$ *is given by*

$$f(\mathbf{z}|r, s) = (\sqrt{2\pi})^{-rs}\frac{W(N-m-r, s)}{W(N-m, s)}|(\mathbf{I}_s - \mathbf{z}'\mathbf{z})|^{(N-m-r-s-1)/2}, \tag{2.5}$$

where $W(.,.)$ *is the Wishart constant defined by*

$$[W(t, p)]^{-1} = \pi^{\frac{p(p-1)}{4}}2^{tp/2}\prod_{j=1}^{p}\Gamma(\frac{t-j+1}{2}). \tag{2.6}$$

In the density (2.5), p is a positive integer and t is a real number with $t > p - 1$. When $r \leq s$, the density of $\mathbf{Z}$ is obtained by interchanging r and s in the Wishart constant defined in (2.6). Moreover, the notation $|\mathbf{A}|$ is used to denote the determinant of the matrix $\mathbf{A}$.

As a direct consequence of the previous result, Diaconis et al. (1992) show that if $r + s \leq N - m$ and $\mathbf{C}$ and $\mathbf{D}$ are full rank matrices, then $\mathbf{V}$ has a density given by

$$g(\mathbf{v}|r, s) = |\mathbf{C}|^{-s/2}|\mathbf{S}|^{-r/2}f(\mathbf{C}^{-1/2}(\mathbf{v} - \mathbf{A}\mathbf{P}_M\mathbf{B}')\mathbf{B}^{-1/2}|r, s),$$

where $f(.|r, s)$ is as given in (2.5). In the special case of $s = 1$, $r = k$ $(k < N - m)$ it follows from (2.5) that the density of $\mathbf{V}$ is given by

(2.7)
$$f(\mathbf{v}|r, s) = \frac{\Gamma(\frac{n-m}{2})}{\Gamma(\frac{N-m-k}{2})\pi^{k/2}}|r^2\mathbf{C}|^{-1/2}\{1 - \frac{(\mathbf{v} - \mathbf{AE})'\mathbf{C}^{-1}(\mathbf{v} - \mathbf{AE})}{r^2}\}^{\frac{N-m-k-2}{2}},$$

provided $(\mathbf{v} - \mathbf{AE})'\mathbf{C}^{-1}(\mathbf{v} - \mathbf{AE}) \leq r^2$ with $\mathbf{E} = \mathbf{P}_M\mathbf{B}'$ and $r^2 = \mathbf{B}\mathbf{Q}_M\mathbf{B}'$. The distribution of $\mathbf{V}$ in this case is known in the literature as the k-variate Pearson type II distribution with location vector $\boldsymbol{\mu} = \mathbf{P}_M\mathbf{B}'$ and scale matrix $\boldsymbol{\Sigma} = r^2\mathbf{C}$. It follows from (2.1) that $E[\mathbf{V}] = \mathbf{AE}$ and $Var[\mathbf{V}] = (N-m)^{-1}r^2\mathbf{C}$. A formal definition of the k-variate Pearson type II distribution is presented next.

Definition 2.2. *A k-dimensional random vector $\mathbf{Y}$ is said to have a symmetric k-variate Pearson type II distribution with parameters $\boldsymbol{\mu}$, which is N-dimensional, and $\boldsymbol{\Sigma}$, a positive definite $N\times N$ matrix and shape parameter r, $r \geq 1$, if its density is given by*

$$f(\mathbf{y}) = \frac{\Gamma(\frac{k}{2}+r+1)}{\Gamma(r+1)\pi^{k/2}}|\boldsymbol{\Sigma}|^{-1/2}(1-(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}))^r,$$

which we denote by $\mathbf{Y} \sim MPII_k(\boldsymbol{\mu},\boldsymbol{\Sigma},r)$ provided $0 \leq (\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \leq 1$, $r > -1$. Moreover,

$$E[\mathbf{Y}] = \boldsymbol{\mu} \quad \text{and} \quad Var[\mathbf{Y}] = \frac{1}{k+2r+2}\boldsymbol{\Sigma}.$$

As a direct consequence of Proposition 2.1 we have that if the k-dimensional random vector $\mathbf{Y}$ is such that $\mathbf{Y} \sim MPII_k(\boldsymbol{\mu},\boldsymbol{\Sigma},r)$ and $\mathbf{C}$ is a $m\times k$ matrix of known constants, then $\mathbf{W} = \mathbf{CY} \sim MPII_k(\mathbf{C}\boldsymbol{\mu},\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}',r)$.

Detailed discussions of the properties of the Pearson type II distribution can be found in Fang et al. (1990). See also Gasco (1997). In the following we provide representations for projections of probability measures which are O_N-invariant on $\Re^N$. Let $\mathbf{Y}$ be a N-dimensional random vector which is $O_N(M)$-invariant and $\boldsymbol{\Pi}$ is a $n\times N$ projection matrix, with $n < N$. As before, denote by P the law of $\mathbf{Y}$ and by $P^{(n)}$ the law of $\mathbf{Y}_s = \boldsymbol{\Pi}\mathbf{Y}$.

Corollary 2.1. *Let P and $P^{(n)}$ be as defined above. Then, for each $n < N - m$,*

$$P^{(n)} = \int_{M\times[0,\infty)} MPII_k(\boldsymbol{\Pi}\mathbf{c}, v\mathbf{Q}_1, r)\mu_n(d\mathbf{c}, dv), \tag{2.8}$$

where μ_n is the P-law of $\mathbf{t}(\mathbf{Y}) = (\mathbf{P}_M\mathbf{Y}, \mathbf{Y}'\mathbf{Q}_M\mathbf{Y})$, $r = (N-m-n-2)/2$ and $\mathbf{Q}_1 = \boldsymbol{\Pi}\mathbf{Q}_M\boldsymbol{\Pi}'$.

Proof. The result follows directly from representation (2.4) and Proposition 2.1 with $s = 1$ and $r = n$.

We present next an example to illustrate the results derived previously.

Example 2.2. $O_N(M) - $ *invariance*. We consider now the case where M is the column space generated by the columns of the $N\times p$-dimensional matrix $\mathbf{X}$, where p is the rank of $\mathbf{X}$. We assume that the N-dimensional random vector $\mathbf{Y}$ is $O_N(M)$ invariant and consider $\mathbf{Y}_s = \boldsymbol{\Pi}\mathbf{Y}$ and $\mathbf{X}_s = \boldsymbol{\Pi}\mathbf{X}$, with rank p. Thus, it follows that $\boldsymbol{\Pi}\mathbf{P}_M\mathbf{Y} = \mathbf{X}_s\mathbf{B}_N$, $\boldsymbol{\Pi}\mathbf{Q}_M\boldsymbol{\Pi}' = \mathbf{X}_s(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_s'$, where $\mathbf{B}_N = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, as defined in the Introduction. Hence, using Corollary 2.1 it follows for $1 \leq n < N-p$ that

$$P^{(n)} = \int_{\Re^p\times[0,\infty)} MPII_n(\mathbf{X}_s\mathbf{b}, v\mathbf{X}_s(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_s', \frac{N-n-p-2}{2})\mu_N(d\mathbf{b}, dv), \tag{2.9}$$

where μ_N is the P-law of the maximal invariant $t(\mathbf{Y}) = (\mathbf{B}_N, (\mathbf{Y} - \mathbf{X}\mathbf{B}_N)'(\mathbf{Y} - \mathbf{X}\mathbf{B}_N)$, as seen in Example 2.1. The conditional model for $\mathbf{Y}_s$ which is implied by the invariance condition is traditionally represented in the form

$$\mathbf{Y}_s = \mathbf{X}_s\mathbf{b} + \mathbf{e}_s,$$

where $\mathbf{e}_s$ is a n-dimensional random vector with distribution

$$MPII_n(\mathbf{0}, v\mathbf{X}_s(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_s', \frac{N-n-p-2}{2}).$$

Moreover, if $\mathbf{B}_s = (\mathbf{X}_s'\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{Y}_s$, then conditional on $\mathbf{B}_N = \mathbf{b}, (\mathbf{Y} - \mathbf{X}\mathbf{B}_N)'(\mathbf{Y} - \mathbf{X}\mathbf{B}_N) = v$, the distribution of $\mathbf{B}_s$ is

$$MPII_p(\mathbf{b}, v(\mathbf{X}'\mathbf{X})^{-1}, \frac{n-2p-2}{2}),$$

and $(\mathbf{B}_s - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{B}s - \mathbf{b})$ has the same distribution as the random variable vW, where $W \sim Beta(p/1, (n-p)/2)$.

Corollary 2.1 provides an exact representation for marginal distributions of $O_N(M)$-invariant distributions for finite populations. Proximity (in terms of the variation distance) of such representation to a mixture of $O_N(M)$ invariant normal distributions has been studied in Diaconis et al. (1992) in the case $M = \{0\}$ and in general by Eaton (1989). Such results are known in the literature as finite forms of de Finetti type theorems and are extended within a general context in Diaconis et al. (1992). The results considered previously and the multivariate extension of Smith (1981) are actually special cases of those general results.

3 The operational structure for finite populations

A systematic approach for proposing operational parameters is considered in Iglesias (1993), Mendel (1994) and Mendel and Kempthorne (1996) and is used in the formulation that follows.

3.1 Preliminary definitions

Let $(\Re^N, \mathcal{B}_N, \wp)$ the statistical space where $\mathcal{B}_N$ is the Borel *sigma*-field of $\Re^N$ and $\wp$ the class $O_N(M)$-invariant probability measures defined on $(\Re^N, \mathcal{B}_N)$, where M is a subspace of $\Re^N$. Moreover, let $\mathcal{F}_0$ the sub-field of $\mathcal{B}_N$, the measurable sets in $\Re^N$, which are invariant under the action of $O_N(M)$, that is, $A \in \mathcal{F}_0 \Leftrightarrow A \in \mathcal{F}$ and $\mathbf{\Gamma}^{-1}A = A$, $\forall \mathbf{\Gamma} \in O_N(M)$. It follows from the topological and algebraic properties of $O_N(M)$ that the *sigma*-field $\mathcal{F}_0$ is sufficient for the class $\wp$, that is,

$$\text{if } A \in \mathcal{B}_N, \quad P(A|\mathcal{F}_0) \text{ is the same } \forall P \in \wp,$$

which follows from Theorem 3 in Farrell (1966). A formal definition of operational parameter is stated next.

Definition 3.1. *Let* $(\Re^N, \mathcal{B}_N, \wp)$ *as before,* $(\mathcal{Y}, \mathcal{F})$ *a measurable space and* $\theta : \Re^N \to \mathcal{Y}$*, a measurable mapping. Then,* θ *is said to be an operational parameter for* $\wp$ *if* $\mathcal{B}_\theta$*, the sigma-field generated by* θ *is sufficient for the class* $\wp$*, that is,*

$$\forall A \in \mathcal{B}_N, \quad P(A|\mathcal{B}_\theta) \text{ is the same } \forall P \in \wp.$$

Recall that $\mathcal{B}_\theta$, is the sub-field generated by θ, that is, $\mathcal{B}_\theta = \{A \in \mathcal{B}_N | A = \theta^{-1}(B), B \in \mathcal{F}\}$. Thus, if the distribution P of the vector $\mathbf{Y}$ is $O_N(M)$-invariant (that is, $P \in \wp$), then $\theta(\mathbf{Y})$ is an operational parameter provided the sub-field $\mathcal{B}_\theta$ is sufficient for $\wp$. In particular, as seen in Section 2, $t(\mathbf{Y}) = (\mathbf{P}_M\mathbf{Y}, \mathbf{Y}'\mathbf{Q}_M\mathbf{Y})$ is a maximal invariant under the action of $O_N(M)$ and then a measurable mapping from $(\Re^n, \mathcal{B}^N, \wp)$ to $(M \times (0,\infty), \mathcal{B}_t, \wp_t)$, where $\mathcal{B}_t$ is the sub-field of measurable sets in $M \times (0,\infty)$ and $\wp_t$ the class of probability measures Q induced by $t(\mathbf{Y})$. Moreover, being $t(\mathbf{Y})$ a measurable and invariant mapping under the action of $O_N(M)$, it follows that $\mathbf{\Gamma}^{-1}(t^{-1}(A)) = t^{-1}(A), \quad \forall A \in \mathcal{B}_t, \quad \forall \mathbf{\Gamma} \in O_N(M)$. Thus, $t^{-1}(A) \in \mathcal{F}_0$ (following from the definition of $\mathcal{F}_0$), which implies that the sub-field generated by the maximal invariant $t(\mathbf{Y})$ is $\mathcal{F}_0$ and then $t(\mathbf{Y})$ is an operational parameter for the class $\wp$ of $O_N(M)$-invariant distributions on $(\Re^N, \mathcal{B}_N)$. Consequently, as seen in Example 2.5, if M is the subspace spanned be the columns of the matrix $\mathbf{X}$ then $t(\mathbf{Y}) = (\mathbf{B}_N, \sigma_N^2)$ is an operational parameter.

Now, if $P = \mathcal{L}(\mathbf{Y})$ is $O_N(M)$-invariant, then we can write

$$\text{(3.1)}, \qquad \mathcal{L}(\mathbf{Y}) = \int_{M \times (0,\infty)} \nu_{(\mathbf{c},\mathbf{r})} Q(d\mathbf{c}, dr) \quad .$$

where $\nu_{(\mathbf{c},r)}$ is the uniform distribution on the orbit $O_\mathbf{y}$ with $\mathbf{y}$ such that $P_M\mathbf{y} = \mathbf{c}$, $\mathbf{y}'P_M\mathbf{y} = r^2$ and Q the law of $t(\mathbf{Y}) = (\mathbf{P}_M\mathbf{Y}, \mathbf{Y}'Q_M\mathbf{Y})$, which is the operational parameter. Hence, the operational parameter provides an indexation of the uniform distributions $\nu_{t(\mathbf{y})}$, which is involved in the representation of $P = \mathcal{L}(\mathbf{Y})$. Moreover, the uniform distribution is common for representing any $O_N(M)$-invariant distribution P. What makes the representation unique (make models distinguishable) is the mixing measure Q in (3.1), the measure associated with the operational parameter $t(\mathbf{Y}) = (\mathbf{P}_M\mathbf{Y}, \mathbf{Y}'\mathbf{Q}_M\mathbf{Y})$ and induced by the particular distribution $P = \mathcal{L}(\mathbf{Y})$ under consideration. We define now what is understood by an operational statistical model for the observed part $\mathbf{Y_s}$ of $\mathbf{Y}$.

Definition 3.2. *An operational statistical model for* $\mathbf{Y_s}$ *under the class* $\wp$ *with respect to the operational parameter* θ *is given by the statistical space* $(\Re^n, \mathcal{B}_n, \wp_c)$*, where* $\wp_c$ *is the family of conditional distributions of* $\mathbf{Y_s}$ *given* θ*.*

Hence, with respect to the operational parameter $\theta(\mathbf{Y}) = (\theta_1(\mathbf{Y}), \theta_2(\mathbf{Y})) = (\mathbf{P}_M\mathbf{Y}, \mathbf{Y}'\mathbf{Q}_M\mathbf{Y})$, the operational statistical model for the class of the $O_N(M)$-invariant distributions is given by $\wp_c = \{\mathbf{\Pi}\nu_{(\theta_1,\theta_2)}, \ (\theta_1, \theta_2) \in M \times (0,\infty)\}$, where

$\mathbf{\Pi}$ is the projection matrix such that $\mathbf{\Pi Y} = \mathbf{Y}_s$ and $\nu_{(\theta_1,\theta_2)}$ is the uniform distribution on

$$S_N(\theta_1,\theta_2) = \{\mathbf{y} \in \Re^N; P_M\mathbf{y} = \theta_1, \mathbf{y}'Q_M\mathbf{y} = \theta_2\}.$$

Thus, it follows from Corollary 2.1 that the operational statistical model corresponding to the observed part $\mathbf{Y}_s$ of $\mathbf{Y}$ is formed by the Pearson type II distributions with parameters $\left(\mathbf{\Pi}\theta_1, \theta_2 Q_1, \frac{N-m-n-2}{2}\right)$. This, implies the operational likelihood function for (θ_1,θ_2).

Note that the predictive law $P^{(n)} = \mathcal{L}(\mathbf{Y}_s)$ can be represented as

$$\mathcal{L}(\mathbf{Y}_s) = \int_{M\times(0,\infty)} MPII_n\left(\mathbf{\Pi}\theta_1, \theta_2 Q_1, \frac{N-m-n-2}{2}\right) Q(d\theta_1, d\theta_2),$$

where $Q \in \mathcal{Q}$, with $\mathcal{Q}$ being the family of probability measures on $M \times (0,\infty)$ induced by the operational parameter $\theta(\mathbf{y})$. The measure Q corresponds to the prior distribution in the Bayesian operational structure. Here, the parameter is a quantity with a very precise meaning like the maximal invariant under the action of $O_N(M)$. Moreover, as pointed out above, difference among $O_N(M)$-invariant models is captured by the measure Q of the maximal invariant $t(\mathbf{Y})$.

Definition 3.3. *Considering the statistical space $(\Re^N, \mathcal{B}_{\Re^N}, \wp)$, θ the operational parameter for the class $\wp$ and $\wp_\theta$ the probability measures family induced by θ, then $Q \in \wp_\theta$, is the prior distribution for θ.*

Being $\theta(\mathbf{Y}) = (P_M\mathbf{Y}, \mathbf{Y}'Q_M\mathbf{Y})$ the operational parameter and $P = \mathcal{L}(\mathbf{Y})$ then the probability measure $Q \in \wp_\theta$ induced by P on $(M \times (0,\infty), \mathcal{B}_\theta)$, is the prior distribution for the operational parameter $\theta(\mathbf{Y})$. That is, since there is a unique Q associated with each P, we have a linear operator

$$L^* : \wp_\theta \to \wp$$

such that

$$Q \to L^*(Q) = P$$

which maps each distribution $Q \in \wp_\theta$ with a probability distribution $P \in \wp$.

3.2 The operational parameter and the likelihood function

Here, we describe in detail the operational parameter and the likelihood function for the finite population model when the subspace M which describes the group $O_N(M)$ is generated by the columns of the $N \times m$- dimensional matrix $\mathbf{X}$.

- *The operational parameter*

After obtaining the projection of $\mathbf{Y}$ on the space M which is generated by the columns of the matrix $\mathbf{X}$, it follows that (see also Example 2.2) $\theta(\mathbf{Y},\mathbf{X}) =$

$(\mathbf{B}_N, \sigma_N^2)$, where

$$\mathbf{B}_N = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \text{ and } \sigma_N^2 = \frac{\mathbf{Y}'\mathbf{Q}_M\mathbf{Y}}{N-m} = \frac{(\mathbf{Y}-\mathbf{X}\mathbf{B}_N)'(\mathbf{Y}-\mathbf{X}\mathbf{B}_N)}{N-m}$$

is a maximal invariant with respect to the class $\wp$. Moreover, since the sub-field $\mathcal{B}_\theta$ generated by $\theta(\mathbf{Y},\mathbf{X})$ is sufficient for $\wp$ (see Farrell, 1966), it follows, according to Definition 3.1, that $\theta(\mathbf{Y},\mathbf{X})$ is an operational parameter for the class $\wp$.

Note that the operational parameter $\mathbf{B}_N$ corresponds to the (unobserved) least square estimator in the finite population and σ_N^2 to the (unobserved) mean square error.

- *The operational likelihood function*

As defined before, $\mathbf{Y}_s$ is the observed part of the vector $\mathbf{Y}$ and being $n < N-m-1$ and $\mathbf{V}_s = \mathbf{\Pi}\mathbf{Q}_M\mathbf{\Pi}' = \mathbf{I}_n - \mathbf{X}_s(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_s'$, it follows from Corollary 2.1 that the marginal distribution of $\mathbf{Y}_s$ is

$$P^{(s)} = \mathcal{L}(\mathbf{Y}_s) = \int_{\Re^N\times(0,\infty)} MPII_n\left(\mathbf{X}_s\mathbf{B}_N, (N-m)\sigma_N^2\mathbf{V}_s, \frac{N-m-n-2}{2}\right)$$

$$\times Q(d\mathbf{B}_N, d\sigma_N^2),$$

where the support of the n-variate Pearson type II distribution is

$$F = \{\mathbf{v}\in\Re^n \mid (\mathbf{v}-\mathbf{X}_s\mathbf{B}_N)'[(N-m)\sigma_N^2\mathbf{V}_s]^{-1}(\mathbf{v}-\mathbf{X}_s\mathbf{B}_N) \le 1\},$$

with Q being the P-law of $(\mathbf{B}_N, \sigma_N^2)$. Thus, it follows from Definition 3.2, that the corresponding density of $\mathbf{Y}_s$, which is the operational likelihood function, is given by

$$f(\mathbf{y}_s|\mathbf{B}_N,\sigma_N^2) = \frac{\Gamma(\frac{n}{2}+r+1)}{\Gamma(r+1)\pi^{\frac{n}{2}}}[(N-m)\sigma_N^2\mathbf{V}_s]^{-1/2} \tag{3.2}$$

$$\times[1-(\mathbf{y}_s-\mathbf{X}_s\mathbf{B}_N)'[(N-m)\sigma_N^2\mathbf{V}_s]^{-1}(\mathbf{y}_s-\mathbf{X}_s\mathbf{B}_N)]^r I_F(\mathbf{y}_s),$$

with $r = \frac{(N-m-n-2)}{2}$, a Pearson type II density with location vector $\mathbf{X}_s\mathbf{B}_N$ and dispersion matrix $\mathbf{\Sigma} = (N-m)\sigma_N^2\mathbf{V}_s$.

Using the observed sample $\mathbf{Y}_s$, which is of size n, we define

$$\mathbf{B}_s = (\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s \text{ and } s_n^2 = \frac{(\mathbf{y}_s-\mathbf{X}_s\mathbf{B}_s)'\mathbf{V}_s^{-1}(\mathbf{y}_s-\mathbf{X}_s\mathbf{B}_s),}{n-m}$$

and we can write the operational likelihood function (3.2) as

$$f(\mathbf{y}_s|\mathbf{B}_N,\sigma_N^2) \propto \left(\frac{1}{\sigma_N^2}\right)^{\frac{n}{2}}\left[1-\frac{h(\mathbf{B}_N)}{\sigma_N^2}\right]^{\frac{N-m-n-2}{2}} I_{[h(\mathbf{B}_N),\infty)}(\sigma_N^2), \tag{3.3}$$

where

$$h(\mathbf{B}_N) = \frac{1}{N-m}[(n-m)s_n^2 + (\mathbf{B}_N - \mathbf{B}_s)'\mathbf{C}(\mathbf{B}_N - \mathbf{B}_s)],$$

and $\mathbf{C} = \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s$.

In the next section the Bayesian inference is developed for the operational parameters and functions of them.

4 Inference for the operational parameters in finite populations

In finite population sampling, the quantities of interest can be linear functions of $\mathbf{Y}$ such as $\theta_l = \mathbf{c}'\mathbf{Y}$, with $\mathbf{c}$ a vector of dimension N, or quadratic, such as $\theta_Q = \mathbf{Y}'\mathbf{A}\mathbf{Y}$, where $\mathbf{A}$ is a matrix of dimension $N \times N$. Important linear functions are the populational total, $T = \mathbf{1}_N'\mathbf{Y}$, or the finite population regression coefficient $\mathbf{B}_N = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Straightforward algebraic manipulations show that we can write $T = \mathbf{c}'\mathbf{B}_N$ with $\mathbf{c} = (N, N\bar{X}_1, \ldots, N\bar{X}_p)'$, where $\bar{X}_i = \sum_{j=1}^n X_{ij}/n$, is the i-th column mean of the matrix $\mathbf{X}$. By considering a prior density π for $(\mathbf{B}_N, \sigma_N^2)$ we have from (3.2), that the posterior density is given by

(4.1)

$$\pi(\mathbf{B}_N, \sigma_N^2|\mathbf{y_s}) \propto \left(\frac{1}{\sigma_N^2}\right)^{\frac{n}{2}} \left[1 - \frac{h(\mathbf{B}_N)}{\sigma_N^2}\right]^{\frac{N-m-n-2}{2}} I_{[h(\mathbf{B}_N),\infty)}(\sigma_N^2)\pi(\mathbf{B}_N, \sigma_N^2),$$

with $h(\mathbf{B}_N)$ and $\mathbf{C}$ as in (3.2).

To obtain the posterior distribution of $(\mathbf{B}_N, \sigma_N^2)$ which we denote by $\pi(\mathbf{B}_N, \sigma_N^2|\mathbf{y}_s$ the following prior distributions are considered:

(1) $\pi(\mathbf{B}_N, \sigma_N^2) \propto \frac{1}{\sigma_N^2}$,

and

(2) $\pi(\mathbf{B}_N, \sigma_N^2) \propto \frac{a m_0^a}{(\sigma_N^2)^{a+1}} I_{(m_0,\infty)}(\sigma_N^2)$.

That is, in case (2), $\sigma_N^2 \sim$ Pareto(a, m_0), where $a > 0$, $0 < m_0 < h(\mathbf{B}_N)$, with $h(\mathbf{B}_N)$ in (3.3). We call attention to the fact that the above prior distributions may not yield proper $O_N(M)$-invariant probability measures. Thus, from (3.3) and (1) and (2) above, the following (proper) posterior results:

$$\pi(\mathbf{B}_N, \sigma_N^2|\mathbf{y}_s) \propto \left(\frac{1}{\sigma_N^2}\right)^{\frac{n}{2}+b} \left[1 - \frac{h(\mathbf{B}_N)}{\sigma_N^2}\right]^{\frac{N-m-n-2}{2}} I_{[h(\mathbf{B}_N),\infty)}(\sigma_N^2), \tag{4.2}$$

with $b = 1$, in case (1), and $b = a + 1$, in case (2).

The main interest is on predicting the populational total T and the finite population regression coefficient $\mathbf{B}_N$. We consider the squared error loss function which yields the the posterior mean as the predictor of $\mathbf{B}_N$. Credibility intervals are also considered.

• *Marginal posterior densities for* $\mathbf{B}_N$ *and* T

We can compute the marginal density of $\mathbf{B}_N$, by integrating the joint posteriori (4.2) with respect to σ_N^2, yielding

$$\pi(\mathbf{B}_N|\mathbf{y}_s) \propto [(n-m) + (\mathbf{B}_N - \mathbf{B}_s)'\mathbf{X}_s' \frac{\mathbf{V}_s^{-1}}{s_n^2}\mathbf{X}_s(\mathbf{B}_N - \mathbf{B}_s)]^{-\frac{n}{2}+c} \tag{4.3}$$

$$\propto \left[(n-m-2c) + (\mathbf{B}_N - \mathbf{B}_s)'\mathbf{X}_s' \frac{\mathbf{V}_s^{-1}(n-m-2c)\mathbf{X}_s}{s^2(n-m)}(\mathbf{B}_N - \mathbf{B}_s)\right]^{-\frac{n-m-2c+m}{2}},$$

where $c = 0$, in case (1), $c = -a$, in case (2), and $\frac{n-m-2c}{n-m}\mathbf{X}_s(s_n^2\mathbf{V}_s)^{-1}\mathbf{X}_s'$ is positive definite. In (4.3), it can be recognized the kernel of a multivariate Student-t distribution with parameters $(\mathbf{B}_s, \frac{\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{X}_s}{s_n^2}, n-m, m)$, in case (1) and a multivariate truncated Student-t distribution $(\mathbf{B}_s, \frac{\frac{n-m+2a}{n-m}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s}{s_n^2}, n-m, m+2a)$, in case (2). Thus, explicit results can be obtained in the case (1). In fact, being $\frac{n-m-2c}{n-m}\frac{\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s}{s_n^2}$ a positive definite matrix, then there exists the unique mode $\mathbf{B}_N = \mathbf{B}_s$, which is also a posterior mean (when defined) since the posterior distribution is symmetric around $\mathbf{B}_N = \mathbf{B}_s$. Thus, the predictor of $\mathbf{B}_N$ minimizing the risk under the squared error loss is given by

$$\widehat{\mathbf{B}}_N = E[\mathbf{B}_N|\mathbf{y}_s] = \mathbf{B}_s, \tag{4.4}$$

with the corresponding Bayes risk given by

$$Var[\mathbf{B}_N|\mathbf{y}_s] = \frac{n-m}{n-m-2}\left(\frac{\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s}{s_n^2}\right)^{-1} = \frac{n-m}{n-m-2}(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}s_n^2.$$

The study presented above generalizes results in Bolfarine and Zacks (1992) where it is considered Bayes estimation of $\mathbf{B}_N$ under normality. Moreover, Bolfarine et al. (1992) show that the predictor $\mathbf{B}_s$ in (4.4) is the optimal predictor in the class of all linear unbiased predictors of $\mathbf{B}_N$. Thus, the predictor of the population total under squared error loss is given by

$$\widehat{T}_N = \mathbf{c}'\widehat{\mathbf{B}}_N = \mathbf{c}'\mathbf{B}_s, \tag{4.5}$$

with $\mathbf{c} = (N, N\bar{X}_1, \ldots, N\bar{X}_p)'$ an $m \times 1$-dimensional vector, so that $\hat{T}_N = N\bar{Y}_s + (N-n)(\bar{\mathbf{X}}_r - \mathbf{X}_s)'\mathbf{B}_s$. Moreover, the risk function of $\hat{T}_N$ is given by

$$R_{\widehat{T}_N} = \mathbf{c}'Var[\mathbf{B}_N|\mathbf{y}_s]\mathbf{c} = \frac{n-m}{n-m-2}\mathbf{c}'(\mathbf{X}_s\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}s_n^2\mathbf{c}. \tag{4.6}$$

Predictor $\hat{T}_N$ in (4.5) is also the posterior mode of the posterior distribution of T since since $\mathbf{B}_s$ is the mode of the posteriori distribution of $\mathbf{B}_N$. From (4.3), it follows that the distribution of $\mathbf{B}_N$ given $\mathbf{y}_s$ is a multivariate t-Student distribution. Thus, the posterior distribution of T given $\mathbf{y}_s$ that is, the distribution of $\mathbf{c}'\mathbf{B}_N$ given $\mathbf{y}_s$ is a Student-t distribution with $\nu = n - m - 2c$ degrees of freedom, with mean $\widehat{T}_N = \mathbf{c}'\mathbf{B}_s$ and variance $R_{\hat{T}_N}$ given in (4.6). Thus, the credibility interval with symmetric probability $1 - \alpha$ for the populational total T is given by

$$\widehat{T}_N \pm t_{\nu,\frac{\alpha}{2}}\sqrt{R_{\widehat{T}_N}} = \mathbf{c}'\mathbf{B}_s \pm t_{\nu,\frac{\alpha}{2}}\sqrt{R_{\widehat{T}_N}},$$

where $0 < \alpha < 1$, $t_{\nu,\frac{\alpha}{2}} > 0$ such that $P[T^* > t_{\nu,\frac{\alpha}{2}}] = \frac{\alpha}{2}$, with $T^* \sim t_\nu$.

- *Marginal posterior density for σ_N^2*

To make inference on the population variance σ_N^2 we consider first the joint distribution of $(\mathbf{B}_N, \sigma_N^2)$ given $\mathbf{y}_s$, using (4.2) in case (1), which can be written as

$$\pi(\mathbf{B}_N, \sigma_N^2|\mathbf{y}_s) \propto \left(\frac{1}{\sigma_N^2}\right)^{\frac{n}{2}+1}\left[1 - \frac{(n-m)s_n^2}{(N-m)\sigma_N^2}\right]^{\frac{N-m-n}{2}-1} \times[1 - (\mathbf{B}_N - \mathbf{B}_s)'\mathbf{C}_1(\mathbf{B}_N - \mathbf{B}_s)]^{\frac{N-m-n}{2}-1}, \tag{4.7}$$

with $\mathbf{C}_1 = \frac{\mathbf{X}_s'\mathbf{V}_s\mathbf{X}_s}{(N-m)\sigma_N^2-(n-m)s_n^2}$, where the expression on the right hand side of the above expression is found to be the Kernel of a Pearson type II distribution. Thus, integrating (4.7) with respect to $\mathbf{B}_N$, it follows that the posterior distribution of σ_N^2, is given by

$$\pi(\sigma_N^2|\mathbf{y}_s) \propto \left(\frac{(n-m)s_n^2}{(N-m)\sigma_N^2}\right)^{\frac{n-m+4}{2}-1}\left(1 - \frac{(n-m)s_n^2}{(N-m)\sigma_N^2}\right)^{\frac{N-n}{2}-1} I_G, \tag{4.8}$$

with $G = \{\sigma_N^2 > 0 \mid (N-m)\sigma_N^2 > (n-m)s_n^2\}$.

Denoting $Z = \frac{(n-m)s_n^2}{(N-m)\sigma_N^2}$, it follows from (4.8) that

$$Z \sim \mathcal{B}\left(\frac{n-m+4}{2}, \frac{N-n}{2}\right). \tag{4.9}$$

Thus, under the squared error loss function the predictor of σ_N^2 which minimizes the squared error loss is

$$\hat{\sigma}_N^2 = E[\sigma_N^2|\mathbf{y}_s].$$

Moreover, since

$$\sigma_N^2 = \frac{n-m}{N-m}\frac{s_n^2}{Z},$$

it then follows from (4.8) that

$$\hat{\sigma}_N^2 = \frac{n-m}{N-m} s_n^2 E\left[\frac{1}{Z}\right] = \frac{n-m}{N-m} \frac{\Gamma(\frac{N-m-4}{2})\Gamma(\frac{n-m+2}{2})}{\Gamma(\frac{N-m-2}{2})\Gamma(\frac{n-m+4}{2})} s_n^2,$$

with risk function given by

$$Var[\sigma_N^2|\mathbf{y}_s] = \left(\frac{n-m}{N-m} s_n^2\right)^2 \frac{\Gamma(\frac{N-m-4}{2})\Gamma(\frac{n-m}{2})}{\Gamma(\frac{N-m+4}{2})\Gamma(\frac{N-m}{2})}$$

$$- \left[\frac{n-m}{N-m} s_n^2 \frac{\Gamma(\frac{N-m-4}{2})\Gamma(\frac{n-m+2}{2})}{\Gamma(\frac{N-m-2}{2})\Gamma(\frac{n-m+4}{2})}\right]^2$$

$$= \left(\frac{n-m}{N-m} s_n^2\right)^2 \frac{\Gamma(\frac{N-m-4}{2})}{\Gamma(\frac{n-m+4}{2})} \left[\frac{\Gamma(\frac{n-m}{2})}{\Gamma(\frac{N-m}{2})} - \frac{\Gamma(\frac{N-m-4}{2})\Gamma^2(\frac{n-m+2}{2})}{\Gamma(\frac{N-m-2}{2})\Gamma^2(\frac{n-m+4}{2})}\right].$$

A symmetric credibility interval with confidence coefficient $\gamma = 1 - \alpha$ is

$$\left[\frac{(n-m)s_n^2}{(N-m)B_{1-\alpha/2}}, \frac{(n-m)s_n^2}{(N-m)B_{\alpha/2}}\right],$$

with $B_\gamma > 0$ such that $P[Z \leq B_\gamma] = \gamma$, $0 < \gamma < 1$. Similar results can be obtained in case (2). In fact, in this case it can be shown that

$$Z \sim \mathcal{B}\left(\frac{n-m+2a+4}{2}, \frac{N-n}{2}\right).$$

The alternative predictors of the population variance considered in Bolfarine and Zacks (1992) are derived under the normality assumption.

Acknowledgments

The authors thank Dr. Reinaldo Arellano-Valle and Dr. Nelson Tanaka for helpful discussions and suggestions. The authors acknowledge FONDECYT-Chile under grant 1971128 and CNPq-Brasil for partial financial support.

References

BARLOW, R.E. and MENDEL, M.B. (1993). The operational Bayesian approach in reliability theory. *Resenhas*, **1**, 46-56.

BARLOW, R.E. and MENDEL, M.B. (1992). De Finetti-type representation for life distribution. *Journal of the American Statistical Association*, **87**, 1116-1122.

BASU, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyã*, A, **31**, 441–454.

BASU, D. (1971). An essay on the logical foundations of survey sampling. Foundations of Statistical Inference. Edited by V.P. Godambe and D.A. Sprott. Hoet, Rinehart and Winston, Toronto, 203–242.

BASU, D. (1975). Statistical information and likelihood. *Sankhyã*, A, **37**,1–71.

BOLFARINE, H. and ZACKS, S. (1992). *Prediction theory for finite populations. Springer Series in Statistics.* Springer-Verlag.

CASSEL, C.M.; SARNDAL, C.E. and WRETMAN, J.H. (1977). *Foundations of Inference in Sample Surveys.* Wiley, New York.

COCHRAN, W.G. (1973). Sampling Techniques. Wiley, New York..

DIACONIS, P.; EATON, M.L. and LAURITZEN, S.I. (1992). Finite de Finetti theorems in linear models and multivariate analysis. *Scandinavian Journal of Statistics*, **19**, 289–315.

EATON, M. (1989). Group invariance applications in statistics. C.B.M.S. Regional Conference Series in Probability and Statistics, **v.1**, IMS, Hayward, California.

ERICSON, W.A. (1969). Subjective Bayesian models in sampling finite populations. *Journal of the Royal Statistical Society*, **B31**, 195–224.

FANG, K.T., KOTZ, J. and NG, K.W. (1990). *Symmetric multivariate and related distributions.* Chapman and Hall, N.Y.

FARRELL, R.H. (1966). *Multivariate calculation. Use of the continuous groups.* Springer-Verlag, New York.

GASCO, L.B. (1997). *Previsão e preditivismo em modelos lineares com e sem erros nas variáveis.* (In Portuguese.) Doctoral Thesis. Instituto de Matemática e Estatística da Universidade de São Paulo.

GODAMBE, V.P. (1966). A new approach to sampling from finite populations, I, II. *Journal of the Royall Statistical Society*, **B28**, 310-328.

IGLESIAS, P.L. (1993). *Finite forms of de Finetti's theorem: a predictivistic approach to statistical inference in finite populations.* (In Portuguese.) Doctoral Thesis. Instituto de Matemática e Estatística da Universidade de São Paulo.

MENDEL, M.B. (1992). Bayesian parametric models for lifetimes. *Bayesian Statistics*, **4**, 697–705. (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.) Oxford University Press.

MENDEL, M.B. (1994). Operational parameters in Bayesian models. *Test*, **3**, 2, 195–206.

MENDEL, M.B. and KEMPTHORNE, P.J. (1996). Operational parameters for the use of Bayesian methods in engineering. *REBRAPE*, **10**, 1–13.

NACHBIN, L. (1965). *The Haar integral.* Van Nostrand, Princeton, New York.

SMITH, A.M.F. (1981). On random sequences with centered spherical symmetry. *Journal of Royal Statistical Society*, **B43**, 208–305.

Heleno Bolfarine and Loretta B. Gasco
Departamento de Estatística, USP,
Caixa Postal 66281, CEP 05315-970
São Paulo, Capital,
hbolfar@ime.usp.br and lgasco@pucp.edu.pe
Brazil

Pilar L. Iglesias
Departamento de Estadística,
Pontificia Universidad Católica de Chile
Casilla 306, Santiago 22,
pliz@mat.puc.cl
Chile