RT-MAE-9321

MAXIMUM LIKELIHOOD PREDICTION IN FINITE POPULATIONS

by

Josemar Rodrigues

and

Silvia Nagib Elian

Palavras Chaves: Finite Population; Superpopulation Model; Maximum Likelihood Prediction; Predictive Likelihood Function.
(key words)

Classificação AMS: 62D05
(AMS Classification)

- Outubro de 1993 -

Maximum Likelihood Prediction in Finite Populations

Josemar Rodrigues
Instituto de Ciências Matemáticas de São Carlos
Universidade de São Paulo
C.P. 668
13560 — São Carlos — SP — Brasil

Silvia Nagib Elian
Instituto de Matemática e Estatística
Universidade de São Paulo
C.P. 20570
01452-990 — São Paulo — SP — Brasil

Summary

In this paper we use the predictive likelihood approach (Bjornstad, 1990) to get some results concerned to the prediction of $\underline{l}'\underline{y}$ and the population regression coefficient B_N in a finite population under the superpopulation model $\underline{y} \sim N_N(X\underline{\beta}, V)$.

Key Words: Finite Population; Superpopulation Model; Maximum Likelihood Prediction; Predictive Likelihood Function.

1. Introduction

Let us consider a finite population with N units, where N is known. Associated with the i-th unit there are $y_i, x_{i1}, x_{i2}, \dots x_{ip}$, where y_i is the value of the variable Y and $x_{i1}, x_{i2}, \dots x_{ip}$ is an observed set of p variables.

Under the superpopulation models approach, we assume that $\underline{y} = (y_1, \dots y_N)'$ has a N-variate normal distribution with mean $X\underset{\sim}{\beta}$ and covariance matrix V, where

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \dots & x_{Np} \end{pmatrix},$$

$\underset{\sim}{\beta} = (\beta_0, \beta_1 \dots \beta_p)'$ is a vector of unknown parameters

and

V is a known nondiagonal matrix.

In order to get information, a sample of n units is selected and we will reorder the elements of $\underline{y}$, X and V as

$$\underline{y} = \begin{pmatrix} \underline{y}_s \\ \underline{y}_r \end{pmatrix}, \quad X = \begin{pmatrix} X_s \\ X_r \end{pmatrix}, \text{ and } V = \begin{pmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{pmatrix},$$

with $\underline{y}_s$ containing the observed sample elements, $\underline{y}_r$ the unobserved elements, $V_s = \mathrm{Var}(\underline{y}_s)$, $V_r = \mathrm{Var}(\underline{y}_r)$, and $V_{sr} = \mathrm{Cov}(\underline{y}_s, \underline{y}_r)$.

The paper is organized as follows: Section 2 we consider the likelihood prediction of $\underline{y}_r$. Section 3 is devoted to likelihood predicition of $\underline{l}'\underline{y} = \sum_{i=1}^{N} l_i y_i$, $\underline{l}' = (l_1, l_2 \dots l_N)$ a known vector of constants and $B_N = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$, the finite population regression coefficient. Prediction of the population total $T = \sum_{i=1}^{N} y_i$, a particular case of $\underline{l}'\underline{y}$, was considered in Royall(1976a) and Tam(1987). Estimation of B_N has been the subject of papers like Hung(1990), Särndal(1982), Hartley and Sielken(1975) and Fuller(1975).

2. Maximum Likelihood Prediction of $\underset{\sim}{y}_r$

Under the superpopulation model

$$\underset{\sim}{y} \sim N_N(X\underset{\sim}{\beta}, V), \tag{2.1}$$

given $\underset{\sim}{y}_s$, the data vector, we obtain the profile predictive of $\underset{\sim}{y}_r$ given $\underset{\sim}{y}_s$, a particular kind of predictive likelihood function discussed in details by Bjornstad(1990).

It is easy to see from model (2.1) that the joint density function of $(\underset{\sim}{y}_s, \underset{\sim}{y}_r)$ is given by

$$q_{\underset{\sim}{\beta}}(\underset{\sim}{y}_s, \underset{\sim}{y}_r) = \frac{1}{(2\pi)^{N/2} \mid V \mid^{1/2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{y} - X\underset{\sim}{\beta})'V^{-1}(\underset{\sim}{y} - X\underset{\sim}{\beta})\right\}.$$

The profile predictive $L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s)$, which is obtained eliminating the nuisance parameter $\underset{\sim}{\beta}$ by maximization, is

$$L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s) = \sup_{\underset{\sim}{\beta}} q_{\underset{\sim}{\beta}}(\underset{\sim}{y}_s, \underset{\sim}{y}_r) = q_{B_N}(\underset{\sim}{y}_s, \underset{\sim}{y}_r),$$

where $B_N = (X'V^{-1}X)^{-1}X'V^{-1}\underset{\sim}{y}$, that is,

$$L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s) = \frac{1}{(2\pi)^{N/2} \mid V \mid^{1/2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{y} - XB_N)'V^{-1}(\underset{\sim}{y} - XB_N)\right\}.$$

We define the maximum likelihood predictor of $\underset{\sim}{y}_r$ as the value of $\underset{\sim}{y}_r$ which maximizes $L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s)$. In order to obtain our main result, we introduce the following expressions:

$$B_N = H^{-1}BC^{-1}\underset{\sim}{y}_s + H^{-1}DE^{-1}\underset{\sim}{y}_r,$$

where

$$\begin{aligned} &B = X_s' - X_r'V_r^{-1}V_{rs}, \quad C = V_s - V_{sr}V_r^{-1}V_{rs}, \\ &D = X_r' - X_s'V_s^{-1}V_{sr}, \quad E = V_r - V_{rs}V_s^{-1}V_{sr} \\ &\textit{and} \;\; H = BC^{-1}X_s + DE^{-1}X_r = X'V^{-1}X. \end{aligned} \tag{2.2}$$

Noting that C and E are symmetric matrices, and, since V is symmetric,

$$V^{-1} = \begin{pmatrix} C^{-1} & -V_s^{-1}V_{sr}E^{-1} \\ -V_r^{-1}V_{rs}C^{-1} & E^{-1} \end{pmatrix} \tag{2.3}$$

is also symmetric, then

$$V_s^{-1}V_{sr}E^{-1} = (C^{-1})'V_{sr}V_r^{-1} = C^{-1}V_{sr}V_r^{-1}. \tag{2.4}$$

The result of the next lemma, given in Bolfarine et al. (to appear), is used in the proof of Theorem 2.1.

Lemma 2.1 *If* $\hat{\underset{\sim}{\beta}} = (X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}\underset{\sim}{y}_s$ *is the weighted least squares estimator of* $\underset{\sim}{\beta}$ *and* H, B, C, D *and* E *are the matrices defined in (2.2), then*

$$H^{-1}BC^{-1}\underset{\sim}{y}_s + H^{-1}DE^{-1}\left[X_r\hat{\underset{\sim}{\beta}} + V_{rs}V_s^{-1}(\underset{\sim}{y}_s - X_s\hat{\underset{\sim}{\beta}})\right] = \hat{\underset{\sim}{\beta}}. \quad \square$$

Theorem 2.1 *Under the superpopulation model (2.1), the maximum likelihood predictor of* $\underset{\sim}{y}_r$ *is*

$$\hat{\underset{\sim}{y}}_{r,mv} = V_{rs}V_s^{-1}\underset{\sim}{y}_s + D'\hat{\underset{\sim}{\beta}}.$$

Proof:

Maximizing $L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s)$ with respect to $\underset{\sim}{y}_r$ is equivalent to minimize

$$S = (\underset{\sim}{y} - XB_N)'V^{-1}(\underset{\sim}{y} - XB_N) = \underset{\sim}{y}'V^{-1}\underset{\sim}{y} - 2\underset{\sim}{y}'V^{-1}XB_N$$

$$+\underset{\sim}{y}'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}XB_N = \underset{\sim}{y}'V^{-1}\underset{\sim}{y} - \underset{\sim}{y}'V^{-1}XB_N.$$

From (2.2), (2.3) and (2.4) it follows that

$$\underset{\sim}{y}'V^{-1}\underset{\sim}{y} = \underset{\sim}{y}_s'C^{-1}\underset{\sim}{y}_s - 2\underset{\sim}{y}_s'V_s^{-1}V_{sr}E^{-1}\underset{\sim}{y}_r + \underset{\sim}{y}_r'E^{-1}\underset{\sim}{y}_r$$

and

$$y'V^{-1}XB_N = B_N'X'V^{-1}XB_N = y_s'C^{-1}B'H^{-1}BC^{-1}y_s$$

$$+2y_s'C^{-1}B'H^{-1}DE^{-1}y_r + y_r'E^{-1}D'H^{-1}DE^{-1}y_r.$$

Hence,

$$S = y_s'C^{-1}y_s - 2y_s'V_s^{-1}V_{sr}E^{-1}y_r + y_r'E^{-1}y_r - y_s'C^{-1}B'H^{-1}BC^{-1}y_s$$

$$-2y_s'C^{-1}B'H^{-1}DE^{-1}y_r - y_r'E^{-1}D'H^{-1}DE^{-1}y_r$$

and

$$\begin{aligned}\frac{\delta S}{\delta y_r} &= -2V_r^{-1}V_{rs}C^{-1}y_s - 2E^{-1}D'H^{-1}BC^{-1}y_s \\ &\quad +2E^{-1}y_r - 2E^{-1}D'H^{-1}DE^{-1}y_r.\end{aligned}$$

Taking $\frac{\delta S}{\delta y_r} = 0$, we find that $\hat{y}_{r,mv}$, the maximum likelihood predictor of y_r, satisfies the condition

$$\left[V_r^{-1}V_{rs}C^{-1} + E^{-1}D'H^{-1}BC^{-1}\right] y_s = \left[E^{-1} - E^{-1}D'H^{-1}DE^{-1}\right] \hat{y}_{r,mv}. \tag{2.5}$$

By (2.4), this condition turns

$$E^{-1}V_{rs}V_s^{-1}y_s + E^{-1}D'\left[H^{-1}BC^{-1}y_s + H^{-1}DE^{-1}\hat{y}_{r,mv}\right] = E^{-1}\hat{y}_{r,mv}$$

or

$$V_{rs}V_s^{-1}y_s + D'\left[H^{-1}BC^{-1}y_s + H^{-1}DE^{-1}\hat{y}_{r,mv}\right] = \hat{y}_{r,mv}.$$

Using the result of Lemma 2.1, we note that

$$\hat{y}_{r,mv} = V_{rs}V_s^{-1}y_s + D'\hat{\beta}$$

satisfies this equation, and so, is one maximum likelihood predictor of y_r.

Further, we will compute $[E^{-1} - E^{-1}D'H^{-1}DE^{-1}]^{-1}$, and hence we note that the predictor is unique. □

Example

Let us consider the model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \varepsilon$$

with

$$X = 1_N, \quad V = (1-\rho)I_N + \rho 1_N 1'_N = (1-\rho)I_N + \rho J_N$$

$$\textit{and} \;\; \underset{\sim}{y} \sim N_N(X\underset{\sim}{\beta}, V),$$

where 1_N is a vector of one matrix of dimension N, I_N is the identity matrix of order N and J_N is a $N \times N$ matrix of ones. Further, we will also consider that $J_{N-n,n}$ is the matrix $(N-n) \times n$ of ones.

In this model, it's easy to see that

$$V_s = (1-\rho)I_N + \rho J_N$$

which implies that

$$V_s^{-1} = \frac{1}{1-\rho}\left[I_N - \frac{\rho}{1+(n-1)\rho}J_N\right], \qquad \hat{\underset{\sim}{\beta}} = \overline{y}_s = \frac{\sum_{i=1}^{n} y_i}{n},$$

$$V_{rs} = \rho J_{N-n,n} \;\; \textit{and} \;\; D = \frac{1-\rho}{1+(n-1)\rho}1'_{N-n}.$$

Also,

$$V_{rs}V_s^{-1}y_s = \frac{\rho}{1+(n-1)\rho}1_{N-n}n\overline{y}_s,$$

$$D'\hat{\underset{\sim}{\beta}} = \frac{1-\rho}{1+(n-1)\rho}1_{N-n}\overline{y}_s$$

and thus,

$$\hat{y}_{r,mu} = 1_{N-n}\overline{y}_s. \; \square$$

In the next theorem, we show that $L_p(y_r \mid y_s)$ can be normalized resulting in a function proportional to a normal density. Since two predictive likelihoods are equivalent if they are proportional to each other, no loss of information is incorred by this normalization.

We say that $L_p(y_2 \mid y_1) \; \alpha \; N_m(\mu, \Sigma)$ if

$$L_p(y_2 \mid y_1) = k(y_1) h_m(y_2, \mu, \Sigma),$$

where $h_m(y_2, \mu, \Sigma)$ is the m-variate normal density with mean μ and covariance matrix Σ and $k(y_1)$ does not depend on y_2.

Theorem 2.2 *Under the superpopulation model (2.1),*

$$L_p(y_r \mid y_s) \; \alpha \; N_{N-n}(\mu, W^{-1})$$

where

$$W = E^{-1} - E^{-1}D'H^{-1}DE^{-1} \quad and \quad \mu = \hat{y}_{r,mv} = V_{rs}V_s^{-1}y_s + D'\hat{\beta}.$$

Proof:

$$L_p(y_r \mid y_s) = \frac{1}{(2\pi)^{N/2} \mid V \mid^{1/2}} \exp\left\{-\frac{1}{2}S\right\},$$

where

$$S = (y - XB_N)'V^{-1}(y - XB_N) = y_s'\left[C^{-1} - C^{-1}B'H^{-1}BC^{-1}\right]y_s$$

$$-2y_r'\left[E^{-1}V_{rs}V_s^{-1} + E^{-1}D'H^{-1}BC^{-1}\right]y_s +$$

$$+y_r'\left[E^{-1} - E^{-1}D'H^{-1}DE^{-1}\right]y_r.$$

From (2.4),

$$S = y_s'\left[C^{-1} - C^{-1}B'H^{-1}BC^{-1}\right]y_s -$$

$$-2y_r' \left[V_r^{-1}V_{rs}C^{-1} + E^{-1}D'H^{-1}BC^{-1}\right] y_s +$$

$$+ y_r' W y_r.$$

By (2.5), it follows that

$$S = y_s' \left[C^{-1} - C^{-1}B'H^{-1}BC^{-1}\right] y_s - 2y_r' W \mu + y_r' W y_r$$

$$= (y_r - \mu)' W (y_r - \mu) - \mu' W \mu + y_s' \left[C^{-1} - C^{-1}B'H^{-1}BC^{-1}\right] y_s.$$

We note that $S_1 = y_s' [C^{-1} - C^{-1}B'H^{-1}BC^{-1}] y_s - \mu' W \mu$ does not depend on y_r, so

$$L_p(y_r \mid y_s) = \frac{1}{(2\pi)^{N/2} \mid V \mid^{1/2}} \exp\left\{-\frac{1}{2}S_1\right\} \exp\left\{-\frac{1}{2}(y_r - \mu)' W (y_r - \mu)\right\}$$

$$\alpha \; N_{N-n}(\mu, W^{-1}). \quad \square$$

Since the maximum likelihood predictor of y_r is the mode of $L_p(y_r \mid y_s)$, it follows from the Theorem 2.2 that $\hat{y}_{r,mv} = \mu = V_{rs}V_s^{-1}y_s + D'\hat{\beta}$, the same result we got in Theorem 2.1.

3. The Maximum Likelihood Predictor of $l'y$ and B_N

Some other quantities besides y_r has been considered in the literature. Under the model (2.1), Tam(1987) derives the optimal predictor of the population total, $T = 1'_N y = \sum_{i=1}^{N} y_i$, where $1'_N$ is a $N \times 1$ vector of ones. It follows directly by his result that the optimal predictor of $l'y$, any linear combination of $y_1, y_2, \ldots y_N$, is

$$\hat{T}^* = l_s' y_s + l_r' \left[X_r\hat{\beta} + V_{rs}V_s^{-1}(y_s - X_s\hat{\beta})\right],$$

where l_s' and l_r' are vectors $1 \times n$ and $1 \times (N-n)$ such that $l' = [l_s' \;\; l_r']$.

Also under this model, Bolfarine et al. (to appear) show that the optimal predictor of the finite population regression coefficient

$$B_N = (X'V^{-1}X)^{-1}X'V^{-1}\underset{\sim}{y}$$

is

$$\hat{\underset{\sim}{\beta}} = (X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}\underset{\sim}{y}_s.$$

In this section we find the predictive likelihood function of $l'\underset{\sim}{y}$ and B_N and the maximum likelihood predictor of these quantities. Royall(1976b) works with another kind of likelihood function to predict $l'\underset{\sim}{y}$ under the superpopulation model (2.1) with known diagonal covariance matrix V.

Theorem 3.1 *Under the superpopulation model $\underset{\sim}{y} \sim N_N(X\underset{\sim}{\beta}, V)$,*

$$L_p(l'\underset{\sim}{y}/\underset{\sim}{y}_s) \ \alpha \ N_1(\hat{T}^*, Var(\hat{T}^* - l'\underset{\sim}{y}))$$

where

$\hat{T}^* = l_s'\underset{\sim}{y}_s + l_r'\underset{\sim}{\mu} = l_s'\underset{\sim}{y}_s + l_r'\left[V_{rs}V_s^{-1}\underset{\sim}{y}_s + D'\hat{\underset{\sim}{\beta}}\right]$ *is the optimal predictor of* $l'\underset{\sim}{y}$;

$Var(\hat{T}^* - l'\underset{\sim}{y}) = l_r'W^{-1}l_r$ *and*

$W = E^{-1} - E^{-1}D'H^{-1}DE^{-1}.$

Proof:

Since $L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s) \ \alpha \ N_{N-n}(\underset{\sim}{\mu}, W^{-1})$ and $l'\underset{\sim}{y} = l_s'\underset{\sim}{y}_s + l_r'\underset{\sim}{y}_r$, it follows directly by properties of conditional distributions and the normal multivariate distribution that

$$L_p(l'\underset{\sim}{y} \mid \underset{\sim}{y}_s) \ \alpha \ N_1(l_s'\underset{\sim}{y}_s + l_r'\underset{\sim}{\mu}, l_r'W^{-1}l_r).$$

After some algebraic manipulations we note that

$$\begin{aligned} MSE(\hat{T}^*) &= E(\hat{T}^* - l'\underset{\sim}{y})^2 = Var(\hat{T}^* - l'\underset{\sim}{y}) = l_r'(V_r - V_{rs}V_s^{-1}V_{sr})l_r \\ &\quad + l_r'(X_r - V_{rs}V_s^{-1}X_s)(X_s'V_s^{-1}X_s)^{-1}(X_r' - X_s'V_s^{-1}V_{sr})l_r = \\ &= l_r'El_r + l_r'D'(X_s'V_s^{-1}X_s)^{-1}Dl_r. \end{aligned}$$

According to this result, to prove that $Var(\hat{T}^* - l'y) = l_r'W^{-1}l_r$, it's enough showing that

$$W^{-1} = E + D'(X_s'V_s^{-1}X_s)^{-1}D.$$

From inverse matrices properties,

$$W^{-1} = (E^{-1} - E^{-1}D'H^{-1}DE^{-1})^{-1} =$$
$$= E - EE^{-1}D'(DE^{-1}EE^{-1}D' - H)^{-1}DE^{-1}E$$
$$= E + D'(H - DE^{-1}D')^{-1}D,$$

$$E^{-1} = (V_r - V_{rs}V_s^{-1}V_{sr})^{-1} = V_r^{-1} - V_r^{-1}V_{rs}(V_{sr}V_r^{-1}V_{rs} - V_s)^{-1}V_{sr}V_r^{-1}$$
$$= V_r^{-1} + V_r^{-1}V_{rs}C^{-1}V_{sr}V_r^{-1} = V_r^{-1} + V_r^{-1}V_{rs}V_s^{-1}V_{sr}E^{-1}$$

$$C^{-1} = (V_s - V_{sr}V_r^{-1}V_{rs})^{-1} = V_s^{-1} - V_s^{-1}V_{sr}(V_{rs}V_s^{-1}V_{sr} - V_r)^{-1}V_{rs}V_s^{-1}$$
$$= V_s^{-1} + V_s^{-1}V_{sr}E^{-1}V_{rs}V_s^{-1}$$

and so,

$$H - DE^{-1}D' = BC^{-1}X_s + DE^{-1}X_r - DE^{-1}(X_r - V_{rs}V_s^{-1}X_s) =$$
$$= BC^{-1}X_s + DE^{-1}V_{rs}V_s^{-1}X_s =$$
$$= (X_s' - X_r'V_r^{-1}V_{rs})(V_s^{-1} + V_s^{-1}V_{sr}E^{-1}V_{rs}V_s^{-1})X_s + DE^{-1}V_{rs}V_s^{-1}X_s =$$
$$= X_s'V_s^{-1}X_s +$$
$$+(X_s'V_s^{-1}V_{sr} - X_r'V_r^{-1}V_{rs}V_s^{-1}V_{sr} + D)E^{-1}V_{rs}V_s^{-1}X_s - X_r'V_r^{-1}V_{rs}V_s^{-1}X_s.$$

Since $D = X_r' - X_s'V_s^{-1}V_{sr}$ and $V_r^{-1} = E^{-1} - V_r^{-1}V_{rs}V_s^{-1}V_{sr}E^{-1}$,

$$H - DE^{-1}D' = X_s'V_s^{-1}X_s + X_r'(E^{-1} - V_r^{-1}V_{rs}V_s^{-1}V_{sr}E^{-1})V_{rs}V_s^{-1}X_s -$$
$$-X_r'V_r^{-1}V_{rs}V_s^{-1}X_s$$
$$= X_s'V_s^{-1}X_s + X_r'V_r^{-1}V_{rs}V_s^{-1}X_s - X_r'V_r^{-1}V_{rs}V_s^{-1}X_s$$
$$= X_s'V_s^{-1}X_s,$$

which implies that

$$W^{-1} = E + D'(X_s'V_s^{-1}X_s)^{-1}D. \quad \square$$

The next lemma is proved in Bolfarine et al. (to appear) and will be useful in the proof of the Theorem 3.2.

Lemma 3.1 *Under the superpopulation model (2.1), the optimal predictor of B_N is*

$$\hat{\underset{\sim}{\theta}}^* = H^{-1}BC^{-1}\underset{\sim}{y}_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}\underset{\sim}{y}_s + D'\hat{\underset{\sim}{\beta}}],$$

where $\hat{\underset{\sim}{\beta}}$ is the weighted least squares estimator.

Furthermore,

$$\hat{\underset{\sim}{\theta}}^* = \hat{\underset{\sim}{\beta}}$$

and

$$Var(\hat{\underset{\sim}{\beta}} - B_N) = H^{-1}DE^{-1}[E + D'(X_s'V_s^{-1}X_s)^{-1}D]E^{-1}D'H^{-1}$$

$$= (X_s'V_s^{-1}X_s)^{-1} - (X'V^{-1}X)^{-1}.\square$$

Theorem 3.2 *Under the superpopulation model (2.1), the predictive likelihood function of B_N is given by*

$$L_p(B_N \mid \underset{\sim}{y}_s) \;\alpha\; N_{p+1}(\hat{\underset{\sim}{\beta}}, Var(\hat{\underset{\sim}{\beta}} - B_N))$$

where

$$Var(\hat{\underset{\sim}{\beta}} - B_N) = H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1}$$

and

$$W = E^{-1} - E^{-1}D'H^{-1}DE^{-1}.$$

Proof:

From Theorem 2.2,

$$L_p(\underset{\sim}{y}_r \mid \underset{\sim}{y}_s) \;\alpha\; N_{N-n}(\underset{\sim}{\mu}, W^{-1})$$

and since $B_N = H^{-1}BC^{-1}\underset{\sim}{y}_s + H^{-1}DE^{-1}\underset{\sim}{y}_r$, then

$$L_p(B_N \mid y_s) \propto N_{p+1}(H^{-1}BC^{-1}y_s + H^{-1}DE^{-1}\mu, H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1}).$$

Also, by Lemma 3.1,

$$H^{-1}BC^{-1}y_s + H^{-1}DE^{-1}\mu = \hat{\beta}.$$

To prove that $Var(\hat{\beta} - B_N) = H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1}$, it is enough to note that

$$Var(\hat{\beta} - B_N) = H^{-1}DE^{-1}\left[E + D'(X_s'V_s^{-1}X_s)^{-1}D\right]E^{-1}D'H^{-1},$$

from Lemma 3.1, and

$$W^{-1} = E + D'(X_s'V_s^{-1}X_s)^{-1}D,$$

from the proof of Theorem 3.1. □

An important consequence of Theorem 3.2 is that the weighted least squares predictor $\hat{\beta}$ is the maximum likelihood predictor of B_N. It follows from Theorem 3.1 that the maximum likelihood predictor of $l'y$ is $\hat{T}^*$, that is also the optimal predictor of $l'y$ under the considered model.

References

Bjornstad, J. F. (1990). *Predictive likelihood: a review.* Statistical Science, 5, 242-265.

Bolfarine, H., Elian, S. N., Rodrigues, J. and Zacks, S. (to appear). *Optimal prediction of the finite population regression coefficient.* Sankhya, Series B.

Fuller, W. A. (1975). *Regression analysis for sample survey.* Sankhya, Series C, 37, 117-132.

Hartley, H. O. and Sielken, R. L. Jr. (1975). *A "superpopulation viewpoint" for finite population sampling.* Biometrics, 31, 411-422.

Hung, H. M. (1990). *Nonlinear regression analysis for complex surveys.* Communication in Statistics, Theory and Methods, 19, 3447-3468.

Royall, R. M. (1976a). *The linear least-squares prediction approach to two-stage sampling.* Journal of the American Statistical Association, 71, 657-664.

Royall, R. M. (1976b). *Likelihood functions in finite population sampling theory.* Biometrika, 63, 605-614.

Sarndal, C. E. (1982). *Implications of survey design for generalized regression estimation of linear functions.* Journal of Statistical Planning and Inference, 7, 155-170.

Tam, S. M. (1987). *Optimality of Royall's predictor under a Gaussian superpopulation model.* Biometrika, 7, 659-660.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1993

9301 - BUENO, V.C. & ARIZONO, H. Comparisons for Maintenance Policies Involving Complete and Minimal Repair Through Compensator Transform, 12p.

9302 - BUENO, V.C. Maintenance Comparisons Through Compensator Transform: Block Policies, 13p.

9303 - FERRARI, P.A. & FONTES, L.R.G. Shock Fluctuations in the Asymmetric Simple Exclusion Process, 17p.

9304 - CORDEIRO, G. M., BOTTER, D.A. & FERRARI, S.L.P. Nonnull Asymptotic Distributions of Three Classic Criteria in Generalized Linear Models, 21p.

9305 - TANAKA, N.I. & TAVARES, H.R. Introdução à Teoria da Percolação, 38p.

9306 - BOLFARINE, H., ZACKS, S. & SANDOVAL, M.C. On Predicting the Population Total Under Regression Models with Measurement Erros, 10p.

9307 - FERRARI, P.A. & FONTES, L.R.G. The Net Output Process of a System with Infinitely Many Queues, 20p.

9308 - HO, L.L. & BUSSAB, W.O. Decomposição de Dados Categóricos com Auxílio de Árvores Dicotômicas, 23p.

9309 - BOLFARINE, H. & VALLE, R.B.A. Robust Modelling in Measurement Error Models Using the t Distribution, 14p.

9310 - FERRARI, S.L. de P. & ARELLANO, R.B. Bartlett Corrected Test for Regression Models with Student-t Independent Errors, 30p.

9311 - PAULINO, C.D.M. & PEREIRA, C.A. de B. Bayesian Methods for Categorical Data Under Informative General Censoring, 13p.

9312 - INOUE, L.Y.T., PASCHOA, O.E.D. & WECHSLER, S. Bayesian Assessment of Bioequivalence, 9p.

9313 - MADRUGA, M.R., PEREIRA, C.A. de B. & RABELLO-GAY, M.N. Bayesian Dosimetry: Radiation Dose Versus Frequencies of Cells with Aberrations, 15p.

9314 - MENTZ, R.P., MORETTIN, P.A. & TOLOI, C.M.C. On Residual Variance Estimation in Arma Models, 47p.

9315 - PAULINO, C.D.M. & PEREIRA, C.A. de B. On Identifiability of Parametric Statistical Models, 35p.

9316 - BARLOW, R. E. & PEREIRA, C.A. de B. The Bayesian Operation and Probabilistic Influence Diagrams, 50p.

9317 - PAULA, G.A. & CORDEIRO, G.M. Bias Correction and Improved Residuals for Non-Exponential Family Nonlinear Models, 18p.

9318 - IGLESIAS, P.L.Z., PEREIRA, C.A. de B. & TANAKA, N.I. Finite Forms of de Finetti's Type Theorem for Univariate Uniform Distributions, 25p.

9319 - BOLFARINE, H., NASCIMENTO, J.A. & ROJAS, M.G. Maximum Likelihood Estimation of Linear Structural Relationships Via the EM Algorithm, 13p.

9320 - BOLFARINE, H. & ROJAS, M.G. Structural Comparative Calibration Using the EM Algorithm, 17p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
IME-USP
Caixa Postal 20.570
01452-990 - São Paulo, Brasil