RT-MAE-9212

ON PREDICTING THE FINITE POPULATION DISTRIBUTION FUNCTION

by

Heleno Bolfarine

and

Mônica C. Sandoval

Palavras Chaves: efficiency, optimal predictor, superpopulation
(Key words) model, totally sufficient statistics.

Classificação AMS: 62D05
(AMS Classification)

- Agosto de 1992 -

ON PREDICTING THE FINITE POPULATION DISTRIBUTION FUNCTION

Heleno Bolfarine and Mônica C. Sandoval
Departamento de Estatística - Universidade de São Paulo
Caixa Postal 20570 - Agência Iguatemi
São Paulo - SP - BRASIL

Abstract

In this article, we consider the optimal prediction of the finite population distribution function under general linear regression models with normally distributed errors. Emphasis is placed on the case where the error variance is unknown. Large sample approximations to the prediction variance of the optimal predictors are also derived.

1. Introduction

Consider a finite population $\mathcal{P} = \{1, \ldots, N\}$, where N is known. Associated with unit k of $\mathcal{P}$, there are $p+1$ quantities $y_k, x_{k1}, \ldots, x_{kp}$, which are all known, except for y_k, $k = 1, \ldots, N$. The quantity y_k is considered to be a realization of a random variable Y_k, $k = 1, \ldots, N$. But, since both are unknown, it is not distinguished between them. Let $\mathbf{y} = (y_1, \ldots, y_N)'$ and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$ a known Nxp matrix where $\mathbf{X}_j' = (x_{j1}, \ldots, x_{jp})$, $j = 1, \ldots, N$. Relating the two sets of variables $\mathbf{y}$ and $\mathbf{X}$, we consider the linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \tag{1}$$

where $E[\mathbf{e}] = 0$ and $Var[\mathbf{e}] = \sigma^2\mathbf{W}$, being $\mathbf{W}$ a known and diagonal matrix. Let $\psi = (\beta, \sigma^2)$. Notice that after a sample s of size n has been selected, we may rewrite $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{W}$ so that

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{W}_s & 0 \\ 0 & \mathbf{W}_r \end{pmatrix},$$

where $\mathbf{y}_s$ $(\mathbf{y}_r)$ corresponds to the (un)observed elements of $\mathbf{y}$ and so on. The finite population distribution function is defined by

$$F_N(t) = \frac{1}{N}\sum_{j=1}^{N} \Delta(t - y_j),$$

where

$$\Delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Prediction of $F_N(t)$ has been the subject of several papers in the literature. See, for example, Sedransk and Sedransk (1979), Chambers and Dunstan (1986) and more recently Rao et al. (1990). Chambers and Dustan (1986) sugested a nonparametric predictor for $F_N(t)$ under the assumption that $\mathbf{X}' = (x_1, \ldots, x_N)$ and $\mathbf{W} = diag(v(x_1), \ldots, v(x_N))$. Rao et al. (1990) suggested ratio and difference estimators within the classical sampling approach, which incorporates information from the auxiliary vector $\mathbf{X}' = (x_1, \ldots, x_n)$. In the present paper, we derive the optimal (best unbiased) predictor of $F_N(t)$, under model (1), with the additional assumption that the vector $\mathbf{e}$ is normally distributed. Optimal prediction of population quantities under Gaussian superpopulation models has been the subject of several recent papers in the literature. See, for example, Tam (1987) and Zacks and Bolfarine (1991).

The optimal predictors (denoted by BUP in the sequel) of $F_N(t)$ are obtained by using a theorem which appears in Rodrigues et al. (1985) and some results about the estimation of the distribution function in infinite populations (see, for example, Olkin and Ghurye, 1969).

Section 2 considers some general results about the prediction of $F_N(t)$ under model (1) with σ^2 unknown. Section 3 is devoted to the prediction of $F_N(t)$ under the simple location model. The best unbiased predictor of $F_N(t)$ and its asymptotic distribution are derived. We also derive an expression for the asymptotic relative efficiency of the best unbiased predictor relative to the ordinary sample empirical distribution function, its natural competitor under the simple location model. Situations where $\mathbf{W} = diag(v_1, \ldots, v_N)$ with $v_i > 0$, $i = 1, \ldots, n$ are studied in Section 4. Large sample approximations are derived for the prediction variances of the BUP for some important special models.

2. Optimal Predictors of $F_N(t)$

As considered in Rodrigues et al. (1985), after the sample s has been selected, we may write $\theta(\mathbf{y})$ as

$$\theta(\mathbf{y}) = F_N(t) = \theta_s + \theta_{sr},$$

where

$$\theta_s = \theta(\mathbf{y}_s) = \frac{n}{N} \sum_{i \in s} \frac{\Delta(t - y_i)}{n},$$

and

$$\theta_{sr} = \theta(\mathbf{y}_s, \mathbf{y}_r) = \frac{(N-n)}{N} \sum_{i \notin s} \frac{\Delta(t - y_i)}{N - n}.$$

Predicting $F_N(t)$ is therefore equivalent to predicting θ_{sr}. Hence, predictors of $F_N(t)$ which are considered in this paper are of the form

$$\hat{\theta}_s = \hat{F}_N(t) = \theta_s + \hat{\theta}_{sr},$$

where $\hat{\theta}_{sr} = \hat{\theta}_{sr}(\mathbf{y}_s)$ is a predictor of θ_{sr}.

As considered in Rodrigues et al. (1985), a predictor $\hat{F}_N(t)$ of $F_N(t)$ is unbiased with respect to the superpopulation model (1) if

$$E_\psi[\hat{F}_N(t) - F_N(t)] = 0,$$

for all $\psi = (\beta, \sigma)$. Let $S = S(\mathbf{y}_s)$ be a complete and totally sufficient statistic for the family $\mathcal{F} = \{F_\psi, \psi \in \Psi\}$, where Ψ is the parameter space as defined, for example, in Rodrigues et al. (1985). The following is a consequence of a more general result which appears in Rodrigues et al. (1985).

Theorem 2.1. *Let $\hat{F}_N(t)$ be any unbiased predictor of $F_N(t)$. Then, under model (1), the BUP of $F_N(t)$ is*

$$\hat{F}_{BU}(t) = \theta_s + \hat{\theta}_{sr}(S),$$

where

$$\hat{\theta}_{sr}(S) = E_\psi[\hat{\theta}_{sr}(\mathbf{y}_s)|S].$$

Since $\mathbf{W}$ is considered to be diagonal, if S is sufficient then it is also totally sufficient. Therefore, under model (1) with normally distributed errors, the complete and sufficient statistic

$$(\hat{\beta}_s, S_y^2) \tag{2}$$

where

$$\hat{\beta}_s = (\mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{y}_s$$

and

$$S_y^2 = (\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)'\mathbf{W}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)/(n-p),$$

also is a complete and totally sufficient statistic. The predictors derived in the next sections are to be based on the statistic $S = (\hat{\beta}_s, S_y^2)$.

3. The Simple Location Model

The simple location model follows from (1) by making

$$\mathbf{X} = \mathbf{1}_N \text{ and } \mathbf{W} = \mathbf{I}_N, \tag{3}$$

where $\mathbf{1}_N$ is a vector of ones of dimension N and $\mathbf{I}_N$ is the N–dimensional identity matrix. In this case, β is an unknown location parameter. According to (2), the complete and totally suficient statistics is then $(\hat{\beta}_s, S_y^2)$, where

$$\hat{\beta}_s = \bar{y}_s, \quad \text{and} \quad S_y^2 = \sum_{i\in s}(y_i - \bar{y}_s)^2/(n-1) \tag{4}$$

and the $\hat{F}_{BU}(t)$ is to be based on it. In the sequel, we denote by

$$F_{Ns}(t) = \frac{1}{n}\sum_{i\in s}\Delta(t - y_i) \tag{5}$$

the ordinary sample empirical distribution function.

Theorem 3.1. *Under the superpopulation model specified by (1) and (3) with errors normally distributed,*

$$\hat{F}_{BU}(t) = \frac{n}{N}F_{Ns}(t) + (1 - \frac{n}{N})T_{n-2}(U_n(t)), \tag{6}$$

where

$$U_n(t) = \frac{\sqrt{n-2}(t-\bar{y}_s)}{\sqrt{\frac{(n-1)^2}{n}S_y^2 - (t-\bar{y}_s)^2}},$$

and $T_{n-2}(.)$ is the distribution function of the Student's t distribution with $n-2$ degrees of freedom.

Proof. According to Olkin and Ghurye (1969), it follows that

$$E_\psi[T_{n-2}(U_n(t))] = E_\psi[\Delta(t-y_i)] = \phi(\frac{t-\beta}{\sigma}),$$

which proves that $\hat{F}_{BU}(t)$ is unbiased. $\phi(.)$ denotes the density of the standard normal density. Moreover, since $\hat{F}_{BU}(t)$ is a function of the totally sufficient and complete statistics $(\hat{\beta}_s, S_y^2)$, the proof is complete.

In the next theorem, the asymptotic distribution of the BUP predictor is considered. We assume that our finite population is embbeded in an infinite sequence of increasing finite populations.

Theorem 3.2. *Consider a sequence of finite populations $\{\mathcal{P}_\nu, \nu \geq 1\}$, satisfying $N_\nu \leq N_{\nu+1}$. From population $\mathcal{P}_\nu$ a sample of size n_ν $(n_\nu < n_{\nu+1})$ is selected in such a way that*

$$N_\nu - n_\nu \to \infty.$$

Moreover, we consider that

$$f_\nu = \frac{n_\nu}{N_\nu} \to f,$$

where $0 < f < 1$. Then,

$$\sqrt{n_\nu}(\hat{F}_{BU_\nu}(t) - F_{N_\nu}(t)) \xrightarrow{D} N(0, (1-f)^2[(1 + \frac{(t-\beta)^2}{2\sigma^2})\phi^2(\frac{t-\beta}{\sigma})$$

(7)
$$+\frac{f}{1-f}\Phi(\frac{t-\beta}{\sigma})(1-\Phi(\frac{t-\beta}{\sigma}))])$$

where " $\xrightarrow{D}$ " means convergence in distribution and $\Phi(.)$ is the distribution function of the standard normal distribution.

Proof. To simplify notation, we drop the subscript ν. We may write

(8)
$$\sqrt{n}(\hat{F}_{BU}(t) - F_N(t)) = (1 - \frac{n}{N})\{\sqrt{n}[T_{n-2}(U_n(t)) - \Phi(\frac{t-\beta}{\sigma})]$$

$$-\sqrt{\frac{n}{N-n}}\sqrt{N-n}[\sum_{j\in r}\frac{\Delta(t-y_j)}{N-n} - \Phi(\frac{t-\beta}{\sigma})]\}.$$

Using the Central Limit Theorem for iid random variables (Serfling, 1980). it follows that

$$(9) \qquad \sqrt{N-n}[\sum_{j\in r}\frac{\Delta(t-y_j)}{N-n} - \Phi(\frac{t-\beta}{\sigma})] \xrightarrow{D} N(0, \Phi(\frac{t-\beta}{\sigma})(1-\Phi(\frac{t-\beta}{\sigma}))),$$

which takes care of the limit distribution of the second term on the right hand side of (8). Moreover, it may be shown that

$$(10) \qquad \sqrt{n}(T_{n-2}(U_n(t)) - \Phi(\frac{t-\beta}{\sigma})) = \sqrt{n}(\Phi(U_n(t)) - \Phi(\frac{t-\beta}{\sigma})) + O_p(n^{-1/2})$$

and that

$$(11) \qquad \sqrt{n}(U_n(t) - \frac{t-\beta}{\sigma}) \xrightarrow{D} N(0, 1 + \frac{(t-\beta)^2}{2\sigma^2}).$$

Thus, by applying the delta method, it follows from (10) and (11) that

$$\sqrt{n}(T_{n-2}(U_n(t)) - \Phi(\frac{t-\beta}{\sigma}) \xrightarrow{D} N(0, (1 + \frac{(t-\beta)^2}{2\sigma^2})\phi^2(\frac{t-\beta}{\sigma})),$$

which together with (8) and (9) concludes the proof.

Bolfarine and Sandoval (1990) have shown that if σ^2 is known then the BUP of $F_N(t)$ is given by

$$\hat{F}_{BU,\sigma} = \frac{n}{N}F_{N_s}(t) + (1-\frac{n}{N})\Phi(\sqrt{\frac{n}{n-1}}(\frac{t-\bar{y}_s}{\sigma})).$$

Thus, we may define, in the case of σ^2 unknown, the following alternative predictor

$$\hat{F}_N^*(t) = \frac{n}{N}F_{N_s} + (1-\frac{n}{N})\Phi(\sqrt{\frac{n}{n-1}}(\frac{t-\bar{y}_s}{S_y})).$$

The result that follows next stablishes an important relationship $\hat{F}_{BU}(t)$ and $\hat{F}_N^*(t)$.

Theorem 3.3. *Under the assumptions described in Theorem 3.2,*

$$\hat{F}_{BU}(t) - \hat{F}_N^*(t) \xrightarrow{a.s.} 0.$$

Proof. Notice that

$$\hat{F}_{BU}(t) - \hat{F}_N^*(t) = (1-\frac{n}{N})[T_{n-2}(U_n(t)) - \Phi(\frac{t-\bar{y}_s}{S_y})].$$

We assert without proof the following facts:

$$\Phi(\frac{t-\bar{y}_s}{S_y}) \xrightarrow{a.s.} \Phi(\frac{t-\beta}{\sigma}),$$

$$(12) \qquad U_n(t) \xrightarrow{a.s.} \frac{t-\beta}{\sigma} \quad (\Longrightarrow \Phi(U_n(t)) \xrightarrow{a.s.} \Phi(\frac{t-\beta}{\sigma}))\cdot$$

and

$$T_{n-2}(x) \longrightarrow \Phi(x) \quad \text{(uniformly)},$$

which implies that

$$(13) \qquad T_{n-2}(U_n(t)) - \Phi(U_n(t)) \xrightarrow{a.s.} 0.$$

Moreover, we may write

$$|T_{n-2}(U_n(t)) - \Phi(\frac{t-\beta}{\sigma})| \leq |T_{n-2}(U_n(t)) - \Phi(U_n(t))| + |\Phi(U_n(t)) - \Phi(\frac{t-\beta}{\sigma})|,$$

which together with (12) and (13) concludes the proof.

Thus, under the assumptions of Theorem 3.2, $\hat{F}_N^*$ and $\hat{F}_{BU}(t)$ have the same asymptotic predictive distribution.

With respect to simple random sampling without replacement, the ordinary predictor of $F_N(t)$ is the sample empirical distribution function, $F_{N_s}(t)$, given by (5). Clearly $F_{N_s}(t)$ is unbiased and it may be considered as a distribution-free predictor. That is, it should be used in situations where there is little or no knowledge at all about the form of the distribution function of the error vector **e**. Indeed it is the optimal predictor when the model under consideration is the class of all continuous distributions with density. In this case, the complete and totally sufficient statistic is the vector of order statistic and $F_{N_s}(t)$ being simmetric in $\mathbf{y}_s$ certanly is a function of it. A direct application of the Central Limit Theorem for iid random variables yields

$$(14) \qquad \sqrt{n_\nu}(F_{N_{s,\nu}}(t) - F_{N_\nu}(t)) \xrightarrow{D} N(0, (1-f)\Phi(\frac{t-\beta}{\sigma})(1-\Phi(\frac{t-\beta}{\sigma}))).$$

From (7) and (14) it follows that the asymptotic relative efficiency of $F_{N_s}(t)$ to $\hat{F}_{BU}(t)$ is

$$ARE(F_{N_s}(t); \hat{F}_{BU}(t)) = f + (1-f)\frac{(1+\frac{(t-\beta)^2}{2\sigma^2})\phi^2(\frac{t-\beta}{\sigma})}{\Phi(\frac{t-\beta}{\sigma})(1-\Phi(\frac{t-\beta}{\sigma}))}.$$

For the case where $t = \beta$,

$$ARE(F_{N_s}(t), \hat{F}_{BU}(t)) = f + (1-f)0.637.$$

Further, it can be shown that as $|t-\beta| \to \infty$,

$$ARE(F_{N_s}(t); \hat{F}_{BU}(t)) \to f.$$

Thus, the above efficiency can be as low as f.

4. The Linear Regression Model

Suppose now that model (1) is such that

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix} \quad \text{and} \quad \mathbf{W} = diag(v_1, \ldots, v_N), \tag{15}$$

where $v_i > 0$, $i = 1, \ldots, n$. In this general case, the totally sufficient and complete statistic is given by (2).

Theorem 4.1. *Under model (1) with* $\mathbf{X}$ *and* $\mathbf{W}$ *as in (15) and unknown* σ^2, *it follows that the BUP of* $F_N(t)$ *is given by*

$$\hat{F}_{BU}(t) = \frac{n}{N} F_{N_s}(t) + (1 - \frac{n}{N}) \sum_{i \notin s} \frac{1}{N-n} T_{n-p-1}(U_n^i(t)),$$

where T_ν *is the distribution function of the Student t distribution with* ν *degrees of freedom and*

$$U_n^i(t) = \frac{\sqrt{n-p-1}(t - \mathbf{x}_i'\hat{\beta}_s)}{\sqrt{(1 - \frac{1}{v_i}\mathbf{x}_i'(\mathbf{X}_s'\mathbf{W}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{x}_i)(n-p)S_y^2 - \frac{1}{v_i}(t - \mathbf{x}_i'\hat{\beta}_s)^2}}.$$

Proof. We consider first the case of constant variance, that is $v_i = 1$, $i = 1, \ldots, N$. According to Olkin and Ghurye (1969),

$$E_\psi[T_{n-p-1}(U_n^i(t))] = E_\psi[\Delta(t - y_i)],$$

which shows that $\hat{F}_{BU}(t)$ is an unbiased predictor of $F_N(t)$. Since it is a function of the totally sufficient and complete statistics, the result follows from Theorem 2.1. Although the results of Olkin and Ghurye (1969) are intended to random vectors with constant variance, the proof for the case $\mathbf{W} = diag(v_1, \ldots, v_N)$ follows directly from the case $\mathbf{W} = \mathbf{I}_N$ by making the transformations $y_i^* = y_i/v_i^{1/2}$, $x_i^* = x_i/v_i^{1/2}$ and $e_i^* = e_i/v_i^{1/2}$, $i = 1, \ldots, N$.

Example 4.1. Consider the simple regression model where $\mathbf{X}' = (x_1, \ldots, x_N)$ and $\mathbf{W} = \mathbf{I}$, that is,

$$y_i = x_i\beta + e_i,$$

and $e_i \sim N(0, \sigma^2)$, with unknown σ^2, $i = 1, \ldots, N$. In this case, it follows from (2) that

$$\hat{\beta}_s = \frac{\sum_{i\in s} x_i y_i}{\sum_{i \in s} x_i^2} \quad \text{and} \quad S_y^2 = \frac{1}{n-1}\{\sum_{i\in s} y_i^2 - \frac{(\sum_{i\in s} x_i y_i)^2}{\sum_{i\in s} x_i^2}\}.$$

It follows from Theorem 4.1 that the BUP of $F_N(t)$ is given by

$$\hat{F}_{BU}(t) = \frac{n}{N} F_{N_s}(t) + (1 - \frac{n}{N}) \sum_{i \notin s} \frac{1}{N-n} T_{n-2}(U_n^i(t)),$$

where

$$U_n^i(t) = \frac{\sqrt{n-2}(t - x_i\hat{\beta}_s)}{\sqrt{(1 - \frac{x_i^2}{\sum_{j\in s} x_j^2})(n-1)S_y^2 - (t - x_i\hat{\beta}_s)^2}},$$

$i \notin s$. Using some results in Hurt (1976), it can be shown, after a reasonable amount of mathematical manipulations, that the prediction variance of $\hat{F}_{BU}(t)$ is given by

$$Var[\hat{F}_{BU}(t) - F_N(t)]$$

$$= (1 - \frac{n}{N})^2 \{ \frac{1}{(N-n)^2} \sum_{i\notin s}\sum_{k\notin s} t_{n-2}(\frac{t - x_i\beta}{\sigma}) t_{n-2}(\frac{t - x_k\beta}{\sigma}) [\frac{x_i x_k}{\sum_{j\in s} x_j^2} + \frac{(t - x_i\beta)(t - x_k\beta)}{2\sigma^2(n-1)}]$$

$$+ \frac{1}{(N-n)^2} \sum_{i\notin s} \Phi(\frac{t - x_i\beta}{\sigma})(1 - \Phi(\frac{t - x_i\beta}{\sigma}))\} + O(n^{-2}), \tag{16}$$

where t_{n-2} denotes the density of the t distribution with $n-2$ degrees of freedom. Notice that if $x_i = 1$, $i = 1, \ldots, N$ and n is large, the prediction variance (16) and the asymptotic prediction variance given in (7) are close.

Example 4.2. In the special case where $\mathbf{X}' = (x_1, \ldots, x_N)$ and $v_i = x_i$, $i = 1, \ldots, N$, it follows that the totally sufficient and complete statistic $(\hat{\beta}_s, S_y^2)$ are such that

$$\hat{\beta}_s = \frac{\bar{y}_s}{\bar{x}_s} \quad \text{and} \quad S_y^2 = \frac{1}{n-1}[\sum_{i\in s} \frac{y_i^2}{x_i} - (\frac{\sum_{i\in s} y_i)^2}{\sum_{i\in s} x_i}],$$

where $\bar{y}_s = \sum_{i\in s} y_i/n$ and $\bar{x}_s = \sum_{i\in s} x_i/n$, so that $\hat{\beta}_s$ is the usual estimator of the ratio of the population means $\bar{y}/\bar{x}$. It follows from Theorem 4.1 that the BUP of $F_N(t)$ is given by

$$\hat{F}_{BU}(t) = \frac{n}{N} F_{Ns}(t) + (1 - \frac{n}{N}) \sum_{i\notin s} \frac{1}{N-n} T_{n-2}(U_n^i(t)), \tag{17}$$

where

$$U_n^i(t) = \frac{\sqrt{n-2}(t - x_i\hat{\beta}_s)}{\sqrt{x_i}\sqrt{(1 - \frac{x_i}{\sum_{j\in s} x_j})(n-1)S_y^2 - \frac{(t - x_i\hat{\beta}_s)^2}{x_i}}}.$$

Furthermore, we can use the approach considered in Example 4.1 to derive an approximation for the prediction variance of predictor (17). After some algebraic manipulations, it can be shown that

(18)

$$Var_\psi[\hat{F}_{BU}(t) - F_N(t)] = (1 - \frac{n}{N})^2 \{ \frac{1}{(N-n)^2} \sum_{i\notin s}\sum_{k\notin s} t_{n-2}(\frac{t - x_i\beta}{\sigma\sqrt{x_i}}) t_{n-2}(\frac{t - x_i\beta}{\sigma\sqrt{x_i}}) [\frac{\sqrt{x_i x_k}}{\sum_{j\in s} x_j}$$

$$+\frac{(t-x_i\beta)(t-x_k\beta)}{2\sqrt{x_ix_k}\sigma^2(n-1)}] + \frac{1}{(N-n)^2}\sum_{i\notin s}\Phi(\frac{t-x_i\beta}{\sigma\sqrt{x_i}})(1-\Phi(\frac{t-x_i\beta}{\sigma\sqrt{x_i}}))\} + O(n^{-2}).$$

Notice that if $x_i = 1$, $i = 1, \ldots, n$, the prediction variances (16) and (18) coincide.

Simulation studies conducted by the authors using populations generated according to the normal superpopulation models considered in Examples 4.1 and 4.2 have indicated that the optimal predictors may present substantial improvement over its competitors proposed by Chambers and Dustan (1986) and Rao et al. (1990). Moreover, the studies also seem to indicate that the prediction variance (18) decreasis as $\sum_{i\in s} x_i$ increasis, which provides some indication of which purposive sample should be used in conjunction with the optimal predictor $\hat{F}_{BU}(t)$.

References

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73, 597-604.

Hurt, J. (1976). Asymptotic expansions of functions of statistics. *Aplikace Matematiky*, 21, 444-456.

Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley.

Olkin, I. and Ghurye, S.G. (1969). Unbiased estimators of some multivariate densities and related functions. *The Annals of Mathematical Statistics*, 40, 1261-1271.

Rao, J.N.K., Kovar, J.G. and Mantel, H.J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77, 365-375.

Rodrigues, J., Bolfarine, H. and Rogakto, A. (1985). A general theory of prediction in finite populations. *International Statistics Review*, 53, 239-254.

Sedransk N. and Sedransk J. (1979). Distinguishing among distributions using data from complex sample designs. *Journal of the American Statistical Association*, 74, 754-760.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*. Wiley.

Tam, S.M. (1987). Optimality of Royall's predictor under a Gaussian superpopulation model. *Biometrika*, 74, 659-660.

Zacks, S. and Bolfarine, H. (1991). Equivariant prediction of the population variance in finite populations. *Sankhya, B*, 53, 3, 288-296.

ULTIMOS RELATORIOS TECNIÇOS PUBLICADOS

1992

9201 - BOLFARINE, H., NASCIMENTO, J.A. & RODRIGUES, J. Comparing Several Regression Models with Measurement Errors. A Bayesian Approach, 16p.

9202 - BOLFARINE, H. & SANDOVAL, M.C. Empirical Bayesian Prediction in the Location Error in Variables Superpopulation Model, 26p.

9203 - BUSSAB, W.O. & BARROSO, L.P. Painel Multivariado - Análise Através do Modelo de Componentes de Variância, 07p.

9204 - LEITE, J.G. & PEREIRA, C.A.B. Urn Scheme to Obtain Properties of Stirling Numbers of Second Kind, 09p.

9205 - BELITSKY, V. A Stochastic Model of Deposition Processes with Nucleation, 21p.

9206 - BOLFARINE, H. & NASCIMENTO, J.A. Bartlett Correction Factors for the Structural Regression Model with Known Reliability Ratio, 11p.

9207 - FERRARI, P.A. Growth Processes on a Strip, 23p.

9208 - FERRARI, P.A., GALVES, J.A. & LANDIM, C. Exponential Waiting Time for a Big Gap in a One Dimensional Zero Range Process, 8p.

9209 - LOSCHI, R.H. Coerência e Probabilidade, 17p.

9210 - CRIBARI-NETO, F. & FERRARI, S.L.P. An Improved Lagrange Multiplier Statistic for the Test of Heteroskedasticity, 22p.

9211 - LEITE, J.G. & BOLFARINE, H. Bayesian Estimation of the Number of Equally Likely Classes in a Population, 10p.

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
  01498 - São Paulo, Brasil