YOKNHAMA MATHEMATICAL
JOURNAL VOL. 32, 1984

THE BEST UNBIASED PREDICTORS IN FINITE POPULATION

By

YASUSHI TAGA AND MASAYUKI HONDA

(Received July 20, 1984)

ABSTRACT. A method for constructing the best unbiased predictors in finite populations are stated by using the complete sufficient statistic in the sense of prediction under super-population models.
KEY WORDS: Super-population; best unbiased predictor

1. Introduction

In the latest decade various methods for inferences on finite populations have been proposed and discussed under super-population models, e.g. Smith (1976), Fuller and Isaki (1981), Hansen, Madow and Tepping (1983), etc.

In this paper methods for constructing the best unbiased predictors in finite populations are stated by using the complete sufficient statistic in the sense of prediction under some super-population models as follows.

Let Π_N be a finite population of size N, the i-th element of which has the labels $(\boldsymbol{x}_i, y_i)$ where $\boldsymbol{x}_i$ denotes k-dimensional covariate vector $(x_{1i}, x_{2i}, \cdots, x_{ki})^t$ and y_i objective variate for $1 \leqq i \leqq N$.

Besides let us assume Π_N may be considered to be a sample of size N taken independently from a super-population Π with density function $f(\boldsymbol{x}, y|\theta)$.

Further let us take a sample of fixed sample size n from Π_N according to a suitably chosen sampling design $p(s)$ which may depend on the values of covariates $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ observed in advance of sampling.

Then the conditional distribution of $\boldsymbol{y}_s=(y_{i_1}, y_{i_2}, \cdots, y_{i_n})^t$ given $\boldsymbol{x}_s=(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \cdots, \boldsymbol{x}_{i_n})$ is represented by

$$p(s)=\prod_{j=1}^{n} f(y_{i_j}|\boldsymbol{x}_{i_j}, \theta)\,, \qquad s=(i_1, i_2, \cdots, i_n) \tag{1.1}$$

where $f(y|\boldsymbol{x}, \theta)$ denotes the conditional density function of y given $\boldsymbol{x}$.

In the following we shall construct predictors of parameters in Π_N utilizing the correlation between $\boldsymbol{x}$ and y, such that predictors may depend only on observed values of $\boldsymbol{y}_s=(y_{i_1}, y_{i_2}, \cdots, y_{i_n})^t$ and $X=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)^t$.

Finally some definitions and a lemma are stated below.

Definition 1.1. A statistic U is said to be θ-unbiased or p-unbiased for a parameter z in Π_N if

$$E_\theta\{U\}=E_\theta\{z\} \quad \text{or} \quad E_p\{U\}=z$$

holds for all θ or $\boldsymbol{y}=(y_1, y_2, \cdots, y_N)'$ respectively, where E_θ or E_p denote expectation operations with respect to the density function f or sampling design p respectively. Moreover U is said to be $p\theta$-unbiased for z if

$$E_\theta E_p\{U\}=E_\theta E_p\{z\}$$

holds for all θ.

Definition 1.2. A statistic T is said to be sufficient in the sense of prediction (pred.-sufficient in short) if T is sufficient for θ in the usual sense and if $\boldsymbol{y}_s$ and z are conditionally independent under any given T.

Lemma 1.1. *Let U be an unbiased predictor of z and T be pred.-sufficient for z, and define U_T by*

$$U_T=E_\theta\{U|T\}\,. \tag{1.1}$$

Then U_T is independent of θ and the relations

$$E_\theta\{U_T\}=E_\theta\{z\}\,, \quad \textit{and} \tag{1.2}$$

$$E_\theta\{(U_T-z)^2\}\leqq E_\theta\{(U-z)^2\} \tag{1.3}$$

holds for all θ, i.e. U_T may be considered to be a θ-unbiased predictor for z obtained by improving U.

If T is complete and sufficient for θ in addition to the above conditions, then U_T is the best unbiased predictor for z which has the smallest mean square error of prediction among all unbiased predictors.

The proof of this lemma may be easily achieved by considering that the quadratic loss $(U-z)^2$ is convex in U for any fixed z and that $\boldsymbol{y}_s$ and z is conditionally independent under any given T. The optimality of U_T for complete T can be shown by the fact that U_T is determined uniquely as a function of T because of its completeness.

In the following let us adopt the quadratic loss $(U-z)^2$ as the criterion of goodness for a predictor U for z.

Note that U is necessarily $p\theta$-unbiased if U is p-unbiased or θ-unbiased, and that expectation operations E_p and E_θ may be exchanged under the condition that $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ are all fixed.

2. Main Results (Normal regression case)

Let us suppose the normal linear regression of η on $\boldsymbol{\xi}$ in the super-population Π such that

$$\eta=\beta_0+\boldsymbol{\xi}^t\boldsymbol{\beta}+\varepsilon \tag{2.1}$$

where β_0, $\boldsymbol{\beta}=(\beta_1, \beta_2, \cdots, \beta_k)^t$ are regression coefficients of η on $\boldsymbol{\xi}=(\xi_1, \xi_2, \cdots, \xi_k)^t$ and $\varepsilon\sim N(0, \sigma^2)$.

Then the linear regression model of $\boldsymbol{y}$ on $X=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)^t$ in Π_N may be represented such that

$$\boldsymbol{y}=W\boldsymbol{\gamma}+\boldsymbol{\varepsilon}\,, \tag{2.2}$$

where $W=(\mathbf{1}_N, X)$, $\boldsymbol{\gamma}=(\beta_0, \boldsymbol{\beta}^t)^t$, $\mathbf{1}_N=(1, 1, \cdots, 1)^t$ and $\boldsymbol{\varepsilon}=(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N)^t\sim N(\mathbf{0}, \sigma^2 I_N)$, N-variate normal distribution with mean vector $\mathbf{0}=(0, 0, \cdots, 0)^t$ and convariance matrix $\sigma^2 I_N$ (I_N: the unit matrix).

As to the sample observations $\boldsymbol{y}_s=(y_{i_1}, y_{i_2}, \cdots, y_{i_n})^t$, the similar linear regression holds:

$$\boldsymbol{y}_s=W_s\boldsymbol{\gamma}+\boldsymbol{\varepsilon}_s\,, \tag{2.3}$$

where $\boldsymbol{y}_s=D_s\boldsymbol{y}$, $W_s=D_sW$, $\boldsymbol{\varepsilon}_s=D_s\boldsymbol{\varepsilon}$ and D_s denotes a diagonal matrix having 1 as the i_j-th element for $j=1, 2, \cdots, n$ and 0 as the other diagonal elements.

Then it is well-known that the least square estimator

$$\hat{\boldsymbol{\gamma}}=(W_s^tW_s)^{-1}W_s^t\boldsymbol{y}_s \tag{2.4}$$

is (conditionally) θ-unbiased for $\boldsymbol{\gamma}$, given $X=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)^t$, and its (conditional) covariance matrix is given by

$$V_\theta(\hat{\boldsymbol{\gamma}}|X)=\sigma^2(W_s^tW_s)^{-1}\,, \tag{2.5}$$

if $W_s^tW_s$ is non-singular.

Now let us construct the best θ-unbiased predictor of the population mean $\bar{y}$ of Π_N using $\hat{\boldsymbol{\gamma}}$ given by (2.4) in the following way.

Since the population mean $\bar{y}$ may be represented by (2.2) as

$$\bar{y}=(N)^{-1}\mathbf{1}_N^t\boldsymbol{y}=(1, \bar{\boldsymbol{x}}^t)\boldsymbol{\gamma}+\bar{\varepsilon}\,, \tag{2.6}$$

then it is easily seen that the predictor T_0 given by

$$T_0=h(\boldsymbol{y}_s)=(1, \bar{\boldsymbol{x}}^t)\hat{\boldsymbol{\gamma}} \tag{2.7}$$

is (conditionally) θ-unbiased for $\bar{y}$, where $\bar{\boldsymbol{x}}=(N)^{-1}\mathbf{1}_N^tX$ and $\bar{\varepsilon}=(N)^{-1}\mathbf{1}_N^t\boldsymbol{\varepsilon}$.

It is noted that the statistic

(2.8) $$T=(W_s\boldsymbol{y}_s,\ \boldsymbol{y}_s{}^t\boldsymbol{y}_s)$$

is (conditionally) complete sufficient for $\theta=(\boldsymbol{\gamma}, \sigma^2)$ in the usual sense under the normal regression model (2.3). Besides it is easily seen that $\boldsymbol{y}_s$ and $\bar{y}$ are independent (see Appendix A), i.e. T is complete pred.-sufficient for θ.

Since T_0 is a function of such T, T_0 is proved to be the best unbiased predictor of $\bar{y}$ by Lemma 1. Namely T_0 has the smallest mean square error of prediction among all unbiased predictors of $\bar{y}$, and it is shown after some calculations to be

(2.9) $$E_\theta\{(T_0-\bar{y})^2|X\}=(\sigma^2/n)\{(1-n/N)+(\boldsymbol{x}-\bar{\boldsymbol{x}}_s)^t\hat{\Sigma}_s^{-1}(\bar{\boldsymbol{x}}-\bar{\boldsymbol{x}}_s)\}$$

where $\bar{\boldsymbol{x}}_s=(n)^{-1}1_N{}^tX_s$ and $\hat{\Sigma}_s=(n)^{-1}X_s{}^tX_s-\bar{\boldsymbol{x}}_s\bar{\boldsymbol{x}}_s{}^t$ (see Appendix B).

Now let us get the best θ-unbiased predictor of S_y^2, the population variance of y in Π_N, as follows.

From (2.2) it is easily seen that

(2.10)
$$\begin{aligned}E_\theta\{S_y^2|X\}&=E_\theta\left\{(N)^{-1}\sum_{i=1}^N y_i^2-\bar{y}^2|X\right\}\\&=(N)^{-1}\sum_{i=1}^N(\boldsymbol{w}_i{}^t\boldsymbol{\gamma})^2+\sigma^2-\left((N)^{-1}\sum_{i=1}^N\boldsymbol{w}_i{}^t\boldsymbol{\gamma}\right)^2-(N)^{-1}\sigma^2\\&=\boldsymbol{\beta}^t\hat{\Sigma}\boldsymbol{\beta}+(1-(N)^{-1})\sigma^2\,,\end{aligned}$$

where

$$\boldsymbol{\gamma}^t(N)^{-1}\sum_{i=1}^N(\boldsymbol{w}_i-\bar{\boldsymbol{w}})(\boldsymbol{w}_i-\bar{\boldsymbol{w}})^t\boldsymbol{\gamma}=(\beta_0,\ \boldsymbol{\beta}^t)\begin{bmatrix}0 & \mathbf{0}^t\\ \mathbf{0} & \hat{\Sigma}\end{bmatrix}\begin{pmatrix}\beta_0\\ \boldsymbol{\beta}\end{pmatrix}=\boldsymbol{\beta}^t\hat{\Sigma}\boldsymbol{\beta}$$

and

$$\boldsymbol{\gamma}^t=(\beta_0,\ \boldsymbol{\beta}^t)\,,\quad \boldsymbol{w}_i{}^t=(1,\ \boldsymbol{x}_i{}^t)\quad\text{for}\quad 1\leqq i\leqq N\,,$$

$$\bar{\boldsymbol{w}}=(N)^{-1}\sum_{i=1}^N w_i\quad\text{and}\quad\hat{\Sigma}=(N)^{-1}\sum_{i=1}^N(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^t\,.$$

Further it is well-known that the residual variance

(2.11) $$\hat{\sigma}^2=(\boldsymbol{y}_s-W_s\hat{\boldsymbol{\gamma}})^t(\boldsymbol{y}_s-W_s\hat{\boldsymbol{\gamma}})/(n-k-1)$$

is unbiased for σ^2 if $W_s{}^tW_s$ is non-singular.

From (2.10) and (2.11) we can get a θ-unbiased predictor as

(2.12) $$\hat{S}_y^2=\hat{\boldsymbol{\beta}}^t\hat{\Sigma}\hat{\boldsymbol{\beta}}+(1-(N)^{-1}-(n)^{-1}\operatorname{tr}[\hat{\Sigma}\hat{\Sigma}_s^{-1}])\hat{\sigma}^2\,,$$

where $\hat{\Sigma}_s=n^{-1}\sum_{j=1}^n(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_s)(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_s)^t$ and $\operatorname{tr}[\hat{\Sigma}\hat{\Sigma}_s^{-1}]\fallingdotseq k$ for large n. (see Appendix C)

3. Discussions

Our method for constructing the best unbiased predictors in finite populations may be applicable under the super-population models other than normal regression case, e.g. log-normal or exponential, if there exists a complete pred.-sufficient statistic under those models.

In the case where the covariate vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ may be designated arbitrarily in a limited region R, it is an interesting problem how to allocate those N vectors in R such that the mean square error of prediction can be minimized, i.e. the so-called optimal design. It should be noted, however, that such optimal design could not be best if the assumed model is far from the true model, because the "best unbiased predictor" under the assumed model can not be unbiased including bigger bias as the difference becomes larger between the assumed models and true ones.

Thus it is important to check the validity of the assumed model or to select a suitable model among the family of assumed models.

We will discuss these points in full detail in near future.

References

[1] T. M. F. Smith, *The foundation of survey samplig : a Review*, J. R. Statist. Soc. A, 139, (1976), 183–195.

[2] W. A. Fuller and C. T. Isaki, *Survey design under super-population models*, Current Topics in Survey Sampling, ed. by D. Krewski R. Platek and J. N. K. Rao, (1981), 199–226.

[3] C. E. Särndal, *Implications of survey design for generalized regression estimation of linear functions*, J. Statist. Plan. Inf., 7, (1982), 155–170.

[4] M. H. Hansen, W. G. Madow and B. J. Tepping, *An evaluation of model-dependent and probability-sampling inferences in sample surveys*, J. Amer. statis. Ass., (1983), 776–793.

Department of Mathematics and Division of Medical Informatics
Yokohama City University Chiba University Hospital
Yokohama 236, Japan Chiba 280, Japan

Appendix A. Since T contains $\sum_{j=1}^{n} y_{i_j}$ as its first component and $\bar{y}=(N)^{-1}\{\sum_{j=1}^{n} y_{i_j}+\sum' y_{i_j}\}$ where $\sum'$ denote the summation over all y_{i_j}'s other than the sample $\boldsymbol{y}_s=(y_{i_1}, y_{i_2}, \cdots, y_{i_n})^t$, then it is easily seen that $\boldsymbol{y}_s$ and $\bar{y}$ are conditionally independent under any fixed T by the assumption that $\boldsymbol{y}=(y_1, y_2, \cdots, y_N)^t$ be considered as a sample of N independent observations from Π.

Appendix B. Since

$$\begin{aligned} T_0-\bar{y} &= (1, \bar{\boldsymbol{x}}^t)(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})-\bar{\varepsilon} \\ &= (1, \bar{\boldsymbol{x}}^t)(W_s^t W_s)^{-1} W_s^t \boldsymbol{e}_s-\bar{\varepsilon} \end{aligned}$$

then

$$E_\theta\{(T_0-\bar{y})^2|X\}=\sigma^2(1,\bar{\boldsymbol{x}}^t)(W_s^tW_s)^{-1}\begin{pmatrix}1\\ \bar{\boldsymbol{x}}\end{pmatrix}+\sigma^2/N \\ -2(1,\bar{\boldsymbol{x}}^t)(W_s^tW_s)^{-1}W_s^tE_\theta\{\boldsymbol{e}_s\bar{\varepsilon}|X\}\,,$$

where

$$(W_s^tW_s)^{-1}=\begin{bmatrix}n & n\bar{\boldsymbol{x}}_s^t\\ n\bar{\boldsymbol{x}}_s & X_s^tX_s\end{bmatrix}^{-1}=(n)^{-1}\begin{bmatrix}1+\bar{\boldsymbol{x}}_s^t\hat{\Sigma}_s^{-1}\bar{\boldsymbol{x}}_s & -\bar{\boldsymbol{x}}_s^t\hat{\Sigma}_s^{-1}\\ -\hat{\Sigma}_s^{-1}\bar{\boldsymbol{x}}_s & \hat{\Sigma}_s^{-1}\end{bmatrix}$$

and

$$E_\theta\{\boldsymbol{e}_s\bar{\varepsilon}|X\}=(\sigma^2/N)D_s\mathbf{1}_N\,.$$

Therefore we can get

$$E_\theta\{(T_0-\bar{y})^2|X\}=\sigma^2/n\{1+\bar{\boldsymbol{x}}_s^t\hat{\Sigma}_s^{-1}\bar{\boldsymbol{x}}_s-2\bar{\boldsymbol{x}}_s^t\hat{\Sigma}_s^{-1}\bar{\boldsymbol{x}}+\bar{\boldsymbol{x}}^t\hat{\Sigma}_s^{-1}\bar{\boldsymbol{x}}\}+\sigma^2/N-2\sigma^2/N \\ =(\sigma^2/n)\{(1-n/N)+(\bar{\boldsymbol{x}}_s-\bar{\boldsymbol{x}})^t\hat{\Sigma}_s^{-1}(\bar{\boldsymbol{x}}_s-\bar{\boldsymbol{x}})\}\,,$$

which proves (2.9).

Appendix C. Since

$$E_\theta\{\hat{\boldsymbol{\beta}}^t\hat{\Sigma}\hat{\boldsymbol{\beta}}|X\}=E_\theta\{\mathrm{tr}(\hat{\Sigma}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^t)|X\} \\ =\mathrm{tr}[\hat{\Sigma}\boldsymbol{\beta}\boldsymbol{\beta}^t+n^{-1}\sigma^2\hat{\Sigma}\hat{\Sigma}_s^{-1}]\,,$$

then

$$E_\theta\{\hat{S}_y^2|X\}=\boldsymbol{\beta}^t\hat{\Sigma}\boldsymbol{\beta}+n^{-1}\sigma^2\,\mathrm{tr}[\hat{\Sigma}\hat{\Sigma}_s^{-1}] \\ +(1-N^{-1}-n^{-1}\,\mathrm{tr}[\hat{\Sigma}\hat{\Sigma}_s^{-1}])\sigma^2 \\ =\boldsymbol{\beta}^t\hat{\Sigma}\boldsymbol{\beta}+(1-N^{-1})\sigma^2\,,$$

which is equal to the conditional expectation of S_y^2 given by (2.10).