Qatar Univ. Sci. J. (2004), 24 : 23 - 36

Some Inferential Problems in Finite Population Sampling

Parimal Mukhopadhyay
Department of Mathematics
University of Qatar
Doha, Qatar

بعض المسائل الاستقرائية في معاينة المجتمعات المنتهية

باريمال موكوباداياى
قسم الرياضيات، جامعة قطر

نراجع في هذا المقال بعض المسائل حول تقدير مجموع (معدل) مجتمع منته من خلال مسح إحصائي يناقش الجزء الثاني مسألة التقدير بالنظر إلى نموذج المجتمع الثابت في حين يناقش الجزء الثالث نفس المسألة على اعتبار أن مجتمعنا هو عينة من المجتمع الأكبر بأستخدم نظرية التنبئو.

وبما أن دالة الكثافة الاحتمالية المستقاة من العينة تعادل احتمالية اختبار العينة، فهذا يلزمنا في حالة استخدام تصميم لا معلوماتي بالاعتماد على الاستقراء النموذجي في هذه الحالة فإن التقارير تكون كافية - دنيا (ولكنها ليست كافية - كاملة) ويمكن تحسينها بطريقة Rao - Blackwell.

وبما أن التقدير غير المنحاز ذا التباين الأصغر المتسق غير موجود لمجموع المجتمع، فقد تمت مراجعة بعض النتائج حول جدارة التقادير في حالة تصميم المعاينة الثابت، أما إذا نظرنا لمجتمعنا على أنه من عينة المجتمع الأكبر، فإن بعض الإستراتيجيات المثلى متوفرة في بعض الحالات، ومن خلال طريقة التنبؤ النظرية، يعتبر تصميم المعاينة الهادف مثالياً لكثير من النماذج وهذا يناقض نظرية المعاينة الاحتمالية التقليدية، ومع ذلك فإن الإستراتيجيات المثلى (النموذجية) تفشل في حالة اختبار النموذج الخاطىء وفي مثل هذه الحالات تنقذنا المعاينة الاحتمالية، لذلك نقترح في هذا البحث بعض الاستراتيجيات التي تعتمد على نموذج المجتمع الأكبر وتصميم المعاينة، كما تم في هذا البحث مراجعة مسألة تقدير تباين التصميم الناتج عن هذه الإستراتيجيات.

Keywords: *Sampling strategies, Likelihood function, Sufficient statistic, UMVUE, Superpopulation models, Prediction-theoretic approach, Estimation, General regression model.*

ABSTRACT

We review some results in problems of estimating a finite population total (mean) through a sample survey. Section 2 considers inference under a fixed population model and Section 3 addresses the same problem when the finite population is looked upon as a sample from a superpopulation and technique of theory of prediction are used. Since the probability density function of data obtained from a sample survey equals the selection probability of the sample, thus making the likelihood function 'flat', use of the likelihood, when a prior is assumed for the finite population parameters, restricts one to model-based inference, in case a non-informative sampling design (s.d.) is used for the survey. The data obtained through a set (sample) are minimal sufficient (though not complete sufficient) for inference and hence the use of Rao-Blackwellization provide improved estimators. Noting the non-existence of a uniformly minimum variance unbiased estimator for population total in general, review is made of the results on admissibility of estimators for a fixed s.d. in the relevant classes. If, however, the survey population is looked upon as a sample from a superpopulation ξ, optimum strategies are available in certain classes. Under the prediction-theoretic approach, a purposive sampling design becomes an optimal one under a wide class of superpopulation models. This is in direct conflict with the classical probability sampling-based theory. However, these model-dependent optimal strategies fail (invoke large bias or large mean square error (mse)) if the assumed models turn out to be wrong. Use of probability sampling salvages the situation. A class of strategies, which depend both on superpopulation model and sampling design, have been suggested. Finally, the problem of asymptotic unbiased estimation of design variance of these strategies under multiple regression superpopulation models have been reviewed.

1. Introduction

Let $P = \{1, \ldots, i \ldots, N\}$ denote a finite population of N (a known number) identifiable units labelled 1, 2, … , N. Associated with each i are two real quantities (y_i, x_i), the value of the main variable 'y' and a closely related auxiliary variable 'x' respectively, $\mathbf{y} = (y_1, \ldots, y_N)$; $\mathrm{x} = (x_1, \ldots, x_N)$ are points in R^N. The quantity $p_i = \mathrm{x}_i / X$; where $X = \Sigma^{N}_{i=1} x_i$, is often called the size-measure of unit i. The problem is that of estimating a parametric function $\theta(\mathbf{y})$, often the population total $\mathrm{Y} = \Sigma^{N}_{i=1} y_i$(population mean $\bar{y} = Y/N$), population variance $\mathrm{S}^2 = \Sigma^{N}_{i=1} (y_i - \bar{y})^2 / N$, through survey sampling, for which a sample is selected from P according to a sampling design (s.d.) p.

Let $S = \{i_1, \ldots, i_{n(S)}\}$, $(1 \le i_t \le N, t = 1, \ldots, N)$, denote a sample (sequence) of units, obtained by $n(S)$ draws from P, i_t denoting the unit selected at the tth draw, i_t not necessarily unequal to $i_{t'}$ even if $t \neq t'$ (which happens in a with replacement (wr) sampling). We denote by $\mathrm{s} = (j_1, \ldots, j_n)$, a sample (set) of n units, $1 \le j_1 < j_2 \ldots \le j_n \le N$. The following concepts are equally valid for S and s. A s.d. is a combination (S, p), where $S = \{s\}$; $p(s)$ denoting the probability of selecting s, $p(s) \ge 0$, $\Sigma_{s\varepsilon S}\, p(s) = 1$.

Let $\pi_i = \Sigma_{s \ni i}\, p(s)$, $\pi_{ij} = \Sigma_{s \ni i,j}\, p(s)$ denote the first order and second order inclusion probabilities, respectively. A s.d. with $\pi_i \,\alpha\, p_i$ is often called a π-ps design. Let $\rho_n = \{p : p(s) > 0\ ) \Rightarrow \nu\,(s) = n\}$, the class of all fixed effective size (number of distinct units, ν (s)) [F.S.] (n) designs. Data obtained through S are $d' = \{(k, Y_k), k \in \mathrm{S}\}$ and through s are $d = \{(k, Y_k), k \in \mathrm{s}\}$. Clearly any d' can be summarized to d by obtaining s corresponding to S.

An estimator $e\,(s, \mathbf{y})$ is a function defined on $S \times R^N$ such that for a given s, its value depends on $\mathbf{y}$ only through those i for which $i \in s$. A combination (p, e) is called a sampling strategy. e is unbiased for θ if $E\,(e\,(s, \mathbf{y}) = \theta(\mathbf{y})\ \forall\ \mathbf{y} \in R^N$($E$ denoting expectation with respect to (wrt) s.d. p). The variance of e is $V(e) = E('e - \theta(\mathbf{y}))^2$ provided e is unbiased for $\theta(\mathbf{y})$. We shall denote the *Horvitz-Thompson* estimator (HTE) as $e_{HT} = \Sigma_{i \in s}\, y_i/\pi_i$. The customary estimator in probability proportional to size wr ($ppswr$) sampling design would be denoted as $e_{pps} = \frac{1}{n} \Sigma_{i\in S}\, y_i/p_i$.

A sampling strategy (a design, estimator pair) $H\,(p, e)$ is said to be better than a sampling strategy H' (p', e') in the sense of variance $(H\, H')$ if

$$V\{H(p, e)\} \le V\{H'(p', e')\} \forall\ \mathbf{y} \in R^N,$$

with strict inequality holding for at least one $\mathbf{y}$. If p is kept fixed, an unbiased estimator e is better than another estimator e' if $V_p(e') \le V_p(e')$ with strict inequality holding for at least one $\mathbf{y}$. For a fixed p, an estimator e* is uniformly minimum variance unbiased estimator (UMVUE) in a class of unbiased estimators $\eta = \{e\}$ if it is better than any other estimator in η.

2. Inference Under A Fixed Population Set-up

We review in this section some inferential aspects in sampling under a fixed population model, i.e. when the population P along with the associated $\mathbf{y}$, $\mathbf{x}$ values are considered as fixed entities and no probability distribution is assumed. Subsection 2.1 deals with the probability density function of data, the likelihood function of $\mathbf{y}$ for a given d (d') and use of the likelihood function in making inference. The concept of sufficiency and Rao-Blackwellization is considered in the next subsection. Uniformly minimum variance unbiased estimation, admissibility of estimators and average variance of estimators under superpopulation models are considered subsequently.

2.1 The p.d.f of Data and the Likelihood Function of y

Let D' (D) be a random variable corresponding to data d' (d). Let also $\psi'(\psi)$ be a random variable having values $S(s)$ in S. A data point $d'(d)$ is said to be consistent with a chosen parameter vector **y** if d' (d) can be obtained from **y**.

For example $d' = \{(1, 8), (3, 7), (4, 9)\}$ is consistent with $\mathbf{y} = \{8, 11, 7, 9, 11\}$.

Let for a given d' (d), $\Omega_{d'}(\Omega_d)$ be the set of **y** for which d' (d) is consistent.

The probability density function (*pdf*) of D' is

$$\begin{aligned} f_{D'}(d';\mathbf{y}) &= P[\{(k,y_k),k \in s\} = d';\mathbf{y} \\ &= P[\psi' = S]\,P[D' = d' \mid \psi' = S;\mathbf{y}] \\ &= p(S)\,\delta(d';\mathbf{y}), \end{aligned} \tag{2.1}$$

where $\delta(d'; \mathbf{y}) = 1(0)$ if d' is consistent (inconsistent) with **y**. Thus

$$f_{D'}(d'; \mathbf{y}) = p(S)\ (0) \text{ for } \mathbf{y} \in \Omega_{d'} \text{ (otherwise)}$$

Similarly, pdf of D is

$$f_D(d; \mathbf{y}) = p(s)\ (0) \text{ for } \mathbf{y} \in \Omega_d \text{ (otherwise)} \tag{2.2}$$

2.1.1 Likelihood Function of y

Given the data $D' = d'$, the likelihood function of parameter vector **y** is

$$L(\mathbf{y} \mid d') = f_{D'}(d'; \mathbf{y}) = p(S)\ (0) \text{ if } \mathbf{y} \in \Omega_{d'} \text{ (otherwise)} \tag{2.3}$$

Similarly, the likelihood function of **y** given $D = d$ is

$$L(\mathbf{y} \mid d) = f_D(d; \mathbf{y}) = p(s)\ (0) \text{ if } \mathbf{y} \in \Omega_d \text{ (otherwise)} \tag{2.4}$$

The likelihood functions are, therefore, 'flat', taking values $p(S)(p(s))$ for $\mathbf{y} \in \Omega_d(\in_{d'})$ and zero elsewhere. There does not exist any unique maximum of the likelihood and hence no maximum likelihood solution of any parametric function $\theta(\mathbf{y})$ exists. The likelihood functions (2.3) and (2.4) only tell us that all $\mathbf{y} \in \Omega_d(\Omega_{d'})$ are equi-probable and tell nothing about the unobserved components of the **y**-vector. The likelihood functions (2.3) and (2.4), first considered by Godambe (1966), are therefore, non-informative. However, if a superpopulaton model ξ is postulated for the population vector **y**, the likelihood function becomes informative (see Royall (1976); Brecking and Chambers (1990), for example).

2.2 Sufficiency, Rao-Blackwellization

The concept of sufficiency and Rao-Blackwellisation in survey sampling in connection with resolving the problem – whether one is required to consider the whole body of data available through a sequence sample or the data obtained from a set sample is enough for making inference was first considered by Basu (1958). As in the traditional statistical theory, if a sufficient statistic is available, any estimator can be improved upon by Rao-Blackwellization.

Definition 2.1 A statistic u (D') is a sufficient statistic for **y** if the conditional distribution of D' given

$u(D') = u_0$ is independent of **y**, provided the conditional distribution is well-defined. Let $z\ (D')$ be a statistic defined over the range space of D' such that $z\ (D') = d$, i.e., z reduces d' obtained through a sequence sample s to the data d for the corresponding set. As an example, if $d' = \{(2, 5), (4, 7), (2, 5)\}$, $z\ (d') = \{(2, 5), (4, 7)\}$.

Theorem 2.1 (Basu and Ghosh, 1967) For any ordered design p, the statistic $z\ (D')$ is sufficient for **y**.

Thus, one need not consider the sequence sample and look at the sequenced data d'. The set sample s and the corresponding data d should be sufficient for making inference. It follows that $z\ (D')$ is minimal sufficient for **y**.

For any estimator $e\ (D')$ for θ define $e\ (d) = E\ \{e\ (D') \mid z\ (D') = d\}$. Since $z\ (D')$ is sufficient for **y**, $e\ (d)$ is independent of any unknown parameter, and depends on D' only through $z\ (D')$ and as such can be taken as an estimator of θ. Since D is a sufficient statistic $e\ (D)$ will be a better estimator than $e\ (D')$.

Theorem 2.2 (Basu and Ghosh, 1967) Let $e\ (D')$ be an estimator of θ. The estimator $e_I(d) = E\ \{e\ (D') \mid z\ (D') = d\}$ has the properties:

- (i) $E(e) = E(e_I)$
- (ii) $MSE(e_I) \leq MSE(e)$ with strict inequality $\forall\ \mathbf{y} \in R^N$ iff $P\ \{e \neq e_I : \mathbf{y}\} > 0$.

2.3 Uniformly Minimum Variance Unbiased Estimation

Godambe (1955) first observed that in survey sampling no UMV-estimator in the class of all linear unbiased estimators of population total exists for any p in general.The proof was subsequently improved by Hege (1965), Hanurav (1966), Ericson (1974) and Lanke (1974).

Definition 2.2 A s.d. p is said to be a *unicluster* design if for any two samples s_1, s_2

$$\{\ p\ (s_1) > 0, p\ (s_2) > 0;\ s_1 \neq s_2\} \Rightarrow s_1 \cap s_2 = \phi,$$

i.e. either two samples are identical or disjoint.

Theorem 2.3 (Lanke, 1974) A s.d. p admits a UMVU-estimator in the class of all linear unbiased estimators *iff* p is a *unicluster* design with $\pi_i > 0\ \forall\ i$.

Theorem 2.4 (Basu, 1971) For any non-census design p (with $\pi_i > 0\ \forall\ i$), there does not exist any UMVUE of Y in the class of all unbiased estimators.

Thus, in general, there does not exist any UMVU-estimator for any s.d.p. Hence, there does not exist any UMVU sampling strategy in general.

2.4 Admissibility Of Estimators

Definition 2.3 For a fixed s.d. p, an estimator e is said to be an admissible estimator of Y within a class C of estimators *iff* there does not exist any estimator in C which is uniformly better than e.

Clearly, within the same class C there may exist more than one admissible estimator. Admissibility ensures that an estimator is uniquely best in C at least at some point y in the parametric space. In the absence of a UMVU-estimator, one should choose an estimator within the class of all admissible estimators. However, a slightly inadmissible estimator may sometimes possess some practical advantages over an admissible estimator and may be used in preference to the later. An important theorem is stated below without proof.

Theorem 2.5 (Cassel, Sarndal and Wretman, 1977) For any s.d. p, with $\pi_i > 0$ ($\forall i$), the generalized difference estimator $e_{GD}(a) = \sum_{i \in s} \frac{y_i - a_i}{\pi_i} + A$ ($A = \sum_{i=1}^{N} a_i$) is admissible in the class of all unbiased estimators of Y (Here, a_i are some known constants).

A corollary to this theorem is that the Horvitz-Theorem estimator is admissible in the class of all unbiased estimators of Y.

2.5 Average Variance Of A Strategy Under A Superpopulation Model

Assume that the value of y on i is a realization of a random variable Y_i ($i = 1, \ldots, N$). Hence the value $\mathbf{y}$ of a survey population P may be looked upon as a particular realization of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_N)$ having a superpopulation distribution ξ_θ, indexed by a parameter vector θ, $\theta \in \Theta$ (the parameter space). The class of priors $\{\xi_\theta, \theta \in \Theta\}$ is called a superpopulation model.

A good deal of inference in survey sampling emerges from the postulation of a suitable prior ξ for $\mathbf{Y}$ and methodologies have been developed to produce optimal sampling strategies based on ξ. Some of these will be reviewed in the next section.

2.5.1 Average Variance under ξ

Since in most cases the expression for variances of different strategies are complicated in nature and are not amenable for comparison, one may take the average values of variance under an assumed appropriate superpopulation model ξ and compare their average variances. In this section we shall consider ξ only for getting the average variance of a strategy.

The average variance (AV) of an unbiased strategy (p, e) under ξ is given by $\varepsilon V_p(e)$. A strategy H_1 will be better than an unbiased strategy $H_2 (H_1 \top H_2)$ if $AV(H_1) < AV(H_2)$.

We recall a very important result due to Godambe and Joshi (1965). The following theorem shows that there exists a lower bound to the average variance of p-unbiased strategies under a very general superpopulation model ξ.

Theorem 2.6 (Godambe and Joshi, 1965) Consider model ξ : $Y_1, \ldots, Y_N$ are independent with $\varepsilon(Y_i) = \mu_i$, $v(Y_i) = \sigma^2_i$ ($i = 1, \ldots, N$). For any unbiased sampling strategy (p, e), with the value of first order inclusion probability π_i,

$$\varepsilon V(p, e) \geq \sum_{i=1}^{N} \sigma^2_i \left(\frac{1}{\pi_i} - 1\right) \tag{2.5}$$

Corollary 2.1 The lower bound (2.5) is attained by e_{HT} applied to a FS(n)-design with $\pi_i \propto \varepsilon(Y_i)$; $i = 1, \ldots, N$. In particular, if $\mu_i = \beta x_i$(β, a constant), any FS-π - ps design applied to e_{HT} attains the lower bound in (2.5).

3 Inference Under Superpopulation Based Approach

Brewer (1963), Royall (1970, 1976), Royall and Herson (1973) and their coworkers considered the survey population as a random sample from a superpopulation and attempted to draw inference about the population parameter from a prediction-theorist's viewpoint.

3.1 Principles Of Inference Based On Theory Of Prediction

We assume that the value y_i on i is a realization of a random variable Y_i (i = 1, ... , N). For simplicity, we shall, henceforth, use the same symbol y_i to denote the population value as well as the random variable of which it is a particular realization, the actual meaning being clear from the context.

We assume throughout that there is a superpopulation distribution ξ of **y**.

Let $\hat{T}_s$ denote a predictor of Y or $\bar{y}$ based on s (the specific one being clear from the context). Note that population total Y (mean $\bar{y}$) are now random variables and not fixed quantities.

Definition 3.1 $\hat{T}_s$ is model-unbiased or ξ-unbiased or m-unbiased predictor of $\bar{y}$ if

$$\varepsilon(\bar{T}_s) = \varepsilon(\bar{y}) = \mu \text{ (say) } \forall\ \theta \in \Theta \text{ and } \forall s : p(s) > 0 \tag{3.1}$$

Definition 3.2 A predictor $\hat{T}_s$ is design-model unbiased (or $p\xi$-unbiased or pm-unbiased) predictor of $\bar{y}$ if

$$E\,\varepsilon(\bar{T}_s) = \mu\ \forall\ \theta \in \Theta \tag{3.2}$$

Clearly, a m-unbiased predictor is necessarily pm-unbiased.

For a non-informative design where $p(s)$ does not depend on the y-values, order of operation E, ε can always be interchanged.

Two types of mean square errors (mse's) of a sampling strategy (p, $\hat{T}_s$) for predicting Y has been proposed in the literature:

$$(a)\ \varepsilon\ MSE\,(p,\hat{T}) = \varepsilon\, E\,(\hat{T} - T)2 = M\,(p,\hat{T}) \text{ (say)}$$

$$(b)\ EMS\ \varepsilon\,(p,\hat{T}) = E\,\varepsilon\,(\hat{T} - \mu)2 \text{ where } \mu = \Sigma\mu_k = \varepsilon(T) = M_1\,(p,\hat{T}) \text{ (say)}$$

It has been recommended that if one's main interest is in predicting the total of the current population from which the sample has been drawn, one should use M as the measure of uncertainty of (p, $\hat{T}$). If one's interest is in predicting the population total for some future population, which is of the same type as the present survey population (having the same μ), one is really concerned with μ, and here M_l should be used (Sarndal, 1980 a). In finding an optimal predictor, one minimizes M or M_l in the class of predictors of interest.

For a given s, the optimal m-unbiased predictor of T (in the minimum ε ($\hat{T}$ - T)^2-sense), as derived by Royall (1970) is,

$$\hat{T}_s^+ = \sum_s y_k + \hat{U}_s^+ , \tag{3.3}$$

where

$$\varepsilon\hat{U}_s^+ = \varepsilon\,(\sum_{\bar{s}} y_k) = \mu_{\bar{s}}, \tag{3.4.1}$$

$$\nu\,(\bar{U}_s^+) \geq \nu\,(\bar{U}_s'), \tag{3.4.2}$$

for any $\hat{U}_s'$ satisfying (3.4.1), where $\bar{s} = P - s$. It is clear that $\hat{T}^+_s$, when it exists, does not depend on the sampling design. An optimal design-predictor pair (p, T_s^+) in the class (ρ, $\hat{\tau}$) is, therefore, one for which

$$M\,(p^+, \hat{T}_s^+) \leq M\,(p\ , \hat{T}_s'),$$

for any $p \in \rho$ a class of sampling designs and $\hat{T}'$, any other m-unbiased predictor $\in \tau$.

Example 3.1

Consider the polynomial regression model:

$$\varepsilon\,(y_k \mid x_k) = \sum_{j=0}^{J} \delta_j \beta_j x_k^j,$$

$$\nu\,(y_k \mid x_k) = \sigma^2\, v(x_k),\ k = 1, \ldots, N, \tag{3.5}$$

$$C\,(y_k, y_{k'} \mid x_k, x_{k'}) = 0,\ k \neq k' = 1, \ldots, N, \tag{3.6}$$

where x_k's are assumed fixed (non-stochastic) quantities, $\beta_j\,(j = 1, \ldots, J)$; σ^2 are unknown quantities, $v(x_k)$ is a known function of x_k, $\delta_j = 1(0)$ if the term x_k^j is present (absent) in μ_k. The model (3.5), (3.6) has been denoted as $\xi\,(\delta_0, \delta_1, \ldots, \delta_J\,;\, v(x))$ by Royall and Herson (1973). The best linear unbiased predictor (BLUP) of Y under this model is

$$\hat{T}_s^*\,(\delta_0, \ldots, \delta_J) = \sum_s y_k + \sum_{j=0}^{J} \delta_j \hat{\beta}_j \sum_{\bar{s}} x_k^j, \tag{3.7}$$

where $\hat{\beta}_j^*$ is the BLUP of β_j under $\xi\,(\delta_0, \ldots, \delta_J;\, v(x))$ as obtainable from Gauss-Markoff theorem.

3.1.1 Special Cases

Under model $\xi\,(0, 1, v(x))$,

$$\hat{T}_s^*\,(0, 1; v\,(x)) = \sum_s y_k + \{(\sum_s x_k\, y_k / v\,(x_k))\,(\sum_s x_k^2 / v\,(x_k))^{-1}\}\, \sum_{\bar{s}} x_k \tag{3.8}$$

$$\varepsilon\,(\hat{T}_s - Y)^2 = \sigma^2\,(\sum_{\bar{s}} x_k)^2 / \sum_s \frac{x_k^2}{v(x_k)} + \sigma^2 \sum_{\bar{s}} v\,(x_k)$$

It follows, therefore, that if

- $v(x_k)$ is monotonically non-decreasing function of x,
- $v(x) / x^2$ is monotonically non-increasing function of x,

the strategy $(p^*, \hat{T}^*)$ will have minimum average variance in the class of all strategies $(p, \hat{T})$, $p \in \rho_n$, $\hat{T} \in L_m$, the class of all linear m-unbiased predictors under ξ, where the sampling design p^* is such that

$$p^*(s) = 1(0) \text{ for } s = s^* \text{ (otherwise)} \tag{3.10}$$

s^* having the property

$$\sum_{s^*} x_k = \max_{s \in S_n} \sum_s x_k \tag{3.11}$$

where

$$S_n = \{s : \nu\,(s) = n\}, \tag{3.12}$$

Considering the particular case, $v(x) = x_k^g$ and writing $\hat{T}^*\,(0, 1; x^g)$ as $\hat{T}^*_g$,

$$\hat{T}_0^* = \sum_s y_k + \{\sum_s x_k\, y_k)\,(\sum_{\bar{s}} x_k)\} / \sum_s x_k^2,$$

$$\hat{T}_1^* = \sum_s y_k + \{\sum_s y_k)\,(\sum_{\bar{s}} x_k)\} / \sum_s x_k = \frac{\bar{y}_s}{\bar{x}_s} X, \tag{3.13}$$

$$\hat{T}_2^* = \sum_s y_k + \{\sum_s y_k\, x_k)\,(\sum_{\bar{s}} x_k)\} / \nu\,(s).$$

Example 3.2

Consider now prediction under multiple regression models as follows. Assume that apart from main variable y we have $(r + 1)$ closely related auxiliary variables x_j ($j = 0, 1, \ldots, r$) with known values $x_{kj} \forall k = 1, \ldots, N$: The variables $y_1, \ldots, y_N$ are assumed to have a joint distribution ξ such that

$$\begin{aligned} \varepsilon(y_k \mid x_k) &= \beta_0 x_{k0} + \beta_1 x_{k1} + \ldots + \beta_r x_{kr}, \\ v(y_k \mid x_k) &= \sigma^2 v_k, \\ v(y_k, y'_k \mid x_k, x'_k) &= 0, \end{aligned} \tag{3.14}$$

where $x_k = (x_{k0}, x_{k1}, \ldots, x_{kr})'$ $\beta_0, \beta_1, \ldots \beta_r$; and δ^2 (> 0) are unknown parameters, v_k is a known function of x_k. If $x_{k0} = 1 \forall k$, the model has an intercept term β_0. Assuming without loss of generality that $s = (1, \ldots, n)$, we shall write

$$\mathbf{y} = (y_s\, y_{\bar{s}})',\ \beta = (\beta_0, \ldots, \beta_r)',$$

$$X = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1r} \\ x_{20} & x_{21} & \cdots & x_{2r} \\ \cdot & \cdot & \cdots & \cdot \\ x_{N0} & x_{N1} & \cdots & x_{Nr} \end{bmatrix} = \begin{bmatrix} X_s \\ X_{\bar{s}} \end{bmatrix} \tag{3.15}$$

$$V = \begin{bmatrix} V_s & 0 \\ 0 & V_{\bar{s}} \end{bmatrix},$$

X_s being a $n \times (r + 1)$ submatrix of X corresponding to $k \in s$; ($X_{\bar{s}}$ defined similarly) V_s ($V_{\bar{s}}$) being a n x $n((N - n) \times (N - n))$ submatrix of V corresponding to $k \in s$ ($k \in \bar{s}$). The multiple regression model (3.15) is, therefore,

$$\varepsilon(y) = X\beta,\ D(y) = \sigma^2 V, \tag{3.16}$$

where D(.) denoting model-dispersion matrix of (.). We shall denote

$$X_j = \sum_k x_{kj},\ x_{js} = \sum_{k \in s} x_{kj},\ \bar{x}_{js} = x_{js} / n,$$

$$x_s = (x_{0s}, \ldots, x_{rs})',\ \bar{x}_s = (\bar{x}_{0s}, \ldots, \bar{x}_{rs})',$$

and $x_{j\bar{s}}$, $\bar{x}_{j\bar{s}}$, $x_{\bar{s}}$, $\bar{x}_{\bar{s}}$ similarly. The model (3.16) will be denoted as ξ (X, v).

For a given s, the BLUP of T is

$$\hat{T}_s^* (X, v) = \sum_s y_k + x'_{\bar{s}} \hat{\beta}_s^* \tag{3.17}$$

where $\hat{\beta}_s^*$ is the generalized least square predictor of β,

$$\hat{\beta}_s^* = (X_s' V_s^{-1} X_s)^{-1} (X_s' V_s^{-1} y_s) \tag{3.18}$$

Hence (Royall, 1971)

$$\hat{T}_s^* (X, v) = [1_n' + (1'_{N-n} X_{\bar{s}}) (X_s' V_s^{-1} X_s)^{-1} X_s' V_s^{-1}]\, y_s, \tag{3.19}$$

where $1'_q = (1, \ldots, 1)'_{q \times 1}$

$$\begin{aligned} M(p, \hat{T}^*) &= E[v(x'_{\bar{s}} \hat{\beta}^*) + v(\textstyle\sum_{\bar{s}} y_k)] \\ &= \sigma^2 E[\{x'_{\bar{s}} (X_s' V_s^{-1} X_s)^{-1} x_{\bar{s}}\} + \textstyle\sum_{\bar{s}} v_k]. \end{aligned} \tag{3.20}$$

3.2 Robustness Of Model-Dependent Optimal Strategies

The model-dependent optimal predictor $\hat{T}(\xi)$ will, in general, be biased and not optimal under an alternative model ξ'. Suppose from practical considerations we assume that the model is $\xi\,(\delta_0, \ldots; \delta_j; v(x))$ and use the predictor $\hat{T}\,(\xi)$ which is BLUE under ξ. The bias of this predictor under a different model ξ' for a particular sample s is

$$\varepsilon_{\xi'}\{\hat{T}_s^*(\xi) - T\} = B\{\hat{T}_s^*(\xi), \xi'\}$$

$$= B_{\delta'_0, \ldots, \delta_J; v'(x)}\, \hat{T}_s^*(\delta_0, \ldots, \delta_J; v(x)), \tag{3.21}$$

ξ'-bias of $\hat{T}^*(\xi)$ for a particular sampling design p is

$$\sum_{s\in S} \varepsilon_{\xi'}\{\hat{T}_s^* - T\}\, p\,(s). \tag{3.22}$$

To preserve the property of unbiasedness of $\hat{T}^*(\xi)$ even when the true model is ξ', we may choose the sampling design in such a way that $\hat{T}^*(\xi)$ remains also unbiased under ξ'. With this end in view, Royall and Herson (1973) introduced the concept of balanced sampling design.

Another way to deal with the situation may be as follows. Of all the predictors belonging to a subclass $\hat{\tau}\,(\xi)$ say, we may choose one $\hat{T}^0(\xi)$, which is least subject to bias even when the model is ξ'. Thus, for a given s, we should use $\hat{T}^0(\xi)$ such that

$$|B\,(\hat{T}_s^0(\xi), \xi')| < |B\,(\hat{T}_s(\xi), \xi')|$$

$\forall \hat{T}\,(\xi) \in \hat{\tau}\,(\xi)$, the choice of subclass $\hat{\tau}\,(\xi)$ being made from other considerations, e.g. from the point of view of mse, etc.

3.2.1 Bias of $\hat{T}_g^*$

We have

$$B\,\{\hat{T}_g^*, \xi'\,(\delta'_0, \ldots, \delta'_J; v(x))\}$$

$$= \sum_{j=0}^{J} \delta'_j\, H_g(j,s)\, \beta_j\,,$$

where

$$H_g(j,s) = [\Sigma x_k^{j+1-g}\, \Sigma x_k - \Sigma x_k^{2-g}\, \Sigma x_k^{g}] \,/\, \sum_s x_k^{2-g} \tag{3.23}$$

$$= I_g(j, s) \,/\, \sum_s x_k^{2-g}, \tag{3.24}$$

which is independent of the form of the variance function in ξ'. Note that $H_g(1, s) = 0$.

Definition 3.3 (Royall and Herson, 1973). A sampling design $\bar{p}\,(L)$ is a balanced s.d. of order L (if it exists) if $\bar{p}\,(s) = 1(0)$ for $s = s_b(L)$ (otherwise), where $s_b\,(L)$, called a balanced sample of order L is such that

$$\bar{x}^{\,j}_{s_b(L)} = \bar{x}^{\,j}_{\bar{s}_b(L)}, \; j = 1, \ldots, L$$

where

$$\bar{x}_s^{(h)} = \sum_s x_k^{(h)}/\, n,\ \bar{x}_{\bar{s}}^{(h)} = \sum_{\bar{s}} x_k^{(h)}/\,(N-n). \tag{3.25}$$

If there are K such samples (3.25), $\bar{p}$ chooses each such sample with probability K^{-1}.

It follows that the ratio predictor $\hat{T}_1^* = \hat{T}$ (0, 1; 1) which is optimal under the model ξ (0, 1; x) remains unbiased under alternative class of models ξ (δ_0, …, δ_J; v (x)), when used along with a balanced sampling design $^{-}{}_{L}$.

In general, consider the bias of $\hat{T}^*$(0, 1; $v(x)$) under ξ (δ_0, …, δ_J; V (x)), which is

$$\sum_0^J \delta_j\, \beta_j\, [(\sum_{\bar{s}} x_k)\,(\frac{\sum_s x_k^{j+1}/\,v(x_k)}{\sum_s x_k^2\,/\,v(x_k)}) - \sum_{\bar{s}} x_k^j\,]. \tag{3.26}$$

Hence, if a sample s*(J) satisfies

$$(\sum_s x_k^j\,)\,/\sum_s x_k = (\sum_s x_k^{j+1}/\,v(x_k))\,/\,(\sum_s x_k^2\,/\,v(x_k)), \quad j = 0, 1, \dots, J, \tag{3.27}$$

the predictors $\hat{T}^*$ (0, 1; v (x)) remains unbiased under ξ (δ_0, …, δ_J; v (x)), on s*(J) for any V (x).

Samples s*(J) satisfying (3.27) may be called generalized balanced samples of order J (Mukhopadhyay, 1996).

For $v(x) = x^2$, (3.27) reduces to

$$n^{-1} \sum_s x_k^{j-1} = (\sum_s x_k^j)\,/\,(\sum_s x_k), j = 0, \dots, J \tag{3.28}$$

Samples satisfying these conditions have been termed 'over-balanced samples' $s_0(J)$ by Scott et al (1978).

The following theorem shows that $\hat{T}^*$(0, 1; v (x)) becomes BLUP under ξ (δ_0, …, δ_J; v (x)) for $s = s^*(J)$ when V (x) is of a particular form.

Theorem 3.1 (Scott, Brewer and Ho, 1978) For $s = s^*(J)$, $\hat{T}^*$(0, 1; v (x)) is BLUP under ξ (δ_0, …, δ_J; v (x)) provided

$$V(x) = v(x) \sum_0^J \delta_j a_j x^{j-1}, \tag{3.29}$$

where a_j's are arbitrary non-negative constants.

It is obvious that all types of balanced samples are rarely available in practice. Royall and Herson (1973), Royall and Pfeffermann (1982) recommended simple random samples as approximately balanced samples $s_b(J)$. Mukhopadhyay (1985a) showed that simple random sampling and *pps* $\sqrt{x}$ sampling are asymptotically equivalent to balanced sampling designs $\bar{p}$ (J) for using the ratio predictor. Mukhopadhyay (1985b) suggested a post-sample predictor which remains almost unbiased under alternative polynomial regression models. For further details on the robustness of the model-dependent optimal predictors the reader may refer to Mukhopadhyay (1977, 1996), among others.

3.3 Projection Predictors

In (3.4) we have considered predicting $\sum_{k \in \bar{s}} Y_k$ only, since the part $\sum_{k \in s} y_k$ of Y is completely known when the data are given and found optimal strategies that minimize $M\,(p, \hat{T}\,)$. However, in predicting the total of a finite population of the same type as the current survey population, one's primary interest is in estimating the superpopulation total $\mu = \sum_k \mu_k$. For a given s, a m-unbiased predictor of T will, therefore, be

$$\hat{T} = \sum \hat{y}_k, \tag{3.30}$$

where

$$\varepsilon(\hat{y}_k) = \mu_k,\ k = 1, \ldots, N. \tag{3.31}$$

The predictors (3.30) are called *projection predictors*

Under $\xi(\delta_0, \ldots, \delta_J; v(x))$, BLU-projection predictor of T is

$$\hat{T}^*(\delta_0, \ldots, \delta_J; v(x)) = \sum_{k=1}^{N} \sum_{0}^{J} \delta_j \beta_j^* x_k^j, \tag{3.32}$$

where $\hat{\beta}^*_j$ is as defined in (3.7). Under $\xi(\mathbf{X}, v)$,

$$\hat{T}^*(\mathbf{X}, v) = x_0' \beta^*_s = 1' \mathbf{X} \beta^*_s, \tag{3.33}$$

where $x'_0 = (x_{00}, x_{01}, \ldots, x_{0r})$; $x_{0j} = \sum X_{kj}$, β^*_s is given in (3.18).

4 Generalised Predictors Under Model ξ (X, v)

Sarndal (1980 b) considered different choices of $\hat{\beta}$ under $\xi(X, v)$. Considering $\hat{\beta}^*$ in (3.19) (dropping the suffix s), we note that a predictor $\hat{\beta}$ of β is of the form

$$\beta = (Z_s' X_s)^{-1} Z_s' X_s \tag{3.35}$$

where Z'_s is a $n \times (r + 1)$ matrix of weight z_{kj} to be suitably chosen such that a predictor $\hat{\beta}$ of β has some desirable properties. $Z_s' X_s$ is of full rank. Different choices of Z_s are:

(a) π^{-1} *weighted.* Z_s and the corresponding β may be called π^{-1} weighted if

$$\Pi_s^{-1} 1_n \in C(Z_s), \tag{3.36}$$

i.e. if

$$\Pi_s^{-1} 1_n = Z_s \lambda, \tag{3.37}$$

$C(Z_s)$ denoting the column space of Z_s, $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_r)'$, a vector of constants λ_j. Here Π = Diag $(\pi_k, k = 1, \ldots, N)$, Π_s = Diag $(\pi_k, k \in s)$.

If $Z_s = \Pi_s^{-1} X_s$

$$\beta = (X_s' \Pi_s^{-1} X_s)^{-1} X_s' \Pi_s^{-1} y_s = \beta\ (\Pi^{-1}) = \beta_\Pi \text{ (say)} \tag{3.38}$$

(b) *BLU – weighted.* Here $Z_s = V_s^{-1} X_s$ when

$$\beta = \hat{\beta}^* = \beta(V^{-1}) \text{ (say)} \tag{3.39}$$

(c) *weighted by an arbitrary matrix Q.* Here $Z_s = Q_s X_s$, where Q is an arbitrary $N \times N$ diagonal matrix of weights and Q_s is a submatrix of Q corresponding to units $k \in s$. Therefore,

$$\beta = (X_s' Q_s X_s)^{-1} (X_s' Q_s y_s) = \beta\,(Q) \text{ (say)}. \tag{3.40}$$

For $\xi(X, v)$, Cassel, Sarndal and Wretman (1976), Sarndal (1980 b) suggested a generalised predictor of T,

$$\sum_s \frac{y_k}{\pi_k} + \sum_{j=0}^{r} (X_j - \sum_s \frac{x_{kj}}{\pi_k}) \hat{\beta}_j^* ,$$
$$= 1_n' \Pi_s^{-1} y_s + (1'X - 1_s' \Pi_s^{-1} X_s) \beta^*,$$
$$= \hat{T}_{GR}(V^{-1}) = \hat{T}^*_{GR} \text{ (say)}, \quad (3.41)$$

where β^*_j is the BLUP of β_j obtained from the jth element of $\hat{\beta}^*_j$ given in (3.18).

For arbitrary weights Q in $\hat{\beta}(Q)$, generalided regression estimator of T is

$$\hat{T}_{GR}(Q) = \sum_s y_k/\pi_k + \sum_0^r \hat{\beta}_j(Q)(X_j - \sum_s x_{kj}/\pi_k) = 1_n' \Pi_s^{-1} Y_s + (1'X - 1_s' \Pi_s^{-1} X_s) \beta(Q). \quad (3.42)$$

Wright (1983) considered a (p, Q, R) strategy for predicting T as a combination of sampling design p and a predictor

$$\hat{T}(Q,R) = \Sigma r_k \{y_k - \sum_0^r x_{kj} \hat{\beta}_j(Q)\} + \sum_0^r X_j \hat{\beta}_j(Q)\}, \quad (3.43)$$

where $R = Diag\ (r_k,\ k = 1, \ldots, N)$, $R_s = Diag\ (r_k,\ k \in s)$, r_k being a suitable weight. For different choices of Q, R one gets different predictors. Some choices of Q are, as before, V^{-1}, Π^{-1}, $(V\Pi)^{-1}$ and of R are 0, I, and Π^{-1}. The choice $R = 0$ gives projection predictors of the type (3.30); $R = \Pi^{-1}$ gives the class of generalised regression predictors considered by Cassel, et al (1976, 1977), Sarndal (1980 b).

We shall now address the problem of asymptotic unbiased estimation of design -variance of $T_{GR}(\Pi^{-1}V^{-1})$ under $\xi\ (X,\ v)$. Consider a more general problem of estimation of A linear functions $\mathbf{F} = (F_1, F_2, \ldots, F_A)' = C'\mathbf{y}$, where $F_q = C_a'\ \mathbf{y}$, $C_a = (C_{1a}, \ldots, C_{Nq})'$, C a $N \times Q$ matrix $((C_{kq}))$, C_{kq} being known constants. Consider the following estimates of F_a:

$$\hat{T}_q = C'_{qs} \Pi_s^{-1} (y_s - y_s) + C_a' \hat{\mathbf{y}}, \quad (3.44)$$

where C'_{qs} is the row vector $(C_{kq},\ k \in s)$ and $\hat{y}_k = x'_k \beta_s$ with

$$\hat{\beta}_s = \hat{\beta}(V^{-1}\Pi^{-1}) = (X_s' V_s^{-1} \Pi_s^{-1} X_s)^{-1} X_s' V_s^{-1} \Pi_s^{-1} y_s. \quad (3.45)$$

The estimator (3.44) is an extension of generalized regression estimator $\hat{T}_{GR}(V^{-1}\Pi^{-1})$ of T. Let $\hat{T} = (\hat{T}_1, \ldots, \hat{T}_A)'$: Then

$$\hat{T} = C_s' \Pi_s^{-1} (y_s - \hat{y}_s) + C'\hat{\mathbf{y}}, \quad (3.46)$$

where C_s is the part of C corresponding to $k \in s$. Now

$$\hat{T} = G_s' \Pi_s^{-1} y_s, \quad (3.47)$$

where

$$\begin{aligned} G_s' &= C_s' - M_s' H_s^{-1} X_s' V_s^{-1}, \\ M_s' &= C_s' \Pi_s^{-1} X_s - C'X, \\ H_s &= X_s' V_s^{-1} \Pi_s^{-1} X_s, \end{aligned} \quad (3.48)$$

Thus $\hat{T}_a = \Sigma\ g_{ska} Y_k / \pi_k$, g_{ska} being the (k, a) th component of G_s. The following two methods have been suggested by Sarndal (1982) for estimating the dispersion matrix $D(\hat{T}) = (\text{cov}_p(\hat{T}_a, \hat{T}_b))$

(a) *Taylor expansion method* : An estimate of $\text{Cov}_p(\hat{T}_a, \hat{T}_b)$ is approximately the Yates-Grundy estimator of covariance,

$$\sum_{k<l\in s}\sum (\pi_{kl}/\pi_k\pi_l - 1)(z_{ka}/\pi_k - z_{la}/\pi_l)(z_{kb}/\pi_k - z_{lb}/\pi_l), \tag{3.49}$$

$$= v_T(a,b) \text{ (say,)}$$

where

$$z_{ka} = C_{ka}e_k,\ e_k = Y_k - x'_k\beta_s. \tag{3.50}$$

Writing

$$YG_s(d_{ka}, d_{kb}) = \sum_{k\le l\in s}\sum (\pi_{kl}/\pi_k\pi_l - 1)(d_{ka}/\pi_k - d_{la}/\pi_l)(d_{kb}/\pi_k - d_{lb}/\pi_l), \tag{3.51}$$

as the YG-transformation of (d_{ka}, d_{kb}), $k \in s$, we have

$$v_T(a,b) = YG_s(z_{ka}, z_{kb}), \tag{3.52}$$

(b) Model method. Here an approximate estimate of $Cov_p(\hat{T}_a, \hat{T}_b)$ is

$$v_M(a,b) = YG_s(z^*_{ka}, z^*_{kb}), \tag{3.53}$$

where

$$z^*_{ka} = g_{ska}e_k. \tag{3.54}$$

For further details on inferential problems in survey sampling the reader may refer to Mukhopadhyay (1996, 1998, 2000).

REFERENCES

[1] **Basu, D. (1958):** On sampling with and without replacement. Sankhya **20**, 287-294.

[2] **Basu, D. (1971):** An essay on the logical foundation of survey sampling. Part one. In *Foundations of Statistical Inference*, Ed. V.P. Godambe and D.A.Sprott, New York:Holt, Rinehart and Winson.

[3] **Basu,D. and Ghosh, J.K. (1967):** Sufficient statistics in sampling from a finite universe. Bull. Int. Statist. Inst. **42** (2), 850-859.

[4] **Brecking, J. and Chambers, R.L.(1990):** Design adjusted estimation with longitudinal survey data. Canberra Statistical Tech. Rep. **001-90**.

[5] **Brewer, K.R.W. (1963):** Ratio estimation in finite populations: some results deducible from the assumptions of an underlying stochastic process. Australian J. Stat., **5**, 93 - 105.

[6] **Cassel, C.M., Sarndal, C.E. and Wretman, J.H. (1976):** Some results on generalized difference estimator and generalized regression estimator for finite population. Biometrics, **63**, 614-620.

[7] **Cassel, C.M., Sarndal, C.E. and Wretman, J.H. (1977):** *Foundations of Inference in Survey Sampling*, Wiley, N.Y.

[8] **Ericson,W.A. (1974):** A note on non-existence of minimum variance unbiased estimation for a finite population total. Sankhya, **38**, 181-184.

[9] **Godambe, V.P. (1955):** A unified theory of sampling from finite populations. J. Roy. Stat. Soc. **B 17**, 269-278.

[10] **Godambe, V.P. (1966):** A new approach to sampling from finite populations. I,II, J. Roy. Stat. Soc., **B 28**, 310-328.

[11] **Godambe, V.P. and Joshi, V.M. (1965):** Admissibility and Bayes estimation in sampling from finite populations I. Ann. Math. Statist., **36**, 1707- 1722.

[12] **Hanurav, T.V. (1966):** Some aspects of unified sampling theory. Sankhya A **28**, 175-203.

[13] **Hege, V.S. (1965):** Sampling dsigns which admit uniformly minimum variance unbiased estimates. Cal. Stat. Assoc. Bull., **14**, 160-162.

[14] **Lanke, J. (1974):** On UMV estimators in survey sampling. Metrika **20**, 196-202.

[15] **Mukhopadhyay, P. (1977):** Robust estimation of finite population total under certain linear regression models. Sankhya C, **39**, 71-87.

[16] **Mukhopadhyay, P. (1985 a):** Asymptotic properties of some BLU-estimators. Orissa Math. Science J. 4, 105-113.

[17] **Mukhopadhyay, P. (1985 b):** Estimation under linear regression models. Metrika **32**, 339-349.

[18] **Mukhopadhyay, P. (1996):** *Inferential Problems in Survey Sampling*, New Age International, New Delhi and London.

[19] **Mukhopadhyay, P. (1998):** *Theory and Methods of Survey Sampling* Prentice- Hall of India, New Delhi.

[20] **Mukhopadhyay, P. (2000):** *Topics in Survey Sampling*, L.N.Ser., 153, Springer Verlag, N.Y.

[21] **Royall,R.M. (1970):** On finite population sampling theory under certain linear regression models. Biometrika, **57**, 377-387.

[22] **Royall, R.M. (1976):** The linear least square prediction approach in twostage sampling. J. Amer. Stat. Assoc.,.**71**, 657-64.

[23] **Royall, R.M. and Herson, J.(1973):** Robust estimation in finite populations I. J. Amer. Stat. Assoc., **68**, 880-889.

[24] **Royall, R.M. and Pfe(r)ermann, D. (1982):** Balanced samples and robust Bayesian inference in finite population samling. Biometrika, **69**, 241- 249.

[25] **Sarndal,C.E. (1980 a):** Two model-based inference argument in survey sampling. Australian J. Stat., **22**, 314-318.

[26] **Sarndal, C.E. (1980 b):** On 1/4-inverse weighting versus best linear weighting in probability sampling. Biometrika, **67**(3), 639-650.

[27] **Sarndal, C.E. (1982):** Implications of survey designs for estimation of linear functions. J.Stat.Pl.Inf., **7**, 155-170.

[28] **Scott, A.J., Brewer, K., Ho, E.W.H. (1978):** Finite population sampling and robust estimation. J. Amer. Stat. Assoc., **73**, 359-361.

[29] **Wright, R.L. (1983):** Sampling designs with multivariate auxiliary information. J. Amer. Stat. Assoc., **78**, 879-884.