Jour. Ind. Soc. Ag. Statistics
51(2&3), 1998 : 315-330

Estimation of Parameters in Survey Sampling

V.P. Godambe
University of Waterloo, Ontario, Canada

SUMMARY

A methodology for estimation of two types of parameters, one of superpopulation and the other of survey population is discussed with some historical context.

Keywords: Estimating function, Gauss-Markoff theorem, Induced parameter, Optimality, Parameter, Population score function, Superpopulation model, Survey population.

1. Introduction

As is 'commonly understood' today, unlike in other areas of statistics in survey sampling we consider *two* types of parameters: enumerative and analytical. Today's understanding is but a gradual historical process.

Around what time did statisticians start thinking of *two* types of parameters, enumerative/survey population parameters on one hand and analytical/ superpopulation parameters on the other? The importance of the question just raised can be seen in the context of the *concept* of 'parameter' itself. This concept as distinct from the earlier 'frequency constants' is the basis of the modern theory of estimation and the consequent notions of likelihood, sufficiency, etc. (Stigler [29]).

It is illuminating to see historically, how the distinction between two types of parameters one of superpopulation (analytical) and the other of survey populations (enumerative) got established in statistics. That this distinction in the past was not often at all understood can be easily seen from the discussion that followed in The Royal Statistical Society after Neyman [27] read his important paper and also from some of the subsequent literature on this topic.

We initially note that by the third decade of this century the statisticians had developed a fruitful *hypothetical population model:* Here the population is supposed to be generated by outcomes (x) of hypothetical indefinite, independent repetitions of some chance experiment. In this population the frequency distribution of the variate x is completely specified up to an unknown constant called a *parameter* θ (Fisher [6], [8]). This unknown parameter could be estimated by drawing a *random sample* from the hypothetical population.

Now a 'random sample' simply means observations obtained by a number of independent repetitions of the chance experiment, which defined the hypothetical population. Note here randomization is axiomatic. For instance let the variate x in the hypothetical population be so distributed that the unknown parameter θ is the mean value, $\varepsilon(x) = \theta$. Further let the sample consists of n independent observations $x_1, \ldots, x_n$. Then the Gauss-Markoff theorem implies that the sample mean $\bar{x}$ is the linear unbiased minimum variance estimate of θ.

Now consider a survey population P on N individuals i, $P = \{i : i = 1, \ldots, N\}$. A variate x takes some specified values x_i for different individuals $i = 1, \ldots, N$. Further let a *hypothetical population* of x values be generated by indefinitely repeated, independent random draws, of individuals from P, with replacement. In this hypothetical population the mean value of x i.e. $\varepsilon(x) \equiv \theta$, equals the mean value of x for the survey population P namely $\bar{X} = \sum_i^N x_i / N$. In retrospect, it appears that Neyman's [27] optimal estimation based on the (Gauss-) Markoff theorem[1] was primarily for θ, the parameter of the hypothetical population just mentioned; it is related to the survey population only through the numerical equality, $\theta = \bar{X}$, *for simple random sampling design*, (Godambe [13]).

Such relationships as discussed above between the hypothetical population (later called super population) parameter θ and the survey population mean $\bar{X}$, would play a central role subsequently in this article.

Neyman's implicitly 'identifying' the hypothetical population parameter $\varepsilon(x) \equiv \theta$ with the survey population mean $\bar{X}$ proved to be a step in the right direction, for bringing survey sampling into the main stream of statistics. Now $\bar{X}$ obtained the theoretical status of a *parameter!*

Yet all was not quite well. Inadequacies, even contradictions, in Neyman's approach started becoming apparent as survey samplers began to use sampling designs more complex than stratified simple random sampling. The new designs depended more essentially than the earlier designs on the *individual labels*, a distinguishing feature of the survey populations. The hypothetical populations do not have these 'individuals' or their 'labels'.

---

1 Neyman [27] refers to the theorem as Markoff's theorem. In the discussion following Neyman's paper Fisher [7] pointed out that the theorem was originally due to Gauss.

For these complex designs, the (Gauss-) Markoff methodology of estimation looked very artificial and unintuitive. The 'individual labels' provided a far wider class of estimates for $\overline{X}$ than the class underlying the (Gauss-) Markoff methodology, (Horvitz and Thompson [25]). This methodology actually appeared to be of doubtful relevance for survey sampling for, in the class of all unbiased estimates for $\overline{X}$, *none* has uniformly minimum variance; a new criterion of optimal estimation provides a solution (Godambe [9]). This result essentially implied that no satisfactory solution of the problem of optimal estimation in survey sampling can be given without a reference to some underlying hypothetical population.

The *confusion* caused by Neyman's introduction of (Gauss-) Markoff methodology is vivid in the research works published during a couple of decades or so following his 1934 paper. However this confusion apparently does not seen to have affected the practice of some survey statisticians. They were guided by pragmatic considerations. For instance Duncan and Shelton [4] in relation to the works of Hansen and Hurwitz say, "They make great use of ratio estimates instead of best linear unbiased estimates."

Thus, though Neyman's invoking the (Gauss-) Markoff theorem in the context of survey sampling proved to be misleading for the *immediate* purpose of estimation, as said before, it was a conceptual advance to *parameterize* the survey population characteristics like mean, variance etc. Parametric inference had already been initiated by Fisher [6] and Neyman [27]. With its appropriate extensions, now we could investigate for the just mentioned 'survey population parameters', confidence intervals (Woodruff [30]), sufficiency (Basu [1]), likelihood function (Godambe [11]), fiducial distribution (Godambe [12], Godambe and Thompson [22]). Though much less direct than for the just referred to results, the above mentioned 'parameterization' had a deeper relevance for the 'optimal estimations' established by Godambe [9] and Godambe and Joshi [21].

Now all these developments subsequent to Neyman's paper clearly established the distinction between *two types* of parameters; one of survey populations and the other of hypothetical populations. Also, these developments, in retrospect seem to have provided the necessary apparatus for the much needed later extension of the theory of estimating functions (Godambe and Thompson [24]). This extension could deal with estimation problem of *both types* of parameters, jointly and separately.

A further clarification of the distinction between the two types of parameters came from survey samplers who were investigating, rather than mean values, the relationships such as causal or sociological between different variates, (cf. Deming [3], Kendall and Lazarsfeld [26]). Here underlying the survey population being sampled, was a well defined 'hypothetical population model'. The parameters of this model were to be estimated from the survey data. The present day concept of a *superpopulation* is a clearer formulation of this 'hypothetical population'.

A large area of applications is covered by the generalization given below.

The survey population is supposed to have been drawn as a random sample from a superpopulation. The problem is to estimate both or either of the parameters of the superpopulation and survey population, on the basis of the survey data and the design. In *analytical* surveys the emphasis is on estimating the superpopulation parameters while in *enumerative* surveys it is on estimating survey population parameters, (Godambe and Thompson [24], Godambe [9]).

In the following I shall briefly indicate how estimating functions provide a methodology to deal with both the above problems.

2. *Superpopulation Parameters*

Following commonly used notation let P denote the survey population of N labelled individuals i, $P = \{i : i = 1, \ldots, N\}, |P| = N$. With each individual i in P is associated a survey variate y_i and an auxiliary variate z_i; let $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_N)$ and $\mathbf{z} = (z_1, \ldots, z_i, \ldots, z_N)$. Both y_i and z_i $(i = 1, \ldots, N)$ can themselves be vectors. As is usually the case, we assume the auxiliary variate $\mathbf{z}$ to be completely known.

A useful generalization of a common superpopulation model is as follows. The joint (superpopulation) distribution f of $(\mathbf{y}, \mathbf{z})$ is assumed to be such that $(y_i, z_i), i = 1, \ldots, N$ are independent with density

$$f_i(y_i, z_i; \theta, \eta_i) = f_{2i}(z_i; \theta, \eta_i) f_{1i}(y_i | z_i; \theta) \quad (1)$$

$i = 1, \ldots, N$. Here θ is the *parameter of interest* and $\eta = (\eta_1, \ldots, \eta_N)$ is a *nuisance parameter*, θ can be a scaler or a vector. Note here the density f and the parameters θ, η define the *superpopulation model.*

With the above model, where the function f is *specified*, up to the unknown parameters θ, η one can study the relationships between the variates y and z in the superpopulation's distribution. For instance θ could be a regression parameter.

To estimate θ a sample s, i.e. a subset of P, $s \subset P$ is drawn with a sampling design p: p is a specified probability distribution on $S = \{s : s \subset P\}$. The sampling design p generally depends on the known auxiliary vector **z** i.e. $p \equiv p(\,.\,|\,\mathbf{z})$. Further the values y_i are observed for all individuals $i \in s$. Thus we are supposed to estimate θ given the model f, the sampling design p, the auxiliary vector **z** and the survey data

$$d = \{(i, y_i) : i \in s\}$$

The probability of the data d is given by

$$\text{Prob}\,(d;\theta,\eta) = \left\{\prod_{i=1}^{N} f_{2i}\,(z_i;\theta,\eta_i)\right\} p\,(s\,|\,\mathbf{z}) \left\{\prod_{i \in s} f_{1i}\,(y_i\,|\,z_i;\theta)\right\} \tag{2}$$

Now any function g (d, θ), with the indicated arguments is said to be an *estimating function* for θ. The dimensionality of g is the same as that of θ; for simplicity of presentation here we assume θ to be one-dimensional. Further the estimating function g is said to be *unbiased* if the expectation of g with respect to the distribution (2) is zero; $\varepsilon E\,(g) = 0$, for all θ, η. Here 'ε' stands for the expectation with respect to the model (1) and 'E' for the expectation with respect to the sampling design 'p' given above. The unbiased estimating functions g are standardized as $\left\{g \,/\, \varepsilon E\left(\frac{\partial g}{\partial \theta}\right)\right\}$ for comparisons of variances. In the class of all unbiased estimating functions g, g^* is said to be optimal if the variance of any standardized g is minimized for g^*

$$\varepsilon E\left\{g^* \,/\, \varepsilon E\left(\frac{\partial g^*}{\partial \theta}\right)\right\}^2 \leq \varepsilon E\left\{g \,/\, \varepsilon E\left(\frac{\partial g}{\partial \theta}\right)\right\}^2 \tag{3}$$

for all g, θ, η, (Godambe [10], Godambe and Thompson [23]). If g^* is an optimal estimating function, the 'solution' of the equation $g^* = 0$ is 'conventionally' called the *optimal estimate*. This convention is justified in terms of the shortest confidence intervals around the 'solution', provided by the distribution of g^*. (Godambe and Heyde [20]).

Under very general conditions, in Prob (d; θ, η) in (2) above, **z** is a complete sufficient statistic for the nuisance parameter η; (θ fixed). It then follows (Godambe [14]) that for estimating θ, in the class of all unbiased estimating functions g (d, θ), the *optimal* is given be the *conditional score*

$$g^*\,(d;\theta) = \sum_{i \in s} \frac{\partial \log f_{1i}\,(y_i\,|\,z_i;\theta)}{\partial \theta} \tag{4}$$

Thus the estimate of θ obtained by solving the equation $g^*\,(d,\theta) = 0$, for the given data d, does not depend on the sampling design p. This conclusion is in line with the model-based theory of survey sampling, (Royall [28]). However in the model theory the conclusion is drawn by conditioning Prob (d; θ, η) on the 'sample s' in *addition* to **z**. We condition only on **z**; it being the complete sufficient statistic for η and the ancillary for θ, (Godambe [14], [16]).

The difference between the two conditionings, the one on (s, **z**) and the other on **z** shows up in the confidence intervals: These, as indicated before, are obtained by inverting the distribution of the estimating function g^*. Now the var. $(g^* \mid s, \mathbf{z}) \neq$ var. $(g^* \mid \mathbf{z})$. The var. $(g^* / \mathbf{z})$ depends on the sampling design while var. $(g^* / s, \mathbf{z})$ does not. Our confidence intervals based on var. $(g^*/\mathbf{z})$ though in conflict with the model-based theory, are in line with the general statistical practice of using the ancillary statistics to assess the accuracy of the estimation. In the present context var. $(g^*/\mathbf{z})$ can be called the *design-effect*, though not altogether in the conventional sense.

In (4), if θ is the regression parameter then assuming normality we have,

$$\frac{\partial \log f_{1i}(y_i \mid z_i; \theta)}{\partial \theta} = \frac{(y_i - \theta z_i) z_i}{\sigma_i^2} \tag{5}$$

Now the 'efficiency' of an estimating function g is defined as the inverse of the variance of its standardized version, that is, the inverse of the right-hand side of the inequality (3). Hence the

$$\text{efficiency}\,(g^* \mid \mathbf{z}) = \sum_1^N (z_i^2 / \sigma_i^2)\, \pi_i$$

π_i being the design inclusion probabilities, $\pi_i = \sum_{s \ni i} p(s)$, $i = 1, \ldots, N$. For the sampling designs p having the expected sample size namely $\sum_1^N \pi_i$ fixed, it is easy to see that the efficiency $(g^*/\mathbf{z})$ is maximum for a purposive design p which selects a sample s for which $\sum_{i \in s} z_i^2 / \sigma_i^2$ is maximum with $p(s \mid \mathbf{z}) = 1$.

If z_i, $i = 1, \ldots, N$ are all equal then the purposive sample should consist of smallest σ_i^2. This is quite intuitive for an analytical survey, where the parameter to be estimated θ, is the parameter of superpopulation; we select the most accurate y's. The situation is quite different if we were to estimate a parameter of survey population (cf Godambe [15]). Some further comments in this respect are given in the next section.

In the literature a sampling design p is called *noninformative*, when it is fully specified by the auxiliary vector $\mathbf{z}$ and when $\mathbf{z}$ is completely known. With the additional condition that the distributions of $\mathbf{z}$ form a complete class, for each θ, the optimal estimating function g^* in (3) is independent of the sampling design p. Yet we hesitate to call the design 'p noninformative', for as stated above our confidence intervals essentially depend on p.

Now in (4) the optimal estimating function g^* depended on the entire form f_1 of the conditional distribution of y given z. Suppose one wants to avoid this dependence of g^* on f_1 and concentrate only on the relationship between y and z, implied by the distribution f_1; for instance $\varepsilon\,\{(y - \theta z) \mid z\} = 0$ or more generally $\varepsilon\,\{(y - \mu(\theta)) \mid z\} = 0$, μ being a specified function of θ. In such a case one can replace the 'conditional score function' given by (4) and (5) by the *quasi-score function* (Godambe [17]) and proceed as before.

The quasi-score function just mentioned above would still in general depend on the variance function of y which may not be known accurately enough. In the following Sections 3 and 4, we develop procedures of the estimations which do not much depend on the variance function of y : Here the strategy is *not* to estimate the 'superpopulation parameter' θ *directly*. Instead we concentrate on estimation of a *survey population parameter* θ_N which generally is not only numerically close to θ ($\theta \cong \theta_N$) but is a natural survey population version of it. We will call θ_N the parameter *induced* by θ. If the survey population size (N) is sufficiently large the induced parameter θ_N admits of very convenient approximations.

3. Survey Population Parameters

To illustrate the concept of an induced parameter let the individual score function in the superpopulation be given by (5). Further we define a function H_0 with arguments y, z and θ as

$$H_0(\mathbf{y}, \mathbf{z}; \theta) = \sum_{i=1}^{N} \frac{(y_i - \theta z_i)\, z_i}{\sigma_i^2}$$

That is H_0 is the survey population based, or, the census score function. If the survey population parameter θ_N^0 solves the equation

$$H_0(\mathbf{y}, \mathbf{z} : \theta_N^0) = 0$$

then by definition (Godambe and Thompson [24], Godambe [19]) θ_N^0 is the parameter *motivated* or *induced* by the superpopulation parameter θ. It is easy to see that $\theta_N^0 \to \theta$ in probability, as $N \to \infty$.

Now consider a survey population parameter θ_N obtained as the solution of the equation

$$H(\mathbf{y}, \mathbf{z}; \theta_N) = 0$$

where

$$H(\mathbf{y}, \mathbf{z}; \theta) = \sum_{i=1}^{N} (y_i - \theta z_i) z_i \tag{7}$$

Clearly for a sufficiently large survey population, i.e. N large, for large variations of the vector $\sigma = (\sigma_1, \ldots, \sigma_N)$ the survey population parameter

$$\theta_N \cong \theta_N^0 \cong \theta$$

with practical certainty. Further since 'θ_N^0' is the induced parameter, the survey population parameter θ_N could be called a parameter 'nearly' or 'approximately' induced by the superpopulation parameter θ.

Here we make three points: (i) In (7), the function H and the survey population parameter θ_N unlike H_0 and θ_N^0, are independent of the variances $\sigma_i^2, i = 1 \ldots, N$; (ii) Given the parameter θ and the auxiliary vector z, H is a linear function of θ_N; (iii) For every fixed θ, H and hence θ_N can be estimated *optimally* on the basis of the data d and the sampling design p.

Let as before the data $d \equiv \{(i, y_i) : i \in s\}$ and $h(d, \theta)$ be a 'design unbiased' estimating function for H in (7) i.e. $E(h - H) = 0$, where 'E' as before stands for the expectations with respect to the sampling design p. Now in the class of estimating functions $\{h : E(h - H) = 0\}$, the estimating function h^* satisfies the optimality criterion (3), if in the class, $\varepsilon E(h - H)^2$ is minimized for $h = h^*$, ε as before denoting the expectation with respect to the superpopulation model. In the present contest the optimal estimating function is given by h^* where

$$h^* = \sum_{i \in s} (y_i - \theta z_i) z_i / \pi_i \tag{8}$$

π_i as before being the design inclusion probabilities, $i = 1, \ldots, N$. This is a special case of a more general result given in (12) of Section 4, due to Godambe and Thompson [24].

Note that for every design unbiased estimating function h, $E\,h(d, \theta_N) = 0$ and for sufficiently large N, $\varepsilon E\,h^2(d, \theta_N)$ is minimized for $h = h^*$ in (8). Thus for the given data d, the estimating equation $h^* = 0$, provides *initially* an estimate for the survey population parameter θ_N. Further since for large N, θ_N is a survey population parameter (approximately) induced by θ ($\theta_N \cong \theta_N^0 \cong \theta$), the equation $h^* = 0$ also provides a *near optimal* estimation for the superpopulation parameter θ; 'near' for $\theta_N \cong \theta$. This estimation is *robust* in the sense of being *independent* of the variance function of y, in the superpopulation model.

Now the 'efficiency' of the estimating function h^* in (8) given by

$$\text{efficiency}\,(h^*) = \left\{ \sum_{i=1}^{N} \frac{\sigma_i^2 z_i^2}{\pi_i} \right\}^{-1} \tag{9}$$

where π_i, $i = 1, \ldots, N$ as before are the design inclusion probabilities. It is interesting to compare the efficiencies of g^* and h^* given by (6) and (9) respectively. For 'fixed expected sample size designs', (i.e. $\sum_1^N \pi_i$ fixed), as noted earlier the efficiency of g^* is maximized for a 'purposive' sampling design which selects sample s for which $\sum_{i \in s} z_i^2 / \sigma_i^2$ is maximum, with probability 1. On the other hand the efficiency of h^* in (9) is maximized for a sampling design for which $\pi_i \propto \sigma_i z_i$, $i = 1, \ldots, N$. A special case when z_i = constant, $i = 1, \ldots, N$ is illuminating. The efficiency of g^* is maximized by selecting a sample having individuals with smallest variances σ_i and hence providing most accurate observations y_i. This, as said before, is quite intuitive for an *analytical survey*. In contrast the efficiency of h^* is maximized for a sampling design providing a sample with most variable y-values. This again is intuitive for an *enumerative survey*, (Godambe and Thompson [24]).

As a variation of the problem, assuming in the superpopulation model (1) the variates y_i, $i = 1, \ldots, N$, are iid, we might try to estimate the mean value parameter $\varepsilon(y) = \mu$. Since $\mu = \mu(\theta, \eta)$, its maximum likelihood estimate $\hat{\mu} = \mu(\hat{\theta}, \hat{\eta})$, where $\hat{\theta}$ and $\hat{\eta}$ are the maximum likelihood estimates of θ and

η obtained from (2). Obviously this estimate $\hat{\mu}$ would be very much *model dependent*. Alternatively we may try to estimate the 'survey population parameter' μ_N induced by μ. This is given as the solution of the equation

$$\sum_{i=1}^{N} (y_i - \mu_N) = 0$$

To estimate μ_N, we use the regression relationship $\varepsilon\,\{(y_i - \theta\, z_i) \mid z_i\} = 0$ implying another (approximately) induced parameter θ_N; $\sum_1^N (y_i - \theta_N\, z_i) = 0$. Further as before θ_N is *optimally* estimated by $\hat{\theta}_N$ where $\left\{\sum_{i \in s} (y_i - \hat{\theta}_N\, z_i)/\pi_i\right\} = 0$. Thus we have $\hat{\mu}_N = \hat{\theta}_N \sum_1^N z_i / N$. Here the optimality of estimation is both 'conditional' on holding z fixed as well as 'unconditional' (Godambe [19]). This is important, for in the unconditional probability sense, $\mu_N \to \mu$, as $N \to \infty$.

4. *A General Method*

In the following we explain the theoretical generality and practical versatility of the 'optimal estimation' illustrated so far. Let y_i, $i = 1, \ldots, N$ be *independent* random variates drawn from a superpopulation model with a parameter θ. Note both θ and y_i $(i = 1, \ldots, N)$ can be vectors. Further for some specified real functions $\phi_i\,(y_i, \theta)$ of the indicated variables

$$\varepsilon\, \phi_i\,(y_i, \theta) = 0 \tag{10}$$

$i = 1, \ldots, N$; 'ε' denoting the expectation with respect to the superpopulation model.

For any fixed value of the parameter θ, consider the problem of estimating the function

$$H = \sum_{i=1}^{N} \phi_i\,(y_i, \theta) \tag{11}$$

Denoting as before the data $\{(i, y_i) : i \in s\}$ by d and the design expectation by E let B be the class of all design unbiased estimates $h(d, \theta)$, of H. That is

$$B = \{h : E\,(h - H) = 0, \forall \theta\}$$

Now the estimating function $h^* \in B$ satisfies the optimality criterion (3) if

$$\varepsilon E(h^* - H)^2 \leq \varepsilon E(h - H)^2, \forall h \in B, \forall \theta$$

A very general theorem (Godambe and Thompson [24], Godambe [19]) asserts that for any sampling design with inclusion probabilities π_i, $i = 1, \ldots, N$ all positive (> 0), the optimal estimating function is given by

$$h^* = \sum_{i \in s} \phi_i (y_i, \theta) / \pi_i \tag{12}$$

The practical flexibility of the optimality of h^* in (12) is made clear in the next paragraph.

For the semiparametric superpopulation model defined by (10) the survey population parameter θ_N^0 induced by 'θ' is given by the equation

$$\sum_{i=1}^{N} \phi_i (y_i, \theta_N^0) \left[\frac{\varepsilon \dfrac{\partial \phi_i}{\partial \theta}}{\varepsilon \phi_i^2} \right]_{\theta_N^0} = 0 \tag{13}$$

$\theta_N^0 \to \theta$ in probability granting some regularity conditions. Now suppose the parameter θ is a k-vector, $\theta = (\theta_1, \ldots, \theta_k)$. Here an 'approximately induced' parameter θ_N is given replacing the equations (13) by the equations

$$\sum_{1}^{N} \phi_i (y_i, \theta_N) = 0 \text{ and } \sum_{1}^{N} \phi_i (y_i, \theta_N) \left[\varepsilon \frac{\partial \phi_i}{\partial \theta_j} \right]_{\theta_N} = 0, \; j = 2, \ldots, k \tag{14}$$

Note generally, $\theta_N \cong \theta_N^0 \cong \theta$ for large N. Unlike the induced parameter θ_N^0, θ_N is *independent* of the variance functions $\varepsilon(\phi_i^2)$, which as said before, often are not known precisely enough. In (12) h^* denotes the optimal estimating function for H in (11). Now H in (11) is a real function. However in the general theorem (Godambe and Thompson [24]) H, can be a vector with real components. Suppose now h^* denotes the set of jointly optimum estimating functions for all the k functions

$$\sum_{1}^{N} \phi_i (y_i, \theta), \; \sum_{1}^{N} \phi_i (y_i, \theta) \, \varepsilon \left(\frac{\partial \phi_i}{\partial \theta_j} \right), \; j = 2, \ldots, k$$

underlying the equations (14). If the estimates $\hat{\theta}_N$ of θ_N and hence of θ are obtained by solving the equations $h^* = 0$ then generally, $\hat{\theta}_N$ are 'independent' of the variance functions $\varepsilon(\phi_i^2)$, $i = 1, \ldots, k$.

The above theory provides optimal or approximately optimal estimation, of the survey population totals (parameters)

$$\sum_{i=1}^{N} f_i(y_i)$$

for some arbitrary function of y_i namely $f_i(y_i)$, $i = 1, \ldots, N$. The only requirement is that the superpopulation expectation $\varepsilon f_i(y_i)$ be a known function of θ, say $\alpha_i(\theta)$, $i = 1, \ldots, N$.

Let in (10)

$$\phi_i(y_i, \theta) = f_i(y_i) - \alpha_i(\theta) \tag{15}$$

$i = 1, \ldots, N$. Then it follows from (14) that

$$\sum_{i=1}^{N} f_i(y_i) = \sum_{i=1}^{N} \alpha_i(\theta_N)$$

Further both sides of the equation are (nearly) *optimally* estimated by $\sum_1^N \alpha_i(\hat{\theta}_N)$.

To elaborate on some aspects of (near) optimal estimation just mentioned, let in (15),

$$\alpha_i(\theta) = \alpha(z_i, \theta) \tag{16}$$

where as before z is the auxiliary variate assumed known. Now the optimality of the estimation obtains both *conditionally* on holding the auxiliary variate z fixed and *unconditionally*. In the unconditional case, the sampling design, particularly the inclusion probabilities π_i, $i = 1, \ldots, N$, in (12) can be random variates as they generally are functions of z.

Now since the (near) optimality of estimation holds, also conditionally, the estimation itself has some kind of in-built conditioning. For instance it includes estimation based on what is generally known as *postsampling stratification* (Godambe [19], p. 237).

Alternatively, when the conditional model does not hold, we can work with an unconditional model where the auxiliary vector z is *not* held fixed. This can be best illustrated when the survey population P is stratified in strata P_j, $j = 1, \ldots, k$. Suppose the parameter $\theta = (\theta_1, \ldots, \theta_k)$. Now we replace the model (10), using (15) and (16) by

$$\varepsilon\,\{f_i(y_i) - \alpha(z_i, \theta_j)\} = 0, \quad i \in P_j, \ j = 1, \ldots, k \tag{17}$$

where ε denotes the superpopulation expectation *unconditional* on z; that is, z along with y is allowed to vary.

In the superpopulation model we assume (y_i, z_i) are iid for $i \in P_j$, $j = 1, \ldots, k$. In this case the induced parameters θ_N^0 given by (13) and the approximately induced parameter θ_N given by (14) are identical. They solve the equations

$$\sum_{i \in P_j} \{f_i(y_i) - \alpha(z_i, \theta_{jN})\} = 0, \qquad j = 1, \ldots, k \tag{18}$$

where $\theta_N = (\theta_{1N}, \ldots, \theta_{kN})$.

Let h^* denote the set of jointly optimum estimating function for the k functions

$$\sum_{i \in P_j} \{f_i(y_i) - \alpha(z_i, \theta_j)\}, \quad j = 1, \ldots, k$$

underlying the equations (18). Then the solution $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ of the equations $h^* = 0$ provide jointly optimal estimation for θ_N and consequently for θ. Hence $\sum_1^N f_i(y_i)$ is estimated by

$$\sum_{j=1}^{k} \sum_{i \in P_j} \alpha(z_i, \hat{\theta}_j) \tag{19}$$

It is easy to see that as a special case when $f_i(y_i)$, $\alpha(z_i, \theta_j) = z_i\,\theta_j$, the estimate (19) reduces to the usual ratio estimate.

Finally we briefly mention some common problems of estimation that can be treated within the framework of optimal estimating functions:

(i) Domain estimation. Here we put in (15) $f_i(y_i) = y_i$, for individuals i in the domain otherwise $f_i(y_i) = 0$.

(ii) Estimation of the change of total on two successive occasions. Now we let y_i be a vector $y_i = (y_i^0, y_i')$, y_i^0 and y_i' denoting values of y on two different occasions. The estimation to the 'change' is achieved by putting in (15), $f_i(y_i) = y_i^0 - y_i'$.

(iii) Estimation under within cluster correlations. As is (ii) we assume y_i to be the vector of all y's in a cluster and let $f_i(y_i)$ to be the cluster total.

(iv) In some practical situations the sampling design though fully specified (known) must be considered to have come from a distribution depending on the superpopulation parameter of interest. This situation is covered by the unconditional optimal estimation discussed in the foregoing. For more details see Godambe [18].

ACKNOWLEDGEMENT

I am grateful to M.P. Singh for his encouraging the historical investigation underlying the discussion in Section 1. Thanks are also due to K.R. Shah for important comments on the first draft of the paper. Two very interesting historical articles though not mentioned in the text, which provided valuable insights are by Bellhouse [2] and by Fienberg and Tanur [5].

REFERENCES

[1] Basu, D., 1958. On sampling with replacement. *Sankhya*, **20**, 287-294.

[2] Bellhouse, D.R., 1988. A brief history of random sampling methods. *Handbook of Statistics*, **6**, (Eds. P.R. Krishnaiah and C.R. Rao), North-Holland, Amsterdam, 1-14.

[3] Deming, W.E., 1950. *Some Theory of Sampling*. John Wiley and Sons Inc., New York. Chapman & Hall, London.

[4] Duncan, J.W. and Shelton, W.C.,1992. U.S. government contributions to probability sampling and statistical analysis. *Statistical Science*, **7**, 320-328.

[5] Fienberg, S.E. And Tanur, J.M., 1996. Reconsidering the fundamental contributions of Fisher and Neyman on experimentation and sampling. *International Statistical Review*, **64**, 237-253.

[6] Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, London, **A222**, 309-368.

[7] Fisher, R.A., 1934. Discussion of Neyman, J., On two different aspects of representative method: the method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.*, **97**, 616.

[8] Fisher, R.A., 1956. *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.

[9] Godambe, V.P., 1955. A unified theory of sampling from finite populations. *J. Roy. Statist. Soc.*, **B17**, 369-378.

[10] Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208-1212.

[11] Godambe, V.P., 1966. A new approach to sampling from finite populations. *J. Roy. Statist. Soc.*, **B28**, 310-328.

[12] Godambe, V.P., 1969. A fiducial argument with applications to survey-sampling. *J. Roy. Statist. Soc.*, **B31**, 246-260.

[13] Godambe, V.P., 1976. A historical perspective of recent developments in the theory of sampling from actual populations. *J. Ind. Soc. Agril. Statist.*, **28**, 1-12.

[14] Godambe, V.P., 1976a. Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, **63**, 277-284.

[15] Godambe, V.P., 1982. Estimation in survey sampling: robustness and optimality (with discussion). *J. Amer. Statist. Assoc.*, **77**, 393-406.

[16] Godambe, V.P., 1984. On ancillarity and Fisher information in the presence of a nuisance parameter. *Biometrika*, **71**, 626-629.

[17] Godambe, V.P., 1985. The foundations of finite sample estimation in stochastic processes. *Biometrika*, **72**, 419-428.

[18] Godambe, V.P., 1993. Inference in analytical surveys. *Sankhya*, **55**, 343-352.

[19] Godambe, V.P., 1995. Estimation of parameters in survey sampling: optimality. *Canadian Journal of Statistics*, **23**, 227-343.

[20] Godambe, V.P. and Heyde, C.C., 1987. Quasi-likelihood and optimal estimation. *International Statistical Review*, **55**, 231-244.

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

[22] Godambe, V.P. and Thompson, M.E., 1971. Bayes, fiducial and frequency aspects of statistical inference in regression analysis in survey-sampling (with discussion). *J. Roy. Statist. Soc.*, **B33**, 361-390.

[23] Godambe, V.P. and Thompson, M.E., 1974. Estimating equations in presence of a nuisance parameter. *Annals of Statistics*, **2**, 568-571.

[24] Godambe, V.P. and Thompson, M.E., 1986. Parameters of superpopulation and survey population: their relationships and estimation. *International Statistical Review*, **54**, 127-138.

[25] Horvitz, D.G. and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47**, 663-685.

[26] Kendall, P.L. and Lazarsfeld, P.F., 1950. Problems of survey analysis. In: Continuities in Social Research : Studies in the Scope and Method of *'The American soldier'* eds. R.K. Merton and P.F. Lazarsfeld. Free Press, Chicago.

[27] Neyman, J., 1934. On two defferent aspects of representative method: the method of stratified sampling and the method of purposive selection (with discussion). *J. Roy. Statist. Soc.*, **97**, 558-652.

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

[29] Stigler, S.M., 1976. Discussion of Savage, L.J., On rereading R.A. Fisher. *Annals of Statistics*, **4**, 441-500.

[30] Woodruff, R.S., 1952. Confidence intervals for medians and other position measures. *J. Amer. Statist. Assoc.*, **47**, 635-646.