Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 268-281

UTILIZING SURVEY FRAMEWORK IN SCIENTIFIC INVESTIGATIONS

By V.P. GODAMBE
University of Waterloo, Canada

SUMMARY. Statistical investigation of any scientific question such as the effect of a treatment on a disease or the effect of traffic intensity on pollution starts generally with an assumption of a probabilistic parametric model. The estimation of these parameters from the data is supposed to suggest an answer to the question under investigation. Now underlying the model just mentioned is the concept of a 'hypothetical population' of results (observations) obtained from a large number of independent repetitions of a chance experiment. However in many situations, 'implicit' with the 'hypothetical population' there is an 'actual' population of individuals such as humans, households, towns. This actual population is called a 'survey population', for the ultimate 'data' is obtained by observations on the individuals sampled from this population. Yet in many statistical investigations (biostatistics, ecology), a general tendency is to ignore, for simplicity of analysis, the underlying survey population. Much is lost by so doing. For taking into account 'explicitly' the survey sampling aspect of the data collection can often enhance the efficiency of estimation. This would be demonstrated in connection with weighted distributions as they arise in some common ecological setting. (The problem within a biostatistical setting was discussed previously by Godambe & Vijayan, 1996). This note, though self contained, is based on and is a continuation of the paper by Godambe and Rajarshi (1989). Here we further emphasize the contribution to the efficiency of estimation in situations where along with the hypothetical population model a corresponding survey population framework is available.

1. Introduction

In spite of different approaches to statistical inference such as Bayesian, non-Bayesian with different combinations, emphasis and interpretations, it can be said that underlying common statistical practice is an assumption

Paper received October 2000; revised August 2001.
AMS (2000) subject classification. Primary 62G05; secondary 62P10.
Keywords and phrases. Estimating function, hypothetical population, maximum likelihood estimation; survey population, weighted distribution.

of a *hypothetical population model*: A population of results, observations or variates generated by a large (infinite) number of *independent* repetitions of a chance experiment like spinning of a roulette, throwing a die or measurement on an instrument. Without such a model, it would be difficult to interpret the common use of the law of large numbers or the central limit theorem, (cf. Godambe, 1976). In this hypothetical population the variate x say, is supposed to have a known density $f(x)$ or a partially known density that is known up to some unknown parameter θ, $f(x;\theta)$. It follows that the density of a *sample* of size n, $x_1, ..., x_n$ is given by

$$\prod_{i=1}^{n} f(x_i;\theta). \tag{1}$$

Generally a statistical investigation of a *scientific question*, such as response of a disease to a treatment or relationship of pollution to traffic intensity, starts with a 'hypothetical population model' just mentioned. The estimation of the unknown parameter θ, based on (1), is supposed to lead to an answer of the 'scientific question'.

The evolution of the above hypothetical population model is associated over a long period of history with the development of statistical theory and practice itself. Now in some situations, with hypothetical population model is *implicit*, some *actual* population of individuals like humans or households or towns. For instance in some scientific investigation, in biostatistics a hypothetical population may be of the responses (disease conditions) obtained as a result of some treatment; here the actual population consists of the individuals to whom the treatment could be given. Or for instance, in ecology the hypothetical population may be that of pollution level obtained as a result of the intensity of traffic; here the actual population consists of towns where the pollution level and traffic intensity are recorded. In both these instances the samples are drawn from actual populations. In the former case the population was of individuals with disease condition while in the latter it was population of towns. The development of a model for sampling and inference in relation to 'actual populations', alternatively called *survey populations* is much more recent, than that of the hypothetical population model, mentioned earlier. The development of the survey sampling model was gradual. From the latter model follow some results that are deeply in conflict with the corresponding results following from the former (Godambe, 1955, 1976). Yet in some scientific investigations, as indicated above, a hypothetical population model has 'implicit' in it, a survey population. In these situations, a general tendency in statistical practice (biostatistics, ecology

etc.), is to ignore the survey population aspect of the model, for simplicity of analysis. (Of course here important exceptions are some analytical surveys.) Much is lost by so doing. In fact when the 'survey sampling aspect' of the model is 'explicitly' taken into account, the efficiency of estimation is often enhanced considerably. This is demonstrated in the sections to follow. Actually in Illustration 5.1, the estimating function which incorporates the survey sampling aspect provides satisfactory estimation; while ignoring this aspect leads to vacuous estimating function providing no estimation at all. Similarly in Illustration 5.2, the estimating function utilizing survey sampling aspects provides reasonable estimation of both the parameters while the one ignoring the aspect, is vacuous for one parameter. This thus sheds a new light on the old controversy 'Do individual labels of a survey population themselves carry any information useful for estimation?' (Godambe, 1975). The context, here is weighted distributions as they arise in some common ecological sampling. Weighted distributions also arise in biostatistical sampling (cf. Godambe and Vijayan, 1996) and other areas, but with different structures.

2. Notation

The formalism below is introduced to provide a very simplified picture of how weighted distributions are obtained in ecological sampling. For illustrations see Section 5.

Let a survey population $\mathcal{P}$, be of N labeled individuals i, $\mathcal{P} = \{i : \ i = 1, 2, ..., N\}$. With each individual i is associated a random (response) variate y_i with distributions given by $f_i(y_i; \boldsymbol{\theta})$ which is completely specified up to an unknown parameter, possibly vector valued, $\boldsymbol{\theta}$. If $\boldsymbol{y} = (y_1, y_2, ..., y_N)$, the probability density of $\boldsymbol{y}$ is given by

$$\prod_{1}^{N} f_i(y_i; \boldsymbol{\theta}). \tag{2}$$

One situation in which response dependent sampling or a weighted distribution arises in an ecological setting, is from a postulated experiment as follows.

Experiment 1. After the variate $\boldsymbol{y}$ is *realized* (to be distinguished from *observed*) from the distribution (2), for each individual $i \in \mathcal{P}$, nature performs a Bernoulli experiment with chance of success $w_i(y_i)$, the function w_i

being completely specified. Only if the experiment results in a 'success', (i, y_i), is included in the data

$$d = \{(i, y_i) : \ i \in s\}, \ s \subseteq \mathcal{P}; \tag{3}$$

the subset s is also called a 'sample' in survey sampling.

REMARK 1. Throughout, we assume that the chance functions $w_i(y_i)$ in the Experiment 1 are independent of $\boldsymbol{\theta}$.

Now, denoting generally 'probability density' or 'probability' under $\boldsymbol{\theta}$ by $Prob(\cdot; \boldsymbol{\theta})$, we have

$$Prob(d; \boldsymbol{\theta}) = \int Prob(s, \boldsymbol{y}; \boldsymbol{\theta}) \prod_{i \notin s} dy_i.$$

If further $Prob(\cdot|\cdot)$ denotes the conditional probability density, from (2), we have

$$Prob(d; \boldsymbol{\theta}) = \int Prob(s|\boldsymbol{y}; \boldsymbol{\theta}) \prod_{i=1}^{N} f_i(y_i; \boldsymbol{\theta}) \prod_{i \notin s} dy_i.$$

Now, since in the Experiment 1,

$$Prob(s|\boldsymbol{y}; \boldsymbol{\theta}) = \prod_{i \in s} w_i(y_i) \prod_{i \notin s} (1 - w_i(y_i)),$$

we have

$$Prob(d; \boldsymbol{\theta}) = \prod_{i \in s} \{f_i(y_i; \boldsymbol{\theta}) w_i(y_i) / w_{i0}(\boldsymbol{\theta})\} \prod_{i \in s} w_{i0}(\boldsymbol{\theta}) \prod_{i \notin s} \{1 - w_{i0}(\boldsymbol{\theta})\} \tag{4}$$

where

$$w_{i0}(\boldsymbol{\theta}) = \int w_i(y_i) f_i(y_i; \boldsymbol{\theta}) dy_i, \quad i = 1, 2, ..., N. \tag{5}$$

In the following section, we consider maximum likelihood estimation of the unknown parameters $\boldsymbol{\theta}$ in (2) on the basis of the data d in (3).

REMARK 2. It is important to note that though our entire discussion is within the set-up of a finite population $\mathcal{P}$ above, we will, in general *not* study estimation of the parameters of $\mathcal{P}$, like functions $t(\boldsymbol{y})$. We restrict ourselves to the estimation of the parameter $\boldsymbol{\theta}$ of the distribution in (2), the only exception being estimation of the size of $\mathcal{P}$, $|\mathcal{P}| = N$.

3. Maximum Likelihood (ML) Estimation

For the parametric model (2), the ml equations for $\boldsymbol{\theta}$, based on the data d are obtained as usual from (4). Denoting $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$, where θ_j's are scalars, we have, for $j = 1, 2, ..., k$,

$$\frac{\partial}{\partial \theta_j} \log Prob(d; \boldsymbol{\theta}) = \sum_{i \in s} \frac{\partial}{\partial \theta_j} \log \left(\frac{f_i w_i}{w_{i0}} \right) + \sum_{i \in s} \frac{1}{w_{i0}} \frac{\partial w_{i0}}{\partial \theta_j} - \sum_{i \notin s} \frac{1}{1 - w_{i0}} \frac{\partial w_{i0}}{\partial \theta_j} = 0. \tag{6}$$

An interesting special case is obtained when in (2) and (4), $f_i = f$, $w_i = w$ and $w_{i0} = w_0$. Denoting $|s|$ by n, we have, from (6)

$$\sum_{i \in s} \frac{\partial}{\partial \theta_j} \log \left(\frac{f(y_i; \boldsymbol{\theta}) w(y_i)}{w_0(\boldsymbol{\theta})} \right) + \frac{n - w_0(\boldsymbol{\theta}) N}{w_0(\boldsymbol{\theta})(1 - w_0(\boldsymbol{\theta}))} \frac{\partial w_0}{\partial \theta_j} = 0; \quad j = 1, 2, ..., k. \tag{7}$$

It is important to note that the ml equations (7), even when the population size $|\mathcal{P}| = N$ is unknown, do *not* reduce to the equations

$$\sum_{i \in s} \frac{\partial}{\partial \theta_j} \log \left(\frac{f(y_i; \boldsymbol{\theta}) w(y_i)}{w_0(\boldsymbol{\theta})} \right) = 0; \quad j = 1, 2, ..., k, \tag{8}$$

which are usually found in the literature on weighted distributions (cf. Fisher, 1934; Rao, 1965). The equations (8) are of the type of equations (1) which, can be obtained, as said before from the hypothetical population model; now the $n = |s|$ variates y_i are iid with the common density (fw/w_0). The equations (8) ignore the survey population $\mathcal{P}$ introduced at the beginning of Section 2. Particularly the l.h.s. of the equation (8) should have the additional term

$$\frac{\partial \log Prob(s; \theta)}{\partial \theta_j}.$$

The distinction of estimation arising out of the equations (7) and (8) respectively, is the *main topic* of discussion of this paper. Actually the 'topic' deals more with the estimating functions to the l.h.s. of the equations rather than the equations themselves.

Here we note that the ml estimate of N obtained from the distribution (4) of the *full data*, $\hat{N} = n$. On the other hand, the natural estimate of N, when θ is known, is given by $[n/w_0(\boldsymbol{\theta})]$. This natural estimate is the ml estimate but with respect to the marginal distribution of only a *part* data $|s| = n$ viz., $Prob(n|N, \boldsymbol{\theta})$, (Feldman and Fox, 1968). Thus, the conventional ml theory does *not* provide equations (8) and $n - w_0 N = 0$ for joint estimation of $\boldsymbol{\theta}$

and N. The weakness of ml estimation, just described, is corrected with an *extension* of ml estimation provided by the theory of estimating functions (Godambe and Thompson, 1989). This theory provides equation (7) and the equation $n - w_0(\boldsymbol{\theta})N = 0$ as jointly optimal for $\boldsymbol{\theta}$ and N, implying equation (8), which is in common use.

For our main topic of discussion it is not necessary to go into the details of the extension of ml estimation mentioned in the preceding paragraph. It is enough to note an underlying result. In the present context, the 'Experiment 1' of Section 2 is statistically equivalent to the following Experiment 2. We note here that, in (4), $(f_i w_i / w_{i0})$ due to (5), is a probability density function.

EXPERIMENT 2. With each individual $i \in \mathcal{P}$, nature performs a Bernoulli experiment with the chance of 'success' $w_{i0}(\boldsymbol{\theta})$ given by (5). Only if the Bernoulli experiment results in a success, a value y_i is drawn from the distribution $(f_i w_i / w_{i0})$; (i, y_i) then becomes a part of statistician's data.

THEOREM 1. Experiment 2 above is mathematically equivalent to Experiment 1 of Section 2.

PROOF. A generic outcome in both, Experiment 1 and Experiment 2 is d of (3). Further, the probability density of d in both the experiments is given by (4) and (5).

4. **Applications**

In the following we will assume the parameter θ, in (2) to be scalar. Further we modify the Remark 1 of Section 2 extending the scope of Experiment 1:

REMARK 3. The chance functions $w_i(y_i)$, though, as in Remark 1, do not depend on θ, they may depend on an additional scalar (unknown) *nuisance* parameter λ. Hence chance functions are

$$w_i(y_i; \lambda), \quad i = 1, ..., N. \tag{9}$$

Though θ is mainly the parameter of *interest*, on the basis of data d in (3), generally we would want to estimate both the parameters θ and λ. Note, even with the additional parameter λ, the Theorem 1 of Section 3 establishing the equivalence of Experiment 1 and Experiment 2 remains valid.

Now with the advent of Experiment 2, we can define for each individual $i \in \mathcal{P}$, a variate ϕ_i, such that $\phi_i = 1$ if the Bernoulli experiment for the

individual i results into a 'success' otherwise $\phi_i = 0$, $i = 1, ..., N$. Thus the two ml equations for the parameters θ and λ, corresponding to (6) can now be written as $g_{1\theta} = 0$ and $g_{1\lambda} = 0$ where

$$\begin{aligned} g_{1\theta} &= \sum_{i=1}^{N} \left[\phi_i \left\{ \frac{\partial}{\partial \theta} \log \left(\frac{f_i w_i}{w_{i0}} \right) + \frac{\partial w_{i0}/\partial \theta}{w_{i0}(1 - w_{i0})} \right\} - \frac{\partial w_{i0}/\partial \theta}{(1 - w_{i0})} \right], \\ g_{1\lambda} &= \sum_{i=1}^{N} \left[\phi_i \left\{ \frac{\partial}{\partial \lambda} \log \left(\frac{f_i w_i}{w_{i0}} \right) + \frac{\partial w_{i0}/\partial \lambda}{w_{io}(1 - w_{i0})} \right\} - \frac{\partial w_{i0}/\partial \lambda}{(1 - w_{i0})} \right]. \end{aligned} \tag{10}$$

Similarly the two estimating equations corresponding to (8) can now be written as $g_{2\theta} = 0$ and $g_{2\lambda} = 0$ where

$$\begin{aligned} g_{2\theta} &= \sum_{i=1}^{N} \phi_i \frac{\partial}{\partial \theta} \log \left(\frac{f_i w_i}{w_{i0}} \right), \\ g_{2\lambda} &= \sum_{i=1}^{N} \phi_i \frac{\partial}{\partial \lambda} \log \left(\frac{f_i w_i}{w_{i0}} \right). \end{aligned} \tag{11}$$

REMARK 4. When as a special case, in (2) and (9), for all $i = 1, ..., N$, $f_i = f$, $w_i = w$ and hence $w_{i0} = w_0$, the estimating functions g in (10) and (11) consist of N iid random variates with mean '0'.

Now when, for $i = 1, ..., N$, $f_i = f$, $w_i = w$, $w_{i0} = w_0$, putting $n = |s|$, the functions g in (10) and (11), simplify as follows:

$$\begin{aligned} g_{1\theta} &= \sum_{i \in s} \frac{\partial \log f(y_i; \theta)}{\partial \theta} - \frac{(N - n)}{(1 - w_0)} \frac{\partial w_0}{\partial \theta}, \\ g_{1\lambda} &= \sum_{i \in s} \frac{\partial \log w(y_i; \lambda)}{\partial \lambda} - \frac{(N - n)}{(1 - w_0)} \frac{\partial w_0}{\partial \lambda}; \end{aligned} \tag{12}$$

$$\begin{aligned} g_{2\theta} &= \sum_{i \in s} \frac{\partial \log f(y_i; \theta)}{\partial \theta} - \frac{n}{w_0} \frac{\partial w_0}{\partial \theta}, \\ g_{2\lambda} &= \sum_{i \in s} \frac{\partial \log w(y_i; \lambda)}{\partial \lambda} - \frac{n}{w_0} \frac{\partial w_0}{\partial \lambda}. \end{aligned} \tag{13}$$

Using an obvious abbreviation for estimating functions in (12) and (13) namely $g_1 \equiv (g_{1\theta}, g_{1\lambda})$ and $g_2 \equiv (g_{2\theta}, g_{2\lambda})$ we note the following: (I) At the

expected value of the random variate n, namely $n = Nw_0$, the estimating functions g_1 and g_2 are equal, $g_1 \equiv g_2$. This suggests that in limit, $N \to \infty$, the estimations based on the equations $g_1 = 0$ and $g_2 = 0$ would tend to be identical. Yet in the illustrations to follow, the difference between the two persists for N as large as 1000! (II) Again, as said in Section 3, if the size of the survey population $|\mathcal{P}| = N$ is unknown and to be estimated from the data d in (3), the distinction between the estimations based on the equations $g_1 = 0$ and $g_2 = 0$ disappears. (III) In an event of rare probability when a sample s with size $|s| = n = N$, is drawn, that is when the sample size is equal to the population size $|\mathcal{P}| = N$, the estimation based on the equations $g_1 = 0$ seems logical while that based on the equations $g_2 = 0$ does not.

The above remarks are in general about point estimation; of parameters θ and λ. However, based on each estimating function, g_1 in (12) and g_2 in (13) we can construct separate confidence intervals for the parameter of interest θ. For such a construction, following Remark 4, in case of the estimating function g_1, one can use either of the (asymptotically) standard normal pivots t_1 or t_1':

$$t_1 = [g_{1\theta}/(-\partial g_{1\theta}/\partial\theta)^{1/2}]_{\hat{\lambda}} \quad \text{and} \quad t_1' = [g_{1\theta}/(\text{variance } g_{1\theta})^{1/2}]_{\hat{\lambda}} \tag{14}$$

where $[\ \ ]_{\hat{\lambda}}$ denotes the value of the bracketed expression at the estimated value $\hat{\lambda}$ of the nuisance parameter λ. Similarly in the case of the estimating function g_2, the pivots are t_2 or t_2':

$$t_2 = [g_{2\theta}/(-\partial g_{2\theta}/\partial\theta)^{1/2}]_{\hat{\lambda}} \quad \text{and} \quad t_2' = [g_{2\theta}/(\text{variance } g_{2\theta})^{1/2}]_{\hat{\lambda}}. \tag{15}$$

The superiority of the estimating functions g_1 in (12) that is the ones which take the underlying survey population into account over the estimating functions g_2 in (13), that is the ones which are traditional based on the hypothetical population model (cf. as in (1)), generally ignoring the underlying survey population, is indicated by the illustrations to follow: When weighted sampling is done with some chance function as in (9), often the estimating equations $g_2 = 0$ have no solutions at all, within the prescribed parametric ranges, while estimating functions g_1 provide satisfactory point and interval estimation.

5. Illustrations

ILLUSTRATION 5.1. This is based on Cook and Martin (1974). An aerial survey of animal populations attempts to record as many groups of animals

as possible. The survey purports to estimate the total number of animals of a species in a given area. Let N be the number of groups in a given quadrat and let y_i be the number of animals in the i-th group, $i = 1, 2, ..., N$. The entire group of animals is observed if any member of the group is observed. Cook and Martin (1974) assume that all animals in a group are observed independently of each other. If λ denotes the probability that an animal is observed, we have, as in (9) the chance function

$$\begin{aligned} w_i &= w_i(y_i, \lambda) = w(y_i, \lambda) \\ &= 1 - (1 - \lambda)^{y_i}, \end{aligned} \tag{16}$$

$y_i = 1, 2, ...$: $i = 1, 2, ..., N$. Further, in terms of our notation in (2), Cook and Martin (1974) assume that for $i = 1, 2, ..., N$,

$$f_i(y_i; \boldsymbol{\theta}) = f(y_i, \theta) = e^{-\theta} \frac{\theta^{y_i - 1}}{(y_i - 1)!} \tag{17}$$

$y_i = 1, 2, ...,$: θ being a scalar parameter. Now in (5), for $i = 1, 2, ..., N$,

$$w_{i0} = w_0 = 1 - e^{-\theta\lambda}(1 - \lambda). \tag{18}$$

Tables 5.1.1 and 5.1.2 below are each based on a fixed survey population of size $N = 1000$; the 1000 y-values are drawn independently from the distribution f in (17) for the indicated (true) value of θ. Column 1 of the Table, numbers the samples 1 to 10. These are drawn from the just mentioned survey population using chance function w_i given by (16) for the indicated (true) value of λ, (see Experiment 1 of Section 2, and (9)). The column (3) provides the mean value of the sample. The estimates $\hat{\theta}$ and $\hat{\lambda}$ given in columns (4) and (5) are obtained from the equations $g_{1\theta} = 0$, $g_{1\lambda} = 0$ based on (12). Similarly the confidence intervals in columns (6) and (7) are constructed using the approximate pivotals t_1 and t_1' in (14). Specifically in (t_1'), (Variance $g_{1\theta}$) $= E\{\text{Var}(g_{1\theta}|s)\} + \text{Var}\{E(g_{1\theta}|s)\}$, which after some simplifications yield

$$(\text{Variance } g_{1\theta}) = \frac{N}{\theta}\{1 - e^{-\lambda\theta}(1 - \lambda)^2\}. \tag{19}$$

Now since for the distribution f in (17) the mean value of y is $(1 + \theta)$, each of the (ten) samples in the Tables provide an estimate $(1 + \hat{\theta})$ for the corresponding (true) population mean $\overline{Y}$. These estimates and $\hat{\theta}$ and $\hat{\lambda}$ and the confidence intervals for θ, compare *very well* against the true values given in the Tables. Note only for one sample out of 20, the interval does not cover

the true value. On the other hand in the present situation, (i.e. for f as in (17), w_i as in (16) and θ, λ as in the Tables) the estimation given by the equations $g_{2\theta} = 0$, $g_{2\lambda} = 0$ based on (13) is *vacuous*, without solutions.

TABLE 5.1.1. POINT ESTIMATES FOR (θ, λ) AND CONFIDENCE INTERVALS (CI) FOR θ; $N = 1000$. THE TRUE VALUES ARE $\theta = 3$, $\lambda = 0.01$, $w_0(\theta, \lambda) = 0.0393$, AND $\overline{Y} = 3.9590$

(1) sample	(2) n	(3) $\bar{y}$	(4) $\hat{\theta}$	(5) $\hat{\lambda}$	(6) CI of 95% Coverage Probability based on t_1 in (14)	(7) CI of 95% Coverage Probability based on t_1' in (14) & (19)
1	41	4.9024	3.1595	0.0101	$2.6600 \leq \theta \leq 3.6591$	$2.7162 \leq \theta \leq 3.7308$
2	42	4.8095	3.0721	0.0105	$2.5864 \leq \theta \leq 3.5579$	$2.6408 \leq \theta \leq 3.6272$
3	33	4.6970	2.9626	0.0085	$2.4262 \leq \theta \leq 3.4990$	$2.4938 \leq \theta \leq 3.5882$
4	42	4.5000	2.7812	0.0113	$2.3224 \leq \theta \leq 3.2400$	$2.3752 \leq \theta \leq 3.3075$
5	47	4.7234	2.9929	0.0120	$2.5404 \leq \theta \leq 3.4453$	$2.5889 \leq \theta \leq 3.5063$
6	45	4.3556	2.6475	0.0126	$2.2166 \leq \theta \leq 3.0784$	$2.2652 \leq \theta \leq 3.1402$
7	45	4.4889	2.7720	0.0122	$2.3295 \leq \theta \leq 3.2144$	$2.3788 \leq \theta \leq 3.2770$
8	39	4.6154	2.8882	0.0102	$2.4018 \leq \theta \leq 3.3747$	$2.4590 \leq \theta \leq 3.4486$
9	28	5.1071	3.3486	0.0065	$2.7240 \leq \theta \leq 3.9731$	$2.8061 \leq \theta \leq 4.0831$
10	35	4.6571	2.9259	0.0091	$2.4087 \leq \theta \leq 3.4431$	$2.4724 \leq \theta \leq 3.5265$

TABLE 5.1.2. POINT ESTIMATES FOR (θ, λ) AND CONFIDENCE INTERVALS (CI) FOR θ; $N = 1000$. THE TRUE VALUES ARE $\theta = 5$, $\lambda = 0.01$, $w_0(\theta, \lambda) = 0.0583$, AND $\overline{Y} = 5.9880$

(1) sample	(2) n	(3) $\bar{y}$	(4) $\hat{\theta}$	(5) $\hat{\lambda}$	(6) CI of 95% Coverage Probability based on t_1 in (14)	(7) CI of 95% Coverage Probability based on t_1' in (14) & (19)
1	63	6.6190	4.8183	0.0112	$4.3062 \leq \theta \leq 5.3305$	$4.3482 \leq \theta \leq 5.3806$
2	73	6.9589	5.1536	0.0123	$4.6594 \leq \theta \leq 5.6477$	$4.6965 \leq \theta \leq 5.6911$
3	68	6.9559	5.1484	0.0114	$4.6368 \leq \theta \leq 5.6601$	$4.6765 \leq \theta \leq 5.7068$
4	59	6.2203	4.4299	0.0112	$3.9250 \leq \theta \leq 4.9347^{\otimes}$	$3.9689 \leq \theta \leq 4.9874^{\otimes}$
5	49	6.4898	4.6868	0.0088	$4.1153 \leq \theta \leq 5.2583$	$4.1685 \leq \theta \leq 5.3233$
6	69	6.7246	4.9236	0.0121	$4.4282 \leq \theta \leq 5.4191$	$4.4669 \leq \theta \leq 5.4647$
7	69	7.0145	5.2060	0.0115	$4.6949 \leq \theta \leq 5.7171$	$4.7342 \leq \theta \leq 5.7633$
8	51	6.8235	5.0121	0.0087	$4.4304 \leq \theta \leq 5.5937$	$4.4826 \leq \theta \leq 5.6569$
9	51	6.7059	4.8976	0.0089	$4.3234 \leq \theta \leq 5.4718$	$4.3752 \leq \theta \leq 5.5346$
10	45	6.5999	4.7921	0.0079	$4.1884 \leq \theta \leq 5.3959$	$4.2465 \leq \theta \leq 5.4674$

$^{\otimes}$True value of θ does not lie within the confidence interval.

ILLUSTRATIONS 5.2. The only change here from the previous Illustration 5.1 consists of replacing the chance function: Now w_i instead as in (16), is given by

$$w_i = w(y_i; \lambda) = \lambda y_i \tag{20}$$

$y_i = 1, 2, ...;\ i = 1, ..., N$. As before the distribution f is given by (17), whose mean value as noted earlier is $(1 + \theta)$. Thus now corresponding to (18) we have here

$$w_{i0} = w_0 = \lambda(1 + \theta). \tag{21}$$

Since in (21) w_0 is a 'probability', the permissible values of λ and θ are such that

$$0 \leq \lambda(1 + \theta) \leq 1.$$

An interesting feature of the functions w and w_0 in (20) and (21) above is that the resulting density function (fw/w_0) of Experiment 2 (Section 3) is independent of the parameter λ! (It is important here to note the mathematical equivalence of Experiment 1 (Section 2) and Experiment 2 (Section 3). The performability of Experiment 1 here follows from that of Experiment 2. Now clearly the estimating equation $g_{2\lambda} = 0$ based on (13), is *vacuous* and cannot provide any estimate for λ. This is an important drawback of the estimating function g_2 in (13) compared to g_1 in (12); for the estimating equation $g_{1\lambda} = 0$ does provide reasonable estimation for λ. Though we have considered in (9), λ as a nuisance parameter, θ being the parameter of main interest, in the present context estimation of λ is also of some interest. Now using (17), (20) and (21), it is easy to check from (12) and (13) that the estimating equations $g_{1\theta} = 0$ and $g_{1\lambda} = 0$ together imply the equation $g_{2\theta} = 0$. That is the *point* estimation of the parameter θ, in the two cases is *identical*. The above facts are all borne out by the Tables to follow: Tables 5.2.1, 5.2.2 provide estimation based on the estimating function g_1 in (12) and Tables 5.2.3, 5.2.4 provide estimation based on g_2 in (13). For computing confidence intervals based on t_1' in Tables 5.2.1 and 5.2.2 we, as in (19) use the result

$$(\text{variance } g_{1\theta}) = E\{\text{Var}(g_{1\theta}|s)\} + \text{Var}\{E(g_{1\theta}|s)\}.$$

The above equation after some simplification yields

$$(\text{Variance } g_{1\theta}) = \frac{N\lambda(1+\theta)}{\theta}\left\{\frac{1-\lambda}{(1+\theta)\{1-\lambda(1+\theta)\}} + 1\right\}. \tag{22}$$

Tables 5.2.3, 5.2.4 do not show confidence intervals based on the pivot t_2' in (15); for here the estimating function g_2 in (13) is independent of λ. Further

it is important to note that though the 'point estimation' of θ based on the estimating functions g_1 and g_2, as said before, is 'identical', the 'confidence intervals' are not. The Tables 5.2.1, 5.2.2 provide *shorter* confidence intervals for θ, (based on the pivot t_1) than the corresponding ones in Tables 5.2.3, 5.2.4 (based on the pivot t_2).

TABLE 5.2.1. POINT ESTIMATES FOR (θ, λ) AND CONFIDENCE INTERVALS (CI) FOR θ; $N = 1000$. THE TRUE VALUES ARE $\theta = 4$, $\lambda = 0.05$, $w_0(\theta, \lambda) = 0.2500$, AND $\overline{Y} = 4.9520$

(1) sam-ple	(2) n	(3) $\bar{y}$	(4) $\hat{\theta}$	(5) $\hat{\lambda}$	(6) CI of 95% Coverage Probability based on t_1 in (14)	(7) CI of 95% Coverage Probability based on t_1' in (14) & (19)
1	225	5.9244	4.1198	0.0439	$3.9102 \leq \theta \leq 4.3495$	$3.8878 < \theta \leq 4.3747$
2	255	5.7176	3.9209	0.0518	$3.7307 \leq \theta \leq 4.1282$	$3.7093 \leq \theta < 4.1522$
3	260	5.6500	3.8559	0.0535	$3.6697 \leq \theta \leq 4.0588$	$3.6484 \leq \theta \leq 4.0826$
4	259	5.6757	3.8806	0.0531	$3.6932 \leq \theta \leq 4.0847$	$3.6719 \leq \theta < 4.1085$
5	254	5.7520	3.9538	0.0513	$3.7622 \leq \theta \leq 4.1627$	$3.7408 \leq \theta \leq 4.1867$
6	274	5.5839	3.7926	0.0572	$3.6134 \leq \theta \leq 3.9874$	$3.5925 < \theta \leq 4.0106$
7	231	5.8528	4.0508	0.0457	$3.8463 \leq \theta \leq 4.2747$	$3.8241 \leq \theta < 4.2997$
8	278	5.8741	4.0713	0.0548	$3.8849 \leq \theta \leq 4.2734$	$3.8642 < \theta < 4.2964$
9	245	5.7673	3.9686	0.0493	$3.7729 \leq \theta \leq 4.1824$	$3.7512 < \theta < 4.2067$
10	241	5.7801	3.9809	0.0484	$3.7831 \leq \theta \leq 4.1970$	$3.7612 < \theta \leq 4.2216$

TABLE 5.2.2. POINT ESTIMATES FOR (θ, λ) AND CONFIDENCE INTERVALS (CI) FOR θ; $N = 1000$. THE TRUE VALUES ARE $\theta = 2$, $\lambda = 0.05$, $w_0(\theta, \lambda) = 0.1500$, AND $\overline{Y} = 3.0050$

(1) sam-ple	(2) n	(3) $\bar{y}$	(4) $\hat{\theta}$	(5) $\hat{\lambda}$	(6) CI of 95% Coverage Probability based on t_1 in (14)	(7) CI of 95% Coverage Probability based on t_1' in (14) & (22)
1	160	3.6063	1.9457	0.0543	$1.7955 \leq \theta \leq 2.1171$	$1.7704 < \theta \leq 2.1472$
2	145	3.7931	2.1142	0.0466	$1.9474 \leq \theta \leq 2.3056$	$1.9210 < \theta \leq 2.3374$
3	147	3.5986	1.9389	0.0500	$1.7827 \leq \theta \leq 2.1183$	$1.7567 < \theta < 2.1498$
4	130	3.6846	2.0162	0.0431	$1.8458 \leq \theta \leq 2.2135$	$1.8184 \leq \theta < 2.2471$
5	155	3.6710	2.0039	0.0516	$1.8482 \leq \theta \leq 2.1817$	$1.8227 < \theta < 2.2145$
6	167	3.6647	1.9982	0.0557	$1.8485 \leq \theta \leq 2.1684$	$1.8237 < \theta < 2.1979$
7	136	3.6397	1.9758	0.0457	$1.8114 \leq \theta \leq 2.1656$	$1.7845 < \theta \leq 2.1984$
8	145	3.4345	1.7926	0.0519	$1.6435 \leq \theta \leq 1.9643^{\otimes}$	$1.6177 \leq \theta \leq 1.9958^{\otimes}$
9	142	3.7042	2.0338	0.0468	$1.8697 \leq \theta \leq 2.2226$	$1.8432 \leq \theta \leq 2.2547$
10	154	3.6948	2.0253	0.0509	$1.8680 \leq \theta \leq 2.2051$	$1.8424 < \theta < 2.2360$

$^{\otimes}$True value of θ does not lie within the confidence interval.

TABLE 5.2.3. POINT ESTIMATES FOR (θ, λ) AND CONFIDENCE INTERVALS (CI) FOR θ; $N = 1000$. THE TRUE VALUES ARE $\theta = 4$, $\lambda = 0.05$, $w_0(\theta, \lambda) = 0.2500$, AND $\overline{Y} = 4.952$

(1)	(2)	(3)	(4)	(5)
				CI of 95% Coverage Probability
sample	n	$\bar{y}$	$\hat{\theta}$	based on t_2 in (15)
1	225	5.9244	4.1198	$3.8592 \leq \theta \leq 4.4133$
2	255	5.7176	3.9209	$3.6823 \leq \theta \leq 4.1881$
3	260	5.6500	3.8559	$3.6218 \leq \theta \leq 4.1181$
4	259	5.6757	3.8806	$3.6452 \leq \theta \leq 4.1442$
5	254	5.7520	3.9538	$3.7137 \leq \theta \leq 4.2228$
6	274	5.5839	3.7926	$3.5665 \leq \theta \leq 4.0452$
7	231	5.8528	4.0508	$3.7959 \leq \theta \leq 4.3377$
8	278	5.8741	4.0713	$3.8379 \leq \theta \leq 4.3311$
9	245	5.7673	3.9686	$3.7729 \leq \theta \leq 4.1824$
10	241	5.7801	3.9809	$3.7335 \leq \theta \leq 4.2587$

TABLE 5.2.4. POINT ESTIMATES FOR (θ, λ) AND CONFIDENCE INTERVALS (CI) FOR θ; $N = 1000$. THE TRUE VALUES ARE $\theta = 2$, $\lambda = 0.05$, $w_0(\theta, \lambda) = 0.1500$, AND $\overline{Y} = 3.0050$

(1)	(2)	(3)	(4)	(5)
				CI of 95% Coverage Probability
sample	n	$\bar{y}$	$\hat{\theta}$	based on t_2 in (15)
1	160	3.6063	1.9457	$1.7490 \leq \theta \leq 2.1840$
2	145	3.7931	2.1142	$1.8970 \leq \theta \leq 2.3784$
3	147	3.5986	1.9389	$1.7347 \leq \theta \leq 2.1885$
4	130	3.6846	2.0162	$1.7945 \leq \theta \leq 2.2899$
5	155	3.6710	2.0039	$1.8004 \leq \theta \leq 2.2506$
6	167	3.6647	1.9982	$1.8021 \leq \theta \leq 2.2343$
7	136	3.6397	1.9758	$1.7615 \leq \theta \leq 2.2395$
8	145	3.4345	1.7926	$1.5973 \leq \theta \leq 2.0330$
9	142	3.7042	2.0338	$1.8199 \leq \theta \leq 2.2954$
10	154	3.6948	2.0253	$1.8199 \leq \theta \leq 2.2745$

Acknowledgement. I am indebted to Mary Thompson for her valuable comments on an earlier draft of the paper and to Ker-Ai Lee for her assistance in computations and discussions.

References

COOK, R.D. and MARTIN, F.B. (1974). A model for quadrat sampling with 'visibility bias'. *J. Amer. Statist. Assoc.* **69**, 345-349.

FELDMAN, D. and FOX, M. (1968). Estimation of the parameter n in the Binomial model. *J. Amer. Statist. Assoc.* **63**, 150-158.

Fisher, R.A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugenics* **6**, 13-25.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. B*, **17**, 168-278.

Godambe, V.P. (1975). A reply to my critics. *Sankhyā, Ser. C* **37**, 53-76.

Godambe, V.P. (1976). A historical perspective of the recent developments in the theory of sampling from actual populations. *Jour. Ind. Soc. Agri. Stat.* **28**, 1-12.

Godambe, V.P. and Rajarshi, M.B. (1989). Optimal estimation for weighted distributions: Semi-parametric models. In *Statistical Data Analysis and Inference*, Ed. Y. Dodge, North Holland, Amsterdam, 199-208.

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with discussion). *J. Statist. Plan. Infer.* **22**, 137-172.

Godambe, V.P. and Vijayan, K. (1996). Optimal estimation for response-dependent retrospective sampling. *Jour. Amer. Statist. Assoc.* **91**, 1724-1734.

Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment. In *Classical and Contagious Discrete Distributions*, Ed. G.P. Patil, 320-332, Statistical Publishing Society, Calcutta.

V.P. Godambe

Department of Statistics and Actuarial Science

Faculty of Mathematics

University of Waterloo

200 University Avenue W.

Waterloo, Ontario N2L 3G1

Canada

E-mail: vpgodamb@uwaterloo.ca