TWO-STAGE SAMPLING FROM A PREDICTION POINT OF VIEW

Jan F. Bjørnstad, The University of Tromsø,
Institute of Mathematical and Physical Sciences, N-9000 Tromsø, Norway

KEY WORDS: Predictive likelihood, mean predictor, prediction intervals

ABSTRACT

Estimating the population total in two-stage survey sampling is considered, making use of a (superpopulation) model. The problem is then really one of predicting the unobserved part of the total, and the concept of predictive likelihood is studied. Prediction intervals and a predictor for the population total are derived for the normal case, based on predictive likelihood.

1. INTRODUCTION

Two-stage surveys are used in sampling from finite populations of, say, N primary units or clusters, where each primary unit consists of m_i secondary units. N is assumed known, but the m_i's are unknown before sampling. Let y_{ij} be the value of the variable of interest for secondary unit j of i'th primary unit. The problem is to estimate the total

$$t = \sum_{i=1}^{N} \sum_{j=1}^{m_i} y_{ij} .$$

An example of this situation is considered in Thomsen and Tesfu (1988), with t being the size of a particular population. The primary units are certain administrative units, the secondary units are households and y_{ij} is the number of persons in household j of the i'th administrative unit.

We assume that, before sampling, other measures of the sizes of the primary units are available to us. Let $x_1, ..., x_N$ be these measures and let $X = \sum_{i=1}^{N} x_i$.

The sampling plan is as follows: At stage 1 a sample s of size n_0 of the primary units $(1, ..., N)$ is selected according to some sampling design, and at stage 2 we select for each $i \in s$ a sample s_i of size n_i of secondary units using possibly a different sampling design than at stage 1. The designs are assumed to be non-informative, i.e. they do not depend on the y_{ij}'s and m_i's. E.g., in Thomsen and Tesfu (1988) the two-stage sampling plan is to use pps-sampling at stage 1 (letting selection probabilities of primary units be proportional to the x_i's) and simple random sampling (srs) at stage 2.

The total sample size is $n = \sum_{i \in s} n_i$ and our data now consists of $y(s) = \{y_{ij} : i \in s, j \in s_i\}$ and $m(s) = \{m_i : i \in s\}$. Let $y = (y(s), m(s))$. For the pps-srs sampling plan mentioned above a commonly used design-unbiased estimator of t is the modified Horvitz-Thompson estimator (see for example Cochran (1977), chapter 11)

$$\hat{t}_{HT} = \frac{X}{n_0} \sum_{i \in s} \frac{m_i \bar{y}_i}{x_i} \tag{1}$$

where $\bar{y}_i = \sum_{j \in s_i} y_{ij}/n_i$.

In this paper a (superpopulation) model is adopted, regarding m_i, y_{ij} as realized values of random variables M_i, Y_{ij} for $j = 1, ..., M_i$ and $i = 1, ..., N$. $\mathbf{M} = (M_1, ..., M_N)$ is assumed independent of all Y_{ij}, and further:

$$\begin{aligned} &E(M_i) = \beta x_i , \quad V(M_i) = \sigma^2 v(x_i) , \quad (2)\\ &\mathrm{Cov}(M_i, M_j) = 0 ,\\ &E(Y_{ij}) = \mu , \quad V(Y_{ij}) = \tau^2, \quad \text{and}\\ &\mathrm{Cov}(Y_{ij}, Y_{ik}) = \rho\tau^2 \quad \text{if} \quad k \neq j\\ &\mathrm{Cov}(Y_{ij}, Y_{lk}) = 0 \quad \text{if} \quad l \neq i . \end{aligned}$$

Let $\theta = (\beta, \sigma, \mu, \tau, \rho)$ with $\rho \geq 0$ and let $v_s = \sum_{i \in s} v(x_i)$, $v_{\bar{s}} = \sum_{i \notin s} v(x_i)$. Typically $v(x) = x^g$ with $0 \leq g \leq 2$.

Royall (1976) considers a similar model for Y_{ij}, assuming the m_i's are known, while

Royall (1986) also considers unknown m_i's with a model similar to (2).

The total t is now a realized value of a random variable T, where T can be expressed as $T = \sum_{i\in s}\sum_{j\in s_i} Y_{ij} + Z$ with

$$Z = \sum_{i\in s}\sum_{j\notin s_i} Y_{ij} + \sum_{i\notin s}\sum_{j=1}^{M_i} Y_{ij}\,. \tag{3}$$

Expressing the total T on this form we see that the problem can be described as one of *predicting* the unobserved value z of the random variable Z. It is often clarifying to write a predictor $\hat{T}$ of T on the form

$$\hat{T} = \sum_{i\in s}\sum_{j\in s_i} Y_{ij} + \hat{Z} \tag{4}$$

where $\hat{Z}$ then implicitly is a predictor of Z. From this point of view $\hat{T}_{HT}$, given by (1), does not look like a reasonable predicor. Royall (1970) seems to have been the first one to realize the value in representing any predictor on the prediction form (4), see also Smith (1976).

Modelling the population in survey sampling problems has been and still is controversial. An important aspect of this issue is that the likelihood principle in a sense makes it necessary to model the population. Without a model the only stochastic elements are the samples $\mathbf{s} = \{s, s_i : i \in s\}$, and the likelihood function is then flat (see, e.g., Cassel et al., (1977)), which means that from the likelihood principle point of view the data contains no information about the unobserved y_{ij}'s and m_i's. To make inference we therefore need to relate the data to the unobserved values somehow, and the most natural way of doing so is to formulate a model (see also remarks by Berger and Wolpert (1984, p. 114)).

The random variables observed are $Y(s)$, $M(s)$ and $\mathbf{s}$, where $\mathbf{s}$ now is ancillary. The likelihood principle implies that inference should depend only on the actual $\mathbf{s}$ observed and not on the sampling design. This is called the prediction approach to survey sampling and will be adopted in this paper. Hence everything is considered conditional on $\mathbf{s}$. The prediction approach aims at choosing a predictor that is good for the actual $\mathbf{s}$ obtained and has given significant contributions to a better understanding of several problems in survey sampling, some of which are mentioned in Thomsen and Tesfu (1988). It also enables one to use more conventional statistical methods, although the problem is not to make inference about θ but rather predict Z. Hence θ basically plays the role of a nuisance parameter.

To predict Z we shall use the concept of predictive likelihood, a non-Bayesian likelihood approach to prediction problems in general. One can argue that in the context of a superpopulation model survey sampling provides one of the more natural prediction problems in statistics, and predictive likelihood could therefore serve as a basis for essentially all problems of this kind in survey sampling. Some major references to the general theory of predictive likelihood are Hinkley (1979), Mathiasen (1979) and Butler (1986). A review of some of the suggested likelihoods is given in Bjørnstad (1990).

Section 2 introduces the concept of predictive likelihood and shows how predictors and prediction intervals can be constructed from a predictive likelihood.

In Section 3 a predictive likelihood is derived for the normal model. The usual approaches to obtain a predictive likelihood do not work in two-stage sampling, mainly because Z is a sum of a *stochastic* number of random variables. Therefore a modification is suggested.

The predictor obtained from the predictive likelihood is given by:

$$\hat{Z}_0 = E_{\hat{\theta}}(Z|y) = \sum_{i\in s}(m_i - n_i) \times \left(\frac{1-\hat{\rho}}{1-\hat{\rho}+n_i\hat{\rho}}\hat{\mu} + \frac{n_i\hat{\rho}}{1-\hat{\rho}+n_i\hat{\rho}}\bar{y}_i\right) + \hat{\mu}\sum_{i\notin s}(\hat{\beta}x_i)\,.$$

Here, $\bar{y}_i = \sum_{j\in s_i} y_{ij}/n_i$ and $\hat{\theta} = (\hat{\mu}, \hat{\tau}, \hat{\rho}, \hat{\beta}, \hat{\sigma})$ is the MLE. With $w_s = \sum_{i\in s} x_i^2/v(x_i)$

$$\hat{\beta} = \{\sum_{i\in s} m_i x_i/v(x_i)\}/w_s\,. \tag{5}$$

Since $\hat{\beta}$ is the weighted least squares estimator it is the best unbiased estimator of β. Let $w_i = (1-\hat{\rho})/(1-\hat{\rho}+n_i\hat{\rho})$.

Writing $\hat{Z}_0 = \sum_{i\in s}\sum_{j\notin s_i}(w_i\hat{\mu}+(1-w_i)\bar{y}_i) + \sum_{i\notin s}(\hat{\beta}x_i)\hat{\mu}$ we see from (3) that predicting Z by $\hat{Z}_0$ means that for $i \notin s$ each unobserved Y_{ij} is predicted by $\hat{\mu}$ and M_i is predicted by $\hat{\beta}x_i$. For $i \in s$, $j \notin s_i$, Y_{ij} is predicted by $w_i\hat{\mu} + (1-w_i)\bar{y}_i$.

Three prediction intervals for Z based on similar predictive likelihoods are constructed. They are all of the form $\hat{Z}_0 \pm u(\alpha/2)\sqrt{V_p(Z)}$ where $u(\alpha/2)$ is the upper $(\alpha/2)$-point of $N(0,1)$. $V_p(Z)$ is a measure of the uncertainty in predicting Z of the form $V_p(Z) = V_{\hat{\theta}}(Z|y)+$ (term for parameter uncertainty), see (18).

With $v_{\bar{s}} = \sum_{i\notin s} v(x_i)$,

$$V_\theta(Z|y) = \tau^2 \sum_{i\in s}(m_i - n_i) \times \quad (6)$$
$$\left(1 - \rho\cdot\frac{n_i\rho}{1-\rho+n_i\rho} + (m_i - n_i - 1)\rho\cdot \frac{1-\rho}{1-\rho+n_i\rho}\right) + \tau^2(\beta X_{\bar{s}} + \rho\sigma^2 v_{\bar{s}}$$
$$+\rho\sum_{i\notin s}\beta x_i(\beta x_i - 1)) + \mu^2\sigma^2 v_{\bar{s}}\,.$$

For large n_0, the three intervals are practically identical. However, for small n_0 they differ significantly. To illustrate this confidence levels are estimated by simulation for $1-\alpha = .95$, $n_0 = 6$, $N = 10$, $v(x) = x$ and selected values of $(x_1, ..., x_N)$ and θ.

In a subsequent paper a more comprehensive simulation study for estimating confidence levels will be undertaken, as well as a consideration of optimality for model-unbiased predictors.

2. PREDICTIVE LIKELIHOOD

We shall here give a brief general introduction to the concept of predictive likelihood. For a more complete exposition we refer to Bjørnstad (1990). Let $Y = y$ be the data. The problem is to predict the unobserved or future value z of a random variable Z usually by a predictor and confidence interval for Z. It is assumed that (Y, Z) has a probability density or mass function (pdf) $f_\theta(y,z)$. In general we let $f_\theta(\cdot)$ and $f_\theta(\cdot|\cdot)$ denote the pdf and conditional pdf of the enclosed variables. The joint likelihood function for the two unknown quantities, z and θ, is given by $l_y(z,\theta) = f_\theta(y,z)$. The aim is to develop a likelihood for z, $L(z|y)$, by eliminating θ from l_y. Any such likelihood is called a predictive likelihood.

Different ways of eliminating θ then give rise to different L. The two main type of suggestions are the conditional predictive likelihood L_c, essentially suggested by Hinkley (1979), and the profile predictive likelihood L_p, first considered by Mathiasen (1979). Let $R = r(Y,Z)$ denote a minimal sufficient statistic for (Y,Z). Then

$$L_c(z|y) = f_\theta(y,z)/f_\theta(r(y,z)) \quad (7)$$

$$L_p(z|y) = \max_\theta f_\theta(y,z) = f_{\hat{\theta}_z}(y,z) \quad (8)$$

Typically, L_c and L_p are quite similar when sufficiency provides a genuine reduction and the dimension of θ is small.

In linear normal models, L_p will ignore the number of parameters and can be misleadingly precise. A modification of L_p, L_{mp}, that adjusts for this was suggested by Butler (1986, rejoinder), see also Bjørnstad (1990). Let $Y = (X_1, ..., X_n)$ and $Z = (X_1', ..., X_m')$, and assume that all X_i's and X_j's are independent. Let $\theta = (\theta_1, ..., \theta_k)$. Then L_{mp} is given by

$$L_{mp}(z|y) = L_p(z|y)\cdot|I^z(\hat{\theta}_z)|^{1/2}/|H_z H_z'|^{1/2}\,. \quad (9)$$

Here, $I^z(\theta) = \{I^z_{ij}(\theta)\}$ is the "observed" information-matrix based on (y,z), i.e. $I^z_{ij}(\theta) = -\partial^2 \log f_\theta(y,z)/\partial\theta_i\partial\theta_j$. $H_z = H_z(\hat{\theta}_z)$, and $H_z(\theta)$ is the $k \times (n+m)$ matrix of second-order partial derivatives of $\log f_\theta(y,z)$ with respect to θ and (y,z). We shall assume that any L considered is normalized as a probability distribution in Z. The mean and variance of L are then called the predictive expectation and the predictive variance of Z, denoted by $E_p(Z)$ and $V_p(Z)$. $E_p(Z)$ is then a natural predictor for z, called the mean predictor. $L(z|y)$ also gives us an idea on how likely different z-values are in light of the data, and can be used to construct prediction intervals for z. An interval (a_y, b_y)

is a $(1-\alpha)$ predictive interval based on $L(z|y)$ if $\int_{a_y}^{b_y} L(z|y)dz = 1-\alpha$. A simplified (quasi) $(1-\alpha)$ predictive interval is of the form

$$E_p(Z) \pm u\sqrt{V_p(Z)} \tag{10}$$

where u is the upper $(\alpha/2)$-point in the actual (exact or approximate) conditional distribution, given y, of $(Z - E_\theta(Z|y))/\sqrt{V_\theta(Z|y)}$.

3. PREDICTOR AND PREDICTION INTERVALS IN TWO-STAGE SAMPLING BASED ON PREDICTIVE LIKELIHOOD

In two-stage sampling, Z is given by (3), and is a sum of two mixtures. Therefore, instead of considering a predictive likelihood for Z directly, we look at a joint predictive likelihood for Z and $M(\bar{s}) = (M_i,\ i \notin s)$. It has the following form

$$L(z, m(\bar{s})|y) = L_{m(\bar{s})}(z|y)L(m(\bar{s})|y) \tag{11}$$

$L_{m(\bar{s})}(z|y)$ is a predictive likelihood for z conditional on $M(\bar{s}) = m(\bar{s})$, i.e. based on $f_\theta(y, z|m(\bar{s}))$. $L(m(\bar{s})|y)$ is a predictive likelihood for $m(\bar{s})$ based on $f_\theta(y, m(\bar{s}))$. Then E_p, V_p follow the usual rules for double expectation, i.e.

$$\begin{aligned} E_p(Z) &= E_p\{E_p(Z|M(\bar{s}))\} \\ V_p(Z) &= E_p\{V_p(Z|M(\bar{s}))\} \\ &\quad + V_p\{E_p(Z|M(\bar{s}))\} \end{aligned} \tag{12}$$

In (12) $E_p(Z|m(\bar{s}))$ and $V_p(Z|m(\bar{s}))$ are the predictive mean and variance for Z from $L_{m(\bar{s})}(z|y)$. In principle we can derive $L(z|y)$ as the marginal likelihood from $L(z, m(\bar{s})|y)$. The advantage of (11) is that we are able to obtain $E_p(Z)$ and $V_p(Z)$ without actually deriving $L(z|y)$.

Under the model (2) we can factorize $f_\theta(y, z, m(\bar{s})) = f_{\sigma,\beta}(m(s), m(\bar{s})) \cdot f_{\mu,\tau,\rho}(y(s), z|m(s), m(\bar{s}))$ and it is readily seen that applying L_p, given by (8), to the terms on the right hand side in (11) in fact gives us $L_p(z, m(\bar{s})|y) = \max_\theta f_\theta(y, z, m(\bar{s}))$, i.e.

$$L_p(z, m(\bar{s})|y) = L_{m(\bar{s}),p}(z|y)L_p(m(\bar{s})|y)\,. \tag{13}$$

It follows that $E_p(Z)$ and $V_p(Z)$ based on $L_p(z, m(\bar{s})|y)$ can be derived by (12). We note that L_c, given by (7), has the same property, i.e. $L_c(z, m(\bar{s})|y) = L_{m(\bar{s}),c}(z|y)L_c(m(\bar{s})|y)$.

Normal model

It is now assumed that model (2) holds and that Y_{ij}, M_i are normally distributed.

We shall first consider the second likelihood in (11), $L(m(\bar{s})|y)$, using L_p. Let $t_\nu^{(k)}(\Sigma)$ denote the k-dimensional multivariate t-distribution with ν degrees of freedom (d.f.) and variance-covariance matrix Σ, i.e. $t_\nu^{(k)}(\Sigma)$ is the distribution of $(\mathbf{U}/W)\sqrt{\nu}$ where $\mathbf{U} \sim N_k(0, \Sigma)$ and $W^2 \sim \chi_\nu^2$.

Let $X(\bar{s})$ be the vector $(x_i\ :\ i \notin s)$. Then $L_p(m(\bar{s})|y)$ leads to a multivariate t-distribution, specifically $L_p(m(\bar{s})|y)$ is such that $[M(\bar{s}) - \hat{\beta}X(\bar{s})]/\hat{\sigma} \sim t_{n_0}^{(N-n_0)}(V)$, where the m.l.e. are $\hat{\beta}$, given by (5), and $\hat{\sigma}^2 = \frac{1}{n_0}\sum_{i\in s}(m_i - \hat{\beta}x_i)^2/v(x_i)$. $V = (v_{ij})$ with $v_{ii} = v(x_i) + x_i^2/w_s$ and $v_{ij} = x_i x_j/w_s$ for $i \neq j$. It follows that $E_p(M_i) = \hat{\beta}x_i$, $V_p(M_i) = \frac{n_0}{n_0-2}\hat{\sigma}^2(v(x_i) + x_i^2/w_s)$ and the predictive covariances are $\mathrm{Cov}_p(M_i, M_j) = \frac{n_0}{n_0-2}\hat{\sigma}^2 \cdot x_i x_j/w_s$ for $i \neq j$. This implies that

$$\begin{aligned} E_p(\sum_{i\notin s} M_i) &= \hat{\beta}X_{\bar{s}} \quad \text{and} \\ V_p(\sum_{i\notin s} M_i) &= \frac{n_0}{n_0-2}\hat{\sigma}^2\left(v_{\bar{s}} + \frac{X_{\bar{s}}^2}{w_s}\right). \end{aligned} \tag{14}$$

L_c and L_{mp} (for $M_i/\sqrt{v(x_i)}$, $i \notin s$), lead to moments similar to (14) with $n_0 - 2$ replaced by $n_0 - 5$ and $n_0 - 4$ respectively.

Let us now consider the first term in (11), $L_{m(\bar{s})}(z|y)$ based on $f_\theta(y, z|m(\bar{s}))$. For this likelihood we will restrict attention to L_p, i.e. deriving $L_{m(\bar{s}),p}(z|y)$. The m.l.e. $\hat{\mu}, \hat{\tau}^2, \hat{\rho}$ can be expressed the following way:

$$\hat{\mu} = \sum_{i\in s}\frac{n_i}{1-\hat{\rho}+n_i\hat{\rho}}\bar{y}_i \Big/ \sum_{i\in s}\frac{n_i}{1-\hat{\rho}+n_i\hat{\rho}} \tag{15}$$

$$\hat{\tau}^2 = \frac{1}{n}\left(\frac{SSE}{1-\hat{\rho}} + \sum_{i\in s}\frac{n_i(\bar{y}_i - \hat{\mu})^2}{1-\hat{\rho}+n_i\hat{\rho}}\right)$$

and $\hat{\rho}$ is found numerically, maximizing $-(n/2)\log\hat{\tau}^2 - (n/2)\sum_{i\in s}\log(1-\hat{\rho}+n_i\hat{\rho}) +$

$((n - n_0)/2)\log(1 - \hat{\rho})$. Here, $SSE = \sum_{i\in s}\sum_{j\in s_i}(y_{ij} - \bar{y}_i)^2$. When $n_i = c$, for all $i \in s$, then $\hat{\mu} = \bar{y} = \sum_{i\in s}\bar{y}_i/n_0$, $\hat{\tau}^2 = SS/n$, $\hat{\rho} = \max\left(0, 1 - \frac{c}{c-1}\cdot\frac{SSE}{SS}\right)$, where $SS = \sum_{i\in s}\sum_{j\in s_i}(y_{ij} - \bar{y})^2$.

Consider first the case when ρ and τ are known. Then $\hat{\mu}$ is given by (15) with ρ replacing $\hat{\rho}$. In this case $L_{m(\bar{s}),p}(z|y)$ is such that Z is normally distributed with

$$E_p(Z|m(\bar{s})) = \sum_{i\in s}(m_i - n_i) \times \left(\frac{1-\rho}{1-\rho+n_i\rho}\hat{\mu} + \frac{n_i\rho}{1-\rho+n_i\rho}\bar{y}_i\right) + \hat{\mu}\sum_{i\notin s} m_i, \quad (16)$$

$$V_p(Z|m(\bar{s})) = V(Z|y, m(\bar{s})) + \frac{\tau^2}{\sum_{i\in s}\frac{n_i}{1-\rho+n_i\rho}}\left(\sum_{i\notin s} m_i + \sum_{i\in s}(m_i - n_i) \times \frac{1-\rho}{1-\rho+n_i\rho}\right)^2. \quad (17)$$

When ρ, τ are unknown, $L_{m(\bar{s}),p}(z|y)$ will for large n_0 be approximately such that Z is normally distributed with $E_p(Z|m(\bar{s}))$ and $V_p(Z|m(\bar{s}))$ given by (16) and (17) with $\hat{\rho}, \hat{\tau}^2$ replacing ρ, τ^2. It now follows, from (13), (14), (16) and (17) that, approximately, $L_p(z, m(\bar{s})|y)$ has $E_p(Z) = E_{\hat{\theta}}(Z|y)$ and

$$V_p(Z) = V_{\hat{\theta}}(Z|y) + \frac{\hat{\tau}^2}{\sum_{i\in s}\frac{n_i}{1-\hat{\rho}+n_i\hat{\rho}}} \times \left[\hat{\beta}X_{\bar{s}} + (1-\hat{\rho})\sum_{i\in s}\frac{m_i - n_i}{1-\hat{\rho}+n_i\hat{\rho}}\right]^2 + \hat{\sigma}^2\left(\hat{\mu}^2\cdot\frac{X_{\bar{s}}^2}{w_s} + \hat{\rho}\hat{\tau}^2\cdot\frac{\sum_{i\notin s}x_i^2}{w_s}\right) + h(2). \quad (18)$$

Here, $V_{\hat{\theta}}(Z|y)$ is given by (6) and

$$h(k) = \frac{n_0}{n_0 - k}\hat{\sigma}^2 \times \left(\frac{\hat{\tau}^2}{\sum_{i\in s}\frac{n_i}{1-\hat{\rho}+n_i\hat{\rho}}} + \frac{k}{n_0}\hat{\mu}^2\right)\left(v_{\bar{s}} + \frac{X_{\bar{s}}^2}{w_s}\right) + \frac{k}{n_0 - k}\hat{\rho}\hat{\tau}^2\hat{\sigma}^2\cdot\left(v_{\bar{s}} + \frac{1}{w_s}\sum_{i\notin s}x_i^2\right).$$

The predictive likelihood

$$L^{(p,c)}(z, m(\bar{s})|y) = L_{m(\bar{s}),p}(z|y)L_c(m(\bar{s})|y)$$

leads to the same $E_p(Z)$ while $V_p(Z)$ equals (18) with $h(5)$ instead of $h(2)$. With

$$L^{(p,mp)}(z, m(\bar{s})|y) = L_{m(\bar{s}),p}(z|y)\cdot L_{mp}(m(\bar{s})|y)$$

we get the same $E_p(Z)$ and $V_p(Z)$ equal to (18) with $h(4)$.

It can be shown that, conditional on y, $(Z - E_\theta(Z|y))/\sqrt{V_\theta(Z|y)}$ is asymptotically $N(0,1)$ as $N - n_0 \to \infty$ provided that the x_i's are bounded as $N - n_0 \to \infty$. Hence $Z|y$ is approximately normal for large $N - n_0$, and the quasi $(1-\alpha)$ predictive interval given by (10) becomes

$$E_{\hat{\theta}}(Z|y) \pm u(\alpha/2)\sqrt{V_p(Z)}$$

where $u\left(\frac{\alpha}{2}\right)$ is the upper $\alpha/2$-point in $N(0,1)$. This amounts to regarding $N(E_p(Z), V_p(Z))$ as a predictive distribution for Z. $V_p(Z)$ equals (18) if the interval is based on $L_p(z, m(\bar{s})|y)$, while $L^{(p,c)}$ has (18) with $h(5)$ and $L^{(p,mp)}$ has (18) with $h(4)$. Let us denote these prediction intervals by I_p, I_{pc} and I_{mp}. Clearly $I_p \subset I_{mp} \subset I_{pc}$.

For large n_0 there is practically no difference between these intervals. However, for small n_0 they do differ. To find out how the intervals perform for small n_0 (and small N) a simulation study with $n_0 = 6$ and $N = 10$ was done to estimate the confidence levels $C_p = P(Z \in I_p(Y))$, $C_{pc} = P(Z \in I_{pc}(Y))$ and $C_{mp} = P(Z \in I_{mp}(Y))$, all conditional on **s**. The approximations to $L_{m(\bar{s}),p}$ and to the distribution of Z given y are not valid for small n_0 and small $N - n_0$. Still, it is of interest to find out how the coverage properties of the different intervals are in this case. In a later paper a more comprehensive simulation study will be undertaken, including also large $n_0, N - n_0$ cases.

The simulation study considers the following two main cases, with $s = (1,2,3,4,5,6)$, $1-\alpha = .95$, $v(x) = x$ and $n_i \equiv c$, $\forall i \in s$. (I) $x_1 = x_2 = x_3 = 50$, $x_4 = x_5 = 30$, $x_6 = x_7 = x_8 = 100$, $x_9 = x_{10} = 50$; $c = 3,10$. (II) $x_1 = x_2 = x_3 =$

5000, $x_4 = x_5 = 3000$, $x_6 = x_7 = x_8 = 10000$, $x_9 = x_{10} = 5000$; $c = 10{,}400$.

Case (I): Two values of μ are considered, $\mu = 5, 100$. For $\mu > 100$ the confidence levels seemed to be essentially equal to the confidence levels when $\mu = 100$. With regard to σ, β the levels seemed to depend essentially on the ratio β/σ and we consider $\beta/\sigma = .75$, 1, 1.5, 2, 3.

(Ia): $\mu = 100$, $\tau = 1, 5$ and $\rho = .1, .5, .9$. The confidence levels are approximately constant for all the various chosen values of θ. Based on simulation of 60,000 observations of (y, z) we find $C_p = .924$, $C_{mp} = .973$, $C_{pc} = .992$.

Table 1. Confidence levels for case (I) and $\mu = 5$, $\tau = 1$; $1 - \alpha = .95$.

β/σ	.75	1	1.5	2	3
ρ					
.1	.929	.936	.932	.937	.927
C_p .5	.930	.921	.913	.899	.891
.9	.920	.923	.899	.895	.889
.1	.971	.973	.962	.961	.948
C_{mp} .5	.967	.959	.943	.927	.905
.9	.961	.952	.930	.919	.904
.1	.990	.991	.984	.980	.967
C_{pc} .5	.989	.981	.971	.958	.930
.9	.986	.978	.958	.948	.922

Table 2. Confidence levels for case (II) and $\sigma = \beta = 1$, $1 - \alpha = .95$.

τ/μ	.01	.05	.20	
ρ				
c	10,400	10,400	10	400
.01	.923	.936	.940	.926
C_p .10	.923	.929	.920	.874
.50	.929	.896	.870	.864
.01	.974	.970	.944	.949
C_{mp} .10	.973	.961	.923	.889
.50	.971	.922	.872	.866
.01	.994	.989	.951	.975
C_{pc} .10	.994	.982	.928	.911
.50	.994	.951	.876	.873

(Ib): $\mu = 5$, $\tau = 1$. Table 1 is based on simulation of 5000 observations of (y, z) in each case.

Case (II): We consider $\beta = \sigma = 1$. It seems that the confidence levels depend on μ, τ only through the coefficient of variation, τ/μ. Table 2 is based on simulation of 5000 observations of (y, z) in each case.

When $n_0 = 6$ and $N = 10$ I_p is clearly too short generally. I_{pc} is typically too wide, especially when ρ is only moderately large. Overall, I_{mp} seems to have confidence levels closest to .95 of the three intervals.

REFERENCES

Berger, J.O. and Wolpert, R.L. (1984). The Likelihood Principle. IMS Lecture Notes – Monograph Series, Vol. 6.

Bjørnstad, J.F. (1990). Predictive likelihood: hood: A review (with Discussion). Statistical Science, 5, 242–265.

Butler, R.W. (1986). Predictive likelihood inference with applications (with Discussion). J.R. Statist. Soc., B 48, 1–38.

Cassel, C.-M., Särndal, C.-E. and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. Wiley.

Cochran, W.G. (1977). Sampling Techiques, 3rd ed. Wiley.

Hinkley, D.V. (1979). Predictive likelihood. Ann. Statist., 7, 718–728 (corrig. 8, 694).

Mathiasen, P.E. (1979). Prediction functions. Scand. J. Statist., 6, 1–21.

Royall, R.M. (1970) On finite population sampling theory under certain linear regression models. Biometrika, 57, 377–387.

Royall, R.M. (1976). The linear least-squares prediction approach to two-stage sampling. J. Am. Statist. Ass., 71, 657–664.

Royall, R.M. (1986). The prediction approach to robust variance estimation in two-stage cluster sampling. J. Am. Statist. Ass., 81, 119–123.

Smith, T.M.F. (1976). The foundations of survey sampling: A review (with Discussion). J.R. Statist. Soc., A 139, part 2, 183–204.

Thomsen, I. and Tesfu, D. (1988). On the use of models in sampling from finite populations. Handbook of Statistics, vol. 6, 369–397.