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.

I. 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)

$$i_{HT} = \frac{X}{n_0} \sum_{i \in s} \frac{m_i \bar{y}_i}{x_i} \quad (1)$$

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 $i = 1, ..., N$. $M = (M_1, ..., M_N)$ is assumed independent of all $Y_{ij}$, and further:

$$E(M_i) = \beta x_i, \quad V(M_i) = \sigma^2 v(x_i), \quad (2)$$

$$\text{Cov}(M_i, M_j) = 0,$$

$$E(Y_{ij}) = \mu, \quad V(Y_{ij}) = \tau^2, \quad \text{and}$$

$$\text{Cov}(Y_{ij}, Y_{ik}) = \rho \tau^2 \quad \text{if} \quad k \neq j$$

$$\text{Cov}(Y_{ij}, Y_{ik}) = 0 \quad \text{if} \quad l \neq i.$$

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

\[
Z = \sum_{i \in s} \sum_{j \in s_i} Y_{ij} + \sum_{i \in s} M_i,
\]

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

\[
\hat{T} = \sum_{i \in s} \sum_{j \in s_i} Y_{ij} + \hat{Z}
\]

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 predictor. 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 \( s = \{ s_i : i \in s \} \), 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 \( s \), where \( s \) now is ancillary. The likelihood principle implies that inference should depend only on the actual \( s \) 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 \( s \). The prediction approach aims at choosing a predictor that is good for the actual \( s \) 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 \( \theta \) but rather predict \( Z \). Hence \( \theta \) 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:

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

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 \( w_s = \sum_{i \in s} x_i^2 / v(x_i) \)

\[
\hat{\beta} = \left\{ \sum_{i \in s} m_i x_i / v(x_i) \right\} / w_s.
\]

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

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

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 $V_p(Z) = V_0(Z|y) + (\text{term for parameter uncertainty})$, see (18).

With $v_s = \sum_{i \notin s} v(z_i)$,

$$V_0(Z|y) = \tau^2 \sum_{i \in s} (m_i - n_i) \times$$

$$\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_i^2 + \rho \sigma^2 v_s) + \rho \sum_{i \notin s} (\beta x_i (\beta x_i - 1)) + \mu^2 \sigma^2 v_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 $1 - \alpha = .95$, $n_0 = 6$, $N = 10$, $v(z) = z$ and selected values of $(z_1, \ldots, z_N)$ and $\theta$.

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 $\theta$, is given by $L(y, z|\theta) = f_\theta(y, z|\theta)$. The aim is to develop a likelihood for $z$, $L(z|y)$, by eliminating $\theta$ from $L(y, z|\theta)$. Any such likelihood is called a predictive likelihood.

Different ways of eliminating $\theta$ 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

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

$$L_p(z|y) = \max_\theta f_\theta(y, z) = f_\delta_0(y, z) \quad (8)$$

Typically, $L_c$ and $L_p$ are quite similar when sufficiency provides a genuine reduction and the dimension of $\theta$ 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 $Y = (X_1, \ldots, X_n)$ and $Z = (X_1', \ldots, X_m')$, and assume that all $X_i$'s and $X_j$'s are independent. Let $\theta = (\theta_1, \ldots, \theta_k)$. Then $L_{mp}$ is given by

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

Here, $I^z(\theta) = \{I_{ij}^z(\theta)\}$ is the "observed" information-matrix based on $(y, z)$, i.e. $I_{ij}^z(\theta) = -\partial^2 \log f_\theta(y, z)/\partial \theta_i \partial \theta_j$, $H_z = H_s(\hat{\theta}_z)$, 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 $\theta$ 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 \(\int_{a}^{b} L(z|y) dx = 1 - \alpha\). A simplified (quasi) \((1 - \alpha)\) predictive interval is of the form

\[
E_P(Z) \pm u \sqrt{V_P(Z)}
\]

(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 \(M(\tilde{s}) = (M_i, i \neq s)\). It has the following form

\[
L(z, m(\tilde{s})|y) = L_{m(s)}(z|y)L(m(\tilde{s})|y)
\]

(11)

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

\[
E_P(Z) = E_P\{E_P(Z|M(\tilde{s}))\}
\]

\[
V_P(Z) = E_P\{V_P(Z|M(\tilde{s}))\}
\]

\[
+V_P\{E_P(Z|M(\tilde{s}))\}
\]

(12)

In (12) \(E_P(Z|m(\tilde{s}))\) and \(V_P(Z|m(\tilde{s}))\) are the predictive mean and variance for \(Z\) from \(L_{m(s)}(z|y)\). In principle we can derive \(L(z|y)\) as the marginal likelihood from \(L(z, m(\tilde{s})|y)\).

Under the model (2) we can factorize \(f_\theta(y, z, m(\tilde{s})) = f_\theta(x_\theta(m(s)), m(\tilde{s})) \cdot f_\mu(x_{\mu}(y(s), z|m(s)), m(\tilde{s}))\) 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 \(L_p(z, m(\tilde{s})|y) = \max_s f_\theta(y, z, m(\tilde{s}))\), i.e.

\[
L_p(z, m(\tilde{s})|y) = L_{m(s),p}(z|y)L_p(m(\tilde{s})|y).
\]

(13)

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

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), \(L_p(m(\tilde{s})|y)\), using \(L_p\). Let \(t_{n_0}^{(k)}(\Sigma)\) denote the \(k\)-dimensional multivariate distribution with \(\nu\) degrees of freedom (d.f.) and variance-covariance matrix \(\Sigma\), i.e. \(t_{n_0}^{(k)}(\Sigma)\) is the distribution of \((U/W)/\sqrt{\nu}\) where \(U \sim N_k(0, \Sigma)\) and \(W^2 \sim \chi^2_{n_0}\).

Let \(X(\tilde{s})\) be the vector \((x_i : i \neq s)\). Then \(L_p(m(\tilde{s})|y)\) leads to a multivariate \(t\)-distribution, specifically \(L_p(m(\tilde{s})|y) = \text{max}_{i \neq s} f_\theta(Y, x_i, x_{\tilde{s}})^{-1/2}/v_{x_i}\). \(V = (v_{ij})\) 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 \(E_\theta(M_i) = \mu x_i, V_\theta(M_i) = v_n - 2 \hat{\sigma}^2 (v(x_i) + x_i^2/w_s)\) and the predictive covariances are \(\text{Cov}_\theta(M_i, M_j) = \frac{v_n - 5 - \hat{\sigma}^2 x_i x_j/w_s}{v_n - 4}\)

with \(v_n = v_{ii} - x_i x_j/w_s\) for \(i \neq j\). This implies that

\[
E_P\{\sum_{i \neq s} M_i\} = \hat{\beta} X_{\tilde{s}} \quad \text{and} \quad \text{(14)}
\]

\[
V_P\{\sum_{i \neq s} M_i\} = \frac{n_0 - 2}{n_0 - 4} \hat{\sigma}^2 \left( v_s + X_{\tilde{s}}^2/w_s \right).
\]

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

\[
\hat{\mu} = \sum_{i \neq s} \frac{v_i}{n_i - 1 - \hat{\rho} + n_i \hat{\rho}} \frac{\tilde{y}_i}{\sum_{i \neq s} \frac{n_i}{1 - \hat{\rho} + n_i \hat{\rho}}}
\]

\[
\hat{\tau}^2 = \frac{1}{n} \left( \frac{SSE}{1 - \hat{\rho}} + \sum_{i \neq s} \frac{n_i (\tilde{y}_i - \hat{\mu})^2}{1 - \hat{\rho} + n_i \hat{\rho}} \right)
\]

and \(\hat{\rho}\) is found numerically, maximizing

\[
-(n/2) \log \hat{\tau}^2 - (n/2) \sum_{i \neq s} \log(1 - \hat{\rho} + n_i \hat{\rho}) +
\]
((n - n_0)/2) \log(1 - \hat{\rho}). \text{ Here, } SSE = 
\sum_{i \in s} \sum_{j \in s} (y_{ij} - \tilde{y_i})^2. \text{ When } n_i = c, \text{ for all } i \in s, \text{ then } \hat{\mu} = \tilde{y} = \sum_{i \in s} \tilde{y_i}/n_0, \hat{\tau}^2 = SS/n, \hat{\rho} = \max (0, 1 - \frac{\hat{\tau}^2}{SS}) \text{, where } SS = 
\sum_{i \in s} \sum_{j \in s} (y_{ij} - \tilde{y_i})^2.

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

\[
E_p(Z|m(\hat{\sigma})) = \sum_{i \in s} (m_i - n_i) \times \left( \frac{1 - \rho}{1 - \rho + n_i \hat{\rho}} \hat{\mu} + \frac{n_i \hat{\rho}}{1 - \rho + n_i \hat{\rho}} \tilde{y_i} \right) + \hat{\mu} \sum_{i \in s} m_i,
\]

\[
V_p(Z|m(\hat{\sigma})) = V(Z|y, m(\hat{\sigma})) + \frac{\hat{\tau}^2}{\sum_{i \in s} \frac{n_i}{1 - \rho + n_i \hat{\rho}}} \left( \sum_{i \in s} m_i + \sum_{i \in s} (m_i - n_i) \times \left( \frac{1 - \rho}{1 - \rho + n_i \hat{\rho}} \right)^2 \right).
\]

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

\[
V_p(Z) = V_{\hat{\sigma}}(Z|y) + \frac{\hat{\tau}^2}{\sum_{i \in s} \frac{n_i}{1 - \hat{\rho} + n_i \hat{\rho}}} \left( \sum_{i \in s} \frac{m_i - n_i}{1 - \hat{\rho} + n_i \hat{\rho}} \right)^2 \beta X_s + (1 - \hat{\rho}) \sum_{i \in s} \frac{m_i - n_i}{1 - \hat{\rho} + n_i \hat{\rho}}^2 \mu^2 \cdot X^2_s + \hat{\rho} \hat{\tau}^2 \cdot \sum_{i \in s} \frac{x_i^2}{w_s} + h(2).
\]

\[
E_{\hat{\sigma}}(Z|y) \pm u(\alpha/2)\sqrt{V_{\hat{\sigma}}(Z)}
\]

where \( u(\alpha/2) \) 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 \( L_p(z, m(\hat{\sigma})|y) \), 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 \( I_p, I_pc, I_m, I_mp \). Clearly \( I_p \subset I_m \subset I_pc \). 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 \( L_{m(s),p} \) 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 \( s = (1, 2, 3, 4, 5, 6), 1 - \alpha = .95, u(z) = z \) 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) \( x_1 = x_2 = x_3 = 50, \)
5000, \( z_4 = z_5 = 3000, z_6 = z_7 = z_8 = 10000, \\
z_9 = z_{10} = 5000; c = 10,400. \)

Case (I): Two values of \( \mu \) are considered, \( \mu = 5, 1000. \) For \( \mu > 100 \) the confidence levels seemed to be essentially equal to the confidence levels when \( \mu = 100. \) With regard to \( \sigma, \beta \) the levels seemed to depend essentially on the ratio \( \beta/\sigma \) 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 \( \theta. \) 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 \( \mu = 5, \tau = 1; 1 - \alpha = .95. \)

<table>
<thead>
<tr>
<th>( \beta/\sigma )</th>
<th>.75</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>.1</td>
<td>.929</td>
<td>.936</td>
<td>.932</td>
<td>.937</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.920</td>
<td>.923</td>
<td>.899</td>
<td>.895</td>
</tr>
<tr>
<td>( C_p )</td>
<td>.971</td>
<td>.973</td>
<td>.962</td>
<td>.961</td>
<td>.948</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.961</td>
<td>.952</td>
<td>.930</td>
<td>.919</td>
</tr>
<tr>
<td>( C_{mp} )</td>
<td>.967</td>
<td>.959</td>
<td>.943</td>
<td>.927</td>
<td>.905</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.961</td>
<td>.952</td>
<td>.930</td>
<td>.919</td>
</tr>
<tr>
<td>( C_{pc} )</td>
<td>.990</td>
<td>.991</td>
<td>.984</td>
<td>.980</td>
<td>.967</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.986</td>
<td>.978</td>
<td>.958</td>
<td>.948</td>
</tr>
</tbody>
</table>

Table 2. Confidence levels for case (II) and \( \sigma = 5, \tau = 1; 1 - \alpha = .95. \)

<table>
<thead>
<tr>
<th>( \tau/\mu )</th>
<th>.01</th>
<th>.05</th>
<th>.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>10,400</td>
<td>10,400</td>
<td>10</td>
</tr>
</tbody>
</table>

| \( C_p \) | .923 | .936 | .940 | .926 |
|          | .923 | .929 | .920 | .874 |
|          | .929 | .896 | .870 | .864 |
|          | .974 | .970 | .944 | .949 |
|          | .973 | .961 | .923 | .889 |
|          | .971 | .922 | .872 | .866 |

| \( C_{mp} \) | .994 | .989 | .951 | .975 |
|             | .994 | .982 | .928 | .911 |
|             | .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 \( \mu, \tau \) only through the coefficient of variation, \( \tau/\mu. \) 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. \( I_{mp} \) is typically too wide, especially when \( \rho \) is only moderately large. Overall, \( I_{mp} \) seems to have confidence levels closest to .95 of the three intervals.

REFERENCES


