Article ID: 1003 - 6326(1999)04 - 0852 - 05

Choice of estimation of unknown parameter under contaminated error model[①]

Wang Zhizhong(王志忠), Zhu Jianjun(朱建军)

College of Resource, Environment and Civil Engineering, Central South University of Technology, Changsha 410083, P. R. China

Abstract: The ε -contaminated normal distribution, $\psi(\Delta) = (1-\varepsilon)\psi_0(\Delta) + \varepsilon\psi_1(\Delta)$, was considered as error model occurring in practice (ψ = probability function, Δ = observation error). The variances of the L_1 (Least Absolute Sum) estimation and the L_2 (Least Squares) estimation were compared with each other based on their asymptotic distribution. The revised L_2 estimation was then derived. The conditions that the L_1 estimation is superior to the L_2 estimation and that the revised L_2 estimation is superior to L_1 estimation were discussed.

Key words: contaminated model; least squares estimation; least absolute sum estimation

Document: A

1 INTRODUCTION

Classical least squares approach is based on the error model of Gauss Markov, that is, there are no outlying observations and the observations are independent normally distributed in statistical sense. In such situations, the L_2 estimation is an uniformly minimum variance unbiased estimation. This method becomes very popular because its mathematical formulas and computing algorithm are relatively simpler than the others. In practice, however, often both assumptions are incorrect[1]. The presence of outliers makes the L_2 estimation non-optimum. To describe the situation, the contaminated error model is presented in modern statistics and in modern surveying adjustment, and the robust statistics and the robust adjustment are also established.

At present, the robust estimates depend mainly on their weight functions. Up to now, nearly all weight functions are chosen personally[2,3]. To resolve these problems, the revised L_2 estimation is derived according to statistics. The optimum property of L_1, L_2, and the revised L_2 estimation are compared. The results show that the L_1 estimation is superior to the L_2 estimation and that the revised L_2 estimation is superior to the L_1 estimation under some conditions.

2 ERROR FORMS OF CONTAMINATED ERROR MODEL

In surveying adjustment, there are two kinds of error models to describe the blunders or the departures under the contaminated error model. One is the stochastic error model, the other is the mean shift error model[4,5].

If the outliers are considered as a part of the functional model, which is called the mean shift error model, we may believe that the outlying observations are such a set of observations, which have the same variances as the other observations, but not the same mean values, i. e.

$$L_{i0} \sim N(E(L) + \delta,\ \sigma_0^2) \qquad (1)$$

$$L_j \sim N(E(L),\ \sigma_0^2) \quad (j \neq i) \qquad (2)$$

Hence, an ε -contaminated normal distribution can be obtained by

$$\psi(\Delta_\varepsilon) = (1-\varepsilon)\psi_0(\Delta) + \varepsilon\psi_1(\Delta_s) \qquad (3)$$

where

① Project 49774209 supported by the National Natural Science Foundation of China

Received Jan. 18, 1999; accepted Jul. 19, 1999

$$\psi_0(\Delta) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{\Delta^2}{2\sigma_0^2}\right) \tag{4}$$

$$\psi_1(\Delta_s) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(\Delta_s-\delta)^2}{2\sigma_0^2}\right) \tag{5}$$

and Δ_s denotes the error of contaminated observation, Δ the error of observations, ε the probability of the outlier happened or the proportion of outlier in the observations, δ the outlier, σ_0^2 the variance of the error of observations.

For example, having a set of direct observations, $x_1, \cdots, x_n$, the L_2 estimation for the unknown parameter is the mean $\bar{x}$ of the sample, and its variance is

$$\sigma_{\bar{x}}^2 = \sigma^2/n \tag{6}$$

with

$$\begin{aligned}\sigma^2 &= E[\Delta_\varepsilon - E(\Delta_\varepsilon)]^2 \\ &= E[\Delta_\varepsilon - \varepsilon\delta]^2 \\ &= \sigma_0^2 + \varepsilon(1-\varepsilon)\delta^2\end{aligned} \tag{7}$$

When $\varepsilon = 0.1$, $\sigma_0 = 1$ and $\delta = 3$, according to Eqn. (7) we have $\sigma^2 = 1.81$; and when $\delta = 5$, ε and σ_0 as before, we have $\sigma^2 = 3.25$. This shows that the variance $\sigma_{\bar{x}}^2$ rapidly increases with δ. Therefore, when there are outliers, the L_2 estimation of unknown parameter is very sensitive to them.

If the outliers are considered as a part of the stochastic model, which is called the stochastic error model, we may believe that the outlying observations are such a set of observations, which have the same mean values as the other observations but not the same variances. The probability density in Eqn. (3) is

$$\psi_0(\Delta) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{\Delta^2}{2\sigma_0^2}\right) \tag{8}$$

$$\psi_1(\Delta_s) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{\Delta_s^2}{2\sigma_1^2}\right) \tag{9}$$

where σ_0^2 is the variance of the error of observation and σ_1^2 the variance of outlier. σ^2 in Eqn. (6) is

$$\sigma^2 = (1-\varepsilon)\sigma_0^2 + \varepsilon\sigma_1^2 \tag{10}$$

The variance $\sigma_{\bar{x}}^2$ rapidly increases with σ_1^2 as that we have discussed above.

The L_1 estimation for the unknown parameters has robustness to outlying observations. This estimation has been used in the adjustment of measurements. In the following, based on their asymptotic distribution, the variances of the L_1 estimation and the L_2 estimation are compared.

3 COMPARISON OF VARIANCES

It is assumed that the adjustment model is

$$\underset{n,1}{\boldsymbol{L}} = \underset{n,p\,p,1}{\boldsymbol{AX}} + \underset{n,1}{\boldsymbol{\Delta}} \tag{11}$$

where $\boldsymbol{L}$ denotes the vector of observations, $\boldsymbol{A}$ the known coefficient matrix, $\boldsymbol{X}$ the vector of unkown parameters; $\boldsymbol{\Delta} = [\Delta_1, \Delta_2, \cdots, \Delta_n]^T$, $\Delta_1, \cdots, \Delta_n$ are independent. Considering the mean shift error model, Δ_i is the Δ_s in Eqn. (3). The probability function of Δ_i is determined by Eqns. (3), (4) and (5). Therefore

$$E(\boldsymbol{\Delta}) = \varepsilon\delta\boldsymbol{d}, \ \mathrm{Cov}(\boldsymbol{\Delta}) = \sigma^2\boldsymbol{I}_n \tag{12}$$

where $\boldsymbol{d} = [1, 1, \cdots, 1]^T$, σ^2 can be found in Eqn. (7). Similarly to Refs. [6] and [7], we have the following asymptotic distribution about the L_1 estimation of the unknown parameters

$$\hat{\boldsymbol{X}}_{L_1} \sim N(\boldsymbol{X}, W^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}) \tag{13}$$

Meanwhile, the asymptotic distribution of the L_2 estimation of unknown parameters is

$$\hat{\boldsymbol{X}}_{L_2} \sim N(\boldsymbol{X}, \sigma^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}) \tag{14}$$

where $W = 1/[2\psi(0)]$,

$$\psi(0) = (1-\varepsilon)/(\sqrt{2\pi}\sigma_0) + \varepsilon/[(\sqrt{2\pi}\sigma_0)\exp(-\delta^2/(2\sigma_0^2))]$$

According to Eqns. (13) and (14), we have to compare W^2 with σ^2. We define the relative efficiency as $Re = \sigma^2/W^2$, then we can obtain

$$Re = 2/\{\pi[1 + k^2\varepsilon(1-\varepsilon)] \cdot [1 - \varepsilon + \varepsilon \cdot \exp(-k^2/2)]^2\} \tag{15}$$

where $k = |\delta|/\sigma_0$.

If $Re = 1$, the L_1 and L_2 estimations have the same asymptotic variances. If $Re > 1$, then the L_1 estimation has a smaller asymptotic variance than the L_2 estimation and vice versa.

In general, only big outliers can be detected. Therefore, supposing $|\delta| = 4\sigma_0$[8,9] and from Eqn. (15), we can find:

(1) $\varepsilon < 4.8\%$, $Re < 1$

The L_2 estimation is superior to the L_1 estimation. Especially, when $\varepsilon = 0$, corresponding to the absence of outliers, we have $Re = 1$.

(2) $\varepsilon = 4.8\%$, $Re = 1$

The L_1 estimation and the L_2 estimation have the same asymptotic variances.

(3) $\varepsilon > 4.8\%$, $Re > 1$

The L_1 estimations is superior to the L_2 estimation, and as with the increase of ε , the relative efficiency continuously increases, the L_1 estimation will gradually be far superior to the L_2 estimation.

Considering the stochastic error model, the probability function of Δ_i is determined by Eqns. (3), (8) and (9). Therefore

$$E(\boldsymbol{\Delta}) = \boldsymbol{O},\ \mathrm{Cov}(\boldsymbol{\Delta}) = \sigma^2 \boldsymbol{I}_n \tag{16}$$

where σ^2 can be found in Eqn. (10). The relative efficiency is

$$Re = 2/[\pi(1 - \varepsilon + \varepsilon/k)^2(1 - \varepsilon + \varepsilon k^2)] \tag{17}$$

where $k = \sigma_1/\sigma_0$. Supposing $\sigma_1 = 4\sigma_0$, we have

(1) When $\varepsilon < 4.5\%$, the L_2 estimation is superior to the L_1 estimation;

(2) When $\varepsilon = 4.5\%$, the L_1 and the L_2 estimation have the same asymptotic variances;

(3) When $\varepsilon > 4.5\%$, the L_1 estimation is superior to the L_2 estimation.

In a word, when ε and outliers become very big, then the L_1 estimation is superior to the L_2 estimation, i. e., the L_1 estimation has robustness to outlying observations and has high efficiency.

4 REVISED L_2 ESTIMATION

Supposing that the no contaminated model is

$$\left.\begin{array}{l} \boldsymbol{L} = \boldsymbol{AX} + \boldsymbol{\Delta}, E(\boldsymbol{\Delta}) = \boldsymbol{O}, \\ \mathrm{Cov}(\boldsymbol{\Delta}) = \sigma_0^2 \boldsymbol{I} \end{array}\right\} \tag{18}$$

where $\boldsymbol{L} = [L_1, \cdots, L_n]^{\mathrm{T}}$,

$\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]^{\mathrm{T}}$.

If there are departures from the model (Eqn. (18)), the corresponding observations can be written as:

$$\boldsymbol{L}_\delta = \boldsymbol{L} + \boldsymbol{\delta}$$

where $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_n]^{\mathrm{T}}$ denotes the vector of the departures.

Under the mean shift model, the ith contaminated observation is

$$L_{i0} = L_i + \delta_i \tag{19}$$

where δ_i denotes the outlier of the ith observation, and δ_i is unknown. The variance of the ith contaminated observation is

$$E[L_{i0} - E(L_i)]^2 = \sigma_0^2 + \delta_i^2 \tag{20}$$

From the adjustment process, the estimate of the outlier δ_i can be obtained by

$$\hat{\delta}_i = \frac{V_i}{r_i} \tag{21}$$

where V_i denotes the ith residual error, r_i the ith local redundancy number.

When σ_0 is known, we use Baarda's test statistic,

$$W_i = \frac{V_i}{\sigma_0 \sqrt{Q_{V_iV_i}}} = \frac{V_i}{\sigma_{V_i}} \sim N(0,1) \tag{22}$$

When σ_0 is unknown, we can use τ statistic,

$$\tau_i = \frac{V_i}{\hat{\sigma}_0 \sqrt{Q_{V_iV_i}}} \sim \tau(1, n - p - 1) \tag{23}$$

where $\hat{\sigma}_0^2 = \dfrac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{PV}^{\mathrm{T}}}{(n - p)}$

At confidence level $(1 - \alpha)$, δ_i can be determined by

$$\hat{\delta}_i = \begin{cases} 0 & |W_i| \leqslant W_{\alpha/2} \\ & (|\tau_i| \leqslant \tau_{\alpha/2}) \\ V_i/r_i & |W_i| > W_{\alpha/2} \\ & (|\tau_i| > \tau_{\alpha/2}) \end{cases} \tag{24}$$

When σ_0 is known, the estimate of the variance of ith observation L_{i0} , i. e., the element of the ith main diagonal of Cov($\boldsymbol{\Delta}$) in Eqn. (18) should be changed into

$$\hat{\sigma}_{i0}^2 = \sigma_0^2 + \hat{\delta}_i^2 \tag{25}$$

When σ_0^2 is unknown, σ_0^2 can be replaced by $\hat{\sigma}_0^2$. After being tested, the estimation of Cov ($\boldsymbol{\Delta}$) can be written as $\boldsymbol{\sigma}_0^2 \boldsymbol{P}^{-1}$. Therefore, we can obtain the revised L_2 estimation under the mean shift model:

$$\hat{\boldsymbol{X}}_{L_2} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{PA})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{PL} \tag{26}$$

Under the stochastic error model, according to the theory of adjustment, we can obtain the estimate of the variance σ_{i0}^2 of L_{i0},

$$\hat{\sigma}_{i0}^2 = \frac{V_i^2}{r_i} \tag{27}$$

When σ_0^2 is known, we can choose the statistic,

$$T_i = \frac{V_i^2}{r_i \sigma_0^2} \sim \chi^2(1) \tag{28}$$

When σ_0^2 is unknown, we constitute the statistic,

$$F_i = \frac{V_i^2}{r_i(V^T V - V_i^2/r_i)/(n-p-1)} \sim F(1, n-p-1) \tag{29}$$

At confidence level $(1-\alpha)$, the variance of L_{i0} can be determined by

$$\hat{\sigma}_{i0}^2 = \begin{cases} \sigma_0^2(\hat{\sigma}_0^2) & T_i \leqslant \chi_\alpha^2 (F \leqslant F_\alpha) \\ V_i^2/r_i & T_i > \chi_\alpha^2 (F > F_\alpha) \end{cases} \tag{30}$$

After being tested, the estimation of Cov $(\boldsymbol{\Delta})$ can still be $\sigma_0^2 \boldsymbol{P}^{-1}$. The revised L_2 estimation under the stochastic error model is Eqn. (26). All the above show that the revised L_2 estimation makes use of both the advantage of determining weight of variance that according to the classical least squares theory and the idea of the posterior revised weight that according to the robust estimation theory. Therefore, the revised L_2 estimation has the advantages of above approachs and is a better estimation.

5 CONCLUSION AND EXAMPLE

In order to compare the efficiency of L_1, L_2 and revised L_2 estimation with each other and to investigate the feasibility of the main results in this paper, imitating tests have been taken for the simple average of 5 observations. The imitating errors of the observations are determined by

$$\Delta_i = \frac{1}{2}\left(\sum_{i=1}^{48} \xi_i - 24\right)$$

where ξ_i is the value of the inter Random function.

In order to assure the reliability of the results, the experiment is consisted of 10 groups; each group contains 300 imitating tests; each imitating test contains 5 observations. According to the adjusted value of the L_2 estimation and the given true value, the true error of each imitated test can be determined. According to the true errors of 300 imitating tests, the variance of the L_2 estimation can be determined. When the 5 observations are contaminated, the variances of the corresponding L_1, L_2 and revised L_2 estimation can also be determined as before.

When the contaminated model is the mean shift model, we suppose $\varepsilon = 0.05$, and δ is any value of $[0, 20]$. The results for the variance of the L_2 estimation (I) in Eqn. (18) and the variances of the revised L_2(II), the L_1(III) and the L_2(IV) estimation in the contaminated model are listed in Table 1.

Table 1 Computation of mean shift model

No.	I	II	III	IV
1	0.1840	0.1955	0.3203	1.0580
2	0.1861	0.2057	0.3239	1.0701
3	0.1845	0.2003	0.3211	1.0609
4	0.1876	0.2021	0.3265	1.0787
5	0.2033	0.2167	0.3538	1.1690
6	0.2032	0.2366	0.3537	1.1684
7	0.2062	0.2133	0.3589	1.1857
8	0.2200	0.2347	0.3829	1.2650
9	0.2056	0.2281	0.3568	1.1788
10	0.2302	0.2399	0.4001	1.3236

When the contaminated model is the stochastic error model, we can supposed that $\varepsilon = 0.05$, $\sigma_1 = 5$. The computating results are listed in Table 2.

Table 2 Computation of stochastic error model

No.	I	II	III	IV
1	0.1973	0.2262	0.3362	0.4341
2	0.2283	0.2531	0.3890	0.5023
3	0.2022	0.2230	0.3446	0.4448
4	0.1918	0.2144	0.3268	0.4220
5	0.2256	0.2473	0.3844	0.4963
6	0.2040	0.2312	0.3476	0.4488
7	0.2039	0.2180	0.3466	0.4475
8	0.1746	0.1927	0.2975	0.3841
9	0.1191	0.2042	0.3250	0.4195
10	0.1918	0.2027	0.3268	0.4220

From Table 1 and Table 2, we know:

(1) When the probability of outliers and outliers becomes very big, the L_1 estimation

method may be taken into consideration as an alternative or as a supplement to the conventional L_2 estimation. The L_1 estimation has robustness and higher efficiency.

(2) The revised L_2 estimation is superior to the L_1 estimation. It has higher efficiency than the L_1 and L_2 estimation and also is a robust estimation.

REFERENCES

1 Hampel F R. Robust Statistic: The Approach Based on Influence Functions. New York: Wiley, 1986: 1 ~20.

2 Zhu Jianjun. The Australia Surveyor, 1991, 36(2): 111~115.

3 Zhu Jianjun. Journal of Geodesy, 1996, 20: 586~590.

4 Wang Zhizhong. Journal of Central South University of Technology, 1997, 4(1): 61~64.

5 Wang Zhizhong. Trans Nonferrous Met Soc China, 1997, 7(4): 160~163.

6 Wang Song-Gui. The Theory on Linear Model and its application. Hefei: Educational Publishing House of Anhui, 1987: 1~30.

7 Wang Zhizhong. Engineering of Surveying and Mapping, (in Chinese), 1998, 7(2): 22~27.

8 Wang Zhizhong. Trans Nonferrous Met Soc China, 1999, 9(1):192~196.

9 Qu Ziqiang. PhD thesis, (in Chinese). Changsha: Central South University of Technology, 1991.

10 Wang Zhizhong. Acta Geodaetica et Cartographica Sinica, (in Chinese), 1999, 28(1): 51~56.

(Edited by He Xuefeng)