Comunicação

A NOTE ON THE CONVENTIONAL OUTLIER DETECTION TEST PROCEDURES

Considerações sobre testes convencionais de detecção de ouliers

JIANFENG GUO

Info Engineering University, P.O.Box 603, 62 Kexuedadao Rd, 450001, Zhengzhou, China
Email: jianfeng.guo@gmail.com

Abstract:

Under the assumption of that the variance-covariance matrix is fully populated, Baarda's w-test is turn out to be completely different from the standardized least-squares residual. Unfortunately, this is not generally recognized. In the limiting case of only one degree of freedom, all the three types of test statistics, including Gaussian normal test, Student's t-test and Pope's Tau-test, will be invalid for identification of outliers: (1) all the squares of the Gaussian normal test statistic coincide with the goodness-of-fit (global) test statistic, even for correlated observations. Hence, the failure of the global test implies that all the observations will be flagged as outliers, and thus the Gaussian normal test is inconclusive for localization of outliers; (2) the absolute values of the Tau-test statistic are all exactly equal to one, no matter whether the observations are contaminated. Therefore, the Tau-test cannot work for outlier detection in this situation; and (3) Student's t-test statistics are undefined.

Keywords: Outlier; Gaussian normal test; Student's t-test; Tau-test; RAIM.

Resumo:

Assumindo que a matriz Variância-Covariância das observações foi obtida, a análise dos resíduos obtidos pelos mínimos quadrados não corresponde ao teste w de Baarda. Geralmente, esta consideração não é aceita na comunidade. Em casos especiais, onde os graus de liberdade são igual a 1 (um), todos os 3 (três) tipos de testes estatísticos, incluindo o teste de normalidade de Gauss, o teste t-Students e o teste Tau-Pope, não podem ser aplicados para detecção de *outliers*: (1) todos os quadrados do teste de normalidade de Gauss, coincidem com o teste estatístico global, mesmo para observações correlacionadas. Assim, a falha do teste global implica em observações classificadas como *outliers* e, então, o teste de normalidade de Gauss é inconclusivo para a detecção de *outliers*; (2) os valores absolutos do teste Tau são exatamente iguais a um, não importando se as observações estão eivadas de erros. Portanto, nesta situação, o

teste Tau não é eficiente para detecção de *outliers*, e (3) as estatísticas do teste t-Student são indefinidas.

Palavras-chave: Oultier; Teste de normalidade de Gauss; Teste t-Student; Teste Tau; RAIM.

1. Introduction

Least-squares (LS) estimation is the predominant technique for geodetic data processing. However, LS estimation is sensitive to outliers (Wolf and Ghilani, 1997; Koch, 1999; Leick, 2004). One way to circumvent this difficulty is to adopt the conventional outlier detection test procedures (see e.g., Baarda, 1968; Pope, 1976; Kok, 1984). In these procedures, the test statistics used for outlier identification include those of the Gaussian normal test, Student's t-test and Pope's τ-test. All these test statistics can be based on Baarda's w-test or the standardized LS residual.

It should be pointed out that Baarda's w-test statistic coincides with the standardized LS residual under the diagonality assumption of the weight matrix. However, these two kinds of test statistics may be completely different in the case of correlated observations. Unfortunately, this is not generally recognized (Guo et al., 2010).

Baselga (2011) presented an angular resection problem with one redundancy, in which the observations were supposed to be uncorrelated. Numerical results state that the standardized LS residual flags all observations. One interesting problem is that of determining whether the conventional outlier detection test procedures, for uncorrelated or correlated observations, can work in the limiting case of only one degree of freedom.

The paper is organized as follows: Firstly, the conventional outlier detection test procedures are briefly overviewed. Under the condition of one redundancy, limitations of three types of test statistics used for outlier identification are addressed. An example associated with a GPS baseline RAIM (Receiver Autonomous Integrity Monitoring) scheme concludes the investigations.

2. Model description

Consider the linear Gauss-Markov model defined by Koch (1999)

$$E(\boldsymbol{L}) = \boldsymbol{A}\boldsymbol{X} \ \ \text{with} \ \ \text{Cov}(\boldsymbol{L}) = \sigma^2 \boldsymbol{P}^{-1} \tag{1}$$

where $\boldsymbol{L}$ is the n x 1 vector of normally distributed observations, $\boldsymbol{A}$ the n x u design matrix with full column rank, **X** the u x1 vector of unknowns, **P** the weight matrix of $\boldsymbol{L}$ and σ^2 the a-priori variance factor of unit weight.

The weighted LS estimator of the unknowns in model (1) is obtained as

$$\hat{\boldsymbol{X}} = (\boldsymbol{A}^T \boldsymbol{P} \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{P} \boldsymbol{L} \tag{2}$$

and the LS residual vector as

$$\boldsymbol{V} = \boldsymbol{L} - \boldsymbol{A}\boldsymbol{X} = \boldsymbol{R}\boldsymbol{L} \tag{3}$$

where $R = I - A(A^T P A)^{-1} A^T P$. The matrix R plays a pivotal role in linear adjustment techniques since it contains extremely useful information (Huber and Ronchetti, 2009; Guo, 2014).

The formulation for the most fundamental test statistic in LS estimation is given by

$$\frac{V^T P V}{\sigma^2} = \frac{\hat{\sigma}^2}{\sigma^2}(n-u) \sim \chi^2(n-u) \tag{4}$$

Under the assumption of normality, based on this statistic, the goodness-of-fit test (Wolf and Ghilani, 1997; Leick, 2004) can be performed to find out whether the adjustment results are valid.

3. Mean-shift outlier detection model

In conventional outlier detection test procedures, it is usually assumed that there is only one (possible) outlier exists in the observations (see Baarda, 1968; Pope, 1976; Baselga, 2011). Let us assume that an outlier exists in the i th observation so that the mean-shift outlier model should have read (Chatterjee and Hadi, 1988)

$$E(L) = AX + c_i \nabla_i \quad \text{with} \quad Cov(L) = \sigma^2 P^{-1}, \tag{5}$$

where $[A \quad c_i]$ has full column rank and c_i is the i th canonical unit vector with all zeros except for a one as its i th entry.

By employing the LS principle, one obtains the estimates

$$\begin{cases} \hat{\nabla}_i = (c_i^T P R c_i)^{-1} c_i^T P R L \\ \hat{X}_{[i]} = (A^T P A)^{-1} A^T P (L - c_i \hat{\nabla}_i) \end{cases} \tag{6}$$

since

$$PR = R^T P = R^T P R. \tag{7}$$

Therefore, the residual vector of the above mean-shift outlier model is given by

$$\begin{aligned} V_{[i]} &= L - c_i \hat{\nabla}_i - A\hat{X}_{[i]} \\ &= R(L - c_i \hat{\nabla}_i) = V - R c_i \hat{\nabla}_i \end{aligned} \tag{8}$$

It follows that

$$\begin{aligned} V_{[i]}^T P V_{[i]} &= L^T P (V - R c_i \hat{\nabla}_i) - 0 \\ &= L^T P V - (c_i^T P R L)^2 (c_i^T P R c_i)^{-1} \\ &= V^T P V - \sigma^2 w_i^2 \end{aligned} \tag{9}$$

in which

$$w_i = \frac{c_i^T P R L}{\sigma \sqrt{c_i^T P R c_i}} \sim \mathcal{N}(0,1) \tag{10}$$

is the i th Baarda's w-test statistic (Baarda, 1968).

4. Three kinds of test statistics used for outlier detection

Three different test statistics, including those for the Gaussian normal test, Pope's τ-test and Student's t-test, are used in conventional outlier detection procedures.

With the aid of the mean-shift outlier model (Chatterjee and Hadi, 1988; Guo, 2013), the problem of outlier identification in model (1) can be carried out in terms of the following hypothesis in model (5)

$$H_{0i} : \nabla_i^2 = 0 \ \ \text{vs.} \ \ H_{ai} : \nabla_i^2 > 0 . \tag{11}$$

The test statistic

$$T_i = w_i^2 = \frac{(c_i^T PRL)^2}{\sigma^2 c_i^T PRc_i}, \tag{12}$$

defines the uniformly most powerful test for testing H_{0i} versus H_{ai} at a given level of significance (Baarda, 1968; Kargoll, 2007; Guo et al., 2010), provided that no other outliers exist.

In practice, it is more convenient to use the square root of the test statistic T_i, namely, the Baarda's w-test statistic w_i (Baarda, 1968). If the variance factor is unknown, the parameter σ involved in the w-test may be replaced by its estimate. This will lead to either Pope's τ-test statistic

$$\tau_i = \frac{w_i}{\sqrt{\hat{\sigma}^2 / \sigma^2}} \sim \tau(n-u) . \tag{13}$$

or to Student's t-test statistic

$$t_i = \frac{w_i}{\sqrt{\hat{\sigma}_{[i]}^2 / \sigma^2}} \sim t(n-u-1) \tag{14}$$

with

$$\hat{\sigma}_{[i]}^2 = V_{[i]}^T P V_{[i]} / (n-u-1) . \tag{15}$$

The standardized LS residual

$$\tilde{v}_i = \frac{v_i}{\sigma \sqrt{c_i^T RP^{-1} c_i}} \sim \mathcal{N}(0,1) \tag{16}$$

has been extensively employed to test whether the i th observation is an outlier, both in the geodetic and the statistical literature, always assuming that the variance factor is given.

If the variance factor is unknown, either the so-called internally Studentized residual

$$\tilde{\tau}_i = \frac{\tilde{v}_i}{\sqrt{\hat{\sigma}^2 / \sigma^2}} \sim \tau(n-u) \tag{17}$$

or the externally Studentized residual

$$\tilde{t}_i = \frac{\tilde{v}_i}{\sqrt{\hat{\sigma}_{[i]}^2 / \sigma^2}} \sim t(n-u-1) \tag{18}$$

can be used instead of the standardized LS residual.

5. Limitation of conventional outlier detection test procedures

Conventional outlier detection test procedures have been used in many geodetic applications. However, can these procedures still work if there is only one degree of freedom, namely $n - u = 1$? In this limiting case, there is no redundancy in the mean-shift outlier model (5). This implies immediately that

$$V_{[i]} = 0 \tag{19}$$

which, in combination with (9) and (12) yields

$$w_i^2 = V^T P V / \sigma^2 \tag{20}$$

and thus

$$w_i^2 \equiv (n-u)\hat{\sigma}^2 / \sigma^2 \sim \chi^2(1), \quad \text{for} \quad i = 1, 2, \cdots, n \tag{21}$$

since $n - u = 1$.

By virtue of (9), (20) can be rewritten as

$$L^T \frac{R^T P c_i c_i^T P R}{c_i^T P R c_i} L = \sigma^2 w_i^2 = V^T P V = L^T P R L \tag{22}$$

Since these two quadratic forms coincide for any realization of the random observation vector L, and since both $R^T P c_i c_i^T P R$ and PR are symmetric, one can obtain

$$\frac{P R c_i c_i^T R^T P}{c_i^T P R c_i} = PR \tag{23}$$

It follows immediately that

$$\frac{c_i^T R c_i c_i^T R^T c_i}{c_i^T P R c_i} = c_i^T R P^{-1} c_i \tag{24}$$

and

$$\frac{c_i^T R c_i c_i^T P R L}{c_i^T P R c_i} = c_i^T R L \tag{25}$$

Therefore, with the definition of $\tilde{v}_i$ it follows that

$$\tilde{v}_i = \frac{c_i^T R c_i c_i^T P R L}{\sigma \cdot c_i^T P R c_i} \bigg/ \sqrt{\frac{c_i^T R c_i c_i^T R c_i}{c_i^T P R c_i}} \tag{26}$$
$$= \mathrm{sgn}(c_i^T R c_i) \cdot w_i$$

which in combination with (21) yields

$$w_i^2 = \tilde{v}_i^2 \equiv \hat{\sigma}^2 / \sigma^2 \sim \chi^2(1), \quad \text{for} \quad i = 1, 2, \cdots, n. \tag{27}$$

where $\mathrm{sgn}(\cdot)$ denotes the signum function.

Substituting (21) into (13) gives

$$\tau_i = \text{sgn}(w_i) \tag{28}$$

Analogically, the following result can be obtained

$$\tilde{\tau}_i = \text{sgn}(\tilde{v}_i) \tag{29}$$

With (21), and (27) - (29) one can conclude that, in the limiting case of only one degree of freedom, (1) the Student's t-test statistics are undefined (degenerate); (2) all the squares of Gaussian normal test coincide with the goodness-of-fit (global) test, even for correlated observations. That is, the failure of the global test implies all the observations will be flagged as outliers, and thus the Gaussian normal test is invalid for localization of outliers; and (3) the τ - test also becomes inappropriate, because the absolute values of the τ-test statistics equal to one no matter whether the observations are contaminated by outliers.

These results can be used in GPS Receiver Autonomous Integrity Monitoring (RAIM) scenario. The GPS receiver needs at least four satellites to yield a three-dimensional position (latitude, longitude, and altitude) and time solution. In order for a GPS receiver to perform RAIM or fault detection function, a minimum of five visible satellites with satisfactory geometry must be visible to it. However, six or more satellites are required to identify the faulty satellite and exclude it from the navigation solution (Parkinson and Axelrad, 1988; Brown, 1992; Langley 1999).

6. An illustrative example

The differential GPS positioning technique (Leick, 2004) has been widely used to cancel all error sources common to two GPS receivers on a short baseline. A static experimental data set was collected with a sampling interval of 1 second and a cut-off elevation angle of 15° on 3 August 2005, in Wuhan, China. The length of the baseline is approximately 6 km. At a special measurement epoch, only five satellites were visible. The satellites were PRN 19, 27, 8, 11 and 28. The numerical details of the linearized GPS double-differenced (DD) pseudorange equations are presented in Table 1.

Table 1: Numerical details of the linearized GPS DD pseudorange equations

Design matrix			Observations	Variance-covariance matrix			
0.5090	0.8925	0.4924	-1.3633	7.6103	-2.2186	0	0
0.3399	-0.3014	-0.4005	1.8606	-2.2186	4.7472	-2.5285	0
-0.8703	0.1303	0.5344	-1.5239	0	-2.5285	5.5016	-2.9731
1.2330	-0.4236	-0.5118	1.9904	0	0	-2.9731	8.8210

The calculated value of the goodness-of-fit test statistic (0.1637) is less than the critical value that corresponds to the 0.01 significance level. Therefore, the underlying adjustment appears valid. The calculated values of the Gaussian normal and Pope's τ-test statistics are listed in Table 2.

Table 2: Calculated values of some one-dimensional test statistics

Observation no.	Residual	w_i	τ_i	$\tilde{v}_i$	$\tilde{\tau}_i$
1	-0.0739	0.4046	1	-0.4046	-1
2	0.6852	0.4046	1	0.4046	1
3	0.0566	0.4046	1	0.4046	1
4	-0.4073	-0.4046	-1	-0.4046	-1

For purposes of illustration, an artificial outlier of +20 (m) is introduced to the first observation. In this situation, the calculated value of the goodness-of-fit test statistic of 10.5651 is clearly unacceptable; for, 6.6349 is the critical value at the 0.01 significance level. All the observations are flagged as outliers by using the Gaussian normal test at the 0.001 significance level (cf. Table 3).

Table 3: Results after outlier simulation of +20 m in obs. #1

Observation no.	Residual	w_i	τ_i	$\tilde{v}_i$	$\tilde{\tau}_i$
1	-0.5938	3.2504	1	-3.2504	-1
2	5.5053	3.2504	1	3.2504	1
3	0.4550	3.2504	1	3.2504	1
4	-3.2725	-3.2504	-1	-3.2504	-1

As demonstrated before, in this limiting case of only one redundancy, neither Baarda's w-test statistic nor the standardized LS residual can be used to identify the single outlying observation. Worse still, the calculated value of Pope's τ-test statistic is equal to either 1 or -1, no matter whether the observations are contaminated by outliers or not.

7. Conclusions

Following the pioneering work by W. Baarda, conventional outlier detection test procedures have been extensively adopted and investigated in geodesy. The test statistics involved in these test procedures include those of Gaussian normal test, Student's t-test, and Pope's τ-test. All these test statistics can be based on Baarda's w-test or the standardized LS residual.

However, in the situation of only one redundancy, (1) Student's t-test statistics are undefined; (2) Pope's τ-test statistics are also invalid, because their absolute values equal to one no matter whether the observations are contaminated by outliers or not; and (3) failure of the global test implies that all the observations will be flagged as outliers and, thus, the Gaussian normal test is inappropriate for the localization of outliers.

A fundamentally efficient approach to control the influence of outliers is based on robust estimation techniques. Obviously, robust estimation methods will perform inconclusively in this limiting case, if the downweighting strategy involved therein is based on any of the aforementioned test statistics.

As a result, without additional information or new observation(s), it is unrealistic to expect that the deteriorating effect of an observation with outlier on the parameter estimates can be reduced in the case of only one degree of freedom.

ACKNOWLEDGEMENT

REFERENCES

Baarda W. *A testing procedure for use in geodetic networks*. Netherlands Geod. Comm., Publ. on Geodesy, 1968, New Series, 2(5), Delft, The Netherlands.

Baselga, S. "Nonexistence of rigorous tests for multiple outlier detection in least-squares adjustment." *Journal of Surving Engineering*, 137(2011): 109-12.

Brown, R. "A baseline GPS RAIM scheme and a note on the equivalence of three RAIM methods." *Navigation*, 39(1992): 301-6.

Chatterjee, S., and Hadi, A. *Sensitivity analysis in linear regression*. New York: Wiley, 1988.

Guo, J. "Analytical quality assessment of iteratively reweighted least-squares (IRLS) method." *Boletim de Ciencias Geodesicas*, 20(2014): 132-41.

Guo, J. "The case-deletion and mean-shift outlier models: equivalence and beyond." *Acta Geodaetica et Geophysica*, 48(2013): 191-7.

Guo, J., Ou, J., and WANG, H. "Robust estimation for correlated observations: two local sensitivity-based downweighting strategies." *Journal of Geodesy*, 84(2010): 243-250.

Huber, P., and Ronchettie, M. *Robust statistics*, 2nd Ed. New York: Wiley, 2009.

Kargoll, B. "On the theory and application of model misspecification tests in geodesy." *PhD diss.*, University of Bonn, 2007.

Koch, K. *Parameter estimation and hypothesis testing in linear models*, 2nd Ed. Berlin: Springer, 1999.

Kok, J. *On data snooping and multiple outlier testing*. NOAA Technical Report, NOS NGS. 30, U.S. Department of Commerce, Rockville, MD, 1984.

Langley, R. "The integrity of GPS." *GPS World*, 10(1999): 60-3.

Leick, A. *GPS satellite surveying*, 3rd Ed. New York: Wiley, 2004.

Parkinson, B., and Axelrad, P. "Autonomous GPS integrity monitoring using the pseudorange residual." *Navigation*, 35(1988): 255-74.

Pope, A. *The statistics of residuals and the detection of outliers*. NOAA Technical Report, NOS 65, NGS 1 Rockville, MD, 1976.

Wolf, P., and Ghilani, C. *Adjustment computations: statistics and least squares in surveying and GIS*, 3rd Ed. New York: Wiley, 1997.