第41卷 第1期
2012年2月

测 绘 学 报
Acta Geodaetica et Cartographica Sinica

Vol.41, No.1
Feb., 2012

GUO Jianfeng, ZHAO Jun. Comparative Analysis of Statistical Tests Used for Detection and Identification of Outliers[J]. Acta Geodaetica et Cartographica Sinica, 2012, 41(1): 14-18.(郭建锋,赵俊. 粗差探测与识别统计检验量的比较分析[J]. 测绘学报,2012,41(1):14-18.)

粗差探测与识别统计检验量的比较分析

郭建锋[1,2],赵 俊[1]

1. 信息工程大学 理学院,河南 郑州 450001; 2. 中国科学院 测量与地球物理研究所,湖北 武汉 430077

Comparative Analysis of Statistical Tests Used for Detection and Identification of Outliers

GUO Jianfeng[1,2], ZHAO Jun[1]

1. Institute of Science, Information Engineering University, Zhengzhou 450001, China; 2. Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China

Abstract: The least-squares (LS) adjustment approach is very susceptive to outliers. A comparative analysis of statistical tests used for detection and identification of outliers, including Gaussian normal test, Tau test and Student's t-test, were addressed in details. Both the standardized local sensitivity indicator and the standardized LS residual can be served as Gaussian normal test statistics for outlier detection. However, the former is superior to the latter one for correlated observations in the sense of detection power. When the variance factor is not known, the internally Studentized residual and externally Studentized residual can be employed to perform Tau test and Student's t-test, respectively. In a parallel manner, the internally Studentized local sensitivity indicator and externally Studentized local sensitivity indicator were constructed and investigated. Since the probability of not rejecting the null hypothesis when it is false (type Ⅱ error) may be too high by using Student's t-test, and since the Tau test has a limitation in itself, both of them are not appropriate statistical tests. To circumvent this difficulty, the standard deviation involved in the standardized local sensitivity indicator can be replaced by its normalized robust least median of squares estimator, and then perform the Gaussian normal test for detection and identification of outlying observations.

Key words: outlier; sensitivity analysis; Gaussian normal test; Student's t-test; Tau test

摘 要:LS估计对粗差非常敏感。在粗差的探测与识别理论体系中,通常采用正态检验、学生氏t检验以及τ检验等,笔者对此进行比较分析。标准化局部敏感度指标与标准化LS残差均可用来作正态检验,但研究表明,当观测量相关时,前者的检验功效大于后者。先验单位权方差因子未知时,可依据内部学生化残差及外部学生化残差分别进行τ检验和学生氏t检验。与此对照,构造了内部学生化敏感度指标及外部学生化敏感度指标以代替标准化敏感度指标。由于τ检验理论本身存在固有缺陷,而学生氏t检验或将造成"纳伪"错误的增加。为此,较为稳妥的方案是仍然采用正态检验,但将标准化局部敏感度指标中的单位权中误差以其抗差LMS估计代替。

关键词:粗差;敏感度分析;高斯正态检验;学生氏t检验;τ检验

中图分类号:P227　　文献标识码:A　　文章编号:1001-1595(2012)01-0014-05

基金项目:国家自然科学基金(40874007; 41004005);国家杰出青年科学基金(40625013)

1 引 言

为保证测量成果达到设计要求,在完成实测任务后,必须进行测量数据的质量分析。大量研究表明,粗差仅仅占到观测量总数的1%至10%左右。粗差的存在往往对最小二乘(least-squares, LS)估计造成不良的影响,即LS估计的抗差性(robustness,又译为稳健性)非常差[1-12]。

对于工程技术与应用领域而言[2],抗差性可以理解为统计推断中的敏感度分析理论(或者称为扰动分析、稳定性分析理论)。换言之,抗差性,即估计量抵御粗差影响的能力,表现为平差结果对观测异常的敏感程度[5]。拟合优度检验[5-7]是检验平差成果的一项重要指标。因此,通过对拟合优度检验量进行敏感度分析构造探测与识别观测异常的统计量,具有显著的物理意义。

在粗差探测法中,假定随机模型能够客观反映观测量之间的(相对)精度及统计相关性,把粗

差问题归结为函数模型与实际模型的偏离。如果拟合优度检验结果不显著,说明在一定显著性水平上,平差成果达到了要求,可以采纳;否则就表明当前的函数模型不能准确描述观测量之间或观测量与未知参数之间的物理或者几何关系[6,13]。需要指出的是,拟合优度检验虽然能够检验出粗差的存在与否,但却不能探测和准确定位有几个观测量以及具体是哪几个观测量受到了多大量级的粗差污染[13,15]。

在粗差的探测与识别中,通常采用正态检验、学生氏 t 检验以及 τ 检验等,而构造相应的统计检验量既可基于局部敏感度指标,亦可基于 LS 残差。本文对实施粗差探测与识别的统计检验量进行了比较分析,得到如下结论:① 相关观测情形,局部敏感度指标比 LS 残差的检验功效大,若单位权中误差精确已知,宜采用基于标准化局部敏感度指标的正态检验;② 单位权中误差未知时,τ 检验理论本身存在固有缺陷,而学生氏 t 检验或将造成"纳伪"错误的增加,较为稳妥的方案是仍然采用正态检验,但将标准化局部敏感度指标中的单位权中误差以其抗差 LMS(least median of squares)估计代替。

2　模型描述

考虑如下线性 Gauss-Markov 模型[5-7]

$$\boldsymbol{L}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{e}\quad \boldsymbol{e}\sim \mathrm{N}(\boldsymbol{O},\ \sigma_0^2\boldsymbol{P}^{-1}) \tag{1}$$

式中,$\boldsymbol{A}$ 为 $n\times u(n-u>1)$ 列满秩设计阵;$\boldsymbol{X}$ 为 $u\times 1$ 未知参数向量;$\boldsymbol{L}$ 为 $n\times 1$ 观测向量;$\boldsymbol{e}$ 为相应的误差(噪声)向量,其方差-协方差阵为 $\sigma_0^2\boldsymbol{P}^{-1}$。这里对称正定阵 $\boldsymbol{P}$ 为观测量的先验权阵,而 σ_0^2 通常称为单位权方差因子。

基于 LS 原理,可得到模型(1)中未知参数的 LS 估计为[5-7]

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

相应的残差向量为

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

式中,$\boldsymbol{R}=\boldsymbol{I}-\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}$ 以矩阵形式反映平差结构,是质量的全面度量,称为平差因子阵[12]。

容易验证平差因子阵 $\boldsymbol{R}$ 幂等,并具以下有用性质

$$\boldsymbol{R}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{R}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{R} \tag{4}$$

基于此,LS 残差的加权平方和 $\Omega=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}$ 亦可表示为[13-16]

$$\Omega=\boldsymbol{L}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{R}\boldsymbol{L} \tag{5}$$

3　基于敏感度分析的统计量

3.1 标准化局部敏感度指标

将 LS 残差的加权平方和对第 i 个观测量 l_i 进行微分,得到[13-14]

$$\partial\Omega/\partial l_i=2\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V} \tag{6}$$

式中,$\boldsymbol{h}_i$ 表示第 i 个分量为 1,其余分量皆为 0 的 n 维单位向量。

显然,$\partial\Omega/\partial l_i$ 衡量的是 Ω 对第 i 个观测值的扰动的敏感程度。注意到

$$\partial\Omega/\partial l_i\sim \mathrm{N}(0,4\sigma_0^2\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{R}\boldsymbol{h}_i) \tag{7}$$

因此统计量

$$w_i=\frac{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}}{\sigma_0\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{R}\boldsymbol{h}_i}} \tag{8}$$

可用于检验 Ω 对第 i 个观测量的扰动是否"敏感"。

根据已知数据可以计算出 w 统计检验量的取值,其绝对值越大,表明 Ω 对第 i 个观测量的扰动越"敏感",故而 l_i 受到粗差污染的可能性就越大。因此,称 $|w_i|$ 为第 i 个观测量的标准化局部敏感度指标[13-14]。

应该指出的是,这里的 w_i 即为可靠性理论中 Baarda[17] 导出的 w 统计检验量。

3.2　外部学生化局部敏感度指标

通过对局部敏感度指标进行标准化,可以有效消除量纲的影响,这对于多源数据融合的质量控制问题意义尤为重要。然而,得到标准化局部敏感度指标的前提是先验单位权中误差精确已知,否则就无法利用 w 统计量进行假设检验。

在测量实践中,先验单位权中误差往往未知[5-7,12]。为此,本文提出如下服从自由度为 $n-u-1$ 的学生氏分布的统计检验量

$$t_i=\frac{w_i}{\sqrt{(\Omega/\sigma_0^2-w_i^2)/(n-u-1)}} \tag{9}$$

并将 $|t_i|$ 称为第 i 个外部学生化局部敏感度指标。

当存在多个粗差时,LS 残差的加权平方和 Ω 往往偏大。由定义不难知道,这或将导致外部学生化局部敏感度指标 $|t_i|$ 普遍偏小,进而造成"纳伪"错误的增加。因此,基于外部学生化局部敏感度指标探测和识别粗差潜在一定的风险。

3.3　内部学生化局部敏感度指标

单位权中误差未知时,还可通过构造统计量

$$\tau_i=\frac{w_i}{\sqrt{\hat{\sigma}_0^2/\sigma_0^2}} \tag{10}$$

进行粗差的探测与识别，这里 $\hat{\sigma}_0^2=\Omega/(n-u)$ 为平差模型式(1)中的验后单位权方差因子。$|\tau_i|$ 称为内部学生化局部敏感度指标。

统计量式(10)亦可表达为服从学生氏分布的统计量 t_i 的函数，即

$$\tau_i=\sqrt{n-u}\frac{t_i}{\sqrt{n-u-1+t_i^2}} \tag{11}$$

Thompson 将统计量 τ_i 服从的分布称为自由度为 $n-u$ 的 τ 分布[18]。在测量质量控制中，τ 检验是应用最为广泛的统计量之一[5,18-24]。

可证明，相互独立的统计量 w_i^2 与 $\Omega/\sigma_0^2-w_i^2$ 均服从 χ^2 分布，并且其自由度分别为 1 和 $n-u-1$。由统计理论知，服从自由度为 p 的 χ^2 分布的统计量与服从自由度分别为 $1/2$、$p/2$ 的 Gamma 分布具有相同概率分布密度[5]，因此 w_i^2 服从自由度分别为 $1/2$、$1/2$ 的 Gamma 分布，而 $\Omega/\sigma_0^2-w_i^2$ 则服从自由度分别为 $1/2$、$(n-u-1)/2$ 的 Gamma 分布。于是，可以构造出如下服从自由度分别为 $1/2$、$(n-u-1)/2$ 的 Beta 分布的统计量

$$\frac{w_i^2}{\Omega/\sigma_0^2}=\frac{w_i^2}{w_i^2+(\Omega/\sigma_0^2-w_i^2)} \tag{12}$$

Beta 分布的一个显著特点是其仅仅在单位区间(0,1)内取值[5]，于是得到

$$0\leqslant\frac{w_i^2}{\Omega/\sigma_0^2}\leqslant 1 \tag{13}$$

注意到关系式

$$\tau_i^2=(n-u)\frac{w_i^2}{\Omega/\sigma_0^2} \tag{14}$$

有

$$|\tau_i|\leqslant\sqrt{n-u} \tag{15}$$

式(15)表明，服从 τ 分布的统计量之绝对值存在上界，而且该上界仅取决于该统计量的自由度。

如果 τ 检验的临界值大于 $\sqrt{n-u}$，那么在给定的显著性水平上，无论观测量受到了多大量级的粗差污染，都不可能利用内部学生化局部敏感度指标将其检验出来，这是 τ 检验理论本身固有的一个缺陷[18,20]。

4 基于残差的统计量

在经典测量平差中仅涉及独立等权观测数据，这种情况下，线性最小二乘平差理论中最基本的关系式 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{PV}=\boldsymbol{O}$ 退化为

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{O} \tag{16}$$

因此，在传统的粗差探测与识别中，均以残差作为对象研究问题。

4.1 标准化残差

若先验单位权中误差精确已知，可构造如下称为标准化残差的统计量[17]

$$\tilde{v}_i=\frac{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{V}}{\sigma_0\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}}\sim \mathrm{N}(0,1) \tag{17}$$

探测和识别粗差。

当第 i 个观测量发生大小为 $|\delta_i|$ 的均值漂移时，统计检验量 w_i 和标准化残差 $\tilde{v}_i$ 的分布函数引起的平移量(即非中心化参数)分别为

$$|E(w_i)|=\frac{|\delta_i|}{\sigma_0}\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{PRh}_i} \tag{18}$$

及

$$|E(\tilde{v}_i)|=\frac{|\delta_i|}{\sigma_0}\frac{|r_i|}{\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}} \tag{19}$$

式中，r_i 表平差因子阵 $\boldsymbol{R}$ 的第 i 个对角元。

依据 Cauchy-Schwarz 不等式，有

$$\begin{aligned}&\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{PRh}_i=\\&(\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1/2}\cdot\boldsymbol{P}^{-1/2}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i)(\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{P}^{1/2}\cdot\boldsymbol{P}^{1/2}\boldsymbol{Rh}_i)\geqslant\\&(\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1/2}\cdot\boldsymbol{P}^{1/2}\boldsymbol{Rh}_i)^2=r_i^2\end{aligned} \tag{20}$$

由此得到

$$|E(w_i)|\geqslant|E(\tilde{v}_i)| \tag{21}$$

一个统计检验量的检验功效是显著性水平和非中心化参数的单调增函数[25]，因此统计量 w_i 比标准化残差 $\tilde{v}_i$ 的检验功效要大，或者等价的，$T_i=w_i^2$ 比统计量 $\widetilde{T}_i=\tilde{v}_i^2$ 的检验功效要大。

事实上，统计量 T_i 为一致最大检验功效统计量[24]。也就是，在给定的显著性水平上，利用 T_i(或 w_i)进行假设检验犯“纳伪”错误的概率比使用任何其他的统计量都要小。

相反，若事先指定显著性水平和检验功效，统计量所对应的非中心化参数将唯一确定，由式(21)立即可知：一致最大检验功效统计量 T_i(或 w_i)对应的最小可探测粗差指标[16-17,25]不会超过统计量 $\widetilde{T}_i$(或标准化残差 $\tilde{v}_i$)，换言之，一致最大检验功效统计量 T_i(或 w_i)较统计量 $\widetilde{T}_i$(或 $\tilde{v}_i$)对粗差更敏感。

由于相关观测情形下统计量 w_i 比标准化残差 $\tilde{v}_i$ 的检验功效要大，而在独立观测情形二者则完全一致，因此建议采用统计检验量 w_i 进行粗差的探测和识别。

4.2 外部学生化残差

先验单位权中误差未知时，可构造如下统计检验量

$$\tilde{t}_i=\frac{\tilde{v}_i}{\sqrt{(\Omega/\sigma_0^2-\tilde{v}_i^2)/(n-u-1)}} \tag{22}$$

这称为外部学生化残差[4]。

根据关系式 $\boldsymbol{RP}^{-1}=\boldsymbol{RP}^{-1}\boldsymbol{R}^{\mathrm{T}}$，容易验证矩阵

$$\left(\sigma_0^{-2}\boldsymbol{PR}-\sigma_0^{-2}\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}\right)cov(\boldsymbol{L})=\boldsymbol{R}^{\mathrm{T}}-\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}$$

为幂等阵，注意到

$$tr\left(\boldsymbol{R}^{\mathrm{T}}-\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}\right)=n-u-1$$

并顾及 $cov(\boldsymbol{L})=\sigma_0^2\boldsymbol{P}^{-1}$ 可逆，因此 $\sigma_0^{-2}\boldsymbol{PR}-\sigma_0^{-2}\dfrac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}$ 的秩亦为 $n-u-1$。

综合上款，二次型

$$\Omega/\sigma_0^2-\tilde{v}_i^2=\boldsymbol{L}^{\mathrm{T}}\left(\sigma_0^{-2}\boldsymbol{PR}-\sigma_0^{-2}\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}\right)\boldsymbol{L} \tag{23}$$

服从自由度为 $n-u-1$ 的 χ^2 分布[5]。

由于

$$\frac{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\sigma_0\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}}cov(\boldsymbol{L})\left(\sigma_0^{-2}\boldsymbol{PR}-\sigma_0^{-2}\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}\right)=\boldsymbol{O}$$

依据正态随机向量的线性组合与其二次型相互独立的判定定理[5]可知，标准化残差 $\tilde{v}_i$ 与 $\Omega/\sigma_0^2-\tilde{v}_i^2$ 相互统计独立，因而，外部学生化残差 $\tilde{t}_i$ 服从自由度为 $n-u-1$ 的学生氏分布。

与外部学生化局部敏感度指标类似，当存在多个粗差时，统计量 $\tilde{t}_i$ 或潜在一定的风险。

4.3　内部学生化残差

若单位权中误差未知，还可构造如下称之为内部学生化残差的统计量[26]

$$\tilde{\tau}_i=\frac{\tilde{v}_i}{\sqrt{\hat{\sigma}_0^2/\sigma_0^2}} \tag{24}$$

进行粗差的探测与识别。

由于

$$\left(\sigma_0^{-2}\boldsymbol{PR}-\sigma_0^{-2}\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}\right)cov(\boldsymbol{L})\sigma_0^{-2}\frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{h}_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{R}}{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}=\boldsymbol{O}$$

根据正态随机向量的两个二次型相互独立的判定定理[5]，$\tilde{v}_i^2$ 与 $\Omega/\sigma_0^2-\tilde{v}_i^2$ 相互统计独立。因此，统计量

$$\frac{\tilde{v}_i^2}{\Omega/\sigma_0^2}=\frac{\tilde{v}_i^2}{\tilde{v}_i^2+(\Omega/\sigma_0^2-\tilde{v}_i^2)} \tag{25}$$

服从自由度分别为 $1/2$、$(n-u-1)/2$ 的Beta分布。进而，内部学生化残差统计量 $\tilde{\tau}_i$ 服从自由度为 $n-u$ 的 τ 分布[18]。

假定除第 i 个观测位置受到了大小为 $|\delta_i|$ 的粗差污染外，其余均为“干净”的观测值，则有

$$\tilde{v}_i(\delta_i)=\frac{\boldsymbol{h}_i^{\mathrm{T}}\hat{\boldsymbol{V}}}{\sigma_0\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}}+\frac{r_i\delta_i}{\sigma_0\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i}} \tag{26}$$

及

$$\Omega(\delta_i)=\boldsymbol{L}^{\mathrm{T}}\boldsymbol{PRL}+\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Ph}_i\delta_i+\delta_i\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{PRL}+\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{PRh}_i\delta_i^2 \tag{27}$$

于是，当扰动量 δ_i 趋于无穷大时，第 i 个内部学生化残差的绝对值之极限为

$$\lim_{\delta_i\to\infty}|\tilde{\tau}_i(\delta_i)|=\sqrt{n-u}\lim_{\delta_i\to\infty}\frac{|\tilde{v}_i(\delta_i)|/\delta_i}{\sqrt{\Omega(\delta_i)/\sigma_0^2}/\delta_i}=$$

$$\sqrt{n-u}\frac{|r_i|}{\sqrt{\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{RP}^{-1}\boldsymbol{h}_i\cdot\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{PRh}_i}} \tag{28}$$

这个结果由 Baselga[20]给出。

若顾及不等式(20)，还可以进一步求出上述极限值的上界

$$\lim_{\delta_i\to\infty}|\tilde{\tau}_i(\delta_i)|\leqslant\sqrt{n-u} \tag{29}$$

式(29)再次验证了这样一个事实，即 τ 检验理论本身确乎存在缺陷。因而，使用 τ 统计量探测和识别粗差存在一定风险。

5　结论与建议

(1) 若单位权中误差精确已知，可采用正态检验。w 统计量反映的是 χ^2 拟合优度检验量对观测值扰动的敏感程度，因而具有明确的物理意义；作为一致最大检验功效统计量，对于给定的显著性水平和检验功效，$T_i=w_i^2$(或 w_i)能够探测出量级最小的粗差。

因此，进行正态检验时，w 统计量为首选，标准化残差次之。

(2) 若 σ 未知，可采用 τ 检验或 t 检验。τ 检验理论本身固有缺陷；而存在多个粗差时，t 检验或将造成“纳伪”错误的增加，亦存在一定的风险。

从检验功效的角度考虑，无论进行 τ 检验抑或 t 检验，均建议采用基于局部敏感度指标的检验量。

(3) Robust 正态检验。由于 τ 检验和 t 检验均存在一定缺陷，因此尚需对单位权方差因子未知时的粗差探测与识别作进一步的讨论。

一种较为稳妥的解决方案是，采用具有明确物理意义的 w 统计量，而统计量中的未知参数 σ 则以其抗差 LMS 估计代替之[1-3,12-15]。即以

$$\tilde{\sigma}=1.482\,6\sqrt{\underset{i}{\mathrm{median}}(\sigma_0^2w_i^2)} \tag{30}$$

代替统计量 w_i 中的先验单位权中误差 σ。

将基于修正的 w 统计量的检验称为 Robust 正态检验。从数学上说，修正的 w 统计量并不严格服从正态分布。然而经验表明，该统计量具有较强的抗差性，当冗余观测较多时尤为如此[1-3,12-15]。

参考文献：

[1] HUBER P J, RONCHETTI E M. Robust Statistics[M]. 2nd ed. New York: Wiley, 2009.

[2] HAMPEL F R, RONCHETTI E M, ROUSSEEUW P J, et al. Robust Statistics: The Approach Based on Influence Functions [M]. New York: Wiley, 1986.

[3] ROUSSEEUW P J, LEROY A M. Robust Regression and Outlier Detection[M]. New York: Wiley, 1987.

[4] CHATTERJEE S, HADI A S. Sensitivity Analysis in Linear Regression[M]. New York: Wiley, 1988.

[5] KOCH K R. Parameter Estimation and Hypothesis Testing in Linear Models[M]. 2nd ed. Berlin: Springer-Verlag, 1999.

[6] LEICK A. GPS Satellite Surveying[M]. 3rd ed. New York: Wiley, 2004.

[7] WOLF P R, GHILANI C D. Adjustment Computations: Statistics and Least Squares in Surveying and GIS[M]. 3rd ed. New York: Wiley, 1997.

[8] ZHOU Jiangwen. Classical Theory of Errors and Robust Estimation[J]. Acta Geodaetica et Cartograghica Sinica, 1989, 18(2): 115-120. (**周江文. 经典误差理论与抗差估计**[J]. **测绘学报**, 1989, 18(2): 115-120.)

[9] OU Jikun. Quasi-accurate Detection of Gross Errors (QUAD)[J]. Acta Geodaetica et Cartograghica Sinica, 1999, 28(1): 15-20. (**欧吉坤. 粗差的拟准检定法**(QUAD**法**)[J]. **测绘学报**, 1999, 28(1): 15-20.)

[10] SONG Lijie, YANG Yuanxi. Comparison between Data Snooping and LEGE[J]. Acta Geodaetica et Cartograghica Sinica, 1999, 28(4): 295-300. (**宋力杰, 杨元喜. 均值漂移模型粗差探测法与** LEGE **法的比较**[J]. **测绘学报**, 1999, 28(4): 295-300.)

[11] LI Xinna, GUI Qingming, XU Apei. Besian Method for Detection of Gross Errors Based on Classification Variables[J]. Acta Geodaetica et Cartograghica Sinica, 2008, 37(3): 355-360. (**李新娜, 归庆明, 许阿裴. 基于识别变量的粗差探测的** Bayes **方法**[J]. **测绘学报**, 2008, 37(3): 355-360.)

[12] ZHOU Jiangwen, HUANG Youcai, YANG Yuanxi, et al. Robust Least Squares Method [M]. Wuhan: Huazhong University of Science and Technology Press, 1997. (**周江文, 黄幼才, 杨元喜, 等. 抗差最小二乘法**[M]. **武汉: 华中理工大学出版社**, 1997.)

[13] GUO Jianfeng. Theory of Model Errors and its Applications in GPS Data Processing[D]. Wuhan: Institute of Geodesy and Geophysics of Chinese Academy of Sciences, 2007. (**郭建锋. 模型误差理论若干问题研究及其在** GPS **数据处理中的应用**[D]. **武汉: 中科院测量与地球物理研究所**, 2007.)

[14] GUO J F, OU J K, WANG H T. Quasi-accurate Detection of Outliers for Correlated Observations[J]. Journal of Surveying Engineering, 2007, 133(3): 129-133.

[15] GUO J K, OU J K, WANG H T. Robust Estimation for Correlated Observations: Two Local Sensitivity-based Downweighting Strategies [J]. Journal of Geodesy, 2010, 84(4): 243-250.

[16] GUO J K, OU J K. Variation Characteristics of MDBs in Robust Estimation[J]. AllgVerm-Nachr, 2010, 117(2): 49-52.

[17] BAARDA W. A Testing Procedure for Use in Geodetic Networks[J]. Netherlands Geod Comm Publ on Geod, 1968, 2(5): 1-97.

[18] POPE A J. The Statistics of Residuals and the Detection of Outliers[R]. Rockville: NOAA Technical Report, Nos 65, NGS 1, 1976.

[19] KOK J J. On Data Snooping and Multiple Outlier Testing[R]. Rockville: NOAA Technical Report, Nos NGS 30, 1984.

[20] BASELGA S. Critical Limitation in Use of τ Test for Gross Error Detection[J]. Journal of Geodesy, 2007, 133(2): 52-55.

[21] CROSS P A, PRICE D R. A Strategy for the Distinction between Single and Multiple Gross Errors in Geodetic Networks[J]. Manuscr Geod, 1985, 10(3): 172-178.

[22] DING X, COLEMAN R. Multiple Outlier Detection by Evaluating Redundancy Contributions of Observations [J]. Journal of Geodesy, 1996, 70(8): 489-498.

[23] SNOW K B, SCHAFFRIN B. Three-dimensional Outlier Detection for GPS Networks and Their Densification via the BLIMPBE Approach [J]. GPS Solutions, 2003, 7(2): 130-139.

[24] KARGOLL B. On the Theory and Application of Model Misspecification Tests in Geodesy[D]. Bonn: University of Bonn, 2007.

[25] TEUNISSEN P J G. Quality Control in Integrated Navigation Systems [C] // Proceedings of the IEEE PLANS90, Nevada: IEEE, 1990: 158-165.

[26] COOK R D. Detection of Influential Observations in Linear Regression[J]. Technometrics, 1977, 19(1): 15-18.

(责任编辑:丛树平)

收稿日期: 2010-12-31

修回日期: 2011-06-03

第一作者简介: **郭建锋**(1972—), **男, 博士, 副教授, 硕士生导师, 研究方向为误差理论与 GNSS 数据处理。**

First author: GUO Jianfeng (1972—), male, PhD, associate professor, MA students' supervisor, majors in errors theory and GNSS data processing.

E-mail: jianfeng.guo@gmail.com