文章编号: 1000-050X(2002)06-0607-04　　文献标识码: A

可靠性分析与数据探测

陶本藻[1]　姚宜斌[1]

(1　武汉大学测绘学院, 武汉市珞喻路 129 号, 430079)

摘　要: 首先从方差比检验出发, 建立了用于可靠性分析和数据探测的统一基本关系式, 由此导出了在相关权情况下的可靠性指标 R_i 的表达式, 证明在权阵为对角阵时其特例就是常用的多余观测分量内可靠性指标 r_i, 其一致性说明了 R_i 的合理性。同时, 提出了与可靠性度量密切联系的一种新的数据探测方法, 实现了可靠性分析和数据探测在理论上的统一和实用上的一致。

关键词: 相关权; 可靠性; 数据探测; 模型误差
中图法分类号: P207.2

可靠性分析和数据探测是平差系统质量控制的重要内容, 其定义和基本理论于 1968 年由巴尔达(Baarda)提出[1], 并已得到公认、发展。随着 GPS 等现代技术的发展, 平差系统中的权阵经常并非对角阵, 为此, 可靠性理论也需要相应地扩展至相关权的一般情形。文献[3] 在相关权的情况下, 对多余观测分量 r_i 的值域进行了讨论, 提出再用 r_i 度量相关权下的可靠性是不合适的, 论证了相关权的情况下, 内、外可靠性的一致性, 并给出了统一的可靠性指标 R_i, 它与文献[2] 给出的局部可靠性指标相同, 但在 R_i 与 r_i 的关系等方面, 文献[3] 比文献[2] 的研究进了一步。文献[4] 在文献[3] 的基础上着重讨论了在 GPS 中的应用。文献[5] 也论证了内、外可靠性的一致性, 也建议用一个指标来度量平差系统的可靠性, 讨论了统一可靠性指标的建立方法。

巴尔达提出的主要是一维数据探测, 后经许多学者研究, 模型误差的统计检验方法已比较成熟。由于一维数据探测简单实用, 许多平差软件包括 GPS 商用软件大都已引用这种方法, 但数据探测的研究没有与可靠性理论相联系。

本文遵循巴尔达的可靠性和数据探测的基本思想, 在最优统计检验原则下, 导出了用于数据探测和可靠性分析的基本关系式。在此基础上, 分别给出了可靠性指标和数据探测方法, 将两者从理论上和实用上统一起来。此时讨论的平差系统的权阵不只局限于对角阵。由此, 可靠性指标和数据探测方法均适用于相关权阵情况。

收稿日期: 2002-08-16。
项目来源: 国家自然科学基金资助项目(40004001)。

1　基本关系式

平差线性模型为:

$$\boldsymbol{l} = \boldsymbol{AX} + \Delta \quad \boldsymbol{D}_l = \sigma_0^2 \boldsymbol{Q} \tag{1}$$

平差残差为:

$$\boldsymbol{V} = \boldsymbol{AX} - \boldsymbol{l} = -\boldsymbol{Q}_{VV}\boldsymbol{Pl} \tag{2}$$

单位权方差估值为:

$$\hat{\sigma}_0^2 = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{PV}/f \tag{3}$$

式中, f 为平差系统的自由度。假设 Δ 为随机误差, 则 $E(\Delta)=0$, 否则, 存在模型误差(粗差、系统误差) $E(\Delta)=\varepsilon$。作统计检验:

$$H_0: E(\Delta) = 0 \quad H_1: E(\Delta) = \varepsilon$$

在 H_0 成立下, 方差比

$$T = \frac{\hat{\sigma}_0^2}{\sigma_0^2} = \frac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{PV}}{f\sigma_0^2}$$

为自由度(f, ∞)的中心 F 分布统计量, 且 T 的期望 $E(T)=1$。

现讨论 H_1 成立的情况。设存在的模型误差为 ε, 由式(2)可得残差的影响项:

$$\boldsymbol{V}_\varepsilon = -\boldsymbol{Q}_{VV}\boldsymbol{P}\varepsilon \tag{4}$$

此时, 相应的方差比为:

$$E\left(\frac{\hat{\sigma}_0^2}{\sigma_0^2}\right) = 1 + \frac{\boldsymbol{V}_\varepsilon^{\mathrm{T}}\boldsymbol{PV}_\varepsilon}{f\sigma_0^2}$$

式中顾及了 $E(\boldsymbol{V})=0$ 和 $E\left(\frac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}}{f\sigma_0^2}\right)=1$，令

$$\lambda^2=\frac{\boldsymbol{V}_\varepsilon^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}_\varepsilon}{\sigma_0^2} \tag{5}$$

则上式为：

$$E\left(\frac{\hat{\sigma}_0^2}{\sigma_0^2}\right)=1+\frac{\lambda^2}{f} \tag{6}$$

式中，λ^2/f 为方差比的期望偏离。在 H_1 成立的条件下，统计量 T 服从非中心 F 分布，即

$$T\sim F(f,\infty,\lambda^2)$$

选定显著水平 α 和检验功效 $1-\beta$，相应的分位值为 T_α，则其概率表达式为：

$$P(T<T_\alpha/H_0)=1-\alpha \tag{7}$$

$$P(T>T_\alpha/H_1)=\beta \tag{8}$$

式(4)以 σ_0 为单位表示，则有：

$$V_\varepsilon/\sigma_0=-\boldsymbol{Q}_{VV}\boldsymbol{P}\varepsilon/\sigma_0 \tag{9}$$

并令

$$\varepsilon/\sigma_0=\boldsymbol{g}k \tag{10}$$

式中，k 为尺度因子，则式(5)可写成：

$$\lambda^2=\left(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{g}\right)k^2 \tag{11}$$

式中顾及了 $\boldsymbol{Q}_{vv}\boldsymbol{P}$ 为幂等阵，即 $(\boldsymbol{Q}_{vv}\boldsymbol{P})^2=\boldsymbol{Q}_{vv}\boldsymbol{P}$，或

$$k=\lambda\Big\backslash\left(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{g}\right)^{1/2} \tag{12}$$

代入式(10)得：

$$\varepsilon=\boldsymbol{g}k\sigma_0=\boldsymbol{g}\lambda\sigma_0\Big\backslash\left(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{g}\right)^{1/2} \tag{13}$$

令 $\boldsymbol{g}^{\mathrm{T}}=(0,0,\cdots,g_i=1,\cdots,0,0)=\boldsymbol{e}_i^{\mathrm{T}}$，即 $\boldsymbol{g}$ 中仅第 i 行元素为1，其余为零，记为 $\boldsymbol{e}_i$，表示仅局限于讨论一个模型误差情况，则式(12)和式(13)为：

$$k_i=\lambda\Big\backslash\left(\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{e}_i\right)^{1/2} \tag{14}$$

$$\varepsilon_i=k_i\sigma_0 \tag{15}$$

式(15)表示第 i 个模型误差为 k_i 倍单位权标准差，而 k_i 的决定取决于 λ 和平差系统的结构。式(15)就是本文给出的讨论可靠性和数据探测的基本关系式。

2 可靠性度量

由式(15)可知，ε_i 未知，为估计可发现的模型误差 ε_i，必须用统计方法估计 λ。

根据统计检验理论，由于抽样的随机性，弃真和纳伪错误不可避免，因此，在作检验时，必须合理地选取显著水平 α，在不改变 α 的条件下尽可能增大检验功效 $1-\beta$，这就是所谓的最优假设检验原则。

现选取显著水平 $\alpha=\alpha_0$，检验功效 $1-\beta=1-\beta_0$，$1-\beta_0$ 一般定为 0.8，由此，就可求得参数 $\lambda^2=\lambda_0^2$(如 $\alpha_0=0.05$，$1-\beta_0=0.8$，$\lambda_0=2.8$，见文献[6])。代入式(14)、式(15)得：

$$k_{0i}=\lambda_0\Big\backslash\left(\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{e}_i\right)^{1/2} \tag{16}$$

$$\varepsilon_{0i}=k_{0i}\sigma_0 \tag{17}$$

式(17)就是在选定 α_0、$1-\beta_0$ 前提下，所估计的可发现的最小模型误差 ε_{0i}，小于 ε_{0i} 的模型误差是不可能发现的。将

$$\sigma_i^2=\sigma_0^2\boldsymbol{q}_{ii} \tag{18}$$

代入式(17)，$\boldsymbol{q}_{ii}$ 为 l_i 的权倒数，为 $\boldsymbol{Q}$ 阵中第 i 行对角线元素，对于相关权阵，$\boldsymbol{q}_{ii}\neq1/\boldsymbol{p}_{ii}$，可得：

$$\varepsilon_{0i}=\lambda_0\sigma_i\Big\backslash\left(\boldsymbol{q}_{ii}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{e}_i\right)^{1/2} \tag{19}$$

令

$$R_i=\boldsymbol{q}_{ii}\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{VV}\boldsymbol{P}\boldsymbol{e}_i \tag{20}$$

则式(19)改写为：

$$\varepsilon_{0i}=\lambda_0\sigma_i\Big\backslash\sqrt{R_i} \tag{21}$$

式中，λ_0、σ_i 为常量；ε_{0i} 与 $\sqrt{R_i}$ 成反比，R_i 愈大，ε_{0i} 愈小，表示可发现的最小模型误差愈小，该观测愈可靠，可靠性大，反之亦然。

因此，采用 R_i 作为可靠性指标是适宜的，这就是文献[3]建议的基于相关权阵的可靠性指标。

特殊地，当 $\boldsymbol{P}$ 为对角阵时，式(20)为：

$$R_i=\boldsymbol{q}_{ii}\boldsymbol{p}_i\boldsymbol{Q}_{VV}\boldsymbol{p}_i=\boldsymbol{p}_i\boldsymbol{Q}_{V_iV_i}=r_i$$

r_i 为多余观测分量，具有 $\sum_{i=1}^{n}r_i=r=f$(r 为平差系统的多余观测数)及 $0\leqslant r_i\leqslant1$。相应地式(21)为：

$$\varepsilon_{0i}=\lambda_0\sigma_i\Big\backslash\sqrt{r_i} \tag{22}$$

式中，r_i 为权阵为对角阵时的可靠性指标。式(22)就是巴尔达给出的讨论内可靠性的关系式。

由上述可见，可靠性一般指标为 $\boldsymbol{R}_i$，$\boldsymbol{r}_i$ 为其特殊(对角权阵)情况的指标。这种一致性说明在相关权情况下，采用 $\boldsymbol{R}_i$ 作为相关权阵时的可靠性指标是 $\boldsymbol{r}_i$ 的自然推广。

但文献[3]已证明在相关权阵情况下，$\boldsymbol{R}_i$ 的值域不是[0,1]，能否改进和是否需要改进是一个值得研究的问题。由于 λ_0 和 $\sigma_i=\sigma_0\sqrt{\boldsymbol{q}_{ii}}$ 与实际观测无关，所以在平差前根据已设计的控制网结构和施测精度计算 r_i，也就确定了每一观测中可发现的最小模型误差。r_i 的大小反映了观测值在平差结构中相对可靠性大小，而 ε_{0i} 则表示可发现的误差绝对值大小。

因此，可以考虑在计算可靠性指标 R_i 后，再计算 ε_{0i}(式(21))，可更清楚地看出平差系统的抗模型误差的能力，不必要求可靠性指标一定在[0,1]之间。

3　数据探测

常用的一组粗差检验统计量为:

$$u = v_j \Big\backslash \sigma_0 \sqrt{Q_{v_j v_j}} \tag{23}$$

当 H_0: $E(v_j)=0$ 时, $u \sim N(0, 1)$, 而拒绝域为 $|u| > u_{\alpha/2}$。

如 $\boldsymbol{P}$ 非对角, 则可采用统计量:

$$u = |\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{v}| \Big\backslash \sigma_0 \left(\boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}_{vv}\boldsymbol{P}\boldsymbol{e}_j\right)^{1/2} \tag{24}$$

实用上式(23)最为简便。但利用式(24)检验时, 仅考虑合理地选取 $\alpha=\alpha_0$, 并未考虑检验功效 $1-\beta$ 的选取, 从统计检验观点看, 它不是最优检验。此外, 在平差系统中, 经可靠性分析, 每一观测的可发现最小误差已经确定, 可以认为如果平差后该观测的误差大于可发现的最小误差, 就可判为模型误差, 反之, 则认为是随机误差。于是可靠性和数据探测对于同一平差系统在理论上得到了统一, 实用上也一致。

为此, 笔者建议数据探测采用如下判别式:

$$\left.\begin{aligned} H_0: E(v_j)=0, \text{则}\ |v_j| < \varepsilon_{0j} \\ H_1: E(v_j)\neq 0, \text{则}\ |v_j| > \varepsilon_{0j} \end{aligned}\right\} \tag{25}$$

式(25)中的 ε_{0i} 按式(22)($\boldsymbol{P}$ 对角)或式(21)计算。

以上导出的式(25)就是笔者提出的一维数据探测公式。它是直接利用可靠性分析的结果进行检验, 使平差结构的可靠性分析与数据探测统一了起来。平差前在计算 $R_i(r_i)$后, 再计算 ε_{0i}, 假定平差系统只有一个模型误差, 像用式(23)那样逐次检验一个模型误差, 但这里要选取的最大模型误差是$\left(|v_j|-\varepsilon_{0j}\right)$, 而不是$|v_j|$的最大值。

4　结　语

本文从平差系统方差比检验统计量出发, 导出用于可靠性和数据探测(一维)的统一基本关系式(14)、式(15)。将模型误差表示为与系统结构、显著水平、检验功效和施测精度有关的关系式。

当 $\boldsymbol{P}$ 为非对角阵时, 式(21)就是用于可靠性分析和一维数据探测的统一关系式, 其中 R_i 为可靠性指标, 可发现最小模型误差 ε_{0i}同时作为判别残差 v_i 是否存在模型误差的分位值。式(22)为式(21)中 $\boldsymbol{P}$ 对角阵时的特例, r_i 为可靠性指标, $|v_j| > \varepsilon_{0j}$, 表示第 i 观测存在模型误差。这里要注意的是, 平差系统一次仅假定只有一个模型误差。

由于 R_i 与 r_i 的一致性, 说明选用 R_i 为相关权阵时可靠性指标是合理的。

本文提出数据探测新公式(25), 在理论上是优化的, 既考虑最优统计检验原则, 又与可靠性分析所得到的可发现的最小误差相联系, 且权阵不局限于对角阵。

以上是延续巴尔达思想体系进行讨论和发展的, 近年来, 相关权时可靠性分析的讨论有所发展。施闯和刘经南则从另一途径提出了基于相关分析的粗差理论(不同于巴尔达出发思想)[7, 8], 也给出了相应的可靠性指标和粗差检验方法, 并在国家高精度 GPS 网整体平差中得到了应用。

平差系统的相关可靠性分析和数据探测尚需要进一步研究。

参　考　文　献

1　Baarda W. A Testing Procedure for Use in Geodesy Networks Neth. Geod. Publ. on Geodesy, New Series., 1968, 2(5)

2　Pelzer H. Some Criteria for the Accuracy and Reliability of Networks. DGK, 1980, 252: 55~67

3　王金岭, 陈永奇. 论观测值的可靠性度量. 测绘学报, 1994, 23(4): 252~258

4　黄声享, 王金岭. 多余观测分量与可靠性度量指标研究. 武汉测绘科技大学学报, 1997, 22(2): 114~118

5　欧吉坤. 一种统一的可靠性指标研究. 见: 周江文, 杨元喜, 欧吉坤编. 测量误差理论新探. 北京: 地震出版社, 1999. 45~57

6　陶本藻. 测量数据统计分析. 北京: 测绘出版社, 1992. 135~140, 170~174

7　施　闯, 刘经南. 基于相关分析的粗差理论. 武汉测绘科技大学学报, 1998, 23(1): 5~9

8　施　闯, 刘经南. 国家高精度 GPS 网整体平差中的粗差分析. 武汉测绘科技大学学报, 1999, 24(2): 107~111

作者简介: 陶本藻, 教授, 博士生导师。现主要从事现代测量数据处理和地壳形变地球动力学解释的研究。代表成果: 青藏高原地壳运动监测及地球动力学机制研究; 测量平差模型和模型误差的理论;《测量数据统计分析》;《自由网平差与变形分析》, 等。发表论文百余篇。

E-mail: tbz0204@yeah.net.cn

Reliability Analysis and Data Snooping

TAO Benzao[1] *YAO Yibin*[1]

(1 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

Abstract: Reliability analysis and data snooping are important contents in quality control of adjustment system. Baarda's reliability analysis and data snooping method is designed for independent observations and it is quite consummate both in theory and in practice. But they are researched respectively and constitute different systems. In this paper, we establish a uniform basic equation used for reliability analysis and data snooping from verifying the ratio of variance, then the expression of reliability index based on the correlation weight is obttained. We also prove that the expression applied to find minimal model errors and the weight matrix is not restricted to diagonal matrix. When the weight matrix is diagonally, the redundant observations' inner reliability index r_i is a particular case of R_i. The consistency proves that R_i is reasonable. At the same time, a new data snooping method is presented, which is related to reliability analysis measurement. The consistency of reliability analysis and data snooping both in theory and in practice is a new viewpoint, which will be used in other fields of surveying data processing.

Key words: correlation weight; reliability; data snooping; model error

About the author: TAO Benzao, professor, Ph. D supervisor. His interested fields are theoretic research on data processing and geophysics interpretation of crustal defomation. His main achievements are the research project of monitoring the present-day crustal movements and studying its geodynamical mechanism in Qinghai-Tibet Plateau; the research project of the theory of adjustment model and its error; "The Monograph of Statistical Analysis of Surveying Data"; "The Monograph of Deformation Analysis of Free Net Adjustment". etc. His published papers are more than 100.

E-mail: tbz0204@yeah. net. cn

(上接第 606 页)

Unbiased Estimation Formula of Unit Weight Mean Square Error in Regularization Solution

SHEN Yunzhong[1] *LIU Dajie*[1]

(1 Department of Surveying and Geoinformatics, Tongji University, 1239 Siping Road, Shanghai, China, 200092)

Abstract: Regularization method is an effective method in solving ill-posed equation. In this paper the unbiased estimation formula of unit weight mean square error in the regularization solution is derived and the formula is verified with numerical test of 1 000 sample data by use of the typical ill-posed equation, i. e. the Fredholm integration equation of the first kind.

Key words: regularization solution; mean square of unit weight; unbiased estimation

About the author: SHEN Yunzhong, professor. He is engaged in the researfch on geodesy.