文章编号: 1000-050X(2000)05-0409-13

关于平差残差和单位权中误差的统计分析

陶本藻[1]

(1　武汉测绘科技大学地学测量工程学院, 武汉市珞喻路 129 号, 430079)

摘　要　从残差的定量关系出发评述了检测粗差的数据探测法和抗差估计的选权迭代法, 指出了其中用残差代替真误差所产生的问题及解决办法, 并且导出了单位权方差估值、单位权中误差估值的概率分布, 给出了后者的数字特征, 对平差问题的精度体系作了统计分析。

关键词　平差残差; 数据探测; 抗差估计; 单位权中误差
分类号　P207　文献标识码　A

在测量平差理论中, 平差残差 $\boldsymbol{V}$ 是一个重要的统计量, 在误差估计、检验和分析中都要应用残差的概率分布和数字特征。残差与真误差有严格的关系式, 残差的概率分布可由真误差的概率分布转换而来。从定性上分析, 可认为残差是相应真误差的估值, 但从定量观点看, 残差 v_i 的大小不一定能反映 Δ_i 的大小, 而是一组真误差的联合影响。从理论上研究残差与相应观测值的真误差间的关系, 这是本文要讨论的第一个问题。

粗差检测的数据探测法以 v_i 来检测粗差 Δ_i, 抗差估计的迭代权法以 $\boldsymbol{V}$ 代替 Δ 建立权函数。对此存在的问题并作出评述是本文讨论的第二个问题。

现行平差精度评定体系是以单位权方差估值无偏为基础的, 而此时单位权中误差有偏。本文导出了单位权中误差估值的概率分布和数字特征以及区间估计和假设检验的表达式, 对已有的应用公式作了理论推导, 对现有这方面的统计理论作了补充。

1　残差 $\boldsymbol{V}$ 与真误差 Δ

高斯-马尔柯夫模型为:

$$\begin{aligned} \boldsymbol{l} &= \boldsymbol{A}\boldsymbol{X} + \Delta \\ D(\boldsymbol{l}) &= \sigma_0^2\boldsymbol{Q} = \sigma_0^2\boldsymbol{P}^{-1} \end{aligned} \tag{1}$$

误差方程为:

$$\boldsymbol{V} = \boldsymbol{A}\hat{\boldsymbol{X}} - \boldsymbol{l} \tag{2}$$

在最小二乘准则 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V} = \min$ 下, 其解为:

$$\hat{\boldsymbol{X}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{l} = \boldsymbol{N}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{l} = \boldsymbol{Q}_{\hat{x}\hat{x}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{l} \tag{3}$$

将式(1)代入式(3)得:

$$\hat{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{Q}_{\hat{x}\hat{x}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\Delta \tag{4}$$

上式表明, 无偏估计 $\hat{\boldsymbol{X}}$ 与真值 $\boldsymbol{X}$ 之差取决于真误差 Δ 的大小, 并与网形结构有关。

残差为:

$$\boldsymbol{V} = (\boldsymbol{A}\boldsymbol{Q}_{\hat{x}\hat{x}}\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{Q})\boldsymbol{P}\boldsymbol{l} = -\boldsymbol{Q}_{vv}\boldsymbol{P}\boldsymbol{l} \tag{5}$$

令 $\boldsymbol{R} = \boldsymbol{Q}_{vv}\boldsymbol{P}$, 则 $\boldsymbol{V} = -\boldsymbol{R}\boldsymbol{l}$, 且

$$\boldsymbol{R} = \begin{bmatrix} r_1 & \cdots & r_{1n} \\ \cdots & & \cdots \\ r_{n1} & \cdots & r_n \end{bmatrix}$$

式中, $r_{ii} = r_i\ (1 \leqslant i \leqslant n)$。考虑函数模型式(1), 则有:

$$\boldsymbol{V} = -\boldsymbol{Q}_{vv}\boldsymbol{P}\Delta = -\boldsymbol{R}\Delta \tag{6}$$

式中,

$$\begin{aligned} v_i = &-r_{i1}\Delta_1 - r_{i2}\Delta_2 - \cdots - r_i\Delta_i - \cdots - \\ &r_{in}\Delta_n = -\sum_{j=1}^{n} r_{ij}\Delta_j = -\boldsymbol{R}_i^{\mathrm{T}}\Delta \end{aligned} \tag{7}$$

亦即任一观测值的残差 v_i 是通过系数 r_{ij} 作用的所有观测值真误差 $\Delta_j\ (j = 1, 2, \cdots, n)$ 的线性函数。这是平差问题中残差 $\boldsymbol{V}$ 与真误差 Δ 的基本关系式。$\boldsymbol{R}_i$ 为 $\boldsymbol{R}$ 阵中的第 i 列阵。

当观测值间不相关, 即 $\boldsymbol{P}$ 为对角阵时, 已经证明, $\boldsymbol{R}$ 中主对角线元素分别是各观测值的多余观测分量, 且有:

$$\sum_{j=1}^{n} r_j = f$$

收稿日期: 2000-03-07.

式中，$0\leqslant r_j\leqslant 1$。一般 $r_j> r_{ij}$，f 为平差问题的自由度，即多余观测数。r_{ij}的计算式为：

$$r_{ij} = I_{ij} - \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{Q}_{\hat{x}\hat{x}}\boldsymbol{a}_j\boldsymbol{P}_j \tag{8}$$

如果 $\Delta_j \sim N(0, \sigma_j^2)$，各 Δ_j 间不相关，则由式(7)得残差 v_i 的概率分布为：$v_i \sim N(0, \sigma_{v_i}^2)$，且

$$\sigma_{v_i}^2 = \boldsymbol{R}_i^{\mathrm{T}}D(\Delta)\boldsymbol{R}_i = \sigma_0^2\boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}_i = \sigma_0^2\boldsymbol{Q}_{v_iv_i}$$

以文献[1]中一个水准网实例进一步剖析残差与真误差的线性关系，如图1。现仅列出 $\boldsymbol{R}_1$ 的计算结果：$r_1=0.51$，$r_{12}=-0.18$，$r_{13}=-0.33$，$r_{14}=-0.30$，$r_{15}=-0.08$，$r_{16}=-0.26$，$r_{17}=-0.18$，$r_{18}=0.10$，$r_{19}=0.08$。

由此结果大致可以了解：$r_1> r_{1j}$ $(j=2, 3, \cdots, 9)$，与路线 l_1 的两端点高程 x_1、x_2 有关联的 r_{ij}数值较大，如 r_{13}、r_{14}、r_{16}、r_{12}、r_{17}；无关联的如 r_{15}、r_{18}、r_{19}很小，趋于零。此网残差 v_1 与 Δ_j 的关系为：

$$\begin{aligned} v_1 = &0.51\Delta_1 - 0.18\Delta_2 - 0.33\Delta_3 - \\ &0.30\Delta_4 - 0.08\Delta_5 - 0.26\Delta_6 - \\ &0.18\Delta_7 + 0.10\Delta_8 + 0.08\Delta_9 \end{aligned}$$

可见，残差 v_1 是各 Δ_j 综合作用的结果，很难看出 v_1 中主成分是 Δ_1 的影响项。

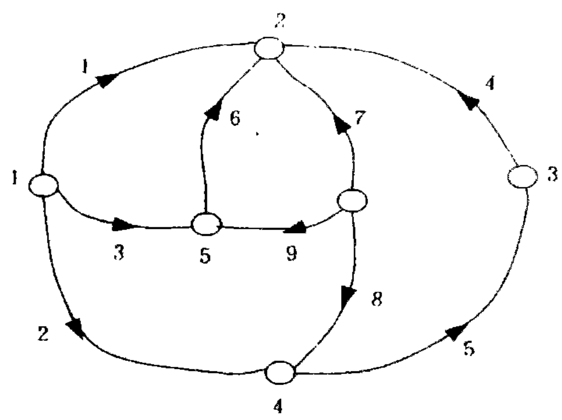

图 1　水准网

Fig. 1　Leveling Network

综上所述，在线性模型平差问题中，残差 v_i 与真误差 Δ_j $(j=1, 2, \cdots, n)$间存在线性关系，残差的概率分布与数字特征取决于真误差的概率分布与数字特征，两者具有严格的函数和统计关系，残差和真误差是两个不同的统计量。在现有文献中，都提到 v_i 是 Δ_i 的估值这个概念，实际上是指函数模型(1)经过平差求得参数 $\boldsymbol{X}$ 的估值 $\hat{\boldsymbol{X}}$，将其代入函数模型求得残差，对比式(1)与式(2)，相应的 Δ 用 $\boldsymbol{V}$ 代替，称为 $\boldsymbol{V}$ 是 Δ 的估值。用 $\hat{\boldsymbol{X}}$ 代替 $\boldsymbol{X}$ 后，$\boldsymbol{V}$ 是 Δ 的估值，这是一个先决条件，所以这仅是一种描述式(1)和式(2)中 $\boldsymbol{V}$ 与 Δ 关系的概念性说法，而不是指它们之间的定量关系。

2　对数据探测法、抗差估计迭代权法的评述

在检测粗差实践中，常用巴尔达的数据探测法和抗差估计的迭代权法，前者以残差 v_i 为统计量检测相应观测值的粗差 Δ_i，后者以 v_i 代替真误差 Δ_i 建立权函数，对具有大到一定程度的大残差赋予小权，使该观测值在平差中不起作用，达到平差结果中不受粗差影响的目的。这两种方法都牵涉到 v_i 与 Δ_i 的关系问题，为此需要加以评述。

粗差检测理论基于将粗差视为一种模型误差，其方法是将粗差归入函数模型，此时平差问题中的 n 个观测误差分为两部分：随机误差 Δ_t 和粗差 Δ_s，其概率分布为：

$$\Delta_t \sim N(0, \sigma_r^2),\ \sigma_r^2 = \sigma_0^2\boldsymbol{Q}_{rr}$$

$$\Delta_s \sim N(\varepsilon_s, \sigma_s^2),\ \sigma_s^2 = \sigma_0^2\boldsymbol{Q}_{ss}$$

设前 q 个观测误差为粗差，后 $n-q$ 个为随机误差，则由式(7)知，残差 v_i 的分布为：

$$v_i \sim N(E(v_i), \sigma_{v_i}^2)$$

式中，

$$E(v_i)=\sum_{s=1}^{q}(-r_{is}\varepsilon_s) \tag{9}$$

相应地，式(7)可改为：

$$v_i=-r_{i1}\varepsilon_1-\cdots-r_{iq}\varepsilon_q-\sum_{t=q+1}^{n}r_{it}\Delta_t \tag{10}$$

因 $\varepsilon\gg\Delta$，故上式可记为：

$$v_i=-r_{i1}\varepsilon_1-\cdots-r_{iq}\varepsilon_q \tag{11}$$

此时的残差 v_i 综合了多个粗差的影响，其大小取决于多个粗差通过相应系数 r 的作用结果。所以 v_i 中并不反映单一的ε_i 结果，其大小显然不等于ε_i 之值。

巴尔达提出的数据探测法以只存在一个粗差为前提，且假定观测值间不相关。如果 v_i 中不含粗差，$E(v_i)=0$，则 $v_i/\sigma_{v_i}\sim N(0, 1)$，检验的概率表达式为：

$$P(u_{-\frac{\alpha}{2}}< (v_i/\sigma_{v_i})< u_{\frac{\alpha}{2}}) = 1-\alpha \tag{12}$$

即检验统计量 ω_i 的拒绝域为：

$$\omega_i = |v_i/\sigma_{v_i}| > u_{\frac{\alpha}{2}} \tag{13}$$

如果平差问题确定只有一个粗差如 ε_1，由式(11)知，$v_i=-r_{i1}\varepsilon_1$，在 n 个残差中均有反映。由于 $r_i> r_{ij}$，此时 $v_1=-r_1\varepsilon_1$，而 v_1/σ_{v_1} 比其他的标准化残差要大的概率极大，由 v_1 检测 Δ_1 可认为合理。

如果实际存在少数几个粗差如 2 ~3 个, 此时反映粗差情况就比较复杂, 而且有相互抵消的可能, 但只要含有粗差的观测个数很少, 统计量 v_i/σ_{v_i} 中最大的、次大的一个值, 由于 $r_i > r_{ij}$ 以及有的 $r_{ij} \propto 0$, 其中相应观测值含有粗差的概率是最大的。巴尔达提出一次只检验一个粗差的检验顺序, 一般可保证检验前两个粗差的正确性。3 个以上粗差的检验能否正确, 理论上就很难说了。从前面的水准网实例看, 如果观测值 l_1、l_5、l_8、l_9 存在粗差, 由于 ε_5、ε_8、ε_9 对 v_1 影响极小, 用 v_1 检测 ε_1 不存在问题。如果存在 ε_1、ε_3(或 ε_4) 两个粗差, 检验 ε_1 也不会存在多大问题。如果存在 ε_1、ε_3、ε_4 三个粗差, 检验正确性就可能得不到保证, 因此粗差在网中的分布也是一个因素。

从理论上分析, 只要观测值间不相关, 平差问题只存在很少量含有粗差的观测值, 通过规定的一次只检测一个粗差的程序, 有可能保证检验的正确性, 这不失为一个较好的检测粗差的方法。虽然 v_i 不等于 $\Delta_i(\varepsilon_i)$, 如果检验 $E(v_i) \neq 0$, 则其中存在 ε_i 可信。

抗差估计选权迭代法, 认为是将粗差归入随机模型的方法, 即其概率分布为:

$$\Delta_t \sim N(0, \sigma_r^2),\ \Delta_s \sim N(0, \sigma_s^2)$$

其中, $\sigma_s^2 \gg \sigma_r^2$。当 σ_s^2 很大, 超出随机模型的方差限值时, 其权很小, 该观测粗差实际在平差中予以删除。理论上, 权函数应是 Δ 的函数, 但 Δ 未知, 用其相应的残差代替, 这种代替是抗差估计迭代权法公认的主要弱点。

其实归入随机模型的提法仅指用迭代权方法消除粗差影响。用残差代替真误差定权, 实际上已考虑了残差与粗差、真误差关系式(10)。也就是说, 顾及了归为函数模型的情形。抗差估计迭代权法用残差定权函数存在的问题与数据探测法所分析的基本一致, 当只存在少量粗差时, 由于粗差 ε_q 主要反映在残差 v_q 上, 权函数的确定比较合理, 权函数的迭代也相当于数据探测法逐次检测粗差, 抗差估计理论也是基于有少量粗差的平差问题。

在抗差估计中, 若残差 $\boldsymbol{V}$ 的初值来自于最小二乘平差, 由式(10)或式(11)知, 可能出现几个较大残差, 第一次定权不太正确可能会影响最后平差结果, 这个问题需要研究改进。而在数据探测法中, 一次只检测一个粗差, 就要好得多。

数据探测和抗差估计要研究的问题很多, 这里笔者仅从残差与粗差的基本关系着手进行初步探讨。

3　单位权方差估值的数字特征和概率分布

对于高斯 -马尔柯夫线性模型, 利用二次型分布定理: 设 $\boldsymbol{X}$ 服从 $N(\xi, \Sigma)$ 分布, $\boldsymbol{M}$ 为对称阵, $\boldsymbol{M}\Sigma$ 幂等, 则二次型 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}$ 服从非中心化的 χ^2 分布, 即

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X} \sim \chi^{2\prime}(R(\boldsymbol{M}), \xi^{\mathrm{T}}\boldsymbol{M}\xi)$$

式中, $\boldsymbol{M}$ 的秩为 χ^2 的自由度; $\lambda = \xi^{\mathrm{T}}\boldsymbol{M}\xi$ 为非中心化参数。已证明, 测量平差中的带权残差平方和 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}$ 服从 $\lambda = 0$、自由度为 f 的中心化 χ^2 分布[2], f 即为平差中的多余观测数, 即

$$\frac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}}{\sigma_0^2} \sim \chi^2_{(f)} \tag{14}$$

而且

$$E(\hat{\sigma}_0^2) = E\left(\frac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}}{f}\right) = \sigma_0^2$$

$$D(\hat{\sigma}_0^2) = \frac{2}{f}\sigma_0^4$$

亦即平差中单位权方差估计无偏。

统计量 $\hat{\sigma}_0^2$ 是一随机量, 可导出其概率分布, 由式(14)知:

$$\hat{\sigma}_0^2 = \frac{\sigma_0^2}{f}\chi^2_{(f)} \tag{15}$$

设随机量 x 的密度函数为 $f(x)$, y 的密度函数为 $g(y)$, 则有分布变换公式:

$$f(x)\,|\mathrm{d}x| = g(y)\,|\mathrm{d}y| \tag{16}$$

已知 y 服从 $\chi^2_{(f)}$ 分布, 其密度函数为:

$$g(y) = \frac{1}{2^{f/2}\,\Gamma\left(\frac{f}{2}\right)}(y)^{\frac{f}{2}-1}\mathrm{e}^{-\frac{1}{2}y}$$

令 $x = \hat{\sigma}_0^2$, 则 $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{f}{\sigma_0^2}$, 代入式(16), 即得 $\hat{\sigma}_0^2$ 的密度函数为:

$$f(\hat{\sigma}_0^2) = \frac{f}{\sigma_0^2}g(y) = \frac{f}{\sigma_0^2}g\left(\frac{f}{\sigma^2}\hat{\sigma}_0^2\right) = \frac{1}{\Gamma\left(\frac{f}{2}\right)}\left(\frac{f}{2\sigma_0^2}\right)^{\frac{f}{2}}\hat{\sigma}_0^{2\left(\frac{f}{2}-1\right)}\mathrm{e}^{-\frac{f}{2\sigma_0^2}\hat{\sigma}_0^2} \tag{17}$$

式中, $\Gamma(\ )$ 为 Gamma 函数, 并且 $0 < \hat{\sigma}_0^2 < \infty$。

由式(15)即可得单位权方差估值的 χ^2 检验公式:

$$P\left(\frac{f\hat{\sigma}_0^2}{\sigma_0^2} > \chi^2_{\alpha(f)}\right) = \alpha \tag{18}$$

χ^2_α 为显著水平 α 的右尾分位值, $f\hat{\sigma}_0^2 = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{V}$。

4　单位权中误差估值的概率分布及其性质

考虑单位权中误差估值：

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

由分布变换公式(16)，设 $x = \hat{\sigma}_0$, $y = \hat{\sigma}_0^2$, $\frac{\mathrm{d}y}{\mathrm{d}x} = 2\hat{\sigma}_0$，则 $\hat{\sigma}_0$ 的密度函数 $f(\hat{\sigma}_0)$ 就可以从方差估值密度函数 $f(\hat{\sigma}_0^2)$ 变换而得：

$$f(\hat{\sigma}_0) = \frac{2}{\Gamma(\frac{f}{2})}\left(\frac{f}{2\sigma_0^2}\right)^{\frac{f}{2}}\hat{\sigma}_0^{(f-1)}\mathrm{e}^{-\frac{f}{2\sigma_0^2}\hat{\sigma}_0^2} \quad (19)$$

现求 $\hat{\sigma}_0$ 概率分布的 r 阶原点矩，顾及密度函数式(19)，由 r 阶原点矩定义式得：

$$a_r = E(\hat{\sigma}_0^r) = \int_0^{\infty}\hat{\sigma}_0^r f(\hat{\sigma}_0)\mathrm{d}\hat{\sigma}_0$$

$$= \frac{2}{\Gamma(\frac{f}{2})}\left(\frac{f}{2\sigma_0^2}\right)^{\frac{f}{2}}\int_0^{\infty}\hat{\sigma}_0^{(f+r-1)}\mathrm{e}^{-\frac{f}{2\sigma_0^2}\hat{\sigma}_0^2}\mathrm{d}\hat{\sigma}_0 \quad (20)$$

令 $$Z = \frac{f}{2\sigma_0^2}\hat{\sigma}_0^2,\ \hat{\sigma}_0 = \left(\frac{2\sigma_0^2}{f}\right)^{\frac{1}{2}}Z^{\frac{1}{2}}$$

则 $$\mathrm{d}\hat{\sigma}_0 = \frac{1}{2}Z^{-\frac{1}{2}}\left(\frac{2\sigma_0^2}{f}\right)^{\frac{1}{2}}\mathrm{d}Z$$

代入式(20)即得：

$$a_r = \frac{1}{\Gamma(\frac{f}{2})}\left(\frac{2\sigma_0^2}{f}\right)^{\frac{r}{2}}\int_0^{\infty}Z^{\frac{1}{2}(f+r)-1}\mathrm{e}^{-Z}\mathrm{d}Z$$

$$= \left(\frac{2\sigma_0^2}{f}\right)^{\frac{r}{2}}\frac{\Gamma(\frac{f+r}{2})}{\Gamma(\frac{f}{2})} \quad (21)$$

当 $r=1$，可得 $\hat{\sigma}_0$ 的数学期望为：

$$a_1 = E(\hat{\sigma}_0) = \left(\frac{2\sigma_0^2}{f}\right)^{\frac{1}{2}}\frac{\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f}{2})} = H_f\sigma_0 \quad (22)$$

式中， $$H_f = \left(\frac{2}{f}\right)^{\frac{1}{2}}\Gamma\left(\frac{f+1}{2}\right)\Big\backslash\Gamma\left(\frac{f}{2}\right) \quad (23)$$

将 H_f 展成级数，采用较精确的司特令公式可得：

$$H_f = 1 - \frac{1}{4(f+1)} - \frac{7}{32(f+1)^2} - \cdots \approx \sqrt{1-\frac{1}{2f}} \quad (24)$$

所以，$E(\hat{\sigma}_0) \neq \sigma_0$，$\hat{\sigma}_0$ 是 σ_0 的有偏估计量。

在测量平差精度评定理论中，是以单位权方差估值的无偏估计为统计基础的，此时单位权中误差估值有偏。由式(24)大致可以看出，随着平差问题多余观测数的增多，单位权中误差估值愈来愈接近无偏，就是 f 不太大，如 $f=6$，仅偏 5%。若多余观测数太少，偏的百分比可能要大一些，这也说明平差问题中多余观测数的重要性。

如果要使单位权中误差估值无偏，其单位权方差估值必为有偏，这就要改变现有的估计公式，测量平差问题以及其他领域数据统计处理，均采用单位权方差估值无偏这一统计前提。

当 $r=2$ 时，由式(20)得 $\hat{\sigma}_0$ 的二阶原点矩为：

$$a_2 = \frac{2\sigma_0^2}{f}\Gamma\left(\frac{f+2}{2}\right)\Big\backslash\Gamma\left(\frac{f}{2}\right) = \sigma_0^2 \quad (25)$$

这里顾及 Gamma 函数性质：

$$\Gamma\left(\frac{f+2}{2}\right) = \frac{f}{2}\Gamma\left(\frac{f}{2}\right)$$

由此得单位权中误差估值的方差为：

$$D(\hat{\sigma}_0) = a_2 - a_1^2 = \sigma_0^2 - H_f^2\sigma_0^2 = (1-H_f^2)\sigma_0^2 \quad (26)$$

或 $$D(\hat{\sigma}_0) \approx \frac{1}{2f}\sigma_0^2 \quad (27)$$

$\hat{\sigma}_0$ 的中误差为：

$$\sigma_{\hat{\sigma}_0} \approx \frac{1}{\sqrt{2f}}\sigma_0 \quad (28)$$

如果 $f=2$，$\sigma_{\hat{\sigma}_0} = \frac{1}{2}\sigma_0$；$f=4$，$\sigma_{\hat{\sigma}_0} = 0.35\sigma_0$。可见，当 f 太小时，估值 $\hat{\sigma}_0$ 很不正确，又一次说明平差问题多余观测数的重要性。

由式(15)知：

$$\hat{\sigma}_0 = \sqrt{\frac{\sigma_0^2}{f}\chi^2_{(f)}} \quad (29)$$

设 $\chi^2_{\alpha(f)}$ 为 $\chi^2_{(f)}$ 的右尾分位值，则有：

$$\alpha = P(\chi^2_{(f)} > \chi^2_{\alpha(f)}) = P\left(\sqrt{\chi^2_{(f)}} > \sqrt{\chi^2_{\alpha(f)}}\right)$$

由此得：

$$\alpha = P\left(\sqrt{\frac{\sigma_0^2}{f}\chi^2_{(f)}} > \sqrt{\frac{\sigma_0^2}{f}\chi^2_{\alpha(f)}}\right) = P\left(\hat{\sigma}_0 > \sqrt{\frac{\sigma_0^2}{f}\chi^2_{\alpha(f)}}\right) \quad (30)$$

式中右端最后一项是 $\hat{\sigma}_0$ 的右尾分位值：

$$\hat{\sigma}_{0\alpha} = \sqrt{\frac{\sigma_0^2}{f}\chi^2_{\alpha(f)}} \quad (31)$$

同理可得左尾分位值：

$$\hat{\sigma}'_{0\alpha} = \sqrt{\frac{\sigma_0^2}{f}\chi^{2\prime}_{\alpha(f)}} \quad (32)$$

由此可得概率表达式为：

$$P(\hat{\sigma}'_{0\alpha/2} < \hat{\sigma}_0 < \hat{\sigma}_{0\alpha/2}) = 1-\alpha$$

此式可用于 $\hat{\sigma}_0$ 的区间估计和假设检验。

参 考 文 献

1　李庆海，陶本藻．概率统计原理和在测量中的应用．北京：测绘出版社，1990

2　陶本藻．测量数据统计分析．北京：测绘出版社，1992

3　陶本藻．自由网平差与变形分析．北京：测绘出版社，1984

陶本藻，男，65 岁，教授，博士生导师。现从事现代测量数据处理理论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》；《自由网平差变形分析》等。发表论文百余篇。

E-mail：Tbz@0204yeah.net.cn

Statistical Analysis of Residual Error and Mean Square of Weight

TAO Benzao[1]

(1　School of Geo-science and Surveying Engineering, WTUSM, 129 Luoyu Road, Wuhan, China, 430079)

Abstract　As an important statistic, the residual error of adjustment is the estimation of the true error of observation in nature, but quantitatively it is a linear function of a group of true errors, whose values reflect the corporate effects of many true errors. By considering the quantitative relation between the residual error and the true error, this paper focuses on the data snooping and the iterative methods of weight defining for the robust estimation which are used to detect gross error. We point out the problems that will be met during the course of substituting the true error for residual error and how to resolve them. The statistical feature of mean square of weight is also discussed in this paper, so we deduce the estimation of the variance of weight and the statistical distribution of the estimation of mean square of weight, and also prompt the mathematical features of the latter. The statistical analysis of adjustment precise system is discussed.

Key words　residual error; data snooping; robust estimation; mean square of weight

TAO Benzao, male, 65, professor, Ph. D supervisor. His interested fields are theory research on data process, geophysics interpret of crustal deformation. 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: Tbz@0204yeah.net.cn