

高等学校教学用书

测量误差检验与数据处理

姚应生 董烈镛 编

煤炭工业出版社

TD172
Y-932

高等学校教学用书

测量误差检验与数据处理

姚应生 董烈镛 编

煤炭工业出版社

739923

(京)新登字042号

内 容 提 要

本书着重采用概率论和数理统计原理,阐述测量数据处理与检验分析的方法。内容包括:观测数据的统计检验、粗差和异常数据取舍、统计运算与最小二乘拟合以及相关观测值、高程网、交会测量、导线测量、三边测量、边角测量、三维网、形变监测网等的数据处理及控制网的精度及可靠性理论。

本书可作为高等院校测量专业本科生选修课或研究生必修课的教材,也可作为测绘工作者的参考书。

责任编辑:洪 敏

高 等 学 校 教 学 用 书 测量误差检验与数据处理

姚应生 董烈镛 编

煤炭工业出版社 出版

(北京安定门外和平里北街21号)

煤炭工业出版社印刷厂 印刷

新华书店北京发行所 发行

开本 $787 \times 1092 \text{mm}^1/16$ 印张 $20^{1/2}$

字数481千字 印数1—1,505

1991年9月第1版 1991年9月第1次印刷

ISBN 7-5020-0564-1/TD·519

书号 3339

定价 5.20元

前 言

本书是高等院校矿山测量或工程测量专业学生，在学习了高等数学、线性代数、概率论与数理统计以及测量平差等课程的基础上，进一步提高实际测量数据处理能力的本科生选修课或研究生必修课的教材。

测量的数据处理，从理论到方法都在更新，而且发展迅速，目前有从一维、二维观测的数据处理到三、四维观测的数据处理；从静态到动态的数据处理。本书着重介绍各种常规测量方法的数据处理，以静态为主。

为了便于学生的自学和扩大知识面，本书取材广泛，论述详细，文字通俗易懂，插图形象、醒目，计算表格简明直观，而且与目前广泛采用的计算器具相适应。因此，本书也可作为广大测量技术人员的参考书与自学用书。

本书第1~5、7章，第6章第1~3节及第8章第3节由西安矿业学院姚应生同志编写，第9~12章、第6章第4节和第8章第1、2节由山东矿业学院董烈镛同志编写，初稿完成后，互相交换审阅，提出修改意见，最后执笔者修改定稿。书稿编写完成后，山东矿业学院施允春副教授审阅了全书，韩晓冬同志帮助做了一些具体工作。在此特向给予帮助和提供方便的同志表示衷心的感谢。

本书的编写是在教学工作之暇进行的，时间仓促，加之编者水平有限，难免存在缺点和错误，恳切希望读者予以批评、指正。

编 者

1990.7

113 4108

目 录

第一章 观测误差与误差统计	1
第一节 观测数据的随机误差与误差统计	1
第二节 观测数据的系统误差	10
第三节 误差的合成	14
第四节 粗差及异常数据的取舍	16
第二章 观测数据的统计检验	23
第一节 几种常用的概率分布	23
第二节 参数的假设检验	27
第三节 观测数据合理性检验	30
第四节 统计分布检验	32
第五节 观测数据相关性检验	38
第三章 统计运算与最小二乘拟合	43
第一节 方差分析	43
第二节 回归分析	51
第三节 相关分析	61
第四节 曲线拟合	64
第五节 最小二乘拟合的递推计算	74
第六节 最小二乘拟合推估	76
第四章 基本测量问题和相关观测值的数据处理	79
第一节 重复观测值的数据处理	79
第二节 双观测值的数据处理	80
第三节 随机值的相关描述	82
第四节 子样相关系数及其估计	83
第五节 随机向量的相关阵	84
第六节 相关观测值的广义算术平均值	86
第七节 相关观测量与非相关观测量的一起处理	88
第五章 一维高程网的数据处理	90
第一节 高差的权	90
第二节 水准网的相关平差	94
第三节 顾及起始数据误差影响的水准网平差	97
第四节 水准测量误差传播及精度估算	101
第六章 交会测量的数据处理	105
第一节 角度交会测量及其误差椭圆	105
第二节 测边交会测量及其误差椭圆	113
第三节 边、角交会测量及其误差椭圆	117
第四节 边、角交会典型网平差	118
第七章 导线测量的数据处理	138

第一节	导线测量成果的归算	138
第二节	导线及导线网的精度评定	140
第三节	导线和导线网的相关平差	142
第四节	顾及起始数据误差影响的导线和导线网平差	152
第五节	顾及系统误差影响的导线网平差	155
第六节	顾及起始数据误差和系统误差影响的导线和导线网平差	159
第七节	导线网分期平差	160
第八节	无定向导线的平差	167
第八章	测边网的数据处理	172
第一节	独立测边网的数据处理	172
第二节	附合测边网的数据处理	178
第三节	用邻边比率法进行三边测量的数据处理	191
第九章	边角网的数据处理	207
第一节	边角网条件平差	207
第二节	边角网坐标平差	215
第三节	边角网的方差估计	223
第十章	矿区三维控制网的数据处理	232
第一节	三维控制网间接平差	232
第二节	三维控制网条件平差	245
第三节	误差椭球	249
第十一章	矿区控制网的精度和可靠性	252
第一节	控制网的精度准则	252
第二节	粗差对平差改正数的影响	257
第三节	单个备选假设下的可靠性理论	262
第四节	粗差定位方法	268
第十二章	矿山形变监测网的数据处理	277
第一节	概述	277
第二节	加权秩亏网平差	279
第三节	秩亏网平差	288
第四节	拟稳平差	292
第五节	附加阵 G 的确定	295
第六节	自由网平差特点	298
第七节	自由网平差的实际应用	302
附表1	标准正态分布表	304
附表2	χ^2 分布表	306
附表3	t 分布表	308
附表4	F 分布表	309
附表5	夏皮罗-威尔克 a_{in} 系数表	312
附表6	夏皮罗-威尔克 $w(n, \alpha)$ 表	314
附表7	相关系数检验法的临界值	314
附表8	循环序列相关系数 R 分布的临界值	315
参考文献		316

第一章 观测误差与误差统计

如果我们对某量进行重复观测,就会发现这些观测值之间往往存在一些差异。如果已经知道某几个量之间应满足某一理论关系,但实际观测结果往往不能满足应有的理论关系,例如平面三角形三个内角之和不等于 180° 。为什么会产生这些差异呢?不难理解,这是观测值中包含有观测误差的缘故。

观测误差的产生,原因很多,概括起来有三个方面的原因,即测量仪器本身有误差,观测者感觉器官的鉴别能力有一定局限性,外界环境条件如湿度、温度、风力、大气折光等因素的影响。由于这三方面的原因,使观测结果产生误差。

根据观测误差对观测结果的影响性质,可将观测误差分为系统误差和随机误差两种。在测量工作的整个过程中,除了这两种误差外,还可能产生错误。错误将大大影响测量成果的可靠性,必须采取适当的方法和措施,加以消除。

系统误差与随机误差在观测过程中,总是同时产生的。当观测中有显著的系统误差时,随机误差就居于次要地位,观测误差就呈现出系统的性质。反之,即呈现出随机性质。系统误差对于观测结果的影响,一般具有累积的作用,它对成果精度的影响也特别显著。

如观测列中已排除系统误差的影响,或者与随机误差相比,系统误差已处于次要地位,则该观测列中主要是存在着随机误差。这样的观测列,称为带有随机误差的观测列。测量数据处理的对象,主要是带有随机误差的观测列。

本章主要介绍随机误差和系统误差的性质、特点、规律、误差估计、误差合成等方法及有关理论。

第一节 观测数据的随机误差与误差统计

一、随机误差

对某已知真值的量在相同条件下进行多次观测,以测量次数 n 为横坐标,误差值 Δ 为纵坐标,按测量序号逐点作图,就得到一个离散的误差序列图,如图1-1所示。如果是自动连续记录,则得到一条连续曲线,这条曲线就是误差序列曲线。如果这条曲线没有任何确定的规律,即前一个误差出现后,不能预定下一个误差出现的大小和方向,它们的出现完全是随机的,则这类误差称为随机误差。

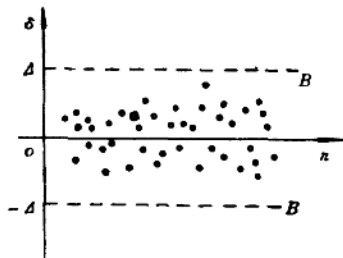


图 1-1

二、随机误差的统计规律

随机误差在大小和符号上都表现出偶然性,即从单个误差来看,该列误差的大小和符

号没有规律性,但就大量误差的总体而言,具有一定的统计规律。

如某测区,在相同条件下,独立观测了358个三角形的全部内角。由于观测值带有随机误差,三角形三个内角之和不等于 180° ,各个三角形内角和的真误差可按下式计算:

$$\Delta_i = (L_1 + L_2 + L_3) - 180^\circ \quad (1-1)$$

式中 L_1 、 L_2 、 L_3 ——三个内角观测值。

现取误差区间的间隔 $d\Delta$ 为 $0.20''$ 。将这一组误差按其正负号与数值大小顺序排列,统计误差出现在各区间内的个数 v_i ,以及“误差出现在某个区间内”这一事件的频率 $\frac{v_i}{n}$ (此处 $n=358$),结果列于表1-1中。

表 1-1 三角形闭合差统计

误差的区间 (")	Δ 为 负 值			Δ 为 正 值			备 注
	个 数 v_i	频率 $\frac{v_i}{n}$	$\frac{v_i}{n} / d\Delta$	个 数 v_i	频率 $\frac{v_i}{n}$	$\frac{v_i}{n} / d\Delta$	
0.00~0.20	45	0.126	0.630	46	0.128	0.640	$d\Delta = 0.20''$ 等于区间 左端值的误 差算入该区 间内
0.20~0.40	40	0.112	0.560	41	0.115	0.575	
0.40~0.60	33	0.092	0.460	33	0.092	0.460	
0.60~0.80	23	0.064	0.320	21	0.059	0.295	
0.80~1.00	17	0.047	0.235	16	0.045	0.225	
1.00~1.20	13	0.036	0.180	13	0.036	0.180	
1.20~1.40	6	0.017	0.085	5	0.014	0.070	
1.40~1.60	4	0.011	0.055	2	0.006	0.030	
1.60以上	0	0	0	0	0	0	
Σ	181	0.505		177	0.495		

从表1-1看出,随机误差的分布,具有下列性质和特点:(1)误差的绝对值有一定的限值,我们称这一特性为有界性;(2)绝对值小的误差比绝对值大的误差出现的可能性大,这一特性称为单峰性;(3)绝对值相等的正负误差出现的可能性相同,这一特性称为对称性;(4)在相同条件下,对同一量进行观测,其误差的算术平均值,随着测量次数 n 的无限增大而趋近于零,这一特性称为抵偿性,即 $\lim_{n \rightarrow \infty} \frac{[\Delta]}{n} = 0$ 。

误差分布情况,除了用上述列表以外,还可用图形表达。例如以横坐标表示误差的大小,纵坐标表示各误差出现在各区间内的频率除以间距的间隔值,即 $v/n / d\Delta$ 。根据表1-1数据绘制出图1-2。

在一组试验中,随机变量是对应着一定的概率取值。因此在相同观测条件下,当一组独立观测值误差总数足够多时,那么,误差出现在各区间内的频率就会稳定在某一常数(理论频率)附近,观测个数愈多,稳定程度也就愈大。如表1-1,当 $n \rightarrow \infty$ 时,各频率也就趋于一个完全确定的数值,就是误差出现在各区间内的概率。当 $n \rightarrow \infty$ 时,如把误差区间无限缩小,则图1-2中各长方条顶边将连接成曲线,如图1-3所示。

这条曲线也就是误差出现在各区间内的概率分布曲线,简称误差分布曲线。由高斯误差定律知,误差分布曲线的密度函数 $f(\Delta)$ 为

$$f(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2 \Delta^2} \quad (1-2)$$

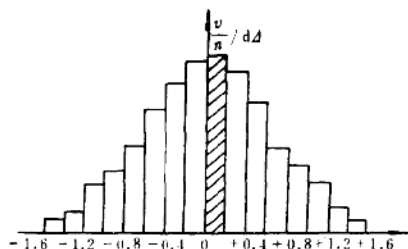


图 1-2

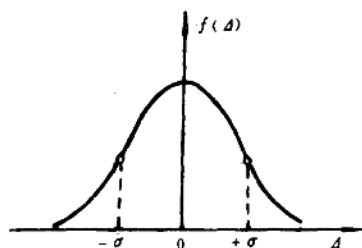


图 1-3

式中 h ——精度值；可用最大或然法求得

$$h^2 = \frac{1}{2\sigma^2} \quad (1-3)$$

$$\sigma^2 \text{——随机误差 } \Delta \text{ 的方差； } \sigma^2 = \lim_{n \rightarrow \infty} \frac{[\Delta\Delta]}{n} \quad (1-4)$$

将式 (1-3) 代入式 (1-2) 得

$$f(\Delta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\Delta^2}{2\sigma^2}} \quad (1-5)$$

式 (1-5) 与正态分布的密度函数相同，因此，测量中观测值和观测误差大都是服从正态分布。本书在以后的理论研究中，都是以正态分布作为描述随机误差分布的数学模型。

由图 1-3 可以看出，误差分布曲线在纵轴两旁各有一个转向点（拐点），如令 $f(\Delta)$ 的二阶导数等于零，可求出转向点的横坐标，即

$$f'(\Delta) = -\frac{\Delta}{\sigma^2\sqrt{2\pi}} e^{-\frac{\Delta^2}{2\sigma^2}}$$

$$f''(\Delta) = -\frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{\Delta^2}{2\sigma^2}} \left\{ 1 - \frac{\Delta^2}{\sigma^2} \right\} = 0$$

由此得

$$1 - \frac{\Delta^2}{\sigma^2} = 0$$

曲线的拐点坐标为

$$\Delta_0 = \pm \sigma \quad (1-6)$$

从图 1-3 可知，如果观测条件好，观测精度高，误差分布紧密，误差分布曲线较高而陡峭；反之误差分布较离散，分布曲线较低而平缓。因此往往用代表误差分布的一个特征值 σ （称标准差）作为衡量一组观测值精度的标准。标准差（或中误差）为

$$\sigma = \lim_{n \rightarrow \infty} \sqrt{\frac{[\Delta\Delta]}{n}} \quad (1-7)$$

但在实践中观测个数是有限的，可采用下式求其估计值

$$\hat{\sigma} = \pm \sqrt{\frac{[\Delta\Delta]}{n}} \quad (1-8)$$

三、正态分布

无论是理论上或实用上, 正态分布都是一种很重要的分布。自然界有许多现象都近似服从正态分布, 测量中的观测误差一般服从正态分布。

(一) 一维正态分布的概率密度

若随机变量 x 的分布密度函数为 $f(x)$

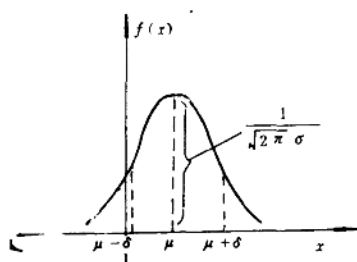


图 1-4

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1-9)$$

则称随机变量 x 服从正态分布, 记为 $x \sim N(\mu, \sigma^2)$, 正态分布曲线如图 1-4 所示。随机变量 x 在区间 (a, b) 内取值的概率 $P(a \leq x \leq b)$ 称为随机变量的分布函数, 记为

$$F(x) = P(a \leq x \leq b) = \int_a^b f(x) dx \quad (1-10)$$

式中 μ, σ^2 ——正态随机变量的数学期望和方差。

参数 μ 和 σ 决定正态分布曲线的形状和位置, 如图 1-5、图 1-6 所示。

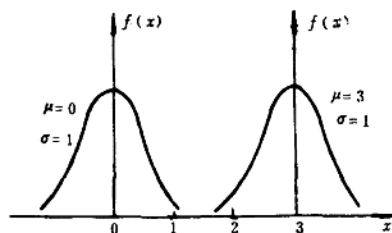


图 1-5

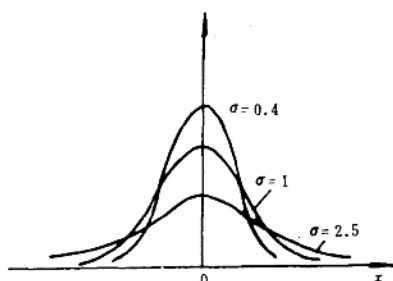


图 1-6

(二) 分布函数标准化及概率计算

按照式 (1-10) 计算概率很麻烦, 为此将式 (1-10) 定积分作变量变换, 把 x 按式 (1-11) 换为 y , 使 y 仍是正态分布, 且使它的数学期望 $\mu = 0$, 方差 $\sigma^2 = 1$, 记为 $y \sim N(0, 1)$ 。这种把随机变量 $x \sim N(\mu, \sigma^2)$ 变为 $y \sim N(0, 1)$ 分布的过程称为正态变量标准化。 y 称为标准化随机变量。

$$y = \frac{x - \mu_x}{\sigma_x} \quad (1-11)$$

$$\left. \begin{aligned} E(y) &= E\left(\frac{x - \mu_x}{\sigma_x}\right) = \frac{1}{\sigma_x} E(x - \mu_x) = \frac{1}{\sigma_x} (\mu_x - \mu_x) = 0 \\ D(y) &= D\left(\frac{x - \mu_x}{\sigma_x}\right) = \frac{1}{\sigma_x^2} D(x - \mu_x) = \frac{\sigma_x^2}{\sigma_x^2} = 1 \end{aligned} \right\} \quad (1-12)$$

标准化后随机变量 y 的密度函数及分布函数为

$$\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad (1-13)$$

$$\phi(Z_\alpha) = P(y \leq Z_\alpha) = \int_{-\infty}^{Z_\alpha} \phi(y) dy = 1 - \alpha \quad (1-14)$$

标准化后的正态分布函数 $\phi(Z_\alpha)$ 已经编制成用表(见附表1),直接查取即可。式(1-14)计算的概率可用小面积表示(图1-7)。 Z_α 为横坐标轴上点位,称为百分点位。

如果要求区间,即求两个百分点位置之间的面积

$$P(x_1 \leq x \leq x_2) = P(x \leq x_2) - P(x \leq x_1) = F(x_2) - F(x_1) \quad (1-15)$$

公式(1-15)计算的概率用面积表示,见图1-8。

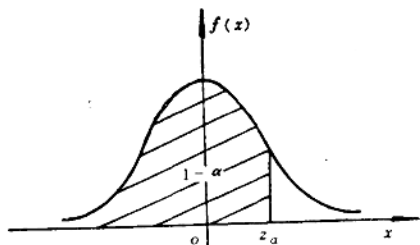


图 1-7

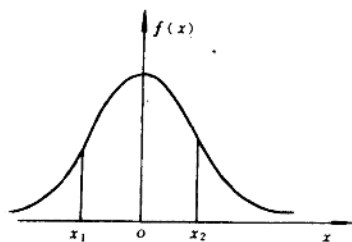


图 1-8

例[1-1] 已知随机变量 x 服从标准正态分布,求该变量 x 落入区间($x_2 = 0.5$, $x_1 = -0.25$)内的概率。

解:

$$\begin{aligned} P(x_1 \leq x \leq x_2) &= F(x_2) - F(x_1) \\ &= F(0.5) - F(-0.25) \\ &= \int_{-\infty}^{0.5} f(y) dy - \int_{-\infty}^{-0.25} f(y) dy \\ &= 0.6915 - 0.4013 = 0.2902 \end{aligned}$$

例[1-2] 已知服从正态分布的随机变量 y 落入区间(c , $-c$)中概率为0.95,求 c 。

解:

$$\begin{aligned} P(-c \leq y \leq c) &= 0.95 \\ P(-c \leq y \leq c) &= P(y \leq c) - P(y \leq -c) = \phi(c) - [1 - \phi(c)] \end{aligned}$$

即

$$2\phi(c) - 1 = 0.95, \phi(c) = 0.975$$

反查正态分布表得 $c = 1.96$,则(c , $-c$)为(1.96 , -1.96)。

例[1-3] 已知观测误差 Δ 服从正态分布,试计算 Δ 出现在($-\sigma$, $+\sigma$), (-2σ , $+2\sigma$), (-3σ , $+3\sigma$)内的概率。

解:

$$\phi(\Delta) = \int_{-\infty}^{\Delta} \phi(y) dy - \int_{-\infty}^{-\Delta} \phi(y) dy$$

当 $k=1$ 时, $\phi(\Delta) = 0.8413 - 0.1587 = 0.6826$

当 $k=2$ 时, $\phi(\Delta) = 0.9772 - 0.0228 = 0.9544$

当 $k=3$ 时, $\phi(\Delta) = 0.9987 - 0.0013 = 0.9973$

即观测误差出现在1倍 $\pm\sigma$ 标准差内的概率为68.3%，出现在2倍 $\pm2\sigma$ 标准差内的概率为95.4%，出现在3倍 $\pm3\sigma$ 标准差内的概率为99.7%。

(三) 置信区间

在正态分布中，观测误差落入 $\pm k\sigma$ 中的概率可表示为

$$P(|\Delta| < k\sigma) = P(-k\sigma < \Delta < +k\sigma) = P(\Delta < k\sigma) - P(\Delta < -k\sigma)$$

$$P(|\Delta| < k\sigma) = 1 - \alpha$$

当 $k=1$ 时， $1-\alpha=0.683$

当 $k=2$ 时， $1-\alpha=0.954$

当 $k=3$ 时， $1-\alpha=0.997$

式中 $(-k\sigma, k\sigma)$ ——置信区间；

$1-\alpha$ ——置信水平（或置信概率，置信度）；

α ——显著性水平或显著度；

$k\sigma$ ——置信限；

k ——置信系数。

表 1-2 置信系数表

α	$1-\alpha$ $P(\Delta < k\sigma)$	k
0.3173	0.6827	1
0.05	0.95	1.96
0.0455	0.9545	2
0.01	0.99	2.58
0.0027	0.9973	3

若已知 k 值可求得 α ，反之，已知 α 可求得 k ，其常用值列于表1-2。

由表1-2知，服从正态分布的随机误差 Δ 落入 $\pm 1.96\sigma$ 内的概率为95%，95%称为置信概率，1.96称为置信概率为95%的情况下随机误差 Δ 的置信系数。随机误差 Δ 落入 $\pm 2.58\sigma$ 内的概率为99%，称 $\pm 2.58\sigma$ 为置信概率99%的情况下随机误差 Δ 的置信限（即误差 Δ 的界限）。

当检验一组观测误差是否服从正态分布时，均在一定的置信度（ $1-\alpha$ ）的条件下进行。通常取 $1-\alpha=0.95$ 或 $1-\alpha=0.99$ （即 $\alpha=0.05$ 或 0.01 ），就是采用 $\pm 1.96\sigma$ 或 $\pm 2.58\sigma$ 作为置信限。凡是提及置信限时，总是要结合概率意义的。当置信限为 $\pm 3\sigma$ 时，置信概率为99.7%，即1000个观测值中，只有3个误差可能超出置信限，故测量中常把 $\pm 3\sigma$ 作为极限误差，即

$$\Delta_{\text{限}} = \pm 3\sigma \quad (1-16)$$

四、误差估计

设对某量进行了 n 次等精度观测，其独立观测值为 $x_1, x_2, \dots, x_n$ ，均值为 $\bar{x}$ ，每一次观测值的标准差估值 $\hat{\sigma}$ 可按以下方法计算：

(一) 标准法——白塞尔 (Bessel) 公式

$$\hat{\sigma} = \pm \sqrt{\frac{\sum [(x_i - \bar{x})^2]}{n-1}} \quad (1-17)$$

例[1-4] 对某量进行8次测量，观测值如下：

$$x_1 = 802.40, x_2 = 802.50, x_3 = 802.38, x_4 = 802.48$$

$$x_5 = 802.42, x_6 = 802.46, x_7 = 802.45, x_8 = 802.43$$

试求每次观测值的标准差。

解：

$$\bar{x} = \frac{1}{n} \sum x_i = 802.44$$

$$\hat{\sigma} = \pm \sqrt{\frac{0.0114}{8-1}} = \pm \sqrt{0.0016} = \pm 0.04$$

(二) 较差法

$$\hat{\sigma} = \{[(x_1 - x_2)^2 + (x_2 - x_3)^2 + \dots + (x_{n-1} - x_n)^2] / (2n - 2)\}^{\frac{1}{2}} \quad (1-18)$$

(三) 绝对值法——彼特斯 (Петерс) 公式

$$\begin{aligned} \hat{\sigma} &= h \Sigma |x_i - \bar{x}| / \sqrt{(n^2 - n)} \\ &\approx h \Sigma |x_i - \bar{x}| / \left(n - \frac{1}{2}\right) \end{aligned} \quad (1-19)$$

$$\text{正态分布时, } h = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \approx \frac{5}{4}$$

五、极差及用极差评定精度的问题

在观测值的精度评定中, 一般常用均方根法, 即用白塞尔公式计算观测值的标准差作为一组观测值的精度衡量标准。但由于均方根法计算标准差时需将各残差平方相加, 还要进行除法及开方运算, 比较复杂, 容易出错, 因此产生了极差法。

设服从正态分布的重复独立观测值为 $x_1, x_2, \dots, x_n$, 其中最大值为 $\max x_i$, 最小值为 $\min x_i$, 它们之间的差称为极差, 用符号 “ R ” 表示

$$R = \max x_i - \min x_i \quad (1-20)$$

R 是子样观测值的函数, 所以也是一服从正态分布的随机变量, 设标准化后为 $w^{(4)}$ 。

$$w = R / \sigma \quad (1-21)$$

$$\text{密度函数为} \quad f(w) = n(n-1)\sigma \int_{-\infty}^{\infty} G(w, x) \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{-\frac{(x+\sigma_w-\mu)^2}{2\sigma^2}} dx \quad (1-22)$$

$$\text{式中} \quad G(w, x) = \left\{ \int_{-\infty}^{x+\sigma_w} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \right\}^{n-1} \quad (1-23)$$

公式 (1-22) 与 (1-23) 的推导详见文献^[4]。

$$\text{数学期望为} \quad E(w) = \frac{n(n-1)}{(\sqrt{2\pi})^n} \int_{-\infty}^{\infty} w dw \int_{-\infty}^{\infty} e^{H_1(w, x_1)} G_1(w, x_1) dx_1 \quad (1-24)$$

$$\text{式中} \quad H_1(w, x_1) = -\frac{1}{2}[x_1^2 + (x_1 + w)^2],$$

$$x_1 = \frac{x - \mu}{\sigma},$$

$$t_1 = \frac{t - \mu}{\sigma},$$

$$G_1(w, x_1) = \left[\sigma \int_{x_1}^{x_1 + w} e^{-\frac{1}{2}t_1^2} dt_1 \right]^{n-1}.$$

设 $E(w) = \alpha_n$, α_n 根据 n 的不同取值, 可由式 (1-24) 算出, 编制成表 1-3, 称极差系数 α_n 数值表。

由于

$$E\left(\frac{R}{\sigma}\right) = \alpha_n$$

$$\text{故} \quad E\left(\frac{R}{\alpha_n}\right) = \sigma \quad (1-25)$$

由式 (1-25) 知, $\frac{R}{\alpha_n}$ 是 σ 的无偏估值, 即

$$\hat{\sigma} = \frac{R}{\alpha_n} \quad (1-26)$$

式 (1-26) 为用极差计算标准差估值 $\hat{\sigma}$ 的公式。

如以例 [1-4] 题意, 按极差法求标准差 $\hat{\sigma}$ 。

$$R = \max x_i - \min x_j = 802.50 - 802.38 = 0.12$$

以 n 为引数, 查表 1-3 得 $\alpha_n = 2.847$, 则 $\hat{\sigma} = 0.12/2.847 = \pm 0.042$ 。

用极差估计 $\hat{\sigma}$, 迅速方便。与白塞尔公式计算的 $\hat{\sigma}$ 比较, 当 $n=5 \sim 10$ 时, 两者相差不大, 当 $n < 4$ 或 $n > 15$ 时, 两者相差较大, 此时还是用白塞尔公式计算 $\hat{\sigma}$ 为好。

表 1-3 极差系数 α_n 表

n	α_n	n	α_n	n	α_n	n	α_n
2	1.128	7	2.704	12	3.258	17	3.588
3	1.693	8	2.847	13	3.336	18	3.640
4	2.059	9	2.970	14	3.407	19	3.689
5	2.326	10	3.078	15	3.472	20	3.735
6	2.534	11	3.173	16	3.532		

六、随机不确定度的精确估计

由图 1-1 可见, 观测值误差一般都落在由 B 和 B' 两条直线所夹的范围内, 这个范围称为误差带。误差带两边界值为 Δ 和 $-\Delta$, Δ 称为误差界限或极限误差, 也称为随机不确定度 ($k\sigma$)。通常用 $\Delta = k_s \sigma$ 表示。系数 k_s 与随机变量分布及子样容量 n 有关, 当 $\Delta \sim N(0, 1)$, $1 - \alpha = 0.997$ 时, $k = 3$ 。置信水平与变量个数 n 有关, $n \rightarrow \infty$ 时, $1 - \alpha = 0.9973$; $n = 14$ 时, $1 - \alpha = 0.99$; $n = 8$, $1 - \alpha = 0.98$ 。标准差 σ 是 $n \rightarrow 8$ 时的取值。实践中, n 为有限个数, 按白塞尔公式计算, 估值为 $\hat{\sigma}$ 。由于 σ 与 $\hat{\sigma}$ 不同, 实际采用的随机不确定度 $\Delta = k \hat{\sigma}$ 。因此, 极限误差采用 $k = 3$, $\Delta_n = 3\hat{\sigma}$ 是不严密的计算公式。如果采用 $\hat{\sigma}$ 代替 σ , 则 k 值应按 $\Delta = x - \mu \sim (0, \sigma^2)$, $\frac{\Delta}{\sigma} = \frac{x - \mu}{\sigma} \sim N(0, 1)$ 计算。如将 σ 换成 $\hat{\sigma}$, 则 $\frac{\Delta}{\hat{\sigma}} = \frac{x - \mu}{\hat{\sigma}}$, 变成了服从 t 分布, 置信系数 $k = t_{\alpha/2}(n-1)$ 。简要证明如下:

根据 t 分布定义的 t (变量) $= \frac{x}{\sqrt{y/n}}$, $x \sim N(0, 1)$, $y \sim \chi^2(n)$, n 为自由度。又知

$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$ [1], 按 t 变量定义, 可作一个变量

$$\frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} / \sqrt{\left(\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right) / (n-1)}$$

上式变量简化后为 $\frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{\alpha}(n-1)$

是服从 $t_{\alpha}(n-1)$ 分布的 t 变量。因此平均值的随机不确定度应为

$$\Delta \bar{x} = \bar{x} - \mu = t_{\alpha}(n-1) \frac{\hat{\sigma}}{\sqrt{n}} \quad (1-27)$$

随机变量 x 的随机不确定度应为

$$\Delta x = t_{\alpha}(n-1) \hat{\sigma}$$

如以例(1-4)数值, 求平均值 $\bar{x}$ 的随机不确定度。

$$\bar{x} = 802.44 \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.04$$

取 $\alpha = 0.01$, $R = n - 1 = 7$, 查 t 分布表 (表1-4) 得 $t_{\alpha}(k) = t_{0.01}(7) = 3.5$

表 1-4 t 分布置信系数 $t_{\alpha}(n-1)$ 数值表

$n-1 \backslash \alpha$	0.01	0.05	$n-1 \backslash \alpha$	0.01	0.05	$n-1 \backslash \alpha$	0.01	0.05
1	63.7	12.71	12	3.06	2.18	23	2.81	2.07
2	9.92	4.30	13	3.01	2.16	24	2.80	2.06
3	5.84	3.18	14	2.98	2.14	25	2.79	2.06
4	4.60	2.78	15	2.95	2.13	26	2.78	2.06
5	4.03	2.57	16	2.92	2.12	27	2.77	2.05
6	3.71	2.45	17	2.90	2.11	28	2.76	2.05
7	3.50	2.36	18	2.88	2.10	29	2.76	2.04
8	3.36	2.31	19	2.86	2.09	30	2.75	2.04
9	3.25	2.26	20	2.84	2.09	40	2.70	2.02
10	3.17	2.23	21	2.83	2.08	60	2.66	2.00
11	3.11	2.20	22	2.82	2.07	120	2.62	1.98
						∞	2.58	1.96

$$\Delta \bar{x}_m = t_{\alpha}(7) \frac{\hat{\sigma}}{\sqrt{n}} = 3.5 \left(\frac{0.04}{\sqrt{8}} \right) = 0.05$$

最后平均值 $\bar{x}$ 的随机不确定度为 0.05。

七、单位权方差

假设对某条水准路线采用 t 台不同精度的水准仪进行了 t 次水准测量, 测得 t 个高差 $h_i (i = 1, 2, \dots, t)$ 。如用这 t 个高差取平均值, 则不同精度的高差所占的份量是不同的。这个“份量”在测量中用“ p ”表示, 称为权。可按下式定义

$$p_i = \frac{\mu^2}{\hat{\sigma}_i^2} \quad (1-28)$$

即权与方差成反比。这样, 精度高的观测值权大, 在平均值中所占“份量”也大。 μ^2 为任意常数, 当取 $\mu^2 = \hat{\sigma}_i^2$ 时, $p_i = 1$, 这时称 μ^2 为单位权方差。权为 1 的观测值称为单位权观测值。单位权方差计算式为

$$\mu^2 = \pm \frac{[pvv]}{n-1} \quad (1-29)$$

式中 $v = x_i - \bar{x}$; 观测值 x 的平均值 $\bar{x}$ 计算公式为

$$\bar{x} = \frac{p_1 x_1 + p_2 x_2 + \cdots + p_n x_n}{p_1 + p_2 + \cdots + p_n}$$

例[1-5] 如图1-9所示, 由5条水准路线测定待定点G的高程, 水准路线长度及测定的G点高程见表1-5。以10公里水准测量高差为单位权观测值。试求G点高程平均值及其标准差, 10公里及每公里的观测高差标准差。

解: G点高程平均值

$$H_G = x_0 + \frac{P[\Delta x]}{[P]} = 112.810 + \frac{192}{38.5} = 112.8150\text{m}$$

$$10\text{公里观测高差的标准差 } \mu = \pm \sqrt{\frac{[pvv]}{n-1}} = \pm \sqrt{\frac{592}{4}} = \pm 12.2\text{mm}$$

$$\text{平均值标准差 } m_{\bar{x}} = \pm \frac{12.2}{\sqrt{38.5}} = \pm 2.0\text{mm}$$

$$\text{每公里观测高差的标准差 } \hat{\sigma}_{\text{公里}} = \pm \frac{\mu}{\sqrt{10}} = \pm 3.9\text{mm}$$

表 1-5 观测数据表

路线	观测高差 (m)	距 离 (km)	$P = \frac{10}{s}$	$v_1(\text{mm})$
1	112.814	2.5	4.0	+1
2	112.807	4.0	2.5	+8
3	112.802	5.0	2.0	+13
4	112.817	0.5	20.0	-2
5	112.816	1.0	10.0	-1

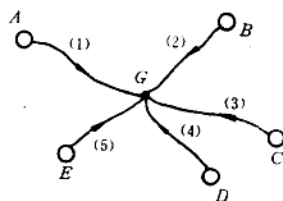


图 1-9

第二节 观测数据的系统误差

一、系统误差的性质、特点和规律

随机误差是随机变量相对于母体均值的偏差。在测量中又由于设备、环境、观测方法等原因而产生误差, 使母体的均值发生位移, 这种位移为系统误差。在数据处理中, 是以不含系统误差的观测值为前提。但实际上, 随机误差与系统误差同时存在于观测值中, 有时系统误差的数值还比较大, 因此测量结果的精度不仅取决于随机误差的影响, 还取决于系统误差的影响。为此, 研究系统误差的规律性及其减弱的方法十分重要。

在相同条件下进行一系列的观测, 如果误差在大小、符号上表现出系统性, 或在测量过程中, 按一定的规律变化、或者为某一常数, 那么, 这种误差就称为系统误差。

图1-10为各种系统误差E在测量过程中的变化规律, 其横轴的变量x可以是时间、测量的顺序号等。

当系统误差与随机误差同时存在时, 误差按图1-11的规律分布。图中 y_0 为真值, E为系统误差, σ 为随机误差的标准差, y_i 为多次重复测量时任一次的测量值, $\bar{y}$ 为子样平均值。

现就图1-10所示的系统误差的规律性说明如下。

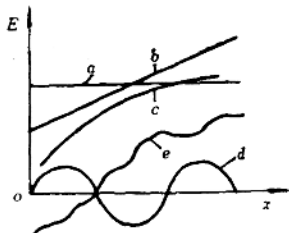


图 1-10

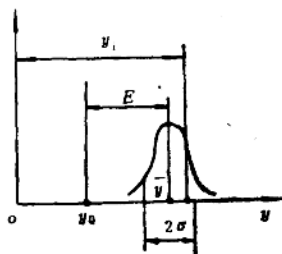


图 1-11

(一) 固定误差 (或称常数误差或定值系统误差)

在整个测量过程中, 误差的大小和符号都固定不变, 这种系统误差称为固定误差。其变化规律如图1-10中的直线 a 。

(二) 线性误差 (或称可变系统误差)

在测量过程中, 随着测量值或时间的变化, 误差值成比例地增大或减小, 这种系统误差称为线性误差。其变化规律如图1-10中的直线 b 。

(三) 多项式误差

当误差与因素不符合线性关系时, 可采用多项式来描述误差与因素的关系, 这种系统误差称为多项式误差, 如图1-10中的曲线 c 。

(四) 周期误差

在整个测量过程中, 随着测量值或时间的变化, 误差是按周期性规律变化的, 称为周期性误差。其变化曲线如图1-10中的曲线 d 。

如激光测距仪, 由于仪器内部电信号的干扰而产生的周期误差, 其改正数按下式计算:

$$v_i = A \cdot \sin(\varphi_0 + \theta_i)$$

式中 v_i ——周期误差改正数;

A ——周期误差的振幅;

φ_0 ——仪器的初相角;

θ_i ——与待测距离尾数相应的相位角差。

(五) 复杂规律误差

在整个测量过程中, 若误差是按确定的且复杂的规律变化, 称为复杂规律误差。其变化曲线如图1-10中的曲线 e 。

二、系统误差对测量结果的影响

(一) 定值系统误差的影响

设等精度观测列 x_i ($i=1, 2, \dots, n$), 含有定值系统误差 ε 和随机误差 δ_i ($i=1, 2, \dots, n$)。

$$\begin{aligned} x_i &= \tilde{x} + \delta_i + \varepsilon \\ \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^n (\tilde{x} + \delta_i + \varepsilon) \end{aligned} \quad (1-30)$$