

测量实践的数据处理

张世英 刘智敏 编著

科学出版社

51.734
549

测量实践的数据处理

张世英 刘智敏 编著

(3661/53)
268

科学出版社

内 容 简 介

本书利用概率论与数理统计的知识来讨论测量结果的数据处理问题。全书共十四章。第二、三章讲概率、统计的基本知识；第四章讲基本测量问题的处理；第五章讲最小二乘法；第六、七章讲条件测量与交会测量的数据处理；第八、九章讲动态测量的精度分析与数据滤波；第十章讲回归方程的建立；第十一章讲数据性质的统计检验；第十二章讲系统误差；第十三章讲实验误差的计算；最后一章介绍数据运算与处理步骤。内容密切联系实际。

本书可供科学工作者、工程技术人员与高校师生参考。

测量实践的数据处理

张世英 刘智敏 编著

*

科 学 出 版 社 出 版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1977年11月第 一 版 开本：787×1092 1/32

1977年11月第一次印刷 印张：16

印数：0001—15,170 字数：367,000

统一书号：13031·629

本社书号：916·13—1

定 价： 1.55 元

前 言

“和其他一切科学一样,数学是从人的需要中产生的。”测量结果数据处理方法的研究,就是根据人们在生产斗争和科学实验的测量实践中对测量数据分析、处理的需要而产生的。

在生产斗争和科学实验中,常常通过测量而获得大量的感性材料——测量数据,然而,对于测量所获得的大量感性材料的认识有待于提高。毛主席教导我们:“要完全地反映整个的事物,反映事物的本质,反映事物的内部规律性,就必须经过思考作用,将丰富的感觉材料加以去粗取精、去伪存真、由此及彼、由表及里的改造制作工夫,造成概念和理论的系统,就必须从感性认识跃进到理性认识。”因此,测量结果数据处理的任务就在于运用数学方法对测量实践中所获得的大量数据“加以去粗取精、去伪存真、由此及彼、由表及里的改造制作工夫”,从而由实验结果引出科学的结论,并得出反映事物内部规律性的东西。

随着生产斗争和科学实验的不断发展,测量结果数据处理方法的研究也在不断地深入进行,目前它已成为现代统计学的一个重要组成部分;它的一些基本方法已为许多科研、生产部门广泛应用。本书仅就实际工作中常用的一些基本的数据处理方法作一较为全面而系统的介绍,书中各章是以某一类型实际问题的需要为中心来介绍一系列精度分析和数据处理方法的。因此,我们着重从应用的角度对这些数据处理方法进行讨论和总结,而不去过多地强调数学本身的严谨性。

书中第二章、第三章介绍了概率论和数理统计的基本知

识,这是为缺少这方面知识的读者准备的,因而基本上是按照阅读本书后面章节的需要而编写的,这些内容已满足了本书对概率论和数理统计知识的要求.书末所列参考资料仅供读者进一步了解书中内容时参考.

本书初稿完成后,请中国科学院数学研究所概率统计室进行了全面的审阅,审阅中概率统计室的同志对初稿提出了许多宝贵意见,作者又根据这些意见对初稿作了修改,原稿虽经修改,但由于作者水平所限,难免会有错误之处,切盼读者批评指正.

张世英 刘智敏

1973年10月

目 录

第一章 引论	1
§ 1 测量在生产斗争和科学实验中的意义	1
§ 2 测量的分类	2
§ 3 测量中的误差	7
§ 4 观测结果的精密性和准确性	11
第二章 概率论的基本知识	14
§ 1 事件的概率	14
§ 2 随机变量及其分布	16
§ 3 随机变量的数学期望、方差及协方差	18
§ 4 正态分布和均匀分布	24
§ 5 随机向量及其分布	29
§ 6 正态随机向量	34
§ 7 极限定理	42
§ 8 平稳随机过程	46
第三章 数理统计的基本知识	52
§ 1 子样及其分布	52
§ 2 参数估计	55
§ 3 χ^2 分布、 t 分布和 F 分布	59
§ 4 假设检验概说	61
第四章 基本测量问题的处理	65
§ 1 概论	65
§ 2 算术平均值及其精度	66
§ 3 等精度观测数据的精度估计	69
§ 4 权的概念。加权算术平均值及其精度	74
§ 5 权的确定	78

§ 6	单位权概念。单位权方差	81
§ 7	例	84
§ 8	成组观测结果的处理	87
§ 9	误差传递的一般法则	96
第五章	最小二乘法	106
§ 1	线性参数的最小二乘估计	106
§ 2	线性参数最小二乘估计的性质	113
§ 3	用最小二乘估计的残差估计观测数据的精度	120
§ 4	不等精度观测时, 线性参数的最小二乘估计	129
§ 5	正规方程的解	134
§ 6	线性参数的无偏最优估计及与最小二乘估计的关系	147
第六章	条件测量的处理	162
§ 1	问题的提出	162
§ 2	条件方程	163
§ 3	利用条件方程消去某些未知量的方法处理条件观测问题	171
§ 4	利用求解条件极值的方法处理条件观测问题	182
§ 5	条件观测中被测量的量的线性函数的估计	195
§ 6	条件观测问题中两种处理方法的比较	197
第七章	交会测量的处理	198
§ 1	交会测量的一般性讨论	198
§ 2	误差椭圆	205
§ 3	前方交会测量	216
第八章	观测数据偶然误差的统计估计	225
§ 1	利用最小二乘曲线拟合的残差估计观测数据的偶然误差	225
§ 2	利用变量差分方法估计观测数据的偶然误差	244
第九章	观测数据的滤波处理	257
§ 1	引言	257
§ 2	多项式加平稳噪声的最优线性数值滤波问题	258

§ 3	多项式加白噪声的最优线性数值滤波问题	264
§ 4	中心平滑	271
§ 5	被观测过程的实时及预报估计	280
§ 6	多项式加平稳相关噪声的最优线性数值滤波问题	292
§ 7	方法误差的概念及其计算	295
第十章	实验结果的数学表示	304
§ 1	回归方程的意义	304
§ 2	两个变量线性回归方程的讨论	307
§ 3	可以化为线性回归方程的某些类型	315
§ 4	两个变量都具有误差时, 线性回归方程的确定	339
§ 5	多个变量线性回归方程的讨论	343
§ 6	多个变量非线性回归方程的线性化处理	353
§ 7	多个变量线性回归方程中变量的取舍——逐步回归方程的建立	357
第十一章	观测数据的统计检验	365
§ 1	引言	365
§ 2	观测数据的合理性检验	366
§ 3	误差统计分布的检验	371
§ 4	观测数据误差间相关性的检验	382
第十二章	系统误差	387
§ 1	引言	387
§ 2	系统误差的特征	389
§ 3	系统误差的种类	392
§ 4	发现系统误差的一些简单方法	395
§ 5	系统误差的假设检验	406
§ 6	多组测量的方差分析	417
§ 7	系统误差的限制	420
§ 8	系统误差的消除——回归分析	424
第十三章	实验误差的计算	428
§ 1	引言	428

§ 2 均方根差计算的白塞尔 (Bessel) 法	431
§ 3 均方根差计算的彼得 (Пeters) 法	433
§ 4 均方根差计算的最大误差法	435
§ 5 均方根差计算的极差法	440
§ 6 误差分析	444
§ 7 综合误差计算	451
§ 8 相关系数	462
§ 9 实验设计	468
第十四章 数据运算与处理步骤	475
§ 1 数据位数	475
§ 2 数据修约规则	476
§ 3 运算中的凑整	478
§ 4 平均值及其误差的计算校核	480
§ 5 粗差	484
§ 6 实验数据处理步骤	491
参考资料	495
附表 1 正态概率积分表	497
附表 2 χ^2 分布的临界值	498
附表 3 t 分布的临界值	500
附表 4 F 分布的临界值	501
附表 5 相关系数 r 检验表	503
附表 6 循环序列相关系数 R 分布的临界值	504

第一章 引 论

§ 1 测量在生产斗争和科学实验中的意义

阶级斗争、生产斗争和科学实验,是建设社会主义强大国家的三项伟大革命运动。测量工作是在生产斗争和科学实验中,获得实践知识的重要手段;也是生产斗争和科学实验本身一项不可少的内容。

在生产斗争和科学实验中,人们要经常对各种类型的物理量进行观测测量,并对观测结果运用数学方法“**加以去粗取精、去伪存真、由此及彼、由表及里的改造制作工夫**”,以形成对实践结果认识的深化,并得出反映事物内部规律性的东西。

机械工人在完成某一工件的制作过程中,按照设计的要求,及时对加工中的工件进行测量,校核加工的质量,以保证对加工件的尺寸及其公差的要求。

对一批产品,通过测量,检验某些技术指标是否符合质量要求。同时,通过测量检查、分析生产的稳定情况,或考验某种技术革新的实际效果。

为了进行大规模的社会主义建设及国防建设,则需要进行规模巨大的大地测量工作、航空摄影测量工作以及工程测量工作。

科学实验的过程,也就是按照一定的科研项目所组织的实验、测量及其对测量结果的分析、认识的过程。

例如,在我们进行飞机的设计工作中,为了确定翼型的设计方案,则需要将初步设计的翼型放到特定的风洞中进行一

系列空气动力学的实验,分析它的受力情况,从而修改设计,进而定型翼型;同时掌握某种翼型的空气动力规律,为进一步的设计工作积累资料。这项空气动力学实验实际上是一系列空气动力的力学测量工作。

长期观测的气候资料是分析、研究天气和气候变化规律的基础,也是农林、水利、建筑等部门在进行生产、规划、设计时所必须参考的。为了有助于农业生产的稳产、高产,我们需要掌握某一地区的气候变化规律,除了能进行短期气象预报外,还应逐步作到中期以至长期的气象预报,这就需要根据某一地区,某一特定气候参数(象水位、降水量、温度……)多年来的观测数据,分析这些时间过程的数据资料所揭示的内部规律性,从而对这一地区特定气候参数今后的变化情况提供参考,为制定气候参数的变化模式提供依据。

对物质物理、化学性能的精密物理、化学测定是了解物质性能的基础。

还可以举出大量的来自生产斗争和科学实验中的测量、测量结果的精度分析及其数据处理的问题。可以说,测量及其结果的精度分析和数据处理这方面的问题几乎在我们生产斗争和科学实验的各个方面随时都可以遇到。

§ 2 测量的分类

为了对在生产斗争和科学实验中所遇到的各种类型的测量及其结果的精度分析、数据处理工作进行系统的研究,我们必须研究各种类型测量问题的某些特殊性。我们从千百次测量实践中首先认识到,从测量的方式上来讲,测量可分作直接测量和间接测量。

为了测定某一物理量,如圆轴的直径和长度,空气的湿度,气流的压力等等,我们可直接用卡尺或千分表、湿度计、压

力表测得所要测定的量,这是最简单的测量问题,叫做直接测量,就是说这时被测量的量可以直接和其单位尺度进行比较。

有些测量则不然,如为了测量某一建筑物的高度 H ,我们在某个观测点上用测角仪测量了该建筑物的仰角 θ ,并测量了从观测点到建筑物的距离 D ,则由几何关系

$$H = D \operatorname{tg} \theta$$

来确定该建筑物的高度 H 。在这个例子中,被测量的量 H 是由直接测量的量 θ 和 D 通过确定的函数关系而求得。这样,高度 H 的测量就是间接测量。

间接测量在测量实践中是非常普遍的,绝大多数测量问题都是通过间接手段而解决的。

机械加工中,为了测量燕尾槽的底面宽度 W (图 1.1),可用两个等直径 d 的量棒放入槽内,用块规量出距离 G ,如燕尾槽角度为 α ,则从几何关系可求出燕尾槽的底面宽度。由图显然有

$$B = \frac{d}{2} \operatorname{ctg} \frac{\alpha}{2},$$

$$W = G + d + d \operatorname{ctg} \frac{\alpha}{2}$$

$$= G + d \left(1 + \operatorname{ctg} \frac{\alpha}{2} \right).$$

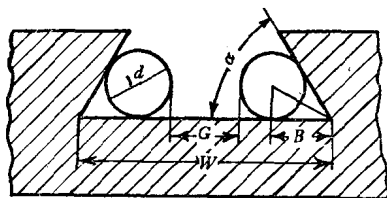


图 1.1

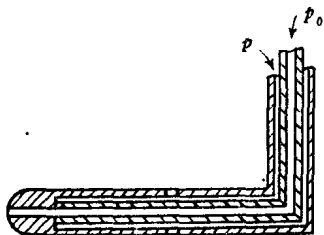


图 1.2

在气体力学实验中常常需要测定气流的速度。例如,为了测定超音速风洞流场中的马赫数 M ,则要利用风洞中的测速装置——测风管 (图 1.2),测定滞止压强 p_0 和来流静止压强 p ,根据超音

速气流的力学关系,由下面关系确定来流马赫数 M :

$$\frac{p_0}{p} = \frac{166.7 \cdot M^7}{(7 \cdot M^2 - 1)^{2.5}},$$

上式中, 气流介质为空气.

上述这些都是间接测量的例子.

如果我们按被测量的量在测量过程中的状态来分, 测量又可分为静态测量和动态测量.

所谓静态测量(或称为重复测量)是指在测量过程中被测量的量是不变的; 而动态测量(或称为过程测量)是指在测量过程中被测量的量是变化的.

前面所举的几个例子中, 某种产品的技术指标, 建筑物的高度, 风洞流场稳定状态下的气流速度等等, 这些被测量的物理量在测量过程中都是不变的, 因而都是静态测量.

我们再看下面这个例子: 由雷达观测数据来确定雷达对某一个瞄准塔的测距能力, 显然, 这样的测量是属于静态测量. 然而, 如果这个雷达对一个慢慢离开它上空的气球进行测距, 这时, 被测量的量——气球到雷达的距离则是变化的, 它就属于动态测量. 由于雷达有直接测距的能力, 因而这两者又都是直接测量.

尽管我们将测量问题根据其反映矛盾的特殊性分为直接测量和间接测量; 静态测量和动态测量, 然而, 它们都蕴含着普遍性的问题, 即需要解决测量结果的精度分析和数据处理这些一般性的问题.

在上面这个例子中, 雷达对某个瞄准塔进行测距, 测得一组观测数据 $R_r, r = 1, 2, \dots, N$, 这样的测量实际上是对距离这个物理量进行的多次重复观测. 按照一般的认识, 我们对这组观测数据 R_r 取算术平均值

$$\bar{R} = \frac{1}{N} \sum_{r=1}^N R_r \quad (N \text{ 为观测次数})$$

来给出雷达到瞄准塔距离 R 的估计。对于这类问题的详细讨论将在第四章内给出。

对于上例的另一种情况,被测量的量的性质就不同了,这时气球到雷达的距离是在不断改变的。这时,仍然取一段观测数据的算术平均值 $\bar{R}$ 就没有多大实际意义了,充其量不过是气球在观测时间 t_1 到 t_N 之间平均距离的一个估计量,这还是在假定气球飞行轨迹近于直线的情况下才有这个意义。在一般情况下,连上述解释都不存在,因此就不能用取算术平均值的办法处理,因为这时所反映的测量的量的本质不同了,这时的测量是对一个飞行过程而言,每个观测点只是飞行过程在特定时间的某一个状态, N 个观测点 $(t_r, R_r), r = 1, 2, \dots, N$ 反映了一段飞行轨迹,我们对这组观测数据的分析就反映为对这一段飞行轨迹整体的认识,而不是对某一个特定观测点的认识。从整体的概念出发,我们对观测数据利用最小二乘法进行曲线拟合来代替静态测量中对观测数据取算术平均值的作法。

在上面这个例子中,由于气球飞行缓慢,我们假定距离随着时间的变化是线性的。这样,用一条直线来拟合这组观测数据,设该直线方程为

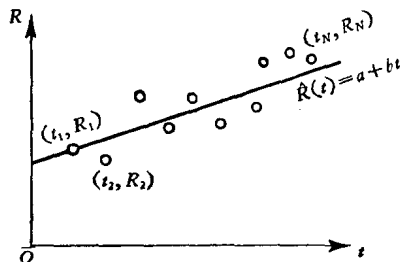


图 1.3

$$\hat{R}(t) = a + bt, \quad (1.1)$$

如图 1.3 所示。这里选择 a 和 b , 使得该直线到诸观测点的偏差的平方和 Q 达到最小,即使

$$Q = \sum_{r=1}^N (\hat{R}(t_r) - R_r)^2$$

达到最小。式中 $\hat{R}(t_r) - R_r$ 称为残差¹⁾，并用 $\hat{\theta}_r$ 表示，即

$$\hat{\theta}_r = \hat{R}(t_r) - R_r. \quad (1.2)$$

因此

$$Q = \sum_{r=1}^N \hat{\theta}_r^2. \quad (1.3)$$

使 Q 达到最小的必要条件是 Q 对 a 与 b 的偏导数都为零，即

$$\frac{\partial Q}{\partial a} = \frac{\partial Q}{\partial b} = 0. \quad (1.4)$$

由 (1.4) 式确定的 a 和 b 是使 Q 最大还是最小呢？由于在 Q 的表达式中的每一项都是非负的，因此， Q 必定有一个非负的最小值，解这两个线性方程仅仅得到一组解 a 和 b ，这必定是使 Q 成为最小的解。

(1.4) 式计算如下：

$$\left. \begin{aligned} \frac{\partial Q}{\partial a} &= -2 \sum_{r=1}^N (R_r - a - b t_r) = 0, \\ \frac{\partial Q}{\partial b} &= -2 \sum_{r=1}^N t_r (R_r - a - b t_r) = 0. \end{aligned} \right\} \quad (1.5)$$

如取

$$\begin{aligned} \bar{t} &= \frac{1}{N} \sum_{r=1}^N t_r, & \bar{R} &= \frac{1}{N} \sum_{r=1}^N R_r, \\ s_1 &= \frac{1}{N} \sum_{r=1}^N R_r t_r, & s_2 &= \frac{1}{N} \sum_{r=1}^N t_r^2, \end{aligned}$$

就有

$$b = \frac{s_1 - \bar{R}\bar{t}}{s_2 - (\bar{t})^2}, \quad a = \bar{R} - b\bar{t}. \quad (1.6)$$

1) 有时也用 $R_r - \hat{R}(t_r)$ 作为残差，这两者在讨论中是一样的。

这样,利用一组观测数据 $(t_r, R_r), r = 1, 2, \dots, N$ 就可估计出飞行轨迹的线性方程(1.1):

上述利用一组观测数据来确定某些参数(如这里的 a, b)的方法称为最小二乘法,最小二乘法的详细讨论,将在第五章给出。

将测量进行了分类,就便于从某些测量问题的特殊性出发,针对各类测量问题特有的矛盾,来讨论各种测量问题中共同存在的观测结果的精度分析、观测结果的数据处理等方面的问题。

§ 3 测量中的误差

通过测量工作的实践我们认识到,任何测量不管测量仪器多么精密,测量进行得多么精细,测量结果总是不可避免地带误差。例如,多次测量一个工件的长度(象轴的长度或直径)时,其所得的结果并不相同;对一段地面距离进行往返丈量,两次丈量的结果往往也不一致;在大地测量中我们知道,三角形是构成三角网的基本图形,而测量平面三角形的三个内角的和一般并不等于 180° ,产生上述这些情况是因为测量进行中必然带有误差所致。正因为进行测量必然带有误差这个共同的特点,这就决定了测量问题的处理不同于一般的数学问题。象上面提到的三角形的测量中,从几何上来讲,一个平面三角形有两个内角就可完全决定其形状,然而,从测量这个特定的问题来讲,就必须测定三个内角,在尽可能消除误差的前提下进行处理,决定三角形的形状。这是测量问题所特有的、区别于一般数学问题的本质所在。

测量中误差的产生有内在的原因和外界的原因两个方面。从内因来讲,测量仪器不准确存在着误差;当对测量数据需进行远距离传输时还存在着传输过程中的象误码那样的误

差以及因为观测者的感觉器官鉴别得不准确而带来的误差等等。属于影响测量误差的外界原因,象测量进行中所处的外界条件,随着温度、湿度、风等自然条件的变化而改变,因而对观测结果产生影响;在利用无线电或光学仪器作为测量手段的测量中,大气对电波或光波的折射也必然影响观测的效果。这些外界条件的影响使得观测结果必然含有误差。

全部测量工作的中心,不外采用种种办法尽可能地消除或减少误差对观测结果的影响,提高结果的精度。从这意义上来讲,可以说整个测量及其结果的处理过程也就是与误差作斗争的过程。

为了有效地向测量误差进行斗争,需要掌握误差出现的规律。尽管对测量来讲都是误差,都对测量结果的准确性产生影响,然而,误差性质却各有不同,将误差进行分类,以便对测量的结果按照误差的不同性质采用不同的方法加以处理。下面我们研究测量中出现的各种误差。

随机误差 在一定的观测条件下进行了一系列的观测,如果观测的误差从表面上可看出其数值和符号都是不定的,不存在任何确定的规律性,它只具有统计性的规律,这种误差称为随机误差。

随机误差的产生取决于测量进行中一系列随机性因素的影响。所谓随机性因素是指实验者不能加以严格控制的因素。

测量中的外界条件,如温度、湿度、空气振动等的瞬间变化,使得每次观测都受这些外界条件的随机扰动而带来随机误差。用无线电测量设备测量时,电波的随机干扰,光学测量的照准误差以及进行观测时不定的读数误差等等,都是产生随机误差的来源。

随机误差尽管从表面上看来没有什么规律,但根据大量的测量实践就总结出在一定的观测条件下随机误差所呈现的