高等学校教材

Basics of Errors Theory and Surveying Adjustment

误差理论与测量平差基础

丁安民 编

测绘出版社

高等学校教材

误差理论与测量平差基础

Basics of Errors Theory and Surveying Adjustment

丁安民　编

测绘出版社
·北京·

内容简介

本书是测绘工程本科专业必修的专业基础课通用教材。本书系统、全面地阐述了测量误差的基本理论、测量平差的基本原理，以及测量平差的具体应用。全书共分九章。主要内容包括：测量误差理论与最小二乘原理、测量平差基本方法、测量平差函数模型和随机模型的概念及建立、测量数据的统计假设检验方法和测量平差具体应用等。本书内容充实、结构严谨、体系完整、理论与应用并重，不仅包括了测量数据处理的经典理论，而且反映了测量平差的当代进展。

本书可作为高等学校测绘工程专业本科教材，也可供相关专业的工程技术人员参考。

图书在版编目(CIP)数据

误差理论与测量平差基础 / 丁安明编. —北京：测绘出版社，2012.8
ISBN 978-7-5030-2614-0

Ⅰ. ①误… Ⅱ. ①丁… Ⅲ. ①误差理论—高等学校—教材 ②测量平差—高等学校—教材 Ⅳ. ①0241.1 ②P207

中国版本图书馆 CIP 数据核字(2012)第 119104 号

责任编辑 贾晓林　**执行编辑** 余易举　**封面设计** 李　伟　**责任校对** 董玉珍

出版发行	测绘出版社	**电　话**	010－83060872(发行部)
地　址	北京市西城区三里河路 50 号		010－68531609(门市部)
邮政编码	100045		010－68531160(编辑部)
电子邮箱	smp@sinomaps.com	**网　址**	www.chinasmp.com
印　刷	北京建筑工业印刷厂	**经　销**	新华书店
成品规格	184mm×260mm		
印　张	11.5	**字　数**	280 千字
印　次	2012 年 8 月第 1 次印刷	**版　次**	2012 年 8 月第 1 版
印　数	0001－2000	**定　价**	28.00 元

书　号 ISBN 978-7-5030-2614-0/P・592

前　言

“误差理论与测量平差基础”是测绘类本科专业的基础核心课程之一。它应用概率和数理统计方法来分析观测数据，为观测数据的处理提供理论基础；以最小二乘法作为处理观测数据的基本原则，讲解测量平差的基本原理、方法和技能；论述近代测量平差的基本理论与方法，介绍测量数据处理的最新研究成果。目前，测绘工程专业涵盖了大地测量、工程测量、摄影测量与遥感和地图制图学等方向，原矿山测量专业现已归并到大地测量与测量工程专业。为贯彻落实卓越工程师教育培养计划，本教材编写组召集河南省有关院校教师对该课程的教学内容进行了修订，使得教学内容更加注重测量平差的基本概念、基础理论、基本知识和基本技能的讲解。

虽然目前各类教材的教学大纲相差不多，但考虑到新专业目录中测绘工程专业与旧专业目录中相应的细分专业的差异，培养模式和课程设置的不同，仍需对该教材的内容从基础化和综合化的角度进行修订。本教材的主要教学内容和教学体系与国内已出版的同类教材基本一致，但也有以下两个特点：

(1)除加强误差理论基础知识的讲解外，增加了矩阵知识在本课程中的应用，例如：矩阵的秩、迹的性质在本课程公式推导中的应用等。

(2)删去了一些陈旧的、与现代科技发展不相称的内容，例如：删去了收舍误差、线性方程组解算等。增加了参数加权平差，间接平差中的分组平差等内容。

在本书的编写中，力求范例的多样性和实用性，以达到培养宽口径人才的要求。本教材第1章到第9章由河南理工大学丁安民教授和张捍卫教授编写，第10章由河南城建学院卫柳艳副教授编写，河南理工大学魏峰远教授对第10章进行了修订。全书由张捍卫教授统稿。在本教材的出版过程中，河南理工大学测绘与国土信息工程学院领导给予了高度重视和关心，并提供了教材建设出版基金，在此表示真诚的感谢。

由于编著者水平所限，文中如有错误和不当之处，敬请读者予以指正。

丁安民

2012年3月

目　录

第1章 绪 论

地球科学的测量数据或观测数据是指用一定的仪器、工具、传感器或其他手段获取的反映地球与其他实体的空间分布有关的信息数据。任何观测数据总是包含有信息和干扰两部分，采集数据就是为了获取有用的信息。干扰也称为误差，是除了有用信息以外的部分。在实际工作中，需要进行大量观测数据的处理，它是测量工作重要环节之一。“误差理论与测量平差基础”课程的任务，就是介绍这一方面的有关理论和方法。高斯(Gauss)和勒让德(Legendre)于18世纪末创立了解决这一问题的基本理论和方法，即最小二乘法。两个世纪以来，随着科学与技术的不断进步，特别是近代科学与技术的发展，最小二乘法也增添了许多新的内容，理论更趋全面严谨，方法更加灵活多样，应用也更为广泛。

本章将说明观测数据总是不可避免地带有误差，以及测量平差所研究的内容，最后介绍本书的任务和内容。

§1.1 测量平差的基本概念

在测量工作中，由于受测量过程中客观存在的各种因素的影响，使得一切测量结果都不可避免地带有误差。例如，对一段距离进行重复观测时，各次观测的长度总不可能完全相同；又如，一个平面三角形三个内角之和理论上应等于180°，实际上，如果对这三个内角进行观测，其观测值之和一般不等于180°。这种差异的产生，是因为观测值中含有观测误差。于是，研究观测误差的内在规律，对带有误差的观测数据进行数学处理并评定其精确程度等，就成为测量工作中需要解决的重要实际问题。

1.1.1 误差来源

观测误差产生的原因很多，概括起来主要有以下四个方面。

(1)观测者：由于观测者的感觉器官的鉴别能力有一定的局限性，因此在仪器的安置、照准、读数等方面都会产生误差。同时，观测者的工作态度、技术水平以及情绪的变化，也会对观测成果的质量产生影响。

(2)测量仪器：所谓测量仪器，是指采集数据时所使用的任何工具和手段。由于每一种仪器只具有一定限度的准确度，由此观测所得的数据必然带有误差。测量是利用测量仪器进行的，由于测量仪器的不完善，因而使观测值产生误差。例如，在用只刻有厘米分划的普通水准尺进行水准测量时，就难以保证在估读厘米以下的尾数时正确无误。同时，仪器本身也有一定的误差。例如：光学经纬仪理论上要求主光轴、俯仰轴和垂直轴三轴相互正交，但实际上不可能严格正交；水准仪的视准轴不平行于水准轴；电磁波测距仪的零位误差、电路延迟；经纬仪和测距仪度盘的刻划误差等。这些因素都会使测量结果产生误差。

(3)外界环境：观测过程所处的客观环境，例如温度、湿度、风力、风向和大气折光等因素都

会对观测结果产生影响。同时，随着这些外界因素的变化，例如温度的高低、湿度的大小、风力的强弱和大气折光的不同，其对观测结果的影响也不同。在多样而变化的外界自然条件下进行观测，就必然使观测结果产生误差。目前，特别是高精度测量，更重视外界环境产生的观测误差，例如，GPS接收机所接收的是来自高空的卫星信号，经过电离层、中性大气层时都会发生信号延迟而产生误差等。

（4）观测对象：观测目标本身的结构、状态和清晰程度等，也会对观测结果直接产生影响。例如，三角测量中的观测目标觇标和圆筒由于风吹日晒而产生了偏差，GPS导航定位中的卫星星历误差、卫星钟误差和设备延迟误差等，都会使测量结果产生误差。

上述的观测者、测量仪器、外界环境和观测对象四个方面的因素是使测量过程中产生误差的主要来源，我们把这四个因素合称为测量条件。显然，测量条件好，观测中产生的误差就会小，观测成果的质量就会高；测量条件差，产生的观测误差就会大，观测成果的质量就会低；如果测量条件相同，观测误差的量级应该相同。我们把测量条件相同的观测称为等精度观测，在相同测量条件下所获取的观测值称为等精度观测值；而测量条件不同的观测称为非等精度观测，相应的观测值称为非等精度观测值。

由于测量条件不尽完善，测量误差是客观存在的。为了检验观测结果的精确性和提高观测结果的可靠性，实践中得出的有效方法是进行多余观测（也称过剩观测）。所谓多余观测就是多于必要观测的观测，例如，直接测定某一段距离的大小时，不是只观测一次，而是观测多次。这时，其中一次是必要观测，其他则为多余观测。又如，在确定平面三个角形形状时，不只是观测任意两个角，由此推算第三个角，而是三个角都观测，这时有两个是必要观测，另一个就是多余观测。

多余观测可以揭示测量误差，同时多余观测又使观测结果产生了矛盾。如平面三角形三个内角观测值之和不等于180°，即闭合差不等于0。为了消除矛盾，必须对观测结果进行平差，为此我们给出测量平差的基本概念。在多余观测的基础上，依据一定的数学模型和某种平差原则，对观测结果进行合理的调整，从而求得一组没有矛盾的最可靠结果，并评定精度，这一过程称为平差。测量平差所依据的原则一般为最小二乘原理。

1.1.2 观测误差及其分类

不难理解，误差是相对于绝对准确而言的。反映一个量真正大小的绝对准确的数值，称为这一量的真值。通过测量直接或间接得到的一个量的大小称为这个量的观测值。与真值相对应的，凡以一定的精确程度反映这一个量大小的数值，都统称为此量的近似值或估计值，简称估值。一个量的观测值或平差值，都是此量的估值。

设以 $\widetilde{L}$ 表示一个量的真值，L 表示它的某一观测值，Δ 表示观测误差，则有

$$\Delta = L - \widetilde{L} \quad (\text{或 } \Delta = \widetilde{L} - L)$$

这里 Δ 是相对于真值的误差，称为真误差。

真值通常是无法测知的，自然真误差也无法获得。但是在一些情况下，有可能预知由观测值构成的某一函数的理论真值。例如：以 L_1、L_2、L_3 表示平面三角形三个内角的观测值，三个内角和的理论真值为180°是已知的。若以 w 表示三内角和的真误差（即三角形闭合差），则

$$w = (L_1 + L_2 + L_3) - 180^\circ$$

因为闭合差真值为0,因此 $\Delta w=w-0=w$,故 w 也可理解为三角形闭合差的真误差。

当对同一个量观测两次,设观测值为 L_1 和 L_2,该量真值为 $\tilde{L}$,且以 d 表示两次观测值的差值,则

$$d=(L_1-\tilde{L})-(L_2-\tilde{L})=(L_1-L_2)-(\tilde{L}-\tilde{L})=L_1-L_2$$

表示两次观测值的差值 d,就是差值的真误差。d 又称较差。

一般情况下,测量误差按其对观测结果的影响性质可分为以下三类。

1. 粗差

是指比在正常观测条件下所可能出现的最大误差还要大的误差。一般情况下,粗差主要是由失误引起的,一般以异常值或孤值形式表现出来。例如,测错、读错、记录错、计算错和仪器故障等所引起的偏差等。

经典测量中,这类粗差一般采取变更仪器或操作程序、重复观测和检核验算、分析等方式,检出粗差并予以剔除。因此,可认为观测值中已基本没有粗差。由于在现代测量技术的自动化数据采集过程中,粗差经常混入信息之中。所以,在观测方案的设计、实施和观测中的检核及测后的分析处理中,采取有效措施进行粗差的探测和消除是非常重要的。

2. 系统误差

是指误差的大小、符号在观测过程中有一定规律的误差。例如,用具有某一尺长误差的钢尺量距时,由尺长误差所引起的距离误差与所测距离长度成正比例增加。另外,有的系统误差源对一部分观测呈规律性影响,而对另一部分观测的影响,其符号又有正有负,呈现出随机性,从总体上看,这种系统误差属于随机性系统误差,在测量中称为半系统误差。

在测量结果中,应尽量消除或减弱系统误差对观测成果的影响,通常采取如下措施:①找出系统误差出现的规律并设法求出它的数值,然后对观测结果进行改正;②改进仪器结构并制订有效的观测方法和操作程序;③综合分析观测资料,发现系统误差,在平差计算中将其消除。

从实际的测量结果中,完全消除系统误差是不可能的。对于在具体测量工作中,系统误差所引起的问题如何处理,将在各有关专业课中讨论,经典平差中不过多涉及这方面内容。作为平差对象的观测数据,一般认为已经消除了系统误差。

3. 偶然误差

是指单个误差的数值和符号没有规律性可循,但对于大量误差总体却存在一定统计规律性的误差。偶然误差是由各种随机因素的偶然性影响而产生的误差。例如,测量时环境变化对观测数据产生的微小影响等。因此,偶然误差就总体而言,都具有一定的统计规律性,故有时把偶然误差又称为随机误差。

在一切测量中,偶然误差是不可避免的。经典最小二乘平差就是在认为观测值仅含有偶然误差的情况下,调整误差,消除矛盾,求出最可靠值,并进行精度评定。

1.1.3 最小二乘原理

设 $L_1,L_2,\cdots,L_n$ 表示 n 个独立不等精度观测结果,且赋予它们相对可信赖程度的数值分别为 $p_1,p_2,\cdots,p_n$,为消除矛盾而赋予观测值对应的改正数为 $v_1,v_2,\cdots,v_n$。最小二乘原理指的是

$$\sum_{i=1}^{n} p_i v_i^2 = [pvv] = \min$$

在本教材中一般以矩阵形式表示，即

$$V^{\mathrm{T}}PV=\min$$

式中

$$V=\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix},\ P=\begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_n \end{bmatrix}$$

这里 P 称为观测值向量的权矩阵，$Q=P^{-1}$ 称为观测值向量的权逆矩阵或协因数矩阵。

依最小二乘原理，平差计算所得的观测值改正数，称为最或然改正数（简称改正数），也称残差；经最小二乘平差计算求得的有关量的最可靠值，称为该量的最或然值，又叫平差值。不仅是通过观测直接或间接得到的数据要进行平差处理，有时对一些已平差过的值也要进行再平差。因此参加平差的数据不仅包括直接观测值，还包括间接观测值和在一定范围内已进行过平差的数据，今后无特别注明时，总是概括地称它们为观测值。

§1.2 测量平差发展简史

如何消除由于观测误差引起的观测值之间的矛盾，从多于待估量的观测值中求出其最优值一直是科技工作者高度关注的问题。1794 年，年仅 17 岁的高斯首先提出了解决这个问题的方法——最小二乘法。他是以算术平均值为待求量的最或然值，观测误差服从正态分布这一假设导出了最小二乘原理。高斯当时利用最小二乘原理成功地预测了谷神星运行轨道，使天文学家又很快找到了这颗行星，但高斯并没有及时发表他的成果。直到 1809 年，高斯才在《天体运动理论》一书中，从概率的观点详细地叙述了他所提出的最小二乘原理。而在此之前，1806 年，勒让德发表了《决定彗星轨道的新方法》一文，从代数的观点独立地提出了最小二乘法原理，并定名为最小二乘法。所以后人称它为高斯-勒让德方法。

自高斯提出最小二乘原理到 20 世纪五六十年代以来，许多学者对测量平差的理论和方法进行了大量的研究，提出了一系列解决各类测量问题的平差方法。这些平差方法都是基于观测值随机独立的假设，一般称为经典最小二乘平差。这一时期，由于计算工具的限制，测量平差的主要研究方向是如何少解线性方程组。自 20 世纪六七十年代开始，随着空间测地技术的发展，电子计算机的广泛应用，以及近代数学在测量平差中的应用，测量平差理论取得了很大发展，出现了许多新的平差理论和平差方法。例如：田斯特拉于 1947 年提出相关观测值的平差理论，将经典平差中对观测值随机独立的要求，推广到随机相关的观测值；克拉鲁普在 1969 年提出最小二乘滤波、推估和配置理论；迈塞尔于 1962 年将高斯的最小二乘平差模型中的列满秩系数矩阵推广到奇异矩阵，提出了解决非满秩平差问题的秩亏自由网平差方法，之后其他学者综合各种情况得到了广义高斯-马尔可夫平差模型，并把广义高斯-马尔可夫模型的参数估计称为最小二乘统一理论；20 世纪 80 年代以来，有人将经典的先验定权方法改进为后验定权方法，提出了多种方差-协方差分量的验后估计法；20 世纪中期以来，很多学者致力于系统误差和粗差的研究，提出了附加系统参数的平差方法和粗差探测理论。

总之，自 20 世纪 70 年代以来，随着全球定位系统（GPS）、地理信息系统（GIS）和遥感系统（RS）在测绘中的应用，测量平差理论和方法得到了飞速发展，出现了许多新的测量数据处理

理论和方法，也推动了测量平差理论的发展。

§1.3 本书的研究任务和内容

测量平差的主要任务有两个：一是依最小二乘原理求出待定量的最可靠值；二是评定观测值和平差结果的精度。本书主要任务是系统介绍最小二乘原理与测量平差的基本理论和方法，为以后的专业课学习和科学研究打下基础，其主要内容为：

(1)偶然误差概率特性，随机变量的数字特征，衡量观测值质量优劣的标准，精度指标及其估计，中误差与权的定义，方差矩阵及权逆矩阵传播规律等。

(2)测量平差函数模型和随机模型的概念与建立，参数估计概念和最小二乘原理。

(3)间接平差、条件平差、附有参数的条件平差和具有约束条件的间接平差方法等，按照最小二乘原理导出平差结果和精度评定公式，以及平差值的统计性质。作为平差方法的扩充，介绍了参数加权和分组的平差方法。

(4)测量平差中必要的统计假设检验方法。

第2章　误差理论基础

所谓测量就是用计量仪器对被观测量进行量度。测量值就是用测量仪器测定待测量所得到的数值。任一待测量都有它的客观大小，这个客观量称为真值。最理想的测量就是能够测得真值，但由于测量都是在一定条件下进行的，由于各种因素的影响不可避免地带有误差。测量平差的基本任务就是处理一系列带有误差的观测值，求出未知量的最佳估值，并评定测量成果的精度。解决这两个问题的基础，就是要研究观测误差的理论，简称误差理论。偶然误差是一种随机变量，就其总体来说具有一定的统计规律。本章将从随机变量的概念着手，引出测量中常用的精度含义，定义评定观测精度的指标，介绍平差理论中的权、协因数矩阵(即权逆矩阵)等重要概念，以及测量数据处理中常用的由观测值函数的真误差估计中误差的方法。

§2.1　随机变量及其概率分布

在一定测量条件下对某个量进行观测时，并不总是出现相同结果的现象称为随机现象。这种在一定条件下由于偶然因素影响，其观测结果可能取各种不同的值，具有不确定性和随机性，但这些取值落在某个范围的概率是一定的，此种变量称为随机变量。例如，某一时间内公共汽车站等车乘客人数，电话交换台在一定时间内收到的呼叫次数等，都是随机变量的实例。

2.1.1　一维随机变量

要全面了解一个随机变量，不但要知道它取哪些值，而且要知道它取这些值的规律，即要掌握它的概率分布。一个随机实验的可能结果(称为基本事件)的全体组成一个基本空间 Ω。随机变量 x 是定义在基本空间 Ω 上的取值为实数的函数，即基本空间 Ω 中每一个点，也就是每个基本事件都有实轴上的点与之对应。例如，随机投掷一枚硬币，可能的结果有正面朝上、反面朝上两种，若定义 x 为投掷一枚硬币时正面朝上的次数，则 x 为一随机变量。当正面朝上时，x 取值1；当反面朝上时，x 取值0。这样 x 就是一个离散型的随机变量，且 $\Omega=\{0,1\}$。如果在确定的条件下，已知硬币正面朝上的概率为 p(硬币正面朝上的次数与投币总次数之比)，则 x 的概率分布为

$$P\{x=1\}=p,\ P\{x=0\}=1-p$$

如果随机变量 y 取有限个可能值：$\Omega=\{y_1,y_2,\cdots,y_{N-1},y_N\}$，$N$ 为自然数，则其概率分布为

$$\left.\begin{aligned}&P\{y=y_j\}=p_j\\&\sum_{j=1}^{N}p_j=1\end{aligned}\right\}\tag{2.1.1}$$

随机变量不但可以是离散型的，也可以是连续型的。如分析测试中的测定值就是一个以概率取值的随机变量，被测定量的取值可能在某一范围内随机变化，具体取什么值在测定之前是无法确定的，但测定的结果是确定的，多次重复测定所得到的测定值具有统计规律性。随机

变量与模糊变量不确定性的本质差别在于，后者的测定结果仍具有不确定性，即模糊性。如果 x 是一个连续型的随机变量，而它的可能值 X 可以是区间 $(-\infty,+\infty)$ 上的任何值。此时，要把随机变量的所有可能值都列出来是不可能的。对于连续型的随机变量 x，其概率分布是 X 的一个函数，记为

$$F(X)=P\{x<X\} \tag{2.1.2}$$

这就是在实验中，随机变量 x 落在 X 左端的概率。它有下列性质：

(1)当 $X_2>X_1$ 时，有 $F(X_2)\geqslant F(X_1)$，即 $F(X)$ 是所有自变量的非减函数；

(2)在负无穷远处，$F(-\infty)\equiv 0$，即将 X 点无限向左移动，则随机事件 x 落在 X 左端成为不可能事件；

(3)在正无穷远处，$F(+\infty)\equiv 1$，即将 X 点无限向右移动，则事件 $x<X$ 在极限过程中成为必然事件。

若 $F(X)$ 连续可微，则引进概率密度函数 $f(x)$ 的定义，即

$$f(X)=\frac{\mathrm{d}F(X)}{\mathrm{d}X} \tag{2.1.3}$$

它明确表达了密度的性质。如果用 $f(x)\mathrm{d}x$ 表示概率元素，则有下列关系式：

(1)随机变量 x 落入区间 (a,b) 的概率为 $P\{a<x<b\}=\int_a^b f(x)\mathrm{d}x$。

(2)随机变量 x 落入区间 $(-\infty,X)$ 的概率为 $F(X)=\int_{-\infty}^{X} f(x)\mathrm{d}x$。

(3) 随机变量 x 落入区间 $(-\infty,+\infty)$ 的概率为 $F(+\infty)=\int_{-\infty}^{+\infty} f(x)\mathrm{d}x\equiv 1$。

2.1.2　二维随机变量

有些随机现象需要同时用多个随机变量来描述。例如，子弹着点的位置需要两个坐标才能确定，它是一个二维随机变量。类似地，需要 n 个随机变量来描述的随机现象中，这 n 个随机变量组成 n 维随机向量。描述随机向量的取值规律，用联合分布函数。随机向量中每个随机变量的分布函数，称为边缘分布函数。如果联合分布函数等于边缘分布函数的乘积，则称这些单个随机变量之间是相互独立的。独立性是概率论所独有的一个重要概念。

称两个事件 $x<X,y<Y$ 同时出现的概率为二维随机变量 (x,y) 的概率分布，即

$$F(X,Y)=P\{x<X,y<Y\} \tag{2.1.4}$$

它有下列性质。

(1)当 $X_2>X_1,Y_2>Y_1$ 时，分别有

$$F(X_2,Y)\geqslant F(X_1,Y),\ F(X,Y_2)\geqslant F(X,Y_1)$$

即概率分布 $F(X,Y)$ 为两个变量 X、Y 的非减函数；

(2)当 X、Y 有一个趋于负无穷大时，有

$$F(-\infty,Y)=F(X,-\infty)=F(-\infty,+\infty)\equiv 0$$

即概率分布中有一个变量取 $-\infty$，其值为 0；

(3)当 X、Y 有一个趋于正无穷大时，有

$$F(X,+\infty)=F_1(X),\ F(+\infty,Y)=F_2(Y)$$

(4)当 X、Y 两个都趋于正无穷大时，有

$$F(+\infty,+\infty)\equiv 1$$

即随机点(x,y)落入全平面中，成为必然事件。

如果 $F(X,Y)$连续可微，则定义二维随机变量的概率密度函数为

$$f(X,Y)=\frac{\partial^2 F(X,Y)}{\partial X\partial Y} \tag{2.1.5}$$

如果用 $f(x,y)\mathrm{d}x\mathrm{d}y$ 表示二维概率元素，则随机点(x,y)落在任意区域 D 内的概率表达式为

$$P\{(x,y)\subset D\}=\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$$

而

$$F(X,Y)=\int_{-\infty}^{X}\int_{-\infty}^{Y}f(x,y)\mathrm{d}x\mathrm{d}y$$

因此

$$F_1(X)=F_1(X,+\infty)=\int_{-\infty}^{X}\mathrm{d}x\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$$

从而得

$$f_1(X)=\frac{\partial F_1(X)}{\partial X}=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$$

同理可得

$$f_2(Y)=\frac{\partial F_2(Y)}{\partial Y}=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x$$

且易知

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y\equiv 1$$

以上是二维随机变量的概率分布，多维可类推。

2.1.3 条件概率密度

设 A、B 为事件，则$(A\cap B)$代表 A 和 B 同时发生的事件。如 N 代表实验总数，$N(A)$为事件 A 出现的次数，$N(B)$为事件 B 出现的次数，而 $N(A\cap B)$为事件$(A\cap B)$出现的次数，则其概率分别为

$$P(A)=\frac{N(A)}{N},\ P(B)=\frac{N(B)}{N},\ P(A\cap B)=\frac{N(A\cap B)}{N}$$

在给定事件 B 的条件下，事件 A 出现的概率为

$$P(A\backslash B)=\frac{N(A\cap B)}{N(B)}=\frac{\dfrac{N(A\cap B)}{N}}{\dfrac{N(B)}{N}}=\frac{P(A\cap B)}{P(B)} \tag{2.1.6}$$

定义为条件概率。

现在讨论两个有联合概率密度的随机变量 x、y，令 A、C 分别代表下列两个事件

$$A=\{x<X\},\ C=\{Y\leqslant y\leqslant Y+\Delta Y\}$$

由上述定义可知，在给定事件 C 的条件下，事件 A 的条件概率为

$$P(A\backslash C)=\frac{P(A\cap C)}{P(C)}$$

即

$$P\{(x<X)\backslash(Y\leqslant y\leqslant Y+\Delta Y)\}=\frac{\int_{-\infty}^{X}\mathrm{d}x\int_{Y}^{Y+\Delta Y}f(x,y)\mathrm{d}y}{\int_{Y}^{Y+\Delta Y}f_2(y)\mathrm{d}y}$$

其中

$$f_2(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x$$

为随机变量 y 的概率密度函数。

如果利用积分中的中值定理，并令 $\Delta Y\to 0$，则得

$$P\{(x<X)\backslash(y=Y)\}=\frac{\int_{-\infty}^{X}f(x,y)\mathrm{d}x}{f_2(y)}$$

如令 $F\{X\backslash(Y=y)\}=\{(x<X)\backslash(y=Y)\}$，则就是在给定 Y 的条件下，随机变量 x 的条件概率分布。从而可得条件概率密度为

$$f(x\backslash y)=\frac{\partial F\{X\backslash(Y=y)\}}{\partial X}=\frac{f(x,y)}{f_2(y)} \tag{2.1.7}$$

同理可得

$$f(y\backslash x)=\frac{f(x,y)}{f_1(x)} \tag{2.1.8}$$

或者

$$f(x,y)=f(x\backslash y)f_2(y)=f(y\backslash x)f_1(x) \tag{2.1.9}$$

这就是贝叶斯(Bayes)公式。

两个随机变量 x、y，如果在给定其中一个的条件下，而另一个的条件概率密度不依赖于给定的条件，即

$$f(x\backslash y)=f_1(x),\ f(y\backslash x)=f_2(y) \tag{2.1.10}$$

由贝叶斯公式可得两个随机变量 x、y 相互独立的充分必要条件为

$$f(x,y)=f_1(x)f_2(y) \tag{2.1.11}$$

如果 x、y 皆为随机向量，上式结果仍然正确。

§2.2 偶然误差的概率特性

偶然误差从表面上看没有任何规律性，但是随着对同一量观测次数的增加，大量的偶然误差就表现出一定的统计规律性。为寻求偶然误差的规律性，下面首先通过测量实例来总结其规律，然后上升到理论进行分析。

某测区，在相同测量条件下，独立地观测了 358 个三角形的全部内角，由

$$\Delta_i=(L_1+L_2+L_3)_i-180^\circ\quad(i=1,2,\cdots,358)$$

算得各三角形的闭合差。由于作业中已尽量剔除了粗差和系统误差影响，这些三角形闭合差，

就整体而言，都是偶然因素所至，故为偶然误差。现将误差出现的范围分为若干相等的小区间，每个区间的长度 $d\Delta$ 为 0.20″。将一组误差按其正负号与误差值的大小排列，统计误差出现在各区间内误差的个数 v_i，以及“误差出现在某个区间内”这一事件的频率 v_i/n（此处 $n=358$），其结果列于表 2.2.1 内（等于区间左端值的误差算入该区间内）。

表 2.2.1 三角形闭合差分布

误差的区间/(″)	Δ为负值			Δ为正值		
	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
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
总和	181	0.505		177	0.495	

从表 2.2.1 中，不难发现如下的规律：①这些闭合差数值上不会超出一定界限；②绝对值小的闭合差比绝对值大的闭合差出现的个数要多；③绝对值相等的正负闭合差个数大致相等。上述情况不仅表现于这个例子里，在大量的测量结果中，偶然误差都有与此完全一致的规律性。

误差分布的情况，除了采用上述误差分布表的形式描述外，还可以利用图形来表达。例如，以横坐标表示误差的大小，纵坐标代表各区间内误差出现的频率除以区间的间隔值，即

$$\frac{v_i}{n d\Delta}$$

此处间隔值均取均 $d\Delta=0.20''$。根据表 2.2.1 给出的数据，可绘制出图 2.2.1。可见，此时图中每一误差区间上的长方条面积就代表误差出现在该区间内的频率。图 2.2.1 中画有斜线的长方条面积，就是代表误差现在 0.00″～+0.20″区间内的频率 0.128。通常称这样的图为直方图，它形象地表示了误差的分布情况。

可以预期，在相同的观测条件下所得到的一组独立的观测误差，只要观测误差的总个数 n 足够大，那么误差出现在各区间内的频率就总是稳定在某一常数（理论频率）附近，而且当观测个数愈多时，稳定的程度也就愈高。例如，就表 2.2.1 的一组误差而言，在观测条件不变的情况下，如果再继续观测更多的三角形，则可预见，随着观测的个数愈来愈多，误差出现在各区间内的频率及其变动的幅度也就愈来愈小。当 $n\to\infty$ 时，各频率也就趋于一个完全确定的数值，这就是误差出现在各区间的概率。这就是说，在一定的观测条件下，对应着一种确定的误差分布。

当 $n\to\infty$ 时，由于误差出现的频率已趋于完全稳定，如果此时把误差区间的间隔无限缩

小，则可想象到，图 2.2.1 中各长方条顶边所形成的折线将变成如图 2.2.2 所示的一条光滑曲线。这种曲线也就是误差的概率分布曲线。通常也称偶然误差的频率分布为其经验分布，当 $n\to\infty$ 时，经验分布的极限称为误差的理论分布。误差的理论分布是多种多样的，但是根据概率论的中心极限定理，大多数测量误差都服从正态分布。因而通常将正态分布看做偶然误差的理论分布。在以后的讨论中，我们都是以正态分布作为描述偶然误差分布的数学模型，这不仅可以带来工作上的便利，而且基本上也是符合实际情况的。

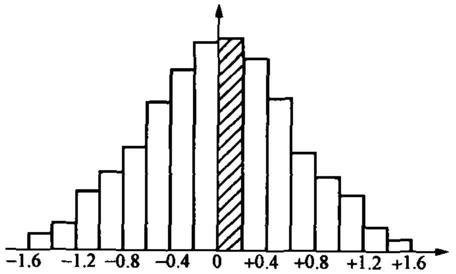

图 2.2.1

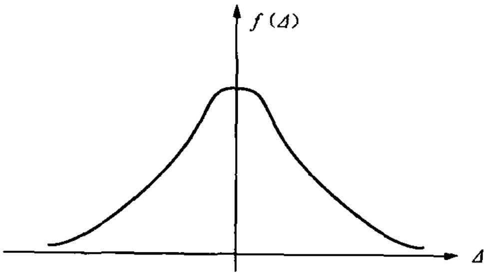

图 2.2.2

中心极限定理指出：若随机变量 y 是众多随机变量 $x_i(i=1,2,3,\cdots,n)$ 之和。如果 x_i 相互独立，且对 y 的影响均匀的小，则当 n 相当大时，随机变量 y 趋于服从正态分布。偶然误差正是这一类型的随机变量。通过上面的分析讨论，可用概率的术语将偶然误差的规律性阐述如下。

(1)在一定的测量条件下，偶然误差的数值不超过一定限值，或者说超出一定限值的误差出现的概率为 0；

(2)绝对值小的误差比绝对值大的误差出现的概率大；

(3)绝对值相等的正负误差出现的概率相同。

这就是偶然误差的三个概率特性，这三个特性可简要概括为：界限性、聚中性和对称性。它们充分揭示了表面上似乎并无规律性的偶然误差的内在规律。

偶然误差的界限性表明，在一定测量条件下，偶然误差的数值是有一定范围的。因此我们可以根据测量条件来确定偶然误差出现的界限。显然测量条件愈好，可能出现的最大偶然误差愈小；反之，则愈大。所以界限性是以后讨论极限误差的理论依据。

偶然误差的聚中性表明，偶然误差愈接近 0，其分布愈密，而且易知，对于较好的测量条件这一特性也必相对明显和突出。

偶然误差的对称性表明，正负偶然误差的分布对称于 0，故其密度函数必为偶函数，有相互抵消性。当误差个数足够多时，其算术平均值应趋于 0，即

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Delta_i = 0 \tag{2.2.1}$$

由此又知，偶然误差的分布是以 0 为对称中心，此中心常称做离散中心或扩散中心。

图 2.2.1 中各长方条的纵坐标为

$$\frac{v_i}{n\mathrm{d}\Delta}$$

其面积即为出现在该区间内的频率。如果将这个问题提到理论上来讨论，则以理论分布取代经验分布。此时，图 2.2.1 中各长方条的纵坐标就是 Δ 的密度函数 $f(\Delta)$，而长方条的面积为