误差理论与实验数据处理

周 秀 银 编

北京航空航天大学出版社

误差理论与实验数据处理

周 秀 銀 编

北京航空航天大學出版社

内 容 简 介

本书主要介绍实验数据处理的基本理论和方法。主要包括误差理论、基本测量问题的处理、系统误差、误差的合成与分配、静态实验和动态实验数据的处理、有效数字及计算法则、随机过程及其数据处理。

本书可作为高等院校有关专业的教材，也可供一般从事测试专业的有关工程技术人员参考。

误差理论与实验数据处理
WUCHA LILUN YU SHIYAN SHUJU CHULI
周秀銀 编
负任编辑 刘 忠 曾昭奇
北京航空航天大学出版社出版
新华书店总店科技发行所发行 各地新华书店经售
昌平振兴胶印厂印装
787×1092 1/16 印张：16.5 字数：409千字
1986年3月第一版 1989年3月第二次印刷 印数：6001～11000册
ISBN 7-81012-097-2/TB·023 定价：4.90元

前　言

实验数据的处理历来就是测试技术学科的重要内容之一。近年来，测试手段的不断完善及计算机的广泛应用，为实验数据的处理提供了坚实的物质基础。随着科学技术的发展，实验数据处理的重要性日益突出。因此，很多高等院校有关测试的专业陆续开设了实验数据处理课程。北京航空学院从1981年起为仪表及测试专业的大学本科生开设了《误差理论与实验数据处理》课程，当时所用教材是北京航空学院黄俊钦教授1979年为硕士研究生编写的讲义。在1982年航空院校专业教材第三编委会上，决定由北京航空学院编写本科生用的《误差理论与实验数据处理》课程教材，该教材的初稿由本书的编者于1983年编出。

本书是以1983年的教材为基础，并根据使用情况作了修改。该教材编写时，注意了与主要前修课（自动控制原理、概率与数理统计等）的联系，并将重点放在何如运用基本理论解决具体数据处理问题上，力求作到体系完整、概念清晰、叙述简明、便于自学。在编写本书过程中，参考了1964年以来国内外有关误差理论及数据处理的主要书籍、期刊及教材，并得到南京航空学院王厚枢、清华大学魏平田等同志的关怀和帮助，在此一并表示感谢。

由于编者水平有限，难免有不妥之处，恳請读者批評、指正。

目　　录

第一章 绪　论

§1-1　测量的任务与测量的分类

任何一个被测量，如钢球的直径、密闭容器中的压力、飞机的飞行速度等，都有一客观存在的值——真值A_0，测量的任务或者说最终目的是测出被测量的真值A_0。但是，由于测量仪器、测量方法、测量环境、人们的观测能力、测量程序等等都不能做到完美无缺，故真值是无法测得的。因此，在生产与科学实验中，通常用有限次测量的算术平均值$\bar{x}$作为真值A_0的最佳估计值，已知

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{1-1-1}$$

式中　n为重复测量的次数；x_i为第i次的测量值。

测量的分类方法很多。如按被测量获得的方式分为直接测量与间接测量；按被测量的状态分为静态测量与动态测量；按被测量是电量还是非电量分为电测量与非电测量等等。

直接测量是将被测量与作为标准的量直接进行比较，或者用标准标定了的仪器对被测量进行测量，从而直接（不需要通过方程式计算）求得被测量。例如，用尺子量长度、用温度计测量温度、用电流表测量电流等。间接测量是指通过直接测量与被测量有确定函数关系的其它量后经过计算得到被测量的测量方法。例如，在测量导线的导电率ρ时，由于没有直接读出ρ值的仪器，便可利用下列关系式

$$\rho=\frac{4l}{\pi d^2 R} \tag{1-1-2}$$

先测出导线长度l、导线电阻R及导线直径d，然后将l、R及d代入式（1-1-2）求出ρ值。d、l及R的测量属于直接测量，并称它们为直接测量量；ρ的测量属于间接测量，并称ρ为间接测量量。

静态测量是指在测量过程中被测的量是不变的。动态测量（或瞬态测量）是指在测量过程中被测的量是变化的，例如化学爆炸周围的压力波的测量就属于动态测量。随着科学技术的发展，动态测量的问题在各个领域都逐渐提到日程上来了，这是值得我们注意的。

§1-2　误差与误差的分类

在任何一种测量中，不可避免地总有误差。这首先表现为，在同样条件下对同一对象进行重复测量时，无论所用的仪器多么精确，方法多么完善，实验者多么细心，最后得到的各次测量结果大多数是不一样的。这是因为众多因素在影响测量过程，而且各种影响因素还在经常不断地变化着。其次，被测之量要对仪器施加作用才能使仪器给出测量结果，因而测量结果并不完全准确地反映被测对象本来的面貌。

通常将误差Δ_x定义为某被测量的给出值x与真值A_0之差，即

$$\Delta_x = x - A_0 \tag{1-2-1}$$

给出值包括测量值、示值、标称值及计算近似值等。

误差乘以（−1），称为修正量 c，即

$$c = -\Delta_x = A_0 - x \tag{1-2-2}$$

经整理

$$A_0 = x + c = x - \Delta_x \tag{1-2-3}$$

由于测不到真值 A_0，故得不到按式（1-2-1）定义的误差 Δ_x。在实际测量中，用得较多的是偏差Δ。偏差Δ是被测量的给出值x与某约定值T之差，即

$$\Delta = x - T \tag{1-2-4}$$

在处理具体问题时，由于x与T的取法不一样，因此Δ可有不同的含义。例如 $\Delta = x - A_0$ 表示测量值与真值的差；$\Delta = \bar{x} - A_0$ 表示算术平均值与真值的差；$\Delta = x - \bar{x}$ 表示测量值与算术平均值的差。

从不同角度出发，误差有各种分类方法：按误差的表达形式分为绝对误差、相对误差；按误差的性质及产生的原因可分为系统误差、随机误差（又称偶然误差）、过失误差（又称粗差、巨差）。

一、绝对误差与相对误差

（一）绝对误差

以被测参数的量纲所表示的误差及偏差都是绝对误差，如式（1-2-1）与式（1-2-4）所示。

（二）相对误差

绝对误差与同量纲的某参考值之比称为相对误差，一般用百分数表示。由于绝对误差及参考值的取法不同，相对误差又可分为：

$$\text{标称相对误差} = \frac{\Delta_x}{x} \times 100\% \tag{1-2-5}$$

$$\text{实际相对误差} = \frac{\Delta_x}{A_0} \times 100\% \tag{1-2-6}$$

$$\text{额定相对误差} = \frac{\Delta_x}{x_{f.s}} \times 100\% \tag{1-2-7}$$

式中　$x_{f.s}$ 为被测量的满量程值。

对于传感器，通常取最大额定相对误差作为传感器的精度指标，即有

$$\text{最大额定相对误差} = \frac{\Delta_{x\,max}}{x_{f.s}} \times 100\% \tag{1-2-8}$$

其中，$x_{f.s}$ 为传感器的满量程值，$\Delta_{x\,max}$ 为满量程范围内的最大测量误差。例如，某压力传感器的测量范围是2～50公斤/厘米^2，在该范围内的最大绝对误差是0.14公斤/厘米^2，则该传感器的精度为

$$\text{精度} = \frac{0.14}{50-2} \times 100\% \approx 0.3\% \tag{1-2-9}$$

即精度为千分之三。

对于多档和用示值表示被测量大小的仪表，通常按最大引用误差划分仪表的精度等级，引用误差的定义如下：

$$引用误差=\frac{示值误差}{满刻度值} \tag{1-2-10}$$

$$最大引用误差=\frac{最大示值误差}{满刻度值} \tag{1-2-11}$$

可见，此处的引用误差与上述的额定相对误差是一致的。

电工类仪表分为0.1，0.2，0.5，1.0，1.5，2.5和5.0七级。对s级仪表（s指七级中的任一级），表明合格仪表的最大引用误差不会超过$s\%$，例如精度为0.5级、量程为300毫伏的电压表，则有：最大引用误差$\leqslant 0.5\%$；最大绝对误差$\leqslant 300\times 0.5\%$毫伏。在对某电工类仪表进行检定时，要根据算得的最大引用误差确定属于七级中的某一级，而不能根据最大引用误差的数值定级。例如最大引用误差为0.15%时，因沒有0.15这一级，故应定为0.2级。

二、系統误差、随机误差及过失误差

（一）系统误差

系统误差是指服从某一函数规律的误差。这里，将固定不变也理解为一种规律，因此常值误差也是系统误差。

系统误差根据需要可以有不同的分类方法：

1. 根据对误差掌握的程度分为确定性系统误差（误差的大小和方向均确切掌握）和不确定系统误差；

2. 根据误差的变化规律分为常值、累进性的、周期性的以及按复杂规律变化的系统误差；

3. 按误差产生的原因可分为：

(1) 工具误差——工具误差也称为仪器误差（简称仪差）。这是由于测量所用工具（仪器、量具等）本身不完善而产生的误差。

(2) 装置误差——这是由于测量设备和电路的安装、布置及调整不得当而产生的误差。

(3) 人身误差——这是由于测量人员的感觉器官和运动器官不完善而产生的误差。这类误差往往因人而异，并与个人当时的生理和心理状态密切有关。

(4) 外界误差——外界误差也称为环境误差。这是由于外界环境（溫度、湿度、电磁場等等）的影响而产生的误差。

(5) 方法误差——方法误差也称为理论误差。这是由于测量方法本身所形成的误差，或者由于测量所依据的理论本身不完善等原因而导致的误差。

上述的分类并不严谨，一个具体的误差往往可以归入这一类，也可以归到另一类。

重要的是系统误差的出现一般具有规律性，其产生的原因往往可知或能掌握。一般地说，首先应尽可能设法预见到各种系统误差的具体来源，并且极力设法消除其影响；其次，是设法确定或估计出未能消除的系统误差之值。

如果存在着某项系统误差而不知道，这是极不利的。因为通过对测量数据的统计处理不一定能发现它是否存在。特别是常值误差，仅凭数据的统计处理是既不能发现，也不能消除

该项误差。

（二）随机误差

当在同一条件下对同一对象反复进行测量时，在极力消除或改正一切明显的系统误差之后，每次测量结果仍会出现一些无规律的随机性变化，这些变化是由随机误差造成的。随机误差的出现，从表面上看毫无规律，似乎纯属偶然，故随机误差亦称为偶然误差。

产生随机误差的因素很多，有些因素虽然知道，但无法准确控制。例如温度、湿度及空气的净化程度等对测量都有影响，在测量时虽力求将它们控制为某个定值，然而在每一次测量时，它们都存在微小的变化。另外，还有一些产生随机误差的因素，尚未发现。总之，产生随机误差的大多数因素和产生系统误差的因素是一样的，只不过是由于变化的因素太多，且各种因素的影响太微小或太复杂，以致无法掌握其具体规律。

对同一被测量，在相同的条件下进行多次重复测量时，随机误差有大有小、有正有负。就其个体而言，随机误差没有规律，不可预料、也不可控制。但就随机误差的总体而言（若测量次数足够多），则可以发现它服从统计规律，从而可以从理论上对随机误差进行研究。随着对某一物理量重复测量的次数的增加，随机误差的算术平均值将逐渐接近于零。因此，多次测量结果的算术平均值将更接近于真值（如果不存在系统误差）

（三）随机误差与系统误差的辩证关系

实际上，常常把某些掌握不到的具有复杂规律的系统误差看作随机误差，甚至把某些虽可以掌握而过于复杂的系统误差也当作随机误差来处理。

此外，有不少系统误差，它的出现也往往带有随机性，这种系统误差一般当作随机误差来处理。

有时，系统误差与随机误差的区分，往往取决于时间因素，在短时间内（例如在一次测量过程中）基本上不变的误差，显然可示为一个常值误差。但时间一长（例如在两次检定周期之间），该项误差则很可能出现随机的变化。

总之，系统误差与随机误差之间并不存在不可逾越的鸿沟。随着人们对误差来源及其变化规律的认识加深，就有可能把以往认识不到而归为随机误差的某项误差予以澄清而明确为系统误差，并加以技术上的适当处理而使其减小或消除掉。反之，当认识不足时，也常把系统误差当作随机误差。

最后，应该指出，在任何一次测量中，系统误差与随机误差一般都同时存在，在具体处理时，常按其对测量结果的影响分别三种情况对待。

(1) 系统误差的影响远大于随机误差的影响，相对地说，随机误差可以忽略不计，此时基本上按纯系统误差来处理。

(2) 系统误差小得可忽略，或经技术上的处理后已经改正（这是正常的测量状况），此时基本上可按纯随机误差来处理。

(3) 系统误差与随机误差的影响相差不多，两者均不可忽略，此时应分别按不同方法来处理。

（四）过失误差

它是一种与事实不符的误差，它主要是由于测试人员粗枝大叶、过度疲劳或操作不正确引起的。例如，读错刻度尺，反读游标尺、记录错误、计算错误等。此类误差只要多方警惕、细心操作，一般都可以避免。

§1-3 测量的精密度、准确度及精确度

精密度表示在多次重复测量中所测数据的重复性或分散程度。随机误差小，重复测量的结果就密集，即重复性好。

准确度表示测量结果与被测量的眞值之间的偏离程度。系统误差小，准确度就高。

精确度又称精度，它是测量结果的精密度与准确度的综合反映。

可见，精密度高而准确度不一定高，反之，准确度高而精密度不一定高。如果精密度和准确度都高时，测量的精度一定高。

下面用图 1-3-1 对精密度、准确度进行说明。图中 D_t 为圆柱体的眞实直径，D_i 为每次的测量值，$\overline{D}$ 为测量的算术平均值，重复测量了 n 次，则 $\overline{D}=\frac{1}{n}\sum_{i=1}^{n} D_i$。以 i 为横坐标，$i=1$ 表示第一次测量，$i=n$ 表示第 n 次测量。以 D_i 作散点图，根据测量情况可有图 1-3-1 中所示的四种情况：(a) 精密而不准确；(b) 准确而不精密；(c) 既精密又准确；(d) 既不精密又不准确。

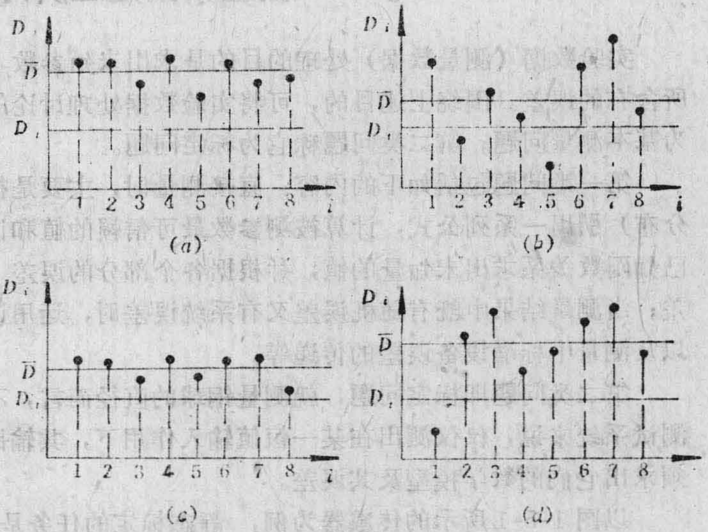

图1-3-1 重复测量时的散点图

(a) 精密而不准确；(b) 准确而不精密；

(c) 既精密又准确；(d) 既不精密又不准确。

测量中，下述情况要区分清楚：如果被测对象是稳定的，则重复测量所得到的一组数据主要反映测量仪器（或说测量条件）的情况；如数据的重复性好则说明测量条件稳定。如果测量对象不稳定而测量条件是稳定的，则通过测量得到的一组数据主要反映被测对象的情况。如果被测对象和测量条件都不稳定，则测试数据就是两者变化的综合反映。

§1-4 不确定度与置信概率

在§1-2节中曾讲到，由于眞值 A_0 不可知，因此无法根据式（1-2-1）将 Δ_x 计算出来。然而，常常可以由各种依据估计 Δ_x 的绝对值的一个上界 U，即

$$U=sup|\Delta_x| \qquad (1\text{-}4\text{-}1)$$

或

$$|\Delta_x|=|x-A_0|\leqslant U \qquad (1\text{-}4\text{-}2)$$

这个上界值 U 通常称为不确定度（有时亦称为置信限），也就是估计出来的一个误差限（极限误差）。所谓估计，是指可能值或大概值。这就涉及到这样一个问题：这种估计有多大的把握，可信程度如何？而可信程度一般用概率给出，这种概率称为置信概率。

显然，对于同一个测量结果，如果估计一较小的U值，则$|\Delta_x|$实际上不小于U的可能性就较大，我们将冒着估计不足的风险。反之，若估计一个较大的U，则$|\Delta_x|$不大于U这件事就较为可信。如果说$|\Delta_x|<\infty$，这件事就100%可信。

由此可见，置信限愈宽，则置信概率愈大。然而，如果置信限给得很大，则测量的结果就变得沒有意义了。能够容许的不确定度，取决于测量的目的及测量结果的用途，不能一概而论。置信概率取多大値才算合适，也同样取决于具体情况，按一般习惯，置信概率可取68%，90%，95%，99%，99.5%，99.73%等。其取值大小，要根据具体要求及该项测量的重要性而定。

§1-5　实验数据处理所讨论的问题

实验数据（测量数据）处理的目的是求出未知参数（被测的量）的数值和评定这一数值所含有的误差。围绕上述目的，可将实验数据处理讨论的问题分为两大类：第一类问题称它为基本测量问题；第二类问题称它为标定问题。

第一类问题包括如下的內容：直接测量时，主要是按古典误差理论（随机误差服从正态分布）引出一系列公式，计算被测参数最可信赖的值和该值所含的误差；间接测量时，根据已知函数关系求出未知量的値，并根据各个部分的误差（称为误差分量）求出间接测量的误差；当测量结果中既有随机误差又有系统误差时，运用误差合成的方法求出综合误差指标；以及测量中标准设备误差的传递等。

第二类问题即标定问题，就测量钢球的直径而言，不存在标定问题。对仪表、传感器及测试系统来说，仅仅测出在某一恒值输入作用下，其输出值的最可信赖的值是不够的，还必须求出它们的数学模型及其误差。

以图1-5-1所示的传感器为例，静态标定的任务是根据对应不同输入x_i所获得的输出测量值y_i，求出传感器的数学模型，即静特性$y=f(x)$以及用$y=f(x)$来表征传感器静特性时的性能指标。动态标定的任务，是根据动态测得的数据求出传感器的动态数学模型：传递函数$W(s)=Y(s)/X(s)$。动态标定问题，与目前广泛研究的系统辨识关系极为密切，辨识问题涉及的理论较多，因此本课不予详述。

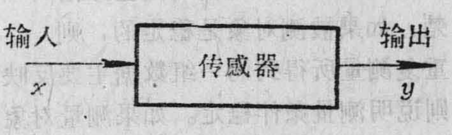

图1-5-1　线性系统

第二章　有效数字与计算法则

§2-1 有 效 数 字

在测量和数字计算中，应该用几位有效数字来代表被测量或计算的结果，是一件很重要的事情。若认为在一个数值中小数点后面的位数愈多，这个数值就愈准确；或者在计算结果中，保留的位数愈多，这个数愈准确。这两种想法都是错误的。第一种想法的错误，在于沒有弄清楚小数点的位置不是决定准确与否的标准，而仅与所用单位的大小有关。例如，记长度为21.3毫米与0.0213米，其准确程度完全相同。第二种想法的错误，在于不了解所有的测量，由于仪器和人们的感官只能作到一定的准确程度。这个准确程度一方面决定于所用仪器刻度的精细程度；另一方面也与所用的方法有关。因此，在计算结果中，无论写多少位数，绝不可能把准确程度增加到超过测量所允许的范围。反之，表示一个数字时，如果书写的位数过少，以至于低于测量所达到的准确程度，同样是错误的。正确的写法应该是在我们写出的位数中，除末位数字可疑或不确定外，其余各位数字都是准确知道的。除有特殊规定，一般认为末位数字上下可能有一个单位的误差，或其下一位的误差不超过±5。

在科学实验中有两类数：一类数其有效位数均可认为无限制，即它们的每一位数都是确定的，例如各种计算式中的 $\sqrt{2}$、$\frac{1}{2}$、π 及自然数等。另一类数是用来表示测量结果的，这一类数的末一位往往由估计得来，因此具有一定的误差或不确定性。在正常测量时，一般只能估计到测量仪器的最小刻度的十分之一，故在记录测量结果时，只允许末一位是由估计得到的不准确的数字，其余数字均为准确数字，我们称此时所记的数字为有效数字。

例如游标卡尺的最小刻度为絲米（十分之一毫米），用此卡尺测得某圆柱体直径为32.47毫米。该数的第一位是厘米位，第二位是毫米位，第三位是絲米位。由于卡尺的最小刻度为絲米位，故前三位数都是准确的数字。该数的第四位为7，它是估计的。因为，换一个人进行测量，测得的直径值可能为32.46毫米或者32.48毫米。因此第四位上可能有上下一个单位的出入，故有效数字为四位。如果将32.47写为32.470，从有效数字来看这样写法是错误的，因为32.470表示五位有效数字。

关于数字“0”，需要特别提一下，它可以是有效数字，也可以不是有效数字。例如在30.05克中以及1.02010克中，所有的0都是有效数字。而在长度为0.00320中，前三个0均非有效数字，因为这些0只与所取的单位有关，而与测量的准确与否无关。又如12000米，我们很难区别0是有效数字还是非有效数字。如果按12000米，则为五位有效数字，故所有的0都是有效数；如果将12000米表示为12×10^3米，或12千米，则有效数只有二位，故所有的0都不是有效数。因此，对于数字后面的0要特别的注意。一般约定，末位数的0指的是有效数字，故1.230×10^4厘米不能书写为1.23×10^4厘米；同样，32.47毫米不能书写为32.470毫米。

§2-2 计 算 法 则

常用的基本法则有以下几种：

一、记录测量数值时，要首先确定有效位数，只允许末一位数字为可疑数字。可疑数字即表示本位数可有±1个单位的误差，或者下一位可有±5个单位的误差。

二、当有效位数确定后，例如为N，则$N+1$位后的数字舍入法则如下：

1. 如第$N+1$位以下的数不到第N位的一个单位的一半（即$N+1$位的数字小于5），则舍去。

例3.2249，要求三位有效数，则为3.22。

2. 如第$N+1$位以下的数超过第N位的一个单位的一半，则第N位增加一单位。

例7.36与7.3501，各要求二位有效数，则皆为7.4。

又例7.3967，要求三位有效数，则为7.40。

3. 如第$N+1$位以下的数正好是第N位的一个单位的一半，则舍入法则如下：

(1) 第N位数如为0，2，4，6，8，则第$N+1$位舍去。例如，

12.305，要求四位有效数，则为12.30；

12.450，要求三位有效数，则为12.4；

0.165，要求小数点后两位有效数，则为0.16。

(2) 第N位数如为1，3，5，7，9，则第N位增加一个单位。例如，

0.0935，要求二位有效数，则为0.094；

12.550，要求三位有效数，则为12.6；

2.55，要求二位有效数，则为2.6。

2.095，要求小数点后二位有效，则为2.10。

4. 如果第$N+1$位以下的数本身就是通过舍入来的，则根据舍入前的数据应用上面的法则。

例4.35是从4.346舍入得到，要求两位有效数，此时应根据4.346来处理，得4.3。

又例4.35是从4.353舍入得到的，要求两位有效数，此时应根据4.353处理，得4.4。

三、加减计算法则

法则A：在加减计算中，各数所保留的小数点后的位数，应与所给各数中、其小数点后位数最少的相同。

例如将13.65，0.0082，1.633三个数相加时，应写为

$$13.65+0.01+1.63=15.29 \tag{2-2-1}$$

法则B：在加减运算中，以小数点后位数最少的为准，其余各数均按舍入法则使其小数点后的位数多一位。仍以上面的数据为例，13.65小数点后两位，其余为四位和三位，故有

$$13.65+0.008+1.633=15.291 \tag{2-2-2}$$

式（2-2-1）计算结果为15.29，小数点后第二位的9是可疑数字。式(2-2-2)计算的结果为15.291，其小数点后的第二位的9及第三位的1皆是可疑数字。

法则A符合有效数中只允许末一位为可疑数字的规定。因此，一般情况用法则A。当加减运算的项数较多，且小数点后位数较少的项数相对较少时，可以用法则B。这样可使运

算的精度稍高一些。

例如60.4，2.02，0.222，0.0467四个数相加，其中三个数在其小数点后有两位或两以上，因此用法则B较适宜。按法则B算得结果为

$$60.4+2.02+0.22+0.05=62.69$$

位若按法则A计算，则算得结果为

$$60.4+2.0+0.2+0.0=62.6$$

四、乘除计算法则

法则A：在乘除运算中，各因子保留的位数以有效数字位数最少的为标准而与小数点的位数无关。

例如进行（603.21×0.32）÷4.011运算时，各数中以0.32的有效数字位数最少（为二位），因此应将运算式表示为

$$[(60\times 10^{1})\times 0.32]\div 4.0=48$$

法则B：在乘除运算中，以有效数字位数最少的为准，其余各数按舍入法则保留的有效数字位可多一位。按此法则，上面所举之例的运算为

$$(603\times 0.32)\div 4.01=48.1$$

五、在对数计算中，所取对数的位数应与真数有效数字位相等，故在查对数表时，真数为几位则应查几位对数表。

六、计算平均値时，若为四个数或超过四个数相平均，则平均値的有效数字位可增加一位。

七、在所有的计算式中，π、e的数以及乘子，如$\sqrt{2}$、$\frac{1}{2}$等的有效数字的位数，可认为无限制，即在运算中，需要几位就可以写几位。

八、在表示精密度时，在大多数情况下，只取一位有效数，最多取两位有效数。

第三章 误差理论基础

§3-1 随机误差的理论分布规律

占典误差理论认为随机误差服从正态分布，这个结论对于绝大多数的测量误差来说是正确的。下面以测量某圆柱体直径为例，来说明随机误差服从正态分布。

测量某圆柱体直径时，假定其系统误差小到可以忽略不计，重复测量50次 ($N=50$)。每次测得的直径为 x_i，见表3-1-1中的第二列。

假定测量仪器的最小刻度为1毫米，故在测量时能判读的最小值是1/10毫米，即0.01厘米。由于只能分辨0.01厘米的差别，故在50次重复测量中，某一数值就有可能多次出现。通常把测量值按区间分组，设子区间的取值为Δx，Δx的最小值是测量仪器最小刻度的1/10 这里取 $\Delta x=0.01$ 厘米。为了避免某具体测量值正好是子区间的边界值而不便确定它落入哪一个子区间，故把最小测量值减去$\frac{1}{2}\Delta x$ 作为第一个子区间的左端点，把最大测量值加上$\frac{1}{2}\Delta x$ 作为最后一个子区间的右端点。这样一来，其测量值正好处于所属子区间的中心。子区间的划分见表3-1-1第三列。

表3-1-1第四列表示测量值出现在某子区间中的次数，用n_i表示，n_i称为频数。

表3-1-1第五列是n_i与总重复测量次数N的比，即n_i/N，称为相对频数或者频率。

表3-1-1第六列是$n_i/(N\Delta x)$，称为频率密度。由此可见，频率=频率密度×Δx。

表3-1-1 $N=50$， $\Delta x=0.01$

序号 i	观测值 x_i（厘米）	子区间端点值 左端点	子区间端点值 右端点	频数 n_i	相对频数 n_i/N	频率密度 $n_i/(N\Delta x)$
1	6.31	6.305	6.315	1	0.02	2
2	6.32	6.315	6.325	1	0.02	2
3	6.33	6.325	6.335	3	0.06	6
4	6.34	6.335	6.345	6	0.12	12
5	6.35	6.345	6.355	8	0.16	16
6	6.36	6.355	6.365	11	0.22	22
7	6.37	6.365	6.375	10	0.20	20
8	6.38	6.375	6.385	6	0.12	12
9	6.39	6.385	6.395	2	0.04	4
10	6.40	6.395	6.405	1	0.02	2
11	6.41	6.405	6.415	1	0.02	2

下面根据表 3-1-1 的数据，给出以频率密度 $n_i/(N\Delta x)$ 为纵坐标，以测量值 x 为横坐标的曲线，见图 3-1-1 中的折线。当改进测量技术（如量具的最小刻度更为精细，以使 x 的有效位数更多和 Δx 小），并在其同时增加测量次数 N，则可使图 3-1-1 中的折线图形向光滑曲线 $f(x)$ 靠近。图中的光滑曲线是按统计理论计算出来的概率密度曲线，它是折线图形的极限。

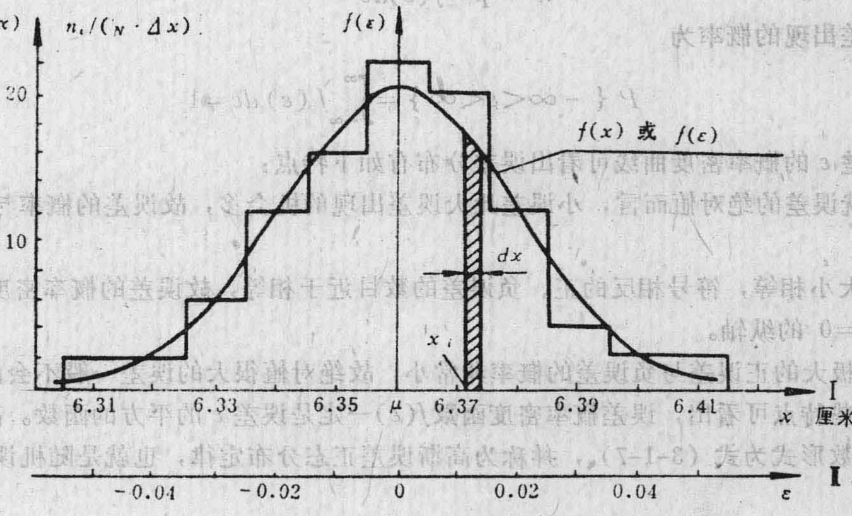

图3-1-1 频率密度与概率密度

下面对测量值（亦称观测值）的分布进行研究，所谓分布就是指某随机变量的概率密度函数或概率积分（简称概率）。

已知某测量值x出现在 $x_i-\dfrac{\Delta x}{2}$ 与 $x_i+\dfrac{\Delta x}{2}$（或者x_i与$x_i+\Delta x$）之间的频率为

$$\frac{n_i}{N}=\frac{n_i}{N\Delta x}\Delta x \tag{3-1-1}$$

当 $\Delta x\to 0$（即子区间取得很小）、$N\to\infty$（即测量次数很大）时，可以预见频率密度趋于概率密度，频率趋于概率，即

$$\frac{n_i}{N\Delta x}\approx f(x) \tag{3-1-2}$$

$$\frac{n_i}{N}\approx f(x)\Delta x=f(x)dx=p \tag{3-1-3}$$

可见图 3-1-1 中的 $f(x)$ 曲线是测量值 x 的概率密度曲线，曲线下的阴影部分所示的面积为某测量值 x 出现在 x_i与x_i+dx 之间的概率：$p=f(x)dx$。因此，若设在 $N(N\to\infty)$ 次测量中，测出的任一测量值都算数。则所有测量值出现的概率为

$$P\{-\infty<x<\infty\}=\int_{-\infty}^{\infty}f(x)dx=1 \tag{3-1-4}$$

该积分对应着图 3-1-1 光滑曲线下的全部曲边梯形的面积。

以上是根据测量值讨论的，若用 ε 表示随机误差，则有

$$\varepsilon=x-\mu$$