

B

普通高等教育机电类规划教材

检测与转换技术

第3版

吉林工业大学 常健生 主编

吉林工业大学 石要武 副主编

北京印刷学院 常 瑞

4-43

机械工业出版社
China Machine Press

本书是根据 1998 年全国高等学校电气工程与自动化专业教学指导委员会修订的“测试技术”课程汇总大纲编写的。涉及的内容比较广泛,阐述了测试技术的理论基础,电工仪表及电磁测试技术,传感器及非电量测试技术,现代测试技术。每章后均附有思考题与习题。

本书为电气工程与自动化专业的技术基础课教材,也可供有关专业师生、从事测试工作的工程技术人员参考。

图书在版编目 (CIP) 数据

检测与转换技术/常健生主编. —3 版. —北京: 机械工业出版社, 2000. 2

普通高等教育机电类规划教材

ISBN 7-111-03106-7

I. 检... I. 常... II. ①自动检测-高等学校-教材②转换单元-高等学校-教材 IV. TP274

中国版本图书馆 CIP 数据核字 (2000) 第 53574 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

责任编辑: 曹俊玲 版式设计: 张世琴 责任校对: 李秋荣

封面设计: 李雨桥 责任印制: 路琳

北京机工印刷厂印刷·新华书店北京发行所发行

2001 年 1 月第 3 版·第 1 次印刷

1000mm×1400mm B5·8.75 印张·336 千字

230 451—240 450 册

定价: 20.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换
本社购书热线电话 (010) 68993821、68326677-2527

前 言

本书系根据1998年全国高等学校电气工程与自动化专业教学指导委员会制订的“测试技术”课程汇总大纲编写的。

本书为电气工程与自动化专业学生技术基础课教材,也可供有关专业师生、从事测试工程工作的工程技术人员参考。

全书分前、后两篇共十一章:前篇中第一章是测试技术的理论基础;第二章、第三章是电工仪表及电磁测试技术;后篇中第四章至第八章是传感器及非电量测试技术;第九章至第十一章是现代测试技术。每章后均附有思考题与习题。

全书由吉林工业大学信息科学与工程学院常健生任主编,吉林工业大学石要武、北京印刷学院常瑞任副主编。参加本书编写的有:石要武(第一章、第三章第四节第六部分)、常健生(第二章、第三章)、常瑞(第四章至第八章)、吉林工业大学翟翌立(第九章至第十一章)。全书由上海大学自动化学院郎文鹏教授主审。

本书部分内容参考了兄弟院校有关自动检测技术、电磁计量技术等方面的教材,编者在此致以谢意。

由于编写时间仓促,书中难免有不妥和错误之处,殷切希望各院校师生及广大读者提出宝贵意见。

编者

1999年12月

绪 论

一、本课程研究的内容及性质

检测与转换技术是自动检测技术和自动转换技术的总称，是信息技术的重要组成部分。它是以研究自动检测系统中的信息提取、信息转换以及信息处理的理论和技术为主要内容的一门应用技术学科。信息提取即从自然界、社会、生产过程中或科学实验中获取人们需要的信息。信息处理是测试的真正目的，是指人们把已经获得到的信息进行加工、运算、分析或综合，以便进行预报、报警、检测、计量、保护、控制、调度和管理等，以期达到预防自然灾害、防止事故发生、提高劳动生产率、正确计量、改善产品质量、顺利进行科学实验、进行文明生产和科学管理等目的。

虽然检测与转换技术服务的领域非常广泛，但是从这门课程的研究内容来看，不外乎是电工仪表及传感器技术、误差理论及数据处理、自动测试系统组成技术等。

综上所述，检测与转换技术是既一门综合性很强，又是在物理学、电工学、电子学、计算机、自动控制等先修课的基础上开设的一门重要技术基础课。

二、本门课程在国民经济中的地位和作用

人类步入信息社会的今天，人们对信息的提取、处理、传输以及综合等要求愈加迫切。作为信息提取的功能器件——仪表及传感器与人类的关系愈来愈密切。例如，用电压表对电压进行监视，用一氧化碳气敏传感器对煤气溢出进行监视等等。

采用自动测试系统实时测量及分析产品性能，采用自动控制系统对产品加工过程实时控制，是提高产品质量的现代方法。

目前，检测与转换技术已成为一些发达国家最重要的热门技术之一，其主要原因是它可以给人们带来巨大的经济效益和社会效益。可以说，一个国家的现代化水平是用自动化水平来衡量的，而自动化水平是用仪表及传感器的种类和数量多少来衡量的。由此看来，检测与转换技术在国民经济中占有极其重要的地位和作用。

三、检测与转换技术的发展方向

(1) 研制新型电工仪表和传感器以及组成新型测试系统。

(2) 提高测试系统的检测分辨率、精度、稳定性和可靠性，一直是本门技术的研究课题。

(3) 微电子技术、微型计算机技术与仪表和传感器相结合可以构成新一代的智能化测试系统，可使测量精度、自动化和多功能水平进一步提高。

(4) 采用多种及多个仪表或传感器去探索测试某一物理量的线的、面的和体的空间参数及综合参数，以构成特殊的自动测试系统。例如，温度场的自动测试；复杂机械零件的形状和位置误差的自动测试；粮食、烟草、酒类以及药材等产品质量综合参数的自动测试等等。

四、本课程的任务和目的

本课程的任务在于使学生掌握检测与转换技术的基本概念、基本理论和常用电工仪表及传感器的工作原理、结构、应用及现代测试技术的一般概念。

本课程的目的在于培养学生具有以下能力：

- (1) 具有选择电工仪表及传感器的能力。
- (2) 具有组建一般测试系统的能力。
- (3) 对一般测试系统中的技术问题具有一定的分析和处理能力。

由于本课程具有综合性的技术基础课性质，在考虑先修课的同时，应尽量围绕上述任务和目的进行讲授。也希望学生在学习时，注意自己上述能力的培养。

前篇 电工仪器与仪表

第一章 检测与转换技术的理论基础

第一节 检测与转换技术的基本概念

检测与转换技术是自动检测技术和自动转换技术的总称，它是以研究自动检测系统中的信息提取、信息转换以及信息处理的理论和技术为主要内容的一门应用技术学科。

信息提取是指用组成的测试系统，从自然界诸多的被检查与测量量（物理量、化学量、生物量与社会量）中提取出有用的信息（一般都是电信号）。

信息转换是将所提取的有用信息，根据下一单元需要，在幅值、功率及精度等方面进行处理和转换。

信息处理的任务，视输出环节的需要，将变换后的电信号进行数字运算、A/D变换等处理。

信息传输的任务是，在排除干扰的情况下经济地、准确无误地把信息进行传递。

第二节 测量误差的概念和分类

一、有关测量技术中的部分名词

(1) 等精度测量。在同一条件下所进行的一系列重复测量称为等精度测量。

(2) 非等精度测量。在多次测量中，如对测量结果精确度有影响的一切条件不能完全维持不变称为非等精度测量。

(3) 真值。被测量本身所具有的真正值称之为真值。量的真值是一个理想的概念，一般是不知道的。但在某些特定情况下，真值又是可知的，如一个整圆周角为 360° 等。

(4) 实际值。误差理论指出，在排除系统误差的前提下，对于精密测量，当测量次数无限多时，测量结果的算术平均值极接近于真值，因而可将它视为被测量的真值。但是测量次数是有限的，故按有限测量次数得到的算术平均值只是统计平均值的近似值。而且由于系统误差不可能完全被排除，故通常只能把精度更高

一级的标准器具所测得的值作为“真值”。为了强调它并非是真正的“真值”，故把它称为实际值。

(5) 标称值。测量器具上所标出来的数值。

(6) 示值。由测量器具读数装置所指示出来的被测量的数值。

(7) 测量误差。用器具进行测量时，所测量出来的数值与被测量的实际值之间的差值。

任何测试系统的测量结果都有一定的误差，即所谓精度。一般来说，不存在没有误差的测量结果，也不存在没有精度要求的测试系统。精度（误差）是一项重要的技术指标。

二、误差的分类

1. 按表示方法分类

(1) 绝对误差。绝对误差是示值与被测量真值之间的差值。设被测量的真值为 A_0 ，器具的标称值或示值为 x ，则绝对误差 Δx 为：

$$\Delta x = x - A_0 \quad (1-1)$$

由于一般无法求得真值 A_0 ，在实际应用时常用精度高一级的标准器具的示值（作为实际值） A 代替真值 A_0 。必须指出， A 并不等于 A_0 ，一般来说 A 总比 x 更接近于 A_0 。

x 与 A 之差常称为器具的示值误差。记为：

$$\Delta x = x - A \quad (1-2)$$

通常以此值来代表绝对误差。

绝对误差一般只适用于标准器具的校准。

与绝对值 Δx 相等，但符号相反的值，称为修正值，常用 C 表示，如：

$$C = -\Delta x = A - x \quad (1-3)$$

通过检定，可以由上一级标准给出测试系统的修正值。利用修正值便可求出测试系统的实际值：

$$A = x - C \quad (1-4)$$

修正值给出的方式不一定是具体的数值，也可以是一条曲线、公式或数表。在某些测试系统中，为了提高测量精度，修正值预先编制成有关程序贮存于仪器中，所得测量结果自动对误差进行修正。

(2) 相对误差。相对误差是绝对误差 Δx 与被测量的约定值之比，它较绝对误差更能确切地说明测量质量。

在实际中，相对误差有下列表示形式：

1) 实际相对误差。实际相对误差 γ_A 是用绝对误差 Δx 与被测量的实际值 A 的百分比值来表示的相对误差。记为：

$$\gamma_A = \Delta x / A \times 100\% \quad (1-5)$$

2) 示值相对误差。示值相对误差 γ_x 是用绝对值误差 Δx 与器具的示值 x 的百分比值来表示的相对误差。记为：

$$\gamma_x = \Delta x/x \times 100\% \quad (1-6)$$

3) 满度(引用)相对误差。满度相对误差 γ_m 又称满度误差, 是用绝对误差 Δx 与器具的满度值 x_m 的百分比值来表示的相对误差。记为：

$$\gamma_m = \Delta x/x_m \times 100\% \quad (1-7)$$

这是应用最多的表示方法。

(3) 容许误差。容许误差是根据技术条件的要求, 规定某一类器具误差不应超过的最大范围。

2. 按误差出现的规律分类

(1) 系统误差(系差)。其变化规律服从某种已知函数。系统误差主要由以下几方面因素引起: 材料、零部件及工艺缺陷; 环境温度、湿度、压力的变化以及其它外界干扰等。

系统误差表明了一个测量结果偏离真值或实际值的程度。系统误差越小, 测量就越正确, 所以还经常用正确度一词来表征系统误差的大小。

(2) 随机误差(随差)。又称偶然误差, 其变化规律未知。随机误差是由很多复杂因素的微小变化的总和所引起的, 因此分析比较困难。但是, 随机误差具有随机变量的一切特点, 在一定条件下服从统计规律。因此, 通过多次测量后, 对其总和可以用统计规律来描述, 则可从理论上估计对测量结果的影响。

随机误差表现了测量结果的分散性。在误差理论中, 常用精密度一词来表征随机误差的大小。随机误差愈小, 精密度愈高。如果一测量结果的随机误差和系统误差均很小, 则表明测量既精密又正确, 简称精确。

(3) 粗大误差。粗大误差简称粗差, 是指在一定条件下测量结果显著地偏离其实际值所对应的误差。

在测量及数据处理中, 如发现某次测量结果所对应的误差特别大或小时, 应认真判断该误差是否属于粗大误差, 如属粗差, 该值应舍去不用。

3. 按误差来源分类

(1) 工具误差。工具误差是指测量工具本身不完善引起的误差。主要包括:

1) 读数误差。由以下几种原因产生: ① 校准误差。该误差通常是指, 测试系统在定标时, 用标准器具对其指定的某些定标点进行定标时所产生的误差。② 测试系统分辨率不高引起的误差。

2) 内部噪声引起的误差。内部噪声包括各种电子器件产生的热噪声、散粒噪声、电流噪声, 以及因开关或插接件接触不良、继电器动作、电动机转动、电源不稳等引起的噪声。

此外, 还有器件老化引起的误差; 测试系统工作条件变化引起的误差等。

(2) 方法误差。方法误差是指测量时方法不完善、所依据的理论不严密以及对被测量定义不明确等诸因素所产生的误差，有时也称为理论误差。

4. 按照被测量随时间变化的速度分类

(1) 静态误差。静态误差是指在测量过程中，被测量随时间变化很缓慢或基本不变时的测量误差。

(2) 动态误差。动态误差是指在被测量随时间变化很快的过程中，测量所产生的附加误差。动态误差是由于有惯性、有纯滞后，因而不能让输入信号的所有成分全部通过；或者输入信号中不同频率成分通过时受到不同程度衰减时引起的。该误差是在动态测量时产生的。

5. 按使用条件分类

(1) 基本误差。基本误差是指测试系统在规定的标准条件下使用时所产生的误差。所谓标准条件，一般是测试系统在实验室标定刻度时所保持的工作条件，如电源电压 $(220 \pm 5\%)$ V，温度 (20 ± 5) °C，湿度小于 80%，电源频率 50Hz 等。

基本误差是测试系统在额定条件下工作时所具有的误差，测试系统的精确度是由基本误差决定的。

(2) 附加误差。当使用条件偏离规定的标准条件时，除基本误差外还会产生附加误差。例如，由于温度超过标准引起的温度附加误差以及使用电压不标准而引起的电源附加误差等。这些附加误差使用时叠加到基本误差上去。

6. 按误差与被测量的关系分类

(1) 定值误差。指误差对被测量来说是一个定值，不随被测量变化。这类误差可以是系统误差，如直流测量回路中存在热电势等，也可以是随机误差，如测试系统中执行电机的起动引起的电压误差等。

(2) 累积误差。在整个测试系统量程内，误差值 Δx 与被测量 x 成比例地变化，即：

$$\Delta x = \gamma_s x \quad (1-8)$$

式中， γ_s 为比例常数。

由上式可见， Δx 随 x 的增大而逐步累积，故称累积误差。

第三节 随机误差概率密度的正态分布

一、随机误差的实验结果——频率直方图

现在我们来研究一组无系统误差且无粗差的独立的等精度实验结果。所谓独立和等精度测量，是指在相同条件下，对某量重复进行的独立测量。设总的测量次数 $n=150$ 次。现将 150 个测量值 (x_i) 由小到大排列分成 11 个区间，或按误差大小排列，并取等间隔值，详见表 1-1。表中， $\delta_i = x_i - x_0$ ($x_0 = 5.26$)； n_i 表示测

量值在 $(x_i \pm \Delta x_i/2)$ 范围内, 或误差在 $(\delta_i \pm \Delta \delta_i/2)$ 范围内出现的次数; 等间隔值 $\Delta x_i = \Delta \delta_i = 0.01$ 。今在直角坐标图上以频率 (n_i/n) 为纵坐标, 以随机误差 (δ_i) 为横坐标画出它们的关系曲线, 得到频率直方图, 或称统计直方图, 如图 1-1 所示。

表 1-1 随机误差实验结果

分区号	测量值 (x_i)	误差值 (δ_i)	出现次数 (n_i)	频率 (n_i/n)	概率密度 ($n_i/(n\Delta\delta_i)$)
1	5.21	-0.05	1	0.007	0.7
2	5.22	-0.04	3	0.020	2.0
3	5.23	-0.03	8	0.058	5.8
4	5.24	-0.02	18	0.120	12.0
5	5.25	-0.01	28	0.187	18.7
6	5.26	0	34	0.227	22.7
7	5.27	+0.01	29	0.193	19.3
8	5.28	+0.02	17	0.113	11.3
9	5.29	+0.03	9	0.060	6.0
10	5.30	+0.04	2	0.013	1.3
11	5.31	+0.05	1	0.007	0.7

对于不同的间隔值 $\Delta \delta_i$ 或 Δx_i , 频率 n_i/n 值也不同, 间隔值越大, 频率值也越大。因此, 对同一组实验数据, 频率直方图也将不同。若取量 $(n_i/(n\Delta\delta_i))$ 作为纵坐标, 则可避免此问题。但其图形仍类似图 1-1。

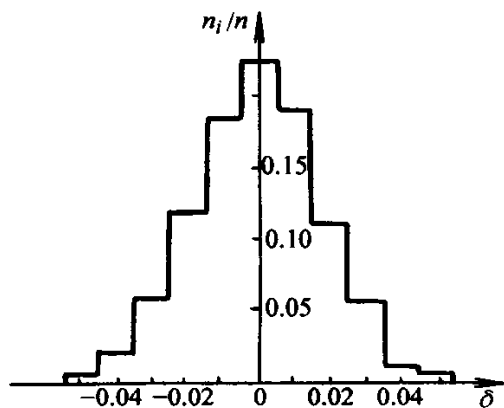


图 1-1 随机误差的频率直方图

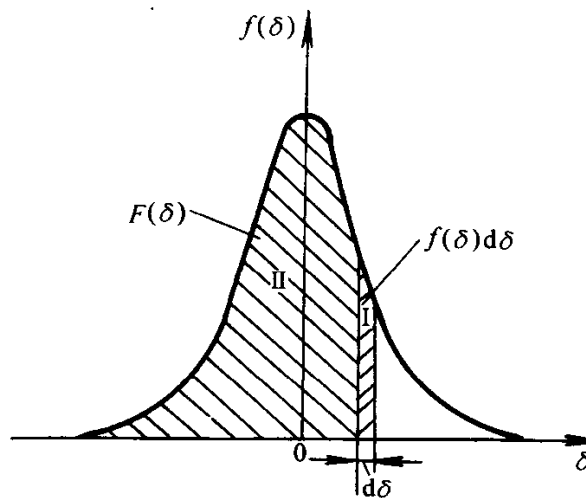


图 1-2 随机误差的概率密度分布曲线

当测量次数 $n \rightarrow \infty$ 时, 令 $\Delta \delta_i \rightarrow d\delta$, $n_i \rightarrow dn$ (均为无穷小量), 则随机误差的概率密度 $f(\delta)$ 可定义为:

$$f(\delta) = \lim_{n \rightarrow \infty} \frac{n_i}{n\Delta\delta} = \frac{1}{n} \frac{dn}{d\delta} \quad (1-9)$$

此时, 随机误差的频率直方图将变为一条光滑的概率密度分布曲线。 $f(\delta)-\delta$ 的图形如图 1-2 所示。显然, $f(\delta)d\delta = dn/n$ 就是曲线下面的阴影部分 (I) 的面积, 或称之为概率元, 用符号表示为:

$$P\{\delta, \delta + d\delta\} = f(\delta)d\delta \quad (1-10)$$

而随机变量的概率分布函数 $F(\delta)$ 定义为随机变量在区间 $(-\infty, \delta)$ 内取值的概率, 它相当于曲线下面阴影部分 (II) 的面积 (见图 1-2)。可表示为如下形式:

$$F(\delta) = P\{-\infty, \delta\} = \int_{-\infty}^{\delta} f(\delta)d\delta \quad (1-11)$$

反之有:

$$f(\delta) = \lim_{\Delta\delta \rightarrow 0} \frac{[F(\delta + \Delta\delta) - F(\delta)]}{\Delta\delta} = F'(\delta) = \frac{dF(\delta)}{d\delta} \quad (1-12)$$

由此可见, 随机变量的概率密度 $f(\delta)$ 与其分布函数 $F(\delta)$ 互为微积分关系。

根据表 1-1 的实验结果及图 1-1 和图 1-2, 可以总结出如下的统计特点:

(1) 对称性。随机误差可正可负, 但绝对值相等的正、负误差出现的次数相同, 或者是概率密度分布曲线 $f(\delta)-\delta$ 对称于纵轴。

(2) 抵偿性。相同条件下, 当测量次数 $n \rightarrow \infty$ 时, 全体误差的代数和为 0, 亦即 $\lim_{n \rightarrow \infty} \sum_{i=1}^n \delta_i = 0$, 或者说, 正误差与负误差相互抵消。

(3) 单峰性。绝对值小的误差出现的次数多, 绝对值大的误差出现的次数少。换言之, 绝对值小的误差比绝对值大的误差的概率密度大, 在 $\delta=0$ 处概率最大, 即 $f(0) = f_{\max}(\delta)$ 。

(4) 有界性。绝对值很大的误差几乎不出现, 故可认为随机误差有一定的界限。

下面进一步研究具有上述统计特性的随机变量的概率密度分布曲线的数学描述。

二、概率密度的正态分布

表 1-1 的实验结果是一组无系差、无粗差、等精度的独立测量结果 (以下不说明, 皆属此情况)。此时, 被测量的真值有恒定的数值, 每次测量也应有一恒定的测量值。然而实际情况并非如此, 就具体测量的各个不同时刻而言, 各个影响因素均有微小的、随机的及独立的变化, 但没有一个起决定作用的因素。这些大量的、微小的、独立的及随机的因素综合影响就产生了随机误差。根据概率论的中心极限定理知: 大量的、微小的及独立的随机变量的总和服从正态分布。显然, 随机误差必然服从正态分布。实践和理论都证明了这一点。严格的理论证明, 可以得到图 1-2 所示的概率密度分布曲线的数学表达式为:

$$f(\delta) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) \quad (1-13)$$

或为：
$$f(\delta) = \frac{h}{\sqrt{\pi}} \exp(-h^2 \delta^2) \quad (1-14)$$

式(1-13)或(1-14)称之为概率方程或高斯误差方程。式中, σ 是均方根误差, 或称标准误差, $h = 1/(\sigma \sqrt{2})$ 是精密度指数。 σ 和 h 是正态分布中的重要特征量, 它们的大小决定后, 概率密度 $f(\delta)$ 就是随机变量 δ 的单值函数, 概率密度分布曲线也就完全确定了。

凡是概率密度可由高斯方程描述的随机变量必然遵循正态分布, 而服从正态分布的随机变量, 其概率密度也一定可由高斯方程描述。随机误差和无系差、无粗差的测量值就是这样的随机变量, 它们的概率密度分布曲线又称之为正态分布曲线。如图 1-3 所示。

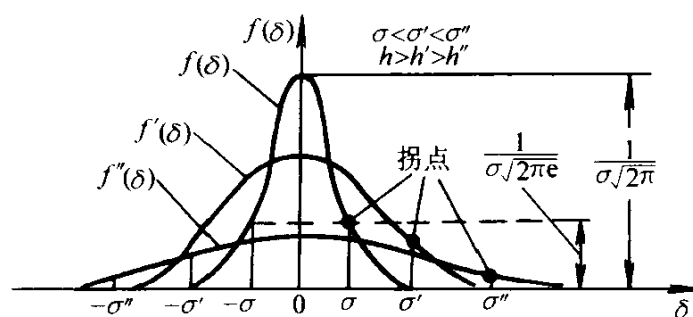


图 1-3 随机误差的正态分布曲线

随机误差的正态分布曲线除具有上述对称性、抵偿性、单峰性及有界性等特点外, 还应注意以下几点:

1) 从图 1-3 中的三条正态分布曲线比较可知: 标准误差 σ 越小, 精密度指数 h 越大, 正态分布曲线越陡, 则小误差的概率密度越大; 相对于大误差而言, 小误差出现的概率也越大, 这意味着测量值越集中, 测量精密度越高。因此, σ 的大小说明了测量值的离散性, 即测量值对于真值的离散程度。故等精度测量是一种 σ 值相同的测量。

2) 取 $df(\delta)/d\delta = 0$, 可得峰值点的坐标为

$$\delta = 0 (x_i = x_0); f(0) = f_{\max}(\delta) = \frac{1}{\sigma \sqrt{2\pi}}$$

3) 取 $d^2f(\delta)/d\delta^2 = 0$ 可得到正态分布曲线上拐点的坐标 f_g 为

$$\delta = \pm \sigma; f_g(\delta) = f(\pm \sigma) = \frac{1}{\sigma \sqrt{2\pi e}}$$

4) $P\{-\infty, +\infty\} = \int_{-\infty}^{+\infty} f(\delta) d\delta = 1$, 这说明随机误差在 $(-\infty, +\infty)$ 区间取值的概率为 1。

正态分布在误差理论中占有重要地位, 很多随机变量是服从正态分布的。尽管如此, 有些误差并不服从正态分布, 而按其它规律分布。例如, 计算中的舍入误差、数字式仪表末位的读数误差等是按均匀分布的; 圆盘偏心引起的角度误差是按反正弦分布的; 放射性元素的原子衰变则遵从泊松分布等。

第四节 算术平均值与标准误差

现若以测量值 x 作为随机变量, 如果它遵从正态分布, 那么它的概率密度 $f(x)$ 可由下式表示:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x - x_0)^2}{\sigma^2}\right] \quad (1-15)$$

其概率密度分布曲线如图 1-4 所示。式中, 被测量真值 x_0 及标准误差 σ 为测量值的正态分布中的两个重要特征量 (它们已不是随机变量了)。如果它们确定了, 则正态分布曲线可以完全确定。现在的问题是: 在已知一组测量数据 (被测量的测量值) 后, 我们如何确定 x_0 和 σ 值?

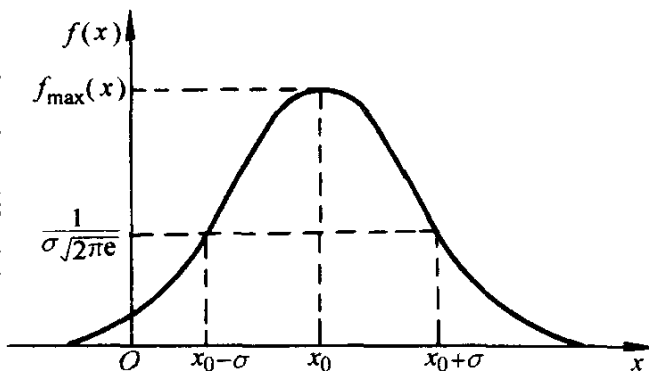


图 1-4 测量值的概率密度分布曲线

一、算术平均值 (\bar{x}) 与数学期望 $M(x)$

假设有一组等精度测量值 x_1, x_2, \dots, x_n , 根据长期的实际经验, 人们已经公认: 其算术平均值 (\bar{x}) 是这组测量数据的最好代表, 故常以 \bar{x} 表示测量结果。这就是算术平均值公理。它可由下式计算:

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_n)}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (1-16)$$

现在来分析算术平均值 \bar{x} 与真值 x_0 的关系。由概率论可知, 随机变量的数学期望定义为随机变量的一阶原点距。它表示了随机变量的中心位置。记为:

$$M_x = \int_{-\infty}^{\infty} x f(x) dx \quad (1-17)$$

如果测量值为 x , 它在微小区间 dx 内取值的概率为 $f(x) dx$, 则上式说明数学期望实际上就是全体测量值依概率的平均数。对于正态分布, 上式积分后可得:

$$M_x = \int_{-\infty}^{\infty} \frac{x}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x - x_0)^2}{\sigma^2}\right] dx = x_0 \quad (1-18)$$

因此, 全体测量值的数学期望就是测量值的真值, 它是正态分布重要特征之一。这点已被严格地证明了。

从几何意义上更容易理解数学期望的概念。在图 1-4 中, 显然可见: 正态分布曲线上概率密度 $f(x)$ 的最大值的横坐标, 或者曲线下面积重心的横坐标就是数学期望。

一般情况下, 测量值多取离散值 (离散型随机变量), 此时 $f(x)$ 不是连续的。

同时对于等精度的无限测量列来说, 测量次数 $n \rightarrow \infty$, 而每个测量值出现的频率为 $1/n$, 此时, 数学期望可表示为:

$$M_x = x_0 = \lim_{n \rightarrow \infty} \left(\frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n} \right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{x_i}{n} \quad (1-19)$$

现在我们引进母体和子样的概念。当测量次数 $n \rightarrow \infty$ 时, 得到的所有可能测量值的全体或总体, 称为母体。如上述等精度无限测量列的情况。它说明母体的数学期望就是被测量的真值。然而实际上任何实验及测量均是有限的, 测量中只能得到母体的若干测量值, 称之为子样, 有限测量列属此情况。

今假定有一等精度的有限测量列: x_1, x_2, \dots, x_n , 它为子样情况, 如测量次数为 n , 每个测量值出现的频率为 $1/n$, 那么这组有限测量值依概率的平均数就是子样的数学期望。人们常称之为算术平均值, 其表示式与式 (1-16) 经验公式完全一样。

当 $n \rightarrow \infty$ 时, 有限测量列 \rightarrow 无限测量列, 子样 \rightarrow 母体。根据误差正态分布的特性 (抵偿性), 并注意到 $\delta_i = x_i - x_0$, 则有:

$$\sum_{i=1}^n \frac{\delta_i}{n} = \left(\sum_{i=1}^n \frac{x_i}{n} \right) - x_0 = (\bar{x} - x_0) \rightarrow 0$$

即 $\bar{x} \rightarrow x_0$, 这就表明: 当 $n \rightarrow \infty$ 时, 子样的算术平均值 \bar{x} 也就趋近于被测量的真值 x_0 。因此, 子样的算术平均值 \bar{x} 就是被测量的真值 x_0 的最佳估计值 \hat{x}_0 。

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \hat{x}_0 \quad (1-20)$$

总之, 在未知 x_0 的情况下, 对于有限测量列 (子样情况), 可以利用算术平均值 \bar{x} 代替真值 x_0 , 用测量偏差或残余误差 (简称残差) $v_i = x_i - \bar{x}$ 代替测量误差 $\delta_i = x_i - x_0$ 。

二、方差与标准误差

由概率论可知, 方差定义为随机变量 (ζ) 的二阶中心距, 它更好地表征了随机变量相对于其中心位置 (数学期望) 的离散程度。对于连续型随机变量, 母体的方差可表示为

$$D_\zeta = \int_{-\infty}^{+\infty} (\zeta - M_\zeta)^2 f(\zeta) d\zeta \quad (1-21)$$

或
$$D_\zeta = \left[\int_{-\infty}^{+\infty} \zeta^2 f(\zeta) d\zeta \right] - (M_\zeta)^2$$

对于全体测量值来说, 母体的方差 D_x 表征了测量值相对于其真值 x_0 的离散程度。方差越小, 正态分布曲线越瘦高, 测量误差越小, 测量精密度越高。

因
$$D_x = \int_{-\infty}^{+\infty} (x - x_0)^2 f(x) dx = \int_{-\infty}^{+\infty} \frac{(x - x_0)^2}{\sigma \sqrt{2\pi}}$$

$$\times \exp\left[-\frac{1}{2} \frac{(x-x_0)^2}{\sigma^2}\right] dx - \sigma^2 \quad (1-22)$$

故 $\sigma = \sqrt{D_x}$

由此可见, 标准误差 σ 是方差 D_x 的均方根值, 这也是标准误差 σ 又称均方根误差的原因。

对于等精度的无限测量列, 随机变量是离散型的, 测量值取离散值, 母体均方根误差可表示为:

$$\sigma = \sqrt{D_x} = \lim_{n \rightarrow \infty} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x_0)^2} = \lim_{n \rightarrow \infty} \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_i^2} \quad (1-23)$$

对于等精度的有限测量列, 因属子样情况, 标准误差的计算方法略有不同。当可知真值 x_0 时, 标准误差的计算公式与母体情况类似, 仅 n 为有限值而已。但当不知真值 x_0 时, 必须取子样的算术平均值 \bar{x} (真值的最佳估计值 $\bar{x} = \hat{x}_0$) 来代替真值 x_0 , 取测量偏差 v_i 来代替测量误差 δ_i 。因此, 不能再用上述公式了。对于独立的、无系统误差的等精度测量来说, n 次测量值有 n 个自由度, 当求其算术平均值时已失去一个自由度, 此时, 可按如下公式计算标准误差的估计值 $\hat{\sigma}$:

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n v_i^2} \quad (1-24)$$

这个公式称之为贝塞尔公式。它是子样在未知真值情况下计算标准误差 σ 的重要公式。当测量次数 $n \rightarrow \infty$, $(n-1) \rightarrow n$, 子样 \rightarrow 母体, $\hat{\sigma} \rightarrow \sigma$ 时, 子样和母体情况下标准误差的计算公式也就趋于一致了。母体的 σ 称为标准误差, 子样的 $\hat{\sigma}$ 称为标准偏差, 但在一般情况下, 对于 $\hat{\sigma}$ 及 σ 的符号并不加以严格区分。可是对有限次测量, 特别是 n 较小时, 必须使用贝塞尔公式计算 σ 值。

下面简单介绍一下算术平均值 \bar{x} 的标准误差的估计方法。由于测量次数有限, 因此 \bar{x} 与 x_0 仍有一定误差。可以证明, 算术平均值的标准偏差 (记为 \hat{S}) 是测量值的标准偏差 ($\hat{\sigma}$) 的 $1/\sqrt{n}$ 倍。即:

$$\hat{S} = \frac{\hat{\sigma}}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n v_i^2} \quad (1-25)$$

式中, 残余误差 $v_i = x_i - \bar{x}$

在子样的情况下, 当测量次数 n 增多时, 随机误差的抵偿性显著, 其影响减弱, 故 \hat{S} 值下降, 即 \bar{x} 作为 x_0 的估计值的标准误差下降, \bar{x} 的精密度提高。但由上式可知, 当 n 增加时, \hat{S} 按 \sqrt{n} 减少的, 故 \hat{S} 下降的速度比 n 增加的速度要慢。当 $n=50$ 后, n 再增加时, \hat{S} 降低的效果不明显, 故一般取 $n=10 \sim 20$ 就可以了, 测量次数很少越过 50 次。

下面介绍另一形式的贝塞尔公式, 它在实际计算中, 特别是 n 较大时, 计算

比较方便，因而被经常使用。

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]$$

所以

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]}$$

如果 x_i 值太大，可任选一与 x_i 接近的 B ，作变换： $y_i = x_i - B$ ，因为 $y_i - \bar{y} = x_i - \bar{x} = v_i$ ，故有：

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right]} \quad (1-26)$$

这种形式的贝塞尔公式具有以下优点：在实际计算中，不会因求算术平均值除不尽而产生舍入误差，在去除坏值过程中，不需要重复计算每个 v_i 及 v_i^2 值，大大简化了计算。当 n 较大时，特别是在设计电子计算机程序时，不需要准备 n 个单元存放 x_i 值，因而能较大地节约内存单元。

第五节 置信区间与置信概率

置信区间定义为：随机变量取值的范围，用符号 $\pm l$ 或 $(-l \sim +l)$ 表示。由于标准误差 σ 是正态分布的重要特征，为此，置信区常以 σ 的倍数来表示，即 $\pm l = \pm Z\sigma$ ，式中， Z 为置信系数， $Z = l/\sigma$ 或 $l = Z\sigma$ (l 称为置信限)。

置信概率定义为：随机变量 (ζ) 在置信区间 ($\pm l$) 内取值的概率，用下列符号表示

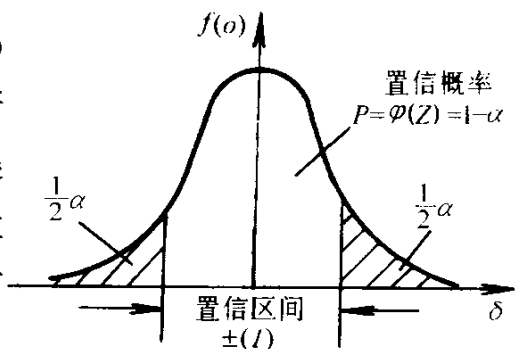
$$p(Z) = P\{|\zeta| \leq Z\sigma\} = \int_{-l}^{+l} f(\zeta) d\zeta \quad (1-27)$$

把置信区间及置信概率两者结合起来称之为置信度，即可信程度。由此可见，从统计学的角度正确说明一个测量结果，必须指明其可信度，亦即必须有置信区间及置信概率这两个指标。

置信水平表示随机变量在置信区间以外取值的概率，又称之为显著性水平，记为：

$$\alpha(Z) = 1 - p(Z) = P\{|\zeta| > Z\sigma\} \quad (1-28)$$

正态分布的置信区间与置信概率如图 1-5 所示。显然，置信区间越宽，置信概率越大，随机误差的范围也越大，对测量精度要求越低，反之，置信区间越窄，置信概率越小，误差范围也变小，对测量精度的要求变高。



下面分析一下，当置信区间取不同大小，或是 图 1-5 置信区间与置信概率

置信系数 Z 取不同典型值时, 置信概率的数值是多少, 它们有何意义。

设置信区间为: $\pm l = \pm Z\sigma$, 置信概率为:

$$\begin{aligned}\varphi(Z) &= P\{|\delta| \leq Z\sigma\} = \int_{-Z\sigma}^{+Z\sigma} f(\delta) d\delta = \int_{-Z\sigma}^{+Z\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\delta^2/2\sigma^2) d\delta \\ &= \frac{2}{\sigma\sqrt{2\pi}} \int_0^{Z\sigma} \exp(-\delta^2/2\sigma^2) d\delta\end{aligned}\quad (1-29)$$

变量置换, 令 $\delta = Z\sigma$, 则 $d\delta = \sigma dZ$, 积分限 $0 \sim Z\sigma$ 变为 $0 \sim Z$, 故有

$$\varphi(Z) = \frac{2}{\sqrt{2\pi}} \int_0^Z \exp(-Z^2/2) dZ \quad (1-30)$$

函数 $\varphi(Z)$ 又称为拉普拉斯函数, 它是置信系数 Z 的函数。表 1-2 列出了置信系数 Z 取不同的典型数值时, 置信概率或拉普拉斯函数 $\varphi(Z)$ 的数值。

应当注意 Z 取 n 个特殊典型值的情况:

表 1-2 正态分布下的置信概率数值表

Z	$\varphi(Z)$	Z	$\varphi(Z)$	Z	$\varphi(Z)$	Z	$\varphi(Z)$
0	0.00000	0.9	0.63188	1.9	0.94257	2.7	0.99307
0.1	0.07966	1.0	0.68269	1.96	0.95000	2.8	0.99489
0.2	0.15852	1.1	0.72867	2.0	0.95450	2.9	0.99627
0.3	0.23585	1.2	0.76986	2.1	0.96427	3.0	0.99730
0.4	0.31084	1.3	0.80640	2.2	0.97219	3.5	0.99535
0.5	0.38293	1.4	0.83849	2.3	0.97855	4.0	0.99937
0.6	0.45149	1.5	0.86639	2.4	0.98361	4.5	0.99993
0.6745	0.50000	1.6	0.89040	2.5	0.98758	5.0	0.99999
0.7	0.51607	1.7	0.91087	2.58	0.99012	∞	1.00000
0.8	0.57629	1.8	0.92814	2.6	0.99068		

当 $Z=1$ 时, 置信区间为 2 倍的标准误差的宽度, 即 $\pm\sigma$; 置信概率为, $\varphi(Z) = P\{|\delta| \leq \sigma\} = \int_{-\sigma}^{+\sigma} f(\delta) d\delta = 0.6827 \approx 2/3$; 置信水平为 $\alpha(Z) = P\{|\delta| > \sigma\} = 1 - \varphi(Z) = 0.3173 \approx 1/3$ 。这种情况说明, 随机误差落在 $\pm\sigma$ 区间内的机会为 68.3%, 而落在 $\pm\sigma$ 区间外的机会为 31.7%。或者从统计意义上说, 三个测量值中有二个落在 $(x \pm \sigma)$ 区间内, 有一个落在此区间之外。

当 $Z=2$ 时, 置信区间为 $\pm 2\sigma$, $\varphi(Z) = P\{|\delta| \leq 2\sigma\} = 0.9545 \approx 21/22$, $\alpha(Z) = P\{|\delta| > 2\sigma\} \approx 0.0455 \approx 1/22$ 。

当 $Z=3$ 时, 置信区间为 $\pm 3\sigma$, $\varphi(Z) = P\{|\delta| \leq 3\sigma\} \approx 0.9973 \approx 369/370$, $\alpha(Z) = P\{|\delta| > 3\sigma\} \approx 0.0027 \approx 1/370$ 。

后两种情况表明, 对于一组既无系差, 又无粗差的等精度测量来说, 当置信