

普通高等学校规划教材

船体建造测量及数据处理技术

叶家玮 编



华南理工大学出版社

出版说明

根据国务院(1978)23号文件批转试行的《关于高等学校教材编审出版若干问题的暂行规定》,我们开展了全国高等学校船舶类专业规划教材编审、出版的组织工作。

为了做好教材编审组织工作,中国船舶工业总公司相应地成立了“船舶与海洋工程”、“船舶动力”、“船电自动化”、“惯性导航及仪器”、“水声电子工程”、“流体传动与控制”、“水中兵器”七个教材委员会,聘请了有关院校的教授、专家 50 余人参加编审指导工作。船舶类专业教材委员会是有关船舶类专业教材建设研究、指导、规划和评审方面的专家组织,主要任务是协助政府机关做好高等学校船舶类专业教材的编审工作,对提高教材质量起审查把关作用。

经过前四轮教材建设,共出版教材 300 余种,建立了较完善的规章制度,扩大了出版渠道,在教材的编审依据、计划体制、出版体制等方面实行了卓有成效的改革,这些改革措施为“九五”期间船舶类专业教材建设奠定了良好基础。根据原国家教委对“九五”期间高校教材建设的要求:“抓好重点教材,全面提高质量,继续增加品种,整体优化配套,深化管理体制和运行机制的改革,加强组织领导”,船舶总公司于 1996 年组织制定了“全国高等学校船舶类专业教材(九五)选题规划”。列入规划的选题共 129 种,其中部委级重点选题 49 种,一般选题 80 种。

“九五”教材规划是在我国发展社会主义市场经济条件下第一个教材规划,为适应社会主义市场经济外部环境,“九五”船舶类专业教材建设实行指导性计划体制。即在指导性教材计划指导下,教材编审出版由主编学校负责组织实施,教材委员会进行质量审查,船舶工业教材编审室组织协调。

“九五”期间要突出抓好重点教材,全面提高教材质量,为此教材建设引入竞争机制,通过教材委员会评审、择优确定主编,实行主编负责制。教材质量审查实行主审、复审制,聘请主编学校以外的专家审稿,最后教材委员会复审,复审合格后由有关教材委员会发给编者出版推荐证书,作为出版依据。全国高校船舶类专业规划教材,就是通过严密的编审程序和高标准、严要求的审稿工作来保证教材质量。

为完成“九五”教材规划,主编学校应充分发挥主导作用。规划教材的立项由学校申报,立项后由主编学校组织实施,教材出版后由学校组织选用,学校是教材编写与教材选用的行为主体,教材计划的执行主要取决于主编学校工作情况。希望有关高校切实负起责任,各有关方面积极配合,为完成“九五”船舶类专业教材规划,为编写出版更多的精品教材而努力。

由于水平和经验局限,教材的编审出版工作和教材本身还会有很多缺点和不足,希望各有关高校、同行专家和广大读者提出宝贵意见,以便改进提高。

船舶工业教材编审室

1999 年 3 月

前 言

本书是根据原中国船舶工业总公司船舶与海洋工程教材委员会意见,并参照兄弟院校和有关资深专家学者的意见和建议编写的。

船体建造过程中几何量的测量、造船公差(即中国造船质量标准中船体建造精度标准部分)及精度管理等技术是船舶建造技术的重要组成部分。改革开放以来,工业产品质量明显提高,重点工程及出口船舶的建造加快了技术进步且提出了更高的技术要求;同时,船体及金属结构件制造时某些工艺阶段也对工件几何精度提出特殊的要求(如某些装配、焊接工艺对几何精度有更高的要求);另外,测量技术的改进及计算机的应用,促进了对误差及数据处理理论和技术的依赖。为了教学和满足工程需要,有必要将有关理论、应用技术(包括目前最新的某些研究成果)及有关标准作较详细介绍。

事实上,技术进步已经要求许多专业(尤其是工程类专业)掌握或了解误差理论及测量数据处理技术。因船体及金属结构件工程涉及测量、尺寸控制和精度标准等,制造工程技术人员理应掌握这方面的知识。这也是我们对船舶与海洋工程专业的学生开设本课程的原因。

本书内容主要包括:静态几何量测量过程中误差理论及数据处理(误差的基本性质与数据处理、误差的合成、最小二乘法等);大、中尺寸几何量测量技术(含船厂专用测量技术),误差分析与评定;船体建造精度标准,尺寸补偿及精度控制技术,有关尺寸链理论及应用等。

本教材对从事船舶与海洋工程或相关专业的教学、科研人员具有一定的参考价值,同时还可作为造船以及相关行业技术人员和管理人员再学习的参考书。

本书承武汉理工大学王勇毅教授审阅,船舶与海洋工程教材委员会委员李洁雅教授复审。在此谨致以深切谢意。

由于编者学术水平有限,恳请读者对书中欠妥和谬误之处批评指正。

编 者

2000年9月于华南理工大学

目 录

第一章 误差理论基础与数据处理	(1)
第一节 误差	(1)
第二节 随机误差	(4)
第三节 系统误差	(25)
第四节 粗大误差	(36)
第五节 误差的传递	(41)
第六节 测量结果的处理	(50)
第七节 误差的合成	(55)
第二章 大、中尺寸测量技术	(61)
第一节 大、中尺寸测量的特殊性	(61)
第二节 大、中尺寸测量技术	(62)
第三节 经纬仪三坐标测量系统	(85)
第四节 摄影测量简介	(87)
第三章 造船精度标准及尺寸精度控制	(98)
第一节 精度标准及造船几何量检测	(98)
第二节 尺寸精度的补偿	(135)
第三节 尺寸精度控制	(152)
第四章 尺寸链	(172)
第一节 基本概念	(172)
第二节 尺寸链的分析	(174)
第三节 极值法解尺寸链	(178)
第四节 统计法解尺寸链	(182)
第五节 尺寸链计算的一般公式	(185)
附录 I (CSQS)1988《中国造船质量标准》摘录—船体建造精度标准	(189)
附录 II 分段焊接变形计算图谱	(210)
附录 III 连续型随机变量的分布函数简表	(217)
参考文献	

第一章 误差理论基础与数据处理

由于测量用的仪器和量具本身精度、测量方法、测量环境等客观因素和操作人员技术水平、经验等主观因素的影响,以及人们认识能力所限等原因,测量和试验所获数值与真实值之间,总存在一定的差异,这在数值上表现为误差。

事实上,过分地追求高精度在现实中是没有意义的。研究误差是要达到以下几个目标:

- ① 认识误差的性质,分析其产生的原因,尽可能地减小误差;
- ② 正确处理测量数据,以便在一定的条件下得到更接近真实值的数据;
- ③ 合理地设计或选用量仪和测量方法,以便在最经济的条件下,得到适用的结果。

第一节 误差

一、误差的表示

1. 绝对误差

误差就是给出值与真值之差。这是在广义的意义上而言的,对于测量误差,可定义为

$$\text{测量误差} = \text{测得值} - \text{真值} \quad (1-1)$$

这实际上是一个绝对误差的概念,其中误差的正或负则表示测得值在正或负的方向上偏离了真值。真值是一个理想的概念,通常指理论真值(例如平面三角形三内角之和为 180°)、计量学约定真值(例如国际计量大会决议对长度单位“米”的长度的约定)、标准器具相对真值(高一级精度标准器具的误差与低一级标准器具或普通计量仪器的误差比值为 $\frac{1}{10} \sim \frac{1}{3}$ 时,则称前者给出值为后者的相对真值)。

在实际工作中,可通过修正值对测得值进行修正,使其更接近真值。

$$\text{修正值} = -\text{误差} = \text{真值} - \text{测得值} \quad (1-2)$$

或

$$\text{真值} = \text{测得值} + \text{修正值} = \text{测得值} - \text{误差}$$

对此式的讨论已涉及实际工作范围,由于修正值和测得值均隐含着误差,这里的“真值”只是更接近真值的值。

2. 相对误差

相对误差的定义为

$$\text{相对误差} = \frac{\text{误差}}{\text{真值}} \quad (1-3)$$

当误差较小时可由下式表示:

$$\text{相对误差} = \frac{\text{误差}}{\text{测得值}} \quad (1-4)$$

相对误差为无量纲值,通常用百分数(%)来表示。

对于相同的被测量量,可以用上述绝对误差来进行比较或评定测量精度的高低。但是对于不同的被测量量,采用绝对误差来比较就不够科学,而相对误差对两种情况却都能适应,因而具有较广泛的意义。

例如用钢卷尺分别测量 10 m 和 100 m 的距离,测得值分别为 10.001m 和 100.003m。绝对误差分别为 1mm 和 3mm,而相对误差分别为 0.01% 和 0.003% (或写为 10^{-4} 和 3×10^{-5})。显然,在这里用相对误差来比较被测量尺寸准确程度更显得确切一些。

二、误差的来源

误差的来源很复杂,具体列项分析时还须注意不要重复和遗漏,下面先作定性讨论。

1. 测量装置误差

(1)标准器误差。标准器是提供标准量值的器具,如氦 86 灯管(长度基准)、激光管、标准量块、标准线纹尺等,其自身体现的量值与真值之间一般存在着差异。

(2)仪器误差。凡是用来直接或间接将被测量量和测量单位进行比较的设备称为仪器或仪表。如何贝比长仪、激光比长仪、钢卷尺等。仪表本身都具有误差。

(3)附件误差。仪器的附件及附属工具如光学仪器瞄准用的标靶等,也会引起误差。

上述误差的具体表现形式为:标准器的结构性误差,如线纹尺分划不准、量块的平面度或平行度误差等;仪器的调整误差,如经纬仪使用时未调整到要求的水平状态;因时间或空间变化引起的各种误差,各种标准器因时间的推移和空间的变化会引起标准量值的变化,如时效会引起线纹尺的弯曲及尺长的变化,长时间使用和温度的变化会使激光管输出激光波长的稳定性下降等。另外,仪器及量具在使用中变形等也会引起一定的误差。

2. 环境误差

仪器或量具的使用环境与对其要求的环境或检定的环境往往是不一致的。环境的变化一般意味着温度、湿度、气压等条件的不同,且额外增加诸如可能的震动、气流效应、阳光照射等不利因素,上述变化可导致测量和被测量本身发生变化,这些变化的结果无疑将导致误差。后面将详细叙述环境误差中尤其是温度误差对大、中尺寸几何量测量的影响。

仪器或量具在规定条件下使用产生的示值误差称为基本误差。所谓规定条件是指检定规程中检定时所规定的工作条件,超出此规定条件使用时所增加的误差则称为附加误差。

3. 方法误差

由于采用近似的测量方法或计算方法而引起的误差,如用钢卷尺测量大尺寸轴的圆周长 s ,再经计算求出该轴的直径($d = s/\pi$)。由于 π 取值的不同、钢卷尺不易位于轴的横截面、钢卷尺厚度、所受拉力不等于检定时拉力等众多导致误差的因素的影响,此近似方法将引起较大的误差。

4. 人员误差

这是一项与测量者视觉分辨能力、反应速度(工作疲劳还将导致上述能力下降)、固有习惯诸如读数时始终偏左或偏右等有关的误差,这类误差常称为人差。

三、误差的分类

根据误差的特点和性质,可将其分为系统误差、随机误差和粗大误差。

1. 系统误差

在偏离规定测量条件下多次测量同一量时,误差的绝对值和符号保持不变;或在测量条件改变时,按某一确定的规律变化的误差。

实验或测量条件一经确定,系统误差就获得一个客观上的确定值,尽管经多次测量取平均也不能减弱其影响。改变实验条件,就可能发现系统误差随“条件”而变化的规律,这也是用物理方法发现系统误差的措施。

系统误差可按下列方法分类:

(1)按对误差认识的程度分为已定系统误差,即已知误差的大小和方向,是一种能够修正的误差;未定系统误差,即误差的大小和方向未知,但可估计出误差的范围。

(2)按误差出现的规律分为恒定系统误差,即误差的大小和方向为一固定值;可变系统误差,即误差的大小和方向是变化的,它包括线性系统误差、周期性系统误差和复杂规律系统误差。

2. 随机误差

在同一条件下,多次重复测量同一量时,绝对值和符号以不可预定方式变化着的误差。通常可以用统计方法估计出随机误差的界限 L ,称 L 为极限误差。对应 $(-L, L)$,则称之为随机误差的置信区间。由于随机误差的变化不能预先确定,因此,这类误差不能修正,而只能对其进行估计。随机误差是具有统计规律的误差。

虽经努力稳定测量过程(如等精度重复测量),但依然存在大小和符号不可预知的随机误差。通常,导致随机误差的原因都较复杂,多为引起微小变化的随机因素。

(1)基准件和测量装置的因素。如基准件实际尺寸与名义尺寸间的差距、量具的变形等。

(2)环境方面的因素。温度、湿度与气压的微小变化、现场的气流扰动等。

(3)测量人员方面的因素。瞄准和读数的不稳定等。

为了方便有关知识应用,下面对与误差分析有关的概率论作必要复习和扩充。

3. 粗大误差

明显歪曲测量结果的误差称为粗大误差。如测量时对错了标志读错了数,或实验条件未达到规定的要求而进行测量等。这里提出的明显歪曲并非是指所有的粗大误差都能直接识别,要除掉它们往往需经过必要的数据处理过程。

四、误差的转化

随机误差和系统误差在一定的条件下或观察问题的出发点发生改变时,它们可能互相转化。例如线纹尺的尺长误差,对于制造来说它是随机误差,但将线纹尺作为基准尺检定成批尺子时,则其固有的尺长误差将使得成批测量结果较真值长些或短些,这就成为系统误差。

另一方面,系统误差也有向随机误差转化的。对于经纬仪水平角刻度盘(具有 360° 刻度)的某一分度线具有一个固定的系统误差,但所有分度线的误差却有大有小,有正有负。

当用经纬仪某一固定位置一定角度时,则具有一固定的系统误差;如果多次采用经纬仪不同的水平角位置(其水平角刻度起始位置可以调整)测量该角,则误差将时大时小、时正时负而随机化了。由此测量所得平均值的误差将得以减小。这样类似的方法称为随机化技术。

在实际测量过程中,可利用上述误差的特点采用合适的方法减小误差。当测量条件稳定且系统误差已知时,就可采用修正的方法去掉系统误差。如果系统误差未知,就可考虑采用随机化技术,使部分系统误差经随机化后而抵消,以求获得较正确的测量结果。

五、精度

精度是和误差相对的概念。

(1) 精密度,表示测量结果中随机误差大小的程度。

(2) 正确度,表示测量结果中系统误差大小的程度。

(3) 准确度(或精确度),表示测量结果与约定真值的一致程度,在量值上为系统误差和随机误差的综合。

由于精密度和正确度反映误差的性质不同,因而精密度高的正确度就不一定高。但准确度高者,则精密度和正确度都高。

打靶弹着点是一个典型而又直观的误差示例。

如图 1-1 所示,图 a 的系统误差小而随机误差大,即正确度高而精密度低。图 b 的系统误差大而随机误差小,即正确度低而精密度高。图 c 的系统误差和随机误差都小,即准确度高。

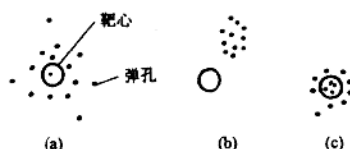


图 1-1 弹着点偏离靶心示意图

第二节 随机误差

一、正态分布

1. 正态分布特性

实际测量中的许多随机变量,它们是由许多相互独立的随机因素综合影响而形成的。其中每个单独的因素,在总的影响中所起的作用都是微小的,这种随机变量一般服从正态分布。属于正态分布的随机误差,一般具有下列特性(设总体数学期望为零):

(1) 对称性。绝对值相等(或大致相等)的正误差和负误差出现的次数大致相等。

(2) 单峰性。绝对值小的误差较绝对值大的误差出现的次数多。

(3) 有界性。对于一定测量条件下的有限个测值中,其误差的绝对值不超过某一界限。

(4) 抵偿性。在不变的测量条件下对同一量重复测量,其误差的算术平均值随着测量次数 n 的无限增加而趋于零。

2. 统计直方图

随机误差的分布规律如何确定呢?其分布规律确定的信息可取自等精度重复的测量数据。首先作统计直方图以确定误差的分布规律趋势,进而采用统计的方法进行分布规律的

检验(分布规律的检验在后面讨论)。下面以一具体例子说明统计直方图问题。

例 1 某船平行中体甲板平面分段甲板边缘直线度误差测量值共 $n = 214$ 个,误差极值为 $+12\text{mm}$ 和 -6mm ,将误差分组且以每组组中值代表该组误差的处理结果列入表 1-1 中,试绘制直方图。

表 1-1 甲板边缘直线度误差分组统计表

组号	分组区间	组中值 $x_i(\text{mm})$	频数 n_i	频率 n_i/n
1	-7.5~-4.5	-6	3	0.0140
2	-4.5~-1.5	-3	14	0.0654
3	-1.5~1.5	0	50	0.2336
4	1.5~4.5	3	66	0.3084
5	4.5~7.5	6	61	0.2850
6	7.5~10.5	9	19	0.0888
7	10.5~13.5	12	1	0.0047
			214	

解 (1) 确定误差的极差

$$R = x_{\max} - x_{\min} = 12 - (-6) = 18 \text{ mm}$$

式中 R ——极差;

$x_{\max}$ ——误差最大值;

$x_{\min}$ ——误差最小值。

(2) 按样本(抽样的全体)容量确定对误差分组的组数。查表 1-2,由 $n = 214$,取 $m = 7$ 。

(3) 由误差极差及组数确定误差分组距

$$\Delta x = \frac{R}{m} = \frac{18}{7} = 2.53 \approx 3$$

(4) 将误差原始数据按表 1-1 形式列表并确定每组误差的边界及组中值 x_i 。

(5) 按表 1-1 的数据绘制直方图,如图 1-2 所示。

为方便深入讨论,将图 1-2 绘成图 1-3 所示的形状。图中 y_i 是对应区间为单位长度时的频率,又称频率密度。误差落在小区间 $\Delta\delta_i$ 中的频率为 $f_i = y_i \Delta\delta_i$,即图中小矩形的面积。当 $n \rightarrow \infty$ 时,令 $\Delta\delta_i \rightarrow 0$,则图 1-3 中直方图的上边缘就形成一条连续曲线 $p(\delta)$ 。此曲线称为概率密度曲线。

对于正态分布概率密度,其函数形式为

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

式中 a ——总体(研究对象的全体)数学期望;

表 1-2 分组数与样本容量关系表

样本容量 n	组数 m
< 50	5~7
50~100	6~10
100~250	7~12
> 250	10~20

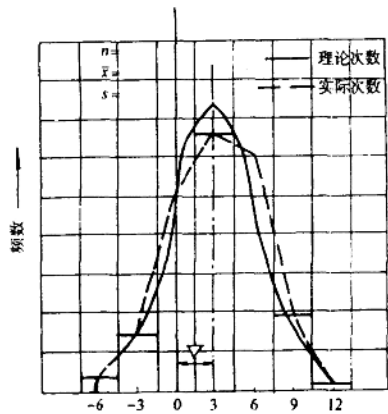


图 1-2 甲板直线度误差分布曲线

σ ——总体方差。

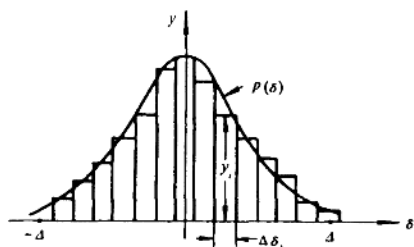
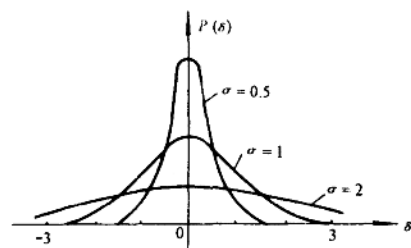


图 1-3 统计直方图



1-4 对应不同 σ 值的截尾正态分布曲线

在图 1-3 中,数学期望为零,上式可写成

$$p(\delta) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{\delta^2}{2\sigma^2}}$$

如果用界限 $\pm \Delta$ 去限定图 1-3 中的曲线范围,则可得如图 1-4 所示的截尾正态分布曲线,其概率密度为

$$p(\delta) = \begin{cases} \frac{c}{\sigma \sqrt{2\pi}} e^{-\frac{\delta^2}{2\sigma^2}}, & |\delta| \leq \Delta \\ 0 & |\delta| > \Delta \end{cases}$$

式中

$$\frac{1}{c} = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\Delta}^{\Delta} e^{-\frac{\delta^2}{2\sigma^2}} d\delta$$

3. 统计特征

(1) 数学期望及其性质

数学期望是随机变量总体的统计平均值。设测得值为 $x_i (i = 1, 2, \dots)$, 数学期望则为其算术平均值的极限

$$a = \lim_{n \rightarrow \infty} \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1-5)$$

设 x_j 为第 j 个小区间的组中值。 $y_j \Delta x_j$ 为其频数(小区间的频率), n 个测得值共分为 m 组, 对于第一组测得值设有 k 个, 则可写为

$$\frac{x_1 + x_2 + \dots + x_k}{n} = x_1 \cdot \frac{k}{n}$$

$\frac{k}{n}$ 为第一组的频数, 即 $y_1 \Delta x_1$:

$$\frac{x_1 + x_2 + \dots + x_k}{n} = x_1 y_1 \Delta x_1$$

仿此可以写出其余各组, (1-5) 式可写为

$$\begin{aligned} a &= \lim_{n \rightarrow \infty} (x_1 y_1 + x_2 y_2 + \dots + x_m y_m) \Delta x \\ &= \lim_{\Delta x \rightarrow 0} \sum x_j y_j \Delta x \\ &= \int_{-\infty}^{\infty} x p(x) dx \end{aligned}$$

数学期望的表达式最后可写为

$$E(x) = \int_{-\infty}^{+\infty} xp(x)dx$$

对于正态分布的数学期望

$$\begin{aligned} E(x) &= \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}} dt + \frac{a}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt, (t = x - a) \\ &= a, \left(\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi} \right) \end{aligned}$$

上式经代换后函数积分为 a 。即参数 a 就是正态分布的数学期望。

数学期望的性质归纳如下:

- ① $E(c) = c$, c 为常数;
- ② $E(cx) = cE(x)$, x 为随机变量;
- ③ $E(x+y) = E(x) + E(y)$, x, y 为任意两随机变量;
- ④ $E(xy) = E(x) \cdot E(y)$, x, y 为两个互相独立的随机变量。

(2) 方差及其性质

方差是描述随机变量与其数学期望偏离程度的数字特征,在数值上为无穷多个真误差 $(x_i - a)$ 平方的算术平均值。利用上面类似的讨论方法有

$$\begin{aligned} \sigma^2 &= \lim_{n \rightarrow \infty} \frac{(x_1 - a)^2 + (x_2 - a)^2 + \cdots + (x_n - a)^2}{n} \\ &= \lim_{\Delta x \rightarrow 0} \sum_{i=1}^n (x_i - a) y_i \Delta x \\ &= \int_{-\infty}^{+\infty} (x - a)^2 p(x) dx \end{aligned}$$

方差可记为 $D(x)$, a 为数学期望,即

$$\begin{aligned} D(x) &= \int_{-\infty}^{+\infty} (x - a)^2 p(x) dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} dt, (t = \frac{x-a}{\sigma}) \\ &= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t d(e^{-\frac{t^2}{2}}) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left(-te^{-\frac{t^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt \right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} (0 + \sqrt{2\pi}) = \sigma^2 \end{aligned}$$

即参数 σ^2 就是正态分布的方差。

方差的性质归纳如下:

- ① $D(c) = 0$, c 为常数;
- ② $D(cx) = c^2 D(x)$, x 为随机变量;

③ $D(x+y) = D(x) + D(y)$, x, y 为两个相互独立的随机变量。
方差有一个重要的计算公式,为

$$D(x) = E(x^2) - [E(x)]^2$$

(3) 协方差和相关系数

下面仅罗列与两随机变量有关的数字特征,以便随后引用和读者查阅。

协方差定义为 $\{[x - E(x)][y - E(y)]\}$, 记为 $\text{Cov}(x, y)$, 即

$$\text{Cov}(x, y) = \{[x - E(x)][y - E(y)]\}$$

当随机变量 x 与 y 相互独立时, $\text{Cov}(x, y)$ 为 0, 协方差有下述性质:

- ① $\text{Cov}(x, y) = \text{Cov}(y, x)$;
- ② $\text{Cov}(ax, by) = ab \cdot \text{Cov}(x, y)$, a, b 为常数;
- ③ $\text{Cov}(x_1 + x_2, y) = \text{Cov}(x_1, y) + \text{Cov}(x_2, y)$ 。

相关系数定义如下

$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

与相关系数有关内容如下:

- ① $|\rho_{xy}| \leq 1$;
- ② $|\rho_{xy}| = 1$ 的充要条件是随机变量 x 与 y 依概率 1 线性相关, 即 $P\{y = ax + b\} = 1$ (a, b 为常数)。另外, 当 x, y 互相独立时, 必有 $\rho_{xy} = 0$ 。反之, 不一定成立。

两随机变量乘积的数学期望为

$$E(xy) = E(x) \cdot E(y) = \text{Cov}(x, y)$$

两随机变量和的方差为

$$D(x \pm y) = D(x) + D(y) \pm 2\text{Cov}(x, y)$$

(4) 正态分布的概率计算

前面已经说明, 概率密度曲线所包含的面积等于随机变量在曲线相应区间发生的概率。

图 1-5 为正态分布概率密度曲线。

当误差落在 $(-\infty, \infty)$ 时, 概率 $P = 1$, 即

$$P = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = 1$$

如果我们研究误差落在 $(-\infty, x)$ 中的概率, 则上式应写为

$$\begin{aligned} P &= \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{z^2}{2}} dz, \left(z = \frac{x-a}{\sigma} \right) \\ &= \Phi\left(\frac{x-a}{\sigma}\right) \end{aligned}$$

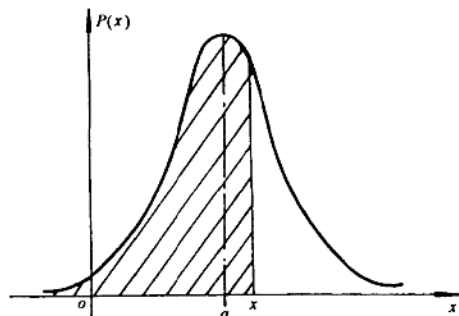


图 1-5 概率计算示意图

应用时将由 $\Phi\left(\frac{x-a}{\sigma}\right)$ 表示的概率积分表达式制成数表, 以供查用(见附录)。计算时须首先计算随机变量的数学期望与方差, 再将其标准化(即经 $z = (x-a)/\sigma$ 计算)后查 $\Phi(z)$ 表, 即可求得。另外, 经标准化的正态分布随机变量称其 $\sim N(0, 1)$ 的正态分布(对于一般正

态分布可写为 $\sim N(a, \sigma^2)$ 。

对于 $\sim N(0, 1)$ 分布的误差落在 $(-\delta, \delta)$ 间的概率, 可写为

$$\begin{aligned} P(|x| < \delta) &= P(-\delta < x < \delta) \\ &= P(x < \delta) - P(x < -\delta) \\ &= \Phi(\delta) - \Phi(-\delta) \quad (1.7) \end{aligned}$$

由图 1-6 可知

$$\Phi(\delta) = 1 - \Phi(-\delta)$$

由此, (1.7) 式可写为

$$P(|x| < \delta) = 2\Phi(\delta) - 1$$

在图 1-6 中, 若令 $Z_\alpha = \delta$, 设 $x \sim N(0, 1)$, 若 Z_α 满足条件

$$P(x > Z_\alpha) = \alpha, \quad (0 < \alpha < 1)$$

则称点 Z_α 为标准正态分布的上 100α 百分位点。

改变上述假设, 若令 $-Z_{\alpha/2} = -\delta$, $Z_{\alpha/2} = \delta$, 有

$$P(|x| > Z_{\alpha/2}) = 2P(x > Z_{\alpha/2}) = 2 \cdot \alpha/2 = \alpha$$

则称 $Z_{\alpha/2}$ 为标准正态分布的两侧 100α 百分位点。 α 称为显著性水平, 而 $1 - \alpha$ 则称为置信水平或置信度。

对应 $x \sim N(a, \sigma^2)$, 考察 $\Phi(\frac{x-a}{\sigma})$, 当 $x - a = n\sigma$ 时, 有

$$\Phi(\frac{x-a}{\sigma}) = \Phi(n)$$

即

$$P(x - a < n\sigma) = \Phi(n)$$

参照 $P(|x| < \delta) = 2\Phi(\delta) - 1$

$$\begin{aligned} P(|x - a| < n\sigma) &= \Phi(n) - \Phi(-n) \\ &= 2\Phi(n) - 1 \end{aligned}$$

当 $n = 1, 2, 3$ 时, 查表计算后有

$$P(|x - a| < \sigma) = 0.6826$$

$$P(|x - a| < 2\sigma) = 0.9544$$

$$P(|x - a| < 3\sigma) = 0.9973$$

上述数值粗略表明: 若进行 100 次等精度测量, 取误差限 $|\delta| = |x - a| = \sigma$ 时, 由上式可能有 32 次测得值的误差超出 $|\delta|$ 的范围; 当取 $|\delta| = 2\sigma$ 时, 有 4 ~ 5 次测得值的误差超出 $|\delta|$ 范围; 而当 $|\delta| = 3\sigma$ 时, 测得值的误差超出 $|\delta|$ 范围的次数可能不到一次。由于超出 $\delta = \pm 3\sigma$ 范围的概率已经很小, 故在实践中常将 $\delta = \pm 3\sigma$ 作为随机误差的极限误差。

二、算术平均值

1. 算术平均值的意义

在同一条件下对某一量值 x 进行一系列等精度、独立、无系统误差的测量, 测得值为 $x_1, x_2, \dots, x_n$, 则其算术平均值为

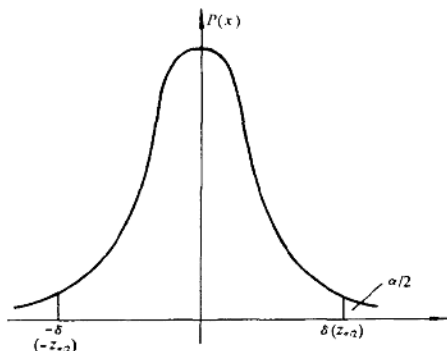


图 1-6 正态分布概率计算

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1-8)$$

设被测量量的真值为 μ , 任一次测量中的随机误差为 δ_i

$$\delta_i = x_i - \mu$$

对此式求和, 有

$$\delta_1 + \delta_2 + \cdots + \delta_n = (x_1 + x_2 + \cdots + x_n) - n\mu$$

即

$$\begin{aligned} \sum_{i=1}^n \delta_i &= \sum_{i=1}^n x_i - n\mu \\ \frac{1}{n} \sum_{i=1}^n x_i &= \frac{1}{n} \sum_{i=1}^n \delta_i + \mu \end{aligned}$$

由前述正态分布随机变量抵偿性可知, 当 $n \rightarrow \infty$ 时, 有 $\frac{1}{n} \sum_{i=1}^n \delta_i \rightarrow 0$. 因此, 可写出下式

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow \mu$$

由此说明, 若对某一量值进行无限多次等精度测量, 其算术平均值最接近真值。由于算术平均值的上述特性, 虽然实际上不可能进行无限次测量, 我们仍可以把有限次测量所获算术平均值近似地作为被测量量的真值。

2. 算术平均值的性质

由于被测量量的真值未知, 因而无法获得某一次测量的随机误差, 但是我们可以用算术平均值取代真值, 用被称为残余误差的量代替随机误差, 表达式为

$$v_i = x_i - \bar{x}, \quad (i = 1, 2, \cdots, n)$$

式中 v_i —— 残余误差(或残差)。

残余误差有下述特性:

① 残余误差的代数和为 0, 即

$$\sum_{i=1}^n v_i = 0$$

② 残余误差的平方和最小, 即

$$\sum_{i=1}^n v_i^2 \rightarrow \text{最小}$$

此为最小二乘法表达式, 由此式求得的测量值即为算术平均值。设最小二乘法测得值为 $\hat{x}$, 则

$$\frac{\partial \sum_{i=1}^n v_i^2}{\partial \hat{x}} = 0$$

$$\begin{aligned} \frac{\partial \sum_{i=1}^n v_i^2}{\partial \hat{x}} &= \sum_{i=1}^n \frac{\partial v_i^2}{\partial v_i} \cdot \frac{\partial v_i}{\partial \hat{x}} = \sum_{i=1}^n [2v_i(-1)] \\ &= -2 \left(\sum_{i=1}^n x_i - n\hat{x} \right) \end{aligned}$$

所以

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

上述性质可用来检验算术平均值和残余误差计算得是否正确。

3. 算术平均值的有关算法

(1) 直接计算算术平均值。

例2 对表1-3第一、二列数据计算其算术平均值,其中 n_j 为各组测量数据的频数。

表1-3 计算算术平均值数据列表

测得值 x_j	测量次数 n_j	$n_j \cdot x_j$	$x_j = x_j - 3.5$	$n_j x_j$
0.5	1	0.5	-3	-3
1.5	5	7.5	-2	-10
2.5	7	17.5	-1	-7
3.5	9	31.5	0	0
4.5	10	45.0	1	10
5.5	8	44.0	2	16
6.5	4	26.0	3	12
Σ	44	172.0		18

解

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_m x_m}{n_1 + n_2 + \cdots + n_m} = \frac{\sum_{j=1}^m n_j x_j}{\sum_{j=1}^m n_j}$$

$$\bar{x} = \frac{172.0}{44} = 3 \frac{10}{11} \approx 3.9$$

(2) 计算算术平均值若采用下列方法,可使运算简化且误差减小,设

$$x_j = x'_j + M$$

式中 M ——计算均值,选择其值接近算术平均值即可。由

$$n_1 x_1 + n_2 x_2 + \cdots + n_m x_m = n_1 x'_1 + n_2 x'_2 + \cdots + n_m x'_m + M(n_1 + n_2 + \cdots + n_m)$$

$$\frac{n_1 x_1 + n_2 x_2 + \cdots + n_m x_m}{n_1 + n_2 + \cdots + n_m} = \frac{n_1 x'_1 + n_2 x'_2 + \cdots + n_m x'_m}{n_1 + n_2 + \cdots + n_m} + M$$

即

$$\bar{x} = \overline{x'} + M$$

式中 $\overline{x'}$ —— x'_j 的算术平均值。

例3 对例2采用计算均值辅助计算算术平均值。

解 检查数据,观察其均值在3.5~4.5之间,暂取 $M = 3.5$ 。将 $x'_j = x_j - M$ 值列于表1-3中第4列,最后一列为 $n_j x'_j$ 之值:

$$\begin{aligned} \bar{x} &= \overline{x'} + M \\ &= 3.5 + \frac{18}{44} \approx 3.9 \end{aligned}$$

计算结果与上例相同。这样计算可以避免当 x_j 很大或测量数据很多时, $\sum n_j x_j$ 过大不便计算或可能引入较大的误差。适当选择 M 可使 $\overline{x'}$ 的计算较直接计算 $\bar{x}$ 简便。当计算均值 M 选择得接近 $\bar{x}$ 时, $\overline{x'}$ 将很小。

(3) 用递推算法计算算术平均值。

当测量数据很多或处于实时采样状态时,为节省计算机内存单元,可采用递推算法计算算术平均值。设 $\bar{x}_m$ 为计入第 m 次测得值 x_m 后的算术平均值,则

$$\bar{x}_m = \bar{x}_{m-1} + \frac{x_m - \bar{x}_{m-1}}{m}$$

三、方差的估计

在数理统计中,数学期望和方差是对总体(讨论问题的全体)而言。由于实验和测量的有限性,上面引进的与数学期望对应的算术平均值,它实际是样本对总体数学期望的估计。下面在样本中进一步引进标准偏差(或称样本方差),作为对总体方差的估计(而单次测量的随机误差则由残余误差估计)。

1. 贝塞尔(Bessel)公式

由于真值未知,对某一测量列 $x_1, x_2, \dots, x_n$, 能够获取的数值只是算术平均值 $\bar{x}$ 和残余误差 $v_1, v_2, \dots, v_n$ 。由残余误差 $v_i = x_i - \bar{x}$ 计算标准偏差的表达式为

$$\hat{\sigma}^2 = \frac{v_1^2 + v_2^2 + \dots + v_n^2}{n-1} \quad (1-9)$$

这即是贝塞尔公式。下式为贝塞尔公式的变形,它有利于减小舍入误差。

$$\hat{\sigma}^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] \quad (1-10)$$

下面从估计量的无偏性(估计量的数学期望等于被估计参数的真值)的角度出发,证明由残余误差平方和计算标准偏差的数学期望,就是测量总体的方差。

$$\begin{aligned} E\{\hat{\sigma}^2\} &= E\left\{ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right\} \\ &= \frac{1}{n-1} E\left\{ \sum_{i=1}^n [(x_i - a) - (\bar{x} - a)]^2 \right\} \\ &= \frac{1}{n-1} E\left\{ \sum_{i=1}^n (x_i - a)^2 - 2 \sum_{i=1}^n (x_i - a)(\bar{x} - a) + n(\bar{x} - a)^2 \right\} \\ &\quad - 2 \sum_{i=1}^n (x_i - a)(\bar{x} - a) = -2 \left(\sum_{i=1}^n x_i - na \right) (\bar{x} - a) \\ &\quad = -2n(\bar{x} - a)^2 \end{aligned}$$

代入 $E\{\hat{\sigma}^2\}$ 式

$$\begin{aligned} E\{\hat{\sigma}^2\} &= \frac{1}{n-1} E\left\{ \sum_{i=1}^n (x_i - a)^2 - n(\bar{x} - a)^2 \right\} \\ &= \frac{1}{n-1} \left\{ \sum_{i=1}^n E[(x_i - a)^2] - nE[(\bar{x} - a)^2] \right\} \\ &= \frac{1}{n-1} \left(n\sigma^2 - n \frac{\sigma^2}{n} \right) \\ &= \sigma^2 \end{aligned}$$

式中 $E[(\bar{x} - a)^2] = \frac{\sigma^2}{n}$ ——算术平均值的方差,随后将给予讨论。

2. 极差法

用贝塞尔公式计算标准偏差需首先计算算术平均值,再求残余误差,然后计算标准偏

差。若希望简化计算或实际测得值较少时,可采用极差法。

对于等精度重复测得值系列 $x_1, x_2, \cdots, x_n$, 选取最大值 $x_{\max}$ 和最小值 $x_{\min}$ 计算极差

$$R = x_{\max} - x_{\min}$$

标准偏差则为

$$\hat{\sigma} = \frac{R}{d_n}$$

式中 d_n ——按测量次数 n 由表 1-4 查得。

表 1-4 d_n 数值表

测量次数 n	1	2	3	4	5	6	7	8	9	10
d_n		1.13	1.69	2.06	2.33	2.53	2.70	2.85	2.97	3.08
测量次数 n	11	12	13	14	15	16	17	18	19	20
d_n	3.17	3.26	3.34	3.41	3.47	3.53	3.59	3.64	3.69	3.74

当测量次数 $n < 10$ 时采用本法比较适宜。

3. 最大误差法

对于服从正态分布的等精度测量列 $x_1, x_2, \cdots, x_n$, 若已知约定真值 μ , 则可由 $\delta_i = x_i - \mu$ 计算出各测得值的随机误差(设无系统误差), 取其绝对值最大值 $|\delta|_{\max}$, 则标准偏差为

$$\hat{\sigma} = \frac{|\delta|_{\max}}{K_n}$$

若约定真值未知, 则取其绝对值最大值 $|v|_{\max}$, 则

$$\hat{\sigma} = \frac{|v|_{\max}}{K'_n}$$

式中 K_n, K'_n ——按测量次数 n 由表 1-5 查得。

此方法计算方便, 也适宜于 $n < 10$ 时的场合。

表 1-5 K_n, K'_n 数值表

n	1	2	3	4	5	6	7	8	9	10
$\frac{1}{K_n}$	1.25	0.88	0.78	0.68	0.64	0.61	0.58	0.56	0.55	0.53
n	11	12	13	14	15	16	17	18	19	20
$\frac{1}{K_n}$	0.52	0.51	0.50	0.50	0.49	0.48	0.48	0.47	0.47	0.46
n	21	22	23	24	25	26	27	28	29	30
$\frac{1}{K_n}$	0.46	0.45	0.45	0.45	0.44	0.44	0.44	0.44	0.43	0.43

n	2	3	4	5	6	7	8	9	10	15	20	25	30
$\frac{1}{K'_n}$	1.77	1.02	0.83	0.74	0.68	0.64	0.61	0.59	0.57	0.51	0.48	0.46	0.44