目　录

第一章 绪　　论

§1-1 工程测量与数据处理

工程测量与数据处理的任务是观察装置系统中的物理过程，定量测试有关物理量，应用统计分析方法，抑制噪声，校正部分误差，求取接近于真值的最佳估计值及噪声的统计特性，并分析与综合各物理量之间的相互关系，建立系统过程的静态或动态数学模型。

工程测量数据处理在各个领域都有广泛应用。根据研究的对象及条件不同，采用不同的数据处理方法。下面介绍几个方面的实际应用及其数据处理方法。

一、过程参数测量与整理

根据专业知识、误差理论与随机量数学，对大量实时测量数据进行可疑数据判别及剔除、误差分析、计算与估计，求得最佳参数估计值及误差分布的概率密度函数，并根据需要打印有关参数及其精度。例如利用风洞对单根圆管发热强迫对流传热参数进行测试时，应用微型计算机实时采入风洞流速压差、热电偶测量的温差、圆管加热电压，进行在线数据处理，迅速打印出6个参量：圆管壁面与气流间温差、风洞风速、热功率、雷诺数、努谢尔特数和放热系数，并给出相应误差与精度。这样，便可解决手工测试与处理速度慢、计算量繁重、容易出差错等问题，实现了自动测试与数据实时处理[9]。

二、动态测试智能化仪表

通过数据处理，可解决热工动态测量中一些特殊问题，如检测快速变化的瞬时值，非线性动态补偿等。在工程的实际应用中，有用气泡的概率密度检测两相流中的气泡份额；有利用燃烧时的微压脉动的功率谱分析研究燃烧效率；有基于不稳定导热原理测定绝缘材料导热系数等。为了使用方便，通常都制成专门的智能化仪表。

三、最佳运行方案

建立系统过程稳态数学模型，确定最佳工况的运行参数。例如风力发电中，对风速、转速、扭矩等参数的测试与处理，可得到不同风速下的气动功率与转速的关系。由此可确定风力发电机按最佳参数的运行方案。在核动力装置运行过程中，以装置热效率最高为目标，确定装置回路中各主要热工参数最佳的运行方案。

四、系统参数预报与故障诊断

应用所建立的预报模型，依据实时测量的有关参数，可进行系统参数预报与故障诊断。例如，由所测量的振动频率与振幅大小变化，可判断汽轮发电机是否正常工作；应用两相流模型，由锅炉水冷壁中两相流变化，可判断水冷壁是否将发生爆管等。

五、系统辨识

系统辨识是通过观察系统的输入输出关系来建立系统的动态模型。动态模型辨识包括验前结构假定及模型参数辨识两部分。验前结构一般由机理分析、实验研究与近似技巧来确定，应用模型鉴别方法来检验。模型结构确定后，模型参数辨识是在存在测量噪声、系统

中有干扰的条件下，通过辨识的方法，找出一个能与具有噪声的观察数据最优拟合的数学模型。经典辨识方法有阶跃响应法、脉冲响应法及幅相特性法。近代辨识方法有基于随机过程理论的相关谱分析方法、滤波技术、最小二乘法、最大似然法、辅助变量法等。应用微型机，制作辨识用软件包，可对热工系统进行动态系统模型辨识。例如，可根据运行数据，求取过热器、给水系统、燃烧系统等传递函数及系统滞后时间。根据热工系统动态模型，可以制作成训练仿真器，以研究装置系统动态，故障分析与危险事故仿真，培训操纵人员。

§1-2 测量误差的基本概念

任何测量都是一种实验过程。实验的结果都含有误差与噪声，即存在各种干扰量。因此，误差自始至终存在于一切科学实验的过程中，称为误差公理。在工程测试中，被测参数，如系统中的温度、压力、流量等，都存在一个客观真值 A_0。由于测量仪器的效率、测量方法误差、环境干扰量、人的观察能力限制等，真值是无法测到的。在稳态测量时，采用有限次测量的算术平均值作为真值的最佳估计值。在动态测量中，采用递推式滤波后的值作为最佳估计值。

从计量意义上讲，测量就是将被测量直接或间接地与一个同类已知量相比较，把已知量作为计量单位，定出两者的比值作为测量值。计量单位即参考标准，可以为理论值，如真空中光速；可以是定义值，如三角形内角之和为 180°；也可以由国际或国家标准器设定。我们试验所用的仪器都由上一级计量机关的标准器检定，因此由这些仪器确定的真值称为相对真值。

按获得方式，测量方法可分为直接测量与间接测量。直接测量是将被测量与参考标准直接进行比较，例如用卡尺测钢珠直径；用温度计测汽机轴温。间接测量是通过测量与被测参数有确定函数关系的其它量，再通过函数关系式得到。例如金属导热系数 k 的函数关系式

$$k=\frac{(q/A)\Delta h}{(T_1-T_2)} \tag{1-1}$$

由直接测量得到两块金属板温度 T_1、T_2，板厚度 Δh，传热速率 q，传热面积 A，估计它们各自误差，再根据(1-1)求出 k 值及其精度估计。

按测量状态可分为稳态与动态测量。稳态测量是指测量过程中被测参数不改变；动态测量指瞬态测量，测量过程中被测参数随时都在变化。

测量过程中存在的误差称为测量误差。按表示形式，测量误差可分为绝对误差与相对误差。按误差性质，可分为系统误差、随机误差及粗差。

一、绝对误差与相对误差

1．绝对误差　绝对误差可用被测参数的量纲表示

$$\Delta x=x-\mu=\text{给出值}-\text{真值（量纲）} \tag{1-2}$$

给出值包括测量值、仪表指示值、实验值、计算近似值（间接测量值）等。

2．相对误差　相对误差一般用百分数表示，无量纲。由于所取参考值不同，可分为

$$\text{实际相对误差} = \frac{\Delta x}{\mu} \times 100\% \tag{1-3}$$

$$\text{标称相对误差} = \frac{\Delta x}{x} \times 100\% \tag{1-4}$$

$$\text{额定相对误差} = \frac{\Delta x}{x_{max}} \times 100\% \tag{1-5}$$

其中，x_{max} 为满量程值。

在工程中，采用最大额定相对误差作为测量传感器的精度等级，即

$$\text{最大额定相对误差} = \frac{\Delta x_{max}}{x_{max}} \times 100\% \tag{1-6}$$

其中，Δx_{max} 为全量程范围内的最大测量误差。

例1-1 某差压传感器测量范围为0～0.2MPa,量程范围内最大绝对误差为0.0004MPa，则该传感器的精度为

$$\frac{0.0004}{0.2-0} \times 100\% = 0.2\%$$

故该传感器精度等级为0.2级。

在工程中，采用最大引用误差划分测量仪器的精度等级，其定义为

$$\text{最大引用误差} = \text{最大示值误差}/\text{满刻度值} \tag{1-7}$$

显然，这与最大额定相对误差意义是相同的。

例1-2 检定2.5级电压表，量程100V，检定中发现最大示值误差在50V处为2V，故该表精度2.5级是合格的，检定的最大引用误差为2%。

二、系统误差、随机误差与粗差

1．系统误差 系统误差指按一定规律变化的误差，其中包括常值误差。在相同条件下，多次测量同一量，其误差保持不变，只有条件变化了，才按确定规律变化。系统误差产生的原因往往可知，须尽量消除产生的根源。例如，调试好仪表零点与量程；按规定安装好仪表位置；采用对称法、替代法等减少系统误差，通过加修正值方式消除确定性残存的系统误差。在数据处理时，还应采用统计检验方法检查是否还残留未被注意的系统误差，若还有，需估计其范围。

2．粗差 粗差是由于测量时疏忽大意造成的误差。这类误差很难估计。带有这类误差的测量结果是不符合事实的，应舍弃不用。例如，反读游标卡尺值、记录错误值、仪表损坏后指示值等。要注意的是对可疑值须仔细分析，确定为粗差后才能剔除，谨防由于某种原因剔除合理的数据。可依基本法则剔除含粗差的数据，最好几种法则同时使用，都符合后才剔除。若被剔除的数据百分比较大，应怀疑数据的误差分布不是正态分布，而是某种非正态分布。

3．随机误差 在同一条件下，对同一对象反复多次测量，在尽量消除系统误差及粗差条件下，每次测量结果仍然存在没有一定规律的误差，这是由于测量过程中的随机因素造成的，称为随机误差。

从每次测量结果看，随机误差数值或大或小、或正或负，没有一定规律性，不可预料，也无法控制，但从多次重复测量的总体看，服从一定的统计规律。一般，随机误差服

从正态分布，反映测量值自身的离散程度，它具有对称性，故其平均值随测量次数增加而逐渐趋于零。这样，通过多次测量，取测量值的平均值，便可减少随机误差，作为真值的最佳估计值。

随机误差的产生是由大量、均匀小的因素共同影响的结果。例如汽机轴温测量，有关因素的微小变化都会对结果数据产生影响，而有些小因素往往无法知道，也无法控制。

随机误差与系统误差虽是两类不同性质的误差，但它们都是误差，都有确定的界限，都存在于一切实验之中，虽可削弱、减少，但无法彻底消除。它们之间也不是绝对不可逾越的，应该辩证地看待这两类误差。有些系统误差的出现带有随机性，往往可按随机误差处理；有些系统误差很小的场合，总误差可按纯随机误差处理。同样，有些随机误差随人们认识深化，归为系统误差。在随机误差很小、系统误差很大的情况下，总误差可按纯系统误差处理。若系统误差与随机误差均不可忽略时，应用两类误差合成的方法处理。

三、精密度、准确度与精确度

工程实验的测量结果，往往用精确度来描述（即精度），它是上述三种误差的综合结果。精密度表征多次重复测量数据的分散程度，一般用标准误差(σ)表示。若测量的随机误差小，重复性就高，分布密集，其精密度高，σ 值小。准确度是表示测量结果偏离真值的程度，可用系统误差(e)表示。若测量的系统误差小，其准确度就高。精确度则是测量结果的精密度与准确度的综合反映，只有随机误差与系统误差都很小时，精确度才高。测量中粗差应该剔除。产生粗差的原因往往是随机的，但对测量结果影响却是固定的，直接影响测量的精确度。

具体测量中应注意区别下列情况：若被测对象是稳定的，则重复测量所得的数据主要反映测量仪器的效率或周围条件的波动。若测量数据重复性好，说明测量条件稳定。如果被测对象不稳定，测量条件稳定，测量数据主要反映对象的不稳定情况。若两者都不稳定，则是两者变化的综合反映。

四、不确定度与置信概率

在工程测量中，因为真值μ不可知，误差 Δx的具体准确数值不可能得到。然而，可依据统计估计，定出 Δx绝对值的上界值U，即

$$U = \mathrm{SUP}\,|\Delta x| \tag{1-8}$$

或

$$|\Delta x| = |x-\mu| < U \tag{1-9}$$

上界值U称为不确定度，这是估计的极限误差。所谓估计是指可能值，可能程度的大小用置信概率P_U（或ξ）给出

$$P\{|\Delta x| = |x-\hat{\mu}| \leqslant U\} = P_U \tag{1-10}$$

即测量误差 Δx落在$[-U,\ U]$范围内的概率。这里，$[-U,\ U]$为置信限，U为不确定度，$\hat{\mu}$为真值μ的最佳估计值。若$U = k\hat{\sigma}$，则k为置信系数，$\hat{\sigma}$为估计的标准误差。

对同一测量结果，若估计的U值较小，$|\Delta x|$ 不小于U的概率就大，这要冒估计不足的风险。反之，估计的U值较大，则$|\Delta x|$ 不大于U的置信概率就较大。考虑极端情况，若取 $|\Delta x|$ 为无穷大，则 $|\Delta x| < \infty$的置信概率$P_U = 100\%$。在具体测量中，P_U一般取68%、90%、95%、97.5%、99.7%等几种，视具体情况确定P_U值。

§1-3 有效数字与计算法则

一、数据的表示方法

试验数据是实验信息与结果的记录，要准确、简明、形象地表示，通常有列表、作图和经验公式等3种。

1．列表法 试验数据表的项目与名称要简明，数字写法要统一、正确，有效数字取舍要合适，自变数要按序排列，间距数值要适当。列表法简单、易作，常用的列表法有统计式、定性式、函数式、定量式。

2．作图法 作图法常用直角坐标，也用单对数坐标、双对数坐标、极坐标等。作图法形象显明，便于直观比较。作图时坐标分度起点不一定为零，应使图形占满整个坐标纸。一般，坐标纸的最小分格相应于试验数据的精确度。

3．经验公式法 根据曲线形状，大多数可应用最小二乘法对数据进行处理，得到经验公式，具体见第五章所述。

二、有效数字

工程实验数据有两种数，一种数是它们的每一位数都是确定的，例如$\sqrt{2}$、π、e等，其有效位可认为是无限的。另一种数是用来表示测量值的，其末位往往由估计而来，具有一定误差或不确定性。正常测量时，一般能估计到测量仪器最小刻度的1/10，故记录数据时，允许末一位数字是由估计得到的。例如，用最小刻度为毫米的尺量某长度为24.51cm，其第4位数1是估计的，有效数字为4位。若写成24.510cm，则错了，因为有5位有效数字。对于"0"，在20.05℃、1060H_2O柱中所有0均是有效数字；但在0.00150m中，前3个0均非有效数字；在15×10^2kg中有效数字为2位，而在1500kg中，有效数字为4位，因此，记录书写时要正确写下有效数字位数。

三、数据计算与处理方法

数据处理时，常用的基本计算方法有

1．记录测量值时 要确定有效数字位数，只允许末位为估计数字。

2．四舍五入、奇收偶弃法则 有效数字确定为N位，则$N+1$位后的数字大于0.5时，末位进1；小于0.5时舍去，末位不变；等于0.5时，则使末位凑成偶数，末位为奇数进1，为偶数时舍弃。

例如将下列数据保留4位有效数字

3．14159 为3.142； 2.71729 为2.717； 4.51050 为4.510； 3.21650 为3.216； 6.378501 为6.379； 7.691499 为7.691

若第$N+1$位以下的数本身就是通过舍入而来的，则应以舍入前的数据按上述法则来处理。采用奇收偶弃的目的，不仅使末位成偶数，便于计算，更主要的是使舍入误差为随机误差，而不造成系统误差，即使

$$\sum_{i=1}^{p}\text{舍入误差}=0 \qquad (p\text{为数据总数})$$

3．加减计算法则　加减运算中，有效数字以小数点位数最少的为准。在运算量大的计算中，为使误差不迅速积累，可多取一位有效数字。例如

$$50.4+2.02+0.051=52.47$$

4．乘除计算法则　乘除计算中，有效数字以最少有效数字位数为准。例如

$$[601.11\times0.32+4.016]=60\times10^{1}\times0.32\div4.0=48$$

5．对数计数　真数与对数的有效位数应相同，故查对数表时，真数为 n 位，就查 n 位对数表。

6．平均值计算　若有 4 个或多于 4 个数取平均，则平均数的有效位数可增加 1 位。

7．常数 π、e、$\sqrt{2}$、$\sqrt{3}$ 等的有效数字位数，需要几位就取几位。表示测量值精度时，一般只取 1 位有效数字，至多不超过 2 位。

8．界限数值不得修约　例如 X^{+A}_{-B} 表示为在 $(X-B)$～$(X+A)$ 范围内均允许；小于 $(X-B)$，大于 $(X+A)$ 为不合格。具体讲，若界限数值为16.5～17.5mm，出现 17.51mm 时，应判为超差；但若出现17.509mm，应视为17.5mm。

§1-4　实验数据分析的基本问题

工程中的实验数据处理，按性质可分为确定性数字信号分析处理与统计信号分析处理。本书讨论的是统计信号数据分析处理的基本问题。

一、基本测量的误差分析

在热工装置系统运行与实验过程中，要求测量许多参数量。取得的测量数据中包含着随机误差、粗差与系统误差。对于直接测量数据误差的处理，首先要检验可疑值，剔除粗差。因为含有粗差的测量值，无法进行统计分析。其次是分析随机误差，检验数据是否是正态分布，估计最佳可信值及其误差区间与置信概率。第三是分析系统误差，根据数据是否偏离正态分布，顺序分析固定、累进、周期性系统误差。最后校正系统误差。作间接测量时，需将各直接测量的物理参数的误差，通过已知的函数关系传递到间接测量参数中来，故间接测量中的误差分析，称为函数误差分析，又称误差传递。若已知函数关系及函数误差，要求允许的直接测量参数的误差范围，或者说要求函数误差最小，求直接测量的最佳条件，这就是测量精度的实验设计。

二、统计检验问题

就其本质来讲，基本测量问题是一种统计推断过程。例如测量的随机误差是在认为基本符合正态分布的条件下进行统计推断的。在工程实验数据处理中，还存在大量数据，其分布并不知道，只能假定它为某个分布，通过样本数据来进行统计推断，这一类问题称为统计检验。统计检验大致分为 3 种类型：显著性检验（参数检验）、非参数检验和相关性检验。

1．显著性检验　检验分布中某参数假定值是否与观察样本有显著矛盾的一类检验称为参数显著性检验。当工程中所取样本很小时，为提高检验的效率，采用复合假设方法进行参数显著性检验。

2．非参数检验　检验样本随机变量的分布是否为某种理论的分布形式，称为非参数检

验或拟合优度检验。应用非参数检验可检验两个随机变量是独立还是相关，而不必知道其分布类型与形状。

3．相关性检验　工程中，应用回归分析方法建立对象输入、输出量之间的静态数学模型，或称经验方程。实质上，这是通过对两个随机变量的相关性检验，来确定两个随机变量的线性相关程度。

要正确地、合理地进行统计检验，首先要进行实验设计，以取得所需的数据，满足检验条件。

三、建模与辨识

热工系统最优设计、系统最佳运行参数确定、实时最优控制策略研究、系统参数预测、故障分析与安全措施制定，都要先建立对象的模型，然后进行计算与仿真研究。模型分为物理模型与数学模型。物理模型是将对象作成小尺度实物模型，应用相似原理进行研究。数学模型是以数学式来描述对象的特征或本质，在计算机上进行仿真研究。数学式一般分为代数方程、微分方程、传递函数、状态方程等函数关系式。建立系统模型称为建模。建模方法有机理法和实验法2种。机理法是应用专业知识，根据系统过程的物理机制，利用有关定律、定理及各种简化假设，导出相应的数学式来近似描述系统过程，也称为理论建模法。实验法即系统辨识法，是在对系统输入、输出量进行定量测试的基础上，从一类系统中确定一个与所测系统等价的系统。对于静态数学模型，一般采用回归分析法，识到各种回归方程；对于动态模型，一般采用传递函数、微分方程或差分方程等。系统辨得主要是针对动态模型辨识而言的。动态模型辨识首先要确定“一类系统”，即模型结构预前假定，一般可通过机理法或以往经验来确定，然后再辨识模型结构参数，故又称为系统参数辨识。

要想在有限条件下，最大限度地得到实验数据的信息内容，尽可能减少试验次数，简化数据处理过程，节省人力、物力与时间，就要合理地进行试验设计。例如系统激励（输入）信号形式、幅度、时间长度的选择，试验点安排等。其次要确定用什么方法估计据数，例如相关法、最小二乘法、广义最小二乘法、最大似然法等，以找到一个与测量数参拟合得最好的数学模型。

第二章　随机变量与随机过程

§2-1　随机变量及理论分布

一、事件与随机变量

在工程实验中，把在一定条件下发生的现象、状态及测试结果都称作事件。在一定条件下必然发生的与不可能发生的事件称为必然事件与不可能事件，统称为确定性事件。对一次实验中可能发生也可能不发生，而在大量重复试验中有其内在必然性的事件称为随机事件。例如工程测量中，用同一仪器对同一参数在相同条件下进行重复多次测量，其测量值总是各不相同的，而每一个可能出现的量值都有一定机会出现（称这种试验为随机试验），一次测量得到的量值就是一个随机事件。为了描述随机现象，引入表述随机试验结果的量X，称为随机变量。

物理参数的实际测量值的可能值全体是随机变量。对某一次测量，随机变量取该量测量值的一个随机数，故随机变量是随机事件的单值函数。

随机变量分为离散型与连续型两类。例如产品合格与报废是离散型，随机误差则是连续型。通常用概率密度函数与分布函数描述随机变量。

例2-1　对某光纤粗细作100次测量，研究其测量值的分布形状。测量数据如表2-1所示。试问光纤粗细的真值最大可能是多少？能不能以99%或95%的把握确定其区间？

要正确回答这些问题，就须了解测量值随机变量取各种值的概率，即研究它的分布规律，步骤如下：

表2-1　光纤测量数据　　单位[10μm]

1.36	1.49	1.43	1.41	1.37	1.40	1.32	1.42	1.47	1.39
1.41	1.36	1.40	1.34	1.42	1.42	1.45	1.35	1.42	1.39
1.44	1.42	1.39	1.42	1.42	1.30	1.34	1.42	1.37	1.36
1.37	1.34	1.37	1.37	1.44	1.45	1.32	1.48	1.40	1.45
1.39	1.46	1.39	1.53	1.36	1.48	1.40	1.39	1.38	1.40
1.36	1.45	1.50	1.43	1.38	1.43	1.41	1.48	1.39	1.45
1.37	1.37	1.39	1.45	1.31	1.41	1.44	1.44	1.42	1.47
1.35	1.36	1.39	1.40	1.38	1.35	1.42	1.43	1.42	1.42
1.42	1.40	1.41	1.37	1.46	1.36	1.37	1.27	1.37	1.38
1.42	1.34	1.43	1.42	1.41	1.41	1.44	1.48	1.55	1.37

1．从表中找出最大值为1.55，最小值为1.27。

2．决定组距和组数。一般分为10～20组，视数据多少而定。本例中取10组，等距分组，确定分组点，如表2-2所示。

3．计算各组频数，即落在各组中数据的数目。各组频数与样本容量（本例中为100）

之比称频率。

4. 计算各组累积频数和累积频率，填入表2-2中，画出频数（频率）分布直方图，见图2-1。

表2-2 频数、频率分布表

组 次	组 区 间	频 数	组 频 率	累积频数	累积频率
1	1.265～1.295	1	0.01	1	0.01
2	1.295～1.325	4	0.04	5	0.05
3	1.325～1.355	7	0.07	12	0.12
4	1.355～1.385	22	0.22	34	0.34
5	1.385～1.415	24	0.24	58	0.58
6	1.415～1.445	24	0.24	82	0.82
7	1.445～1.475	10	0.10	92	0.92
8	1.475～1.505	6	0.06	98	0.98
9	1.505～1.535	1	0.01	99	0.99
10	1.535～1.565	1	0.01	100	1.00

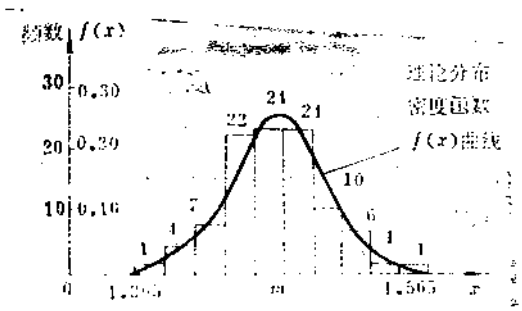

图2-1 频数（频率）分布直方图

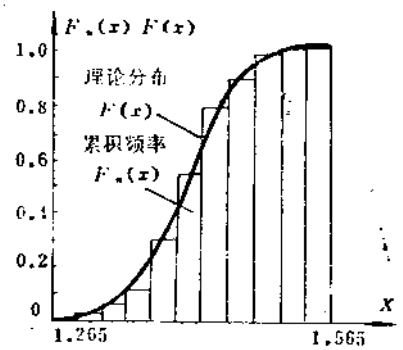

图2-2 经验分布图

5. 画出经验分布（累积频率分布）图，如图2-2，其分布数据来自样本数据。理论分布是指随机变量X的总体分布，即样本容量n趋于无穷大时的极限分布。在图2-2上画出理论分布曲线$F(x)$，图2-1上画出理论分布密度函数曲线$f(x)$。

二、概率分布与密度函数

1. 一维随机变量　由例2-1可知，当逐步增加样本容量，并相应地加大组数后，各组频率将逐步稳定在某确定值。当样本容量$n\to\infty$时，各组频率任意地接近某一定值，此定值称为概率。此时，直方图演变为一光滑曲线，称为概率分布密度曲线，用$f(x)$表示。经验分布则趋于理论分布，以$F(x)$表示分布函数。

若X为随机变量，x为一个任意实数，x_1，x_2，…，x_n为具体的随机变量值，则落在x与$x+dx$区间的随机变量的概率为

$$p(x<X\leqslant x+dx)=F(x+dx)-F(x)$$
$$=f(x)\,dx \qquad (2\text{-}1)$$

式中，$f(x)$为概率密度函数。

定义随机变量X的分布函数为

$$F(x)=P(X\leqslant x_1)=\int_{-\infty}^{x_1}f(x)\,dx \tag{2-2}$$

式中，$P(X\leqslant x_1)=P(-\infty<X\leqslant x_1)$，且有

$$f(x)=\frac{dF(x)}{dx} \tag{2-3}$$

例如在例2-1中，取$x_1=1.415$，则

$$F_n(x)=P_n(X\leqslant 1.415)=0.58$$

2. 二维随机变量　多输入多输出系统的概率模型要用多维随机变量表示，其中以二维随机变量为最常见，它由x、y 两个随机坐标值描述。如雷达显示屏幕上的随机点可形象地理解为$(X,\ Y)$两维随机变量。

设有两个任意实数x_1，y_1，事件$(X\leqslant x_1)$与$(Y\leqslant y_1)$同时出现的概率为

$$F(x,y)=P(X\leqslant x_1,Y\leqslant y_1)=\int_{-\infty}^{x_1}\int_{-\infty}^{y_1}f(x,\ y)\,dy\,dx \tag{2-4}$$

称为二维随机变量$(X,\ Y)$联合分布函数。

同样，类似(2-3)可得

$$f(x,\ y)=\frac{\partial^2F(x,\ y)}{\partial x\partial y} \tag{2-5}$$

若X的分布密度与Y的出现值有关，即在$Y=y$条件下，X的分布密度记为$f(x/y)$，称为条件分布密度。这种性质称为两随机变量之间相关。这种相关性可能强，也可能弱。若弱到可以忽略，则可认为X、Y两随机变量是相互独立的。此时，它们的条件分布密度函数

$$f(x/y)=f_1(x)\qquad (X\text{独立于}Y)$$
$$f(y/x)=f_2(y)\qquad (Y\text{独立于}X)$$

成立，它们的联合分布密度函数为

$$f(x,\ y)=f_1(x)f_2(y)\qquad (X、Y\text{相互独立}) \tag{2-6}$$

它们的联合分布函数为

$$\begin{aligned}F(x,\ y)&=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,\ y)\,dy\,dx\\&=\int_{-\infty}^{x}\int_{-\infty}^{y}f_1(x)f_2(y)\,dy\,dx\\&=\int_{-\infty}^{x}f_1(x)\,dx\int_{-\infty}^{y}f_2(y)\,dy=F_1(x)F_2(y)\end{aligned} \tag{2-7}$$

连续随机变量分布函数概念同样适用于离散随机变量，有

$$\left.\begin{aligned}F(x)&=P(X\leqslant x)=\sum_{x_i<x}P(X=x_i)\\F(x,\ y)&=\sum_{x_i<x}\sum_{y_i<y}P(X=x_i,\ Y=y_i)\end{aligned}\right\} \tag{2-8}$$

三、正态分布函数

大量的实验资料表明，测量的随机误差是服从正态分布的。正态分布是连续型随机变量的一种理论分布，由高斯（F·Gauss）1795年首先提出，也称高斯分布。

由例2-1可见，通过样本建立直方图。对含有随机误差的量值分布进行总结，可得如下性质：

1．单峰性　误差出现的概率只有一个峰值，绝对值越小的误差，出现的概率越大。测量值$x_i=\mu+\varepsilon_i$，故随机误差ε_i的概率密度曲线是把图2-1中的纵坐标移到m处。

2．对称性　绝对值相等的正负误差出现的概率相等。

3．抵偿性　由对称性可知，在等精度测量条件下，测量次数$n\to\infty$时，随机误差之和应该趋于零。

4．有界性　在等精度测量条件下，误差的绝对值不会超过一定的界限。

由以上性质可推导出正态分布的概率密度函数。

令μ为真值，$m_1, m_2, \cdots, m_n$为测量值，$x_1, x_2, \cdots, x_n$为对应的随机误差。由(2-1)可知，组距为$dx_1, dx_2, \cdots, dx_n$，出现误差$x_1, x_2, \cdots, x_n$的概率分别为$f(x_1)dx_1$，$f(x_2)dx_2, \cdots, f(x_n)dx_n$，由概率乘法法则，$x_1, x_2, \cdots, x_n$同时出现的概率为

$$P=\prod_{i=1}^{n} f(x_i)\,dx_i \tag{2-9}$$

在工程实验中，一次测试得$x_1, x_2, \cdots, x_n$值，说明其出现的概率最大。为方便运算，取

$$\ln P=\sum_{i=1}^{n}\ln f(x_i)+\sum_{i=1}^{n}\ln dx_i$$

令
$$\frac{d\ln P}{d\mu}=\sum_{i=1}^{n}\frac{d[\ln f(x_i)]}{dx_i}\cdot\frac{dx_i}{d\mu}=0$$

注意到dx_i与μ无关，又$dx_i=-d\mu$，故

$$\sum_{i=1}^{n}\frac{d[\ln f(x_i)]}{x_i dx_i}x_i=0 \tag{2-10}$$

按随机误差性质3，样本容量$n\to\infty$时，有

$$\sum_{i=1}^{n}x_i=0 \tag{2-11}$$

要使(2-10)和(2-11)同时成立，只有下式成立。

$$\frac{1}{x_1}\cdot\frac{d[\ln f(x_1)]}{dx_1}=\cdots=\frac{1}{x_n}\cdot\frac{d[\ln f(x_n)]}{dx_n}=k \tag{2-12}$$

式中k为任意常数。故对任一个误差x，有

$$\frac{d[\ln f(x)]}{dx}=kx$$

积分得　$\ln f(x)=\frac{1}{2}kx^2+\ln C$

取反对数，得

$$f(x)=Ce^{\frac{1}{2}kx^2} \tag{2-13}$$

由误差分布性质1，误差大时概率密度应减小，即k为负值，令$\frac{1}{2}k=-h^2$。又据全概率$\int_{-\infty}^{\infty}f(x)dx=1$，可得

$$C\int_{-\infty}^{+\infty}e^{-h^2x^2}dx=1$$

令 $u=\sqrt{2}hx$，代入上式

$$\frac{C}{\sqrt{2}h}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du=\frac{C}{\sqrt{2}h}\sqrt{2\pi}=1$$

故

$$C=\frac{h}{\sqrt{\pi}}$$

于是

$$f(x)=\frac{h}{\sqrt{\pi}}e^{-h^2x^2} \tag{2-14}$$

将(2-14)代入(2-9)，得

$$P=\prod_{i=1}^{n}f(x_i)dx_i=\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2\left(\sum_{i=1}^{n}x_i^2\right)}\prod_{i=1}^{n}dx_i$$

除去与P极值无关量，令

$$P^0=h^n e^{-h^2\left(\sum_{i=1}^{n}x_i^2\right)}$$

求

$$\frac{dP^0}{dh}=h^{n-1}e^{-h^2\left(\sum_{i=1}^{n}x_i^2\right)}\left[-2h^2\left(\sum_{i=1}^{n}x_i^2\right)+n\right]=0$$

得

$$h=\frac{1}{\sqrt{2}}\sqrt{\frac{n}{\sum_{i=1}^{n}x_i^2}} \tag{2-15}$$

定义

$$\sigma^2=\frac{\sum_{i=1}^{n}x_i^2}{n} \tag{2-16}$$

称为标准误差，其值大小反映测量误差x_i的分散性。

将(2-16)代入(2-15)得

$$h=\pm\frac{1}{\sqrt{2}\sigma} \tag{2-17}$$

代入(2-14)可得

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \tag{2-18}$$

此即随机误差的概率密度函数。

若以x表示测量值，μ为真值，$x-\mu$为误差值，以$x-\mu$替换(2-18)中的x，得

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

$$=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad (2\text{-}19)$$

这就是正态分布的常用型式。式中μ和σ^2称为正态分布的数学期望与方差，是决定正态分布的两个特征参数。图2-3给出了μ取定值，σ取不同值时的正态分布密度曲线。显然，μ表示分布的集中位置，σ表示分布的离散程度。

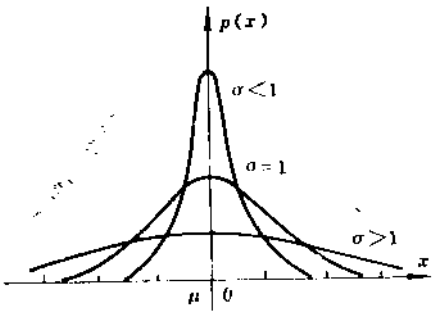

图2-3 μ、σ的直观意义

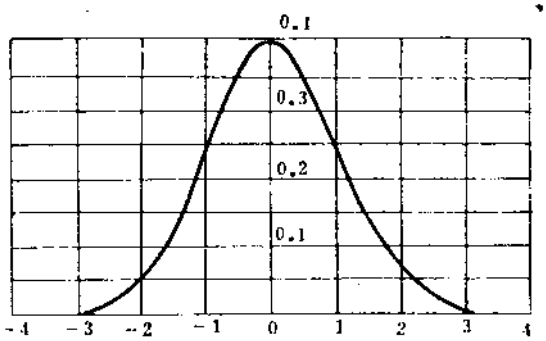

图2-4 $N(0,1)$曲线

记$\mu=0$，$\sigma=1$的正态分布为$N(0,)1$称为标准正分态布，如2-4图所示。其分布密度函数$f(x)$与分布函数$F(x)$为

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

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy \qquad (2\text{-}20)$$

对任何正态分布，只要进行变换，令$\frac{x-\mu}{\sigma}=u$,就可化成标准正态分布，其函数值已制成表（附录表1）。

例2-2 $\mu=20$，$\sigma=30$，计算$P\{x>80\}$值。

解：令$u=\frac{x-\mu}{\sigma}=\frac{80-20}{30}=2.0$

则 $$P\{x>80\}=P\{u>2.0\}=2.275\%\ (\text{查表})$$

例2-3 计算$k=1$，2，3时的$P\{|x-\mu|<k\sigma\}$值。

解： $P\{|x-\mu|<\sigma\}=P\{x-\mu>-\sigma\}-P\{x-\mu>\sigma\}$

$$=P\{\frac{x-\mu}{\sigma}>-1\}-P\{\frac{x-\mu}{\sigma}>1\}\quad (\text{查表})$$

$$=(1-0.1587)-(1-0.8413)=68.26\%$$

同样可求得

$$P\{|x-\mu|<2\sigma\}=95.4\%$$

$$P\{|x-\mu|<3\sigma\}=99.7\%$$

由此可见，当随机误差不确定度取$U=2\sigma$时，置信概率为95.4%，而随机误差$\geqslant 3\sigma$的概率为0.3%以下，故可用误差$\geqslant 3\sigma$作为判别存在粗差的准则。

四、其它分布函数

这里仅介绍几种。

1. 二项分布　作n次重复独立试验，出现事件A的概率为p；出现非A事件的概率为$q=1-p$；在n次试验中事件A出现k次的概率为$f(n,\ k,\ p)$。

例如，$n=5$，$k=2$时，可理解为从产品中任取5件，问5件中有2件为废品的概率是多少？已知废品率为p，按概率乘法法则，取废品2次合格品3次的概率为p^2q^3，而其排列次序共有C_5^2种。根据概率加法定理有

$$f(5,\ 2,\ p)=C_5^2p^2(1-p)^3$$

故二项分布的密度函数为

$$f(n,\ k,\ p)=\binom{n}{k}p^k(1-p)^{n-k} \tag{2-21}$$

式中，$C_n^k=\binom{n}{k}$

分布函数为

$$F(n,\ k,\ p)=\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i} \tag{2-22}$$

例2-4　某探测器测粒子效率为$p=0.55$，探测到的粒子数k服从二项分布，探测到至少1个粒子的概率为90%，现1个粒子未记录到，试估计在T时间内穿过探测器的粒子数。

解：由(2-21)得

$$f(n,\ 0,\ 0.55)=\binom{n}{0}p^0(1-p)^n=1-0.9=0.1$$

即

$$(1-0.55)^n=0.10$$

$$n=\frac{\lg 0.1}{\lg(1-0.55)}=2.9$$

故以90%的置信概率估计，穿过探测器的粒子数不大于2.9个。

2. 泊松分布　在二项分布中，k可取0，1，…，n。若出现事件A的概率p不是常数，而是$p=\frac{\lambda}{n}$，其中，$n=1,\ 2,\ 3,\cdots$，λ为大于零的常数，则随机变量k的分布规律可导出

$$\lim_{n\to\infty} f(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \tag{2-23}$$

这是泊松分布的密度函数，λ为期望值。

将$p=\frac{\lambda}{n}$代入(2-21)便可得

$$f(X=k)=\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{1}{\left(1-\frac{\lambda}{n}\right)^k}$$

$$= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \frac{1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k}$$

又

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\lim_{n\to\infty}\frac{1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k} = 1$$

便得到(2-23)。

例2-5　若例2-4中不存在仪器误差，探测器效率为 $p=100\%$，但仍有可能测不到粒子。因为粒子本身是随机变量，且服从泊松分布。若至少有1个粒子穿过仪器的概率为90%，试分析实际穿过仪器的粒子数为 N 的概率。

解：将 $k=N$ 代入(2-23)，得

$$f(X=N) = \frac{\lambda^N}{N!}e^{-\lambda}$$

$$f(X=0) = 1-0.9 = 0.1$$

$$1-f(X=0) = 1-e^{-\lambda} = 0.9$$

故

$$f(X=0) = 0.1 = e^{-\lambda}$$

$$\lambda = -\ln 0.1 = 2.3$$

以90%的置信概率估计，穿过仪器的粒子平均数 λ 不大于2.3个。

若既考虑仪器测量效率造成的误差，又考虑对象本身的随机性，则需将两者综合考虑。

例2-6　考虑粒子数 N 为随机变量，服从泊松分布，$f(X=N)=\frac{\lambda^N}{N!}e^{-\lambda}$，又考虑探测器效率。在给定 N 的条件下 k 为随机变量，服从二项分布，得 $f(k/N)=\frac{N!}{k!(N-k)!}\times p^k(1-p)^{N-k}$（条件密度函数），推证随机变量 k 的分布密度函数。

推证　应用全概率公式有

$$f(k) = \sum_{N=k}^{\infty} f(k/N)f(N) = \frac{(\lambda p)^k}{k!}e^{-\lambda}\sum_{N=k}^{\infty}\frac{[(1-p)\lambda]^{N-k}}{(N-k)!}$$

$$= \frac{(\lambda p)^k}{k!}e^{-\lambda}\sum_{i=0}^{\infty}\frac{[(1-p)\lambda]^i}{i!}$$

$$= \frac{(\lambda p)^k}{k!}e^{-\lambda},$$