研究生教材
YANJIUSHENG JIAOCAI

数学方法与应用

谷根代 主编
杨晓忠 刘敬刚 苏岩 副主编

中国电力出版社
CHINA ELECTRIC POWER PRESS

研究生教材
YANJIUSHENG JIAOCAI

数学方法与应用

主　编　谷根代
副主编　杨晓忠　刘敬刚　苏　岩
编　写　李　鹏
主　审　牛东晓

中国电力出版社
CHINA ELECTRIC POWER PRESS

内 容 提 要

本书为研究生教材，介绍了解决科学与工程中常见问题的数学方法，并强调方法在计算机上如何实现。全书共 7 章，主要包括数学基本知识回顾、矩阵理论、矩阵计算、数值逼近与数值积分、常微分方程初始问题的数值解、线性回归分析、时间序列分析等内容。

本书可作为理工科院校工程硕士和工学硕士研究生教材，还可作为工程技术人员和科研人员的参考用书。

图书在版编目（CIP）数据

数学方法与应用/谷根代主编．—北京：中国电力出版社，2011.12

研究生教材

ISBN 978-7-5123-2476-3

Ⅰ.①数… Ⅱ.①谷… Ⅲ.①高等数学—数学方法—研究生—教材 Ⅳ.①O13-0

中国版本图书馆 CIP 数据核字（2011）第 260831 号

中国电力出版社出版、发行

（北京市东城区北京站西街 19 号 100005 http://www.cepp.sgcc.com.cn）

北京市同江印刷厂印刷

各地新华书店经售

*

2012 年 1 月第一版 2012 年 1 月北京第一次印刷

787 毫米×1092 毫米 16 开本 11 印张 232 千字

定价 **19.00** 元

前　言

科学和工程技术的进步和发展，对在计算机上使用的数学方法提出了迫切的要求。

为了既方便从事电力研究的工程技术人员使用，又方便工程硕士和工学硕士研究生的数学教学，我们编写了此书。在内容上尽量选用电力领域的数学问题和常用的、行之有效的数学方法，并力求阐明算法思想和计算步骤。各章具有相对独立性，读者可以根据需要选用有关内容。

本书由华北电力大学的谷根代教授主编，杨晓忠、刘敬刚、苏岩副主编，李鹏编写。全书由谷根代统稿，华北电力大学牛东晓教授主审。牛东晓教授认真审阅了书稿，纠正了书中的一些不妥之处，并提出了许多宝贵的修改意见。华北电力大学数学建模与仿真中心的教师和同学们，对书中内容的讨论，保证了科学性。这里，向他们及本书所列参考文献的作者们表示衷心的感谢。

本书的出版得到国家自然科学基金（批准号 10771065）、河北省自然科学基金（批准号 A2007001027）和华北电力大学“211 工程”创新人才核心课程建设基金资助。

出版精品书籍，使莘莘学子受益，一直是作者追求的目标。但限于作者水平，加之时间仓促，书中错漏和不足之处在所难免，衷心期望广大读者和同行专家、学者批评指正。

编　者

2011 年 11 月

目　录

第1章 数学基本知识回顾

人类从数数开始逐渐建立了自然数和常数的概念，形成了常量数学或初等数学体系. 初等数学大约持续了两千年的历史，直到16世纪末结束. 17世纪以来，随着资本主义社会的发展，以机器为主的大工业占主导地位. 与此同时，对运动的研究变成了自然科学的中心问题. 在对各种变化过程和各种变化着的量之间的依赖关系的研究中，产生了变量和函数这两个新的数学概念. 直到17世纪后半叶，牛顿和莱布尼兹建立了微积分，开创了变量数学或高等数学的新篇章. 高等数学，以及后来发展起来的线性代数、概率论与数理统计等学科已成为当代大学本科生必备知识，也是进一步学习现代数学方法的基础.

本章主要回顾大学本科期间学过的并将在后面章节用到的知识，目的是为学习本书提供方便.

1.1 函数的微积分

对各种变化过程或各种变化着的量之间的依赖关系的观察、预测和跟踪，是人们研究“运动”的根本目的. 也就是说，研究函数的变化规律是微积分的主要内容. 本节不想在微积分的基本概念（如极限、连续、可微或导数、可积或积分、级数等）上泼洒笔墨，只介绍一些用于观察、预测和跟踪的概念、理论和方法，比如连续函数的介值定理、微分中值定理、泰勒（Taylor）公式、积分中值定理等.

定理1.1 （介值定理）设 $f(x)$在 $[a, b]$ 上连续，且 $f(a)=A$，$f(b)=B$，$A\neq B$，则对于 A 与 B 之间的任意一个常数 C，则至少存在一个 $\xi\in(a, b)$，使 $f(\xi)=C$.

推论1.1 （零点定理）设 $f(x)$在 $[a, b]$ 上连续，且 $f(a)$，$f(b)$异号，则至少存在 $\xi\in(a, b)$，使 $f(\xi)=0$.

介值定理是连续函数没有缝隙的最好解释. 介值定理和零点定理的直观描述如图1-1和图1-2所示.

定理1.2 （积分中值定理）设 $f(x)$在$[a, b]$上连续，则在$[a, b]$上至少存在一点ξ，使

$$\int_a^b f(x)\mathrm{d}x = f(\xi)(b-a) \tag{1-1}$$

积分中值定理表明：区间$[a, b]$上连续函数的积分值总可以规范到某个矩形的面积，如图1-3和图1-4所示.

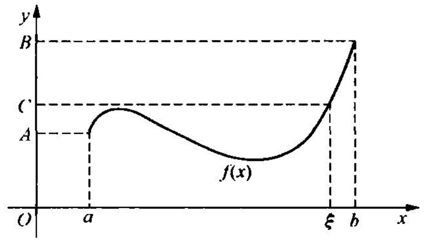

图 1-1 介值定理示意图

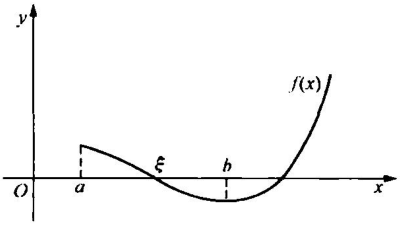

图 1-2 零点定理示意图

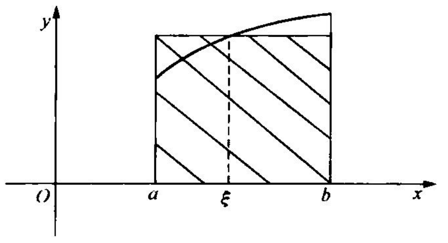

图 1-3 积分中值定理示意图（一）

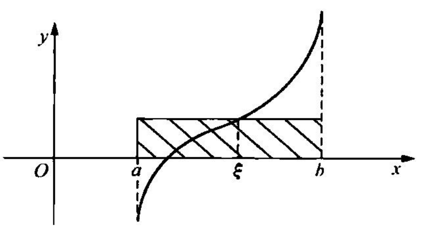

图 1-4 积分中值定理示意图（二）

定理 1.3 （加权积分中值定理）设 $f(x)$在$[a, b]$上连续，$g(x)$在$[a, b]$上不变号且可积，则在$[a, b]$上至少存在一点 $\xi(a\leqslant\xi\leqslant b)$，使

$$\int_a^b f(x)g(x)\mathrm{d}x = f(\xi)\int_a^b g(x)\mathrm{d}x \tag{1-2}$$

显然，$g(x)\equiv 1$ 时，加权积分中值定理就是积分中值定理.

可微函数（也称为光滑函数）往往比具有许多尖点特征的函数（连续但不可导）表现得更加容易预测. 它不仅保有连续函数的性质，而且显示出更多特性.

定理 1.4 （拉格朗日中值定理）设 $f(x)$在$[a, b]$上连续，(a, b)内可导，则至少存在一点 $\xi\in(a, b)$，使

$$f'(\xi)=\frac{f(b)-f(a)}{b-a} \tag{1-3}$$

拉格朗日中值定理的另一种表现形式为：对于任意的 $x\in[a, b]$ 和 $\Delta x+x\in[a, b]$，有$0<\theta<1$，使

$$f(x+\Delta x)-f(x)=f'(x+\theta\Delta x)\Delta x \tag{1-4}$$

它称之为有限增量定理或微分中值定理.

有限增量定理是估计预测精度（误差估计）的重要手段. 比如：如果 $f(x)$在 x_0 处有连续的导数，则有 $f(x)=f(x_0)+f'(x_0)(x-x_0)+o(|x-x_0|)$. 于是，当$|\Delta x|=|x-x_0|$充分小时，可预测出 x_0 的邻近点 x 的值

$$f(x)\approx f(x_0) \tag{1-5}$$

有误差 $f'(x_0)(x-x_0)$；或者

$$f(x)=f(x_0)+f'(x_0)(x-x_0) \tag{1-6}$$

有误差估计 $o(|x-x_0|)$.

【例 1-1】 计算$\sqrt{2}$的近似值.

解 $\sqrt{2}\approx\sqrt{1.4^2}=1.4$，有误差

$$\frac{1}{2\sqrt{x_0}}(x-x_0)=\frac{1}{2\sqrt{1.4^2}}(2-1.4^2)=0.0143$$

$$\sqrt{2}\approx\sqrt{1.4^2}+\frac{1}{2\sqrt{1.4^2}}(2-1.4^2)=1.4143，有误差\ o(0.04).$$

本例说明，估计式（1-6）比式（1-5）提供了更大的预测范围．可惜的是，式（1-6）没有明确给出像式（1-5）那样的误差估计式．

泰勒公式提供了更好的函数预测和误差估计方法，并显示了一般函数与多项式类函数的依赖关系．

定理 1.5 （泰勒定理）设 f：R→R 在 $x_0\in\mathrm{R}$ 及其邻域 $N_\delta(x_0)$ 有直到 $n+1$ 阶的导数，则对于任意 $x\in N_\delta(x_0)$，存在介于 x_0 和 x 之间的数 ξ，使得

$$\begin{aligned}f(x)=&f(x_0)+f'(x_0)(x-x_0)+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n\\&+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}\end{aligned}\tag{1-7}$$

实际中，要求函数具有 3 阶及 3 阶以上导数是很苛刻的．因此，最常用的是函数在 x_0 及其邻域 $N_\delta(x_0)$ 内的一阶泰勒公式（局部线性化公式）

$$f(x)\approx f(x_0)+f'(x_0)(x-x_0)\tag{1-8}$$

有误差为 $\frac{f''(\xi)}{2}(x-x_0)^2$．

局部线性化公式（1-8）是解决非线性问题的很重要的手段．图 1-5 和图 1-6 显示了不同曲率半径下局部线性化公式（1-8）的逼近效果．

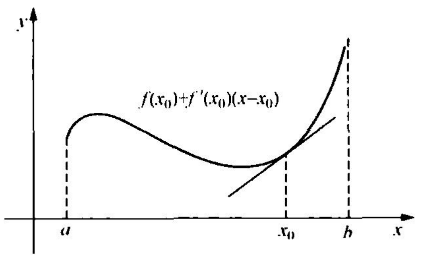

图 1-5 曲率半径小（$|f''(x_0)|$ 大）

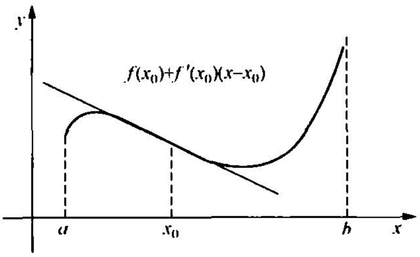

图 1-6 曲率半径大（$|f''(x_0)|$ 小）

注意，二阶泰勒公式

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2}(x-x_0)^2+\frac{f''(\xi)}{6}(x-x_0)^3\tag{1-9}$$

也很有用．

【例 1-2】 用一阶和二阶泰勒公式计算 $\sqrt{2}$ 的近似值．

解 一阶泰勒公式

$$\sqrt{2}\approx\sqrt{1.4^2}+\frac{1}{2\sqrt{1.4^2}}(2-1.4^2)=1.4143$$

有误差

$$e=\left|\frac{f''(\xi)}{2}(x-x_0)^2\right|\approx\frac{1}{8\times1.4^3}(2-1.96)^2=7.2886\times10^{-5}$$

二阶泰勒公式

$$\sqrt{2}\approx\sqrt{1.4^2}+\frac{1}{2\sqrt{1.4^2}}(2-1.96)-\frac{1}{8\times1.4^3}(2-1.96)^2=1.4142$$

有误差

$$e=\left|\frac{f'''(\xi)}{6}(x-x_0)^3\right|\approx\frac{3}{8\times1.4^5}\times0.04^3=4.4624\times10^{-6}$$

除了上面讨论的连续函数之外，还有科学与工程领域经常要面对的两类函数，一是周期函数，二是平方可积函数（能量有限函数）．对这两类函数的性质，法国数学家 J. Fourier（傅里叶）给出了明确的解答．

定理 1.6 设 $f(x)$是周期为 T 的函数，且在区间$\left[-\frac{T}{2},\ \frac{T}{2}\right]$上满足条件：(1) 连续或只有有限个第一类间断点；(2) 至多有有限个极值点．则 $f(x)$有傅里叶级数展开式

$$\frac{f(x-0)+f(x+0)}{2}=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos\frac{2n\pi}{T}x+b_n\sin\frac{2n\pi}{T}x\right)\tag{1-10}$$

其中

$$a_n=\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\cos\left(\frac{2n\pi}{T}x\right)f(x)\mathrm{d}x,\quad n=0,\ 1,\ 2,\ \cdots\tag{1-11}$$

$$b_n=\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\sin\left(\frac{2n\pi}{T}x\right)f(x)\mathrm{d}x,\quad n=1,\ 2,\ \cdots\tag{1-12}$$

定理 1.6 明确指出：一个周期为 T 的函数均为一系列简谐振荡函数

$$\begin{cases}u_n(x)=a_n\cos\frac{2n\pi}{T}x+b_n\sin\frac{2n\pi}{T}x=A_n\sin\left(\frac{2n\pi}{T}x+\theta_n\right)\\ n=0,\ 1,\ \cdots,\ m\end{cases}$$

的叠加．

定理 1.7 设 $f(x)\in L^2(R)=\left\{f(x)\,\middle|\,\int_{-\infty}^{\infty}[f(x)]^2\mathrm{d}x<\infty\right\}$（称之为平方可积函数），则有

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F(\omega)\mathrm{e}^{\mathrm{i}\omega x}\mathrm{d}\omega\tag{1-13}$$

这里，

$$F(\omega)=\int_{-\infty}^{\infty}f(x)\mathrm{e}^{-\mathrm{i}\omega x}\mathrm{d}x\tag{1-14}$$

称为 $f(x)$ 的傅里叶变换．

定理 1.7 中，$\mathrm{i}^2=-1$．式（1-14）称为 $f(x)$的傅里叶变换，显示了能量有限函数的谱分布，它是信号分析的基础；式（1-13）称为 $f(x)$的傅里叶逆变换或傅里叶积分．

1.2 向量与矩阵

1.1 节介绍了一种最简单的量与量之间的依赖关系，即一对一的关系，或者说是一元

函数. 实际上量与量之间的依赖关系要比这复杂得多，经常要观察、预测一对多和多对多的关系. 因此，必须把“数”扩展到向量和矩阵，才能方便地描述复杂的问题，如“系统”和“网络”等.

有关 n 维向量 $x=(x_1, x_2, \cdots, x_n)^T$ 和矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

的概念、运算和性质这里不再赘述，只对科学与工程实践中所面对的几个问题加以讨论.

1.2.1 线性代数方程组

形如

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1 \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2 \\ \qquad\vdots \\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=b_n \end{cases} \tag{1-15}$$

称为 n 元线性代数方程组. 若记

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},\quad b=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix},\quad x=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

则线性代数方程组（1-15）可简写为

$$Ax=b \tag{1-16}$$

线性代数的理论明确指出：线性代数方程组（1-16）有解的充分必要条件是系数矩阵 A 的秩与增广矩阵 $B=(Ab)$ 的秩相等. 并且，当 $\det(A)\neq 0$ 时，方程组(1-16)有唯一解 $x=A^{-1}b$.

1.2.2 方阵的特征值

设 A 为一个 n 阶实矩阵，满足方程

$$Ax=\lambda x \tag{1-17}$$

的数 λ 和非零向量 x，分别称为矩阵 A 的特征值和相应的特征向量.

如式（1-17）所表示的问题称为矩阵 A 的特征值问题，其用处很大，是系统的稳定性分析基础.

线性代数的理论指出：特征多项式方程

$$\det(A-\lambda E)=0 \tag{1-18}$$

包含了 A 的全部特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$（可以有重根），相应于每个特征值 λ_i 的特征向量包含在下列齐次线性代数方程组中.

$$(A-\lambda_i E)x=0 \tag{1-19}$$

1.2.3 向量范数和矩阵范数

观察或预测空间两点 x、x^* 之间的距离，需要下面的范数概念.

定义 1.1 在 $\mathbf{R}^n$ 上定义的一个实函数 $\|x\|$，如果对于任意的 $\alpha\in\mathbf{R}$，x，$y\in\mathbf{R}^n$，满足

(1) 非负性，即 $\|x\|\geqslant 0$，$\|x\|=0$ 当且仅当 $x=0$；

(2) 齐次性，即 $\|\alpha x\|=|\alpha|\cdot\|x\|$；

(3) 三角不等式，即 $\|x+y\|\leqslant\|x\|+\|y\|$；

则称 $\|x\|$ 为向量 x 的范数.

特别地，$\mathbf{R}^n$ 上两种常用的范数是

2—范数

$$\|x\|_2=\sqrt{x_1^2+x_2^2+\cdots+x_n^2} \tag{1-20}$$

∞—范数

$$\|x\|_\infty=\max_{1\leqslant i\leqslant n}(|x_i|) \tag{1-21}$$

相应地，两点 x^*、x 的距离

$$\|x^*-x\|_2=\sqrt{(x_1^*-x_1)^2+(x_2^*-x_2)^2+\cdots+(x_n^*-x_n)^2} \tag{1-22}$$

$$\|x^*-x\|_\infty=\max_{1\leqslant i\leqslant n}|x_i^*-x_i| \tag{1-23}$$

定义 1.2 在 $\mathbf{R}^{n\times n}$ 上定义一个实函数 $\|A\|$，如果对于任意的 A，$B\in\mathbf{R}^{n\times n}$，$x\in\mathbf{R}^n$，以及 $\alpha\in\mathbf{R}$，满足下列性质

(1) 非负性，即 $\|A\|\geqslant 0$，$\|A\|=0$ 当且仅当 $A=0$；

(2) 齐次性，即 $\|\alpha A\|=|\alpha|\cdot\|A\|$；

(3) 三角不等式，即 $\|A+B\|\leqslant\|A\|+\|B\|$；

(4) 相容性，即 $\|AB\|\leqslant\|A\|\cdot\|B\|$，$\|Ax\|\leqslant\|A\|\cdot\|x\|$；

则称 $\|A\|$ 为矩阵 A 的范数.

特别地，$\mathbf{R}^{n\times n}$ 中常用的矩阵范数是

∞—范数

$$\|A\|_\infty=\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}|a_{ij}| \tag{1-24}$$

1—范数

$$\|A\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|a_{ij}| \tag{1-25}$$

2—范数

$$\|A\|_2=\sqrt{\rho(A^{\mathrm{T}}A)} \tag{1-26}$$

其中，$\rho(A^{\mathrm{T}}A)=\max\limits_{1\leqslant i\leqslant n}(|\lambda_i(A^{\mathrm{T}}A)|)$ 称为矩阵 $A^{\mathrm{T}}A$ 的谱半径，$\lambda_i(A^{\mathrm{T}}A)$ 表示矩阵 $A^{\mathrm{T}}A$ 的特征值.

1.2.4 多元函数和向量值函数泰勒公式

把一元函数泰勒公式推广到多元函数 f：$D\subset\mathbf{R}^n\to\mathbf{R}$，即

$$y=f(x)=f(x_1, x_2, \cdots, x_n)$$

和向量值函数 F：$D\subset\mathbf{R}^n\to\mathbf{R}^n$，即

$$F(x)=\begin{pmatrix} f_1(x)\\ f_2(x)\\ \vdots\\ f_n(x)\end{pmatrix}=\begin{pmatrix} f_1(x_1, x_2, \cdots, x_n)\\ f_2(x_1, x_2, \cdots, x_n)\\ \vdots\\ f_n(x_1, x_2, \cdots, x_n)\end{pmatrix}$$

更有实际意义.

首先将一元函数可微的定义

$$\lim_{t\to 0}\frac{1}{t}[f(x_0+t)-f(x_0)-at]=0$$

推广到向量值函数 $F(x)$. 这里 $a=f'(x_0)$.

定义 1.3 向量值函数 F: $D\subset \mathrm{R}^n\to\mathrm{R}^n$ 在点 $x^{(0)}\in D$ 称为可微的，如果存在 n 阶矩阵 A，满足

$$\lim_{\|h\|\to 0}\frac{1}{\|h\|}\|F(x^{(0)}+h)-F(x^{(0)})-Ah\|=0\quad (h\in\mathrm{R}^n)\tag{1-27}$$

可以证明，这个矩阵 A 是唯一的.

记 $A=F'(x_0)$，它称为 $F(x)$ 在点 $x^{(0)}$ 的导数，且

$$F'(x^{(0)})=\begin{pmatrix}\dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n}\\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_n}\end{pmatrix}_{x=x^{(0)}}\tag{1-28}$$

多元函数可以看成向量值函数的特例. 因此，多元函数 f: $\mathrm{R}^n\to\mathrm{R}$ 的导数

$$f'(x^{(0)})=\left(\frac{\partial f}{\partial x_1},\ \frac{\partial f}{\partial x_2},\ \cdots,\ \frac{\partial f}{\partial x_n}\right)_{x=x^{(0)}}\tag{1-29}$$

记 $\nabla f(x^{(0)})=[f'(x^{(0)})]^{\mathrm{T}}$，称为多元函数 $f(x)$ 在点 $x^{(0)}$ 处的梯度.

多元函数的梯度函数 $\nabla f(x)$ 是向量值函数. 若 $\nabla f(x)$ 在点 $x^{(0)}$ 处可微，则 $f(x)$ 有二阶导数

$$f''(x^{(0)})=\nabla^2 f(x^{(0)})=\begin{pmatrix}\dfrac{\partial^2 f}{\partial x_1\partial x_1} & \dfrac{\partial^2 f}{\partial x_2\partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_1}\\ \dfrac{\partial^2 f}{\partial x_1\partial x_2} & \dfrac{\partial^2 f_i}{\partial x_2\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{\partial^2 f}{\partial x_1\partial x_n} & \dfrac{\partial^2 f}{\partial x_2\partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_n}\end{pmatrix}_{x=x^{(0)}}\tag{1-30}$$

$f''(x)$ 又称为 $f(x)$ 在 $x^{(0)}$ 处的海色（Hesse）矩阵.

定义 1.4 称向量值函数 F: $D\subset\mathrm{R}^n\to\mathrm{R}^n$ 在点 $x^{(0)}\in D$ 及其邻域二阶可微，如果

$$\lim_{\|h\|\to 0}\frac{1}{\|h\|}\|\mathrm{F}'(x^{(0)}+h)-F'(x^{(0)})-F''(x^{(0)})h\|=0\tag{1-31}$$

成立. 这里 $F''(x^{(0)})$ 为二阶导数.

如果 $F(x)$ 在 $x^{(0)}\in D$ 二阶可微，则有

$$F'(x^{(0)}+h)-F'(x^{(0)})-F''(x^{(0)})h=o(\|h\|)E$$

推出

$$F''(x^{(0)})h = F'(x^{(0)}+h) - F'(x^{(0)}) + o(\|h\|)E$$
$$= \left(\frac{\partial f_i(x^{(0)}+h)}{\partial x_j} - \frac{\partial f_i(x^{(0)})}{\partial x_j}\right)_{n\times n} + o(\|h\|)E$$
$$= \left(\sum_{k=1}^{n} \frac{\partial^2 f_i(x^{(0)})}{\partial x_k \partial x_j} h_k\right)_{n\times n} + o(\|h\|)E$$

于是，对于充分小的 $\|h\|$ 有

$$F''(x^{(0)})h = (H_1(x^{(0)})h \quad H_2(x^{(0)})h \quad \cdots \quad H_n(x^{(0)})h)^{\mathrm{T}}$$

其中

$$H_i(x^{(0)}) = f_i''(x^{(0)}) = \begin{pmatrix} \frac{\partial^2 f_i}{\partial x_1 \partial x_1} & \frac{\partial^2 f_i}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f_i}{\partial x_n \partial x_1} \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_2} & \frac{\partial^2 f_i}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f_i}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_n} & \frac{\partial^2 f_i}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f_i}{\partial x_n \partial x_n} \end{pmatrix}_{x=x^{(0)}} \tag{1-32}$$

$i=1,\ 2,\ \cdots,\ n.$

定理 1.8 （多元函数的泰勒公式）设多元函数 f：$\mathrm{R}^n \to \mathrm{R}$ 在 $x^{(0)} \in \mathrm{R}^n$ 及其邻域二阶可微，则对于此邻域中的任意 x，存在 $0 \leqslant t \leqslant 1$，使

$$\begin{aligned} f(x) = & f(x^{(0)}) + f'(x^{(0)})^{\mathrm{T}}(x-x^{(0)}) \\ & + (x-x^{(0)})^{\mathrm{T}} \frac{f''(x^{(0)}+t(x-x^{(0)}))}{2}(x-x^{(0)}) \end{aligned} \tag{1-33}$$

定理 1.9 （向量值函数的泰勒公式）设向量值函数 F：$\mathrm{R}^n \to \mathrm{R}$ 在 $x^{(0)} \in \mathrm{R}^n$ 及其邻域二阶可微，则对于此邻域中的任意 x，有

$$\begin{aligned} F(x) = & F(x^{(0)}) + F'(x^{(0)})(x-x^{(0)}) \\ & + \int_0^1 (1-t)F''(x^{(0)}+t(x-x^{(0)}))(x-x^{(0)})(x-x^{(0)})\mathrm{d}t \end{aligned} \tag{1-34}$$

1.3 随机变量及其分布

自然界和社会上发生的现象是多种多样的. 有一类现象，在一定条件下必然发生，例如向上抛一枚石子必然下落，同性电荷必然相互排斥等，这类现象称为确定性现象. 另一类现象表现为：一次试验或观察之前不能预知确切的结果，但大量重复试验的统计结果却呈现出某种规律性. 比如，在相同条件下抛一枚硬币，观察其哪面朝上？大量重复试验观察的结果是硬币的每一面大致各占一半. 这类现象，称为随机现象.

研究随机现象的手段就是随机试验，它的特点是：①可以在相同条件下重复进行；②每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；③一次试验之前不能确定哪个结果会出现. 如果将随机试验的每个结果记为样本点 e，把所有可能结果的集合记为样本空间 $S=\{e\}$，则随机试验的某事件 A 就是样本空间的子集（$A \subset S$）.

1.3.1 随机变量的概念

研究随机现象的目的就是确定事件 A 在一次实验中发生的可能性大小，称为事件 A 的概率，记为 $P(A)$. 为了更方便地实现这个目标，引入随机变量和随机变量分布函数的概念.

定义 1.5 对每一个样本 $e\in S$，规定一个实数 $X(e)$与之对应，这样就得到一个定义在样本空间 S 上单值实值函数 $X=X(e)$，称之为随机变量.

定义 1.6 设随机变量 X，$\forall x\in \mathrm{R}$，定义函数

$$F(x)=P\{X\leqslant x\} \tag{1-35}$$

称为随机变量 X 的分布函数.

若随机变量的值是有限个或可列个点时，称为离散型随机变量. 设离散型随机变量 X 的所有可能取值为 $x_k(k=1,2,\cdots)$，则有概率分布（或分布律）

$$P\{X=x_k\}=p_k,\ k=1,2,\cdots \tag{1-36}$$

凡不是离散型的随机变量称为非离散型随机变量. 在非离散型随机变量类中，人们只关注连续型随机变量.

定义 1.7 随机变量 X 的分布函数为 $F(x)$，若存在非负函数 $f(x)\geqslant 0$，对 $\forall x\in \mathrm{R}$ 有

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t \tag{1-37}$$

则称 X 是连续型随机变量. $f(x)$称为 X 的概率密度函数（概率密度）.

引入随机变量及其分布函数，事件 $A=\{e\mid X(e)\in I,\ I\subset \mathrm{R}\}$或$\{X\in I\}$的概率就转化为常规的数的计算问题. 即，当 X 为离散型随机变量时，$P\{X\in I\}=\sum\limits_{x_k\in I} p_k$；当 X 为连续型随机变量时，$P\{X\in I\}=\int_I f(x)\mathrm{d}x$.

1.3.2 几个重要的随机变量及其概率分布

（1）（0，1）分布$(0<p<1)$

$$P\{X=x_k\}=p^k(1-p)^{1-k},\ k=0,1$$

满足（0，1）分布的随机变量 X 描述的是只有两个可能结果的随机试验.

（2）二项分布 $X\sim b(n,p)$ $(0<p<1,\ q=1-p)$

$$P\{X=k\}=C_n^k p^k q^{n-k},\ k=0,1,2,\cdots$$

满足二项分布的随机变量 X 描述的是 n 次重复独立试验（n 重贝努利试验）中事件 A 出现的次数.

（3）泊松分布 $X\sim\pi(\lambda)$ $(\lambda>0)$

$$P\{X=k\}=\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!},\ k=0,1,2,\cdots$$

满足泊松分布的随机变量 X 描述的是在一个时间间隔内事件 A 出现的次数. 它与二项分布的关系是：$\lim\limits_{n\to\infty} b(n,p)=\pi(\lambda)$，其中 $\lambda=np$.

（4）均匀分布 $X\sim U(a,b)$

概率密度函数

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b \\ 0, & x\notin(a, b)\end{cases}$$

均匀分布的随机变量表示 X 落在(a, b)中任意等长度的子区间内的概率相同.

(5) 指数分布

概率密度函数（$\theta>0$ 为常数）

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0 \\ 0, & x\leqslant 0\end{cases}$$

指数分布的随机变量 X 具有无记忆性的特性，即

$$\forall s, t>0, P\{X>s+t|X>s\}=P\{X>t\}$$

(6) 正态分布 $X\sim N(\mu, \sigma^2)$

概率密度函数为

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

其中 μ，σ 为常数，且 $\sigma>0$.

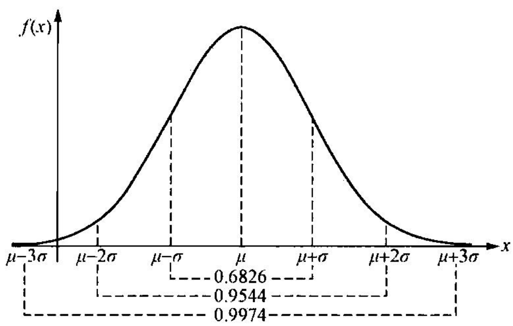

图 1－7 服从参数为 μ，σ 的正态分布

特别地，当 $\mu=0$，$\sigma=1$ 时，正态分布称为标准正态分布，记为 $X\sim N(0, 1)$.

如图 1－7 所示，正态分布是对称的，且具有下列特性.

(1) 渐近性 $\lim\limits_{x\to\infty}f(x)=0$；

(2) σ 原则

$$P\{\mu-\sigma<x<\mu+\sigma\}=68.26\%$$
$$P\{\mu-2\sigma<x<\mu+2\sigma\}=95.44\%$$
$$P\{\mu-3\sigma<x<\mu+3\sigma\}=99.74\%$$

这说明，x 落在区间$(\mu-3\sigma, \mu+3\sigma)$内几乎是肯定的. 这就是人们所说的“$3\sigma$”原则.

在自然和社会现象中，我们观察的指标 X 是由大量相互独立的随机变量之和生成的，它们往往服从或近似服从正态分布. 比如：在指定时刻，一个城市的耗电量是大量用户耗电量的总和；一个物理试验的测量误差是由许多观察不到的微小误差合成的. 这就是为什么正态随机变量占有重要地位的一个原因.

1.3.3 随机变量的数字特征

(1) 随机变量的数学期望（均值）

设离散型随机变量 X 的分布为 $P\{X=x_k\}=p_k$（$k=1, 2, \cdots$），X 取值的中心

$$E(X)=\sum_{k=1}^{\infty}x_kp_k$$

称为离散型随机变量的 X 数学期望.

设连续型随机变量 X 的概率密度函数为 $f(x)$，X 取值的中心

$$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$$

称为连续型随机变量的数学期望.

(2) 随机变量的方差

$D(X)=\mathrm{var}(X)=E\{[X-E(X)]^2\}$称为随机变量 X 的方差；$\sigma(X)=\sqrt{D(X)}$ 称为随机变量 X 的标准差或均方差. 它刻画了 X 取值的分散程度.

(3) 协方差与相关系数

我们知道，当 X，Y 相互独立时，有

$$E\{(X-E(X))(Y-E(Y))\}=E(X-E(X))\cdot E(Y-E(Y))=0$$

因此，$E\{(X-E(X))(Y-E(Y))\}\neq 0$，说明 X，Y 不相互独立，即相关. 关于 X，Y 相关性的度量有定义：称

$$\mathrm{cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

为 X，Y 的协方差；

$$\rho_{XY}=\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为 X 与 Y 的相关系数.

显然，$|\rho_{XY}|\leqslant 1$. 特别地，当 $|\rho_{XY}|=1$ 时，存在常数 a，b，使 $P\{Y=a+bX\}=1$.

几个重要随机变量的均值和方差如表 1－1 所示. 表中结果说明，数字特征决定了随机变量的分布.

表 1－1　重要随机变量的均值和方差

随机变量	分布模型	数学期望	方　差
正态分布	$N(\mu, \sigma^2)$	μ	σ^2
均匀分布	$U(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
二项分布	$b(n, p)$	np	$np(1-p)$
泊松分布	$\pi(\lambda)$	λ	λ

1.3.4 统计量及其抽样分布

1. 常用的统计量

设 X_1，X_2，…，X_n 是来自总体 X 的样本，相应观察值为 x_1，x_2，…，x_n，根据实际统计需要，常采用如下统计量进行参数估计和检验，如表 1－2 所示.

表 1－2　常 用 的 统 计 量

名　称	统 计 量	观 察 值
样本均值	$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$	$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$
样本方差	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$	$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$