简明线性代数讲义

【前言】本讲义取材于编者 2003—2004 年编写的讲义。讲义内容虽历经十八年教学修订成稿，但错漏之处难免；请读者不吝赐教（联系方式：zijunyxq@163.com）！本次修订中未曾增添的内容如：线性规划的单纯形法、投入产出分析及层次分析法等，请读者参考有关教材，敬请谅解！

【序言—代数学的发展简述】

代数学的发展基础—算术，即为我们熟悉的自然数、正分数等的四则运算。代数学成为独立分支是从初等代数阶段开始的，其中心内容是方程理论，至十六世纪已基本完备。代数方程理论在初等代数中由一元一次方程向两个方向扩展：①

增加未知量的个数（二元、三元方程组）；②增加未知量的幂次（一元二次方程）。韦达曾经这样地描述过“算术”与“代数”：所谓“算术”，即仅研究关于具体数的计算方法；所谓“代数”，即是研究关于事物的类或形式的运算方法—字母表示数的思想方法是代数学发展史上的一个重大转折。代数学的深化阶段即是高等代数阶段。十七世纪下半叶，从研究线性方程组的解出发，在莱布尼茨、凯莱等人的努力下，建立了以行列式、矩阵和线性方程组为主要内容的线性代数，标志着高等代数理论体系的建立。由于计算机的飞速发展与广泛应用，许多实际问题可以通过离散化的数值计算加以解决；作为处理离散问题的线性代数，已成为科研与设计等的必备数学基础。代数学的抽象化阶段—近世代数（抽象代数）产生于十九世纪，其研究各种抽象的合理化的代数系统，包括群论、环论、线性代数等许多分支。一般认为，其形成的时间为 **1926** 年；从此代数学的研究对象由代数方程根的计算与分布，进入到研究数字、文字和更一般元素的代数运算规律和各种代数结构。

第一章　行列式（determinant）

行列式源于求解线性方程组，其基础内容于 **19** 世纪由 **Cauchy** 所奠定；行列式的应用主要体现在：①求解线性方程组；②求矩阵的秩；③判断向量的相关性；④求矩阵的特征值等等。目前，其理论也早已超出求解线性方程组的范围，而广泛应用于经济学、力学、工程数学等其他领域。

§1　行列式的定义

对于二元方程组$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases}$，当$a_{11}a_{22} - a_{12}a_{21} \neq 0$时，由消元法可求出$x_1 = \dfrac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{12}a_{21}}$，$x_2 = \dfrac{b_2a_{11} - b_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}}$；注意到：解的分子、分母均是方程组中未知量的系数及常数项的 **6** 个元素中 **4** 个元素的二次齐次多项式，其均由两项组成，而每一项都是不同行、不同列的两个元素的乘积！

若记$D_2 = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$，则方程组的解可写为：

$x_1=\dfrac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$，$x_2=\dfrac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$；这里将上述$D_2$中$a,b,c,d$排成的两行两列定义为$ad-bc$的符号$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$称为二阶行列式。三元方程组的情形类似，也可通过定义所谓的“三阶行列式”，来描述方程组的解；这里，不再累述！

从上述二、三阶行列式对于求解二、三元方程组的作用的得到启发，能否“引进”“n阶行列式”来讨论n元方程组的求解；我们先来分析二、三阶行列式D_2,D_3的特点：

①D_2,D_3均是一个数，它们分别是$2!,3!$项的代数和，而每一项又是来自不同行、不同列的2个因子和3个因子的乘积；②带正号的项与带负号的项各占一半。

若规定：每一项的各因子a_{ij}按这样的顺序写出—第一个下标排成自然序；则易见：对于D_2，带正号的项第二个下标排列为12，带负号的第二个下标的排列为21，与自然序12颠倒；对于D_3，带正号的项第二个下标的排列分别是123、231和312，带负号的项则为321、213和132；它们颠倒的个数分别为偶数和奇数；从而，颠倒个数的奇偶性决定该项的正负性！

【**排列（Arrangement）**】由$1,2,\cdots,n$这n个数（码）组成的任一个有序数组$i_1i_2\cdots i_n$称为$1,2,\cdots,n$的一个n阶（级）排列。

显然，n个数码$1,2,\cdots,n$的不同排列有$n!$种！一般地，在一个排列里，如果一个较大的数排在一个较小的数之前，则称这两个数构成一个逆（反）序；称一个排列中出现的逆序总数为该排列的逆序数，记之为$N(i_1i_2\cdots i_n)$或$\tau(i_1i_2\cdots i_n)$。

给定$1,2,\cdots,n$的一个排列，可如下计算逆序数：先看有多少个比1大的数排在1之前，设为m_1个；再看有多少个比2大的数排在2之前，设为m_2个；依此类推，最后看有多少个比$n-1$大的数排在$n-1$之前，设为m_{n-1}个；从而，该排列的逆序数为$\sum\limits_{i=1}^{n-1}m_i$个。

【**例1.1.1**】**1**） 求排列523164879的逆序数；**2**）在自然数$1,2,\cdots,n$

的所有排列中哪个排列的逆序数最大？**3**）自然数$1,2,\cdots,n$的任一排列$i_1i_2\cdots i_n$的逆序数与正序数之和为多少；若$N\left(i_1i_2\cdots i_n\right)=k$，则$N\left(i_ni_{n-1}\cdots i_1\right)$为多少？

所谓偶（奇）排列，即逆序数为偶（奇）数的排列。

【定义】设有n个数码组成的一个排列，若将排列里的任意两个数码i,j交换一次，而其余数码保持不变，则得到一个新的排列；对于排列施行的这样一个变换，称之为一个“对换”，常用符号(i,j)表示。

【定理】排列经一次对换后奇偶性改变；任一个排列$i_1i_2\cdots i_n$总可由排列$12\cdots n$经一系列对换得到，且对换的次数与$N\left(i_1i_2\cdots i_n\right)$奇偶性相同。

【定理】$n\geq 2$时，n个数码的奇、偶排列的个数相等！

已经定义过$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=a_{11}a_{22}-a_{12}a_{21}$；这里，$a_{ij}$称为行列式的元素，并约定：在一个行列式里，称横排为行（**row**），纵排为列（**column**），元素a_{ij}的第一个下标表示元素所在行的序数，称为行标，第二个下标表示元素所在列的序数，称为列标。

【行列式的逆序定义】用符号$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$表示的$n$阶行列式指的是$n!$项的代数和，这些项是一切可能取自不同行、不同列上$n$个元素的乘积$a_{1j_1}a_{2j_2}\cdots a_{nj_n}$，每一项的符号为$(-1)^{N\left(j_1j_2\cdots j_n\right)}$，即：若排列为偶（奇）排列，该项符号为正（负），记作：

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}=\left|\left(a_{ij}\right)_{n\times n}\right|=\sum_{\left(j_1j_2\cdots j_n\right)}(-1)^{N\left(j_1j_2\cdots j_n\right)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}=\sum_{\left(i_1i_2\cdots i_n\right)}(-1)^{N\left(i_1i_2\cdots i_n\right)}a_{i_1,1}a_{i_2,2}\cdots a_{i_n,n}$$

$$=\sum_{\left(\begin{smallmatrix} i_1i_2\cdots i_n \\ j_1j_2\cdots j_n \end{smallmatrix}\right)}(-1)^{N\left(i_1i_2\cdots i_n\right)+N\left(j_1j_2\cdots j_n\right)}a_{i_1j_1}a_{i_2j_2}\cdots a_{i_nj_n}$$，这里，$\left(j_1j_2\cdots j_n\right)$表示求和取遍$1,2,\cdots,n$的所有可能排列。

一个n阶行列式正是二、三阶行列式的推广；当$n=1$时，一阶行列式$\left|a_{11}\right|$即数a_{11}。

【例 1.1.2】1）试求$D_4=\begin{vmatrix} a & 0 & 0 & b \\ 0 & c & d & 0 \\ 0 & e & f & 0 \\ g & 0 & 0 & h \end{vmatrix}$；2）求下列$f(x)$中$x^4,x^3$的系数及$\dfrac{d^3 f(x)}{dx^3}$，这里，$f(x)=\begin{vmatrix} 5x & 1 & 2 & 3 \\ 2 & 1 & x & 3 \\ x & x & 2 & 3 \\ 1 & 2 & 1 & -x \end{vmatrix}$；

【例 1.1.3】若行列式$D_n=\left|\left(a_{ij}\right)_{n\times n}\right|$的每个元素$a_{ij}$同乘以$c^{i-j}\,(c\neq 0)$，则行列式$\left|\left(c^{i-j}a_{ij}\right)_{n\times n}\right|$的值如何变化？；

对于含零元素较多的行列式可通过定义计算，因为行列式各项中若有一个因子为零，该项即为零；特别地，若n阶行列式中零元素的个数多于n^2-n，行列式必为零！

【例 1.1.4】试求下列行列式：

$$\begin{vmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & 0 & a_{23} & 0 & 0 \\ 0 & 0 & 0 & a_{34} & 0 \\ 0 & 0 & 0 & 0 & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{vmatrix},\begin{vmatrix} 0 & a_{12} & a_{13} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ 0 & a_{42} & a_{43} & 0 & 0 \\ 0 & a_{52} & a_{53} & 0 & 0 \end{vmatrix};$$

由行列式的逆序定义，

①上（下）三角行列式等于主对角线上元素的乘积，即：

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix}=\begin{vmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}=a_{11}a_{22}\cdots a_{nn};$$

②对角行列式也等于主对角线上元素的乘积；

③下反三角、反对角、上反三角行列式：

$$\begin{vmatrix} 0 & \cdots & 0 & a_{1n} \\ 0 & \cdots & a_{2,n-1} & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{n,n-1} & a_{nn} \end{vmatrix}=\begin{vmatrix} 0 & \cdots & 0 & a_{1n} \\ 0 & \cdots & a_{2,n-1} & 0 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & 0 & 0 \end{vmatrix}=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & 0 & \cdots & 0 & 0 \end{vmatrix}=(-1)^{\frac{n(n-1)}{2}}a_{1n}a_{2,n-1}\cdots a_{n1}$$

§2 行列式的性质

理论上，总可由行列式的（逆序）定义求出任一n阶行列式，但当阶数n较大时，这种做法并不太可行；如：n阶行列式有$n!$项，每一项都是n个元素的乘积，这样需作$(n-1)n!$次

乘法运算，从而需要讨论行列式的基本性质来简化计算！

【行列式的性质】①设行列式$D=\left|\left(a_{ij}\right)_{n\times n}\right|=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$，若把$D$的行（列）变成列（行），得到一个新的行列式$D^{\mathrm{T}}\left(D'\right)=\begin{vmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{vmatrix}$，称为$D$的转置行列式；行列式与它的转置（**transpose**）行列式相等；由①，行列式中对行成立的性质对列均成立，从而行与列地位均等；

②交换行列式的任意两行（列），行列式改变符号；由②，即有：行列式对行（列）有交错性，即行列式中若有两行（列）相同，则行列式为零；③（规范性）$\det I=\begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix}=1$；④把行列式的某一行（列）的所有元素同乘以一个数，等于以数去乘这个行列式；从而，由④，即有：行列式中某一行（列）中所有元素的公因子可以提到行列式符号外，且若行列式中有两行（列）对应元素成正比，则行列式为零；⑤（多线性）设行列式中第i行所有元素可表示成$b_{ij}+c_{ij}$之和，$j=1,2,\cdots,n$；则有：

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{i1}+c_{i1} & b_{i2}+c_{i2} & \cdots & b_{in}+c_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{i1} & b_{i2} & \cdots & b_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{i1} & c_{i2} & \cdots & c_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

；⑥行列式中某一行（列）的元素乘以同一个数后加至另一行（列）的对应元素时，行列式不变。（课堂证明）

【例 **1.2.1**】**1**）已知$204,527,255$均能被17整除，试证：$D_3=\begin{vmatrix} 2 & 5 & 2 \\ 0 & 2 & 5 \\ 4 & 7 & 5 \end{vmatrix}$

能被17整除；**2）**计算n阶行列式$D_n=\begin{vmatrix}0&1&1&\cdots&1\\1&0&1&\cdots&1\\\cdots&\cdots&\cdots&\cdots&\cdots\\1&1&1&\cdots&0\end{vmatrix}$；

【例 1.2.2】已知$f(x)=\begin{vmatrix}x+1&x+2&x+3&2\\1&2x+2&7&5\\3x+1&3&3x+3&8\\4x+1&8&1&4x+4\end{vmatrix}$，则$f(x)$是多少次多项式？

【例 1.2.3】求下列行列式的和：$\sum\limits_{(j_1j_2\cdots j_n)}\left(\begin{vmatrix}a_{1j_1}&a_{1j_2}&\cdots&a_{1j_n}\\a_{2j_1}&a_{2j_2}&\cdots&a_{2j_n}\\\cdots\cdots&\cdots\cdots&\cdots\cdots&\cdots\cdots\\a_{nj_1}&a_{nj_2}&\cdots&a_{nj_n}\end{vmatrix}\right)$，求和取遍$1,2,\cdots,n$的所有排列$j_1j_2\cdots j_n$；这里，$\sum\limits_{(j_1j_2\cdots j_n)}$亦可写作：$\sum\limits_{j_1=1}^{n}\sum\limits_{j_2=1}^{n}\cdots\sum\limits_{j_n=1}^{n}$；

【例 1.2.4】试由拆项法求下列行列式：

1）$\begin{vmatrix}1+x&1&1&1\\1&1-x&1&1\\1&1&1+z&1\\1&1&1&1-z\end{vmatrix}$；**2）**$\begin{vmatrix}x+1&x&\cdots&x\\x&x+\frac{1}{2}&\cdots&x\\\cdots&\cdots&\cdots&\cdots\\x&x&\cdots&x+\frac{1}{n}\end{vmatrix}$；

3）$\begin{vmatrix}x&a_1&a_2&\cdots&a_n\\a_1&x&a_2&\cdots&a_n\\a_1&a_2&x&\cdots&a_n\\\cdots&\cdots&\cdots&\cdots&\cdots\\a_1&a_2&a_3&\cdots&x\end{vmatrix}$，$\begin{vmatrix}1+x_1y_1&1+x_1y_2&\cdots&1+x_1y_n\\1+x_2y_1&1+x_2y_2&\cdots&1+x_2y_n\\\cdots&\cdots&\cdots&\cdots\\1+x_ny_1&1+x_ny_2&\cdots&1+x_ny_n\end{vmatrix}$；

【例 1.2.5】1）设$f(x)=\begin{vmatrix}1&x-1&2x-1\\1&x-2&3x-2\\1&x-3&4x-3\end{vmatrix}$，试证：$\exists\xi\in(0,1)$，使得$f'(\xi)=0$；**2）**对于奇数阶的行列式$\left|(a_{ij})_{n\times n}\right|$，若$\forall i,j,a_{ij}=-a_{ji}$，则$\left|(a_{ij})_{n\times n}\right|=0$；**3）**试证：

$$\frac{d}{dt}\left(\begin{vmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1}(t) & a_{i2}(t) & \cdots & a_{in}(t) \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{vmatrix}\right) = \sum_{j=1}^{n}\left(\begin{vmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{da_{j1}(t)}{dt} & \frac{da_{j2}(t)}{dt} & \cdots & \frac{da_{jn}(t)}{dt} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{vmatrix}\right)$$，并由此求极限 $\lim\limits_{x\to 0}\dfrac{\begin{vmatrix} x^3 & x^2 & 1 \\ 3 & 2 & 1 \\ x & \sin x & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 3 \\ \sin x & \cos x & 1 \\ 0 & 1 & 1 \end{vmatrix}}$。

§3 行列式的按行（列）展开与 Laplace 展开

由已知的二、三阶行列式的定义，容易得到：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

$+a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = \cdots\cdots$；从而，三阶行列式的计算至多可归结为三个二阶行列式的计算，故而实现了行列式的降阶计算！

【定义】在 n 阶行列式中任意取定 k 行和 k 列，位于这些行、列相交处的元素所构成的 k 阶行列式称为行列式的一个 k 阶子式；$n(n>1)$ 阶行列式 $D=\left|(a_{ij})_{n\times n}\right|$ 的某一元素 a_{ij} 的余子式 M_{ij} 指的是在 D 中划去 a_{ij} 所在的行和列后所余下的 $n-1$ 阶子式；称 $(-1)^{i+j}M_{ij}$ 为 a_{ij} 的代数余子式，记作：$A_{ij}=(-1)^{i+j}M_{ij}$。

【定理】①若在一个 n 阶行列式 $D=\left|(a_{ij})_{n\times n}\right|$ 中，第 i 行（第 j 列）元素除 a_{ij} 外都是零，则该行列式等于 a_{ij} 与它的代数余子式 A_{ij} 的乘积，即：$D=a_{ij}A_{ij}$；

②行列式 $D=\left|(a_{ij})_{n\times n}\right|$ 等于它任意一行（列）的所有元素与它们对应的代数余子式乘积的和，即：行列式有依行（列）的展开式：$D=a_{i1}A_{i1}+a_{i2}A_{i2}+\cdots+a_{in}A_{in}=\sum\limits_{j=1}^{n}a_{ij}A_{ij}, i=1,2,\cdots,n$ 或

$D=a_{1j}A_{1j}+a_{2j}A_{2j}+\cdots+a_{nj}A_{nj}=\sum\limits_{i=1}^{n}a_{ij}A_{ij}, j=1,2,\cdots,n$；

③行列式$D=\left|\left(a_{ij}\right)_{n\times n}\right|$的某一行（列）的元素与另外一行（列）对应元素代数余子式乘积的和等于零，也即：

$a_{i1}A_{j1}+a_{i2}A_{j2}+\cdots+a_{in}A_{jn}=\sum_{k=1}^{n}a_{ik}A_{jk}=0,i\neq j$　　或

$a_{1i}A_{1j}+a_{2i}A_{2j}+\cdots+a_{ni}A_{nj}=\sum_{k=1}^{n}a_{ki}A_{kj}=0,i\neq j$。

【例 1.3.1】求多项式$f\left(x\right)=\begin{vmatrix} x & 0 & x \\ 2 & 3 & x \\ 0 & -1 & 1 \end{vmatrix}=ax^2+bx+c$的各项系数之和$a+b+c$；

【例 1.3.2】1）已知五阶行列式$D_5=\begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 2 & 2 & 1 & 1 \\ 3 & 1 & 2 & 4 & 5 \\ 1 & 1 & 1 & 2 & 2 \\ 4 & 3 & 1 & 5 & 0 \end{vmatrix}=27$，试求：$A_{44}+A_{45}$与$A_{41}+A_{42}+A_{43}$；2）设$D_n=\begin{vmatrix} 1 & 3 & 5 & \cdots & 2n-1 \\ 1 & 2 & 0 & \cdots & 0 \\ 1 & 0 & 3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & \cdots & n \end{vmatrix}$，求其代数余子式$A_{1i}$之和$A_{11}+A_{12}+\cdots+A_{1n}$；

【例 1.3.3】1）设有n阶行列式$D=\left|\left(a_{ij}\right)_{n\times n}\right|$，行列式

$D^*=\begin{vmatrix} a_{11}+x & a_{12}+x & \cdots & a_{1n}+x \\ a_{21}+x & a_{22}+x & \cdots & a_{2n}+x \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}+x & a_{n2}+x & \cdots & a_{nn}+x \end{vmatrix}$，试证：$D^*=D+x\cdot\sum_{i,j=1}^{n}A_{ij}$，$A_{ij}$为$a_{ij}$在$D$中的代数余子式；

2）试证：$\begin{vmatrix} a_{11}-a_{12} & a_{12}-a_{13} & \cdots & a_{1,n-1}-a_{1n} & 1 \\ a_{21}-a_{22} & a_{22}-a_{23} & \cdots & a_{2,n-1}-a_{2n} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1}-a_{n2} & a_{n2}-a_{n3} & \cdots & a_{n,n-1}-a_{nn} & 1 \end{vmatrix}=\sum_{i,j=1}^{n}A_{ij}$，这里$A_{ij}$为$a_{ij}$在$D$中的代数余子式。

【定义】取行列式$D=\left|\left(a_{ij}\right)_{n\times n}\right|$中第$i_1,i_2,\cdots,i_k$行及第$j_1,j_2,\cdots,j_k$列交点处的元素，按原来$D$中的相对位置构成一个$k$阶行列式，

称之为D的一个k阶子式，记之为：$A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$；将$D$中剩下的元素按原来的相对位置构成一个$n-k$阶行列式，称之为上述$k$阶子式的余子式，记之为：$M\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$；令

$p=i_1+i_2+\cdots+i_k$，$q=j_1+j_2+\cdots+j_k$，记：

$A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}=(-1)^{p+q}\cdot M\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$；称之为上述$k$阶子式的代数余子式。

【**Laplace 定理**】设$D=\left|(a_{ij})_{n\times n}\right|$是$n$阶行列式，在$D$中任取$k$行（列），那么含于这$k$行（列）的全部$k$阶子式与它们所对应的代数余子式的乘积之和等于$D$；即若取定$k$行：

$1\le i_1<i_2<\cdots<i_k\le n$，则$D=\sum\limits_{1\le j_1<j_2<\cdots<j_k\le n} A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}\cdot A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$；

若取定k列：$1\le j_1<j_2<\cdots<j_k\le n$，则

$$D=\sum_{1\le i_1<i_2<\cdots<i_k\le n} A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}\cdot A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}。$$

不妨先设$i_1i_2\cdots i_k=12\cdots k$；

$$D=\sum_{(t_1t_2\cdots t_n)}(-1)^{N(t_1t_2\cdots t_n)}a_{1t_1}a_{2t_2}\cdots a_{nt_n}=\sum_{1\le j_1<j_2<\cdots<j_k\le n}\ \sum_{\substack{t_1,t_2,\cdots,t_k\text{是}\\ j_1,j_2,\cdots,j_k\text{的排列}}}\ \sum_{\substack{t_{k+1},\cdots,t_n\text{是}\\ j_{k+1},\cdots,j_n\text{的排列}}}(-1)^{N(t_1t_2\cdots t_n)}a_{1t_1}a_{2t_2}\cdots a_{nt_n}$$

$$=\sum_{1\le j_1<j_2<\cdots<j_k\le n}\left[\left(\sum_{\substack{t_1,t_2,\cdots,t_k\text{是}\\ j_1,j_2,\cdots,j_k\text{的排列}}}(-1)^{\tau\begin{pmatrix} j_1 & j_2 & \cdots & j_k \\ t_1 & t_2 & \cdots & t_k \end{pmatrix}}a_{1t_1}a_{2t_2}\cdots a_{kt_k}\right)\cdot\left(\sum_{\substack{t_{k+1},\cdots,t_n\text{是}\\ j_{k+1},\cdots,j_n\text{的排列}}}(-1)^{\tau\begin{pmatrix} j_{k+1} & \cdots & j_n \\ t_{k+1} & \cdots & t_n \end{pmatrix}}a_{k+1,t_{k+1}}\cdots a_{nt_n}\right)(-1)^u\right]$$

$$\underline{\underline{u=1+\cdots+k+j_1+\cdots+j_k}}\sum_{1\le j_1<j_2<\cdots<j_k\le n} A\begin{pmatrix} 1 & 2 & \cdots & k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}\cdot A\begin{pmatrix} 1 & 2 & \cdots & k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$$，这里，$\tau\begin{pmatrix} j_1 & j_2 & \cdots & j_k \\ t_1 & t_2 & \cdots & t_k \end{pmatrix}$表示把$j_1j_2\cdots j_k$变为$t_1t_2\cdots t_k$所施行的对换次数。

【**例 1.3.4**】试由 **Laplace** 定理计算如下的行列式：

$$D_5=\begin{vmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & x_1 & x_2 & x_3 & x_4 \\ 0 & x_1^2 & x_2^2 & x_3^2 & x_4^2 \end{vmatrix}，\quad D_{2n}=\begin{vmatrix} a_n & & & & & b_n \\ & \ddots & & & \iddots & \\ & & a_1 & b_1 & & \\ & & c_1 & d_1 & & \\ & \iddots & & & \ddots & \\ c_n & & & & & d_n \end{vmatrix}；$$

【范德蒙（Vandermonde）行列式】n**阶范德蒙行列式的形式和结果为：**

$$V_n\left(x_1,x_2,\cdots,x_n\right)=\begin{vmatrix}1&1&\cdots&1\\x_1&x_2&\cdots&x_n\\x_1^2&x_2^2&\cdots&x_n^2\\\cdots&\cdots&\cdots&\cdots\\x_1^{n-1}&x_2^{n-1}&\cdots&x_n^{n-1}\end{vmatrix}=\begin{vmatrix}1&x_1&x_1^2&\cdots&x_1^{n-1}\\1&x_2&x_2^2&\cdots&x_2^{n-1}\\\cdots&\cdots&\cdots&\cdots&\cdots\\1&x_n&x_n^2&\cdots&x_n^{n-1}\end{vmatrix}=\prod_{1\le j<i\le n}\left(x_i-x_j\right)\text{；}$$

分析其特点：n**阶行列式中的每行（列）都是同一个数的不同方幂，且自上（左）而下（右）方幂次数由零逐一递增至**$n-1$**。常可通过范德蒙行列式去计算其他一些行列式！**

【例 1.3.5】试求下列行列式：1）$\begin{vmatrix}1&1&\cdots&1\\2&2^2&\cdots&2^n\\3&3^2&\cdots&3^n\\\cdots&\cdots&\cdots&\cdots\\n&n^2&\cdots&n^n\end{vmatrix}$**；2）**

$\begin{vmatrix}a^n&(a-1)^n&\cdots&(a-n)^n\\a^{n-1}&(a-1)^{n-1}&\cdots&(a-n)^{n-1}\\\cdots&\cdots&\cdots&\cdots\\a&a-1&\cdots&a-n\\1&1&\cdots&1\end{vmatrix}$**；3）**$\begin{vmatrix}a_1^n&a_1^{n-1}b_1&\cdots&a_1b_1^{n-1}&b_1^n\\a_2^n&a_2^{n-1}b_2&\cdots&a_2b_2^{n-1}&b_2^n\\\cdots&\cdots&\cdots&\cdots&\cdots\\a_{n+1}^n&a_{n+1}^{n-1}b_{n+1}&\cdots&a_{n+1}b_{n+1}^{n-1}&b_{n+1}^n\end{vmatrix}$**；**

【例 1.3.6】试求 1）$\begin{vmatrix}\dfrac{x_1}{x_1-1}&\dfrac{x_2}{x_2-1}&\cdots&\dfrac{x_n}{x_n-1}\\x_1&x_2&\cdots&x_n\\x_1^2&x_2^2&\cdots&x_n^2\\\cdots&\cdots&\cdots&\cdots\\x_1^{n-1}&x_2^{n-1}&\cdots&x_n^{n-1}\end{vmatrix}$**，**

2）$\begin{vmatrix}1&1&1&1\\1+\sin\varphi_1&1+\sin\varphi_2&1+\sin\varphi_3&1+\sin\varphi_4\\\sin\varphi_1+\sin^2\varphi_1&\sin\varphi_2+\sin^2\varphi_2&\sin\varphi_3+\sin^2\varphi_3&\sin\varphi_4+\sin^2\varphi_4\\\sin^2\varphi_1+\sin^3\varphi_1&\sin^2\varphi_2+\sin^3\varphi_2&\sin^2\varphi_3+\sin^3\varphi_3&\sin^2\varphi_4+\sin^3\varphi_4\end{vmatrix}$**；**

【例 1.3.7】利用加（镶）边法求如下的超范德蒙行列式：

$\begin{vmatrix}1&1&1&1\\a&b&c&d\\a^2&b^2&c^2&d^2\\a^4&b^4&c^4&d^4\end{vmatrix}$**；**

§4 行列式的计算

现将行列式的计算方法总结如下：

1）利用行列式的逆序定义；

2）利用行列式的性质，比如：析因子法—运用行列式的性质找出全部因子，最后确定最高次项的系数；

①$\begin{vmatrix} 1 & 1 & 2 & 3 \\ 1 & 2-x^2 & 2 & 3 \\ 2 & 3 & 1 & 5 \\ 2 & 3 & 1 & 9-x^2 \end{vmatrix}$；②$\begin{vmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{vmatrix}$；

3）化行列式为三角形行列式：

①箭形行列式：$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 0 & \cdots & 0 \\ 1 & 0 & 3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & \cdots & n \end{vmatrix}$；②可化为箭形的行列式：

$\begin{vmatrix} x_1 & a_2 & a_3 & \cdots & a_n \\ a_1 & x_2 & a_3 & \cdots & a_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & a_3 & \cdots & x_n \end{vmatrix}$；

4）行和、列和相等的行列式：$\begin{vmatrix} x & a & \cdots & a \\ a & x & \cdots & a \\ \cdots & \cdots & \cdots & \cdots \\ a & a & \cdots & x \end{vmatrix}$；

5）降阶法；

6）升阶法（又称加边法或镶边法）—在原行列式的基础上增加一行（列），且保持原行列式不变的情况下，计算行列式的一种方法。一般地，利用此法计算的行列式应满足各行（列）含有共同元素的特点，且化简后常变成箭形行列式！

①$\begin{vmatrix} a_1+b_1 & a_2 & \cdots & a_n \\ a_1 & a_2+b_2 & \cdots & a_n \\ \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & \cdots & a_n+b_n \end{vmatrix}$；②$\begin{vmatrix} x_1^2+1 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2^2+1 & \cdots & x_2x_n \\ \cdots & \cdots & \cdots & \cdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2+1 \end{vmatrix}$；

7）递推法：对于n阶行列式D_n，若能找出D_n与D_{n-1}或D_{n-2}之间的一种关系（即递推公式），然后由此公式求出通项公式D_n，即为递推法。

①$D_n=\begin{vmatrix} a+b & ab & & & \\ 1 & a+b & ab & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & a+b & ab \\ & & & 1 & a+b \end{vmatrix}$；②$D_n=\begin{vmatrix} x & -1 & 0 & \cdots & 0 & 0 \\ 0 & x & -1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x & -1 \\ a_n & a_{n-1} & a_{n-2} & \cdots & a_2 & a_1 \end{vmatrix}$；

8）第一（二）数学归纳法；如：试证：

$D_n=\begin{vmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{vmatrix}=n+1$；再如：

试分别由递推法和归纳法求如下行列式：

$\begin{vmatrix} 2\cos\theta & 1 & & & \\ 1 & 2\cos\theta & \ddots & & \\ & \ddots & \ddots & & \\ & & \ddots & & 1 \\ & & & 1 & 2\cos\theta \end{vmatrix}$，$\begin{vmatrix} \cos\theta & 1 & & & \\ 1 & 2\cos\theta & \ddots & & \\ & \ddots & \ddots & & \\ & & \ddots & & 1 \\ & & & 1 & 2\cos\theta \end{vmatrix}$；

【附录——第一、二数学归纳法】

由特殊情况而推出一般规律，即所谓“归纳法”；对于一个与自然数有关的命题，当它对有限个自然数成立时，不一定对所有的自然数成立，我们不可能对每个自然数逐个加以验证，判断这类命题的正确与否，就可用“数学归纳法”！

数学归纳法的理论依据是自然数集的一个最基本性质—最小数原理：自然数集N的任一非空子集S都有最小数，即$\exists m\in S$，使得$\forall s\in S$，都有$m\le s$。

简证：由于$\Phi\ne S\subset N$，故$\exists n\in S$，从1到n有n个自然数；于是S中不超过n的自然数最多有n个，在这n个自然数中必有一个最小数m。

【注】最小数原理不仅对自然数集成立，对于整数集的如下子集$Z_c=\{x|x\in Z,x\ge c\}$（c为一给定整数）也成立；但其并非对任意数集都成立；比如：Z就没有最小数，$A=\{x|x\in Q,0<x<1\}$也是如此！

【第一数学归纳法】设$P(n)$是一个与自然数有关的命题，如果$P(n)$满足：①当$n=1$时，$P(n)$成立；②假设$n=k$时，$P(n)$成

立；则当$n=k+1$时，$P(n)$也成立；则$P(n)$对一切自然数成立！

反证法：假设命题$P(n)$并不对一切自然数成立，令$S=\left\{s\,\middle|\,s\in N,\text{命题对}s\text{不成立}\right\}$，故$\Phi\neq S\subset N$；由最小数原理，$S$中有最小数$m$，$m\neq 1$；从而，$m-1\in N$，且$m-1\notin S$；于是命题对$m-1$成立，从而对$m$亦成立，矛盾！

若仅由命题$P(k)$成立还导不出$P(k+1)$成立，则需要更强的假设条件，从而有，

【第二数学归纳法】设$P(n)$是一个与自然数有关的命题，如果$P(n)$满足：①当$n=1$时，$P(n)$成立；②假设$P(n)$对一切$n\leq k$的自然数成立；那么$P(n)$对$n=k+1$也成立；则$P(n)$对一切自然数成立！

【注】一般地，能用第一数学归纳法证明的命题都可用第二数学归纳法来论证，但反之则不一定；而且使得命题成立的最小自然数不必非得是 **1**！

9）利用 **Vandermonde** 行列式的结果等等。

§5 Cramer 法则

Cramer 法则是 **Cramer** 于 **1750** 年发现的，实际上，最早发现的是 **Leibnitz**，其比 **Cramer** 要早 **50** 年，但该法则并没有完全解决线性方程组的求解问题。

【定理（**Cramer** 法则）】设n个变量（元）$x_1,x_2,\cdots,x_n$的线性方程组$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=b_n\end{cases}$的系数行列式$D=\left|\left(a_{ij}\right)_{n\times n}\right|\neq 0$，则此方程组有非零解$x_j=\dfrac{D_j}{D},j=1,2,\cdots,n$，其中$D_j$是把$D$的第$j$列换成$\left(b_1,b_2,\cdots,b_n\right)^{\mathrm{T}}$所得的行列式；

易见，$D_j=\sum\limits_{k=1}^{n}b_kA_{kj}$，$a_{i1}D_1+a_{i2}D_2+\cdots+a_{in}D_n=\sum\limits_{j=1}^{n}a_{ij}D_j=\sum\limits_{j=1}^{n}a_{ij}\left(\sum\limits_{k=1}^{n}b_kA_{kj}\right)$
$=\sum\limits_{k=1}^{n}b_k\left(\sum\limits_{j=1}^{n}a_{ij}A_{kj}\right)=\sum\limits_{k=1}^{n}b_k\delta_{ik}D=b_iD$；另外，若$\left(x_1,x_2,\cdots,x_n\right)$与$\left(y_1,y_2,\cdots,y_n\right)$均为上述方程组的解，令$z_j=x_j-y_j$，则$\sum\limits_{j=1}^{n}a_{ij}z_j=0,i=1,2,\cdots,n$；从而

有 $0=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}a_{ij}z_j\right)A_{ik}=\sum_{j=1}^{n}z_j\left(\sum_{i=1}^{n}a_{ij}A_{ik}\right)=\sum_{j=1}^{n}z_j\delta_{jk}D=z_kD\Rightarrow z_k=0$。也可如下验证：

$$Dx_1=\begin{vmatrix} a_{11}x_1 & a_{12} & \cdots & a_{1n} \\ a_{21}x_1 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}x_1 & a_{n2} & \cdots & a_{nn} \end{vmatrix} \overset{c_j\times x_j+c_1,\ j=2,3,\cdots,n}{=\!=\!=\!=\!=\!=} \begin{vmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{vmatrix}=D_1;$$

类似有：$Dx_j=D_j, j=2,3,\cdots,n$；从而，$x_j=\dfrac{D_j}{D}, j=1,2,\cdots,n$。事实上，利用 $n+1$ 阶行列式

$$\begin{vmatrix} b_i & a_{i1} & a_{i2} & \cdots & a_{in} \\ b_1 & a_{11} & a_{12} & \cdots & a_{1n} \\ b_2 & a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_n & a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}=0, i=1,2,\cdots,n$$

也可推得 $\left(\dfrac{D_1}{D},\dfrac{D_2}{D},\cdots,\dfrac{D_n}{D}\right)$ 是上述线性方程组的解。

【例 1.5.1】试证：如果多项式 $f(x)=c_0+c_1x+\cdots+c_nx^n$ 对 $n+1$ 个不同的 x 值都等于零，则 $f(x)\equiv 0$；

【例 1.5.2】1）写出过三点 $(1,1,1),(2,3,-1),(3,-1,-1)$ 的平面方程；**2）**试求通过 xOy 平面上不在同一直线上的三点 $P_1(x_1,y_1)$，$P_2(x_2,y_2)$，$P_3(x_3,y_3)$ 的圆的方程；**3）**写出不在同一平面上的四点 $(1,1,1),(1,1,-1),(1,-1,1),(-1,0,0)$ 的球面方程，并写出其球心及半径。

第二章 矩阵（Martrix）

矩阵概念源于十八世纪，它也是由求解方程组的需要而逐步形成的，**1856** 到 **1857** 年，才由 **Cayley**（凯莱）提出“矩阵”这一术语而沿用至今；矩阵是用代数方法解决实际问题的基本工具，如今已广泛应用于自然科学、工程技术、社会科学特别是经济学等领域。

§1 矩阵的概念和性质

矩阵具有广泛的实际背景；比如物理背景：作用于物体的力，既有大小又有方向，必须用几个数来刻划；在工程技术领域，如水利、土木工程中大坝、桥梁的受力情况、建筑结构的受力状况等也不能仅由一个数来说明！在计算机与通

信领域，则矩阵的应用更是广泛；而在经济管理方面，也常常用统计表来描述公司、部门的经营状况。

一个n阶行列式从形式上看正是n^2个元素按一定规则排成的n行n列；实际上，常会遇到由若干个数排成行与列的矩形阵列；研究这类问题，常将这样一个阵列当成一整体考虑，即为矩阵。

【矩阵】由mn个数$a_{ij}(i=1,2,\cdots,m; j=1,2,\cdots,n)$排成的$m$行$n$列的如下矩形阵列：$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{matrix}$，称为$m$行$n$列矩阵，简称$m\times n$阵；常用大写英文字母表示，且为了紧凑、易分辨起见，往往用括弧将上述矩形阵列括起来，即为：$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$，有时简记为：$A=\left(a_{ij}\right)_{m\times n}$；称$a_{ij}$为矩阵$A$的元素，$i$称为行标，$j$称为列标。有序数组$\left(a_{i1},a_{i2},\cdots,a_{in}\right)$称为$A$的第$i$行（**row**），纵列数组$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$称为$A$的第$j$列（**column**）。

【注 1】1）$n\times n$矩阵也称为n阶方阵，称$a_{ii}(i=1,2,\cdots,n)$为其主对角线上元素；若矩阵A的元素全是实数（复数、非负数、零），称矩阵A是实矩阵（复矩阵、非负矩阵、零矩阵）；

2）若矩阵$A=\left(a_{ij}\right)_{m\times n}$只有一行，即：$m=1$，则称$A=\left(a_{11},a_{12},\cdots,a_{1n}\right)$为行矩阵或行向量；若矩阵$A$只有一列，即：$n=1$，则称$A=\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$为列矩阵或列向量；当$m=n=1$时$A=\left(a_{11}\right)=a_{11}$,，即为数$a_{11}$；

3）若方阵$A=\left(a_{ij}\right)_{n\times n}$除主对角线外其余元素均为零，称$A=\left(a_{ij}\right)_{n\times n}$

为对角（矩）阵，即有：$A=\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{mn} \end{pmatrix}$，简记为：

$A=diag\left\{a_{11},a_{22},\cdots,a_{nn}\right\}$；若 $a_{ii}=1,i=1,2,\cdots,n$，称对角（矩）阵 A 为 n 阶单位矩阵，常记为 E_n 或 E，亦可记为：$E_{n\times n},I_n,I_{n\times n},I$；若方阵 A 的主对角线以上（下）元素全为零，称 A 为下（上）三角矩阵，二者统称为三角形矩阵；

4）在矩阵中，若一行元素全为零，称该行为零行；若一行元素不全为零，称该行为非零行；在非零行中，从左往右数，第一个不为零的元素称为首非零元素；若矩阵 A 满足：（**a**）各非零行首非零元素分布在不同列；（**b**）若有零行，零行在矩阵的最下端，则称 A 为阶梯形矩阵（各行首非零元素依次后移的矩阵）；

5）由定义，所谓 n 阶方阵，即为 n^2 个元素组成的 n 行 n 列的矩形数表；所谓 n 阶行列式，即为 n^2 个元素根据行列式运算法则计算所得的一个数值。

由矩阵的定义，矩阵可以是各种各样的，为辨别两矩阵的不同，引入“矩阵相等”概念：设有矩阵 $A=\left(a_{ij}\right)_{m\times n}$，$B=\left(b_{ij}\right)_{k\times l}$，若其满足：**1**）$m=k,n=l$；**2**）$a_{ij}=b_{ij},i=1,2,\cdots,m;j=1,2,\cdots,n$；则称 A,B 相等，记为：$A=B$；比较：$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$和$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$。

矩阵是从大量的实际问题中抽象出来的数学概念，引进矩阵的概念当然不是研究一个孤立的矩阵，而是为了讨论它们之间的关系。由实际问题提出的矩阵之间存在着密切的关系，其中最主要的是它们之间可进行代数运算，矩阵代数运算的概念正是实际存在的客观事物相互关系的一种抽象（引例可参见教材【重视】）。

矩阵的基本运算如下：

1、 矩阵的加（减）法运算：设有矩阵 $A=\left(a_{ij}\right)_{m\times n}$，$B=\left(b_{ij}\right)_{m\times n}$，矩阵 $C=\left(a_{ij}+b_{ij}\right)_{m\times n}$ 称为 A 与 B 的和，记作：$C=A+B$；

【注 **2**】**1**）只有行列数对应相同的矩阵才能进行加法运算；**2**）定义 $A=\left(a_{ij}\right)_{m\times n}$ 的负矩阵 $-A=\left(-a_{ij}\right)_{m\times n}$，矩阵的减法运算可视

作加法运算的逆运算，故而仅有行列数对应相同的矩阵才能进行减法运算；若 $A=\left(a_{ij}\right)_{m\times n}$， $B=\left(b_{ij}\right)_{m\times n}$，则应有 $A-B=A+\left(-B\right)$ $=\left(a_{ij}\right)_{m\times n}+\left(-b_{ij}\right)_{m\times n}=\left(a_{ij}-b_{ij}\right)_{m\times n}$；

3）矩阵的加、减法满足如下如下运算律：

a）【交换律】$A+B=B+A$；**b**）【结合律】$\left(A+B\right)+C=A+\left(B+C\right)$；**c**） $A+O=O+A=A,O=\left(0\right)_{m\times n}$。

2、矩阵的数乘运算：已知数 k 与矩阵 $A=\left(a_{ij}\right)_{m\times n}$，用数 k 去乘矩阵 A 的每个元素，所得到的 m 行 n 列矩阵 $\left(ka_{ij}\right)_{m\times n}$ 称为数（纯量）k 与矩阵 A 的乘积，记作： $kA=\left(ka_{ij}\right)_{m\times n}$；

【注 **3**】**1**）不应混淆数与矩阵的乘积和行列式的性质，如：$2\begin{pmatrix}1&2\\3&4\end{pmatrix}=\begin{pmatrix}2&4\\6&8\end{pmatrix}$， $\left|\begin{pmatrix}2&4\\6&8\end{pmatrix}\right|=\begin{vmatrix}2&4\\6&8\end{vmatrix}=2^2\begin{vmatrix}1&2\\3&4\end{vmatrix}=2^2\left|\begin{pmatrix}1&2\\3&4\end{pmatrix}\right|$，推广至一般情形：若 k 为常数， $A=\left(a_{ij}\right)_{m\times n}$ 为 n 阶方阵，则有：$\left|kA\right|=\left|\left(ka_{ij}\right)_{n\times n}\right|=k^n\left|\left(a_{ij}\right)_{n\times n}\right|=k^n\left|A\right|$；**2**）矩阵的数乘运算有如下性质：$k,l$ 为常数， A,B 均为 $m\times n$ 阵，**a**）【数乘关于加法的分配律】$k\left(A+B\right)=kA+kB$；**b**）【加法关于数乘的分配律】$\left(k+l\right)A=kA+lA$；**c**）【数乘与乘法的结合律】$k\left(lA\right)=\left(kl\right)A$；**d**） $1A=A,0A=O_{m\times n}$。

【例 **2.1.1**】若有矩阵 $A=\begin{pmatrix}1&2&3\\4&0&7\end{pmatrix}$， $B=\begin{pmatrix}2&3&1\\5&7&1\end{pmatrix}$，且矩阵 X 满足矩阵方程 $2X-A=2B$，试求 X；

3、 矩阵的乘法运算（矩阵与矩阵的乘法运算）：设 $A=\left(a_{ij}\right)_{m\times s}$，$B=\left(b_{ij}\right)_{s\times n}$，其乘积定义为 $m\times n$ 矩阵 $C=AB=\left(c_{ij}\right)_{m\times n}$，其中 c_{ij} 是左乘矩阵 A 的第 i 行与右乘矩阵 B 的第 j 列对应元素乘积的和，即：$c_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{is}b_{sj}=\sum\limits_{k=1}^{s}a_{ik}b_{kj}$； 该乘法运算可形象地表示为：

$$\begin{pmatrix}\cdots&\cdots&\cdots&\cdots\\a_{i1}&a_{i2}&\cdots&a_{is}\\\cdots&\cdots&\cdots&\cdots\end{pmatrix}\begin{pmatrix}\vdots&b_{1j}&\vdots\\\vdots&b_{2j}&\vdots\\\vdots&\vdots&\vdots\\\vdots&b_{sj}&\vdots\end{pmatrix}=\begin{pmatrix}&\vdots&\\\cdots&c_{ij}&\cdots\\&\vdots&\end{pmatrix}。$$

【注 **4**】由定义，A,B 的乘积矩阵 C 的第 j 列是由 A 的诸行与 B 的第 j 列逐一相乘所得， C 的第 i 行是由 A 的第 i 行与 B 的诸列

逐一相乘所得；并且若A,B可相乘，左乘矩阵A的列数必等于右乘矩阵B的行数，且乘积矩阵的行（列）数等于左（右）乘矩阵A（B）的行（列）数。

【例 **2.1.2**】利用矩阵的乘法与矩阵相等将线性方程组

写成矩阵方程的形式$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1 \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m \end{cases}$；

【例 **2.1.3**】**1**）设$A=\left(1,\frac{1}{2},\frac{1}{3}\right)$，$B=\begin{pmatrix}1\\2\\3\end{pmatrix}$，试求$AB,BA$；**2**）设$A=\begin{pmatrix}1&1\\0&0\end{pmatrix}$，$B=\begin{pmatrix}0&1\\0&1\end{pmatrix}$，试求$AB,BA$；若$A=(1,2)$，$B=\begin{pmatrix}0&1\\0&1\end{pmatrix}$，结果又如何？

【注 **5**】**1**）矩阵的乘法运算不满足交换律；一则交换次序后，二矩阵不一定能相乘，再则是即使能相乘，结果也未必相等！**2**）对于数乘运算，若有$kA=O$，则必有$k=0$或$A=O$；而矩阵的乘法运算是有零因子的，即：$A\neq O$，$B\neq O$，但可能有$AB=O$，从而矩阵的乘法运算不满足消去律，其与交换律称为矩阵乘法区别于数的乘法的两个重要特点；

3）矩阵的乘法运算满足如下性质：设以下矩阵可相乘，

a）【结合律】$(AB)C=A(BC)$，可推广至任意有限个的情形；从而矩阵作乘法运算时，可任意地加上或去掉括号（证明）；**b**）【（乘法关于加法的）左分配律】$A(B+C)=AB+AC$；【（乘法关于加法的）右分配律】$(B+C)A=BA+CA$；**c**）【矩阵乘法对数乘的结合律】$\forall k$，$k(AB)=(kA)B=A(kB)$；

可以验证：$\forall A=(a_{ij})_{m\times n}$，有$I_mA_{m\times n}=A_{m\times n}I_n=A_{m\times n}$；若$m=n$，即有$I_nA=AI_n=A$；即：单位矩阵起着数字乘法中数“**1**”相同的作用（单位 **1**）；称$\lambda I_n=\begin{pmatrix}\lambda&&&\\&\lambda&&\\&&\ddots&\\&&&\lambda\end{pmatrix}$为纯量（数量）方阵，并且$\forall A=(a_{ij})_{m\times n}$，有$(\lambda I_m)A_{m\times n}=A_{m\times n}(\lambda I_n)=A_{m\times n}$；若$m=n$，即有$(\lambda I_n)A=A(\lambda I_n)=\lambda A$。

【例 2.1.4】若实矩阵 $A=\left(a_{ij}\right)_{n\times n}$，证明：$A=O$ 当且仅当 $A^{\mathrm{T}}A=O$；

【例 2.1.5】若 A,B 均为 n 阶上（下）三角阵，AB 则也是上（下）三角阵，亦可推广至有限个的情形。

以下通过一个例子来看看矩阵的乘法运算在实际中的应用！

考虑直角三角形 ΔMON，其顶点坐标为 $M\begin{pmatrix}1\\1\end{pmatrix}$，$O\begin{pmatrix}0\\0\end{pmatrix}$ 和 $N\begin{pmatrix}1\\0\end{pmatrix}$，

设矩阵 $A=\begin{pmatrix}0&1\\-1&0\end{pmatrix}$，$B=\begin{pmatrix}1&0\\0&\frac{1}{2}\end{pmatrix}$，即有：$A\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}1\\-1\end{pmatrix}$，$A\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$，

$A\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\-1\end{pmatrix}$，得到三个点 $M^{/}\begin{pmatrix}1\\-1\end{pmatrix}$，$O\begin{pmatrix}0\\0\end{pmatrix}$ 和 $N^{/}\begin{pmatrix}0\\-1\end{pmatrix}$；从而，用 A 左乘 ΔMON 各顶点的结果即是将三角形绕原点顺时针旋转 90°；

$B\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}1\\\frac{1}{2}\end{pmatrix}$，$B\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$，$B\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$，得到三个点 $M^{//}\begin{pmatrix}1\\\frac{1}{2}\end{pmatrix}$，$O\begin{pmatrix}0\\0\end{pmatrix}$ 和 $N^{//}\begin{pmatrix}1\\0\end{pmatrix}$；从而，用 B 左乘 ΔMON 各顶点的结果即是将该三角形 ON 边上的高压缩为原来的一半。进一步地；

①先压缩后旋转：由矩阵乘法结合律，先压缩后旋转相当于用 AB 左乘 ΔMON 各顶点坐标，$AB=\begin{pmatrix}0&\frac{1}{2}\\-1&0\end{pmatrix}$，$AB\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}\frac{1}{2}\\-1\end{pmatrix}$，

$AB\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$，$AB\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\-1\end{pmatrix}$，得到的三个点构成 ΔM_1ON_1（图示）；

②先旋转后压缩：同理，先旋转后压缩相当于用 BA 左乘 ΔMON 各顶点坐标，$BA=\begin{pmatrix}0&1\\-\frac{1}{2}&0\end{pmatrix}$，$BA\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}1\\-\frac{1}{2}\end{pmatrix}$，$BA\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$，$BA\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\-\frac{1}{2}\end{pmatrix}$，

得到的三个顶点构成 ΔM_2ON_2（图示）；由此，交换 A,B 相乘的次序，结果不尽相同；究其因，则是 $AB\neq BA$。

基于上述的讨论，可以看出：平面上的点绕原点旋转或平移，是可以通过矩阵的乘法实现的；更一般地，可用矩阵 $\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}$ 左乘平面上的任意点 $M=\begin{pmatrix}x\\y\end{pmatrix}$，相当于将点 M 绕原点 O 逆时针旋转 θ 角。在使用计算机时，数组的输入要用到矩阵形式，动画的设计则更是大量涉及矩阵的运算；这里不再累述，有兴趣的读者可参看《计算机图形学》！

【例 2.1.6】若 n 阶方阵 A,B 满足 $AB=BA$，则称 A,B 是可交换的，试求所有与 $A=\begin{pmatrix}1&0\\2&1\end{pmatrix}$ 可交换的矩阵；

由上例，单位矩阵、数量方阵与任何同阶方阵均是可交换的。

由矩阵的乘法运算可定义同一方阵的乘幂（方幂）运算；记：$A^2=AA$，$A^3=AAA$，$\cdots$，$A^k=\underbrace{AA\cdots A}_{k个},k\geq 1$，称之为 A 的 k 次（乘、方）幂；可以验证：基于乘法运算的结合律，方幂运算有如下运算律：1）$A^kA^l=A^{k+l}$；2）$\left(A^k\right)^l=A^{kl}$，这里 $k,l\in Z^+$，学过求逆运算后，k,l 可以为负数。由于矩阵乘法不满足交换律，故 $k>1$ 时，一般有 $(AB)^k\neq A^kB^k$。

以下，我们通过几个例题来了解求方阵幂的方法！

【例 2.1.7】设 $\alpha=\left(1,\frac{1}{2},\frac{1}{3}\right)$，$\beta=\begin{pmatrix}1\\2\\3\end{pmatrix}$，$A=\beta\alpha$，试求 $A^n,n\geq 1$；

【例 2.1.8】1）试由递推公式法验证：若 $A=\begin{pmatrix}1&1\\1&1\end{pmatrix}$，则 $A^n=2^{n-1}A$；

2）试通过数学归纳法来计算 $\begin{pmatrix}\cos\varphi&\sin\varphi\\-\sin\varphi&\cos\varphi\end{pmatrix}^n$；

【例 2.1.9】设 A,B 为同阶方阵，且 $AB=BA$，则有 $(A+B)^n=\sum\limits_{k=0}^{n}C_n^kA^kB^{n-k}$，这里，$A^0=B^0=I$；由其试求：

$\begin{pmatrix}1&1\\0&1\end{pmatrix}^n$，$\begin{pmatrix}a&b\\b&a\end{pmatrix}^n$，$\begin{bmatrix}1&1&0\\0&1&1\\0&0&1\end{bmatrix}^n$；

【矩阵多项式】设 A 是方阵，$f(\lambda)=\sum\limits_{i=0}^{n}a_i\lambda^i$ 是关于 λ 的 n 次多项式，记：$f(A)=\sum\limits_{i=0}^{n}a_iA^i$，这里，$A^0=I$，则称 $f(A)$ 为方阵 A 的多项式（又称矩阵多项式或多项式矩阵）。

若多项式 $f(\lambda),g(\lambda),h(\lambda),l(\lambda)$ 满足：$f(\lambda)+g(\lambda)=g(\lambda)+f(\lambda)=l(\lambda)$，$f(\lambda)g(\lambda)=g(\lambda)f(\lambda)=h(\lambda)$，则有：1）$f(A)g(A)=g(A)f(A)=h(A)$；2）$f(A)+g(A)=g(A)+f(A)=l(A)$。

由上述，易知：由 $\lambda^2-2\lambda-3=(\lambda-3)(\lambda+1)=(\lambda+1)(\lambda-3)$ 即有：$A^2-2A-3I=(A-3I)(A+I)=(A+I)(A-3I)$，也即有所谓矩阵多项式

的“因式分解”，其在以后抽象矩阵的求逆中作用突出！

4、 矩阵的转置（transpose）运算：将矩阵$A=\left(a_{ij}\right)_{m\times n}$的行与列互换，所得的$n\times m$矩阵，称为$A$的转置矩阵，简称$A$的转置，记作：$A^{\mathrm{T}}$（或$A'$）；若令$A^{\mathrm{T}}=\left(a_{ij}{}'\right)_{n\times m}$，则有$a_{ij}{}'=a_{ji},\forall i,j$。

【注 6】1）可以验证矩阵的转置运算满足如下性质：

i）$\left(A^{\mathrm{T}}\right)^{\mathrm{T}}=A$，从而偶数次转置等同没有转置；

ii）$(A+B)^{\mathrm{T}}=A^{\mathrm{T}}+B^{\mathrm{T}}$，即：和的转置等于转置的和，其可推广至有限和的情形；iii）$(kA)^{\mathrm{T}}=kA^{\mathrm{T}}$；iv）$(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$，即乘积的转置等于转置的反序积（重视证明），推广至一般有限的情形有：$\left(A_1A_2\cdots A_k\right)^{\mathrm{T}}=A_k{}^{\mathrm{T}}A_{k-1}{}^{\mathrm{T}}\cdots A_1{}^{\mathrm{T}}$；

2）若$A=\left(a_{ij}\right)_{n\times n}$，则$|A|=\left|A^{\mathrm{T}}\right|$,（行列式等于转置行列式）亦可写为：$\det A=\det A^{\mathrm{T}}$；

3）若方阵A满足$A^{\mathrm{T}}=A$，称A为对称（矩）阵；若$A^{\mathrm{T}}=-A$，则称A为反对称（矩）阵（分析其特点）；试着讨论$A+A^{\mathrm{T}}$，AA^{T}，并注意：$A=\frac{1}{2}\left(A+A^{\mathrm{T}}\right)+\frac{1}{2}\left(A-A^{\mathrm{T}}\right)$。

【例 2.1.10】在一项住房建筑计划中，一位社会学家设计了一张表格，询问每个居民哪些是他最好的朋友；假定仅有 5 名居民，他们的答案用矩阵A表示：

$$A=\begin{array}{c|c} & \begin{array}{ccccc} x & y & z & w & u \end{array} \\ \begin{array}{c} x\\ y\\ z\\ w\\ u \end{array} & \begin{pmatrix} 0&1&0&1&0\\ 1&0&1&0&0\\ 0&0&0&1&1\\ 0&0&0&0&1\\ 1&1&0&1&0 \end{pmatrix} \end{array}$$

，这里，若x认为y是他最好的朋友，则记为 1，否则为 0；A的每一行表示该行左边所示居民的选择，试计算矩阵AA^{T}和$A^{\mathrm{T}}A$，并解释其每个元素的意义；

【例 2.1.11】设有方阵$A=\left(a_{ij}\right)_{n\times n}$，$A$的迹（trace）定义为其主对角线上$n$个元素之和，即：$tr(A)=\sum_{i=1}^{n}a_{ii}$；可以验证：

$\forall A=\left(a_{ij}\right)_{n\times n}$，$\forall B=\left(b_{ij}\right)_{n\times n}$，**1）**$tr(A+B)=tr(A)+tr(B)$；**2）**$tr(kA)=ktr(A)$；**3）**$tr\left(A^{\mathrm{T}}\right)=tr(A)$；**4）**$tr(AB)=tr(BA)$；试由此证明：$\forall A_{n\times n}$，$\forall B_{n\times n}$，等式$AB-BA=I_n$必不成立！

5、 矩（方）阵的求逆运算：在通常的数的乘法中，对数 $a(a \neq 0)$，总存在数 b，使得 $ab = ba = 1$，数 b 即为 a 的倒数，记为 a^{-1}；利用倒数概念，数的除法就可归结为数的乘法，即：$a \div b = ab^{-1}$，$b \neq 0$。在考察 n 阶方阵中是否引进除法时，自然应把倒数概念推广！

设 A 是 n 阶方阵，若存在一个 n 阶方阵 B，使得 $AB = BA = I_n$，则称 B 是 A 的逆（矩）阵，记作：$B = A^{-1}$。

凡是有逆矩阵的矩阵称为可逆阵（非（奇）异阵或非退化阵），否则称为不可逆阵（奇异阵或退化阵）。

常常将矩阵的求逆运算视作“除法运算”，即为矩阵乘法运算的逆运算；正如数的除法中“**0**”不能作为除数，并非任意非零方阵都有逆矩阵，如：$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$；再由于矩阵的乘法运算不满足交换律，故 $A^{-1}BA = B$ 一般不成立，而只有 $(BA)A^{-1} = B$。

一个 n 阶方阵 A 若有逆矩阵，则逆矩阵必唯一。事实上，若 B,C 均是 n 阶方阵 A 的逆矩阵，即：$AB = BA = AC = CA = I_n$，则有：$B = BI_n = B(AC) = (BA)C = I_nC = C$。

【例 **2.1.12**】**1**）设 A 为 n 阶幂等阵，即：$A^2 = A$，试证：$2E - A$ 可逆，并求其逆；**2**）设 n 阶方阵 A 满足方程 $A^2 - 3A - 10E = O$，试证：$A, A - 4E$ 均可逆，并求其逆；**3**）A 为 n 阶方阵，若有 $k \in N$，使 $A^k = O$（幂零阵），则 $I_n - A$ 可逆，试求其逆；**4**）若 n 阶矩阵 A 满足 $A^2 + aA + bE = O$，其中 a,b 均为常数，试讨论 A 可逆的充要条件；

【例 **2.1.13**】上（下）三角非异阵的逆矩阵仍是上（下）三角矩阵；

【例 **2.1.14**】设 $A, B, A+B, A^{-1}+B^{-1}$ 均为 n 阶非奇异阵，试求 $\left(A^{-1}+B^{-1}\right)^{-1}$；

方阵的求逆运算适合如下规则：

1） 若 A 非异，则 $\left(A^{-1}\right)^{-1} = A$；**2**）若 A,B 均是 n 阶非异阵，则 AB 也是 n 阶非异阵，且 $(AB)^{-1} = B^{-1}A^{-1}$；可推广至一般情形：若 $A_1, A_2, \cdots, A_k$ 均是 n 阶非异阵，则 $A_1A_2\cdots A_k$ 也是 n 阶非异阵，且

$(A_1A_2\cdots A_k)^{-1}=A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}$（穿脱原理）——乘积的逆等于逆的反序积；**3**）若 A 非异，$k\neq 0$，则 kA 也非异，且 $(kA)^{-1}=k^{-1}A^{-1}$；**4**）若 A 非异，则 A^{T} 也非异，且 $\left(A^{\mathrm{T}}\right)^{-1}=\left(A^{-1}\right)^{\mathrm{T}}$。

【定义】设 $A=\left(a_{ij}\right)_{n\times n}$，$A_{ij}$ 为 a_{ij} 在 $\det A$ 中的代数余子式，则称矩阵 $\left(A_{ij}\right)_{n\times n}^{\mathrm{T}}=\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}^{\mathrm{T}}$ 为 A 的伴随矩阵，记作：A^*（或 $adjA$）。

【定理】设 $A=\left(a_{ij}\right)_{n\times n}$，则 A 可逆当且仅当 $\det A\neq 0$；

可以验证：$A^* A = A^* A = |\ \ | A_n$，且 $|A^*|=|A|^{n-1}$；

【推论】可逆矩阵 A 的逆矩阵 $A^{-1}=\dfrac{1}{\det A}A^*$；

上述推论可作为公式法求可逆矩阵的逆矩阵，此法适合低阶矩阵求逆的情形!

【例 2.1.15】1）试求 $\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}^{-1}$；**2**）已知 $A^6=E_2$，试求 A^{11}，其中 $A=\begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$；

【例 2.1.16】已知 $\det A=\det\left(a_{ij}\right)_{3\times 3}=a\neq 0$，试求：**1**）$\left|A^{-1}\right|$；**2**）$\left|A^*\right|$；**3**）$|-2A|$，$\left|(3A)^{-1}\right|$；**4**）$\left|\dfrac{1}{2}A^*-3A^{-1}\right|$；**5**）$\left(A^*\right)^{-1}$；

【例 2.1.17】（矩阵方程的求解）**1**）设矩阵 A 的伴随矩阵 $A^*=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -3 & 0 & 8 \end{pmatrix}$，且 $ABA^{-1}=BA^{-1}+3E$，试求 B；**2**）设 $A=\begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix}$，且 X 满足：$A^*X=A^{-1}+2X$，试求 X；

§2　分块矩阵、矩阵的初等变换与初等矩阵

为了简化矩阵的运算，以下引入分块矩阵及其运算；其并不是一种新的运算，只是矩阵运算的简化形式。

所谓“矩阵的分块”，就是用若干条横线和竖线将一个

矩阵分成若干快，即得“分块矩阵”。如：$A=\begin{pmatrix} 1 & 3 & 4 & \vdots & 0 \\ 2 & 1 & 5 & \vdots & 3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 3 & \vdots & 7 \end{pmatrix}$ 是一个分块矩阵，若记：$A_{11}=\begin{pmatrix} 1 & 3 & 4 \\ 2 & 1 & 5 \end{pmatrix}$，$A_{12}=\begin{pmatrix} 0 \\ 3 \end{pmatrix}$，$A_{21}=(1,0,3)$，$A_{22}=(7)$；于是，$A$ 可表示为：$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$，即为分块矩阵；这里，即为分块矩阵的子块（块矩阵）。一个矩阵可以有各种各样的分块方法，但如何分块得视具体情况来定！例如：

$$A=\begin{pmatrix} 1 & 1 & \vdots & 0 & 0 & \vdots & 0 \\ 0 & 1 & \vdots & 0 & 0 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 2 & 1 & \vdots & 0 \\ 0 & 0 & \vdots & 1 & 1 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 0 & 0 & \vdots & 1 \end{pmatrix}$$ ，若记 $A_1=\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$，$A_2=\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$，$A_3=(1)$，

则形式上有：$A=\begin{pmatrix} A_1 & O & O \\ O & A_2 & O \\ O & O & A_3 \end{pmatrix}$；作为分块矩阵，它是一个“对角阵”，常称之为分块对角阵（准对角阵），也可记为：$A=diag\{A_1,A_2,A_3\}$。

两个分块矩阵 $A=(A_{ij})_{r\times s}$，$B=(B_{ij})_{l\times k}$ 称为相等，若 $r=l$，$s=k$，且 $A_{ij}=B_{ij}, i=1,2,\cdots,r; j=1,2,\cdots,s$；也即：两个分块矩阵相等，不仅它们的分块方式相同，而且每一子块也相等。

矩阵分块须把握如下原则：**1**）适当分块后，在矩阵运算中，可将子块（子矩阵）当作“数”，像普通的数为元素的矩阵一样运算；**2**）尽可能使运算简单、方便，因此分块时须注意到矩阵的运算规则。以下讨论分块矩阵的运算：

（**1**） 加、减法运算：若 $m\times n$ 阵 A,B 具有相同的分块，即：$A=(A_{ij})_{r\times s}$，$B=(B_{ij})_{r\times s}$，且 A_{ij},B_{ij} 作为矩阵行、列数分别相等，则 $A\pm B=(A_{ij}\pm B_{ij})_{r\times s}$；

易见，两个分块矩阵之和（差）仍是分块矩阵，其和（差）与通过的普通矩阵 A,B 的加、减法结果是一致的。

（2） 数乘运算：定义常数k与分块矩阵$A=\left(A_{ij}\right)_{r\times s}$的数乘为：$kA=\left(kA_{ij}\right)_{r\times s}$；

（3） 乘法运算：设矩阵A,B可相乘，且A,B的分块如下：

$$A=\begin{array}{c@{}c} \begin{matrix} n_1 & n_2 & \cdots & n_s \end{matrix} & \\ \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{bmatrix} & \begin{matrix} m_1 \\ m_2 \\ \vdots \\ m_r \end{matrix} \end{array}，\quad B=\begin{array}{c@{}c} \begin{matrix} l_1 & l_2 & \cdots & l_t \end{matrix} & \\ \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ B_{21} & B_{22} & \cdots & B_{2t} \\ \cdots & \cdots & \cdots & \cdots \\ B_{s1} & B_{s2} & \cdots & B_{st} \end{bmatrix} & \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_s \end{matrix} \end{array}$$

，则有：

$$C=AB=\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1t} \\ C_{21} & C_{22} & \cdots & C_{2t} \\ \cdots & \cdots & \cdots & \cdots \\ C_{r1} & C_{r2} & \cdots & C_{rt} \end{pmatrix}$$

，其中：$C_{ij}=\sum_{k=1}^{s}A_{ik}B_{kj}$是$m_i\times l_j$阵。

【注1】 左乘矩阵A列的划分必须等同于右乘矩阵B行的划分！

（4） 转置运算：设$A=\left(A_{ij}\right)_{r\times s}$，则$A$的转置矩阵$A^{\mathrm{T}}$为$s\times r$分块阵，且

$$A^{\mathrm{T}}=\begin{pmatrix} A_{11}^{\mathrm{T}} & A_{21}^{\mathrm{T}} & \cdots & A_{r1}^{\mathrm{T}} \\ A_{12}^{\mathrm{T}} & A_{22}^{\mathrm{T}} & \cdots & A_{r2}^{\mathrm{T}} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1s}^{\mathrm{T}} & A_{2s}^{\mathrm{T}} & \cdots & A_{rs}^{\mathrm{T}} \end{pmatrix}。$$

【例 2.2.1】将矩阵$A=\left(a_{ij}\right)_{m\times n}$按列分块，即$A=\left(A_1,A_2,\cdots,A_n\right)$，其中$A_i$为的第$i$列，即$A_i=\left(a_{1i},a_{2i},\cdots,a_{mi}\right)^{\mathrm{T}}$，设$y=\left(y_1,y_2,\cdots,y_n\right)^{\mathrm{T}}$为列向量，则有：

$$Ay=\left(A_1,A_2,\cdots,A_n\right)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}=y_1A_1+y_2A_2+\cdots+y_nA_n$$

；若将矩阵A按行分块，即：

$$A=\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}$$

，其中α_i为A的第i行，即：$\alpha_i=\left(a_{i1},a_{i2},\cdots,a_{in}\right)$，设$x=\left(x_1,x_2,\cdots,x_m\right)$为行向量，则有：

$$xA=\left(x_1,x_2,\cdots,x_m\right)\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}$$

$=x_1\alpha_1+x_2\alpha_2+\cdots+x_m\alpha_m$；若 $A=\left(a_{ij}\right)_{m\times n}=\left(A_1,A_2,\cdots,A_n\right)=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\end{pmatrix}$，

$B=\left(b_{ij}\right)_{n\times s}=\left(B_1,B_2,\cdots,B_s\right)=\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{pmatrix}$，则有：（1）$AB=A\left(B_1,B_2,\cdots,B_s\right)$

$=\left(AB_1,AB_2,\cdots,AB_s\right)$，而 $AB_j=\left(A_1,A_2,\cdots,A_n\right)\begin{pmatrix}b_{1j}\\ b_{2j}\\ \vdots\\ b_{nj}\end{pmatrix}=b_{1j}A_1+b_{2j}A_2+\cdots+b_{nj}A_n$；

（2）$AB=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\end{pmatrix}B=\begin{pmatrix}\alpha_1B\\ \alpha_2B\\ \vdots\\ \alpha_mB\end{pmatrix}$，其中 $\alpha_iB=\left(a_{i1},a_{i2},\cdots,a_{in}\right)\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{pmatrix}=a_{i1}\beta_1+a_{i2}\beta_2+\cdots$

$+a_{in}\beta_n$；

综上所述：（a）矩阵 A 与列向量 y 的乘积，是 A 的各列的线性组合，组合系数为 y 的各分量；（b）行向量 x 与 A 的乘积，是 A 的各行的线性组合，组合系数为 x 的各分量；（c）AB 的第 j 列是 A 的各列的线性组合，AB 的第 i 行是 B 的各行的线性组合。

【例 2.2.2】设 $A=\left(a_{ij}\right)_{m\times n}$，试证：若对任一 n 维列向量 x，恒有 $Ax=O_{m\times 1}$，则 $A=O_{m\times n}$；

【例 2.2.3】设 n 阶方阵 $A=\begin{pmatrix}O_{(n-1)\times 1} & E_{n-1}\\ 0 & O_{1\times(n-1)}\end{pmatrix}$，试证：$A^n=O_{n\times n}$；

【例 2.2.4】1）设有分块矩阵 $A=\begin{pmatrix}B & D\\ O & C\end{pmatrix}$，其中 A 的子块 B,C 分别为 m,n 阶可逆阵，试求 A^{-1}；2）设有分块矩阵 $A=\begin{pmatrix}B & O\\ D & C\end{pmatrix}$，其中 A 的子块 B,C 分别为 m,n 阶可逆阵，试求 A^{-1}；3）设有分块矩阵 $A=\begin{pmatrix}B & O\\ O & C\end{pmatrix}$，其中 A 的子块 B,C 分别为 m,n 阶可逆阵，试求 A^{-1}；4）设有分块矩阵 $A=\begin{pmatrix}O & B\\ C & O\end{pmatrix}$，其中 A 的子块 B,C 分别为 m,n

阶可逆阵，试求A^{-1}。

已经知道，矩阵的概念源于求解线性方程组，求解线性方程组的常用的一种方法是（高斯）消元法，这种方法可归结为对系数矩阵及其增广矩阵的变换（举例说明）；类似于消元法的步骤，

【定义】称以下三类矩阵变换为矩阵的第一、二、三类初等行（列）变换：（1）对换矩阵的某两行（列）；（2）用某一非零常数乘以矩阵的某一行（列）；（3）将矩阵的某一行（列）乘以常数后加至另一行（列）；上述三种变换统称为矩阵的初等变换。

【定义】对单位矩阵施行一次第一、第二、第三类初等变换后得到的矩阵分别称为第一类、第二类及第三类初等矩阵。

第一类初等矩阵（对换阵）：$P(ij)=\begin{array}{cc} \begin{pmatrix} 1 & & & & \\ & 0 & \cdots & 1 & \\ & \vdots & \ddots & \vdots & \\ & 1 & \cdots & 0 & \\ & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \text{第}i\text{行} \\ \\ \text{第}j\text{行} \\ \\ \end{array} \end{array}$，其由单位矩阵第$i$行（列）与第$j$行（列）对换而得；

第二类初等矩阵（倍乘阵）：$P(i(k))=\begin{array}{cc} \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & k & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \text{第}i\text{行} \\ \\ \\ \end{array} \end{array}$，其由单位矩阵的第$i$行（列）乘以常数$k(k\neq 0)$而得；

第三类初等矩阵（倍加阵）：$P(ij(c))=\begin{array}{cc} \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & c & \\ & & & & & \\ & & 0 & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \text{第}i\text{行} \\ \\ \text{第}j\text{行} \\ \\ \\ \end{array} \\ \begin{array}{ccc} & \text{第}i\text{列} & \text{第}j\text{列} \end{array} & \end{array}$，

其由单位矩阵的第j行（第i列）乘以常数c加至第i行（第j列）而得；

可以验证：$|P(ij)|=-1$，$|P(i(k))|=k$，$|P(ij(c))|=1$；从而，初

等矩阵都是可逆的，且有：$\left[P(ij)\right]^{-1}=P(ij)$，$\left[P(i(k))\right]^{-1}=P\left(i\left(\frac{1}{k}\right)\right)$，$\left[P(ij(c))\right]^{-1}=P(ij(-c))$。

以下定理揭示了初等矩阵与初等变换之间的密切联系，我们可以通过分块矩阵的运算证明之！

【定理】设$A=(a_{ij})_{m\times n}$，则对A施行一次初等行变换后所得的矩阵等价于用一个m阶相应的初等矩阵左乘A所得的乘积；对A施行一次初等列变换后所得的矩阵等价于用一个n阶相应的初等矩阵右乘A所得的乘积。

【例 2.2.5】对于下列的初等变换，写出相应的初等矩阵及A,B之间的关系：$A=\begin{pmatrix}1&0&-1&2\\2&3&1&2\\1&-1&2&1\end{pmatrix}\xrightarrow{r_1\times(-2)+r_2}\begin{pmatrix}1&0&-1&2\\0&3&3&-2\\1&-1&2&1\end{pmatrix}\xrightarrow{c_1\times 1+c_3}$
$\begin{pmatrix}1&0&0&2\\0&3&3&-2\\1&-1&3&1\end{pmatrix}=B$；

【例 2.2.6】利用初等变换求如下矩阵方程的解：

$$\begin{pmatrix}1&1&\cdots&1\\0&1&\cdots&1\\\cdots&\cdots&\cdots&\cdots\\0&0&\cdots&1\end{pmatrix}X=\begin{pmatrix}1&2&\cdots&n\\0&1&\cdots&n-1\\\cdots&\cdots&\cdots&\cdots\\0&0&\cdots&1\end{pmatrix}；$$

【注】以下我们用初等变换法来解释线性方程组求解的高斯（**Gauss**）消元法：考虑线性方程组求解程序，

$\begin{cases}x+y=2\\x+2y=3\end{cases}\Leftrightarrow\begin{cases}x+y=2\\y=1\end{cases}\Leftrightarrow\begin{cases}x=1\\y=1\end{cases}$，此即，$\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}2\\3\end{pmatrix}\Leftrightarrow\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}2\\1\end{pmatrix}$
$\Leftrightarrow\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}$，也即：$\begin{pmatrix}1&1&2\\1&2&3\end{pmatrix}\xrightarrow{r_1\times(-1)+r_2}\begin{pmatrix}1&1&2\\0&1&1\end{pmatrix}\xrightarrow{r_2\times(-1)+r_1}$
$\begin{pmatrix}1&0&1\\0&1&1\end{pmatrix}$。

以下考虑用初等变换法求可逆矩阵的逆矩阵，

【定理】任一$m\times n$阵$A=(a_{ij})_{m\times n}$，必可经过有限次初等变换化为如下形式的$m\times n$阵$\begin{pmatrix}I_s&O\\O&O\end{pmatrix}$，亦即存在初等矩阵$P_1,P_2,\cdots,P_r$及$Q_1,Q_2,\cdots,Q_t$，使得$P_r\cdots P_2P_1AQ_1Q_2\cdots Q_t=\begin{pmatrix}I_s&O\\O&O\end{pmatrix}$；这里，$0\le s\le\min\{m,n\}$。

【定理】A为非异阵（可逆阵）的充要条件是A可表示为有限个初等矩阵的乘积。

设n阶矩阵A可逆，其逆矩阵为A^{-1}，则必存在初等矩阵$P_1,P_2,\cdots,P_k$，使得$A=P_1P_2\cdots P_k$，从而$A^{-1}=P_k^{-1}P_{k-1}^{-1}\cdots P_1^{-1}=Q_kQ_{k-1}\cdots Q_1$，这里$P_i^{-1}=Q_i$；$A^{-1}A=\left(Q_kQ_{k-1}\cdots Q_1\right)A=I_n$。由分块矩阵的运算，构造如下的分块矩阵$(A\ \vdots\ I_n)$，即有：

$Q_kQ_{k-1}\cdots Q_1\left(A\ \vdots\ I_n\right)=\left(Q_kQ_{k-1}\cdots Q_1A\ \vdots\ Q_kQ_{k-1}\cdots Q_1I_n\right)=\left(I_n\ \vdots\ A^{-1}\right)$；此即表明：若对分块矩阵$(A\ \vdots\ I_n)$施行初等行变换，当左半块$A$化为同阶单位矩阵时，右半块则化为$A^{-1}$；于是，可如下利用初等行变换求可逆矩阵$A$的逆：

先构造$n\times 2n$阵$(A\ \vdots\ I_n)$；再对$(A\ \vdots\ I_n)$连续地施行初等行变换，化A为单位阵I_n，则I_n化为A^{-1}，$(A\ \vdots\ I_n)\xrightarrow{\text{初等行变换}}\left(I_n\ \vdots\ A^{-1}\right)$；

类似地，也可由初等列变换法求矩阵的逆：

1）构造$2n\times n$阵$\begin{pmatrix}A\\ \cdots\\ I_n\end{pmatrix}$；**2**）对$\begin{pmatrix}A\\ \cdots\\ I_n\end{pmatrix}$连续地施行初等列变换，使得上半块$A$化为单位矩阵$I_n$，则$I_n$部分化为$A^{-1}$，即：

$$\begin{pmatrix}A\\ \cdots\\ I_n\end{pmatrix}Q_kQ_{k-1}\cdots Q_1\underline{\underline{\text{初等列变换}}}\begin{pmatrix}AQ_kQ_{k-1}\cdots Q_1\\ \cdots\\ I_nQ_kQ_{k-1}\cdots Q_1\end{pmatrix}=\begin{pmatrix}I_n\\ \cdots\\ A^{-1}\end{pmatrix}。$$

基于以上求可逆矩阵的逆的办法，可如下求如下矩阵方程的解；

1）在矩阵方程$AX=B$中，若$A_{n\times n}$可逆，B为$n\times m$阵，则有$X=A^{-1}B$；实际上，也可构造$n\times(n+m)$块矩阵$(A\ \vdots\ B)$，对其连续地施行初等行变换，当左半块A化为I_n时，右半块B化为$A^{-1}B$；即：$(A\ \vdots\ B)\xrightarrow{\text{初等行变换}}\left(I_n\ \vdots\ A^{-1}B\right)$；

2）在矩阵方程$XC=B$中，若$C_{n\times n}$可逆，B为$m\times n$阵，则有$X=BC^{-1}$；实际上，也可构造$(m+n)\times n$块矩阵$\begin{pmatrix}B\\ \cdots\\ C\end{pmatrix}$，对其连续地施行初等列变换，当下半块$C$化为$I_n$时，上半块$B$化为$BC^{-1}$；即：

$$\begin{pmatrix} B \\ \cdots \\ C \end{pmatrix} \xrightarrow{\text{初等列变换}} \begin{pmatrix} BC^{-1} \\ \cdots \\ I_n \end{pmatrix}。$$

【例 2.2.7】求解矩阵方程： $AX = A + 2X$，**其中** $A = \begin{pmatrix} 4 & 2 & 3 \\ 1 & 1 & 0 \\ -1 & 2 & 3 \end{pmatrix}$；

【例 2.2.8】解方程： $AXB = C$，**其中** $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 4 & 3 \end{pmatrix}$，$B = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}$，

$C = \begin{pmatrix} 1 & 3 \\ 2 & 0 \\ 3 & 1 \end{pmatrix}$。

我们也可尝试由分块矩阵的初等变换法求其逆，如对**【例 2.2.4】**！以下讨论分块矩阵的初等变换！首先介绍分块初等矩阵；

【分块对换阵】$\begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix}$，它由 $m+n$ 阶单位矩阵 $\begin{pmatrix} I_m & O \\ O & I_n \end{pmatrix}$，交换两“行”或两“列”所得；

【分块倍乘阵】以下 C_1, C_2 分别是 m, n 阶可逆阵，$\begin{pmatrix} C_1 & O \\ O & I_n \end{pmatrix}$ 或 $\begin{pmatrix} I_m & O \\ O & C_2 \end{pmatrix}$，它由 $m+n$ 阶单位矩阵 $\begin{pmatrix} I_m & O \\ O & I_n \end{pmatrix}$，将第一“行”（第二“列”）左（右）乘非异阵 C_1（C_2）所得；

【分块倍加阵】$\begin{pmatrix} I_m & O \\ C_3 & I_n \end{pmatrix}$ 或 $\begin{pmatrix} I_m & C_4 \\ O & I_n \end{pmatrix}$，它由 $m+n$ 阶单位矩阵 $\begin{pmatrix} I_m & O \\ O & I_n \end{pmatrix}$，将第一“行”左乘 C_3 加至第二“行”或将第一“列”右乘 C_4 加至第二“列”所得。

【注1】 分块初等矩阵有着与初等矩阵类似的定理，以下以“矩阵分块零化技巧”为例来说明之！

设 M 为 $m+n$ 阶方阵，分块为 $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$，其中 A, D 分别为 m, n 阶非异阵，B, C 不必是方阵，可将 B, C 零化，化为准三（对）角阵。

1）若A可逆，$\begin{pmatrix} I_m & O \\ -CA^{-1} & I_n \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}=\begin{pmatrix} A & B \\ O & D-CA^{-1}B \end{pmatrix}$，

$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I_m & -A^{-1}B \\ O & I_n \end{pmatrix}=\begin{pmatrix} A & O \\ C & D-CA^{-1}B \end{pmatrix}$；

2）若D可逆，$\begin{pmatrix} I_m & -BD^{-1} \\ O & I_n \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}=\begin{pmatrix} A-BD^{-1}C & O \\ C & D \end{pmatrix}$，

$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I_m & O \\ -D^{-1}C & I_n \end{pmatrix}=\begin{pmatrix} A-BD^{-1}C & B \\ O & D \end{pmatrix}$；还可进一步化为准对角阵，

3）若A可逆，$\begin{pmatrix} I_m & O \\ -CA^{-1} & I_n \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I_m & -A^{-1}B \\ O & I_n \end{pmatrix}=\begin{pmatrix} A & O \\ O & D-CA^{-1}B \end{pmatrix}$；若$D$可逆，$\begin{pmatrix} I_m & -BD^{-1} \\ O & I_n \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I_m & O \\ -D^{-1}C & I_n \end{pmatrix}=\begin{pmatrix} A-BD^{-1}C & O \\ O & D \end{pmatrix}$。

§3 矩阵、行列式及其应用

先来看行列式的函数定义；

【定义】数域F（实数集或复数集）上的方阵$A=\left(a_{ij}\right)_{n\times n}$的行列式定义为数域$F$中的数 $|A|=\det\left(a_{ij}\right)_{n\times n}=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}=$ $\sum\limits_{(j_1j_2\cdots j_n)}(-1)^{N(j_1j_2\cdots j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}$；这里，令$M_n(F)$为数域$F$上$n$阶方阵的全体；对于每个$n$阶方阵$A\in M_n(F)$，有唯一的$\det A\in F$；也即：行列式（运算）$\det(\bullet)$实际上是$M_n(F)$到$F$上的一个映射（函数），即：$\det:M_n(F)\to F$，$A\mapsto \det A$，亦称行列式映射。

行列式的性质可重述如下：

1）行列式对行（列）有多线性，也即：$\forall i\in\{1,2,\cdots,n\}$，$\alpha_i,\alpha_i^*\in F^n$，$\lambda,\mu\in F$，有：$\det\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{i-1} \\ \lambda\alpha_i+\mu\alpha_i^* \\ \alpha_{i+1} \\ \vdots \\ \alpha_n \end{pmatrix}=\lambda\det\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{i-1} \\ \alpha_i \\ \alpha_{i+1} \\ \vdots \\ \alpha_n \end{pmatrix}+\mu\det\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{i-1} \\ \alpha_i^* \\ \alpha_{i+1} \\ \vdots \\ \alpha_n \end{pmatrix}$；**2**）行列式对

行（列）有交错性：即：$\det\begin{pmatrix}\vdots\\ \alpha_i\\ \vdots\\ \alpha_i\\ \vdots\end{pmatrix}=0$；**3**）规范性：$\det I_n=1$；

事实上，行列式的这三条性质完全定义了行列式！

【定理】定义在n阶方阵集上，对于方阵的行（列）有多线性、交错性和规范性的函数，存在且唯一，即为行列式；设有映射$D:M_n(F)\to F$，且D满足：（**1**）对方阵的行有多线性，即：$D\begin{pmatrix}\vdots\\ \lambda\alpha_i+\mu\alpha_i^*\\ \vdots\end{pmatrix}=\lambda D\begin{pmatrix}\vdots\\ \alpha_i\\ \vdots\end{pmatrix}+\mu D\begin{pmatrix}\vdots\\ \alpha_i^*\\ \vdots\end{pmatrix}$；（**2**）对方阵的行有交错性，即：$D\begin{pmatrix}\vdots\\ \alpha_i\\ \vdots\\ \alpha_i\\ \vdots\end{pmatrix}=0$；（**3**）规范性，即：$D(I_n)=1$；则必有$D=\det$，也即：$\forall A\in M_n(F)$，有$D(A)=\det A$。

这里，记$A=\left(a_{ij}\right)_{n\times n}$，$\alpha_i=\left(a_{i1},a_{i2},\cdots,a_{in}\right)$为$A$的第$i$行，并记

$e_i=\left(0\quad\cdots\quad \underset{\text{第}i\text{个}}{1}\quad 0\quad\cdots\quad 0\right)$，$D(A)=D\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix}=D\begin{pmatrix}\sum_{j=1}^n a_{1j}e_j\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix}=\sum_{j=1}^n D\begin{pmatrix}e_j\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix}a_{1j}$

$$=\sum_{j=1}^n\sum_{k=1}^n a_{1j}a_{2k}D\begin{pmatrix}e_j\\ e_k\\ \alpha_3\\ \vdots\\ \alpha_n\end{pmatrix}=\cdots=\sum_{(j_1j_2\cdots j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}D\begin{pmatrix}e_{j_1}\\ e_{j_2}\\ \vdots\\ e_{j_n}\end{pmatrix}=\sum_{(j_1j_2\cdots j_n)}(-1)^{N(j_1j_2\cdots j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}D\begin{pmatrix}e_1\\ e_2\\ \vdots\\ e_n\end{pmatrix}$$

$$=\sum_{(j_1j_2\cdots j_n)}(-1)^{N(j_1j_2\cdots j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}=\det A;$$

另：定义于n阶方阵集上，对方阵的行有多线性、交错性的函数$D(\bullet)$必为行列式的倍数$c\cdot\det(\bullet)$。

【定理】若$A=\left(a_{ij}\right)_{n\times n}$，$B=\left(b_{ij}\right)_{n\times n}$，则$\det(AB)=\det A\cdot\det B$；

易见，$\det(AB)=\det\left(\sum_{k=1}^n a_{ik}b_{kj}\right)_{n\times n}$，将其依此按第$1,2,\cdots,n$列拆开，

即有：$\det(AB)=\begin{vmatrix}\sum_{j_1=1}^{n}a_{1j_1}b_{j_11} & \sum_{j_2=1}^{n}a_{1j_2}b_{j_22} & \cdots & \sum_{j_n=1}^{n}a_{1j_n}b_{j_nn}\\ \sum_{j_1=1}^{n}a_{2j_1}b_{j_11} & \sum_{j_2=1}^{n}a_{2j_2}b_{j_22} & \cdots & \sum_{j_n=1}^{n}a_{2j_n}b_{j_nn}\\ \cdots & \cdots & \cdots & \cdots\\ \sum_{j_1=1}^{n}a_{nj_1}b_{j_11} & \sum_{j_2=1}^{n}a_{nj_2}b_{j_22} & \cdots & \sum_{j_n=1}^{n}a_{nj_n}b_{j_nn}\end{vmatrix}$

$$=\sum_{1\le j_1,j_2,\cdots,j_n\le n}b_{j_11}b_{j_22}\cdots b_{j_nn}\begin{vmatrix}a_{1j_1} & a_{1j_2} & \cdots & a_{1j_n}\\ a_{2j_1} & a_{2j_2} & \cdots & a_{2j_n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{nj_1} & a_{nj_2} & \cdots & a_{nj_n}\end{vmatrix}=\sum_{1\le j_1,j_2,\cdots,j_n\le n}(-1)^{N(j_1j_2\cdots j_n)}b_{j_11}b_{j_22}\cdots b_{j_nn}\begin{vmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{vmatrix}$$

$=\det B\cdot\det A$；

上述结论推广至为长方阵的情形！

【定理（Binet—Cauchy）】设 $A=\left(a_{ij}\right)_{m\times n}$， $B=\left(b_{ij}\right)_{n\times m}$，则

$$\det(AB)=\begin{cases}0, & m>n;\\ \det A\cdot\det B, & m=n;\\ \sum\limits_{1\le j_1<j_2<\cdots<j_m\le n}A\begin{pmatrix}1 & 2 & \cdots & m\\ j_1 & j_2 & \cdots & j_m\end{pmatrix}\cdot B\begin{pmatrix}j_1 & j_2 & \cdots & j_m\\ 1 & 2 & \cdots & m\end{pmatrix}, & m<n;\end{cases}$$。

【例 2.3.1】设 n 阶方阵 A,B 满足 $A^{\mathrm{T}}A=B^{\mathrm{T}}B=I_n$， 且 $|A|+|B|=0$，则 $|A+B|=0$；

【例 2.3.2】A 为五阶方阵，且 $|A|=3$，试求 $\left|AA^{\mathrm{T}}\right|,\left|A^*\right|,\left|\left(A^*\right)^*\right|,\left|(3A)^*\right|$；

【例 2.3.3】求 $D_4=\begin{vmatrix}a & b & c & d\\ -b & a & -d & c\\ -c & d & a & -b\\ -d & -c & b & a\end{vmatrix},D_n=\begin{vmatrix}1+x_1y_1 & 1+x_1y_2 & \cdots & 1+x_1y_n\\ 1+x_2y_1 & 1+x_2y_2 & \cdots & 1+x_2y_n\\ \cdots & \cdots & \cdots & \cdots\\ 1+x_ny_1 & 1+x_ny_2 & \cdots & 1+x_ny_n\end{vmatrix}$；

【例 2.3.4】设 $S_k=\lambda_1^k+\lambda_2^k+\cdots+\lambda_n^k$， $k=1,2,\cdots$， 试证：

$$\begin{vmatrix}n & S_1 & S_2 & \cdots & S_{n-1}\\ S_1 & S_2 & S_3 & \cdots & S_n\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ S_{n-1} & S_n & S_{n+1} & \cdots & S_{2n-2}\end{vmatrix}=\prod_{1\le j<i\le n}\left(\lambda_i-\lambda_j\right)^2$$；

【例 2.3.5】n 阶方阵 $A=\begin{pmatrix}a_0 & a_1 & a_2 & \cdots & a_{n-1}\\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ a_1 & a_2 & a_3 & \cdots & a_0\end{pmatrix}$称为轮迴方阵（循

环方阵），试求$\det A$；

【例 2.3.6】1）已知三阶行列式$|(\alpha_1,\alpha_2,\alpha_3)|=2$，求行列式$|(\alpha_1+2\alpha_3,\alpha_3,\alpha_1+\alpha_2)|$；　　**2）**设$\alpha_1,\alpha_2,\cdots,\alpha_n$均为$n$维列向量，$A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$，$B=(\alpha_1+\alpha_2,\alpha_2+\alpha_3,\cdots,\alpha_{n-1}+\alpha_n,\alpha_n+\alpha_1)$，若$|A|=a\neq 0$，试求$|B|$；

【例 2.3.7】设三阶方阵$A=\begin{pmatrix}1&2&-2\\4&t&3\\3&-1&1\end{pmatrix}$，$B$为三阶非零矩阵，且$AB=O_{3\times 3}$，求$t$；

【再论 Cramer 法则】设$A=(a_{ij})_{n\times n}$，$X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，$b=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$；若方程组$AX=b$中$\det A\neq 0$，则有：

$$X=A^{-1}b=\frac{1}{\det A}A^*b=\frac{1}{D}\begin{pmatrix}A_{11}&A_{21}&\cdots&A_{n1}\\A_{12}&A_{22}&\cdots&A_{n2}\\\cdots&\cdots&\cdots&\cdots\\A_{1n}&A_{2n}&\cdots&A_{nn}\end{pmatrix}\begin{pmatrix}b_1\\b_1\\\vdots\\b_1\end{pmatrix}；\ x_i=\frac{1}{D}\sum_{j=1}^{n}b_jA_{ji}=\frac{D_i}{D}$$

，其中D_i是由系数行列式D的第i列换为$b=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$所得。

以下讨论“行列式的降阶公式”!

对于分块矩阵$\begin{pmatrix}A&B\\O&C\end{pmatrix}$，$A,C$为方阵，固定$A,B$，将$C$的元素$c_{ij}$视为常数，则$\det\begin{pmatrix}A&B\\O&C\end{pmatrix}$是方阵$C$的行向量的多线性、交错性函数；从而，$\det\begin{pmatrix}A&B\\O&C\end{pmatrix}=\det C\cdot\det\begin{pmatrix}A&B\\O&I\end{pmatrix}=\det C\cdot\det A\cdot\det\begin{pmatrix}I&B\\O&I\end{pmatrix}=\det C\cdot\det A$；类似有$\det\begin{pmatrix}A&O\\D&C\end{pmatrix}=\det C\cdot\det A$，这里$A,C$均为方阵。

【降阶公式】设$M=\begin{pmatrix}A&B\\C&D\end{pmatrix}$是方阵，其中$A$是非异阵，则$|M|=\left|\begin{pmatrix}A&B\\C&D\end{pmatrix}\right|=\begin{vmatrix}A&B\\C&D\end{vmatrix}=|A|\cdot|D-CA^{-1}B|$；若$D$是非异阵，则$|M|=\left|\begin{pmatrix}A&B\\C&D\end{pmatrix}\right|=\begin{vmatrix}A&B\\C&D\end{vmatrix}=|D|\cdot|A-BD^{-1}C|$；其中，若$A$可逆且（**1**）$AC=CA$，则$|M|=|AD-CB|$；（**2**）$AB=BA$，则$|M|=|DA-CB|$；若$D$可逆且（**1**）$DB=BD$，则$|M|=|DA-BC|$；（**2**）$CD=DC$，则$|M|=|AD-BC|$。

【降阶公式的注】当上述矩阵A,D均不可逆时，我们可用“摄

动法”验证降阶公式仍成立！

【例 2.3.8】试计算 $D_5=\begin{vmatrix}2&1&0&1&-1\\3&2&1&0&2\\1&1&2&-1&5\\0&-1&6&1&4\\4&5&8&7&-1\end{vmatrix}$；

【例 2.3.9】设 $\alpha=(a_1,a_2,\cdots,a_n)$， $a_i\neq 0$， $a_i\in R$， $i=1,2,\cdots,n$；$A=diag\{a_1,a_2,\cdots,a_n\}$，试求 $\det(A-\alpha^{\mathrm{T}}\alpha)$；

【例 2.3.10】设 $A=(a_{ij})_{m\times n}$， $B=(b_{ij})_{n\times m}$ $m\geq n$， $\lambda\neq 0$，试证：$|\lambda I_m-AB|=\lambda^{m-n}\cdot|\lambda I_n-BA|$；

【例 2.3.11】设 $a_i\neq 0$， $i=1,2,\cdots,n$，试求 $\begin{vmatrix}0&a_1+a_2&\cdots&a_1+a_n\\a_2+a_1&0&\cdots&a_2+a_n\\\cdots&\cdots&\cdots&\cdots\\a_n+a_1&a_n+a_2&\cdots&0\end{vmatrix}$；

第三章 线性空间与线性方程组

§0 近世（抽象）代数初步（预备知识）

近世代数的主要内容就是研究所谓的“代数系统”，即带有运算的集合。表示集合的方法通常有两种：列举（枚举）法与描述法，即用它的元素所具有的特性来刻画。

设 A,B 是两个集合，作一个新的集合：$\{(a,b)|a\in A,b\in B\}$，称之为 A,B 的笛卡尔积（简称为卡氏积或直积），记作：$A\times B$；由定义，一般的 $A\times B\neq B\times A$，例如：$A=\{1,2,3\},B=\{4,5\}$；卡氏积的概念可推广为：n 个集合 $A_1,A_2,\cdots,A_n$ 的卡氏积定义为：$A_1\times A_2\times\cdots\times A_n=\{(a_1,a_2,\cdots,a_n)|a_i\in A_i,i=1,2,\cdots,n\}$，常记之为：$\prod_{i=1}^{n}A_i$。

【定义 1】1）设 A,B 是两个集合，则 $A\times B$ 的子集称为 A,B 间的一个二元关系；$\forall a\in A,\forall b\in B$，若 $(a,b)\in R$，称 a,b 有关系 R，记作：aRb；若 $(a,b)\notin R$，称 a,b 不具有关系 R，记作：$a\bar{R}b$；2）$A\times A$ 的子集 R 称为集合 A 上的一个二元关系；3）设～是集合 A 上的一个二元关系，若～满足：

①自反性（反身性）：$\forall a\in A,a\sim a$；②对称性：$\forall a,b\in A,a\sim b\Rightarrow b\sim a$；③传递性：$\forall a,b,c\in A,a\sim b,b\sim c\Rightarrow a\sim c$；则称～是 A 上的一个等价关系；此时，若 $a\sim b$，称 a,b 等价。

【例 3.0.1】1、$\forall a,b\in Z$，规定$aRb\Leftrightarrow 4|(a+b)$，问：$R$是否为$Z$上的一个二元关系；若是，是否为一个等价关系？2、判断如下的二元关系是否为等价关系？

① $A=\{平面上的所有直线\}$，$\forall l_1,l_2\in A, l_1Rl_2\Leftrightarrow l_1\perp l_2$；

② $\forall A,B\in M_n(F), A\sim B\Leftrightarrow$存在可逆矩阵$P,Q$使得$A=PBQ$；

【定义 2】若把集合A的全体元素分成若干互不相交的子集，即任何两个不同的子集都无共同的元素，则每个这样的子集叫做A的一个类；类的全体叫做A的一个分类；

【注】集合的分类与集合的等价关系有着密切的联系—集合A的一个分类可以决定A上的一个等价关系；反之，集合A上的一个等价关系也可以决定A的一个分类！

$\Rightarrow$规定$a\sim b\Leftrightarrow a,b$在同一类；易见$a\sim a$，$a\sim b\Rightarrow b\sim a$，$a\sim b, b\sim c\Rightarrow a\sim c$，从而，～是$A$上的一个等价关系；

$\Leftarrow$将A中所有同固定元素a等价的元素都放在一起，记A的此子集为$[a]$，所有这样的子集就组成A的一个分类：① $a\sim b\Rightarrow [a]=[b]$；② A的每个元素a的确属于某一类；③ A的每个元素a只能属于一个类。

【定义 3】设～是集合A上的一个等价关系，由A在等价关系下的全体不同等价类组成的集合称为A关于的商集，记作：$A/\sim$。举例说明！

【定义 4】一个$A\times B$到D的映射称为一个$A\times B$到D的二元代数运算，简称代数运算；特别地，当$A=B=D$时，$A\times A$到A的代数运算简称为A上的代数运算。通常用“$\circ$”来表示代数运算，并将(a,b)在$\circ$下的像记作：$a\circ b$。

【注】一个代数运算只是一种特殊的映射；由定义，$\forall a\in A, \forall b\in B$，就可以通过这个代数运算，得到$D$一个的元素$d$，即：$(a,b)\to d=a\circ b$；也可以说，所给的代数运算能够对$a,b$进行运算而得到一个结果$d$，这正是普通计算法的特征。如：① $A=Z, B=Z/\{0\}, D=Q$，则$\circ:(a,b)\mapsto \dfrac{a}{b}=a\circ b$是一个$A\times B$到$D$的代数运算，此即普通的除法运算；② $A=Z$，则$\circ:(a,b)\mapsto \sqrt{a^2+b^2}$不是$A$上的代数运算。

【例 3.0.2】设$A=Z^+$，下列哪些法则是A上的代数运算？

①$a\circ b=a^b$；②$a\circ b=a+b-2$；③$a\circ b=a$；

一个代数运算可以是任意规定的，并不一定要有多大的意义；如果任取几个集合，任意规定几个代数运算，很难希望由此会得到什么满意的结果；故常常须对集合的代数运算加以限制。事实上，数、多项式、矩阵、函数等的普通运算，一般都满足通常的运算规则，如：结合律、交换律或分配律等！

【定义 5】设$\circ$是集合A上的一个代数运算，若$\forall a,b,c\in A$，都有：$(a\circ b)\circ c=a\circ(b\circ c)$，则称$\circ$满足结合律。

数、多项式、函数与矩阵对普通的加法与数乘运算都满足结合律，但一般的代数运算则不一定满足！讨论：满足结合律的代数运算对元素的运算会带来何种影响——对于满足结合律的代数运算来说，任意个元素只要不改变元素的前后次序，就可以任意结合而不必再加括号！

【定义 6】设$\circ$是集合A上的一个代数运算，若$\forall a,b\in A$，都有：$a\circ b=b\circ a$，则称$\circ$满足交换律。

【例 3.0.3】判定有理数集上的下列代数运算$\circ$是否满足结合律、交换律；①$a\circ b=a+b+ab$；②$a\circ b=(a+b)^2$；③$a\circ b=a$；④$a\circ b=b^3$。

当A,B是有限集时，$A\times B$到D的代数运算常常用一个矩形表给出；若 $A=\{a_1,a_2,\cdots,a_m\},B=\{b_1,b_2,\cdots,b_n\}$，则$A\times B$到$D=\{d_{ij}\,|\,i=1,2,\cdots,m;\,j=1,2,\cdots,n\}$的一个代数运算可以表示为：

$$\begin{array}{ccccc}\circ & b_1 & b_2 & \cdots & b_n\\ a_1 & d_{11} & d_{12} & \cdots & d_{1n}\\ a_2 & d_{21} & d_{22} & \cdots & d_{2n}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ a_m & d_{m1} & d_{m2} & \cdots & d_{mn}\end{array}$$

，这个表通常称为运算表或 **Cayley** 表。

【例 3.0.4】1）设$A=\{a,b,c,d\}$，由表
$$\begin{array}{ccccc}\circ & a & b & c & d\\ a & a & b & c & d\\ b & b & d & a & c\\ c & c & a & b & d\\ d & d & c & a & b\end{array}$$
所给的代数运算是否满足交换律？**2）**设$A=\{a,b,c\}$上的代数运算$\circ$满足结

合律和交换律，试完成运算表中的运算：

$\circ$	a	b	c
a	a		
b	b	c	a
c	c		

。

结合律和交换律都只同一种代数运算发生关系，而分配律是同两种代数运算发生关系的一种规律。

【定义 7】①设$\odot$是$B\times A$到A的代数运算，$\oplus$是A上的代数运算，若：$\forall a_1,a_2\in A,b\in B$，都有：$b\odot(a_1\oplus a_2)=(b\odot a_1)\oplus(b\odot a_2)$，则称$\odot$对$\oplus$满足左分配律；②设$\otimes$是$A\times B$到$A$的代数运算，$\oplus$是$A$上的代数运算，若$\forall a_1,a_2\in A,b\in B$，都有：$(a_1\oplus a_2)\otimes b=(a_1\otimes b)\oplus(a_2\otimes b)$，则称$\otimes$对$\oplus$满足右分配律。

【定义 8】设$\circ$是集合A上的一个代数运算，①若$\forall a,b,c\in A$，$a\circ b=a\circ c\Rightarrow b=c$，则称$\circ$满足左消去律；②若$\forall a,b,c\in A$，$b\circ a=c\circ a\Rightarrow b=c$，则称$\circ$满足右消去律。

易知：数的加法运算与非零数的乘法运算都适合左、右消去律。

群只有一种代数运算；一个代数运算用什么符号来表示，可以自由决定，有时可用“$\circ$”表示，有时可用“$\bar{\circ}$”表示；为方便起见，常用普通的乘法符号来表示，即：不写$a\circ b$而写ab；因此，也常把一个群的代数运算称作“乘法”。

【定义 9—群】称一个非空集合G关于一个叫做乘法的代数运算构成一个群，如果：①G对于乘法是封闭的；②结合律成立：$\forall a,b,c\in G$，$a(bc)=(ab)c$；③G中有一个元素e，满足：$ae=ea=a$，称e是G的单位元（素）；④G中每一个元素都有逆元，即：$\forall a\in G,\exists b\in G$，使得$ab=e$。

【例 3.0.5】1）整数加群$(Z,+)$、实数加群$(R,+)$、$(R^+,\cdot)$；2）设$G=\{(a,b)|a,b\in R,a\neq 0\}$，在$G$上定义乘法：$(a,b)(c,d)=(ac,ad+b)$，试证：$G$关于该乘法构成一个群。

§1 线性空间的概念

线性代数就是线性空间的理论，线性空间是线性代数研究的最基本的几何对象，矩阵则是研究线性空间诸多几何问题的最有效的代数工具。线性空间的许多几何问题最终可归

结为相应的矩阵分类问题，这个关系也恰好构建了线性代数两大理论—矩阵理论和线性空间理论之间的联系！

已经认识了这样一些数集：自然数集、整数集、有理数集、实数集和复数集，而且已在其上定义了加、减、乘、除运算；将这些已知的数集的特征抽象出来，有如下定义，

【数域】设 F 是复数集的子集，且至少有两个不同的元素，若 F 中任意两个数的加法、减法、乘法及除法（除数不为零）运算结果仍属于 F，则称 F 是一个数域。

由于任意两个相同数差是 **0**，任意两个非零数商是 **1**；因此，任意数域必包含 **0**、**1** 两个数；可以验证：有理数域是一个最小的数域；由于 **1** 属于任一数域，将 **1** 连加任意次，则有任意正整数属于数域；同理可推知任一有理数都在数域里，从而任一数域必包含有理数域！

【例 **3.1.1**】试验证所有形如 $a+b\sqrt[3]{2}+c\sqrt[3]{4}$ 的数构成一个数域，其中 a,b,c 均为有理数；

下面讨论向量！如果建立一个平面直角坐标系，平面上任一点 C 均可由一个有序数组 (a,b) 来表示，这里 a,b 分别为点 C 的横、纵坐标；反之，若给定一个有序数组 (a,b)，必有平面上一点与之对应；这样，平面上的点与有序实数偶之间可建立起一一对应关系。平面上的向量即指平面上的一根有向线段，平面上任一以原点 O 为始点的向量均可对应一个实数序偶；从而，可用实数序偶来代替平面上以原点为始点的全体向量；亦可直接把实数序偶定义为起点为原点的平面向量。将平面向量“代数化”的好处就是：（**1**）可以用代数工具研究几何对象；（**2**）易于推广至一般情形，即所谓“**n** 维向量”；这种推广已不仅仅是形式上的，它对于数学的发展及其应用起着极其重要的作用！

【定义】设 F 是一数域，$a_i \in F$，$i=1,2,\cdots,n$，由 $a_1,a_2,\cdots,a_n$ 组成的有序数组 $\alpha=(a_1,a_2,\cdots,a_n)$ 称为数域 F 上的一个 **n** 维向量。

【注**1**】**1**） 有时也将行向量表示成列向量的形式，并统称二者为向量；**2**）两个行（列）向量相等，当且仅当二者的分量对应相等；**3**）**n** 维行（列）向量实际上是 $1\times n\,(n\times 1)$ 阵，即为行（列）矩阵，从而对数域 F 上的 **n** 维行（列）向量定

义的加（减）法运算及数乘运算，其与矩阵的同类运算并无二致；

4）零向量与负向量即为特殊的零矩阵与非负矩阵。

向量的运算有如下运算规则：

（1）加法交换律：$\alpha+\beta=\beta+\alpha$；（2）加法结合律：$(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$；（3）$\alpha+\theta=\alpha$；（4）$\alpha+(-\alpha)=\theta$；（5）$1\alpha=\alpha$；（6）$k(\alpha+\beta)=k\alpha+k\beta$；（7）$(k+l)\alpha=k\alpha+l\alpha$；（8）$k(l\alpha)=(kl)\alpha$；这里，$k,l\in F$，$\alpha,\beta,\gamma$是数域$F$上的同维行（列）向量。

【例 3.1.2】求解向量方程$2x+3\beta=3x+\gamma$，$\beta=(2,0,1)$，$\gamma=(3,1,-1)$；

基于直观的实二维（三维）空间，抽象其可作进一步推广，

【线性空间】设F是数域，V是一集合，若在V上定义一个加法“$+$”，即对V中任意两个元素α与β，总存在V中唯一的元素γ与之对应，记为：$\gamma=\alpha+\beta$；也即：$\forall\alpha,\beta\in V$，$\alpha+\beta\in V$；在数域F与V之间还定义一种运算，称为数乘“$\bullet$”，即对数域F中任一数k及V中任一元素α，在V中总存在唯一的元素δ与之对应，记为：$\delta=k\cdot\alpha=k\alpha$；也即：$\forall k\in F$，$\forall\alpha\in V$，$k\cdot\alpha=k\alpha\in V$；若上述运算有如下运算规则：1）加法交换律：$\alpha+\beta=\beta+\alpha,\ \alpha,\beta\in V$；2）加法结合律：$(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma),\alpha,\beta,\gamma\in V$；3）存在元素$\theta\in V$，使得$\forall\alpha\in V$，有$\alpha+\theta=\theta+\alpha=\alpha$；4）$\forall\alpha\in V$，$\exists\beta\in V$，使得$\alpha+\beta=\theta$，记$\beta=-\alpha$；5）$\forall\alpha\in V$，$1\alpha=\alpha$；6）$k(\alpha+\beta)=k\alpha+k\beta$；7）$(k+l)\alpha=k\alpha+l\alpha$；8）$k(l\alpha)=(kl)\alpha$；则称$V$是数域$F$上的线性空间（向量空间），$V$中元素称为向量，元素$\theta$称为零向量，满足 4）的$\beta$称为$\alpha$的负向量。

【例 3.1.3】1）在$C[0,1]$上定义函数的加法运算及数乘运算，$(f+g)(x)=f(x)+g(x)$，$x\in[0,1]$，$f,g\in C[0,1]$；$(cf)(x)=cf(x)$，$c\in R$；试验证：$C[0,1]$是实数域R上的线性空间；2）试验证全体n阶上三角实矩阵在矩阵的加法及数乘运算下是实数域R上的线性空间；3）设V是以零为极限的实数数列全体，即：$V=\left\{\{a_n\}\middle|\lim\limits_{n\to\infty}a_n=0\right\}$，定义两个数列的加法及数乘为：$\{a_n\}+\{b_n\}=\{a_n+b_n\}$，$c\{a_n\}=\{ca_n\}$；试验证在$V$上述运算下是实数

域R上的线性空间；

【例 3.1.4】试检验下列集合对指定的加法及数乘运算是否构成实数域上的线性空间：**1）** $R^{+}, a\oplus b=ab, \lambda\circ a=a^{\lambda}$；**2）** R^{2}，向量加法，$\lambda\circ(x,y)=\begin{cases}(0,0),\lambda=0;\\ \left(\lambda x,\dfrac{y}{\lambda}\right),\lambda\neq 0;\end{cases}$；**3）** $R^{2},(x,y)+(x_1,y_1)=(x+x_1,0),$ $\lambda(x,y)=(\lambda x,0)$；**4）**；**n** 维行向量空间$R^{n}$，$\alpha\oplus\beta=\alpha-\beta,\lambda*\alpha=-\lambda\alpha$；

【例 3.1.5】1）无穷多个向量的和是否是向量？若不是，试举出反例；**2）**已知$V=\left\{x\middle|x=(x_1,x_2,x_3)\in R^{3},且x_1+x_2=a\right\}$是向量空间，试求$a$；

以下简单地讨论线性空间的基本性质；

①线性空间V的零向量是唯一的；

不妨设θ_1,θ_2均是零向量，易见，$\theta_1=\theta_1+\theta_2=\theta_2+\theta_1=\theta_2$。

②线性空间的每个向量的负向量是唯一的；

不妨设β,γ均是α的负向量，即：$\alpha+\beta=\alpha+\gamma=\theta$；易见，$\beta=\beta+\theta=\beta+(\alpha+\gamma)=(\beta+\alpha)+\gamma=\theta+\gamma=\gamma$。

③加法有消去律，即由$\alpha+\beta=\alpha+\gamma$可推出$\beta=\gamma$；两边同时加上$-\alpha$，即有：$(-\alpha)+(\alpha+\beta)=(-\alpha)+(\alpha+\gamma)$，即：$(-\alpha+\alpha)+\beta=(-\alpha+\alpha)+\gamma$，$\beta=\gamma$。

④加法可移项，即由$\alpha+\beta=\gamma$可推出$\alpha=\gamma-\beta$，

$\alpha=\alpha+\theta=\alpha+(\beta+(-\beta))=(\alpha+\beta)+(-\beta)=\gamma-\beta$；

⑤$0\cdot\alpha=\theta$，$(-1)\alpha=-\alpha$，$k\cdot\theta=\theta$；

⑥若$k\alpha=\beta,k\neq 0$，则$\alpha=\dfrac{1}{k}\beta$；特别地，若$k\alpha=\theta$，则$k=0$或$\alpha=\theta$。

“抽象的线性空间”的引入扩大了我们的视野，它把众多不同研究对象的共同特点用“线性空间”加以概括，从而极大地扩大了代数学理论的应用范围；线性空间的理论是线性代数（或高等代数）的核心！

以下接着来讨论向量的线性关系。在线性空间理论中，向量之间最基本的关系为：线性相关、线性无关和线性组合。

【定义】设V是数域F上的线性空间，$\alpha_1,\alpha_2,\cdots,\alpha_m$是$V$中$m$个向量，若存在$m$个不全为零的数$k_1,k_2,\cdots,k_m\in F$，使得$k_1\alpha_1+k_2\alpha_2+\cdots+k_m\alpha_m=\theta$（零向量），则称$\alpha_1,\alpha_2,\cdots,\alpha_m$线性相关；反之，若$F$中

不存在m个不全为零的数使得上式成立，则称$\alpha_1,\alpha_2,\cdots,\alpha_m$线性无关。

【注1】1） 在上述定义中，$k_1,k_2,\cdots,k_m$必须取自数域F；例如：若把复数集视作实数域上的线性空间，则向量$1,i$线性无关；**2）** 线性无关可等价地定义：若存在$k_1,k_2,\cdots,k_m\in F$，使得$k_1\alpha_1+k_2\alpha_2+\cdots+k_m\alpha_m=\theta$，则必有$k_1=k_2=\cdots=k_m=0$；线性无关又称线性独立。

【定义】设V是域F上的线性空间，$\alpha_1,\alpha_2,\cdots,\alpha_m,\beta\in V$，若存在$k_1,k_2,\cdots,k_m\in F$，使得$\beta=k_1\alpha_1+k_2\alpha_2+\cdots+k_m\alpha_m$，则称$\beta$是$\alpha_1,\alpha_2,\cdots,\alpha_m$的线性组合或称$\beta$可由$\alpha_1,\alpha_2,\cdots,\alpha_m$线性表示（出）。

【例 3.1.6】1）$R_2[x]$中下列各向量是否对R线性相关：$f_1=2x^2-3x+2,\ f_2=x^2-1, f_3=x, f_4=4$；**2）**试判断$M_2(R)$中下列向量是否对$R$线性相关：$A_1=\begin{pmatrix}1&2\\3&4\end{pmatrix},A_2=\begin{pmatrix}-1&1\\1&1\end{pmatrix},A_3=\begin{pmatrix}0&2\\-1&-1\end{pmatrix}$；**3）**若向量$\alpha_1=\begin{pmatrix}1\\0\\2\end{pmatrix},\alpha_2=\begin{pmatrix}-1\\2\\1\end{pmatrix},\alpha_3=\begin{pmatrix}2\\k\\5\end{pmatrix}$线性无关，试求$k$；**4）**设向量$\alpha,\beta,\gamma$线性无关，试证：向量$\alpha+\beta,\beta+\gamma,\gamma+\alpha$也线性无关。

【注 2】1）若向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$线性相关，则任一包含这组向量的向量组必线性相关；若$\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$线性无关，则从这组向量中任取一组向量必线性无关；也即："部分相关则整体相关，整体无关则部分无关"；**2）**设$\alpha_1,\alpha_2,\cdots,\alpha_m$是域$F$上线性空间$V$的$m$个向量，则$\alpha_1,\alpha_2,\cdots,\alpha_m$线性相关的充要条件是其中至少有一个向量可由其余向量线性表示；**3）**若向量组线性无关，则其"加长向量组"也线性无关；若向量组线性相关，则其"截断向量组"也线性相关。

来看这样一些简单事实：①向量组$\alpha_1,\alpha_2,\cdots,\alpha_r$中任一向量$\alpha_i\,(1\le i\le r)$都可由这组向量线性表出； ②若向量$\gamma$可由向量$\beta_1,\beta_2,\cdots,\beta_r$线性表出，而每一向量$\beta_i$又可由向量$\alpha_1,\alpha_2,\cdots,\alpha_s$线性表出，则$\gamma$必可由向量$\alpha_1,\alpha_2,\cdots,\alpha_s$线性表出； ③若向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$线性无关，而向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_r,\beta\}$线性相关，则$\beta$必可由$\alpha_1,\alpha_2,\cdots,\alpha_r$线性表出；

【定义】　设$\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$和$\{\beta_1,\beta_2,\cdots,\beta_s\}$是线性空间$V$的两个向量组；若每一$\alpha_i$都可由$\beta_1,\beta_2,\cdots,\beta_s$线性表出且每一$\beta_i$都可由$\alpha_1,\alpha_2,\cdots,\alpha_r$线性表出，也即：两向量组可相互线性表出，则称两向量组等价。

可以验证：向量组的等价关系具有自反性、对称性和传递性，从而其是二元关系中的等价关系！

【定理】设向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$线性无关，且每一向量$\alpha_i$都可由向量组$\{\beta_1,\beta_2,\cdots,\beta_s\}$线性表示，则$r\le s$；（简而言之，此即：一个线性无关组不可能由比其数量更少的一组向量线性表出—课堂论证）

【推论 1】两个等价的线性无关的向量组含有相同个数的向量；

【定义】设线性空间V中有一组向量，若该组向量中存在一组向量$\alpha_1,\alpha_2,\cdots,\alpha_m$满足：（ⅰ）$\alpha_1,\alpha_2,\cdots,\alpha_m$线性无关；（ⅱ）在该组向量中任取一个向量$\beta$，则向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_m,\beta\}$线性相关；则称$\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$是这组向量的极大（线性）无关组。

【推论 2】等价的向量组的极大无关组含有相同个数的向量，一个向量组的任意两个极大无关组（若存在）含有相同个数的向量。

【定义】向量组的极大无关组所含向量的个数称为向量组的秩。

易知，全由零向量组成的向量组秩规定为零；向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$的秩记作：　$r\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$或$rank\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$，$R\{\alpha_1,\alpha_2,\cdots,\alpha_r\}, Rank\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$。易见，向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$线性无关的充要条件是$r\{\alpha_1,\alpha_2,\cdots,\alpha_r\}=r$。

【推论 3】若向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_r\}$可由向量组$\{\beta_1,\beta_2,\cdots,\beta_s\}$线性表出，则$r\{\alpha_1,\alpha_2,\cdots,\alpha_r\}\le r\{\beta_1,\beta_2,\cdots,\beta_s\}$；

【推论 4】等价的向量组具有相同的秩。

【例 3.1.7】设向量组$\alpha_1=\begin{pmatrix}2\\0\\0\end{pmatrix},\alpha_2=\begin{pmatrix}-1\\3\\0\end{pmatrix},\alpha_3=\begin{pmatrix}7\\-4\\0\end{pmatrix}$，试求$\{\alpha_1,\alpha_2,\alpha_3\}$的一个极大无关组及其秩；（注：该题的解法应需推广至更一

般的情形！）

【定义】称由矩阵的各列向量组成的向量组的秩为矩阵的列秩，称由矩阵的各行向量组成的向量组的秩为矩阵的行秩。

【引例】试求矩阵 $A=\begin{pmatrix}1&0&1&2\\2&1&0&1\\1&3&0&2\end{pmatrix}$ 的行秩及列秩；

【定理】矩阵的行秩与列秩在初等变换下不变；

不妨先来验证：<u>矩阵的行秩在初等行变换下保持不变</u>

—将矩阵按行分块，$A=\left(a_{ij}\right)_{m\times n}=\begin{pmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_m\end{pmatrix}$；由于 $P_m(ij)A=\begin{array}{c}\begin{pmatrix}\alpha_1\\\vdots\\\alpha_j\\\vdots\\\alpha_i\\\vdots\\\alpha_m\end{pmatrix}\end{array}\begin{array}{l}\\\\\text{第}i\text{行}\\\\\text{第}j\text{行}\\\\\\\end{array}$

$=A_1$，易见 A 的行向量组与 A_1 的行向量组可相互线性表出，即为等价，$P_m(i(k))A,P_m(ij(c))A$ 的情形类似，不再累述！再来验证：<u>矩阵的列秩在初等行变换下保持不变</u>—将矩阵按列分块，$A=\left(a_{ij}\right)_{m\times n}=\left(A_1,A_2,\cdots,A_n\right)$，$Q\left(A_1,A_2,\cdots,A_n\right)=\left(QA_1,QA_2,\cdots,QA_n\right)$，$Q$ 为初等矩阵；不妨设 A 的列向量组的一个极大无关组 $\left\{A_1,A_2,\cdots,A_t\right\}$，则：**1**）设 $k_1,k_2,\cdots,k_t\in F$，且有：$k_1QA_1+k_2QA_2+\cdots+k_tQA_t=\theta$，即：$Q\left(k_1A_1+k_2A_2+\cdots+k_tA_t\right)=\theta$，由 Q 非异及 $A_1,A_2,\cdots,A_t$ 线性无关，即有：$k_1=k_2=\cdots=k_t=0$；从而，$QA_1,QA_2,\cdots,QA_t$ 线性无关；**2**）由于 $\left\{A_1,A_2,\cdots,A_t\right\}$ 是 $\left\{A_1,A_2,\cdots,A_n\right\}$ 的极大无关组；从而，$\forall A_j,1\le j\le n$，$A_j=\sum_{k=1}^{t}l_{jk}A_k$，即有：$\forall QA_j,1\le j\le n$，$QA_j=\sum_{k=1}^{t}l_{jk}QA_k$；也即：$A,QA$ 的列向量组的极大无关组具有相同个数的向量，且列向量之间也具有相同的线性关系；其他情形类似，不再累述！

【推论 **1**】任一矩阵的行秩等于列秩；

以后，通称矩阵的行秩、列秩为矩阵的秩，记作：$r(A)$、$R(A),rank(A),Rank(A)$。

【推论 **2**】$r(A)=r\left(A^{\mathrm{T}}\right)$；

【推论 **3**】任一矩阵与一非异阵相乘，其秩不变；

【推论 4】n阶可逆阵的秩等于n，故可逆阵又称满秩阵；

【推论 5】n阶方阵A的n个行（列）向量线性无关的充要条件是$|A|\neq 0$。

【矩阵秩的行列式判别定理】若$m\times n$阵$A=\left(a_{ij}\right)_{m\times n}$有一个$r$阶子式不为零，且$A$中任意$r+1$阶子式（若存在）都等于零，则$r(A)=r$；若$r(A)=r$，则$A$中必有一个$r$阶子式不为零，而所有的$r+1$阶子式（若存在）都等于零。

【例 3.1.8】用初等变换法求矩阵$A=\begin{pmatrix}1&2&0\\0&1&1\\-1&2&3\end{pmatrix}$的秩；

【例 3.1.9】试证：若A,B为同型矩阵，则$r(A+B)\leq r(A)+r(B)$，举例说明可能出现$r(A+B)=r(A)+r(B)$；

【例 3.1.10】设$A=\left(a_{ij}\right)_{m\times n}$，$B=\left(b_{ij}\right)_{n\times p}$，试证：$r(AB)\leq \min\{r(A),r(B)\}$；

【例 3.1.11】1）设$a_i,b_i\neq 0$，$i=1,2,\cdots,n$，试求矩阵$A=\left(a_{ij}\right)_{n\times n}=\left(a_ib_j\right)_{n\times n}$的秩；2）设$m>n$，$A=\left(a_{ij}\right)_{m\times n}$，$B=\left(b_{ij}\right)_{n\times m}$，试证：$|AB|=0$。

【例 3.1.12】设$r\left(A_{3\times 3}\right)=2$，则存在实数$k$，使得$\left(A^*\right)^2=kA^*$；

§2 线性方程组及其解

先来看线性方程组解的判别定理：

线性方程组$\boxed{\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2\\\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m\end{cases}}$，即：$\boxed{AX=\beta}$（矩阵方程），

这里，$A=\left(a_{ij}\right)_{m\times n}$，$X=\left(x_1,x_2,\cdots,x_n\right)^{\mathrm{T}}$，$\beta=\left(b_1,b_2,\cdots,b_m\right)^{\mathrm{T}}$；若将系数矩阵按列分块为：$A=\left(\alpha_1,\alpha_2,\cdots,\alpha_n\right)$，线性方程组即如下向量方程：$\boxed{x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n=\beta}$，$A$的增广矩阵$A=\left(A\ \vdots\ \beta\right)=\left(\alpha_1,\alpha_2,\cdots,\alpha_n,\beta\right)$，则：1）若$r(A)=r\left(A\right)=n$，方程组有且只有一组解；

2）若$r(A)=r\left(A\right)<n$，方程组有无穷多解；3）若$r(A)\neq r\left(A\right)$，方程组无解。

在方程组$AX=\beta$中，若$\beta=(0,0,\cdots,0)^{\mathrm{T}}$，称之为齐次方程组；否则为非齐次方程组。对于齐次方程组$AX=\theta$，$r(A)=r\left(A\right)$，

$A=(A\ \vdots\ \theta)$；从而，**1**）若$r(A)<n$，齐次方程组有无穷多解；**2**）若$r(A)=n$，齐次方程组有唯一解—零解；此时，称齐次组只有平凡解。

以下，<u>从线性空间的观点</u>讨论齐次方程组$AX=\theta$的解，它的任一解都可视作 **n** 维列向量空间的一个元素（向量）：若α,γ是解，即：$A\alpha=A\gamma=\theta$，从而$A(\alpha+\gamma)=\theta$，即：$\alpha+\gamma$也是解；若α是解，则$\forall k\in F$，$A(k\alpha)=k(A\alpha)=\theta$，即：$k\alpha$也是解。可以验证：$\boxed{\{X\,|AX=\theta,X\in F_n\}}$对于向量的加法及数乘运算构成数域$F$上的一个线性空间，称之为$F_n$的线性子空间，又可称为齐次方程组的解空间！$\{\theta\}$和$F_n$称为$F_n$的平凡子空间，其余的子空间称为非平凡子空间。

接着我们讨论方程组$AX=\theta$的<u>解空间的结构</u>！

【定义】齐次线性方程组$AX=\theta$有非零解时，若$\eta_1,\eta_2,\cdots,\eta_t$是其解且满足：**1**）$\eta_1,\eta_2,\cdots,\eta_t$线性无关；**2**）$AX=\theta$的每一解都可由解向量$\eta_1,\eta_2,\cdots,\eta_t$线性表出；则称$\eta_1,\eta_2,\cdots,\eta_t$是$AX=\theta$的一个基础解系！若齐次组$AX=\theta$有一个基础解系：$\eta_1,\eta_2,\cdots,\eta_t$，则它的解集$W=\{k_1\eta_1+k_2\eta_2+\cdots+k_t\eta_t\,|k_i\in F,i=1,2,\cdots,t\}$的代表元素$k_1\eta_1+k_2\eta_2+\cdots$ $+k_t\eta_t$称为齐次线性方程组的通解，其中$k_i\in F,i=1,2,\cdots,t$。

【定理】数域F上n元齐次线性方程组$AX=\theta$的系数矩阵的秩$r(A)=r<n$，它一定有基础解系；且它的每一个基础解系所含解向量的个数等于$n-r$。

由于$r(A)=r$，适当交换方程式的次序并除去多余的方程，得到的一个同解方程组$\begin{cases}a_{11}x_1+\cdots+a_{1r}x_r+a_{1,r+1}x_{r+1}\cdots+a_{1n}x_n=0\\ a_{21}x_1+\cdots+a_{2r}x_r+a_{2,r+1}x_{r+1}\cdots+a_{2n}x_n=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ a_{r1}x_1+\cdots+a_{rr}x_r+a_{r,r+1}x_{r+1}\cdots+a_{rn}x_n=0\end{cases}(1)$；由于$r(A)=r$，（**1**）的系数矩阵必有一个$r$阶子式不为零，不妨设$\begin{vmatrix}a_{11}&a_{12}&\cdots&a_{1r}\\ a_{21}&a_{22}&\cdots&a_{2r}\\ \cdots&\cdots&\cdots&\cdots\\ a_{r1}&a_{r2}&\cdots&a_{rr}\end{vmatrix}\neq0$，即有：$\begin{cases}a_{11}x_1+\cdots+a_{1r}x_r=-a_{1,r+1}x_{r+1}-\cdots-a_{1n}x_n\\ a_{21}x_1+\cdots+a_{2r}x_r=-a_{2,r+1}x_{r+1}-\cdots-a_{2n}x_n\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ a_{r1}x_1+\cdots+a_{rr}x_r=-a_{r,r+1}x_{r+1}-\cdots-a_{rn}x_n\end{cases}(2)$，（**2**）中共有$n-r$个参数（自由未知量）$x_{r+1},x_{r+2},\cdots,x_n$，若（**2**）

可解得：$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} c_{1,r+1} & c_{1,r+2} & \cdots & c_{1n} \\ c_{2,r+1} & c_{2,r+2} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{r,r+1} & c_{r,r+2} & \cdots & c_{rn} \end{pmatrix} \begin{pmatrix} x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix} (2')$，依次取$\begin{pmatrix} x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix}$为

$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$，从而有（1）的$n-r$个解：

$\eta_1 = \begin{pmatrix} c_{1,r+1} \\ \vdots \\ c_{r,r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$，$\eta_2 = \begin{pmatrix} c_{1,r+2} \\ \vdots \\ c_{r,r+2} \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$，……，$\eta_{n-r} = \begin{pmatrix} c_{1n} \\ \vdots \\ c_{rn} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$；易见，$\eta_1, \eta_2, \cdots, \eta_{n-r}$线性无关；现设$\eta$是$AX = \theta$的任一解向量，且设

$\eta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$，则$\eta$须满足上述$(2')$，即：$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} = \begin{pmatrix} c_{1,r+1} & c_{1,r+2} & \cdots & c_{1n} \\ c_{2,r+1} & c_{2,r+2} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{r,r+1} & c_{r,r+2} & \cdots & c_{rn} \end{pmatrix} \begin{pmatrix} a_{r+1} \\ a_{r+2} \\ \vdots \\ a_n \end{pmatrix}$，

从而，

$$\eta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} \\ \begin{pmatrix} a_{r+1} \\ a_{r+2} \\ \vdots \\ a_n \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} c_{1,r+1} & c_{1,r+2} & \cdots & c_{1n} \\ c_{2,r+1} & c_{2,r+2} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{r,r+1} & c_{r,r+2} & \cdots & c_{rn} \end{pmatrix} \begin{pmatrix} a_{r+1} \\ a_{r+2} \\ \vdots \\ a_n \end{pmatrix} \\ E_{n-r} \begin{pmatrix} a_{r+1} \\ a_{r+2} \\ \vdots \\ a_n \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} c_{1,r+1} & c_{1,r+2} & \cdots & c_{1n} \\ c_{2,r+1} & c_{2,r+2} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{r,r+1} & c_{r,r+2} & \cdots & c_{rn} \end{pmatrix} \\ E_{n-r} \end{pmatrix} \begin{pmatrix} a_{r+1} \\ a_{r+2} \\ \vdots \\ a_n \end{pmatrix}$$

$= (\eta_1, \eta_2, \cdots, \eta_{n-r}) \begin{pmatrix} a_{r+1} \\ a_{r+2} \\ \vdots \\ a_n \end{pmatrix} = a_{r+1}\eta_1 + a_{r+2}\eta_2 + \cdots + a_n\eta_{n-r}$，也即：$AX = \theta$的任一解均可表示为$\eta_1, \eta_2, \cdots, \eta_{n-r}$的线性组合；从而，$\eta_1, \eta_2, \cdots, \eta_{n-r}$即为$AX = \theta$的一个基础解系！

最后，我们来讨论非齐次线性方程组$AX = \beta$的解的结构！

易见，若α, γ是$AX = \beta$的两个解向量，即：$A\alpha = A\gamma = \beta$，即

有：$A(\alpha-\gamma)=A\alpha-A\gamma=\theta$，也即：$\alpha-\gamma$是导出组$AX=\theta$的解；反之，若$A\alpha=\beta$，$A\gamma=\theta$，则$A(\alpha+\gamma)=A\alpha+A\gamma=\beta$，即：$\alpha+\gamma$是非齐次方程组$AX=\beta$的解。于是，

【定理】若数域F上n元非齐次线性方程组$AX=\beta$有解，则其解集为$U=\{\gamma_0+\eta|\eta\in W\}$，其中$\gamma_0$是$AX=\beta$的一个解（称$\gamma_0$是特解），$W$是其导出组$AX=\theta$的解集；也将$U$记为$\gamma_0+W$。

【推论】若$AX=\beta$有解，则其解唯一的充要条件是其导出组$AX=\theta$只有零解。

当$AX=\beta$有无穷多解时，$AX=\theta$必有非零解；取$AX=\theta$的一个基础解系：$\eta_1,\eta_2,\cdots,\eta_{n-r}$，$r=r(A)$；从而$AX=\beta$的解集$U$为$U=\{\gamma_0+k_1\eta_1+k_2\eta_2+\cdots+k_{n-r}\eta_{n-r}\,|k_i\in F,i=1,2,\cdots,n-r\}$，解集$U$的代表元素$\gamma_0+k_1\eta_1+k_2\eta_2+\cdots+k_{n-r}\eta_{n-r}$为$AX=\beta$的通解。

【例 3.3.1】1）设$A=(a_{ij})_{m\times n}(m\le n)$，试证：对任意$m$维列向量$b$，方程组$AX=b$都有解的充要条件是$R(A)=m$；2）设$A\in M_3(F)$，则$r(A^TA)=r(A)$；3）如果方程组$A_{m\times n}X=b$对任何向量$b\in R^m$都有解，则$A_{m\times n}$的行向量组线性无关；

【例 3.3.2】求一齐次方程组，使$\alpha_1=(-1,1,1,0,0)^{\mathrm{T}},\alpha_2=(7,5,0,2,6)^{\mathrm{T}}$构成它的一个基础解系；

【例 3.3.3】求线性方程组$\begin{cases}x_1-x_2+2x_3+x_4=1\\2x_1-x_2+x_3+2x_4=3\\x_1-x_3+x_4=2\\3x_1-x_2+3x_4=5\end{cases}$的通解，并求出满足$x_1^2=x_2^2$的全部解；

【例 3.3.4】设$A=\begin{pmatrix}1&1&\cdots&1\\a_1&a_2&\cdots&a_n\\a_1^2&a_2^2&\cdots&a_n^2\\\cdots&\cdots&\cdots&\cdots\\a_1^{n-1}&a_2^{n-1}&\cdots&a_n^{n-1}\end{pmatrix}$，$X=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$，$b=\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$，其中，$a_i\ne a_j,i\ne j$，$i,j=1,2,\cdots,n$，求解线性方程组$A^{\mathrm{T}}X=b$；

【例 3.3.5】已知线性方程组$(1)\begin{cases}x_1+x_2=0\\x_2-x_4=0\end{cases},(2)\begin{cases}x_1-x_2+x_3=0\\x_2-x_3+x_4=0\end{cases}$，试求①求方程组（1）、（2）的基础解系；②求方程组（1）、（2）的公共解；

【例 3.3.6】已知方程组(1) $\begin{cases} -2x_1+x_2+ax_3-5x_4=1 \\ x_1+x_2+x_3+bx_4=4 \\ 3x_1+x_2+x_3+2x_4=c \end{cases}$ (2) $\begin{cases} x_1+x_4=1 \\ x_2-2x_4=2 \\ x_3+x_4=-1 \end{cases}$ 是同解方程组，试确定(1)中的系数 a,b,c；

【选读内容】矛盾线性方程组的近似解

对于非齐次线性方程组 $AX=b, A=\left(a_{ij}\right)_{m\times n}$，如果 $r(A)\neq r(A|b)$，则方程组无解；这样的方程组称为矛盾方程组或不相容方程组。在处理一些实际问题时，即使线性方程组不存在精确解，有时却需要研究其近似解。

当线性方程组 $AX=b$ 无解时，理论上我们可选择任意的 n 维列向量作为近似解，那么该选择怎样的“近似标准”来得到“最优”近似解呢？我们从几何上看，若设 $A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$，$X=(x_1,x_2,\cdots,x_n)^T$，则 $AX=(\alpha_1,\alpha_2,\cdots,\alpha_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n$；当向量 $X=(x_1,x_2,\cdots,x_n)^T$ 变化时，AX 生成一个子空间：$W=\left\{AX \,\middle|\, X\in R^n\right\}$ $=\left\{x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n \,\middle|\, x_1,x_2,\cdots,x_n\in R\right\}=L(\alpha_1,\alpha_2,\cdots,\alpha_n)$；从而，$b\in W\Leftrightarrow AX=b$ 有解，也即方程组 $AX=b$ 无解的原因即 $b\notin W$。这里我们不妨把 W 想象为一个平面，向量 b 不在该平面；但我们可在 W 上寻找一个与 b 最“接近”的向量来近似它，易见此向量即为向量 b 在 W 上的投影向量 β。设 $\beta=AX_0$，则 X_0 即为我们所要求的近似解；AX_0 也是 W 中与 b “距离”最近的向量，即：$\|b-AX_0\|=\min\limits_{AX\in W}\|b-AX\|$。从此意义上说，$X_0$ 是上述方程组的最优近似解；此方法称为“最小二乘法”。以下，我们来讨论如何求这个最优近似解！

在上述讨论中，由于 $(b-\beta)\perp W$，故 $(b-\beta)\perp\alpha_i, i=1,2,\cdots,n$；从而，$\alpha_i^T(b-\beta)=0, i=1,2,\cdots,n$，也即：$A^T(b-\beta)=0$，展开后即得，$A^TAX_0=A^Tb$，即最优近似解 X_0 是方程组 $A^TAX=A^Tb$ 的解。因此，我们有如下定理：

设有线性非齐次方程组 $AX=b, A=\left(a_{ij}\right)_{m\times n}$，当 $r(A)\neq r(A|b)$ 时，

1）可由所谓“正规方程组” $A^TAX=A^Tb$ 的解来作为 $AX=b$ 的最

优近似解；2）正规方程组必有解；且当$r\left(A^TA\right)=n$时，A^TA可逆，其解可表示为：$\boxed{X_0=\left(A^TA\right)^{-1}A^Tb}$。

关于 2），一则$r\left(A^TA,A^Tb\right)=r\left[A^T\left(A,b\right)\right]\le r\left(A^T\right)=r\left(A\right)$，再则$r\left(A^TA,A^Tb\right)\ge r\left(A^TA\right)=r\left(A\right)$，从而，$r\left(A^TA,A^Tb\right)=r\left(A^TA\right)$。

【例 3.3.7】试给出下列线性方程组的最优近似解：

1）$\begin{cases}x-y=1\\x+y=2\\2x-y=3\end{cases}$；2）$\begin{cases}x+y+z=1\\x+y+z=2\end{cases}$。

§3 线性空间及线性映射

若V是数域F上n维行向量空间F^n，$e_i=(0\ \cdots\ \underset{\text{第}i\text{个}}{1}\ 0\ \cdots\ 0)$，$i=1,2,\cdots,n$；易见，$e_1,e_2,\cdots,e_n$是$V\left(F^n\right)$中$n$个线性无关的向量；而且，$\forall\alpha=\left(a_1,a_2,\cdots,a_n\right)\in V$，$a_i\in F,i=1,2,\cdots,n$，则有：$\alpha=a_1e_1+a_2e_2+\cdots+a_ne_n$。

【定义】若V是数域F上的线性空间，若存在V中n个线性无关的向量$\alpha_1,\alpha_2,\cdots,\alpha_n$，使得$V$中任一向量均可表示这组向量的线性组合，则称$\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$是$V$的一组基，线性空间$V$称为$n$维线性空间；称$n$为线性空间$V$的维数，记作：$\dim V=n$；若不存在有限个向量组成$V$的一组基，则称$V$为无限维线性空间。

【注1】1） 一般地，线性空间的基不是唯一的；如：在二维行向量空间中，$\{(1,0),(0,1)\}$与$\{(1,0),(1,1)\}$均是基；2）线性空间V的一个极大无关组即为V的一组基；由向量组的等价性，基中向量的个数是确定的。

【例 3.3.1】可以验证：$\dim\left[M_n\left(F\right)\right]=n^2$，$\dim\left[F_n\left(x\right)\right]=n$；复数集$C$是实数域$R$上的 2 维线性空间，$R$是$R$上的 1 维线性空间；$\forall n\in N$，$R[0,2\pi]$中向量组$\{1,\cos x,\cos 2x,\cdots,\cos nx,\sin x,\sin 2x,\cdots,\sin nx\}$对$R$线性无关，从而$R[0,2\pi]$是无限维的；

【例 3.3.2】1）求出下列实线性空间的维数和一组基：

$V_1=\left\{A\middle|A\in M_n\left(R\right),A=A^{\mathrm T}\right\}$，　　$V_2=\left\{A\middle|A\in M_n\left(R\right),A=-A^{\mathrm T}\right\}$，

$V_3=\left\{\left(a_{ij}\right)_{n\times n}\middle|a_{ij}\in R,a_{ij}=0,i>j\right\}$；　　2）在$M_2\left(R\right)$中找出$A_3,A_4$，使得

A_1,A_2,A_3,A_4构成一个基，这里，$A_1=\begin{pmatrix}1&1\\1&-1\end{pmatrix}$，$A_2=\begin{pmatrix}-1&1\\-1&1\end{pmatrix}$；

以后，为书写及定理证明的简洁，引进一种“形式向量”；设V是域F上的线性空间，$\alpha_1,\alpha_2,\cdots,\alpha_n$是$V$中任意$n$个向量，则$(\alpha_1,\alpha_2,\cdots,\alpha_n)$即是由$\alpha_1,\alpha_2,\cdots,\alpha_n$构成的形式向量；形式向量的各分量也是一些抽象向量，这里对它的一些“运算”作如下规定：一个自然的做法，它应与普通向量和矩阵的运算保持一致，①$(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\beta_1,\beta_2,\cdots,\beta_n)\Leftrightarrow\alpha_i=\beta_i,i=1,2,\cdots,n$；

②$(\alpha_1,\alpha_2,\cdots,\alpha_n)\pm(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1\pm\beta_1,\alpha_2\pm\beta_2,\cdots,\alpha_n\pm\beta_n)$； ③$A=(a_{ij})_{n\times n}\in M_n(F)$，$(\alpha_1,\alpha_2,\cdots,\alpha_n)A=\left(\sum\limits_{j=1}^n a_{j1}\alpha_j,\sum\limits_{j=1}^n a_{j2}\alpha_j,\cdots,\sum\limits_{j=1}^n a_{jn}\alpha_j\right)$；若再有$B=(b_{ij})_{n\times n}\in M_n(F)$，$C=(c_{ij})_{n\times m}$，则应有，

④$(\alpha_1,\alpha_2,\cdots,\alpha_n)(A+B)=(\alpha_1,\alpha_2,\cdots,\alpha_n)A+(\alpha_1,\alpha_2,\cdots,\alpha_n)B$；

⑤$[(\alpha_1,\alpha_2,\cdots,\alpha_n)A]C=(\alpha_1,\alpha_2,\cdots,\alpha_n)(AC)$。

设V是域F上的n维线性空间，若给定V的一组基$\alpha_1,\alpha_2,\cdots,\alpha_n$，则$\forall\beta\in V$，$\exists k_1,k_2,\cdots,k_n\in F$，使得$\beta=k_1\alpha_1+k_2\alpha_2+\cdots+k_n\alpha_n$；可以验证，$k_1,k_2,\cdots,k_n$由$\beta$唯一确定！

联系解析几何中，在R^2中，$\forall(a,b)\in R^2$，$(a,b)=a(1,0)+b(0,1)$，这里$(1,0),(0,1)$是R^2的一组基，a,b又可称为横坐标、纵坐标。以下，在抽象的线性空间中引入坐标概念，

【定义】设$\alpha_1,\alpha_2,\cdots,\alpha_n$是域$F$上$n$维线性空间$V$的一组基，$\beta\in V$，且有：$\beta=k_1\alpha_1+k_2\alpha_2+\cdots+k_n\alpha_n=(a_1,a_2,\cdots,a_n)\begin{pmatrix}k_1\\k_2\\\vdots\\k_n\end{pmatrix}$，由该线性表出式所唯一确定的有序数组$(k_1,k_2,\cdots,k_n)$称为在基$\alpha_1,\alpha_2,\cdots,\alpha_n$下的坐标；称$n$维列向量$(k_1,k_2,\cdots,k_n)^{\mathrm{T}}$为$\beta$在基$\alpha_1,\alpha_2,\cdots,\alpha_n$下的坐标向量。

一般地，同一向量在同一空间的不同基下的坐标是不同的！

【例 3.3.3】1）已知三维线性空间的一组基（底）为$\alpha_1=(1,1,0)^{\mathrm{T}}$，$\alpha_2=(1,0,1)^{\mathrm{T}}$，$\alpha_3=(0,1,1)^{\mathrm{T}}$，求向量$\beta=(2,0,0)^{\mathrm{T}}$在上述基（底）下的

坐标；2）在$M_2(R)$中，求$A=\begin{pmatrix}-3 & 5\\ 4 & 2\end{pmatrix}$在基$A_1=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$、$A_2=\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$、$A_3=\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$和$A_4=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$下的坐标；

以下讨论：若线性空间的基发生了变动，同一向量的坐标有何变化？设$(e_1,e_2,\cdots,e_n)$和$(f_1,f_2,\cdots,f_n)$分别是域F上线性空间V的两组基；易见，$f_1,f_2,\cdots,f_n$可由$e_1,e_2,\cdots,e_n$线性表出，不妨设：$\begin{cases} f_1=a_{11}e_1+a_{12}e_2+\cdots+a_{1n}e_n\\ f_2=a_{21}e_1+a_{22}e_2+\cdots+a_{2n}e_n\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ f_n=a_{n1}e_1+a_{n2}e_2+\cdots+a_{nn}e_n\end{cases}$，采取“形式写法”有

$(f_1,f_2,\cdots,f_n)=(e_1,e_2,\cdots,e_n)\begin{pmatrix}a_{11} & a_{21} & \cdots & a_{n1}\\ a_{12} & a_{22} & \cdots & a_{n2}\\ \cdots & \cdots & \cdots & \cdots\\ a_{1n} & a_{2n} & \cdots & a_{nn}\end{pmatrix}$，称上述矩阵$A^{\mathrm{T}}=\left(a_{ji}\right)_{n\times n}$为基$(e_1,e_2,\cdots,e_n)$到基$(f_1,f_2,\cdots,f_n)$的过渡矩阵；若$V$中向量$\alpha$有如下坐标表示式：$\alpha=(e_1,e_2,\cdots,e_n)\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}=(f_1,f_2,\cdots,f_n)\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}$，由过渡矩阵及基的坐标表示式的唯一性，可得，

$(e_1,e_2,\cdots,e_n)\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}=(e_1,e_2,\cdots,e_n)\left[A^{\mathrm{T}}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}\right]$；从而，$\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}=A^{\mathrm{T}}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}$；又若有：$(e_1,e_2,\cdots,e_n)=(f_1,f_2,\cdots,f_n)B^{\mathrm{T}}$；可以验证：$A,B$互为逆矩阵。

【例 3.3.4】在线性空间$M_2(R)$中，1）求基$E_1=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},E_2=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix},E_3=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$和$E_4=\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$到基$F_1=\begin{pmatrix}2 & 1\\ -1 & 1\end{pmatrix},F_2=\begin{pmatrix}0 & 3\\ 1 & 0\end{pmatrix},F_3=\begin{pmatrix}5 & 3\\ 2 & 1\end{pmatrix}$和$F_4=\begin{pmatrix}6 & 6\\ 1 & 3\end{pmatrix}$的过渡矩阵；2）求向量$M=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$在两组基下的坐标；3）求一非零向量$A$，使其在两组基下坐标相等。

以下讨论“线性空间”的同构！

给定n维线性空间V的一组基$e_1,e_2,\cdots,e_n$；$\forall\alpha\in V$，若有：

$\alpha = a_1e_1 + a_2e_2 + \cdots + a_ne_n = (e_1, e_2, \cdots, e_n)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$；于是$\alpha$与坐标向量$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$相互唯一确定，即：存在$V$与$F_n$之间的一一映射，

$\varphi: \alpha = a_1e_1 + a_2e_2 + \cdots + a_ne_n \to \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$；不难验证：$\forall \alpha, \beta \in V$，$\varphi(\alpha+\beta) = \varphi(\alpha)+\varphi(\beta)$；须注意的是：$\alpha+\beta$中的“$+$”是$V$中的加法运算，而$\varphi(\alpha)+\varphi(\beta)$中的“$+$”是$F_n$中的加法运算。以上的分析表明：在$V$中引进一组基后，可以建立起$V$中向量与域$F$上的$n$维列向量空间$F_n$之间的一一映射，<u>且保持了线性运算</u>，即：

$\varphi(\alpha+\beta) = \varphi(\alpha)+\varphi(\beta), \varphi(k\alpha) = k\varphi(\alpha), k \in F, \alpha, \beta \in V$。

【定义】设V,U是域F上的两个线性空间，若存在一一映射（双射）$\varphi: V \to U$，使得$\forall \alpha, \beta \in V$及$\forall k \in F$，$\varphi(\alpha+\beta) = \varphi(\alpha)+\varphi(\beta)$，$\varphi(k\alpha) = k\varphi(\alpha)$，则称线性空间$V$与$U$同构，记为：$V \overset{\varphi}{\cong} U$。

由此，数域F上的任一n维线性空间$V \overset{\varphi}{\cong} F_n$。

<u>线性空间的同构有如下性质</u>：

1）若$V \overset{\varphi}{\cong} U$，则$\varphi(\theta_1) = \theta_2$，其中零向量$\theta_1 \in V, \theta_2 \in U$；

2）φ将V中线性相（无）关的向量变成U中线性相（无）关的向量；

3）同构关系是一种等价关系，即满足自反性、对称性和传递性；

4）域F上两个有限线性空间V,U同构当且仅当$\dim V = \dim U$。

综上所述，两个同构的线性空间不仅元素之间有一一对应关系，且这种对应保持了线性关系；可以认为：<u>两个同构的线性空间“结构”是相同的</u>；这样<u>在一个线性空间中由线性关系所得的性质在与之同构的线性空间中也成立</u>！

§4 欧几里得空间

线性空间的向量有加法及数乘运算；若在实数域上的n维向量空间R^n中，规定一个“内积”，即可引进长度、正交

等度量概念。

【定义】设V是实数域R上的任一线性空间，V上的一个二元函数$(\cdot,\cdot)$，若满足：$\forall\alpha,\beta,\gamma\in V$，$\forall k\in R$，

1）对称性：$(\alpha,\beta)=(\beta,\alpha)$；2）线性性：$(\alpha+\gamma,\beta)=(\alpha,\beta)+(\gamma,\beta)$，$(k\alpha,\beta)=k(\alpha,\beta)$；3）正定性：$(\alpha,\alpha)\geq 0$，　等号成立当且仅当$\alpha=\theta$；则称$(\cdot,\cdot)$为$V$上的一个内积。

【例 3.4.1】1）在$M_n(R)$中，定义二元函数$(A,B)\underline{\underline{def}}\,tr(AB^{\mathrm{T}})$，可以验证：$(A,B)$是$M_n(R)$上的一个内积；2）在$C[a,b]$中，规定$(f,g)\underline{\underline{\Delta}}\int_a^b f(x)g(x)dx$，可以验证：$(f,g)$是$C[a,b]$上的一个内积。

【定义】实数域R上的线性空间V若给定一个内积，则称V是一个实内积空间，有限维的实内积空间V称为欧几里得空间；V的维数称为欧几里得空间V的维数。

称非负实数$\sqrt{(\alpha,\alpha)}$为向量α的长度（范数），记作：$|\alpha|$或者$\|\alpha\|$；易见，$\forall k\in R$，$\forall\alpha\in V$，$\|k\alpha\|=|k|\cdot\|\alpha\|$。长度为 1 的向量为单位向量；若$\alpha\neq\theta$，将$\alpha$化为$\frac{1}{\|\alpha\|}\alpha$称为单位化。

【定理】（<u>Cauchy—Buniakowski 不等式</u>）V是实内积空间，$\forall\alpha,\beta\in V$，$|(\alpha,\beta)|\leq\|\alpha\|\cdot\|\beta\|$，等号成立当且仅当$\alpha,\beta$线性相关。

易见，$\forall t\in R$，$\forall\alpha,\beta\in V$，$(\alpha+t\beta,\alpha+t\beta)\geq 0$，也即：$(\beta,\beta)t^2+2(\alpha,\beta)t+(\alpha,\alpha)\geq 0$，从而，$\Delta=\left[2(\alpha,\beta)\right]^2-4(\beta,\beta)(\alpha,\alpha)\leq 0$，即有不等式成立！等号成立当且仅当$\exists!t_0\in R$，使得$(\alpha+t_0\beta,\alpha+t_0\beta)=0$，也即：$\alpha+t_0\beta=\theta$，即：$\alpha,\beta$线性相关。

【定义】在实内积空间，两个非零向量α,β的夹角$\langle\alpha,\beta\rangle$规定为$\langle\alpha,\beta\rangle\underline{\underline{def}}\arccos\frac{(\alpha,\beta)}{\|\alpha\|\cdot\|\beta\|}$，$0\leq\langle\alpha,\beta\rangle\leq\pi$；若$(\alpha,\beta)=0$即$\langle\alpha,\beta\rangle=\frac{\pi}{2}$，称$\alpha,\beta$垂直（正交），记作：$\alpha\perp\beta$。由 **Cauchy—Buniakowski** 不等式，

【推论】在实内积空间V中，三角形不等式成立，即：$\forall\alpha,\beta\in V$，$\|\alpha+\beta\|\leq\|\alpha\|+\|\beta\|$；在实内积空间$V$中，也有勾股定理成立，即：若$\alpha\perp\beta$，$\|\alpha+\beta\|^2=\|\alpha\|^2+\|\beta\|^2$；同样亦可知余弦定理成立！

【定义】在实内积空间V中，$\forall\alpha,\beta\in V$，规定$d(\alpha,\beta)\underline{\underline{def}}\|\alpha-\beta\|$，称$d(\alpha,\beta)$是$\alpha$与$\beta$的距离；可以验证：$d(\alpha,\beta)$满足<u>距离公理</u>。

在欧几里得空间（简称欧氏空间），由两两正交的非零向量组成的向量组称为正交向量组；由两两正交的单位向量组成的向量组称为正交单位向量组；可以验证：<u>欧氏空间中的正交向量组必线性无关</u>！

在n维欧氏空间V中，n个向量组成的正交向量组必是V的一个基，称为正交基；若其中每个向量都是单位向量，则称为标准正交基。

【定理】设$\alpha_1,\alpha_2,\cdots,\alpha_s$是欧几里得空间$V$的一个线性无关的向量组，令$\begin{cases}\beta_1=\alpha_1\\ \beta_2=\alpha_2-\dfrac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \beta_s=\alpha_s-\displaystyle\sum_{j=1}^{s-1}\dfrac{(\alpha_s,\beta_j)}{(\beta_j,\beta_j)}\beta_j,\end{cases}$，则$\beta_1,\beta_2,\cdots,\beta_s$是正交向量组，且向量组$\{\alpha_1,\alpha_2,\cdots,\alpha_s\}$与$\{\beta_1,\beta_2,\cdots,\beta_s\}$等价；上述过程称为施密特（**Schmidt**）正交化；将上述正交向量组单位化后，即可得一组标准正交向量。

对线性无关的向量组中向量的个数s作数学归纳法；$s=1$，即向量组为$\alpha_1,\alpha_1\neq\theta,\beta_1=\alpha_1$，则$\beta_1$是正交向量组；显然$\{\alpha_1\}\underline{\underline{\sim}}\{\beta_1\}$；假设$s=k$，定理成立，即$\beta_1,\beta_2,\cdots,\beta_k$是正交向量组，且与$\alpha_1,\alpha_2,\cdots,\alpha_k$等价；则当$s=k+1$时，由于
$\beta_{k+1}=\alpha_{k+1}-\displaystyle\sum_{j=1}^{k}\frac{(\alpha_{k+1},\beta_j)}{(\beta_j,\beta_j)}\beta_j$；从而，$\forall i\in\{1,2,\cdots,k\}$，有

$(\beta_{k+1},\beta_i)=(\alpha_{k+1},\beta_i)-\displaystyle\sum_{j=1}^{k}\frac{(\alpha_{k+1},\beta_j)}{(\beta_j,\beta_j)}(\beta_j,\beta_i)=(\alpha_{k+1},\beta_i)-\frac{(\alpha_{k+1},\beta_i)}{(\beta_i,\beta_i)}(\beta_i,\beta_i)=0$，即：与正交，…………。需要注意的是施密特正交化的思路！以下简单讨论向量正交化的几何意义：以$\alpha_1=\begin{pmatrix}2\\0\end{pmatrix},\alpha_2=\begin{pmatrix}1\\1\end{pmatrix}\in R^2$为例，$\alpha_1,\alpha_2$不共线，即：线性无关；令$\beta_1=\alpha_1=\begin{pmatrix}2\\0\end{pmatrix}$，$\beta_2=\alpha_2-\dfrac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}\beta_1=\begin{pmatrix}0\\1\end{pmatrix}$，$\beta_1\perp\beta_2$；从而，通过施密特正交化方法将$R^2$中两个不共线的

向量α_1,α_2化为相互垂直的两个向量β_1,β_2。

由上述，将线性空间的一组基经施密特正交化后即可得一组正交基，再经单位化后，即得一组标准正交基。<u>利用标准正交基计算向量内积比较方便</u>！若$\eta_1,\eta_2,\cdots,\eta_n$是$n$维线性空间$V$的一组标准正交基，且$\alpha=\sum_{i=1}^{n}x_i\eta_i$，$\beta=\sum_{i=1}^{n}y_i\eta_i\in V$，则

$(\alpha,\beta)=\left(\sum_{i=1}^{n}x_i\eta_i,\sum_{j=1}^{n}y_j\eta_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}x_iy_j\left(\eta_i,\eta_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}x_iy_j\delta_{ij}=\sum_{i=1}^{n}x_iy_i$；由此，可导出：$\forall\alpha,\beta\in V$，不妨设$\alpha=(a_1,a_2,\cdots,a_n)$，$\beta=(b_1,b_2,\cdots,b_n)$，$(\alpha,\beta)=\sum_{i=1}^{n}a_ib_i=\alpha\beta^{\mathrm{T}}$（标准内积）。

利用标准正交基，向量的坐标分量可用内积表达；设α在标准正交基$\eta_1,\eta_2,\cdots,\eta_n$下的坐标为$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，即：$\alpha=\sum_{i=1}^{n}x_i\eta_i$，两边与$\eta_j$作内积，即得：$\left(\alpha,\eta_j\right)=\left(\sum_{i=1}^{n}x_i\eta_i,\eta_j\right)=\sum_{i=1}^{n}x_i\left(\eta_i,\eta_j\right)=\sum_{i=1}^{n}x_i\delta_{ij}=x_j$；从而，$\boxed{\alpha=\sum_{i=1}^{n}(\alpha,\eta_i)\eta_i}$，称之为向量$\alpha$的傅立叶（**Fourier**）展开，每个系数$(\alpha,\eta_i)$称为$\alpha$的傅立叶系数。

可以验证：在$[a,b]$上（黎曼）可积函数的全体$R[a,b]$按照通常定义的方式定义加法及（实）数乘运算构成一（实）线性空间；且若定义$(f,g)=\int_a^b f(x)g(x)dx$，则$(f,g)$是其上的内积。设$\{\varphi_0,\ \varphi_1,\ \varphi_2;\cdot\}$是内积空间$R[\ a\ \ ]$中的一组标准正交基，假设$f\in R[a,b]$且能展开成如下形式的级数：$f(x)=\sum_{k=0}^{\infty}c_k\varphi_k(x)$，两边同乘以$\varphi_l(x)$，$f(x)\varphi_l(x)=\sum_{k=0}^{\infty}c_k\varphi_k(x)\varphi_l(x)$；假设该级数可逐项积分，则有：$(f,\varphi_l)=\sum_{k=0}^{\infty}c_k(\varphi_k,\varphi_l)=\sum_{k=0}^{\infty}c_k\delta_{kl}=c_l,l=0,1,2,\cdots$；该式称为关于正交函数系的欧拉—傅立叶公式；从而，级数$\sum_{k=0}^{\infty}(f,\varphi_k)\varphi_k$称为$f$关于正交函数系的傅立叶级数。

考察三角函数系$\left\{\frac{1}{\sqrt{2}},\cos x,\sin x,\cos 2x,\sin 2x,\cdots,\cos kx,\sin kx,\cdots\right\}$，并定义如下内积：$(f,g)=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)g(x)dx$；从而有函数$f$关于规

范（标准）正交函数系的傅立叶级数$A_0\frac{1}{\sqrt{2}}+\sum_{k=1}^{\infty}\left(a_k\cos kx+b_k\sin kx\right)$，其中：$A_0=\left(f,\frac{1}{\sqrt{2}}\right)=\frac{1}{\sqrt{2}\pi}\int_{-\pi}^{\pi}f\left(x\right)dx$，$a_k=\left(f,\cos kx\right)=\frac{1}{\pi}\int_{-\pi}^{\pi}f\left(x\right)\cos kxdx$，$b_k=\left(f,\sin kx\right)=\frac{1}{\pi}\int_{-\pi}^{\pi}f\left(x\right)\sin kxdx$。

【例 3.4.2】在$R_2[x]$上给定一个内积：$\left(f\left(x\right),g\left(x\right)\right)\underline{\underline{def}}\int_{-1}^{1}f\left(x\right)g\left(x\right)dx$，试求$R_2[x]$的一个标准正交基；

【例 3.4.3】设$\alpha_1=\begin{pmatrix}1\\1\end{pmatrix},\alpha_2=\begin{pmatrix}1\\-1\end{pmatrix}$为二维列向量空间$R^2$的一组基，对于内积$(\alpha,\beta)=\alpha^{\mathrm{T}}A\beta$，其中$\alpha,\beta\in R^2$，$A=\begin{pmatrix}2&1\\1&1\end{pmatrix}$，试由施密特正交化方法，求出与$\alpha_1,\alpha_2$等价的一组标准正交基；

【定义】实数域上的n阶方阵A，若满足$AA^{\mathrm{T}}=A^{\mathrm{T}}A=I_n$，则称$A$是正交矩阵。

可以验证：①I是正交阵；②若A是正交阵，则$|A|=1$或-1，且$A^{-1}\left(A^{\mathrm{T}}\right)$也是正交阵；③若$A,B$是正交阵，则$AB$也是正交阵，且可推广至有限个的情形；④若将$n$阶方阵按行、列分块有：$A=(A_1,A_2,\cdots,A_n)=\begin{pmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_n\end{pmatrix}$，则有：（ⅰ）$A$为正交阵$\Leftrightarrow$ $\alpha_i\alpha_j^{\mathrm{T}}=\delta_{ij}$；（ⅱ）$A$为正交阵$\Leftrightarrow A_i^{\mathrm{T}}A_j=\delta_{ij}$；从而，正交矩阵$A$的行（列）向量组都是标准正交向量，且是$R^n\left(R_n\right)$的一组标准正交基。

【定理】1）欧几里得空间中的一个标准正交基到另一个标准正交基的过渡矩阵是正交矩阵；2）设$\alpha_1,\alpha_2,\cdots,\alpha_n$是$n$维欧氏空间$V$的一个标准正交基，向量组$\{\beta_1,\beta_2,\cdots,\beta_n\}$满足：$(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)P$，$P$是正交矩阵，则$\beta_1,\beta_2,\cdots,\beta_n$也是$V$的一个标准正交基。

对于实数域R上的一个线性空间V，可以指定不同的内积，从而使V成为不同的内积空间；以下就来讨论实内积空间之间的关系。

【定义】设V,U均是实数域R上的内积空间，φ是V到U的线

性映射；若$\forall x,y\in V$，$(\varphi(x),\varphi(y))=(x,y)$，则称$\varphi$是$V$到$U$的保持内积的线性映射；若$\varphi$是同构映射，则称$\varphi$是内积空间$V$到$U$上的同构；

【注 1】①若φ保持内积，则φ必保持向量的范数（保范），即：$\|\varphi(x)\|=\|x\|,x\in V$；故保持内积的线性映射必是单射；②两个内积空间的同构即保持内积的同构，同构关系也是一个等价关系；

设V,U均是n维实内积空间，若φ是V到U的线性映射，则以下命题等价：（ⅰ）φ保持内积；（ⅱ）φ是同构；（ⅲ）φ将V中任一组标准正交基变为U中的一组标准正交基。

我们曾经在欧氏空间R^2,R^3中引入向量α的模$\|\alpha\|$（长度或范数），事实上，向量α的模$\|\alpha\|$是R^2,R^3到R的映射，且满足如下性质：①正定性：$\|\alpha\|\ge 0,\|\alpha\|=0\Leftrightarrow\alpha=\theta$（零向量）；②齐次性：$\|k\alpha\|=|k|\cdot\|\alpha\|,k\in R$；③三角不等式：$\|\alpha+\beta\|\le\|\alpha\|+\|\beta\|$。以此向量模的基本性质为出发点，可以用公理形式在一般线性空间引入向量范数的概念，从而建立一类具有拓扑性质的线性空间——线性赋范空间。

【范数】设V是数域F上的线性空间，如果映射$\|\cdot\|:V\to R$满足下列条件（称为范数公理）：①正定性：$\|\alpha\|\ge 0,\|\alpha\|=0\Leftrightarrow\alpha=\theta$；②齐次性：$\|k\alpha\|=|k|\cdot\|\alpha\|,k\in R$；③三角不等式：$\|\alpha+\beta\|\le\|\alpha\|+\|\beta\|$，$\alpha,\beta\in V$；则称映射$\|\cdot\|$是$V$上的一个范数，$\|\cdot\|$在$\alpha$处的像$\|\alpha\|$称为$\alpha$的范数。定义了范数的线性空间称为线性赋范空间，记为$(V,\|\cdot\|)$；若无需特别说明，即简记为V。

设$(V,\|\cdot\|)$是线性赋范空间，定义映射$d:V\times V\to R$为：$d(\alpha,\beta)=\|\alpha-\beta\|,\alpha,\beta\in V$；则映射$d$具有如下性质（称为距离公理或度量公理）：①正定性：$d(\alpha,\beta)\ge 0,d(\alpha,\beta)=0\Leftrightarrow\alpha=\beta$；②对称性：$d(\alpha,\beta)=d(\beta,\alpha),\alpha,\beta\in V$；③三角不等式：$d(\alpha,\beta)\le d(\alpha,\gamma)+d(\gamma,\beta),\alpha,\beta,\gamma\in V$。上述的映射$d(\cdot,\cdot)$称为$V$中由范数诱导的距离（或度量），非负实数$d(\alpha,\beta)=\|\alpha-\beta\|$称为两点$\alpha,\beta$之间的距离。从而，线性赋范空间中由范数诱导的距离是欧氏空间中两点间距离概念的推广，且距离公理也反映了距离

的基本特性。

设U是线性赋范空间V的非空子集，$\alpha\in V$，记：$d(\alpha,U)=\inf\limits_{\beta\in U}\|\alpha-\beta\|$，称之为点$\alpha$到集合$U$的距离；如果$\beta_0\in U$且$\|\alpha-\beta_0\|=\inf\limits_{\beta\in U}\|\alpha-\beta\|$，则称$\beta_0$是$U$中对于$\alpha$的最佳逼近元。

线性赋范空间的例：1）若在n维向量空间R^n中定义：$\|\alpha\|=\sqrt{\sum\limits_{i=1}^{n}x_i^2},\alpha=(x_1,x_2,\cdots,x_n)\in R^n$，则易知$\|\cdot\|$满足范数公理，称之为欧氏范数，而相应的$d(\alpha,\beta)=\|\alpha-\beta\|=\sqrt{\sum\limits_{i=1}^{n}(x_i-y_i)^2}$称为欧氏距离；若在$R^n$上定义：$\|\alpha\|_1=\sum\limits_{i=1}^{n}|x_i|,\|\alpha\|_\infty=\max\{|x_1|,|x_2|,\cdots,|x_n|\}$，易知$\|\alpha\|_1,\|\alpha\|_\infty$都是$R^n$上的范数；2）若在连续函数空间$C[a,b]$定义：$\|f\|=\max\limits_{a\le t\le b}|f(t)|$，$f=f(t)\in C[a,b]$，易知$C[a,b]$是线性赋范空间，且$d(f,g)=\max\limits_{a\le t\le b}|f(t)-g(t)|,f,g\in C[a,b]$；若定义$\|f\|_0=\int_a^b|f(t)|dt,f=f(t)\in C[a,b]$，则可验证$(C[a,b],\|f\|_0)$也是线性赋范空间。

应用范数诱导的距离，可以在线性赋范空间刻划逼近的概念，即空间点列的收敛性，进而给出抽象映射的连续性的定义；应用范数还可以定义邻域和开集，构造出线性赋范空间的拓扑结构。

设V是线性赋范空间，$\{\alpha_n,n\ge1\}$是V中的点列，$\alpha_0\in V$，如果$\lim\limits_{n\to\infty}\|\alpha_n-\alpha_0\|=0$，则称点列$\{\alpha_n,n\ge1\}$依范数收敛于$\alpha_0$，简称收敛于$\alpha_0$；记为：$\lim\limits_{n\to\infty}\alpha_n=\alpha_0$或$\alpha_n\to\alpha_0(n\to\infty)$。可以验证：线性赋范空间中的收敛点列也有数学分析（微积分）中收敛实数列的一些类似性质，如：极限的唯一性、收敛必有界等，这里不再累述！

【例】在$C[a,b]$中，函数列$\{f_n(t),n\ge1\}$收敛于$f(t)$，则$\lim\limits_{n\to\infty}f_n=f\Leftrightarrow\lim\limits_{n\to\infty}\|f_n-f\|=\lim\limits_{n\to\infty}\left[\max\limits_{a\le t\le b}|f_n(t)-f(t)|\right]=0\Leftrightarrow\forall\varepsilon>0,\exists N\in\mathrm{N}$，使得当$n>N$时，对于一切$t\in[a,b]$都有$|f_n(t)-f(t)|<\varepsilon\Leftrightarrow$ 在$[a,b]$上函数列$f_n(t)\Rightarrow f(t)(n\to\infty)$，即$C[a,b]$中的收敛序列是一致收敛函数列。事实上，在一些具体空间中的各不相同的收敛概念，可以统一于线性赋范空间中依范数收敛的概念，这就为深入研究和

统一处理各类的极限提供了方便！

【定义】设$\left(V,\|\cdot\|_V\right),\left(U,\|\cdot\|_U\right)$是两个线性赋范空间，映射$T:V\to U$，称$T$在点$\alpha_0\in V$连续，若对于$V$中任意收敛于$\alpha_0$的点列$\{\alpha_n,n\ge 1\}$：$\{\alpha_n,n\ge 1\}\subset V$且$\lim\limits_{n\to\infty}\|\alpha_n-\alpha_0\|_V=0$，$U$中相应的点列$\{T\alpha_n,n\ge 1\}$收敛于$T\alpha_0$：$\lim\limits_{n\to\infty}\|T\alpha_n-T\alpha_0\|_U=0$；若$T$在$V$中每一点都连续，则称$T$在$V$上连续。

由此定义，线性赋范空间$(V,\|\cdot\|)$中的范数$\|\cdot\|$是$V\to R$的连续映射！

在实数理论中，**Cauchy** 收敛准则（**Cauchy** 原理）对于数列及函数列收敛性的研究起着重要的作用！我们也已在线性赋范空间中定义了收敛点列，那么是否也有类似的结论呢？

设V是线性赋范空间，$\{\alpha_n,n\ge 1\}\subset V$，若$\forall\varepsilon>0,\exists N\in\mathrm{N}$，使得$\forall n,m>N$，总有$\|\alpha_n-\alpha_m\|<\varepsilon$，则称$\{\alpha_n,n\ge 1\}$是$V$中的 **Cauchy** 列或基本列。可以验证：线性赋范空间中的 **Cauchy** 列具有与R中的 **Cauchy** 列完全相同的性质！如果线性赋范空间V中的每个 **Cauchy** 列都收敛于V中的点，则称V是完备的。完备的线性赋范空间称为 **Banach** 空间。实数集的完备性是建立严密极限理论的基础，线性赋范空间的完备性是正是R中的这个基本属性的抽象与扩充，它同样在泛函分析中占有重要的位置（不完备的线性赋范空间是大量存在的）。前面已提及内积空间，完备的内积空间则称为 **Hilbert** 空间，以内积为基础，可以研究该类空间的几何性质，有兴趣的读者可参考有关泛函分析的教材！

第四章 矩阵的相似、特征值（向量）及二次型

§1 矩阵的相似

已经知道，线性变换在不同基下的矩阵相似；而且，若有可逆矩阵U，使得$U^{-1}AU=\Lambda$（对角阵），即有：

$A^m=\left(U\Lambda U^{-1}\right)^m=U\Lambda U^{-1}U\Lambda U^{-1}\cdots U\Lambda U^{-1}=U\Lambda^m U^{-1}$。

【定义】设A,B为数域F上的n阶方阵，若存在数域F上的一个n阶可逆矩阵U，使得$U^{-1}AU=B$，则称A,B是相似的，记作：$A\sim B$；称非异阵U为A（到B的）相似变换阵；若B为对角阵，

称之为A的相似标准形，此时亦称A可对角化。

可以验证：数域F上的n阶矩阵间的相似关系满足：①反身性（自反性）：$A=I_n^{-1}AI_n$，即：$A\sim A$；②对称性：若$A\sim B$，则$B\sim A$；③传递性：若$A\sim B$且$B\sim C$，则$A\sim C$。

矩阵的相似关于矩阵的运算还有如下一些简单关系；若$B_1=U^{-1}A_1U$，$B_2=U^{-1}A_2U$，则有：$B_1+B_2=U^{-1}\left(A_1+A_2\right)U$，$B_1B_2=U^{-1}\left(A_1A_2\right)U$，$B_i^m=U^{-1}A_i^mU,i=1,2$。

相似的矩阵有如下一些基本性质：①相似的矩阵有相同的行列式；②相似的矩阵同可逆或同不可逆；若可逆，其逆矩阵也相似；③相似的矩阵有相同秩；④相似的矩阵有相同的迹；

【例 4.1.1】1）若A可逆，则$AB\sim BA$；**2）**若$A_1\sim B_1$，$A_2\sim B_2$，则$\begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix}\sim\begin{pmatrix} B_1 & O \\ O & B_2 \end{pmatrix}$；

【例 4.1.2】数域F上的n阶方阵A,B满足$AB-BA=A$，则A不可逆；

【例 4.1.3】1）若$A\sim B$，则$kA\sim kB$且$A^{\mathrm{T}}\sim B^{\mathrm{T}}$；**2）**设$f\left(x\right)=a_0+a_1x+\cdots+a_mx^m$是数域$F$上的一元多项式，$A$是数域$F$上的$n$阶方阵，定义矩阵多项式：$f\left(A\right)=a_0I_n+a_1A+\cdots+a_mA^m$；若$A\sim B$，则$f\left(A\right)\sim f\left(B\right)$。

§2 矩阵的特征值与特征向量

假设数域F上的n阶方阵A相似于对角阵$\Lambda=diag\left\{\lambda_1,\lambda_2,\cdots,\lambda_n\right\}$，即存在一个$n$阶可逆阵$U=\left(\alpha_1,\alpha_2,\cdots,\alpha_n\right)$，使得$U^{-1}AU=\Lambda$，即：$AU=U\Lambda$；从而，$A\left(\alpha_1,\alpha_2,\cdots,\alpha_n\right)=\left(\alpha_1,\alpha_2,\cdots,\alpha_n\right)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$，也即：存在域$F$上的$n$个线性无关的列向量$\alpha_1,\alpha_2,\cdots,\alpha_n$，使得$A\alpha_i=\lambda_i\alpha_i$，$i=1,2,\cdots,n$；因此，一个方阵$A$是否可对角化，关键在于能否找到一组线性无关的向量$\alpha_1,\alpha_2,\cdots,\alpha_n$满足：$A\alpha_i=\lambda_i\alpha_i,i=1,2,\cdots,n$。

【定义】设A是数域F上的n阶方阵，若n维列向量空间F_n有

非零向量α，使得$A\alpha=\lambda\alpha,\lambda\in F$，则称$\lambda$是$A$的一个特征值，$\alpha$是$A$的属于（对应于）$\lambda$的一个特征向量。

易知：若$A\alpha=\lambda\alpha$，$\forall k\in F\left(k\neq 0\right)$，$A\left(k\alpha\right)=k\left(A\alpha\right)=\lambda\left(k\alpha\right)$，$k\alpha$也是$A$属于$\lambda$的特征向量；从而，属于一个特征值的特征向量不是唯一的，α只是其中的“代表”。可以验证：设λ是A的一个特征值，α是A属于λ的一个特征向量$\Leftrightarrow A\alpha=\lambda\alpha,\alpha\neq\theta$ $\Leftrightarrow\left(\lambda I-A\right)\alpha=\theta,\alpha\neq\theta\Leftrightarrow\alpha$是齐次线性方程组$\left(\lambda I-A\right)X=\theta$的一个非零解$\Leftrightarrow\left|\lambda I-A\right|=0,\alpha$是齐次线性方程组的一个非零解。

以后，称$\left|\lambda I-A\right|$（或$\left|A-\lambda I\right|$）为矩阵A的特征多项式；称特征方程$\left|\lambda I-A\right|=0$（或$\left|A-\lambda I\right|=0$）的根为$A$的特征值（根）；从而，可按如下步骤求方阵$A$的特征值和特征向量：①计算$A$的特征多项式$\left|\lambda I-A\right|$；②判断$\left|\lambda I-A\right|=0$在数域$F$中有无根；若它在数域中无根，则$A$没有特征值，从而无对应的特征向量；若在域$F$中有根，则它在域$F$中的全部根即$A$的全部特征值；③对$A$的每个特征值$\lambda_i$，求齐次方程组$\left(\lambda_i I-A\right)X=\theta$的一个基础解系$\eta_1,\eta_2,\cdots,\eta_{n-r}$，这里$r=r\left(\lambda_i I-A\right)$；于是，$A$的属于特征值$\lambda_i$的全部特征向量组成的集合为

$$\left\{k_1\eta_1+k_2\eta_2+\cdots+k_{n-r}\eta_{n-r}\left|k_i\in F,\text{且不全为零}\right.\right\}。$$

【例 4.2.1】设$A=\begin{pmatrix}1&-1\\1&1\end{pmatrix}$，若把$A$视作实数域$R$上的矩阵，求$A$的特征值；若把$A$视作复数域$C$上的矩阵，求$A$的特征值及特征向量；

【例 4.2.2】设n阶方阵A有特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$，其对应的特征向量分别为$\alpha_1,\alpha_2,\cdots,\alpha_n$，则：①$kA\left(k\neq 0\right)$；②$A^2,A^k,k\geq 1$；③$A^{\mathrm{T}}$；④$A$可逆，$A^{-1},A^*$的特征值、特征向量各是什么？若有$f\left(x\right)=a_0+a_1x+\cdots+a_mx^m$，试求$f\left(A\right)$的特征值及特征向量。

由高次方程的韦达定理，可以验证：若A是数域F上的n阶方阵，则A的特征多项式$f\left(\lambda\right)=\left|\lambda I-A\right|$是$\lambda$的$n$次多项式，且$f\left(\lambda\right)=\lambda^n-trA\cdot\lambda^{n-1}+\cdots+\left(-1\right)^n\cdot\det A$；从而有：$\sum\limits_{i=1}^{n}\lambda_i=trA$，$\prod\limits_{i=1}^{n}\lambda_i=\det A$。

同样可知，⑤相似的矩阵有相同的特征多项式；⑥相似的矩阵有相同的特征值（包括重数相同）。

【例 4.2.3】设n阶方阵A有n个特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$，且$B$与$A$相似，则$\forall a,b\in R,a\neq 0$，试求$|aA+bI_n|$；

【例 4.2.4】设四阶方阵满足$AA^{\mathrm{T}}=2I_n$，且$|A|<0$，试求A^*的一个特征值；

§3 矩阵的对角化、实对称矩阵

【定理】数域F上阶方阵A可对角化的充要条件是：A有n个线性无关的特征向量$\alpha_1,\alpha_2,\cdots,\alpha_n$；此时，若令$U=(\alpha_1,\alpha_2,\cdots,\alpha_n)$，则有：$U^{-1}AU=diag\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$，其中$\lambda_i$是$\alpha_i$所属的特征值，$i=1,2,\cdots,n$；称上述的对角阵为$A$的相似标准形。

【注1】　除主对角线上元素的排列次序外，A的相似标准形唯一确定！

欲判断域F上的n阶方阵A是否有n个线性无关的特征向量，首先得求出n阶矩阵A的全部特征值；设A的所有不同的特征值是$\lambda_1,\lambda_2,\cdots,\lambda_m$，其中$\lambda_i$的重数是$n_i$，即：$n_1+n_2+\cdots+n_m=n$；再求$(\lambda_i I-A)X=\theta$的解，$i=1,2,\cdots,m$；若以上$m$个齐次方程组的基础解系中分别含有$n_1,n_2,\cdots,n_m$个线性无关的解向量，则$A$必可对角化！这是由于：

【定理】设λ_1,λ_2分别是域F上n阶方阵A的不同的特征值，$\alpha_1,\alpha_2,\cdots,\alpha_s$和$\beta_1,\beta_2,\cdots,\beta_r$分别是$A$的属于$\lambda_1,\lambda_2$的线性无关的特征向量，则$\alpha_1,\alpha_2,\cdots,\alpha_s,\beta_1,\beta_2,\cdots,\beta_r$线性无关。

由上述定理即知：1）n阶方阵A的属于不同特征值的特征向量线性无关；从而，2）域F上的n阶方阵A若有n个线性无关的特征向量，则A必可对角化！

【例 4.3.1】1）设数列$\{a_n\},\{b_n\}$满足：$\begin{cases}a_n=a_{n-1}+2b_{n-1}\\ b_n=-a_{n-1}+4b_{n-1}\end{cases}$，且$a_0=b_0=1$，试求$\{a_n\},\{b_n\}$；2）设可微函数$x=x(t),y=y(t)$满足：$\begin{cases}\dfrac{dx}{dt}=x+2y\\ \dfrac{dy}{dt}=-x+4y\end{cases}$，且$x(0)=y(0)=1$，试求$x(t),y(t)$；

【例 4.3.2】若n阶方阵A满足$A^2-A=2I$，则A可对角化；

【例 4.3.3】已知三阶矩阵A与三维向量x，使得向量组x,Ax,A^2x线性无关，且满足$A^3x=3Ax-2A^2x$，1）记$P=(x,Ax,A^2x)$，

求三阶方阵B，使得$A=PBP^{-1}$；2）试求$\det(A+I)$；

【定义】实数域上的矩阵又称实矩阵，实数域上的对称矩阵简称为实对称矩阵。

【定理】实对称矩阵的特征多项式在复数域中每一个根都是实数，即：实对称矩阵只有实特征值；而且，实对称矩阵属于不同特征值的特征向量是正交的。（课堂证明）

【定理】实对称矩阵必正交相似于对角阵；

以下对实对称阵的阶作数学归纳法！

当$n=1$时，$I_1^{-1}(a_{11})I_1=1(a_{11})1=(a_{11})$；假设任一$n-1$阶实对称阵都正交相似于对角阵；则对于$n$阶实对称阵$A$，取$A$的一个特征值$\lambda_1$及其对应的特征向量$\eta_1$，且$\|\eta_1\|=1$；将$\eta_1$扩充为$R_n$的一组基，经施密特正交化和单位化处理，即可得$R_n$的一个标准正交基$\eta_1,\eta_2,\cdots,\eta_n$，令$T_1=(\eta_1,\eta_2,\cdots,\eta_n)$，则$T_1$为$n$阶正交矩阵，且$T_1^{-1}AT_1=T_1^{-1}A(\eta_1,\eta_2,\cdots,\eta_n)=T_1^{-1}(A\eta_1,A\eta_2,\cdots,A\eta_n)=T_1^{-1}(\lambda_1\eta_1,A\eta_2,\cdots,A\eta_n)$，由于$T_1^{-1}T_1=T_1^{-1}(\eta_1,\eta_2,\cdots,\eta_n)=(e_1,e_2,\cdots,e_n)=I_n$，从而有$T_1^{-1}AT_1=\begin{pmatrix}\lambda_1 & \alpha\\ O & B_1\end{pmatrix}$，由$T_1^{-1}=T_1^{\mathrm{T}}$，$A=A^{\mathrm{T}}$，可知$\alpha=O_{1\times(n-1)}$，$B_1$为$n-1$阶实对称阵；由归纳假设，存在$n-1$阶正交阵$T_2$，使得$T_2^{-1}B_1T_2=diag\{\lambda_2,\lambda_3,\cdots,\lambda_n\}$；记$T=T_1\cdot\begin{pmatrix}1 & O\\ O & T_2\end{pmatrix}$—正交矩阵，可知

$T^{-1}AT=\begin{pmatrix}1 & O\\ O & T_2\end{pmatrix}^{-1}\cdot T_1^{-1}AT_1\cdot\begin{pmatrix}1 & O\\ O & T_2\end{pmatrix}=diag\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$，定理得证。

【例 4.3.4】设n阶实对称阵A满足$A^3+A^2+A=3I$，试求A；

【例 4.3.5】设三阶实对称阵A的特征值为$\lambda_1=-1,\lambda_2=\lambda_3=1$，对应于$\lambda_1$的特征向量为$\xi_1=(0,1,1)^{\mathrm{T}}$，试求$A$；

【例 4.3.5 的注】不妨设对应于特征值$\lambda=1$的特征向量为$\eta=(x_1,x_2,x_3)^{\mathrm{T}}$，由于$\xi_1$与$\eta$正交，故有：$x_2+x_3=0$；该方程组基础解系中有两个线性无关的解向量，不妨任取其解空间的一组基$\{\xi_2,\xi_3\}$，取$U=(\xi_1,\xi_2,\xi_3)$，则有：$U^{-1}AU=diag\{-1,\lambda,\lambda\}=\begin{pmatrix}-1 & O\\ O & \lambda I_2\end{pmatrix}$，

$AU=U\begin{pmatrix}-1 & O\\ O & \lambda I_2\end{pmatrix}$，若$\begin{vmatrix}k_{11} & k_{12}\\ k_{21} & k_{22}\end{vmatrix}\neq 0$，$(\eta_2,\eta_3)=(\xi_2,\xi_3)\begin{pmatrix}k_{11} & k_{12}\\ k_{21} & k_{22}\end{pmatrix}=$

$(k_{11}\xi_2+k_{21}\xi_3,k_{12}\xi_2+k_{22}\xi_3)$，从而，$\{\eta_2,\eta_3\}$亦可作为解空间的一组基；

设$T=\begin{pmatrix}k_{11}&k_{12}\\k_{21}&k_{22}\end{pmatrix}$，易知：若记$\zeta=(\xi_2,\xi_3)$，则$AU=U\begin{pmatrix}-1&O\\O&\lambda I_2\end{pmatrix}$

等价于$A(\xi_1,\zeta)\begin{pmatrix}1&O\\O&T\end{pmatrix}=(\xi_1,\zeta)\begin{pmatrix}1&O\\O&T\end{pmatrix}\begin{pmatrix}1&O\\O&T\end{pmatrix}^{-1}\begin{pmatrix}-1&O\\O&\lambda I_2\end{pmatrix}\begin{pmatrix}1&O\\O&T\end{pmatrix}$，此即：

$A(\xi_1,\eta_2,\eta_3)=(\xi_1,\eta_2,\eta_3)\begin{pmatrix}-1&O\\O&\lambda I_2\end{pmatrix}$；从而，解空间基的选择并不影响求解矩阵$A$！

【例 4.3.6】设α,β是两个n维非零列向量，且α,β正交，试证：矩阵$A=\alpha\beta^T$的特征值都是零，且A不可对角化；

§4　二次型、矩阵的合同

二次型的“型”（**Form**）意即齐次多项式，故二次型即二次齐次多项式。

【定义】设f是数域F上的n元二次齐次多项式：

① $$\begin{aligned}f(x_1,x_2,\cdots,x_n)=\ &a_{11}x_1^2+2a_{12}x_1x_2+\cdots+2a_{1n}x_1x_n\\&+a_{22}x_2^2+\cdots+2a_{2n}x_2x_n\\&+\cdots+\cdots\\&+a_{nn}x_n^2\end{aligned}$$，则称f是数域F上的n元二次型，简称二次型；上式也常写成：②

$$\begin{aligned}f(x_1,x_2,\cdots,x_n)=\ &a_{11}x_1^2+a_{12}x_1x_2+\cdots+a_{1n}x_1x_n\\&+a_{21}x_2x_1+a_{22}x_2^2+\cdots+a_{2n}x_2x_n\\&+\cdots+\cdots+\cdots+\cdots\\&+a_{n1}x_nx_1+a_{n2}x_nx_2+\cdots+a_{nn}x_n^2\end{aligned}=\sum_{j=1}^n\sum_{i=1}^n a_{ij}x_ix_j=\sum_{i,j=1}^n a_{ij}x_ix_j$$，

其中：$a_{ij}=a_{ji},i,j=1,2,\cdots,n$；上述的②也常写为矩阵形式：

③ $$f(x_1,x_2,\cdots,x_n)=(x_1,x_2,\cdots,x_n)\begin{pmatrix}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\\\cdots\cdots\cdots\cdots\cdots\cdots\\a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n\end{pmatrix}$$

$$=(x_1,x_2,\cdots,x_n)\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$，在矩阵$A=(a_{ij})_{n\times n}$中，$a_{ij}=a_{ji},i,j=1,2,\cdots,n$；若令$X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，③可简写为：

$f(X)=X^{\mathrm{T}}AX$。

可以验证：给定域F上的一个n元二次型，即得域F上的一个

n阶对称矩阵A（又称该二次型的相伴矩阵）；反之，给定域F上一对称矩阵A，亦可得到一个二次型，称之为对称矩阵A的相伴二次型。由此，数域F上的n元二次型与域F上的n阶对称矩阵之间存在着一一对应关系；从而，使得利用矩阵的理论和方法研究二次型成为可能。

【例 4.4.1】写出下列二次型的矩阵表示式：1）$f(x_1,x_2,\cdots,x_n)$ $x_1x_2+x_2x_3+\cdots+x_{n-1}x_n$；2）$f(x_1,x_2,x_3)=(a_1x_1+a_2x_2+a_3x_3)^2$；

【例 4.4.2】试写出矩阵$A=\begin{pmatrix}1&0&2\\0&1&-1\\2&-1&0\end{pmatrix}$的相伴二次型；

如下的多项式：$f(x,y)=x^2+2xy+4y^2-1$，$f(x,y)=x^3+2xy-3x+4y$，则不是二次型（二次齐次多项式）。

二次型理论的一个基本问题是：化二次型为标准形，即通过变量的线性替换化二次型为只含平方项的二次型；二次型的研究源于解析几何中化二次曲线和二次曲面的方程为标准形的问题。

设有二次型$f(X)=X^{\mathrm{T}}AX$，若令$X=CY$，C为n阶可逆阵，$Y=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$，称之为变量（元）$x_1,x_2,\cdots,x_n$到变量（元）$y_1,y_2,\cdots,y_n$的一个非退化（满秩）线性变换（替换）；从而，$n$元二次型$X^{\mathrm{T}}AX$经非退化线性替换$X=CY$，有$X^{\mathrm{T}}AX=(CY)^{\mathrm{T}}A(CY)$ $=Y^{\mathrm{T}}(C^{\mathrm{T}}AC)Y$；即新二次型（与原二次型等价）的相伴矩阵是$C^{\mathrm{T}}AC$。

【定义】设A,B是域F上的n阶对称矩阵，若存在n阶非异阵，使得$B=C^{\mathrm{T}}AC$，则称B与A是合同的，或称B与A具有合同关系；记作：$A\simeq B$；

可以验证：合同关系是一种等价关系，即：①自反性：$A=I^{\mathrm{T}}AI$，$A\simeq A$；②对称性：若$A\simeq B$，则$B\simeq A$；③传递性：若$A\simeq B$且$B\simeq C$，则$A\simeq C$。

由此，化二次型为只含平方项的二次型等价于对于对称矩阵A，寻找非异阵C，使$C^{\mathrm{T}}AC$为对角阵（A的合同标准形）。以下证明：数域F上任一n元二次型必等价于一只含平方项的二次型；也即：数域F上任一对称矩阵必合同于一个对角矩阵。

对方阵的阶n作数学归纳法；当$n=1$时，$(a)\simeq(a)$；假设$n-1$阶对称矩阵必合同于一$n-1$阶对角阵；则对于n阶对称矩阵$A=\left(a_{ij}\right)_{n\times n},a_{ij}=a_{ji}\in F$；**1**）若$a_{11}\neq 0$，对$A$分块，$A=\begin{pmatrix} a_{11} & \alpha \\ \alpha^{\mathrm{T}} & A_1 \end{pmatrix}$，$A_1$为$n-1$阶对称阵；利用分块矩阵的初等变换，采用矩阵零化技巧，则有：$\begin{pmatrix} 1 & O \\ -a_{11}^{-1}\alpha^{\mathrm{T}} & I_{n-1} \end{pmatrix}\begin{pmatrix} a_{11} & \alpha \\ \alpha^{\mathrm{T}} & A_1 \end{pmatrix}\begin{pmatrix} 1 & -a_{11}^{-1}\alpha \\ O & I_{n-1} \end{pmatrix}=\begin{pmatrix} a_{11} & O \\ O & A_1-a_{11}^{-1}\alpha^{\mathrm{T}}\alpha \end{pmatrix}$，并且$\begin{pmatrix} 1 & -a_{11}^{-1}\alpha \\ O & I_{n-1} \end{pmatrix}^{\mathrm{T}}=\begin{pmatrix} 1 & O \\ -a_{11}^{-1}\alpha^{\mathrm{T}} & I_{n-1} \end{pmatrix}$，$\left(A_1-a_{11}^{-1}\alpha^{\mathrm{T}}\alpha\right)^{\mathrm{T}}=A_1-a_{11}^{-1}\alpha^{\mathrm{T}}\alpha$；从而，$A\simeq\begin{pmatrix} a_{11} & O \\ O & A_2 \end{pmatrix}$，$A_2=A_1-a_{11}^{-1}\alpha^{\mathrm{T}}\alpha$；由归纳假设，存在域$F$上$n-1$阶非异阵$C_2$，使得$C_2^{\mathrm{T}}A_2C_2=D_2$—对角阵；从而，$\begin{pmatrix} 1 & O \\ O & C_2 \end{pmatrix}^{\mathrm{T}}\begin{pmatrix} a_{11} & O \\ O & A_2 \end{pmatrix}\begin{pmatrix} 1 & O \\ O & C_2 \end{pmatrix}=\begin{pmatrix} a_{11} & O \\ O & D_2 \end{pmatrix}$，即：$\begin{pmatrix} a_{11} & O \\ O & A_2 \end{pmatrix}\simeq\begin{pmatrix} a_{11} & O \\ O & D_2 \end{pmatrix}$；从而，$A\simeq\begin{pmatrix} a_{11} & O \\ O & D_2 \end{pmatrix}$；**2**）若$a_{11}=0$，存在$a_{ii}\neq 0$，此时令$B=P_n^{\mathrm{T}}(1i)AP_n(1i)$；即有$B$的$(1,1)$元为$a_{ii}\neq 0$，以下情形类似**1**）；　**3**）若$a_{11}=a_{22}=\cdots=a_{nn}=0$，存在$a_{ij}\neq 0,i\neq j$，令$H=P_n^{\mathrm{T}}\left(ji(1)\right)AP_n\left(ji(1)\right)$，即有$H$的$(i,i)$元为$2a_{ij}\neq 0$，以下情形类似于**2**）；
4）若$A=O$，结论显然！

以下，我们依次讨论化简二次型的三种方法：①配方法；②正交线性替换法；③成对的初等行、列变换法。

【配方法】又称为 **Lagrange** 方法，此法的缺点是：不能在$C^{\mathrm{T}}AC$化为对角阵的同时即找出C，而常常须多次方阵的乘积或求逆方可得到；因而计算常复杂而费时。

配方法的基础是运用如下公式：

$$\left(x_1+x_2+\cdots+x_n\right)^2=\begin{matrix} x_1^2 & +2x_1x_2 & +\cdots & +2x_1x_n \\ & +x_2^2 & +\cdots & +2x_2x_n \\ & & +\cdots & +\cdots \\ & & & +x_n^2 \end{matrix}$$

。以下，通过几个例题我们来说明如何正确地使用配方法！

【例 **4.4.3**】**1**）将二次型$f\left(x_1,x_2,x_3\right)=x_1^2+2x_1x_2-4x_1x_3-3x_2^2-6x_2x_3+x_3^2$化为标准形；**2**）将无平方项的二次型$f\left(x_1,x_2,x_3\right)=x_1x_2+x_1x_3-3x_2x_3$化为标准形，并写出所作的非退化的线性替换；

【例 4.4.4】用配方法化二次型为标准形时，应如何配方才能保证使用的是满秩线性替换？例如：$f\left(x_1,x_2,x_3\right)=4x_1^2-4x_1x_2+6x_2^2$ $=2x_1^2+2\left(x_1-x_2\right)^2+4x_2^2=2y_1^2+2y_2^2+4y_3^2$，其中线性替换为$\begin{cases}y_1=x_1\\y_2=x_1-x_2\\y_3=x_2\end{cases}$，以上配方是否正确？

【正交线性替换法】利用矩阵理论可得到化实二次型（实数域上的二次型）为标准形的方法：$f(X)=X^{\mathrm{T}}AX$ 为实二次型，A为实对称阵，即存在正交矩阵U，使得$U^{-1}AU=U^{\mathrm{T}}AU=diag\left\{\lambda_1,\lambda_2,\cdots,\lambda_n\right\}$，$\lambda_i$为$A$的特征值；用二次型语言描述即：任一实二次型$f(X)=X^{\mathrm{T}}AX$都可经非退化实线性替换$X=UY$（或称正交线性替换）化为$\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2$，$\lambda_i$为$A$的特征值。

【引例】用正交线性替换化实二次型$f(x,y)=3x^2+4xy+3y^2$为标准形，并写出正交线性替换。

可以验证：正交线性替换阵为$C=\begin{pmatrix}\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\\-\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\end{pmatrix}$，从而，

$f(X)=X^{\mathrm{T}}\begin{pmatrix}3&2\\2&3\end{pmatrix}X\underline{\underline{X=CY}}Y^{\mathrm{T}}\begin{pmatrix}1&0\\0&5\end{pmatrix}Y$，这里，$X=\begin{pmatrix}x\\y\end{pmatrix}$，$Y=\begin{pmatrix}x'\\y'\end{pmatrix}$，也即：原二次型可化为标准形$g(Y)=x'^2+5y'^2$，$1,5$分别为$\begin{pmatrix}3&2\\2&3\end{pmatrix}$的两个特征值。

我们接着考虑平面上的坐标变换以及坐标变换下的二次曲线方程。设$Oxy,Ox'y'$分别是平面上两个笛卡尔直角坐标系，有共同的原点O，从Ox轴沿逆时针方向转到Ox'轴的角是θ，以$(x,y),(x',y')$分别表示同一点P在两坐标系下的坐标，r为O,P两点的距离，则有：

$\begin{cases}x'=r\cos\varphi\\y'=r\sin\varphi\end{cases}$，$\begin{cases}x=r\cos(\theta+\varphi)=x'\cos\theta-y'\sin\theta\\y=r\sin(\theta+\varphi)=x'\sin\theta+y'\cos\theta\end{cases}$，① 类似有：

$\begin{cases}x'=x\cos\theta+y\sin\theta\\y'=-x\sin\theta+y\cos\theta\end{cases}$，② ①、②常称为坐标系的旋转变换公式；也可写作：$\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}\begin{pmatrix}x'\\y'\end{pmatrix}$、$\begin{pmatrix}x'\\y'\end{pmatrix}=\begin{pmatrix}\cos\theta&\sin\theta\\-\sin\theta&\cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$，

上述两矩阵均为正交矩阵。

基于此，二次曲线 $f(x,y)=3x^2+4xy+3y^2=5$ 实际上是一条斜椭圆！

【成对的初等行、列变换法】对于数域 F 上的 n 元二次型，由于 $P^{\mathrm{T}}(ij)=P(ij)$、 $P^{\mathrm{T}}(i(k))=P(i(k))$、 $P^{\mathrm{T}}(ij(c))=P(ji(c))$，且任一非异阵 C 总可表示为若干个初等矩阵的乘积，不妨设 $C=P_1P_2\cdots P_s$， P_i 为初等矩阵；

$$\begin{pmatrix} A \\ \cdots \\ I \end{pmatrix} \xrightarrow[\text{对}I\text{只施行其中的初等列变换}]{\text{对}A\text{施行成对的初等变换}} \begin{pmatrix} P_s^{\mathrm{T}} & O \\ O & I \end{pmatrix}\begin{pmatrix} P_{s-1}^{\mathrm{T}} & O \\ O & I \end{pmatrix}\cdots\begin{pmatrix} P_1^{\mathrm{T}} & O \\ O & I \end{pmatrix}\begin{pmatrix} A \\ I \end{pmatrix}P_1P_2\cdots P_s$$

$=\begin{pmatrix} C^{\mathrm{T}} & O \\ O & I \end{pmatrix}\begin{pmatrix} A \\ I \end{pmatrix}C=\begin{pmatrix} C^{\mathrm{T}}AC \\ C \end{pmatrix}$，这里， $C^{\mathrm{T}}AC=diag\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$，从而有 $f(X)=X^{\mathrm{T}}AX$ 的标准形 $\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2$；

$$(A \ \vdots \ I) \xrightarrow[\text{对}I\text{只施行其中的初等行变换}]{\text{对}A\text{施行成对的初等变换}} P_s^{\mathrm{T}}P_{s-1}^{\mathrm{T}}\cdots P_1^{\mathrm{T}}(A \ \vdots \ I)\begin{pmatrix} P_1 & O \\ O & I \end{pmatrix}\begin{pmatrix} P_2 & O \\ O & I \end{pmatrix}\cdots\begin{pmatrix} P_s & O \\ O & I \end{pmatrix}$$

$=C^{\mathrm{T}}(A \ \vdots \ I)\begin{pmatrix} C & O \\ O & I \end{pmatrix}=(C^{\mathrm{T}}AC \ \vdots \ C^{\mathrm{T}})$。

【例 4.4.5】将实二次型 $f(x_1,x_2,x_3)=x_1x_2+x_1x_3-3x_2x_3$ 化为标准形；

§5 实二次型、正定二次型与正定矩阵

同一个二次型，其标准形往往并不唯一；但标准形中系数非零的平方项个数却相同，这是它们的共性。

【定理】二次型 $f(x_1,x_2,\cdots,x_n)=f(X)=X^{\mathrm{T}}AX$ 的标准形中，系数非零的平方项个数等于其相伴矩阵 A 的秩。

设非退化线性替换 $X=CY$ 可化二次型 $X^{\mathrm{T}}AX$ 为标准形 $d_1y_1^2+d_2y_2^2+\cdots+d_ry_r^2$，即有： $C^{\mathrm{T}}AC=diag\{d_1,d_2,\cdots,d_r,0,\cdots,0\}$， $d_i\neq 0$，显然有 $r(C^{\mathrm{T}}AC)=r$，即： $r(A)=r$；以后常称 A 的秩为二次型 $X^{\mathrm{T}}AX$ 的秩。

以下，我们仅讨论实数域上规范形的唯一性问题！

设 $f(x_1,x_2,\cdots,x_n)=f(X)$ 是秩为 r 的实二次型，则经适当非退化实线性替换可化为标准形；标准形中非零的 r 个平方项系数可正可负，适当调整即有：

$f(x_1,x_2,\cdots,x_n)=d_1y_1^2+\cdots+d_py_p^2-d_{p+1}y_{p+1}^2-\cdots-d_ry_r^2$，其中：$d_i>0,i=1,2,\cdots,r$；再作非退化实线性替换： $z_1=\sqrt{d_1}y_1,\cdots,z_p=\sqrt{d_p}y_p,\cdots,z_r=\sqrt{d_r}y_r,$

$z_{r+1}=y_{r+1},\cdots,z_n=y_n$；即得：$f\left(x_1,x_2,\cdots,x_n\right)=z_1^2+\cdots+z_p^2-z_{p+1}^2-\cdots-z_r^2$，称之为实二次型 $f\left(x_1,x_2,\cdots,x_n\right)$ 的规范形。

【惯性定理】n 元实二次型 $X^{\mathrm{T}}AX$ 的规范形是唯一的。

设 $r(A)=r$，$f(X)=X^{\mathrm{T}}AX$，其经非退化线性替换 $X=CY$，$X=BZ$，变成两个规范形：

$X^{\mathrm{T}}AX=y_1^2+\cdots+y_p^2-y_{p+1}^2-\cdots-y_r^2=z_1^2+\cdots+z_q^2-z_{q+1}^2-\cdots-z_r^2$，不妨设 $p>q$，则经过非退化的线性替换 $Z=\left(B^{-1}C\right)Y$，有

$z_1^2+\cdots+z_q^2-z_{q+1}^2-\cdots-z_r^2=y_1^2+\cdots+y_p^2-y_{p+1}^2-\cdots-y_r^2$，（$*$）；设 $D=B^{-1}C=\left(d_{ij}\right)_{n\times n}$，试找一组 $y_1,y_2,\cdots,y_n$，使得上述（$*$）式右端大于零而左端小于等于零，从而产生矛盾；为此，令 $\left(y_1,y_2,\cdots y_p,y_{p+1},\cdots,y_n\right)=\left(k_1,k_2,\cdots,k_p,0,\cdots,0\right)$，$k_i$ 不全为零，并且使 $z_1,z_2,\cdots z_q$ 取值全为零；

由于 $\begin{pmatrix} z_1\\ \vdots\\ z_q\\ z_{q+1}\\ \vdots\\ z_n\end{pmatrix}=\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n}\\ d_{21} & d_{22} & \cdots & d_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ d_{n1} & d_{n2} & \cdots & d_{nn}\end{pmatrix}\begin{pmatrix} k_1\\ \vdots\\ k_p\\ 0\\ \vdots\\ 0\end{pmatrix}$，故，$\begin{cases} z_1=d_{11}k_1+d_{12}k_2+\cdots+d_{1p}k_p\\ z_2=d_{21}k_1+d_{21}k_2+\cdots+d_{2p}k_p\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ z_q=d_{q1}k_1+d_{q1}k_2+\cdots+d_{qp}k_p\end{cases}$；

$p>q$，而齐次方程组

$\begin{cases} d_{11}x_1+d_{12}x_2+\cdots+d_{1p}x_p=0\\ d_{21}x_1+d_{21}x_2+\cdots+d_{2p}x_p=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ d_{q1}x_1+d_{q1}x_2+\cdots+d_{qp}x_p=0\end{cases}$ 有非零解，于是，$k_1,k_2,\cdots,k_p$ 可以取到一组不全为零的实数，使得 $z_1=z_2=\cdots=z_q=0$，矛盾！即有：$p\le q$；类似可证 $q\le p$，从而 $p=q$。

【定义】在实二次型 $X^{\mathrm{T}}AX$ 的规范形中，系数为 1 的平方项个数 p 称为 $X^{\mathrm{T}}AX$ 的正惯性指数，系数为 -1 的平方项个数 $r-p$ 称为 $X^{\mathrm{T}}AX$ 的负惯性指数，正、负惯性指数之差 $2p-r$ 称为 $X^{\mathrm{T}}AX$ 的符号差。

由该定义，惯性定理可如下叙述，

【二次型语言】两个 n 元实二次型等价 $\Leftrightarrow$ 它们的规范形相同 $\Leftrightarrow$ 它们的秩相等且正惯性指数也相等，且秩及正（负）惯性指数唯一；【矩阵语言】任一 n 阶实对称阵 A 必合同于对角阵

$\begin{pmatrix} I_p & O & O \\ O & -I_{r-p} & O \\ O & O & O \end{pmatrix}$，其中 $p, r-p$ 分别为正、负惯性指数，该对角阵称为 A 的合同规范形；由此，若 A, B 均为阶实对称阵，则 A, B 在实数域上合同 $\Leftrightarrow A, B$ 有相同的秩和相同的正惯性指数。

【注1】　合同关系是一种等价关系，它可将二阶实对称阵组成的集合分成哪几类，我们可尝试写出每类中一个最简单的矩阵，即：合同规范形；①秩为 **0** 的矩阵：$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$；②秩为 **1** 的矩阵：正惯性指数有两种可能 **0**、**1**，合同规范形分别为 $\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$；③秩为 **2** 的矩阵：正惯性指数有三种可能 **0**、**1**、**2**，它们各成一类，合同规范形分别为 $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$；综上所述共分成 **6** 类！若将 n 阶实对称阵按合同关系分类，即两个 n 阶实对称阵属于同一类当且仅当它们在实数域上合同。

二次型（函数）$f(x) = x^2$ 是最简单的二次型，$\forall a \in R, a \neq 0$，都有 $f(a) > f(0)$，这就表明 $f(x) = x^2$ 的最小值问题与一元二次型 x^2 的性质密切相关，由此可猜测：n 元二次函数的极值问题与 n 元二次型的性质有着联系！

【定义】$X^{\mathrm{T}}AX$ 为 n 元二次型，若 $\forall X \in R_n$，$X \neq (0,0,\cdots,0)^{\mathrm{T}}$，都有：①$f = X^{\mathrm{T}}AX > 0$，则称 f 为正定的；②$f = X^{\mathrm{T}}AX < 0$，则称 f 为负定的；③$f = X^{\mathrm{T}}AX \geq 0$，则称 f 为半正定（非负定）的；④$f = X^{\mathrm{T}}AX \leq 0$，则称 f 为半负定（非正定）的；⑤若 f 既不满足③，又不满足④，则称 f 为不定的。

例如：3 元实二次型 $X^{\mathrm{T}}AX = x_1^2 + x_2^2 + x_3^2$ 是正定的，2 元（3 元）实二次型 $X^{\mathrm{T}}BX = x_1^2 + x_2^2$ 是（半）正定的，而 3 元实二次型 $X^{\mathrm{T}}CX = x_1^2 + x_2^2 - x_3^2$ 是不定的。

由上述定义，易知：

1）$f(x_1, x_2, \cdots, x_n)$ 负定 $\Leftrightarrow -f(x_1, x_2, \cdots, x_n)$ 正定；**2**）$f(x_1, x_2, \cdots, x_n)$ 半负定 $\Leftrightarrow -f(x_1, x_2, \cdots, x_n)$ 半正定。

【定理】n 元实二次型 $X^{\mathrm{T}}AX$ 正定 $\Leftrightarrow X^{\mathrm{T}}AX$ 的正惯性指数等于 n；

从而，n元实二次型$X^{\mathrm{T}}AX$正定$\Leftrightarrow$规范形为$y_1^2+y_2^2+\cdots+y_n^2\Leftrightarrow$标准形中$n$个系数全大于零。

【定义】实对称阵A称为正定的，如果实二次型$X^{\mathrm{T}}AX$是正定的；即：对于R_n中任意非零列向量α，有$\alpha^{\mathrm{T}}A\alpha>0$；正定的实对称矩阵简称为正定矩阵。　由此定义，即有，

【定理】n阶实对称阵A正定$\Leftrightarrow A$的正惯性指数为$n\Leftrightarrow A\simeq I_n$ $\Leftrightarrow A$的合同标准形中，主对角元全大于 **0**；

【推论】①n阶实对称阵A正定当且仅当A的特征值全大于 **0**，从而正定矩阵的行列式大于 **0**；②正定矩阵的合同对称阵也是正定矩阵③与正定二次型等价的实二次型也是正定的，也即：非退化线性替换不改变实二次型的正定性。

【例 **4.5.1**】**1**）设A是实对称阵，试证：当实数t充分大时，$tI+A$是正定矩阵；**2**）若A,B均是正定矩阵，则$A+B$也是正定阵；**3**）若A是正定阵，则A^{-1},A^*也是正定阵；

【例 **4.5.2**】设A是正定实对称阵，试证：对R_n中任意两个列向量x,y，有$\left(x^{\mathrm{T}}Ay\right)^2\le\left(x^{\mathrm{T}}Ax\right)\left(y^{\mathrm{T}}Ay\right)$，等号成立当且仅当$x,y$线性相关；

【例 **4.5.3**】试证：若$\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j\left(a_{ij}=a_{ji}\right)$是正定二次型，则

$$f\left(y_1,y_2,\cdots,y_n\right)=\begin{vmatrix} a_{11} & \cdots & a_{1n} & y_1 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} & y_n \\ y_1 & \cdots & y_n & 0 \end{vmatrix}$$是负定二次型；

【例 **4.5.4**】**1**、（**1**）若实二次型$f(X)=X^{\mathrm{T}}AX$的矩阵$A=\left(a_{ij}\right)_{n\times n}$的行列式$|A|<0$，试证：存在实$n$维向量$X$，使得$f(X)<0$；（**2**）设$f(X)=X^{\mathrm{T}}AX$是一实$n$元二次型，若有$n$维向量$X_1,X_2$，使得$f(X_1)>0$，$f(X_2)<0$，则：**a**）$X_1,X_2$线性无关；**b**）存在$n$维向量$X_0\ne\theta$，使得$f(X_0)=0$；（**3**）设二次型$f(X)=X^{\mathrm{T}}AX$的正、负惯性指数都不为零，试证：存在非零向量$X_1,X_2,X_3$，使得$f(X_1)>0$，$f(X_2)=0$，$f(X_3)<0$；

【例 **4.5.5**】试求函数$f(X)=X^{\mathrm{T}}AX+2\beta X+b$的最小值，其中$A$是$n$阶正定阵，$\beta=\left(b_1,b_2,\cdots,b_n\right)$，$X=\left(x_1,x_2,\cdots,x_n\right)^{\mathrm{T}}$，$b,b_i\in R$；

【例 4.5.6】（1）设 $A=\left(a_{ij}\right)_{n\times n}$ 正定，则有：$|A|\le\prod_{i=1}^{n}a_{ii}$；（2）若 A 是 n 阶正定阵，则 $|A|\le\left(\frac{1}{n}trA\right)^{n}$；

【例 4.5.7】（1）设实对称阵 A 满足 $A^3-3A^2+5A-3E=O$，则 A 正定；（2）设 $A=\left(a_{ij}\right)_{n\times n}$ 正定，$b_i\ne 0$ 为任意实数，则 $B=\left(a_{ij}b_ib_j\right)_{n\times n}$ 也是正定阵；（3）设 $A=\left(a_{ij}\right)_{m\times n}$ 为实矩阵，$B=\lambda E+A^{\mathrm{T}}A$，$E$ 为 n 阶单位阵，则当 $\lambda>0$ 时，矩阵 B 为正定阵；（4）设实矩阵 $A=\left(a_{ij}\right)_{m\times n}$，且 $r(A)=n$，则二次型 $f(X)=X^{\mathrm{T}}\left(A^{\mathrm{T}}A\right)X$ 正定，其中 $X=\left(x_1,x_2,\cdots,x_n\right)^{\mathrm{T}}$；

【例 4.5.8】（1）试证：对于任意实 n 元二次型 $f(X)=X^{\mathrm{T}}AX$（A 不必对称），存在一个正数 M，使得对任意 n 维列向量 X，$|f(X)|\le MX^{\mathrm{T}}X$；（2）试求实 n 元二次型 $f(X)=X^{\mathrm{T}}AX$ 在条件 $\sum_{i=1}^{n}x_i^2=1$ 下的最大、小值，其中 $X=\left(x_1,x_2,\cdots,x_n\right)^{\mathrm{T}}$；

行列式大于零的实对称阵，却不一定是正定的，如：$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$，以下从“子式”的角度讨论实对称阵。

【定义】设 A 是一 n 阶方阵，A 的一个子式称为主子式，如果它的行指标与列指标相同，即形如：$A\begin{pmatrix}i_1&i_2&\cdots&i_k\\i_1&i_2&\cdots&i_k\end{pmatrix}$，$A$ 的主子式 $A\begin{pmatrix}1&2&\cdots&k\\1&2&\cdots&k\end{pmatrix}$ 称为 A 的 k 阶顺序主子式，$k=1,2,\cdots,n$。

可以验证：$A=\left(a_{ij}\right)_{n\times n}$ 的主子式共有 $C_n^1+C_n^2+\cdots+C_n^n=2^n-1$ 个，而顺序主子式只有 n 个。

【定理】实对称阵 A 正定的充要条件是 A 的所有顺序主子式全大于零。

必要性：设 n 阶实对称阵 A 正定，将 A 分块 $A=\begin{pmatrix}A_k&B_1\\B_1^{\mathrm{T}}&B_2\end{pmatrix}$，则 $|A_k|$ 是 A 的 k 阶顺序主子式；以下证明 A_k 正定，$\forall\delta(\ne\theta)\in R_k$，由 A 正定，$\left(\delta^{\mathrm{T}},\theta\right)\begin{pmatrix}A_k&B_1\\B_1^{\mathrm{T}}&B_2\end{pmatrix}\begin{pmatrix}\delta\\\theta^{\mathrm{T}}\end{pmatrix}=\delta^{\mathrm{T}}A_k\delta>0$；

从而 A_k 正定，故而 $|A_k|>0,k=1,2,\cdots,n$；

充分性：对实对称阵的阶数 n 作数学归纳法！当 $n=1$ 时，$|(a)|>0$

即 $a>0$，从而 (a) 正定；假设对于 $n-1$ 阶实对称阵命题成立，则对于 n 阶实对称阵 $A=\left(a_{ij}\right)_{n\times n}$，将 A 分块 $A=\begin{pmatrix} A_{n-1} & \alpha \\ \alpha^{\mathrm{T}} & a_{nn} \end{pmatrix}$，$A_{n-1}$ 为 $n-1$ 阶实对称阵，且 $|A_{n-1}|>0$，由行列式的降阶定理，有：

$|A|=|A_{n-1}|\cdot\left|a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha\right|$，从而由 $|A|>0$，知 $a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha>0$；可以验证：

$A\simeq\begin{pmatrix} A_{n-1} & O \\ O & a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha \end{pmatrix}$，由于 A_{n-1} 正定，故存在 $n-1$ 阶非异阵 C_1，使得 $C_1^{\mathrm{T}}A_{n-1}C_1=I_{n-1}$；从而，$\begin{pmatrix} C_1 & O \\ O & 1 \end{pmatrix}^{\mathrm{T}}\begin{pmatrix} A_{n-1} & O \\ O & a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha \end{pmatrix}\begin{pmatrix} C_1 & O \\ O & 1 \end{pmatrix}=$

$\begin{pmatrix} I_{n-1} & O \\ O & a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha \end{pmatrix}$，即有：$\begin{pmatrix} A_{n-1} & O \\ O & a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha \end{pmatrix}\simeq\begin{pmatrix} I_{n-1} & O \\ O & a_{nn}-\alpha^{\mathrm{T}}A_{n-1}^{-1}\alpha \end{pmatrix}$，而右端矩阵是正定的，故 A 是正定的！

【推论】实二次型 $X^{\mathrm{T}}AX$ 正定的充要条件是 A 的所有顺序主子式全大于零。

【例 4.5.4】试证：实对称矩阵负定的充要条件是：它的偶数阶顺序主子式全大于零，奇数阶顺序主子式全小于零；

【例 4.5.5】1）若 A 是 n 阶正定矩阵，则 $\left|A+A^{-1}\right|\geq 2^n$；2）已知二次型 $f=f\left(x_1,x_2,x_3\right)=x_1^2+x_2^2+x_3^2-k^2\left(ax_1+bx_2+cx_3\right)^2$，其中 a,b,c 不全为零，$k\neq 0$，试问：a,b,c,k 满足何条件，f 是正定二次型？

概率论与数理统计讲义

授课教师：郭志军（2011，8－第一次修订）

【前言】本讲义虽经一次修订，但仍系仓促成稿，错漏之处在所难免；请读者不吝赐教！（联系方式：guozhijun20062006@yahoo.com.cn）讲义中所选习题系作者多年教学积累；讲义的内容几经锤炼且有别于绝大多数讲义、教材；其最大特色在于架起了初等概率论与高等概率论之间的友好“桥梁”，特别适合期待进一步深造的经管类学生！讲义后所附的阅读材料可供学有余力时选读，其所选内容会在以后的修订中有所调整或增删！

第五章 数理统计的基本概念（见教材）
第六章 参数估计、假设检验
§1（参数）点估计
§2（参数）区间估计
§3 假设检验

【附录部分】

阅读材料一－－关于正态分布的推导（基于极大似然方法）
阅读材料二－**R**－**S** 积分及其性质
阅读材料三－条件期望的另一种计算方法
阅读材料四－微分法求（联合）概率密度
阅读材料五－期望向量、协方差矩阵与多元正态分布

【参考阅读书目】

1． 随机数学－钱敏平、叶俊著，高等教育出版社；
2． 随机数学引论－林元烈著，清华大学出版社；
3． 概率论－杨振明著，科学出版社；
4． 概率与统计－陈家鼎、郑忠国著，北京大学出版社；
5． 概率论基础教程－**Sheldon M.Ross** 著（中译本：郑忠国、詹从赞译），人民邮电出版社；
6． 概率论与数理统计教程－茆诗松著，高等教育出版社；
7． 应用概率统计－刘嘉焜等著，科学出版社；
8． 测度与积分－赵荣侠等著，西安电子科技大学出版社；

【概率论部分】

第一章　随机事件及其概率

§1 概率论的发展简史

概率论是研究随机现象数量规律性的一门学科，它源于对机会游戏（赌博问题）的研究。概率概念的要旨只是在 **17** 世纪中叶法国数学家帕斯卡**(Pascal)**与费马**(Fermat)**的讨论中才比较明确，他们在往来的信函中讨论"合理分配赌注问题"；该问题可以简化为：

甲、乙两人同掷一枚硬币，各出相同的赌注；规定：掷出正面者获胜；先胜满三局者赢取全部赌注。假定在甲胜二局乙胜一局时，赌局由于某种原因中止了，问应该怎样分配赌注才算公平合理。

帕斯卡：若再掷一次甲胜，甲获全部赌注， 两种情况可能性相同，所以这两种情况平均一下；若乙胜，甲、乙平分赌注；甲应得总赌注的$3/4$，乙得总赌注的$1/4$。

费马：结束赌局至多还要2局，结果为四种等可能情况：

情况　1　2　3　4

赌局　甲甲　甲乙　乙甲　乙乙　；前3种情况，甲获全部赌注，仅第四种情况，乙获全部赌注。所以甲分得总赌注的$3/4$，乙得总赌注的$1/4$。

帕斯卡与费马各自用不同的方法解决了这个问题。虽然他们在解答中没有明确定义概念；但是他们定义了使该赌徒取胜的机遇，也就是赢得情况数与所有可能情况数的比，这实际上就是概率，所以概率的发展被认为是从帕斯卡与费马开始的。在概率问题早期的研究中，逐步建立了事件、概率和随机变量等重要概念以及它们的基本性质。后来许多社会问题和工程技术问题，如：人口统计、保险理论、天文观测、误差理论、产品检验和质量控制等地提出均促进了概率论的发展。从 **17** 世纪到 **19** 世纪，贝努利、棣莫弗、拉普拉斯、高斯、泊松、切比雪夫、马尔可夫等著名数学家都对概率论的发展做出了杰出的贡献。在这段时间里，概率论的发展简直到了使人着迷的程度。但是，随着概率论中各个领域获得大量成果以及在其他基础学科和工程技术上的应用，由拉普拉斯给出的概率定义（古典概率）的局限性很快便暴露了出来，其甚至无法适用于一般的随机现象。因此可以说，到 **20** 世纪初，概率论的一些基本概念，诸如概率等尚没有确切的定义，概率论作为一个数学分支缺乏严格的理论基础。概率论的第一本专著是 **1713** 年问世的雅各布·贝努利的《推测术》。经过二十多年的艰难研究，贝努利在书中表述并证明了著名的“大数定律”。所谓“大数定律”，简单地说就是，当试验次数很大时，事件出现的频率与概率有较大偏差的可能性很小。这一定理第一次在单一的概率值与众多现象的统计度量之间建立了演绎关系，构成了从概率论通向更广泛应用领域的桥梁。因此，贝努利被称为概率论的奠基人。

为概率论首先建立严格理论基础的是前苏联数学家柯

尔莫哥洛夫（**Kolmogorov**）。**1933** 年，他发表了著名的《概率论的基本概念》并于其中建立了概率论的公理化体系，成为概率论发展史上的一个里程碑，为其以后的迅速发展奠定了基础。

20 世纪以来，由于物理学、生物学、工程技术、农业技术和军事技术发展的推动，概率论获得了飞速发展，其理论课题不断扩大与深入，应用范围大大拓宽。在最近几十年中，概率论的方法被引入各个工程技术学科和社会学科。目前，概率论在近代物理、自动控制、地震和气象预报、质量控制、农业试验和公用事业等方面都得到了重要应用；其方法已被越来越多地引入经济、金融和管理科学领域，出现了诸如：统计物理学，生物统计学，随机微分方程（随机微积分或随机分析），随机信号分析，随机振动分析，随机运筹学，金融随机分析等等等交叉学科；特别是二十世纪以来，作为概率统计的一个新兴领域随机过程，获得了迅猛的发展，迄今，已成为经济金融等学科不可或缺的重要工具！

现在，概率论已发展成为一门与实际紧密相连的理论严谨的数学科学。它内容丰富，结论深刻，有别开生面的研究课题，有自己独特的概念和方法，已经成为了近代数学一个有特色的分支。

§2 随机事件及其概率

$$\text{自然界中的现象}\begin{cases}\text{必然（确定性）现象}\\ \text{随机（偶然性）现象}\begin{cases}\text{个别随机现象}\\ \text{大量随机现象}\end{cases}\end{cases}$$

【**必然现象**】在一定的条件下，一定会出现（发生）的现象。

【**随机现象**】在一定的条件下，可能出现也可能不出现的现象。

【注 **1**】在概率论中，“出现”与“发生”同义。

概率论主要研究大量随机现象（统计规律性），但是也不排斥个别随机现象。

$$\text{随机性的作用}\begin{cases}\text{很小以致可以忽略}\\ \text{基本的，如布朗运动}\end{cases}$$

【注 **2**】随机现象中的“不确定性”有两层含义，一则指客

观结果的不确定性；一则指主观猜测的不确定性，后者融入了观察者个人的信念。

【统计规律性】大量随机现象所具有的规律性，概率论主要研究此。

对于随机现象，即使条件完全相同，它们的出现所产生的结果也不尽相同，此之谓“现象的随机性”；那么随机性产生的原因是什么呢？

任何随机现象都是相互联系和相互影响的，它的行为受许多因素地支配或制约；而控制所有这些因素原则上做不到，往往只限于决定该现象状态的最基本因素。并且除此之外还有许多时隐时现，转瞬即逝，无法控制的偶然因素；当随机现象重复出现时，这些因素产生的效应是不同的，不确定的，从而使得现象带有随机性。

【试验】凡对现象的观察或为此而进行的实验。

【决定性试验】凡对决定性（必然）现象的观察或为此而进行的实验。

【随机试验】凡对随机现象的观察或为此而进行的实验，常记为 E **(experiment)**或 E_i。

【事件（试验的结局）】试验观察的结果。

无论何种试验，都包含两个方面，即试验的条件和试验的结果。随机试验的条件有的是“人为的”，如在一定的条件下观察“射击是否命中目标”；有的是“不依人的意志为转移的”，如花粉微粒的无规则运动（布朗运动）。

【统计规律性】大量重复试验中随机现象所呈现的固有规律。

【随机事件】随机试验的结果，常简称为事件。

【注 **3**】为以后研究方便，有时把有固有结果的试验看作是随机试验的极端情形；有时需把几次试验作为一个整体看成一次随机试验，如连续地掷三次骰子。同理，也将必然事件和不可能事件视作随机事件的极端情形。

【必然事件】每次试验中一定会出现的事件，记作 Ω；

【不可能事件】每次试验中一定不会出现的事件，记作 Φ；

【注 **4**】任何随机试验都伴随有必然事件和不可能事件，如

E：对某目标进行两次射击，“至多命中目标两次”就是必然事件，“命中目标三次”就是不可能事件。常用大写英文（拉丁）字母 A,B,C 等或 A_i,B_j,C_k 等表示；有时也用**{……}**、**“……”**表示事件，花括弧中和双引号下指明事件的内容。

随机试验的共同特点为；

1、在相同的条件下可重复进行；

2、每次试验的结果可能不止一个，但事先明确所有可能的结果；

3、试验之前不能确定那个结果会出现。

概率论只关心在随机试验中可能会观察到的那些事件以及每次具体的试验中出现了的事件；因此，与每个随机试验相联系的有一个事件的集合，即在试验中可以观察到的事件的全体。至于这个事件集应该具备什么性质，以后将会讨论。既然数学本身从来不只研究那些只由孤立元素组成的集合，我们就有必要在上述事件集中定义事件之间的各种关系与运算。

【注 **5**】自从集合论进入了概率论，概率论才真正进入了现代化门槛。

事件的关系

1.【包含关系】若事件 A 出现必然会导致事件 B 出现，则称“A 是 B 的特款”或“A 包含于 B”，记作 $A\subset B$；

易见对任意事件 $A\subset\Omega$，这里规定 $\Phi\subset A$；

2.【等价（相等）关系】若事件 A,B 满足 $A\subset B$ 且 $B\subset A$,则称事件 A,B 等价或相等；

【注 **6**】在概率论中，对同一事件给出不同的等价表示是一种主要的技巧。

事件的运算

1【事件的并运算】设 A,B 为两事件，则“A,B 至少一个发生”这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A,B 的并（事件），记作：$A\cup B$；

若 $A_1,A_2,\cdots,A_n$ 均为事件，则“$A_1,A_2,\cdots,A_n$ 至少一个发生”作为事件，称之为 $A_1,A_2,\cdots,A_n$ 的并事件，记作：$A_1\cup A_2\cup\cdots\cup A_n$，简

记为：$\bigcup\limits_{i=1}^{n}A_i$；

若 $A_1,A_2,\cdots,A_n,\cdots$ 为一列事件，则“$A_1,A_2,\cdots,A_n,\cdots$ 至少一个发生”作为事件，称之为 $A_1,A_2,\cdots,A_n,\cdots$ 的并事件，记作：

$A_1\cup A_2\cup\cdots\cup A_n\cup\cdots$，简记为：$\bigcup\limits_{n=1}^{\infty}A_n$；（可列并）

2【事件的交运算】设 A,B 为两事件，则“A,B 同时发生” 这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A,B 的交（积）（事件），记作：$A\cap B$；

若 $A_1,A_2,\cdots,A_n$ 均为事件，则“$A_1,A_2,\cdots,A_n$ 同时发生”作为事件，称之为 $A_1,A_2,\cdots,A_n$ 的交事件，记作：$A_1\cap A_2\cap\cdots\cap A_n$，简记为：$\bigcap\limits_{i=1}^{n}A_i$；

若 $A_1,A_2,\cdots,A_n,\cdots$ 为一列事件，则“$A_1,A_2,\cdots,A_n,\cdots$ 同时发生”作为事件，称之为 $A_1,A_2,\cdots,A_n,\cdots$ 的交事件，记作：$A_1\cap A_2\cap\cdots\cap A_n\cap\cdots$，简记为：$\bigcap\limits_{n=1}^{\infty}A_n$；（可列交）

若 A,B 两事件不可能同时发生，则称 A,B 互不相容（互斥），记作：$A\cap B=\Phi$；若 $A_1,A_2,\cdots,A_n$ 互不相容（两两互不相容），则又称 $A_1,A_2,\cdots,A_n$ 的并为 $A_1,A_2,\cdots,A_n$ 的和，记作：$\sum\limits_{i=1}^{n}A_i$；即有：$\bigcup\limits_{i=1}^{n}A_i=\sum\limits_{i=1}^{n}A_i$；若 $A_1,A_2,\cdots,A_n,\cdots$ 互不相容（两两互不相容），则又称 $A_1,A_2,\cdots,A_n,\cdots$ 的并（可列并）为 $A_1,A_2,\cdots,A_n,\cdots$ 的和，记作：$\sum\limits_{n=1}^{\infty}A_n$；即有：$\bigcup\limits_{n=1}^{\infty}A_n=\sum\limits_{n=1}^{\infty}A_n$；

3【事件的差运算】设 A,B 为两事件，则“A 发生而 B 不发生”这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A,B 的差事件，记作：$A-B(A/B)$；

利用差运算，可以将“有交并”转化为“不交并”，如：$A\cup B=(A-B)\cup B=(A-B)+B\left[(B-A)\cup A=(B-A)+A\right]$；

4【事件的逆（余）运算】设 A 为事件，则“A 不发生” 这种情况可能出现也可能不出现，其作为一个随机事件，称之为 A 的逆（余）（事件），记作：$\overline{A}\left(A^c\right)$；

【差积转化律】 $A-B=A\overline{B}$ ；

5【事件列的极限（运算）】

规定： $\boxed{\bigcup_{n=1}^{\infty} A_n = \lim_{n\to\infty}\bigcup_{i=1}^{n} A_i}$ ， $\boxed{\bigcap_{n=1}^{\infty} A_n = \lim_{n\to\infty}\bigcap_{i=1}^{n} A_i}$ ；依此若事件列$\{A_n, n\ge 1\}$，

(a)单调递增，即：$A_1\subset A_2\subset\cdots\subset A_n\subset\cdots$；则定义事件列的极限为

$$\boxed{\lim_{n\to\infty} A_n = \lim_{n\to\infty}\bigcup_{i=1}^{n} A_i = \bigcup_{n=1}^{\infty} A_n} \quad ;$$

(b)单调递减，即：$A_1\supset A_2\supset\cdots\supset A_n\supset\cdots$；则定义事件列的极限为

$$\boxed{\lim_{n\to\infty} A_n = \lim_{n\to\infty}\bigcap_{i=1}^{n} A_i = \bigcap_{n=1}^{\infty} A_n} \quad ;$$

对于任意的随机事件列$\{A_n, n\ge 1\}$，令，

(1) $B_n = \bigcup_{k=n}^{\infty} A_k, n\ge 1$，则$\{B_n, n\ge 1\}$单调递减，从而有：

$\lim_{n\to\infty} B_n = \lim_{n\to\infty}\bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$ ；称此极限为$\{A_n, n\ge 1\}$的<u>上极限</u>，记作：$\limsup_{n\to\infty} A_n$，即有：$\boxed{\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k}$；

(2) $C_n = \bigcap_{k=n}^{\infty} A_k, n\ge 1$，则$\{C_n, n\ge 1\}$单调递增，从而有：

$\lim_{n\to\infty} C_n = \lim_{n\to\infty}\bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$ ；称此极限为$\{A_n, n\ge 1\}$的<u>下极限</u>，记作：$\liminf_{n\to\infty} A_n$，即有：$\boxed{\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k}$；

(3) 若$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$，则称此极限为$\{A_n, n\ge 1\}$的<u>极限</u>，也即：

$$\boxed{\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \lim_{n\to\infty} A_n} \quad ;$$

【注 7】在概率论中，所谓“定义一个事件”，即指其是否发生；因此，对于很多表述，往往有“名词性”和“动词性”两种意味；如：A既表示A事件，又可表示A事件发生；其他情形类似！

【例 1.2.1】设A, B, C为三个事件，试表示如下事件；

1、“A发生”；2、“只有A发生”；3、“三事件中至少两个发生”；4、“三事件中不多于一个发生”；

【例 1.2.2】在某系任选一名学生，A＝“被选学生是男生”，B＝{被选学生是三年级男生},C＝{被选学生是运动员}，

1、试述$AB\bar{C}$的意义；　　2、在何情况下，$ABC = C$？

3、在何情况下，$\bar{A} = B$？

【例 1.2.3】若A表示“产品甲畅销且产品乙滞销”，试述$\overline{A}$；

【例 1.2.4】求事件列$\left\{\left(a-\frac{1}{n},b\right),n\geq 1\right\},\left\{\left[a,b-\frac{1}{n}\right],n\geq 1\right\}$的上(下)极限；

事件的概率

几何学中，线段的长度，平面（空间）图形的面积（体积），物理学中物质的某些量等都可用数值来度量；我们观察一个随机试验的诸事件，总发现“有些事件出现的可能性大些，有些事件出现的可能性小些”，而这些“可能性大小”自然也可用数值来度量，这是由于事件出现的“可能性大小”是客观存在的。这个刻划事件发生可能性大小的数值至少应该满足两个要求；

1、它具有一定的客观性，不能随意改变，而且理论上应可通过在“相同条件下”大量的重复试验加以识别和检验；

2、它必须符合一般常理，如：事件发生可能性大（小）的，这个值就应该大（小）些；必然事件的值最大为1，不可能事件的值最小为0。

【概率的一般（直观）定义】刻划（随机）事件发生（出现）可能性大小的数值（数量指标），又称或然率或几（机）率，它介于0与1之间。

在概率论的发展史上，人们曾针对不同的问题，从不同的角度给出了定义概率和计算概率的各种方法；然而这些“定义的概率”都存在一定的缺陷——与其称它们为概率的定义，弗如称它们为计算概率的方法。下面我们将陆续介绍之。

以下介绍两种等可能概型（概率模型），先考虑一种特殊的随机试验，譬如掷一枚（质地均匀）硬币，人们自然会想到硬币正反两面对称，故出现正面与反面应是一样，从而有理由认为出现正面和反面的可能性都是0.5；在其他一些例子中，人们得出类似的结论也多是由于人们利用了研究对象的物理均衡性或者几何对称性。

古典概型【等可能概型 **1**】－古典概率：概率的古典定义（计算概率的古典方法）

由于它的简单，曾是历史上最早被研究的一类概型，归纳它的特点如下：如果一个随机试验的所有可能结果只有有限个，而且每个结果的出现都是等可能的，则称这个随机试验是古典概型。

设有古典概型E,以$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$表示它的样本空间；则对于任意的事件$A$，若它恰好包含其中的$m$个样本点，则称事件$A$（发生）的（古典）概率（即在古典概型背景下计算概率的古典方法）为$\frac{m}{n}$,记作：$P(A)=\frac{m}{n}$;也即在古典概型下，由古典方法计算出的事件A的（古典）概率为

$$\boxed{P(A)=\frac{\text{事件}A\text{所包含的样本点数（有利于}A\text{的样本点数）}}{\text{样本空间所包含的样本点数}}}。$$

【例 1.2.5】（错例分析）掷两枚硬币，求“出现一正一反”的概率；

【例 1.2.6】1、从$1,2,\cdots,n$中任取两数，求A＝{两数之和为偶数}的概率；**2**、从 **N** 中任取两数，求A＝“两数之和为偶数”的概率；

【例 1.2.7】袋中有a只黑球和b只白球，将球随机一一取出，取后不放回，试球求A＝{第k（$1\le k\le a+b$）次取出黑球}的概率；（四种方法）

“古典概率”有如下性质：

1【非负性】设是A古典概型中任一事件，则$0\le P(A)\le 1$；

2【规范性】（又称规一性或正则性）对必然事件 Ω，$P(\Omega)=1$；

3【有限可加性】设同一古典概型中的事件$A_1,A_2,\cdots,A_n$互不相容，则$P\left(\bigcup_{i=1}^{n}A_i\right)=P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)$；

对于古典概型，“样本点出现的等可能性”是个基本假设； 在实际问题中，往往并不知道这个假设是否满足，而只凭主观对物理均衡性质，几何对称性质来判断是不完全确切的。通常，人们认为“判断事件发生的可能性大小最可靠的办法”即是通过大量重复试验；特别是当样本点不可能判定为等可能出现时尤其要采取这个方法，这便有了“统计概率”的称谓。

统计概率－计算概率的统计方法

【频数】在相同的条件下进行n次试验，在这n次试验中A发生的次数n_A，即为事件A的频数；

【频率】比值称为n_A/n事件发生的频率，记之为$f_n(A)=n_A/n$；

频率具有如下特点：

1、 频率的大小能体现事件发生的可能性大小，频率大（小）的发生的可能性应该大（小）些；

2、 频率具有一定的随机波动性；

3、 随着重复试验次数的增加，频率呈现出相对的稳定性。

事实上，频率具有稳定性这一事实，说明了刻划事件发生可能性大小的概率具有一定的客观存在性，而且频率的稳定性也正反映出大量随机现象的统计规律性。

基于人们不自觉的共识，人们常用频率(frequency) $f(A)$作为事件A的概率的一种量度，这样计算的概率称为统计概率，即为计算概率的统计方法；不妨将$f(A)$写作$P(A)$，从而"统计概率"满足如下性质：

1【非负性】设A是任一事件，则$0 \le P(A) \le 1$；

2【规范性】$P(\Omega)=1$；

3【有限可加性】设事件$A_1, A_2, \cdots, A_n$互不相容，则

$$P\left(\bigcup_{i=1}^{n} A_i\right)=P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i);$$

"统计概率"同样也有理论和应用上的缺陷性，我们没有理由可以这样认为：

1、 取试验次数为$n+1$来计算频率总会比取试验次数为n来计算频率更准确；

2、 实际情况是，我们不知道试验次数究竟取多大；甚至很难保证；在多次（大量）重复试验时，每次试验的条件完全一样。

几何概型【等可能概型 2】：几何概率—概率的几何定义（计算概率的几何方法）

在概率论的发展早期，人们就已注意到只考虑随机现象的可能结果只有有穷个是不够的，还需考虑无穷个的情形；事实上，当试验的可能结果无穷多时，当然不能简单地通过样本点的计数来计算概率；如

【引例】在区间$(0,1)$内任取两个数，求事件A={两数之和小于$6/5$}和B={两数之积不小于$3/16$}的概率；

归纳这类例子的共同特点，即可以通过空间集合的几何度量（如；长度，面积，体积等）来计算概率。

【几何概型（几何概率）】设试验E的样本空间为某可度量的几何区域 Ω，且 Ω中任一子区域（事件）出现的可能性大小与该区域的几何度量成正比，而与该区域的位置和形状无关，则称试验E 为几何概型。若A是Ω 中一区域，且A可度量，则定义事件A的概率$\boxed{P(A)=\frac{A\text{的几何度量}}{\Omega\text{的几何度量}}=\frac{\mu(A)}{\mu(\Omega)}}$，其中若$\Omega$是一维的，二维的或三维的，那么$\Omega$的几何度量分别是长度，面积或体积，称这样定义的概率为几何概率（即计算概率的几何方法）。

由定义，计算概率的几何方法和古典方法类似，也是由一个比值来刻划，只是前者是后者的推广。

【注 8 】求解几何概型归纳起来一般有如下关联的四个步骤：

1 、明确问题的实质，即是否为几何概型；

2 、明确等可能性的几何元素－任何一个几何概型其样本点都可归纳为具有某种等可能性的几何元素；

3 、用几何区域（如；区间，平面区域，空间区域等）来表示样本点数的总和；

3、 利用初等几何或微积分知识求出样本空间Ω 的几何区域的几何度量$\mu(\Omega)$ 和随机事件A的几何区域的几何度量$\mu(A)$，最终由几何方法得到$P(A)=\frac{\mu(A)}{\mu(\Omega)}$,即为$A$的（几何）概率。

【注 9 】通常几何概型大体可分为两类；

(a)样本空间具有明显的几何意义，这类问题结构简单，易于求得；

(b)样本空间所对应的几何区域没有直接给出，这类问题比较复杂；一般可从代数方法入手，引入适当变量，利用代数和几何的联系找出几何区域，再依几何方法计算概率。

【例 **1.2.8**】（会面问题）甲乙两人相约7点到8点的某个时刻在某地会面，先到者等20分钟，过时即离去，试求{两人能

会面}的概率；

【例 1.2.9】(蒲丰[Buffon]投针问题)平面上画着一些平行线，它们的距离都是a,现在向此平面随机地投掷一长度为L的针，试求{此针与任一平行线相交}的概率；

【例 1.2.9 的注】投针问题是法国数学家 G.Buffon 于 1777 年提出的，它成为历史上的著名问题是因为由其结论可以计算的圆周率的近似值。现在利用计算机进行投针的模拟试验就方便多了！需注意的是：要计算一个量，先建立一个概率模型，使得某事件的概率与此量有关，并找出它们之间的关系式，然后进行试验，统计该事件的频率，从而计算出这个量的近似值；这种思想后来发展成为一种数值计算方法，称之谓“蒙特卡罗（Monte-Carlo）方法”。

【例 1.2.10】(贝特朗[Betrand]奇论)在一半径为r的圆内“任意”地作一炫，试求“此炫长度l大于圆内接正三角形边长$\sqrt{3}\,r$”的概率；

【例 1.2.11】(几乎不可能事件）向$[0,1]$区间随机地投掷一点，令A＝{点落在“$1/2$”处}，A_n＝{点落在$\left(\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n}\right)$}，$n\geq 2$，试求$P(A)$；

类似地，“几何概率”具有如下性质：

1【非负性】设A是几何概型中任一事件，则$0\leq P(A)\leq 1$；

2【规范性】对必然事件 Ω，有$P(\Omega)=1$；

3【可列可加性】(又称完全可加性或σ 一可加性）设同一几何概型中一列事件$\{A_n,n\geq 1\}$互不相容，则

$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=P\left(\sum_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)\quad ;$$

几何概型提出了许多新课题；比如：第一、在几何概型中只有样本空间的可测（可度量）子集才能计算概率，因此只有可测子集才能称为事件，不可测子集不能称为事件，它们不是概率论的研究对象；第二、在几何概型中，概率为 **0** 的事件（即零概率事件）未必是不可能事件，概率为 **1** 的事件未必是必然事件等等。

主观概率－计算概率的主观方法

在现实生活中，有些随机现象是不能重复或不能大量重复的，此时该如何确定概率呢。统计界的贝叶斯学派认为：一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信念，这样给出的概率称为“主观概率”。

如：用主观方法确定“大学生考试作弊”的概率；

【注 8】(1)主观概率不同于主观臆造，前者要求当事人对所考察的事件有丰富的经验和透彻的了解，并能对现在和历史信息进行仔细分析；(2)用主观方法得出的可能性大小本质上是对随机事件概率的一种推断和估计，虽然结论有待检验和修正，但可信性在统计上是有价值的；况且在遇到随机现象无法大量重复时，用主观方法作出决策和推断是适合的，其至少是统计（频率）方法的一种补充。当然，主观方法也应符合以下的公理化性质。

§3 概率模型与公理化结构

已经给出的计算概率的古典方法，统计方法，几何方法和主观方法，均存在一定的适用范围以及理论缺陷和应用局限；但不能因此而否定它们在各自的具体问题中计算概率的作用；只是由上述任何“定义”来作为概率的数学定义去建立概率理论是不可能的。广义上讲，一切数学学科都应是公理化的。

公理化思想朴素而又简单，公理化方法使数学合情合理地升离实际，摆脱实践中各种说不清的“纠缠”；正是公理化方法抽象出了数学模型，公理化更是数学严格推理的前提保证。

上述概率的各种“定义”都有共同的属性，这些从客观现实中提炼出的共性，理应作为概率数学理论的基础。

已经知道，联系于一个随机试验有一个样本空间，它是由所有代表“基本事件”的样本点的全体组成；自然会问：

1） 随机事件可由样本空间的一个子集来表示，那么是否样本空间的任何子集都代表一个事件呢？

2） 人们总是希望用最简单的事件的概率来推算复合事件的概率；那么是否所有的基本事件的概率可求；如果可求，是否可由此计算复合事件的概率？

通常的做法是：根据具体情况来确定一组事件，记它们组成的集合为F，要求F包含与问题有关的事件，而且关于可列并，交及余等运算均封闭。

【（事件）σ域（代数）】设Ω是某试验E的样本空间，F是由Ω的一些子集所成的集族（类），若F满足：

（1） $\Omega\in F$； （2）若$A\in F$，则$\overline{A}\in F$；

（3） 若$\forall n\geq 1,\ A_n\in F$,也即$\{A_n,n\geq 1\}\subset F$,，则$\bigcup\limits_{n=1}^{\infty}A_n\in F$；

则称F为（事件）σ域（代数），F中的集称为事件（F可测集）。

易证：F关于有限并（交）、可列交、差、上（下）极限等诸多（**Borel**）运算均封闭。

【注 1】1、$F=\{\Phi,\Omega\}$，这是Ω上的最小σ–域；2、$F=\{A:A\subset\Omega\}$，这是由Ω的所有子集构成的，是Ω上的最大σ–域；3、$F=\{\Phi,A,\overline{A},\Omega\}$，其中$A$是$\Omega$的任一子集，常称此$\sigma$域为由$A$生成的$\sigma$域，记作：$\sigma(A)$；它是包含$A$的最小$\sigma$–域；**4**、基本事件不一定是事件。

【可测空间】将任一样本空间Ω以及建立在其上由其若干子集组成的σ域（代数）F组成的二元体(Ω,F)称为具有σ一域（代数）结构的样本空间或简称可测空间。

【概率(函数，测度)与概率空间(概率场)】设(Ω,F)为一可测空间，对每一集合（事件）$A\in F$，定义实值集（合）函数$P(A)$，使之满足：

（1）（非负性）对每一$A\in F$，有$0\leq P(A)\leq 1$；

（2）（规范性）对必然事件 Ω，有$P(\Omega)=1$；

（3）（可列可加性）若 $\forall n\geq 1, A_n\in F$，且$A_iA_j=\Phi$， 则$P\left(\bigcup\limits_{i=1}^{\infty}A_i\right)=P\left(\sum\limits_{i=1}^{\infty}A_i\right)=\sum\limits_{i=1}^{\infty}P(A_i)$； 称$P(\cdot)$为定义于$(\Omega,F)$上的概率测度（函数），简称概率；并称三元体$(\Omega,F,P)$为概率空间。

【例 **1.3.1**】袋中有大小，质量相同的红，黄，蓝，白球各1个，从中有放回地任取3次，每次1个，并记录颜色；1）写出概率空间；2）求“取到3个球都不同”的概率；

【例 1.3.2】1）设对应于某试验E的样本空间$\Omega=\{0,1,2,\cdots\}$，F是由Ω的一切子集组成的集族，则F是一$\sigma-$代数，$\forall A\in F$，定义：$P(A)=\sum_{k\in A}\frac{\lambda^k}{k!}e^{-\lambda},\lambda>0,P(\Phi)=0$，则$P(\cdot)$是可测空间$(\Omega,F)$上的概率； 2）设$\Omega=R,F=\sigma\left(\left\{(a,b)\middle|+\infty<a<b<+\infty\right\}\right)$，对于每个区间$I$，定义：$P(I)=\int_I\frac{1}{\pi(1+x^2)}dx$，问：$P(\cdot)$是否为$(\Omega,F)$上的概率测度？

【例 1.3.3】某人有n把钥匙，其中仅一把能打开他的锁。他随机地用这些钥匙去开，打不开仍放回去。试问："该人恰好第r次打开这把锁"的概率为多大；试构造一个概率空间，描述这一随机现象。

【注 2】概率空间其实就是随机试验的数学模型；有了概率空间，我们的研究对象就理所当然地由随机试验转向概率空间。

由概率的公理化定义可导出概率的如下性质：

1、 $P(\Phi)=0$ ；　　　　2、可列可加性$\Rightarrow$有限可加性；

由性质 2 可得如下常用公式：（1） $A\in F,P(\overline{A})=1-P(A)$；（2）若$A,B\in F$，则$P(A-B)=P(A)-P(AB)$；特别地，若$A\supset B$，则有$P(A-B)=P(A)-P(B)$（减法公式或可减性）；由此也可得，若$A\supset B$，则$P(A)\ge P(B)$（概率的单调性）；

3 【概率的连续性】

（1）【下连续性】若$A_n\in F,n\ge1,$且$A_n\uparrow$ ，则有

$$\boxed{P\left(\bigcup_{n=1}^{\infty}A_n\right)=P\left(\lim_{n\to\infty}\bigcup_{i=1}^{n}A_i\right)=P\left(\lim_{n\to\infty}A_n\right)=\lim_{n\to\infty}P(A_n)}\ ;$$

【注 3】所谓"连续性"，即是指可以交换极限运算与概率运算的次序。

（2）【上连续性】若$A_n\in F,n\ge1,$且$A_n\downarrow$，则有

$$\boxed{P\left(\bigcap_{n=1}^{\infty}A_n\right)=P\left(\lim_{n\to\infty}\bigcap_{i=1}^{n}A_i\right)=P\left(\lim_{n\to\infty}A_n\right)=\lim_{n\to\infty}P(A_n)};$$

（3）【连续性】若$A_n\in F,n\ge1,$且$\lim_{n\to\infty}A_n$存在，则有

$$\boxed{P\left(\lim_{n\to\infty}A_n\right)=\lim_{n\to\infty}P(A_n)};$$

【注 4】由上述，不难验证：可列可加性$\Leftrightarrow$ 连续性＋ 有限

可加性 ；

4.【加法公式（多除少补原理）—Jordan（若尔当）公式】

设 $A_i\in F,i\geq 1$，则 $P\left(\bigcup_{i=1}^{n}A_i\right)=S_1-S_2+S_3-\cdots+(-1)^{n-1}S_n$，其中：$S_1=\sum_{i=1}^{n}P(A_i)$，$S_2=\sum_{1\leq i\leq j\leq n}P(A_iA_j)$，$S_3=\sum_{1\leq i\leq j\leq k\leq n}P(A_iA_jA_k)$，$\cdots,S_n=P(A_1A_2\cdots A_n)$；特别地，若 $A,B,C\in F$，则有：a） $P(A\cup B)=P(A)+P(B)-P(AB)$ ；

b） $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$ ；

【例 1.3.4】 设 $P(A)=0.4,P(B)=0.5,P(A\cup B)=0.7$， 试求 $P(A-B)$ 和 $P(B-A)$；

【例 1.3.5】 设 $A,B\in F$， $P(A)=0.5,P(B)=0.7$， 试求 $P(AB)$ 和 $P(A\cup B)$ 的最值；

【例 1.3.6】 给定一概率空间 (Ω,F,P)， $A,B\in F$，且 $P(A)=P(B)=1$（几乎必然事件），试求： $P(AB),P\left(A\bar{B}\right),P(A\cup B),P\left(A\cup\bar{B}\right)$；

【例 1.3.7】 设 $A,B,C\in F$， 且 $P(A)=P(B)=P(C)=\frac{1}{4}$， $P(AB)=0$，$P(AC)=P(BC)=\frac{1}{6}$， 试求 “ A,B,C 全不发生” 的概率；

【例 1.3.8】 已知 $P(A\cup B)=0.7$， $P\left(A\cup\bar{B}\right)=0.9$， 试求 $P(A)$；

【例 1.3.9】 试证明概率不等式： $\boxed{|P(AB)-P(A)P(B)|\leq\frac{1}{4}}$；

【例 1.3.10】 设 A,B,C 为三事件，满足 $C\supset AB,\bar{C}\supset\overline{AB}$， 1）试证 $AC=C\bar{B}\cup AB$； 2）若还有 $P(AB)=P(A)P(B)$，试证： $P(AC)\geq P(A)P(C)$；

【例 1.3.11】 1）若 $AB=\overline{AB}$， 且 $P(A)=a$， 则 $P(B)$ 为多少；

2）已知 $P(AB)=P\left(\overline{AB}\right)$，且 $P(A)=a$， 求 $P(B)$ ；

【例 1.3.12】 试证： $\forall A,B\in F$， 有 $P(A\cup B)P(AB)\leq P(A)P(B)$；

【例 1.3.13】 如果事件 $A_1,A_2,\cdots\cdots$ 发生的概率都是 1，试证：

$P\left(\bigcap_{j=1}^{\infty}A_j\right)=P\left(\bigcap_{j=1}^{n}A_j\right)=1$；

【例 1.3.14】 设 $A_1\subset A_2\subset\cdots$， 且 $P(A_j)=\left(1-\frac{1}{j}\right)^j,j=1,2,\cdots$， 试求 $P\left(\bigcup_{j=1}^{\infty}A_j\right)$；

【例 1.3.15】（配对问题）某人先写了n封信，又写了n个信封，他随意地将信装入信封；如果信封和装入其中的信是寄给同一个收信人的，就称为一个“配对”；试求：1）“至少有一个配对”的概率；2）“没有一个配对”的概率；3）“恰好有k个配对”的概率。

§4　条件概率

至此，我们一直是在没有试验其他信息的情况下计算事件的概率的；但一旦知道试验中某个情况已经发生，总是希望在试验中能利用新得到的信息，从而重新去认识随机事件的不确定性。那么，如何去认识利用这些信息呢?

【引例】E：掷一颗骰子，令A={掷出“2”点}，B={掷出偶数点}；易知，$P(A)=\frac{1}{6}$，$P(B)=\frac{1}{2}$；假若已经知道某次试验“已掷出偶数点”，那么此时“掷出‘2’点”的“概率”应为$\frac{1}{3}$；记之为：$P(A|B)$或$P_B(A)=\frac{1}{3}$；它并不同于原先的概率$P(A)=\frac{1}{6}$；原因何在；

有这样一种观点：事件B的发生改变了试验E原先的样本空间$\Omega=\{\omega_1,\omega_2,\cdots\omega_n\}$，（其中$\omega_i$—掷出“$i$”点）或者确切地说缩减了样本空间；由于事件$B$的发生，原样本空间$\Omega$中的可能结果$\omega_1,\omega_3,\omega_5$已被排除在外，此时缩减后的样本空间应该是$\Omega_B=B=\{\omega_2,\omega_4,\omega_6\}$，基于$\Omega_B$的等可能性分析，事件$A$发生的概率应是$\frac{1}{3}$.

以上观点即是关于条件概率的直观解释，以下我们遵循这种直观解释（尽管有些以偏概全），基于古典概型和几何概型来讨论“条件概率的定义”；

（1）　由古典概型入手：设事件A,B和AB分别包含样本空间总数为n的样本点中$n_A,n_B和n_{AB}$个，由条件概率的直观解释，

$$\boxed{P(A|B)=\frac{n_{AB}}{n_B}=\frac{n_{AB}/n}{n_B/n}=\frac{P(AB)}{P(B)}}\ ;$$

（2）　由几何概型入手：考虑试验E：向平面有界区域Ω内随机地投掷一点，令A＝{随机点落入子区域A}，B＝{随机点落入子区域B}，AB＝{随机点落入子区域AB}，分别以

$\mu(\Omega),\mu(A),\mu(B),\mu(AB)$表示区域$\Omega,A,B,AB$的面积；则有：

$P(B)=\dfrac{\mu(B)}{\mu(\Omega)},P(AB)=\dfrac{\mu(AB)}{\mu(\Omega)}$，$\boxed{P(A|B)=\dfrac{\mu(AB)}{\mu(B)}=\dfrac{\mu(AB)/\mu(\Omega)}{\mu(B)/\mu(\Omega)}=\dfrac{P(AB)}{P(B)}}$；

【条件概率的一般定义】设(Ω,F,P)为一概率空间，$A,B\in F$，且$P(B)>0$，在已知事件B发生的条件下，事件A发生的条件概率定义为：$P(A|B)=P(AB)/P(B)$；

【注 1】（1）在初等概率论中，没有测度论的支持，我们无法给出条件概率的严格定义，也无法给出$P(B)=0$时的条件概率$P(A|B)$；

（2）为了区别于$P(A|B)$，常称$P(A)$为事件A的无条件概率；

（3）取$B=\Omega$,则$P(A/\Omega)=\dfrac{P(A\Omega)}{P(\Omega)}=P(A)$，故而无条件概率只是条件概率的一种特殊情形。

由上述定义并不能确切地给出$P(A)$与$P(A|B)$的必然联系；我们可以借助维恩（文氏 Venn ）图 来讨论二者的关系：

(a) $A\cap B=\Phi,P(A/B)=0\le P(A)$；(b) $A\subset B,P(A/B)=P(A)/P(B)\ge P(A)$；

(c) $A\supset B,P(A/B)=1\ge P(A)$；(d) $A\cap B\ne\Phi,P(A/B),P(A)$无法比较；

【注 2】关于条件概率$P(A|B)$可视具体情况运用如下两种方法来计算：

1） 在缩减后的样本空间Ω_B中依条件概率的直观解释计算

2） 在原来的样本空间Ω中直接依定义计算；

【例 1.4.1】若$P(\overline{A})=0.3$，$P(B)=0.4$，$P(A\overline{B})=0.5$，试求$P(B|A\cup\overline{B})$；

【例 1.4.2】现有5件产品，3件正品，2件次品，现进行不放回抽样，每次取一件，令A_i＝{第i次取到次品}，$i=1,2,\cdots,5$，试求$P(A_2|\overline{A_1})$；

【例 1.4.3】甲，乙对同一目标射击一次，其命中率分别为0.6和0.5，已知“目标被命中”，求“它是甲命中”的概率；

【例 1.4.4】已知$0<P(A),P(B)<1$，$P(B|A)=1$，试求$P(\overline{A}|\overline{B})$；

【例 1.4.5】（抽签问题）竹筒中有10支签，其中有1支好签，10人依次抽取1支，取后不放回，令A_i＝{第i人抽得好签}，

$i=1,2,\cdots,10$，试求$P(A_i)$；若好签有2支呢?

【例 1.4.6】以A_t表示“一分子在$(0,t]$内不与其他分子碰撞”，假设“分子在$(0,t]$内不发生碰撞的条件下，在$(t,t+\Delta t]$内发生碰撞”的概率为$\lambda\Delta t+o(\Delta t)$，试求$P(A_t)$；

可以验证：条件概率$P(\cdot|B)$满足如下性质：

（1）【非负性】对每一$A\in F$，有$0\le P(A|B)\le 1$；

（2）【规范性】$P(\Omega|B)=1$；

（3）【可列可加性】若 $\forall n\ge 1, A_n\in F$，且$A_iA_j=\Phi$，则

$$P\left(\bigcup_{i=1}^{\infty}A_i|B\right)=P\left(\sum_{i=1}^{\infty}A_i|B\right)=\sum_{i=1}^{\infty}P(A_i|B);$$

从而，$P(\cdot|B)$为定义于(Ω,F)上的概率测度（函数）；它理应满足概率的所有性质，如：有限可加性，加法公式，单调性，连续性等等。记$P(\cdot|B)=P_B(\cdot)$,即有(Ω,F,P_B)也是一个概率空间，称之为条件概率空间；此即表明：将概率空间(Ω,F,P)限制在事件B上则得到条件概率空间(Ω,F,P_B)；因此，可以用两种不同的观点去认识条件概率：

<u>第一种观点</u>：可测空间(Ω,F)没有变，只是概率由$P(\cdot)$变为$P_B(\cdot)$，它对任何$A\in F$都有定义，并且$P_B(A)=\dfrac{P(AB)}{P(B)}$，概率空间由$(\Omega,F,P)$变为$(\Omega,F,P_B)$；

<u>第二种观点</u>：可测空间(Ω,F)变了，Ω变为$\Omega_B=B$，而F变为$F_B=\{B\cap C:C\in F\}$，概率也由$P(\cdot)$变为$P_B^{/}(\cdot)$，且$P_B^{/}(A)=P(A)/P(B)$，即：概率空间由(Ω,F,P)变为$(\Omega_B,F_B,P_B^{'})$。

【注 3】两种观点其实并无本质不同，前者较简单，它将“样本空间缩小了”这一事实掩盖了，把应当除去的样本点保留了下来；对Ω中的任何A，Ω中的$A\cap B$ 即是Ω_B中的A,故两个不同的概率空间定义的概率 P_B与$P_B^{'}$并无二致。

以下，我们<u>尝试在条件概率空间计算事件的条件概率，其可转化为原概率空间中求事件的概率：</u>

设A为概率空间(Ω,F,P)上的正概率事件，即：$P(A)>0$,若$B\in F$，且$P_A(B)>0$,则对条件概率空间(Ω,F,P_A)而言，$\forall C\in F$，有

$\boxed{P_A(C/B)=\dfrac{P_A(BC)}{P_A(B)}=\dfrac{P(BC/A)}{P(B/A)}=P(C/A\bigcap B)}$ ；

条件概率有三大用途：乘法公式、全概率公式和贝叶斯公式。

【乘法公式】设(Ω,F,P)为一概率空间，若$A_1,A_2,\cdots,A_n\in F$，且$P(\bigcap_{j=1}^{n-1}A_j)>0$，则：$P(\bigcap_{i=1}^{n}A_i)=P(A_1)P(A_2/A_1)P(A_3/A_1A_2)\cdots P(A_n/\bigcap_{i=1}^{n-1}A_i)$；

特别地，若$A,B\in F,P(AB)=P(A)P(B/A)$ ；

【例 1.4.7】袋中有6只白球和4只红球，现任意一一取出，取后不放回，令A_k ={第k次取出红球}，$1\le k\le 10$，试求$P(A_1A_2)$，$P(A_2)$；（本题也可用古典方法计算）

【例 1.4.8】一批产品共100件，其中有次品10件，合格品90件；现从中任取一件，取后不放回，接连取三次，试求“第三次才取到合格品”的概率；（乘法公式，古典方法）

【例 1.4.9】1．利用概率方法证明下列恒等式：设$a,b(a<b)$为任意正整数，则恒有$1+\dfrac{b-a}{b-1}+\dfrac{(b-a)(b-a-1)}{(b-1)(b-2)}+\cdots+\dfrac{(b-a)\times\cdots\times 2\times 1}{(b-1)\cdots(a+1)a}=\dfrac{b}{a}$；

2.试构造概率模型证明恒等式：

$1+\dfrac{N-1}{N}\dfrac{n+1}{n}+\dfrac{(N-n)(N-n-1)}{N^2}\dfrac{n+2}{n}+\cdots+\dfrac{(N-n)(N-n-1)\cdots 2\times 1}{N^{N-n}}\dfrac{N}{n}=\dfrac{N}{n}$。

全概率公式是一个重要且实用的公式，它的思想贯穿了概率论的始终，成为概率论的一个重要特色；其思想很简单，即：分解；意指引入一个适当的划分将复杂事件分解为若干简单事件，通过计算出简单事件的概率而得到复杂事件的概率；这种思想也是我们通常解决复杂问题的一般思路。

【（离散型）全概率公式】

【划分（分割，剖分）】设(Ω,F,P)为概率空间，若$A_i\in F$，$i\ge 1$，且$A_i\cap A_j=\Phi,i\ne j$；并有$\bigcup_{i=1}^{n}A_i=\Omega$；则称$A_1,A_2,\cdots,A_n$为样本空间$\Omega$的一个（有限，穷）划分（分割，剖分），或称{$A_1,A_2,\cdots,A_n$}为样本空间$\Omega$的一个完备事件组。类似，可给出可列情形的定义。

【1】 设(Ω,F,P)为概率空间，$A_1,A_2,\cdots,A_n$ 为样本空间Ω的一个有限划分，且$P(A_i)>0,i=1,2,\cdots,n$；则$\forall B\in F$，有

$$P(B)=P(B\Omega)=P(B\bigcap\{\bigcup_{i=1}^{n}A_i\})=P(\bigcup_{i=1}^{n}BA_i)=\sum_{i=1}^{n}P(BA_i)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$$

【2】 设(Ω,F,P)为概率空间，$A_1,A_2,\cdots,A_n,\cdots$为样本空间的一个可列划分，且 $P(A_i)>0,i=1,2,\cdots$ ；则$\forall B\in F$，有

$$P(B)=P(B\Omega)=P(B\cap\{\bigcup_{i=1}^{\infty}A_i\})=P(\bigcup_{i=1}^{\infty}BA_i)=\sum_{i=1}^{\infty}P(BA_i)=\sum_{i=1}^{\infty}P(A_i)P(B|A_i)$$

【3】 设(Ω,F,P)为概率空间， $B,A_i\in F$ 且$P(A_i)>0,i=1,2,\cdots,n$；当$i\neq j$时， $A_i\cap A_j=\Phi$，并有$B\subset\bigcup_{i=1}^{n}A_i$，则有

$$P(B)=P(B\cap\{\bigcup_{i=1}^{n}A_i\})=P(\bigcup_{i=1}^{n}BA_i)=\sum_{i=1}^{n}P(BA_i)=\sum_{i=1}^{n}P(A_i)P(B|A_i)\quad ;$$

【4】 设(Ω,F,P)为概率空间， $B,A_i\in F$ 且$P(A_i)>0,i=1,2,\cdots,n$；当$i\neq j$时， $A_i\cap A_j=\Phi$，并有$P(\sum_{i=1}^{n}A_i)=1$；则有

$$P(B)=P(B\Omega)=P(\{B\cap\{\bigcup_{i=1}^{n}A_i\}\}\cup\{B\cap\{\overline{\bigcup_{i=1}^{n}A_i}\})=P(\bigcup_{i=1}^{n}BA_i)=\sum_{i=1}^{n}P(BA_i)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$$

【5】（条件概率的全概率公式）设(Ω,F,P)为概率空间，$A_1,A_2,\cdots,A_n$ 为样本空间Ω的一个有限划分，且$P(A_i)>0,$ $i=1,2,\cdots,n$；若$B,D\in F$，且$P(D)>0$，则有 $P_D(B)=P_D(B\Omega)=$

$$P_D(B\cap\{\bigcup_{i=1}^{n}A_i\})=P_D(\bigcup_{i=1}^{n}BA_i)=\sum_{i=1}^{n}P_D(BA_i)=\sum_{i=1}^{n}P_D(A_i)P_D(B|A_i)=\sum_{i=1}^{n}P(A_i|D)P(B|A_iD)$$

以上均可称为离散型全概率公式，我们将在第三章给出连续型（广义）全概率公式。这里我们提前声明－在以下两种情况下常用全概率公式来计算概率：其一是将一个复杂事件分解，其二是一个随机试验分成“两步”，“第一步”结果没有说明，而求的是“第二步”中某事件的概率。

【例 1.4.10】（敏感问题调查）在调查家庭暴力（或婚外恋，吸毒等敏感问题）所占家庭的比例p时,被调查者往往不愿回答真相，这使得调查数据失真。为得到实际的p同时又不侵犯个人隐私，调查人员将袋中放入比例是p_0的红球和比例是$q_0=1-p_0$的白球。被调查者在袋中任取一球窥视后放回，并承诺取得红（白）球就讲真（假）话。被调查者只需在匿名调查表中填“是”（即有家庭暴力）或“否”，然后将调查表放入投票箱，没人知道被调查者是否讲真话和回答的是什么。（1）如果声称有家庭暴力的家庭比例是p_1，求p；（2）如果袋中装有30个红球，50个白球，调查了320个家庭，其中有195

个家庭回答“是”，试估算p；

【例 1.4.11】袋中有5只白球和5只黑球，从中任取5球放入一空袋，再从此5球中取出3球放入另一空袋，求 “从第三只袋中取出白球”的概率；

【例 1.4.12】（末步分析法）

（1） 连续n次掷一枚硬币，第一次掷出正面的概率为a，“第二次以后每次出现与前一次相同面”的概率为b ，求“第n次掷出正面”的概率；

（2）设有n只袋子，每只袋中有a只黑球和b只白球，现从第一只袋中任取一球放入第二只袋中，然后从第二只袋中任取一球放入第三只袋子，依此下去，问：“从第n只袋中任取一球是黑球”的概率；

【例 1.4.13】（首步分析法）

（1）（赌徒破产问题）设某赌徒有赌本$i(i\geq 1)$元，其对手有赌本$a-i$ (≥ 0)元，每赌一次该赌徒以p（$q=1-p$）的概率赢（输）一元，赌博一直进行到两赌徒中有一人破产为止，试求“该赌徒破产”的概率；

（2）（玻利亚『**polya**』概型）罐中有a只黑球和b只白球，每次从中任取一球并连同c个同色球一起放回，如此反复进行，试求“第n次取球时取出黑球”的概率；【$c=0(-1)$时即有（无）放回抽样】

（3）【抽签原理】－在古典型的不放回抽样中，一个事件在任何一次抽取时发生的概率，都与初始状态时该事件包含的样本点数与样本空间所包含的样本点数之比相同：设有n个签，其中有k个好签，n个人依次抽取一个，取后不放回，试求“第$i(1\leq i\leq n)$人抽取好签”的概率；

（4）在n次 **Bernoulli** 试验中，如果某事件A在每次试验中发生的概率都为p，试求“n次试验中A发生偶数次”的概率；

【贝叶斯(**Bayes**)公式】（又称逆概率公式）设(Ω,F,P)为概率空间，$A_1,A_2,\cdots,A_n$为样本空间Ω的一个有限划分，且$P(A_i)>0,i=1,2,\cdots,n$；　　　则$\forall B\in F$，且满足$P(B)>0$，有

$$P(A_j/B)=\frac{P(A_jB)}{P(B)}=\frac{P(A_j)P(B/A_j)}{\sum_{i=1}^{n}P(A_i)P(B/A_i)};$$

关于其他情形的贝叶斯公式，这里不再累述！

【例 1.4.14】（伊索寓言－孩子与狼）试由 Bayes 公式分析大人（村民）对放羊小孩的信任程度是如何下降的；

【例 1.4.15】设有来自三个地区的分别有 10 名，15 名和 25 名考生的报名表，其中女生报名表分别有 3 份，7 份和 5 份；现随机地抽取一个地区的报名表，从中先后抽出两份，

（1） 求“先抽到的是一份女生表”的概率；

（2） 已知后抽到的是一份男生表，求“先抽到的是一份女生表”的概率

【例 1.4.16】在回答有四个选项 A,B,C,D 的选择题时，由于题目较难，全班只有百分之五的学生能解出正确答案；假设能解出答案的学生回答正确的概率是 0.99，不能解出答案的学生随机猜测答案；试求“答题正确的学生是猜对答案”的概率并评价这样出题是否合适；

【本例的注】由该例，一般来讲，难题不应该出成选择题！

【例 1.4.17】（疾病普查问题）艾滋病是人类健康的大敌，发病率还在逐年上升。为了有效防止艾滋病（HIV）流入我国，保护我国公民的健康，1995 年在我国的出入境管理处曾制订了对 HIV 的普查规定：对于在国外生活或工作两个月以上的中国公民回国入境时进行 HIV 的验血检查，但是没实行多久该规定就被叫停了。假设当时符合被检查条件的公民中携带 HIV 的比例是十万分之一，验血检查的准确率是百分之九十五（有病被正确诊断和没病被正确诊断的概率都是百分之九十五）。若甲在检查后被通知带有 HIV，试求“甲的确带有 HIV”的概率；

【本例的注】由该例，可以说明检查出携带 HIV 距离真正携带 HIV 还差得很远，亦可说明这样的普查没有什么实质意义－造成这个结果的原因是发病率较低和诊断的准确性还不够高；正是基于此，可以说对于发病率很低的疾病进行普查意义是不大的，特别是在检查的准确性也不太高的情况下！

【例 1.4.18】（吸烟与肺癌问题）某地区曾对 50－60 岁的男性公民进行调查，肺癌病人中吸烟的比例是 0.997，无肺癌人群中吸烟的比例是 0.958；如果整个人群的发病率是，试求（不）吸烟人群中的肺癌发病率，吸烟人群的肺癌发病率是不吸烟人群的肺癌发病率的多少倍？

【注 4 】有时，我们干脆把贝叶斯公式中的A_i看作是导致事件B发生的各种可能原因或者影响事件B发生的各种可能因素，那么全概率公式提示了我们：事件B发生的概率恰好是事件B在这些"原因"事件发生下的条件概率的加权平均，权重为这些"原因"事件的概率$P(A_i)$；常称这些$P(A_i)$为"先验（验前）概率"；贝叶斯公式告诉我们：倘若在一次试验中，某正概率事件B已经发生，而事件 $A_1,A_2,\cdots$ 都可能导致或影响事件B的发生。"事件B已发生"这个信息自然应该被用来重新估计先验概率，贝叶斯公式正是从数量上刻划了这个变化，有时称$P(A_i/B)$为事件A_i的"后验（验后）概率"。一言以蔽之，如果说全概率公式是由"原因"求"结果"(由因及果)的话，那么贝叶斯公式则是由"结果"求"原因"（由果溯因）。

§5 事件（试验）的独立性

在概率空间(Ω,F,P) 中，$A,B\in F,P(A)与P(A/B)(P_B(A))$常常是有所区别的；从而意识到：事件$B$的发生（已知信息）有可能会改变事件$A$发生的可能性大小。

考虑一种情形：$P(A)=P(A|B)\left[P_B(A)\right]$ **(a)**，此时即有：$P_B(A)=\dfrac{P(AB)}{P(B)}=P(A)$，也即：$P(AB)=P(A)P(B)$ **(b)**；进一步，若有$0<P(B)<1$，则有：$P\left(A\middle|\overline{B}\right)=P_{\overline{B}}(A)=\dfrac{P\left(A\overline{B}\right)}{P\left(\overline{B}\right)}=\dfrac{P(A)-P(AB)}{1-P(B)}=P(A)$ **(c)**；因此，若事件B的发生没有影响A发生的可能性（**(a)**式成立），则事件B不发生也不会影响A发生的可能性（**(c)**式成立）；从而可以说，事件B的发生与否不影响事件A的发生。反之，若又有$P(A)>0$，则由 **(b)** 式易得：$P(B|A)\left[P_A(B)\right]=\dfrac{P(AB)}{P(A)}=P(B)$ **(d)**；若再有条件：$P(A)<1$，则

又有：

$$P\left(B\middle|\overline{A}\right)\left[P_{\overline{A}}(B)\right]=\frac{P\left(\overline{A}B\right)}{P\left(\overline{A}\right)}=\frac{P(B)-P(AB)}{1-P(A)}=P(B)\qquad \textbf{(e)}$$

；因此，若以上所提及的条件若均满足，则有：若事件B的发生与否不影响事件A的发生，则事件A的发生与否不影响事件B的发生；也就是所谓A,B的发生与否相互不影响，以后常称之为（相互）独立！

【事件的独立性 1】设(Ω,F,P)为概率空间，若$A,B\in F$，且满足：$P(AB)=P(A)P(B)$，则称A,B（统计）相互独立（又称两两独立），简称独立；否则称为（统计）相依。

【注 1】（1）在独立性的定义式中采用了**(b)**式，这是由于在该式中，A,B的地位是对等的，更能体现出独立的相互性；（2）**(b)**式的内涵也比其他四式更加丰富；可以验证：Φ,Ω，几乎不可能事件，几乎必然事件与任意事件都是相互（统计）独立的。

【注 2】可以验证：若$\{A,B\},\{\overline{A},B\},\{A,\overline{B}\},\{\overline{A},\overline{B}\}$这四组事件中有一组相互独立，则其他三组必相互独立。

【引例】一只盒中装有a只红球和b只黑球，每次随机地取出一球，记：R_i＝“第次取出红球”，$i=1,2,\cdots,a+b$，问：事件R_1,R_2是否独立？

【引例的注】不同的抽样方式对独立性是有影响的，可以验证：当总体很大时，这种影响又是可以忽略不计的：

$$\left|P(R_1R_2)-P(R_1)P(R_2)\right|=\frac{ab}{(a+b)^2(a+b-1)}\le\frac{1}{2}\frac{1}{a+b-1}\xrightarrow{a+b\to\infty}0$$

,也即：当a或b很大时，R_1与R_2几乎是独立的（不放回抽样）；由此得到启发：不妨以$\left|P(AB)-P(A)P(B)\right|$来刻划事件$A$与$B$的相关程度；为了避免“小概率事件”带来的影响，作“规范化处理”，定义$\boxed{\rho(A,B)=\dfrac{P(AB)-P(A)P(B)}{\sqrt{P(A)P(B)\left[1-P(A)\right]\left[1-P(B)\right]}}}$，称之为事件$A$和$B$的相关系数；可以验证：**1**）$\rho(A,B)=0\Leftrightarrow$事件$A,B$独立；**2**）$\left|\rho(A,B)\right|\le 1$；这在第三章可由柯西不等式得出！

【事件的独立性 2】设(Ω,F,P)为概率空间，$A,B,C\in F$，若以下

条件：1）$P(AB)=P(A)P(B)$；　2）$P(AC)=P(A)P(C)$；　3）$P(BC)=P(B)P(C)$；　4）$P(ABC)=P(A)P(B)P(C)$；则称 A,B,C 相互独立；

【注 3】（1）对于三个或三个以上的事件来说，相互独立必两两独立，但两两独立未必相互独立；（2）可以验证：任选三事件中的 A 作为主体，其与 $B\cup C, B\cap C, B-C$ 等等均（相互，两两）独立。

【事件的独立性 3】设 (Ω, F, P) 为概率空间，$A_1, A_2, \cdots, A_n \in F$，如果对于任意的 $s, 1\le s\le n$，以及任意的 $1\le i_1<i_2<\cdots<i_s\le n$，都有 $P\left(A_{i_1}A_{i_2}\cdots A_{i_s}\right)=P\left(A_{i_1}\right)P\left(A_{i_2}\right)\cdots P\left(A_{i_s}\right)$，则称事件 $A_1, A_2, \cdots, A_n$ 相互独立。

【注 4】(1)该定义中给出的“概率等式”实际上包含 2^n-1-n 个等式；同样，对于多个事件的情形，相互（整体）独立必两两独立，反之则不成立！

(2)已经知道，$\sigma(A)=\{\Phi, A, \bar{A}, \Omega\}$，从而我们有：$A, B$ 独立 $\Leftrightarrow \sigma(A), \sigma(B)$ 独立，即 $\sigma(A)$ 中任一事件与 $\sigma(B)$ 中任一事件相互独立；此结论也可推广到多个事件的情形！

【例 1.5.1】设 A 事件与其自身独立，则 $P(A)=0$ 或 1；

【例 1.5.2】设 A, B 独立，且两事件中“仅 A 发生”和“仅 B 发生”的概率均为 $\frac{1}{4}$，求 $P(A)$ 和 $P(B)$；

【例 1.5.3】试证：若两正概率事件互斥，则必不独立；若两正概率事件独立，则必相容；

【试验的独立性】若试验 E_1 和试验 E_2 的所有可能结果（事件）之间相互独立，则称试验 E_1 与试验 E_2 独立；

　　事实上，要严格地刻划两试验的独立性，必须适当地构造一个能同时描述两个试验的新的概率空间；以下举例！

【引例】连续“独立”地掷两次骰子，试求 A=“第一次掷出‘1’点”和 B=“第二次掷出‘3’点”的概率；

如果孤立地看这两次试验：E_1：掷一次骰子，其样本空间为 $\Omega_1=\{\omega_1, \omega_2, \cdots, \omega_6\}$，　ω_i 一第一次掷出点，$i=1, 2, \cdots, 6$；　取 $F_1=2^{\Omega_1}=\{C: C\subset\Omega_1\}$，且基于等可能性分析，可在 (Ω_1, F_1) 上建立概率测度：$P_1(\{\omega_i\})=\frac{1}{6}, i=1,2,\cdots,6$；因此，$P_1(A)=\frac{1}{6}$；$E_2$：再掷

一次骰子，其样本空间为$\Omega_2=\{\omega_1',\omega_2',\cdots,\omega_6'\}$，$\omega_i'$一第二次掷出点，$i=1,2,\cdots,6$；取$F_2=2^{\Omega_2}=\{C:C\subset\Omega_2\}$，同样基于等可能性分析，可在$(\Omega_2,F_2)$上建立概率测度：$P_2(\{\omega_i'\})=\frac{1}{6},i=1,2,\cdots,6$；因此，$P_2(B)=\frac{1}{6}$；这样，两次试验就相应建立了两个概率空间：$(\Omega_i,F_i,P_i),i=1,2$。

如果将两次试验视作一次试验：E：连续“独立”地掷两次骰子（姑且将E视作两次试验的笛卡尔积，记作：$E=E_1\times E_2$），则其样本空间为：$\Omega=\Omega_1\times\Omega_2=\{(\omega_i,\omega_j'):i,j=1,2,\cdots,6\}$，其样本点$\omega_{ij}=(\omega_i,\omega_j')$一第一次掷出‘$i$’点，第二次掷出‘$j$’点，$i,j=1,2,\cdots,6$；取$F=F_1\times F_2=\{D:D\subset\Omega=\Omega_1\times\Omega_2\}$，基于所谓的独立性（样本点的等可能性），可在$(\Omega,F)$上建立概率测度$P$：$P(\{\omega_{ij}\})=\frac{1}{36},i,j=1,2,\cdots,6$；即得一概率空间$(\Omega,F,P)$；姑且称$P$为乘积概率测度，姑且称$(\Omega,F,P)$为乘积概率空间，分别记之为：$\boxed{P=P_1\times P_{2,}(\Omega,F,P)=(\Omega_1,F_1,P_1)\times(\Omega_2,F_2,P_2)}$；在乘积概率空间考虑$A,B$的概率：

$$P(A\times\Omega_2)=P_1\times P_2(A\times\Omega_2)=\frac{6}{36}=P_1(A)\times P_2(\Omega_2)=P_1(A)\quad,$$

$$P(\Omega_1\times B)=P_1\times P_2(\Omega_1\times B)=\frac{6}{36}=P_1(\Omega_1)\times P_2(B)=P_2(B)\quad;$$

从而，

$$P(A\times B)=P_1\times P_2(A\times B)=P_1\times P_2[(A\times\Omega_2)\cap(\Omega_1\times B)]=P_1\times P_2(A\times\Omega_2)\bullet P_1\times P_2(\Omega_1\times B)$$

$=P_1(A)P_2(B)$；　$A\times\Omega_2=\{\text{第一次掷出“1”点}\}$ 与 $\Omega_1\times B=\{\text{第二次掷出“3”点}\}$关于概率测度$P=P_1\times P_2$独立；这样，就严格描述了源于两个试验的两个事件的独立性！很多时候，我们虽然没有正式地建立乘积概率空间，但是我们都不自觉地在脑海里如此地处理之，以致可以得出完全一致的结论；故而，以后我们就将“错”就“错”，就不再提及并写出所谓的乘积试验，乘积概率，乘积概率空间！（有兴趣者，可参考杨振明编著的《概率论》或严士健等编著的《概率论基础》等书）

【独立重复试验一**Bernoulli** 试验】若试验$E_1,E_2,\cdots,E_n$相互独立，且$E_i=E,\Omega_i=\Omega$，则称$E_1,E_2,\cdots,E_n$为n次独立重复试验；若每

次试验只关心某个事件A发生与否，则称之为“n重 **Bernoulli** 试验”。

设在n重 **Bernoulli** 概型中，事件A在此n次独立重复试验中指定的某k次试验（不妨设前k次）中发生，而在其余的$n-k$次试验中不发生，则其概率为：

$$P\left(A_1A_2\cdots A_k\overline{A_{k+1}}\cdots\overline{A_n}\right)=P\left(A_1\right)P\left(A_2\right)\cdots P\left(A_k\right)P\left(\overline{A_{k+1}}\right)\cdots P\left(\overline{A_n}\right)=p^kq^{n-k},q=1-p,$$

这里，$A_i=\{第i次试验中A事件发生\},i=1,2,\cdots,n$；从而，在$n$重 **Bernoulli** 试验中，“事件$A$恰好发生$k$次”的概率$p$为：

$$P=P\left(\bigcup_{1\le i_1<i_2<\cdots<i_k\le n}A_{i_1}A_{i_2}\cdots A_{i_k}\overline{A_{i_{k+1}}}\cdots\overline{A_{i_n}}\right)=\sum_{1\le i_1<i_2<\cdots<i_k\le n}P\left(A_{i_1}A_{i_2}\cdots A_{i_k}\overline{A_{i_{k+1}}}\cdots\overline{A_{i_n}}\right)=\sum_{1\le i_1<i_2<\cdots<i_k\le n}p^kq^{n-k}$$

$$=\frac{n!}{k!(n-k)!}p^kq^{n-k},k=0,1,\cdots,n;$$

【例 1.5.4】（小概率事件原理）设随机试验E中事件A出现的概率$\varepsilon(>0)$，则无论ε多么地小，当我们不断地重复做试验时，“A迟早会出现”的概率为1；（利用概率的连续性和可加性分别做）

【注 5】1）虽然小概率事件在一次试验中不太可能发生，但在不断重复该试验时，它却必定迟早会发生；正所谓“只要功夫深，铁杵磨成针”，“智者千虑，必有一失”，“多行不义必自毙”…！**2**）在生产和生活的许多实际中，人们常不约（言）而同（喻）地认定独立性满足；即使对于系统中看上去有明显相依性的各个环节，也常常近似地假定它们是独立运行，因为“独立性假设”常常可以大大简化概率的计算！

【例 1.5.5】设有两名射手对同一目标进行射击，甲命中目标的概率为p_1，乙命中目标的概率为p_2，甲先射并规定谁先命中谁获胜，求“甲获胜”的概率；（利用可加性和全概率公式分别做）

【例 1.5.6】（**Poisson** 分布在随机选择下的不变性）设某天到图书馆的人数恰好为k的概率为$\frac{\lambda^k}{k!}e^{-\lambda},\lambda>0,k\in N\cup\{0\}$，每位到图书馆的人借书的概率为$p(0<p<1)$，且借书与否相互独立，试证：“借书的人数恰为$r$”的概率为$\frac{(\lambda p)^r}{r!}e^{-\lambda p},r=0,1,\cdots$。

第二章　离散型随机变量

§1　离散型随机变量及其分布

已经知道，概率论是从数量的角度来研究随机现象的统计规律性的；其通过建立一系列的定理和公式，借以更好地描述处理和解决各种与随机现象有关的理论和应用问题。在处理实际问题时，为全面研究随机试验的结果以揭示随机事件中包含的客观存在的统计规律，常须考虑那些与试验结果联系的某些量；对于那些不表现为数量的结果，也可通过数量处理而得以数量化。这些量是随着试验结果的变化而变化的，因而是“随机的”。这里的量化过程对应着一个法则，其使样本空间的“样本点”与（实）数集的“数”一一对应起来，这个“随机的数”即为随机变量。

【注 1】 利用随机变量（由俄国的彼得堡学派引入）可以表示随机事件（这是随机事件的第三种表示法），可以将不同的随机事件的规律及其相互关系进行统一处理，且有助于研究随机事件的本质规律；它也极大地丰富了概率论的研究对象与范围，使得微积分等分析工具得以引入从而进一步地拓宽了概率论的应用领域（可以形象地说：随机变量架起了概率论与分析学之间的桥梁）。

以下，我们举例说明试验结果的数量化处理过程；

【引例】(1) E_1：掷一枚均匀的硬币；其样本空间为 $\Omega=\{\omega_1,\omega_2\}$，其中：$\omega_1$－掷出正面，$\omega_2$－掷出反面； 指定 1 与 ω_1 对应，指定 0 与 ω_2 对应，则我们建立起 $\Omega=\{\omega_1,\omega_2\}$ 与 $\{0,1\}$ 之间的一一对应关系，不妨记之为：X，即有：　　$X(\omega_1)=1, X(\omega_2)=0$；

(2) E_2：掷一颗骰子；其样本空间为 $\Omega=\{\omega_1,\omega_2,\cdots\cdots,\omega_6\}$，其中：$\omega_i$－掷出“$i$”点，$i=1,2,\cdots,6$；指定 i 与 ω_i 对应，则我们建立起 $\Omega=\{\omega_1,\omega_2,\cdots\cdots,\omega_6\}$ 与 $\{1,2,\cdots\cdots,6\}$ 之间的一一对应关系，不妨记之为：Y，即有：　　$Y(\omega_i)=i, i=1,2,\cdots\cdots,6$；

【随机变量的直观定义】设 E 为某随机试验，其样本空间 Ω 为至多可列集（有穷集或可列无穷集）；若对于样本空间 Ω 的每个样本点 ω，都有唯一的实数 $X(\omega)$ 与之对应，即：$X(\omega)$ 是定

义于Ω上取值于R上的单值实值函数（映射），记之为：$X(\omega):\Omega\to R$，称 $X(\omega)$ 为（实）（**离散型**）**随机变量**（**discrete random varible**），简记为：**d.r.v.** ；

【注 2】 常用希腊字母ξ,η,ζ等或者大写英文（拉丁）字母X,Y,Z等（或加以下标）来表示随机变量，而随机变量的取值则用相应的小写字母（若有的话）。

【注 3】 对于一个随机试验，我们不仅关心出现了哪些结果，更关心这些结果出现的可能性大小；因此，建立 了“样本点”与“实数”之间的对应关系，更须知道随机变量取这些值（解释为何是事件）的概率；

【注 4】 原像及其性质（补充）

【原像】设f是非空集A到B上的一个映射，即：$f:A\to B$，$\forall B_1\subset B$；定义：$\boxed{f^{-1}(B_1)=\left\{f^{-1}(y):y\in B_1\right\}(f^{-1}(y)\text{表示}y\text{在}f\text{下的原像})}$，称之为$B_1$（在$f$下）的原像； 类似，$\forall A_1\subset A$,定义:
$\boxed{f(A_1)=\left\{f\ (x):x\in A_1\right\}(f(x)\text{表示}x\text{在}f\text{下的像})}$;(举例说明)

由试述定义，$f(A)\subset B$，$f^{-1}(B)=A$；讨论上述原像的性质；

原像保持集合运算的所有性质 ：

（1） $\forall B_1,B_2\subset B,B_1\cup B_2\subset B,f^{-1}(B_1\cup B_2)=f^{-1}(B_1)\cup f^{-1}(B_2)$ ；

（2） $\forall B_1,B_2\subset B,B_1\cap B_2\subset B,f^{-1}(B_1\cap B_2)=f^{-1}(B_1)\cap f^{-1}(B_2)$ ；

（3） $\forall B_1,B_2\subset B,B_1-B_2\subset B,f^{-1}(B_1-B_2)=f^{-1}(B_1)-f^{-1}(B_2)$ ；

（4） $\forall B_1\subset B,\overline{B_1}=B-B_1\subset B,f^{-1}(\overline{B_1})=A-f^{-1}(B_1)=\overline{f^{-1}(B_1)}$ ；可以验证：上述的性质（1），（2）可推广到有限（可列）并（交）甚至是不可列并（交）的情形，这里不再累述！

【例 2.1.1】独立地掷两颗骰子，记掷得的点数之和为X,试讨论其可能取值情况及相应的概率；

【例 2.1.2】连续独立地掷一枚硬币，记首次掷出正面（不妨称之为“成功”）时已掷的次数为Y，试讨论其可能取值情况及相应的概率；

【注 5】几何分布的无记忆性－又称无后效性或永远年青性；几何分布具有无记忆性的根本原因在于每次试验中“成功”的概率不随试验次数而改变；常有一类产品在使用过程中没有明显损耗，其寿命就具有这种性质。如：喝水用的玻璃杯

——每喝一次水，对杯子几乎没有一点磨损 ，它往往是由于偶然失手才损坏的，而不是在使用中磨碎的，当使用多次后杯子没有破碎，它仍然如同新的一般被使用，此即“无记忆性”。因此无记忆性的实际意义就是“没有明显的损耗”，故服从几何分布的随机变量常用来表示无明显损耗的一类产品的使用寿命。需注意的是：一个取正整数值的离散型随机变量具有无记忆性的充要条件是它服从几何分布！

【例 2.1.3】将一颗骰子连续地掷n次，记 X（Y）为n个点数的最小（大）值，试求X,Y的“概率分布”；

【例 2.1.4】从$1,2,3,4$中任取一个数X，再从$1,2,\cdots,X$中任取一个数，记之为Y，试求$P(\{Y=2\})$；

【注 6】一般地，一个离散型随机变量（$d.r.v.$）的（概率）分布可表示为：$\begin{pmatrix} X & x_1 & x_2 & \cdots & x_k & \cdots \\ P & p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$， 其中$\{x_1,x_2,\cdots\cdots,x_k,\cdots\cdots\}$为$r.v.X$的取值范围（不妨记之为$R(X)$），而且：

1. $P(\{\omega: X(\omega)=x_k, \omega\in\Omega\}) = P(X^{-1}(\{x_k\})) = P(\{X=x_k\}) = p_k \ge 0$

$k=1,2,\cdots$； **2.** $\sum\limits_k^{?} p_k = P(\bigcup\limits_k \{\omega: X(\omega)=x_k\}) = P(\Omega)=1$ ；事实上，可以验证：凡是满足上述两条件的一组数$\{p_1,p_2,\cdots,p_k,\cdots\}$必可作为某个离散型随机变量的概率分布（又称分布律或分布列）。

【例 2.1.5】（<u>二项分布</u>）设已做了n重贝努利试验（结果：“成功”和“失败”），每次试验中“成功”的概率为p，令Z表示n次试验中“成功”的次数，试求Z的概率分布；

【例 2.1.6】设$X\sim\begin{pmatrix} 0 & 1 \\ 1/4 & 3/4 \end{pmatrix}$，$P\left(\left\{Y=-\dfrac{1}{2}\right\}\right)=1$，$n$维向量$\alpha_1,\alpha_2,\alpha_3$线性无关，试求“向量$\alpha_1+\alpha_2,\alpha_2+2\alpha_3,X\alpha_3+Y\alpha_1$线性相关”的概率；

【注 7】**d.r.v.的概率分布反映了它取各种可能值的概率分配，包含了它的全部概率信息；其描述了 d.r.v.取值的统计规律性**。实际情况是：我们常常不得不同时考虑多个随机变量。一个自然的作法是：将<u>来自同一概率空间</u>的多个随机变量放在一起，作为一个随机取值的向量，即随机向量；其取各种可能的向量值的统计规律性自然就是它取各可能向量值的

概率！

【例 2.1.7】一袋中装有2只白球和3只红球，现不放回地依次抽取2球，设X表示“第一次取到白球的只数”，Y表示“第二次取到白球的只数”，试求(X,Y)的“联合（概率）分布”；

【例 2.1.8】将一枚均匀的硬币连续抛掷次，以X表示3次中出现正面的次数，以Y表示3次中出现正面与反面之差的绝对值，试写出(X,Y)的联合分布及关于X和Y的边缘（边际）分布；

【联合分布】设$d.r.v.X$的$R(X)=\{x_1,x_2,\cdots,x_m\}$，$d.r.v.Y$的$R(Y)=\{y_1,y_2,\cdots,y_n\}$，$(X,Y)$的所有可能取的向量值为$\{(x_i,y_j):i=1,\cdots,m;j=1,\cdots,n\}$，且

$P(\{(X,Y)=(x_i,y_j)\})=P(\{X=x_i,Y=y_j\})=P((X,Y)^{-1}(\{(x_i,y_j)\}))$

$=P(X^{-1}(\{x_i\})\cap Y^{-1}(\{y_j\}))=p_{ij},i=1,\cdots,m;j=1,\cdots,n;$称之为$(X,Y)$的联合分布；通常将其表示为如下的矩阵形式：$\begin{pmatrix} X/Y & y_1 & \cdots & y_n \\ x_1 & p_{11} & \cdots & p_{1n} \\ \vdots & \cdots & \cdots & \cdots \\ x_m & p_{m1} & \cdots & p_{mn} \end{pmatrix}$，

从而 $\boxed{\begin{aligned}&\forall x_i\in R(X),\{X=x_i\}=\{X=x_i\}\cap\Omega\\&=\{X=x_i\}\cap(\bigcup_{j=1}^{n}\{Y=y_j\})\\&=\bigcup_{j=1}^{n}\{X=x_i,Y=y_j\}\end{aligned}}$ $\boxed{\begin{aligned}&\forall y_i\in R(Y),\{Y=y_j\}=\{Y=y_j\}\cap\Omega\\&=\{Y=y_j\}\cap(\bigcup_{i=1}^{m}\{X=x_i\})\\&=\bigcup_{i=1}^{m}\{X=x_i,Y=y_j\}\end{aligned}}$

从而有：$\boxed{p_{i\cdot}=P(\{X=x_i\})=\sum_{j=1}^{n}P(\{X=x_i,Y=y_j\})=\sum_{j=1}^{n}p_{ij}}$，

$\boxed{p_{\cdot j}=P(\{Y=y_j\})=\sum_{i=1}^{m}P(\{X=x_i,Y=y_j\})=\sum_{i=1}^{m}p_{ij}}$；

$\{p_{1\cdot},p_{2\cdot},\cdots,p_{m\cdot}\}$（$\{p_{\cdot 1},p_{\cdot 2},\cdots,p_{\cdot n}\}$），即为$(X,Y)$关于$X(Y)$的边缘（边际）分布，它实际上$X(Y)$就是的概率分布！

【注 8】基于上述，<u>由联合分布求边缘分布</u>，只须将上述联合分布矩阵通过行（列）求和即可；但反之则不可！稍后见相关例题。

【注 9】上述的$m(n)$可推广至无穷的情形,这里不再累述！对

于三维随机向量，则须通过三维矩阵来刻划其联合分布（它有6个边缘分布）；更高维的依此类推。

【例 2.1.9】同例 **2.1.7**，试分别在以下两种情况下求X,Y的联合分布与边缘分布：**1**）第一次取后不放回，再取第二次；**2**）第一次取后放回，再取第二次。

【例 2.1.9 的注】通过该例，可说明以下问题：

（**1**） 两种情况下的联合分布不同，但边缘分布相同；联合分布不同的原因是因为试验的方法不同（无放回抽样、有放回抽样），因而造成了X,Y之间的相互关系不同：无放回抽样时，取值的概率受到取值的影响；而在有放回抽样时则不然；此即说明联合分布确实反映出X与Y之间的某种联系，这种联系是边缘分布所不能体现的；一般地，边缘分布不能决定联合分布，除非加上某种条件，如：独立性等；

（**2**）事实上，无放回抽样下的边缘分布相同是基于“抽签原理”，有放回抽样下的边缘分布相同是基于“试验的独立性”**!**

以下，我们来讨论“随机变量的严格定义”**!**

【随机变量的严格定义】设(Ω,F,P)是一概率空间，对于$\omega\in\Omega,\xi(\omega)$是一个单值实函数；若$\forall x\in R,\{\omega:\xi(\omega)\le x\}=\{\omega:\xi(\omega)\in(-\infty,x]\}=\xi^{-1}((-\infty,x])=\{\xi\le x\}$是一随机事件，也即**:**$\forall x\in R,\{\xi\le x\}\in F$**,**则称$\xi=\xi(\omega)$ 为随机变量（数）。

【注 10】（**1**）定义于Ω或(Ω,F)上的随机变量$\xi(\omega)$，有时也称作可测空间(Ω,F)到(R,β)的可测映射；

（**2**）由定义，随机变量$\xi(\omega)$总联系着一个概率空间(Ω,F,P)；为方便计不必每次都写出概率空间，甚至还常常将$\xi(\omega)$写成ξ，将$\{\omega:\xi(\omega)\le x\}$写成$\{\xi\le x\}$；须强调的是：要求$\forall x\in R,\{\xi\le x\}\in F$，可使$P(\{\xi\le x\})$总有意义！也即：$\{\xi\le x\}$总是“可度量”的随机事件；这里，$\{\xi\le x\}$是指所有满足$\xi(\omega)\le x$的样本点$\omega$的集合

（**3**）同一概率空间可定义不只一个的随机变量，但均须满足定义中的条件；

（**4**）定义于(Ω,F,P)单值实函数是否为随机变量，要视(Ω,F,P)中σ一域F的结构而定，也即F能否包含一切形如$\{\xi\le x\}$的集

合；可以验证：若 $\forall x\in R,\{\xi\le x\}\in F$， 则

（a） $\forall x\in R,\{\xi>x\}=\xi^{-1}((x,\infty))=\xi^{-1}(R-(-\infty,x])=\xi^{-1}(R)-\xi^{-1}((-\infty,x])$ $=\Omega-\{\xi\le x\}=\overline{\{\xi\le x\}}\in F$ ；（b） $\forall x\in R,\{\xi<x\}=\xi^{-1}((-\infty,x))$ $=\xi^{-1}(\bigcup_{n=1}^{\infty}(-\infty,x-\frac{1}{n}])=\bigcup_{n=1}^{\infty}\xi^{-1}\left((-\infty,x-\frac{1}{n}]\right)=\bigcup_{n=1}^{\infty}\left\{\xi\le x-\frac{1}{n}\right\}\in F$；（c） $\forall x\in R,\{\xi=x\}$ $=\xi^{-1}(\{x\})=\xi^{-1}((-\infty,x]-(-\infty,x))=\xi^{-1}((-\infty,x])-\xi^{-1}((-\infty,x))=\{\xi\le x\}-\{\xi<x\}\in F$；（d）类似有： $\forall x\in R,\{\xi\ge x\}=\overline{\{\xi<x\}}\in F$ ；且有， $\forall a,b\in R,a<b$ $\{a<\xi\le b\}=\xi^{-1}((a,b])=\xi^{-1}((-\infty,b]-(-\infty,a])=\xi^{-1}((-\infty,b])-\xi^{-1}((-\infty,a])=\{\xi\le b\}-\{\xi\le a\}\in F$ $\{a\le\xi<b\}=\{\xi<b\}-\{\xi<a\}\in F$ ；$\{a<\xi<b\}=\{\xi<b\}-\{\xi\le a\}\in F$ ； $\{a\le\xi\le b\}=\{\xi\le b\}-\{\xi<a\}\in F$ ；至此，我们知道：若 $\forall x\in R,\{\xi\le x\}\in F$，满足作为 ω 的函数 $X(\omega)$ 在任何类型的区间以及单点集,有限集或可列集（实际上是一切 **Borel** 集）内取值的 Ω 的子集都隶属于 F，从而都有概率可言！

（5）常见的随机变量有两种类型：离散型和非离散型；在非离散型中又有（绝对）连续型和奇异连续型以及它们的混合型；

在下一章我们将致力于讨论连续型！

（6）定义中的条件 $\{\xi\le x\}\in F$ 也可置换成条件 $\{\xi<x\}\in F$，理由不再累述！（课堂论证）

（7）【离散型随机变量的随机运算】

设 X,Y 均为定义于同一概率空间 (Ω,F,P) 的离散型随机变量，且 $R(X)=\{x_1,x_2,\cdots\},R(Y)=\{y_1,y_2,\cdots\}$，定义：

1． $(X+Y)(\omega)=X(\omega)+Y(\omega)$； 2． $(X-Y)(\omega)=X(\omega)-Y(\omega)$ ；

3． $(XY)(\omega)=X(\omega)Y(\omega)$ ； 4． $\left(\frac{X}{Y}\right)(\omega)=\frac{X(\omega)}{Y(\omega)},\omega\in\{\omega|Y(\omega)\ne0\}$；

则 $X\pm Y$，XY，$\frac{X}{Y}$ 均为定义于 Ω 上的函数，并且，$\forall x\in R$ 有，

$\{X+Y\le x\}=\bigcup_{i,j:x_i+y_j\le x}\{X=x_i,Y=y_j\}=\bigcup_{i,j:x_i+y_j\le x}\left[\{X=x_i\}\cap\{Y=y_j\}\right]$（至多可列并）$\in F$；其他情形类似，课堂论证！因此，定义于同一概率空间的（离散型）随机变量的和，差，积，商仍是该概率空间的（离散型）随机变量！事实上，若 $\{X_n,n\ge1\}$ 是此概率空间的随机（变量）列，定义：$\boxed{\left(\sup_{n\ge1}X_n\right)(\omega)=\sup_{n\ge1}X_n(\omega)}$，

$\boxed{\left(\inf_{n\ge1} X_n\right)(\omega)=\inf_{n\ge1} X_n(\omega)}$；则由于$\forall x\in R$，$\left\{\sup_{n\ge1} X_n\le x\right\}=\bigcap_{n\ge1}\{X_n\le x\}\in F$，$\left\{\inf_{n\ge1} X_n> x\right\}=\bigcap_{n\ge1}\{X_n> x\}\in F$；从而有$\sup_{n\ge1} X_n,\inf_{n\ge1} X_n$均是此概率空间上的（离散型）随机变量！依此，该随机序列的上极限$\boxed{\limsup_{n\to\infty} X_n=\inf_{n\ge1}\sup_{k\ge n} X_k}$和下极限$\boxed{\liminf_{n\to\infty} X_n=\sup_{n\ge1}\inf_{k\ge n} X_k}$都是随机变量；特别地，若其极限存在，其也是此概率空间一随机变量！（课堂讨论）

【示性函数】设(Ω,F,P)是一概率空间，$\forall A\in F$,定义A的示性函数为：$I_A(\omega)[\chi_A(\omega)]=\begin{cases}1,\omega\in A;\\0,\omega\notin A;\end{cases}$；易见，定义于$\Omega$上$\omega$的实函数$I_A=I_A(\omega)$确是$(\Omega,F)$上的随机变量：

$\forall x\in R,\{I_A\le x\}=\Phi\in F,x<0;\{I_A\le x\}=\overline{A}\in F,0\le x<1;\ \{I_A\le x\}=\Omega\in F,x\ge1;$；

实际上，可以验证：Ω的子集A为随机事件当且仅当集合A的示性函数为随机变量！

【注 11】引入了示性函数,可使事件的表示，关系与运算转化为数量的表示，关系与运算！这里不加证明地给出如下关系（部分）：

（1）$A\subset B\Leftrightarrow I_A\le I_B,A=B\Leftrightarrow I_A=I_B;$ （2）$I_{A\cup B}=I_A+I_B-I_{A\cap B},I_{A\cap B}=I_AI_B;$

（3）$I_A=1-I_{\overline{A}},A\supset B,I_{A-B}=I_A-I_B;$ （4）$I_{A\cup B}=I_A\vee I_B,I_{A\cap B}=I_A\wedge I_B;$

（这里不再推广！）

【注 12】基于示性函数，可有如下 **d.r.v.的分解与表示**：

给定概率空间(Ω,F,P)，若$A_n\in F,n\ge1$，且$i\ne j,A_iA_j=\Phi,\sum_{n=1}^{\infty}A_n=\Omega$，

易证：$X(\omega)=\sum_{n=1}^{\infty}x_nI_{A_n}(\omega)$为$(\Omega,F)$上一离散型随机变量，且有$X(\omega)=x_n,\omega\in A_n;$也即$X(\omega)$有分布列：$P(\{X=x_n\})=P(A_n),n\ge1;$反之，对于任一离散型随机变量$X$，若其有分布列：$P(\{X=x_n\})=p_n,n\ge1;$则只须记：$A_n=\{\omega:X(\omega)=x_n\}=\{X=x_n\},n\ge1$；从而有：$X(\omega)=\sum_{n=1}^{\infty}x_nI_{A_n}(\omega),,\omega\in\Omega$。总之，任一离散型随机变量总可以分解为互斥事件示性函数的线性组合！用示性函数的线性组合来表示离散型随机变量这种分解的思想与方法在现代概率论中起着巨大的基础作用！

【例 2.1.10】试将二项（分布随机）变量用示性函数表示；

【例 2.1.11】设ξ,η是定义于同一概率空间(Ω,F,P)上的随机变量，试证：（1）$|\xi|$、$\xi\vee\eta=\max(\xi,\eta)$、$\xi\wedge\eta=\min(\xi,\eta)$、$\zeta(\omega)=\xi(\omega)I_A(\omega)+\eta(\omega)I_{\bar{A}}(\omega)[A\in F]$均是该概率空间上的随机变量；（2）$\{\xi=\eta\}=\{\omega|\xi(\omega)=\eta(\omega)\},\{\xi\ge\eta\}=\{\omega|\xi(\omega)\ge\eta(\omega)\}\in F$。

【例 2.1.12】箱中装有次品a_1,a_2与正品b_1,b_2,b_3，现从中一次取出两件产品，1）写出此试验的样本空间；2）设ξ表示所取两件产品中次品的件数，标出ξ在每个样本点上的值；3）写出原像$\{\xi=0\}$、$\{\xi\le 1\}$、$\{\xi\ge 2\}$都由那些样本点组成；4）写出这些事件的概率。

在这一节结束前，我们来重新回顾一下“条件概率”！我们已讨论过在事件B发生的条件下，事件A（发生）的条件概率$P(A|B)$；我们试着拓展这个概念。条件概率$P(A|B)$本质上与随机试验出现的样本点ω有关；这是因为只有当$\omega\in B$时，事件B才发生；所以“在事件B发生的条件下”的条件概率$P(A|B)$应该与$\omega\in B$联系起来：设$B\in F$，且$0<P(B)<1$，则$\forall A\in F$，条件概率$P(A|B)$与$P(A|\bar{B})$均有意义；又设$\vartheta=\{\Phi,\Omega,B,\bar{B}\}=\sigma(B)$，即由$B$生成的$\sigma$域，定义——

$$P(A|\vartheta)=\begin{cases}\dfrac{P(AB)}{P(B)},\omega\in B;\\[2mm]\dfrac{P(A\bar{B})}{P(\bar{B})},\omega\in\bar{B};\end{cases}$$

，则$P(A|\vartheta)$是个随机变量；因为它在B与$\bar{B}$上取常数值，所以$P(A|\vartheta)$是关于ϑ可测的随机变量，称$P(A|\vartheta)$是事件A关于σ域ϑ的条件概率。最后需要说明的是，关于σ域的条件概率（条件期望）【随机变量】在现代金融理论中有着重要的运用，有关它的进一步介绍可见于高等概率论、随机过程和随机分析的教材与讲义中，读者亦可参考作者所写的《2011 应用随机过程讲义》（可在“新浪共享”中搜索）。

§2 离散型随机变量的数字特征

已经知道，离散型随机变量的概率分布完全刻划了随机变量取值的概率（统计）规律，但是常常无须对其描述得那

么精细（有时是不能）；从而我们希望用一个或多个实数来描述随机变量取值的某种（些）特征。

所谓随机变量的数字特征，即指联系于它的概率分布的某个（些）数，如：平均值，最大值等，其反映了随机变量某个（些）方面的特征。数学期望（**expectation** or **expected value**）（又称均值[**mean**]）即是这样一个数，它实际是加权平均概念的推广。

【引例】已知某班的数学成绩（按5分制）统计如下：

$\begin{pmatrix} \text{分数} & 1 & 2 & 3 & 4 & 5 \\ \text{人数} & 0 & 5 & 8 & 9 & 9 \end{pmatrix}$；易见该班平均数学成绩为：

$\boxed{\bar{x}=1\times\frac{0}{31}+2\times\frac{5}{31}+3\times\frac{8}{31}+4\times\frac{9}{31}+5\times\frac{9}{31}}$;换个角度考虑此问题:现从该班随机地选出一名学生,记其数学成绩为 X ,则其有如下的概率分布; $\begin{pmatrix} X & 1 & 2 & 3 & 4 & 5 \\ P & \frac{0}{31} & \frac{5}{31} & \frac{8}{31} & \frac{9}{31} & \frac{9}{31} \end{pmatrix}$;从而可以认为随机变量 X 的平均值即是它的所有可能取值的加权平均,权重为取相应值的概率;以后我们常常称之为“概率平均值”。

【数学期望】设 X 为概率空间 (Ω,F,P) 上的一个离散型随机变量，其概率分布为：$P(\{X=x_n\})=p_n,n=1,2,\cdots\cdots$; ；若级数 $\sum_{n\ge1}x_np_n$ 绝对收敛，即：$\sum_{n\ge1}|x_n|p_n$ 收敛，则随机变量 X 的(数学)期望定义为 $\sum_{n\ge1}x_np_n$，并记之为 $E(X(\omega))$，简记为 $E(X)$ 或 EX 。

【例 2.2.1】设 $d.r.v.X$ 有分布列：$P\left(\left\{X=(-1)^{j+1}\frac{3^j}{j}\right\}\right)=\frac{2}{3^j},j=1,2,\cdots$，试求 EX ;

【例 2.2.2】设 $A\in F,I_A$ 为 A 的示性函数，试求 EX ;

【注 1】（1）定义中要求级数绝对收敛，是为了保证求和不受求和次序的影响；（2）一个随机变量的数学期望（若存在）已不再是随机变量，而是一个确定的实数－随机因素已被加权平均掉了；另外，若 X 为仅取非负整值的随机变量，则有：

$$EX=\sum_{n\ge1}nP(\{X=n\})=\sum_{n\ge1}\sum_{k=1}^{n}P(\{X=n\})\underline{\underline{\text{交换求和次序}}}\sum_{k\ge1}\sum_{n\ge k}P(\{X=n\})$$

$$=\sum_{k\ge1}P(\{X\ge k\})=\sum_{n\ge1}P(\{X\ge n\})$$;

【例 2.2.3】若 $r.v.X\sim B(n,p)$， $r.v.Y\sim Geo(p)$，试求 EX 和 EY；

【例 2.2.4】设 $d.r.v.X$ 有分布列： $\begin{pmatrix} X & -1 & 0 & 1 & 2 \\ P & \frac{1}{4} & \frac{1}{6} & \frac{1}{3} & \frac{1}{4} \end{pmatrix}$； 试求 $Y=X^2,Z=X+1$ 的数学期望 EY 和 EZ；

【注 2】设 X 为概率空间 $(\Omega,\ \mathcal{F},P)$ 上的一个离散型随机变量，$g(\cdot)$ 为一元 $Borel$ 函数，令 $Y=g(X)$，则有 Y 的数学期望（若存在）为：

$$EY=\sum_{j:y_j\in R(Y)}y_jP(\{Y=y_j\})=\sum_{j:y_j\in R(Y)}y_jP\left(\bigcup_{i:g(x_i)=y_j}\{X=x_i\}\right)$$

$$=\sum_{j:y_j\in R(Y)}\sum_{i:g(x_i)=y_j}g(x_i)P(\{X=x_i\})=\sum_{i:x_i\in R(X)}g(x_i)P(\{X=x_i\})=E[g(X)]$$ ；

【例 2.2.5】设 $d.r.v.X$ 有分布列： $P(\{X=k\})=\frac{1}{2^{k+1}},k=0,1,\cdots$ ， 求 $Y=\cos(\pi X)$ 的数学期望；（两种方法）

【例 2.2.6】假设一部机器在一天内发生故障的概率为 0.2，机器发生故障时全天停止工作。若一周 5 个工作日（无故障、发生 1 次故障、发生 2 次故障）发生 3 次或 3 次以上故障，则获得利润（10 万、5 万、0 万）-2 万元，求机器一周产生的平均利润；

为简化数学期望的计算与讨论，以下考虑其性质；设 X 为 $(\Omega,\mathcal{F},P)$ 上一随机变量，可以验证：

（1） $|E(X)|\le E(|X|)$ （若存在）；

（2）若 $\exists a,b\in R$, 使得 $a\le X\le b$ ， 则 $a\le EX\le b$ ；

【线性性质】设 $X_1,X_2,\cdots,X_n$ 是 n 个存在期望的随机变量，$c_0,c_1,\cdots,c_n$ 是 $n+1$ 个实数，则

$$E(c_0+c_1X_1+\cdots+c_nX_n)=c_0+c_1EX_1+\cdots+c_nEX_n。$$

（<u>可以二维为例证明</u>）由该性质，自然有：

（a）常数作为一退化的随机变量其期望即其本身；

（b）对于多维随机向量，即便不知其联合分布，仅知其边缘分布，亦可通过该性质求出其和的数学期望；

【例 2.2.7】将 n 只球独立地放入 M 只盒子中，若每只球放入各只盒子是等可能的，问：平均有多少只盒子有球？

【例 2.2.8】求连续独立地掷100颗骰子所得点数之和的数学期望；

【例 2.2.9】某仪器由A,B,C三个元件组成，A,B,C发生故障的概率分别为0.1，0.2和0.3，问：某时刻平均有多少元件正常工作？

【例 2.2.10】设$r.v.X\sim B(n,p)$，试由"随机变量的分解法"求EX；

【例 2.2.11】一袋中装有60只黑球和40只红球，现从中任取20只，则平均取到多少只红球？

已经知道离散型随机变量的数学期望定义为它的概率加权平均值，而以此作为随机变量的某种特征这种作法是否"最为合适"，并且又该如何理解这里的"最为合适"？

设X是一个随机变量，且EX存在记之为μ；对于任一实数a，$X-a$表示X与a的离差（误差），$(X-a)^2$表示平方离差（误差），$E(X-a)^2$表示平均平方离差（误差）；$E(X-a)^2=E(X-\mu+\mu-a)^2=E(X-\mu)^2+(\mu-a)^2\geq E(X-\mu)^2$；等号成立当且仅当$\mu=a$，也即：$\boxed{\min_{a\in R}E(X-a)^2=E(X-\mu)^2;}$这就意味着：若用一个"实数"去代表一个"随机变量"，它的数学期望（若存在）即是使得平均平方误差最小的那个实数；而此时的平均平方误差$E(X-\mu)^2$也恰好刻划了随机变量在其均值附近取值的分散程度，以后记之为：DX（或$D(X)$、$VarX$），称之为方差。

【注 3】(1) 可以验证：随机变量X的方差为零当且仅当X以概率 1 取值为某个常数（数学期望）；也即：$D(X)=0\Leftrightarrow P(\{X=EX\})=1$，或$X=EX,a.s.$；这是由于：$\overline{\{X=EX\}}=\{X\neq EX\}=\{|X-EX|>0\}=\bigcup_{n=1}^{\infty}\left\{|X-EX|\geq\frac{1}{n}\right\}$；再由概率函数的连续性以及切比雪夫不等式即可得证！

(2) 若EX^2存在，则DX存在，且 $DX=EX^2-(EX)^2$；(由期望的线性性质即可得)

(3) 若DX存在，则有：$D(X+c)=DX,D(aX+b)=D(aX)=a^2DX$；

【例 2.2.12】若$r.v.X\sim B(n,p)$，$r.v.Y\sim Geo(p)$，试求 DX和DY；

【例 2.2.13】袋中有n张卡片，分别标有号码$1,2,\cdots,n$，现从中

不放回抽出k张卡片，设ξ表示所抽出的号码之和，求$E\xi, D\xi$。

§3 离散型随机变量的条件分布、独立性

条件分布是研究随机变量间“相依关系”的一个有力工具，其广泛被应用于概率统计的各个分支。

设离散型随机变量X和Y的可能取值范围分别为$\{x_1, x_2, \cdots, x_m, \cdots\}$和$\{y_1, y_2, \cdots, y_n, \cdots\}$，且已知二维离散型随机向量$(X,Y)$的联合分布为：$P\left(\{(X,Y)=(x_i,y_j)\}\right)=p_{ij}, i=1,\cdots,m,\cdots; j=1,\cdots,n,\cdots;$ 其边缘分布为：$p_{i.}=P\left(\{X=x_i\}\right)=\sum_{j\ge1} p_{ij}, i=1,\cdots,m,\cdots;$

$p_{.j}=P\left(\{Y=y_j\}\right)=\sum_{i\ge1} p_{ij}, j=1,\cdots,n,\cdots;$；$\forall y_j\in R(Y)$，且$P\left(\{Y=y_j\}\right)>0$，在事件$\{Y=y_j\}$发生的条件下，考虑$X$的概率分布；也即：条件概率$P\left(X=x_i \middle| Y=y_j\right), i=1,\cdots,m,\cdots;$（为方便计，以后不再写为：$P\left(\{X=x_i\}\middle|\{Y=y_j\}\right)$）若令$A_j=\{Y=y_j\}, B_i=\{X=x_i\}, i\ge1;$

则：$P\left(X=x_i \middle| Y=y_j\right)=P\left(B_i \middle| A_j\right)=\dfrac{P\left(A_jB_i\right)}{P\left(A_j\right)}=\dfrac{p_{ij}}{p_{.j}}, i=1,\cdots,m,\cdots;$因此，对于固定的$j$，$\left\{P\left(X=x_i \middle| Y=y_j\right), i=1,\cdots,m,\cdots\right\}$给出了一个分布列，记之为：$\left\{P_{X|Y}\left(x_i \middle| y_j\right), i=1,\cdots,m,\cdots\right\}$ ；

【条件分布】设(X,Y)是一个离散型随机向量，若$P\left(\{Y=y_j\}\right)>0$，则对于固定的$j$，$P_{X|Y}\left(x_i \middle| y_j\right)=P\left(X=x_i \middle| Y=y_j\right)=\dfrac{p_{ij}}{p_{.j}}, i=1,\cdots,m,\cdots$；称之为$X$在给定 $Y=y_j$时（或$\{Y=y_j\}$发生的情况下）的条件分布（列）；若$P\left(\{X=x_i\}\right)>0$，则对固定的$i$，$P_{Y|X}\left(y_j \middle| x_i\right)=P\left(Y=y_j \middle| X=y_i\right)=\dfrac{p_{ij}}{p_{i.}}$， $j=1,\cdots,n,\cdots$；称之为Y在给定$X=x_i$时（或$\{X=x_i\}$发生的情况下）的<u>条件分布（列）</u>；常将以上的条件分布列表示为：

$X\mid_{Y=y_j}$	x_1	x_2	……
P	$P_{X\|Y}\left(x_1 \middle\| y_j\right)$	$P_{X\|Y}\left(x_2 \middle\| y_j\right)$	……

和

$Y\mid_{X=x_i}$	y_1	y_2	……
P	$P_{Y\|X}\left(y_1 \middle\| x_i\right)$	$P_{Y\|X}\left(y_2 \middle\| x_i\right)$	……

【例 2.3.1】设(X,Y)有如下的联合分布：

$$\begin{pmatrix} (X,Y) & (1,1) & (1,4) & (2,1) & (2,2) & (2,3) & (3,2) \\ P & \frac{2}{27} & \frac{1}{27} & \frac{6}{27} & \frac{6}{27} & \frac{6}{27} & \frac{6}{27} \end{pmatrix}$$；试求：给定

$X(Y)=1$时，$Y(X)$的条件分布列；

【例 2.3.2】假设有一枚并不均匀的硬币，出现正面的概率为p，将其独立地掷n次，记Y为掷出正面的次数，首次掷出正面是在第X次，求在n次抛掷中仅掷出一次正面的条件下，X的条件概率分布；

【例 2.3.3】若$r.v.X \sim Geo(p)$，则有：

（1）$P(X>m+n|X>n)=P(\{X>m\}),m,n=0,1,\cdots;$

（2）$P(X=m+n|X>n)=P(\{X=m\}),m,n=1,2,\cdots;$（1），（2）即几何分布的本征性质：无记忆性；也即，给定$X>n$时几何随机变量的条件分布等价于无条件分布！

【注 1】分析几何分布的无记忆性的缘由：几何分布的概率背景是独立重复试验，这是“不学习，不总结经验”的一系列试验，故对已进行过的试验无所记忆；正所谓只有深刻认识和领会“失败是成功之母”，方才“功夫不负有心人”。

有关二维随机向量的条件分布可推广到三维及三维以上的情形，这里不再累述，详见“随机数学引论”（林元烈编著）；以下，我们将利用随机事件的独立性来定义随机变量的独立性；

【随机变量的独立性】 称离散型随机变量X与Y是相互独立的，若$\forall x_i \in R(X)$和$\forall y_j \in R(Y)$，随机事件$\{X=x_i\}$与$\{Y=y_j\}$相互独立，即：$\forall x_i \in R(X)和\forall y_j \in R(Y), P(\{X=x_i,Y=y_j\})=P(\{X=x_i\})P(\{Y=y_j\})$

【注 2】多个离散型随机变量的独立性可类似定义；一般地，由一个随机向量的联合分布可唯一地求出相应的边缘分布；但仅由边缘分布的有关信息还不足以导出联合分布，若加上“独立性”的假设，则可唯一地确定联合分布！

【例 2.3.4】设(X,Y)的联合分布为：$\begin{pmatrix} X/Y & 1 & 2 & 3 \\ 1 & 1/6 & 1/9 & 1/18 \\ 2 & 1/3 & \alpha & \beta \end{pmatrix}$；且$X,Y$独立，试求$\alpha$与$\beta$；

【例 2.3.5】设随机变量X_1,X_2,X_3,X_4独立同分布，且有分布列$\begin{pmatrix} 0 & 1 \\ 0.6 & 0.4 \end{pmatrix}$，试求行列式$D=\begin{vmatrix} X_1 & X_2 \\ X_3 & X_4 \end{vmatrix}$的概率分布；

【例 2.3.6】(分布的再生性『可加性』)设 $X\sim B(n,p)$，$Y\sim B(m,p)$，且 X,Y 独立，试求 $X+Y$ 的分布；

【同分布与独立同分布】若离散型随机变量 X 与 Y 值域满足 $R(X)=R(Y)=\{a_1,a_2,\cdots\cdots\}$，且 $\forall a_i\in\{a_1,a_2,\cdots\cdots\}$，$P(\{X=a_i\})=P(\{Y=a_i\})$，则称 X 与 Y 同分布；若 $\{X_n,n\geq 1\}$ 是一个独立的随机变量序列，且 $X_n,n\geq 1$ 同分布，则称 $\{X_n,n\geq 1\}$ 是一个独立同分布随机序列，简记为：$X_n,n\geq 1$,**i.i.d.(independent identically distributed)**;

【注 3】在概率论中，对于同分布的随机变量常常不加以区别，因为它们具有“完全相同的概率性质”；同时，我们也有：$\boxed{X=Y,a.s.\Rightarrow X\underline{\underline{d}}Y}$ ；并且，对于独立的随机变量，我们有：

（1）若随机变量 $X_1,X_2,\cdots\cdots,X_n$ 相互独立，则有

$$E(X_1X_2\cdots X_n)=EX_1EX_2\cdots EX_n;$$

（2）若随机变量 $X_1,X_2,\cdots\cdots,X_n$ 相互独立，则 $\forall a_i\in R,i=1,2,\cdots,n;$

$$D\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_i^2D(X_i);$$

§4 条件数学期望

【引例】将一颗骰子连续地抛掷两次，易知第一次掷出的点数 X 的数学期望为 3.5；但若已知两次掷出的点数之和 $Y=5$，那么在此条件下第一次掷出的点数 X 的期望还会是 3.5 吗？易见，$\{Y=5\}=\{(\omega_1,\omega_4'),(\omega_2,\omega_3'),(\omega_3,\omega_2'),(\omega_1,\omega_1')\}$ ，这里，ω_i－第一次掷出“i”点，ω_i' 一第二次掷出“i”点；基于等可能性分析，以上四个样本点出现的概率均为 $\frac{1}{4}$；从而给定 $Y=5$ 时，X 的条件分布为：$P_{X|Y}(i|5)=P(X=i|Y=5)=\frac{1}{4},i=1,2,3,4;$ 即有：在给定 $Y=5$ 时，X 的（条件分布的）数学期望为：$\frac{1}{4}\times(1+2+3+4)=\frac{5}{2}$，这个数学期望也正是 $r.v.X$ 在给定 $Y=5$ 的条件下的条件分布 $\{P_{X|Y}(i|5),i=1,2,3,4\}$ 的数学期望。

【条件数学期望】设二维离散型随机变（向）量 (X,Y) 的所有可能取“值”为：$\{(x_i,y_j),i=1,2,\cdots;j=1,2,\cdots;\}$，对于一切满足 $P(\{Y=y\})>0$ 的 y，$y\in R(Y)$；则 X 在给定 $Y=y$ 时（或 $\{Y=y\}$ 发生的条件下）的条件分布 $\{P_{X|Y}(x_i|y),i=1,2,\cdots\}$ 也是一个概率分布。

从而，$r.v.X$ 在给定 $Y=y$ 时的"数学期望"即为给定 $Y=y$ 时，$r.v.X$ 的条件数学期望（若存在）；也即条件分布 $\{P_{X|Y}(x_i|y), i=1,2,\cdots\}$ 的数学期望，记作：$E(X|Y=y), y\in R(Y)$; 也即：

$$\boxed{E(X|Y=y)=\sum_i x_i P_{X|Y}(x_i|y)=\sum_i x_i P(X=x_i|Y=y)}$$

【例 2.4.1】已知 (X,Y) 的联合分布为：

$$\begin{pmatrix} (X,Y) & (1,1) & (1,4) & (2,1) & (2,2) & (2,3) & (3,2) \\ P & \frac{2}{27} & \frac{1}{27} & \frac{6}{27} & \frac{6}{27} & \frac{6}{27} & \frac{6}{27} \end{pmatrix}$$

，试求：$E(X|Y=1)$ 和 $E(Y|X=1)$；

【注 1】（1）以上定义给出的只是狭义的"条件数学期望"，这在初等概率论中已经足够了；

（2）比较（无条件）数学期望与条件数学期望，有如下异同：$\boxed{EX=\sum_i x_i P(\{X=x_i\})=\sum_i x_i P(\{\omega: X(\omega)=x_i\})}$，它是对所有的 $\omega\in\Omega$，$X(\omega)$ 取值全体的加权平均（$\boxed{EX=\sum_{\omega\in\Omega} X(\omega)P(\{\omega\})}$）；

$$\boxed{E(X|Y=y)=\sum_i x_i P_{X|Y}(x_i|y)=\sum_i x_i P(\{\omega: X(\omega)=x_i\}|\{\omega: Y(\omega)=y\})}$$

，它是对 $\omega\in\{\omega: Y(\omega)=y\}$ 时 $X(\omega)$ 的局部加权平均。

如果对于任意的 $y\in R(Y)$，$E(X|Y=y)$ 均存在，即：$y\in R(Y), y\to E(X|Y=y)$，从而可将 $E(X|Y=y)$ 看作是定义于 $R(Y)$ 上的函数，以后不妨记之为 $g(y)$，即有：$g(y)=E(X|Y=y), y\in R(Y)$；

【注 2】这里的 $g(y)$ 不是普通的函数，这是由于 y 是随机地在 $R(Y)$ 内取值，也即 $E(X|Y=y)$ 依赖于 $\{Y=y\}$，即依赖于 $\omega\in\{\omega: Y(\omega)=y\}$。故从全局看，有必要引入一个新的随机变量 $g(Y)=g[Y(\omega)]=E(X|Y)(\omega)$，使它在"$Y=y$"时，即 $\omega\in\{\omega: Y(\omega)=y\}$ 时取值为 $E(X|Y=y)$，$y\in R(Y)$；称 $E(X|Y)$ 为给定 Y 时 X 的条件数学期望。

【注 3】$E(X|Y=y)$（$y\in R(Y)$）是一个数（函数），而 $E(X|Y)$ 是一个随机变量！

【（重）全（数学）期望公式】若 (X,Y) 有如下的联合分布：

$P\left(\left\{\left(X,Y\right)=\left(x_i,y_j\right)\right\}\right)=p_{ij},i=1,\cdots,m,\cdots;j=1,\cdots,n,\cdots;$ 且由于 $E\left(X|Y\right)$ 是一个随机变量，故可取期望（若存在）；

$$E\left[E\left(X|Y\right)\right]=E\left[g\left(Y\right)\right]=\sum_j g\left(y_j\right)P\left(\left\{Y=y_j\right\}\right)=\sum_j E\left(X|y=y_j\right)P\left(\left\{Y=y_j\right\}\right)$$

$$=\sum_j\sum_i x_i P_{X|Y}\left(x_i|y_j\right)P\left(\left\{Y=y_j\right\}\right)=\sum_j\sum_i x_i P\left(\left\{X=x_i,Y=y_j\right\}\right)$$

$$=\sum_i\left(x_i\sum_j P\left(\left\{X=x_i,Y=y_j\right\}\right)\right)=\sum_i x_i P\left(\left\{X=x_i\right\}\right)=EX$$

；也即：$\boxed{E\left[E\left(X|Y\right)\right]=EX}$，此即：全（数学）期望公式！该公式体现的是两次加权平均达一次加权平均的效果（举例说明），这种思想对应着第一章全概率公式的思想；二者是否有所联系？

【注 4】 全概率公式是全期望公式的特殊情形，而后者是前者的推广！在上述全期望公式中，令 $X=I_B,B\in F;\ A_i=\left\{Y=y_i\right\}$，$y_i\in R(Y)$；即有 $\left\{A_1,A_2,\cdots\right\}$ 是一个完备事件组；从而有，

$$P\left(B\right)=E\left(I_B\right)=E\left[E\left(I_B|Y\right)\right]=E[g(Y)]=\sum_i g\left(y_i\right)P\left(\left\{Y=y_i\right\}\right)=\sum_i E\left(I_B|A_i\right)P\left(A_i\right)$$

$$=\sum_i P\left(B|A_i\right)P\left(A_i\right)$$

！

【注 5】事实上，作为离散型随机变量 Y 的**(Borel)**函数 $g\left(Y\right)$ 的 $E\left(X|Y\right)$，其也是离散型随机变量，且有如下的示性函数表示式：$E\left(X|Y\right)=E\left(X|Y\right)(\omega)=\sum_i E\left(X|Y=y_i\right)I_{\{Y=y_i\}}(\omega)$；　直观上看，随机变量 $E\left(X|Y\right)$ 是 ω 的函数；当 $\omega\in\left\{\omega:Y\left(\omega\right)=y\right\}$ 时，$E\left(X|Y\right)$ 的取值为 $E\left(X|Y=y\right)$；因而，条件数学期望更深入地描述了 X,Y 之间的某种关系！

【注 6】设 $B_1,B_2,\cdots$ 是样本空间 Ω 的一个划分，且 $P\left(B_j\right)>0,j=1,2,\cdots$，则有：$\boxed{EX=\sum_j E\left(X|B_j\right)P\left(B_j\right)}$；亦有教材将此公式称为"全期望公式"，而将上述全期望公式称为"重期望公式"！

【例 2.4.2】从编号为 $1,2,\cdots,n$ 的 n 张卡片中随机地取一张，若抽出的卡片号码为 k，则第二张卡片从编号为 $1,2,\cdots,k$ 的 k 张卡片中抽取，记 X 为抽出的第二张卡片号码，求 EX；

【例 2.4.3】一个工人照看 n 台同一类型的机床，n 台机床排

成一行，相邻两台间距为a，试求工人从已照看的机床到下一台待照看的机床行走距离的数学期望；

【例 2.4.4】从一只装有a只白球和b只黑球的袋中任取n只球（$1 \le n \le a+b$）放入另一只袋中，然后从该袋中任取一球，问："如此方式取得白球"与"直接从袋中取出白球"的概率是否相同？

【例 2.4.5】（巴格达窃贼问题）一窃贼被关在有3个门的地牢中，其中有1个门通向自由，从这个门走3个小时便回到地面；第2个门通向一个地道，在此地道中走5个小时又回到地牢；第3个门通向一个更长的地道，在此地道中走7个小时又回到地牢。若窃贼每次选择这3个门的概率相等，求他为获得自由而奔走的平均时间；

【例 2.4.6】（营业额问题）设某商店从中午12点到下午3点到达的顾客数η服从200人/小时的$Poisson$分布，到达商店的顾客所发的钱数$\xi_i \sim B(100, 0.15)$，且诸ξ_i相互独立，与η也独立，求该段时间商店的平均营业额－随机和的 **Wald** 公式；

【例 2.4.7】袋中有编号为$1,2,\cdots,n$的n个球，现从中任取一球；如取到1号球，则得1分，且停止取球；若取到$i(i \ge 2)$号球，则得i分，且将此球放回，再继续取球；依此进行下去，设ξ表示得到的总分数，试求$E\xi$；

【例 2.4.8】现进行 **Bernoulli** 试验，且每次试验中成功的概率为p；假设试验一直进行到接连两次成功出现为止，ξ表示试验停止时已进行的试验次数，试求$E\xi$；

【例 2.4.9】设(X,Y)的联合分布为：$\begin{pmatrix} X/Y & 1 & 2 & 3 \\ 1 & \frac{2}{27} & \frac{5}{27} & \frac{1}{27} \\ 2 & \frac{4}{27} & \frac{7}{27} & \frac{2}{27} \\ 3 & \frac{1}{27} & \frac{3}{27} & \frac{2}{27} \end{pmatrix}$；试求$E(X|Y)$的分布列，并求$EX$；

【例 2.4.10】医生手中有m种不同的退烧药，看病时为发烧的病人以概率p_i选用第i种药；如果服用第i种药的病人在第j

小时恢复体温的概率为$\frac{\mu_i^j}{j!}e^{-\mu_i}$， $j=0,1,\cdots$，**(a)**计算服用药物i的病人恢复体温平均需要的时间；**(b)**计算下一位病人恢复体温平均需要的时间；

【例 **2.4.11**】甲有8万元可以投资两个项目；项目A需要投资至少5万，成功的概率是0.8，失败的概率是0.2，成功后收回本金并获利0.5，失败将损失2万；项目B需要投资至少6万，成功的概率是0.6，失败的概率是0.4，成功后收回本金并获利0.7，失败将损失3万。假设甲总是将手中的资金全部用于投资，且只能对各项目投资一次；**(a)**先投资项目A，然后投资项目B时，平均收益是多少，方差是多少？**(b)**先投资项目B，然后投资项目A时，平均收益是多少，方差是多少？

【该例的注】平均收益越大越值得投资；方差刻划的是投资风险，这是因为方差越大，投资收益就越不稳定！

对于更一般的情形－关于子σ域的条件期望，其是随机过程与鞅论的基础，也是金融（数学）工程特别是金融衍生品定价理论的基础；对于其的讨论仅限于高等概率论、随机过程和随机分析等课程，这里不再累述！

第三章 连续型随机变量

§1 连续型随机变量及其分布

【引例】向$[0,1]$区间随机地掷一点，由“等可能性”可知点落入其任一子区间的概率等于其长度（几何概型）；不妨设点的位置（坐标）为X，则X是一个“随机变量”，$R(X)=[0,1]$，$\forall a,b\in[0,1],a<b,P\left(\left\{\omega:a\le X(\omega)\le b\right\}\right)=P\left(\left\{a\le X\le b\right\}\right)=b-a=\int_a^b I_{[0,1]}(x)dx$；如何去刻划$r.v.X$取值的统计规律呢？在上式中，令$a=b$， 则有： $P\left(\left\{\omega:X(\omega)=a\right\}\right)=P\left(\left\{X=a\right\}\right)=0$ ；由此说明，$r.v.X$不同于第二章的离散型随机变量，他们有着本质的区别！

【连续型随机变量（**continuous random variable or c.r.v.**）】设X为概率空间(Ω,F,P)上的随机变量，若存在非负可积函数$f(x)$， 使得$\forall a,b\in R,a<b$，都有：$P\left(\left\{\omega:a<X(\omega)\le b\right\}\right)=P\left(\left\{a<X\le b\right\}\right)=\int_a^b f(x)dx$；则称$X$为连续型随机变量，简记为：**c.r.v.** ;称$f(x)$为X的概率密度函数（概率密度、分布密度、密度）（**probability**

density function）简记为：**p.d.f.**；

【注 1】（1）定义中的条件等式可等价地描述为：（验证之）$\forall a\in R, P(\{\omega: X(\omega)\le a\})=P(\{X\le a\})=P(\{-\infty<X\le a\})=\int_{-\infty}^{a} f(x)dx$；（2）由定义，概率密度函数不是唯一的；这是由于改变其作为被积函数在至多可列个点处的函数值不改变积分值！（3）若 $f(x)$ 是一个概率密度函数，令：$A_n=\{\omega:-n<X(\omega)\le n\}=\{-n<X\le n\}, n\ge 1,$ 易见，随机事件列 $\{A_n, n\ge 1\}\uparrow\Omega$，并且，

$$\int_{-\infty}^{+\infty} f(x)dx=\lim_{n\to\infty}\int_{-n}^{+n} f(x)dx=\lim_{n\to\infty}P(\{-n<X\le n\})=\lim_{n\to\infty}P(A_n)=P\left(\lim_{n\to\infty}A_n\right)=P(\Omega)=1$$

反之，若函数 $f(x)$ 满足:1)非负可积；2) $\int_{-\infty}^{+\infty} f(x)dx=1$；则 $f(x)$ 必可作为某连续型随机变量的概率密度函数；（4）若随机变量 X 具有密度 $f(x)$，则有：$\frac{P(\{x_0-\Delta x<X\le x_0\})}{\Delta x}=\frac{1}{\Delta x}\int_{x_0-\Delta x}^{x_0} f(x)dx$；若 $f(x)$ 在 x_0 点连续，则有：$\lim_{\Delta x\to 0}\frac{P(\{x_0-\Delta x<X\le x_0\})}{\Delta x}=f(x_0)$；这类似于物理学中的“质量的线（面，体）密度”，故而称 $f(x)$ 为密度；但密度不一定是连续函数（密度的几何意义）；

（5）比较：离散型随机变量与连续型随机变量；

(a)若 X 为离散型随机变量，则有：

$$P(\{X=a\})>0, a\in R(X); P(\{X=a\})=0, a\notin R(X);$$

(b)若 X 为连续型随机变量，且 $X\sim f(x)$，则有：

$\forall n\ge 1, \{X=a\}\subset\left\{a-\frac{1}{n}<X\le a+\frac{1}{n}\right\}=A_n\downarrow$；一方面，$P(\{X=a\})=P\left(\lim_{n\to\infty}A_n\right)$ $=\lim_{n\to\infty}P(A_n)=\lim_{n\to\infty}\int_{a-\frac{1}{n}}^{a+\frac{1}{n}} f(x)dx=0$；一方面，

$P(\{X=a\})\le P\left(\left\{a-\frac{1}{n}<X\le a+\frac{1}{n}\right\}\right)=\int_{a-\frac{1}{n}}^{a+\frac{1}{n}} f(x)dx\to 0, n\to\infty$；从而，$\forall a\in R, P(\{X=a\})=0$；也即：连续型随机变量取任意值的概率总为 0！进而有：

$$\boxed{P(\{a<X<b\})=P(\{a<X\le b\})=P(\{a\le X<b\})=P(\{a\le X\le b\})}$$

【例 3.1.1】设连续型随机变量 X,Y 具有相同的概率密度，且 $X\sim f_X(x)=\frac{3}{8}x^2 I_{\{0<x<2\}}(x)$，其中 $I_{\{0<x<2\}}(x)$ 是 $\{0<x<2\}$ 的示性函数；若已知 $A=\{X>a\}$ 与 $B=\{Y>a\}$ 相互独立，且 $P(A\cup B)=\frac{3}{4}$，求常数 a；

【例 3.1.2】设 $c.r.v.X$ 的 $p.d.f.$为 $f(x)=\begin{cases}\frac{1}{3},0<x<1;\\ \frac{2}{9},3\le x\le 6;\end{cases}$ ，且k使得 $P(\{X\ge k\})=\frac{2}{3}$ ，求k的取值范围；

现在，我们已经知道：概率分布（分布列）和概率密度分别是用来刻划离散型随机变量和连续型随机变量取值的统计规律的；为了统一刻划各类随机变量，也为了避免集（合）函数使用和研究的不便；以下引入（概率）分布函数，它只是普通的点函数。

【(概率）分布函数】设 $X=X(\omega)$是定义于概率空间(Ω,F,P)上的一个随机变量，$\forall x\in R$，定义：$F(x)=P(\{X\le x\})$，称之为$r.v.X$的（概率）分布函数，简记为：**d.f. (distribution function)**。

（**1**） 若X为$d.r.v.$，且有分布列：$P(\{X=x_i\})=p_i,i=1,2,\cdots$，则有：

$$F(x)=P(\{X\le x\})=P\left(\bigcup_{i:x_i\le x}\{X=x_i\}\right)=\sum_{i:x_i\le x}P(\{X=x_i\})=\sum_{i:x_i\le x}p_i \quad ;$$

（**2**） 若X为$c.r.v.$， 且有概率密度$f(x)$；则有：

$$F(x)=P(\{X\le x\})=P(\{-\infty<X\le x\})=\int_{-\infty}^{x}f(t)dt$$

；若 $f(x)$ 连续，则（几乎处处）有：$\boxed{\frac{d}{dx}F(x)=f(x)}$ ；

【注 2】(a)由定义，分布函数$F(x)$是定义于实数集R上的取值于$[0,1]$上的点函数，它表示事件$\{\omega:X(\omega)\le x\}$发生的概率，也即：$\{\omega:X(\omega)\in(-\infty,x]\}$的概率；当右端点$x$发生变化时，$P(\{X\le x\})$也相应发生变化，故$P(\{X\le x\})$是右端点$x$的函数；$(b)$给定概率空间$(\Omega,F,P)$，$P$是定义于$F$上的集函数，为使$P(\{X\le x\})$有意义，必须要求：$\forall x\in R,\{X\le x\}\in F$，这正符合随机变量定义中的限制条件；

(c)对于多个随机变量的情形，常常附之以下标，如：$F_X(x),F_Y(y),\cdots\cdots$， 自变量用相应的小写，这是为了避免混淆；

(d)随机变量X的分布函数$F(x)$包含了它的全部概率特征的信息；如：

1）任取数列$\{b_n,n\ge 1\}$，满足：$b_1>b_2>\cdots>b_n\cdots\to 0$，再令：

$A_n = X^{-1}((-\infty, x-b_n]) = \{X \le x-b_n\}, n \ge 1$，则有 $\{A_n, n \ge 1\} \uparrow A = \{X < x\}$

从而有：$P(\{X < x\}) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} F(x-b_n) = F(x-)$（由数列的任意性），$x \in R$；

2） $\forall x \in R$，$\{X = x\} = \{X \le x\} - \{X < x\}$，从而有：

$\boxed{P(\{X = x\}) = P(\{X \le x\}) - P(\{X < x\}) = F(x) - F(x-)}$；类似有：

$\forall x \in R$，$\overline{\{X \le x\}} = \{X > x\}, \overline{\{X \ge x\}} = \{X < x\}$，从而有：

$P(\{X > x\}) = 1 - P(\{X \le x\}) = 1 - F(x)$；$P(\{X \ge x\}) = 1 - P(\{X < x\}) = 1 - F(x-)$；

3） $\forall a, b \in R, a < b$，有如下结论：

$P(\{a < X < b\}) = P(\{X < b\}) - P(\{X \le a\}) = F(b-) - F(a)$；

$P(\{a < X \le b\}) = P(\{X \le b\}) - P(\{X \le a\}) = F(b) - F(a)$；

$P(\{a \le X \le b\}) = P(\{X \le b\}) - P(\{X < a\}) = F(b) - F(a-)$；

$P(\{a \le X < b\}) = P(\{X < b\}) - P(\{X < a\}) = F(b-) - F(a-)$；至此，我们知道，无论随机变量在怎样的区间（开的，闭的，半开半闭的，还是有界的，无界的）取值，都可以通过分布函数来刻划！

(e) 可以证明：同分布的随机变量具有相同的分布函数，从而具有相同的概率特征；因此，在概率论中同分布的随机变量常常不加以区别，视作一体！

分布函数具有如下性质：

（**1**）【单调不减性】 $\forall a, b \in R. a < b, \{X \le a\} \subset \{X \le b\}$，从而，

$$F(a) = P(\{X \le a\}) \le P(\{X \le b\}) = F(b)$$；

（**2**）【右连续性】任取数列 $\{b_n, n \ge 1\}$ 满足：$\forall a \in R$，$b_1 > b_2 > \cdots > b_n \cdots \to a$，令 $A_n = \{a < X \le b_n\}, n \ge 1$，易见随机列 $\{A_n, n \ge 1\} \downarrow \bigcap_{n=1}^{\infty} A_n = \Phi$，则

$$0 = P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty}\left[P(\{X \le b_n\}) - P(\{X \le a\})\right] = \lim_{n\to\infty} F(b_n) - F(a);$$

由数列 $\{b_n, n \ge 1\}$ 的任意性，即有：$\forall a \in R$，$F(a) = \lim_{x\to a+} F(x)$；

（**3**）【正则性】令 $B_n = \{X \le -n\}, C_n = \{X \le n\}$，则有事件列：$\{B_n, n \ge 1\} \downarrow \Phi$，$\{C_n, n \ge 1\} \uparrow \Omega$；从而由概率的连续性知：

$$F(-\infty) = \lim_{n\to\infty} F(-n) = \lim_{n\to\infty} P(B_n) = P\left(\lim_{n\to\infty} B_n\right) = P(\Phi) = 0$$；

$$F(+\infty) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right) = P(\Omega) = 1$$；

【例 3.1.3】设随机变量 X 有分布列：$\begin{pmatrix} X & 0 & 1 & 3 \\ P & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$，试求 X 的（概率）分布函数；

【例 3.1.4】设 $r.v.X$ 的 $d.f.$ 为 $\frac{1}{4}I_{\{2\le x<3\}}(x)+\frac{1}{2}I_{\{3\le x<5\}}(x)+I_{\{x\ge 5\}}(x)$,试求 X 的分布（列、律)；

【例 3.1.5】设 X 具有密度 $f(x)=\frac{1}{2}e^{-|x|},x\in R$，试求 X 的分布函数；

【例 3.1.6】若 X 的分布函数为 $xI_{\{0\le x<1\}}(x)+I_{\{x\ge 1\}}(x)$，试求 X 的 $p.d.f.$；

【例 3.1.7】一个使用了 t 小时的热敏电阻，在 Δt 内失效的概率为 $\lambda\Delta t+o(\Delta t)$，设其使用寿命是连续型随机变量，求 X 的分布；

【例 3.1.8】设 X 的 $d.f.$ 为：$F(x)=\frac{x}{2}I_{\{0\le x<1\}}(x)+\frac{2}{3}I_{\{1\le x<3\}}(x)+I_{\{x\ge 3\}}(x)$，试求事件 $\{X<2\},\{X=1\},\left\{X>\frac{1}{2}\right\}$ 和 $\{2<X<3\}$ 的概率；

【例 3.1.9】设 X 的分布函数为 $F(x)$,则 $F(ax+b)$，$F(x^2+a)$，$F(x^3-a)$，$F(|x|)$ 哪些可作为随机变量的分布函数？

【例 3.1.10】设 $r.v.X$ 的 $d.f.$ 和 $p.d.f.$ 分别为 $F(x)$ 和 $f(x)$，当 $x\le 0$ 时，$F(x)=0$；当 $x>0$ 时，$F(x)+f(x)=1$，试求 $f(x)$；

【例 3.1.11】设 $r.v.X=c,a.s.$（几乎必然），Y 为 $c.r.v.$，则 $Z=X+Y$ 的分布函数为连续函数；

【例 3.1.12】在 ΔABC 中任取一点 P，设 ξ 为点 P 到底边 AB 的距离，又已知 AB 上的高为 h，求 ξ 的分布；

【例 3.1.13】设 $F(x)=\frac{1+2x}{3}I_{\{0<x\le 1\}}(x)+I_{\{x>1\}}(x)$，试证明：1）$F(x)$ 是分布函数；2）$F(x)$ 既不是离散型，也不是连续型，但却可以表示为两类分布函数的线性组合！

【例 3.1.14】假设 $r.v.\xi$ 的绝对值不大于 1，$P(\{\xi=1\})=2P(\{\xi=-1\})=\frac{1}{4}$，在事件 $\{-1<\xi<1\}$ 出现的条件下，ξ 在 $(-1,1)$ 内任一子区间上取值的条件概率与该区间的长度成正比，试求 ξ 的分布函数；

【本例的注】例中的 $r.v.\xi$ 是混合型的随机变量；它有离散部分，因为它取 -1 与 1 的概率均大于 0；它又有连续部分，它的可能取值充满 $(-1,1)$；于是其分布函数可以分解。易验证：以上的分布

函数$F(x)$满足：$\forall x\in R$，$F(x)=\frac{3}{8}F_1(x)+\frac{5}{8}F_2(x)$，其中：

$$F_1(x)=\begin{cases}0,x<-1;\\ \frac{1}{3},-1\le x<1;\\ 1,x\ge 1;\end{cases}，\quad F_2(x)=\begin{cases}0,x<-1;\\ \frac{x+1}{2},-1\le x<1;\\ 1,x\ge 1;\end{cases}。$$

【例 3.1.15】若$F(x)$是一分布函数，试证：$\forall h>0,\Phi(x)=\frac{1}{h}\int_x^{x+h}F(t)dt$也是一分布函数；

【注 3】（补充）设(Ω,F,P)是概率空间，则任意 **F**-可测随机变量的全体是线性空间；不仅如此，随机变量对于通常的运算都是封闭的！（课堂推导）

§2 连续型随机变量的数学期望

已经知道离散型随机变量的数学期望即是其所有可能取值的概率平均值；对于连续型随机变量，其密度$f(x)$与dx的作用相当于离散型随机变量的分布列p_i，受此启发：利用微元法的思想，用$f(x)dx$替代p_i，用积分号替代求和号来定义连续型随机变量的数学期望。

【分析】设连续型随机变量X的值域$R(X)$为有限区间$(a,b]$在此区间任意插入$n-1$个分点$x_1,x_2,\cdots,x_{n-1}$，使得

$a=x_0<x_1<x_2<\cdots<x_{n-1}<x_n=b$，将$(a,b]$分割成$n$个子区间，其中第$i$个子区间为$(x_{i-1},x_i]$，$i=1,2,\cdots,n$；相应有：

$P(\{X\in(x_{i-1},x_i]\})=\int_{x_{i-1}}^{x_i}f(x)dx=f(\xi_i)(x_i-x_{i-1}),\xi_i\in(x_{i-1},x_i]$；若$x_i-x_{i-1}$充分小，则有$P(\{X\in(x_{i-1},x_i]\})\approx f(x_i)(x_i-x_{i-1})$，$i=1,2,\cdots,n$；此时可将随机变量$X$近似视作离散型随机变量，且有分布列：

X	ξ_1	ξ_2	$\cdots\cdots$	ξ_n
P	$f(\xi_1)\Delta x_1$	$f(\xi_2)\Delta x_2$	$\cdots$	$f(\xi_n)\Delta x_n$

，其中：

$\Delta x_i=x_i-x_{i-1},i=1,2,\cdots,n$；故有：$EX\approx\sum_{i=1}^{n}\xi_i f(\xi_i)\Delta x_i$；注意到，$n$越大，和式作为期望的近似值精确度就越高；从而，可定义：

$EX=\lim_{n\to\infty}\sum_{i=1}^{n}\xi_i f(\xi_i)\Delta x_i=\int_a^b xf(x)dx$；去掉最初关于随机变量值域的局限，即有：$\boxed{EX=\int_{-\infty}^{+\infty}xf(x)dx}$。

【数学期望】设 X 是一个连续型随机变量，且有密度 $f(x)$，如果 $\int_{-\infty}^{+\infty}|x|f(x)dx<\infty$，则称 $\int_{-\infty}^{+\infty}xf(x)dx$ 为 X 的（数学）期望（均值），记之为：$E(X)$ 或 EX；若 $\int_{-\infty}^{+\infty}|x|f(x)dx=\infty$，则称 X 的数学期望不存在。

【例 3.2.1】（柯西分布）设 $r.v.X\sim f(x)=\frac{1}{\pi(1+x^2)},x\in R$，判断 EX 是否存在？

【例 3.2.2】设随机变量 X 具有概率密度

$$f(x)=xI_{\{0<x\le1\}}(x)+(2-x)I_{\{1<x\le2\}}(x)\ ;$$ 试求 EX；

【例 3.2.3】设 $r.v.X\sim f(x)=\frac{1}{2a}e^{-\frac{|x-b|}{a}},x\in R$，试求 EX（结论加以推广）；

【例 3.2.4】设 G 为曲线 $y=2x-x^2$ 与 x 轴所围区域，在区域 G 内任取一点，该点到 y 轴的距离为 ξ，求 $E\xi$；

下面讨论几类重要的连续型分布。

【均匀分布（**Uniform distribution**）】若随机变量 X 具有概率密度 $f(x)=\frac{1}{b-a}I_{\{a\le x\le b\}}(x)$，则称 X 服从 $[a,b]$ 上的均匀分布，记作：$X\sim U[a,b]$；讨论均匀分布的概率背景：几何概型，E：向 $[a,b]$ 上随机地投掷一点，记点的位置（坐标）为 X；易见，$R(X)=[a,b]$，且 $\forall x\in R$，有：

$x<a,F(x)=P(\{X\le x\})=0;a\le x\le b,F(x)=\frac{x-a}{b-a};x>b,F(x)=1$，进一步有：$X$ 的概率密度为 $f(x)=\frac{1}{b-a}I_{\{a\le x\le b\}}(x)$；反之，若 $X\sim U[a,b]$，则有：$\forall[c,d]\subset[a,b],P(\{X\in[c,d]\})=\frac{d-c}{b-a}$，也即 X 取值于 $[a,b]$ 的任一子区间的概率与该区间的长度成正比而与其位置无关，即为几何概型的情形！

【例 3.2.5】设 $r.v.X\sim U[1,6]$，求“方程 $t^2+Xt+1=0$ 有实根”的概率；

【例 3.2.6】设 $r.v.\xi\sim U[0,1]$，求：1）$X=[n\xi]+1$；2）$Y=\left[\frac{\ln\xi}{\ln q}\right]+1,0<q<1$ 的分布；

【例 3.2.7】（1）设 $r.v.X$ 在 $(0,1)$ 内取值，其 $d.f.$ 为 $F(x)$，且 $\forall 0<x<y<1$，$F(y)-F(x)$ 与 $y-x$ 成正比，试求 $F(x)$；（2）设 $r.v.X$ 在 $(0,1)$ 内取值，且 $\forall 0<x<y<1$，$P(\{x<X\le y\})$ 只依赖于 $y-x$，试求 X 的分

布；

【例 3.2.8】设$\{\xi_n, n\ge 1\}$是定义于概率空间(Ω,F,P)上的一列独立同$B(1,0.5)$分布的随机变量，且$\forall\omega\in\Omega$，定义：$\eta(\omega)=\sum_{n=1}^{\infty}\frac{\xi_n(\omega)}{2^n}$；则$\eta$为定义于$(\Omega,F,P)$上的服从$U[0,1]$分布的随机变量；

【例 3.2.9】设随机变量$Y\sim U[0,1]$，令$Y_n=\sum_{k=0}^{2^n-1}\frac{k}{2^n}I_{\left\{\frac{k}{2^n}\le Y<\frac{k+1}{2^n}\right\}}(\omega), n\ge 0$；求：（1）$Y_1,Y_2$的分布列；（2）$Y_n$的分布列；（3）$Y_1=\frac{1}{2}$给定的条件下，$Y_2$的条件分布列；

【指数分布（**Exponential distribution**）】若随机变量X具有概率密度$f(x)=\lambda e^{-\lambda x}I_{\{x>0\}}(x)$，$\lambda>0$为常数，则称$X$服从参数为$\lambda$的指数分布，记作：$X\sim E(\lambda)$（或 $Ex(\lambda),Exp(\lambda)$）；

【复习伽玛函数及其性质】

【例 3.2.10】设$r.v.X\sim E(1)$，$Y=\min(X,2)$，试求Y的分布及EY；（正误辨析一顺便介绍 **Stieltjes[斯蒂尔杰斯]积分**）

【无记忆性】可验证以下命题的等价性：设X为非负连续型随机变量，则（1）X服从指数分布；（2）$P(X>x+y|X>x)=P(\{X>y\})$，$x,y>0$；

【例 3.2.11】设在时间$(0,t]$内经搜索发现沉船的概率为$P(t)=1-e^{-\lambda t}$，$t>0$，求发现沉船所需的平均时间；

【例 3.2.12】设$r.v.Y\sim E(1)$，$X_k=I_{\{Y>k\}}(\omega),k=1,2$，试求$X_1,X_2$的联合（边缘）分布；

【正态分布（**Normal distribution**）】（又称高斯[**Gauss**]分布）若离散型随机变量X有概率密度为：$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in R$，其中$\mu,\sigma(\sigma>0)$为常数，则称$X$服从参数为$(\mu,\sigma)$的正态分布，记作：$X\sim N(\mu,\sigma^2)$；若$\mu=0,\sigma=1$，则称$X$服从标准正态分布。

易见，密度$f(x)$具有如下性质：1）关于直线$x=\mu$对称，此即表明：$\forall h>0, P(\{\mu-h<X\le\mu\})=P(\{\mu<X\le\mu+h\})$；2）$f(x)$的图像呈倒立的钟形，且在$x=\mu$处取得最大值，曲线以$x$轴为水平渐近线；对于同样长度的区间，其离$\mu$越远，$X$落于其上的概率就越小；3）$f(x)$在$x=\mu\pm\sigma$处有拐点，且存在任意阶导数；且$\sigma$

越大（小），曲线越平坦（陡峭）；

可以验证：$\int_{-\infty}^{+\infty} f(x)dx=1$（**Poisson 积分**）。

【注1】 正态分布是自然界中最常见的一种分布，如：测量的误差，炮弹的落点等诸多现象都服从或近似服从正态分布。一般地，若影响某一数量的指标很多，且每种影响都相互独立，作用很小，即可认为这个数量服从或近似服从正态分布；许多分布都以正态分布来逼近；许多重要的分布，如三大抽样分布：χ^2（卡方）分布，t（学生氏）分布和F（费歇尔）分布都是由正态分布所导出的；同时，正态分布还有许多优良的性质也使其无论是在理论上，还是在应用上都有重要的作用！

常将标准正态分布的分布函数和密度函数即为：$\Phi(x)$和$\varphi(x)$；

1）若$r.v.X\sim N(\mu,\sigma^2)$，则

$P(\{X\le\mu-x\})=\int_{-\infty}^{\mu-x} f(t)dt=\int_{x}^{+\infty} f(\mu-t)dt=1-F(\mu+x)$；特别地，若 $r.v.X\sim N(0,1^2)$，则有：$\boxed{\Phi(-x)=1-\Phi(x)}$，$x\in R$ ；

2）若$r.v.X\sim N(\mu,\sigma^2)$，令 $Y=\dfrac{X-\mu}{\sigma}$， 则有：$\forall y\in R, F_Y(y)=P(\{Y\le y\})=$

$P(\{X\le\mu+\sigma y\})=\dfrac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma y} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{t^2}{2}}dt=\Phi(y)$，此即说明$Y$服从标准正态分布（任何正态分布的概率计算均可转化与标准正态分布有关的概率计算）！

【例 3.2.13】设$X\sim N(1,0.4^2)$，试求$P(\{X>0\})$和$P(\{0.2<X<1.8\})$；

【例 3.2.14】已知$\ln X\sim N(1,2^2)$ ，试求 $P\left(\left\{\frac{1}{2}<X<2\right\}\right)$ ；

【例 3.2.15】假设数学的考试成绩近似服从正态分布$N(70,10^2)$，按考分从高到低排第100名的成绩恰为60分（其余的不及格），问：按考分从高到低排第20名的成绩约为多少？

【例 3.2.16】设某电子元件在工作中其两端电压$V\sim N(220,20^2)$，当$V\in[200,240]$，失效的概率为0.05；当$V<200$，失效的概率为0.1；当$V>240$，失效的概率为0.5；求：

（1）“此元件失效”的概率；

（2）“当元件失效时，电压超过240”的概率；

【例 3.2.17】设$r.v.X\sim N(2,1^2)$，$B=\{\omega:X(\omega)\ge 0\}=\{X\ge 0\}$，试求$X$在$B$

发生的条件下的条件概率分布函数 $F_{X|B}(x)$ 和条件概率密度函数 $f_{X|B}(x)$；

【例 3.2.18】设随机变量 $X\sim N(\mu,\sigma^2)$，则有：$E(|X-EX|)=\sqrt{\frac{2}{\pi}}\sigma$；

【例 3.2.19】国际市场每年对我国某种商品的需求量是随机变量 X（单位：吨），且 $X\sim U[2000,4000]$；已知每售出一吨，可挣得外汇 3 千元，但若售不出去，则每吨需支付存储费及其他损失 1 千元，问：需组织多少货源，才能使国家期望收益最大？

【例 3.2.20】报童每天卖报量 ξ 是一个随机变量，且 $\xi\sim N(1000,100^2)$（单位：份）；设每份报纸的购进价为 0.3 元，卖出价为 0.5 元。如果当天卖不完，可退还报社，每份可得退回费 0.2 元；试问：该报童每天应取多少份报纸才能使平均收益最大？

【例 3.2.21】医院对某地区的人进行验血，该地区一个人化验结果为阳性的概率为 p，原方法需每人化验一次；现采用把 k 个人的血样合在一起化验的方法，若为阴性，则每个人的结果皆为阴性；若为阳性，再把这 k 个人的血样逐个化验；

1）试求 k 个人一组中，一个人化验次数的分布列；

2）若 $p=0.01$，$k=10$，医院的工作量平均会减少百分之几？

【例 3.2.22】某分布的分布函数为 $F(x)=\begin{cases}0, x<0;\\ \frac{2x+1}{5}, 0\le x<1;\\ 1, x\ge 1;\end{cases}$，试由 $R-S$ 积分的双线性性质求该分布的数学期望；

【例 3.2.23】在可靠性与生存分析中，所研究的寿命现象是非负随机变量，记作 ξ，其分布函数为 $F(x)$，密度函数为 $f(x)$，称 $S(x)=1-F(x)$ 为生存函数，这时常引入失效率函数 $\lambda(x)$，其定义为：$\lambda(x)=\frac{f(x)}{S(x)}$；

1） 给出失效率函数 $\lambda(x)$ 的直观解释，并推导用 $\lambda(x)$ 表示 $S(x)$ 的公式；

2） 某放射性物质在初始时刻的质量为 m_0，在单位时间内每个原子产生分裂核的概率为常数 λ，试求经过时间 x 后该放射性物质质量的期望；

§3 连续型随机变量的独立性

以下介绍二维连续型随机向量，$n(n>2)$ 维的情形依此类推！

【连续型随机向量】设 X,Y 是定义于同一概率空间 (Ω,F,P) 上的随机变量，称二维随机向量 (X,Y) 是连续型的（联合连续的），若存在非负可积函数 $f(x,y)$，$(x,y)\in R^2$，使得对于任意的二维矩形域 $B=\{(x,y):a<x\le b,c<y\le d\}$，都有

$\boxed{P(\{(X,Y)\in B\})=P(\{a<X\le b,c<Y\le d\})=\iint_B f(x,y)dxdy}$，并称 $f(x,y)$ 为 (X,Y) 的联合（概率）密度（函数）。

【注 1】1）可以验证联合密度 $f(x,y)$ 满足：a) 非负可积；

b) $\iint_{R^2} f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy=1$；（反常或广义二重积分）

2）若 $f(x,y)$ 在 (x_0,y_0) 点连续，则有

$$\lim_{\Delta x\to0,\Delta y\to0}\frac{P(\{(X,Y)\in(x_0-\Delta x,x_0]\times(y_0-\Delta y,y_0]\})}{\Delta x\Delta y}=f(x_0,y_0)$$

，由该式求联合密度的方法称为“微元法”；

3)事实上，若 $G\in\beta^2$（二维 Borel 域）是 xoy 平面上任一区域，则以 X,Y 为横，纵坐标的随机点落入区域 G 的概率为：

$\boxed{P(\{(X,Y)\in G\})=\iint_G f(x,y)dxdy}$；并且也可由该式作为上述定义中的条件式。

【注2】 若 (X,Y) 具有联合密度 $f(x,y)$，易知：$\forall a,b\in R,a<b$，

$$P(\{a<X\le b\})=P(\{a<X\le b,-\infty<Y<+\infty\})=\iint_{(a,b]\times(-\infty,+\infty)} f(x,y)dxdy$$

$=\int_a^b\left(\int_{-\infty}^{+\infty} f(x,y)dy\right)dx$；由 a,b 的任意性，可知 (X,Y) 的分量 X 的概率密度函数为 $\int_{-\infty}^{+\infty} f(x,y)dy$，常称之为 (X,Y) 关于 X 的边缘（边际）密度（函数）；记之为：$f_X(x)$。同理，可给出 (X,Y) 关于 Y 的边缘（边际）密度函数：$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx$。总之，边缘密度可由联合密度唯一确定！

【例 3.3.1】函数 $f(x,y)=k\exp\{-(ax^2+2bxy+cy^2)\}$ 为二元密度函数的充要条件是：$a>0$，$c>0$，$ac-b^2>0$，$k=\frac{1}{\pi}\sqrt{ac-b^2}$。

【例 3.3.2】设 $f(x,y)=\begin{cases} x^2+\frac{xy}{3},(x,y)\in[0,1]\times[0,2]; \\ 0,\text{其他}; \end{cases}$，试求：$P(\{X+Y\geq 1\})$；

【例 3.3.3】设 $(\xi,\eta)\sim f(x,y)=cx^2y,0<x<2,0<y<1$；求：1)常数 c；2) “ξ,η 至少有一个小于 $\frac{1}{2}$” 的概率；

【例 3.3.4】利用概率思想来证明：$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{-\frac{x^2}{2}}dx\leq\sqrt{1-e^{-a^2}}$ ；

【例 3.3.5】试举出反例说明 X,Y 是连续型随机变量，但 (X,Y) 却不是连续型随机向量；且该情形不适用于离散型随机变量；

【例 3.3.6】设 $f(x,y)=\begin{cases} x^2+\frac{xy}{3},(x,y)\in[0,1]\times[0,2]; \\ 0,\text{其他}; \end{cases}$，试求：

（1） $f_X(x),f_Y(y)$ ；（2） $P(\{X=Y\})$ ；（3） $P\left(Y<\frac{1}{2}\middle|X<\frac{1}{2}\right)$ ；

【例 3.3.7】（二维均匀分布）(a)设 $(X,Y)\sim U(D)D:x^2+y^2\leq R^2$，试求 X,Y 的 $p.d.f.$ $f_X(x),f_Y(y)$;(b)向区域 $G=\{(x,y)\,|\,|x|+|y|\leq 2\}$ 内随机地投掷一点，记之为 (X,Y)，求 (X,Y) 的 $J.p.d.f.$和 X,Y 的 $M.p.d.f.$；

【注 3】（二维）均匀分布的边缘分布不一定是（一维）均匀分布；

【注 4】掌握由联合密度求边缘密度的两种作法！

【例 3.3.8】设 $(\xi,\eta)\sim U(D),D=\{(x,y)\,|\,0<x^2<y<x<1\}$ ，试求：

1） ξ 的边缘密度函数；2） $P\left(\left\{\eta>\frac{1}{2}\right\}\right)$ ；

以下将要介绍的联合（概率）分布函数是统一刻划各类随机向量（包括联合离散，联合连续等情形）取（向量）值的概率规律一个重要、有效工具！

【联合（概率）分布函数】设 (X,Y) 为二维随机向量，$\forall(x,y)\in R^2$，称 $F(x,y)=P(\{X\leq x,Y\leq y\})$ 为 (X,Y) 的联合（概率）分布函数；更高维的情形依此！（讨论其几何意义）

若 (X,Y) 是联合离散的，且有：

$P(\{(X,Y)=(x_i,y_j)\})=p_{ij},i=1,2,\cdots;j=1,2,\cdots;$ 则有：$\forall(x,y)\in R^2$，

$$F(x,y)=P(\{X\leq x,Y\leq y\})=P\left(\bigcup_{i:x_i\leq x}\bigcup_{j:y_j\leq y}\{X=x_i,Y=y_j\}\right)=\sum_{i:x_i\leq x}\sum_{j:y_j\leq y}P(\{X=x_i,Y=y_j\})$$

$$=\sum_{i:x_i\le x}\sum_{j:y_j\le y}p_{ij}\quad ;$$

【例 3.3.9】设(X,Y)的联合分布为：$\begin{pmatrix} X/Y & 1 & 2 \\ 1 & \frac{1}{4} & \frac{1}{2} \\ 3 & \frac{1}{8} & \frac{1}{8}\end{pmatrix}$；求$(X,Y)$的联合分布函数；

若(X,Y)是联合连续的，且有：$(X,Y)\sim f(x,y)$，则有：$\forall (x,y)\in R^2$

$$F(x,y)=P(\{X\le x,Y\le y\})=P(\{(X,Y)\in(-\infty,x]\times(-\infty,y]\})=\int_{-\infty}^{x}\int_{-\infty}^{y}f(s,t)dtds\quad ;$$

【例 3.3.10】设$(X,Y)\sim f(x,y)=3xI_{\{0<y<x<1\}}(x,y)$，试求$(X,Y)$的联合分布函数；

【例 3.3.11】问：$F(x,y)=I_{\{x+y>0\}}(x,y),(x,y)\in R^2$是否为一随机向量的联合分布函数？

【例 3.3.12】设(X,Y)的联合分布函数为

$$F(x,y)=A\left(B+\arctan\frac{x}{2}\right)\left(C+\arctan\frac{y}{2}\right),-\infty<x,y<+\infty$$ ，试求A,B,C；

<u>联合分布函数具有如下一些性质</u>：

设(X,Y)的联合分布函数为：$F(x,y)=P(\{X\le x,Y\le y\})$ ，

（1）$F(x,y)$是关于$x(y)$的单调递增函数；（简证）

（2）$0\le F(x,y)\le 1$且 $\forall x,y\in R$ ，

$$F(-\infty,y)=\lim_{x\to-\infty}F(x,y)=\lim_{x\to-\infty}P(\{X\le x,Y\le y\})=P(\Phi)=0\quad ;$$

$$F(x,-\infty)=\lim_{y\to-\infty}F(x,y)=\lim_{y\to-\infty}P(\{X\le x,Y\le y\})=P(\Phi)=0\quad ;$$

$$F(-\infty,-\infty)=\lim_{x\to-\infty,y\to-\infty}F(x,y)=\lim_{x\to-\infty,y\to-\infty}P(\{X\le x,Y\le y\})=P(\Phi)=0\,;$$

$$F(+\infty,+\infty)=\lim_{x\to+\infty,y\to+\infty}F(x,y)=\lim_{x\to+\infty,y\to+\infty}P(\{X\le x,Y\le y\})=P(\Omega)=1$$ ；而且，我们还有：$\forall x,y\in R,$

$$F_X(x)=P(\{X\le x\})=P(\{X\le x,Y<+\infty\})=\lim_{y\to\infty}P(\{X\le x,Y\le y\})=F(x,+\infty)\quad ;$$

$$F_Y(y)=P(\{Y\le y\})=P(\{X<+\infty,Y\le y\})=\lim_{x\to\infty}P(\{X\le x,Y\le y\})=F(+\infty,y)$$ ；称由此得到的$F_X(x),F_Y(y)$为(X,Y)关于$x(y)$的边缘分布函数！其即是X,Y的分布函数。需要说明的是这里由联合分布函数得到边缘分布函数的公式常称为分布函数族$\{F(x,y),F_X(x),F_Y(y)\}$的相容性

条件—边缘分布函数只能反映各分量单个变化的概率规律，而联合分布函数却能反映随机向量整体变化的概率规律，且还能反映分量之间的相互关系，这是边缘分布所不能反映出的；另外，我们还能看到边缘分布相同而联合分布不同的例子！

（3）$\forall x\in R$，$F(x,y)$ 是 y 的右连续函数；$\forall y\in R$，$F(x,y)$ 是 x 的右连续函数;

（4）$\forall a,b,c,d\in R, a<b, c<d$；

$\boxed{P(\{a<X\le b, c<Y\le d\})=F(b,d)-F(b,c)-F(a,d)+F(a,c)}$ （简证）;

若二维随机向量 (X,Y) 有联合分布函数 $F(x,y)$ 和联合密度函数 $f(x,y)$，则应有：$\boxed{\frac{\partial^2}{\partial x\partial y}F(x,y)=f(x,y)}$（几乎处处成立）。

【例 3.3.13】用 (X,Y) 的联合分布函数 $F(x,y)$ 来表示如下事件的概率： $\forall a,b,c,d\in R, a<b, c<d$，

1）$P(\{X<a, Y\le c\})$；2）$P(\{X<a, Y<c\})$；3）$P(\{a<X\le b, Y\le c\})$；

4）$P(\{X=a, Y\le c\})$；5）$P(\{a\le X<b, c\le Y<d\})$；6）$P(\{X<a, Y<+\infty\})$；

7）$P(\{X=a, Y<+\infty\})$；8）$P(\{X=a, Y=c\})$；（详细讲解）

（二维联合分布函数完全刻划了二维随机向量取“值”的概率规律）

【独立性】设 X,Y 是定义于同一概率空间 (Ω,F,P) 的两个随机变量，且 (X,Y) 的联合分布函数为 $F(x,y)$，X,Y 的边缘分布函数分别为 $F_X(x),F_Y(y)$，若 $\forall(x,y)\in R^2$，$\boxed{F(x,y)=F_X(x)F_Y(y)}$**(a)**，则称 X 与 Y 是（相互）独立的。

（1）若 X,Y 是联合离散的，则有**(b)**式：

$$\boxed{X,Y\text{相互独立}\Leftrightarrow \forall x\in R(X)(R),\forall y\in R(Y)(R), P(\{X=x,Y=y\})=P(\{X=x\})P(\{Y=y\})}$$

（简证**(a)**与**(b)**等价）先证：**(a)** $\Rightarrow$ **(b)**

由**(a)**式，即有：$\forall x,y\in R, P(\{X\le x,Y\le y\})=P(\{X\le x\})P(\{Y\le y\})$，**1)**先证 $\forall x,y\in R, P(\{X<x,Y\le y\})=P(\{X<x\})P(\{Y\le y\})$；**2)** $\forall x\in R$，任取数列 $\{x_n, n\ge1\}$ 使得 $x_1<x_2<\cdots<x_n<\cdots\to x$，从而 $\bigcup_{n=1}^{\infty}(-\infty,x_n]=(-\infty,x)$，即有：$\bigcup_{n=1}^{\infty}\{X\le x_n\}=\{X<x\}$，也即有：$\bigcup_{n=1}^{\infty}\{X\le x_n, Y\le y\}=\{X<x, Y\le y\}$；于是，

$$P(\{X<x,Y\le y\})=P\left(\bigcup_{n=1}^{\infty}\{X\le x_n,Y\le y\}\right)=\lim_{n\to\infty}P(\{X\le x_n,Y\le y\})=\lim_{n\to\infty}P(\{X\le x_n\})P(\{Y\le y\})$$

$=P\left(\bigcup_{n=1}^{\infty}\{X\le x_n\}\right)P(\{Y\le y\})=P(\{X<x\})P(\{Y\le y\})$ ；**1）－2），即得：** $\forall x,y\in R, P(\{X=x,Y\le y\})=P(\{X=x\})P(\{Y\le y\})$，重复以上的作法，即可得**(b)**式！再证：**(b)** $\Rightarrow$**(a)**　$\forall x,y\in R$，

$$P(\{X\le x,Y\le y\})=P\left(\bigcup_{i:x_i\le x}\{X=x_i,Y\le y\}\right)=P\left(\bigcup_{i:x_i\le x}\bigcup_{j:y_j\le y}\{X=x_i,Y=y_j\}\right)=\sum_{i:x_i\le x}\sum_{j:y_j\le y}P(\{X=x_i,Y=y_j\})$$

$=\sum_{i:x_i\le x}\sum_{j:y_j\le y}P(\{X=x_i\})P(\{Y=y_j\})=P(\{X\le x\})P(\{Y\le y\})$ ；即得**(a)**式！

（2）若 X,Y 是联合连续的，则有**(c)**式：

$\boxed{X,Y\text{相互独立}\Leftrightarrow f(x,y)=f_X(x)f_Y(y)[\text{几乎处处成立，在}f(x,y)\text{的连续点}]}$ ；

【例 3.3.14】设 $X\sim f(x)=\frac{1}{2}e^{-|x|},-\infty<x<+\infty$，试证：$X$ 与 $|X|$ 不独立；

【例 3.3.15】设 $(X,Y)\sim f(x,y)=\frac{1+xy}{4}I_{\{|x|<1,|y|<1\}}(x,y)$ ；试证：X 与 Y 不独立，但 X^2 与 Y^2 独立；

【例 3.3.16】设 (X,Y) 的 $J.d.f.$ 为：

$F(x,y)=xyI_{\{0\le x<1,0\le y<1\}}(x,y)+yI_{\{x\ge1,0\le y<1\}}(x,y)+xI_{\{0\le x<1,y\ge1\}}(x,y)+I_{\{x\ge1,y\ge1\}}(x,y)$ ；

问：X,Y 是否独立？

【例 3.3.17】设 $r.v.X=c,a.s.$，则 X 与任一随机变量 Y 必独立！

【例 3.3.18】设 X,Y 独立同 $N(0,1^2)$ 分布，试求脱靶量 $R=\sqrt{X^2+Y^2}$ 的概率密度；

【本例的注】这里之所以称 $R=\sqrt{X^2+Y^2}$ 为脱靶量，是因为若将 (X,Y) 视为弹落点，则 R 是弹落点到目标 $(0,0)$ 的距离。易知这里的 R 服从瑞利（**Rayleigh**）分布；且瑞利概率密度 $f_R(r)=re^{-\frac{r^2}{2}},r>0$ 在 $r=1$ 取最大值；如果画出以原点为圆心，若干宽度为 2ε 的圆环，则子弹落在圆环 $\left\{(x,y)\middle|1-\varepsilon<\sqrt{x^2+y^2}<1+\varepsilon\right\}$ 内的概率较大；这也解释了为什么优秀射击运动员在比赛时打出 9 或 8 环的机会较多，打出 10 环或 $7,6$ 环的机会较少。

【例 3.3.19】（正态分布的推导）在打靶问题中，设靶心位于所在平面的原点；由于各因素的影响，子弹落在点 **A** 的坐标 (ξ,η) 是一个二维随机向量，若它同时满足：**1.** ξ,η 分别具有连续密

度函数 $f_1(x)$ 和 $f_2(y)$，且 $f_1(0)f_2(0)>0$；**2.** ξ,η 相互独立；**3.** (ξ,η) 的密度函数在 (x,y) 的值只依赖于该点到靶心的距离；试求 ξ 与 η 的分布；

【例 3.3.20】（随机变量的随机加权平均）设 X,Y,ξ 是定义于同一概率空间 (Ω,F,P) 上的 3 个相互独立的随机变量，设 X,Y 的分布函数分别是 $F_1(x),F_2(x)$，而 $\xi\sim B(1,p),0<p<1$，令 $Z=\xi X+(1-\xi)Y$，称之为 X 与 Y 的随机加权平均，即：

$\forall\omega\in\Omega,Z(\omega)=\xi(\omega)X(\omega)+[1-\xi(\omega)]Y(\omega)$，　　试求 $Z=Z(\omega)$ 的分布；

【例 3.3.21】设 $(X,Y)\sim f(x,y)=\frac{12}{7}x(x+y)I_{\{0<x,y<1\}}(x,y)$，　令 $U=X\wedge Y,V=X\vee Y$，试求 (X,Y) 的联合密度函数；

【顺（次）序统计量】（补充）设 $X_1,X_2,\cdots\cdots,X_n$ 独立同分布且有分布函数 $F(\cdot)$，对每个 $\omega\in\Omega$，将 $X_1(\omega),X_2(\omega),\cdots\cdots,X_n(\omega)$ 从小到大排序：$X_{(1)}(\omega)\le X_{(2)}(\omega)\le\cdots\cdots\le X_{(n)}(\omega)$；这样，即可得到 n 个新的随机变量 $X_{(1)},X_{(2)},\cdots\cdots,X_{(n)}$，称它们为 $X_1,X_2,\cdots\cdots,X_n$ 的顺（次）序统计量；其中，$X_{(1)}$（$X_{(n)}$）称为最小（大）顺（次）序统计量，$X_{(i)}$ 称为第 i 个顺（次）序统计量。（举例说明）

易见，$\boxed{X_{(1)}=\min\{X_1,X_2,\cdots\cdots,X_n\}}$，$\boxed{X_{(n)}=\max\{X_1,X_2,\cdots\cdots,X_n\}}$，且有：

$$F_{X_{(n)}}(x)=P(\{X_{(n)}\le x\})=P(\{X_1\le x,X_2\le x,\cdots,X_n\le x\})=\prod_{i=1}^{n}P(\{X_i\le x\})=[F(x)]^n$$

$$F_{X_{(1)}}(x)=P(\{X_{(1)}\le x\})=1-\prod_{i=1}^{n}P(\{X_i>x\})=1-[1-F(x)]^n$$；下面考虑：

（1）$X_{(k)}$ 的分布（$1\le k\le n$）；

分布函数法：$F_{X_{(k)}}(x)=P(\{X_{(k)}\le x\})=P(\{X_1,X_2,\cdots,X_n\text{中至少有}k\text{个不大于}x\})$

$$=\sum_{i=k}^{n}P(\{X_1,X_2,\cdots,X_n\text{中恰有i个不大于}x\})$$

$$=\sum_{i=k}^{n}\frac{n!}{i!(n-i)!}P(\{X_{j_1}\le x,X_{j_2}\le x,\cdots,X_{j_i}\le x,X_{j_{i+1}}>x,\cdots,X_{j_n}>x\})$$

$$=\sum_{i=k}^{n}\frac{n!}{i!(n-i)!}[F(x)]^i[1-F(x)]^{n-i}$$　；

微元密度法：$\forall x\in R$，考虑：$h>0$ 充分小，$P(\{x-h<X\le x\})$

$$\approx P(\{X_1,X_2,\cdots,X_n\text{中恰有}k-1\text{个落入}(-\infty,x-h],\text{恰有1个落入}(x-h,x],\text{恰有n-k个落入}(x,+\infty\})$$

………；再由 $f_{X_k}(x)=\lim\limits_{h\to0}\dfrac{P(\{x-h<X_{(k)}\le x\})}{h}$ 即可得到密度函数！（这里暂略过程，课堂详述）

（2）$(X_{(k)},X_{(l)})(k<l)$ 和 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 的联合分布（微元密度法，部分结合分布函数法）——【课堂论证】

【例 3.3.22】在长为 a 的线段上任取 n 个点，则相距最远的两点之间距离的数学期望为 $\dfrac{n-1}{n+1}a$；

【例 3.3.23】设 $r.v.X$ 有连续可微的分布函数 $F(x)$，将 X 的 n 次独立观察所得的样本按从小到大的顺序排列得到：

$X_{(1)}<X_{(2)}<\cdots<X_{(n)}$；试求 $Y=\dfrac{F(X_{(n)})-F(X_{(2)})}{F(X_{(n)})-F(X_{(1)})}$ 的概率密度；

§4 条件分布与条件数学期望

【第一类条件分布】设 $(X,Y)\sim f(x,y)$，若 $A\subset R$，且 $P(\{X\in A\})>0$，则可定义给定 $X\in A$ 时关于 Y 的条件概率：$\forall a,b\in R,a<b$，

$P(a<Y\le b|X\in A)=\dfrac{P(\{a<Y\le b,X\in A\})}{P(\{X\in A\})}$；若 $\forall a,b\in R,a<b$，存在非负可积函数 $f_{Y|X\in A}(y)$，使得 $P(a<Y\le b|X\in A)=\int_a^b f_{Y|X\in A}(y)dy$，则称 $f_{Y|X\in A}(y)$ 为给定 $X\in A$ 时 Y 的<u>条件概率密度</u>；若 $\forall y\in R$，

$P(Y\le y|X\in A)=\int_{-\infty}^{y} f_{Y|X\in A}(y)dy$，则称 $F_{Y|X\in A}(y)=P(Y\le y|X\in A)$ 为给定 $X\in A$ 时 Y 的<u>条件分布函数</u>。

【例 3.4.1】设 $(X,Y)\sim f(x,y)=(x+y)I_{\{0<x,y<1\}}(x,y)$，试求在 $\left\{0<X<\dfrac{1}{n}\right\}$ 发生的条件下，Y 的条件分布函数和条件密度函数；

【例 3.4.2】设随机变量 $X\sim N(\mu,\sigma^2)$，$\forall a<b$，求 $F_{X|a<X<b}(x),f_{X|a<X<b}(x)$ 和 $E(X|a<X<b)$；

【例 3.4.3】设某元件的寿命 $X\sim E(\lambda)$（单位：小时），设已知 $X>a$ 时，求它的平均余寿；（两种方法）

【第二类条件分布】设 $(X,Y)\sim f(x,y)$ 即有：$\forall x\in R,P(\{X=x\})=0$ 从而 $\forall a,b\in R,a<b$，无法直接定义 $P(a<Y<b|X=x)$；自然会想到：

$\boxed{P(a<Y\le b|X=x)=\lim\limits_{\delta x\to0}P(a<Y\le b|x-\delta x<X\le x+\delta x)}$；

设 $(X,Y)\sim f(x,y),x\in\{x|f_X(x)>0\},\forall y\in R$，则有：

$$P(Y\le y|X=x)=\lim_{\delta x\to 0}P(Y\le y|x-\delta x<X\le x+\delta x)=\lim_{\delta x\to 0}\frac{P(\{x-\delta x<X\le x+\delta x,Y\le y\})}{P(\{x-\delta x<X\le x+\delta x\})}$$

$$=\lim_{\delta x\to 0}\frac{\int_{-\infty}^{y}\int_{x-\delta x}^{x+\delta x}f(s,t)dsdt}{\int_{x-\delta x}^{x+\delta x}f_X(s)ds}=\int_{-\infty}^{y}\frac{f(x,t)}{f_X(x)}dt\quad;$$

【条件密度函数】设$(X,Y)\sim f(x,y),x\in\{x|f_X(x)>0\}$，称$\frac{f(x,y)}{f_X(x)}$为给定$X=x$时$Y$的条件密度函数；同理，$y\in\{y|f_Y(y)>0\}$，称$\frac{f(x,y)}{f_Y(y)}$为给定$Y=y$时$X$的条件密度函数。

【注 1】（1）以上两个条件密度函数分别记为：$f_{Y|X}(y|x)$ 和 $f_{X|Y}(x|y)$；即有：$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}I_{\{x|f_X(x)>0\}}(x)$ 和

$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}I_{\{y|f_Y(y)>0\}}(y)$ ；常称之为密度的贝叶斯公式；

（2）由定义，$f(x,y)=f_{X|Y}(x|y)f_Y(y)$，$f(x,y)=f_{Y|X}(y|x)f_X(x)$，常称之为密度的乘法公式；

（3）设 $(X,Y)\sim f(x,y),\forall y\in R$ ，

$$P(\{Y\le y\})=P(\{-\infty<X<+\infty,Y\le y\})=\int_{-\infty}^{+\infty}\int_{-\infty}^{y}f(x,y)dydx$$

$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}f_{Y|X}(y|x)dy\right]f_X(x)dx=\int_{-\infty}^{+\infty}P(Y\le y|X=x)f_X(x)dx\quad\textbf{(a)};$$

事实上，类似可得到：$\boxed{P(\{a<Y\le b\})=\int_{-\infty}^{+\infty}P(a<Y\le b|X=x)f_X(x)dx}$ **(b)**；

更一般地，若 $A\in F$ 且与某连续型随机变量X有关，则有：

$\boxed{P(A)=\int_{-\infty}^{+\infty}P(A|X=x)f_X(x)dx}$ **(c)**；上述三式均可称作<u>连续型（广义）全概率公式</u>，其实质是离散型全概率公式的连续版本！

（4）在联合密度$f(x,y)$中，x与y的地位是“对等”的，它们都是变量，故x与y的变化范围等都并列地标注在$f(x,y)$表达式后面；而在条件密度$f_{X|Y}(x|y)\left[f_{Y|X}(y|x)\right]$中，$x$与$y$的地位就不对等了：$\{y|f_Y(y)>0\}\left[\{x|f_X(x)>0\}\right]$是$f_Y(y)>0\left[f_X(x)>0\right]$的范围，它是$f_{X|Y}(x|y)$ $\left[f_{Y|X}(y|x)\right]$存在的“前提条件”，必须标在前面，$f_{X|Y}(x|y)\left[f_{Y|X}(y|x)\right]$只是$x[y]$的函数，故后面只能标注$x[y]$的变化范围，此时的$y[x]$只能视作“常数”。

【例 3.4.4】设$X_1,X_2,\cdots,X_n$独立同$U(0,1)$分布，(a)在$X_{(n)}=y,0<y<1$条

件下，试求 $X_{(1)}$ 的概率密度；**(b)**试求$\left(X_{(1)},X_{(2)},\cdots X_{(n-1)}\right)\Big|_{X_{(n)}=y}$ 的联合密度；

【例 3.4.5】设 X_1,X_2 独立，且 $X_i\sim E(\lambda_i),i=1,2,\forall t>0,$ 试求：

$$\boxed{P\left(X_1<X_2\middle|X_1\wedge X_2=t\right)}\quad;$$

【例 3.4.6】设随机变量 X 与 U 相互独立，X 的密度函数为 $p(x)$，U 服从$[0,1]$上的均匀分布，又函数 $q(x)$ 满足条件：

（1） $q(x)\geq 0$，且 $\int_{-\infty}^{+\infty}q(x)dx=1$；

（2）存在 $a>0$，使得 $\dfrac{p(x)}{q(x)}\geq a$ （当 $p(x)>0$ 时），令 $r(x)=a\dfrac{q(x)}{p(x)}$ （当 $p(x)=0$ 时，规定 $r(x)=0$）；又记 $M=\{U\leq r(X)\}$，试证明：$P(X\leq z|M)=\int_{-\infty}^{z}q(x)dx$，即 X 在 M 发生的条件下的条件密度函数恰是 $q(x)$；

【例 3.4.7】设 $r.v.X\sim U(0,1)$ ，当观察到 $X=x(0<x<1)$ 时，Y 是 $(x,1)$ 上的均匀分布，求： **(1)** (X,Y) 的 $J.p.d.f.$； **(2)** Y 的 $p.d.f.$；

【例 3.4.8】设 $(X,Y)\sim U(D),D:x^2+y^2\leq R^2$，试求：

（1） $f_{Y|X}(y|x)$；（2） $P\left(Y>0\middle|X=\dfrac{R}{2}\right)$；（3） $P\left(Y>0\middle|X>\dfrac{R}{2}\right)$；

【例 3.4.9】设 X,Y 相互独立，且 $P(\{X=1\})=0.3,P(\{X=2\})=0.7$，$Y\sim f_Y(y)$，试求 $Z=X+Y$ 的分布；（分别使用离散型和连续型全概率公式求解）

【例 3.4.10】设 $X\sim B(n,Y)$，其中 $Y\sim U(0,1)$ ，试求 X 的分布；

【例 3.4.11】设 $X_1,X_2,\cdots\cdots,X_n,i.i.d.$，且 $X_i\sim f(x)$，试求

$$\boxed{P\left(\{X_n>\max\left(X_1,X_2,\cdots\cdots,X_{n-1}\right)\}\right)}\quad;$$

【例 3.4.12】设 $r.v.X\sim U(0,1)$，$\forall x\in(0,1)$，当 $X=x$ 时，$Y\sim E\left(\dfrac{x}{2}\right)$，试求 $P(\{X\geq Y\})$ ；

【例 3.4.13】设 $(X,Y)\sim f(x,y)=\dfrac{1}{x}I_{\{0<y<x<1\}}(x,y)$，试求：（1） $f_{Y|X}(y|x)$ ；

（2） $P\left(X^2+Y^2\leq 1\middle|X=x\right),0<x<1$ ，并由此求 $P\left(\{X^2+Y^2\leq 1\}\right)$；

【例 3.4.14】在圆周上任取三点 A,B,C，试求$\{\Delta ABC为锐角三角形\}$的概率；

【例 3.4.15】假设甲，乙两个产品的寿命为X,Y，分别服从参数为λ,μ的指数分布，假定X,Y是独立的，试求$P(\{X<Y\})$；

【例 3.4.16】设随机变量X,Y相互独立，且X有分布函数$F(x)$和概率密度$f(x)=F'(x)$；如果$P(\{Y>y\})=\left[P(\{X>y\})\right]^{\beta}$，$\beta$是正常数，试求$P(\{X\ge Y\})$；

【条件数学期望 1】设(X,Y)为二维连续型随机向量，$f_Y(y)$为Y的概率密度，$f_{X|Y}(x|y)$为给定$Y=y$时X的条件概率密度；若$\int_{-\infty}^{+\infty}|x|f_{X|Y}(x|y)dx<+\infty$，则称$\boxed{E(X|Y=y)=\int_{-\infty}^{+\infty}xf_{X|Y}(x|y)dx}$为给定$Y=y$时$X$的条件数学期望。

【例 3.4.17】从$t=0$开始，一部手机等待第一个短信到达的时间X服从参数为λ的指数分布；已知$X=x$的条件下，第二个短信的到达时间Y有概率密度$g(y)=\lambda e^{-\lambda(y-x)}$，$y>x$；**(a)**计算$(X,Y)$的联合密度；**(b)**试求$Y$的概率密度；**(c)**已知$Y=y$时，计算$X$的概率分布和数学期望；

【例 3.4.18】保险公司一项理赔损失X具有分布密度$f(x)=\frac{3}{8}x^2$，$0\le x\le 2$，假定处理损失为x的一项理赔需要花费的时间T（单位：月）服从$U[x,2x]$，则“处理一项理赔的时间需3个月以上”的概率是多少？

【条件数学期望 2】若$P(\{Y\in B\})>0,B\in\beta$，则给定$Y\in B$时$X$的条件分布函数为：$\forall x\in R,P(X\le x|Y\in B)=\dfrac{P(\{X\le x,Y\in B\})}{P(\{Y\in B\})}$；若$\forall B_1\in\beta$，存在非负可积函数$f_{X|Y\in B}(x)$，满足：$P(X\in B_1|Y\in B)=\int_{B_1}f_{X|Y\in B}(x)dx$，则称$f_{X|Y\in B}(x)$为给定$Y\in B$时$X$的条件密度函数； 若再有：$\int_{-\infty}^{+\infty}|x|f_{X|Y\in B}(x)dx<\infty$，则称$\int_{-\infty}^{+\infty}xf_{X|Y\in B}(x)dx$为给定$Y\in B$时$X$的条件数学期望，记作：$E(X|Y\in B)$；即有：$\boxed{E(X|Y\in B)=\int_{-\infty}^{+\infty}xf_{X|Y\in B}(x)dx}$。

【例 3.4.19】设$r.v.X\sim N(2,1^2),B=\{X\ge 0\}$，求$X$在$B=\{X\ge 0\}$下的条件数学期望；

【例 3.4.20】设$r.v.\xi,\eta$独立同$E(\lambda)$分布，令$X=\begin{cases}3\xi+1,\xi\ge\eta;\\6\eta,\xi<\eta;\end{cases}$，试求$EX$；

【例 3.4.21】设随机变量X,Y独立同$U[0,1]$分布，令

$Z=\begin{cases} X+Y, 0\le X+Y\le 1; \\ (X+Y)-1, 1< X+Y\le 2; \end{cases}$，试求 Z 的分布及 EZ；

【例 3.4.22】设 X_1,X_2 独立，且 $X_i\sim E(\lambda_i), i=1,2$，试求 $E(X_1|X_1<X_2)$；

【注 2】指数分布有一个有趣性质：$\boxed{E(X_1|X_1\le X_2)=E[\min(X_1,X_2)]}$；这个性质甚至可推广为：若 $X_1,X_2,\cdots,X_n$ 独立，且 $X_i\sim E(\lambda_i), i=1,2,\cdots,n$；则 $\min(X_1,X_2,\cdots,X_n)\sim E\left(\sum_{i=1}^{n}\lambda_i\right)$；学生试做：

设 X_1,X_2,X_3 独立，$X_i\sim E(\lambda_i), i=1,2,3$；试求 $E(X_2|X_1\le X_2\le X_3)$；

由上述定义，条件数学期望 $E(X|Y=y)$ 是 X 局限于 $\{\omega: Y(\omega)=y\}$ 上的局部加权平均，它是 y 的“函数”，不妨记之为 $g(y)$，即有：$g(y)=E(X|Y=y), y\in R(Y)$；类似于离散情形的条件数学期望，定义一个新的随机变量 $E(X|Y)=E(X|Y)(\omega)$，使其在 $\omega\in\{\omega|Y(\omega)=y\}$ 时取值为 $g(y)$，不妨记之为：$\boxed{g(Y)=E(X|Y)=E(X|Y)(\omega)}$，称之为 X 关于 Y（或给定 Y 时 X）的条件数学期望！

【注 3】上述条件期望为一随机变量，故可取期望：

$$E[E(X|Y)]=E[g(Y)]=\int_{-\infty}^{+\infty}g(y)f_Y(y)dy=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}xf_{X|Y}(x|y)dx\right]f_Y(y)dy$$

$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_{-\infty}^{+\infty}x(\int_{-\infty}^{+\infty}f(x,y)dy)dx=\int_{-\infty}^{+\infty}xf_X(x)dx=EX$；即为连续情形的全（数学）期望公式！事实上，上述公式中的 X,Y 可以是任意类型的随机变量，只须 X 的期望存在即可（条件期望存在的充分条件）！由上述公式也可导出连续型（广义）全概率公式，这里不再累述。

【例 3.4.23】设 $r.v.\xi_1\sim U(0,1)$，且 $\forall k\ge 1, \xi_{k+1}|_{\xi_k=x_k}\sim U(x_k,1)$，试求 $E(\xi_n)$；

【例 3.4.24】从一个含有 a 只黑球和 b 只白球的袋中任取 n 只球，（$1\le n\le a+b$）再从这 n 只球中任取 1 球，求“该球是黑球”的概率；

【例 3.4.25】设 $r.v.\eta\sim U(0,a), X\sim U[\eta,a]$，则 $E(X|\eta)\sim U\left(\frac{a}{2},a\right)$；

【例 3.4.26】设供电公司每月可供应某工厂的电力服从 $[10,30]$（单位：万度）上的均匀分布，该工厂每月实际生产所需电力服从 $[10,20]$ 上的均匀分布。若工厂能从供电公司得到充足的电力，每 1 万度电可创造 30 万元的利润；若得不到足够的电力，不

足部分通过其他途径解决，此时每1万度电产生10万元的利润；求工厂每月的平均利润。（三种方法）

【例 3.4.27】若随机变量 X,Y 满足：$E(X|Y)=Y,a.s.,E(Y|X)=X,a.s.$，则 $X=Y,a.s.$；

【例 3.4.28】若 (X,Y) 服从 $U(D)$ 分布，$D=[0,1]^2$，试求 $E(X|XY)$；

【例 3.4.29】设随机序列 $\{\xi_n,n\ge 1\}$ 独立同 $U[0,1]$ 分布，令 $\tau=\min\{k:\xi_1+\cdots+\xi_k>1\}$，则有 $E\tau=e$；

【例 3.4.30】（1）设 $X_1,X_2,\cdots,X_n$ 独立同分布，$S_n=X_1+X_2+\cdots+X_n$，如果 $P(\{X_i>0\})=1$，则 $E(X_i|S_n)=\frac{1}{n}\sum_{j=1}^{n}X_j$；（2）设 $X_1,X_2,\cdots,X_n$ 独立同分布，$S_k=X_1+X_2+\cdots+X_k$，$k\le n$，如果 $E(|X_1||S_n)$ 存在，试求 $E(S_k|S_n)$；

以下，我们通过例题来讨论“基于条件期望的直观定义”来计算条件期望的方法！

【例 3.4.31】1）设 X,Y 独立同 $E(1)$ 分布，试求 $E[\cos(X+Y)|X]$；2）设 $(X,Y)\sim U(D)$，$D:x^2+y^2\le 1,x,y\ge 0$，试求 $E(X|Y)$；3）设 $(X,Y)\sim U(D)$，$D=\{(x,y):0\le y\le x\le 1\}$，求 $E(X|Y)$、$E(Y|X)$、$E[(X-Y)^2|X]$；4）设 X,Y 独立同 $U[0,1]$ 分布，试求 $E[X^2|X+Y]$；5）设 X,Y 独立同 $U[0,1]$ 分布，试求 $E(X|X<Y)$；6）设 $X_1,X_2,\cdots,X_n$ 独立同 $U[0,1]$ 分布，且 $T=\max\{X_1,X_2,\cdots,X_n\}$，试求 $E(X_1|T)$。

【例 3.4.32】设随机向量 (X,Y) 以概率0.4取（向量）值 $(0.75,0.25)$，以概率0.6均匀取（向量）值于 $[0,1]^2$；试求 $E(X|Y)$，并由此求 EX；

§5 随机向量的函数、随机变量的函数和随机向量的向量值函数的分布

【随机向量函数的分布】设 $(X,Y)\sim f(x,y)$，$g(\cdot,\cdot)$ 为二元 **Borel** 函数，$Z=g(X,Y)$，考虑 Z 的分布；$\forall z\in R$，$F_Z(z)=P(\{Z\le z\})=$ $P(\{g(X,Y)\le z\})=\iint\limits_{g(x,y)\le z}f(x,y)dxdy$；以下，仅讨论一些简单情形：

（1）$g(x,y)=x+y,Z=g(X,Y)=X+Y$；（3）$g(x,y)=xy,Z=g(X,Y)=XY$；

（2）$g(x,y)=x-y,Z=g(X,Y)=X-Y$；（4）$g(x,y)=\frac{x}{y},Z=g(X,Y)=\frac{X}{Y}$；

一．求 $Z=X+Y$ 的分布；设 $(X,Y)\sim f(x,y)$，则，

法 1 ）$\forall z\in R, F_Z(z)=P(\{Z\le z\})=P(\{X+Y\le z\})=\int_{-\infty}^{+\infty}P(X+Y\le z|X=x)f_X(x)dx$
$=\int_{-\infty}^{+\infty}P(Y\le z-x|X=x)f_X(x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x}f_{Y|X}(y|x)dyf_X(x)dx$
$=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x}f(x,y)dydx=\int_{-\infty}^{+\infty}\int_{-\infty}^{z}f(x,u-x)dudx=\int_{-\infty}^{z}\int_{-\infty}^{+\infty}f(x,u-x)dxdu$；由 z 的任意性，可知：$Z=X+Y\sim f_{X+Y}(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\int_{-\infty}^{+\infty}f(z-y,y)dy$；

法 2 ）$\forall z\in R, F_Z(z)=P(\{Z\le z\})=P(\{X+Y\le z\})=\iint\limits_{x+y\le z}f(x,y)dxdy$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x}f(x,y)dydx=\int_{-\infty}^{+\infty}\int_{-\infty}^{z}f(x,u-x)dudx$$，后略！

法 3 ）$\forall a,b\in R, a<b, P(\{a<X+Y\le b\})=\iint\limits_{a<x+y\le b}f(x,y)dxdy$，后类似略！

至此，我们有：$(X,Y)\sim f(x,y)$，则有，
$X+Y\sim f_{X+Y}(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\int_{-\infty}^{+\infty}f(z-y,y)dy$；特别地，若 X,Y 相互独立，则有，$X+Y\sim f_{X+Y}(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx=\int_{-\infty}^{+\infty}f_X(z-y)f_Y(y)dy$；

【例 3.5.1】设 X,Y 独立同 $U(0,1)$ 分布，试求 $X+Y$ 的分布；

【例 3.5.2】设 $(X,Y)\sim f(x,y)=(2-x-y)I_{\{0<x,y<1\}}(x,y)$；试求：
（1）$P(\{X\ge 2Y\})$；（2）$Z=X+Y$ 的概率密度 $f_Z(z)$；

【例 3.5.3】设 $X_1,X_2,\cdots,X_n$ 独立同 $E(\lambda)$ 分布，试求 $X_1+X_2+\cdots+X_n$ 的分布；

【例 3.5.4】设 $\{X_n,n\ge 1\}$ $i.i.d.$ 且 $X_i\sim E(\lambda)$，记 $N=\min\left\{n\left|\sum_{i=1}^{n}X_i>1,n\ge 1\right.\right\}$，试求 N 的分布；

二.求 $Z=X-Y$ 的分布；设 $(X,Y)\sim f(x,y)$，则有，$\forall z\in R$，
$F_Z(z)=P(\{Z\le z\})=P(\{X-Y\le z\})=\iint\limits_{x-y\le z}f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{x-z}^{+\infty}f(x,y)dydx$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{z}f(x,x-u)dudx=\int_{-\infty}^{z}\int_{-\infty}^{+\infty}f(x,x-u)dxdu$$；从而有，

$X-Y\sim\int_{-\infty}^{+\infty}f(x,x-z)dx=\int_{-\infty}^{+\infty}f(z+y,y)dy$；特别地，若 X,Y 独立，则有，
$X-Y\sim\int_{-\infty}^{+\infty}f_X(x)f_Y(x-z)dx=\int_{-\infty}^{+\infty}f_X(y+z)f_Y(y)dy$；

【例 3.5.5】设 $(X,Y)\sim f(x,y)=3xI_{\{0<y<x<1\}}(x,y)$；试求 $X-Y$ 的概率密度函数；

【例 3.5.6】设 $r.v.X\sim E(\lambda),Y\sim U(0,h),h>0$，$X$ 与 Y 独立，试求 $X-Y$ 的概率密度函数；

三．求 $Z=XY$ 的分布；设 $(X,Y)\sim f(x,y)$，则有，$\forall z>0$，

$F_Z(z)=P(\{Z\le z\})=P(\{XY\le z\})=\iint\limits_{xy\le z} f(x,y)dxdy=\cdots\cdots=\int_{-\infty}^{z}\left[\int_{-\infty}^{+\infty} f\left(x,\frac{u}{x}\right)dx\right]du$；

从而有，$XY\sim f_{XY}(z)=\int_{-\infty}^{+\infty} f\left(x,\frac{z}{x}\right)\frac{1}{|x|}dx$；特别地，若 X,Y 独立，则有，

$XY\sim f_{XY}(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y\left(\frac{z}{x}\right)\frac{1}{|x|}dx$ ； $z<0$ 的情形类似！

【例 3.5.7】设 X,Y 独立同 $U(0,1)$ 分布，试求 XY 的分布；

四．求 $Z=\frac{X}{Y}$ 的分布；设 $(X,Y)\sim f(x,y)$，则有，$\forall z>0$ ，

$F_Z(z)=P(\{Z\le z\})=P\left(\left\{\frac{X}{Y}\le z\right\}\right)=\iint\limits_{\frac{x}{y}\le z} f(x,y)dxdy=\int_{-\infty}^{z}\int_{-\infty}^{+\infty} f(yu,y)|y|dydu$ ；

从而有，$\frac{X}{Y}\sim f_{\frac{X}{Y}}(z)=\int_{-\infty}^{+\infty} f(yz,y)|y|dy$；特别地，若 X,Y 独立，则有，

$\frac{X}{Y}\sim f_{\frac{X}{Y}}(z)=\int_{-\infty}^{+\infty} f_X(yz)f_Y(y)|y|dy$ ；

【例 3.5.8】设 ξ,η 独立，且 $\xi\sim U(1,3),\ \eta\sim f_\eta(y)=e^{-(y-2)}I_{\{y>2\}}(y)$，试求 $\frac{\xi}{\eta}$ 的分布；

【随机变量函数的分布】设 $X\sim f(x),Y=g(X)$ ，$g(\cdot)$ 为一元 **Borel** 函数，则 $\forall y\in R$，$F_Y(y)=P(\{Y\le y\})=P(\{g(X)\le y\})=\int\limits_{\{x|g(x)\le y\}} f(x)dx$ ；

即得 Y 的分布；

【引例】设 $X\sim U(0,1)$，$\Phi^{-1}(x)$ 是 $\Phi(x)$ 的反函数，试求 $Y=\Phi^{-1}(X)$ 的分布；

以下介绍两个重要定理；

【定理 1】设 $c.r.v.X\sim f_X(x)$ ，若 $y=g(x)$ 是严格单调函数且可导，则 $Y=g(X)$ 是一个连续型随机变量，且有概率密度函数为：

$\boxed{f_Y(y)=f_X[h(y)]|h'(y)|I_{\{\alpha<y<\beta\}}(y)}$，其中：$h(y)$ 是 $g(x)$ 的反函数，

$\alpha=\min\{g(-\infty),g(+\infty)\},\beta=\max\{g(-\infty),g(+\infty)\}$ ， $g(-\infty)=\lim\limits_{x\to-\infty}g(x)$，$g(+\infty)=\lim\limits_{x\to+\infty}g(x)$；（课堂论证）

【定理 2】设 $c.r.v.X\sim f_X(x),D\subset R,Y=g(X),P(\{Y\in D\})=1$，如果：

1）$\forall y\in D,\{Y=y\}=\bigcup\limits_{i=1}^{n}\{X=h_i(y)\}$，即：$Y^{-1}(D)=\bigcup\limits_{i=1}^{n}X^{-1}[h_i(D)]$；

2）每个 $h_i(y)$ 都是 D 到某区域 D_i 的可逆映射（函数），且在 D 内有

连续的导数；

3） $D_1, D_2, \cdots\cdots, D_n$ 互不相交；则 Y 有如下的概率密度函数：

$$\boxed{f_Y(y)=\sum_{i=1}^{n} f_X\left[h_i(y)\right]\left|h_i^{'}(y)\right| I_{D}(y)}$$ ；（课堂证明）

【例 3.5.9】设 $X \sim U\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$，求 $Y=\tan X$ 的概率密度函数；

【例 3.5.10】设 ξ 与 η 独立同 $U(0,1)$ 分布，试求 $3\xi+2\eta$ 的密度函数；

【例 3.5.11】设 $\xi \sim f_\xi(x)=2(1-x)I_{\{0<x<1\}}(x)$，求一单调增函数 $h(x)$，使得 $\eta=h(\xi)\sim E(\lambda)$；

【例 3.5.12】设 $X \sim U\left[-\frac{\pi}{2},\frac{3\pi}{2}\right]$， $Y=X^4$，求 Y 的密度函数；

【例 3.5.13】设 $\xi \sim U\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$， $\eta=\cos\xi$，求 η 的 $p.d.f.$ ；

【例 3.5.14】 $\xi \sim f_\xi(x), -\infty<x<+\infty, \eta=\sin\xi$，试由分布函数法和密度公式分别求出 η 的概率密度；

【<u>随机向量的向量值函数的分布</u>】考虑二重积分的如下性质：对于二重积分 $\iint_D f(x,y)dxdy$，作变量代换： $x=x(u,v), y=y(u,v)$；

则：<u>被积函数</u>　$f(x,y)\to f\left[x(u,v),y(u,v)\right]=h(u,v)$

　　<u>积分区域</u>　$D(\subset xoy\text{平面})\to Q(\subset uov\text{平面})$　；记此变换为

$g:(x,y)\to(u,v)$　；若 $g\in C^1(D)$，且 $\left|\frac{D(x,y)}{D(u,v)}\right|=\left\|\frac{\partial(x,y)}{\partial(u,v)}\right\|\neq 0$，则，

$$\boxed{\iint_D f(x,y)dxdy=\iint_Q f\left[x(u,v),y(u,v)\right]\left|\frac{D(x,y)}{D(u,v)}\right|dudv}$$ 。由此， $\forall D\subset xoy$ 平面，在映射 g（一一映射，可微）的作用下，若有 $g: D\to Q\subset uov$ 平面，则： $P\left(\{(X,Y)\in D\}\right)=P\left(\{(U,V)\in Q\}\right)$；这里， $g: X=X(U,V), Y=Y(U,V)$，即有： $\forall Q\subset uov\text{平面}, g^{-1}(Q)=D\subset xoy\text{平面}$，

$$P\left(\{(U,V)\in Q\}\right)=P\left(\{(X,Y)\in D\}\right)=\iint_D f(x,y)dxdy=\iint_Q f\left[x(u,v),y(u,v)\right]\left|\frac{D(x,y)}{D(u,v)}\right|dudv$$

；由 Q 的任意性，即知：

$$\boxed{(U,V)\sim f\left[x(u,v),y(u,v)\right]\left\|\frac{D(x,y)}{D(u,v)}\right\|}$$ ；

【例 3.5.15】设 X,Y 独立同 $E(1)$ 分布，令 $U=X+Y, V=\frac{X}{Y}$，试求 (U,V) 的联合密度函数，并验证 U,V 是独立地；

【例 3.5.16】设 X,Y 独立同 $E(\lambda)$ 分布，求 $\left(X+Y,\dfrac{X}{X+Y}\right)$ 的联合密度及两个边缘密度；

【例 3.5.17】若随机变量 X,Y 独立，且 $X\sim E(\lambda),Y\sim U(0,h)$，求 $X+Y,X-Y$ 的密度函数；

【例 3.5.18】若随机变量 X,Y 独立，且服从 $N(0,\sigma^2)$ 分布，试证明：X^2+Y^2 与 $\dfrac{X}{Y}(Y\neq 0)$ 独立；

【例 3.5.19】设二维随机向量 $(X,Y)\sim f(x,y)=\dfrac{4}{7}(1+x+y)I_{\{0<x,y<1\}}(x,y)$，求 $Y(1+X)$ 的密度；

【例 3.5.20】设 ξ 和 η 独立同 $U(0,1)$ 分布，令 $U=\sqrt{-2\ln\xi}\cos(2\eta\pi)$，$V=\sqrt{-2\ln\xi}\sin(2\eta\pi)$，证明：$U$ 和 V 独立同 $N(0,1^2)$ 分布；

【例 3.5.21】设随机变量 ξ 和 η 的联合密度函数形如 $f(x,y)=h\left(\sqrt{x^2+y^2}\right)$，1）试求随机变量 $R=\sqrt{\xi^2+\eta^2}$ 和 $S=\arctan\left(\dfrac{\eta}{\xi}\right)$ 的联合密度函数，并证明 R 和 S 独立；2）设 $U=\xi\cos\alpha+\eta\sin\alpha$ 与 $V=-\xi\sin\alpha+\eta\cos\alpha$，试证明：随机向量 (U,V) 的联合密度函数与 $f(x,y)$ 相同。

【例 3.5.22】设随机向量 (ξ,η) 服从单位圆上的均匀分布，又设 $W=\xi^2+\eta^2$，$U=\xi\sqrt{\dfrac{-2\ln W}{W}}$，$V=\eta\sqrt{\dfrac{-2\ln W}{W}}$，1）证明：随机变量 W 与随机变量 $Z=\arctan\left(\dfrac{\eta}{\xi}\right)$ 相互独立，分别服从区间 $[0,1]$ 与 $[0,2\pi]$ 上的均匀分布；2）证明：随机变量 U 和 V 相互独立且皆服从标准正态分布；

【例 3.5.23】设 X,Y 独立同 $E(\lambda)$ 分布，试求 $\boxed{E(X|X+Y=z),z>0}$；

§6 概率不等式、协方差和相关系数、二元正态分布

【概率不等式】以下仅讨论 **Markov** 不等式、切比雪夫不等式和柯西（**Cauchy**）不等式。

【马尔可夫（**Markov**）不等式】

（一）若非负随机变量 X 的期望 EX 存在，则 $\forall\varepsilon>0$，

1）若 X 是离散型随机变量且有分布列：$P(\{X=x_i\})=p_i,i=1,2,\cdots$，

$$EX=\sum_i x_i P(\{X=x_i\})\ge\sum_{i:x_i\ge\varepsilon}x_iP(\{X=x_i\})\ge\sum_{i:x_i\ge\varepsilon}\varepsilon P(\{X=x_i\})\ge\varepsilon P\left(\bigcup_{i:x_i\ge\varepsilon}\{X=x_i\}\right)$$

$=\varepsilon P(\{X\ge\varepsilon\})$ ；也即： $P(\{X\ge\varepsilon\})\le\frac{EX}{\varepsilon}$ ；

2）若 X 是连续型随机变量且有密度函数 $f(x)$，

$EX=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^{+\infty}xf(x)dx\ge\int_\varepsilon^{+\infty}xf(x)dx\ge\int_\varepsilon^{+\infty}\varepsilon f(x)dx=\varepsilon P(\{X\ge\varepsilon\})$；也即：

$$P(\{X\ge\varepsilon\})\le\frac{EX}{\varepsilon};$$

（二）若随机变量 X 满足： $E\left(|X|^k\right)<+\infty$ ，则 $\forall\varepsilon>0$，即有：

$P\left(\left\{|X|^k\ge\varepsilon^k\right\}\right)\le\frac{E\left(|X|^k\right)}{\varepsilon^k}$，也即： $P(\{|X|\ge\varepsilon\})\le\frac{E\left(|X|^k\right)}{\varepsilon^k}$ ；

【切比雪夫不等式】将马尔可夫不等式中的 $|X|^k$ 置换成 $(X-c)^2$，

即有： $P(\{|X-c|\ge\varepsilon\})\le\frac{E\left(|X-c|^2\right)}{\varepsilon^2}$ ；它有如下常见的形式：

（1） $P(\{|X-c|\ge\varepsilon\})\le\frac{E\left((X-c)^2\right)}{\varepsilon^2}$ ；

（2）取 $c=EX$ ，即有： $P(\{|X-EX|\ge\varepsilon\})\le\frac{E\left((X-EX)^2\right)}{\varepsilon^2}=\frac{DX}{\varepsilon^2}$ ；

（3）取 $c=EX$ ， $\varepsilon=k\sigma=k\sqrt{DX}$ ，即有： $P(\{|X-EX|\ge k\sigma\})\le\frac{1}{k^2}$；

【注 1】切比雪夫不等式常用来：

a)证明切比雪夫大数定律；　　**b)**估算概率的上，下界；

【例 3.6.1】若随机变量 $X\sim U(-1,3)$,试将由切比雪夫不等式得到的概率 $P\left(\left\{|X-EX|\ge2\sqrt{DX}\right\}\right)$ 的上界与其精确值作比较；

【例 3.6.2】设 $r.v.X\sim f(x)=\frac{x^ne^{-x}}{n!}I_{\{x>0\}}(x)$, 即： $X\sim\Gamma(n+1,1)$ ；试证：

$$P(\{0<X<2(n+1)\})\ge\frac{n}{n+1};$$

【柯西不等式（**Cauchy-Schwarz**）】见过的柯西不等式版本有：

（1） $\left(a_1^2+a_2^2+\cdots+a_n^2\right)\left(b_1^2+b_2^2+\cdots+b_n^2\right)\ge\left(a_1b_1+a_2b_2+\cdots+a_nb_n\right)^2$ ；

（2） $f(x),g(x)\in R[a,b],\int_a^bf^2(x)dx\int_a^bg^2(x)dx\ge\int_a^bf(x)g(x)dx$ ；

（3） $\alpha,\beta\in V,|(\alpha,\beta)|\le\|\alpha\|\|\beta\|$，这里，$V$ 是向量空间，$(\cdot,\cdot),\|\cdot\|$ 分别是向量的内积与模；

（4）【概率版本】设 $E(X^2),E(Y^2)<+\infty$，则 $[E(XY)]^2\le E(X^2)E(Y^2)$；

【注 2】设 X 为随机变量，则 $DX=0 \Leftrightarrow P(\{X=EX\})=1$（证明）；

由上述【注 2】可得到柯西（内积）不等式等号成立的条件：

$\forall t\in R, E\left[(tX+Y)^2\right]\ge 0$， $E(X^2)t^2+2E(XY)t+E(Y^2)\ge 0$， 从而，

$[2E(XY)]^2-4E(X^2)E(Y^2)\le 0$，等号成立当且仅当存在唯一的实数 t_0，使得 $E\left[(t_0X+Y)^2\right]=0$， $E(t_0X+Y)=0$， $D(t_0X+Y)=0$， 从而有

$P(\{t_0X+Y=E(t_0X+Y)\})=P(\{t_0X+Y=0\})=1$，（置于向量空间【线性相关，内积】的范畴来解释等号成立的充要条件）。

【协方差】设 (X,Y) 为二维随机向量，若 $E|(X-EX)(Y-EY)|<\infty$，则称 $E[(X-EX)(Y-EY)]$ 为随机变量 X 与 Y 的协方差（**Covariance**），记作： $Cov(X,Y)$， 即： $Cov(X,Y)=E[(X-EX)(Y-EY)]$；

由期望的线性性质，即有： $Cov(X,Y)=E(XY)-EXEY$； 由上述定义，我们有：

（1）若 X,Y 是联合离散的，且有联合分布列：

$$P(\{(X,Y)=(x_i,y_j)\})=p_{ij}, i=1,2,\cdots; j=1,2,\cdots;$$ 则

$$Cov(X,Y)=\sum_i\sum_j(x_i-EX)(y_j-EY)p_{ij};$$

（2）若 X,Y 是联合连续的，且有联合分布密度 $f(x,y)$；则

$$Cov(X,Y)=\iint_{R^2}(x-EX)(y-EY)f(x,y)dxdy;$$ 若 $Cov(X,Y)=0$，称 X,Y 不相关或零相关！

通常，协方差 $Cov(X,Y)$ 有如下一些性质：

1. （对称性） $Cov(X,Y)=Cov(Y,X)$； **2.** $Cov(X,X)=DX$；

3. $Cov(aX,bY)=abCov(X,Y)$； **4.** $Cov(X+c,Y+d)=Cov(X,Y)$；

5. $Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$； **6.** $Cov(X,a)=0, a\in const$；

7. $D(X\pm Y)=DX+DY\pm 2Cov(X,Y)$，并且更一般地，

$$\boxed{D\left(\sum_{i=1}^n a_iX_i\right)=\sum_{i=1}^n a_i^2DX_i+2\sum_{1\le i<j\le n}a_ia_jCov(X_i,X_j)};$$

【例 3.6.3】设 $X\sim f(x)$（偶函数），且 $E(|X|^3)<\infty$，试证： X 与 $Y=X^2$ 不相关但也不独立；

【例 3.6.4】设 $(X,Y)\sim U(D)$，其中 $D=\{(x,y)|x^2+y^2\le 1\}$，试证： X,Y 既不相关也不独立；

【注 3】协方差的数值虽然一定程度上反映了 X 与 Y 之间的相互关系，但它仍然受 X,Y 本身数值（量纲）大小的影响，再如：直观上，kX 与 kY（k 为常数）之间的联系和 X 与 Y 之间的联系并无二致，但 $Cov(kX,kY)=k^2Cov(X,Y)$，协方差却增加了 k^2 倍；为了克服以上不足，先对随机变量“标准化”。所谓“标准（化）随机变量”，即指 $r.v.X$ 满足 $EX=0,DX=1$；对于任何随机变量 X，若 $EX=\mu,DX=\sigma^2$，只须令：$\boxed{Y=\frac{X-EX}{\sqrt{DX}}=\frac{X-\mu}{\sigma}}$，则有 $EY=0,DY=1$，于是 ，则称 Y 是 X 的标准化，这也是 σ 被称为“标准差”的一个原因。

【（线性）相关系数】设 (X,Y) 是一个二维随机向量，若 DX,DY 存在，则称 $Cov\left(\frac{X-EX}{\sqrt{DX}},\frac{Y-EY}{\sqrt{DY}}\right)=E\left[\left(\frac{X-EX}{\sqrt{DX}}\right)\left(\frac{Y-EY}{\sqrt{DY}}\right)\right]=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$ 为 X 与 Y 的（线性）相关系数，记之为：$\rho,\rho_{X,Y},\rho(X,Y)$ ；

【注 4】（1）相关系数消除了计量单位对随机变量的影响，因此它比协方差更能精确地刻划随机变量之间的关系；（2）经过标准化处理的 $r.v.X^*=\frac{X-EX}{\sqrt{DX}}$ 是把原分布中心 EX 移至原点，不使分布中心偏左或偏右，然后缩小或扩大坐标轴，使分布不致过疏或过密；在排除这些干扰后，原随机变量的一些性质就会显现出来，故标准化处理技术在概率统计中会经常使用。

由定义，相关系数具有如下性质：

1.（对称性） $\rho(X,Y)=\rho(Y,X)$；　**2.** $\rho(X,Y)=\rho(aX,bY)$ ；

3. $\rho(X,Y)=\rho(X+c,Y+d)$ ；　**4.** $|\rho(X,Y)|\le 1$ ；

性质 **4.**可由柯西不等式加以验证：

$$\boxed{|\rho(X,Y)|=\left|Cov\left(\frac{X-EX}{\sqrt{DX}},\frac{Y-EY}{\sqrt{DY}}\right)\right|=\left|E\left[\left(\frac{X-EX}{\sqrt{DX}}\right)\left(\frac{Y-EY}{\sqrt{DY}}\right)\right]\right|\le\sqrt{E\left(\frac{X-EX}{\sqrt{DX}}\right)^2E\left(\frac{Y-EY}{\sqrt{DY}}\right)^2}=1}$$

其等号成立的条件为等价于柯西不等式等号成立的条件，稍后再作讨论。可以验证以下四个命题等价：

1. $Cov(X,Y)=0$ ；　**2.** $\rho(X,Y)=0$ ；

3. $E(XY)=EXEY$ ；　**4.** $D(X\pm Y)=DX+DY$ ；

【定理】设 $r.v.X,Y$ 的方差存在，其相关系数为 ρ_{XY}，则：

$|\rho_{XY}|=1\Leftrightarrow X$ 与 Y 以概率 1 线性相关，即存在非零实数 a 与实数 b，

使得 $P(\{Y=aX+b\})=1$ ；易见，

$|\rho_{XY}|=1 \Leftrightarrow \left|E\left(\frac{X-EX}{\sqrt{DX}},\frac{Y-EY}{\sqrt{DY}}\right)\right|=\sqrt{E\left(\frac{X-EX}{\sqrt{DX}}\right)^2 E\left(\frac{Y-EY}{\sqrt{DY}}\right)^2} \Leftrightarrow$ 存在非零实数 a 和 b，使得 $P\left(\left\{\frac{Y-EY}{\sqrt{DY}}=a\frac{X-EX}{\sqrt{DX}}\right\}\right)=1 \Leftrightarrow P(\{Y=aX+b\})=1$，这里，$b=EY-\frac{aEY\sqrt{DY}}{\sqrt{DX}}$ ；

【注 5】 $|\rho_{XY}|=1$ 时，即除去一个零概率事件外，X 与 Y 之间存在着一个线性关系；$\rho_{XY}=1$ 时，称 X 与 Y 正相关；$\rho_{XY}=-1$ 时，称 X 与 Y 负相关；$\rho_{XY}=0$ 时，称 X 与 Y 不（零）相关；因此，相关系数是描述随机变量间线性关系强弱的一个数字特征，确切地应称为线性相关系数。

【注 6】独立与不相关的比较：若

$r.v.X,Y$ 独立 $\Rightarrow E(XY)=EXEY \Rightarrow Cov(X,Y)=0 \Rightarrow \rho(X,Y)=0$ ；即：独立必不相关；反之则不成立！这是由于 X,Y 不相关，只是说明 X,Y 之间“不存在”线性关系，但却可能存在别的函数关系，此时就可能不再相互独立（如**【例 3.6.4】**）。

【例 3.6.5】将一枚均匀硬币重复掷 n 次，并以 X 和 Y 分别表示正面和反面朝上的次数，试求 X 和 Y 的相关系数；

【例 3.6.6】设 $r.v.X_2,X_2,\cdots\cdots,X_n$ 中任两个相关系数均为 ρ，试证：

$$\rho \ge -\frac{1}{n-1} \quad ;$$

【例 3.6.7】若 $r.v.X,Y$ 满足 $EX=EY=0,VarX=VarY=1.Cov(X,Y)=\rho$，试证：$E\left[\max\left(X^2,Y^2\right)\right]\le 1+\sqrt{1-\rho^2}$ ；

【例 3.6.8】相关系数在线性变换下保持不变 $U=aX+b,V=cY+d$，则：**(1)** $ac>0$ 时，$\rho(U,V)=\rho(X,Y)$ ； **(2)** $ac<0$ 时，$\rho(U,V)=-\rho(X,Y)$ ；

【例 3.6.9】试验证：在某些特殊场合，如：两点分布，二元(维)正态分布等，“独立”与“不相关”是等价的；

【例 3.6.10】设随机变量 X,Y 满足 $D(2X+Y)=0$，试求 X,Y 的相关系数 ρ_{XY}；

【例 3.6.11】设 $X\sim f_X(x)=\begin{cases}1/2,-1\le x<0;\\ 1/4,0\le x<2;\\ 0,\text{其他};\end{cases}$，$Y=X^2$，且 $Y\sim f_Y(y)$，试求：

1） $f_Y(y)$； 2） $Cov(X,Y)$； 3） (X,Y)的 $J.d.f.$ 的值 $F\left(-\frac{1}{2},4\right)$；

【例 3.6.12】已知 $X\sim N(0,1^2)$，在 $X=x$ 时， $Y\sim N(x,1^2)$，试求 Y 的分布及 ρ_{XY}；

【例 3.6.13】设 $X\sim U[-1,1]$， $Y=|X-a|$， $a\in[-1,1]$， $\rho_{XY}=0$， 求 a；

【例 3.6.14】设 $(X,Y)\sim f(x,y)=cxe^{-y}$， $0<x<y<+\infty$； 试求 $P(X<1|Y=2)$ 及 $E[Y-(aX+b)]^2$ 的最小值；

【二元（维）正态分布】若 $(X,Y)\sim f(x,y)=$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

，其中， $-\infty<x,y<+\infty(\sigma_1,\sigma_2,>0,\mu_1,\mu_2\in R)$； 则称 (X,Y) 服从参数为 $(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$ 的二元（维）正态分布，记作：

$$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)\qquad;$$

【注 7】二元（维）正态分布有如下的典型分解式：

（1） $f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{\left[y-\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2}{2\sigma_2^2(1-\rho^2)}}$ ；

（2） $f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{\left[x-\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right]^2}{2\sigma_1^2(1-\rho^2)}}$ ；

由以上的典型分解式，易得二元（维）正态分布的如下性质：

设 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1,\sigma_2;\rho)$， 则，

1）两个边缘分布： $X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$（课堂推导）；

2）两个条件分布： $Y|_{X=x}\sim \mathrm{N}\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),(1-\rho^2)\sigma_2^2\right)$，

$X|_{Y=y}\sim \mathrm{N}\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),(1-\rho^2)\sigma_1^2\right)$（课堂推导）；由上述 2），可知，

$E(X|Y=y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$， **(a)** $E(Y|X=x)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$； **(b)**

【注 8】设 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1,\sigma_2;\rho)$，讨论 ρ 的意义； X,Y 的相关系数为： $\frac{Cov(X,Y)}{\sqrt{DXDY}}=\frac{E(XY)-\mu_1\mu_2}{\sigma_1\sigma_2}=\frac{E[E(XY|X)]-\mu_1\mu_2}{\sigma_1\sigma_2}=\frac{E[XE(Y|X)]-\mu_1\mu_2}{\sigma_1\sigma_2}$

$=\cdots\cdots=\rho$（由注 7 的(b)式）；若令二元正态密度中的 $\rho=0$，即有：

$f(x,y)=f_X(x)f_Y(y)$，从而 X,Y 独立；故在二元正态分布场合，独立与不相关等价！

【例 3.6.15】若随机向量$(X,Y)\sim f(x,y)=ke^{-\left(ax^2+2bxy+cy^2\right)},-\infty<x,y<+\infty$，在什么条件下，$X$ 与 Y 相互独立？

【例 3.6.16】 (a) 设随机向量(X,Y)服从二元正态分布 $f(x,y)=\dfrac{1}{2\pi ab}e^{-\frac{1}{2}\left(\frac{1}{a^2}x^2+\frac{1}{b^2}y^2\right)}$，试求“$(X,Y)$取值于椭圆$\dfrac{1}{a^2}x^2+\dfrac{1}{b^2}y^2=k^2$内”的概率；(b) 设随机向量$(X,Y)\sim N\left(\mu_1,\mu_2;a^2,b^2;\rho\right)$，求“$(X,Y)$取值于椭圆$\dfrac{1}{a^2}(x-\mu_1)^2-\dfrac{2\rho}{ab}(x-\mu_1)(y-\mu_2)+\dfrac{1}{b^2}(y-\mu_2)^2=k^2$内”的概率；

第四章极限定理

极限定理包括大数定律（法则）（弱大数定律与强大数定律）和中心极限定理（局部中心极限定理与积分中心极限定理）。通俗地说，<u>凡描述在一定的条件下随机序列的前若干项的算术平均值收敛于其均值的算术平均值的一类定理都可统称为大数定理</u>；<u>凡描述在一定的条件下，大量随机变量之和的极限分布是正态分布的一类定理都可统称为中心极限定理</u>！

§1 （弱）大数定律

【<u>依概率收敛</u>】设$\{X_n n\geq 1\}$为概率空间(Ω,F,P)上的随机序列，若存在随机变量X，使得：$\forall\varepsilon>0,\lim\limits_{n\to\infty}P\left(\{|X_n-X|\geq\varepsilon\}\right)=0$，或等价地 $\lim\limits_{n\to\infty}P\left(\{|X_n-X|<\varepsilon\}\right)=1$，则称随机序列$\{X_n n\geq 1\}$依概率收敛于随机变量$X$；记作：$\lim\limits_{n\to\infty}X_n=X(P)$，$P-\lim\limits_{n\to\infty}X_n=X$ 或 $X_n\xrightarrow{P}X$；若 $X_n\xrightarrow{P}X$，即有：$X_n-X\xrightarrow{P}0$。

【直观解释】$X_n\xrightarrow{P}X$ 也即$\forall\varepsilon>0$，n充分大时，事件$\{|X_n-X|\geq\varepsilon\}$发生的概率很小（收敛为$0$），这是在概率意义下的收敛性；因此，$n$很大时，有很大把握保证$X_n$很接近于$X$；

【切比雪夫大数定律】 设$\{X_n,n\geq 1\}$为$i.i.d.$随机序列，且 $EX_i=\mu,DX_i=\sigma^2$，记$S_n=\sum\limits_{i=1}^{n}X_i,\overline{X}=\dfrac{1}{n}S_n$，则$\forall\varepsilon>0,\lim\limits_{n\to\infty}P\left(\left\{\left|\overline{X}-\mu\right|\geq\varepsilon\right\}\right)=0$，即：$\overline{X}\xrightarrow{P}\mu$；有时，也将该定律的条件写为：$\{X_n,n\geq 1\}$为独立随机序列，若存在常数$c$，使得$\forall n,DX_n\leq c$，即$X_n$的方差一致有

界，则$\forall\varepsilon>0,\lim\limits_{n\to\infty}P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\right|\geq\varepsilon\right\}\right)=0$；若随机序列$\{X_n,n\geq1\}$满足该式，则称它服从（弱）大数定律！

【直观解释】当试验次数趋于无穷时，$\frac{1}{n}\sum_{i=1}^{n}X_i$依概率收敛于其期望，这也是对算术平均值稳定性的较确切的解释；它说明大数次重复试验下所呈现的客观规律，故称大数定律！

【贝努利（**Bernoulli**）大数定律】设n_A表示在n重 **Bernoulli** 试验中事件A发生的频数，且在每次试验中A发生的概率为p，则$\forall\varepsilon>0$，有：$\lim\limits_{n\to\infty}P\left(\left\{\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right\}\right)=0$，即：$\frac{n_A}{n}\xrightarrow{P}p$；

【直观解释】频率依概率收敛于概率；这也说明在大数次重复试验下，用频率去估计概率是可行的。

以上的定律都是假定随机变量的二阶矩存在且方差一致有界的前提下成立大数定律；然而有些前提却是不必要的，

【辛钦大数定律】设$\{X_n,n\geq1\}$为$i.i.d.$随机序列，且$EX_n=\mu$，则$\forall\varepsilon>0,\lim\limits_{n\to\infty}P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geq\varepsilon\right\}\right)=0$，即：$\boxed{\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu}$。

【例 4.1.1】设$\{\xi_n,n\geq1\}$独立同$U[0,a]$分布，令

$X_n=\max\{\xi_1,\xi_2,\cdots,\xi_n\},n\geq1$，则$\{X_n,n\geq1\}$依概率收敛于$a$，即：

$X_n\xrightarrow{P}a$ ；

【例 4.1.2】设$X_1,X_2,\cdots,X_n,\cdots,i.i.d.$，其$p.d.f.$为$f(x)=e^{-(x-a)}I_{\{x>a\}}(x)$，令$Y_n=\max\{X_1,X_2,\cdots,X_n\},n\geq1$，试证：$\{Y_n,n\geq1\}$依概率收敛于$a$，即：

$Y_n\xrightarrow{P}a$ ；

【例 4.1.3】设$\{X_n,n\geq1\}$为独立的随机序列，且$P\left(\left\{X=\pm3^n\right\}\right)=\frac{1}{3^{2n+2}}$，$P\left(\left\{X=0\right\}\right)=1-\frac{2}{3^{2n+2}}$ ；问：$\{X_n,n\geq1\}$是否服从大数定律？

【例 4.1.4】设$\{X_n,n\geq1\}$是$i.i.d.r.v.$序列，且$X_n\sim U(0,1),n\geq1$，若

$Y_n=\left(\prod_{k=1}^{n}X_k\right)^{\frac{1}{n}},n\geq1$，试证：$Y_n\xrightarrow{P}c$，并求出$c$；

【注1】 可以验证：若$X_n\xrightarrow{P}c$，$g(\cdot)$为连续函数，则

$g(X_n)\xrightarrow{P}g(c)$；（课堂证明）

§2 中心极限定理

【依分布收敛】设 $F_n(x), n\ge 1, F(x)$ 分别为 $X_n, n\ge 1, X$ 的分布函数，若对于 $F(x)$ 的任一连续点 x，有 $\lim\limits_{n\to\infty} F_n(x)=F(x)$，则称随机序列 $\{X_n, n\ge 1\}$ 依分布收敛于 X，$F(x)$ 为 $\{F_n(x), n\ge 1\}$ 的极限分布函数。

【列维－林德伯格中心极限定理】设 $\{X_n, n\ge 1\}$ 为 $i.i.d.$ 随机序列，且 $EX_n=\mu, DX_n=\sigma^2, n\ge 1$，则 $\forall x\in R$，有

$$\lim_{n\to\infty} P\left(\left\{\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\le x\right\}\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)$$

；也即：$\dfrac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$ 的极限分布是标准正态分布；此时也称随机序列服从中心极限定理！

【注 1】 定理表明：独立同分布的随机变量之和的标准化变量（规范和）依分布收敛于标准正态变量！定理既验证了“只要独立同分布的随机变量方差存在，无论其原来是何分布，其极限分布均是正态分布”，从而在理论上支持了正态分布的重要性，初步说明了为什么实际应用中会经常遇到正态分布；又提供了一种“计算独立同分布随机变量和的分布的近似方法”，从而实际应用时十分简洁有效；只要和式中相加各项的个数充分地大，即可不必关心每个随机变量原先为何分布，都可用正态分布逼近！

将该定理应用到 **Bernoulli** 试验的背景，即有：

【德莫佛－拉普拉斯中心极限定理】设在 n 重 **Bernoulli** 试验中，事件 A 在每次试验中出现的概率为 p $(0<p<1)$，n_A 为 n 次试验中 A 出现的次数，则，

$$\forall x\in R, \lim_{n\to\infty} P\left(\left\{\frac{n_A-np}{\sqrt{np(1-p)}}\le x\right\}\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)$$ ；

【注 2】（1）一般地，

a. 在 np 较小的情况下，由 **Poisson** 定理知用 **Poisson** 分布近似较有效，即：

$$P\left(\left\{\frac{n_A-np}{\sqrt{np(1-p)}}\le x\right\}\right)\approx\sum_{k=0}^{[r]}\frac{\lambda^k}{k!}e^{-\lambda}, \lambda=\lim_{n\to\infty} np, r=np+x\sqrt{np(1-p)}$$ ；

b. 在 np 较大的情况下，常用正态分布来近似：

$P\left(\left\{\frac{n_A-np}{\sqrt{np(1-p)}}\le x\right\}\right)\approx\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$；需强调，**Poisson 分布逼近是分布列的逼近**，而正态分布逼近是分布函数的逼近；二者并不同！

（2）一般地，大数定律和中心极限定理之间并没有确定的关系也即是：$\{X_n,n\ge1\}$服从前（后）者，也可能不服从后（前）者；但是在独立同分布的场合，两者都存在，而且中心极限定理比大数定律更精确！可以证明由德莫佛－拉普拉斯中心极限定理可以推出 **Bernoulli** 大数定律；易见，

$$\forall\varepsilon>0,\lim_{n\to\infty}P\left(\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}\right)=\lim_{n\to\infty}P\left(\left\{\left|\frac{n_A-np}{\sqrt{np(1-p)}}\right|\le\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right\}\right)$$

$$=\lim_{n\to\infty}P\left(\left\{-\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}<\frac{n_A-np}{\sqrt{np(1-p)}}<\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right\}\right)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=1\quad;$$

【注 3】大数定律和中心极限定理都是描述随机变量的“和”的分布的极限过程，但它们却有质的区别：大数定律只刻划了“频率稳定于概率”，“平均值稳定于期望值”，即随机序列算术平均值的取值发展的趋向，属于定性的描述；而中心极限定理肯定地是在某些条件下，随机变量之和服从正态分布，即给出了过程的定量的描述，其在实用上往往具有更大的价值（大样本统计推断的理论基础）。

【例 **4.2.1**】分别用切比雪夫不等式与德莫佛－拉普拉斯中心极限定理确定当掷一枚均匀硬币时，需要掷多少次才能保证出现正面的频率在0.4到0.6之间的概率不小于0.9？

【例 **4.2.2**】假设某大学报名选修统计课程的学生人数$X\sim P(100)$，负责开课的老师决定：如果选课的学生人数不少于120人 ,就分成两个班讲授；如果少于120人，就集中一个班授课。试问：该老师讲授两个班的概率是多少？（0.0228）

【例 **4.2.3**】设$g(x)\in C[0,1]$，并记$I=\int_0^1 g(x)dx$；若$r.v.X\sim U[0,1]$，随机序列$\{X_n,n\ge1\}$独立且与X同分布，并且

$\eta_n=\frac{1}{n}\sum_{i=1}^{n}g(X_i),D[g(X_i)]=\sigma^2$，（1）求$E\eta_n,D\eta_n$，并证明：$\eta_n\xrightarrow{P}I$；

（2）$\forall\varepsilon>0$，利用中心极限定理估计概率$P(\{|\eta_n-I|<\varepsilon\})$；

【例 4.2.4】试利用中心极限定理证明： $\lim\limits_{n\to\infty}\sum\limits_{k=0}^{n}\frac{n^k}{k!}e^{-n}=\frac{1}{2}$ ；

【数理统计部分】

第五章 数理统计的基本概念（见教材）

第六章 参数估计、假设检验

§1（参数）点估计

这里的参数通常指：**1.**总体分布中的未知参数可以是向量；**2.**分布中未知参数的函数；**3.**分布中的各种数字特征。

【矩估计的基本思想】即替换原理：用样本矩（原点矩或中心矩）来替换相应的总体矩；用样本矩的函数来替换相应的总体矩的函数。

【替换原理的理论依据】辛钦大数定律，即样本矩依概率（以概率1）收敛于相应的总体矩。

【注**1**】 基于替换原理，在总体分布未知的场合下，常对以下各种参数进行如下估计：

（**1**） 以样本均值估计总体均值 $EX=\mu$，即：$\mu_{ME}=\overline{X}$（矩估计量），$\mu_{ME}=\overline{x}$（矩估计值）；

（**2**） 以样本方差（未修正）$S_n^{\ 2}=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 估计总体方差 $DX=\sigma^2$，即：$\sigma_{ME}^{\ \ 2}=S_n^{\ 2}$；

（**3**） 以样本的标准差估计总体的标准差，即：$\sigma_{ME}=S_n$；

（**4**） 若 $((x_1,y_1),(x_2,y_2),\cdots\cdots,(x_n,y_n))$ 是取自二元总体 (X,Y) 的一个样本，ρ 为 X,Y 的相关系数，则：$\rho_{ME}=\frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{S_{1n}S_{2n}}$；

矩估计法的步骤如下：

假设总体 X 的密度函数或概率函数为 $f(x;\theta)$，其中 $\theta=(\theta_1,\theta_2,\cdots,\theta_n)$，$\theta_i$ 为未知参数，$i=1,2,\cdots,n$；若总体 X 的 **k** 阶原点矩存在，则 $v_k=EX^k$ 也是 $\theta_1,\theta_2,\cdots,\theta_n$ 的函数， 记之为：$v_k=v_k(\theta_1,\theta_2,\cdots,\theta_n),k=1,2,\cdots,n$； 假设由参数方程 $v_k=v_k(\theta_1,\theta_2,\cdots,\theta_n),k=1,2,\cdots,n$ 可解出 $\theta_i=\theta_i(v_1,v_2,\cdots,v_n),i=1,2,\cdots,n$；再用相应的样本矩 $\overline{v_k}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^{\ k},k=1,2,\cdots,n$ 来替换 $v_k=EX^k,k=1,2,\cdots,n$，即有：

$\theta_{iME} = \theta_i\left(\overline{v_1}, \overline{v_2}, \cdots, \overline{v_n}\right), i = 1, 2, \cdots, n$ 。

【例 6.1.1】设总体 $X \sim f(x;\theta) = \theta(\theta+1)x^{\theta-1}(1-x)I_{\{0<x<1\}}(x)$，试求 θ_{ME}；

【例 6.1.2】设总体 $X \sim U(a,b), a,b$ 未知且 $-\infty < a, b < +\infty$，试求 a_{ME}, b_{ME}；

【例 6.1.3】设总体 $X \sim E(\lambda), \lambda > 0$ 未知，试求 λ_{ME}；（两种方法）

【例 6.1.4】设总体 $X \sim B(1,\theta)$，试求 $\theta, \theta(1-\theta)$ 的矩估计量；

【矩估计法总结】

1. 直观，简便，无须使用总体 X 的分布类型；
2. 矩估计可能不唯一，这也是其缺点之一，故尽量使用低阶矩替换；
3. 有时，在求解上述方程组时，不能或很难得出解析式，则只能通过数值计算求解；
4. 若总体相应的矩不存在，则此时矩法失效；
5. 由于样本矩表达式与总体的分布类型没有任何联系，此即表明矩估计有时没有充分利用总体的分布类型中所含未知参数的信息，因此矩估计法并不是很好的估计方法。

【最(极)大似然估计的基本原理】最(极)大似然原理。

为了说明该原理，先看如下引例，

【引例】设总体 $X \sim B(1,p)$，p 未知，(X_1, X_2, X_3) 为取自 X 的样本，(x_1, x_2, x_3) 为样本的一个观测值，令

$$L(x_1,x_2,x_3;p) = P\left(\left\{(X_1,X_2,X_3) = (x_1,x_2,x_3)\right\}\right) = \prod_{i=1}^{3} P\left(\{X_i = x_i\}\right) = p^{\sum_{i=1}^{3} x_i}(1-p)^{3-\sum_{i=1}^{3} x_i}$$

（1） 如果需要从 p 的两个值 $\frac{1}{4}, \frac{3}{4}$ 中选一个，即：

$p \in \Theta = \left\{\frac{1}{4}, \frac{3}{4}\right\}$； 将仅依赖于 $\sum_{i=1}^{3} x_i$ 的 $L(x_1,x_2,x_3;p)$ 列表如下：

$x_1+x_2+x_3$	0	1	2	3
$L\left(x_1,x_2,x_3;\frac{1}{4}\right)$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{1}{64}$
$L\left(x_1,x_2,x_3;\frac{3}{4}\right)$	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{27}{64}$

；若观测到 $x_1 + x_2 + x_3 = 0(3)$，自然应该认为 p 取 $\frac{1}{4}(\frac{3}{4})$，此时其是使观测值取最大概率（出现可能性最大）的参数值 。

（2） 如果参数 p 的取值范围为 $[0,1]$，即：$p \in \Theta = [0,1]$，设 $(x_1, x_2, \cdots\cdots, x_n)$ 为样本 $(X_1, X_2, \cdots\cdots, X_n)$ 的观测值，此时有，

$$L(x_1, x_2, \cdots\cdots, x_n; p) = P(\{(X_1, X_2, \cdots\cdots, X_n) = (x_1, x_2, \cdots\cdots, x_n)\}) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i},$$

依（1），此时应选取 p 的估计值为何值才能使该样本的观测值出现的可能性最大？易见，

1.若 $\sum_{i=1}^n x_i = 0$，则 $p_{MLE} = 0$；　　**2.**若 $\sum_{i=1}^n x_i = n$，则 $p_{MLE} = 1$；

3.若 $0 < \sum_{i=1}^n x_i < n$，令：$l(p) = L(x_1, x_2, \cdots\cdots, x_n; p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$，则，

$\ln l(p) = \left(\sum_{i=1}^n x_i\right)\ln p + \left(n - \sum_{i=1}^n x_i\right)\ln(1-p)$，令 $\frac{d}{dp}\ln l(p) = 0$，则有：

$p_{MLE} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$。

以下，我们分别就总体离散和连续两种情形来讨论未知（待估）参数的最大似然估计！

【离散型总体】设总体 X 的分布列为：

$P(\{X = x_i\}) = p(x_i; \theta_1, \theta_2, \cdots, \theta_k)$，其中 $\theta_1, \theta_2, \cdots, \theta_k$ 为未知（待估）参数，且 $(\theta_1, \theta_2, \cdots, \theta_k) \in \Theta$（参数空间），$(X_1, X_2, \cdots\cdots, X_n)$ 为取自总体 X 的一个样本，$(x_1, x_2, \cdots\cdots, x_n)$ 为样本的一个观测值，则样本的联合分布为 $l(\theta_1, \theta_2, \cdots, \theta_k) = L(x_1, x_2, \cdots\cdots, x_n; \theta_1, \theta_2, \cdots, \theta_k) = P(\{(X_1, X_2, \cdots\cdots, X_n) = (x_1, x_2, \cdots\cdots, x_n)\})$

$= \prod_{i=1}^n P(\{X_i = x_i\}) = \prod_{i=1}^n p(x_i; \theta_1, \theta_2, \cdots, \theta_k)$，称之为 $(\theta_1, \theta_2, \cdots, \theta_k)$ 的似然函数，则参数 $(\theta_1, \theta_2, \cdots, \theta_k)$ 的最大似然估计值 $(\theta_{1MLE}, \theta_{2MLE}, \cdots, \theta_{kMLE})$ 满足：

$$L(x_1, x_2, \cdots\cdots, x_n; \theta_{1MLE}, \theta_{2MLE}, \cdots, \theta_{kMLE}) = \max_{(\theta_1, \theta_2, \cdots, \theta_k) \in \Theta} L(x_1, x_2, \cdots\cdots, x_n; \theta_1, \theta_2, \cdots, \theta_k)$$

$\left(= \sup_{(\theta_1, \theta_2, \cdots, \theta_k) \in \Theta} L(x_1, x_2, \cdots\cdots, x_n; \theta_1, \theta_2, \cdots, \theta_k)\right)$；这里，$\Theta$ 为参数空间，即为 $(\theta_1, \theta_2, \cdots, \theta_k)$ 的取值范围。因此，只须根据具体问题得到似然函数，求出其最（极）大值点，即可得到待估参数的最大似然估计值，进而得到最大似然估计量。

【例 **6.1.5**】设总体 X 服从 $\{1, 2, \cdots, N\}$ 上的均匀分布，$(x_1, x_2, \cdots\cdots, x_n)$ 为样本的一个观测值，试求：N_{MLE}；

【例 **6.1.6**】**(1)**设总体 X 有分布：$P(\{X = 1\}) = \theta^2, P(\{X = 2\}) = 2\theta(1-\theta)$

，$P(\{X=3\})=(1-\theta)^2$，$\theta(0<\theta<1)$为未知参数；若已知取得的样本值为$(1,2,1)$，试求θ_{ME},θ_{MLE}；

(2)设一个试验有三种可能结果，其发生的概率分别为：$p_1=\theta^2, p_2=2\theta(1-\theta), p_3=(1-\theta)^2$，现做$n$次试验，观测到这三种结果发生的次数分别为$n_1,n_2,n_3(n_1+n_2+n_3=n)$，试求$\theta_{ME},\theta_{MLE}$；

【连续型总体】设总体X的概率密度函数为$f(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中$(\theta_1,\theta_2,\cdots,\theta_k)$为待估参数，且$(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta$，$(X_1,X_2,\cdots\cdots,X_n)$为取自总体 X 的一个样本；设$(x_1,x_2,\cdots\cdots,x_n)$为样本的一个观测值，则样本落在$(x_1,x_2,\cdots\cdots,x_n)$的附近（邻域）

$(x_1-\delta_1,x_1+\delta_1)\times(x_2-\delta_2,x_2+\delta_2)\times\cdots\times(x_n-\delta_n,x_n+\delta_n)$的概率为：

$$P(\{(X_1,X_2,\cdots\cdots,X_n)\in(x_1-\delta_1,x_1+\delta_1)\times(x_2-\delta_2,x_2+\delta_2)\times\cdots\times(x_n-\delta_n,x_n+\delta_n)\})$$

$$=\prod_{i=1}^n P(\{X_i\in(x_i-\delta_i,x_i+\delta_i)\})=\prod_{i=1}^n\int_{x_i-\delta_i}^{x_i+\delta_i} f(x;\theta_1,\theta_2,\cdots,\theta_k)dx\approx\prod_{i=1}^n[f(x_i;\theta_1,\theta_2,\cdots,\theta_k)2\delta_i]$$

$$=\left[\prod_{i=1}^n f(x_i;\theta_1,\theta_2,\cdots,\theta_k)\right]2^n\delta_1\delta_2\cdots\delta_n$$ 此概率的大小仅依赖于

$\prod_{i=1}^n f(x_i;\theta_1,\theta_2,\cdots,\theta_k)$；因此，可选取似然函数为：

$$l(\theta_1,\theta_2,\cdots,\theta_k)=L(x_1,x_2,\cdots\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)=\prod_{i=1}^n f(x_i;\theta_1,\theta_2,\cdots,\theta_k)\text{；}\quad(\theta_1,\theta_2,\cdots,\theta_k)$$

选取应使似然函数取最大值，从而有$(\theta_1,\theta_2,\cdots,\theta_k)$的最大似然估计值满足：

$$L(x_1,x_2,\cdots\cdots,x_n;\theta_{1MLE},\theta_{2MLE},\cdots,\theta_{kMLE})=\max_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}L(x_1,x_2,\cdots\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)$$

$$\left(=\sup_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}L(x_1,x_2,\cdots\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)\right)$$

【例 6.1.7】设总体$X\sim U[0,\theta]$，$\theta>0$为未知参数，试求θ_{MLE}；

【例 6.1.8】设总体$X\sim U[\theta,\theta+1]$，其中$\theta\in R$为未知参数，试求θ_{MLE}；

【例 6.1.9】设总体X的$p.d.f.$为$f(x;a,b)=bxI_{\{0\le x<1\}}(x)+axI_{\{1\le x<2\}}(x)$，样本的观测值为$(0.5,0.8,1.5,1.5)$，试求：（**1**）$a_{MLE},b_{MLE}$；（**2**）设$Y=e^X$，求$P(\{Y<2\})$的最大似然估计值；

【例 6.1.10】设总体X的$p.d.f.$为$f(x;\theta)=\theta I_{\{0<x<1\}}(x)+(1-\theta)I_{\{1\le x<2\}}(x)$，其中$\theta(0<\theta<1)$是未知参数；若$(X_1,X_2,\cdots\cdots,X_n)$为取自总体$X$的一个样本，$N$为样本值$(x_1,x_2,\cdots\cdots,x_n)$中小于$1$的个数，试求：

θ_{ME},θ_{MLE}；

【最大似然估计法总结】

1. 最大似然估计可能不止一个，有时甚至为无穷多个；

2. 有时，即是最大似然估计值唯一确定，却求不出或很难求出它的明显解析式，只能根据其样本观测值数值求解；

3. 对数似然方程的解不一定使得对数似然函数取得最大值；

4. 最大似然估计有许多优良的性质，其中有一条即：最大似然估计的不变性，也即若a的最大似然估计为a_{MLE}，则$g(a)$的最大似然估计为$g(a_{MLE})$。

【基于截尾样本的最大似然估计】设某种产品的寿命$X\sim E(\lambda)$，$\lambda>0$未知,由于时间或财力所限，不能做完全寿命试验，即：从总体中随机抽取n件，在$t=0$时同时投入试验，直到每个产品都失效为止；只能做截尾寿命试验。第一类：将n个产品在$t=0$时同时投入试验，试验进行到时刻t_0为止，称为定时截尾寿命试验；第二类：在n个同时投入试验的产品中，一旦失效数目达到预先规定的失效r个时就停止，称为定数截尾寿命试验；现在我们来求λ_{MLE}。

定时截尾寿命试验：设试验进行到规定的时刻t_0为止，设在$[0,t_0]$内有r个产品失效，其失效的时间为$0\le x_{(1)}\le x_{(2)}\le\cdots\le x_{(r)}\le t_0$，其余的$n-r$个产品均超过$t_0$；确定似然函数需要知道观测到上述结果出现的概率：　一个产品在$\left[x_{(i)},x_{(i)}+dx_{(i)}\right]$（$dx_{(i)}$很小）失效的概率近似为$f\left(x_{(i)}\right)dx_{(i)}$，而其余$n-r$个产品寿命超过$t_0$的概率为$\left[e^{-\lambda t_0}\right]^{n-r}$，故上述概率的近似值为

$$nf\left(x_{(1)}\right)dx_{(1)}(n-1)f\left(x_{(2)}\right)dx_{(2)}\cdots(n-r+1)f\left(x_{(r)}\right)dx_{(r)}\left[e^{-\lambda t_0}\right]^{n-r}$$

$$=\frac{n!}{(n-r)!}\lambda^r e^{-\lambda\sum_{i=1}^{r}x_{(i)}-\lambda(n-r)t_0}$$，此时取λ的似然函数为：

$$l(\lambda)=L\left(x_{(1)},x_{(2)},\cdots,x_{(r)};\lambda\right)=\lambda^r e^{-\lambda\sum_{i=1}^{r}x_{(i)}-\lambda(n-r)t_0}$$；令$\frac{d}{d\lambda}\ln l(\lambda)=0$，即有最大似然估计值：$\lambda_{MLE}=\frac{r}{\sum_{i=1}^{r}x_{(i)}+(n-r)t_0}$，从而有最大似然估计量：$\lambda_{MLE}=\frac{r}{\sum_{i=1}^{r}X_{(i)}+(n-r)t_0}$。

定数截尾寿命试验：设试验进行到第r个产品失效为止，其失

效的时间依次为 $0 \le x_{(1)} \le x_{(2)} \le \cdots \le x_{(r)}$，其余的 $n-r$ 个产品均超过 $x_{(r)}$；确定似然函数需要知道观测到上述结果出现的概率：一个产品在 $\left[x_{(i)}, x_{(i)}+dx_{(i)}\right]$ $\left(dx_{(i)}\text{很小}\right)$ 失效的概率近似为 $f\left(x_{(i)}\right)dx_{(i)}$，而其余 $n-r$ 个产品寿命超过 $x_{(r)}$ 的概率为 $\left[e^{-\lambda x_{(r)}}\right]^{n-r}$，故上述概率的近似值为

$$nf\left(x_{(1)}\right)dx_{(1)}(n-1)f\left(x_{(2)}\right)dx_{(2)}\cdots(n-r+1)f\left(x_{(r)}\right)dx_{(r)}\left[e^{-\lambda x_{(r)}}\right]^{n-r}$$

$=\dfrac{n!}{(n-r)!}\lambda^r e^{-\lambda\sum_{i=1}^{r}x_{(i)}-\lambda(n-r)x_{(r)}}$，此时取 λ 的似然函数为：

$l(\lambda)=L\left(x_{(1)},x_{(2)},\cdots,x_{(r)};\lambda\right)=\lambda^r e^{-\lambda\sum_{i=1}^{r}x_{(i)}-\lambda(n-r)x_{(r)}}$；令 $\dfrac{d}{d\lambda}\ln l(\lambda)=0$，即有最大似然估计值 $\lambda_{MLE}=\dfrac{r}{\sum_{i=1}^{r}x_{(i)}+(n-r)x_{(r)}}$，从而有最大似然估计量 $\lambda_{MLE}=\dfrac{r}{\sum_{i=1}^{r}X_{(i)}+(n-r)X_{(r)}}$。

§2（参数）区间估计

点估计虽然给出了未知参数的估计量（值），但并未给出估计量（值）的可靠度和精确度；而区间估计正好弥补了这一不足！

【区间估计（置信区间）】设 θ 是总体 X 的分布 $F(x;\theta)$ 中的未知参数，样本 $(X_1,X_2,\cdots\cdots,X_n)$ 取自总体 X，对于给定的 α（通常 $\alpha=0.1,0.05,0.01$ 等），若存在统计量 $\theta_1=\theta_1(X_1,X_2,\cdots\cdots,X_n)$ 和 $\theta_2=\theta_2(X_1,X_2,\cdots\cdots,X_n)$，使得 $P(\{\theta_1\le\theta\le\theta_2\})=1-\alpha$，则称 $[\theta_1,\theta_2]$（随机区间）为 θ 的置信度（置信水平）为 $1-\alpha$ 的置信区间（区间估计）。

【注 1】**1.** $\theta_1(\theta_2)$ 常称为置信下（上）限，α 又称为显著性水平；

2. $[\theta_1,\theta_2]$ 是随机区间（左，右端点都是随机变量），其长度一般也是随机变量；故 $[\theta_1,\theta_2]$ 包含 θ 的真值的概率 $1-\alpha$ 刻划了这个区间估计的可靠度（性）；

3.区间长度的期望 $E(\theta_2-\theta_1)$ 刻划了这个估计的精确度（性）；其值越小，则精确度越高；

4.在样本容量固定的情况下，不可能既提高可靠度又提高精确度；同时提高二者的途径一般是增加样本容量；

5.正确理解区间估计的“涵义”：若进行 100 次的抽样，进而得到 100 个具体的区间 $\left[\theta_1(x_1,x_2,\cdots,x_n),\theta_2(x_1,x_2,\cdots,x_n)\right]$，那么这其中大约有 $100(1-\alpha)$ 个区间包含了未知参数 θ 的真值。

构造区间估计常有如下一些方法：

1）　枢轴量法；2）统计量法；3）基于假设检验的接受域法；

以下仅讨论枢轴量法建立区间估计的一般步骤：

1. 先给出一个统计量$T=T(X_1,X_2,\cdots\cdots,X_n)$，一般选择未知参数$\theta$的点估计量；
2. 构造T和θ的函数$G(T,\theta)$，使其分布与待估参数θ无关（$G(T,\theta)$一枢轴量）；
3. 选择两个常数c_1,c_2，使得$P(\{c_1\le G(T,\theta)\le c_2\})=1-\alpha$；
4. 对不等式$c_1\le G(T,\theta)\le c_2$作恒等变形，得：$\theta_1\le\theta\le\theta_2$，即有：$P(\{c_1\le G(T,\theta)\le c_2\})=P(\{\theta_1\le\theta\le\theta_2\})=1-\alpha$，从而有，$[\theta_1,\theta_2]$为$\theta$的$1-\alpha$的区间估计。

【例 6.2.1】正态总体的参数区间估计：

1. 设总体$X\sim N(\mu,\sigma^2)$，σ^2已知，求μ的$1-\alpha$的区间估计；
2. 设总体$X\sim N(\mu,\sigma^2)$，σ^2未知，求μ的$1-\alpha$的区间估计；
3. 设总体$X\sim N(\mu,\sigma^2)$，μ已知，求σ^2的$1-\alpha$的区间估计；
4. 设总体$X\sim N(\mu,\sigma^2)$，μ未知，求σ^2的$1-\alpha$的区间估计。

【注 2】a) 区间估计可能有无穷多个，但当枢轴量的分布对称时，往往选择对称的区间；b) 随机区间的长度也可能为常数。

【例 6.2.2】非正态总体的参数区间估计：

设总体$X\sim E(\lambda),\lambda>0$未知,试求$\lambda$的置信水平为$1-\alpha$的置信区间；

【例 6.2.3】区间估计的大样本方法：

设总体X，其$EX=\mu$未知，$DX=\sigma^2$已知，试求μ的$1-\alpha$区间估计。

§3 假设检验

假设检验可分为：参数假设检验和非参数假设检验。

参数假设检验一对总体分布中的未知参数提出某种假设，然后利用样本提供的信息对所提出的假设进行检验，根据检验的结果对所提出的假设作出拒绝或接受的判断。

非参数假设检验一对总体分布的形式或总体的性质提出某种假设所进行的检验。

以下仅考虑参数假设检验！

基本思想、基本概念：

在数理统计中，常把关于总体分布的某个命题称为假设；如：

“总体服从均匀分布”、“总体均值大于200”等；一般地，假设的提出要有一定的理由和依据，但这种理由和依据往往是不充分的；这时就需要抽样调查，根据抽样的结果对假设的正确性作出判断，进而决定接受还是拒绝这个假设，即为“假设检验问题”。以下我们通过例题来说明其基本思想！

例：某灯泡厂在正常情况下，所生产的灯泡的使用寿命 $X\sim N\left(1800,100^2\right)$（单位:小时）今从该厂生产的一批灯泡中随机抽取25个检测，测得其平均寿命为 $\bar{x}=1730$，假设标准差保持不变，问：能否认为这批灯泡的平均寿命仍为1800小时？

注：表面上看，该问题即判断总体均值 $\mu=1800$ 是否成立？μ 的估计值 $\hat{\mu}=\bar{x}=1730\neq1800$，似乎应否定“$\mu=1800$”；如此做又未免武断，因为：即使 $\mu=1800$ 成立，样本均值 $\bar{x}=1800$ 出现的可能性也不大；事实上，由于 $\overline{X}$ 为 $c.r.v$，$P\left(\left\{\overline{X}=1800\right\}\right)=0$，即事件 $\left\{\overline{X}=1800\right\}$ 几乎不发生。因此参数点估计无法回答该问题，须另寻他法。

问题再描述：已知总体 $X\sim N\left(\mu,100^2\right)$，现抽取一容量为25的样本，$\bar{x}=1730$，检验假设 $H_0:\mu=1800\leftrightarrow H_1:\mu\neq1800$ 哪一个成立？

问题的一般形式：总体 $X\sim N\left(\mu,\sigma_0^{\ 2}\right)$，现抽取一容量为 n 的样本，得 $\bar{x}$，检验假设 $H_0:\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0$ 哪一个成立？

常称 H_0 是原（零）假设，H_1 是备择（对立）假设。

一般形式的讨论：现从 μ 的点估计量 $\overline{X}$ 出发研究该问题。令 $U=\dfrac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}$，当 H_0 成立时，$U\sim N\left(0,1^2\right)$，即有：$|U|$ 的取值一般偏小；当 H_1 成立时，$U\sim N\left(*,1^2\right)$，统计量 U 的取值主要集中于 $*$ 的周围，即：$|U|$ 的取值一般偏大；由此：$|U|$ 的取值偏小对 H_0 有利；反之对 H_1 有利；也即：$|U|$ 偏大且超过一定界限时，应拒绝 H_0；而当 $|U|$ 没有超过一定界限时，应接受或保留 H_0；从而，H_0 的拒绝域具有如下形式：

$\boxed{W=\left\{\left(x_1,x_2,\cdots,x_n\right):|U|>c\right\}}$，$c$ 待定又如何定？

强调：由于假设 $H_0:\mu=\mu_0$ 是正常情况下的状态（是以往的事实），故若没有充分的证据，就不应轻易怀疑其正确性。

一般地，要事先给定 H_0 成立时被拒绝的概率 α（显著性水平），

即：$P\left(X\in W\left|H_0\text{为真}\right.\right)=\alpha(\le\alpha)$；从而，$P_{\mu_0}\left(\left\{\left|\frac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}\right|>c\right\}\right)=P\left(\left|\frac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}\right|>c\left|\mu=\mu_0\right.\right)=\alpha$，故 $c=u_{\frac{\alpha}{2}}$，H_0 的拒绝域为 $W=\left\{(x_1,x_2,\cdots,x_n):|U|>u_{\frac{\alpha}{2}}\right\}$。上述推理可以说是一种"反证法"—为了检验 H_0 是否成立，先假设 H_0 成立，再运用统计方法观察由此会导致何种后果，如果（对 H_0 不利的）小概率事件在一次试验中发生了，即表明 H_0 很可能不正确，从而拒绝 H_0；反之，则没有理由拒绝 H_0，应接受它！此概率性质的反证法的依据是小概率事件原理，其区别于纯数学的反证法！

以下基于上例，给出（参数）假设检验的基本步骤：

一、建立假设：

假设检验中，常把一个被检验的假设称为原（零）假设，用 H_0 表示；常将不应轻易加以否定的假设作为原假设；当 H_0 被拒绝时而接受的假设称为备择（对立）假设，用 H_1 表示，它们常成对出现；如上例，可建立如下假设：

$H_0:\mu=1800\leftrightarrow H_1:\mu\ne1800$ 或 $H_0:\mu=1800 vs H_1:\mu\ne1800$，这里，**vs** 是"**versus**（对）"的缩写；表示 H_0 对 H_1 的假设检验问题。

二、选择检验统计量，给出拒绝域形式：

由样本对原假设进行判断总是通过一个统计量完成，该统计量称为检验统计量。使原假设 H_0 被拒绝的样本观测值所在区域称为拒绝域；它是样本空间的一个子集，并用 W 表示。

在上例中，选取的检验统计量 $U=\frac{\overline{X}-1800}{100}\sqrt{25}$；当 H_0 成立时，$|U|$ 一般偏小，假设显著性水平 $\alpha=0.05$，则由 $P(\{|U|>c\})=0.05$ 可知的拒绝域为：$W=\{(x_1,x_2,\cdots,x_n):|U|>u_{0.025}\}$。

当拒绝域确定了，检验准则也确定了：

1）如果 $(x_1,x_2,\cdots,x_n)\in W$，则认为 H_0 不成立；

2）如果 $(x_1,x_2,\cdots,x_n)\notin W$，则认为 H_0 成立。

由此，一个拒绝域唯一确定一个检验准则；反之，一个检验准则也唯一确定一个拒绝域。

检验的结果与真实情况可能吻合也可能不吻合；因此，检验是

可能犯错误的！检验可能犯的错误有两类：

1）H_0为真但由于随机性使样本观测值落在拒绝域中，从而拒绝原假设H_0，这种错误称为第一类错误，其发生的概率称为犯第一类错误的概率，又称弃真概率，常记之为α，即：

$\alpha=P\left(拒绝H_0\middle|H_0为真\right)$；

2）H_0不真（即H_1为真）但由于随机性使样本观测值落在接受域中，从而接受原假设H_0，这种错误称为第二类错误，其发生的概率称为犯第二类错误的概率，又称存（纳）伪概率，常记之为β，即：$\beta=P\left(接受H_0\middle|H_1为真\right)$。

三、作出判断：

由样本的观测值计算检验统计量的观测值，进而观察“小概率事件”是否发生，若发生（即样本观测值落入拒绝域内），则拒绝H_0，否则接受H_0。由上例，由于

$|U|=\left|\frac{1730-1800}{100}\sqrt{25}\right|=3.5>1.96=u_{0.025}$，即对$H_0$不利，应拒绝$H_0$，即认为$\mu\neq1800$。

【注 1】一个好的检验准则总希望犯两类错误的概率都较小；但一般场合下很难实现这种要求，除非增大样本容量！

【关于检验假设的进一步说明】在上例中，已经指出：假设的提出要有一定的理由和依据；在实际问题中，还有许多形式的假设，常见的关于一维参数的假设有以下8种形式：

1）$H_0:\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0$；2）$H_0:\mu\le\mu_0\leftrightarrow H_1:\mu>\mu_0$；

3）$H_0:\mu\ge\mu_0\leftrightarrow H_1:\mu<\mu_0$；4）$H_0:\mu=\mu_0\leftrightarrow H_1:\mu>\mu_0$；

5）$H_0:\mu=\mu_0\leftrightarrow H_1:\mu<\mu_0$；6）$H_0:\mu=\mu_0\leftrightarrow H_1:\mu=\mu_1(\mu_1\neq\mu_0)$；

7）$H_0:\mu\le\mu_1或\mu\ge\mu_2\leftrightarrow H_1:\mu_1<\mu<\mu_2(\mu_1<\mu_2)$；

8）$H_0:\mu_1\le\mu\le\mu_2(\mu_1<\mu_2)\leftrightarrow H_1:\mu<\mu_1或\mu>\mu_2$；由如上形式，参数假设的本质在于：将参数空间$\Theta$分解成互不相容的两个非空子集$\Theta_0$及$\Theta_1=\Theta-\Theta_0$；检验参数属于哪一个子集；故而，所有的参数假设都可写为如下的形式：$H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1=\Theta-\Theta_0$。不论是$H_0$还是$H_1$，若其中只含有一个参数值，则称为简单假设；否则称为复合假设。

在实数轴上，若与中任何一个都位于另一个的一侧，这种假设

称为单侧假设；相应的假设检验称为单侧检验；反之，若有一个位于另一个的两侧，这种假设称为双侧假设，相应的假设检验称为双侧检验。

一、 单个正态总体的参数假设检验：

正态总体的参数无非两个：期望 μ 与方差 σ^2 。设 $(X_1,X_2,\cdots,X_n)$ 是取自正态总体 $X\sim N(\mu,\sigma^2)$ 的样本，

1. 均值 μ 的检验：

1） σ^2 已知时，对 μ 主要有如下三种假设，

a）双侧假设 $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\neq\mu_0$；

b）单侧假设 $H_0:\mu\le\mu_0 \leftrightarrow H_1:\mu>\mu_0$；

c）单侧假设 $H_0:\mu\ge\mu_0 \leftrightarrow H_1:\mu<\mu_0$；

对于 **a）**，选取检验统计量 $U=\dfrac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}$， H_0 的拒绝域为：$W=\left\{(x_1,x_2,\cdots,x_n):|U|>u_{\frac{\alpha}{2}}\right\}$； 对于 **b）**，选取检验统计量 $U=\dfrac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}$，由于 H_0 成立时，U 值偏小；H_1 成立时，U 值偏大；故的 H_0 拒绝域形如：$W=\{(x_1,x_2,\cdots,x_n):U>c\}$；

$$P(X\in W|H_0\text{为真})=P\left(\frac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}>c|\mu\le\mu_0\right)=1-\Phi\left(c+\frac{\mu_0-\mu}{\sigma_0}\sqrt{n}\right)\le1-\Phi(c),$$

给定显著性水平 α，欲使 $P(X\in W|H_0\text{为真})\le\alpha$，只需使 $1-\Phi(c)=\alpha$，故 $c=u_\alpha$，即拒绝域为 $W=\{(x_1,x_2,\cdots,x_n):U>u_\alpha\}$； 对于 **c）**，选取检验统计量 $U=\dfrac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}$，由于 H_0 成立时，U 值偏大；H_1 成立时，值偏小；；故 H_0 的拒绝域形如：$W=\{(x_1,x_2,\cdots,x_n):U<c\}$；

$$P(X\in W|H_0\text{为真})=P\left(\frac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n}<c|\mu\ge\mu_0\right)=\Phi\left(c-\frac{\mu-\mu_0}{\sigma_0}\sqrt{n}\right)\le\Phi(c),$$

给定显著性水平 α，欲使 $P(X\in W|H_0\text{为真})\le\alpha$，只需使 $\Phi(c)=\alpha$，故 $c=u_{1-\alpha}=-u_\alpha$，即拒绝域为 $W=\{(x_1,x_2,\cdots,x_n):U<-u_\alpha\}$。

2） σ^2 未知时，对 μ 主要有如下三种假设，

a）双侧假设 $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\neq\mu_0$；

b）单侧假设 $H_0:\mu\le\mu_0 \leftrightarrow H_1:\mu>\mu_0$；

c）单侧假设 $H_0:\mu\ge\mu_0 \leftrightarrow H_1:\mu<\mu_0$； 由于 σ^2 未知，可由 S^{*2} 代替 σ^2，

选择检验统计量$T=\frac{\overline{X}-\mu_0}{S^*}\sqrt{n}$（$H_0$为真时，$T\sim t(n-1)$）。

对于 a)，给定显著性水平α，可得的拒绝域

$W=\left\{(x_1,x_2,\cdots,x_n):|T|>t_{\frac{\alpha}{2}}(n-1)\right\}$；对于 b)，由于$H_0$成立时，$T$值偏小；$H_1$成立时，$T$值偏大;故$H_0$的拒绝域形如：

$W=\{(x_1,x_2,\cdots,x_n):T>c\}$，故

$P(X\in W|H_0\text{为真})=P\left(\frac{\overline{X}-\mu_0}{S^*}\sqrt{n}>c|\mu\le\mu_0\right)=1-F\left(c-\frac{\mu-\mu_0}{S^*}\sqrt{n}\right)\le 1-F(c)$，这里$F(\cdot)$为分布函数，欲使$P(X\in W|H_0\text{为真})\le\alpha$，只需$1-F(c)=\alpha$，$c=t_\alpha(n-1)$，即$H_0$的拒绝域为$W=\{(x_1,x_2,\cdots,x_n):T>t_\alpha(n-1)\}$；对于 c)，$H_0$的拒绝域形如：$W=\{(x_1,x_2,\cdots,x_n):T<c\}$，故

$P(X\in W|H_0\text{为真})=P\left(\frac{\overline{X}-\mu_0}{S^*}\sqrt{n}<c|\mu\ge\mu_0\right)=F\left(c-\frac{\mu-\mu_0}{S^*}\sqrt{n}\right)$，这里$F(\cdot)$为分布函数，欲使$P(X\in W|H_0\text{为真})\le\alpha$，只需$F(c)=\alpha$，

$c=t_{1-\alpha}(n-1)=-t_\alpha(n-1)$，即$H_0$的拒绝域为

$W=\{(x_1,x_2,\cdots,x_n):T<t_{1-\alpha}(n-1)\}$。

【例 6.3.1】设$(X_1,X_2,\cdots,X_n)$是取自正态总体$N(\mu,1)$的一个样本，考虑如下假设检验问题：$H_0:\mu=2\leftrightarrow H_1:\mu=3$，若检验由拒绝域$W=\left\{(x_1,x_2,\cdots,x_n):\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\ge 2.6\right\}$来确定；**1**）试求当$n=20$时检验犯两类错误的概率；**2**）如果要使犯第二类错误的概率$\beta\le 0.01$，n最小应取多少？**3**）证明：当$n\to\infty$时，$\alpha,\beta\to 0$。（$\alpha=0.0037,\beta=0.0367$）

【例 6.3.2】设$(X_1,X_2,\cdots,X_{16})$是取自正态总体$N(\mu,4)$的样本，考虑检验问题：$H_0:\mu=6 vs H_1:\mu\ne 6$，其拒绝域为$W=\{(x_1,x_2,\cdots,x_n):|\bar{x}-6|\ge c\}$，试求$c$使得检验的显著性水平为$0.05$，并求检验在$\mu=6.5$处犯第二类错误的概率；（$c=0.98,\beta=0.8299$）

【例 6.3.3】设总体为均匀分布$U(0,\theta)$，$(X_1,X_2,\cdots,X_n)$是样本，考虑检验问题：$H_0:\theta\ge 3 vs H_1:\theta<3$，拒绝域取为$W=\{(x_1,x_2,\cdots,x_n):x_{(n)}\le 2.5\}$，求检验犯第一类错误的概率$\alpha$的最大值；若要使得$\alpha$不超过$0.05$，$n$至少应取多大？（$\alpha\le\left(\frac{5}{6}\right)^n,n\ge 17$）

【例 6.3.4】设总体 X 服从 $N(a,2.6^2)$ 分布，样本 $(X_1,X_2,\cdots,X_n)$ 取自总体 X，考虑检验问题：$H_0:a\le 12 \leftrightarrow H_1:a=13$，若拒绝域为 $W=\{(x_1,x_2,\cdots,x_n):\bar{x}\ge 12.4277\}$；

1）当 $n=100$ 时，求 α,β；2）当 $n\to\infty$ 时，$\alpha,\beta\to 0$。（$\alpha\le 0.05,\beta=0.0139$）

【附录部分】

【阅读材料一】—关于正态分布的推导（基于极大似然方法）

德国数学家高斯在研究误差理论时，利用极大似然估计方法，导出了正态分布！

设某物体的真实长度为 μ_0，它是未知的；现对其长度进行 n 次测量，得到样本观测值 $(x_1,x_2,\cdots,x_n)$，其平均值为 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，又设 $\delta_i=x_i-\mu_0$ 为测量误差，诸随机变量 δ_i 相互独立且同分布；现考虑一个问题：随机变量 δ_i 的概率分布为什么时，才能使观测值的平均值 $\bar{x}$ 等于真值 μ_0 有最大的可能性（概率）？

假设随机变量 δ_i 的密度函数为 $f(\delta_i)(i=1,2,,n)$，从而待估参数 μ_0 的似然函数为：$L(\mu_0)=L(x_1,x_2,\cdots,x_n;\mu_0)=\prod_{i=1}^{n}f(\delta_i)$，这里 μ_0 为未知（待估）参数。为求此似然函数的极大值，对 $L(\mu_0)$ 取对数，再求导并令之为 0；

$\frac{d[\ln L(\mu_0)]}{d\mu_0}=\sum_{i=1}^{n}\frac{d[\ln f(\delta_i)]}{d\mu_0}\underline{\underline{(d\mu_0=-d\delta_i)}}-\sum_{i=1}^{n}\frac{d[\ln f(\delta_i)]}{d\delta_i}=0$，从而有

$\boxed{\sum_{i=1}^{n}\left\{\frac{d[\ln f(\delta_i)]}{d\delta_i}\cdot\frac{1}{\delta_i}\cdot\delta_i\right\}=0}$（1）；欲使平均值 $\bar{x}$ 等于真值 μ_0 有最大的可能性，必须当 $\mu_0=\bar{x}$ 时，上面的（1）式成立；而 $\mu_0=\bar{x}$ 必须满足 $\boxed{\sum_{i=1}^{n}\delta_i=0}$（2）；于是（1）、（2）两式关于一切 i 以及每个 δ_i 的可能值都成立，所以两式中 δ_i 的系数必须成比例；由此可知对于一切 i 皆有，$\boxed{\frac{d[\ln f(\delta_i)]}{d\delta_i}\cdot\frac{1}{\delta_i}=k}$；此处 k 为常数。欲求密度函数 $f(\cdot)$，只需求解<u>微分方程</u>：$\boxed{\frac{d[\ln f(x)]}{dx}\cdot\frac{1}{x}=k}$；即有 $f(x)=Ce^{\frac{k}{2}x^2}$。由 $\int_{-\infty}^{+\infty}f(x)dx=1$ 可得，$k=-\frac{1}{\sigma^2},\sigma>0$ 为常数；从而 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$。

【阅读材料二】—R—S 积分及其性质

黎曼—斯蒂尔杰斯积分（Riemann—Stieltjes）是黎曼（定）积分的推广，常简记为 R—S 积分；在概率论中有着广泛应用！

【定义 1】设 $f(x),F(x)$ 是定义于 $[a,b]$ 上的实值函数，在 $[a,b]$ 上任取分点：$a=x_0<x_1<\cdots<x_{n-1}<x_n=b$，在每个子区间 $(x_{k-1},x_k]$ 上任取一点 $x_k^*(k=1,2,\cdots,n)$，作和式 $\sigma=\sum_{k=1}^{n}f(x_k^*)[F(x_k)-F(x_{k-1})]$；记：$\lambda=\max_k(x_k-x_{k-1})$，若 $\lambda\to 0$ 时，无论 $[a,b]$ 如何分割，也无论点 $x_k^*(k=1,2,\cdots,n)$ 如何选取，上述和式的极限总存在，则称该极限为 $[a,b]$ 上 $f(x)$ 关于 $F(x)$ 的 **R—S 积分**，记为：$\int_a^b f(x)dF(x)$，也即：

$$\boxed{\int_a^b f(x)dF(x)=\lim_{\lambda\to 0}\sum_{k=1}^{n}f(x_k^*)[F(x_k)-F(x_{k-1})]}。$$

【定义 2】设 $f(x),F(x)$ 在 $(-\infty,+\infty)$ 上均有定义，且对于任意的闭区间 $[a,b]$，积分 $\int_a^b f(x)dF(x)$ 总存在；若 $\lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_a^b f(x)dF(x)$ 存在，则记此极限为 $\boxed{\int_{-\infty}^{+\infty}f(x)dF(x)}$，称之为 $f(x)$ 关于 $F(x)$ 在 $(-\infty,+\infty)$ 上的广义 **R—S 积分**。

【定理】若 $f(x)$ 在 $[a,b]$ 上<u>连续</u>，且 $F(x)$ 在 $[a,b]$ 上<u>单调有界</u>，则在 $[a,b]$ 上 $f(x)$ 关于 $F(x)$ **R—S 可积**，即 $\int_a^b f(x)dF(x)$ 存在。

【**R—S 积分的双线性**】若下列出现的 **R—S** 积分都有意义，则

$$\boxed{\int_a^b[kf_1(x)+lf_2(x)]dF(x)=k\int_a^b f_1(x)dF(x)+l\int_a^b f_2(x)dF(x)},$$

$$\boxed{\int_a^b f(x)d[kF_1(x)+lF_2(x)]=k\int_a^b f(x)dF_1(x)+l\int_a^b f(x)dF_2(x)}。$$

【分部积分公式】若 $f(x)$ 在 $[a,b]$ 上<u>连续</u>，且 $F(x)$ 在 $[a,b]$ 上<u>单调有界</u>，则 $\boxed{\int_a^b f(x)dF(x)+\int_a^b F(x)df(x)=f(b)F(b)-f(a)F(a)}$。

在概率论中使用 **R—S** 积分时，其中的 $F(x)$ 一般是某个随机变量 X 的分布函数，因而 $dF(x)$ 具有概率意义，比如我们就有：$\boxed{P(\{a<X\le b\})=\int_a^b dF(x)}$；以下我们给出随机变量的数学期望的一个一般性定义（这里，不讨论关于概率测度的勒贝格积分定义）。

【定义 3】设随机变量 X 的分布函数为 $F(x)$，若 $\int_{-\infty}^{+\infty}|x|dF(x)<\infty$，则称 X 的数学期望存在，且有 $EX=\int_{-\infty}^{+\infty}xdF(x)$。

【定义 3 的注】我们就离散型随机变量和连续型随机变量的情形分别来讨论定义 3：

1） 设 X 为离散型随机变量，其分布列为 $\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$，分布函数为 $F(x)$；且 $x_1 < x_2 < \cdots < x_n < \cdots$（为方便计），数列 $\{x_n\}$ 无聚点，则 $\boxed{EX = \int_{-\infty}^{+\infty} xdF(x) = \sum_n x_n p_n}$。

考虑 R－S 积分定义中的和式：$\sum_{k=1}^{n} f(x_k^*)[F(x_k) - F(x_{k-1})]$：由于数列 $\{x_n\}$ 无聚点，当分割加细时，总可以使每个子区间最多只包含 X 的一个可能取值；若子区间 $(x_{k-1}, x_k]$ 不包含 X 的任何可能取值 x_i，则 $F(x_k) - F(x_{k-1}) = 0$，即此项对和式无贡献；若子区间 $(x_{k-1}, x_k]$ 包含 X 的一个可能取值 x_i，则

$F(x_k) - F(x_{k-1}) = F(x_i) - F(x_i -) = p_i$；当 $\lambda \to 0$ 时，包含 x_i 的这个子区间的相应项 $x_k^*[F(x_k) - F(x_{k-1})]$ 趋于 $x_i p_i$，故上述结论成立！

2） 设 X 为连续型随机变量，其概率密度为 $f(x)$，则易有 $\boxed{EX = \int_{-\infty}^{+\infty} xdF(x) = \int_{-\infty}^{+\infty} xf(x)dx}$。

该定义将离散型和连续型随机变量（当然也包括其他类型随机变量）的数学期望统一成一个表达式；不仅如此，该定义还可以用来计算混合型随机变量的数学期望！

【例题 1】设 $f(x) = x$，$F(x) = \sin x, x \in [0, \pi]$，试求

$$\int_0^{\pi} f(x)dF(x) = \int_0^{\pi} xd\sin x;$$

【例题 2】设 $f(x) = x$，$F(x) = e^{|x|}, -1 \le x \le 1$，试求

$$\int_{-1}^{1} f(x)dF(x) = \int_{-1}^{1} xd\left(e^{|x|}\right);$$

【例题 3】设 $f(x) = x$，$F(x) = k, k-1 < x \le k, k = 1, 2, 3, \cdots$，试求

$$\int_1^4 f(x)dF(x) = \int_1^4 xdF(x);$$

【例题 4】设随机变量 X 的分布函数为 $F(x) = \begin{cases} 0, x < -1; \\ \dfrac{5x+7}{16}, -1 \le x < 1; \\ 1, x \ge 1; \end{cases}$，试求 EX；

【法 1】 $EX = \int_{-\infty}^{+\infty} xdF(x) = \int_{\{-1\}} xdF(x) + \int_{(-1,1)} xdF(x) + \int_{\{1\}} xdF(x) =$

$$(-1)\left[F(-1)-F(-1-)\right]+\int_{-1}^{1}xd\left(\frac{5x+7}{16}\right)+1\cdot\left[F(1)-F(1-)\right]=$$

$$(-1)\frac{2}{16}+\int_{-1}^{1}x\frac{5}{16}dx+1\cdot\frac{4}{16}=\frac{1}{8};$$

【法 2】由 $F(-1)-F(-1-)=\frac{2}{16}$， $F(1)-F(1-)=\frac{4}{16}$，即有：

$F(x)=\frac{3}{8}F_1(x)+\frac{5}{8}F_2(x)$，其中，$F_1(x)$为离散型随机变量的分布函数，其分布列为$\begin{pmatrix}-1 & 1\\ \frac{1}{3} & \frac{2}{3}\end{pmatrix}$； $F_2(x)$为区间$(-1,1)$上均匀分布的分布函数；

故由 **R－S** 积分的双线性可得， $EX=\int_{-\infty}^{+\infty}xdF(x)=$

$$\frac{3}{8}\int_{-\infty}^{+\infty}xdF_1(x)+\frac{5}{8}\int_{-\infty}^{+\infty}xdF_2(x)=\frac{3}{8}\left[(-1)\frac{1}{3}+1\cdot\frac{2}{3}\right]+\frac{5}{8}\int_{-1}^{1}x\frac{1}{2}dx=\frac{1}{8}。$$

【例题 5】在$[0,1]$上随机地取一数X，其以概率0.2取值0.3，以概率0.3取值0.7， 又以概率0.5均匀取值于$[0.2,0.5]\cup[0.6,0.8]$； 试求 $EX,DX,P\left(X\le 0.3\middle|X\le 0.7\right)$；

【阅读材料三】－条件期望的另一种计算方法

有时为了计算随机变量及其函数的数学期望，我们还有如下的结论：

【命题 1】设X是取非负整数值的随机变量，则

$$\boxed{EX=\sum_{k=1}^{\infty}P(\{X\ge k\})=\sum_{k=0}^{\infty}P(\{X>k\})};$$

简证： $EX=\sum_{j=1}^{\infty}j\cdot P(\{X=j\})=\sum_{j=1}^{\infty}\left[\sum_{k=1}^{j}P(\{X=j\})\right]\overset{\text{交换求和次序}}{=\!=\!=}\sum_{k=1}^{\infty}\left[\sum_{j=k}^{\infty}P(\{X=j\})\right]$

$$=\sum_{k=1}^{\infty}\left[P\left(\bigcup_{j=k}^{\infty}\{X=j\}\right)\right]=\sum_{k=1}^{\infty}P(\{X\ge k\})。$$

【命题 2】设X是非负随机变量，则$\boxed{EX=\int_0^{\infty}P(\{X>x\})dx}$；

简证：该命题适合一切期望存在的非负随机变量，我们仅讨论连续型的情形！设X具有密度$f(x)$，则

$$EX=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^{\infty}xf(x)dx=\int_0^{\infty}\left[\int_0^{\infty}I_{\{y<x\}}(y)dy\right]f(x)dx=$$

$$\int_0^{\infty}\left[\int_0^{\infty}f(x)I_{\{y<x\}}(y)dx\right]dy=\int_0^{\infty}\left[\int_y^{\infty}f(x)dx\right]dy=\int_0^{\infty}P(\{X>y\})dy。$$

类似于命题 1、2，我们有如下命题；

【命题 3】设A是正概率事件，且给定A时，随机变量X取非负

整数值，则 $\boxed{E(X|A)=\sum_{k=1}^{\infty}P(X\ge k|A)}$。

【命题 4】设 A 是正概率事件，且给定 A 时，随机变量 X 取非负值，则 $\boxed{E(X|A)=\int_0^{\infty}P(X>x|A)dx}$。

【命题 5】设 A 是正概率事件， X 是非负随机变量，则

$$\boxed{E(X|A)=\frac{E(XI_A)}{P(A)}}。$$

简证：易知， $\{X>x\}\cap A=\{XI_A>x\},x>0$； 从而，

$$E(X|A)=\int_0^{\infty}P(X>x|A)dx=\frac{1}{P(A)}\int_0^{\infty}P(\{X>x\}\cap A)dx=\frac{1}{P(A)}\int_0^{\infty}P(\{XI_A>x\})dx$$

$$\underline{\underline{命题2}}\frac{E(XI_A)}{P(A)}。$$

【例题 1】设 X 服从参数为 λ 的指数分布， $a>0$， 试求

$E(X-a|X>a)$；

【例题 2】设 X 服从 $U(2,9)$ 分布，试求 $X|_{X<7}$ 的概率密度和数学期望；

【例题 3】假设某设备的使用寿命 Y 服从威布尔分布，其概率密度为 $f(y)=aby^{b-1}\exp(-ay^b),y\ge0$； 对 $b=\frac{1}{2}$，计算平均剩余寿命 $E(Y-t|Y>t)$；

【例题 4】设 X,Y 独立同 $U[0,1]$ 分布，试求 $E(X|X<Y)$。

【阅读材料四】－微分法求（联合）概率密度

本材料所介绍的微分法是现首都师范大学（原北京大学）教授何书元引入的计算随机变量与随机向量函数分布的简洁新方法，该材料也整理自何老师的著作《概率引论》!

一、 连续型随机变量函数的概率密度

设随机变量 X 有分布函数 $F(x)$，用 dx 表示 x 的微分，由微积分知识，当 $F'(x)$ 在 x 连续时，有

$P(\{X=x\})=\lim_{\Delta x\to0}|F(x)-F(x-\Delta x)|=dF(x)=F'(x)dx$ ； 故而我们可用 $\boxed{P(\{X=x\})=g(x)dx}$ 表示 X 在 x 有概率密度 $g(x)$。

【开集】如果 D 是开区间或开区间的并集，则称 D 是开集。

【定理 1】如果开集 D 使得 $P(\{X\in D\})=1$， $g(x)$ 在 D 中连续，使得 $P(\{X=x\})=g(x)dx,x\in D$（a）； 则 X 有概率密度 $f(x)=g(x),x\in D$（b）。

简证：仅对构成D的开区间的长度都大于某正数δ的情况来证明！对$x\notin D$，$P(\{X=x\})\le P(\{X\notin D\})=0$，说明$F(x)=P(\{X\le x\})$在$D$外连续；对$x\in D$，（**a**）式说明$F(x)$在$x$有连续的导数，且有$F'(x)=g(x)$；于是$F(x)$连续,故有概率密度（**b**）。

如果$h(y)$在y可微，$F(x)$在$x=h(y)$有连续的导数，则

$P(\{X=h(y)\})=\lim\limits_{\Delta y\to 0}\left|F[h(y)]-F[h(y-\Delta y)]\right|=\left|dF[h(y)]\right|=\left|f[h(y)]d[h(y)]\right|$

$=f[h(y)]|h'(y)|dy$；于是有

$$P(\{X=h(y)\})=\left|f[h(y)]d[h(y)]\right|=f[h(y)]|h'(y)|dy。$$

【**定理 1 的推论 1**】设X有概率密度$f(x)$，$Y=g(X)$是X的函数，开集D使得$P(\{Y\in D\})=1$；如果D上的分段严格单调可微函数$h(y)$使得$P(\{Y=y\})=P(\{X=h(y)\})=f[h(y)]|h'(y)|dy, y\in D$，则$Y$有概率密度$\boxed{f_Y(y)=f[h(y)]|h'(y)|, y\in D}$。

【**例题 1**】设$X\sim N(\mu,\sigma^2)$，则$Y=\dfrac{X-\mu}{\sigma}\sim N(0,1^2)$。

由于$X\sim f_X(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，且对任意的$y$有，

$P(\{Y=y\})=P\left(\left\{\dfrac{X-\mu}{\sigma}=y\right\}\right)=P(\{X=\sigma y+\mu\})=f_X(\sigma y+\mu)|d(\sigma y+\mu)|$

$=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\sigma y+\mu-\mu)^2}{2\sigma^2}}\sigma dy=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$；即$Y$有概率密度$f_Y(y)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$，也即：$Y\sim N(0,1^2)$。

【**例题 2**】设$X\sim N(0,1^2)$，$b\ne 0$，试求$Y=a+bX$的概率密度；

由于$X\sim f_X(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，且对任意的$y$有，$P(\{Y=y\})=$

$P(\{a+bX=y\})=P\left(\left\{X=\dfrac{y-a}{b}\right\}\right)=\left|f_X\left(\dfrac{y-a}{b}\right)d\left(\dfrac{y-a}{b}\right)\right|=\dfrac{1}{\sqrt{2\pi}|b|}e^{-\frac{(y-a)^2}{2b^2}}dy$；即有：$Y\sim N(a,b^2)$。

【**例题 3**】设X服从参数为$\lambda=1$的指数分布$E(1)$，a,b是正常数，试求$Y=\left(\dfrac{X}{a}\right)^{\frac{1}{b}}$的概率密度；

X有概率密度$f_X(x)=e^{-x}, x>0$，由于$P(\{Y>0\})=1$，且对任意的$y>0$，

有 $P(\{Y=y\})=P\left(\left\{\left(\frac{X}{a}\right)^{\frac{1}{b}}=y\right\}\right)=P(\{X=ay^b\})=\left|f_X(ay^b)d(ay^b)\right|=e^{-ay^b}aby^{b-1}dy$；

故 Y 的概率密度为 $f_Y(y)=e^{-ay^b}aby^{b-1},y>0$；即：$Y$ 服从常数为 (a,b) 的威布尔分布，记作：$Y\sim W(a,b)$。

【例题 4】设 $X\sim N(\mu,\sigma^2)$，试求 $Y=e^X$ 的概率密度；

易见 $P(\{Y>0\})=1$，且对任意的 $y>0$，有 $P(\{Y=y\})=P(\{e^X=y\})$

$=P(\{X=\ln y\})=\left|f_X(\ln y)d(\ln y)\right|=\frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}dy,y>0$；即有：

$Y\sim f_Y(y)=\frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}},y>0$，记作：$Y\sim Ln(\mu,\sigma^2)$，称 Y 服从参数为 (μ,σ^2) 的对数正态分布。

【定理 1 的推论 2】设 X 有概率密度 $f_X(x)$，$Y=g(X)$ 是 X 的函数，开集 D 使得 $P(\{Y\in D\})=1$；如果 $h_j(y)$ 在 D 上分段严格单调可微，使得 $P(\{Y=y\})=\sum_{j=1}^{n}P(\{X=h_j(y)\})=\sum_{j=1}^{n}f_X[h_j(y)]\left|h_j'(y)\right|dy,y\in D$，则 Y 有概率密度 $\boxed{f_Y(y)=\sum_{j=1}^{n}f_X[h_j(y)]\left|h_j'(y)\right|,y\in D}$。

【例题 5】设 $X\sim N(0,1^2)$，试求 $Y=X^2$ 的概率密度；

X 有概率密度 $X\sim f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，由于 $P(\{Y>0\})=1$，故对 $y>0$，

$P(\{Y=y\})=P(\{X=\sqrt{y}\})+P(\{X=-\sqrt{y}\})=\left|f_X(\sqrt{y})d(\sqrt{y})\right|+\left|f_X(-\sqrt{y})d(\sqrt{y})\right|$

$=\frac{1}{\sqrt{y}}f_X(\sqrt{y})dy=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}dy,y>0$；即：$Y$ 有密度 $f_Y(y)=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}},y>0$。

【例题 6】设 $a>0$，$X\sim U[-a,a]$，试求 $Y=\frac{1}{|X|}$ 的概率密度；

易见 $P\left(\left\{Y>\frac{1}{a}\right\}\right)=1$。对 $y>\frac{1}{a}$ 有 $\frac{1}{y}<a$，且 $f_X\left(\frac{1}{y}\right)=f_X\left(-\frac{1}{y}\right)=\frac{1}{2a}$；于是

$P(\{Y=y\})=P\left(\left\{|X|=\frac{1}{y}\right\}\right)=P\left(\left\{X=\frac{1}{y}\right\}\right)+P\left(\left\{X=-\frac{1}{y}\right\}\right)=\frac{1}{y^2}f_X\left(\frac{1}{y}\right)dy$

$+\frac{1}{y^2}f_X\left(-\frac{1}{y}\right)dy=\frac{1}{ay^2}dy$，从而 Y 有概率密度 $f_Y(y)=\frac{1}{ay^2},y>\frac{1}{a}$。

【例题 7】设 $X\sim E(\lambda)$，试求 $Y=\begin{cases}X,X\geq 1;\\-X^2,X<1;\end{cases}$ 的概率密度；

取开集 $D=(-1,0)\cup(1,+\infty)$，则 $P(\{Y\in D\})=P(\{Y>1\})+P(\{-1<Y<0\})$

$=P(\{X>1\})+P(\{0<X<1\})=1$。由于 X 有概率密度 $f_X(x)=\lambda e^{-\lambda x},x>0$，故对于 $y\in(1,+\infty)$，有 $P(\{Y=y\})=P(\{X=y\})=f_X(y)dy=\lambda e^{-\lambda y}dy$；对于 $y\in(-1,0)$，有 $P(\{Y=y\})=P(\{X^2=-y\})=P(\{X=\sqrt{-y}\})=$ $f_X(\sqrt{-y})\left|d(\sqrt{-y})\right|=\dfrac{\lambda}{2\sqrt{-y}}e^{-\lambda\sqrt{-y}}dy$；从而，$Y$ 有概率密度

$$f_Y(y)=\begin{cases}\lambda e^{-\lambda y},y\in(1,+\infty);\\ \dfrac{\lambda}{2\sqrt{-y}}e^{-\lambda\sqrt{-y}},y\in(-1,0);\end{cases}$$ 。

二、 连续型随机向量向量值函数的概率密度

设随机向量 (X,Y) 的联合分布函数 $F(x,y)$ 在点 (x,y) 的邻域内有连续的二阶混合偏导数，由于

$$P(\{X=x,Y=y\})=\lim_{\Delta x\to0+,\Delta y\to0+}P(\{x-\Delta x<X\le x,y-\Delta y<Y\le y\})=$$

$$\lim_{\Delta x\to0+,\Delta y\to0+}\left\{\begin{array}{l}\left[P(\{X\le x,Y\le y\})-P(\{X\le x,Y\le y-\Delta y\})\right]\\ -\left[P(\{X\le x-\Delta x,Y\le y\})-P(\{X\le x-\Delta x,Y\le y-\Delta y\})\right]\end{array}\right\}=$$

$$\lim_{\Delta x\to0+,\Delta y\to0+}\{[F(x,y)-F(x,y-\Delta y)]-[F(x-\Delta x,y)-F(x-\Delta x,y-\Delta y)]\}=$$

$$\lim_{\Delta x\to0+}\left\{\frac{\partial F(x,y)}{\partial y}-\frac{\partial F(x-\Delta x,y)}{\partial y}\right\}dy=\frac{\partial^2F(x,y)}{\partial x\partial y}dxdy$$ 。故而我们用

$\boxed{P(\{X=x,Y=y\})=g(x,y)dxdy}$ 表示 (X,Y) 在点 (x,y) 有联合密度 $g(x,y)$。

【**定理 2**】如果平面上的开集 D 使得 $P(\{(X,Y)\in D\})=1$，且 D 中的连续函数 $g(x,y)$ 使得 $P(\{X=x,Y=y\})=g(x,y)dxdy,(x,y)\in D$；则 (X,Y) 有联合密度 $f(x,y)=g(x,y),(x,y)\in D$。

由微积分，如果 $\begin{cases}x=x(u,v)\\ y=y(u,v)\end{cases}$ 在平面开集 D 内有连续的偏导数，且雅可比行列式 $J=\dfrac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\dfrac{\partial x}{\partial u}&\dfrac{\partial x}{\partial v}\\ \dfrac{\partial y}{\partial u}&\dfrac{\partial y}{\partial v}\end{vmatrix}\ne0$；则有 $dxdy=\left|\dfrac{\partial(x,y)}{\partial(u,v)}\right|dudv=|J|dudv$，$(x,y)\in D$，其中 $|J|$ 是 J 的绝对值。

【**定理 2 的推论 1**】设随机向量 (X,Y) 有联合密度 $f(x,y)$，$\begin{cases}U=u(X,Y)\\ V=v(X,Y)\end{cases}$ 是 (X,Y) 的向量值函数，平面上的开集 D 使得 $P(\{(U,V)\in D\})=1$；如果对 $(u,v)\in D$，有

$P(\{U=u,V=v\})=P(\{X=x,Y=y\})=f(x,y)dxdy=f(x,y)|J|dudv$，其中 $\begin{cases}x=x(u,v)\\y=y(u,v)\end{cases}$ 是 D 上可逆映射，有连续的偏导数，且雅可比行列式 $J\neq 0$，则 (U,V) 有联合密度 $\boxed{g(u,v)=f[x(u,v),y(u,v)]|J|,(u,v)\in D}$。

【例题 8】设随机向量 (X,Y) 有联合密度 $f(x,y)$，(U,V) 由线性变换 $\begin{cases}U=2X-Y\\V=2X+3Y\end{cases}$ 决定，试求 (U,V) 的联合密度；

从 $\begin{cases}u=2x-y\\v=2x+3y\end{cases}$ 可解出 $\begin{cases}x=\dfrac{3u+v}{8}\\y=\dfrac{-u+v}{4}\end{cases}$，且有 $J^{-1}=\dfrac{\partial(u,v)}{\partial(x,y)}=8, J=\dfrac{\partial(x,y)}{\partial(u,v)}=\dfrac{1}{8}$；

对于 (u,v)，由 $P(\{U=u,V=v\})=P\left(\left\{X=\dfrac{3u+v}{8},Y=\dfrac{-u+v}{4}\right\}\right)=$ $f\left(\dfrac{3u+v}{8},\dfrac{-u+v}{4}\right)|J|dudv$，$(U,V)$ 的联合密度 $g(u,v)=\dfrac{1}{8}f\left(\dfrac{3u+v}{8},\dfrac{-u+v}{4}\right)$。

【例题 9】设随机变量 X,Y 独立同 $N(0,1^2)$ 分布，由极坐标变换 $\begin{cases}X=R\cos\Theta\\Y=R\sin\Theta\end{cases}$ 决定，试求 (R,Θ) 的联合密度；

易见 $P(\{R>0,0<\Theta<2\pi\})=1$，且 (X,Y) 有联合密度 $f(x,y)=\dfrac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$。

对于 $\begin{cases}x=r\cos\theta\\y=r\sin\theta\end{cases}$，有 $J=\dfrac{\partial(x,y)}{\partial(r,\theta)}=r$；于是对 $r>0,0<\theta<2\pi$，有 $P(\{R=r,\Theta=\theta\})=P(\{X=x,Y=y\})=f(x,y)dxdy=f(r\cos\theta,r\sin\theta)rdrd\theta$ $=\dfrac{1}{2\pi}re^{-\frac{r^2}{2}}drd\theta$；故 (R,Θ) 的联合密度为 $g(r,\theta)=\dfrac{1}{2\pi}re^{-\frac{r^2}{2}},r>0,0<\theta<2\pi$。

【定理 2 的推论 2】设随机向量 (X,Y) 有联合密度 $f(x,y)$，$\begin{cases}U=u(X,Y)\\V=v(X,Y)\end{cases}$ 是 (X,Y) 的向量值函数，平面上的开集 D 使得 $P(\{(U,V)\in D\})=1$；如果对 $(u,v)\in D$，有

$P(\{U=u,V=v\})=\sum\limits_{i=1}^{n}P(\{X=x_i,Y=y_i\})=\sum\limits_{i=1}^{n}f(x_i,y_i)dx_idy_i=\sum\limits_{i=1}^{n}f(x_i,y_i)|J_i|dudv$，

其中，$\begin{cases}x_i=x_i(u,v)\\y_i=y_i(u,v)\end{cases}$ 是 D 上的可逆映射，且有连续的偏导数，其雅可比行列式 $J_i=\dfrac{\partial(x_i,y_i)}{\partial(u,v)}\neq 0,(u,v)\in D$；则 (U,V) 有联合密度

$\boxed{g(u,v)=\sum_{i=1}^{n} f\left[x_i(u,v),y_i(u,v)\right]|J_i|,(u,v)\in D}$。

【例题 10】设随机向量(X,Y)有联合密度$f(x,y)$，试求$(U,V)=(X\vee Y,X\wedge Y)$的联合密度；

对于开集$D=\{(u,v):u>v\}$，有$P(\{(U,V)\in D\})=P(\{X\neq Y\})=\iint\limits_{R^2} I_{\{x\neq y\}} f(x,y)dxdy=1$。对于$(u,v)\in D$，由$P(\{U=u,V=v\})=P(\{X=u,Y=v\})+P(\{X=v,Y=u\})=f(u,v)dudv+f(v,u)dudv$，可得$(U,V)$的联合密度为$g(u,v)=f(u,v)+f(v,u),u>v$。

【例题 11】设随机变量X,Y独立同$N(0,1^2)$分布，试求$(U,V)=\left(\frac{X}{Y},X^2+Y^2\right)$的联合密度；

设$D=\{(u,v):v>0,-\infty<u<+\infty\}$，即有$P(\{(U,V)\in D\})=1$。对$(u,v)\in D$，定义$\begin{cases} x=u\sqrt{\frac{v}{1+u^2}} \\ y=\sqrt{\frac{v}{1+u^2}} \end{cases}$，则$\begin{cases} u=\frac{x}{y} \\ v=x^2+y^2 \end{cases}$，且$J=\frac{\partial(x,y)}{\partial(u,v)}=\left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1}=\frac{1}{2(1+u^2)}$。

故由$P(\{U=u,V=v\})=P\left(\left\{\frac{X}{Y}=u,X^2+Y^2=v\right\}\right)=P\left(\left\{X=uY,|Y|=\sqrt{\frac{v}{1+u^2}}\right\}\right)=P(\{X=x,Y=y\})+P(\{X=-x,Y=-y\})=2P(\{X=x,Y=y\})=2f(x,y)|J|dudv=\frac{1}{2}e^{-\frac{v}{2}}\frac{1}{\pi(1+u^2)}dudv$，可得$(U,V)$的联合密度$g(u,v)=\frac{1}{2}e^{-\frac{v}{2}}\frac{1}{\pi(1+u^2)},v>0,-\infty<u<+\infty$。

【定理 3】如果R^n的开集D使得$P(\{Y=(Y_1,\cdots,Y_n)\in D\})=1$，且$D$中的连续函数$g(y_1,y_2,\cdots,y_n)$使得$P(\{Y_1=y_1,Y_2=y_2,\cdots,Y_n=y_n\})=g(y_1,y_2,\cdots,y_n)dy_1dy_2\cdots dy_n$，则$f(y_1,y_2,\cdots,y_n)=g(y_1,y_2,\cdots,y_n),(y_1,y_2,\cdots,y_n)\in D$是$Y=(Y_1,\cdots,Y_n)$的联合密度。

【定理 3 的推论 1】设X,Y都是n维随机向量，X有联合密度$f(x)=f(x_1,x_2,\cdots,x_n)$，$Y=h(X)$；如果$R^n$的开集$D$使得$P(\{Y\in D\})=1$，且对$y=(y_1,y_2,\cdots,y_n)\in D$，有$P(\{Y=y\})=\sum_{i=1}^{m}P(\{X=x^i\})=\sum_{i=1}^{m}f(x^i)dx^i$

$=\sum_{i=1}^{m} f\left(x^{i}\right)\left|J_{i}\right| dy$；其中 $x^{i}=x^{i}(y)$ 是 D 上的可逆映射，有连续的偏导数，

且雅可比行列式 $J_{i}=\frac{\partial x^{i}}{\partial y}=\left|\begin{pmatrix} \frac{\partial x_{1}^{i}}{\partial y_{1}} & \cdots & \frac{\partial x_{1}^{i}}{\partial y_{n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_{n}^{i}}{\partial y_{1}} & \cdots & \frac{\partial x_{n}^{i}}{\partial y_{n}} \end{pmatrix}\right| \neq 0, y \in D, 1 \leq i \leq m$，则 Y 有联合

密度 $g(y)=\sum_{i=1}^{m} f\left(x^{i}\right)\left|J_{i}\right|, y \in D$。

【**例题 12**】设 $X_{1}, X_{2}, \cdots, X_{n}$ 独立同分布，有分布函数 $F(x)$ 和概率密度 $f(x)$，试求：**1**）$\left(X_{(1)}, X_{(2)}, \cdots, X_{(n)}\right)$ 的联合密度；**2**）$X_{(k)}$ 的概率密度；**3**）对于 $1 \leq i<j \leq n$，计算 $\left(X_{(i)}, X_{(j)}\right)$ 的联合密度。

对于 **1**），定义 R^{n} 的开子集 $D=\left\{\left(x_{1}, x_{2}, \cdots, x_{n}\right) \mid x_{1}<x_{2}<\cdots<x_{n}\right\}$，易知 $P\left(\left\{\left(X_{(1)}, X_{(2)}, \cdots, X_{(n)}\right) \in D\right\}\right)=1$。对于 $x_{1}<x_{2}<\cdots<x_{n}$，

$\left\{X_{(1)}=x_{1}, X_{(2)}=x_{2}, \cdots, X_{(n)}=x_{n}\right\}=\bigcup_{\left(i_{1} i_{2} \cdots i_{n}\right)}\left\{X_{i_{1}}=x_{1}, X_{i_{2}}=x_{2}, \cdots, X_{i_{n}}=x_{n}\right\}$，从而，

$P\left(\left\{X_{(1)}=x_{1}, X_{(2)}=x_{2}, \cdots, X_{(n)}=x_{n}\right\}\right)=\sum_{\left(i_{1} i_{2} \cdots i_{n}\right)} P\left(\left\{X_{i_{1}}=x_{1}, X_{i_{2}}=x_{2}, \cdots, X_{i_{n}}=x_{n}\right\}\right)=$

$n! P\left(\left\{X_{1}=x_{1}, X_{2}=x_{2}, \cdots, X_{n}=x_{n}\right\}\right)=n! \prod_{i=1}^{n} P\left(\left\{X_{i}=x_{i}\right\}\right)=$

$n! f\left(x_{1}\right) f\left(x_{2}\right) \cdots f\left(x_{n}\right) d x_{1} d x_{2} \cdots d x_{n}$；由此得 $\left(X_{(1)}, X_{(2)}, \cdots, X_{(n)}\right)$ 的联合密度为

$g\left(x_{1}, x_{2}, \cdots, x_{n}\right)=n! f\left(x_{1}\right) f\left(x_{2}\right) \cdots f\left(x_{n}\right), x_{1}<x_{2}<\cdots<x_{n}$；对于 **2**），由于

$P\left(\left\{X_{(k)}=x\right\}\right)=P\left(\left\{X_{1}, X_{2}, \cdots, X_{n} \text{中有} k-1 \text{个} <x，\text{有一个}=x，\text{有} n-k \text{个}>x\right\}\right)$

$=\frac{n!}{(k-1)!(n-k)!} P\left(\bigcap_{i=1}^{k-1}\left\{X_{i}<x\right\}\right) P\left(\left\{X_{k}=x\right\}\right) P\left(\bigcap_{j=k+1}^{n}\left\{X_{j}>x\right\}\right)=$

$n C_{n-1}^{k-1}[F(x)]^{k-1} f(x) d x[1-F(x)]^{n-k}$，所以 $X_{(k)}$ 的概率密度为

$g_{k}(x)=n C_{n-1}^{k-1}[F(x)]^{k-1} f(x)[1-F(x)]^{n-k}$；对于 **3**），由 $P\left(\left\{X_{(i)}<X_{(j)}\right\}\right)=1$，

对于 $x<y$，有 $P\left(\left\{X_{(i)}=x, X_{(j)}=y\right\}\right)=$

$P\left(\left\{X_{1}, \cdots, X_{n} \text{中有} i-1 \text{个}<x，j-i-1 \text{个属于}(x, y)，n-j \text{个}>y，\text{一个}=x，\text{一个}=y\right\}\right)$

$=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j} f(x) f(y) d x d y$；

从而，$\left(X_{(i)}, X_{(j)}\right)$ 的联合密度为 $g(x, y)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times$

$[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j} f(x) f(y), x<y$。

【注】类似地，对条件密度也可作如下形式地推导：

假设$(X,Y)\sim f(x,y)$，对于满足$f_Y(y)>0$的y，可形式地有 $\boxed{P(X=x|Y=y)=\frac{P(\{X=x,Y=y\})}{P(\{Y=y\})}=\frac{f(x,y)dxdy}{f_Y(y)dy}=\frac{f(x,y)}{f_Y(y)}dx}$；根据上述介绍的微分法，给定$Y=y$时，$X$有条件密度$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$。

【阅读材料五】—期望向量、协方差矩阵与多元正态分布

以下仅以二维为例，其有关结论亦可推广至多维的情形！

设$X=(X_1,X_2)$是二维随机向量，如果$\mu_1=EX_1,\mu_2=EX_2$存在，则称X的期望存在，且有$\boxed{EX=(EX_1,EX_2),EX^T=(EX_1,EX_2)^T}$；令$\mu=EX$，即有$\mu=(\mu_1,\mu_2)$。如果每个随机变量$Y_{ij},i,j=1,2$的数学期望$EY_{ij}$都存在，则称以$Y_{ij},i,j=1,2$为元素的矩阵$Y=\begin{pmatrix}Y_{11}&Y_{12}\\Y_{21}&Y_{22}\end{pmatrix}$的数学期望存在，且定义为$\boxed{EY=\begin{pmatrix}EY_{11}&EY_{12}\\EY_{21}&EY_{22}\end{pmatrix}}$。

假设X,Y如上定义，即：$X=(X_1,X_2),Y=\begin{pmatrix}Y_{11}&Y_{12}\\Y_{21}&Y_{22}\end{pmatrix}$；则对任何常数向量$a=(a_1,a_2)$及二阶常数矩阵$A,B$，有如下结论：

$\boxed{E(aX^T)=aEX^T}$；$\boxed{E(AY)=AEY,E(YB)=E(Y)B}$；$\boxed{E(AYB)=AE(Y)B}$；$\boxed{(EY)^T=E(Y^T)}$。

如果随机向量$X=(X_1,X_2)$的数学期望$\mu=EX$存在，每个X_i的方差$\sigma_{ii}<+\infty$，则称$\Sigma=E\left[(X-\mu)^T(X-\mu)\right]=(\sigma_{ij})_{2\times2}$为$X$的协方差矩阵，其中$\sigma_{ij}=Cov(X_i,X_j)$是$X_i$和$X_j$的协方差。由于$\sigma_{ij}=\sigma_{ji}$，故协方差矩阵$\Sigma$是对称矩阵。

【定理】设$X=(X_1,X_2)$有协方差矩阵Σ，$EX=(\mu_1,\mu_2)$，则（1）Σ是半正定矩阵；（2）Σ退化的充要条件是存在一组不全为零的常数a_1,a_2，使得$P\left(\left\{\sum_{i=1}^{2}a_i(X_i-\mu_i)=0\right\}\right)=1$。

易见，任取二维实向量$a=(a_1,a_2)$，有$a\Sigma a^T=\sum_{i=1}^{2}\sum_{j=1}^{2}a_ia_j\sigma_{ij}=$

$$\sum_{i=1}^{2}\sum_{j=1}^{2}a_ia_jE\left[(X_i-\mu_i)(X_j-\mu_j)\right]=E\left[\sum_{i=1}^{2}\sum_{j=1}^{2}a_ia_j(X_i-\mu_i)(X_j-\mu_j)\right]=$$

$E\left[\sum_{i=1}^{2} a_i\left(X_i-\mu_i\right)\right]^2 \geq 0$。

以下讨论二元正态分布！

设随机变量 X_1, X_2 独立且同 $N(0,1^2)$ 分布，则 $X=(X_1,X_2)$ 有联合密度 $\varphi(x)=\varphi(x_1,x_2)=\frac{1}{2\pi}e^{-\frac{x_1^2+x_2^2}{2}}=\frac{1}{2\pi}e^{-\frac{1}{2}xx^T}$；这时称 $X=(X_1,X_2)$ 服从二元（维）标准正态分布，记作：$\boxed{X\sim N(0,I)}$。

定义：$f(x)=\frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right\}$，这里，$\Sigma$ 是正定矩阵，且满足 $\Sigma=A^TA$；我们来说明其是某随机向量的联合密度。

对于 $Y\sim N(0,I)$，定义 $X=YA+\mu$，$X=(X_1,X_2), Y=(Y_1,Y_2)$，利用微分法可得，$P(\{X=x\})=P(\{YA+\mu=x\})=P\left(\left\{Y=(x-\mu)A^{-1}\right\}\right)=$ $\varphi\left[(x-\mu)A^{-1}\right]\left|\frac{\partial(x-\mu)A^{-1}}{\partial(x_1,x_2)}\right|dx_1dx_2=\varphi\left[(x-\mu)A^{-1}\right]\left|\det\left(A^{-1}\right)\right|dx_1dx_2$；由 $\left|\det\left(A^{-1}\right)\right|=\frac{1}{|\det(A)|}=\frac{1}{\sqrt{|\det(\Sigma)|}}$，即有 $X=(X_1,X_2)$ 的联合密度 $\varphi\left[(x-\mu)A^{-1}\right]\left|\det\left(A^{-1}\right)\right|=\frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left\{-\frac{1}{2}(x-\mu)A^{-1}A^{-T}(x-\mu)^T\right\}=f(x)$。如果 $X=(X_1,X_2)$ 有联合密度 $f(x)$，则称 $X=(X_1,X_2)$ 服从二维（元）正态分布，记作：$\boxed{X\sim N(\mu,\Sigma)}$。

以下设 $Y\sim N(0,I)$，A 是可逆方阵，Σ 是正定矩阵，$\mu=(\mu_1,\mu_2)$。

【**定理 1**】设 $Z=(Z_1,Z_2)$ 是随机向量，则

1）$Z\sim N(\mu,\Sigma)$ 当且仅当存在 Y, A, μ 使得 $Z=YA+\mu, \Sigma=A^TA$；

2）二维正态分布的边缘分布是（一维）正态分布；

对于 **1**），如果 $Z=YA+\mu, \Sigma=A^TA$，则由上述的推导可得 $Z\sim N(\mu,\Sigma)$。反之，如果 $Z\sim N(\mu,\Sigma)$，则其和上述的 $X=YA+\mu$ 同分布，且 $\Sigma=A^TA$；从而，$\overline{Y}=(Z-\mu)A^{-1}$ 和 $Y=(X-\mu)A^{-1}$ 同分布，故 $\overline{Y}\sim N(0,I)$，且 $Z=\overline{Y}A+\mu$。

对于 $(a,b)\neq 0$，只要证明线性组合 $X_1=aY_1+bY_2+\mu_1$ 服从正态分布。取 $A=\begin{pmatrix} a & b \\ b & -a \end{pmatrix}$，则 $A^T=A$；令 $\sigma_1^2=a^2+b^2=|\det(A)|$，则 $\Sigma=A^TA=\sigma_1^2I$，$\Sigma^{-1}=\sigma_1^{-2}I, \det(\Sigma)=\sigma_1^4$。将其代入 $(X_1,X_2)=YA+\mu$ 的联合密度，即有，

$$f(x)=\frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left\{-\frac{1}{2}(x-\mu)A^{-1}A^{-T}(x-\mu)^T\right\}=\frac{1}{2\pi\sqrt{\sigma_1^4}}\exp\left\{-\frac{1}{2\sigma_1^2}(x-\mu)(x-\mu)^T\right\}$$

$=\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right\}\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_1^2}\right\}$；此即说明 X_1,X_2 相互独立，且 $X_1\sim N(\mu_1,\sigma_1^2)$。

将正定矩阵 Σ 写成 $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$，则

$\det(\Sigma)=\sigma_1^2\sigma_2^2(1-\rho^2),|\rho|<1$；易知 $\Sigma^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1}\\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2}\end{pmatrix}$；此时，

即有 $(x-\mu)\Sigma^{-1}(x-\mu)^T=\frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$；于是上述的联合密度可写成如下的等价形式

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}。$$

由于该等价形式中只有参数 $\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho$，故又可把 $X\sim N(\mu,\Sigma)$ 记作：$X\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$。

【定理 2】如果 $X=(X_1,X_2)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，则
1） $X_1\sim N(\mu_1,\sigma_1^2),X_2\sim N(\mu_2,\sigma_2^2)$，且 X_1,X_2 独立的充要条件是 $\rho=0$；**2）** 若 Z_1,Z_2 独立，且都服从正态分布，则 (Z_1,Z_2) 服从二维正态分布；**3）** 线性变换 $Z=XB+c$ 服从二维正态分布；**4）** 当 $(b_1,b_2)\neq 0$ 时，线性组合 $Z_1=b_1X_1+b_2X_2+b_3$ 服从正态分布。

【例题 1】设 $X\sim N(1,2),Y\sim N(3,8)$，且 X,Y 相互独立，试求 $U=4X+3Y,V=6X-2Y$ 的联合密度；

由定理 2，(U,V) 服从二维正态分布，且 $EU=13,EV=0,DU=104$，$DV=104,Cov(U,V)=0$；从而有 $(U,V)\sim N(13,0;104,104;0)$。

【例题 2】设 $(X,Y)\sim N(2,3;4,5;\sqrt{0.8})$，$U=X+2Y+1,V=3X-2Y+4$，试求：**1）** U,V 的分布；**2）** (U,V) 的联合分布及协方差矩阵。

【注】更多有关二维（多维）正态分布的内容可参考多元统计分析教材！