第 3 章　态势信息处理的理论基础

本章对信息融合过程中常用的基础理论进行集中描述。这些理论从概率统计理论出发，首先介绍统计判决方面的基础知识，它们在航迹处理的数据关联和航迹相关时会使用到，然后介绍线性滤波理论，并推导 Kalman 滤波器，给出 Kalman 滤波器的工作流程和扩展的 EKF、UKF 滤波器，并用粒子滤波作为对滤波理论的总结。滤波器是航迹跟踪的基础理论，对处理传感器的输出数据、序贯地获得运动目标状态具有重要作用。最后，本章概要描述属性融合的基本理论，包括贝叶斯方法和 DS 证据理论，它们在航迹的身份信息融合时能发挥主要作用。这些基本理论只是信息融合这个大框架中使用到的部分内容，当需要从其他角度对信息融合问题进行描述时，还会用到其他理论工具。

3.1　统计与判决理论

3.1.1　随机数及其生成

定义在样本空间 Ω 上的实数函数 $X=X(w)$ 称为随机变量。仅取有限个或可列个值的随机变量称为离散随机变量；取值充满某个区间 (a,b) 的随机变量称为连续随机变量，这里 a 可为 $-\infty$，也可为 $+\infty$。

设 X 是一个随机变量，对任意实数 x，称 $F(x)=P(X\leqslant x)$ 为 X 的分布函数，记为 $X\sim F(x)$。分布函数具有如下 3 条基本性质。

1）单调性

$F(x)$ 是单调非减函数，即对任意的 $x_1<x_2$，都有 $F(x_1)\leqslant F(x_2)$。

2）有界性

对任意的 x，有 $0\leqslant F(x)\leqslant 1$，且

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,\quad F(+\infty)=\lim_{x\to+\infty}F(x)=1$$

3）右连续性

$F(x)$ 是 x 的右连续函数，即对任意的 x_0，都有

$$\lim_{x\to x_0^+}F(x)=F(x_0),\quad 即\ F(x_0+0)=F(x_0)$$

可以证明：具有上述 3 条性质的函数 $F(x)$ 一定是某一个随机变量的分布函数。

若离散随机变量 X 的可能取值是 $x_1,x_2,\cdots,x_n,\cdots$，则称 X 取 x_i 的概率

$$p_i=p(x_i)=P(X=x_i),i=1,2,\cdots,n,\cdots$$

为 X 的概率分布列，简称分布列。分布列也可用列表方式表示。

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

分布列 $p(x_i)$具有如下两个基本性质。

1）非负性

$p(x_i)\geqslant 0, i=1,2,\cdots$;

2）正则性

$$\sum_{i=1}^{+\infty} p(x_i) = 1$$

离散随机变量 X 的分布函数为 $F(x) = \sum\limits_{x_i \leqslant x} p(x_i)$，它是有限级或可列无限级阶梯函数，离散随机变量 X 取值于区间$(a,b]$上的概率为 $P(a < X \leqslant b) = F(b) - F(a)$。常数 c 可看作仅取一个值的随机变量 X，即 $P(X = c) = 1$，它的分布常称为单点分布或退化分布。

记连续随机变量 X 的分布函数为 $F(x)$，若存在一个非负可积函数 $p(x)$，使得对任意实数 x，都有 $F(x) = \int_{-\infty}^{x} p(t)\mathrm{d}t$，则称 $p(x)$ 为 X 的概率密度函数，简称密度函数。连续随机变量 X 的分布函数 $F(x)$ 是$(-\infty, +\infty)$上的连续函数，它可能在有限个点或可列个点上不可导，除此以外，还有 $F'(x) = p(x)$。

1. 期望与方差

$$EX = \begin{cases} \sum\limits_{i} x_i p(x_i) & X\text{ 为离散随机变量} \\ \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x & X\text{ 为连续随机变量} \end{cases}$$

为 X 的数学期望，简称期望或均值，且称 X 的数学期望存在，否则称数学期望不存在。

数学期望是由分布决定的，它是分布的位置特征。只要两个随机变量同分布，则其数学期望总是相等的。假如把概率看作质量、把分布看作某物体的质量分布，那么数学期望就是该物体的重心位置。

数学期望的性质

（1）X 的某一函数 $g(X)$的数学期望为

$$E[g(X)] = \begin{cases} \sum\limits_{i} g(x_i)p(x_i) & \text{离散场合} \\ \int_{-\infty}^{+\infty} g(x)p(x)\mathrm{d}x & \text{连续场合} \end{cases}$$

（2）若 c 是常数，则 $E(c)=c$。

（3）对任意常数 a，都有 $E(aX)=aE(X)$。

（4）对任意的两个函数 $g_1(x)$和 $g_2(x)$，都有

$$E[g_1(X) \pm g_2(X)] = E[g_1(X)] \pm E[g_2(X)]$$

方差称随机变量 X 对期望 $E(X)$的偏差平方的数学期望（设其存在）

$$\mathrm{Var}(X) = E[X - E(X)]^2$$

为 X 的方差，称方差的正平方根 $\sigma(X)=\sigma_X=\sqrt{\mathrm{Var}(X)}$为 X 的标准差。

方差是由分布决定的，它上分布的散度特征，方差越大，分布越分散；方差越小，分布越集中。标准差与方差的功能相似，只是量纲不同。

方差的性质　以下涉及的方差均假设其存在。

(1) $\mathrm{Var}(X)=E(X^2)-[E(X)]^2$。

(2) 若 c 是常数,则 $\mathrm{Var}(c)=0$。

(3) 若 a,b 是常数,则 $\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$。

(4) 若随机变量 X 的方差存在,则 $\mathrm{Var}(X)=0$ 的充要条件是 X 几乎处处为某个常数 a,即 $P(X=a)=1$。

随机变量的标准化　对任意随机变量 X,如果 X 的数学期望和方差存在,则称

$$X^* = \frac{X-E(X)}{\sqrt{\mathrm{Var}(X)}}$$

为 X 的标准化随机变量,此时 $E(X^*)=0$,$\mathrm{Var}(X^*)=1$。

2. 常用随机变量的分布

二项分布　若 X 的概率分布列为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\cdots,n$$

则称 X 服从二项分布,记为 $X\sim b(n,p)$,其中 $0<p<1$。n 重伯努利试验中成功的次数 X 服从二项分布 $b(n,p)$,其中 p 为一次伯努利试验中成功发生的概率。

泊松分布　若 X 的概率分布列为

$$(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad k=0,1\cdots$$

则称 X 服从泊松分布,记为 $X\sim P(\lambda)$,其中参数 $\lambda>0$。单位时间(或单位面积、单位产品等)上某稀有事件(这里稀有事件是指不经常发生的事件)发生的次数常服从泊松分布 $P(\lambda)$,其中 λ 为该稀有事件发生的强度。在 n 重伯努利试验中,记事件 A 在一次试验中发生的概率为 p_n,当 $n\to+\infty$时,有 $np_n\to\lambda$,二项分布趋近于泊松分布。

几何分布　若 X 的概率分布列为

$$P(X=k)=(1-p)^{k-1}p,\quad k=1,2,\cdots$$

则称 X 服从几何分布,记为 $X\sim Ge(p)$,其中 $0<p<1$。在伯努利试验序列中,事件 A 首次出现时的试验次数 X 服从几何分布 $Ge(p)$,其中 p 为每次试验中事件 A 发生的概率。几何分布的无记忆性:若 $X\sim Ge(p)$,则对任意正整数 m 与 n,都有

$$P(X>m+n\mid X>m)=P(X>n)$$

正态分布　若 X 的密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

则称 X 服从正态分布,记作 $X\sim N(\mu,\sigma^2)$,其中参数$-\infty<\mu<+\infty$,$\sigma>0$。一个变量若是大量微小的、独立的随机因素的叠加结果,则此变量一定是正态变量(服从正态分布的变量)。测量误差是由量具零点偏差、测量环境的影响、测量技术的影响、测量人员的心理影响等随机因素叠加而成的,所以常认为测量误差服从正态分布。

μ 是正态分布的数学期望,即 $E(X)=\mu$,称 μ 为正态参数的位置参数。μ 是正态分布的对称中心,在 μ 的左侧和 $p(x)$下的面积为 0.5;在 μ 的右侧和 $p(x)$下的面积也为 0.5,

所以 μ 也是正态分布的中位数。若 $X\sim N(\mu,\sigma^2)$，则在离越近取值的可能性越大，离越远取值的可能性越小。

σ^2 是正态分布的方差，即 $\mathrm{Var}(X)=\sigma^2$。σ 是正态分布的标准差，σ 越小，正态分布越集中；σ 越大，正态分布越分散。σ 又称为正态分布的尺度参数。若 $X\sim(\mu,\sigma^2)$，则其密度函数 $p(x)$ 在 $\mu\pm\sigma$ 处有两个拐点。

称 $\mu=0,\sigma=1$ 时的正态分布 $N(0,1)$ 为标准正态分布。若 $X\sim N(\mu,\sigma^2)$，则 $U=(X-\mu)/\sigma\sim N(0,1)$，其中 $U=(X-\mu)/\sigma$ 称为 X 的标准化变换。

均匀分布　若 X 的密度函数和分布函数分别为

$$p(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b\\ 0, & 其他\end{cases}\qquad F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\leqslant x<b\\ 1, & x\geqslant b\end{cases}$$

则称服从区间 (a,b) 上的均匀分布，记作 $X\sim U(a,b)$。

向区间 (a,b) 随机投点，落点坐标 X 一定服从均匀分布 $U(a,b)$。这里，“随机投点”是指点落在任意相等长度的小区间上的可能性是相等的。区间 $(0,1)$ 上的均匀分布是导出其他分布随机数的桥梁。

指数分布　若 x 的密度函数为

$$p(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

则称 X 服从指数分布，记作 $X\sim\mathrm{Exp}(\lambda)$，其中参数 $\lambda>0$。

若一个元器件（或一台设备，或一个系统）遇到外来冲击时即告失效，则首次冲击来到的时间 X（寿命）服从指数分布，很多产品的寿命可认为服从或近似服从指数分布。指数分布的无记忆性：若 $X\sim\mathrm{Exp}(\lambda)$，则对任意 $s>0,t>0$，都有

$$P(X>s+t\mid X>s)=P(X>t)$$

伽马分布　$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x$ 为伽马函数，其中参数 $\alpha>0$。伽马函数具有如下性质：

(a) $\Gamma(1)=1$；　(b) $\Gamma(1/2)=\sqrt{\pi}$；

(c) $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$；　(d) $\Gamma(n+1)=n\Gamma(n)=n!(n\in N)$

若 X 的密度函数为

$$p(x)=\begin{cases}\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

则称 X 服从伽马分布，记作 $X\sim\mathrm{Ga}(\alpha,\lambda)$，其中 $\alpha>0$ 为形状参数，$\lambda>0$ 为尺度参数。若一个元器件（或一台设备，或一个系统）能抵挡一些外来冲击，但遇到第 k 次冲击时即失效，则第 k 次冲击来到的时间 X（寿命）服从形状参数为 k 的伽马分布 $\mathrm{Ga}(k,\lambda)$。$\alpha=1$ 时的伽马分布就是指数分布，即 $\mathrm{Ga}(1,\lambda)=\mathrm{Exp}(\lambda)$。称 $\alpha=n/2,\lambda=1/2$ 时的伽马分布为自由度为 n 的 χ^2（卡方）分布，记为 $\chi^2(n)$，其密度函数为

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}e^{-\frac{x}{2}}x^{\frac{n}{2}-1}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

$\chi^2(n)$分布的期望和方差分别为 $E(X)=n$，$\mathrm{Var}(X)=2n$。若形状参数为整数 k，则伽马变量可表示成 k 个独立同分布的指数变量之和，即若 $X\sim \mathrm{Ga}(k,\lambda)$，则 $X=X_1+X_2+\cdots+X_k$，其中 $X_1,X_2,\cdots,X_k$ 是相互独立且都服从指数分布 $\mathrm{Exp}(\lambda)$的随机变量。

3. 随机数的生成

1）同余法

同余法(Linear Congruence Generator，LCG)是 Lehmer 于 1951 年提出的。同余法利用数论中的同余运算原理产生随机数。同余法是目前发展迅速且使用普遍的方法之一。

LCG 的递推公式为

$$x_n=(ax_{n-1}+c)(\bmod m)\quad (n=1,2,\cdots) \tag{3-1}$$

其中 x_n、a、c 均为正整数。只需给定初值 x_0，就可以由式(3-1)得到整数序列$\{x_n\}$，对每一个 x_n，作变换 $u_n=x_n/m$，则$\{u_n\}(n=1,2,\cdots)$就是$[0,1)$上的一个序列。如果$\{u_n\}$通过了统计检验，那么就可以将 u_n 作为$[0,1)$上的均匀分布随机数。

在式(3-1)中，若 $c=0$，则称相应的算法为乘同余法，并称 a 为乘子；若 $c\neq 0$，则称相应的算法为混合同余法。同余法也称为同余发生器，其中 x_0 称为种子。

由式(3-1)可以看出，对于十进制数，当取模 $m=10^k$(k 为正整数)时，求其同余式运算较简便。例如，$36=31236(\bmod 10^2)$，只要对 31236 从右截取 $k=2$ 位数，即得余数 36。同理，对于二进制数，取模 $m=2^k$ 时，求其同余式运算更简便了。

电子计算机通常是以二进制形式表示数的。在整数尾部字长为 L 位的二进制计算机上，按式(3-1)求以 m 为模的同余式时，可以利用计算机具有的整数溢出功能。设 L 为计算机的整数尾部字长，取模 $m=2^L$，若按式(3-1)求同余式时，显然有

当 $ax_{n-1}+c<m$ 时，则 $x_n=ax_{n-1}+c$；

当 $ax_{n-1}+c\geqslant m$ 时，则 $x_n=ax_{n-1}+c-m[(ax_{n-1}+c)/m]$。

这里，$[x]$是取 x 的整数部分。在计算机上由 x_{n-1}求 x_n 时，可利用整数溢出原理，不进行上面的除法运算。实际上，由于计算机的整数尾部字长为 L，机器中可存放的最大整数为 2^L-1，而此时 $a^{x_{n-1}}+c\geqslant m\geqslant 2^L-1$，因此 $a^{x_{n-1}}+c$ 在机器里存放时占的位数多于 L 位，于是发生溢出，只能存放 x_n 的右后 L 位。这个数值恰是模 $m=2^L$ 的剩余，即 x_n。这就减少了除法运算，而实现了求同余式。经常取模 $m=2^L$(L 为计算机尾部字长)，正是利用了溢出原理减少除法运算。

由式(3-1)产生的 $x_n(n=1,2,\cdots)$到一定长度后，会出现周而复始的周期现象，即$\{x_n\}$可以由其某一子列的重复出现而构成，这种重复出现的子列的最短长度称为序列 x_n 的周期。由式(3-1)不难看出，$\{x_n\}$中两个重复数之间的最短距离长度就是它的周期，用 T 代表周期。周期性表示一种规律性，它与随机性是矛盾的。因此，通常只能取$\{x_n\}$的一个周期作为可用的随机序列。这样，为了产生足够多的随机数，就必须使$\{x_n\}$的周期尽可

能大。由前所述，一般取 $m=2^L$，这就是说模 m 已取到计算机能表示的数的最大数值，即使产生的随机数列 $\{x_n\}$ 的周期达到可能的最大数值，如适当地选取参数 x_0、a、c 等，还可能使随机数列 $\{x_n\}$ 达到满周期。

2）组合法

组合法的基本思想是把预定概率密度函数 $f(x)$ 表示为其他一些概率密度的线性组合，而这些概率密度的随机抽样容易产生。通过这种避难就易的手段，也许可以达到较快的输出速度和较好的性能。若分布密度函数 $f(x)$ 可表示为式（3-2）所示的函数项级数的和，

$$f(x)=\sum_{i=1}^{\infty}p_i f_i(x) \tag{3-2}$$

其中 $\sum_{i=1}^{\infty}p_i$、$f(x)$ 皆为概率密度函数。依照如下步骤可得到分布为 $f(x)$ 的一次抽样。

（1）产生一个随机自然数 I，使 I 服从分布律：$P(I=i)=p^i, i=1,2,3,\cdots$

（2）产生服从 $f_I(x)$ 的随机数 X_0。

利用全概率公式，有

$$\begin{aligned}P(x<X\leqslant x+\mathrm{d}x)&=\sum_{i=1}^{\infty}P(I=i)P(x<X\leqslant x+\mathrm{d}x\mid I=i)\\&=\sum_{i=1}^{\infty}p_i f_i(x)\mathrm{d}x\\&=f(x)\mathrm{d}x\end{aligned}$$

故 X 服从 $f(x)$ 分布。下面以产生双指数或拉普拉斯（Laplace）分布的随机数为例简单说明这种方法。双指数分布具有概率密度函数 $f(x)=0.5\mathrm{e}^{-|x|}$，$f(x)$ 可表述为

$$f(x)=0.5f_l(x)+0.5f_r(x)$$

3）舍选法

这种方法的实质是从许多均匀随机数中选出一部分，使之成为具有给定分布的随机数。给定随机变量的密度函数 $f(x)$，设 $f(x)$ 集中于有限区间 (a,b) 且在 (a,b) 上有限，即 $\int_a^b f(x)\mathrm{d}x=1$，于是可选取正常数 α，使得对一切 $x\in(a,b)$ 都满足 $\alpha f(x)<1$。又设 ξ、η 是两个独立的随机变数，ξ 在 (a,b) 上均匀分布，而 η 在 $(0,1)$ 上均匀分布，则

$$P(\xi\leqslant d\mid \alpha f(\xi)\geqslant\eta)=\int_a^d f(x)\mathrm{d}x\quad(a<d<b)$$

产生方法为

（1）产生 (a,b) 上均匀分布的随机数 ξ。

（2）产生 $(0,1)$ 上均匀分布的随机数 η。

（3）若 $\eta\leqslant\alpha f(\xi)$，则返回 ξ，否则重新产生 ξ、η。

若 $f(x)$ 不是集中于有限区间，假设为 $(-\infty,+\infty)$，可选择任意一个在 $(-\infty,+\infty)$ 分布的函数 $g(x)$，若 $\alpha f(x)/g(x)<1$，将上述方法改为产生按 $g(x)$ 分布的随机数 ξ，要求 $\eta\leqslant\alpha f(\xi)/g(\xi)$。

$$P(\xi\leqslant d\mid\eta\leqslant\alpha f(\xi)/g(\xi))=\frac{P(\xi\leqslant d,\eta\leqslant\alpha f(\xi)/g(\xi))}{P(\eta\leqslant\alpha f(\xi)/g(\xi))}$$

$$=\frac{\int_{-\infty}^{d}\int_{0}^{\alpha f(x)/g(x)}g(x)\mathrm{d}x\mathrm{d}y}{\int_{-\infty}^{+\infty}\int_{0}^{\alpha f(x)/g(x)}g(x)\mathrm{d}x\mathrm{d}y}=\int_{-\infty}^{d}f(x)\mathrm{d}x=F(x)$$

随机数产生的效率：按照 $g(x)$产生的随机数，被接受的概率

$$P(\eta\leqslant\alpha f(\xi)/g(\xi))=\int_{-\infty}^{+\infty}\int_{0}^{\alpha f(x)/g(x)}g(x)\mathrm{d}x\mathrm{d}y=\alpha\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\alpha$$

由 $\alpha f(x)/g(x)<1$，得 $\alpha=\min g(x)/f(x)$。由于$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\int_{-\infty}^{+\infty}g(x)\mathrm{d}x=1$，所以必然存在 x，$g(x)\leqslant f(x)$，$\alpha\leqslant 1$。$g(x)$ 与 $f(x)$ 越接近，随机数产生的效率越高。

对无穷区间的随机变量需要求出 $\alpha\leqslant\min g(x)/f(x)$，在很多情况下无法得到 α，使得舍选法无法使用。在无穷区间采样时，舍选法按 $g(x)$产生的随机数 ξ 被选中的概率与 $f(\xi)$成正比，与 $g(\xi)$成反比，基本可以抵消 ξ 产生的概率密度 $g(\xi)$。这样，选中的随机数 ξ 基本上符合分布 $f(\xi)$。按上述思路，改进的舍选法有限样本法采取的策略是：按照按 $g(x)$产生随机数 ξ_i，$i=1,2,\cdots,n$，然后按照与 $f(\xi)$成正比，与 $g(\xi)$成反比的概率从 $\Xi=\{\xi_i\mid i=1,2,\cdots,n\}$中选择 ξ_i，选出的数值 $\eta=\xi_i$ 服从 $f(x)$分布。

有限样本法的具体步骤：按照 $g(x)$ 分布产生随机数 ξ_i，$i=1,2,\cdots,n$，定义 $\omega_i=f(\xi_i)/g(\xi_i)$，将 ω_i 归一化得$q_i=\omega_i/\sum_{j=1}^{n}\omega_j$，定义离散随机变量 η 的分布 $p(\eta=\xi_i)=q_i$，则 η 的分布近似为 $f(x)$，即 $p(\eta\leqslant a)\approx\int_{-\infty}^{a}f(x)\mathrm{d}x$。

3.1.2　假设检验

参数空间 $\Theta=\{\theta\}$的非空子集或有关参数 θ 的命题称为统计假设，简称假设。原假设即根据需要而设立的假设，常记为 H_0 ：$\theta\in\Theta$o。备择假设即在原假设被拒绝后而采用(接受)的假设，常记为 H_1 ： $\theta\in\Theta 1$。对原假设 H_0 ：$\theta\in\Theta$o 作出判断的法则称为检验法则，简称检验。检验有两个结果：“原假设不正确”，称为拒绝原假设，或称检验显著；“原假设正确”，称为接受原假设，或称检验不显著。

两类错误及其发生的错误。原假设 H_0 正确，但被拒绝，这种判断错误称为第一类错误，其发生概率称为犯第一类错误的概率，或称拒真概率，常记为 α；原假设 H_0 不正确，但被接受，这种判断错误称为第二类错误，其发生概率称为犯第二类错误的概率，或称受伪概率，常记为 β。

假设检验的基本步骤如下。

(1) 建立假设。根据要求建立原假设 H_0 和备择假设 H_1。

(2) 选择检验统计量，给出拒绝域 W 的形式。

- 用于对原假设 H_0 作出判断的统计量称为检验统计量。
- 使原假设被拒绝的样本观察值所在区域称为拒绝域，常用 W 表示。
- 一个拒绝域 W 唯一确定一个检验法则，反之，一个检验法则唯一确定一个拒绝

域 W。

（3）选择显著性水平 $\alpha(0<\alpha<1)$。只控制犯第一类错误的概率不超过 α 的检验称为水平为 α 的检验，或称为显著性检验，但也不能使 α 过小，在适当控制 α 中制约 β，最常用的 $\alpha=0.05$，有时也选择 $\alpha=0.10$ 或者 $\alpha=0.01$。

（4）给出拒绝域。由概率等式 $P(W)=\alpha$ 确定具体的拒绝域。

（5）作出判断。

- 当样本 $(x_1,\cdots,x_n)\in W$，则拒绝 H_0，即接受 H_1。
- 当样本 $(x_1,\cdots,x_n)\in \overline{W}$，则接受 H_0。

表 3-1 为单个正态总体均值的假设检验。

表 3-1　单个正态总体均值的假设检验

检验法	条件	原假设 H_0	备择假设 H_1	检验统计量	拒绝域
u 检验	σ 已知	$u\leqslant u_0$	$u>u_0$	$u=\dfrac{\bar{x}-u_0}{\frac{\sigma}{\sqrt{n}}}$	$\{u\geqslant u_{1-\alpha}\}$
		$u\geqslant u_0$	$u<u_0$		$\{u\leqslant u_\alpha\}$
		$u=u_0$	$u\neq u_0$		$\{\lvert u\rvert\geqslant u_{1-\frac{\sigma}{2}}\}$
t 检验	σ 未知	$u\geqslant u_0$	$u>u_0$	$t=\dfrac{\bar{x}-u_0}{\frac{s}{\sqrt{n}}}$	$\{t\geqslant t_{1-\alpha}(n-1)\}$
		$u\leqslant u_0$	$u<u_0$		$\{t\leqslant t_\alpha(n-1)\}$
		$u=u_0$	$u\neq u_0$		$\{\lvert t\rvert\geqslant t_{1-\frac{\alpha}{2}}(n-1)\}$

表 3-2 为单个正态总体均值差的假设检验。

表 3-2　单个正态总体均值差的假设检验

检验法	条件	原假设 H_0	备择假设 H_1	检验统计量	拒绝域
u 检验	σ_1、σ_2 已知	$u\leqslant u_0$ $u\geqslant u_0$ $u=u_0$	$u>u_0$ $u<u_0$ $u\neq u_0$	$u=\dfrac{(\bar{x}-\bar{y})}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$	$\{u\geqslant u_{1-\alpha}\}$ $\{u\leqslant u_\alpha\}$ $\{\lvert u\rvert\geqslant u_{1-\frac{\sigma}{2}}\}$
t 检验	σ_1、σ_2 未知 $\sigma_1=\sigma_2$	$u\leqslant u_0$ $u\geqslant u_0$ $u=u_0$	$u>u_0$ $u<u_0$ $u\neq u_0$	$t=\dfrac{(\bar{x}-\bar{y})}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$	$\{t\geqslant t_{1-\alpha}(m+n-2)\}$ $\{t\leqslant t_\alpha(m+n-2)\}$ $\{\lvert t\rvert\geqslant t_{1-\frac{\alpha}{2}}(m+n-2)\}$
大样本 u 检验	σ_1、σ_2 已知，m、n 充分大	$u\leqslant u_0$ $u\geqslant u_0$ $u=u_0$	$u>u_0$ $u<u_0$ $u\neq u_0$	$u=\dfrac{(\bar{x}-\bar{y})}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}$	$\{u\geqslant u_{1-\alpha}\}$ $\{u\leqslant u_{1-\alpha}\}$ $\{\lvert u\rvert\geqslant u_{1-\frac{\sigma}{2}}\}$
近似 t 检验	σ_1、σ_2 未知，m、n 不很大	$u\leqslant u_0$ $u\geqslant u_0$ $u=u_0$	$u>u_0$ $u<u_0$ $u\neq u_0$	$t=\dfrac{(\bar{x}-\bar{y})}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}$	$\{t\geqslant t_{1-\alpha}(l)\}$ $\{t\leqslant t_\alpha(l)\}$ $\{\lvert t\rvert\geqslant t_{1-\frac{\alpha}{2}}(l)\}$

3.1.3　常用的统计模型

在一个统计问题中，研究对象的全体称为总体，构成总体的每个成员称为个体。若关

心的是总体中每个个体的一个数量指标，则该总体称为一维分布。若关心的是总体中的每个个体的两个数量指标，则该总体称为二维总体。二维总体就是一个二维分布，以此类推。

不含未知参数的样本函数称为统计量。统计量的分布称为抽样分布。从总体中随机抽取的部分个体构成的集合称为样本空间，样本空间中的个体称为样本，样本个数称为样本容量或样本量。样本常用 n 个指标值 $x_1,x_2,\cdots,x_n$ 表示。

若将样本观测值 $x_1,x_2,\cdots,x_n$ 由小到大排列，得有序样本 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，用有序样本定义如下函数

$$F_n(x)=\begin{cases}0, & x<x_{(1)}\\ k/n, & x_{(k)}\leqslant x<x_{(k+1)},k=1,2,\cdots,n-1\\ 1, & x\geqslant x_{(n)}\end{cases}$$

则称 $F_n(x)$ 为该样本的经验分布函数。

格里纹科定理　设 $x_1,x_2,\cdots,x_n$ 是取自总体分布函数为 $F(x)$ 的样本，$F_n(x)$ 是该样本的经验分布函数，则当 $n\to\infty$ 时：$P(\sup\limits_{-\infty<x<\infty}|F_n(x)-F(x)|)=1$。

此定理表明：当 n 相当大时，经验分布函数 $F_n(x)$ 是总体分布函数 $F(x)$ 的一个良好的近似。该定理是经典统计学的一块基石。

样本均值　样本 $x_1,\cdots,x_n$ 的算术平均值称为样本均值，记为 $\bar{x}$。

分组样本均值：$\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{k}x_if_i$，其中 n 为样本量，k 为组数，x_i 与 f_i 为第 i 组的组中值与频数，分组样本均值是完全样本均值的一种较好的近似。样本均值的性质：

(1) $\sum\limits_{i-1}^{n}(x_i-\bar{x})=0$，样本数据 x_i 对样本均值 $\bar{x}$ 的偏差之和为零。

(2) 样本数据 x_i 与样本均值 $\bar{x}$ 的偏差平方之和最小，即对任意实数 c，都有

$$\sum_{i=1}^{n}(x_i-\bar{x})^2\leqslant\sum_{i=1}^{n}(x_i-c)^2$$

(3) 若总体分布为 $N(\mu,\sigma^2)$，则 $\bar{x}$ 的精确分布为 $N(\mu,\sigma^2/n)$。

(4) 若总体分布未知，但其期望 μ 与方差 σ^2 存在，则当 n 较大时，$\bar{x}$ 的渐近分布为 $N(\mu,\sigma^2/n)$，这里渐近分布是指 n 较大时的近似分布。

样本方差与样本标准差　样本方差有两个，样本方差 s^{*2} 与样本无偏方差 s^2 如下。

$$s^{*2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,\quad s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

实际中常用的是无偏样本方差 s^2。这是因为当 σ^2 为总体方差时，总有

$$E[s^{*2}]=\frac{n-1}{n}\sigma^2,\quad E[s^2]=\sigma^2$$

这表明：s^{*2} 有系统偏小的误差，而 s^2 无此系统偏差。今后称 s^2 为样本方差，称 $s=\sqrt{s^2}$ 为样本标准差。

s^2 的计算有如下 3 个公式可供选用。

$$s^2=s^2=\frac{1}{n-1}\sum(x_i-\bar{x})^2=\frac{1}{n-1}\left[\sum x_i{}^2-\frac{(\sum x_i)^2}{n}\right]=\frac{1}{n-1}\left[\sum x_i{}^2-n\bar{x}^2\right]$$

三大抽样分布：χ^2 分步、F 分布、t 分布

设 $x_1,x_2,\cdots,x_n$ 和 $y_1,y_2,\cdots,y_n$ 是来自标准正态分布的两个相互独立的样本，则此 3 个统计量的构造及其抽样分布见表 3-3。

表 3-3　三大抽样分布及其数字特征

统计量的构造	抽样分布密度函数	期望	方差
$\chi^2=x_1^2+x_2^2+\cdots+x_n^2$	$p(y)=\dfrac{1}{\Gamma\left(\frac{n}{2}\right)2^{n/2}}y^{\frac{n}{2}-1}\mathrm{e}^{-\frac{y}{2}}I_{\{y>0\}}$	n	$2n$
$F=\dfrac{(y_1^2+y_2^2+\cdots+y_m^2)/m}{(x_1^2+x_2^2+\cdots+x_n^2)/n}$	$p(y)=\dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\dfrac{m}{n}\right)^{m/2}y^{\frac{m}{2}-1}$ $(1+\dfrac{m}{n}y)^{-\frac{m+n}{2}}I_{\{y>0\}}$	$\dfrac{n}{n-2}(n>2)$	$\dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$
$t=\dfrac{y_1}{\sqrt{(x_1^2+x_2^2+\cdots+x_n^2)/n}}$	$p(y)=\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}(1+\dfrac{1}{n}y^2)^{-\frac{n+1}{2}}$	$0(n>1)$	$\dfrac{n}{n-2}(n>2)$

今后正态总体参数的置信区间与假设检验大多数将基于这三大抽样分布。

一个重要的定理

设 $x_1,\cdots,x_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，其样本均值与样本方差分别为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad s^2=\frac{1}{n-1}\sum(x_i-\bar{x})^2$$

则有①$\bar{x}$ 与 s^2 相互独立；②$\bar{x}\sim N(\mu,\delta^2/n)$；③$\dfrac{(n-1)s^2}{\delta^2}\sim\chi^2(n-1)$。

一些重要推论

(1) 设 $x_1,x_2,\cdots,x_n$ 是来自正态总体 $N(\mu,\delta^2)$ 的样本，则有 $t=\dfrac{\sqrt{n}(\bar{x}-\mu)}{s}\sim t(n-1)$。

其中 $\bar{x}$ 为样本均值，为样本标准差。

(2) 设 $x_1,x_2,\cdots,x_m$ 是来自 $N(\mu_1,{\delta_1}^2)$ 的样本，$y_1,y_2,\cdots,y_n$ 是来自 $N(\mu_2,{\delta_2}^2)$ 的样本，且此两个样本相互独立，则有 $F=\dfrac{{s_x}^2/{\delta_1}^2}{{s_y}^2/{\delta_2}^2}\sim F(m-1,n-1)$。

其中 ${s_x}^2$、${s_y}^2$ 分别是两个样本的方差。若 ${\delta_1}^2={\delta_2}^2$，则 $F={s_x}^2/{s_y}^2\sim F(m-1,n-1)$。

3.2　正交投影与最小二乘估计

3.2.1　引言

首先看一个过原点的直线拟合问题。若 $y=ax+w$，w 为随机误差，给出 n 个观测值 (y_i,x_i)，估计参数 a 使 $e=\sum_{i=1}^{n}(y_i-ax_i)^2$ 最小。

由$\frac{\mathrm{d}e}{\mathrm{d}a}=-2\sum_{i=1}^{n}(y_i-ax_i)x_i=0$得$\sum_{i=1}^{n}(y_ix_i-ax_i^2)=0$，即$\hat{a}=\frac{\sum_{i=1}^{n}y_ix_i}{\sum_{i=1}^{n}x_i^2}$

设$\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$，$\boldsymbol{X}=(x_1,x_2,\cdots x_n)^{\mathrm{T}}$，$e=(\boldsymbol{Y}-\hat{a}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{Y}-\hat{a}\boldsymbol{X})$，若$\boldsymbol{Y}=a\boldsymbol{X}$，则$e=0$。否则，求$a$的估计，使$e$最小。

在n维空间中，$\boldsymbol{Y}=a\boldsymbol{X}$意味着$\boldsymbol{Y}$与$\boldsymbol{X}$在同一维空间中，如图3-1(a)所示。若$\boldsymbol{Y}\neq a\boldsymbol{X}$，如图3-1(b)所示，求$a$的估计，使$e=(\boldsymbol{Y}-\hat{a}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{Y}-\hat{a}\boldsymbol{X})$最小。显然，仅当误差$\boldsymbol{Y}-\hat{a}\boldsymbol{X}$与矢量$\boldsymbol{X}$垂直(正交)时，$e=(\boldsymbol{Y}-\hat{a}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{Y}-\hat{a}\boldsymbol{X})$达到最小。

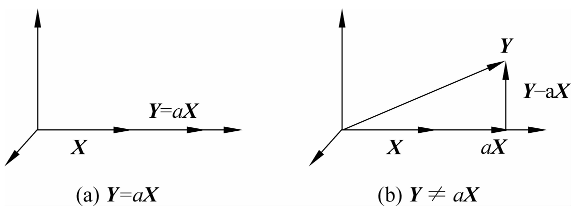

图 3-1　$\boldsymbol{Y}$与$a\boldsymbol{X}$的关系

问题转化为求a的估计$\hat{a}$，使$\boldsymbol{Y}-\hat{a}\boldsymbol{X}$与$\boldsymbol{X}$正交：$(\boldsymbol{Y}-a\boldsymbol{X})^{\mathrm{T}}\boldsymbol{X}=0$，由此得

$$(\boldsymbol{Y}-a\boldsymbol{X})^{\mathrm{T}}\boldsymbol{X}=\sum_{i=1}^{n}y_ix_i-a\sum_{i=1}^{n}x_i^2=0$$

从而有$\hat{a}=\frac{\sum_{i=1}^{n}y_ix_i}{\sum_{i=1}^{n}x_i^2}$。

上面给出了过原点直线的拟合结果，利用n个观测值(y_i,x_i)估计参数a。对于一般的直线，设$y=ax+b+w$，w为随机误差，由n个观测值(y_i,x_i)估计两个参数a、b，使$e=\sum_{i=1}^{n}(y_i-\hat{a}x_i-\hat{b})^2$最小。显然，$e$是待估计参数$\hat{a}$、$\hat{b}$的函数，通常可用求导数的方法求其极值。

由$\begin{cases}\dfrac{\partial e}{\partial a}=-2\sum\limits_{i=1}^{n}(y_i-ax_i-b)x_i=0\\ \dfrac{\partial e}{\partial b}=-2\sum\limits_{i=1}^{n}(y_i-ax_i-b)=0\end{cases}$得$\begin{cases}\sum\limits_{i=1}^{n}y_ix_i-a\sum\limits_{i=1}^{n}x_i^2-b\sum\limits_{i=1}^{n}x_i=0\\ \sum\limits_{i=1}^{n}y_i-a\sum\limits_{i=1}^{n}x_i-nb=0\end{cases}$

因此，$\begin{cases}\hat{a}=\dfrac{\sum\limits_{i=1}^{n}y_ix_i-\left(\sum\limits_{i=1}^{n}x_i\right)\left(\sum\limits_{i=1}^{n}y_i\right)}{n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2}\\ \hat{b}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i-\dfrac{\hat{a}}{n}\sum\limits_{i=1}^{n}x_i\end{cases}$

对于直线$y=ax+b$的参数估计，由n个观测值，可以得到

$$\begin{cases} y_1 = ax_1 + b \\ y_2 = ax_2 + b \\ \quad\vdots \\ y_n = ax_n + b \end{cases}$$

写成矩阵形式为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

设 $\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$，$\boldsymbol{A} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$，$\boldsymbol{Z} = \begin{pmatrix} a \\ b \end{pmatrix}$，则方程可写成 $\boldsymbol{Y} = \boldsymbol{AZ}$。

如果没有观测噪声，则对所有 $1 \leqslant i \leqslant n$，$y_i = ax_i + b$，由 2 个观测值即可求得参数 a、b。若存在观测噪声，则方程 $\boldsymbol{Y} = \boldsymbol{AZ}$ 无解，此时该方程称为矛盾方程。对于矛盾方程，求 $\boldsymbol{Z}$，使得 $(\boldsymbol{Y} - \boldsymbol{AZ})^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{AZ})$ 最小。此时得到的解称为矛盾方程的最小二乘解。

设 $\alpha_1 = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，$\alpha_2 = (1, 1, \cdots, 1)^{\mathrm{T}}$，则 $\boldsymbol{A} = (\alpha_1, \alpha_2)$。由 $\boldsymbol{Y} = \boldsymbol{AZ}$ 得 $\boldsymbol{Y} = a\alpha_1 + b\alpha_2$，即 $\boldsymbol{Y}$ 是矩阵 $\boldsymbol{A}$ 列矢量的线性组合，或者说存在矢量 $\boldsymbol{Z}$，使得 $\boldsymbol{Y} = \boldsymbol{AZ}$，也可以讲 $\boldsymbol{Y}$ 属于 $\boldsymbol{A}$ 列矢量构成的空间，记为 $\boldsymbol{Y} \in R(\boldsymbol{A})$，如图 3-2(a)所示。若对所有 $\boldsymbol{Z}$，$\boldsymbol{Y} \neq \boldsymbol{AZ}$，则方程无解，或者说 $\boldsymbol{Y} \notin R(\boldsymbol{A})$，如图 3-2(b)所示。

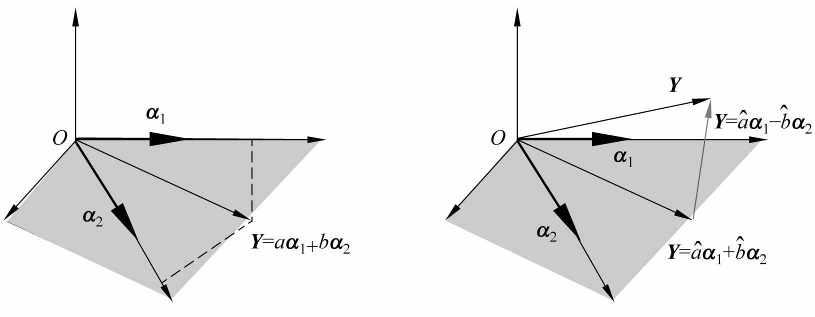

图 3-2　$\boldsymbol{Y}$ 与 $R(\boldsymbol{A})$的关系

显然，求 $\boldsymbol{Z} = (a, b)^{\mathrm{T}}$ 的估计，使 $(\boldsymbol{Y} - \boldsymbol{AZ})^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{AZ})$ 最小，就是使矢量 $\boldsymbol{Y} - (\hat{a}\boldsymbol{\alpha}_1 + \hat{b}\boldsymbol{\alpha}_2)$ 最短(范数最小)。直观上看，仅当矢量 $\boldsymbol{Y} - (\hat{a}\boldsymbol{\alpha}_1 + \hat{b}\boldsymbol{\alpha}_2)$ 与 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$ 所在平面垂直时，$\boldsymbol{Y} - (\hat{a}\boldsymbol{\alpha}_1 + \hat{b}\boldsymbol{\alpha}_2)$ 最短。也就是说，当误差矢量 $\boldsymbol{Y} - (\hat{a}\boldsymbol{\alpha}_1 + \hat{b}\boldsymbol{\alpha}_2)$ 与矢量 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_2$ 都垂直时，或者说当误差矢量 $\boldsymbol{Y} - (\hat{a}\boldsymbol{\alpha}_1 + \hat{b}\boldsymbol{\alpha}_2)$ 与矢量 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_2$ 都正交时，$(\boldsymbol{Y} - \boldsymbol{AZ})^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{AZ})$ 达到最小，有

$$\begin{cases} (\boldsymbol{Y} - \boldsymbol{AZ})^{\mathrm{T}}\alpha_1 = 0 \\ (\boldsymbol{Y} - \boldsymbol{AZ})^{\mathrm{T}}\alpha_2 = 0 \end{cases}，得$$

$$\begin{cases}\sum_{i=1}^{n}(y_i-\hat{a}x_i-\hat{b})x_i=0\\ \sum_{i=1}^{n}(y_i-\hat{a}x_i-\hat{b})=0\end{cases}，解得\begin{cases}\hat{a}=\dfrac{\sum_{i=1}^{n}y_ix_i-(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}\\ \hat{b}=\dfrac{1}{n}\sum_{i=1}^{n}y_i-\dfrac{\hat{a}}{n}\sum_{i=1}^{n}x_i\end{cases}$$

利用正交性原理，得到了相同的结果。显然，利用正交性原理，不仅运算简单，而且更加直观。

这样，利用正交性原理，解决了直线的最小二乘拟合问题。从前面的讨论可知，直线的拟合问题可以归结为矛盾方程组最小二乘解问题。

下面讨论一般的方程组 $\boldsymbol{Y}=\boldsymbol{AX}$ 求解问题。这里，$\boldsymbol{A}=(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 为 $m\times n$ 的矩阵，$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$。我们记 $R(\boldsymbol{A})$ 为变换（矩阵）$\boldsymbol{A}$ 的值域，即 $R(\boldsymbol{A})=\{y\mid\exists_x y=\boldsymbol{A}x\}$。显然，$R(\boldsymbol{A})$ 中的矢量都是 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的线性组合，$R(\boldsymbol{A})$ 就是 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 构成的空间。

若 $\boldsymbol{Y}\in R(\boldsymbol{A})$，则方程有解，表示存在 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，满足 $\boldsymbol{Y}=\sum_{i=1}^{n}x_i\alpha_i$，或者说，$\boldsymbol{Y}$ 是 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的线性组合，或等同地讲，$\boldsymbol{Y}$ 属于 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 构成的空间。

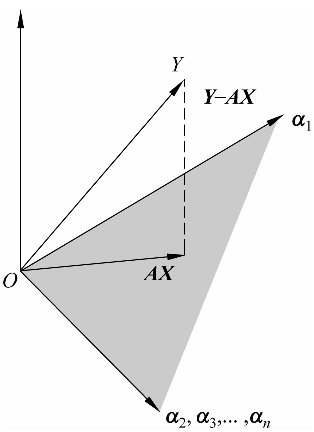

图 3-3　$\boldsymbol{Y}-\boldsymbol{AX}$ 的几何意义

若方程无解，则说明不存在 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，满足 $\boldsymbol{Y}=\sum_{i=1}^{n}x_i\alpha_i$，或者说，$\boldsymbol{Y}$ 不是 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的线性组合（$\boldsymbol{Y}$ 不属于 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 构成的空间）。对于矛盾方程组，我们的目标是求使 $\|(\boldsymbol{Y}-\boldsymbol{AX})\|=(\boldsymbol{Y}-\boldsymbol{AX})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{AX})$ 最小的最小二乘解。

我们用 $P_{R(\boldsymbol{A})}$ 表示将矢量投影到空间 $R(\boldsymbol{A})$ 的投影变换，$P_{R(\boldsymbol{A})}\boldsymbol{Y}$ 就是矢量 $\boldsymbol{Y}$ 在 $R(\boldsymbol{A})$ 上的投影，可以表示成 $\boldsymbol{A}$ 的线性组合，即 $P_{R(\boldsymbol{A})}\boldsymbol{Y}\in R(\boldsymbol{A})$，方程 $\boldsymbol{AX}=P_{R(\boldsymbol{A})}\boldsymbol{Y}$ 有解。显然，最小二乘解 $\boldsymbol{X}$ 满足 $\boldsymbol{AX}=P_{R(\boldsymbol{A})}\boldsymbol{Y}$，如图 3-3 所示。有了投影矩阵 $P_{R(\boldsymbol{A})}$，就可以得到矛盾方程组的最小二乘解。

3.2.2　内积空间

本科线性代数教程已介绍过线性空间的基本概念。但是，基本的线性空间概念仅包含矢量的加法及标量与矢量的乘法。本节介绍矢量空间中另一种基本的运算——矢量的内部乘积（简称内积）。

如图 3-4 所示，设 $\boldsymbol{a}$、$\boldsymbol{b}$ 为二维坐标中的两个矢量，其顶点坐标分别为 (x_1,y_1) 和 (x_2,y_2)，则矢量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的长度分别为 $\|\boldsymbol{a}\|=\sqrt{x_1^2+y_1^2}$ 和 $\|\boldsymbol{b}\|=\sqrt{x_2^2+y_2^2}$。

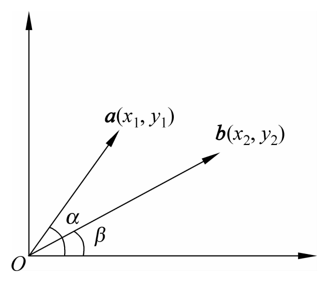

图 3-4　矢量关系

$$x_1=\|\boldsymbol{a}\|\cos\alpha,\ y_1=\|\boldsymbol{a}\|\sin\alpha;$$
$$x_2=\|\boldsymbol{b}\|\cos\beta,\ y_2=\|\boldsymbol{b}\|\sin\beta;$$

矢量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 之间的夹角余弦为

$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta=\frac{x_1x_2+y_1y_2}{||\boldsymbol{a}||\cdot||\boldsymbol{b}||}。$$

数值 $x_1x_2+y_1y_2=||\boldsymbol{a}||\cdot||\boldsymbol{b}||\cos(\alpha-\beta)$ 为矢量 $\boldsymbol{a}$ 在矢量 $\boldsymbol{b}$ 方向投影的长度 $||\boldsymbol{a}||\cos(\alpha-\beta)$ 与矢量 $\boldsymbol{b}$ 长度的乘积，也可以解释为矢量 $\boldsymbol{b}$ 在矢量 $\boldsymbol{a}$ 方向投影的长度 $||\boldsymbol{b}||\cos(\alpha-\beta)$ 与矢量 $\boldsymbol{a}$ 长度的乘积。如果矢量 $\boldsymbol{a}$ 或矢量 $\boldsymbol{b}$ 为单位长度，则数值 $x_1x_2+y_1y_2$ 表示了一个矢量在一个单位矢量方向的投影长度。若矢量 $\boldsymbol{a}$ 和矢量 $\boldsymbol{b}$ 的夹角为 0，则 $x_1x_2+y_1y_2=||\boldsymbol{a}||\cdot||\boldsymbol{b}||$。

$x_1x_2+y_1y_2$ 是两个矢量之间的一种运算，我们将其定义为矢量 $\boldsymbol{a}$ 和矢量 $\boldsymbol{b}$ 的内积 $<\boldsymbol{a},\boldsymbol{b}>=x_1x_2+y_1y_2$。有了内积的定义，矢量 $\boldsymbol{a}$、$\boldsymbol{b}$ 间夹角的余弦可表示为

$$\cos(\alpha-\beta)=\frac{<\boldsymbol{a},\boldsymbol{b}>}{||\boldsymbol{a}||\cdot||\boldsymbol{b}||}$$

若矢量 $\boldsymbol{a}$ 和矢量 $\boldsymbol{b}$ 相互垂直（正交），则 $<\boldsymbol{a},\boldsymbol{b}>=0$。矢量 $\boldsymbol{a}$ 的长度可表示为 $||\boldsymbol{a}||=\sqrt{<\boldsymbol{a},\boldsymbol{a}>}$。

我们推广二维空间的定义，定义 n 维空间的两个矢量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$、$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$ 的内积为 $<\boldsymbol{x},\boldsymbol{y}>=x^{\mathrm{T}}y=x_1y_1+x_2y_2+\cdots x_ny_n$，同样定义矢量间的夹角为 $\arccos\dfrac{<\boldsymbol{x},\boldsymbol{y}>}{||\boldsymbol{x}||\cdot||\boldsymbol{y}||}$。

不难验证，按照上述定义的内积 $<\boldsymbol{x},\boldsymbol{y}>=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=x_1y_1+x_2y_2+\cdots+x_ny_n$ 满足交换律、分配率、齐次性和非负性，即

（1）交换律：$<\boldsymbol{x},\boldsymbol{y}>=<\boldsymbol{y},\boldsymbol{x}>$。

（2）分配率：$<\boldsymbol{x},\boldsymbol{y}+\boldsymbol{z}>=<\boldsymbol{x},\boldsymbol{y}>+<\boldsymbol{x},\boldsymbol{z}>$。

（3）齐次性：$<k\boldsymbol{x},\boldsymbol{y}>=k<\boldsymbol{x},\boldsymbol{y}>$。

（4）非负性：$<\boldsymbol{x},\boldsymbol{x}>\geqslant 0$，当且仅当 $x=0$ 时，$<\boldsymbol{x},\boldsymbol{x}>=0$。

作为上述运算的推广，对线性空间上的任意二元运算，只要满足满足交换律、分配率、齐次性和非负性，都可以将其定义为空间上矢量的内积。

定义 3-1　设 V 是实数域 R 上的线性空间，V 中任意两个矢量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的内积 $<\boldsymbol{x},\boldsymbol{y}>$ 都为满足下列 4 个条件的 $V\times V\to R$ 映射。

（1）交换律：$<\boldsymbol{x},\boldsymbol{y}>=<\boldsymbol{y},\boldsymbol{x}>$。

（2）分配率：$<\boldsymbol{x},\boldsymbol{y}+\boldsymbol{z}>=<\boldsymbol{x},\boldsymbol{y}>+<\boldsymbol{x},\boldsymbol{z}>$。

（3）齐次性：$<k\boldsymbol{x},\boldsymbol{y}>=k<\boldsymbol{x},\boldsymbol{y}>$。

（4）非负性：$<\boldsymbol{x},\boldsymbol{x}>\geqslant 0$，当且仅当 $\boldsymbol{x}=0$ 时，$<\boldsymbol{x},\boldsymbol{x}>=0$。

所谓**欧几里得空间**，就是定义了内积的**实**线性空间，又称为**内积空间**或**欧氏空间**。而定义了内积的**复线性空间**称为**酉空间**。酉空间的理论与欧氏空间的理论很接近，有一套平行的理论。如不特别声明，我们所说的空间主要指欧氏空间。

有了内积的定义，同样可以定义矢量 $\boldsymbol{x}$、$\boldsymbol{y}$ 的夹角为 $\arccos\dfrac{<\boldsymbol{x},\boldsymbol{y}>}{||\boldsymbol{x}||\cdot||\boldsymbol{y}||}$，矢量间的距离为 $\|\boldsymbol{x}-\boldsymbol{y}\|=\sqrt{<\boldsymbol{x}-\boldsymbol{y},\boldsymbol{x}-\boldsymbol{y}>}$。

由定义 3-1 可知，内积是我们直观上三维空间运算的推广，并不特指某一个运算，而是指满足条件的任意二元运算。有了内积的定义，就可以定义矢量的长度（范数）和矢量间的距离和夹角了。

设 x、y 是两个二阶矩随机变量，不难验证，运算 $E(xy)$ 也满足内积性质，即

（1）交换律：$E(xy)=E(yx)$。

（2）分配率：$E(x(y+z))=E(xy)+E(xz)$。

（3）齐次性：$E(kxy)=kE(xy)$。

（4）非负性：$E(x^2)\geqslant 0$，$E(x^2)=0$ 的充要条件是 $x\overset{a.s}{=}0$。

这里，$x\overset{a.s}{=}0$ 表示 x 以概率 1 趋近于 0。

在随机过程中，我们知道，全体二阶矩随机变量组成的集合按照通常定义的随机变量的加法及数乘构成线性空间。如果定义内积 $<\boldsymbol{x},>aE(xy)$，则该线性空间构成欧几里得空间。同时，该空间具有完备性，可以证明它是一个 Hilbert 空间。

3.2.3 正交投影矩阵

1. 线性子空间

在直角坐标系中，通常使用一个 3 元组 (x,y,z) 表示一个点的坐标。实际上，这个点也可以用一个二维坐标 (x,y) 和一个一维坐标 z 表示。很容易验证，XOY 平面和 Z 轴构成了三维空间的二维和一维子空间。实际上，在三维空间中，任何一个过原点的直线都构成了一个一维子空间；同样，任何一个过原点的平面都构成了一个二维子空间。只要过原点的直线不在这个平面上，则这条直线和平面上的任何两条直线都可以构成三维空间的基。这条直线和该平面是三维几何空间的一个部分，它们对于原来的运算也都构成一个线性空间，因此可以给出线性子空间的定义。

定义 3-2 设 V_1 是数域 K 上的线性空间 V 的一个非空子集，且 V 已有的线性运算满足以下条件：

（1）若 $x,y\in V_1$，则 $x+y\in V_1$；

（2）若 $x\in V_1$，$k\in K$，则 $kx\in V_1$。

则称 V_1 为 V 的**线性子空间**或**子空间**。

显然，任何一个线性子空间 V_1 的维数都不大于整个线性空间 V 的维数，即 $\dim V_1\leqslant\dim V$。

例如，设 $\boldsymbol{A}$ 为一个 $m\times n$ 阶的实数矩阵，假设 $m\geqslant n$，其列矢量为 $\alpha_1,\alpha_2,\cdots,\alpha_n$。我们记 $R(\boldsymbol{A})$ 为变换（矩阵）$\boldsymbol{A}$ 的值域，即 $R(\boldsymbol{A})=\{y\mid\exists_x y=\boldsymbol{A}x\}$。显然，$R(\boldsymbol{A})$ 中的矢量都是 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 的线性组合，$R(\boldsymbol{A})$ 就是 m 维空间中 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 构成的一个 n 维子空间。对于任意一个 m 维矢量 $\boldsymbol{b}$，$\boldsymbol{b}\in R(\boldsymbol{A})$，当且仅当方程 $\boldsymbol{A}x=b$ 有解。显然，$R(\boldsymbol{A})\subset R^m$，其维数等于矩阵 $\boldsymbol{A}$ 的秩，即 $\dim R(\boldsymbol{A})=\operatorname{rank}\boldsymbol{A}$。由于 $\operatorname{rank}\boldsymbol{A}=\operatorname{rank}\boldsymbol{A}^{\mathrm{T}}$，则 $\dim R(\boldsymbol{A})=\dim R(\boldsymbol{A}^{\mathrm{T}})=\operatorname{rank}\boldsymbol{A}$。

容易看出，每个非零线性空间至少有两个子空间，一个是它自己，另一个是仅由零矢量构成的子集，通常称后者为**零子空间**。由于零子空间不含线性无关的矢量，因此它没有

基，规定其维数为零。

不难验证，任意 m 阶齐次线性方程组 $\boldsymbol{Ax}=0$ 的解空间也是 R^m 的子空间，称为 $\boldsymbol{A}$ 的**核空间**或**零空间**，记为 $N(\boldsymbol{A})$，$N(\boldsymbol{A})=\{x \mid \boldsymbol{A}x=0\}$。$\boldsymbol{A}$ 的核空间的维数，称为 $\boldsymbol{A}$ 的**零度**，记为 $n(\boldsymbol{A})$。容易证明，$n(\boldsymbol{A})=n-\mathrm{rank}\,\boldsymbol{A}$，$n(\boldsymbol{A}^{\mathrm{T}})=m-\mathrm{rank}\,\boldsymbol{A}^{\mathrm{T}}$。因而有

$$\mathrm{rank}\boldsymbol{A}+n(\boldsymbol{A})=n$$

$$\mathrm{rank}\boldsymbol{A}^{\mathrm{T}}+n(\boldsymbol{A}^{\mathrm{T}})=m$$

$$n(\boldsymbol{A})-n(\boldsymbol{A}^{\mathrm{T}})=n-m$$

2. 子空间的交与和

线性空间 V 的两个子空间 V_1、V_2 是集合 V 的两个子集。V_1、V_2 的并 $V_1\cup V_2$ 与交 $V_1\cap V_2$ 同样也是 V 的子集。那么，$V_1\cup V_2$ 与 $V_1\cap V_2$ 是否是 V 的子空间？

在三维空间中，任何过原点的平面都构成其二维子空间。设 V_1、V_2 是两个过原点的平面，对于 V_1、V_2 中的两个非零矢量 $\boldsymbol{x}_1\in V_1$，$\boldsymbol{x}_2\in V_2$，有 $\boldsymbol{x}_1\in V_1\cup V_2$，$\boldsymbol{x}_2\in V_1\cup V_2$，但 $\boldsymbol{x}_1+\boldsymbol{x}_2\notin V_1\cup V_2$，即 $V_1\cup V_2$ 对运算＋不封闭，不构成空间。因此，$V_1\cup V_2$ 是 V 的子集，但不是 V 的子空间。

对于交运算，我们知道，在三维空间中，任何两个过原点平面的交集都是一个过原点的直线，它是三维空间中的一个一维子空间。

对于一般情况，若 V_1、V_2 是空间 V 的子空间，则 $0\in V_1\cap V_2$。对于 $V_1\cap V_2$ 中任意两个矢量 $\boldsymbol{x}_1\in V_1\cap V_2$，$\boldsymbol{x}_2\in V_1\cap V_2$，有 $\boldsymbol{x}_1+\boldsymbol{x}_2\in V_1$，并且 $\boldsymbol{x}_1+\boldsymbol{x}_2\in V_2$，因此，$\boldsymbol{x}_1+\boldsymbol{x}_2\in V_1\cap V_2$，即 $V_1\cap V_2$ 对运算＋是封闭的。容易验证，$V_1\cap V_2$ 对数乘运算也是封闭的。因此，$V_1\cap V_2$ 是 V 的子空间。

定理 3-1　如果 V_1、V_2 是数域 K 上的线性空间 V 的两个子空间，那么它们的交 $V_1\cap V_2$ 也是 V 的子空间。

在三维空间中，任何一个矢量 (x,y,z) 都可以表示成 XOY 平面上的矢量 $(x,y,0)$ 和 Z 轴上的矢量 $(0,0,z)$ 之和，即 $(x,y,z)=(x,y,0)+(0,0,z)$。实际上，矢量 $(x,y,0)$ 就是 (x,y,z) 沿着 Z 轴在 XOY 平面上的投影；矢量 $(0,0,z)$ 就是 (x,y,z) 沿着 XOY 平面在 Z 轴上的投影。对于任何过原点的直线和平面，它们构成的子空间分别记为 V_1 和 V_2，只要直线不在平面上（V_1 不是 V_2 的子空间），则三维空间中的任何一个矢量都可以表示成空间 V_1 上一个矢量与 V_2 上一个矢量之和。同样，二维空间的任何一个矢量也可以表示成两个一维空间中矢量的和。

对于一般情况，设 V_1、V_2 都是数域 K 上的线性空间 V 的子空间，我们称 V_1 上元素与 V_2 上元素之和构成的集合为 V_1 与 V_2 的和，记为 $V_1+V_2=\{z \mid \exists_{x\in V_1}\ \exists_{y\in V_2}\ z=x+y\}$。

显然，$0\in V_1+V_2$。

对任何 $\boldsymbol{x}_1\in V_1+V_2$，$\boldsymbol{x}_2\in V_1+V_2$，都存在 $\boldsymbol{y}_1\in V_1$，$\boldsymbol{y}_2\in V_2$，$\boldsymbol{z}_1\in V_1$，$\boldsymbol{z}_2\in V_2$，满足 $\boldsymbol{x}_1=\boldsymbol{y}_1+\boldsymbol{y}_2$，$\boldsymbol{x}_2=\boldsymbol{z}_1+\boldsymbol{z}_2$。由于 $\boldsymbol{y}_1+\boldsymbol{z}_1\in V_1$，$\boldsymbol{y}_2+\boldsymbol{z}_2\in V_2$，因此 $\boldsymbol{x}_1+\boldsymbol{x}_2=(\boldsymbol{y}_1+\boldsymbol{z}_1)+(\boldsymbol{y}_2+\boldsymbol{z}_2)\in V_1+V_2$，即集合 V_1+V_2 对运算＋封闭。很容易验证，V_1+V_2 对数乘运算也是封闭的，因此 V_1+V_2 是空间 V 的子空间。

定理 3-2　若 V_1、V_2 都是数域 K 上的线性空间 V 的子空间，则它们的和 V_1+V_2 也

是 V 的子空间。

不难证明，若 $V_1 \subseteq W$，并且 $V_2 \subseteq W$，则 $V_1 + V_2 \subseteq W$，即 $V_1 + V_2$ 是包含 V_1 和 V_2 的最小子空间。

在三维空间中，如果两个平面既不重叠，也不平行，那么这两个平面的交就为一条直线（一维空间），三维空间中任何一个矢量都可以表示成这两个平面中矢量的和。也就是说，这两个二维平面的和空间构成三维空间，两个空间之和的维数等于两个空间维数之和减去其交空间的维数。对于一般情况，有定理 3-3。

定理 3-3（维数公式） 若 V_1、V_2 都是数域 K 上的线性空间 V 的子空间，则 $\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2)$。

证明：设 $\dim V_1 = n_1$，$\dim V_2 = n_2$，$\dim(V_1 \cap V_2) = m$，显然 $m \leqslant \min\{n_1, n_2\}$。不失一般性，假设 $n_1 \leqslant n_2$。

若 $m = n_1$，由 $V_1 \cap V_2 \subseteq V_1$ 知 $V_1 \cap V_2 = V_1$，因此 $V_1 = V_1 \cap V_2 \subseteq V_2$，从而 $V_1 + V_2 = V_2$。所以，$\dim(V_1 + V_2) = \dim V_2 = n_1 + n_2 - m$。

若 $m < n_1$，设 $x_1, x_2, \cdots, x_m$ 为 $V_1 \cap V_2$ 的基，则可以分别找到 $n_1 - m$ 个矢量 $y_1, y_2, \cdots, y_{n_1-m}$ 和 $n_2 - m$ 个矢量 $z_1, z_2, \cdots, z_{n_2-m}$，使得 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}$ 和 $x_1, x_2, \cdots, x_m, z_1, z_2, \cdots, z_{n_2-m}$ 分别构成 V_1 和 V_2 的基。显然，$x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 都是 $V_1 + V_2$ 子空间中的矢量。

如果 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 线性无关，并且 $V_1 + V_2$ 中的任何矢量都可以表示成 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 的线性组合，则 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 构成 $V_1 + V_2$ 的基，$V_1 + V_2$ 的维数为 $n_1 + n_2 - m$。

显然，V_1 中任何矢量都可以表示成 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}$ 的线性组合，V_2 中任何矢量都可以表示成 $x_1, x_2, \cdots, x_m, z_1, z_2, \cdots, z_{n_2-m}$ 的线性组合。因此，V_1 和 V_2 中的任何矢量都可以表示成 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 的线性组合，$V_1 + V_2$ 中的任何矢量也都可以表示成 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 的线性组合。

为了证明 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_{n_1-m}, z_1, z_2, \cdots, z_{n_2-m}$ 线性无关，只需证明，如果存在 $n_1 + n_2 - m$ 个数 $k_1, k_2, \cdots, k_m, p_1, p_2, \cdots, p_{n_1-m}, q_1, q_2, \cdots, q_{n_2-m}$，满足 $\sum_{i=1}^{m} k_i x_i + \sum_{j=1}^{n_1-m} p_j y_j + \sum_{l=1}^{n_2-m} q_l z_l = 0$，则 $k_1, k_2, \cdots, k_m, p_1, p_2, \cdots, p_{n_1-m}, q_1, q_2, \cdots, q_{n_2-m}$ 全为 0。

令 $x = \sum_{l=1}^{n_2-m} q_l z_l = -\sum_{i=1}^{m} k_i x_i - \sum_{j=1}^{n_1-m} p_j y_j$，则由 $x = \sum_{l=1}^{n_2-m} q_l z_l$ 得 $x \in V_2$，由 $x = -\sum_{i=1}^{m} k_i x_i - \sum_{j=1}^{n_1-m} p_j y_j$ 得 $x \in V_1$，所以，$x \in V_1 \cap V_2$。因此，x 可以写成 $x_1, x_2, \cdots, x_m$ 的线性组合，即存在 $t_1, t_2, \cdots, t_m$，$x = \sum_{i=1}^{m} t_i x_i$。由此得 $\sum_{i=1}^{m} t_i x_i - \sum_{l=1}^{n_2-m} q_l z_l = 0$。但 $x_1, x_2, \cdots, x_m, z_1, z_2, \cdots, z_{n_2-m}$ 是 V_2 的基，它们线性无关，因此 $t_1 = t_2 = \cdots = t_m = q_1 = q_2 = \cdots = q_{n_2-m} = 0$，$x = 0$。

由 $x=0$ 可得 $\sum\limits_{i=1}^{m}k_ix_i+\sum\limits_{j=1}^{n_1-m}p_jy_j=0$，但 $x_1,x_2,\cdots,x_m,y_1,y_2,\cdots,y_{n_1-m}$ 是 V_1 的基，得

$$k_1=k_2=\cdots=k_m=p_1=p_2=\cdots=p_{n_1-m}=0。$$

由此，证明了若 $\sum\limits_{i=1}^{m}k_ix_i+\sum\limits_{j=1}^{n_1-m}p_jy_j+\sum\limits_{l=1}^{n_2-m}q_lz_l=0$，则 $k_1,k_2,\cdots,k_m,p_1,p_2,\cdots,p_{n_1-m}$，$q_1,q_2,\cdots,q_{n_2-m}$ 全为 0，即 $x_1,x_2,\cdots,x_m,y_1,y_2,\cdots,y_{n_1-m},z_1,z_2,\cdots,z_{n_2-m}$ 线性无关，它们构成了 V_1+V_2 的基。

证毕

根据和空间 V_1+V_2 的定义，空间 V_1+V_2 中任一个矢量 $\boldsymbol{z}$ 都存在 $x\in V_1$、$y\in V_2$，满足 $z=x+y$。但一般这种表示并不唯一。

例如，设 V 为一个三维空间。在三维空间 V 中，如果两个平面 V_1 和 V_2 既不重叠，也不平行，那么，在 V_1 中任选一个不为 0 的矢量 $\boldsymbol{a}$，在 V_2 中任选两个线性无关的矢量 $\boldsymbol{b}$ 和 $\boldsymbol{c}$，则矢量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 线性无关，它们构成三维空间的一个基。三维空间中任何一个矢量都可以表示成矢量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 的线性组合，即对任意矢量 $\boldsymbol{z}$，都存在 k_1、k_2、k_3，满足 $\boldsymbol{z}=k_1\boldsymbol{a}+k_2\boldsymbol{b}+k_3\boldsymbol{c}$。而 $k_1\boldsymbol{a}\in V_1$，$k_2\boldsymbol{b}+k_3\boldsymbol{c}\in V_2$，所以，$\boldsymbol{z}\in V_1+V_2$。但是，因为 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 是任意选择的，所以可以有多种选择。$x\in V_1$、$y\in V_2$，满足 $\boldsymbol{z}=x+y$。

在上述三维空间中，V_1 和 V_2 为两个不同的平面，维数都是二，其交集为一直线，即 $\dim(V_1\cap V_2)=1$。$\dim V=3$，$\dim V_1=\dim V_2=2$。因此，$\dim V=\dim V_1+\dim V_2-\dim(V_1\cap V_2)$。

在上例中，如果 V_1 为一个不为 0 的矢量，既不在 V_2 上，也不与 V_2 平行，则 $\dim(V_1\cap V_2)=0$，$\dim V=\dim V_1+\dim V_2$。对 V 中的任意矢量 $\boldsymbol{z}$，都存在唯一 $x\in V_1$、$y\in V_2$，满足 $\boldsymbol{z}=x+y$。

对于一般情况，如果 $\dim(V)=\dim(V_1)+\dim(V_2)$，那么，对 V 中的任意矢量 $\boldsymbol{z}$，都存在唯一 $x\in V_1$、$y\in V_2$，满足 $\boldsymbol{z}=x+y$。这种情况下，我们称 V_1+V_2 为 V_1 与 V_2 的直和或直接和，记为 $V_1\oplus V_2$。

定义 3-3　如果 V_1+V_2 中的任一矢量只能唯一表示为子空间 V_1 的一个矢量与子空间 V_2 的一个矢量的和，则称 V_1+V_2 为 V_1 与 V_2 的直和或直接和，记为 $V_1\oplus V_2$。

定理 3-4　和 V_1+V_2 为直和的充要条件是 $V_1\cap V_2=L(0)$。

证明：

充分性。设 $V_1\cap V_2=L(0)$，则对 $\boldsymbol{z}\in V_1+V_2$，若有

$$\boldsymbol{z}=x_1+x_2,x_1\in V_1,x_2\in V_2$$

$$\boldsymbol{z}=y_1+y_2,y_1\in V_1,y_2\in V_2$$

则有 $(x_1-y_1)+(x_2-y_2)=0$，$x_1-y_1\in V_1$，$x_2-y_2\in V_2$，即 $(x_1-y_1)=-(x_2-y_2)\in V_1\cap V_2$。因此，由 $x_1-y_1=0$，$x_2-y_2=0$，得 $x_1=y_1$，$x_2=y_2$。于是，$\boldsymbol{z}$ 的分解式唯一，V_1+V_2 为直和。

必要性。假设 V_1+V_2 为直和，证明必有 $V_1\cap V_2=L(0)$。

如果 $V_1\cap V_2\neq L(0)$，则存在 $x\neq 0$，$x\in V_1\cap V_2$。因为 $V_1\cap V_2$ 为线性空间，则 $-x\in V_1\cap V_2$，由此得 $0=0+0=x+(-x)$，与 V_1+V_2 为直和的假设矛盾。

证毕

由定理 3-4,很容易得到下述推论。

推论 1　设 V_1、V_2 都是线性空间 V 的子空间,令 $U=V_1+V_2$,则 $U=V_1\oplus V_2$ 的充要条件为 $\dim U=\dim(V_1+V_2)=\dim V_1+\dim V_2$。

推论 2　如果 $x_1,x_2,\cdots,x_k$ 为 V_1 的基,$y_1,y_2,\cdots,y_l$ 为 V_2 的基,且 V_1+V_2 为直和,则 $x_1,x_2,\cdots,x_k,y_1,y_2,\cdots,y_l$ 为 $V_1\oplus V_2$ 的基。

在通常的三维空间中,任意 3 个线性无关的矢量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 都可以构成三维空间的一个基,而矢量 $\boldsymbol{b}$ 和 $\boldsymbol{c}$ 可以构成一个二维平面。三维空间中任何一个矢量都可以表示成矢量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 的线性组合,并且这种表示是唯一的。因此,我们讲三维欧氏空间可以表示为由矢量 $\boldsymbol{a}$ 构成的一维空间 V_1 和矢量 $\boldsymbol{b}$、$\boldsymbol{c}$ 构成的二维空间 V_2 的直和 $V_1\oplus V_2$。如果矢量 $\boldsymbol{a}$ 与矢量 $\boldsymbol{b}$ 和 $\boldsymbol{c}$ 都正交,则矢量 $\boldsymbol{a}$ 与 V_2 正交。例如,在通常的直角坐标系中,Z 轴正交(垂直)于 XOY 平面。

用 $V_1^{\perp}$ 表示欧氏空间 V^n 中所有与 V_1 正交的矢量的集合。若 $x\in V_1^{\perp}$,$y\in V_1^{\perp}$,$z\in V_1$,则有

$<x+y,z>=<x,z>+<y,z>=0+0=0$,所以 $x+y\in V_1^{\perp}$

$<kx,z>=k<x,z>=k0=0$,所以 $kx\in V_1^{\perp}$。

因此,$V_1^{\perp}$ 为 V^n 的一个子空间。我们称子空间 $V_1^{\perp}$ 为 V_1 的**正交补空间**或 V_1 的**正交补**。由于 $V_1\cap V_1^{\perp}=L(0)$,因此,如果 $V^n=V_1+V_1^{\perp}$,则由定理 3-4 可得 $V^n=V_1\oplus V_1^{\perp}$。显然,$V_1+V_1^{\perp}\subseteq V^n$。为了证明 $V^n=V_1+V_1^{\perp}$,只需证明 $V^n\subseteq V_1+V_1^{\perp}$,即对任意 $x\in V^n$,都存在 $y\in V_1$ 和 $z\in V_1^{\perp}$ 满足 $x=y+z$。例如,对于三维空间,假设空间 V_1 为 XOY 平面,则与 V_1 正交的空间 $V_1^{\perp}$ 为 Z 轴(一维空间)。对于任意一个三维空间的矢量 $\boldsymbol{x}$,设其坐标为 (x,y,z),x 沿着 Z 方向在 V_1 空间的投影为 $y=(x,y,0)$,则 $z=x-y=(0,0,z)\in V_1^{\perp}$。对于一般情况,取 y 为 x 沿着 $V_1^{\perp}$ 空间在空间 V_1 上的投影,容易证明 $z=x-y$ 与空间 V_1 正交,即 $z\in V_1^{\perp}$。可以得到定理 3-5。

定理 3-5　任意欧氏空间 V^n 都为其子空间 V_1 及 V_1 的正交补空间 $V_1^{\perp}$ 的直和,即 $V^n=V_1\oplus V_1^{\perp}$。

证明:若 $V_1=\{0\}$,则 $V_1^{\perp}=V^n$,从而 $V^n=V_1\oplus V_1^{\perp}$ 成立。

若 $V_1\neq\{0\}$,设 $\dim V_1=m(1\leqslant m\leqslant n)$,且 V_1 的一个标准正交基为 $x_1,x_2,\cdots,x_m$。

对任意 $x\in V^n$,令 $a_i=<x_i,x>(i=1,2,\cdots,m)$,那么,$y=\sum\limits_{i=1}^{n}a_ix_i\in V_1$。再令 $z=x-y$,由于 $<z,x_i>=<x-y,x_i>=<x,x_i>-<y,x_i>=0$,所以 $z\in V_1^{\perp}$。而 $x=y+z$,因此 $x\in V_1+V_1^{\perp}$,从而 $V^n\subseteq V_1+V_1^{\perp}$,得 $V^n=V_1+V_1^{\perp}$。由于 $V_1\cap V_1^{\perp}=L(0)$,因此 $V^n=V_1\oplus V_1^{\perp}$。

证毕

推论　设 V_1 是任意欧氏空间 V^n 的子空间,且 V_1 的维数为 m,则 $V_1^{\perp}$ 的维数为 $n-m$,即有 $n=\dim V=\dim V_1+\dim V_1^{\perp}$。

任意 n 阶齐次线性方程组 $\mathbf{A}x=0$ 的解空间也是 R^n 的子空间,称其为 $\mathbf{A}$ 的**核空间**或**零空间**,记为 $N(\mathbf{A})$,$N(\mathbf{A})=\{x|\mathbf{A}x=0\}$。$\mathbf{A}$ 的核空间的维数称为 $\mathbf{A}$ 的**零度**,记为 $n(\mathbf{A})$。

设 $m\times n$ 阶系数矩阵 $\boldsymbol{A}=(a_{ij})_{m\times n}$ 的秩为 r，$x=(x_1,x_2,\cdots,x_n)$，$\boldsymbol{A}$ 的第 i 个行矢量记为 $\boldsymbol{\beta}_i=(a_{i1},a_{i2},\cdots,a_{in})$。则方程组 $\boldsymbol{A}x=0$ 可改写为

$$<\beta_1,x>=0,<\beta_2,x>=0,\cdots,<\beta_m,x>=0$$

由此可见，求齐次线性方程组的解矢量，就是求所有与矢量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_m$ 正交的矢量。设 $\beta_1,\beta_2,\cdots,\beta_m$ 生成的子空间为 $V_1=L(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}},\cdots,\boldsymbol{\beta}_m^{\mathrm{T}})$，所有与 V_1 正交的矢量的集合也形成一个子空间，称其为齐次线性方程组的解空间。根据定义，齐次线性方程组 $\boldsymbol{A}x=0$ 的解空间 $N(\boldsymbol{A})$ 就是矩阵 $\boldsymbol{A}$ 的行矢量构成的空间 V_1 的正交补空间 $V_1^{\perp}$，V_1 的维数就是矩阵 $\boldsymbol{A}$ 的秩 r，解空间的维数是 $V_1^{\perp}$ 的维数 $n-r$，即 $n(\boldsymbol{A})=n-r$。同样，可以得到 $n(\boldsymbol{A}^{\mathrm{T}})=m-\mathrm{rank}\ \boldsymbol{A}^{\mathrm{T}}=m-\mathrm{rank}\ \boldsymbol{A}$。

因而，对 $m\times n$ 的矩阵 $\boldsymbol{A}$，有 $\mathrm{rank}\ \boldsymbol{A}+n(\boldsymbol{A})=n$，$\mathrm{rank}\ \boldsymbol{A}+n(\boldsymbol{A}^{\mathrm{T}})=m$，$n(\boldsymbol{A})-n(\boldsymbol{A}^{\mathrm{T}})=n-m$。

根据定义，$R(\boldsymbol{A})$ 表示 $\boldsymbol{A}$ 的列矢量构成的空间，则 $\boldsymbol{A}$ 的行矢量构成的空间 V_1 可以表示为 $R(\boldsymbol{A}^{\mathrm{T}})$。因此，$N(\boldsymbol{A})=R^{\perp}(\boldsymbol{A}^{\mathrm{T}})$，$R(\boldsymbol{A}^{\mathrm{T}})\oplus N(\boldsymbol{A})=R^n$，有定理 3-6。

定理 3-6　对于任意矩阵 $\boldsymbol{A}=(a_{ij})_{m\times n}\in R^{m\times n}$，都有

$$N(\boldsymbol{A})=R^{\perp}(\boldsymbol{A}^{\mathrm{T}}),\quad R(\boldsymbol{A}^{\mathrm{T}})\oplus N(\boldsymbol{A})=R^n$$
$$N(\boldsymbol{A}^{\mathrm{T}})=R^{\perp}(\boldsymbol{A}),\quad R(\boldsymbol{A})\oplus N(\boldsymbol{A}^{\mathrm{T}})=R^m$$

证明：设 $\boldsymbol{A}$ 的第 i 个行矢量记为 $\boldsymbol{\beta}_i$，并记 $V_1=R(\boldsymbol{A}^{\mathrm{T}})=L(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}},\cdots,\boldsymbol{\beta}_m^{\mathrm{T}})\subseteq R^n$。于是有

$$\begin{aligned}V_1^{\perp}&=R^{\perp}(\boldsymbol{A}^{\mathrm{T}})=\{y\mid y\perp\boldsymbol{\beta}_i^{\mathrm{T}},i=1,2,\cdots,m\}=\{y\mid\boldsymbol{\beta}_i y=0\}=\{y\mid\boldsymbol{A}y=0\}\\&=N(\boldsymbol{A})\end{aligned}$$

因此，$R^n=V_1\oplus V_1^{\perp}=R(\boldsymbol{A}^{\mathrm{T}})\oplus N(\boldsymbol{A})$。

由 $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{A}$，可得 $N(\boldsymbol{A}^{\mathrm{T}})=R^{\perp}(\boldsymbol{A})$，$R(\boldsymbol{A})\oplus N(\boldsymbol{A}^{\mathrm{T}})=R^m$。

证毕

对于 m 维空间 R^m 中的任意 n 维子空间 V_1，定理 3-6 给出了求其正交补空间 V_1^{T} 的方法。设 V_1 的基为 $\alpha_1,\alpha_2,\cdots\alpha_n$，即 $V_1=L(\alpha_1,\alpha_2,\cdots,\alpha_n)$。令 $\boldsymbol{A}=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$，则 $V_1=R(\boldsymbol{A})$，$V_1^{\perp}=N(\boldsymbol{A}^{\mathrm{T}})$。

3. 投影变换与投影矩阵

在三维直角坐标系中，称矢量 $(x,y,0)$ 为矢量 (x,y,z) 沿着 Z 轴在 XOY 平面上的投影。同样，$(0,0,z)$ 为矢量 (x,y,z) 沿着 XOY 平面在 Z 轴上的投影，矢量 $(x,y,z)=(x,y,0)+(0,0,z)$。

设 L 和 M 都是 R^n 的子空间，并且 $L\oplus M=R^n$。由直和的定义可知，任意 $x\in R^n$ 都存在唯一的 $y\in L$ 和 $z\in M$，$x=y+z$。我们称 y 是 x 沿着 M 到 L 的投影，z 是 x 沿着 L 到 M 的投影。

定义 3-4　将任意 $\boldsymbol{x}\in R^n$ 变为沿着 M 到 L 的投影的变换称为沿着 M 到 L 的投影算子，记为 $P_{L,M}$，即 $P_{L,M}x=y$，$\boldsymbol{y}\in L$。

由定义 34 知，投影算子 $P_{L,M}$ 将整个空间 R^n 变到子空间 L。特别地，若 $x\in L$，则 $P_{L,M}x=x$；若 $x\in M$，则 $P_{L,M}\boldsymbol{x}=0$。因此，$P_{L,M}$ 的值域为 $R(P_{L,M})=L$，零空间为 M。对于任意 $x\in R^n$，$x\in M$，当且仅当 $P_{L,M}\boldsymbol{x}=0$。