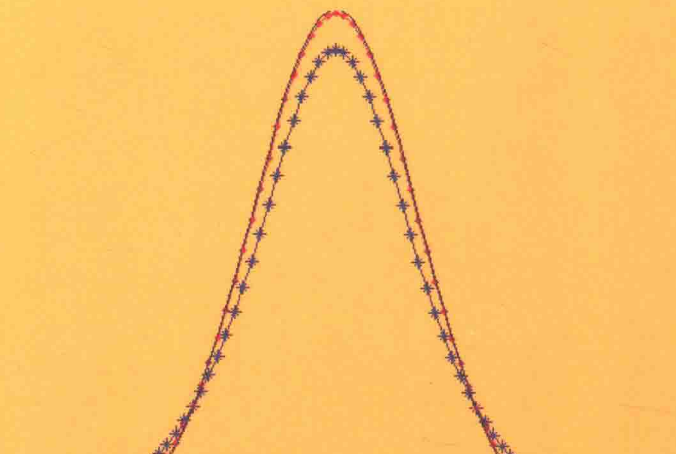
系统与控制中的随机方法

陈 曦 编著



清华大学出版社

系统与控制中的随机方法

陈 曦 编著

清华大学出版社
北京

内容简介

本书内容分为3大部分共7章。第1部分内容包括概率论基础知识、统计量与抽样分布、参数估计及贝叶斯推断；第2部分内容是滤波和随机线性系统的控制及优化方法；第3部分以无线传传感器网络的覆盖与目标探测为例，介绍如何利用前两部分中的知识和方法进行系统的设计、性能分析及优化。第1部分和第2部分中的各章均配有习题。

本书适合作为相关专业的研究生和高年级本科生教材，对从事工程应用研究的人员也大有帮助。

图书在版编目（CIP）数据

系统与控制中的随机方法/陈曦编著. --北京：清华大学出版社，2013

ISBN 978-7-302-32549-9

Ⅰ. ①系…　Ⅱ. ①陈…　Ⅲ. ①随机系统－研究 ②随机控制－研究　Ⅳ. ①O231

中国版本图书馆CIP数据核字(2013)第109946号

责任编辑：刘　颖
封面设计：常雪影
责任校对：王淑云
责任印制：宋　林

出版发行：清华大学出版社
　网　　址：http://www.tup.com.cn，http://www.wqbook.com
　地　　址：北京清华大学学研大厦A座　**邮　　编**：100084
　社 总 机：010-62770175　**邮　　购**：010-62786544
　投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
　质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京国马印刷厂
经　　销：全国新华书店
开　　本：185mm×230mm　**印　　张**：11.75　**字　　数**：242千字
版　　次：2013年7月第1版　**印　　次**：2013年7月第1次印刷
印　　数：1～3000
定　　价：25.00元

产品编号：051157-01

前　言

现代工程系统包括通信、制造和物流等很多领域。我们通常能够从实际的系统中采集到大量的数据。要优化这些系统的运行模式、提高其运行效率，通常需要对系统进行建模，然后根据实际数据估计模型中的参数，进一步进行预测、控制或优化。在此过程中需要许多的随机方法，本书讲解其中常用的方法及其理论基础。

本书的初稿曾作为清华大学自动化系的研究生专业基础课程的讲义。虽然有关随机方法的内容很多，但考虑到课程的时间限制，我们只选取最常用的基础理论和方法。本书内容分为 3 大部分；第 1 部分介绍参数估计及贝叶斯方法；第 2 部分介绍随机系统的滤波、控制及优化方法；第 3 部分以无线传感器网络的覆盖与目标探测为例，介绍如何利用前两部分中的知识和方法进行系统的设计、性能分析及优化。本书适合作为相关专业的研究生和高年级本科生教材。对从事工程应用研究的人员也大有帮助。

感谢阅读本书草稿并对文字表述提出建议、指出排版错误的多位同学：黄永皓、柳哲、聂春凯、郭瑛、王靖远等。特别感谢柳哲同学帮助修改和绘制部分插图。感谢刘颖老师所做的技术编辑工作。

由于编者水平有限，虽经多次修改，书中一定还存有不妥甚至错误之处，恳请读者和专家批评指正。

编　者

2013 年 3 月

于北京清华园

目 录

第1章　概率论基础知识

概率这个概念的形成与赌博活动关系密切。在文艺复兴前，概率还不是一个数学的概念。到16世纪，开始有一些意大利数学家讨论掷骰子中各种情况出现的机遇问题，这种研究结晶出古典概率的定义。把所研究的情况分解为一些看似有同等可能的简单情况，其数目与全部可能结果数之比，取为该情况出现的概率。此定义最初始自何人已不可考，现今为人所知的一位是卡丹诺[①]。他在数学上知名是因为他发现了一般的三次代数方程的解法，而在概率史上的地位是因为他的一本名为《机遇博弈》的著作，该书约成于1564年，但在1663年才得以发表，其时关于概率论的若干重要著作已经问世，这就削弱了该著作在概率史上的地位和影响。

对“分赌本问题”的讨论和研究促进了概率论的发展，该问题可以这样描述：A，B二人赌博，各出a元，每局各人获胜概率都是0.5，二人约定：谁先胜S局，就赢得全部注金$2a$元。在进行到A胜S_1局、B胜S_2局（S_1、S_2都小于S）时赌博因故停止，问此时注金$2a$应如何分配给A和B才公平？

多年之后，一位名叫德梅尔的赌徒向帕斯卡[②]请教几个有关赌博的问题，这促使帕斯卡与费马[③]在1654年7—10月间来往的信中讨论具体的赌博问题。他们引进了赌博的值（value）的概念，值等于赌注乘以概率。3年后，惠更斯[④]改“值”为“期望”（expectation），（数学）期望成为概率论的最重要概念之一。惠更斯的主要著作《机遇的规律》出版于1657年，得到学术界的重视并成为标准教本长达50年之久。伯努利[⑤]与惠更斯长期保持通信联系，仔细研读过《机遇的规律》，因此启发了他对概率的兴趣。在生命的最后两年，伯努利写了《推测术》。该书在他去世时尚未整理定稿，这个工作最后由他的侄儿完成。《推测术》的第一部分对《机遇的规律》一书作了仔细的注解，总量比原书长4倍；

---

① Girolamo Cardano，1501—1576，意大利人。

② Blaise Pascal，1623—1662，法国数学家、物理学家、思想家。

③ Pierre de Fermat，1601—1665，法国数学家。

④ Christiaan Huygens，1629—1695，荷兰物理学家、天文学家、数学家。

⑤ Jacob Bernoulli，1654—1705，瑞士数学家。

在该书的第四部分讨论了概率论在社会、道德和经济领域中的应用，其中包括了该书的精华，那就是以其名字命名的大数定律[①]。

早期概率论的研究对象多是离散的变量，通常称为经典概率论。在 17 世纪后期发展成熟的微积分为研究连续随机变量的概率分布奠定了基础。

我们首先回顾概率论中的一些概念与结论，它们是本书内容的基础。

1.1 随机变量

在一定条件下可能发生也可能不发生的试验结果称为随机事件，简称事件。随机事件中有两个特殊的情况，其中之一是在一定条件下，每次试验都必定发生的事件称为必然事件；另一个是在一定条件下，每次试验都一定不发生的事件称为不可能事件。

最初关于概率的定义与频率紧密相关。随机事件在一次试验中是否发生，固然是无法事先知道的，但当进行多次重复试验时，就可以发现有关其发生的可能性的大小的统计规律。如果在相同条件下进行 n 次重复试验，事件 A 出现了 m 次，那么事件 A 在 n 次试验中出现的频率为 m/n。当 n 增大时，m 也会增大，m/n 在 n 趋于无穷大时会收敛到一个定值。这个统计规律表明事件 A 发生的可能性的大小是事件本身所固有的性质。事件 A 发生的可能性的大小称为事件 A 的概率。当试验的次数足够多时，可以用事件发生的频率近似地表示该事件的概率。

很多随机试验只有有限个不同的基本事件。一般来说，事件总是由几个基本事件组成，而每个基本事件都是等可能的。基本事件的全体记作 Ω，称为基本事件空间。如果事件 A 由 $k(k\leqslant n)$ 个不同的基本事件组成，则 k/n 可以作为事件 A 的概率。这就是概率的古典定义。

随着概率的应用范围的扩大和微积分理论的发展，古典概率的定义及频率的直观解释都不再能满足概率论的发展，于是就有了关于概率的公理化定义。

定义 1.1（σ 代数，σ-algebra）　设 $\Omega=\{\omega\}$，称 ω 为基本事件。$\mathcal{F}=\{A\,|\,A\subseteq\Omega\}$，如果 $\mathcal{F}$ 满足条件：

1. $\Omega\in\mathcal{F}$；
2. 若 $A\in\mathcal{F}$，则 $\overline{A}=\{\omega\in\Omega\,|\,\omega\notin A\}\in\mathcal{F}$；
3. 对于任意 $A_k\in\mathcal{F}, k=1,2,\cdots$，则 $\bigcup\limits_{k=1}^{+\infty}A_k\in\mathcal{F}$。

则称 $\mathcal{F}$ 是 Ω 中的一个 σ 代数。

定义 1.2（概率，probability）　设 $\mathcal{P}(A)(A\in\mathcal{F})$ 是 σ 代数 $\mathcal{F}$ 上的实值函数，如果它满足

① 大数定律的名称首见于泊松在 1837 年的一篇著作中，并非出自《推测术》。

条件：

1. 对任意 $A\in\mathcal{F}$，$0\leqslant\mathcal{P}(A)\leqslant 1$；

2. $\mathcal{P}(\Omega)=1$；

3. 对任意 $A_k\in\mathcal{F}(k=1,2,\cdots)$，$A_i\cap A_j=\varnothing(i\neq j)$，有 $\mathcal{P}(\bigcup\limits_{k=1}^{+\infty}A_k)=\sum\limits_{k=1}^{+\infty}\mathcal{P}(A_k)$。

则称$\mathcal{P}(A)$为$\mathcal{F}$上的概率测度，或简称概率。$A(\in\mathcal{F})$称为事件，$\mathcal{F}$是事件的全体。$\mathcal{P}(A)$称为事件A 的概率，$(\Omega,\ \mathcal{F},\ \mathcal{P})$称为概率空间。

概率的公理化定义方便人们将微积分的理论和技巧用于概率的计算和描述，从而大大促进了概率论的发展。

定义 1.3(随机变量，random variable)　设$(\Omega,\mathcal{F},\ \mathcal{P})$为一个概率空间，且$\mathbb{R}$为实数空间。$X$ 是一个函数，$X:\Omega\to\mathbb{R}$，满足对任意实数 $x\in\mathbb{R}$，集合$\{\omega\in\Omega\mid X(\omega)\leqslant x\}\in\mathcal{F}$，并具有由$\mathcal{P}$定义的概率，则称 X 为随机变量。

定义 1.4(分布函数，probability distribution function)　随机变量 X 的分布函数定义为

$$F(x)=\mathcal{P}(\omega\in\Omega:X(\omega)\leqslant x)=\mathcal{P}(X\leqslant x)。$$

$F(x)$是非减的。$\lim\limits_{x\to-\infty}F(x)=0$ 且 $\lim\limits_{x\to+\infty}F(x)=1$。

非离散随机变量 X 的取值不能一一列出。但在实际中，我们对非离散随机变量所取的值落在某一区间的概率更感兴趣。

$$\mathcal{P}(x_1<X\leqslant x_2)=\mathcal{P}(X\leqslant x_2)-\mathcal{P}(X\leqslant x_1)=F(x_2)-F(x_1)。$$

定义 1.5(离散随机变量，discrete random variable)　如果随机变量 X 定义在离散空间上，$X:\Omega\to\mathcal{S}$，$\mathcal{S}=\{s_1,s_2,\cdots\}$，则称 X 为离散随机变量，其特性由事件$\{X=s_k\}$的概率来描述：

$$\mathcal{P}(X=s_k)=p_k,\quad k=1,2,\cdots。\tag{1.1}$$

由概率的定义，p_k 满足如下两个条件：

$$p_k\geqslant 0,\qquad k=1,2,\cdots,\qquad 并且\qquad \sum_{k=1}^{+\infty}p_k=1。$$

称(1.1)式为离散随机变量的分布律。

定义 1.6(连续随机变量，continuous random variable)　如果对于随机变量 X 的分布函数$F(x)$，存在非负可积函数 $f(x)$，满足

$$F(x)=\int_{-\infty}^{x}f(y)\mathrm{d}y,$$

则称 X 为连续随机变量，$f(x)$称为 X 的概率密度函数，简称分布密度。

如果 $F(x)$在 x 处可导，则有$\dfrac{\mathrm{d}F(x)}{\mathrm{d}x}=f(x)$。另外还满足归一化条件，$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$。

定义 1.7(独立性，independence of random variables)　设 X 和 Y 为随机变量，若对

于任意实数 x 和 y，

$$\mathcal{P}(X \leqslant x, Y \leqslant y) = \mathcal{P}(X \leqslant x)\,\mathcal{P}(Y \leqslant y) \tag{1.2}$$

成立，则称 X 与 Y 相互独立。

1.2 随机变量的性质

1.2.1 期望与方差

定义 1.8(数学期望，expectation) 随机变量 X 的分布函数为 $F(x)$，若

$$\int_{-\infty}^{+\infty} |x| \, \mathrm{d}F(x) < +\infty,$$

定义 X 的数学期望为

$$EX = \int_{-\infty}^{+\infty} x \mathrm{d}F(x)。$$

设 X 为离散随机变量，其分布律为$\mathcal{P}(X=x_k)=p_k, k=1,2,\cdots$，如果 $\sum\limits_k |x_k| p_k < +\infty$，则

$$EX = \sum_k x_k \mathcal{P}(X = x_k) = \sum_k x_k p_k。$$

设 X 为连续随机变量，$f(x)$为其分布密度，如果 $\int_{-\infty}^{+\infty} |x| f(x) \mathrm{d}x < +\infty$，则

$$EX = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x。$$

数学期望也简称期望或均值(mean)。

定义 1.9(方差，variance) 对随机变量 X，如果 $E(X-EX)^2$ 存在，定义 X 的方差为

$$\mathrm{Var}(X) = E(X - EX)^2,$$

$\mathrm{Var}(X)$的正方根 $\sqrt{\mathrm{Var}(X)}$，称为标准差(standard deviation)。

方差是随机变量的另一重要数字特征。它描述随机变量取值的分散程度。

定义 1.10(原点矩，moment) 设 X 为随机变量，若 $E(X^k), k=1,2,\cdots$存在，称它为 X 的 k 阶原点矩，简称 k 阶矩。

定义 1.11(中心矩，central moment) 设 X 为随机变量，若 $E([X-EX]^k), k=2,3,\cdots$存在，称它为 X 的 k 阶中心矩。

定义 1.12(众数，mode) 随机变量 X 的一系列可能的取值中出现可能性最大的数值称为众数。若 X 是离散随机变量，则 X 的最大可能取值为众数；若 X 是连续随机变量，其分布密度为 $f(x)$，使 $f(x)$达到最大值的点就是众数。

值得注意的是，众数不一定存在；若存在，也不一定惟一。

定义 1.13（中位数，median） 对任意随机变量 X，满足

$$\mathcal{P}(X \leqslant m) \geqslant \frac{1}{2} \quad \text{和} \quad \mathcal{P}(X \geqslant m) \geqslant \frac{1}{2}$$

的 m 称为 X 的中位数。

从概率上讲，m 这个点正好居于中央，所以称为中位数。中位数优于期望的地方在于它受个别特大或特小的值的影响很小，有时比均值更说明问题。例如，某社区内人的收入的中位数告诉我们：有一半的人收入低于此值，另一半高于此值。直观上这个值比均值具有代表性。

定义 1.14（分位数，quantile） 对任意随机变量 X，一个数 x_p 如果满足

$$\mathcal{P}(X \leqslant x_p) \geqslant p \quad \text{和} \quad \mathcal{P}(X \geqslant x_p) \geqslant 1-p, \quad 0 < p < 1,$$

则称 x_p 为 X 的 p 分位数。

定义 1.15（变异系数，coefficient of variation） 对任意随机变量 X，标准差与数学期望的比值称为变异系数，记为 ν，即 $\nu = \sqrt{\mathrm{Var}(X)}/EX$。

变异系数是以数学期望为单位计量随机变量的分散程度的一种数字特征。

1.2.2 母函数

定义 1.16（母函数，probability generating function） 设 X 是取非负整数值的随机变量，它的概率分布为

$$\mathcal{P}\{X = k\} = p_k \quad k = 0,1,2,\cdots,$$

则称随机变量函数 θ^X 的均值 $E(\theta^X)$，即

$$g(\theta) = E(\theta^X) = \sum_{k=0}^{+\infty} \theta^k p_k$$

为与随机变量 X 的概率分布对应的母函数，简称 X 的母函数。

由于 $\sum\limits_{k=0}^{+\infty} p_k = 1$，由幂级数的收敛性可知，$g(\theta)$ 至少在 $|\theta| < 1$ 时一致收敛且绝对收敛。因此，母函数对任何整数随机变量都存在，而且概率分布与母函数是一一对应的。利用母函数很容易求得概率分布的数字特征。例如，若数学期望存在，有

$$g'(1) = \sum_{k=0}^{+\infty} k p_k = EX。$$

当方差存在时，因为

$$E[X(X-1)] = \sum_{k=2}^{+\infty} k(k-1) p_k = g''(1),$$

故
$$\mathrm{Var}(X)=E(X^2)-(EX)^2=g''(1)+g'(1)-[g'(1)]^2。$$

例 1.1　几何分布的分布函数为 $P_G(x=k)=pq^{k-1}, k=1,2,\cdots, p>0, q>0, p+q=1$。几何分布的母函数为
$$g(\theta)=\sum_{k=1}^{+\infty}\theta^k pq^{k-1}=\frac{\theta p}{1-\theta q},$$
则
$$g'(1)=\left.\frac{p}{(1-\theta q)^2}\right|_{\theta=1}=\frac{1}{p},\quad g''(1)=\left.\frac{2pq}{(1-\theta q)^3}\right|_{\theta=1}=\frac{2q}{p^2}。$$

设 X 和 Y 是两个相互独立的取非负整数值的随机变量，g_X 和 g_Y 分别为它们的母函数，随机变量 $X+Y$ 的母函数记为 g_{X+Y}，则有 $g_{X+Y}=g_X g_Y$。

母函数的这个性质可以推广到多个独立非负整数值随机变量之和的情况。

定义 1.17(矩母函数，moment generating function)　设 X 是定义在$(\Omega,\mathcal{F},P)$上的随机变量，如果 $E(\mathrm{e}^{tX})$在原点的某个邻域内存在，称函数 $M(t)=E(\mathrm{e}^{tX})$为随机变量 X 的矩母函数。

随机变量 X 的原点矩可以用它的矩母函数 $M(t)$的导数表示。若 $M(t)$对于在$(-t_0, t_0)$中的 t 存在(这里 $t_0>0$)，则 $E(X^k)=M^{(k)}(0)$。

与概率母函数类似，相互独立的随机变量之和的矩母函数是各个随机变量矩母函数的乘积。我们还可以证明，矩母函数惟一地确定了分布函数。但由于有些随机变量的矩母函数可能不存在，这使矩母函数的效用受到限制。

概率母函数和矩母函数都是研究随机变量分布的非常重要的分析工具。

1.2.3　特征函数

定义 1.18(特征函数，characteristic function)　设 X 是定义在$(\Omega,\mathcal{F},P)$上的随机变量，$F(x)$是它的分布函数，称函数
$$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})=\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x),\quad -\infty<t<+\infty,\ \mathrm{i}=\sqrt{-1}$$
为随机变量 X(或分布函数 $F(x)$)的特征函数。

若 X 是离散随机变量，$\mathcal{P}\{X=x_k\}=p_k(k=1,2,\cdots)$，则其特征函数为
$$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})=\sum_k \mathrm{e}^{\mathrm{i}tx_k}p_k。$$
若 X 是连续随机变量，其分布函数密度为 $f(x)$，则其特征函数为
$$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})=\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}tX}f(x)\mathrm{d}x。$$

$\varphi(t)$是 $f(x)$的傅里叶(Fourier)变换。

随机变量的特征函数总是存在,它对一切 t 有定义。特征函数具有如下性质:

1. 有界: $|\varphi(t)| \leqslant \varphi(0)=1(-\infty<t<+\infty)$;
2. 一致连续: $\varphi(t)$在$(-\infty,+\infty)$上一致连续;
3. 非负定: 对于任意正整数 n,任意实数 $t_1,t_2,\cdots,t_n$ 和任意复数 $z_1,z_2,\cdots,z_n$,有
$$\sum_{i=1}^{n}\sum_{j=1}^{n}\varphi(t_i-t_j)z_i\bar{z}_j \geqslant 0;$$
4. $\varphi(-t)=\overline{\varphi(t)}$;
5. 两个相互独立的随机变量的和的特征函数等于它们的特征函数之积;
6. 设 $Y=aX+b$,a,b 为常数,则 $\varphi_Y(t)=\mathrm{e}^{\mathrm{i}tb}\varphi_X(at)$;
7. 如果 X 有 n 阶原点矩,则它的特征函数 $\varphi(t)$有 n 阶导数,并且
$$\varphi^{(k)}(0)=\mathrm{i}^k E(X^k),\quad k=1,2,\cdots,n。$$

反之,若 $\varphi^{(2n)}(0)$存在,则 X 具有 $2n$ 阶原点矩。

显然,分布函数决定特征函数。反之,特征函数也能决定分布函数。这样就得到下面的逆转公式和惟一性定理。

定理 1.1 设分布函数 $F(x)$的特征函数为 $\varphi(t)$,x_1,x_2 是 $F(x)$的连续点,则
$$F(x_2)-F(x_1)=\lim_{T\to+\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}tx_1}-\mathrm{e}^{-\mathrm{i}tx_2}}{\mathrm{i}t}\varphi(t)\mathrm{d}t。$$

定理 1.2 分布函数由其特征函数惟一决定。

1.3 随机向量及其分布

在许多随机现象中,每次试验的结果要同时用几个数来描述,对应于每个样本点,用一个向量$(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 来表示试验的结果。例如,3 位小孩玩"石头、剪子、布",观察并记录每个小孩的手势,每次得到一个三维向量。

1.3.1 随机向量

定义 1.19(随机向量,random vectors) 若随机变量 X_1, X_2, $\cdots$,X_n 定义在同一概率空间$(\Omega, \mathcal{F},\mathcal{P})$上,则称
$$\boldsymbol{X}=(X_1, X_2, \cdots,X_n)^{\mathrm{T}}$$
构成一个 n 维随机向量。

定义 1.20(联合分布函数,joint distribution function) 设 $x_1,x_2,\cdots,x_n$ 是 n 个实数,称 n 元函数

$$F(x_1,x_2,\cdots,x_n)=P(X_1\leqslant x_1,\ X_2\leqslant x_2,\cdots,X_n\leqslant x_n) \tag{1.3}$$

为随机向量 $\boldsymbol{X}=(X_1,\ X_2,\cdots,X_n)^{\mathrm{T}}$ 的联合分布函数。

n 元分布函数 $F(x_1,x_2,\cdots,x_n)$ 具有下列性质：

1. $F(x_1,x_2,\cdots,x_n)$ 对任一 x_i 单调不减；
2. $F(x_1,x_2,\cdots,x_n)$ 对任一 x_i 右连续；
3. 令

$$F(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_n)=\lim_{x_i\to-\infty}F(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n),$$

$$F(+\infty,\cdots,+\infty)=\lim_{x_1\to+\infty,\cdots,x_n\to+\infty}F(x_1,\cdots,x_n),$$

则

$$F(x_1,\ \cdots,x_{i-1},-\infty,x_{i+1},\ \cdots,x_n)=0,$$
$$F(+\infty,\cdots,+\infty)=1;$$

4. 设 $a_i\leqslant b_i(i=1,2,\cdots,n)$，则

$$F(b_1,b_2,\cdots,b_n)-\sum_{i=1}^{n}F_i+\sum_{i<j}F_{ij}-\cdots+(-1)^nF(a_1,\ a_2,\cdots,a_n)\geqslant 0,$$

其中 $F_{ij\cdots k}$ 是当 $x_i=a_i,x_j=a_j,\cdots,x_k=a_k$ 而其余 $x_l=b_l$ 时 $F(x_1,x_2,\cdots,x_n)$ 的值。

以上性质的详细证明留给读者。对于最后一个性质，用数学归纳法可以推导出

$$F(b_1,b_2,\cdots,b_n)-\sum_{i=1}^{n}F_i+\sum_{i<j}F_{ij}-\cdots+(-1)^nF(a_1,a_2\cdots,a_n)$$
$$=\mathcal{P}(a_1<x_1\leqslant b_1,a_2<x_2\leqslant b_2,\cdots,\ a_n<x_n\leqslant b_n)\geqslant 0。$$

随机向量的分布函数也有离散型与连续型的分别。离散型的概率分布集中在有限个或可列个点上。若随机向量为连续型时，则存在非负可积函数 $f(x_1,\ x_2,\cdots,x_n)$，使

$$F(x_1,x_2,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}f(y_1,y_2,\cdots,y_n)\mathrm{d}y_1\cdots\ \mathrm{d}y_n,$$

这里的 $f(x_1,x_2,\cdots,\ x_n)$ 称为密度函数，满足条件

$$f(x_1,x_2,\cdots,x_n)\geqslant 0,\quad 且\quad \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(y_1,y_2,\cdots,y_n)\mathrm{d}y_1\cdots\ \mathrm{d}y_n=1。$$

例 1.2 做 n 次重复独立的试验，每次试验的结果为 $A_1,\cdots,A_m$，$\mathcal{P}(A_i)=p_i(i=1,2,\cdots,m)$，且

$$\sum_{i=1}^{m}p_i=1,\quad p_i\geqslant 0。$$

记 X_i 表示在 n 次试验中事件 A_i 出现的次数，$(X_1,X_2,\cdots,X_m)^{\mathrm{T}}$ 就是一个 m 维随机向量，其分布律为

$$\mathcal{P}(X_1=n_1,\cdots,X_m=n_m)=\frac{n!}{n_1!\cdots n_m!}p_1^{n_1}\cdots\ p_m^{n_m}, \tag{1.4}$$