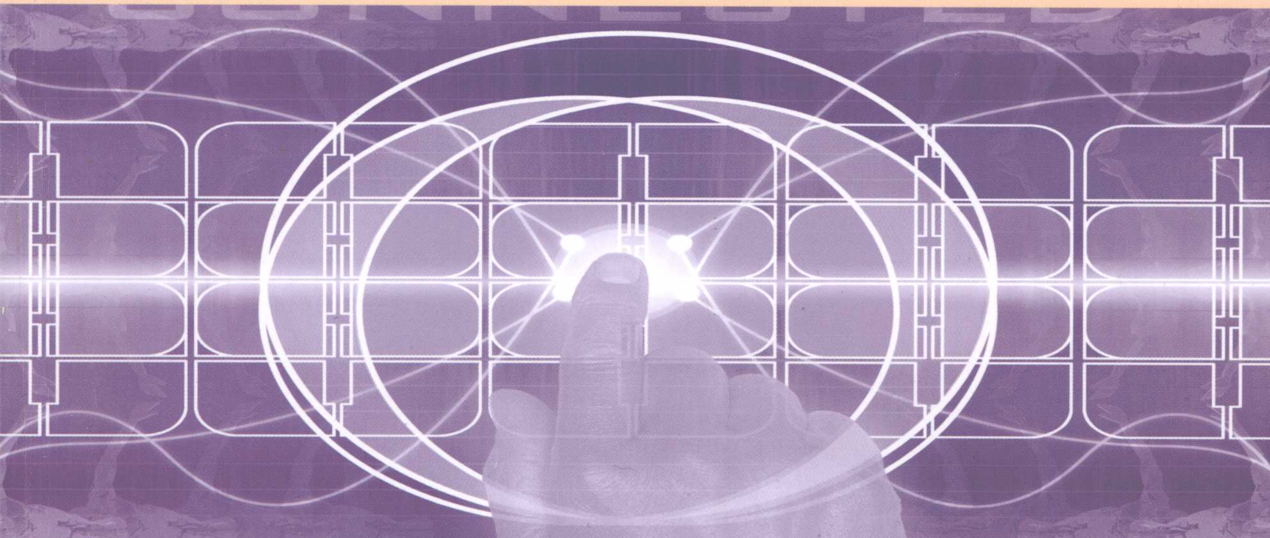
高等学校教材

随机过程

吴群英　编著

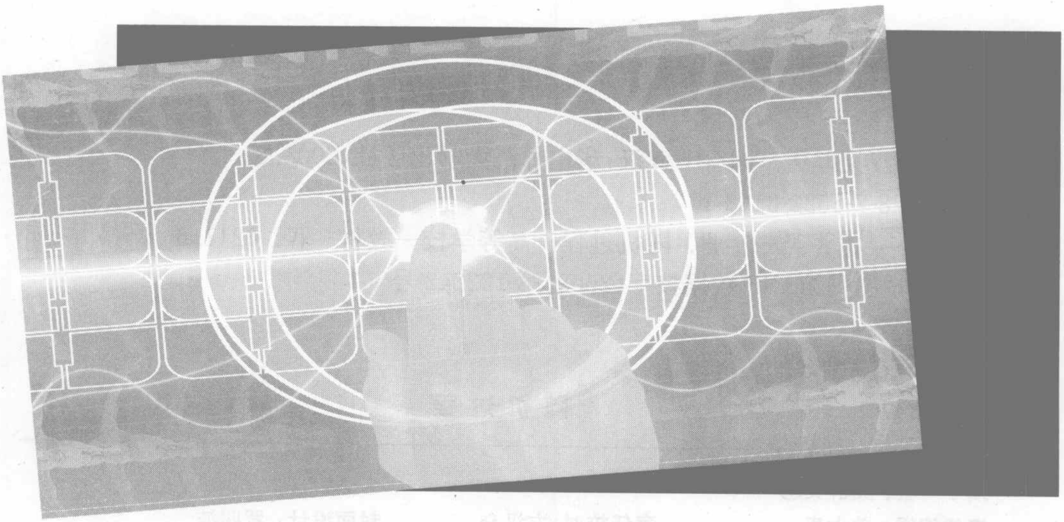


华中师范大学出版社

高等学校教材

随机过程

吴群英　编著



华中师范大学出版社
2008 · 武汉

内容简介

本书是随机过程的入门教材，较为系统地介绍了随机过程的基本概念、基本思想、基本原理和基本方法，主要内容有：概率论补充知识，泊松过程，更新理论，离散时间与连续时间马尔可夫链. 本书的特点是，在系统介绍随机过程基本理论和方法的同时，强调实际应用，着重讲清各种方法的实际背景和思想方法，注重渗透随机过程的思想，并尽力结合自然科学，特别是经济等领域的实际案例介绍随机过程的实用方法，把随机过程的方法与实际应用结合起来. 书中每一章都给出许多具有启发性的实例，一方面便于读者理解，另一方面也便于读者自学.

本书可作为高等学校本科生、研究生的随机过程课程教材，也可供教师以及从事概率统计研究的科研工作者阅读和参考.

新出图证(鄂)字 10 号

图书在版编目(CIP)数据

随机过程/吴群英　编著—武汉：华中师范大学出版社，2008. 8

ISBN 978-7-5622-3143-1

Ⅰ. 随…　Ⅱ. 吴…　Ⅲ. 随机过程—高等学校—教材　Ⅳ. 0211. 6

中国版本图书馆 CIP 数据核字(2008)第 122242 号

随 机 过 程

编　　著：吴群英©

责任编辑：曾太贵　　责任校对：方汉交　　封面设计：罗明波

编 辑 室：二编室　　电　　话：027—67867362

出版发行：华中师范大学出版社　　社　　址：湖北省武汉市珞瑜路 100 号

电　　话：027—67863040(发行部)　027—67861321(邮购)　传　　真：027—67863291

网　　址：http://www.ccnupress.com　　电子信箱：hscbs@public.wh.hb.cn

经　　销：新华书店湖北发行所

印　　刷：海军工程学院印刷厂　　督　　印：章光琼

字　　数：213 千字

开　　本：787mm×960mm　1/16　　印　　张：12.25

版　　次：2008 年 8 月第 1 版　　印　　次：2008 年 8 月第 1 次印刷

印　　数：1—2000　　定　　价：22.00 元

欢迎上网查询、购书

前　言

随机过程是随机数学的一个重要分支，其研究对象虽然与概率论一样是随机现象，但区别在于它主要研究的是随“时间”变化的、“动态”的、“整体”的随机现象的统计规律. 近 40 多年来，随着科学技术的发展，它已被广泛地应用到动态可靠性、生物工程、自动控制、无线电通信及管理学科等领域.

本书是作者多年来在为非数学专业的本科生、研究生开设随机过程课程的教学实践的基础上根据讲稿补充、修改而成. 本书是应用随机过程的入门教材，仅以初等概率论、数学分析(高等数学)、线性代数作为基础，可作为非数学专业的本科生及研究生的必修课教材，亦可作为教师以及工程管理人员的参考书.

本书力求突出以下几个特点：

1. 在不失随机过程系统性的前提下，明显不同于纯理类教材.

2. 着眼于引发读者的兴趣，每章都尽力安排有广泛的应用背景且引人入胜的例子，使读者不但逐步领悟到随机过程的思想方法，还逐步体会到随机过程的广泛应用性，以及随机过程方法的魅力和实用性.

3. 注重揭示随机过程的概念的来源及实际背景，典型随机模型的提炼、特性刻画、应用背景以及发展情况，使读者既能掌握随机过程的基本内容，又能学到解决实际问题的方法、思路与技巧.

4. 竭力以概率的观点来讲述随机过程的理论，而不是过分依赖于分析方法，只要有可能，我们都力求从概率的观点而不是分析的观点去考察随机过程，充分显示出概率分析的特点，并且竭力使用富有启发性，又非常有趣的直观推导的方法.

5. 有意安排并反复使用一些对解决应用概率问题十分有用的数学技巧，便于读者学会使用.

6. 随着现代科技的迅猛发展，条件数学期望将在其中发挥愈来愈重要的作用，本教材将条件数学期望作为现代随机过程的最基本的概念之一，以力求对这种发展趋势予以及时的反应和引导.

7. 突出条件概率公式及其他的基本思想和使用技巧，把它作为贯穿本教材的主导线索之一，并加以阐明和应用.

由于作者水平所限，本教材的讲稿虽经多次使用和修改，但书中一定还存在不少缺点和错误，恳请读者指正.

作　者

2008 年 5 月

数学符号说明

$\equiv$	恒等
$\triangleq$	“定义为”或“记为”
$\forall$	任意的
$\exists$	存在
$\in$	属于
$\prod$	连乘号
$\sum$	求和号
$\cup$	集合的并
$\cap$	集合的交
$\Leftrightarrow$	“等价”或者“充分必要”
$\stackrel{d}{=}$	同分布
$\xrightarrow{d}$	依分布收敛
$\uparrow$	单调递增
$\downarrow$	单调递减
a. s.	“几乎处处”或“以概率 1”
c. f.	特征函数
d. f.	分布函数
i. o.	有无穷多个
i. i. d.	独立同分布
r. v.	随机变量
C_n^m	n 个取 m 个的组合
$d(i)$	状态 i 的周期
$i \to j$	i 可达 j
$i \leftrightarrow j$	i 与 j 互通
I_A	集合 A 的示性函数
max	最大值
min	最小值
$n!$	n 的阶乘 $n! \triangleq n(n-1)\cdots 1$
$n!!$	n 的双阶乘 $n!! \triangleq \begin{cases} n(n-2)\cdots 2, 当\ n=2k, \\ n(n-2)\cdots 1, 当\ n=2k+1. \end{cases}$
$o(1)$	无穷小
$o(x)$	x 的高阶无穷小
$[x]$	不超过 x 的最大整数
$p_{ij}^{(n)}$	从 i 转移到 j 的 n 步转移概率
$F(x)$	X 的分布函数 $F(x) \triangleq P(X \leqslant x)$
$\mathbf{R}$	全体实数的集合
$\overline{\lim}$ 或 lim sup	上极限
$\underline{\lim}$ 或 lim inf	下极限
$\mathrm{cov}(X,Y)$	X 与 Y 的协方差
EX	X 的数学期望
$\mathrm{var}X$	X 的方差
$B(p,n)$	参数为 p,n 的二项分布
$E(\lambda)$	参数为 λ 的指数分布
$B^-(r,p)$	参数为 r,p 的负二项分布
$G(p)$	参数为 p 的几何分布
$N(\mu,\sigma^2)$	参数为 μ,σ^2 的正态分布
$P(\lambda)$	参数为 λ 的泊松分布
$\Gamma(n,\lambda)$	参数为 n,λ 的 Γ- 分布
$X \sim$	X 服从某分布
$\Gamma(\alpha)$	Γ 函数 $\Gamma(\alpha) \triangleq \int_0^\infty x^{\alpha-1} e^{-x} dx$

目 录

第1章　基础知识

概率论研究随机现象及其规律性，主要是通过研究随机变量(r. v.)$X(\omega) \triangleq X$,不考虑时间变化. 例如，抛硬币出现正面的概率. 如果把时间因素加进去，即考虑了过程的情况，则称$X(\omega,t) \triangleq X(t)$为随机过程. 因此，概率论是随机过程的特殊情况. 即当t固定在$t = t_0$时，$X(t_0)$即是随机变量.

随机过程是随机数学的一个重要分支，其研究对象虽然与概率论一样是随机现象，但区别在于它主要研究的是随“时间”变化的、“动态”的、“整体”的随机现象的统计规律. 事实上，在实际生活中有许多随机现象，通常用一个或有限多个随机变量去描述就可以了，这就是概率论知识；而还有另外一类随机现象，仅用一个或有限多个随机变量去描述则不能完全揭示这些随机现象的全部统计规律性. 因此在研究这些现象时，必须考虑其随时间进程发展变化的过程，不得不利用无穷多个随机变量去加以描述，而具有某种属性的无穷多个随机变量的集合就构成一个随机过程，因此，随机过程的诞生和发展，是科技发展的必然产物. 近40多年来，随着科学技术的发展，它已被广泛地应用到动态可靠性、生物工程、自动控制、无线电通信及管理学科等领域.

由于随机过程研究的是一族随机变量的统计规律性，一些基本研究方法借助于概率论方法，所以，学习随机过程，需要概率论的基础知识，因而，这一章我们将概括介绍随机过程需要用到但工科概率论未涉及或涉及不深的内容.

学过概率论的读者可通过思考以下几个概率中的问题，检查自己对概率的掌握情况.

1. 庄家设赌：参赌者抛骰子2次，连续出现2个6为胜，获胜可得20倍赌金. 问此赌博对庄家有利还是对参赌者有利?

2. 为什么普查某种疾病阳性者需复查?

3. 赌球预测：设每晚有2个球队进行比赛，连续6天压中胜队的人可获一笔可观的赌金. 现你每次比赛前都收到自称是某预测公司的一条预测此场比赛胜队的短信，每次都正确，这其中有爆出冷门的，有出人意料的，正当你感到这预测很神的时候，又收到了此预测公司的一条短信，声称本公司拥有顶级的概率统计学家，能100%精确地预测每场球的胜负，如你前面收到的短信就是实证，你只需往某某账号支付1000元，就可定时地获得预测的短信，从而获得一大笔赌金. 现问你

是否会支付这 1000 元的预测费?为什么?

4. 我国的彩票号码预测:某预测公司称该公司收集了以前中奖的号码，进行概率统计分析，并以此可预测未来中奖的号码. 你认为可否?为什么?

1.1 概　　率

1.1.1 事件序列$\{E_n, n \geqslant 1\}$的上极限和下极限

定义 1.1.1 事件序列$\{E_n, n \geqslant 1\}$的上极限定义为：$\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} E_m = \{$有无穷多个$E_n$发生$\} \triangleq \{E_n, \text{i. o.}\}$，记为$\overline{\lim\limits_{n\to\infty}} E_n$或$\limsup\limits_{n\to\infty} E_n$；$\{E_n\}$的下极限定义为$\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} E_m = \{$某时刻以后的所有$E_n$都发生$\}$，记为$\underline{\lim\limits_{n\to\infty}} E_n$或$\liminf\limits_{n\to\infty} E_n$. 显然有$\liminf\limits_{n\to\infty} E_n \subseteq \limsup\limits_{n\to\infty} E_n$. 如$\liminf\limits_{n\to\infty} E_n = \limsup\limits_{n\to\infty} E_n \triangleq E$，则称$\lim\limits_{n\to\infty} E_n = E$为事件序列$\{E_n, n \geqslant 1\}$的极限.

如$E_n \subseteq E_{n+1}$($E_n \supseteq E_{n+1}$)，则称事件序列$\{E_n, n \geqslant 1\}$是递增(减)的，记为$E_n \uparrow$($E_n \downarrow$).

性质 1.1.1 如$E_n \uparrow$，则$\lim\limits_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$；如$E_n \downarrow$，则$\lim\limits_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$.

证明 如$E_n \uparrow$，则有$\forall n \geqslant 1, \bigcap_{m=n}^{\infty} E_m = E_n, \bigcup_{m=n}^{\infty} E_m = \bigcup_{m=1}^{\infty} E_m$，故

$$\underline{\lim_{n\to\infty}} E_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} E_m = \bigcup_{n=1}^{\infty} E_n,$$

$$\overline{\lim_{n\to\infty}} E_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} E_m = \bigcup_{m=1}^{\infty} E_m,$$

因此

$$\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n.$$

1.1.2 概率的连续性

命题 1.1.1(概率的连续性)　设$\{E_n, n \geqslant 1\}$是递增(减)的事件序列，则

$$\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right).$$

即极限号“lim”与概率号“P”可交换.

证明 设$E_n \uparrow$，即$E_n \subseteq E_{n+1}$，定义F_n如下

$$F_1 = E_1, F_n = E_n\left(\bigcup_{i=1}^{n-1} E_i\right)^c = E_n E_{n-1}^c, n \geqslant 2.$$

这样定义的目的是使$F_1, \cdots, F_n$互不相容，且有$\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i = E_n, \bigcup_{i=1}^{\infty} F_i =$

$\bigcup\limits_{i=1}^{\infty}E_i$，所以由性质 1.1.1

$$P\left(\lim_{n\to\infty}E_n\right)=P\left(\bigcup_{i=1}^{\infty}E_i\right)=P\left(\bigcup_{i=1}^{\infty}F_i\right)=\sum_{i=1}^{\infty}P(F_i)$$
$$=\lim_{n\to\infty}\sum_{i=1}^{n}P(F_i)=\lim_{n\to\infty}P\left(\bigcup_{i=1}^{n}F_i\right)$$
$$=\lim_{n\to\infty}P\left(\bigcup_{i=1}^{n}E_i\right)=\lim_{n\to\infty}P(E_n).$$

同理可证当 $E_n\downarrow$ 时的情况.

例 1.1.1　考虑一个群体，它由能产生同类后代的个体构成. 初始的个体数用 X_0 表示，称为第0代的总数. 第0代的全体后代构成第1代，其总数以 X_1 表示. 一般地，以 X_n 表示第 n 代的总数. 求群体最终灭绝的概率.

解　因 $\{X_n=0\}\subseteq\{X_{n+1}=0\}$，所以，$P(X_n=0)\leqslant P(X_{n+1}=0)$，即数列 $P(X_n=0)\uparrow$，且有上界1，故 $\lim\limits_{n\to\infty}P(X_n=0)$ 存在，由概率的连续性有

$$\lim_{n\to\infty}P(X_n=0)=P\left(\lim_{n\to\infty}X_n=0\right)=P\left(\bigcup_{n=1}^{\infty}(X_n=0)\right)$$
$$=P(\exists n,\text{使 }X_n=0)=P(\text{群体迟早灭绝}).$$

即第 n 代没有个体的极限概率等于此群体最终灭绝的概率.

命题 1.1.2　Borel-Cantelli 引理

(1) 设 $\{E_n,n\geqslant 1\}$ 是一列事件，如 $\sum\limits_{n=1}^{\infty}P(E_n)<\infty$，则 $P(E_n,\text{i.o.})=0$；

(2) 如 $E_1,E_2,\cdots$ 相互独立，且 $\sum\limits_{n=1}^{\infty}P(E_n)=\infty$，则 $P(E_n,\text{i.o.})=1$.

证明　(1) 记 $F_n=\bigcup\limits_{k=n}^{\infty}E_k$，则 $F_n\downarrow$，由性质 1.1.1

$$\{E_i,\text{i.o.}\}=\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}E_k=\bigcap_{n=1}^{\infty}F_n=\lim_{n\to\infty}F_n,$$

由命题 1.1.1，即概率的连续性以及条件 $\sum\limits_{k=1}^{\infty}P(E_k)<\infty$，有

$$P(E_i,\text{i.o.})=P\left(\lim_{n\to\infty}F_n\right)=\lim_{n\to\infty}P(F_n)$$
$$=\lim_{n\to\infty}P\left(\bigcup_{k=n}^{\infty}E_k\right)\leqslant\lim_{n\to\infty}\sum_{k=n}^{\infty}P(E_k)=0.$$

(2) 由上式的证明过程以及 $E_1,E_2,\cdots$ 的相互独立性，有

$$P(E_i,\text{i.o.})=\lim_{n\to\infty}P\left(\bigcup_{k=n}^{\infty}E_k\right)=\lim_{n\to\infty}\left(1-P\left(\bigcap_{k=n}^{\infty}E_k^c\right)\right)$$
$$=\lim_{n\to\infty}\left(1-\prod_{k=n}^{\infty}P(E_k^c)\right)$$

$$= 1 - \lim_{n\to\infty}\prod_{k=n}^{\infty}(1-P(E_k)), \tag{1.1.1}$$

因为$1-x\leqslant e^{-x}$，所以$1-P(E_k)\leqslant e^{-P(E_k)}$，再由条件$\sum\limits_{n=1}^{\infty}P(E_n)=\infty$得对每一个$n$，$\sum\limits_{k=n}^{\infty}P(E_k)=\infty$，因此

$$0\leqslant\prod_{k=n}^{\infty}(1-P(E_k))\leqslant\prod_{k=n}^{\infty}e^{-P(E_k)}=e^{-\sum\limits_{k=n}^{\infty}P(E_k)}=0,$$

故由式(1.1.1)得$P(E_i,\text{i.o.})=1$.

1.2 数学期望与条件数学期望

1.2.1 数学期望

在工科概率论中，已分别就离散型随机变量和连续型随机变量定义了数学期望. 我们首先回顾一下.

当X为离散型随机变量，分布列为$p_i=P\{X=x_i\},i=1,2,\cdots$时，若$\sum\limits_{i=1}^{\infty}|x_i|p_i<\infty$，则定义$X$的数学期望为

$$EX=\sum_{i=1}^{\infty}x_ip_i. \tag{1.2.1}$$

当X为连续型随机变量，密度函数为$f(x)$时，若$\int_{-\infty}^{\infty}|x|f(x)\mathrm{d}x<\infty$，则定义$X$的数学期望为

$$EX=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x. \tag{1.2.2}$$

下面我们就一般情况给出随机变量X的数学期望的定义. 设随机变量X的分布函数(d. f.)为$F(x)$，即$F(x)\triangleq P(X\leqslant x)$，若

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

则定义X的数学期望为

$$EX=\int_{-\infty}^{\infty}x\mathrm{d}F(x). \tag{1.2.3}$$

这里所用的积分是 H. L. Lebesgue-Stieltjes 积分. 特别地，当X为离散型时，它的分布函数$F(x)$是阶梯函数，此积分就成为求和的形式，此时，式(1.2.3)就成为式(1.2.1)；当X为连续型时，此时积分就化为$\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$，即成为 G. F.

B. Riemann 积分(就是高等数学中讲过的那种积分)，此时，式(1.2.3) 就成为式(1.2.2). 后面关于随机变量 X 的函数的数学期望也可以用 H. L. Lebesgue-Stieltjes 积分统一表示.

若$\int_{-\infty}^{\infty} |g(x)| \mathrm{d}F(x) < \infty$，则 $g(X)$ 的数学期望 $Eg(X)$ 存在，且

$$Eg(X) = \int_{-\infty}^{\infty} g(x)\mathrm{d}F(x).$$

特别地，当 X 为离散型或连续型随机变量时，我们有公式

$$Eg(X) = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x, \text{当 } X \text{ 是连续型}, \\ \sum_n g(x_n)P(X = x_n), \text{当 } X \text{ 是离散型}. \end{cases}$$

下面给出数学期望的几个重要公式，假设所提到的 r.v. 的数学期望都存在.

1. 如随机变量 X 是非负的，$F(x)$ 是 X 的分布函数，记 $\overline{F}(x) \triangleq 1 - F(x)$，则

$$EX = \int_0^{\infty} \overline{F}(x)\mathrm{d}x = \int_0^{\infty} P(X > x)\mathrm{d}x, \tag{1.2.4}$$

$$EX^p = \int_0^{\infty} px^{p-1}P(X > x)\mathrm{d}x, p > 0, \tag{1.2.5}$$

因此对任意随机变量 X，总有

$$E|X| = \int_0^{\infty} P(|X| > x)\mathrm{d}x.$$

证明　由数学期望的定义式(1.2.3)，交换积分顺序，有

$$\begin{aligned} EX &= \int_0^{\infty} x\mathrm{d}F(x) = \int_0^{\infty}\int_0^{x} \mathrm{d}t\mathrm{d}F(x) = \int_0^{\infty} \mathrm{d}t\int_t^{\infty} \mathrm{d}F(x) = \int_0^{\infty} (1 - F(t))\mathrm{d}t \\ &= \int_0^{\infty} (1 - F(x))\mathrm{d}x = \int_0^{\infty} \overline{F}(x)\mathrm{d}x = \int_0^{\infty} P(X > x)\mathrm{d}x. \end{aligned}$$

或由

$$\int_0^{\infty} \overline{F}(x)\mathrm{d}x = \int_0^{\infty}\int_x^{\infty} \mathrm{d}F(y)\mathrm{d}x = \int_0^{\infty} \mathrm{d}F(y)\int_0^{y} \mathrm{d}x = \int_0^{\infty} y\mathrm{d}F(y) = EX.$$

故式(1.2.4) 成立.

下证式(1.2.5). 由式(1.2.4)，并作代换 $x = y^n$ 即得

$$EX^p = \int_0^{\infty} P(X^p > x)\mathrm{d}x = \int_0^{\infty} P(X > x^{\frac{1}{p}})\mathrm{d}x = \int_0^{\infty} py^{p-1}P(X > y)\mathrm{d}y.$$

2. $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i.$ $\quad$ (1.2.6)

3. 如随机变量 $X_1, \cdots, X_n$ 相互独立，则

$$E\Big(\prod_{i=1}^{n}X_i\Big)=\prod_{i=1}^{n}EX_i. \tag{1.2.7}$$

4. $$\mathrm{var}\Big(\sum_{i=1}^{n}X_i\Big)=\sum_{i=1}^{n}\mathrm{var}X_i+2\sum_{1\leqslant i<j\leqslant n}\mathrm{cov}(X_i,X_j)=\sum_{i=1}^{n}\mathrm{var}X_i+\sum_{i\neq j}\mathrm{cov}(X_i,X_j), \tag{1.2.8}$$

特别地，如 $X_1,\cdots,X_n$ 相互独立，则 $\mathrm{var}\Big(\sum_{i=1}^{n}X_i\Big)=\sum_{i=1}^{n}\mathrm{var}X_i$.

5. $$\mathrm{cov}(X_i,X_j)=E(X_iX_j)-EX_iEX_j,\ \mathrm{cov}(X_i,X_i)=\mathrm{var}X_i. \tag{1.2.9}$$

例 1.2.1 匹配问题. 在一次集会上，n 个人把他们的帽子放到房间的中央混合在一起，而后每个人随机地选取一顶，记 X 为拿到自己的帽子的人数. 求 X 的均值与方差.

解 为了求解，我们利用表示式

$$X=X_1+X_2+\cdots+X_n,$$

其中

$$X_i=\begin{cases}1,\text{如果第 } i \text{ 个人拿到自己的帽子},\\ 0,\text{其他}.\end{cases}$$

因为第 i 个人是等可能地选到 n 顶帽子中的任何一顶，所以得 $P\{X_i=1\}=\dfrac{1}{n}$，从而

$$EX_i=\frac{1}{n},$$

$$\mathrm{var}X_i=\frac{1}{n}\Big(1-\frac{1}{n}\Big)=\frac{n-1}{n^2}.$$

又由于

$$X_iX_j=\begin{cases}1,\text{如果第 } i \text{ 个人与第 } j \text{ 个人都拿到自己的帽子},\\ 0,\text{其他}.\end{cases}$$

所以当 $i<j$ 时，有

$$E(X_iX_j)=P\{X_i=1,X_j=1\}=P\{X_i=1\}P\{X_j=1\mid X_i=1\}=\frac{1}{n}\,\frac{1}{n-1},$$

由公式(1.2.9)

$$\mathrm{cov}(X_i,X_j)=E(X_iX_j)-EX_iEX_j,$$

因此，

$$\mathrm{cov}(X_i,X_j)=\frac{1}{n(n-1)}-\frac{1}{n^2}=\frac{1}{n^2(n-1)},$$

所以，由期望和方差的公式(1.2.6) 和(1.2.8) 得

$$EX = 1,$$
$$\mathrm{var}X = \frac{n-1}{n} + 2C_n^2 \frac{1}{n^2(n-1)} = 1.$$

于是匹配数的均值与方差都是1.

1.2.2　条件数学期望

给定 $Y = y$，定义 X 的条件数学期望为

$$E(X \mid Y = y) \triangleq \int x \mathrm{d}F(x \mid y).$$

特别地，对离散型和连续型 r. v.，有

$$E(X \mid Y = y) = \begin{cases} \sum_n x_n P(X = x_n \mid Y = y), \text{当} X, Y \text{是离散型}, \\ \int_{-\infty}^{\infty} x f(x \mid y) \mathrm{d}x, \text{当} X, Y \text{是连续型}. \end{cases}$$

其中 $P(X = x \mid Y = y) = \dfrac{P(X = x, Y = y)}{P(Y = y)} (P(Y = y) > 0)$，以及

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) \mathrm{d}x}, \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x \neq 0.$$

以 $E(X \mid Y)$ 表示 r. v. Y 的函数，它在 $Y = y$ 时，取值为 $E(X \mid Y = y)$. 条件期望的极其有用的性质是对一切随机变量 X 和 Y，当期望存在时，有以下公式.

(1)
$$EX = E[E(X \mid Y)] = \int_{-\infty}^{\infty} E(X \mid Y = y) \mathrm{d}F_Y(y) \tag{1.2.10}$$

特别地，

$$EX = \begin{cases} \sum_n E(X \mid Y = y_n) P(Y = y_n), \text{当} Y \text{是离散型}, \\ \int_{-\infty}^{\infty} E(X \mid Y = y) f_Y(y) \mathrm{d}y, \text{当} Y \text{是连续型}. \end{cases} \tag{1.2.11}$$

先对一个随机变量取条件，不仅使我们能求得期望，也可以用这种方法计算概率.

(2)
$$P(A) = \int_{-\infty}^{\infty} P(A \mid Y = y) \mathrm{d}F_Y(Y). \tag{1.2.12}$$

特别地，

$$P(A) = \begin{cases} \sum_n P(A \mid Y = y_n) P(Y = y_n), \text{当} Y \text{是离散型}, \\ \int_{-\infty}^{\infty} P[A \mid Y = y] f_Y(y) \mathrm{d}y, \text{当} Y \text{是连续型}. \end{cases} \tag{1.2.13}$$

如 $B_1,\cdots,B_n$ 是一个划分，令 $Y = iI_{B_i}$，其中，$I_B \triangleq \begin{cases}1, B\text{ 发生},\\0, B\text{ 不发生}\end{cases}$ 称为集合 B 的示性函数，则 $Y = i \Leftrightarrow B_i$ 发生，$i = 1,2,\cdots,n$. 所以由公式(1.2.13) 即得熟知的全概率公式：$P(A) = \sum_{i=1}^{n} P(A \mid Y = i)P(Y = i) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$. 因此，公式(1.2.12) 是全概率公式的推广.

证明　(1) 我们仅证当 X,Y 是连续型的情况，其余情况类似. 因为

$$\begin{aligned}\int_{-\infty}^{\infty} E(X \mid Y = y)f_Y(y)\mathrm{d}y &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x \mid y)\mathrm{d}xf_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)}f_Y(y)\mathrm{d}y\mathrm{d}x \\ &= \int_{-\infty}^{\infty} xf_X(x)\mathrm{d}x = EX.\end{aligned}$$

故(1.2.10) 成立.

(2) 记 $X \triangleq X(\omega) \triangleq I_A = \begin{cases}1, A\text{ 发生},\\0, A\text{ 不发生}\end{cases} = \begin{cases}1, \omega \in A,\\0, \omega \notin A.\end{cases}$ 即 X 的分布律为

X	1	0
P	$P(A)$	$P(\overline{A})$

，

故

$$EI_A = EX = P(A),$$
$$E(I_A \mid Y = y) = E(X \mid Y = y) = P(A \mid Y = y),$$

由公式(1.2.10) 得

$$P(A) = EX = \int_{-\infty}^{\infty} E(X \mid Y = y)\mathrm{d}E_Y(y) = \int_{-\infty}^{\infty} P(A \mid Y = y)\mathrm{d}F_Y(y).$$

例 1.2.2　假定进行 n 次独立试验，每次试验的结果为 $1,2,\cdots,r$ 中的一个，它们出现的概率分别为 $p_1,p_2,\cdots,p_r$，$\sum_{i=1}^{r} p_i = 1$，以 N_i 记试验结果为 i 的次数.

(1) 计算 $N_1,\cdots,N_r$ 的联合分布，亦称为多项分布.

(2) 计算 $\mathrm{cov}(N_i,N_j)$.

(3) 计算未出现的结果的个数的均值与方差.

解　(1) 由题意，当 $n_1 + n_2 + \cdots + n_r = n$ 时，利用概率的乘法公式，

$$\begin{aligned}&P(N_1 = n_1, N_2 = n_2, \cdots, N_r = n_r) \\ &= P(N_1 = n_1)P(N_2 = n_2 \mid N_1 = n_1)\cdots P(N_r = n_r \mid N_1 = n_1,\cdots,N_{r-1} = n_{r-1}) \\ &= \mathrm{C}_n^{n_1}p_1^{n_1}\mathrm{C}_{n-n_1}^{n_2}p_2^{n_2}\cdots\mathrm{C}_{n-n_1-n_2-\cdots-n_{r-1}}^{n_r}p_r^{n_r} \\ &= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-n_2-\cdots-n_{r-1})!}{n_r!0!}\prod_{i=1}^{r} p_i^{n_i}\end{aligned}$$

$$= \frac{n!}{\prod_{i=1}^{r} n_i!} \prod_{i=1}^{r} p_i^{n_i}.$$

(2) 把每次试验结果简化为出现结果 i 和不出现结果 i，则 $P(X=i)=p_i$，$N_i \sim B(p_i, n)$，其中，$B(\cdot,\cdot)$ 表示二项分布. 所以，

$EN_i = np_i, \mathrm{var}N_i = np_i(1-p_i), EN_i^2 = \mathrm{var}N_i + (EN_i)^2 = np_i(1-p_i+np_i)$.

当 $i \neq j$ 时，又因为 $N_i \mid (N_j = m)$ 可看成在 $n-m$ 个试验中，可能出现的结果为 $1,2,\cdots,j-1,j+1,\cdots,r$，同样每次试验结果简化为出现 i 和不出现 i，出现 i 的概率为 $p = P(X=i \mid X \neq j) = \dfrac{P(X=i)}{P(X \neq j)} = \dfrac{p_i}{1-p_j}$，即相应出现 i 的概率为 $\dfrac{p_i}{1-p_j}, i = 1,\cdots,j-1,j+1,\cdots,r$，所以 $(N_i \mid (N_j = m)) \sim B\left(\dfrac{p_i}{1-p_j}, n-m\right)$，故

$$E\Big[N_i \mid (N_j = m)\Big] = (n-m)\frac{p_i}{1-p_j},$$

因此

$$\begin{aligned}
E(N_iN_j) &= \sum_{m=0}^{n} E\Big[N_iN_j \mid N_j = m\Big] P(N_j = m) \\
&= \sum_{m=0}^{n} m \cdot (n-m)\frac{p_i}{1-p_j} P(N_j = m) \\
&= \frac{p_i}{1-p_j}\sum_{m=0}^{n}(nm - m^2)P(N_j = m) = \frac{p_i}{1-p_j}\Big[nEN_j - EN_j^2\Big] \\
&= \frac{p_i}{1-p_j}\Big[n^2p_j - np_j + np_j^2 - n^2p_j^2\Big] = np_ip_j(n-1),
\end{aligned}$$

故由公式(1.2.9)得

$$\mathrm{cov}(N_i, N_j) = E(N_iN_j) - EN_iEN_j = np_ip_j(n-1) - n^2p_ip_j = -np_ip_j, i \neq j.$$

$$\mathrm{cov}(N_i, N_i) = \mathrm{var}N_i = np_i(1-p_i).$$

(3) 记 $I_j = \begin{cases} 1, \text{结果 } j \text{ 从未出现}, \\ 0, \text{其余}. \end{cases}$ 则

$$EI_j = EI_j^2 = P(I_j = 1) = P(\text{结果 } j \text{ 从未发生}) = (1-p_j)^n.$$

$$\mathrm{var}I_j = EI_j^2 - (EI_j)^2 = (1-p_j)^n\Big[1-(1-p_j)^n\Big],$$

$$E(I_iI_j) = P(I_iI_j = 1) = P(\text{结果 } i,j \text{ 从未出现}) = (1-p_i-p_j)^n, i \neq j,$$

$$\mathrm{cov}(I_i, I_j) = E(I_iI_j) - EI_iEI_j = (1-p_i-p_j)^n - (1-p_i)^n(1-p_j)^n.$$

故未出现结果的个数 $X = \sum_{j=1}^{r} I_j$，由公式(1.2.6)，(1.2.8)有