高等学校教材

随机过程

苏中根 编著

高等教育出版社

高等学校教材

随机过程

Suiji Guocheng

苏中根 编著

高等教育出版社·北京

内容提要

本书介绍随机过程的基本概念和基本理论，着重讲解 Poisson 过程、Markov 链、Galton-Watson 分枝过程、鞅、Brown 运动和平稳随机过程遍历性．本书选材恰当，内容丰富，深入浅出．除前两章外，各章内容相对独立并且体系完整，便于读者阅读．每章含有附录，包括人物和背景介绍，趣味性和科学性兼顾．每章习题都经过精心挑选，难易适中，可作为正文的有益补充．

本书可作为高等学校理工科类“随机过程”课程教材或参考书，也可供其他科研人员参考．

图书在版编目（CIP）数据

随机过程 / 苏中根编著．-- 北京：高等教育出版社，2016. 2

ISBN 978-7-04-044729-3

Ⅰ．①随… Ⅱ．①苏… Ⅲ．①随机过程 - 高等学校 - 教材 Ⅳ．① O211.6

中国版本图书馆 CIP 数据核字（2016）第 020854 号

策划编辑 胡 颖　　责任编辑 胡 颖　　封面设计 钟 雨　　版式设计 杜微言
插图绘制 郝 林　　责任校对 陈旭颖　　责任印制 毛斯璐

出版发行 高等教育出版社
社　　址 北京市西城区德外大街4号
邮政编码 100120
印　　刷 三河市华骏印务包装有限公司
开　　本 787 mm × 960 mm 1/16
印　　张 16
字　　数 290 千字
购书热线 010-58581118
咨询电话 400-810-0598
网　　址 http://www.hep.edu.cn
http://www.hep.com.cn
网上订购 http://www.hepmall.com.cn
http://www.hepmall.com
http://www.hepmall.cn
版　　次 2016 年 2 月第 1 版
印　　次 2016 年 2 月第 1 次印刷
定　　价 30.10 元

物 料 号 44729-00

前　言

随机过程是一族随机变量, 用于描述随时间而变化的随机现象. 研究随机过程目的在于揭示随机现象变化和发展的内在规律与机制. 由于随机过程广泛存在于各种实际问题中, 其现已成为概率论、数理统计、计量经济、金融数学、商业管理、运筹优化、计算机科学、统计物理等学科中不可缺少的研究内容. 许多高等院校都开设随机过程课程, 深受广大学生欢迎.

作者近年来一直在浙江大学开设随机过程课程, 并在教学过程中不断收集资料, 编写讲义和课件, 倾听同学们意见和建议. 本书正是在此基础上, 经过反复酝酿和修改, 编写而成. 全书共分八章, 概括如下:

第一章回顾概率论的基本知识, 包括 Kolmogorov 的概率空间公理化定义、随机变量及其分布函数、随机变量数字特征和经典极限定理. 为了今后各章的需要, 本章特别强调条件概率、条件分布和条件期望的概念和运算, 给出全概率公式和全期望公式.

第二章介绍随机过程基本概念. 随机过程的概率分布由任意有限维分布来刻画. 本书所讨论的随机过程形式多样, 特色鲜明, 既有离散时间过程, 也有连续时间过程; 既有可数状态过程, 也有不可数状态过程; 既有增量独立过程, 也有条件独立过程.

第三章介绍 Poisson 过程. 该过程用于描述随机服务系统, 统计寻求服务的顾客数, 其增量独立并服从 Poisson 分布. Poisson 过程是连续时间、取非负整数值的 Markov 过程. 除齐次 Poisson 过程外, 本章还介绍非齐次 Poisson 过程和复合 Poisson 过程, 这些常常出现在实际应用问题中.

第四章介绍 Markov 链. 该过程状态空间最多含有可数多个状态, 随着时间推移, 在不同状态之间进行转移. Markov 链具有条件独立性, 在给定当前状态下, 将来处于何种状态与过去所经历的过程无关. Markov 链的分布由初始分布和转移概率矩阵所决定, 从长远来看, 转移概率矩阵起着更为关键的作用. Markov 链可以用于预测和决策问题.

第五章介绍 Galton-Watson 分枝过程. 它是一种特殊的 Markov 链, 用于描述物种繁衍、细胞裂变等增长现象. 生成函数是研究 Galton-Watson 分枝过程的一个有效工具.

第六章介绍鞅. 与其他过程比较, 鞅的直观背景并不明显, 但是它的应用非常广泛, 灵活多变, 技巧性强. 基本收敛定理和停时定理是鞅论中最具特色的基本内容.

第七章介绍 Brown 运动. 该过程是连续时间参数、取实数值的正态过程, 具有独立平稳增量, 从而是连续鞅和 Markov 过程. Brown 运动可以用于描述粒子运动, 其轨迹处处连续但极不规则. 除去一些基本性质之外, 本章还简要介绍 Itô 积分及其在金融数学中的应用, 推导出 Black-Scholes 关于欧式买入期权的定价公式.

第八章介绍平稳随机过程遍历性. 平稳随机过程具有时间平移不变性, 各个时刻的数学期望和方差都是常数. 从而在相当宽的条件下, 可以用时间平均代替统计平均, 为随机过程的统计推断奠定了基础.

本书除前两章外, 其他各章内容相对独立, 读者可以根据实际情况加以选择. 每章都含有附录, 作为正文阅读的延伸和补充. 附录中的人物介绍, 可以帮助读者了解有关随机过程提出和研究的历史背景. 书中配有大量习题, 难易程度适中, 读者通过练习, 可以更好地理解正文中所介绍的基本概念和基本结果.

本书绝大多数内容只涉及经典随机过程的基本知识, 读者仅需要具备一定的线性代数、微积分和概率论基础理论; 少量内容可能需要实变函数和测度论, 读者初学时可以跳过.

在本书编写过程中, 许多同学和老师提出了宝贵意见和建议, 作者深表感谢.
限于作者水平和能力, 书中肯定存在不妥之处, 敬请读者批评指正.

苏中根

中国 · 杭州

2015 年 9 月

目　录

第一章　初等概率论

本章简要回顾初等概率论的基本内容, 包括概率空间的公理化定义, 随机变量分布函数的基本性质, 数学期望的运算, 以及经典概率极限定理, 供今后各章参考.

§1.1　概率空间

一、概率空间

概率空间由三要素组成: 样本空间 Ω, 随机事件域 $\mathcal{A}$ 和概率 P, 通常记为 $(\Omega, \mathcal{A}, P)$. 具体地说, Ω 表示随机试验 (现象) 所有可能的基本结果; $\mathcal{A}$ 由一类随机事件 (简称事件) 所组成, 满足下列条件:

(i) $\Omega, \varnothing \in \mathcal{A}$,

(ii) 如果 $A \in \mathcal{A}$, 那么 $\overline{A} \in \mathcal{A}$,

(iii) 如果 $A_n \in \mathcal{A}, n \geqslant 1$, 那么 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$;

$P : \mathcal{A} \mapsto [0,1]$ 满足下列条件:

(i′) 规范性: $P(\varnothing) = 0$, $P(\Omega) = 1$,

(ii′) 可加性: 如果 $(A_n, n \geqslant 1)$ 是一列互不相容事件, 那么

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

概率空间是随机试验 (现象) 高度抽象化的数学模型, 其核心是概率 P, 不同的 P 代表着不同的概率模型. 建立恰当的概率模型需要对随机试验 (现象) 有着充分了解, 往往需要有关学科的背景知识. 初等概率论主要介绍计算或估计事件概率大小的理论和方法: 从一些简单事件的概率开始, 通过事件运算和概率运算来计算复杂事件的概率.

事件运算类似于集合运算, 如和、差、交、并等. 概率运算性质可以由上述 (i′) 和 (ii′) 推出, 如

(1) 对任意 A, $P(\overline{A}) = 1 - P(A)$;

(2) 如果 $B \subset A$, 那么 $P(A \setminus B) = P(A) - P(B)$;

(3) 多退少补: 对任意 $A_1, A_2, \cdots, A_m$,

$$P\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} P(A_i) - \sum_{1\leqslant i<j\leqslant m} P(A_iA_j) + \cdots + (-1)^{m-1}P(A_1A_2\cdots A_m);$$

(4) 次可加性: 对任意一列事件 $(A_n, n \geqslant 1)$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} P(A_n);$$

(5) 连续性: 假设 $(A_n, n \geqslant 1)$ 是一列单调递增 (或递减) 事件, 则

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n).$$

常见的概率模型包括古典概率模型和几何概率模型. 古典概率模型特点如下: $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$, 其中 $N < \infty$; $\mathcal{A} = 2^{\Omega}$; 每个结果等可能发生, 即 $P(\{\omega_i\}) = \dfrac{1}{N}$. 特别, 对每个事件 $A \in \mathcal{A}$,

$$P(A) = \frac{|A|}{|\Omega|},$$

其中 $|A|$ 表示事件 A 中所含基本结果的个数.

作为上述模型的连续形式, 几何概率模型特点如下: Ω 是 $\mathbb{R}^d$ $(d \geqslant 1)$ 上的可测区域; $\mathcal{A}$ 为 Ω 中 Borel 可测子集的全体; 每个结果等可能发生. 特别, 对每个事件 $A \in \mathcal{A}$,

$$P(A) = \frac{|A|}{|\Omega|},$$

其中 $|A|$ 表示事件 A 中的 Borel 测度. 注意, 给定每个基本结果 ω, $P(\{\omega\}) = 0$.

二、条件概率

假设事件 $B \in \mathcal{A}$, $P(B) > 0$, 定义

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad A \in \mathcal{A}, \tag{1.1}$$

称 $P(A|B)$ 为给定 B 发生的条件下, A 发生的**条件概率**.

注 1.1　给定事件 B, $P(\cdot|B) : \mathcal{A} \mapsto [0,1]$ 定义了 $\mathcal{A}$ 上的一个概率, 满足条件 (i′) 和 (ii′), 从而 $(\Omega, \mathcal{A}, P(\cdot|B))$ 为概率空间.

条件概率是计算事件概率的一个强有力工具. 特别, 下列**全概率公式**经常使用.

定理 1.1 假设 $B_1, B_2, \cdots, B_N$ $(N \leqslant \infty)$ 是一列互不相容事件, $P(B_n) > 0$, 并且 $\Omega = \sum\limits_{n=1}^{N} B_n$. 那么对任意事件 A,

$$P(A) = \sum_{n=1}^{N} P(A|B_n)P(B_n).$$

特别, 任意给定 B, 如果 $0 < P(B) < 1$, 那么

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}).$$

(1.1) 式可以改写成

$$P(AB) = P(B)P(A|B).$$

该式可以推广到多个事件, 得下列**链式法则**:

$$P(A_1A_2\cdots A_m) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1)\cdots P(A_m|A_{m-1}A_{m-2}\cdots A_1).$$

三、独立性

注意, 条件概率 $P(A|B)$ 可能比 $P(A)$ 大, 也可能比 $P(A)$ 小. 如果

$$P(A|B) = P(A), \tag{1.2}$$

则称事件 A 和 B **相互独立**. (1.2) 意味着事件 B 的发生对事件 A 发生的概率大小没有影响. 由 (1.1) 得, A 和 B 相互独立等价于

$$P(AB) = P(A)P(B). \tag{1.3}$$

有时称 (1.3) 为**乘法公式**. 使用公式 (1.3) 作为独立性定义更方便, 它并不要求 $P(A) > 0$ 或 $P(B) > 0$. 特别, 不可能事件 $\varnothing$, 必然事件 Ω 与任何其他事件相互独立.

独立性概念是概率论学科中最重要的概念之一, 可以推广到任意多个事件. 假设 $A_1, A_2, \cdots, A_m$ 满足下列 $2^m - m - 1$ 个方程:

$$\begin{cases} P(A_iA_j) = P(A_i)P(A_j), \quad i < j, \\ P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k), \quad i < j < k, \\ \cdots\cdots\cdots\cdots \\ P(A_{i_1}A_{i_2}\cdots A_{i_{m-1}}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_{m-1}}), \quad i_1 < i_2 < \cdots < i_{m-1}, \\ P(A_1A_2\cdots A_m) = P(A_1)P(A_2)\cdots P(A_m), \end{cases}$$

则称 $A_1, A_2, \cdots, A_m$ **相互独立**. 由定义可以看出, 如果 m 个事件相互独立, 那么其中任何一组事件都相互独立.

更一般地, 假设 $\mathcal{A}_1$, $\mathcal{A}_2$ 为 $\mathcal{A}$ 的两个子 σ–域, 如果任何事件 $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$ 都有 A_1 和 A_2 相互独立, 称 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ **相互独立**.

§1.2 随机变量

一、分布函数

假设 $(\Omega, \mathcal{A}, P)$ 是概率空间, $X : \Omega \mapsto \mathbb{R}$ 是一个映射. 给定事件 $B \in \mathcal{B}$, 记

$$X^{-1}(B) = \{\omega \in \Omega : X \in B\}.$$

如果对任意 $B \in \mathcal{B}$, 下列可测性条件满足:

$$X^{-1}(B) \in \mathcal{A},$$

则称 X 是**随机变量**. 根据 Borel σ–域的构造, X 是随机变量当且仅当

$$X^{-1}((-\infty, x]) \in \mathcal{A}, \quad x \in \mathbb{R}.$$

给定一个随机变量 X, 定义

$$F_X(x) = P(\omega : X \leqslant x), \quad x \in \mathbb{R},$$

称 F_X 为 X 在概率 P 下的**分布函数**. 当上下文明确时, 可以省略下标 X, 仅写作 F. 分布函数 F_X 完全确定了随机变量 X 的值和取每个值的概率大小, 是概率论中最基本的研究对象. 由概率 P 的基本性质, 可以得到

(1) $F_X(-\infty) = 0, F_X(\infty) = 1$;

(2) F_X 关于 x 单调增加;

(3) F_X 关于 x 右连续;

(4) 任意给定 $a < b$,

$$P(a < X \leqslant b) = F_X(b) - F_X(a).$$

注 1.2 如果定义 $F_X(x) = P(\omega : X < x)$, 那么 F_X 是左连续的.

注 1.3 有时称任何满足上述 (1)—(3) 的函数 $F(x)$ 为分布函数. 事实上, 任意给定一个这样的函数 $F(x)$, 总能构造一个概率空间 $(\Omega, \mathcal{A}, P)$ 以及一个随机变量 $X : \Omega \mapsto \mathbb{R}$, 使得 X 在 P 下的分布函数 F_X 恰好就是 F.

最为常见的随机变量包括离散随机变量和连续随机变量. 离散随机变量最多只取可数个值, 一般可表示为

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}, \tag{1.4}$$

其中 $N \leqslant \infty$. 这时, F_X 可写成

$$F_X(x) = \sum_{i:x_i \leqslant x} p_i.$$

它为阶梯形分布函数, x_i 为跳跃点,跳跃高度为 p_i. 典型的离散分布如: 两点分布、Bernoulli 分布、Poisson 分布、几何分布、超几何分布、负二项分布等.

连续随机变量取值为 $\mathbb{R}$ 上一个区间或多个区间, 并存在一个**密度函数** $p(x)$ 使得

$$F(x) = \int_{-\infty}^{x} p(u)\mathrm{d}u, \quad x \in \mathbb{R}.$$

该分布函数为绝对连续函数, 其导数为 $p(x)$. 典型的连续分布如: 均匀分布、指数分布、正态分布、Cauchy 分布, Gamma 分布、t 分布、χ^2 分布等.

典型的离散和连续随机变量如同数学分析中的初等函数, 是概率论中构造随机变量的最基本元素. 假如 $X : \Omega \mapsto \mathbb{R}$ 是随机变量, $f : \mathbb{R} \mapsto \mathbb{R}$ 是 Borel 可测函数, 那么 $Y = f(X)$ 是随机变量. 但计算 Y 的分布并没有一个普遍适用的方法. 一般来说, 如果 X 是离散随机变量, 那么 Y 一定是离散的, 其分布不难计算; 如果 X 是连续的, Y 的取值与分布依赖于函数 f 的定义.

二、随机向量

假设 $(X, Y) : \Omega \mapsto \mathbb{R}^2$ 是二元可测映射, 即对任意事件 $B \in \mathcal{B}^2$, 都有

$$\{\omega \in \Omega : (X, Y) \in B\} \in \mathcal{A},$$

称 (X, Y) 是**随机向量**. 给定随机向量 (X, Y), 定义二元**联合分布函数**

$$F_{X,Y}(x, y) = P(\omega : X \leqslant x, Y \leqslant y).$$

它具有下列基本性质:

(1) 任意给定 $x, y \in \mathbb{R}$,

$$F_{X,Y}(-\infty, y) = F(x, -\infty) = 0, \quad F_{X,Y}(\infty, \infty) = 1;$$

(2) $F_{X,Y}$ 关于 x 和 y 单调增加;

(3) $F_{X,Y}$ 关于 x 和 y 右连续;

(4) 任意给定 $a < b$,

$$P(a < X \leqslant b, c < Y \leqslant d) = F_{X,Y}(b,d) + F_{X,Y}(a,c) - F_{X,Y}(a,d) - F_{X,Y}(b,c).$$

称 F_X 和 F_Y 是 (X,Y) 的**边际分布函数**. 一般地, 联合分布函数 $F_{X,Y}$ 唯一确定边际分布函数, 即

$$F_X(x) = F_{X,Y}(x,\infty), \quad F_Y(y) = F_{X,Y}(\infty, y).$$

但边际分布函数并不能确定联合分布函数.

给定 $X = x$ 的条件下, Y 的**条件分布函数**为

$$F_{Y|X}(Y \leqslant y | X = x) = \lim_{\varepsilon \to 0} \frac{F_{X,Y}(x+\varepsilon, y) - F_{X,Y}(x,y)}{F_X(x+\varepsilon) - F_X(x)};$$

给定 $Y = y$ 的条件下, X 的**条件分布函数**为

$$F_{X|Y}(X \leqslant x | Y = y) = \lim_{\varepsilon \to 0} \frac{F_{X,Y}(x, y+\varepsilon) - F_{X,Y}(x,y)}{F_Y(y+\varepsilon) - F_Y(y)}.$$

如果对任意 $x, y \in \mathbb{R}$,

$$F_{X,Y}(x,y) = F_X(x) F_Y(y),$$

则称 X 和 Y **相互独立**. 这等价于对任意 $A, B \in \mathcal{B}$,

$$P(\omega : X \in A, Y \in B) = P(\omega : X \in A) P(\omega : Y \in B).$$

由此可知, 当 X 和 Y 相互独立时, 对任意 Borel 可测函数 $f, g : \mathbb{R} \mapsto \mathbb{R}$, $f(X)$ 和 $g(Y)$ 相互独立.

假设离散随机向量 (X,Y) 分布如下:

		Y					
		y_1	y_2	$\cdots$	y_j	$\cdots$	
X	x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
	x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
		$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	

那么 X 的**边际分布列**为

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_i & \cdots \\ p_{1\cdot} & p_{2\cdot} & \cdots & p_{i\cdot} & \cdots \end{pmatrix},$$

其中

$$p_{i\cdot} = \sum_{j=1}^{\infty} p_{ij}, \quad i = 1, 2, \cdots;$$

Y 的**边际分布列**为

$$Y \sim \begin{pmatrix} y_1 & y_2 & \cdots & y_j & \cdots \\ p_{\cdot 1} & p_{\cdot 2} & \cdots & p_{\cdot j} & \cdots \end{pmatrix},$$

其中

$$p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij}, \quad j = 1, 2, \cdots.$$

给定 $X = x_i$ 的条件下, Y 的**条件分布列**为

$$P_{Y|X}(y_j|x_i) = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1, 2, \cdots;$$

给定 $Y = y_j$ 的条件下, X 的**条件分布列**为

$$P_{X|Y}(x_i|y_j) = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \cdots.$$

X 和 Y **相互独立**当且仅当

$$p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i, j = 1, 2, \cdots.$$

假设连续随机向量 (X, Y) 的联合密度函数为 $p_{X,Y}(x, y)$, 即

$$F_{X,Y}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p_{X,Y}(u, v)\mathrm{d}u\mathrm{d}v, \quad x, y \in \mathbb{R},$$

那么 X 的**边际密度函数**为

$$p_X(x) = \int_{-\infty}^{\infty} p(x, y)\mathrm{d}y, \quad x \in \mathbb{R};$$

Y 的**边际密度函数**为

$$p_Y(y) = \int_{-\infty}^{\infty} p(x, y)\mathrm{d}x, \quad y \in \mathbb{R}.$$

给定 $X=x$ 的条件下, Y 的**条件密度函数**为

$$p_{Y|X}(y|x)=\frac{p_{X,Y}(x,y)}{p_X(x)},\quad y\in\mathbb{R};$$

给定 $Y=y$ 的条件下, X 的**条件密度函数**为

$$p_{X|Y}(x|y)=\frac{p_{X,Y}(x,y)}{p_Y(y)},\quad x\in\mathbb{R}.$$

X 和 Y **相互独立**当且仅当

$$p_{X,Y}(x,y)=p_X(x)p_Y(y),\quad x,y\in\mathbb{R}.$$

假设 $f:\mathbb{R}^2\mapsto\mathbb{R}$ 是二元 Borel 可测函数, 那么 $Z=f(X,Y)$ 是随机变量. 但计算 Z 的分布函数并没有普适的方法, 通常依赖于 f 的定义. 事实上, 给定 $z\in\mathbb{R}$, 记 $A_z=\{(x,y):f(x,y)\leqslant z\}$, 这是 Borel 可测集, 那么

$$\begin{aligned}F_Z(z)&=P(\omega:Z\leqslant z)\\&=P(\omega:f(X,Y)\leqslant z)\\&=P(\omega:(X,Y)\in A_z).\end{aligned}$$

之后的计算取决于 (X,Y) 的联合分布和 A_z 的形状.

下列定理表明 Bernoulli 分布、Poisson 分布和正态分布分别具有可加性.

定理 1.2 (i) 假设 $X\sim B(n,p)$, $Y\sim B(m,p)$, 并且 X 和 Y 相互独立, 那么 $X+Y\sim B(n+m,p)$;

(ii) 假设 $X\sim\mathcal{P}(\lambda)$, $Y\sim\mathcal{P}(\mu)$, 并且 X 和 Y 相互独立, 那么 $X+Y\sim\mathcal{P}(\lambda+\mu)$;

(iii) 假设 $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$, 并且 X 和 Y 相互独立, 那么 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

二元联合正态分布具有如下基本性质.

定理 1.3 假设 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$, 那么

(i) $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$;

(ii) 给定 $X=x$ 的条件下, Y 的条件分布为 $N(\mu_2+\rho\sigma_2(x-\mu_1)/\sigma_1,(1-\rho^2)\sigma_2^2)$; 给定 $Y=y$ 的条件下, X 的条件分布为 $N(\mu_1+\rho\sigma_1(y-\mu_2)/\sigma_2,(1-\rho^2)\sigma_1^2)$;

(iii) X 和 Y 相互独立当且仅当 $\rho=0$;

(iv) 任意给定 $a,b\in\mathbb{R}$, $aX+bY\sim N(\mu,\sigma^2)$, 其中

$$\mu=a\mu_1+b\mu_2,\quad \sigma^2=a^2\sigma_1^2+b^2\sigma_2^2+2ab\rho\sigma_1\sigma_2.$$

三、多维随机向量

假设 $(X_1,\cdots,X_m)$ 是 m–维随机向量, 可以类似定义联合分布函数、边际分布函数、条件分布函数和独立性等. 特别, $(X_1,\cdots,X_m)$ 相互独立当且仅当

$$F_{X_1,\cdots,X_m}(x_1,\cdots,x_m)=\prod_{i=1}^{m}F_{X_i}(x_i),\quad x_1,\cdots,x_m\in\mathbb{R}.$$

定理 1.4 假设 $(X_1,\cdots,X_m)$ 相互独立, $f_i:\mathbb{R}^{k_i}\mapsto\mathbb{R}$ 是可测函数, $i=1,\cdots,m$, 其中 $k_1+\cdots+k_l\leqslant m$. 令 $I_i\subset\{1,2,\cdots,m\}$ 为互不相交的指标集, 其中 $|I_i|=k_i$, $i=1,\cdots,l$. 定义

$$Y_i=f_i(X_j,j\in I_i),\quad i=1,\cdots,l,$$

那么 $Y_1,\cdots,Y_l$ 相互独立.

多维正态随机向量最为常见. 假设 $(X_1,\cdots,X_m)$ 是 m–维正态随机向量, 联合密度函数为

$$p_{X_1,\cdots,X_m}(x_1,\cdots,x_m)=\frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}},$$

其中 $\boldsymbol{x}=(x_1,\cdots,x_m)$, $\boldsymbol{\mu}=(\mu_1,\cdots,\mu_m)$, $\boldsymbol{\Sigma}$ 为 $m\times m$ 对称可逆矩阵.

定理 1.5 $(X_1,\cdots,X_m)$ 是 m–维正态随机向量当且仅当任意线性组合为正态随机变量. 特别, $(X_1,\cdots,X_m)\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 当且仅当对任意一组实数 $a_1,a_2,\cdots,a_m$,

$$\sum_{i=1}^{m}a_iX_i\sim N(\boldsymbol{a}\boldsymbol{\mu}^{\mathrm{T}},\boldsymbol{a}\boldsymbol{\Sigma}\boldsymbol{a}^{\mathrm{T}}),$$

其中 $\boldsymbol{a}=(a_1,\cdots,a_m)$.

§1.3 数字特征

一、数学期望

令 F_X 是随机变量 X 的分布函数, 如果

$$\int_{-\infty}^{\infty}|x|\mathrm{d}F_X(x)<\infty,\tag{1.5}$$

那么 X 的数学期望存在并且有限, 其值为

$$EX=\int_{-\infty}^{\infty}x\mathrm{d}F_X(x)<\infty,\tag{1.6}$$

其中 (1.5) 和 (1.6) 为 Riemann-Stieltjes 积分 (见第七章附录).

特别, 当 X 是离散随机变量时 (见 (1.4)), (1.6) 式可写成

$$EX = \sum_{i=1}^{N} x_i p_i;$$

当 X 是连续随机变量, 密度函数为 p_X 时, (1.6) 式可写成

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

简单地说, 随机变量的数学期望是随机变量的取值按概率大小进行加权平均, 它是随机变量最重要的数字特征. 典型随机变量的数学期望都与分布参数有直接关系. 下列定理给出了数学期望的基本性质.

定理 1.6 假设以下所涉及的数学期望存在且有限.

(i) 如果 $a \leqslant X \leqslant b$ a.s., 那么 $a \leqslant EX \leqslant b$;

(ii) 假设 a, b 是常数, 那么

$$E(aX + b) = aEX + b;$$

(iii) 对任意随机变量 X, Y,

$$E(X + Y) = EX + EY;$$

(iv) 令 $f : \mathbb{R} \mapsto \mathbb{R}$ 是 Borel 可测函数, 那么

$$Ef(X) = \int_{-\infty}^{\infty} f(x) \mathrm{d}F_X(x).$$

令 $k \geqslant 1$, 如果 $E|X|^k < \infty$, 则称 EX^k 为 X 的**第 k 阶矩**, 记作 $m_k = EX^k$.

例 1.1 (1) 令 $X \sim \mathcal{P}(\lambda)$, 那么

$$E[X(X-1)\cdots(X-k+1)] = \lambda^k, \quad k \geqslant 1. \tag{1.7}$$

由 (1.7) 可以递推得到 m_k. 特别, $m_1 = \lambda$, $m_2 = \lambda^2 + \lambda$.

(2) 令 $X \sim N(0, \sigma^2)$, 那么

$$m_{2k} = (2k-1)!!\sigma^{2k}, \quad m_{2k-1} = 0, \quad k \geqslant 1.$$

注 1.4 一般来说, 随机变量的矩并不能唯一确定其分布函数.