“十三五”移动学习型规划教材

随机过程

李媛 蔡立刚 刘芳 丁宁 编著

STOCHASTIC PROCESS

机械工业出版社
CHINA MACHINE PRESS

向教师提供PPT课件

"十三五"移动学习型规划教材

随机过程

李 媛 蔡立刚 刘 芳 丁 宁 编著

机械工业出版社

本书为理工科研究生、理科高年级本科生与经济类和管理类本科生随机过程课程教材. 全书共分六章，主要内容包括预备知识、基本概念、随机分析、马尔可夫过程、平稳过程和鞅. 书中附有大量实例，加强了随机过程的应用性. 本书还通过移动互联网，为读者提供课外阅读资料. 读者可以通过专用 APP，获得与课程内容相关的人物介绍、概念补充和理论讲解等内容.

书中概念的描述和理论的推导比较详细，便于读者自学，也可作为科技、工程人员的参考书.

图书在版编目(CIP)数据

随机过程/李媛等编著. —北京：机械工业出版社，2018. 12
"十三五"移动学习型规划教材
ISBN 978-7-111-61595-8

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

中国版本图书馆 CIP 数据核字(2018)第 294505 号

机械工业出版社(北京市百万庄大街 22 号　邮政编码 100037)
策划编辑：韩效杰　责任编辑：韩效杰　李　乐
责任校对：张　薇　封面设计：鞠　杨
责任印制：张　博
北京铭成印刷有限公司印刷
2019 年 2 月第 1 版第 1 次印刷
184mm×260mm • 10. 25 印张 • 261 千字
标准书号：ISBN 978-7-111-61595-8
定价：29. 80 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务
服务咨询热线：010-88379833
读者购书热线：010-88379649

网络服务
机 工 官 网：www. cmpbook. com
机 工 官 博：weibo. com/cmp1952
教育服务网：www. cmpedu. com
金 书 网：www. golden-book. com

前　言

随着20世纪社会经济的飞速发展，随机过程作为一门新生的数学学科逐步发展起来. 它内容丰富，应用非常广泛. 随机过程研究的对象与初等概率论一样，也是随机现象的统计规律性. 初等概率论研究的是随机现象的静态特性，而随机过程研究的是随机现象的动态特性，即随机现象的发展与变化过程. 随机过程是初等概率论的重要分支，初等概率论是随机过程的理论基础.

本书为现代随机过程理论的入门教材，可作为高年级本科生及研究生的必修课教材，也可作为教师和科研人员的参考书. 本书共分六章，主要内容包括预备知识、基本概念、随机分析、马尔可夫过程、平稳过程和鞅. 第一章预备知识部分介绍了概率论的基本概念，主要包含概率空间与随机变量、随机变量的数字特征与特征函数、多维正态随机向量和条件数学期望等基本理论. 第二章基本概念部分介绍了随机过程的基本概念、分布、数字特征和多维随机过程、复随机过程等基本理论. 第三章随机分析部分介绍了随机过程在均方意义下的极限、连续、导数和积分问题. 第四章马尔可夫过程部分主要介绍了马尔可夫过程的定义、转移概率及其关系、转移概率的极限性态，并着重讨论了马尔可夫链以及时间连续状态离散的马尔可夫过程. 第五章平稳过程部分主要介绍了平稳过程的基本概念、相关函数、均方遍历性、谱密度和谱分解. 第六章鞅部分主要介绍了鞅的基本概念、鞅的构造方法及分解定理，并引入停时，给出停时定理和鞅收敛定理及其应用.

本书的主要特点：

1. 对读者所需的数学基础要求起点较低，读者只需具备概率论、微积分和线性代数的基本知识.

2. 书中概念的描述和理论的推导比较详细，便于自学.

3. 强调实际应用. 本书配有一些经济、信息、社会、管理和生物科学等领域的相关例题与习题，可帮助学生加深理解，提高应用随机过程解决实际问题的能力.

4. 通过移动互联网，为读者提供课外阅读资料. 读者可以通过专用 APP，获得与课程内容相关的人物介绍、概念补充和理论讲解等内容.

5. 本书着重介绍随机过程的基本概念和相关基本过程，力求用简洁的语言对知识进行准确的描述，适合随机过程短学时教学.

6. 本书的编写者都是长期从事随机过程教学的一线教师，能够积极地吸收国内外最新的优秀教学成果，更能直接地了解学生需求，为随机过程教材的编写提供了教研与教学经验上的保障.

本书由李媛、刘芳统稿，各章执笔的分别是李媛（第四章），刘芳（第一章、第三章、第五

章)，蔡立刚(第二章)，丁宁(第六章). 本书在撰写与出版过程中得到了沈阳工业大学和沈阳理工大学理学院、研究生院、教务处领导和部分教师的大力支持和帮助，在此我们一并表示衷心的感谢. 同时我们也要对审阅本书的专家表示感谢，谢谢你们提出的宝贵意见. 我们希望本书的出版能够为我国随机过程的教学和发展起到积极的作用.

由于编者水平有限，书中错误和不足之处在所难免，恳请读者批评指正. 可通过以下邮箱与作者联系：

syliyuan@sut. edu. cn

liufang5208@sylu. edu. cn

编　者

目　录

第一章 预备知识

概率论中的基本概念和基本理论是学习随机过程的基础.本章主要内容包括概率空间、随机变量及其概率分布、特征函数、多维正态随机变量以及条件数学期望等,读者可以按所学专业选学,为学习随机过程做准备.

第一节 概率空间与随机变量

概率论中的一个基本概念是**随机试验**,随机试验的结果不能预先确定.一个试验所有可能结果的集合称为此试验的**样本空间**,记为Ω.

样本空间中的每个元素称为**样本点**,**事件**是样本空间的一个子集,若某试验的结果是这个子集的一个元素,则称这个事件发生了.特别指出,单个样本点所组成的集合称为**基本事件**,样本空间Ω称为**必然事件**,而空集$\varnothing$称为不可能事件.

定义 1 设样本空间Ω的某些子集构成的集合记为$\mathcal{F}$,若$\mathcal{F}$满足下列性质:

(1) $\Omega\in\mathcal{F}$;

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

(3) 若$A_k\in\mathcal{F}(k=1,2,\cdots)$,则$\bigcup\limits_{k=1}^{\infty}A_k\in\mathcal{F}$,

则称$\mathcal{F}$是一个σ-**域**(或称波莱尔(Borel)域).

我们把任一样本空间Ω与由Ω的子集所组成的一个σ-域$\mathcal{F}$写在一起,记为$(\Omega,\mathcal{F})$,称为具有σ-域结构的样本空间,或简称为可测空间.

定义 2 设A是样本空间Ω的事件,$\mathcal{F}$是一个σ-**域**,定义在$\mathcal{F}$上的实值函数$P(A)$如果满足:

(1) $0\leqslant P(A)\leqslant 1$;

(2) $P(\Omega)=1$;

(3) 若$A_k\in\mathcal{F}(k=1,2,\cdots)$,且$A_iA_j=\varnothing,i\neq j,i,j=1,2,\cdots$,则

$$P\Big(\bigcup_{k=1}^{\infty}A_k\Big)=\sum_{k=1}^{\infty}P(A_k),$$

我们称 $P(A)$ **为事件** A **的概率**.

设 Ω 是一个样本空间，$\mathcal{F}$ 是 Ω 中的 σ-**域**，P 为 $\mathcal{F}$ 上的概率. 我们称具有这样结构的样本空间为**概率空间**，记为 $(\Omega,\mathcal{F},P)$.

概率空间 $(\Omega,\mathcal{F},P)$ 上的概率 P 有如下性质：

性质 1 $P(\varnothing)=0$.

性质 2 对任一事件 A，有 $P(\overline{A})=1-P(A)$.

性质 3 若 $A\supset B$，则 $P(A-B)=P(A)-P(B)$，且 $P(A)\geqslant P(B)$.

性质 4 若 $A_k\in\mathcal{F}(k=1,2,\cdots,n)$，且 $A_iA_j=\varnothing,i\neq j,i,j=1,2,\cdots,n$，则

$$P\left(\bigcup_{k=1}^{n}A_k\right)=\sum_{k=1}^{n}P(A_k).$$

几个基本公式：

(1) 条件概率公式：$P(A|B)=\dfrac{P(AB)}{P(B)},P(B)>0$.

(2) 乘法公式：$P(AB)=P(A|B)P(B),P(B)>0$.

$P(A_1A_2\cdots A_n)=P(A_n|A_{n-1}\cdots A_1)P(A_{n-1}|A_{n-2}\cdots A_1)\cdots P(A_2|A_1)P(A_1)$（推广）.

(3) 全概率公式：$\bigcup\limits_{i=1}^{n}B_i=\Omega,P(\Omega)=1$,

$$B_iB_j=\varnothing,i\neq j,i,j=1,2,\cdots,n,$$

$$P(A)=\sum_{i=1}^{n}P(A|B_i)P(B_i).$$

(4) 贝叶斯公式：$\bigcup\limits_{i=1}^{n}B_i=\Omega,P(\Omega)=1$,

$$B_iB_j=\varnothing,i\neq j,i,j=1,2,\cdots,n,$$

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n}P(A|B_j)P(B_j)}.$$

(5) 事件的独立性：n 个事件 $A_1,A_2,\cdots,A_n$ 独立 $\Leftrightarrow$

$$P(A_{i_1}A_{i_2}\cdots A_{i_s})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_s}),$$

$$1\leqslant i_1\leqslant\cdots\leqslant i_s\leqslant n,\forall s\in\mathbf{Z}^+,2\leqslant s\leqslant n.$$

在随机试验中，若存在一个变量，它依据试验出现的结果而取不同的数值，则称此变量为随机变量. 由于随机试验出现的结果带有随机性，因而随机变量的取值也带有随机性. 从数学角度看，样本空间 Ω 中每一个样本点 ω 对应一个数 $X(\omega)$，这就是随机变量. 或者说随机变量是定义在样本空间上的函数，但是这个函数需要满足一些要求.

定义 3 设 $(\Omega,\mathcal{F},P)$ 为概率空间，对于 $\omega\in\Omega$，$X(\omega)$ 是一个取实值的单值函数；若对于任一实数 x，$\{\omega:X(\omega)<x\}$ 是一随机事件，即 $\{\omega:X(\omega)<x\}\in\mathcal{F}$，则称 $X(\omega)$ 为随机变量.

从定义看出，随机变量 $X(\omega)$ 总是与一个概率空间相联系的，相当于把概率空间中每个元素映射到数轴上，以建立起样本空间与数集的对应关系. 为书写简便，一般不必每次都写出概率空间 $(\Omega,\mathcal{F},P)$，并且可以将 $X(\omega)$ 关于 ω 的依赖性省略，简记为 X，把 $\{\omega:X(\omega)<x\}$ 记为 $\{X<x\}$ 等.

既然对任意一个实数 x，有 $\{\omega:X(\omega)<x\}\in\mathcal{F}$，那么对 Ω 的子集 $\{\omega:X(\omega)<x\}$ 就可以确定概率.

定义 4 设 $(\Omega,\mathcal{F},P)$ 是概率空间，而 $X=X(\omega)$ 是 $(\Omega,\mathcal{F},P)$ 上的随机变量. 若对任意一个实数 x，有概率

$$F(x)=P\{\omega:X(\omega)\leqslant x\}$$

或简写为

$$F(x)=P\{X\leqslant x\},$$

则称 $F(x)$ 是随机变量 X 的分布函数.

随机变量概念的产生是概率论的重大进展，它使概率论研究的对象由简单的事件扩大到更多范畴，并将微积分理论成功地引入到概率论中. 根据随机变量取值的特点可以分为离散型随机变量和连续型随机变量：若随机变量 X 的一切可能取值是有限个或无限可数个，则称 X 为离散型随机变量；若随机变量 X 的一切可能取值充满一个有限或无限区间，则称 X 为连续型随机变量.

若存在有限个或可列个数 $x_1,x_2,\cdots$，使随机变量 X 有

$$P\{X=x_k\}=p_k,k=1,2,\cdots,\text{且满足}\sum_{k=1}^{\infty}p_k=1,p_k\geqslant 0,$$

则称其为离散型随机变量 X 的分布律.

若对任意实数 x，存在非负实函数 $f(x)$，使随机变量 X 的分布函数 $F(x)$ 有

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

则称 $f(x)$ 为连续型随机变量 X 的概率密度函数，简称概率密度.

下面举出一些常见的离散型和连续型随机变量的例子.

1. 离散型

(1) 0—1 分布

设随机变量 X 只可能取 0 和 1 两个值，其分布律为

$$P\{X=1\}=p,P\{X=0\}=1-p,0<p<1,$$

则称 X 服从 0—1 分布.

(2) 二项分布

设试验只有两个可能结果 A 与 $\overline{A}$，且 $P(A)=p$，$P(\overline{A})=1-p=q$，将试验独立重复 n 次，则称该试验为 n 重伯努利试验. 在 n 重伯努利试验中，事件 A 发生 m 次的概率记为

$$\mathrm{C}_n^m p^m q^{n-m},0\leqslant m\leqslant n.$$

令 X 表示 n 重伯努利试验中 A 发生的次数，则 X 为一个随机

变量，它的可能值为 $0,1,2,\cdots,n$，则 X 的分布律为

$$P\{X=m\}=C_n^m p^m q^{n-m},m=0,1,2,\cdots,n,$$

显然

$$\sum_{m=0}^{n} C_n^m p^m q^{n-m}=(p+q)^n=1,$$

则称 X 服从参数为 n,p 的二项分布，记为 $X\sim b(n,p)$.

(3) 泊松分布

设随机变量 X 的可能取值为 $0,1,2,\cdots$，而取各个值的概率为

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!},k=0,1,2,\cdots,$$

其中，$\lambda>0$ 为常数，则称 X 服从参数为 λ 的泊松分布，记为 $X\sim\pi(\lambda)$，且有

$$\sum_{k=0}^{\infty} P\{X=k\}=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1.$$

2. 连续型

(1) 均匀分布

设连续型随机变量 X 在有限区间 $[a,b]$ 内取值，其概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b,\\ 0, & \text{其他},\end{cases}$$

则称 X 在区间 $[a,b]$ 上服从均匀分布，记为 $X\sim U(a,b)$.

(2) 指数分布

设连续型随机变量 X 的概率密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0,\\ 0, & \text{其他},\end{cases}$$

其中，$\lambda>0$ 为常数，则称 X 服从参数为 λ 的指数分布.

(3) 正态分布

设连续型随机变量 X 的概率密度函数为

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

其中 $\mu,\sigma(\sigma>0)$ 为常数，则称 X 服从参数为 μ,σ 的正态分布或高斯(Gauss)分布，记为 $X\sim N(\mu,\sigma^2)$. 特别地，称参数 $\mu=0,\sigma=1$ 的正态分布 $N(0,1)$ 为标准正态分布.

定义 5 设 $(\Omega,\mathcal{F},P)$ 为一概率空间，定义在 Ω 上的 n 个随机变量 $X_1,X_2,\cdots,X_n$，记为向量形式 $(X_1,X_2,\cdots,X_n)$，称为 n 维随机向量或 n 维随机变量.

n 维随机向量的分布函数可定义为

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

同样，n 维随机向量也可分为离散型和连续型进行研究.

第二节 随机变量的数字特征与特征函数

随机变量的分布函数是随机变量概率分布的完整描述,但是要找到随机变量的分布函数有时不是一件容易的事.另一方面,在实际问题中描述随机变量的概率特征,不一定都需要求分布函数,往往只要求出描述随机变量概率特征的几个表征值就够了,这就需要引入随机变量的数字特征和特征函数.

定义 6 设 $f(x)$,$g(x)$是定义在$[a,b]$上的两个有界函数,$a=x_1<x_2<\cdots<x_n=b$ 是区间$[a,b]$的任一分割,$\Delta x_k=x_k-x_{k-1}$,$\lambda=\max\limits_{1\leqslant k\leqslant n}\Delta x_k$,在每一个子区间$[x_{k-1},x_k]$上任意取一点 ξ_k 作和式

$$S=\sum_{k=1}^{n}f(\xi_k)[g(x_k)-g(x_{k-1})].$$

若极限

$$\lim_{\lambda\to 0}S=\lim_{\lambda\to 0}\sum_{k=1}^{n}f(\xi_k)[g(x_k)-g(x_{k-1})]$$

存在且与$[a,b]$的分法和 ξ_k 的取法都无关,则称此极限为函数 $f(x)$ 对函数 $g(x)$在区间$[a,b]$上的斯蒂尔切斯(Stieltjes)积分,简称 S 积分,记为$\int_a^b f(x)\mathrm{d}g(x)$. 此时也称函数 $f(x)$ 对函数 $g(x)$ 在区间$[a,b]$上 S 可积.

定义 7 设 $f(x)$,$g(x)$是定义在$(-\infty,+\infty)$上的两个函数,若在任意有限区间$[a,b]$上,函数 $f(x)$对函数 $g(x)$在区间$[a,b]$上 S 可积,且极限

$$\lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_a^b f(x)\mathrm{d}g(x)$$

存在,则称此极限为函数 $f(x)$对函数 $g(x)$在无穷区间$(-\infty,+\infty)$上的斯蒂尔切斯积分,简称 S 积分,记为$\int_{-\infty}^{+\infty}f(x)\mathrm{d}g(x)$.

在 S 积分中,当 $g(x)$取一些特殊形式时,积分可化为级数或通常积分.

若 $g(x)$是在$(-\infty,+\infty)$上的阶梯函数,它的跳跃点为 x_1,x_2,…(有限或可列无限个),则

$$\int_{-\infty}^{+\infty}f(x)\mathrm{d}g(x)=\sum f(x_k)[g(x_k+0)-g(x_k-0)];$$

若 $g(x)$ 是在$(-\infty,+\infty)$上的可微函数,它的导函数为 $g'(x)$,则

$$\int_{-\infty}^{+\infty}f(x)\mathrm{d}g(x)=\int_{-\infty}^{+\infty}f(x)g'(x)\mathrm{d}x.$$

显然,上两种情况一个是把 S 积分化成和式,另一个是把 S 积分化成黎曼积分.

定义 8 设函数 $g(x)$ 定义在无限区间 $(-\infty,+\infty)$ 上，若积分

$$\int_{-\infty}^{+\infty} e^{itx} dg(x) = \int_{-\infty}^{+\infty} \cos tx \, dg(x) + i\int_{-\infty}^{+\infty} \sin tx \, dg(x)$$

存在，则称此积分为对 $g(x)$ 的傅里叶—斯蒂尔切斯(Fourier-Stieltjes)积分. 简称 F-S 积分

根据上面几个积分定义形式，给出随机变量的数字特征和特征函数的定义.

定义 9 设 X 是一个随机变量，$F(x)$ 是其分布函数，若 $\int_{-\infty}^{+\infty} |x| dF(x)$ 存在，则称

$$E(X) = \int_{-\infty}^{+\infty} x dF(x)$$

为随机变量 X 的数学期望或均值.

若 X 是离散型随机变量，其分布律为 $P\{X=x_k\}=p_k$，$k=1,2,\cdots$，则

$$E(X) = \sum_{k=1}^{\infty} x_k p_k;$$

若 X 是连续型随机变量，其概率密度函数为 $f(x)$，则

$$E(X) = \int_{-\infty}^{+\infty} xf(x) dx.$$

由于随机变量的函数仍为随机变量，因此随机变量的函数也可求数学期望，即设 X 是一随机变量，其分布函数为 $F(x)$，$y=h(x)$ 是连续函数，若 $\int_{-\infty}^{+\infty} h(x) dF(x)$ 存在，则

$$E(Y) = E[h(X)] = \int_{-\infty}^{+\infty} h(x) dF(x),$$

其中 $Y=h(X)$ 为随机变量 X 的函数.

以上结论在概率论中是以定理形式给出的，可以按离散型和连续型分别加以证明(略).

同理可推广到多元随机变量函数的数学期望：设 $(X_1,X_2,\cdots,X_n)$ 是 n 维随机变量，其联合分布函数为 $F(x_1,x_2,\cdots,x_n)$，$g(x_1,x_2,\cdots,x_n)$ 是连续函数，如果积分

$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x_1,x_2,\cdots,x_n) dF(x_1,x_2,\cdots,x_n)$ 存在，则

$$E[g(X_1,X_2,\cdots,X_n)]$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x_1,x_2,\cdots,x_n) dF(x_1,x_2,\cdots,x_n).$$

定义 10 设 X 是一个随机变量，若 $E\{[X-E(X)]^2\}$ 存在，则称

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

为随机变量 X 的方差. 显然方差是非负的，在应用上还引入量 $\sqrt{D(X)}$，称为标准差或均方差.

定义 11　量 $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 X 与 Y 的**协方差**，记为 $\mathrm{Cov}(X,Y)$，即

$$\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\},$$

而

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量 X 与 Y 的**相关系数**. 若 $\rho_{XY}=0$，则称 X 与 Y **不相关**.

随机变量的数学期望和方差具有下列性质：

设 a,b 是常数，随机变量 X 与 Y 的期望与方差都存在，则

(1) $E(a)=a,D(a)=0$.

(2) $E(aX+bY)=aE(X)+bE(Y),D(aX+bY)=a^2D(X)+b^2D(Y)+2ab\mathrm{Cov}(X,Y)$.

特别地，若 X 与 Y 相互独立，则有 $D(X+Y)=D(X)+D(Y)$，即 $\mathrm{Cov}(X,Y)=0$.

(3) 若 X 与 Y 相互独立，则 $E(XY)=E(X)E(Y)$.

(4) 若 $E(X^2)$ 与 $E(Y^2)$ 都存在，则 $[E(XY)]^2\leqslant E(X^2)E(Y^2)$. 此不等式称为**施瓦茨(Schwarz)不等式.**

(5) 若 $E(X^r),r>0$ 存在，则 $\forall\varepsilon>0,P\{|X|\geqslant\varepsilon\}\leqslant\frac{E|X|^r}{\varepsilon^r}$. 此不等式称为**马尔可夫(Markov)不等式.** 特别地，在马尔可夫不等式中，令 $r=2$，将 X 换成 $X-E(X)$，可得重要的**切比雪夫(Chebyshev)不等式**：$P\{|X-E(X)|\geqslant\varepsilon\}\leqslant\frac{D(X)}{\varepsilon^2}$.

定义 12　设 X 和 Y 是随机变量，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**.

若
$$E\{[X-E(X)]^k\},k=2,3,\cdots$$

存在，称它为 X 的 k **阶中心矩**.

若
$$E(X^kY^l),k,l=1,2,\cdots$$

存在，称它为 X 和 Y 的 $k+l$ **阶混合矩**.

若
$$E\{[X-E(X)]^k[Y-E(Y)]^l\},k,l=1,2,\cdots$$

存在，称它为 X 和 Y 的 $k+l$ **阶混合中心矩**.

显然，X 的数学期望 $E(X)$ 是 X 的一阶原点矩，方差 $D(X)$ 是 X 的二阶中心矩，协方差 $\mathrm{Cov}(X,Y)$ 是 X 和 Y 的二阶混合中心矩.

数字特征一般由各阶矩所决定，但随着阶数的增高，求矩的计算会愈加麻烦. 另一方面，随机现象是错综复杂的，往往需要多个随机变量，甚至要由随机变量序列依某种收敛意义的迫近来描述. 要解决复杂得多的问题，就需要有优越的数学工具，而傅里叶变换是数学中非常重要而有效的工具，把它应用于分布函数或密度函数，就产生了“特征函数”.

先引进复随机变量的概念.

定义 13 设 X,Y 都是定义在概率空间 $(\Omega,\mathcal{F},P)$ 上的实值随机变量,则称 $Z=X+\mathrm{i}Y$ 为**复随机变量**,其中 $\mathrm{i}=\sqrt{-1}$,它的数学期望定义为 $E(Z)=E(X)+\mathrm{i}E(Y)$.

定义 14 设 X 是定义在概率空间 $(\Omega,\mathcal{F},P)$ 上的随机变量,它的分布函数为 $F(X)$,称 $\mathrm{e}^{\mathrm{i}tX}$ 的数学期望 $E(\mathrm{e}^{\mathrm{i}tX})$ 为 X 的特征函数,有时也称为分布函数 $F(X)$ 的**特征函数**,其中 $\mathrm{i}=\sqrt{-1}$,$t\in\mathbf{R}$. 记 X 的特征函数为 $\varphi_X(t)$,在不引起混乱的情况下简写为 $\varphi(t)$.

由于

$$\mathrm{e}^{\mathrm{i}tX}=\cos tX+\mathrm{i}\sin tX,$$

故

$$\begin{aligned}\varphi(t) &= E(\mathrm{e}^{\mathrm{i}tX}) = E(\cos tX)+\mathrm{i}E(\sin tX)\\ &= \int_{-\infty}^{+\infty}\cos tx\,\mathrm{d}F(x)+\mathrm{i}\int_{-\infty}^{+\infty}\sin tx\,\mathrm{d}F(x)\\ &= \int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}F(x).\end{aligned}$$

因此,X 的特征函数也可称为对分布函数 $F(x)$ 的傅里叶—斯蒂尔切斯变换,简称 F-S 积分.

随机变量 X 的特征函数 $\varphi(t)$ 的性质:

(1) $|\varphi(t)|\leqslant\varphi(0)=1$.

(2) $\varphi(-t)=\overline{\varphi(t)}$($\overline{\varphi(t)}$表示 $\varphi(t)$ 的共轭复数).

(3) 设随机变量 $Y=aX+b$,其中 a,b 是常数,则 Y 的特征函数为

$$\varphi_Y(t)=\mathrm{e}^{\mathrm{i}bt}\varphi(at).$$

(4) $\varphi(t)$ 在 $(-\infty,+\infty)$ 上一致连续.

(5) 设随机变量 X,Y 相互独立,又 $Z=X+Y$,则

$$\varphi_Z(t)=\varphi_X(t)\varphi_Y(t).$$

(6) 设随机变量 X 的 n 阶原点矩存在,则它的特征函数 $\varphi(t)$ 的 n 阶导数 $\varphi^{(n)}(t)$ 存在,且有

$$E(X^k)=\mathrm{i}^{-k}\varphi^{(k)}(0),k\leqslant n.$$

(7) 随机变量 X 的特征函数 $\varphi(t)$ 是非负定的,即对任意正整数 n,任意复数 $z_1,z_2,\cdots,z_n$ 以及 $t_r,t_s\in\mathbf{R}$,$r,s=1,2,\cdots,n$,有

$$\sum_{r,s=1}^{n}\varphi(t_r-t_s)z_r\overline{z_s}\geqslant 0.$$

由上性质中(1)(4)(7)的结论逆推,即有**波赫纳-辛钦定理**:若函数 $\varphi(t)$($t\in\mathbf{R}$)连续、非负定且 $\varphi(0)=1$,则 $\varphi(t)$ 必为特征函数.

对于给定的一个分布函数 $F(x)$,可唯一确定它的特征函数 $\varphi(t)$,反之,分布函数 $F(x)$ 也能唯一地被其特征函数 $\varphi(t)$ 表达出来,即**反演公式**:设随机变量 X 的分布函数和特征函数分别为 $F(x)$ 和 $\varphi(t)$,则对于 $F(x)$ 的任意连续点 x_1 和 x_2($-\infty<x_1<x_2<$

$+\infty)$，有

$$F(x_2)-F(x_1)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}tx_1}-\mathrm{e}^{-\mathrm{i}tx_2}}{\mathrm{i}t}\varphi(t)\mathrm{d}t,$$

并由唯一性定理可得，随机变量 X 的分布函数 $F(x)$ 和特征函数 $\varphi(t)$ 之间是互相唯一确定的.

在研究只取有穷或无穷非负整数值的随机变量时，用母函数来代替特征函数较为方便.

定义 15 设随机变量 X 的分布律为

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

记实变数 s 的实函数

$$\Psi_X(s)=E(s^X)=\sum_k p_k s^k(-1\leqslant s\leqslant 1),$$

称 $\Psi_X(s)$ 为 X 的母函数，简记为 $\Psi(s)$.

母函数与特征函数也有相应类似的性质.

第三节　多维正态随机向量

正态分布在概率论中扮演极为重要的角色. 一方面，由中心极限定理知实际中许多随机变量服从或近似地服从正态分布；另一方面，正态分布具有良好的分析性质. 下面我们讨论多维正态随机向量.

首先给出 n 维随机向量的一些数字特征.

设随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的分布函数为

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

其数学期望定义为

$$E(\boldsymbol{X})=(EX_1,EX_2,\cdots,EX_n)^{\mathrm{T}}.$$

随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的协方差阵定义为

$$\boldsymbol{B}=\begin{pmatrix}\mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Cov}(X_n,X_n)\end{pmatrix},$$

显然，协方差阵是非负定的.

下面介绍多维正态随机向量.

在概率论中曾经讨论过二维正态随机向量 $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$. 它的概率密度

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

$$(-\infty<x_1<+\infty,-\infty<x_2<+\infty),$$

其中 $\mu_1=EX_1,\mu_2=EX_2,\sigma_1^2=DX_1,\sigma_2^2=DX_2,\rho$ 是随机变量 X 和 Y

的相关系数.

下面用向量和矩阵的形式表示二维正态分布密度函数.

令

$$\boldsymbol{B}=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},\boldsymbol{\mu}=(\mu_1,\mu_2)^{\mathrm{T}},\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}},$$

于是

$$|\boldsymbol{B}|=(1-\rho^2)\sigma_1^2\sigma_2^2.$$

而

$$\boldsymbol{B}^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\dfrac{1}{\sigma_1^2} & -\rho\dfrac{1}{\sigma_1\sigma_2}\\ -\rho\dfrac{1}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2}\end{pmatrix},$$

故

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=\frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right],$$

因此,

$$f(x_1,x_2)=\frac{1}{2\pi|\boldsymbol{B}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

同理可得 n 维正态随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的概率密度为

$$f(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{B}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

其中

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},$$

$$\boldsymbol{\mu}=E\boldsymbol{X}=(EX_1,EX_2,\cdots,EX_n)^{\mathrm{T}},$$

$$\boldsymbol{B}=\begin{pmatrix}\mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Cov}(X_n,X_n)\end{pmatrix},$$

且矩阵 $\boldsymbol{B}$ 是正定的,此时 $|\boldsymbol{B}|>0$,$\boldsymbol{X}$ 为 n **维正态随机向量**,记为 $\boldsymbol{X}\sim N(\boldsymbol{\mu},\boldsymbol{B})$.

n 维正态分布 $N(\boldsymbol{\mu},\boldsymbol{B})$ 的特征函数是

$$\varphi(\boldsymbol{t})=\exp\left\{\mathrm{i}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{t}\right\}.$$

n 维正态随机向量具有如下性质:

(1) n 维正态随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的 $m(m<n)$ 个分量构成 m 维正态随机向量 $\widetilde{\boldsymbol{X}}=(X_1,X_2,\cdots,X_m)^{\mathrm{T}}$.

(2) n 维正态随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的 n 个分量 X_1, $X_2,\cdots,X_n$ 相互独立的充分必要条件是它们两两不相关.

(3) 若 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 服从 n 维正态分布 $N(\boldsymbol{\mu},\boldsymbol{B})$,且

$l_1,l_2,\cdots,l_n$ 是常数，则随机变量 $Y=\sum_{j=1}^{n}l_jX_j$ 服从一维正态分布 $N\left(\sum_{j=1}^{n}l_j\mu_j,\sum_{j=1}^{n}\sum_{k=1}^{n}l_jl_k\mathrm{Cov}(X_j,X_k)\right)$.

此性质说明 n 维正态随机向量分量的线性组合是一个正态随机变量.

(4) 若 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 服从 n 维正态分布 $N(\boldsymbol{\mu},\boldsymbol{B})$，又 m 维随机向量 $\boldsymbol{Y}=\boldsymbol{CX}$，其中 $\boldsymbol{C}$ 是 $m\times n$ 阶矩阵，则 $\boldsymbol{Y}$ 服从 m 维正态分布 $N(\boldsymbol{C\mu},\boldsymbol{CBC}^{\mathrm{T}})$.

此性质说明正态随机向量经过线性变换后仍为正态随机向量.

第四节 条件数学期望

首先我们回忆一下条件概率概念.

设 B 是一个事件，且 $P(B)>0$，则事件 B 发生的条件下事件 A 发生的条件概率为

$$P(A|B)=\frac{P(A\cap B)}{P(B)}.$$

如果 X 与 Y 是离散型随机变量，对一切使得 $P\{Y=y\}>0$ 的 y，给定 $Y=y$ 时，X 的条件概率定义为

$$P\{X=x|Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}.$$

X 的条件分布函数定义为

$$F(x|y)=P\{X\leqslant x|Y=y\}.$$

X 的条件数学期望定义为

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

如果 X 与 Y 有联合概率密度函数 $f(x,y)$，则对一切使得 $f_Y(y)\geqslant 0$ 的 y，给定 $Y=y$ 时，X 的条件概率密度定义为

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}.$$

X 的条件分布函数定义为

$$F(x|y)=P\{X\leqslant x|Y=y\}=\int_{-\infty}^{x}f(u|y)\mathrm{d}u.$$

X 的条件数学期望定义为

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

我们以 $E(X|Y)$ 表示随机变量 Y 的函数，它在 $Y=y$ 时，取值为 $E(X|Y=y)$. 条件期望的一个重要性质是对一切随机变量 X 和 Y，当期望存在时，有

$$E(X)=E[E(X|Y)]=\int E(X|Y=y)\mathrm{d}F_Y(y). \tag{1-1}$$

当 Y 为一个离散型随机变量时，式(1-1)可化为