应用随机分析课程讲义可以通过北大网盘进行下载
请本学期选修的同学点击如下链接或扫描如下二维码获得

https://disk.pku.edu.cn/link/AR6908073B54A144AAAFA89A98F090FDF0

你的北大网盘用户名是 你的姓名_你的学号
如果你未能下载，可能是我还未来得及添加你的学号
请通过邮件与我联系并告诉我你的姓名和学号, 我直接发给你讲义

我的邮箱: liuyong@math.pku.edu.cn

应用随机分析

Applied Stochastic Calculus

部分讲义 2025 版

刘 勇　编著

北京大学数学科学学院

2025 年 2 月 17 日

注意事项

2025 版讲义中蓝色标注部分是对 2024 版有较大调整的地方

参考教材

主要参考教材:

龚光鲁 随机微分方程及其应用概要 清华大学出版社 2014 年 1 月第 3 次印刷

参考书目:

1. Øksendal, B., *Stochastic Differential Equations: An Introduction with Application.* 6th ed. Springer 2005
2. Klebaner, F. C., *Introduction to Stochastic Calculus with Applications.* 3rd ed. 随机分析及其应用 人民邮电出版社 2008 (英文影印版)
3. Lawler, G. F., *Introduction to Stochastic Processes.* 2nd ed. Chapman & Hall/CRC 2006 随机过程导论 张景肖 译 机械工业出版社 2010
4. Pavliotis G. A., *Stochastic Processes and Applications, Diffusion Processes, the Fokker-Planck and Langevin Equtions.* Springer 2014

http://www.math.pku.edu.cn/teachers/liuyong/teachingindex.html

课程设置和要求

所有内容解释权归任课老师

成绩： 平时成绩 20%; 期中测验 30%; 期末考试 50%. 平时成绩包括作业与出勤情况.

作业： 请交纸质版作业

- **收发时间:** 每单周周二上课时交 (从第三周开始), 双周周二上课时发作业.
- **补交规定:** 补交作业只记录不批改, 补交作业的最后期限 2025 年 6 月 10 日期末考试发试卷前, 过此时间点补交的作业一律不计入平时成绩.
- **其它事项:** 如果有同学不愿意做作业, 请写书面声明, 2025 年 3 月 4 日上课时交给任课老师. 总评成绩另行计算.

出勤： 遵守学校相关规定. 可能会有临时点名.

期中测验： 期中测验根据进度定在第 6-9 周，一个学时 (50 分钟). 提前两周通知.

期末考试: 时间: 2025 年 6 月 10 日下午; 地点: 待定.

答疑： 时间: 第二周开始, 如果没有特殊情况一般在每周一 15:15-16:15

地点: https://www.math.pku.edu.cn/teachers/liuyong/teachingindex.html

也可以用邮件与老师预约答疑时间.

预约邮件要求:

- **邮件标题:** 应用随机分析答疑
- **邮件内容:** 姓名, 学号, 下周有空时间段 (缺一不可). 请尽量详述问题和你的困惑
- **截止时间:** 每周五下午 3 点
- **其它事项:** 不留姓名和学号的邮件将被视为垃圾邮件
  更改邮件标题将可能导致你的邮件被遗漏

目录

第一章　概率论和随机过程的基本概念

1.1　引言

我们在概率论和应用随机过程课程中主要讲一些特殊的过程及其性质, 例如: 随机游动、Poisson 过程、Brown 运动和马氏链等. 那么, 对于较为一般的随时间演化的随机现象, 也就是随机过程, 我们将如何研究呢?

中学时我们学习的一些初等函数及其性质, 例如: 二次函数、三角函数; 在大学微积分中, 我们是对较为一般的 "好" 函数进行研究的. 引入积分和微分, 进一步使用牛顿-莱布尼兹公式, 从而我们能研究常微分方程 (ordinary differential equation, ODE)、偏微分方程 (partial differential equation, PDE) 等. 那么这些方法的内在想法又是什么?

我们学过的应用随机过程就相当于中学学的若干初等函数和它们的性质, 而应用随机分析这门课就相当于我们大学学的微积分. 具体地说, 首先, 应该从数学上来定义什么是 "随机性"? 也就是需要用公理化的方法, 对一类现象来定义 "随机性", 并且量化. 在数学上定义一类随机现象关于时间的随机性, 这就是 "鞅 (martingale)". 为了定义 "鞅", 我们需要借助严格公理化意义下条件期望的概念来实现. 因此, 我们必须一开始就要了解概率论的公理化是如何建立的.

接着, 有了鞅之后, 一个自然的问题就是鞅是否可以用已知的一些 "初等" 或是 "特殊" 的过程 "简单" 的表示出来. 这就是随机积分, 例如: 关于 Brown 运动或 Poisson 点过程的随机积分. 有了随机积分之后, 就要有关于随机积分的计算公式, 这就是 Itô 公式. 它相当于微积分中的牛顿-莱布尼兹公式. 有了这些准备之后, 我们就可以把一大类随机现象表示成一个 "确定性部分"+"鞅 (随机积分) 部分", 也就是 Itô 过程, 而 Itô 随机微分方程 (stochastic differential equation, SDE) 是其中一种特殊的过程. 如同常微分方程和偏微分方程, 我们可以通过随机微分方程来描述更多的随机现象, 利用相关的数学方法得到随机微分方程的很多性质.

1.2　概率论的公理化结构

在自然界和社会生活中, 有很多不确定的现象. 在一组固定的条件下, 既可能发生, 又可能不发生的现象称为随机现象. 对随机现象的研究必然要进行 "调查"、"观测" 和 "实验", 称为随机试验 (trial). 在概率论中, 我们假定随机试验可以在相同条件下重复地进行, 每次试验的结果可能不止一个, 并且能事先确定试验的所有可能结果, 但每次试验的结果事先又不可预测. 我们通过对随机试验建立数学模型来研究随机现象的规律性. 这个数学模型应该具有哪些要素?

首先, 需要知道随机试验的所有可能的结果是什么? 因此, 我们引入以下的概念.

定义 1.2.1. *把随机试验每一个可能的结果称为一个***样本点** *(sample point), 通常用 ω 表示; 所有可能的结果组成的集合称为***样本空间** *(sample space), 通常用 Ω 表示.*

例如, 先后掷两次硬币这个随机试验可能出现的结果是 (正, 正)(正, 反)(反, 正)(反, 反), 把这四个结果作为样本点构成样本空间.

接着, 我们感兴趣试验中出现的一些事儿, 比如，在上面的例子中可能感兴趣“两次出现的结果相同”这件事儿. 它是指 (正, 正)(反, 反) 这两个样本点之一出现. 再比如,“第二次不出现反面”这件事儿. 它是指 (正, 正)(反, 正) 这两个样本点之一出现. 实际上, 它们都是一些样本点的集合.

此处给定点的一个集合 A 是指对于任何一个点 ω, 都可以判断它是不是属于 A. 如果是, 则记为 $\omega \in A$; 如果不是, 则记为 $\omega \notin A$[①]. 我们约定不包含任何点的集合也是点集, 称为空集, 记为 $\emptyset$. 我们引入以下的概念.

定义 1.2.2. 我们把**事件** (event) 定义为样本点的某个集合. 称某事件发生当且仅当它所包含的某个样本点出现.

虽然试验的样本点在试验前就很明确, 但是只有试验之后, 才能确定某个给定的事件是否发生.

我们把样本空间 Ω 本身也作为一个事件. 每次试验必然有 Ω 中的某个样本点出现, 即 Ω 必然发生. 因此, 我们称 Ω 为**必然事件**. 我们把空集 $\emptyset$ 也作为一个事件. 每次试验中, 它都不发生, 因此, 称为 $\emptyset$ 为**不可能事件**.

通常我们希望通过一些简单事件来表示复杂事件, 利用简单事件的信息来了解复杂事件的信息. 因此, 我们还需要知道事件的关系和事件的运算. 事件作为样本空间的子集, 本质上它们之间的关系和运算就是集合之间的关系与运算. 我们在此需要强调的是用概率论的语言来解释这些关系及运算.

(1) 若事件 A 中的每一个样本点都包含在事件 B 中, 称事件B **包含了事件** A; 或者说**事件** A **发生蕴含事件** B **发生**, 记为 $A \subset B$. 若 $A \subset B$ 和 $B \subset A$ 同时成立, 称**事件** A **与事件** B **等价**或A **等于** B, 记为 $A = B$.

(2) 由所有不包含在事件 A 中的样本点所组成的事件称为事件 A 的**逆事件**或**对立事件**, 记为 A^c. A^c 表示事件 A 不发生.

(3) 用 $A \cap B$ 或者 AB 表示事件 A 和事件 B 同时发生, 即 $A \cap B$ 或者 AB 表示所有同时属于 A 和 B 的样本点的集合, 称为A **与** B **的交**. 用 $\cap_{n=1}^{\infty} A_n$ 表示事件 $A_n, n = 1, 2, \ldots$ 同时发生, 即所有同时属于 $A_n, n = 1, 2, \ldots$ 的样本点的集合, 称为$A_n, n = 1, 2, \ldots$ **的交**

(4) 用 $A \cup B$ 表示事件 A 和和事件 B 至少有一个发生, 即 $A \cup B$ 表示至少属于 A 或 B 中的一个的所有样本点的集合, 称为A **与** B **的并**. 用 $\cup_{n=1}^{\infty} A_n$ 表示事件 $A_n, n = 1, 2, \ldots$ 中至少一个发生, 即表示至少属于 $A_n, n = 1, 2, \ldots$ 中一个的所有样本点的集合, 称为$A_n, n = 1, 2, \ldots$ **的并** .

(5) 若 $AB = \emptyset$, 表示事件 A 和事件 B 不同时发生, 称A **与** B **为互不相容**, 即属于 A 的样本点必不属于 B, 属于 B 的样本点必不属于 A. 若事件 $A_n, n = 1, 2, \ldots$ 两两互不相容, 即任意的 $n \neq m$, $A_n \cap A_m = \emptyset$, 称 $A_n, n = 1, 2, \ldots$ **互不相容**.

(6) 用 $A \setminus B$ 表示事件 A 发生但事件 B 不发生, 即 $A \setminus B$ 表示所有属于 A 但不属于 B 的样本点的集合, 称为A **与** B **的差**.

在本讲义中, 对于事件的运算顺序做如下的约定: 先进行逆运算, 再进行交运算, 最后是并运算或差运算. 进一步, 事件之间的关系还成立如下的运算律:

交换律: $AB = BA$; $A \cup B = B \cup A$.

结合律: $(AB)C = A(BC)$; $(A \cup B) \cup C = A \cup (B \cup C)$.

分配律: $(A \cup B) \cap C = (AC \cup BC)$; $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

[①] 这段话的理解可以借鉴如下英文表达: A particular set A is well defined if it is possible to tell whether any given point *belongs to* it or not. These two cases are denoted respectively by $\omega \in A$, $\omega \notin A$. Thus a set is determined by specified rule of membership.
英文摘自 K. L. Chung(钟开莱), F. AitSahlia, *Elementary Probability Theory.* 4th ed. Springer 2003. page 2-3

De Morgan 定理 (对偶原理): $(A \cup B)^c = A^c \cap B^c$; $(A \cap B)^c = A^c \cup B^c$.

我们还关心一些事件 (发生) 的**概率**. 所谓概率就是度量事件发生的可能性大小的量, 实际上就是以某些事件为自变量的非负函数, 必然事件的概率是 1, 不可能事件的概率是 0, 互不相容事件并的概率等于每个事件概率之和.

概括起来, 建立随机试验的数学模型时, 我们必须知道: (1) 试验的样本空间 Ω. 它应该是一个非空的集合; (2) 可以 "观测" 到的或感兴趣的事件以及这些事件通过事件的运算得到的事件的全体, 记为 $\mathcal{F}$; (3) 这些事件的概率. 更进一步, 还需要知道 $\mathcal{F}$ 满足怎样的数学结构以及各事件的概率 P 之间的关系.

当研究某个随机试验时, 首先应该有我们感兴趣的事件, 而且知道

- 如果事件 A 发生, 则可以推知 A^c 不发生. 也就是说, 如果 A 是我们感兴趣的事件, 则 A^c 也应该是我们感兴趣的事件.
- 如果 $A_n, n = 1, 2, \ldots$ 之一发生, 则可以推知事件 $\cup_{n=1}^{\infty} A_n$ 发生. 也就是说, 如果 $A_n, n = 1, 2, \ldots$ 是我们感兴趣的事件, 则 $\cup_{n=1}^{\infty} A_n$ 也应该是我们感兴趣的事件.

由此, 我们引入以下的概念.

定义 1.2.3. $\mathcal{F}$ 是由样本空间 Ω 的一些子集组成的集合, 如果满足

(1) $\mathcal{F}$ 非空;

(2) $A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}$, 即 $A \in \mathcal{F}$ 蕴含 $A^c \in \mathcal{F}$;

(3) $A_n \in \mathcal{F}$, $n = 1, 2, \cdots \Longrightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{F}$, 即 $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ 蕴含 $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$,

则称 $\mathcal{F}$ 为事件域 (event field).

由事件域的定义, 有如下简单的性质, 证明留作习题.

性质 1.2.1. $\mathcal{F}$ 是事件域, 则 $\Omega \in \mathcal{F}$; $\emptyset \in \mathcal{F}$.

结合事件域的概念, 我们引入概率的公理化定义.

定义 1.2.4. 定义在事件域 $\mathcal{F}$ 上的一个集合函数 P 称为**概率** (或**概率测度**), 如果它满足

(1) 非负性: 对于任意的 $A \in \mathcal{F}$, $\mathrm{P}(A) \geq 0$;

(2) 规范性: $\mathrm{P}(\Omega) = 1$;

(3) 可列可加性或完全可加性: 若 $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$ 互不相容, 则

$$\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n). \tag{1.2.1}$$

利用概率的性质, 有如下简单的结论, 证明留作习题.

性质 1.2.2. $\mathrm{P}(\emptyset) = 0$.

性质 1.2.3. (概率的连续性)

(1) 若 $A_n \in \mathcal{F}, n = 1, 2, \ldots$ 满足对于任意 $n \geq 1$, $A_n \subset A_{n+1}$, 则

$$\lim_{n \to \infty} \mathrm{P}(A_n) = \mathrm{P}(\cup_{n=1}^{\infty} A_n).$$

(2) 若 $A_n \in \mathcal{F}, n=1,2,\dots$ 满足对于任意 $n \geq 1$, $A_n \supset A_{n+1}$, 则
$$\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}(\cap_{n=1}^{\infty} A_n).$$

定义 1.2.5. 称三元体 $(\Omega, \mathcal{F}, \mathrm{P})$ 为**概率空间**, 其中 Ω 是样本空间, $\mathcal{F}$ 是关于这个样本空间的一个事件域, P 是定义在 $\mathcal{F}$ 上的概率.

这是 Kolmogorov 在 1933 年建立的概率论公理化结构. 一个概率空间就是描述一个随机试验的数学模型. 对于具体的随机试验如何构建概率空间, 需要视具体问题而定. 我们特别强调: 此后, 当概率空间确定好, 只有**事件域 $\mathcal{F}$ 中的元素才称为事件**, 也就是说当表述中涉及到事件时, 一定要明确它是哪个事件域的元素. 例如:

例 1.2.1. 有一个正方形盒子, 盒盖是透明的, 上面画 “十” 字线, 从盒子正上方看将盒子分成四块区域 (如图 1). 盒中有一个小球可以滚动. 随意晃动盒子, 小球停止运动后, 随机地停留在这四块区域中的某块中 (设想是理想状态, 忽略小球处于 “十” 字线上的情况).

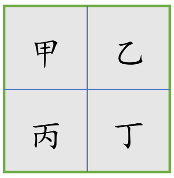


图 1

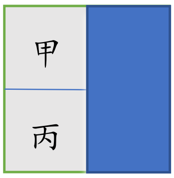


图 2

当考虑盒子经晃动后小球停留在哪些区域这个试验时, 样本空间是 $\Omega = \{$甲, 乙, 丙, 丁$\}$, 事件域 $\mathcal{F}_1$ 为 Ω 的全体子集组成的集合, 其中共有 16 个事件. 对于试验的每一个结果 ω, 对于 Ω 的每个子集 A, 总能判断出 ω 是否属于 A, 也就是说每次试验后, 总能知道事件 A 是否发生.

如果再设想用一块不透明的材料将 “乙” 和 “丁” 区域遮盖住 (如图 2). 同样考虑将盒子晃动后小球停留在哪些区域的试验, 我们也可以将试验的样本空间取为 $\Omega = \{$甲, 乙, 丙, 丁$\}$. 但是, 此时事件域
$$\mathcal{F}_2 = \{\Omega, \emptyset, \{\text{甲}\}, \{\text{丙}\}, \{\text{甲, 丙}\}, \{\text{乙, 丁}\}, \{\text{甲, 乙, 丁}\}, \{\text{丙, 乙, 丁}\}\},$$
其中只包含了 8 个事件.

对于试验的某些结果, 虽然可以看到小球停留在被遮蔽的区域, 但是我们不能判断此时小球是否在 “乙” 区域, 也就是说, 不知道 $\{$乙$\}$ 这事儿是否发生了. 因此, 对于事件域 $\mathcal{F}_2$, Ω 的子集 $\{$乙$\}$ 就不是事件. 同样的, $\{$丁$\}$ 和 $\{$甲, 乙$\}$ 等也不是事件.

在测度论中, 设 X 是一个非空集合, $\mathcal{C}$ 是 X 的一些子集组成的集合, 称为集合系 (或集类、集族). 若 $\mathcal{C}$ 满足定义1.2.3中的 (1)(2)(3), 则称 $\mathcal{C}$ 为**σ-域** (或**σ-代数**). 若 $\mathcal{C}$ 上的集合函数 μ 满足定义1.2.4中的非负性和可列可加性, 则称 μ 为 $\mathcal{C}$ 上的**测度**.

设 $\mathcal{F}$ 为 X 上的 σ-域, 称 $(\mathrm{X}, \mathcal{F})$ 为**可测空间**. X 的子集 A 称为 $\mathcal{F}$ 可测的, 若 $A \in \mathcal{F}$. 设 μ 是 σ-域 $\mathcal{F}$ 上的测度, 则称 $(\mathrm{X}, \mathcal{F}, \mu)$ 为**测度空间**. 概率空间是一类特殊的测度空间.

在测度论或概率论的研究中, 往往可以较直接的知道某个集合系 $\mathcal{C}$ 上的测度 μ, 但 $\mathcal{C}$ 未必是 σ-域. 为了构造可测空间需要增添适量的子集将 $\mathcal{C}$ 扩充成 σ-域, 还需要给新增加的子集合理地赋予数值, 成为一个测度. 这是测度论中测度扩张定理来具体实现的, 需要较多的概念和知识, 已超出了这本书的范围. 但下面的概念对本书的后续讲解是有帮助的.

定义 1.2.6. $\mathcal{C}$ 是 X 的非空集合系, $\mathcal{S}$ 是 X 的 σ-域. 若 $\mathcal{S}$ 满足下面的条件:

(1) $\mathcal{S} \supset \mathcal{C}$;

(2) 对于 X 上的任一 σ-域 $\mathcal{S}'$, 都有

$$\mathcal{S}' \supset \mathcal{C} \Longrightarrow \mathcal{S}' \supset \mathcal{S}.$$

则称 $\mathcal{S}$ 是由集合系 $\mathcal{C}$ **生成的** σ**-域**, 或**包含** $\mathcal{C}$ **的最小** σ**-域**. 记为 $\sigma(C)$.

性质 1.2.4. 由任何集合系 $\mathcal{C}$ 生成的 σ-域 $\sigma(C)$ 存在且唯一.

这个性质的证明留作习题.

我们介绍在概率论中常用的 Borel σ-域. 设 $\mathbb{R}$ 是实数全体, 记 $\mathbb{R}$ 上左开右闭区间组成的集合系为

$$\mathcal{C} = \{(a, b] : a, b \in \mathbb{R}, a < b\}.$$

称 $\sigma(\mathcal{C})$ 为 $\mathbb{R}$ 上的 Borel σ-域, 记为 $\mathcal{B}(\mathbb{R})$, 其中的元素称为 Borel 集. 不难验证对于左闭右开区间组成的集合系、开区间组成的集合系以及闭区间组成的集合等都生成 $\mathcal{B}(\mathbb{R})$.

例 1.2.2. 考虑 $\mathcal{C}$ 上的测度 μ:

$$\mu\big((a, b]\big) = b - a.$$

利用测度扩张定理可以唯一的将 $\mathcal{C}$ 上的测度 μ 扩张到 $\mathcal{B}(\mathbb{R})$ 上. 即, 在 $\mathcal{B}(\mathbb{R})$ 存在唯一的测度 $\widetilde{\mu}$ 满足, 对于任意的 $(a, b] \in \mathcal{C}$,

$$\widetilde{\mu}\big((a, b]\big) = b - a.$$

$\widetilde{\mu}$ 称为 $\mathbb{R}$ 上 Lebesgue 测度.

在 Lebesgue 测度建立过程中, 有两点需要强调一下: (1) 由实变函数知道, 并不是 $\mathbb{R}$ 的任意子集都是 Borel 集. 也就是说, 虽然 $\mathbb{R}$ 上的子集全体 $\mathscr{P}(\mathbb{R})$ 也是一个 σ-域, 但是并不能在其上构造一个测度 $\widetilde{\mu}$ 满足 $\widetilde{\mu}\big((a, b]\big) = b - a$. [1] 这也说明无论是测度论还是概率论, 恰当地确定 σ-域或事件域是非常关键的; (2) 在利用测度扩张定理时, 对于集合系 $\mathcal{C}$ 及其上的测度 μ 都是有要求的. 有兴趣的读者可以参阅相关教材.

在 n-维欧氏空间 $\mathbb{R}^n$ 中, 如果 $\mathbf{a} = (a_1, \ldots, a_n), \mathbf{b} = (b_1, \ldots, b_n)$ 满足 $a_i < b_i, i = 1, \ldots, n$, 记 $(\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^n : a_i < x_i \le b_i, i = 1, \ldots, n\}$. 设 $\mathcal{C} = \{(\mathbf{a}, \mathbf{b}] : \mathbf{a}, \mathbf{b} \in \mathbb{R}^n\}$, 称 $\sigma(\mathcal{C})$ 为 $\mathbb{R}^n$ 上的 Borel σ-域, 记为 $\mathcal{B}(\mathbb{R}^n)$, 其中的元素称为 Borel 集.

f 是从 $\mathbb{R}^n$ 映射到 $\mathbb{R}^d$ 的函数, 如果对于任意的 $\mathbf{a} \in \mathbb{R}^d$, 都有

$$\{\mathbf{x} \in \mathbb{R}^n : f_i(\mathbf{x}) \le a_i,\ i = 1, \ldots, d\} \in \mathcal{B}(\mathbb{R}^n),$$

则称 f 是 Borel 函数. 根据测度论, Borel 函数的定义等价于: 对于任意的 $A \in \mathcal{B}(\mathbb{R}^d)$, $\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \in A\} \in \mathcal{B}(\mathbb{R}^n)$.

1.3 随机变量

当建立好一个随机试验的概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 之后, 如果我们能直接得到每个事件的概率, 那么相当于完全地知道了这个试验的全部信息. 但在现实中这样的试验并不多见, 例如一些简单的古典概型. 然而, 随机现象纷繁复杂丰富多彩, 使得相应的样本空间千差万别. 有些试验结果可以用数字表示, 例如: 测量误差; 有些不是, 例如: 掷硬币的结果用正面和反面来表示. 对于复杂的试验, 我们总是希望利用简单事件的概率来推算出复杂事件的概率. 这样, 需要使用更多统一的数学方法来研究, 比如将试验结果 "数字化", 用一个数字 ξ 来表示, 也就是说 ξ 是以样本空间为定义域取值于 $\mathbb{R}$ 的函数. 但 ξ 仅仅是一个函数是不够的, 对于随机试验除了样本空间, 还需要关注事件域 $\mathcal{F}$, ξ 应该与 $\mathcal{F}$ 发生联系, 即 $\{\omega : \xi(\omega) \le x\}$ 应该是事件. 由此, 我们引入以下的概念:

[1] 严格地说, 利用测度扩张定理可以将 μ 扩张到一个比 $\mathcal{B}(\mathbb{R})$ 更大的 σ- 域 $\mathcal{L}$ 上. $\mathcal{L}$ 中的元素称为 Lebesgue 可测集. 但 $\mathcal{L}$ 也不是 $\mathscr{P}(\mathbb{R})$. 设 $\mathcal{N} = \{N \subset \mathbb{R} : \exists A \in \mathcal{B}(\mathbb{R}), \widetilde{\mu}(A) = 0 \text{ 使得 } A \supset N\}$, 则

$$\mathcal{L} = \{A \cup N : A \in \mathcal{B}(\mathbb{R}), N \in \mathcal{N}\}.$$

定义 1.3.1. 对于给定概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$, ξ 是从 Ω 到 $\mathbb{R}$ 的函数, 如果任意的 $x \in \mathbb{R}$, 都有 $\{\omega : \xi(\omega) \leq x\} \in \mathcal{F}$, 则称 ξ 是**随机变量** (random variable).

根据测度论, 随机变量的定义等价于任意的 $A \in \mathcal{B}(\mathbb{R})$, 都有 $\{\omega : \xi(\omega) \in A\} \in \mathcal{F}$, 即它是一个事件. 需要强调的是, 随机变量的定义依赖于事件域 $\mathcal{F}$. 在本书以后的行文中, 如果需要, 我们会指明随机变量是关于哪个事件域的随机变量. 如果 ξ 是关于 $\mathcal{F}$ 的随机变量, f 是 Borel 函数, 则 $f(\xi)$ 还是关于 $\mathcal{F}$ 的随机变量.

设 $A \subset \Omega$, 集合 A 的**示性函数**定义为:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A. \end{cases}$$

当 $A \in \mathcal{F}$ 时, $\mathbf{1}_A$ 是关于 $\mathcal{F}$ 的随机变量, 否则不是.

例 1.3.1. (接例1.2.1) 设

$$\xi(\omega) = \begin{cases} 1 & \omega = \text{甲}, \\ 2 & \omega = \text{乙}, \\ 3 & \omega = \text{丙}, \\ 4 & \omega = \text{丁}. \end{cases}$$

显然, ξ 是关于 $\mathcal{F}_1$ 的随机变量. 但它不是关于 $\mathcal{F}_2$ 的随机变量, 这是因为 $\{\omega : \xi(\omega) \leq 2.5\} = \{\text{甲, 乙}\} \notin \mathcal{F}_2$.

例 1.3.2. 设 $A_1, \cdots, A_n, \cdots$ 是样本空间的一个划分, 即 $A_n \cap A_m = \emptyset, n \neq m$ 且 $\cup_{n=1}^{\infty} A_n = \Omega$. 设 $\mathcal{F}$ 是包含 $\{A_1, \cdots, A_n, \cdots\}$ 的最小事件域. $(\Omega, \mathcal{F})$ 上的随机变量一定可以写成如下的形式: $\sum_{n=1}^{\infty} a_n \mathbf{1}_{A_n}$. 证明留作习题.

本节中我们仅简单地介绍与随机变量相关的基本概念和本书中常用的数字特征及必要的一些计算公式. 这部分内容在概率论或概率统计的相关教材都有详细的讲解.

定义 1.3.2. 设 ξ 是概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量. 任意的 $A \in \mathcal{B}(\mathbb{R})$, $\mathrm{P}(\omega : \xi(\omega) \in A)$ 构成 $\mathcal{B}(\mathbb{R})$ 上的一个概率, 称为 ξ 的**概率分布** (probability distribution), 简称**分布** (distribution). 两个随机向量 ξ 和 η 如果有相同的概率分布, 则称 ξ 和 η **同分布**.

定义 1.3.3. 设 ξ 是概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量, 称 x 的函数

$$F(x) = \mathrm{P}(\omega : \xi(\omega) \leq x), x \in \mathbb{R}$$

为 ξ 的**分布函数** (distribution function).

ξ 和 η 可以是两个不同概率空间上的随机变量, 但它们可以有相同的分布函数. 由测度论可以知道, ξ 和 η 的分布函数相同等价于它们同分布.

在本书中我们主要涉及到以下两类随机变量[①]:

(i) 如果随机变量 ξ 只取有限个值 $x_1, \ldots, x_n$ 或可列个值 $x_1, x_2, \ldots, x_n, \ldots$, 则称 ξ 是**离散型随机变量**. 记 $p(x_i)$ 为 $\mathrm{P}(\omega : \xi(\omega) = x_i)$, $i = 1, 2, \ldots$, 则 ξ 的概率分布可以通过如下的分布列来表示:

$\xi(\omega)$	x_1	x_2	$\ldots$	x_i	$\ldots$
P	$p(x_1)$	$p(x_2)$	$\ldots$	$p(x_i)$	$\ldots$

(1.3.1)

[①]事实上, 利用测度论中 Lebesgue 分解定理对随机变量的分布作分解, 除离散型和连续型随机变量外, 还可以得到一类奇异型随机变量. 它的分布函数是连续的, 但不存在密度函数使其可以表示成 (1.3.2) 中不定积分的形式. 同时 Lebesgue 分解定理也告诉我们任何随机变量的分布都是这三种类型随机变量分布的混合.

(ii) 对于随机变量 ξ, 如果存在非负可积函数 $p(x)$ 使其分布函数 $F(x)$ 可以表示为

$$F(x)=\int_{-\infty}^{x}p(y)\mathrm{d}y,\ x\in\mathbb{R}, \tag{1.3.2}$$

称 ξ 为**连续型随机变量**. $p(x)$ 称为 ξ 的**密度函数** (density function). $p(x)$ 满足

$$\int_{-\infty}^{\infty}p(y)\mathrm{d}y=1,$$

而且 $F(x)$ 连续且 $p(x)=F'(x)$.

定义 1.3.4. 设 ξ 是 $(\Omega,\mathcal{F},\mathrm{P})$ 上的随机变量, 如果 $\int_\Omega|\xi(\omega)|\mathrm{dP}(\omega)<\infty$, 则称 ξ 的**数学期望存在**, 定义

$$\mathrm{E}\xi=\int_\Omega\xi(\omega)\mathrm{dP}(\omega)$$

为 ξ 的**数学期望** (简称**期望**或**均值**). 如果 $\int_\Omega|\xi(\omega)|\mathrm{dP}(\omega)=\infty$, 则称 ξ 的**数学期望不存在**.

我们这里直接引用 $(\Omega,\mathcal{F},\mathrm{P})$ 上抽象 Lebesgue 积分 $\int_\Omega\xi(\omega)\mathrm{dP}(\omega)$ 给出数学期望的定义. Lebesgue 积分实质上可以理解为加权平均. 直观上, ξ 的值落入 x 的 ε-邻域 $[x-\varepsilon,x+\varepsilon)$ 的概率是 $\mathrm{P}\big(\omega:\xi\in[x-\varepsilon,x+\varepsilon)\big)$, 即 ξ 在 x 附近取值的 "权重" 近似地取成 $\mathrm{P}\big(\omega:\xi(\omega)\in[x-\varepsilon,x+\varepsilon)\big)$. 将 ξ 的值域划分成至多可列个这样互不相交的 ε-邻域 $[x_i-\varepsilon,x_i+\varepsilon)$, 那么, ξ 的加权平均就近似地等于

$$\sum_i x_i\mathrm{P}\big(\omega:\xi(\omega)\in[x_i-\varepsilon,x_i+\varepsilon)\big).$$

当划分越来越细, ε 越来越小时, 如果上面的求和式收敛, 则其极限就是数学期望 (Lebesgue 积分)[①].

由此可以知道, 设 f 是 Borel 函数, 对于随机变量的函数 $f(\xi)$, 如果 $\int_\Omega|f(\xi(\omega))|\mathrm{dP}(\omega)<\infty$, 则 $\mathrm{E}f(\xi)=\int_\Omega f(\xi(\omega))\mathrm{dP}(\omega)$. 本书中更多地是讨论离散型和连续型随机变量, 我们给出这两类随机变量数学期望如下的公式.

性质 1.3.1. 设 ξ 是离散型随机变量, 其分布列为 (1.3.1), 如果 $\sum_i|x_i|p(x_i)$ 收敛, 则

$$\mathrm{E}\xi=\sum_i x_ip(x_i).$$

设 f 是 Borel 函数, 如果 $\sum_i|f(x_i)|p(x_i)$ 收敛, 则

$$\mathrm{E}f(\xi)=\sum_i f(x_i)p(x_i).$$

设 ξ 连续型随机变量, 其密度函数为 $p(x)$, 如果 $\int_{-\infty}^{\infty}|x|p(x)\mathrm{d}x<\infty$, 则

$$\mathrm{E}\xi=\int_{-\infty}^{\infty}xp(x)\mathrm{d}x.$$

如果 $\int_{-\infty}^{\infty}|f(x)|p(x)\mathrm{d}x<\infty$, 则

$$\mathrm{E}f(\xi)=\int_{-\infty}^{\infty}f(x)p(x)\mathrm{d}x.$$

在本节和下一节中, 当讨论某随机变量的数学期望时, 如无特殊说明均表示其数学期望存在.

性质 1.3.2. 数学期望的基本性质:

(1) 对于事件 A, $\mathrm{E}\mathbf{1}_A=\mathrm{P}(A)$.

(2) 若 $\xi\ge\eta$, 则 $\mathrm{E}\xi\ge\mathrm{E}\eta$. 特别地, $\xi\ge0$, 则 $\mathrm{E}\xi\ge0$.

(3) $\mathrm{E}(a\xi+b\eta)=a\mathrm{E}\xi+b\mathrm{E}\eta$, 其中 a,b 是常数.

(4) $|\mathrm{E}\xi|\le\mathrm{E}|\xi|$.

[①] 严格地说, 应该是 $\sum_i|x_i|\mathrm{P}(\omega:\xi\in[x-\varepsilon,x+\varepsilon))$ 收敛, $\sum_i x_i\mathrm{P}(\omega:\xi\in[x-\varepsilon,x+\varepsilon))$ 的极限值定义为数学期望.

(5) Cauchy-Schwarz 不等式: 如果 $E\xi^2 < \infty, E\eta^2 < \infty$, 则

$$\big(E(\xi\eta)\big)^2 \le E\xi^2 E\eta^2.$$

等号成立的充分必要条件是存在不全为 0 的常数 a, b, 使得

$$P\big(\omega : a\xi(\omega) + b\eta(\omega) = 0\big) = 1.$$

可以利用抽象 Lebesgue 积分的理论证明以上性质, 也可以通过离散型或连续型随机变量数学期望的公式验证.

定义 1.3.5. 称 $\mathrm{var}\xi = E(\xi - E\xi)^2$ 为随机变量 ξ 的**方差** (variance); $\sqrt{\mathrm{var}\xi}$ 称为**标准差** (standard deviation).

定义 1.3.6. 设 ξ, η 是 $(\Omega, \mathcal{F}, P)$ 上的两个随机变量.

(1) 称 $\mathrm{cov}(\xi, \eta) = E\big((\xi - E\xi)(\eta - E\eta)\big)$ 为 ξ 和 η 的**协方差** (covariance). 当 $\mathrm{cov}(\xi, \eta) = 0$ 时, 称 ξ 和 η **不相关**.

(2) 当 $0 < \mathrm{var}\xi < \infty, 0 < \mathrm{var}\eta < \infty$ 时, 称

$$\rho_{\xi\eta} = \frac{\mathrm{cov}(\xi, \eta)}{\sqrt{\mathrm{var}\xi}\sqrt{\mathrm{var}\eta}},$$

称为 ξ 和 η 的**相关系数** (correlation coefficient).

由 Cauchy-Schwarz 不等式 $|\rho_{\xi\eta}| \le 1$. 当 $\xi = \eta$ 时, 方差是协方差的特例.

定义 1.3.7. 称 $\phi(\theta) = Ee^{i\theta\xi}, \theta \in \mathbb{R}$ 为 ξ 的**特征函数** (characteristic function).

定义 1.3.8. 如果 $Ee^{s\xi} < \infty$, 则称 $M(s) = Ee^{s\xi}$ 为 ξ 的**矩母函数** (moment generating fuction).

1.4 随机向量

除了随机变量, 我们还经常需要研究随机向量.

定义 1.4.1. 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是定义在同一概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量, 则称

$$\boldsymbol{\xi}(\omega) = \big(\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega)\big)$$

称为 n **维随机向量**. 一维随机向量就是随机变量.

定义 1.4.2. 设 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 维随机变量. 任意的 $A \in \mathcal{B}(\mathbb{R}^n)$, $P(\omega : \boldsymbol{\xi}(\omega) \in A)$ 构成 $\mathcal{B}(\mathbb{R}^n)$ 上的一个概率, 称为 $\boldsymbol{\xi}$ 的**概率分布** (probability distribution), 简称**分布** (distribution). 两个随机向量 $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 如果有相同的概率分布, 则称 $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ **同分布**.

定义 1.4.3. 设 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 维随机变量, 称 $\mathbb{R}^n$ 上的 n 元函数

$$F(x_1, \cdots, x_n) = P(\omega : \xi_1(\omega) \le x_1, \cdots, \xi_n \le x_n), \ (x_1, \cdots, x_n) \in \mathbb{R}^n$$

为 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$ 的**联合分布函数** (joint distribution function).

同样的, $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 可以是两个不同概率空间上的随机向量, 但它们可以有相同的联合分布函数. 由测度论可以知道, $\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 的分布函数相同等价于它们同分布.

定义 1.4.4. 设 $\boldsymbol{\xi} = \{\xi_1, \cdots, \xi_n\}$ 是 $(\Omega, \mathcal{F}, P)$ 上的随机向量, $d < n$, $1 \le i_1 < \cdots < i_d \le n$, 则任意的 $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P\big(\omega : (\xi_{i_1}, \cdots, \xi_{i_d}) \in A\big)$$

构成 $\mathcal{B}(\mathbb{R}^d)$ 上的一个概率, 称为 $\boldsymbol{\xi}$ 关于 $(\xi_{i_1}, \cdots, \xi_{i_d})$ 的**边际分布** (marginal distribution).

称 $(x_{i_1}, \cdots, x_{i_d}) \in \mathbb{R}^d$ 的函数

$$F_{i_1,\cdots,i_d}(x_{i_1}, \cdots, x_{i_d}) = P\big(\omega : \xi_{i_1}(\omega) \le x_{i_1}, \cdots, \xi_{i_d}(\omega) \le x_{i_d}\big)$$

为 $\boldsymbol{\xi}$ 关于 $(x_{i_1}, \cdots, x_{i_d})$ 的**边际分布函数**.

在本书中我们更多关心的是两类随机向量:

(i) 设 $\xi_1,\cdots,\xi_n$ 是离散型随机变量, 则称 $\boldsymbol{\xi}=\{\xi_1,\cdots,\xi_n\}$ 为离散型随机向量. 设 ξ_i 的取值为 $x_i(k)$, $k\geq 1$, $i=1,\cdots,n$, 则 $\boldsymbol{\xi}$ 的取值为 $\big(x_1(k_1),\cdots,x_n(k_n)\big)$, $k_1,\cdots,k_n\geq 1$. $\boldsymbol{\xi}$ 的分布列记为

$$p\big(x_1(k_1),\cdots,x_n(k_n)\big)=\mathrm{P}\big(\omega:\xi_1(\omega)=x_1(k_1),\cdots,\xi_n(\omega)=x_n(k_n)\big).$$

设 $d<n$, $j_1<\cdots<j_{n-d}$, 且 $\{j_1,\cdots,j_{n-d}\}=\{1,\cdots,n\}\setminus\{i_1,\cdots,i_d\}$, 则 $\boldsymbol{\xi}$ 关于 $(\xi_{i_1},\cdots,\xi_{i_d})$ 的边际分布列为

$$p\big(x_{i_1}(k_{i_1}),\cdots,x_{i_d}(k_{i_d})\big)=\sum_{k_{j_1}}\cdots\sum_{k_{j_{n-d}}}p\big(x_1(k_1),\cdots,x_n(k_n)\big).$$

(ii) 对于 n 维随机变量 $\boldsymbol{\xi}$, 如果存在非负可积函数 $p(x_1,\cdots,x_n)$ 使其联合分布函数 $F(x_1,\cdots,x_n)$ 可以表示为

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

称 $\boldsymbol{\xi}$ 为**连续型随机向量**. $p(x_1,\cdots,x_n)$ 称为 $\boldsymbol{\xi}$ 的**联合密度函数** (joint density function). 且满足 $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=1$.

设 $d<n$, $j_1<\cdots<j_{n-d}$, 且 $\{j_1,\cdots,j_{n-d}\}=\{1,\cdots,n\}\setminus\{i_1,\cdots,i_d\}$, 则 $\boldsymbol{\xi}$ 关于 $(\xi_{i_1},\cdots,\xi_{i_d})$ 的边际密度函数为

$$p_{i_1,\cdots,i_d}(x_{i_1},\cdots,x_{i_d})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_n)\mathrm{d}x_{j_1}\cdots\mathrm{d}x_{j_{n-d}}.$$

性质 1.4.1. 设 $f(x_1,\cdots,x_n)$ 是 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的 Borel 函数, $\boldsymbol{\xi}$ 是 n 维连续型随机向量, 如果 $f(\boldsymbol{\xi})$ 的数学期望存在, 则

$$\mathrm{E}f(\boldsymbol{\xi})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

定义 1.4.5. 如果 $\mathrm{E}\xi_i$, $i=1,\cdots,n$ 都存在, 则称

$$\mathrm{E}\boldsymbol{\xi}=(\mathrm{E}\xi_1,\cdots,\mathrm{E}\xi_n)$$

为随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的**数学期望**.

定义 1.4.6. 如果 $\mathrm{E}\xi_i^2<\infty$, $i=1,\cdots,n$, 称

$$\boldsymbol{\Sigma}=\begin{pmatrix}\mathrm{cov}(\xi_1,\xi_1) & \mathrm{cov}(\xi_1,\xi_2) & \cdots & \mathrm{cov}(\xi_1,\xi_n)\\ \mathrm{cov}(\xi_2,\xi_1) & \mathrm{cov}(\xi_2,\xi_2) & \cdots & \mathrm{cov}(\xi_2,\xi_n)\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{cov}(\xi_n,\xi_1) & \mathrm{cov}(\xi_n,\xi_2) & \cdots & \mathrm{cov}(\xi_n,\xi_n)\end{pmatrix}$$

为随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的**协方差矩阵**.

协方差阵是非负定矩阵.

定义 1.4.7. 称 $\phi(\boldsymbol{\theta})=\mathrm{e}^{\mathrm{i}\sum_{i=1}^n\theta_i\xi_i},\boldsymbol{\theta}=(\theta_1,\cdots,\theta_n)\in\mathbb{R}^n$ 为 n 维随机向量 $\boldsymbol{\xi}$ 的**特征函数**.

定义 1.4.8. 如果 $\mathrm{E}\mathrm{e}^{\sum_{i=1}^n s_i\xi_i}<\infty$, $\mathbf{s}=(s_1,\cdots,s_n)\in\mathbb{R}^n$, 称 $\mathrm{M}(\mathbf{s})=\mathrm{E}\mathrm{e}^{\sum_{i=1}^n s_i\xi_i}$ 为 n 维随机向量 $\boldsymbol{\xi}$ 的**矩母函数**.

定义 1.4.9. 设 $\xi_1,\cdots,\xi_n$ 是随机变量. 如果对任意实数 $x_1,\cdots,x_n$,

$$\mathrm{P}\big(\omega:\xi_1(\omega)\leq x_1,\cdots,\xi_n(\omega)\leq x_n\big)=\mathrm{P}\big(\omega:\xi_1(\omega)\leq x_1\big)\cdots\mathrm{P}\big(\omega:\xi_n(\omega)\leq x_n\big),$$

则称随机变量 $\xi_1,\cdots,\xi_n$ **相互独立**.

$\xi_1,\cdots,\xi_n$ 相互独立等价于对于任意的 Borel 集 $A_i\in\mathcal{B}(\mathbb{R})$, $i=1,\cdots,n$,

$$\mathrm{P}\big(\omega:\xi_1(\omega)\in A_1,\cdots,\xi_n(\omega)\in A_n\big)=\mathrm{P}\big(\omega:\xi_1(\omega)\in A_1\big)\cdots\mathrm{P}\big(\omega:\xi_n(\omega)\in A_n\big).$$

定义 1.4.10. 设 $\boldsymbol{\xi}_i$ 是 d_i 维随机向量,$i=1,\cdots,n$. 如果对于任意的 Borel 集 $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$,

$$\mathrm{P}\big(\omega : \boldsymbol{\xi}_1(\omega) \in A_1, \cdots, \boldsymbol{\xi}_n(\omega) \in A_n\big) = \mathrm{P}\big(\omega : \boldsymbol{\xi}_1(\omega) \in A_1\big) \cdots \mathrm{P}\big(\omega : \boldsymbol{\xi}_n(\omega) \in A_n\big).$$

则称随机向量 $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n$ **相互独立**.

定义 1.4.11. 设 $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots$ 是一列随机向量, 如果对于任意的 n, $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n$ 相互独立, 则称 $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots$ **相互独立**.

1.5 随机过程

1.5.1 随机过程的一种构造方式

结合我们前面对概率论公理系统的了解, 从应用随机过程的知识知道, 一个随机过程是定义在同一个概率空间 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的以 T 为指标集的随机变量族 $\{X_t,\ t \in T\}$(或记为 $\{X(t),\ t \in T\}$). T 解释为时间, 可以是 $\{0, 1, 2 \cdots, n\}$、非负整数集、非负实数、整数或实数. 对于任意固定的 $t \in T$, X_t 的取值空间称为状态空间, 记为 S, 其中的元素称为状态.

我们想知道一个随机过程的样本空间和样本点可以是什么样子, 具体而言是想了解 Ω 与 T 和 S 之间关系. 我们可以把 $\{X_t,\ t \in T\}$ 写成 $\{X(t, \omega),\ t \in T, \omega \in \Omega\}$, 因此将过程看成 $T \times \Omega$ 到 S 的函数 (映射). 当 $t \in T$ 固定时, $X(t, \cdot)$ 是一个 $(\Omega, \mathcal{F}, \mathrm{P})$ 上的随机变量; 当 $\omega \in \Omega$ 固定时, $X(\cdot, \omega)$ 是一个从 T 到 S 的函数 (映射), 称为对于 ω 的样本轨道.

例 1.5.1. 考虑一个先后掷 4 次硬币这个过程, 正面记为 1, 反面记为 0, 则整个过程的全体样本点是 $\omega = (\omega_1\omega_2\omega_3\omega_4)$, 其中 $\omega_i = 0$ 或 1, $i = 1, 2, 3, 4$. 即,

$$\Omega = \begin{pmatrix} (0000) & (0001) & (0010) & (0011) \\ (0100) & (0101) & (0110) & (0111) \\ (1000) & (1001) & (1010) & (1011) \\ (1100) & (1101) & (1110) & (1111) \end{pmatrix}.$$

那么过程实际就是 $X(t, \omega) = \omega_t$. 例如: $X\big(3, (\omega_1\omega_2\omega_3\omega_4)\big) = \omega_3$, 更具体地, $X\big(3, (0010)\big) = 1$. 因此, Ω 看成

$$\{0,1\}^4 = \{0,1\} \times \{0,1\} \times \{0,1\} \times \{0,1\} = \{\omega = (\omega_1\omega_2\omega_3\omega_4),\ \omega_i = 0 \text{ 或 } 1,\quad i = 1, 2, 3, 4\}$$

或等价于 $\{1, 2, 3, 4\}$ 到 $\{0, 1\}$ 的映射全体.

一般地, 对于一个状态空间为 S 的随机过程 $\{X_t,\ t \in T\}$, Ω 可以取成 $S^T = \{T\text{到}S\text{的映射}\}$, 进一步利用 Kolmogorov 相容性定理, 可以构造出事件域 $\mathcal{F}$ 以及事件域上的概率测度 P, 使得 P 应该满足 $\{X_t,\ t \in T\}$ 的统计性质, 比如, 独立同分布性质等.

1.5.2 信息、事件域和带流 (滤子) 的概率空间

在 "随机世界" 中, 与 "确定性世界" 不同, 由于不确定性, 人们不能精确预测 "运动" 未来的 "轨迹", 但人们总是希望通过已知过去和现在的信息来预测 "未来". 如何在概率空间的框架下来精确定义这种 "信息"? 这就是事件域流. 在这里, 我们仅用下面简单的例子来说明这一思想.

考虑先后掷三次硬币这一随机过程, 样本空间是

$$\Omega = \{000, 001, 010, 011, 111, 110, 101, 100\} = \{0,1\}^3;$$

样本点表示成 $\omega = (\omega_1\omega_2\omega_3)$, 事件域可以取为 $\mathcal{F} = \mathscr{P}(\Omega)$.

在掷硬币之前, 也就是在试验之前, 我们只能确定所有的样本点是什么, 但并不知道哪个样本点将出现. 我们所能了解的信息仅仅是必然事件 Ω 发生, 不可能事件 $\emptyset$ 不发生, 也就是说, 这时我们能知道的事件域是 $\mathcal{F}_0 = (\Omega, \emptyset)$.

图一

当第一次掷完之后, 虽然, 试验还未完成, 我们不能预测具体某个样本点 ω 是否最终出现, 但这时已经知道 ω 的部分 “信息”. 例如, 若观测到正面出现, 也就是说 $\omega_1 = 1$, 那么我们可以知道什么样的 “信息”? 推断出哪些事件发生或不发生?

图二

$$\omega \in A_1 = \{第一次是正面\} = \{111, 110, 101, 100\},\ \omega \in \Omega,$$

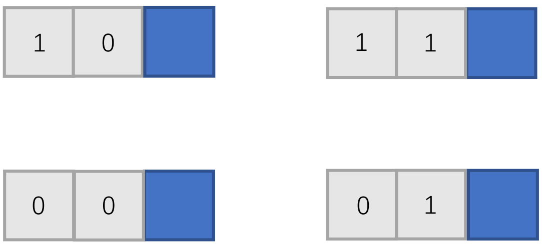
$$\omega \notin A_0 = \{第一次是反面\} = \{000, 001, 010, 011\},\ \omega \notin \emptyset;$$

若 $\omega_1 = 0$, 类似地,

$$\omega \notin A_1 = \{第一次是正面\} = \{111, 110, 101, 100\},\ \omega \in \Omega,$$
$$\omega \in A_0 = \{第一次是反面\} = \{000, 001, 010, 011\},\ \omega \notin \emptyset.$$

当第一次掷完之后, 对于每一个样本点 ω, 对于任意的 $A \in \mathcal{F}_1 = \{\Omega, \emptyset, A_1, A_0\}$ 我们总是能判断出 $\omega \in A$ 是否成立, 也就是说能判断事件 A 是否发生. 此时, 我们还注意到 $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}$.

当第一次, 第二次掷完之后, 若 $\omega_1 = 1$, $\omega_2 = 0$, 则我们可以推断出的信息是



图三

$$\omega \in A_{10} = \{100, 101\},\ \omega \notin A_{11} = \{110, 111\},\ \omega \in \Omega,$$
$$\omega \notin A_{00} = \{000, 001\},\ \omega \notin A_{01} = \{010, 011\},\ \omega \notin \emptyset,$$

并且可以推断出 ω 是否属于这几个事件的交、并、对立事件以及其交、并、对立事件再交、并和对立事件, 如此往复下去, 也就是说总能推断出此时 ω 是否属于 A, A 是事件域 $\mathcal{F}_2$ 中的任一事件.

$$\begin{aligned}\mathcal{F}_2 \quad = \quad & \{\Omega, \emptyset, A_1, A_0, A_{11}, A_{10}, A_{01}, A_{00}, A_{11}^c, A_{10}^c, A_{01}^c, A_{00}^c, \\ & A_{11} \cup A_{01}, A_{11} \cup A_{00}, A_{10} \cup A_{01}, A_{01} \cup A_{00}\}.\end{aligned}$$

若 $\omega_1=1,\omega_2=1$; $\omega_1=0,\omega_2=0$ 或 $\omega_1=0,\omega_2=1$ 我们同样可以推断出它们是否属于 $\mathcal{F}_2$ 中的事件. 即, 此时对于每一个的样本点 ω, 对于任意的 $A\in\mathcal{F}_2$, 我们总能判断出 $\omega\in A$ 是否成立. 此时, 我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}_2\subset\mathcal{F}$.

当三次硬币都掷完之后, 对于任意的样本点 ω, 我们总能判断出对于任意的 $A\in\mathcal{F}_3$, $\omega\in A$ 是否成立, 即事件 A 是否发生, 其中 $\mathcal{F}_3=\mathscr{P}(\Omega)$. 此时, 我们还注意到 $\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}_2\subset\mathcal{F}_3\subset\mathcal{F}$.

若进一步考虑 $T=\mathbb{Z}^+, S=\{0,1\}$, 即先后掷无穷多次硬币, 则 Ω 可以取成

$$\Omega=\{0,1\}^{\mathbb{Z}^+}=\{\omega|\omega=(\omega_1\omega_2\omega_3\cdots),\omega_i=0\text{或}\,1,i=1,2,3,\cdots\}.$$

$\mathcal{F}_0=\{\Omega,\emptyset\}$; $\mathcal{F}_1=\{\Omega,\emptyset,A_1,A_0\}$, 其中, $A_1=\{\omega|\omega_1=1\}=\{\omega|\omega=(1\omega_2\omega_3\cdots),\ \omega_i=0\text{或}1,\ i\geq 2\}$, $A_0=\{\omega|\omega_1=0\}=\{\omega|\omega=(0\omega_2\omega_3\cdots),\ \omega_i=0,\text{或}1,\ i\geq 2\}$. 类似地, 可以定义 $\mathcal{F}_n, n\geq 2$.

设 $\mathcal{F}_\infty$ 是包含 $\cup_{n=1}^\infty\mathcal{F}_n$ 的最小事件域 (但不是 $\mathscr{P}(\Omega)$). 我们可以取 $\mathcal{F}=\mathcal{F}_\infty$. 因此, 我们发现 $(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\geq 1},\mathrm{P})$ 可以比 $(\Omega,\mathcal{F},\mathrm{P})$ 更好地描述掷多次硬币这个随机试验 (随机过程).

因此研究随机过程时, 我们通常要考虑一个带流的概率空间 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in T},\mathrm{P})$ (filtered probability space), 满足 $t_1<t_2$ 时, $\mathcal{F}_{t_1}\subset\mathcal{F}_{t_2}\subset\mathcal{F}$. $(\mathcal{F}_t)_{t\in T}$ 被称为事件域流 (σ-代数流, σ-域流) 或滤子[1,2].

1.6　方差有限的随机变量空间和 Gauss 系

1.6.1　$L^2(\Omega,\mathcal{F},\mathrm{P})$

数学期望是随机变量的一个重要数字特征, 它表示随机变量取值的平均水平, 但是仅用数学期望描述随机变量通常是不够的, 通常还要利用方差. 方差描述了随机变量关于其数学期望的离散程度. 方差越大, 说明这个随机变量越"随机", 可以认为是随机现象"随机性"的一种刻画. 因此, 在一定程度上可以用期望和方差两个数字特征来描述随机现象. 有时, 在金融和经济领域, 方差有时也被理解成对"风险"的一种度量. 我们知道, 当一个随机变量的二阶矩有限时, 它的数学期望和方差一定存在. 因此, 我们下面来看这样一类随机变量有什么样的数学结构.

概率空间 $(\Omega,\mathcal{F},\mathrm{P})$ 上二阶矩有限, 也就是说方差有限的全体随机变量组成的空间记为 $L^2(\Omega,\mathcal{F},\mathrm{P})$(或 $L^2(\Omega)$).

首先, 回忆在高等代数或线性代数中 n 维欧氏空间 (Euclid 空间) 的概念. 在本节中 n 维向量指列向量, t 表示矩阵转置. 欧氏空间有两个基本的要点:

(1) 线性空间;

(2) 内积: $\mathbf{x}=(x_1,\cdots,x_d)^{\mathrm{t}},\mathbf{y}=(y_1,\cdots,y_d)^{\mathrm{t}}$, $\langle\mathbf{x},\mathbf{y}\rangle=\sum_{i=1}^d x_iy_i$. 由内积可以定义欧氏距离: $\|\mathbf{x}-\mathbf{y}\|=\sqrt{\sum_{i=1}^d(x_i-y_i)^2}=\sqrt{\langle\mathbf{x}-\mathbf{y},\mathbf{x}-\mathbf{y}\rangle}$[3,4].

事实上, $L^2(\Omega,\mathcal{F},\mathrm{P})$ 也是一个欧氏空间, 也就是说:

(1) 线性空间: 若 $\xi,\eta\in L^2(\Omega,\mathcal{F},\mathrm{P})$, 即 $\mathrm{E}\xi^2<\infty$, $\mathrm{E}\eta^2<\infty$, 则由 Cauchy-Schwarz 不等式, 对于任意的 $a,b\in\mathbb{R}$,

$$\begin{aligned}\mathrm{E}(a\xi+b\eta)^2&\leq a^2\mathrm{E}\xi^2+b^2\mathrm{E}\eta^2+2|ab|\mathrm{E}(\xi\eta)\\&\leq a^2\mathrm{E}\xi^2+b^2\mathrm{E}\eta^2+2|ab|\sqrt{\mathrm{E}\xi^2\mathrm{E}\eta^2}\\&<\infty.\end{aligned}$$

---

[1] 在实际研究中, 有时 $\mathcal{F}$ 可以为 $\mathcal{F}_\infty$, 也可以取的比 $\mathcal{F}_\infty$ 大.

[2] 因为在学习应用随机过程时遇到的过程简单些, 用的是所谓的"自然 σ-域流", 所以当时并不强调带流的概率空间的概念. 在研究复杂的随机现象时, 考虑带流的概率空间就非常有用了.

[3] 欧氏距离的定义本质上来源于勾股定理, 内积和欧氏距离的关系是通过平行四边形法则相联系.

[4] 一般的实线性空间中抽象的内积定义是:$\langle\mathbf{x},\mathbf{y}\rangle$ 是对称双线性型, 且满足 (1) $\langle\mathbf{x},\mathbf{x}\rangle\geq 0$; (2) $\langle\mathbf{x},\mathbf{x}\rangle=0$ 当且仅当 $\mathbf{x}=0$.

(2) 内积: 可以在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 上定义内积, 对于任意的 $\xi,\eta\in L^2(\Omega,\mathcal{F},\mathrm{P})$,

$$\langle\xi,\eta\rangle=\mathrm{E}(\xi\eta)\le\sqrt{\mathrm{E}\xi^2\mathrm{E}\eta^2}<\infty.$$

很容易验证 $\langle\xi,\eta\rangle$ 满足内积的定义. 此时欧氏距离定义为

$$\|\xi-\eta\|=\sqrt{\langle\xi-\eta,\xi-\eta\rangle}=\sqrt{\mathrm{E}(\xi-\eta)^2}.$$

其实, 这就是概率论中的两个随机变量的均方距离.

在 $L^2(\Omega,\mathcal{F},\mathrm{P})$, 若 $\mathrm{P}\big(\omega:\xi(\omega)=\eta(\omega)\big)=1$, 则 ξ 和 η 视为同一元素. 如果 $\xi\in L^2(\Omega,\mathcal{F},\mathrm{P})$, 且它的标准差 $\sqrt{\mathrm{var}\xi}>0$, 则 $\frac{\xi-\mathrm{E}\xi}{\sqrt{\mathrm{var}\xi}}$ 的数学期望为 0, 标准差为 1. 我们称 $\frac{\xi-\mathrm{E}\xi}{\sqrt{\mathrm{var}\xi}}$ 是 ξ 的标准化, 而且

$$\xi=\mathrm{E}\xi+\sqrt{\mathrm{var}\xi}\,\frac{\xi-\mathrm{E}\xi}{\sqrt{\mathrm{var}\xi}}.$$

在概率论中 $\mathrm{cov}(\xi,\eta)$ 是对随机变量 ξ 与 η 之间 “线性相关程度” 的一种 “度量”, 为什么可以这么说?

在有限维欧氏空间中, $\frac{\langle\mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}$ 表示两个向量 $\mathbf{x}$ 和 $\mathbf{y}$ 之间夹角的余弦, 可以理解为它们之间线性相关性的一种度量. 而相关系数 $\rho_{\xi\eta}=\dfrac{\mathrm{cov}(\xi,\eta)}{\sqrt{\mathrm{var}\xi}\sqrt{\mathrm{var}\eta}}$, $|\rho_{\xi\eta}|\le1$, 事实上就是两个向量之间夹角余弦这一概念在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的推广. 因此, 可以用协方差 $\mathrm{cov}(\xi,\eta)$ 或相关系数 $\rho_{\xi\eta}$ 来表示两个随机变量之间 “线性相关” 的程度.

$L^2(\Omega,\mathcal{F},\mathrm{P})$ 可能是一个无穷维的欧氏空间. 也就是说可能在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中找不到有限个元, 使之构成一组基将其中的元素线性表示出来.

定义 1.6.1. *在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中随机变量序列 $\{\xi_n\}$ 称为收敛到 $\xi\in L^2(\Omega,\mathcal{F},\mathrm{P})$, 如果当 $n\to\infty$ 时, 有 $\|\xi_n-\xi\|\to0$, 即 $\mathrm{E}(\xi_n-\xi)^2\to0$, 也就是随机变量序列的均方收敛.*

定义 1.6.2. *在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的随机变量序列 $\{\xi_n\}$ 称为 Cauchy 序列, 如果当 $n,m\to\infty$ 时, 有 $\|\xi_n-\xi_m\|\to0$.*

性质 1.6.1. *$L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的任意 Cauchy 序列必有极限, 即如果 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的随机变量序列满足 $n,m\to\infty$ 时, 恒有 $\|\xi_n-\xi_m\|\to0$, 则一定存在 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 中的元素 ξ, 使得 $\|\xi_n-\xi\|\to0$. 因此, $L^2(\Omega,\mathcal{F},\mathrm{P})$ 是完备的, 即 $L^2(\Omega,\mathcal{F},\mathrm{P})$ 是一个 Hilbert 空间.*

1.6.2 Gauss 分布、Gauss 系和 Gauss 过程

在方差有限的随机变量 (随机向量) 类中, 有一类特殊的随机变量 (随机向量), 这就是服从 Gauss 分布的随机变量 (向量).

定义 1.6.3. *设 $\eta_1,\cdots,\eta_m$ 为 m 个相互独立服从标准正态分布的随机变量, $(a_{ij})_{1\le i\le d,1\le j\le m}$, μ_i, $1\le i\le d$ 为常数, 设*

$$\begin{cases}\xi_1 &= a_{11}\eta_1+\cdots+a_{1m}\eta_m+\mu_1\\ \cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \xi_d &= a_{d1}\eta_1+\cdots+a_{dm}\eta_m+\mu_d\end{cases},$$

则称 d 维随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_d)^{\mathrm{t}}$ 服从 d 维 Gauss 分布, 其矩阵形式为

$$\begin{pmatrix}\xi_1\\ \vdots\\ \xi_d\end{pmatrix}=\begin{pmatrix}a_{11}&\cdots&a_{1m}\\ \vdots&\ddots&\vdots\\ a_{d1}&\cdots&a_{dm}\end{pmatrix}\begin{pmatrix}\eta_1\\ \vdots\\ \eta_m\end{pmatrix}+\begin{pmatrix}\mu_1\\ \vdots\\ \mu_d\end{pmatrix}.$$

即

$$\boldsymbol{\xi}=A\boldsymbol{\eta}+\mu,\ \ A=(a_{ij})_{1\le i\le d,1\le j\le m},\ \ \boldsymbol{\eta}=(\eta_1,\cdots,\eta_m)^{\mathrm{t}},\ \ \boldsymbol{\mu}=(\mu_1,\cdots,\mu_d)^{\mathrm{t}}.$$

$\boldsymbol{\xi}$ 有如下的一些简单性质:

(1) $\mathrm{E}\boldsymbol{\xi} = \boldsymbol{\mu}$, 协方差矩阵 $\boldsymbol{\Sigma} = AA^{\mathrm{t}}$.

(2) $\boldsymbol{\xi}$ 的特征函数 $\phi(\boldsymbol{\theta}) = \exp\big(\mathrm{i}\langle\boldsymbol{\theta}, \boldsymbol{\mu}\rangle - \frac{1}{2}\boldsymbol{\theta}^{\mathrm{t}}\boldsymbol{\Sigma}\boldsymbol{\theta}\big)$.

(3) $\boldsymbol{\xi}$ 的矩母函数 $\mathrm{M}(\mathbf{s}) = \exp\big(\langle\mathbf{s}, \boldsymbol{\mu}\rangle + \frac{1}{2}\mathbf{s}^{\mathrm{t}}\boldsymbol{\Sigma}\mathbf{s}\big)$.

(4) 若 $\boldsymbol{\Sigma}$ 可逆, 称 $\boldsymbol{\xi}$ 的分布称为 d 维正态分布 (normal distribution), 其密度函数

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{t}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big),$$

其中, $|\boldsymbol{\Sigma}|$ 表示 $\boldsymbol{\Sigma}$ 的行列式, 记 $\boldsymbol{\xi} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(5) 若 $\mathrm{P}\big(\omega : \boldsymbol{\xi}(\omega) = \boldsymbol{\mu}\big) = 1$, $\boldsymbol{\mu}$ 是一个常向量, 则 $\boldsymbol{\xi}$ 是一个协方差阵为 0 的 Gauss 随机向量.

性质 1.6.2. (1) $\boldsymbol{\xi}$ 服从 d 维 Gauss 分布, 则对于任意的 $r < d$, 其任意的 r 维边际分布还是 Gauss 分布. 反之不然.

(2) $\boldsymbol{\xi}$ 服从 d 维正态分布, 则对于任意的 $r < d$, 其任意的 r 维边际分布还是正态分布. 反之不然.

例 1.6.1. 设 $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$ 是标准正态分布的密度函数, $g(x) = \begin{cases} \cos x & |x| < \pi \\ 0 & |x| \geq \pi \end{cases}$. 设二元随机向量 (ξ, η) 的密度函数为

$$p(x, y) = \phi(x)\phi(y) + \frac{1}{2\pi}\mathrm{e}^{-\pi^2}g(x)g(y), \quad (x, y) \in \mathbb{R}^2.$$

(ξ, η) 不是 Gauss 随机向量, 但 ξ 服从标准正态分布, η 亦然.

性质 1.6.3. 下列命题等价.

(1) 随机向量 $\xi = (\xi_1, \cdots, \xi_d)^{\mathrm{t}}$ 服从 d 维 Gauss 分布.

(2) 对于任意 $a_1, \cdots, a_d$, 线性组合 $\sum_{k=1}^{d} a_k\xi_k$ 服从一维 Gauss 分布.

这个性质说明 Gauss 分布的一个重要性质: **关于线性运算封闭**, 也就是说如果 $(\xi_1, \cdots, \xi_d)^{\mathrm{t}}$ 服从多元 Gauss 分布, 则 $\xi_1, \cdots, \xi_d$ 的任意线性组合 $\sum_{k=1}^{d} a_k\xi_k$ 服从一维 Gauss 分布. 另外, 服从 Gauss 分布的随机变量还有一个重要性质: **关于均方收敛意义下封闭**, 也就是说:

性质 1.6.4. 设 $\xi_n, n = 1, 2\cdots$ 是一列服从 Gauss 分布的随机变量, 若存在 ξ 使得 $\lim_{n\to\infty}\mathrm{E}(\xi_n - \xi)^2 = 0$, 则 ξ 也服从 Gauss 分布.

定义 1.6.4. 随机变量族 $\{\xi_t, t \in I\}$ 称为 Gauss 系, 如果对于任意 n 以及任意 $t_1, t_2, \cdots, t_n \in I$, 随机向量 $(\xi_{t_1}, \cdots, \xi_{t_n})^{\mathrm{t}}$ 服从 Gauss 分布. 如果 Gauss 系中的指标集 I 取成时间参数, 则之称为 Gauss 过程.