第一讲：概率论简介

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

参考书目

- 概率论, 张颢, 高等教育出版社
- 概率论 (第二版), 杨振明, 科学出版社
- 概率论基础, 李贤平, 高等教育出版社
- 概率论 (第三版), 苏淳, 冯群强, 科学出版社
- 概率论导论 (翻译版), Joseph K. Blitzstein, Jessica Hwang (张景肖译), 机械工业出版社
- 概率导论 (第二版), Dimitri P. Bertsekas, John N. Tsitsiklis (郑忠国, 童行伟译), 人民邮电出版社
- Probability, A. N. Shiryaev, Springer
- 科学发现纵横谈, 王梓坤, 北京师范大学出版社

参考教材封面

概率论
Probability Theory
张颖 编著

"十二五"普通高等教育本科国家级规划教材
南开大学数学教学丛书
概率论
第二版
杨振明 编
科学出版社

"十二五"普通高等教育本科国家级规划教材
概率论基础
第三版
复旦大学 李贤平
高等教育出版社

中国科学院考研指定参考书
普通高等教育"十一五"国家级规划教材
中国科学技术大学数学教学丛书
概率论
(第三版)
苏淳 冯群强 编著
科学出版社

"十三五"国家重点出版物出版规划项目
名校名家基础学科系列
概率论导论
(翻译版)
Introduction to Probability
[美] 约瑟夫·K. 布利茨斯坦 (Joseph K. Blitzstein) 著
哈佛大学
杰西卡·黄 (Jessica Hwang)
南加州大学
张景肖 译
CRC Press
Taylor & Francis Group
机械工业出版社
CHINA MACHINE PRESS

TURING 图灵数学·统计学丛书
美国工程院院士力作，MIT等全球众多名校教材！
Introduction to Probability, 2E
概率导论
[美] Dimitri P. Bertsekas John N. Tsitsiklis 著
郑忠国 童行伟 译
(第2版·修订版)
从直观、自然的角度阐述概率，是理工科学生入门首选。
中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS

Graduate Texts in Mathematics
A.N. Shiryaev
Probability
Second Edition
Springer

"十一五"国家重点图书出版规划项目
科学素养大家谈丛书
科学发现纵横谈
——王梓坤教育随笔
王梓坤 著
大连理工大学出版社
DALIAN UNIVERSITY OF TECHNOLOGY PRESS

成绩构成

- 20% - 30%：平时成绩，含平时作业、大作业、讨论报告等。
- 20% - 30%：期中随堂测验
- 50% - 60%：期末考试
- 5 分：课堂及平时表现加分项
- 考勤：不定时随机点名，如被点到未来上课者每次**扣总评成绩 5 分**
- 期末考试过后，凡与考试成绩有关的邮件,QQ 留言，微信留言均不回复，敬请谅解

答疑时间与课程 QQ 群

- 答疑
  - 时间：每周二中午 12:00-13:30
  - 地点：新文科楼 528-4 办公室
- 课程 QQ 群
  - 理科试验班：814238326
  - 强基拔尖班：964600973

作业提交

- 自由组合, 每两人一组并确定组长由组长提交每次作业
- 提交方式: Gitee 线上提交 (Markdown 文件与导出的 pdf 文件)
- 提交时间: 每周日晚 12 点之前
- 如何提交:
  - 注册 Gitee 账号并更新个人信息为" 学号-姓名":
    强基拔尖班: 强基拔尖班邀请链接注册 Gitee 账号.
    理科试验班: 理科试验班邀请链接注册 Gitee 账号.
  - 苹果系统已预装 Git, Window 系统需下载安装 Git:
    下载地址: https://git-scm.com/download/
  - 设置 Gitee SSH 公钥, 以便推拉作业仓库

Gitee SSH 公钥设置

生成 SSH 公钥

> Windows 用户建议使用 **Windows PowerShell** 或者 **Git Bash**，在 **命令提示符** 下无 `cat` 和 `ls` 命令。

1. 通过命令 `ssh-keygen` 生成 SSH Key：

```
ssh-keygen -t ed25519 -C "Gitee SSH Key"
```

- `-t` key 类型
- `-C` 注释

输出，如：

```
Generating public/private ed25519 key pair.
Enter file in which to save the key (/home/git/.ssh/id_ed25519):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/git/.ssh/id_ed25519
Your public key has been saved in /home/git/.ssh/id_ed25519.pub
The key fingerprint is:
SHA256:ohDd0OK5WG2dx4gST/j35HjvlJlGHvihyY+Msl6IC8I Gitee SSH Key
The key's randomart image is:
+--[ED25519 256]--+
|     .o          |
|    .+oo         |
|   ...O.o +      |
|    .= * = +.    |
|   .o +..S*. +   |
|. ...o o..+* *   |
|.E. o . ..+.O    |
| . . ... o =.    |
|    ..oo. o.o    |
+----[SHA256]-----+
```

- 中间通过三次**回车键**确定

2. 查看生成的 SSH 公钥和私钥：

```
ls ~/.ssh/
```

输出：

```
id_ed25519  id_ed25519.pub
```

- 私钥文件 `id_ed25519`
- 公钥文件 `id_ed25519.pub`

3. 读取公钥文件 `~/.ssh/id_ed25519.pub`：

```
cat ~/.ssh/id_ed25519.pub
```

输出，如：

```
ssh-ed25519 AAAA***5B Gitee SSH Key
```

复制终端输出的公钥。

设置账户 SSH 公钥

用户可以通过主页右上角 「个人设置」->「安全设置」->「SSH 公钥」->「添加公钥」，添加生成的 public key 添加到当前账户中。

> 需要注意：**添加公钥需要验证用户密码**

通过 `ssh -T` 测试，输出 SSH Key 绑定的**用户名**：

```
$ ssh -T git@gitee.com
Hi USERNAME! You've successfully authenticated, but GITEE.COM does not provide shell acce
```

作业提交: Git 命令首次作业提交

1. Clone 作业仓库
   强基拔尖班: git clone git@gitee.com:xzhangseu/a-prob-2024.git
   理科试验班: git clone git@gitee.com:xzhangseu/b-prob-2024.git
2. 创建创建以个人学号-姓名分支:
   git checkout -b 12345-姓名
3. 将个人新建的分支推送至远程:
   git push --set-upstream origin 12345-姓名
4. 书写作业并提交:
   git add 作业文件名
   git commit -m " 第 k 次作业"
   git push

作业提交：Git 命令后续作业提交

1. 更改本地分支到 master 并拉取作业:
   git checkout master
   git pull
2. 拉取作业后更改至分支"12345-姓名" 完成作业并提交:
   git checkout 12345-姓名
3. 书写作业并提交:
   git add 作业文件名
   git commit -m " 第 k 次作业"
   git push

作业格式要求

- 作业文件命名为: 第 k 次作业
- 作业不能仅有公式, 没有语言表达
- 作业内容: 标题要添加" 第 k 次作业 + 每组两人的学号与姓名", 如下图所示

第k次作业

学号1-姓名1　　学号2-姓名2

作业提交

- 完成作业需要学习
  - Git 知识
  - Gitee 使用
  - Markdown 指南
  - Markdown Preview Enhanced 使用指南

Markdown 文档编辑软件及插件

- Visual Studio Code 插件
  - Markdown Preview Enhanced
  - Better Markdown & Latex Shortcuts
  - Auto-Open Markdown Preview
  - Markdown Preview Github Styling
- Obsidian 插件
  - obsidian-latex-suite
  - obsidian-latex
  - math-booster
  - dataview
  - mathlinks
- logseq

一些期望

- 守时：作业按规定时间提交
- 勇于表现，不做沉默者
- 培养批判性思维
- 相互尊重，教学意见及时与老师沟通
- 注重培养书面语言表达能力
- 实事求是、客观的进行教学评价

学习方法

- 课前预习
- 课上学习
- 课后复习
- 每 2-3 周完整复习前 2-3 周的学习内容

一本书三部电影

约翰·克利斯朵夫

肖申克的救赎

闻香识女人

心灵捕手

什么是概率

- 运气、巧合、随机性、不确定性、风险、时运、机会 - 用法模糊
- 概率普遍存在于科学与日常生活中, 但可能非常反直觉
- 单纯依靠直觉, 会出现预测不准确或面临预测过于自信带来严重风险
- 概率论课程旨在搭建一个有原则的量化不确定性和随机性的逻辑框架, 通过学习强化自身的直觉

几个问题

问题 1.1 假设对于概率论课程成绩有两种不同的方式确定, 你选择哪一种?

- 课程成绩由提交作业的次数来决定, 只要至少提交过 6 次作业就给你 60 分;
- 课程成绩由某随机试验的结果来决定, 但我保证: 你的课程成绩小于 60 分的概率为 0.

几个问题

问题 1.2 某人医院体检报告甲胎蛋白 (AFP) 值偏高, 医生建议进行肝细胞活检, 当问及原因时, 医生说: 我们发现在肝癌患者中, 有 99.9% 的患者其甲胎蛋白数值都偏高. 医生给出活检的理由是否充分?

几个问题

问题 1.3 三门问题: 在某电视节目中, 主持人蒙提霍尔会选择一名观众参与一个游戏.

- 三扇门: 一扇门后是一辆汽车, 另外两扇门后各有一只山羊;
- 参与者可从三扇门中任选一扇, 并可以获得该扇门后的奖品;
- 在揭晓参与者所选门后的奖品之前, 主持人会打开其中一个未被选中的一扇门, 并揭晓门后的奖品是山羊.
- 参与者可以坚持原来的选择或换成剩下的那扇门;
- 问题: 参与者是否要改变自己的选择?
- 如果将上面的问题推广至 n 扇门, 其中 k 扇门后是汽车, 那又做何选择? (此处, $n \geq 3, k \leq n-2$)

注 1.1 主持人蒙提霍尔不是在参与者未选择的两扇门中随机选择, 而是总是选取一个后面是羊的门展示.

几个问题

问题 1.4 假设有这样一个赌博游戏: 赌注是 3000 元, 赌博者可以选择掷几个骰子, 只要掷出的骰子有 6 点且没有 1 点, 则赌博者即可获得 10000 元.

- 赌博者应选几个骰子掷胜率更高?
- 如果换做是你, 你是选择玩这个赌博游戏, 还是不玩?

几个问题

问题 1.5 假设某合法赌场内有胜率 1/2 与胜率 1/20 的两类赌局. 现迫于需要, 24 小时之内要在此合法赌场将 100 万元变成 200 万元.

- 一口气将 100 万下注在胜率为 1/2 的赌局上;
- 将 100 万分成 10 份, 每份 10 万下注在胜率为 1/20 的赌局上, 赌 10 次.

问题 1.6 赌徒破产问题: 假设去某赌徒去某赌场玩轮盘赌游戏, 其获胜的概率是 0.48. 该赌徒现有资产 1000 元, 并决定一直玩下去直到资产翻倍或破产, 试问下面哪一种下注方式对赌徒更为有利

- 每次下注 1 元;
- 每次下注 100 元;
- 每次下注 200 元;
- 每次下注 1000 元.

几个问题

问题 1.7 De Mere 分赌注问题

- 甲乙两人各下赌注 m 元;
- 先胜三局者取得全部赌金;
- 甲胜一局而乙胜零局时赌博被迫中止;
- 问赌注如何分?

其他问题

- 买彩票问题;
- 抽签公平性问题;
- 生日相同的概率问题;
- 小概率事件终将发生;
- “一而再再而三”、“智者千虑, 必有一失”等谚语的概率理解;
- “狼来了”故事的概率解释.

概率论的应用领域

- 统计学：概率论是统计学的基础和语言
- 物理学："上帝不掷骰子"，但量子物理学、统计力学均建立在概率基础之上
- 生物学：遗传学、基因序列、群体的生灭、迁移等随机模型
- 计算机科学：随机算法，算法性能、机器学习、人工智能
- 气象学：天气预报
- 赌博：概率论早期研究旨在解决赌博和机会游戏的问题
- 金融学：概率是计量金融的核心
- 政治科学：预测选举结果，民调等
- 保险精算，风险管理理与控制：保费核定及风险量化等
- 医学：随机临床试验
- 化学化工：化学反应动力学反应的时变率及影响因素问题等
- 天文学：湍流理论及星云密度起伏，辐射传递等
- 运筹与控制，排队论
- 生活：生活是不确定的，在不确定中利用概率做决策。
- ……

概率论课程难不难？

- 概率论本质是分析类课程，研究内容与数学分析有很好的类比性
- 数学分析是确定性分析理论，概率论是随机性分析理论
- 难，也不难？
  - 概率论为解决随机问题提供了理论方法，同时也会产生陷阱和悖论
  - 即使是莱布尼茨和牛顿，也无法避免一些概率论上的基本错误
- 如何避免潜在错误
  - 数值模拟：利用 R 语言运行模拟，看看到底谁是对的
  - 完整性检查：使用不同的方法来解决同样的问题，或者检查所得到的答案在简单和极端的情况下是否都是合理的

概率论课程主要内容

- 概率简介 (第 1,2,6,5 章)
  - 事件及其运算, 概率定义及性质
  - 条件概率、事件独立性
  - 全概率公式、贝叶斯公式
- 随机变量 (第 3,4,8,10,9 章)
  - 随机变量定义、性质、结构、研究方法（分布、分布函数）, 分类
  - 多维随机变量定义、研究方法（联合分布、联合分布函数）, 分类
  - 常见的随机变量具体分布类型：离散型、连续型
  - 条件分布函数、随机变量的独立性
  - 分布函数形式下的全概率公式、贝叶斯公式

概率论课程主要内容

- 随机变量的数字特征：期望、方差、协方差 (第 7,11,12 章)
  - 随机变量的积分的定义及性质
  - 期望的定义以及与随机变量积分之间的关系
  - 方差、协方差的定义及性质、随机变量的矩、特征函数、矩母函数
  - 条件数学期望的定义及性质、条件方差
- 随机变量序列的极限 (第 13,14,15 章)
  - 随机变量序列的几种收敛性及相关关系：几乎必然、依概率, 依分布等
  - 弱大数定律与强大数定律
  - 中心极限定理

概率雏形

- 1470 年, 拉丁文的诗书《De Vetula》出版. 上面有首诗记录了 3 个骰子点数和的排列组合结果.

Qvot Punctaturas, et qvot Cadentias habeat qvilibet numerorũ compositorum.

3	18	Punctatura	1	Cadentia	1
4	17	Punctatura	1	Cadentiæ	3
5	16	Punctaturæ	2	Cadentiæ	6
6	15	Punctaturæ	3	Cadentiæ	10
7	14	Punctaturæ	4	Cadentiæ	15
8	13	Punctaturæ	5	Cadentiæ	21
9	12	Punctaturæ	6	Cadentiæ	25
10	11	Punctaturæ	6	Cadentiæ	27

Latin poem De Vetula (~1250?)
by Ovid pseunonim (Richard de Fournival?)
Taken from Kendall (1956)

- 左图为原始印刷, 右图为阿拉伯数字版本. 可以看出, 三个骰子点数相加等于 10 或者 11, 有 6 种组合情况, 但是排列情况有 27 种.

概率雏形

- 140 年后, 即 1610 年左右, 资助 Galileo(伽利略) 的 Tuscany(托斯卡纳) 大公请教了 Galileo 3 个骰子和的问题.
- Tuscany 大公在赌博时发现三个骰子点数和为 10 比点数和为 9 出现的更频繁一些. 他认为有 6 种方式得到 10, 分别为

$$(6,3,1),(6,2,2),(5,4,1),(5,3,2),(4,4,2),(4,3,3),$$

  同样有 6 种方式得到 9, 分别是

$$(6,2,1),(5,3,1),(5,2,2),(4,4,1),(4,3,2),(3,3,3).$$

  因此,9 和 10 出现的频率不应该有差异.
- Galileo 指出了大公的错误, 三个骰子是不同的,$(6,3,1)$ 和 $(6,1,3)$ 是不同的两种情况。可惜伽利略那个时候只考虑了频次, 还没有形成概率的思想。

概率论思想形成

- 1564 年左右, 意大利博学家 Cardano(卡尔达诺) 写了本书《Liber de Ludo Aleae ("Book on Games of Chance")》.
- Cardano 是第一个系统的推算概率的人, 三个骰子和的问题在《Liber de Ludo Aleae 》的第 13 章有提到并且解决.

Consensus sortis in duabus Aleis.

2	12	1	3	11	2	4	10	3. Æqual.
5	9	4	6	8	5	7	8	18. Ad Frit:

Consensus sortis in tribus Aleis tum Frit.

Sortis			Fritilli.		
3	18	1	3	115	
4	17	3	4	125	
5	16	6	5	126	
6	15	10	6	133	
7	14	15	7	33	Circuitus 216.
8	13	21	8	36	Æqualitas 108.
9	12	25	9	37	
10	11	27	10	36	
			11	38	
			12	26	

- 《Liber de Ludo Aleae 》直到一个世纪后的 1663 年才出版, 伽利略 1642 年去世, 所以也没有机会看到.
- Cardano 在第 14 章中明确定义了“比例”, 若赌局中有利的所有可能的数目为 a, 不利的数目为 b, 则应该根据 a/b 的结果来下注.

French Society in the 1650's

- 赌博是一种流行时尚
- 法律没有禁止
- 当游戏越来越复杂，利益越来越大时，需要数学方法来计算获胜可能性

概率论的诞生

- Cardano 虽说是概率的创派始祖, 但概率论真正变成一门学科的标志性事件则是 1654 年 Pascal(帕斯卡) 和 Fermat(费马) 关于两个赌博问题的通信.
- Pascal(帕斯卡) 就是法国物理学家和数学家帕斯卡, 命名压强单位的物理学家 Pascal.
- Fermat(费马) 就是提出“费马大定理”并且困扰数学家 300 年之久的费马.

两个赌博问题

问题 1.8 De Mere(德梅尔): 事件可能性大小

- 一颗骰子投 4 次至少得到一个六点,
- 两颗骰子投 24 次至少得到一个双六.

问题 1.9 De Mere 分赌注问题

- 甲乙两人各下赌注 m 元;
- 甲乙两人获胜的概率一样, 均为 $\frac{1}{2}$;
- 先胜 t 局者取得全部赌金;
- 甲胜 a 局而乙胜 b 局时赌博被迫中止;
- 问赌注如何分?

Chevalier De Mere

骰子点数问题

- 据经验所知, 一颗骰子投 4 次至少得到一个六点的概率大于 $\frac{1}{2}$;
- 两颗骰子掷一次的结果 6 倍于一颗骰子掷一次的结果, 则两颗骰子投 24 次至少得到一个双六的概率也应大于 $\frac{1}{2}$;
- 但赌场经验却并非如此.
- 考虑成功概率为 p 的随机试验, 连续做 n 次, 我们来确定至少一次成功的概率为 q_0 对应的 n:

$$1-(1-p)^n = q_0 \Rightarrow n = \frac{\ln(1-q_0)}{\ln(1-p)} = \frac{-\ln(1-q_0)}{p+\frac{p^2}{2}+\cdots} \approx \frac{-\ln(1-q_0)}{p};$$

- 两个不同成功概率的随机试验, 连续做 n 次, 要保其至少成功一次的概率相同所需要的 n 近似反比于单次试验成功的概率.

分赌注问题历史

- 分赌注问题最早是由意大利数学家 Paccioli(帕乔利) 在 1494 年提出, 当时 Paccioli 给出的答案是按 $a:b$ 进行分配.
- 50 年后, Tartaglia(塔尔塔利亚) 提出反对意见: 根据这条规则, 如果游戏在一轮之后停止, 那么其中一名玩家就会得到全部赌注.
- Tartaglia 试图修改帕乔利的规则, 以便将上述情况考虑进去, 但最后他怀疑这个问题可能根本没有确定的答案.
- 1537 年, Cardano 也曾经思考过这个问题.
  - Cardano 给出了一个公式 $f(n)=1+\ldots+n$, 甲还剩 $s-a$ 局就可以获胜, 乙还剩 $s-b$ 局可以获胜, 两者的分金比率应该为 $f(s-b):f(s-a)=(s-b)(s-b+1):(s-a)(s-a+1)$.
  - 虽然 Cardano 答案错误, 但 Cardano 已经开始考虑用未来剩余的赌局来决策, 而不是局限于已发生的事件.
- 直到 1654 年, De Mere 向 Pascal 请教, Pascal 和 Fermat 进行书信交流, 并且用不同的解法给出了这个问题的正确答案.
- 后来, 荷兰数学家 Huygens 也参与了 Pascal 和 Fermat 的讨论, 并在 1657 年出版书籍《De ratiociniis in ludo aleae ("On Reasoning in Games of Chance")》.

Blaise Pascal (布莱士·帕斯卡)

- 16 岁时发现著名的帕斯卡六边形定理：内接于一个二次曲线的六边形的三双对边的交点共线;
- 17 岁时写成《圆锥曲线论》，是研究德札尔格射影几何工作心得的论文;
- 设计并制作了一台能自动进位的加减法计算装置，被称为是世界上第一台数字计算器;
- 以积分学的原理解决了摆线问题;
- 与费马相互通信探求赌注分配问题，形成概率论中一个重要的基本概念—数学期望.

Pierre de Fermat (皮耶·德·费玛)

- 独立于笛卡儿发现了解析几何的基本原理;
- 求切线、求极大值和极小值以及定积分方法;
- 费马和帕斯卡在相互通信探求赌注分配问题, 形成概率论中一个重要的基本概念—数学期望;
- 数论中的费马大小定理;
- 光学的最小作用原理.

Christiaan Huygens (克里斯蒂安・惠更斯)

- Christiaan Huygens 在 1657 年写了世界上第一本关于概率论的著作 De ratiociniis in ludo aleae (“On Reasoning in Games of Chance”), 中文译名“论赌博中的计算”.
- 概率论的基本概念、第一次在书中提出期望.

分赌注问题:Fermat 解法

- 显然甲乙距离赢得赌局分别还差 $r = t - a$ 局和 $s := t - b$ 局. 因此, 若赌局不终止, 赌局肯定会在接下来的 $r + s - 1$ 轮内结束.
- 考虑所有 $r + s - 1$ 场赌局的结果, 计算在所有可能结果中, 甲乙最终获胜的概率. 这样一来, 整个问题就简化为一个古典概型的概率计算问题.
- 以上述计算出来的获胜概率作为分配赌注的比例.
- 若 $t = 6, a = 5, b = 3$, 则最多还需要赌 3 场就可以分出胜负. 后续所有可能的结果分别为

$$\{\text{甲甲甲}, \text{甲甲乙}, \text{甲乙甲}, \text{甲乙乙}, \text{乙甲甲}, \text{乙甲乙}, \text{乙乙甲}, \text{乙乙乙}\}.$$

- 所有的 7 种结果中, 乙只有一个结果获胜, 也就是连胜三局, 故赌注分配比例应为 $7:1$.

分赌注问题:Pascal 解法

- 杨辉三角或 Pascal 三角法
  - 后续 $r+s-1$ 场赌局的所有可能结果中:
    使得甲获胜的结果个数为: $n_{甲}=\sum_{k=r}^{r+s-1} C_{r+s-1}^{k}$
    使得乙获胜的结果个数为: $n_{乙}=\sum_{k=s}^{r+s-1} C_{r+s-1}^{k}$
  - 若 $t=6, a=5, b=3$ 即 $r=1, s=3$, 则

$$n_{甲}=C_3^1+C_3^2+C_3^3=7, \quad n_{乙}=C_3^3=1$$

分赌注问题:Pascal 解法

- 递推法
  - 假设甲和乙赌注共 64 个金币, 考虑 $t=6, a=5, b=4$ 的情况:
    如果下一局甲胜, 则甲得到全部 64 枚金币
    如果下一局甲输, 则比分变为 $5:5$, 这时候是平局, 甲乙平分赌注.
  - 不管哪种情况, 甲至少会拿 32 个金币, 剩余的 32 个金币, 可能归甲也可能归乙, 平分. 这样甲应该拿 $32+(64-32)/2=48$ 个金币, 乙则拿 16 个金币.
  - 考虑 $t=6, a=5, b=3$ 的情况:
    - 如果下一局甲胜, 则拿到所有赌注;
    - 如果甲败, 则问题转化为 $t=6, a=5, b=4$ 的情况;
    - 由此可见, 甲至少可拿 48 个硬币, 剩余金币平分;
    - 因此, 甲最终可拿 $48+(64-48)/2=56$ 个金币, 乙拿 8 个金币;
    - 和 Fermat 的 $7:1$ 结果相同.
  - 在递推解法里,Pascal 用到了期望的思想, 尤其是 $48+(64-48)/2$ 这个式子.
    - 下一局甲胜得到 $x=64$ 个金币, 甲输得到 $y=48$ 个金币;
    - 按 Pascal 的式子, 甲拿到的金币为:
      $y+(x-y)/2=(x+y)/2=x*0.5+y*0.5$, 这其实就是条件期望公式.

分赌注问题:Pascal 递推法的一般解法

- 赌博中断时, 甲乙距离赢得赌局分别还差 $r = t - a$ 局和 $s := t - b$ 局;
- 假设中断后赌局继续进行, 甲最终获胜的概率记为 $e(r, s)$, 则有边界条件:

$$e(0, s) = 1, \quad e(r, 0) = 0, \quad e(r, r) = 1/2$$

且有如下递推式

$$e(r, s) = [e(r-1, s) + e(r, s-1)]/2.$$

- 上式用到了全概率公式, 是一个偏差分方程. Pascal 借助算术三角阵, 利用数学归纳法得到

$$e(r, s) = \frac{D_{r+s-1,s-1}}{D_{r+s-1}} (r + s = 2, 3, \cdots)$$

- 这里分子表示算术三角形中第 $r + s - 1$ 行最后 s 项之和，而分母为第 $r + s - 1$ 行的所有数字之和, 故为 2^{r+s-1}.

Jacob Bernoulli(雅各布·伯努利)

- 概率论与变分法先驱之一;
- 第一次使用“积分”一词;
- 首次给出直角坐标和极坐标下的曲率半径公式;
- 1713 年出版《Ars Conjectandi》(可译为《猜度术》): 提出了概率论中的“伯努利大数定律”-大数定律的最早形式.

Abraham de Moivre(亚伯拉罕·棣莫弗)

- 棣莫弗公式:

$$(\cos x + i\sin x)^n = \cos(nx) + i\sin(nx)$$

- 二项分布的正态逼近: 棣莫弗-拉普拉斯中心极限定理;
- 利用概率生成函数计算均匀分布随机变量和的分布;
- 寿险保险数学: 年金的分析与计算.
- 开创了概率论的现代方法: 1718 年发表了《The Doctrine of Chance》. 在此书中统计独立性的定义首次出现. 该书在 1738 年与 1756 年出了扩展版, 生日问题出现在 1738 年的版本中, 赌徒破产问题出现在 1756 年的版本中.

Pierre-Simon Laplace(皮埃尔-西蒙・拉普拉斯)

- Pierre-Simon Laplace 在他 1812 年的著作《Théorie Analytique des Probabilitiés》(可译为《概率论的解析理论》)中介绍了概率的数学理论及其科学应用;
- Laplace 只考虑了古典概型, 对一般的概率及其应用没有介绍;
- 到 1850 年, 许多数学家发现古典概型对一般场合不合理, 开始尝试重新定义概率.

概率的公理化体系

- 从 17 世纪到 19 世纪长达 300 多年的时间里, 以 Laplace(1749 一 1827), De Moivre(1687-1754), J.Bernouil, D.Bernouli(1700–1782). Pascal、Fermat 等为代表的优秀学者从各自不同的角度对概率论的许多方面进行了深入细致的研究, 取得了大量成果, 可是都无法将概率论纳入到严格的公理化体系中. 而概率论作为数学的重要分支, 其逻辑的严密性和完整性始终受到质疑.
- 数学大师 Hilbert(1862—1943) 在 1900 年提出的 23 个著名问题中, 第 6 个问题是关于“物理定律的公理化”, 其中就涉及概率论的公理化问题.
- 激发了以 Borel(1871 一 1956) 和 Lebesgue(1875—1941) 等为代表的一批学者的研究热情. 他们在测度理论 (Measure Theory) 方面得到的成果为后续研究指明了道路.

Andrey Kolmogorov(安德列・柯尔莫哥洛夫)

- Andrey Kolmogorov 第一个在他 1933 年的著作《Grundbegriffe der Wahrscheinlichkeitsrechnun》(《Foundations of the Theory of Probability》) 中严格的定义了概率;
- 类似于 Euclid 基于公理体系建立几何, 他从基本公理建立了概率理论, 从而使概率论称为一门严谨的数学分支;
- 概率理论的现代研究和测度论非常紧密的结合在一起.

1903.4.25-1987.10.20

Kolmogorov 成就

- 20 世纪最伟大的数学家之一, 在概率论、实分析、拓扑学、泛函分析、微分方程、生物数学、等很多领域都有着开创性的贡献.
- 1933 年, 出版《概率论基础》, 首次将概率论建立在严格的公理基础上, 解决了希尔伯特第 6 问题的概率部分。
- 线性拓扑空间理论的创始人之一, 和美国著名数学家亚历山大（J.W.Alexander）同时独立引入上同调群的概念;
- 40 年代, 在遗传学、弹道学、气象学、金属结晶学等有重要贡献;
- 50 年代, 经典力学、遍历理论、函数论、信息论、算法理论等。
- 柯尔莫哥洛夫的研究几乎遍及数论之外的一切数学领域。1963 年, 在第比利斯召开的概率统计会议上, 美国统计学家沃尔夫维茨 (J.Wolfowitz) 说: “我来苏联的一个特别的目的是确定柯尔莫哥洛夫到底是一个人呢, 还是一个研究机构.”

Kolmogorov 轶事

- 历史学要证实自己的观点需要几个甚至几十个正确证明才行, 数学只需要一个证明就行, 于是转行数学.
- 20 年代的莫斯科大学, 学生被要求在十四个不同的数学分支参加十四门考试; 但考试可用相应领域的独立研究代替. Kolmogorov 从来没有参加考试, 他写了十四个不同方向的有新意的文章。
- Kolmogorov 总是以感激的口气提到斯大林: “首先, 他在战争年代为每一位院士提供了一床毛毯; 第二, 原谅了我在科学院的那次打架。” Kolmogorov 一次在选举会上打了 Luzin（卢津）一个耳光, 他说: “（打架）那是我们常用的方式。” Luzin 在实变函数方面有着很重要的贡献, 但是以打架而论, 远非 Kolmogorov 的对手.

第二讲：事件及其运算关系

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

概率: (随机) 事件发生的可能性大小的度量

- 随机现象: 一定条件下, 并不总是出现相同结果的现象, 比如抛一枚硬币, 掷一颗骰子等;
- 随机试验: 随机现象的实现和对它某个特征的观测; 常简称试验
  - 随机试验中要求试验的结果至少 2 个;
  - 每次试验或观测得到其中的一个结果, 在试验或观测之前不能预知是哪个结果发生;
  - 一般还要求试验在相同条件下能够重复.
  - 比如: 观测把硬币抛 4 次后正面向上的次数; 观测某地的温度变化; 某电话总机单位时间内转接的电话次数.

样本点与样本空间

- 样本点 (Sample Point): 随机试验的每个可能结果称为一个样本点，常用ω表示.
- 样本空间 (Sample Space): 全体样本点构成的集合，常用Ω表示.
  - 一个试验的样本空间可能是有限的，可数无限的或不可数的;
  - 当样本空间为有限的时，我们可以将样本点可视化为圆点，做一次试验就等同于随机选取一个圆点;
  - 如果所有的圆点都是相同的，那么所有的圆点被抽中的可能性就相同，即后面将要介绍的古典概型.

事件

- 随机事件: 简称事件 (Event), 由部分样本点组成的试验结果
  - 通常指在随机试验中我们所关心的某一部分可能结果
  - 如果按照样本点可视化为圆点，则事件就是一些圆点的集合
- 基本事件: 仅含有一个样本点的随机事件称为基本事件, 它犹如分子中的原子, 在化学反应中不能再分, 所以有“基本”两字.
- 事件发生
  - 事件 A 发生: 事件 A 包含的样本点在随机试验结果中出现即 $\omega \in A$;
  - 通常把事件 A 所包含的样本点构成的集合与事件 A 视为等同, 也记为 A,
- 两个特殊的事件:
  - 必然事件(Ω): 在随机试验中一定会发生的事件;
  - 不可能事件($\emptyset$): 在随机试验中不可能发生的事件;

硬币抛 3 次的随机试验

- 样本点:正正正、正正反、正反正、反正正、正反反、反正反、反反正、反反反，共 8 种可能结果;
- 基本事件: {正正正}、{正正反}、{正反正}、{反正正}、{正反反}、{反正反}、{反反正}、{反反反}，共 8 个基本事件;
- 样本空间 Ω:

$$\Omega := \{正正正, 正正反, 正反正, 反正正, 正反反, 反正反, 反反正, 反反反\}$$

- 随机事件:

$$\begin{aligned} A &:= \{至少出现两次正面\} \\ &= \{正正正, 正正反, 正反正, 反正正\} \\ B &:= \{三次均为正面\} = \{正正正\} \\ \cdots \end{aligned}$$

样本空间示例

- 掷一枚骰子, 观察出现的点数. 则 $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- 考察某一地区的年降雨量, 则 $\Omega = \{x | 0 \leq x < T\}$, 这里 T 表示某个常数, 表示降雨量不会超过 T.
- 将一枚硬币抛两次, 考察正反面出现的情况, 则

$$\Omega = \{\text{正正, 正反, 反正, 反反}\}.$$

- 将一枚硬币抛两次, 考察正面出现的次数, 则 $\Omega = \{0, 1, 2\}$.
- 先后掷两次骰子, 考察两次骰子的点数, 则

$$\Omega = \{(a_1, a_2), a_i \in \{1, 2, \cdots, 6\}, i = 1, 2\}.$$

事件间的并:$A \cup B$

定义 2.1 (事件 A 与事件 B 的并) 记为 $A \cup B$, 其含义为: 事件 A 和事件 B 至少有一个发生, 即由事件 A 与 B 中所有的样本点组成的新事件.用数学符号来描述即为:

$$A \cup B := \{\omega \in \Omega : \omega \in A, \text{ 或} \omega \in B\}$$

例 2.1 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为奇数, 事件 B 为掷出的点数不超过 3, 则

$$A = \{1, 3, 5\}, \quad B = \{1, 2, 3\}, \quad A \cup B = \{1, 2, 3, 5\}$$

事件间的交:$A \cap B$

定义 2.2(事件 A 与事件 B 的交) 记为 $A \cap B$ 或 AB, 其含义为: 事件 A 和事件 B 同时发生, 即由事件 A 与 B 中公共的样本点组成的新事件.用数学符号来描述即为:

$$A \cap B := \{\omega \in \Omega : \omega \in A \text{ 且} \omega \in B\}$$

例 2.2 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为奇数, 事件 B 为掷出的点数不超过 3, 则

$$A = \{1, 3, 5\}, \quad B = \{1, 2, 3\}, \quad A \cap B = \{1, 3\}$$

事件的差:$A-B$

定义 2.3(事件 A 与 B 的差) 记为 $A-B$ 或 $A\backslash B$, 其含义为: 事件 A 发生且事件 B 不发生, 即由在 A 中而不在 B 中的样本点组成的新事件. 用数学符号来描述即为:

$$A-B:=\{\omega\in\Omega:\omega\in A\ 且\omega\notin B\}$$

例 2.3 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为奇数, 事件 B 为掷出的点数不超过 3, 则

$$A=\{1,3,5\},\quad B=\{1,2,3\},\quad A-B=\{5\}$$

对立事件:$\overline{A}$

定义 2.4 (事件 A 的对立事件) 记为 $\overline{A}$ 或 A^c, 其含义为: 事件 A 不发生, 即由在 Ω 中而不在 A 中的样本点组成的新事件.用数学符号来描述即为:

$$\overline{A} := \{\omega \in \Omega : \omega \notin A\} = \Omega - A$$

例 2.4 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为奇数, 则

$$A = \{1, 3, 5\}, \quad \overline{A} = \{2, 4, 6\}$$

对称差事件:$A\Delta B$

定义 2.5 (事件 A 与 B 的对称差事件) 记为 $A\Delta B$, 其含义为: 事件 A 与 B 中恰有一个事件发生.用数学符号来描述即为:

$$A\Delta B := \{\omega \in \Omega : \omega \in A \cup B \text{ 且} \omega \notin A \cap B\} = (A - B) \cup (B - A)$$

包含关系: $A \subset B$

定义 2.6 (包含关系) 称事件 A 包含于事件 B, 或事件 B 包含事件 A(记为$A \subset B$), 如果事件 A 发生必然导致事件 B 发生, 即属于 A 的样本点必然属于 B.

例 2.5 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为 2 或 4, 事件 B 为掷出的点数为偶数. 则显然有事件 A 发生必然导致事件 B 发生, 同时易知

$$A = \{2, 4\} \subset \{2, 4, 6\} = B$$

相等关系: $A = B$

定义 2.7 (相等关系) 称事件 A 与事件 B 相等 (记为$A = B$), 如果 $A \subset B$ 且 $B \subset A$, 即事情件 A, B 有相同的样本点.

例 2.6 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为 $2, 4, 6$, 事件 B 为掷出的点数为偶数. 则显然有

$$A \subset B, B \subset A,$$

同时易有

$$A = \{2, 4, 6\} = \{2, 4, 6\} = B$$

互不相容关系: $A \cap B = \emptyset$

定义 2.8 (互不相容关系) 称事件 A 与事件 B 互不相容, 如果事件 A, B 不能同时发生, 即事情件 A, B 没有相同的样本点.

例 2.7 考虑掷一枚骰子, 观察出现的点数. 记事件 A 为掷出的点数为奇数, 事件 B 为掷出的点数为偶数. 则显然有事件 A, B 互不相容, 并且

$$A \cap B = \{1, 3, 5\} \cap \{2, 4, 6\} = \emptyset$$

对立事件与互不相容事件

- $\overline{\overline{A}} = A, A \cap \overline{A} = \emptyset, A \cup \overline{A} = \Omega$;
- A, B 互为对立事件的充要条件为 $A \cap B = \emptyset, A \cup B = \Omega$;
- 对立事件一定是互不相容事件, 但互不相容事件不一定是对立事件;
- $A - B = A \cap \overline{B} = A\overline{B}$

多个事件的运算

- 有限并事件: $\bigcup_{k=1}^{n} A_k$
- 有限交事件: $\bigcap_{k=1}^{n} A_k$
- 可列并事件: $\bigcup_{k=1}^{\infty} A_k$
- 可列交事件: $\bigcap_{k=1}^{\infty} A_k$

事件运算的性质

(1) 交换律: $A \cup B = B \cup A, AB = BA$;

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

(3) 分配律: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$,
$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$;

(4) De Morgan 对偶法则:

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A}_i$$

$$\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A}_i$$

事件列的上下极限

定义 2.9 令 $A_n \subset \Omega, n \geq 1$ 为一列集合 (事件), 记

$$\inf_{k\geq n} A_k := \cap_{k=n}^{\infty} A_k, \quad \sup_{k\geq n} A_k := \cup_{k=n}^{\infty} A_k$$

则事件列 A_n 的上下极限定义如下:

$$\liminf_{n\to\infty} A_n := \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k,$$
$$\limsup_{n\to\infty} A_n := \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k.$$

若

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = A$$

则称 A 为事件列 $\{A_n, n \geq 1\}$ 的极限, 并记作 $\lim_{n\to\infty} A_n = A$.

单调事件列的极限

定义 2.10 事件列 $\{A_n\}$ 称为单调不减的，如果对任意的 n 均有 $A_n \subset A_{n+1}$；事件列 $\{A_n\}$ 称为单调不增的，如果对任意的 n 均有 $A_{n+1} \subset A_n$；

定理 2.1 假设 $\{A_n\}$ 是单调事件列，则

1. 如果 $\{A_n\}$ 是单调不减事件列，则 $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$
2. 如果 $\{A_n\}$ 是单调不增事件列，则 $\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n$

示性函数 (indicator function)

定义 2.11 设 $A \subset \Omega$, 我们定义事件 (集合)A 的示性函数如下

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

注 2.1

- $A = \{\omega : 1_A(\omega) = 1\}$;
- $1_A \leq 1_B \Leftrightarrow A \subset B$;
- $1_{A\cap B} = 1_A \cdot 1_B$;
- $1_{A\cup B} = 1_A + 1_B - 1_{A\cap B}$;

事件列上下极限的性质

引理 2.1 令 $\{A_n\}$ 为 Ω 的任一子集列，则

$$\begin{aligned}
\limsup_{n\to\infty} A_n &= \{\omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty\} \\
&= \{\omega : \omega\text{属于无穷多个}A_n\} \\
\liminf_{n\to\infty} A_n &= \{\omega : \sum_{n=1}^{\infty} 1_{\bar{A}_n}(\omega) < \infty\} \\
&= \{\omega : \text{除去有限多个}A_n\text{外}, \omega\text{属于余下所有的}A_n\}
\end{aligned}$$

证明: 由 $\omega \in \limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$ 知，对任意的 n，存在 $k_n \geq n$ 使得 $\omega \in A_{k_n}$，从而 $\sum_{n=1}^{\infty} 1_{A_n} \geq \sum_{n=1}^{\infty} 1_{A_{k_n}} = \infty$.

反之，若 $\omega \in \{\omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty\}$，则存在 $k_n \to \infty$ 使得 $\omega \in A_{k_n}$，从而对任意的 n，$\omega \in \cup_{k=n}^{\infty} A_k$，即

$$\omega \in \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k = \limsup_{n\to\infty} A_n.$$

事件运算表

符　号	集合论意义	概率论意义
Ω	全集或空间	样本空间、必然事件
$\varnothing$	空集	不可能事件
ω	元素	样本点、基本事件
A	可测子集	随机事件
$\omega \in A$	ω 是 A 的元素	试验结果 ω 属于 A, 事件 A 发生
$A \subset B$	A 包含在 B 中	若 A 发生, 则 B 一定发生; 事件 A 蕴涵事件 B
$A = B$	A 与 B 相等	A 与 B 为同一事件; 它们同时发生或同时不发生
$A \cap B$ 或 AB	交集	表示 A 与 B 同时发生
$AB = \varnothing$	不相交	A 与 B 不相容 (互斥), 它们不可能同时发生
$A \cup B$	并集	表示 A 或 B 发生, A 与 B 中至少有一个发生
A^c	余集	对立事件; A^c 发生表示 A 不发生
$A - B$ 或 AB^c	差集	表示 A 发生, 而 B 不发生
$A \triangle B$	对称差	表示 A 与 B 中恰有一个发生
$\limsup\limits_{n\to\infty} A_n = \bigcap\limits_{k=1}^{\infty} \bigcup\limits_{n=k}^{\infty} A_n$	上极限集合	表示事件序列 $\{A_n\}$ 中有无穷多个事件发生
$\liminf\limits_{n\to\infty} A_n = \bigcup\limits_{k=1}^{\infty} \bigcap\limits_{n=k}^{\infty} A_n$	下极限集合	表示事件序列 $\{A_n\}$ 中至多有有限个事件不发生

例 2.8 设 A, B, C 为某个随机现象的三个事件，试用事件的交，差，并，对立事件表示下列事件

- 事件 A, B 发生而 C 不发生:$AB\overline{C}$;
- 事件 A, B, C 中至少有一个发生:$A \cup B \cup C$;
- 事件 A, B, C 中至少有两个发生: $AB \cup AC \cup BC$;
- 事件 A, B, C 中恰好有两个发生: $AB\overline{C} \cup A\overline{B}C \cup \overline{A}BC$;
- 事件 A, B, C 同时发生: ABC;
- 事件 A, B, C 都不发生: $\overline{A} \cap \overline{B} \cap \overline{C}$;
- 事件 A, B, C 不全发生: $\overline{A} \cup \overline{B} \cup \overline{C}$;

例 2.9 (掷硬币) 掷一枚硬币 10 次, 正面用 H 表示反面用 T 表示.

- 一个可能的结果为 HHHTHHTTHT, 且样本空间就是所有长度为 10 的由 H 和 T 组成的字符串的集合
- 若令 H 为 1，T 为 0 , 则一个结果可用数列 $(s_1, s_2, \cdots, s_{10})$ 表示, 其中 $s_j \in \{0, 1\}$, 样本空间即为这些数列的集合
- 下面我们看几个事件:
  - A_i: 第 i 次掷硬币结果为 H，则

$$A_1 = \{(1, s_2, \cdots, s_{10}) : s_j \in \{1, 0\} \text{ 且} 2 \leqslant j \leqslant 10\}.$$

  - B: 至少有一次掷硬币结果为 H, 则$B = \bigcup_{j=1}^{10} A_j$.
  - C: 所有掷硬币结果都为 H, 则$C = \bigcap_{j=1}^{10} A_j$.
  - D: 至少有两个连续的 H, 则$D = \bigcup_{j=1}^{9} (A_j \cap A_{j+1})$.

第三讲：古典概型与几何概型

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

古典概型

定义 2.12 (古典概型:) 有两个条件,

(1) (有限性) 试验结果只有有限个 (记为 n);

(2) (等可能性) 每个基本事件发生的可能性相同.

对于古典概型: $\Omega = \{\omega_1, \cdots, \omega_n\}, \mathcal{F} = \{A : A \subset \Omega\}$, 事件 $A \in \mathcal{F}$ 的概率定义为

$$P(A) = \frac{|A|}{|\Omega|}, \ \forall A \in \mathcal{F}$$

其中, $|B|$ 表示事件 B 所包含的基本事件的个数.

古典概型的两个重要方面

- 确定样本空间与样本点

  例 2.10 掷两个均匀硬币，出现正、反两面的概率是多少？

  - 数学家 d'Alembert(达朗贝尔) 认为掷两个均匀硬币，出现的结果无外乎" 正正"," 反反"," 正反"，因此概率应是 $\frac{1}{3}$;
  - 事实上，该试验样本空间为

$$\Omega = \{(\text{正}, \text{正}), (\text{正}, \text{反}), (\text{反}, \text{正}), (\text{反}, \text{反})\}$$

  - 若硬币不可区分，则样本空间为 $\Omega = \{\{\text{正}, \text{正}\}, \{\text{正}, \text{反}\}, \{\text{反}, \text{反}\}\}$，但样本空间中的样本点并不是等可能出现，所以也不能说出现正、反两面的概率是 $\frac{1}{3}$.

  例 2.11 (Leibnitz 的错误) 掷两个均匀骰子，得到点数和为 11 与 12 的概率相同吗？

- 如何计数：见下面计数原理.

计数原理

加法原理: 考虑试验 A 和 B，若试验 A 有 a 种可能结果，试验 B 有 b 种可能结果，则试验 A 或 B 共有 $a+b$ 种结果

乘法原理 [1]: 考虑一个复合试验，它由两个子试验 A 和 B 构成。假设试验 A 有 a 种可能结果，每一种情况又都对应于试验 B 的 b 种可能结果。那么这个复合试验共有 ab 种可能结果。

[1]通常很容易想当然地认为试验是有先后顺序的，但是乘法原理并没有要求试验 A 必须在试验 B 之前实施。

冰淇淋甜筒购方案计数

例 2.12 (冰淇淋甜筒) 假没某人正在购买一支冰淇淋甜筒，甜筒有蛋卷或华夫饼两种选择，冰淇淋有巧克力 (C)、草莓 (S) 和香草 (V) 三种口味可选。考虑以下问题:

- 该随机试验共有 6 种可能购买方案：无论是先决定甜筒类型还是冰淇淋口味类型，试验都是总共有 $2 \times 3 = 3 \times 2 = 6$ 种可能
- 假设怀特先生某天上下午各买一支冰淇淋甜筒，则该试验共有 $6^2 = 36$ 种可能结果
- 如果只关注这天吃了何种类型的冰淇淋，而不在乎吃冰淇淋的顺序，即 $(cakeC, waffleV)$ 和 $(waffleV, cakeC)$ 之间没有区别，则共有

$$6 * 5/2 + 6 = 21$$

排列组合

- 从 n 个不同的元素中，有放回的取出 r 个元素组成的排列种数为 n^r 种;
- 从 n 个不同的元素中，不放回的取出 r 个元素组成的排列种数为

$$P_n^r = n(n-1)\cdots(n-r+1);$$

- 从 n 个不同的元素中，不放回的取出 r 个元素组成的组合，种数为

$$C_n^r = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!};$$

- 从 n 个不同的元素中，有放回的取出 r 个元素组成的组合，种数为 [2]

$$C_{n+r-1}^r.$$

[2] 又称 Bose-Einstein 值，物理学家玻色和爱因斯坦 1920 年做了不可区分粒子的相关研究，成功预测了玻色 - 爱因斯坦凝聚态陌生物质状态的存在

生日问题

例 2.13 若屋子里有 k 个人，假设每个人的生日都等可能为一年 365 天中的一天 (不考虑 2 月 29 日的情况)，而且每个人的生日是相互独立的 (假设这里不存在双胞胎)。试问：屋内有两个及以上的人在同一天过生日的概率是多少？

解： k 个人中没有人生日相同的概率为

$$P(\text{没有人生日相同}) = \frac{365 \cdot 364 \cdots (365-k+1)}{365^k}$$

故至少有两个人生日相同的概率为

$$P(\text{至少有两个人生日相同}) = 1 - \frac{365 \cdot 364 \cdots (365-k+1)}{365^k}$$

可以绘制至少两个人生日相同的概率随 k 变化的函数图。概率首次超过 0.5 是在 $k = 23$, 当 $k = 57$ 时，这个概率超过了 99%.

例 2.14 甲乙丙丁四人进行乒乓球双打练习，两人一对地结为对打的双方，有多少种不同的结对方式?

解: 简单的组合问题: 从四人中选出两人结为一对，剩下的两人结为一对即可。于是可得:

$$\text{有 } C_4^2 = 6 \text{ 种方式. ????}$$

但事实是否如此呢? 注意此时并不要求两队之间的顺序，所以

$$\text{只有 3 种结对方式: } C_4^2/2 = 3.$$

注 2.2

- 在按组合模式分出的“组内”, 元素之间是没有“顺序”的;
- 但是需要指出的是：在“组”之间却存在着“顺序”, 或者叫做“编号”!
- 在按“组合”模式计算时，我们计算的是“取出两个人”的所有不同取法数目.
- 假如我们把取出的两人算一组，而把留下的两人算另一组，那么由于“取出甲乙，留下丙丁”和“取出丙丁，留下甲乙”是两种不同的取出方式，而在这种计算方法中，被算作两种不同的“分组”方式，从而得到 6 种“分组”方式.
- “组合”是一种“有编号的分组模式”，或者说，按照组合模式计算出的分组方式数目中，已经天然地把组的不同编号方式数目计算在内了.

例 2.15 欲将 6 个人分为 3 组，每组 2 人，分别从事 3 项不同工作，求分配方式数.

解: 先取出两人从事第 1 项工作，有 C_6^2 种方式；再取出两人从事第 2 项工作，有 C_4^2 种方式；剩下的两人从事第 3 项工作。所以一共有:

$$C_6^2 \cdot C_4^2 = \frac{6!}{4! \cdot 2!} \frac{4!}{2! \cdot 2!} = \frac{6!}{2! \cdot 2! \cdot 2!} = 90$$

种分配方式.

例 2.16 要把 7 人分为 3 个小组，执行同一种任务，其中一个组 3 人，另两个组各 2 人，求分组方式数.

解: 显然这也是一个“无编号分组”问题。但是却与上面的情况有所不同。因为其中有一个 3 人组，无论是否编号，它都与其余两个组有所区别 (编号无非是为了对分出的组加以区分), 所以在按“有编号分组模式”算出分组方式数之后，只应再除以 2! (即除去两个不加区分的组的排列顺序数), 故得：共有

$$\frac{7!}{3!\cdot 2!\cdot 2!}\cdot\frac{1}{2!}=\frac{7!}{3!\cdot(2!)^3}$$

种分组方式. 为了适应这种分为多个“不同的”组的问题需求，人们总结出如下的“多组组合模式”：

多组组合与排列模式

- 多组组合模式: 有 n 个不同元素，要把它们分为 k 个不同的组，使得各组依次有 $n_1, n_2, \cdots, n_k$ 个元素，其中 $n_1 + n_2 + \cdots + n_k = n$, 则一共有

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

种不同分法.

- 不尽相异元素的排列模式: 有 n 个元素，属于 k 个不同的类，同类元素之间不可辨认，各类元素分别有 $n_1, n_2, \cdots, n_k$ 个，其中 $n_1 + n_2 + \cdots + n_k = n$, 要把它们排成一列，则一共有

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

种不同排法.

例 2.17 一批产品有 N 个，其中废品有 M 个。现从中随机取出 n 个，在以下两种情形下，分别求“其中恰好有 m 个废品”这一事件的概率。(1) 有放回地选取；(2) 不放回地选取.

解： 记 $A=\{$其中恰好有m个废品$\}$, 则

- 有放回情形: $|\Omega|=N^n, |A|=C_n^m M^m (N-M)^{n-m}$, 所以

$$P(A)=\frac{C_n^m M^m (N-M)^{n-m}}{N^n}=C_n^m\left(\frac{M}{N}\right)^m\left(\frac{N-M}{N}\right)^{n-m}$$

- 不放回情形: $|\Omega|=C_N^n, |A|=C_M^m C_{N-M}^{n-m}$, 所以

$$P(A)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

例 2.18 一个笼子里关着 7 只白猫和 3 只黑猫共 10 只猫，把笼门打开每次只允许钻出 1 只猫。若 10 只猫均钻出笼子，并以 A_k 表示事件“第 k 只出笼的猫是黑猫”，试求 $P(A_k)$，其中，$k = 1, 2, ..., 10$.

方法一：将 10 只猫看成不同的猫

- $|\Omega| = 10!$
- $|A_k| = C_3^1 \cdot 9!$
- $P(A_k) = \dfrac{|A_k|}{|\Omega|} = \dfrac{3 \cdot 9!}{10!} = \dfrac{3}{10}, \quad k = 1, 2, \cdots, 10.$

方法二：仅白猫与黑猫可区别

- 现在只有两种不同的元素，一种有 7 个 (7 只白猫), 另一种有 3 个 (3 只黑猫), 由不尽相异元素的排列模式知

$$|\Omega| = C_{10}^{3} = 120$$

- $|A_k| = C_9^2 = 36$
- $P(A_k) = \frac{36}{120} = \frac{3}{10}, \quad k = 1, 2, \cdots, 10.$

方法三：仅考虑前 k 只出笼猫的排列情况

- $|\Omega| = C_{10}^{k} \cdot k! = \dfrac{10!}{(10-k)!}$
- $|A_k| = C_3^1 \cdot C_9^{k-1} \cdot (k-1)! = \dfrac{3 \cdot 9!}{(10-k)!}$
- $P(A_k) = \dfrac{3 \cdot 9!}{10!} = \dfrac{3}{10}, \quad k = 1, 2, \cdots, 10$

“球-盒”计数问题

- $S(n,k)$: 第二类 Stirling 数, 表示将 n 个元素的集合分解为 k 个非空子集的分法数目;
- 不难验证 $S(n+1,k)=kS(n,k)+S(n,k-1)$, 并可以证明 (枚举空盒的个数，剩下的随便放，盒子是相同的最后要除以 $m!$)

$$S(n,k)=\frac{1}{k!}\sum_{j=0}^{k}(-1)^j C_k^j (k-j)^n$$

- $p(n,k)$: 表示将 n 分解为 k 个正整数之和的分法数目. ($p(n,k)$ 的计算相当复杂, 涉及整数以及集合分拆的问题, 有兴趣的可参阅组合计数方面的文献).

n 个球 r 个盒子	盒可区分		盒不可区分	
	允许空盒	不允许空盒 $n>r$	允许空盒	不允许空盒 $n>r$
球可区分	r^n	$r!S(n,r)$	$\sum_{k=1}^{r}S(n,k)$	$S(n,r)$
球不可区分	C_{n+r-1}^{n}	$C_{(n-r)+r-1}^{n-r}$	$\sum_{k=1}^{r}p(n,r)$	$p(n,r)$

Maxwell-Boltzmann(麦克斯韦-玻尔兹曼) 统计

- 考查热平衡系统以及其中的粒子，在温度充分高且分布的密度足够低，粒子间的量子效应可以忽略.
- 假定能态被标记为 $\{1, 2, \cdots, K\}$，粒子的总数为 N, 那么恰有 N_i 个粒子处于能态 i 上的概率为

$$P_{\mathrm{M-B}} = K^{-N} \frac{N!}{N_1! N_2! \cdots N_K!}, \quad N_1 + \cdots + N_K = N.$$

- Maxwell-Boltzmann 统计最重要的假设是粒子的可区分性.
- 该假设条件下粒子 A 处于能态 1 且粒子 B 处于能态 2，与粒子 A 处于能态 2 且粒子 B 处于能态 1 是不同的两种系统状态.
- 该假设所导出的粒子的能态分布 (即 Boltzmann 分布) 较为符合物理事实.
- 但是其在熵方面却导致了和物理事实严重违背的结果 (所谓的 Gibbs 悖论), 而一旦假设粒子不可区分, 则矛盾会迎刃而解.

Bose-Einstein(博兹-爱因斯坦) 统计

- 当温度下降, 且粒子的空间密度上升时, 粒子间的量子效应变得显著而不可忽略, 粒子在能态间的分布也随之发生变化.
- 根据表现的不同, 可将粒子分为两类: 玻色子 (Boson) 和费米子 (Fermion).
- 无穷多个玻色子能够在同一时间处于 (" 凝聚") 同一能态上, 从而出现统计物理学中著名的 Bose-Einstein 凝聚 (Bose-Einstein condensate) 现象.
- 假定粒子总数为 N 但粒子不可区分, 能态数目为 K, 则玻骰子在各能态上的分布服从 Bose-Einstein 统计,

$$P_{\mathrm{B-E}} = \frac{1}{C_{N+K-1}^{N}} = \binom{N+K-1}{N}^{-1}$$

- Bose-Einstein 统计的粒子能态分布情况与 N 个不可区分的球放入 K 个可区分的盒中的放法形成一一对应.

Fermi-Dirac(费米-狄利克雷) 统计

- 费米子遵守量子力学中的 Pauli 不相容原则，即不同的粒子无法处在相同的能态上 (注意玻色子是不满足这一原则的).
- 因此, 系统中的粒子数目一定比能态数目少, 即 $N < K$.
- 费米子的能态服从 Fermi-Dirac 分布,

$$P_{\mathrm{F-D}} = \frac{1}{C_K^N} = \binom{K}{N}^{-1}$$

- 该分布对应于 N 个不可区分的球放入 K 个可区分的盒中, 且每个盒里不能多于两球的放法数目.
- Fermi-Dirac 统计在帮助人们理解电子的行为方面起到了非常关键的作用.

几何概型的基本思想

1. 可度量性。样本空间 Ω 充满某个区域，其度量 (长度、面积、体积) 大小可用 S_Ω 表示;
2. 等可能性。任意一点落在度量相等的子区域内是等可能的，即任一子区域 A 的概率，只与子区域的度量 S_A 有关，而与子区域的位置无关;
3. 若事件 A 为 Ω 的某个子区域，且其度量大小可用 S_A 表示，则事件 A 的概率为

$$P(A) = \frac{S_A}{S_\Omega}$$

例 2.19 蒲丰投针问题: 平面上画有间隔为 d 的等距平行线，向平面任意投掷一枚长为 l 的针，求针与平行线相交的概率.

解: 注意到

$$\begin{aligned}\Omega &= \{(\varphi, x) : 0 \le x \le d/2, 0 \le \varphi \le \pi\}\\ A &= \{\text{针与平行线相交}\}\\ &= \{(\varphi, x) : x \le \frac{l}{2}\sin\varphi\}\end{aligned}$$

由几何概型知

$$P(A) = \frac{S_A}{S_\Omega} = \frac{\int_0^\pi \frac{l}{2}\sin\varphi d\varphi}{\frac{d}{2}\pi} = \frac{2l}{d\pi}.$$

随机模拟

- l, d 已知，将 π 的值代入即可计算 $P(A)$ 的值;
- 若我们以频率近似概率 $P(A)$, 即

$$\frac{n}{N} \approx P(A) = \frac{2l}{d\pi}$$

则我们可得到 π 的近似值

$$\pi \approx \frac{2lN}{dn}$$

表 1: 历史试验数据

试验者	年份	l/d	投掷次数	数交次数	π 的近似值
Wolf	1850	0.8	5000	2532	3.1596
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415
Reina	1925	0.54	2520	859	3.1795

会面问题

例 2.20 甲乙约定在下午 6 点到 7 点之间在某处会面，并约定先到者等候另一个人 20 分钟，过时即可离去，求两人能会面的概率.

解: 以 x 和 y 分别表示甲乙两人到达约地点的时间（以分钟为单位），则易知:

$$\begin{aligned}\Omega &:= \{(x,y): 0 \le x \le 60, \quad 0 \le y \le 60\}\\ A &:= \{(x,y): |x-y| \le 20\}\end{aligned}$$

故

$$P(A) = \frac{S_A}{S_\Omega} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

例 2.21 在一圆内任取一条弦，问其长度超过该圆内接等边三角形的边长的概率是多少？

解: 易知，圆内接正三角形的边长 $\sqrt{3}R$

1. 注意到弦的长度与它与圆心的距离有关，而与弦的方向无关。显然，当且仅当弦与圆心的距离小于 $R/2$ 时，弦的长超过 $\sqrt{3}R$, 故概率为 $1/2$;
2. 把弦的一端固定，另一端点在圆周上随机移动。若在固定端点作一切线，则与切线交角在 60° 与 120° 之间的弦才能超过 $\sqrt{3}R$. 故所求概率为 $1/3$;
3. 弦长被其中点位置决定，则当且仅当弦的中点落在半径为 $R/2$ 的小圆内时，弦的长度才会超过 $\sqrt{3}R$, 故所求概率为 $1/4$.

第四讲: 黎曼积分、Lebesgue 积分、测度与概率

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

黎曼积分

- 函数 $f\colon [a,b] \mapsto \mathbb{R}$;
- 分割 Δ: $a = x_0 < x_1 < \cdots < x_n = b$;
- 黎曼和:

$$I(f,\Delta) := \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i),\ \text{其中}\ \xi_i \in [x_i, x_{i+1}];$$

- 令 $|\Delta| := \max_{1\le i\le n} |x_i - x_{i-1}|$, 则若

$$\lim_{|\Delta|\to 0} I(f,\Delta) = \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i) = L := \int_a^b f(x)dx,$$

则称 L 为函数 f 在区间 $[a,b]$ 的积分, 并称函数 f 是黎曼可积的.

黎曼积分示意图

黎曼积分

黎曼积分的缺点

- 一些很简单的函数不是黎曼可积的, 如 Dirichlet 函数:

$$D(x) := \begin{cases} 1, & x\text{是有理数}, \\ 0, & x\text{是无理数}. \end{cases}$$

- 需要一致收敛条件来保证函数积分与极限运算的次序交换, 即若 $f_n(x)$ 一致收敛且黎曼可积, 则

$$\lim_{n\to\infty}\int_a^b f_n(x)dx = \int_a^b \lim_{n\to\infty} f_n(x)dx$$

- 黎曼可积函数空间不完备, 即存在黎曼可积函数列 $f_n(x)$ 逐点收敛到函数 $f(x)$, 但 $f(x)$ 不是黎曼可积.

黎曼积分的等价定义 (I)

- 考虑函数 $f\colon [a,b] \mapsto \mathbb{R}$ 及分割 Δ: $a = x_0 < x_1 < \cdots < x_n = b$, 并令 $|\Delta| := \max_{1\le i\le n} |x_i - x_{i-1}|$ 以及

$$M_i(\Delta) := \sup_{x_{i-1}\le x\le x_i} f(x),\ m_i(\Delta) := \inf_{x_{i-1}\le x\le x_i} f(x),$$

$$M(\Delta) := \sum_{i=0}^{n-1} M_i(x_{i+1}-x_i),\ m(\Delta) := \sum_{i=0}^{n-1} m_i(x_{i+1}-x_i).$$

- 上下积分:

$$\overline{\int_a^b} f(x)dx := \lim_{|\Delta|\to 0} M(\Delta) = \inf_{\Delta} M(\Delta),$$

$$\underline{\int_a^b} f(x)dx := \lim_{|\Delta|\to 0} m(\Delta) = \sup_{\Delta} m(\Delta)$$

- $f(x)$ 黎曼可积当且仅当上下积分相等, 即 $\overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx$

黎曼积分的等价定义 (II)

定义 2.13 (分段常值函数或阶梯函数) 函数 $f\colon [a,b] \mapsto \mathbb{R}$ 称为分段常值函数或阶梯函数, 如果区间 $[a,b]$ 存在一个有限区间分割 $I_1, I_2, \cdots, I_n$ 使得 $f(x)$ 在每个区间上 I_i 上恒为常数 c_i, 即

$$f(x) = \sum_{i=1}^{n} c_i 1_{I_i}(x)$$

注 2.3

- 容易验证, $\sum_{i=1}^{n} c_i |I_i|$ 与分段常值函数或阶梯函数 $f(x)$ 表达形式中所选择的区间分割无关.
- 我们记

$$p.c. \int_a^b f(x)dx := \sum_{i=1}^{n} c_i |I_i|$$

并将其称为分段常值函数或阶梯函数 $f(x)$ 在 $[a,b]$ 上的积分.

黎曼积分的等价定义 (II)

定义 2.14 (Darboux integral 达布积分) 设函数 $f\colon [a,b] \mapsto \mathbb{R}$ 为有界函数, 则

- 称
$$\underline{\int_a^b} f(x)dx := \sup_{g\le f,\ g\text{为阶梯函数}} p.c.\int_a^b g(x)dx$$
为函数 $f(x)$ 在 $[a,b]$ 上的达布下积分;
- 称
$$\overline{\int_a^b} f(x)dx := \inf_{h\ge f,\ h\text{为阶梯函数}} p.c.\int_a^b h(x)dx$$
为函数 $f(x)$ 在 $[a,b]$ 上的达布上积分;
- 显然, $\underline{\int_a^b} f(x)dx \le \overline{\int_a^b} f(x)dx$. 如果
$$\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx,$$
则称函数 $f(x)$ 在 $[a,b]$ 上达布可积.

黎曼积分的等价定义 (II)

定理 2.2 设函数 $f\colon [a,b] \mapsto \mathbb{R}$ 为有界函数, 则函数 $f(x)$ 黎曼可积当且仅当函数 $f(x)$ 达布可积.

黎曼积分的本质

- 黎曼积分思想源于古希腊的“穷竭法”.
  - 求圆的面积: 通过圆外切正 n 边形和内接正 n 边形, 来得到圆面积的上下界, 然后取极限发现上下界都相等, 从而得到圆的面积.
  - 求函数 $f(x)$ 围成的面积: 把 $f(x)$ 与坐标轴围成的区域通过对定义域 $[a,b]$ 进行有限分割并利用分割小区间上函数 $f(x)$ 的最大最小值构造函数外接与函数内接长方形, 然后取分割的极限得到函数 $f(x)$ 围成的面积.
  - 黎曼积分整个过程就是: 有限分割–近似求和–取极限.
- 对函数定义域进行有限区间分割后近似求和取极限, 适用于连续函数, 即函数在小区间内振荡不明显, 但对然而对于振荡很厉害的函数, 哪怕分割很细, 在很细的区间内振荡依然很厉害 (典型的例子是狄利克雷函数).

Lebesgue 积分基本思想

- 黎曼积分是对函数的定义域 $[a,b]$ 进行分割，而 Lebesgue 积分则是对函数 $f(x)$ 的值域进行分割.
- 令 $l<L$ 分别为 $f(x)$ 在区间 $[a,b]$ 上的下确界与上确界，考虑 $[l,L]$ 的分割 Δ_y:

$$l=l_0<l_1<l_2<\cdots<l_n=L$$

Lebesgue 积分基本思想

- 令

$$|\Delta_y| := \max_{1\le k\le n} |l_k - l_{k-1}|,$$
$$E_k := \{x : l_k \le f(x) < l_{k+1}\}, k = 0, 1, \cdots, n-1,$$

- 显然

$$\sum_{k=0}^{n-1} l_k 1_{E_k}(x) \le f(x) < \sum_{k=0}^{n-1} l_{k+1} 1_{E_k}(x), \forall x \in [a, b].$$

因此, 相应的积分应保持上述不等关系, 即

$$\sum_{k=0}^{n-1} l_k \int_a^b 1_{E_k}(x)dx \le \int_a^b f(x)dx < \sum_{k=0}^{n-1} l_{k+1} \int_a^b 1_{E_k}(x)dx.$$

- 类似于黎曼积分的想法, 我们希望当值域分割 $|\Delta_y| \to 0$ 时, 上述左右两端的极限存在且相等.

Lebesgue 积分基本思想

- 若对每一个 $E_k, k=0,\cdots,n-1$, 示性函数 $1_{E_k}(x)$ 黎曼可积, 且 $|\Delta_y|<\epsilon$, 则

$$\sum_{k=0}^{n-1}(l_{k+1}-l_k)\int_a^b 1_{E_k}(x)dx \le \epsilon\sum_{k=0}^{n-1}\int_a^b 1_{E_k}(x)dx = \epsilon(b-a).$$

- 若示性函数 $1_{E_k}(x)$ 不是黎曼可积的, 如何定义:

$$m(E_k) := \int_a^b 1_{E_k}(x)dx,$$

也即集合 E_k 的测度 $m(E_k)$ 如何定义?

- 若我们可以定义 $m(E_k)$, 则我们可以类似于“黎曼和”一样, 我们构造“勒贝格和”, 用面积已知的一些区域逼近曲线下方的区域:

$$\int_a^b f(x)dm := \lim_{|\Delta_y|\to 0}\Big[\sum_{k=0}^{n-1} l_k\cdot m(E_k)\Big].$$

黎曼积与与 Lebesgue 积分的直观解释

- 想像一位零售商, 在一天结束营业时汇总今天的营业收入.
  - 黎曼积分法: 从钱箱中依次取出现金累加计算获得总的营业收入;
  - Lebesgue 积分法: 先将钱箱中的现金按面值进行分类, 用面值乘以该面值现金的数量并累加求和获得总的营业收入.

如何测量集合大小

- 基数或势 (Cardinality): 集合元素的个数;
- 称两个集合的基数相等, 如果两个集合之间存在一一映射;
- $A = \{2, 4, 6, \cdots, \}$ 与 $B = \{1, 2, 3, 4, \cdots\}$ 的基数是相等的;
- 基数或势的优缺点:
  - 在有限集上, 基数还是很有用处的;
  - 但对于无限集情形, 这个度量显得过于粗糙, 例如, 实数集上任意两个区间的基数都是相同的;
- 对于连续的集合, 我们又有自然的 "体积" 的概念:
  一维空间: 长度; 二维空间: 面积; 三维空间: 传统的体积;

可数集与不可数集

- 称集合是可数 (countable) 的, 如果该集合和自然数集 $\mathbf{N}$ 的子集之间存在一一映射, 否则称该集合是不可数 (uncountable) 集.
- 可数集包括有限可数集和无穷可数集.
- 有限个可数集的并仍然是可数集, 可数个可数集的并仍然是可数集.
- 两个可数集的笛卡尔积是可数集.
- 实数区间 $(0,1)$ 是典型的不可数集.
  假设 $(0,1)$ 可数, 其元素可以排列为 $a_1, a_2, a_3, \cdots$. 注意到 $(0,1)$ 中的元素 a_n 可以表示为

$$a_n = 0.a_n^{(1)} a_n^{(2)} a_n^{(3)} \cdots = \sum_{k=1}^{\infty} a_n^{(k)} 10^{-k}$$

  构造 $x = 0.x_1 x_2 x_3 \cdots$, 其中, 对任意的 k, 取 $x_k \neq a_k^{(k)}$. 显然, $x \neq a_k$, $\forall k \in \mathbf{N}$, 矛盾. 故 $(0,1)$ 是不可数集合. 由于

$$f(x) = -\cot(\pi x) : (0,1) \to (-\infty, \infty)$$

  是一一映射, 立刻得到 $\mathbf{R}$ 也是不可数集.

集合的测度

- 对于集合大小, 我们希望定义另外的度量标准: 测度.
  - 希望定义一个测度来测量集合的大小, 能够处理各种情况, 包括有限集、无限集甚至是各种抽象的集合;
  - 希望在某些特定的情况下, 和我们熟知的度量相容: 比如区间的长度, 区域的面积, 空间的体积等度量相容;
  - 希望具有常见度量的一些性质: 可加性.
- 如此一来, 设计一个测度就好像变成了制造一把尺子 (或者量筒: 多维情形).
- 尺子有其适用范围: 用一个普通米尺来测量地球的半径估计会很有困难, 或者用它去测量原子的半径也是不可能的.
- 测度的适用范围在测度论里同样重要: 在定义一个测度的时候, 我们需要同时定义哪些集合是 “可测的”(measurable) – 也就是可以用这个测度来衡量.

区间、矩形、基本集及其测度

- 区间 $I \subset \mathbb{R}$: $[a,b]$, $[a,b)$, (a,b), $\cdots$;
- 矩形 $B \subset \mathbb{R}^d$: $B := I_1 \times \cdots \times I_d$, 其中 I_i 为区间;
- 矩形的测度: $|B| := |I_1| \times |I_2| \times \cdots \times |I_d|$
- 基本集 $E \subset \mathbb{R}^d$: 能够表示为有限个矩形的并的集合;
  - 更进一步, 可以证明基本集一定可以表示为有限个不相交矩形的并, 即
    $$E = \cup_{i=1}^k B_i, \text{其中} B_i, i = 1, \cdots, k, \text{为矩形且互不相交};$$
  - 我们将 $m(E) := |B_1| + |B_2| + \cdots + |B_k|$ 称为基本集 E 的基本测度.

约当测度 (Jordan Measure)

定义 2.15 (约当测度) 令 E 为 $\mathbb{R}^d$ 的有界子集.

- 称
$$m_{*,J}(E) := \sup_{A\subset E, A\text{为基本集}} m(A)$$
为约当内测度;
- 称
$$m^{*,J}(E) := \inf_{A\supset E, A\text{为基本集}} m(A)$$
为约当外测度;
- 称 E 为约当可测的, 如果 $m_{*,J}(E) = m^{*,J}(E)$. 此时称 $m(E) := m_{*,J}(E) = m^{*,J}(E)$ 为集合 E 的约当测度.

约当测度的优缺点

设 E, $F \subset \mathbb{R}^d$ 为约当可测集, 则

- $E \cup F$, $E \cap F$, $E \backslash F$, $E \Delta F$ 均为约当可测集;
- 非负性: $m(E) \geq 0$;
- 有限可加性: 如果 E, F 不相交, 则 $m(E \cup F) = m(E) + m(F)$;
- 平称不变性: $m(x + E) = m(E)$;
- $E_i, i = 1, 2, \cdots$ 约当可测, 但 $\cup_{i=1}^{\infty} E_i$, $\cap_{i=1}^{\infty} E_i$ 不一定约当可测

Riemann 积分与 Jordan 测度之间的联系

定理 2.3 (Riemann 积分与 Jordan 测度之间的联系)

- 如果 E 是区间 $[a,b]$ 上的 Jordan 可测集, 则示性函数

$$1_E(x) := \begin{cases} 1, & x \in E \\ 0, & x \notin E \end{cases}$$

是 Riemann 可积的且 $\int_a^b 1_E(x)\,dx = m(E)$.

- 设 $[a,b]$ 为一区间, $f\colon [a,b] \to \mathbb{R}$ 为一有界函数, 则 f 是 Riemann 可积当且仅当

$$E_+ := \{(x,t) : x \in [a,b], 0 \le f(x) \le t\} \tag{1}$$

$$E_- := \{(x,t) : x \in [a,b], f(x) \le t \le 0\} \tag{2}$$

在 $\mathbb{R}^2$ 上是 Jordan 可测的, 且有

$$\int_a^b f(x)\,dx = m_2(E_+) - m_2(E_-)$$

此处的 m_2 表示二维 Jordan 测度.

Lebesgue 测度

- Lebesgue 外测度 $m^*(E)$:

$$m^*(E) := \inf_{\cup_{i=1}^{\infty} B_i \supset E, B_i, i \geq 1, \text{为矩形}} \sum_{i=1}^{\infty} |B_i|$$

- 显然, $m^*(E) \leq m^{*,J}(E)$, 且存在 $E \subset \mathbb{R}^d$ 使得不等号严格成立;
- Lebesgue 可测集的卡拉西奥多里 (Carathéodory) 准则:

称 $E \subset \mathbb{R}^d$ 为 Lebesgue 可测集, 如果对任意的 $S \subset \mathbb{R}^d$ 均有

$$m^*(S) = m^*(S \cap E) + m^*(S \backslash E),$$

并称 $m(E) := m^*(E)$ 为集合 E 的Lebesgue 测度.

Lebesgue 可测集的代数结构

定义 2.16 (事件或集合的 Borel 运算) 我们将对所给出的一些事件或集合所作的各种 (有限次或可列次) 取余、取交和取并运算以及它们的混合运算都称为 事件或集合的 Borel 运算。

定义 2.17 ($\sigma-$ 代数) 设 $\mathcal{F}$ 为 $\mathbb{R}^d$ 的某些子集构成的集合类, 称 $\mathcal{F}$ 为 $\mathbb{R}^d$ 上的 σ-代数, 如果

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

(2) 若 $A \in \mathcal{F}$, 则 $\overline{A} := \mathbb{R}^d - A \in \mathcal{F}$;

(3) 若 $A_i \in \mathcal{F}, i = 1, 2, \cdots,$, 则 $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

命题 2.1 在 $\mathbb{R}^d$ 上的 Lebesgue 可测集具有以下性质:

- 若 E, $F \subset \mathbb{R}^d$ Lebesgue 可测, 则 $E \cup F$, $E \cap F$, $E \backslash F$, $E \Delta F$ 均 Lebesgue 可测;
- 若 $E_i, i \geq 1$,Lebesgue 可测, 则 $\cup_{i=1}^{\infty} E_i$, $\cap_{i=1}^{\infty} E_i$ 也 Lebesgue 可测.

Lebesgue 测度的性质

- 非负性: $m(E) \geq 0$;
- $m(\emptyset) = 0$;
- 可列可加性: 如果 $E_i, i \geq 1$ 互不相交且 Lebesgue 可测, 则

$$m(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} m(E_i);$$

- 平移不变性: $m(x + E) = m(E)$;
- 存在 $F \subset \mathbb{R}^d$ 不 Lebesgue 可测.

不可测集

巴拿赫-塔斯基分球悖论 (Banach-Tarski Paradox)

将一个三维的半径为 1 的实心球用某种巧妙方法分成五等分 (五等分的意思是, 把其中一份旋转平移后可以和另外一份重合), 然后把这五个分块旋转平移后, 可以组合成两个半径为 1 的实心球.

选择公理 (Axiom of Choice): 任意一组 (可能有不可数无限个) 非空集合, 我们都可以从每个集合挑出一个元素.

巴拿赫-塔斯基分球悖论的启示

- 在 $\Omega = \mathbb{R}^3$ 时, 无法构造出一个测度既与常见的体积度量相容且具有平移不变性, 又可以度量 $\mathbb{R}^3$ 的任一子集;
- 类似的, 在 $\Omega = \mathbb{R}$ 时, 同样无法构造出一个测度既与常见的长度度量相容且具有平移不变性, 又可以度量 $\mathbb{R}$ 的任一子集; (不可测集构造详见教材第 33-34 页 2.4 节 "不可测集".)
- 因此, 对于可测空间 $(\Omega, \mathcal{F})$ 中的 $\mathcal{F}$ 通常为 Ω 的某些子集构成的 σ 代数, 而不一定是 Ω 的所有子集构成的集类.

由 Lebesgue 测度到一般抽象测度

- 要度量集合元素的全体: Ω
- 测度的适用范围 (定义域): $\mathcal{F} := \{A \subset \Omega : A\text{可测}\}$, 通常要求 $\mathcal{F}$ 为 σ-代数
- 通常称 $(\Omega, \mathcal{F})$ 为可测空间
- 测度的定义: 称定义在 $\mathcal{F}$ 的集函数 $\mu(\cdot)$ 为测度, 如果
  (1) $\mu(A) \geq 0, \forall A \in \mathcal{F}$;
  (2) $\mu(\emptyset) = 0$;
  (3) μ 具有可列可加性: 若 $A = \cup_{i=1}^{\infty} A_i$, 其中 $A, A_i \in \mathcal{F}$, 且 $i \neq j$ 时, $A_i \cap A_j = \emptyset$, 则

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

- 称 $(\Omega, \mathcal{F}, \mu)$ 为测度空间.

概率与测度

- Ω: 概率论中的样本空间对应于测度中要度量的集合所包含元素的全体
- $\mathcal{F}$: 事件域, 即某些随机事件组成的集合, 通常要求其为 σ-代数;
  - 注意到随机事件是随机试验中我们所关心的可能出现的结果, 它由一个或若干个基本事件组成, 因此事件是样本空间的子集, 从而事件域为样本空间的某些子集构成的集合;
  - 概率论中的事件域对应于测度中要度量的集合全体即可测集的概念
- $(\Omega, \mathcal{F})$:可测空间

概率的公理化定义

定义 2.18 (概率的公理化定义) 设 Ω 为样本空间, $\mathcal{F}$ 为 Ω 的某些子集组成的 σ-代数 (事件域), 若定义在 $\mathcal{F}$ 上的集函数 $P(\cdot)$ 满足

(1) 非负性: $P(A) \geq 0$;

(2) 正则性: $P(\Omega) = 1 \Rightarrow P(\emptyset) = 0$(结合可列可加性);

(3) 可列可加性: 若 $A_i, i = 1, 2, \cdots$ 互不相交, 则

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

则称 P 为 $(\Omega, \mathcal{F})$ 上的概率测度, 简称概率, 称 $(\Omega, \mathcal{F}, P)$ 为概率空间.

关于概率的几个注记

- 从概率的定义可以看出，概率是全集度量为 1 的测度;
- 概率与我们熟知的度量 (长度, 面积, 体积, 质量) 相比, 本质上没有太大的区别, 只不过概率是用来衡量随机事件发生可能性大小的一种度量, 其在全集上的度量为 1;
- 同一随机试验得到的可测空间 $(\Omega, \mathcal{F})$, 可定义不同的概率.
  - 对于抛一枚硬币这一随机试验,

$$\Omega = \{\text{正}, \text{反}\}, \text{ 选取} \mathcal{F} = \{\emptyset, \{\text{正}\}, \{\text{反}\}, \Omega\};$$

  - 我们可以构造概率 $P_1(\cdot)$:

$$P_1(\emptyset) = 0, P_1(\{\text{正}\}) = 1/2, P_1(\{\text{反}\}) = 1/2, P_1(\Omega) = 1;$$

  - 也可以构造概率 $P_2(\cdot)$: 其中 $p + q = 1$,

$$P_2(\emptyset) = 0, P_2(\{\text{正}\}) = p, P_2(\{\text{反}\}) = q, P_1(\Omega) = 1.$$

第五讲：概率的公理化体系

张鑫

Email: xzhangseu@seu.edu.cn

东南大学 quad 数学学院

代数 (algebra) 或域 (field)

定义 2.19 (集合的代数 (algebra)) 设 $\mathcal{A}$ 是由 Ω 的某些子集所构成的集合类, 称 $\mathcal{A}$ 为集合 Ω 上的代数, 如果,

1. $\Omega \in \mathcal{A}, \varnothing \in \mathcal{A}$;
2. 若 $A \in \mathcal{A}$, 则 $\overline{A} := \Omega - A \in \mathcal{A}$;
3. 若 $A, B \in \mathcal{A}$, 则 $A \cup B \in \mathcal{A}$.

例 2.22 设 $\Omega = \{a, b, c, d\}$, 则集族 $\mathcal{A} = \{\varnothing, \Omega, \{a\}, \{b, c, d\}\}$ 是代数.

例 2.23 (有限集上的代数) 设 Ω 为有限集, 则集族 2^{Ω} (即 Ω 的所有子集构成的集合) 是代数. 例如, 当 $\Omega = \{a, b, c\}$ 时,

$$\mathcal{A} = \{\varnothing, \Omega, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

是代数.

注 2.4

- 代数是" 集合的集合", 即其本身是集合, 其中的元素仍然是集合.
- 代数对于集合的求补、有限交和有限并运算封闭。
- 由于代数仅对集合的有限交并封闭, 若样本空间 Ω 是无限集, 将会导致它将无法包含一些重要的集合 (特别是无限集合).

事件 σ 域定义及其性质

定义 2.20 (事件 σ 域或 σ- 代数) 设 $\mathcal{F}$ 为 Ω 的某些子集构成的集合类，称 $\mathcal{F}$ 为 Ω 上的事件 σ 域或 σ- 代数，如果

1. $\Omega \in \mathcal{F}$;
2. 若 $E \in \mathcal{F}$, 则 $\overline{E} := \Omega - E \in \mathcal{F}$;
3. 若 $E_i \in \mathcal{F}, i = 1, 2, \cdots$, 则 $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$.

定理 2.4 若 $\mathcal{F}$ 为 Ω 上的 σ-代数，则

- $\emptyset \in \mathcal{F}$;
- 若 $E_k \in \mathcal{F}, k = 1, 2, \cdots$, 则 $\cap_{k=1}^{\infty} E_k \in \mathcal{F}$;
- 若 $E_1, \cdots, E_n \in \mathcal{F}$, 则 $\cap_{k=1}^{n} E_k \in \mathcal{F}$ 且 $\cup_{k=1}^{n} E_k \in \mathcal{F}$;
- 若 $E_1, E_2 \in \mathcal{F}$, 则 $E_1 - E_2 \in \mathcal{F}$.

几个例子

例 2.24 (代数而非 σ-代数) 设 Ω 是无限集合, 集族 $\mathcal{A} \subset 2^{\Omega}$ ，满足

$$A \in \mathcal{A} \Longleftrightarrow \#A < \infty \text{ 或} \#\overline{A} < \infty,$$

即 $\mathcal{A}$ 中的集合或者自身是有限集，或者其补集是有限集. 容易验证, $\mathcal{A}$ 是代数, 但不是 σ-代数.

例 2.25 (可数集上 σ-代数) 设样本空间 Ω 为可数集, $\mathcal{F}$ 为其上的 σ-代数, 且 $\forall a \in \Omega$, 有 $\{a\} \in \mathcal{F}$, 则 $\forall A \subset \Omega$, 有 $A \in \mathcal{F}$.

例 2.26 设 $\mathcal{F}$ 为实数集 $\mathbf{R}$ 上的 σ-代数. 若对 $\forall x \in \mathbf{R}$, 都有 $(-\infty, x] \in \mathcal{F}$, 则 $\mathcal{F}$ 包含 $\mathbf{R}$ 上的任意区间.

证明: 仅以 $[a, b)$ 为例. 注意到:

$$(-\infty, a) = \bigcup_{k=1}^{\infty} \left(-\infty, a - \frac{1}{k}\right] \in \mathcal{F}, \quad (x, \infty) = \overline{(-\infty, x]} \in \mathcal{F}, \forall x$$

$$[b, \infty) = \bigcap_{k=1}^{\infty} \left(b - \frac{1}{k}, \infty\right) \in \mathcal{F} \quad [a, b) = \overline{(-\infty, a) \cup [b, \infty)} \in \mathcal{F}$$

σ-代数例子及如何构造

- $\mathcal{F} := \{\emptyset, \Omega\}$;
- $\mathcal{F} := \{A : A \subset \Omega\}$;
- $\mathcal{F} := \{\emptyset, A, \bar{A}, \Omega\}$;
- 如何由 $\{A, B\}$(假设 A, B 为非平凡事件，互不包含且不互为对立事件) 扩充成事件 σ 域?

最小 (生成)σ-代数

定理 2.5 假设 $\mathcal{C}$ 为 Ω 的子集组成的非空集类，则存在唯一 σ-代数 $\mathcal{C}_0$ 使得:

(1) $\mathcal{C} \subset \mathcal{C}_0$;

(2) 若 $\mathcal{G}$ 为任一 σ-代数，且 $\mathcal{C} \subset \mathcal{G}$, 则 $\mathcal{C}_0 \subset \mathcal{G}$.

证明:

- 若 $\mathcal{F}_t, t \in T$ 为一族 σ-代数，则 $\cap_{t\in T}\mathcal{F}_t$ 仍为 σ-代数;
- 令 $\mathcal{C}^{\#} := \{\mathcal{B} : \mathcal{B}为\sigma - 代数, \mathcal{C} \subset \mathcal{B}\}$;
- 令 $\mathcal{C}_0 := \cap_{\mathcal{B}\in\mathcal{C}^{\#}}\mathcal{B}$, 则易证 $\mathcal{C}_0$ 即为定理所求之 σ-代数.

注 2.5

- $\mathcal{C}_0 := \cap_{\mathcal{B}\in\mathcal{C}^{\#}}\mathcal{B}$ 是包含集类 $\mathcal{C}$ 的最小 σ-代数;
- 一般也称 $\mathcal{C}_0$ 为集类 $\mathcal{C}$ 的生成 σ-代数，后面我们常用 $\sigma(\mathcal{C})$ 表示 $\mathcal{C}_0$.

实直线上的 σ- 域

例 2.27 令 $\Omega := \mathbb{R}$, 考虑

$$\mathcal{A}_1 = \{(-\infty, x) | x \in \mathbb{R}\},\ \mathcal{A}_2 = \{(a, b) | -\infty < a < b < \infty\},$$

则 $\sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_2)$

解: 注意到，对任意实数 x, 均有

$$(-\infty, x) = \cup_{n=1}^{\infty}(x - n, x).$$

因此 $(-\infty, x) \in \sigma(\mathcal{A}_2)$, 从而 $\mathcal{A}_1 \subset \sigma(\mathcal{A}_2)$, 进而可得 $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_2)$.
反之，注意到 $[a, b) = (-\infty, b) - (-\infty, a) \in \sigma(\mathcal{A}_1)$ 以及

$$\{a\} = \cap_{n=1}^{\infty}[a, a + \frac{1}{n}) \in \sigma(\mathcal{A}_1),$$

因此，$(a, b) = [a, b) - \{a\} \in \sigma(\mathcal{A}_1)$, 即 $\mathcal{A}_2 \subset \sigma(\mathcal{A}_1)$, 从而

$$\sigma(\mathcal{A}_2) \subset \sigma(\mathcal{A}_1).$$

Borel 集

例 2.28 设 $\Omega = \mathbb{R}$, 令

$$
\begin{aligned}
\mathcal{C}_1 &= \{(-\infty, x) : -\infty < x < +\infty\} \\
\mathcal{C}_2 &= \{(a, b] : -\infty \le a \le b < +\infty\} \\
\mathcal{C}_3 &= \{[a, b] : -\infty < a \le b < +\infty\} \\
\mathcal{C}_4 &= \{[a, b) : -\infty < a \le b \le +\infty\} \\
\mathcal{C}_5 &= \{(a, b) : -\infty \le a \le b \le +\infty\} \\
\mathcal{C}_6 &= \{A \subset \mathbb{R} : A\text{为开集}\} \\
\mathcal{C}_7 &= \{A \subset \mathbb{R} : A\text{为闭集}\}
\end{aligned}
$$

则以上 7 个集类的生成 σ-代数均相同，称以上集类的生成 σ-代数中的元素为 Borel 集，并记作 $\mathcal{B}(\mathbb{R})$, 即

$$\sigma(\mathcal{C}_1) = \cdots = \sigma(\mathcal{C}_7) := \mathcal{B}(\mathbb{R}).$$

其他集类

定义 2.21 设 $\mathcal{C}$ 为 Ω 的某些子集构成的非空集类,

- 称 $\mathcal{C}$ 为 π 类, 如果 $\mathcal{C}$ 关于有限交运算封闭, 即

$$A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C};$$

- 称 $\mathcal{C}$ 为代数, 如果 $\mathcal{C}$ 关于有限交与取余运算封闭, 即

$$A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}, \bar{A} \in \mathcal{C};$$

- 称 $\mathcal{C}$ 为单调类, 如果 $\mathcal{C}$ 关于单调极限运算封闭, 即

$$A_n \in \mathcal{C}, A_n \uparrow A \text{ 或} A_n \downarrow A \Rightarrow A \in \mathcal{C}$$

- 称 $\mathcal{C}$ 为 λ 类, 如果 $\Omega \in \mathcal{C}$ 并且 $\mathcal{C}$ 对真差运算及上升极限运算封闭, 即
  - $\Omega \in \mathcal{C}$;
  - $A, B \in \mathcal{C}, B \subset A \Rightarrow A - B \in \mathcal{C}$;
  - $A_n \in \mathcal{C}, n \geq 1, A_n \uparrow A \Rightarrow A \in \mathcal{C}$

集类之间的关系

定理 2.6 设 $\mathcal{C}$ 为 Ω 的某些子集构成的非空集类, 则

1. 若 $\mathcal{C}$ 为 $\sigma-$ 代数, 则 $\mathcal{C}$ 一定是:π 类, 代数,λ 类, 单调类;
2. 若 $\mathcal{C}$ 既是代数又是单调类, 则 $\mathcal{C}$ 一定是 $\sigma-$ 代数;
3. 若 $\mathcal{C}$ 既是 π 类又是 λ 类, 则 $\mathcal{C}$ 一定是 $\sigma-$ 代数.

单调类定理

定义 2.22 与生成 $\sigma-$ 代数类似, 我们可以定义包含集类 $\mathcal{C}$ 的最小单调类 $m(\mathcal{C})$ 及最小 λ 类 $\lambda(\mathcal{C})$, 分别称为由集类 $\mathcal{C}$ 生成的单调类与 λ 类.

定理 2.7 (单调类定理) 设 $\mathcal{C}, \mathcal{G}$ 为 Ω 中的集类且 $\mathcal{C} \subset \mathcal{G}$, 则

1. 若 $\mathcal{C}$ 为一代数, 则 $m(\mathcal{C}) = \sigma(\mathcal{C})$;
2. 若 $\mathcal{C}$ 为一 π 类, 则 $\lambda(\mathcal{C}) = \sigma(\mathcal{C})$;
3. 若 $\mathcal{C}$ 为代数而 $\mathcal{G}$ 为单调类, 则 $\sigma(\mathcal{C}) \subset \mathcal{G}$;
4. 若 $\mathcal{C}$ 为 π 类而 $\mathcal{G}$ 为 λ 类, 则 $\sigma(\mathcal{C}) \subset \mathcal{G}$.

如何应用单调类定理

- 问题: 如何证明某 $\sigma-$ 代数 $\mathcal{F}$ 中的所有元素都具有某种性质 (P)?
- 步骤:
  - $\mathcal{G} := \{A \in \mathcal{F} : A\text{具有性质}(P)\}$;
  - 找到某一集类 $\mathcal{C}$ 为代数或 π 类且具有以下性质:

$$\mathcal{C} \subset \mathcal{G}, \quad \sigma(\mathcal{C}) = \mathcal{F};$$

  - 证明 $\mathcal{G}$ 为单调类或 λ 类;
  - 由单调类定理可知:

$$\mathcal{F} = \sigma(\mathcal{C}) \subset \mathcal{G} \subset \mathcal{F}$$

即

$$\mathcal{G} = \mathcal{F}$$

从而 $\mathcal{F}$ 中所有的元素都具有某种性质 (P).

概率的公理化定义

定义 2.23 设 $\mathcal{F}$ 为样本空间 Ω 上的 σ-代数，如果定义在 $\mathcal{F}$ 上的集函数 $P(\cdot)$ 具有:

1. 非负性: $P(A) \geq 0, \forall A \in \mathcal{F}$;
2. 规范性: $P(\Omega) = 1$;
3. 可列可加性：对 $\mathcal{F}$ 中的任何两两互不相容事件列 $\{A_n\}$ 有

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

则称 $P(\cdot)$ 为 $\mathcal{F}$ 上的概率测度，而 $P(A)$ 称为事件 A 的概率，$(\Omega, \mathcal{F}, P)$ 称为概率空间.

如何定义概率

- 概率定义在 $\sigma-$ 代数 $\mathcal{F}$ 上, 但是由于 $\mathcal{F}$ 中集合的复杂性, 直接对每个集合赋予概率存在困难;
- 首先在 $\mathcal{F}$ 的子族 $\mathcal{A} \subset \mathcal{F}$ 上定义概率（一般 $\mathcal{A}$ 中的集合较为简单), 然后再将其扩张到 $\mathcal{F}$ 上;
- 如何选择恰当的 $\mathcal{A}$, 使得 $\mathcal{F} = \sigma(\mathcal{A})$?

扩张是否唯一?

例 2.29 (有限集合上的概率) 设样本空间 $\Omega = \{a, b, c, d\}$, σ-代数 $\mathcal{F}$ 由集族 $\mathcal{A}$ 生成, 即 $\mathcal{F} = \sigma(\mathcal{A})$, $\quad \mathcal{A} = \{\{a\}, \{a, b\}, \{a, b, c\}, \Omega\}$. 定义

$$P(\{a\}) = \frac{1}{4}, P(\{a, b\}) = \frac{1}{2}, P(\{a, b, c\} = \frac{3}{4}).$$

例 2.30 令 $\Omega = \{a, b, c, d\}$, $\mathcal{F} = 2^{\Omega}$, $\mathcal{A} = \{\{a, b\}, \{a, c\}, \{b, d\}, \{c, d\}\}$.

- 容易看出,$\sigma(\mathcal{A}) = 2^{\Omega}$;
- 定义 P_1 和 P_2 为

$$P_1(\{a\}) = P_1(\{d\}) = \frac{1}{6}, \quad P_1(\{b\}) = P_1(\{c\}) = \frac{1}{3}$$
$$P_2(\{a\}) = P_2(\{d\}) = \frac{1}{3}, \quad P_2(\{b\}) = P_2(\{c\}) = \frac{1}{6}$$

- 显然, 它们并不相同, 但它们在 $\mathcal{A}$ 上却完全一致

$$P_1(\{a, b\}) = \frac{1}{2}, P_1(\{a, c\}) = \frac{1}{2}, P_1(\{b, d\}) = \frac{1}{2}, P_1(\{c, d\}) = \frac{1}{2}$$
$$P_2(\{a, b\}) = \frac{1}{2}, P_2(\{a, c\}) = \frac{1}{2}, P_2(\{b, d\}) = \frac{1}{2}, P_2(\{c, d\}) = \frac{1}{2}$$

概率扩张唯一性定理

定理 2.8 (概率扩张的唯一性) 给定样本空间 Ω 和其上的 σ-代数 $\mathcal{F}$. 若

- P_1 和 P_2 是定义在 $\mathcal{F}$ 上的两个概率;
- 集族 $\mathcal{A}$ 关于交运算封闭, 且 $\sigma(\mathcal{A}) = \mathcal{F}$;
- 对任意的 $A \in \mathcal{A}$ 均有 $P_1(A) = P_2(A)$;

则对任意的 $A \in \mathcal{F}$ 一定有 $P_1(A) = P_2(A)$.

证明: 证明思路如下:

- 令 $\mathcal{G} := \{A \in \sigma(\mathcal{A}) : P_1(A) = P_2(A)\}$;
- 注意到 $\mathcal{A}$ 为 π 类且 $\mathcal{A} \subset \mathcal{G}$, 故只需证明 $\mathcal{G}$ 为 λ 类即验证即
  - $\Omega \in \mathcal{G}$;
  - $A, B \in \mathcal{G}, B \subset A \Rightarrow A - B \in \mathcal{G}$;
  - $A_n \in \mathcal{G}, n \geq 1, A_n \uparrow A \Rightarrow A \in \mathcal{G}$
- 利用单调类定理可知 $\mathcal{G} = \sigma(\mathcal{A})$

(概率) 测度扩张定理

定理 2.9 (Carathéodory 测度扩张定理) 设 Ω 为样本空间,$\mathcal{A}$ 为代数. 如果 P_0 为定义在 $\mathcal{A}$ 上的概率, 且满足可数可加性, 那么存在唯一的定义在 $\sigma(\mathcal{A})$ 上的概率 P, 满足

$$P(A) = P_0(A), \quad \forall A \in \mathcal{A},$$

注 2.6 更一般的测度扩张定理详见测度论教材, 如严加安院士的 "测度论讲义" 等.

如何理解概率

- 概率空间定义中的样本空间 Ω 和概率 P 都很好理解，但是 σ-代数 $\mathcal{F}$ 的出现总是显得有些突兀.
- 概率是一个以事件 (集合) 为自变量的函数.
- 给定一个事件 (集合), 概率函数即赋予它一个非负实数, 用以表示该事件 (集合) 的样本点在统计实验的结果中出现的可能性大小.
- 函数的基本概念中，”定义域”占据重要位置. 概率函数也不例外.
- 函数的定义域中的数对“加减乘除”运算封闭. 类似的, 概率函数应对事件的“交差并补”等运算封闭.
- σ-代数 $\mathcal{F}$ 本质上就是概率函数的“定义域”, 对自变量 (事件) 的“交差并补”等运算封闭..

概率测度的性质 I

定理 2.10 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，则其上的概率 P 具有如下性质:

1. $P(\emptyset) = 0$;
2. 有限可加性：若 $A_1, \cdots, A_n$ 为 $\mathcal{F}$ 中两两互不相容事件，则

$$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$$

3. 可减性：若 $A, B \in \mathcal{F}$ 且 $A \subset B$, 则 $P(B - A) = P(B) - P(A)$;
4. 单调性：若 $A, B \in \mathcal{F}$ 且 $A \subset B$, 则 $P(A) \leq P(B)$;
5. $P(\bar{A}) = 1 - P(A)$;

概率测度的性质 II

6. 加法公式 (容斥原理): 对任意两个事件 A, B, 有

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{3}$$

一般的，对任意 n 个事件 $A_1, \cdots, A_n$, 有

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{1 \le i < j \le n} P(A_i A_j) + \sum_{1 \le i < j < k \le n} P(A_i A_j A_k) \\ &\quad + \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n). \end{aligned} \tag{4}$$

7. 次可加性 (Boole 不等式): 对任意两个事件 A, B, 有

$$P(A \cup B) \le P(A) + P(B).$$

一般的，对任意 n 个事件 $A_1, \cdots, A_n$, 有

$$P(\cup_{i=1}^n A_i) \le \sum_{i=1}^n P(A_i).$$

概率测度的性质 III

8. Bonferroni 不等式:

$$P(\cup_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j),$$

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k), \cdots .$$

9. 下连续性: 若 $A_n \in \mathcal{F}$ 且 $A_n \subset A_{n+1}, n = 1, 2, \cdots$, 则

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

10. 上连续性: 若 $A_n \in \mathcal{F}$ 且 $A_n \supset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\cap_{n=1}^{\infty} A_n) = P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$$

11. 概率 P 的可列可加性等价于 P 有限可加且下连续.

加法公式证明 I

注意到

$$A \cup B = A \cup (B - AB).$$

故由概率的有限可加性及减法公式可得:

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

对于一般情形，可由归纳法证明。当 $n = 2$ 时，(4) 式即为 (3) 式。设 (4) 式对 $n-1$ 成立，则先对两个事件 $\cup_{i=1}^{n-1} A_i$ 与 A_n 应用 (3) 得

$$\begin{aligned} P(\cup_{i=1}^{n} A_i) &= P(\cup_{i=1}^{n-1} A_i) + P(A_n) - P((\cup_{i=1}^{n-1} A_i) \cap A_n) \\ &= P(\cup_{i=1}^{n-1} A_i) + P(A_n) - P((\cup_{i=1}^{n-1} (A_i A_n)) \end{aligned} \tag{5}$$

加法公式证明 II

而

$$
\begin{aligned}
P(\cup_{i=1}^{n-1} A_i) &= \sum_{i=1}^{n-1} P(A_i) - \sum_{1\le i<j\le n-1} P(A_iA_j) + \sum_{1\le i<j<k\le n-1} P(A_iA_jA_k) \\
&\quad + \cdots + (-1)^{n-2} P(A_1A_2\cdots A_{n-1}),
\end{aligned}
$$

$$
\begin{aligned}
P((\cup_{i=1}^{n-1}(A_iA_n)) &= \sum_{1\le i\le n-1} P(A_iA_n) - \sum_{1\le i<j\le n-1} P(A_iA_jA_n) \\
&\quad + \cdots + (-1)^{n-2} P(A_1A_2\cdots A_n) \\
&= \sum_{1\le i<j=n} P(A_iA_n) - \sum_{1\le i<j<k=n} P(A_iA_jA_n) \\
&\quad + \cdots + (-1)^{n-2} P(A_1A_2\cdots A_n)
\end{aligned}
$$

故将上面两式代入 (5) 式即得加法公式.

Boole 不等式的证明: 归纳法

- $n=2$ 时显然有

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leqslant P(A_1) + P(A_2),$$

- 假定 $n=m$ 时结论成立, 那么 $n=m+1$ 时, 有

$$\begin{aligned} P(\bigcup_{i=1}^{m+1} A_i) &\leq P(\bigcup_{i=1}^{m} A_i) + P(A_{m+1}) \\ &\leqslant \sum_{i=1}^{m} P(A_i) + P(A_{m+1}) \\ &= \sum_{i=1}^{m+1} P(A_i) \end{aligned}$$

Boole 不等式证明: 非归纳法

- 注意到:

$$
\begin{aligned}
\cup_{i=1}^{n} A_i &= A_1 \cup (A_2 - A_1) \cup (A_3 - (A_1 \cup A_2)) \cup \cdots \cup (A_n - \cup_{i=1}^{n-1} A_i) \\
&= A_1 \cup (A_2 \overline{A}_1) \cup (A_3 \overline{A}_2 \overline{A}_1)) \cup \cdots \cup (A_n \overline{A}_{n-1} \cdots \overline{A}_1) \\
&= A_1 \cup \cup_{i=2}^{n} (A_i \overline{A}_{i-1} \cdots \overline{A}_1)
\end{aligned}
$$

- 从而

$$
\begin{aligned}
P(\cup_{i=1}^{n} A_i) &= P(A_1) + \sum_{i=2}^{n} P(A_i \overline{A}_{i-1} \cdots \overline{A}_1) \\
&= P(A_1) + \sum_{i=2}^{n} \left[P(A_i) - P(A_i \cap (A_{i-1} \cup \cdots \cup A_1)) \right] \\
&= \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n} P(\cup_{j=1}^{i-1} (A_i A_j)) \leq \sum_{i=1}^{n} P(A_i)
\end{aligned}
$$

Bonferroni 不等式证明: 非归纳法

- 注意到

$$P(\cup_{j=1}^{i-1}(A_iA_j)) \le \sum_{j=1}^{i-1} P(A_iA_j)$$

- 从而

$$\begin{aligned} P(\cup_{i=1}^{n}A_i) &= \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n} P(\cup_{j=1}^{i-1}(A_iA_j)) \\ &\ge \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n}\sum_{j=1}^{i-1} P(A_iA_j) \\ &= \sum_{i=1}^{n} P(A_i) - \sum_{1\le j<i\le n} P(A_iA_j). \end{aligned}$$

- 再由$P(\cup_{j=1}^{i-1}(A_iA_j)) \ge \sum_{j=1}^{i-1} P(A_iA_j) - \sum_{1\le j<k\le i-1} P(A_iA_j \cap A_iA_k)$ 可得

$$P(\cup_{i=1}^{n}A_i) \le \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i \cap A_j) + \sum_{1\le i<j<k\le n} P(A_i \cap A_j \cap A_k)$$

概率 P 的下连续性

设 F_n 是 $\mathcal{F}$ 中的一个单调不减的事件序列，即

$$\lim_{n\to\infty} F_n = \cup_{i=1}^{\infty} F_i$$

若令 $F_0 = \emptyset$, 则

$$\cup_{i=1}^{\infty} F_i = \cup_{i=1}^{\infty}(F_i - F_{i-1}).$$

由于 $F_{i-1} \subset F_i$, 显然诸 $F_i - F_{i-1}$ 互不相容，故由可列可加性得

$$\begin{aligned} P(\cup_{i=1}^{\infty} F_i) &= P(\cup_{i=1}^{\infty}(F_i - F_{i-1})) = \sum_{i=1}^{\infty}(P(F_i) - P(F_{i-1})) \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}(P(F_i) - P(F_{i-1})) = \lim_{n\to\infty} P(F_n). \end{aligned}$$

概率 P 的上连续性

设 $\{E_n\}$ 是一列单调不增的事件序列，则 $\{\bar{E}_n\}$ 为单调不减事件序列，由概率的下连续性可得

$$P(\cup_{i=1}^{\infty}\bar{E}_i) = \lim_{n\to\infty} P(\bar{E}_n) = 1 - \lim_{n\to\infty} P(E_n) \tag{6}$$

另一方面，

$$P(\cup_{i=1}^{\infty}\bar{E}_i) = P(\overline{\cap_{i=1}^{\infty}E_i}) = 1 - P(\cap_{i=1}^{\infty}E_i) = 1 - P(\lim_{n\to\infty} E_n) \tag{7}$$

综合 (6) 式与 (7) 式可得

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

可列可加 $\Leftrightarrow$ 有限可加且下连续

$\Rightarrow$, 显然，下证 $\Leftarrow$。

设 $A_i \in \mathcal{F}, i = 1, 2, \cdots$，是两两互不相容的事件序列，则由有限可加性知，对任意的 n 均有 $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, 等式两边同时令 $n \to \infty$ 可得

$$\lim_{n\to\infty} P(\cup_{i=1}^n A_i) = \lim_{n\to\infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^\infty P(A_i),$$

令 $F_n := \cup_{i=1}^n A_i$, 则易有 F_n 为单调不减事件序列，故有概率的下连续性可得

$$\lim_{n\to\infty} P(\cup_{i=1}^n A_i) = \lim_{n\to\infty} P(F_n) = P(\cup_{n=1}^\infty F_n) = P(\cup_{i=1}^\infty A_i)$$

几个概率空间的例子

例 2.31 (Bernoulli 概率空间) 取 $\mathcal{F}=\{\emptyset, A, \bar{A}, \Omega\}$, 其中 A 为 Ω 的非空真子集。任取两个正数 $p, q(p+q=1)$, 令

$$P(\emptyset)=0, P(A)=p, P(\bar{A})=q, P(\Omega)=1.$$

易证 P 是一个概率测度。从而 $(\Omega, \mathcal{F}, P)$ 是一个概率空间.

例 2.32 (有限概率空间) 样本空间是有限集 $\Omega=\{\omega_1, \cdots, \omega_n\}$. $\mathcal{F}$ 取为 Ω 的一切子集 (共 2^n 个) 组成的集类。设 $\{p_i\}_{1\leq i\leq n}$ 为满足 $\sum_{i=1}^n p_i=1$ 的 n 个非负实数。定义集函数 $P(\cdot)$ 如下

$$P(A)=\sum_{\omega_i\in A} p_i.$$

易证 P 为概率测度，称此 $(\Omega, \mathcal{F}, P)$ 为有限概率空间。特别的若取 $p_i=\dfrac{1}{n}, 1\leq i\leq n$, 则为古典概率空间.

几个概率空间的例子

- 例 2.33 (离散概率空间) 样本空间 $\Omega = \{\omega_1, \omega_2, \cdots\}$ 为可列集，$\mathcal{F}$ 取为 Ω 的所有子集所构成的集类。设 $\{p_i\}_{i\geq 1}$ 为满足 $\sum_{i=1}^{\infty} p_i = 1$ 的一列非负实数。对任意的 $E \subset \Omega$, 定义集函数 $P(\cdot)$ 如下

$$P(E) = \sum_{j:\omega_j \in E} p_j.$$

易证 P 是一个概率测度，于是 $(\Omega, \mathcal{F}, P)$ 是一个概率空间，称为离散概率空间.

- 例 2.34 (一维几何概率空间) 设样本空间 Ω 为实直线上的具有正 Lebesgue 测度的某个 Borel 集，取 $\mathcal{F}$ 为 Ω 的所有 Borel 子集所构成的类。对每个 $E \in \mathcal{F}$, 令

$$P(E) = \frac{m(E)}{m(\Omega)}$$

其中 $m(E)$ 和 $m(\Omega)$ 分别表示集合 E 与 Ω 的 Lebesgue 测度。易证 $P(\cdot)$ 为概率测度.

减法公式妙用

例 2.35 口袋中有编号为 $1, 2, \cdots, n$ 的 n 个球，从中有放回地任取 m 次，求取出 m 个球最大号码为 k 的概率.

解: 令

$$\begin{aligned} A_k &:= \{\text{取出的}m\text{个球的最大号码为}k\}, \\ B_k &:= \{\text{取出的}m\text{个球的最大号码为小于等于}k\} \end{aligned}$$

则易有 $A_k = B_k - B_{k-1}$ 且 $P(B_k) = \frac{k^m}{n^m}$.另一方面，显然有 $B_{k-1} \subset B_k$, 故

$$P(A_k) = P(B_k - B_{k-1}) = P(B_k) - P(B_{k-1}) = \frac{k^m - (k-1)^m}{n^m}$$

对立事件妙用

例 2.36 口袋中有 $n-1$ 个黑球，1 个白球，每次从口袋中随机地摸出一球，并换入一只黑球，求第 k 次摸到黑球的概率.

解: 记 $A := \{第k次取到黑球\}$,则 A 的对立事件为

$$\overline{A} = \{第k次取到白球\}$$

这意味着，第 1 次，$\cdots$，第 $k-1$ 次只能取到黑球，而第 k 次取到白球.

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{(n-1)^{k-1}}{n^k}$$

巧用对立事件解决无空盒问题

例 2.37 将 m 个不同的小球等可能地放入 n 个不同的盒子，$m > n$, 试求无空盒出现的概率.

- 易知 $|\Omega| = n^m$, 记 $E = \{无空盒出现\}$, 但 $|E|$ 不易求得;
- 记 $A_k = \{第k号盒子是空盒\}$, $|A_k|, |A_{j_1} \cdots A_{j_i}|$ 均易求;
- E 与 A_k 的关系: $\overline{E} = \bigcup\limits_{k=1}^{n} A_k$;
- 加法公式:

$$P(\overline{E}) = P\left(\cup_{k=1}^{n} A_k\right) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{1 \le j_1 < \cdots < j_i \le n} P(A_{j_1} \cdots A_{j_i})$$

巧用对立事件解决无空盒问题

- 计算概率 $P(A_{j_1}\cdots A_{j_i})$:

$$P(A_{j_1}\cdots A_{j_i})=\frac{(n-i)^m}{n^m}=\left(1-\frac{i}{n}\right)^m;$$

- $P(\overline{E})=\sum_{i=1}^{n}(-1)^{i-1}C_n^i\left(1-\frac{i}{n}\right)^m$;
- $P(E)=1-P(\overline{E})=\sum_{i=0}^{n}(-1)^{i}C_n^i\left(1-\frac{i}{n}\right)^m$.

Buffon 投针问题的续一

例 2.38 向画满间隔为 a 的平行直线的桌面上任投一直径为 $l(l<a)$ 的半圆形纸片，求事件 $E=\{$纸片与某直线相交$\}$ 的概率.

解: 设想把半圆形纸片拼成一个圆形，并记

$$F=\{\text{新拼的半圆形与某直线相交}\},$$

则

$$\begin{aligned} E\cup F &= \{\text{直径为}l\text{的圆形与某直线相交}\} \\ E\cap F &= \{\text{长为}l\text{的线段 (即公共直径) 与某直线相交}\} \end{aligned}$$

由前面结论可知 $P(E\cap F)=\dfrac{2l}{\pi a}$. 为计算 $P(E\cup F)$, 可令 d 表示圆心到直线的最近距离，则有 $\Omega=\{d\,|0\le d\le \frac{a}{2}\}, E\cup F=\{d\,|0\le d\le \frac{l}{2}\}$. 因此

$$P(E\cup F)=\frac{l}{a}.$$

从而: $P(E)=\dfrac{P(E)+P(F)}{2}=\dfrac{P(E\cup F)+P(E\cap F)}{2}=\dfrac{\pi l/2+l}{\pi a}.$

Buffon 投针问题的续二

例 2.39 在平面上画有间距为 a 与 b 的水平直线以及垂直直线，向平面随机地抛掷长度为 l 的针，$l < \min\{a, b, a+b-\sqrt{(a+b)^2-\pi ab}\}$，试求事件 E={针与某条直线相交 } 的概率.

- 记 A={针与某条水平直线相交 }, B={针与某条垂直直线相交 };
- $E = A \cup B$, 从而 $P(E) = P(A \cup B) = P(A) + P(B) - P(AB)$;
- $P(A) = \dfrac{2l}{\pi a}, \quad P(B) = \dfrac{2l}{\pi b}$;

Buffon 投针问题的续二

- 计算 $P(AB)$:
  - 设 ρ 和 δ 分别表示针的中点与水平直线的最近距离和与垂直直线的最近距离，设 θ 为针与水平直线的夹角
  - 易知

$$\Omega = \{(\rho, \delta, \theta) | 0 \le \rho \le \frac{a}{2}, 0 \le \delta \le \frac{b}{2}, 0 \le \theta \le \frac{\pi}{2}\},$$

$$AB = \{(\rho, \delta, \theta) | (\rho, \delta, \theta) \in \Omega, \rho \le \frac{l}{2}\sin\theta, \delta \le \frac{l}{2}\cos\theta\}.$$

  - 从而

$$P(AB) = \frac{m(AB)}{m(\Omega)} = \frac{\int_0^{\frac{\pi}{2}} d\theta \int_0^{\frac{l}{2}\cos\theta} d\delta \int_0^{\frac{l}{2}\sin\theta} d\rho}{\frac{a}{2} \cdot \frac{b}{2} \cdot \frac{\pi}{2}} = \frac{l^2}{\pi ab}$$

- 故由加法公式可得

$$P(E) = P(A \cup B) = P(A) + P(B) - P(AB) = \frac{2l(a+b) - l^2}{\pi ab}.$$

配对问题一

例 2.40 设有 n 个孩子，每个人戴着一顶帽子. 如果他们将各自的帽子放在桌子上混杂在一起，然后每人从中随机取一顶帽子戴上，试问至少有一人戴对自己的帽子的概率是多少？

解: 若记

$$A_i := \{第i个人戴对自己的帽子\},$$

则所求事件概率为 $P(\cup_{i=1}^n A_i)$，易知有

$$\begin{aligned} P(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n}, \ \sum_{i=1}^n P(A_i) = 1, \\ P(A_iA_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, i \neq j, \\ \sum_{1\le i<j\le n} P(A_iA_j) &= C_n^2 \frac{1}{n(n-1)} = \frac{1}{2!}, \end{aligned}$$

配对问题一（续）

同理可得

$$
\begin{aligned}
P(A_iA_jA_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}, i \neq j \neq k, \\
\sum_{1\le i<j<k\le n} P(A_iA_jA_k) &= C_n^3 \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}, \\
&\cdots \quad , \\
P(\cap_{i=1}^n A_i) &= \frac{1}{n!}.
\end{aligned}
$$

故由加法公式可得

$$
P(\cup_{i=1}^n A_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n-1}\frac{1}{n!} \xrightarrow{n\to\infty} 1 - e^{-1}.
$$

$$
e^x = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}.
$$

配对问题二

例 2.41 接上面例子, 试求下述各事件的概率:
(1) 恰有 k 个孩子戴对帽子; (2) 至少有 m 个孩子戴对帽子.

- 记 E_k ={恰有 k 个孩子戴对帽子 },A_m ={至少有 m 个孩子戴对帽子 };
- $\Omega = n!$, $|E_k| =?$, $|A_m| =?$;
- 若记 D_k ={给定的某 $n-k$ 个孩子均未戴对自己的帽子 },则有 $|E_k| = C_n^k |D_k|$
- D_k 表示给定的某 $n-k$ 个孩子未戴对自己的帽子，不涉及其余 k 个人，故可套用 $n-k$ 情形下的配对问题即

$$\frac{|D_k|}{(n-k)!} = P(D_k) = 1 - P(\overline{D}_k) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!}$$

- 因此,

$$|D_k| = (n-k)!P(D_k) = (n-k)!\left\{\frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!}\right\}$$

配对问题二

- 从而

$$
\begin{aligned}
P(E_k) &= \frac{|E_k|}{n!} = \frac{C_n^k |D_k|}{n!} \\
&= C_n^k \frac{(n-k)!}{n!} \left\{ \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k} \frac{1}{(n-k)!} \right\} \\
&= \frac{1}{k!} \sum_{j=2}^{n-k} (-1)^j \frac{1}{j!} = \frac{1}{k!} \sum_{j=0}^{n-k} (-1)^j \frac{1}{j!}.
\end{aligned}
$$

- 在上式中令 $n \to \infty$, 可得极限概率为 $\dfrac{1}{e \cdot k!}$
- 由于 $A_m = \bigcup\limits_{k=m}^{n} E_k$, 且事件 $E_m, E_{m+1}, \cdots, E_n$ 两两不交，故

$$
P(A_m) = \sum_{k=m}^{n} P(E_k) = \sum_{k=m}^{n} \frac{1}{k!} \sum_{j=0}^{n-k} (-1)^j \frac{1}{j!} = \sum_{k=m}^{n} \sum_{j=0}^{n-k} (-1)^j \frac{1}{k!j!}.
$$

配对问题三

- 允许一个孩子拿多只帽子的情况: 即允许有孩子没有拿到帽子. 某一个孩子没有拿到帽子的概率是

$$P_1 = \frac{(n-1)^n}{n^n} = \left(1 - \frac{1}{n}\right)^n$$

- 孩子人数 n 并不等于帽子数目 m, 即并不是每一个孩子都有自己的帽子. 某一个孩子没有拿到帽子的概率是

$$\widetilde{P}_1 = \frac{(n-1)^m}{n^m} = \left(1 - \frac{1}{n}\right)^m$$

令 $n, m \to \infty$, 且保持 $m/n = \lambda$ 为常数, 得到 $\widetilde{P}_1 \to \exp(-\lambda)$.

- 某一个孩子恰好拿到 k 顶帽子的概率为

$$\widetilde{P}_k = \binom{m}{k} \frac{(n-1)^{m-k}}{n^m} = \frac{1}{k!} \frac{m(m-1)\cdots(m-k+1)}{(n-1)^k} \left(1 - \frac{1}{n}\right)^m$$
$$\to \frac{\lambda^k}{k!} \exp(-\lambda)$$

第六讲：条件概率、全概率公式与贝叶斯公式

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

条件思考的重要性

- 条件概率说明了如何以合乎逻辑且相一致的方式将证据纳入人们对世界的理解当中.
- 添加条件是一种非常有用的解决问题的策略.
- 条件既可以作为更新判断依据的方法，也是解决问题的策略.
- 条件是统计学的灵魂.

条件概率的引入

例 2.42 抛掷一枚均匀的骰子一次，已知掷出的点数为奇数，试求点数大于 1 的概率。

- $\Omega = \{1, 2, \cdots, 6\}, A = \{1, 3, 5\}, B = \{2, 3, 4, 5, 6\}$;
- 令 $P(B|A)$ 表示“已知 A 发生的情况下 B 发生的概率”;
- $A = \{1, 3, 5\}$, B 发生就是 $AB = \{3, 5\}$ 发生，于是

$$P(B|A) = \frac{|AB|}{|A|} = \frac{2}{3}.$$

- 在原样本空间考虑:

$$P(AB) = \frac{|AB|}{|\Omega|} = \frac{2}{6}, \quad P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6}.$$

- 易验证:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{2}{3}.$$

条件概率的引入

例 2.43 从分别写有号码 $1, 2, \cdots, 10$ 的 10 张卡片中随机抽取一张，已知抽出的卡片的号码不小于 3，试求其号码为偶数的概率。

- $\Omega = \{1, 2, \cdots, 10\}, A = \{3, 4, \cdots, 10\}, B = \{2, 4, 6, 8, 10\}$;
- 类似上一例:

$$P(B|A) = \frac{|AB|}{|A|} = \frac{4}{8} = \frac{1}{2}.$$

- 在原样本空间考虑:

$$P(AB) = \frac{|AB|}{|\Omega|} = \frac{4}{10} = \frac{2}{5}, \quad P(A) = \frac{|A|}{|\Omega|} = \frac{8}{10} = \frac{4}{5}.$$

- 易验证:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{1}{2}.$$

条件概率的引入

例 2.44 考察有两个小孩的家庭，其样本空间为 $\Omega := \{bb, bg, gb, gg\}$，其中 b, g 分别表男孩与女孩，bg 表示大的是男孩，小的是女孩，其他样本点类似说明. 在 Ω 中的四个样本点可能情况下，我们讨论如下事件的概率:

- 事件 $A :=$ {家中至少有一个女孩} 发生的概率为 $P(A) = \dfrac{3}{4}$
- 若已知 $B :=$ {家中至少有一个男孩} 发生，再求 A 发生的概率为

$$P(A|B) = \frac{2}{3}$$

- 若对上述条件概率的分子分母各除以 4，则有

$$P(A|B) = \frac{2/4}{3/4} = \frac{P(AB)}{P(B)}$$

条件概率的定义

定义 2.24 设 $(\Omega, \mathcal{F}, P)$ 为概率空间. $A \in \mathcal{F}$, $B \in \mathcal{F}$, 若 $P(B) > 0$, 则称

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

为 "在已知 B 发生的情况下，A 发生的 条件概率", 简称条件概率.

注 2.7 条件概率的两种计算方法:

- 原则性方法: $P(A|B) = \dfrac{P(AB)}{P(B)}$
- 把 B 作为样本空间看待 (经常显得非常方便): $P(A|B) = \dfrac{|AB|}{|B|}$

条件概率的例子

例 2.45 (两张牌) 洗好一副标准扑克后。从中随机抽取两张牌，无放回地一次抽一张。设 A 事件表示第一张牌为红桃，事件 B 表示第二张牌为红色。求 $P(A|B)$ 和 $P(B|A)$.

解: 由题意易知

$$P(A \cap B) = \frac{13 \cdot 25}{52 \cdot 51} = \frac{25}{204}$$
$$P(B) = \frac{26 \cdot 51}{52 \cdot 51} = \frac{1}{2}$$
$$P(A) = \frac{1}{4}$$

从而，

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{25/204}{1/2} = \frac{25}{102}$$
$$P(B|A) = \frac{P(B \cap A)}{P(B)} = \frac{25/204}{1/4} = \frac{25}{51}$$

条件概率的注

- 条件概率大小与原事件无条件概率并没有明确的大小关系，见教材例 6.5.
- 注意哪些事件放在竖线的哪一边是非常重要的，具体来说就是 $P(A|B) \neq P(B|A)$.
- 无论 $P(A|B)$ 还是 $P(B|A)$ 都是有意义的 (直观上或数学上):
  - 牌抽取的时间顺序并不能决定出现何种条件概率.
  - 在计算条件概率时，我们考虑的是一个事件给另一个事件带来的信息，而不是一个事件是否导致了另一个事件.
- 此外，也可以通过条件概率的直接解释得出 $P(B|A) = 25/51$:
  - 如果第一张抽的牌为红桃，那么剩下的牌就由 25 张红色牌和 26 张黑色牌组成 (所有牌被下一次抽中的可能性是相同的)
  - 所以抽取一张红牌的条件概率是 $25/(25+26) = 25/51$.

条件概率的性质

定理 2.11 条件概率是概率，即若设 $P(B) > 0$, 则

- 非负性: $P(A|B) \geq 0, \forall A \in \mathcal{F}$;
- 规范性: $P(\Omega|B) = 1$;
- 可列可加性：若 $A_n \in \mathcal{F}, n = 1, 2, \cdots$, 互不相容，则

$$P(\cup_{n=1}^{\infty} A_n|B) = \sum_{n=1}^{\infty} P(A_n|B)$$

证明: 由条件概率的定义易证非负性与规范性。下面说明可列可加性。假设 $A_n, n = 1, \cdots$, 互不相容，则 $A_nB, n = 1, \cdots$, 也互不相容。从而

$$\begin{aligned} P(\cup_{n=1}^{\infty} A_n|B) &= \frac{P((\cup_{n=1}^{\infty} A_n)B)}{P(B)} = \frac{P(\cup_{n=1}^{\infty}(A_nB))}{P(B)} \\ &= \sum_{n=1}^{\infty} \frac{P(A_nB)}{P(B)} = \sum_{n=1}^{\infty} P(A_n|B). \end{aligned}$$

乘法公式

定理 2.12 (乘法公式)

- 若 $P(A) > 0$, 则

$$P(AB) = P(A)P(B|A) \tag{8}$$

- 若 $P(A_1A_2\cdots A_{n-1}) > 0$, 则

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{9}$$

证明: 由条件概率的定义，移项即得 (8) 式，下证 (9). 因为

$$P(A_1) \geq P(A_2) \geq \cdots \geq P(A_1A_2\cdots A_{n-1}) > 0$$

故 (9) 式中条件概率均有意义。从而由两个事件的乘法公式可得

$$\begin{aligned}
&P(A_1A_2\cdots A_n) = P(A_1A_2\cdots A_{n-1})P(A_n|A_1A_2\cdots A_{n-1}) \\
&= P(A_1\cdots A_{n-2})P(A_{n-1}|A_1\cdots A_{n-2})P(A_n|A_1A_2\cdots A_{n-1}) \\
&= \cdots = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})
\end{aligned}$$

全概率公式的引入

例 2.46 有两个罐子

- 在第一个罐中放有 7 个红球、2 个白球和 3 个黑球
- 在第二个罐中放有 5 个红球、4 个白球和 3 个黑球
- 从第一个罐中随机取出 1 个球放入第二个罐中
- 再从第二个罐中随机取出 1 个球来
- 试求 $B=\{$从第二个罐中随机取出的球为红球$\}$ 的概率

问题简要分析:

- 从第二个罐中随机取出的球这一随机试验结果依赖于第一次从第一个罐中所取球的结果;
- 因此, 无论是计算 $|B|$ 还是直接计算 $P(B)$ 都不很容易
- 若已知第一次从第一个罐中所取球的结果后, 则从第二个罐中随机取出的球为红球的概率很容易计算.

全概率公式的引入

解:

- 以 A_1, A_2, A_3 分别表示由第一个罐子取出的是红球, 白球和黑球;
- 显然 $A_k, k=1,2,3$ 两两互不相容且 $\bigcup_{k=1}^{3} A_k = \Omega$;
- $B = \bigcup_{k=1}^{3} A_k B$;
- $P(B) = \sum_{k=1}^{3} P(A_k B) = \sum_{k=1}^{3} P(A_k) P(B|A_k)$;
  - $P(A_1) = \dfrac{7}{12}, P(A_2) = \dfrac{1}{6}, P(A_3) = \dfrac{1}{4}$;
  - $P(B|A_1) = \dfrac{6}{13}, P(B|A_2) = \dfrac{5}{13}, P(B|A_3) = \dfrac{5}{13}$;
- 故 $P(B) = \dfrac{7}{12} \cdot \dfrac{6}{13} + \dfrac{1}{6} \cdot \dfrac{5}{13} + \dfrac{1}{4} \cdot \dfrac{5}{13} = \dfrac{67}{156}$;

全概率公式

定理 2.13 (全概率公式) 设 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割, 即 $A_1, \cdots, A_n$ 互不相容, 且 $\cup_{k=1}^n A_k = \Omega$, 若 $P(A_k) > 0, k = 1, 2, \cdots, n$, 则对任一事件 B 有

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

证明: 因为

$$B = B\Omega = B(\cup_{k=1}^n A_k) = \cup_{k=1}^n (BA_k)$$

且 $BA_1, BA_2, \cdots, BA_n$ 互不相容, 所以由有限可加性得

$$\begin{aligned} P(B) &= P(\cup_{k=1}^n (BA_k)) = \sum_{k=1}^{n} P(BA_k) \\ &= \sum_{k=1}^{n} P(A_k)P(B|A_k) \end{aligned}$$

全概率公式图示

图 2.3　A_i 构成了样本空间的划分；$P(B)$ 等于 $\sum_i P(B \cap A_i)$。

全概率公式 (续)

对于全概率公式, 我们要注意以下几点

- 全概率公式的最简单形式: 如果 $0 < P(A) < 1$, 则

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

- 条件 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割, 可改为 $A_1, A_2, \cdots, A_n$ 互不相容, 且 $B \subset \cup_{k=1}^{n} A_k$
- 在必要时, 可把分割的概念推广到可列个事件 $(n = \infty)$ 的情形, 相应地有

$$P(B) = \sum_{k=1}^{\infty} P(A_k B) = \sum_{k=1}^{\infty} P(A_k)P(B|A_k).$$

贝叶斯公式

定理 2.14 设 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个分割，即 $A_1, A_2, \cdots, A_n$ 互不相容，且 $\cup_{k=1}^{n} A_k = \Omega$，若

$$P(B) > 0, P(A_k) > 0, k = 1, 2, \cdots, n$$

则

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}.$$

特别的，我们有

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

贝叶斯几率

定义 2.25 (几率) 一个事件的几率 (odds) 为

$$\text{odds}(A) = P(A)/P(\overline{A}) = \frac{P(A)}{1-P(A)}$$

概率的几率表示: $P(A) = \text{odds}(A)/(1+\text{odds(A)})$

定理 2.15 (贝叶斯公式的几率形式) 对于任意两个正概率事件 A 和 B, 给定以 B 为条件的情况下, A 的几率如下:

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

- 后验几率 $\dfrac{P(A|B)}{P(\overline{A}|B)}$ 等于先验几率 $\dfrac{P(A)}{P(\overline{A})}$ 乘以似然比因子 $\dfrac{P(B|A)}{P(B|\overline{A})}$.
- 有时候用上述形式的贝叶斯准则可以更方便地求出后验几率, 如果需要的话还可以将几率形式转换成概率形式形式.

全概率与贝叶斯内涵: 由因到果, 由果推因

- 在现实中把事件 B 看作结果, 把事件 $A_1, A_2, \cdots, A_n$ 看作导致结果 B 的各种原因
- 全概率公式是由各种原因推理出结果事件发生的概率, 是由因到果
- 在日常生活中常常是观察到某种现象, 反推造成这种现象的各种原因的概率, 即由果推因
- 条件概率 $P(A_k|B)$, 就是在观察到结果事件 B 已经发生的情况下, 推断结果事件 B 是由原因 A_k 造成的概率的大小
- $P(A_k)$: 先验概率, 在没有别的前提信息情况下的概率值, 一般借助经验来估计, 或赋予所有原因以相同的先验概率
- $P(A_k|B)$: 后验概率, 在获得"结果事件 B 发生"这个信息之后原因 A_k 出现的概率
- 后验概率可看作先验概率在获取了新信息之后的一种修正, 贝叶斯公式恰恰提供了一种计算后验概率的工具
- 贝叶斯理论对于人工智能、深度学习等理论具有重要的指导意义, 贝叶斯统计受到了从未有过的青睐, 迎来了前所未有的发展机遇

条件全概率公式和条件贝叶斯公式

定理 2.16 (条件全概率公式) 令 $A_1, \cdots, A_n$ 为样本空间 Ω 的一个划分. 设对于所有的 i 满足 $P(A_iE) > 0$, 则

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E)$$

定理 2.17 (条件贝叶斯公式) 设 $P(AE) > 0$ 且 $P(BE) > 0$, 则有

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

条件性思维谬误：控方证人的错误

例 2.47 (控方证人的错误) 1988 年，Sally Clark 由于她的两个孩子在出生不久便死亡，因而被指控谋杀幼童。在审讯期间，控方的一个专家证人证实

- 新生儿因婴儿猝死综合症 (SIDS) 而死亡的概率为 $1/8500$
- 所以两个新生儿由于婴儿猝死综合症死亡的概率为 $(1/8500)^2$，大约为 7300 万分之一
- 因此，他认为 Clark 清白的概率仅为 7300 万分之一.

解: 这个推理过程至少有两个问题

- 一个家庭内部成员之间死于 SIDS 是否相互独立？
  - 家庭内部成员之间死于 SIDS 相互独立时，“第一个孩子死于 SIDS” 且 “第二个孩子也死于 SIDS” 的概率是相应的两个事件概率相乘
  - 如果遗传因素或其他家庭特有的风险因素导致某些家庭内的所有新生儿面临 SIDS 的风险增加，这种独立性就不再成立了.

条件性思维谬误：控方证人的错误

- 这个所谓的专家将两个不同的条件概率混淆了：$P(\text{清白} \mid \text{证据})$ 和 $P(\text{证据} \mid \text{清白})$ 是不一样的.
  - 专家 1 声称：如果在被告人是清白的情况下，两个孩子死亡的概率很低；那就是说 $P(\text{证据} \mid \text{清白})$ 非常小.
  - 但人们感兴趣的是，给定现在所有的证据 (孩子均死) 条件下，报告人仍清白的概率，即 $P(\text{清白} \mid \text{证据})$.
  - 由贝叶斯准则可知，

$$P(\text{清白} \mid \text{证据}) = \frac{P(\text{证据} \mid \text{清白})P(\text{清白})}{P(\text{证据})},$$

  - 所以为了计算 $P(\text{清白} \mid \text{证据})$, 这里需要考虑 $P(\text{清白})$, 也就是被告清白的先验概率。这个概率是很高的；
  - 虽然 SIDS 造成两个婴儿死亡是很罕见的，但是蓄意杀害两个婴儿的情况也很少见！
  - 基于现有证据的后验概率是对很低的 $P(\text{证据} \mid \text{清白})$ 和很高的 $P(\text{清白})$ 的一个平衡。专家的结果 $(1/8500)^2$ 是有问题的，它只是整个计算式中的一部分.

Monty Hall 三门问题

例 2.48 在 Monty Hall 主持的 “Let’s Make a Deal” 节目中，有三扇门，其中有两扇门后面是一只山羊，一扇门后面是一辆车. 选手将获得其所选中的那扇门后面的物品.

- 一位选手从三扇最近的门中选一扇
- Monty 知道车在哪扇门后面，并以不暴露车的位置打开剩下两扇门中的一扇，即他打开的门后面永远是山羊.
- 若剩下的两扇门后都是山羊的话，Monty 会等可能地随机选一扇门.
- 然后 Monty 会让对手选择，是换另一扇没打开的门还是不换.
- 如果对手的目标是得到车，她应该换吗?

Monty Hall 问题 (续)

解:

- 先将三扇门从 1 到 3 编号.
- 不失一般性, 可以假设选手选择的是 1 号门
- 令

$$
\begin{aligned}
A &= \{\text{换门得到车}\} \\
B &= \{\text{不换门得到车}\} \\
C_i &:= \{\text{车在第}i\text{扇门后面}\}, i = 1, 2, 3.
\end{aligned}
$$

- 显然, $P(B) = P(\text{不换门得到车}) = 1/3$
- 则由全概率公式

$$
\begin{aligned}
P(A) &= P(A|C_1) \cdot \frac{1}{3} + P(A|C_2) \cdot \frac{1}{3} + P(A|C_3) \cdot \frac{1}{3} \\
&= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}
\end{aligned}
$$

Monty Hall 三门问题图示

图 2.5　蒙提·霍尔（Monty Hall）问题的树状图，换门赢车的概率为 2/3。

随机徘徊的吸收概率

例 2.49 质点在数轴上整数点运动, 若质点处在整数点 i 上, 则下一时刻

- 以概率 p 向右移动到 $i+1$
- 以概率 q 向左移动到 $i-1$, 其中 $p,q>0, p+q=1$
- 我们把质点的上述运动称之为随机徘徊或随机游动
- 考虑数轴上的两个吸收壁 0 与 $a(a>1)$: 质点运动到 0 或 a 之后就永远不再移动
- 现求自 $i(0<i<a)$ 出发的质点将被 0 或 a 吸收的概率

解: 令

$E_i = \{质点自i出发\}, F = \{质点在0被吸收\}, G = \{质点在a被吸收\}.$

下面对 $i = 0, 1, \cdots, a$ 计算

$$P_i = P(F|E_i), \quad Q_i = P(G|E_i)$$

记 $B = \{质点第一次向左移动\ 1\ 单位\}$ 并以第 1 次可能的运动情况为条件进行全概率展开可得

$$\begin{aligned} Q_i &= P(GB|E_i) + P(G\bar{B}|E_i) \\ &= P(B|E_i)P(G|BE_i) + P(\bar{B}|E_i)P(G|\bar{B}E_i) \\ &= pQ_{i+1} + qQ_{i-1}, \quad i = 1, 2, \cdots a-1. \end{aligned}$$

将上述递推式重写为

$$Q_{i+1} - Q_i = \frac{q}{p}(Q_i - Q_{i-1}), \quad i = 1, \cdots, a-1. \tag{10}$$

1. 当 $p=q=1/2$ 时, 由 $Q_0=0$ 及 $Q_a=1$ 可得

$$Q_i=\frac{i}{a},\quad i=0,\cdots,a.$$

2. 当 $p\neq q$ 时, 反复利用递推式(10)并注意到 $Q_0=0$ 可导出

$$Q_{i+1}-Q_i=\left(\frac{q}{p}\right)^i Q_1,\quad i=1,\cdots,a-1. \tag{11}$$

上式对 $i=1,\cdots,a-1$ 求和并利用 $Q_a=1$ 可得

$$Q_1=\left[\sum_{i=0}^{a-1}\left(\frac{q}{p}\right)^i\right]^{-1}=\frac{1-q/p}{1-(q/p)^a}$$

再将 (11) 式对 $i=1,\cdots,i-1$ 求和并将上述 Q_1 代入可得

$$Q_i=\frac{1-(q/p)^i}{1-(q/p)^a},\quad i=0,\cdots,a.$$

综上可得

$$Q_i = \begin{cases} \dfrac{i}{a}, & p = q, \\ \dfrac{1-(q/p)^i}{1-(q/p)^a}, & p \neq q. \end{cases}$$

类似可得

$$P_i = \begin{cases} \dfrac{a-i}{a}, & p = q, \\ \dfrac{(q/p)^i-(q/p)^a}{1-(q/p)^a}, & p \neq q. \end{cases}$$

Simpson 悖论

例 2.50 有两种治疗肾结石的方案，其治疗效果如下：

- 方案 1: 小结石患者占 25%，大结石患者占 75%，小结石患者的治愈率是 93%，大结石患者的治愈率是 73%
- 方案 2: 小结石患者占 77%，大结石患者占 23%，小结石患者的治愈率是 87%，大结石患者的治愈率是 69%
- 不管是对小结石患者，还是大结石患者，方案 1 的治愈率都要高于方案 2
- 方案 1 优于方案 2 吗?

解: 计算两种方案的治愈率

- 记 A={患者是小结石患者}，B={患者被治愈}
- 根据全概率公式

$$P_1(B) = P_1(A)P_1(B|A) + P_1(\overline{A})P_1(B|\overline{A}) = 0.25\cdot 0.93 + 0.75\cdot 0.73 = 0.78$$

$$P_2(B) = P_2(A)P_2(B|A) + P_2(\overline{A})P_2(B|\overline{A}) = 0.77\cdot 0.87 + 0.23\cdot 0.69 = 0.8286$$

- 方案 2 的治愈率高于方案 1，可见方案 1 并不优于方案 2

敏感性问题调查

敏感性问题调查方案的关键在于要使被调查者愿意作出真实的回答又能保守个人秘密, 如果调查方案有误, 被调查者就会拒绝配合, 所得调查数据将失去真实性.

调查方案: 被调查者只需要加答以下两个问题中的一个问题, 且只需要回答“是”或“否”.
问题 A: 你的生日是否在 7 月 1 日之前?
问题 B: 所调查的敏感性问题.

调查方案的操作:
(1) 被调查者在没有旁人的情况下独自一人在房间内操作回答问题
(2) 被调查者从一个罐子中随机抽一球, 看过颜色后放回, 若抽到白球, 回答问题 A, 抽到红球, 回答问题 B.
(3) 被调查者无论回答问题 A 还是问题 B, 只需在仅有“是”与“否”选项的答卷上作答然后将答卷放入密封的投票箱内.

敏感性问题调查

问题: 假若我们有 n 张问卷, 其中 k 张回答“是”, 我们如何确定选定红球回答问题 B 为“是”的概率?

已知:

- 红白球的比例, 即 $P(\text{红球}) := \pi, P(\text{白球}) = 1 - \pi$;
- $P(\text{是}|\text{白球}) = 0.5$;
- $P(\text{是}) \approx \dfrac{k}{n}$.

待求: $p := P(\text{是}|\text{红球})$.

解: 由全概率公式可得

$$P(\text{是}) = P(\text{白球})P(\text{是}|\text{白球}) + P(\text{红球})P(\text{是}|\text{红球})$$

即

$$\frac{k}{n} \approx (1-\pi)\cdot 0.5 + \pi \cdot p \Rightarrow p = \frac{k/n - 0.5(1-\pi)}{\pi}$$

Bayesian 滤波

- 许多工程科学类问题都可以归结为如下图所示形式

- 系统自身的状态 X 是我们关注的随机事件，但是由于技术手段等方面的限制，无法直接对 X 进行观测，只能获取另一随机事件，即间接反映系统状态 X 的观测量 Z.
- 问题转化为已知观测量 Z 的条件下，如何对系统状态 X 进行推断. 用条件概率的语言讲，就是计算 $P(X \mid Z)$.
- 通常情况下，X 和 Z 随时间变化，分别记作 X_n 和 Z_n，其中 n 表示时间. 那么在实际应用中，到 n 时刻为止，我们掌握的实际观测数据为 $\{Z_1, \cdots, Z_n\}$,

Bayesian 滤波

- 如何实现 $P(X_n|Z_1,\cdots,Z_n) \longrightarrow P(X_{n+1}|Z_1,\cdots,Z_{n+1})$ 的递推计算?
- 注意到, 若已知 X_{n+1} 时,Z_{n+1} 与 $\{Z_1,\cdots,Z_n\}$ 独立, 则

$$
\begin{aligned}
P(X_{n+1}|Z_1,\cdots,Z_{n+1}) &= \frac{P(X_{n+1},Z_1,\cdots,Z_{n+1})}{P(Z_1,\cdots,Z_{n+1})} \\
&= \frac{P(Z_{n+1}|X_{n+1},Z_1,\cdots,Z_n)\,P(X_{n+1},Z_1,\cdots,Z_n)}{P(Z_1,\cdots,Z_{n+1})} \\
&= \frac{P(Z_{n+1}|X_{n+1})\,P(X_{n+1},Z_1,\cdots,Z_n)}{P(Z_{n+1}|Z_1,\cdots,Z_n)P(Z_1,\cdots,Z_{n+1})}
\end{aligned}
$$

Bayesian 滤波

- 再由

$$
\begin{aligned}
& P\left(X_{n+1}, Z_1, \cdots, Z_n\right) \\
= & \sum_{X_n} P\left(X_{n+1}, X_n, Z_1, \cdots, Z_n\right) \\
= & \sum_{X_n} P\left(X_{n+1} | X_n, Z_1, \cdots, Z_n\right) P\left(X_n | Z_1, \cdots, Z_n\right) P\left(Z_1, \cdots, Z_n\right) \\
= & \sum_{X_n} P\left(X_{n+1} | X_n\right) P\left(X_n | Z_1, \cdots, Z_n\right) P\left(Z_1, \cdots, Z_n\right)
\end{aligned}
$$

- 将上式代入前面一页表达式中可得

$$
P\left(X_{n+1} | Z_1, \cdots, Z_{n+1}\right) = \frac{P\left(Z_{n+1} | X_{n+1}\right)}{P\left(Z_{n+1} | Z_n, \cdots, Z_1\right)} \sum_{X_n} P\left(X_{n+1} | X_n\right) P\left(X_n | Z_1, \cdots, Z_n\right)
$$

- 上式便是对系统状态进行递推估计的 Bayesian 滤波公式，在雷达、声呐、通信、导航、机器学习等许多领域有广泛应用.

第七讲：条件概率、全概率与贝叶斯公式的应用

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

条件概率的例子：男孩女孩概率问题

例 2.51 (年长的是女孩 vs 至少一个女孩) 某家庭有两个孩子，已知至少有一个是女孩。两个孩子都是女孩的概率是多少？如果条件改为年长的孩子是女孩，那么两个都是女孩的概率又是多少？

解: 假设每个孩子都是女孩和男孩的可能性相同且不相关，那么

$$P(\text{都是女孩} \mid \text{至少有一个是女孩}) \tag{12}$$

$$= \frac{P(\text{都是女孩，至少有一个是女孩})}{P(\text{至少有一个是女孩})} = \frac{1/4}{3/4} = 1/3 \tag{13}$$

$$P(\text{都是女孩} \mid \text{年长的是女孩}) \tag{14}$$

$$= \frac{P(\text{都是女孩，年长的是女孩})}{P(\text{年长的是女孩})} = \frac{1/4}{1/2} = 1/2 \tag{15}$$

男孩女孩概率问题续：随机的一个孩子是女孩

例 2.52 某家庭有两个孩子。随机遇到其中的一个，发现是女孩。给定这个信息后，两个孩子都是女孩的概率是多少？假设随机遇到两个孩子的可能性相同，且与性别无关.

解:

- 直观来看，结果应为 1/2.
- 令 G_1, G_2, G_3 分别表示年长、年幼、随机的孩子是女孩这三个事件。由对称性可得: $P(G_1) = P(G_2) = P(G_3) = 1/2$
- 根据朴素概率的定义，或者独立性，可得: $P(G_1 \cap G_2) = 1/4$
- 因此，$P(G_1 \cap G_2 | G_3) = P(G_1 \cap G_2 \cap G_3)/P(G_3) = 1/2$
- 又因为 $G_1 \cap G_2 \cap G_3 = G_1 \cap G_2$, 所以概率为 1/2.

注 2.8 假设一个强制性法律规定：如果一个男孩有姐妹则禁止他走出家门。那么这时“随机遇到的孩子是女孩”就等价于“ 至少有一个孩子是女孩”

男孩女孩概率问题续：冬天出生的女孩

例 2.53 某家庭有两个孩子，给定条件至少一个是女孩且在冬天出生，求两个孩子都是女孩的概率。假设四个季节出生的可能性相同且性别和季节是相互独立的.

解: 由条件概率的定义，可得:

$$
\begin{aligned}
&P(\text{两个都是女孩} \mid \text{至少有一个是冬天出生的女孩}) \\
&= \frac{P(\text{两个都是女孩，至少有一个是冬天出生的女孩})}{P(\text{至少有一个是冬天出生的女孩})}
\end{aligned}
$$

由于指定的孩子是在冬天出生的女孩的概率为 1/8, 所以,

$$P(\text{至少有一个是在冬天出生的女孩}) = 1 - (7/8)^2.$$

男孩女孩概率问题续：冬天出生的女孩

利用性别和季节是相互独立的假设，得到:

$$
\begin{aligned}
&P(\text{两个都是女孩，至少有一个是冬天出生的女孩})\\
&= P(\text{两个都是女孩，至少有一个是冬天出生的})\\
&= (1/4)P(\text{至少有一个是冬天出生的女孩})\\
&= (1/4)(1 - P(\text{所有孩子都不是在冬天出生的})
\end{aligned}
$$

合在一起得到，

$$P(\text{两个都是女孩} \mid \text{至少有一个是冬天出生的女孩}) = 7/15$$

结绳问题

例 2.54 n 根绳 $2n$ 个头两两相接，求 $A=\{$恰好结成n个圈$\}$ 的概率.

- 设想 $2n$ 个头排成一行，规定将第 $2k-1$ 个头与第 $2k$ 个端头相接;
- 令 B_i 表示第 i 根绳的头与尾恰好相接，则 $A=B_1B_2\cdots B_n$;
- 若以 $n(A)$ 表示事件 A 所包含的样本点个数，则易知

$$n(\Omega)=(2n)!, n(B_1)=2n(2n-2)!\Rightarrow P(B_1)=\frac{n(B_1)}{n(\Omega)}=\frac{1}{2n-1};$$

- $P(B_2|B_1)$ 可看作 $n-1$ 根绳某根绳头尾相接的概率，类比 n 根绳情形可得 $P(B_2|B_1)=\dfrac{1}{2(n-1)-1}=\dfrac{1}{2n-3}$;
- 同理可得$P(B_k|B_1B_2\cdots B_{k-1})=\dfrac{1}{2[n-(k-1)]-1}=\dfrac{1}{2n-2k+1}, 3\le k\le n$;
- 利用乘法公式可得

$$\begin{aligned}P(A)&=P(B_1B_2\cdots B_n)=P(B_1)P(B_2|B_1)\cdots P(B_n|B_1\cdots B_{n-1})\\&=\frac{1}{(2n-1)!!}\end{aligned}$$

取球问题

例 2.55 在计算机中输入程序，让它自动完成如下操作:

- 在 $1-\frac{1}{2^n}$ 时刻，往盒中放入标号 $10(n-1)+1 \sim 10n$ 的 10 个球，同时取出标号为 $10(n-1)+1$ 的球，$n \geq 1$;
- 在 $1-\frac{1}{2^n}$ 时刻，往盒中放入标号 $10(n-1)+1 \sim 10n$ 的 10 个球，同时取出标号为 n 的球，$n \geq 1$;
- 在 $1-\frac{1}{2^n}$ 时刻，往盒中放入标号 $10(n-1)+1 \sim 10n$ 的 10 个球，同时随机地从盒中取出一个球，$n \geq 1$.

则在时刻 1, 盒中的球数结果如下

- 盒子中有无穷多个球;
- 盒子变为空的;
- 盒子变为空的概率等于 1?

取球问题

解:

- 记 E={在时刻 1 时盒子变空 };
- $\overline{E}$={在时刻 1 时盒中有球未被取出 };
- 记 A_k={在时刻 1 时 k 号球仍在盒中未被取出 }，则 $\overline{E} = \bigcup\limits_{k=1}^{\infty} A_k$;
- 由概率的次可加性知

$$P(\overline{E}) = P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k);$$

- 为证 $P(\overline{E}) = 0$，只需证明 $P(A_k) = 0, k \geq 1$;
- 由于证法类似，仅以证明 $P(A_1) = 0$ 为例;

取球问题

- 记 B_n={在 $1-\dfrac{1}{2^n}$ 时刻 1 号球未被取出 }，易知 $A_1=\bigcap\limits_{n=1}^{\infty}B_n$;
- 令 $C_m=\bigcap\limits_{n=1}^{m}B_n$, 则有

$$C_m=\bigcap_{n=1}^{m}B_n\supset\bigcap_{n=1}^{m+1}B_n=C_{m+1},\text{ 且 }\lim_{m\to\infty}C_m=A_1;$$

- 由概率的上连续性知

$$P(A_1)=P\big(\lim_{m\to\infty}C_m\big)=\lim_{m\to\infty}P\big(C_m\big);$$

- 下面求 $P(C_m)$, 即 $P\big(\bigcap_{n=1}^{m}B_n\big)=\prod_{n=1}^{m}\frac{9n}{9n+1}=\prod_{n=1}^{m}\left(1-\frac{1}{9n+1}\right)$;
  - B_1: 在 1/2 时刻，盒中 10 个球，1 号球未被取出，故 $P(B_1)=\frac{9}{10}$;
  - B_2: 在 3/4 时刻，盒中 19 个球，1 号球未被取出，故 $P(B_2|B_1)=\frac{18}{19}$;
  - B_n: $P(B_n|B_1B_2\cdots B_{n-1})=\frac{9n}{9n+1}$.
- $P(A_1)=\prod_{n=1}^{\infty}\left(1-\frac{1}{9n+1}\right)=0$. 同理可证 $P(A_k)=0$, $k=2,3,\cdots$.

罐子模型（波利亚模型）

例 2.56 设罐中有 b 个黑球，r 个红球，每次随机的取出一球，取出后将原球放回，还加进 c 个同色球和 d 个异色球。记 $B_i := \{第i次取出的是黑球\}, R_j := \{第j次取出的是红球\}$. 若连续从罐子中取出三个球，其中有两个红球，一个黑球，则由乘法公式可得

$$
\begin{aligned}
P(B_1R_2R_3) &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2) \\
&= \frac{b}{b+r}\cdot\frac{r+d}{b+r+c+d}\cdot\frac{r+d+c}{b+r+2c+2d} \\
P(R_1B_2R_3) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2) \\
&= \frac{r}{b+r}\cdot\frac{b+d}{b+r+c+d}\cdot\frac{r+d+c}{b+r+2c+2d} \\
P(R_1R_2B_3) &= P(R_1)P(R_2|R_1)P(B_3|R_1R_2) \\
&= \frac{r}{b+r}\cdot\frac{r+c}{b+r+c+d}\cdot\frac{b+2d}{b+r+2c+2d}
\end{aligned}
$$

显然以上概率与黑球在第几次抽取有关.

罐子模型（波利亚模型）

- 当 $c=-1, d=0$ 时，即为不返回抽样。此时前次抽取结果会影响后次抽取结果，但只要抽取的黑球与红球个数确定，则概率不依赖其抽出球的次序，都是一样的.

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)=\frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}.$$

- 当 $c=0, d=0$ 时，即为返回抽样。此时前次抽取结果不会影响后次抽取结果，故上述三个概率相等且都等于

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)=\frac{br^2}{(b+r)^3}.$$

- 当 $c>0, d=0$ 时，称为传染病模型。此时每次取出球后会增加下一次取出同色球的概率，或换言之，每发现一个传染病患者，以后都会增加再传染的概率。与前两种情况一样，三个概率都等于

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)=\frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}.$$

罐子模型（波利亚模型）

- 从上面的结果可以看出，只要 $d=0$，以上三个概率都相等，即只要抽取的黑球与红球的个数确定，则概率不依赖于抽出黑红球的次序.
- 当 $c=0, d>0$ 时，称为安全模型。此模型可解释为：每当事故发生了 (当红球被取出)，安全工作就抓紧一些，下次再发生事故的概率就会减少，而当事故没有发生时 (黑球被取出)，安全工作就放松一些，下次再发生事故的概率就会增大，此时，上述三个概率分别为

$$
\begin{aligned}
P(B_1R_2R_3) &= \frac{b}{b+r}\cdot\frac{r+d}{b+r+d}\cdot\frac{r+d}{b+r+2d} \\
P(R_1B_2R_3) &= \frac{r}{b+r}\cdot\frac{b+d}{b+r+d}\cdot\frac{r+d}{b+r+2d} \\
P(R_1R_2B_3) &= \frac{r}{b+r}\cdot\frac{r}{b+r+d}\cdot\frac{b+2d}{b+r+2d}
\end{aligned}
$$

罐子模型（波利亚模型）

例 2.57 设罐中有 b 个黑球，r 个红球，每次随机取出一球后将原球放回并加进 c 个同色球，如此反复进行。试证明：在前 $n = n_1 + n_2$ 次取球中，取出了 n_1 个红球和 n_2 个黑球的概率为

$$C_n^{n_1} \frac{a(a+c)(a+2c)\cdots(a+n_1c-c)b(b+c)(b+2c)\cdots(b+n_2c-c)}{(a+b)(a+b+c)(a+b+2c)\cdots(a+b+nc-c)}.$$

摸球问题

例 2.58 有三个罐子, 各装有两个球, 分别为两个白球、一白一黑和两个黑球. 任意取出一个罐子, 摸出一球, 发现是白球. (1) 求该罐中另一个球也是白球的概率; (2) 把摸出的球放回罐中, 再从该罐中随机摸出一球, 求该球也是白球的概率.

解:

- $A_k, k = 1, 2$: 表示第 k 次取球取出的是白球的事件;
- $B_k, k = 1, 2, 3$: 表示取出的是装有两白、一白一黑和两黑球的罐子;
- 问题 (1) 要求的是该白球取自两白的罐子的概率, 即 $P(B_1|A_1)$;
- 由条件概率公式, 得 $P(B_1|A_1) = \frac{P(A_1B_1)}{P(A_1)} = \frac{P(B_1)P(A_1|B_1)}{P(A_1)}$;
  - $P(B_1) = 1/3$, $P(A_1|B_1) = 1$;
  - $P(A_1) = \sum_{k=1}^{3} P(B_k)P(A_1|B_k) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 = \frac{1}{2}$.
- $P(B_1|A_1) = \dfrac{1/3 \cdot 1}{1/2} = 2/3$.

摸球问题

- 问题 (2) 是在同一个罐子两次有放回的取球, 要求的是在第一次取出白球的条件下, 第二次取出的还是白球的条件概率, 即 $P(A_2|A_1)$
- 由条件概率的定义, 知$P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)}$;
  - $P(A_1) = \sum_{k=1}^{3} P(B_k)P(A_1|B_k) = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0 = \frac{1}{2}.$
  - $P(A_1A_2) = \sum_{k=1}^{3} P(B_k)P(A_1A_2|B_k) = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{4} + \frac{1}{3}\cdot 0 = \frac{5}{12}.$
- $P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)} = \frac{5/12}{1/2} = \frac{5}{6}.$

摸球问题

例 2.59 甲盒中有球 5 红 1 黑, 乙盒中有球 5 红 3 黑. 随机取出一个盒子, 从中无放回地相继取出两个球, 试求在第一个球是红球的条件下, 第二个球也是红球的概率.

解:

- $B :=\{$第一个球是红球$\}$, $C :=\{$第二个球是红球$\}$
- $P(C|B) = \dfrac{P(BC)}{P(B)}$
  - $A :=\{$取出的是甲盒$\}$
  - $P(BC) = P(A)P(BC|A) + P(\overline{A})P(BC|\overline{A}) = \dfrac{1}{2}\cdot\dfrac{5}{6}\cdot\dfrac{4}{5} + \dfrac{1}{2}\cdot\dfrac{5}{8}\cdot\dfrac{4}{7} = \dfrac{43}{84}$
  - $P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}) = \dfrac{1}{2}\cdot\dfrac{5}{6} + \dfrac{1}{2}\cdot\dfrac{5}{8} = \dfrac{35}{48}$
- $P(C|B) = \dfrac{43}{84}/\dfrac{35}{48} = \dfrac{172}{245}$

一般摸球模型

例 2.60 袋中有 r 个红球与 b 个黑球. 每次从袋中任摸出 1 球并连同 s 个同色球一起放回. 以 R_n 表示第 n 次摸出红球, 试证 $P(R_n) = \dfrac{r}{r+b}$.

证明: 利用归纳法来证明:$n = 1$ 时, $P(R_1) = \dfrac{r}{r+b}$ 显然成立.

假设 $n-1$ 时命题成立. 为求 $P(R_n)$, 我们以第 1 次取球的可能结果 R_1 与 $\bar{R}_1$ 作为 Ω 的分割, 用全概率公式可得:

$$
\begin{aligned}
P(R_n) &= P(R_1)P(R_n|R_1) + P(\bar{R}_1)P(R_n|\bar{R}_1) \\
&= \frac{r}{r+b} \cdot \frac{r+s}{r+s+b} + \frac{b}{r+b} \cdot \frac{r}{r+b+s} \\
&= \frac{r}{r+b}
\end{aligned}
$$

注 2.9 当 $s = 0$ 时相当于放回摸球, 而 $s = -1$ 相当于不放回摸球.

掷硬币问题

例 2.61 甲、乙二人抛掷一枚均匀的硬币，甲抛了 100 次，乙抛了 101 次. 求事件 $E :=\{$乙抛出的正面次数比甲多$\}$ 的概率.

解:

- 如果甲和乙都抛掷这枚均匀的硬币 100 次，那么当然会有三种不同的可能结果
  - A_0:甲乙抛出的正面次数一样多
  - A_1:甲抛出的正面次数比乙多
  - A_2:乙抛出的正面次数比甲多
  - $P(A_0)+P(A_1)+P(A_2)=1$且 $P(A_1)=P(A_2)$
- 以乙第一次抛出的硬币结果对样本空间进行分割:
  - 若 $B=\{$乙第一次时抛出的是正面$\}$ 发生，则乙只要在接下来的 100 次抛掷中，抛出的正面次数不比甲少即可，即

$$P(E|B)=P(A_0)+P(A_1)$$

  - 若 $\overline{B}:=\{$乙第一次时抛出的是反面$\}$ 发生，则乙只要在接下来的 100 次抛掷中，抛出的正面次数比甲多即可，即 $P(E|\overline{B})=P(A_1)=P(A_2)$
- $P(E)=P(B)P(E|B)+P(\overline{B})P(E|\overline{B})=\frac{1}{2}\left(P(A_0)+P(A_1)+P(A_2)\right)=\frac{1}{2}$.

对战获胜概率问题

例 2.62 甲、乙进行某项对战比赛, 每回合胜者得 1 分, 败者不得分. 比赛进行到有 1 人比另外 1 人多 2 分终止, 多 2 分者获胜. 现知每回合甲胜的概率为 $p \in (0,1)$. 试求 $A :=\{$甲最终获胜$\}$ 的概率.

解: 法一 (经典解法)

- 显然甲只能在偶数个回合后获胜
- 记 $A_{2n}=\{$甲在 $2n$ 个回合后获胜$\}$, 则 $A = \cup_{n=1}^{\infty} A_{2n}$
- $A_{2n}, n \geq 1$ 两两互不相容且$P(A_{2n}) = (2p(1-p))^{n-1}p^2$
- 甲最终获胜的概率

$$
\begin{aligned}
P(A) &= \sum_{n=1}^{\infty} P(A_{2n}) = p^2 \sum_{n=1}^{\infty} (2p(1-p))^{n-1} \\
&= p^2 \sum_{n=0}^{\infty} (2p(1-p))^{n} = \frac{p^2}{1-2p(1-p)}.
\end{aligned}
$$

对战获胜概率问题

解: 法二 (全概率公式)

- 以前两个回合的战绩对样本空间进行分割
- 分别以 B_1, B_2, B_3 表示甲二胜、一胜一败、二败事件
- $P(B_1) = p^2$, $P(B_2) = 2p(1-p)$, $P(B_3) = (1-p)^2$
- $P(A|B_1) = 1$, $P(A|B_2) = P(A)$, $P(A|B_3) = 0$
- 由全概率公式得 $P(A) = \sum_{i=1}^{3} P(B_i)P(A|B_i) = p^2 + 2p(1-p)P(A)$
- $P(A) = \dfrac{p^2}{1-2p(1-p)}$

对战获胜概率问题

解: 法三 (随机游动)

- 考察质点在数轴整数点上的随机游动: 在整数点 $x = n$
  - 以概率 p 向右移动到整数点 $x = n + 1$,
  - 以概率 $1 - p$ 向左移动到整数点 $x = n - 1$
- 以 p_n 表示 "质点由 $x = n$ 出发, 未达 -2 前先到达 2 的概率"
- 显然, $p_{-2} = 0, p_2 = 1$
- 甲获胜的概率即为: p_0
- 由全概率公式易得

$$\begin{cases} p_0 = pp_1 + (1-p)p_{-1}, \\ p_1 = pp_2 + (1-p)p_0 = p + (1-p)p_0, \\ p_{-1} = pp_0 + (1-p)p_{-2} = pp_0. \end{cases}$$

- $p_0 = p(p + (1-p)p_0) + (1-p)(pp_0) = p^2 + 2p(1-p)p_0$
- 故

$$p_0 = \frac{p^2}{1 - 2p(1-p)}.$$

随机抛硬币问题

例 2.63 (随机抛硬币) 假设有一枚均匀的硬币和一枚以概率 3/4 正面朝上的不均匀硬币. 随机选取一枚硬币掷 3 次, 3 次都是正面朝上. 试问:

1. 给定上述信息后, 选取的硬币是均匀硬币的概率有多大?
2. 接下去掷第四次时, 仍是正面朝上的概率是多少?

解: 令

$$A := \{\text{选取的硬币掷 3 次均正面朝上}\}$$

$$F := \{\text{选取的硬币是均匀的}\}, \quad H := \{\text{第 4 次正面朝上}\}$$

则

$$\begin{aligned} P(F|A) &= \frac{P(A|F)P(F)}{P(A)} = \frac{P(A|F)P(F)}{P(A|F)P(F) + P(A|F^c)P(F^c)} \\ &= \frac{(1/2)^3 \cdot 1/2}{(1/2)^3 \cdot 1/2 + (3/4)^3 \cdot 1/2} \approx 0.23 \\ P(H \mid A) &= P(H \mid F, A)\, P(F \mid A) + P(H \mid F^{\mathbf{c}}, A)\, P(F^{\mathbf{c}} \mid A) \\ &= \frac{1}{2} \cdot 0.23 + \frac{3}{4} \cdot (1 - 0.23) \approx 0.69. \end{aligned}$$

罕见病检测诊断问题

例 2.64 考虑以验血结果诊断某种罕见病的患病概率:

- 某罕见病的发病率为 1%
- 通过验血诊断该病的误诊率为 5%, 即非患者中有 5% 的人验血结果为阳性, 患者中有 5% 的人验血结果为阴性
- 现已知某人验血结果为阳性, 试求他患有此病的概率

解:

- 记 $D :=\{$患有此病$\}$, $T_1 :=\{$第一次验血结果为阳性$\}$.
- 要求的概率是: $P(D|T_1) = \dfrac{P(DT_1)}{P(T_1)} = \dfrac{P(D)P(T_1|D)}{P(T_1)}$
- 由题意可知

$$P(D) = 1\%,\ P(\overline{D}) = 99\%,\ P(T_1|D) = 95\%,\ P(T_1|\overline{D}) = 5\%$$

$$\begin{aligned} P(T_1) &= P(D)P(T_1|D) + P(\overline{D})P(T_1|\overline{D}) = 1\% \cdot 95\% + 99\% \cdot 5\% \\ &= 0.0685 \end{aligned}$$

罕见病检测诊断问题

- $P(D|T_1) = \dfrac{P(D)P(T_1|D)}{P(T_1)} = \dfrac{1\% \cdot 95\%}{0.0685} \approx 0.16.$
- 几率方法:

$$\frac{P(D|T_1)}{P(D^c|T_1)} = \frac{P(D)}{P(D^c)}\frac{P(T_1|D)}{P(T_1|D^c)} = \frac{1}{99} \cdot \frac{0.95}{0.05} \approx 0.19$$

  由几率与概率之间的关系可知

$$P(D|T_1) = 0.19/(1+0.19) \approx 0.16,$$

  与上述结果一致.
- 用几率形式的贝叶斯准则计算更迅速的原因是此时不需要计算普通贝叶斯准则的分母.

罕见病检测诊断问题图示

罕见病检测诊断问题

例 2.65 接上例, 检测结果为阳性的某人, 决定进行第二次检测. 假设新的检测结果与之前的结果相互独立, 且有相同的敏感性和特异性. 若第二次检测结果也为阳性, 试求此人患有此病的概率.

解:

- 记 $T_2 :=\{$第二次验血结果为阳性$\}$
- 要求的概率是: $P(D|T_1T_2)$
- 一步法: 将两个检测结果一次性都考虑在内以进行概率更新,
  - 计算几率

$$\frac{P(D|T_1\cap T_2)}{P(D^c|T_1\cap T_2)}=\frac{P(D)}{P(D^c)}\frac{P(T_1\cap T_2|D)}{P(T_1\cap T_2|D^c)}=\frac{1}{99}\cdot\frac{0.95^2}{0.05^2}=\frac{361}{99}\approx 3.646$$

  - $P(D|T_1T_2)=\frac{3.646}{1+3.646}\approx 0.78.$

罕见病检测诊断问题

- 两步法:
  - 在完成第一次检测后, 某人患有此病的后验几率为

$$\frac{P(D|T_1)}{P(D^c|T_1)} = \frac{1}{99} \cdot \frac{0.95}{0.05} \approx 0.19$$

  - 将后验几率作为新的先验几率, 然后基于第二次检测结果更新后验几率

$$\begin{aligned}\frac{P(D|T_1 \cap T_2)}{P(D^c|T_1 \cap T_2)} &= \frac{P(D|T_1)}{P(D^c|T_1)} \frac{P(T_2|D, T_1)}{P(T_2|D^c, T_1)} \\ &= (\frac{1}{99} \cdot \frac{0.95}{0.05}) \frac{0.95}{0.05} \\ &= \frac{361}{99} \approx 3.646\end{aligned}$$

  - $P(D|T_1T_2) = \dfrac{3.646}{1 + 3.646} \approx 0.78$

电邮问题

例 2.66 甲、乙二人之间经常用 E-mail 相互联系，他们约定在收到对方信件的当天即给 E-mail 回复. 由于线路问题，每 n 份 E-mail 中会有 1 份不能在当天送达收件人. 甲在某日发了 1 份 E-mail 给乙，但未在当天收到乙的回音. 试求乙在当天收到了甲发给他的 E-mail 的概率.

解:

- 在这个问题中，包含了两个不确定的环节:
  - 甲发给乙的 E-mail 不一定在当天到达乙处
  - 乙回给甲的 E-mail 不一定在当天到达甲处
- A={乙在当天收到甲的 E-mail}, B={甲在当天收到乙回的 E-mail}
- $P(A|\overline{B}) = \dfrac{P(A)P(\overline{B}|A)}{P(A)P(\overline{B}|A) + P(\overline{A})P(\overline{B}|\overline{A})}$
- 由题中条件知

$$P(A) = \frac{n-1}{n},\ P(\overline{A}) = \frac{1}{n},\ P(\overline{B}|A) = \frac{1}{n},\ P(\overline{B}|\overline{A}) = 1.$$

- $P(A|\overline{B}) = \dfrac{\frac{n-1}{n}\cdot\frac{1}{n}}{\frac{n-1}{n}\cdot\frac{1}{n} + \frac{1}{n}\cdot 1} = \dfrac{n-1}{2n-1} < 1/2$

计算概率的递推方法: 一步分析

例 2.67 (分支过程) 池塘里只有一只变形虫叫作 Bobo.

- 每过 1 分钟, Bobo 有三种结果: 死去、分裂成两个或保持原状
- 三种结果出现的概率相同
- 此后所有活着的 Bobo 都将继续以这种方式相互独立地进行下去
- 那么这个变形虫种族最终灭亡的概率是多少?

分支过程 (续)

解:

- 令 $D := \{最终种族灭绝\}$, 本题希望求出 $P(D)$.
- 我们在第一步结果的基础上即以 1 分钟后的结果进行分析:
  - 令 $B_i := \{1分钟后Bobo变成的变形虫个数\}(i = 0, 1, 2)$
  - 易知 $P(D|B_0) = 1$ 和 $P(D|B_1) = P(D)$, $P(D|B_2) = P(D)^2$.
- 利用全概率公式有

$$\begin{aligned} P(D) &= P(D|B_0) \cdot \frac{1}{3} + P(D|B_1) \cdot \frac{1}{3} + P(D|B_2) \cdot \frac{1}{3} \\ &= 1 \cdot \frac{1}{3} + P(D) \cdot \frac{1}{3} + P(D)^2 \cdot \frac{1}{3}. \end{aligned}$$

- 由上式可解得 $P(D) = 1$, 即变形虫种族最终会以概率 1 灭绝.

对战相遇概率问题

例 2.68 包括甲、乙二人在内的 2^n 名乒乓球运动员参加一场淘汰赛.

- 第一轮任意两两配对比赛, 然后 2^{n-1} 名胜者再任意两两配对进行第二轮比赛, 如此下去, 直至第 n 轮决出一名冠军为止
- 假定每一名运动员在各轮比赛中胜负都是等可能的
- 求 $B :=\{$甲、乙二人在淘汰赛中相遇$\}$ 的概率

对战相遇概率问题

解: 记 $p_n := P(B)$, 即甲、乙二人在 2^n 人参赛的比赛中相遇的概率

- 以甲、乙二人是否在第一轮比赛相遇对样本空间进行分割
- $A :=\{$甲、乙二人在第一轮比赛中相遇$\}$
- 由全概率公式

$$p_{n+1} = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}) = P(A) + (1 - P(A)) \cdot \frac{1}{4}p_n$$

- $P(A)$: 甲、乙二人在第一轮比赛中相遇的概率
  - 采用无编号分组模式考虑
  - 2^{n+1} 个人两两配对的方式一共有 $\dfrac{(2^{n+1})!}{2^{2^n}(2^n)!}$ 种
  - 甲、乙二人配为一对的配对方式有 $\dfrac{(2^{n+1}-2)!}{2^{2^n-1}(2^n-1)!}$ 种
  - 将上述两式相除, 即得 $P(A) = \dfrac{1}{2^{n+1}-1}$
- $p_{n+1} = \frac{1}{2^{n+1}-1} + \frac{1}{4}(1 - \frac{1}{2^{n+1}-1})p_n$
- $p_{n+1} = \frac{1}{2^n}$

掷骰子问题

例 2.69 甲乙轮流掷一均匀骰子. 甲先掷, 以后每当某人掷出 1 点后则交给对方掷, 否则此人继续掷. 试求事件 $A_n = \{第n次由甲掷\}$ 的概率.

解: 记 $p_n = P(A_n)$, 则以 A_{n-} 与 $\bar{A}_{n-1}$ 为分割用全概率公式可得:

$$\begin{aligned} p_n = P(A_n) &= P(A_{n-1})P(A_n|A_{n-1}) + P(\bar{A}_{n-1})P(A_n|\bar{A}_{n-1}) \\ &= p_{n-1}\frac{5}{6} + (1-p_{n-1})\frac{1}{6} = \frac{2}{3}p_{n-1} + \frac{1}{6} \end{aligned}$$

经过整理, 可将上式化为以下递推的形式

$$p_n - \frac{1}{2} = \frac{2}{3}\left(p_{n-1} - \frac{1}{2}\right), n = 2, 3, \cdots.$$

由 $p_1 = 1$ 可得 $p_n - \frac{1}{2} = \left(\frac{2}{3}\right)^{n-1}\left(p_1 - \frac{1}{2}\right) = \frac{1}{2}\left(\frac{2}{3}\right)^{n-1}$. 因此, 我们有

$$p_n = P(A_n) = \frac{1}{2}\left[1 + \left(\frac{2}{3}\right)^{n-1}\right], n = 1, 2, \cdots,$$

结绳问题续

例 2.70 n 根绳 $2n$ 个头两两相接, 求 $A_n = \{$恰好结成n个圈$\}$ 的概率.

解: 此前曾用条件概率解过本题, 现利用全概率公式给出一个解答.

- 将 n 根短绳作编号并记 $p_n = P(A_n)$
- 记 B=$\{1$ 号绳连成 1 个圈$\}$ 并用 B 和 $\overline{B}$ 作为对 Ω 的分划
- 由全概率公式可知

$$p_n = P(A_n) = P(B)P(A_n|B) + P(\overline{B})P(A_n|\overline{B})$$

- $P(B) = \dfrac{1}{2n-1}, P(A_n|\overline{B}) = 0, P(A_n|B) = P(A_{n-1}) = p_{n-1}$
- $p_n = P(A_n) = \frac{1}{2n-1}p_{n-1}$, $n = 2, 3, \cdots$.
- 反复利用上式, 并由 $p_1 = 1$ 可得

$$p_n = \frac{1}{(2n-1)!!}, \ n = 1, 2, \cdots .$$

秘书问题

例 2.71 某公司需招收秘书一名，共有 n 个人报名应聘，公司面试规则与录取策略如下：

- 面试规则：逐个面试，并在面试当时对应聘者表态是否录用，一旦对应聘者表态不录用，不可改变决定
- 录用策略：
  - 不录用前 $k(1 \leq k < n)$ 个面试者
  - 自第 $k+1$ 个开始，只要发现某人比他前面的所有面试者都好，就录用他，否则就录用最后一个
- 试对该公司的策略作概率分析

秘书问题

- $A :=\{$最佳人选被录用$\}$
- $B_j :=\{$最佳人选在面试顺序中排在第 j 位$\}$
- 全概率公式: $P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j)$
  - $P(B_j) = \frac{(n-1)!}{n!} = \frac{1}{n}, j = 1, \cdots, n$
  - 当 $1 \le j \le k$ 时, $P(A|B_j) = 0$ (最佳人选位于前 k 个面试者, 不会被录用)
  - 当 $k+1 \le j \le n$ 时, $P(A|B_j) = \dfrac{k}{j-1}$ (最佳人选被录用当且仅当前 $j-1$ 个面试者中的最佳者在前 k 个人中)
- $P(A) = \dfrac{1}{n}\sum_{j=k+1}^{n}\dfrac{k}{j-1} = \dfrac{k}{n}\sum_{j=k}^{n-1}\dfrac{1}{j} \sim \dfrac{k}{n}\ln\dfrac{n}{k}$
- 令 $g(x) = \dfrac{x}{n}\ln\dfrac{n}{x}, x > 0$, 则 $g'(x) = \dfrac{1}{n}\ln\dfrac{n}{x} - \dfrac{1}{n} = 0 \Longleftrightarrow x = \dfrac{n}{e}$
- 若要 $P(A)$ 达到最大, 只需 k 取最靠近 $\frac{n}{e}$ 的正整数
- 最大概率值为 $g\left(\frac{n}{e}\right) = \frac{1}{e} \approx 0.36788$ 与 n 无关

第八讲：事件的独立性

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

两个事件的独立性

- 直观上来说，两个事件的独立性是指：一个事件的发生不影响另一个事件的发生。比如在掷两颗骰子的实验中，第一颗骰子的点数和第二颗骰子的点数是互不影响的.
- 从概率的角度看，$P(A|B)$ 与 $P(A)$ 的差别在于：事件 B 的发生改变了事件 A 发生的概率，也即事件 B 对事件 A 有某种影响。故如果 A 与 B 的发生是相互不影响的，则有

$$P(A|B) = P(A), \quad P(B|A) = P(B).$$

上面两式均等价于

$$P(AB) = P(A)P(B) \tag{16}$$

- 注意到 (16) 式对 $P(B) = 0$ 或 $P(A) = 0$ 仍然成立，为此，我们用 (16) 作为两个事件相互独立的定义.

两个事件独立性的定义

定义 2.26 如果对事件 A 与 B 有

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

成立，则称 事件 A 与 B 相互独立, 简称 A 与 B 独立. 否则称 A 与 B 不独立或相依.

注 2.10

- 零概率事件 E 与任何事件相互独立，特别的，不可能事件与任何事件相互独立
- 若 A, B 互不相容且独立, 则 A, B 至少有一个零概率事件
- 非零概率不相容事件, 一定不独立; 非零概率独立事件，一定相容
- 若事件 A 与其自身相互独立, 则 $P(A) = 0$, 或 $P(A) = 1$

如何确定事件的独立性:

- 实际问题中，两个事件的独立大多根据经验及相互有无影响的直观性来判断.
- 但对于较复杂事件，有无相互影响并不是很直观，则需要验证 (16) 式是否成立来说明独立性.

对立事件的独立性

定理 2.18 若 A 与 B 独立，则 A 与 $\overline{B}$ 独立，$\overline{A}$ 与 B 独立，$\overline{A}$ 与 $\overline{B}$ 独立.

证明: 我们仅证 $P(A\overline{B}) = P(A)P(\overline{B})$, 其余类似可证.

$$\begin{aligned} P(A\overline{B}) &= P(A-B) = P(A) - P(AB) = P(A) - P(A)P(B) \\ &= P(A)(1-P(B)) = P(A)P(\overline{B}) \end{aligned}$$

对于上面的定理直观上来理解也是很容易的：因 A, B 独立，故 A 的发生不影响 B 的发生，从而也不会影响 B 的不发生，$\cdots$

概率测度与独立性的关系

例 2.72 考虑掷硬币问题, 记正面向上对应的样本点为"H", 反面向上为"T", 那么连续掷三次的结果构成样本空间

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

令事件 A 为" 最后一次是反面", B 为" 三次结果相同", 则有

$$A = \{HHT, HTT, THT, TTT\}, B = \{HHH, TTT\}$$

试讨论 A, B 的独立性.

解: ：设每次抛掷反面向上的概率是 p, 那么

$$P(A) = p^3 + 2p^2(1-p) + p(1-p)^2, P(B) = p^3 + (1-p)^3, P(AB) = p^3,$$

不难验证, $p = 0$、$p = 1$ 和 $p = \frac{1}{2}$ 时,A 和 B 是独立的, 否则两者不独立.

三个事件的独立性

定义 2.27 设 A, B, C 三个事件，如果有

$$\left.\begin{aligned} P(AB) &= P(A)P(B) \\ P(AC) &= P(A)P(C) \\ P(BC) &= P(B)P(C) \end{aligned}\right\} \tag{17}$$

$$P(ABC) = P(A)P(B)P(C) \tag{18}$$

则称 A, B, C 相互独立。如果仅有 (17) 式成立，则称 A, B, C 两两独立.

两两独立与相互独立的关系

- 由定义可知，三个事件相互独立必能推出两两独立.
- 但两两独立未必能推出相互独立，即 (17) 式成立，不一定能推出 (18) 成立
  - 考虑独立投掷两枚均匀硬币的随机试验，设事件 A 代表第一枚硬币正面朝上，事件 B 代表第二枚硬币正面朝上，事件 C 表示两枚硬币结果相同。易知: A B 和 C 是两两独立，但

$$P(A \cap B \cap C) = 1/4 \neq 1/8 = P(A)P(B)P(C).$$

  - 考虑一个均匀的正四面体，第一二三面分别染上红 / 白 / 黑色，第四面同时染上红白黑色。现在以 A, B, C 分别记投一次四面体出现红，白，黑色朝下的事件。则易有

$$\begin{aligned} P(A) = P(B) = P(C) &= 1/2 \\ P(AB) = P(BC) = P(AC) &= 1/4 \\ P(ABC) &= 1/4 \end{aligned}$$

两两独立与相互独立的关系

- 反之，如果 (18) 成立，是否能推出 (17) 成立？
  - 考虑一个均匀的正八面体，第 1, 2, 3, 4 面染上红色，第 1, 2, 3, 5 面染上白色，第 1, 6, 7, 8 面染上黑色。现在以 A, B, C 分别记投一次八面体出现红，白，黑色朝下的事件，则

$$
\begin{aligned}
P(A) = P(B) = P(C) = 4/8 = 1/2 \\
P(ABC) = 1/8 \\
P(AB) = 3/8 \neq 1/4 = P(A)P(B)
\end{aligned}
$$

三个以上事件的独立性

定义 2.28 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，$A_1, A_2, \cdots, A_n \in \mathcal{F}$, 对任意的 $1 \le k \le n$ 及任意的 $1 < j_1 < j_2 < \cdots < j_k \le n$ 均有:

$$P(A_{j_1}A_{j_2}\cdots A_{j_k}) = P(A_{j_1})P(A_{j_2})\cdots P(A_{j_k}) \tag{19}$$

成立，则称事件 $A_1, \cdots, A_n$ 相互独立.

- (19) 式共有多少个等式?

$$\left.\begin{array}{l} P(A_{j_1}A_{j_2}) = P(A_{j_1})P(A_{j_2}) \\ P(A_{j_1}A_{j_2}A_{j_3}) = P(A_{j_1})P(A_{j_2})P(A_{j_3}) \\ \quad\vdots \\ P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n) \end{array}\right\} C_n^2 + \cdots + C_n^n = 2^n - n - 1$$

- 从定义可以看出，n 个相互独立事件中的任取 $m(2 \le m \le n)$ 个事件仍是相互独立的，而且任意一部分与另一部分也是独立的.
- 类似于前面的证明，将相互独立事件中的任一部分换为对立事件，所得诸事件仍是相互独立的.

任意多个事件相互独立

定义 2.29 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，每个 $t \in T$ 有 $A_t \in \mathcal{F}$. 称 $\{A_t, t \in T\}$ 为独立事件族，如果对 T 的任意有限子集 $\{t_1, t_2, \cdots, t_s\}$, 事件 $A_{t_1}, A_{t_2}, \cdots, A_{t_s}$ 相互独立.

例 2.73 $\mathcal{F}$ 中事件序列 $\{A_n\}$ 为相互独立的充分必要条件是，任意 $n \geq 1$, 事件 $A_1, A_2, \cdots, A_n$ 独立；等价的，任意有限个自然数 $k_1, \cdots, k_s$ 有

$$P(A_{k_1} A_{k_2} \cdots A_{k_s}) = P(A_{k_1}) P(A_{k_2}) \cdots P(A_{k_s})$$

条件独立

定义 2.30 称事件 A 和 B 是关于 E 条件独立的，如果

$$P(A \cap B|E) = P(A|E)P(B|E)$$

- 两个事件可以在给定事件 E 的条件下是条件独立的，但它们不是独立的.
- 两个事件可以是独立但却不是关于 E 条件独立的.
- 两个事件可以关于 E 条件独立但关于 $\overline{E}$ 不存在条件独立.

条件独立不意味着独立

例 2.74 假设有两枚硬币，一枚是均匀的，一枚是不均匀的。从两枚硬币中随机的选一枚硬币并进行抛掷 2 次，若令

$$
\begin{aligned}
F &:= \{\text{选取的硬币是均匀的}\} \\
A_1 &:= \{\text{第一次投掷硬币正面朝上}\} \\
A_2 &:= \{\text{第二次投掷硬币正面朝上}\}
\end{aligned}
$$

则给定 F 为条件，A_1 和 A_2, 是相互独立的，A_1 和 A_2 并不是无条件独立的，因为 A_1 会提供关于 A_2 的信息.

独立不意味着条件独立

例 2.75 假设只有我的朋友 Alice 和 Bob 给我打过电话。每天他俩都会相互独立地决定是否给我打电话。若令

$$A := \{\text{Alice 给我打电话}\}$$
$$B := \{\text{Bob 给我打电话}\}$$
$$R := \{\text{听到电话铃响}\}$$

- 显然，A 和 B 是无条件独立的.
- 但现在我听到一声电话铃响，那 A 和 B 就不再独立了：如果这个电话不是 Alice 打的，那就肯定是 Bob 打的。从而

$$P(B|R) < 1 = P(B|\overline{A}R) = \frac{P(B\overline{A}R)}{P(\overline{A}R)} = \frac{P(B\overline{A}|R)}{P(\overline{A}|R)}.$$

显然: $P(B\overline{A}|R) > P(B|R)P(\overline{A}|R)$

- B 与 $\overline{A}$ 关于 R 不条件独立，A, B 亦是如此.

给定 E 条件独立 *vs* 给定 $\overline{E}$ 条件独立

例 2.76 假设有两种课程：好的课程和坏的课程。在好的课上，如果你努力，就很有可能得到 A. 在坏的课上，教授随机分配给学生分数，而不管他们是否努力。若令

$$G := \{\text{这个课程是好的}\}$$
$$W := \{\text{你学习努力}\}$$
$$A := \{\text{你的得分为}A\}$$

这时，给定 $\overline{G}$, A 和 W 是条件独立的，但给定 G, A 和 W 却不是独立的！

集类 (族) 的独立性

定义 2.31 (集类 (族) 的独立性) 考虑样本空间 $\Omega, \mathcal{A}_k \subset \Omega$, $k = 1, \cdots, n$. 称集类 (族)$\{\mathcal{A}_k\}_{k=1}^n$ 是相互独立的, 如果 $\{\mathcal{A}_k\}_{k=1}^n$ 满足

$$P\left(\bigcap_{k\in I} A_k\right) = \prod_{k\in I} P(A_k), \forall A_k \in \mathcal{A}_k, \forall I \subset \{1, 2, \cdots, n\}.$$

定理 2.19 (σ-代数的独立性) 设 $\mathcal{F}_1, \cdots, \mathcal{F}_n$ 为 Ω 上的 σ-代数, 若

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n), \quad \forall A_k \in \mathcal{F}_k, k = 1, 2, \cdots, n$$

则 $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$ 是独立的.

例 2.77 (硬币实验的独立性) 抛掷不均匀硬币的实验，正面 (用 1 表示) 向上的概率是 p, 反面 (用 0 表示) 向上的概率是 q. 假设连抛 n 次, 则样本空间 Ω 为 $\Omega = \{a_1 a_2 \cdots a_n : a_k = 0, 1\}$. 考虑事件 $A_k = \{a_k = 1\}$, $k = 1, \cdots, n$, 构造 σ-代数 $\mathcal{F}_k$: $\mathcal{F}_k = \left\{\Omega, \varnothing, A_k, A_k^C\right\}$.可以验证, 这些 σ-代数是独立的.

独立的集类 (族) 生成的 σ-代数未必独立

- 考虑 $\Omega = \{1,2,3,4\}$, 集类 (族)$\mathcal{A} = \{\{1,2\},\{2,3\}\}$,$\mathcal{B} = \{\{2,4\}\}$
- 令 $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \dfrac{1}{4}$, 则

$$P(\{1,2\} \cap \{2,4\}) = P(\{2\}) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(\{1,2\})P(\{2,4\}),$$

$$P(\{2,3\} \cap \{2,4\}) = P(\{2\}) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(\{2,3\})P(\{2,4\})$$

- 集类 (族)$\mathcal{A}$ 和 $\mathcal{B}$ 独立
- 但由于

$$\sigma(\mathcal{A}) = 2^{\Omega}, \sigma(\mathcal{B}) = \{\{2,4\},\{1,3\},\varnothing,\Omega\}$$

且明显有

$$P(\{2,4\} \cap \{3\}) = 0 \neq \frac{1}{8} = \frac{1}{2} \times \frac{1}{4} = P(\{2,4\})P(\{3\})$$

- 故, $\sigma(\mathcal{A})$ 和 $\sigma(\mathcal{B})$ 不独立.

独立的 π-类 (族) 生成 σ-代数独立

定理 2.20 如果 $\mathcal{A}$ 和 $\mathcal{B}$ 是独立的集类 (族), $\mathcal{B}$ 是 π-类, 那么 $\mathcal{A}$ 和 $\sigma(\mathcal{B})$ 也独立.

证明: 应用单调类定理.

- 任意固定 $A \in \mathcal{A}$, 令

$$\mathscr{A} = \{B \in \sigma(\mathcal{B}) : P(AB) = P(A)P(B)\}$$

  则 $\mathscr{A}$ 是 λ-类 (系统)
- 由 $\mathcal{B} \subset \mathscr{A}$ 及单调类定理可得: $\sigma(\mathcal{B}) \subset \mathscr{A}$.
- $\mathcal{A}$ 和 $\sigma(\mathcal{B})$ 独立.

Borel-Cantelli 引理

定理 2.21 对于事件列 $\{A_j\}$，有

- 如果 $\sum_{j=1}^{\infty} P(A_j) < \infty$，则 $P(A_n\ i.o.) = 0$
- 如果 $\{A_j\}$ 相互独立，$\sum_{j=1}^{\infty} P(A_j) = \infty$，则 $P(A_n\ i.o.) = 1$

证明: 注意到

$$P(A_n\ i.o.) = P(\lim_{n\to\infty} \cup_{j=n}^{\infty} A_j) = \lim_{n\to\infty} P(\cup_{j=n}^{\infty} A_j) \le \lim_{n\to\infty} \sum_{j=n}^{\infty} P(A_j) = 0$$

由于

$$\begin{aligned}
P(\cup_{j=n}^{\infty} A_j) &= \lim_{m\to\infty} P(\cup_{j=n}^{m} A_j) = \lim_{m\to\infty} (1 - P(\cap_{j=n}^{m} \overline{A}_j)) \\
P(\cap_{j=n}^{m} \overline{A}_j) &= \Pi_{j=n}^{m} P(\overline{A}_j) = \Pi_{j=n}^{m} (1 - P(A_j)) \\
&\le \Pi_{j=n}^{m} \exp(-P(A_j)) = \exp(-\sum_{j=n}^{m} P(A_j)) \overset{m\to\infty}{\longrightarrow} 0
\end{aligned}$$

随机试验的独立性

- 先考虑两个随机试验，假定 $(\Omega_i, \mathcal{F}_i, P_i), i = 1, 2$ 为第 i 个随机试验对应的概率空间。按照之前独立性的理解，两个试验的独立性应当叙述为:

$$\text{对任何的}A_i \in \mathcal{F}_i, i = 1, 2, A_1\text{与}A_2\text{同时}$$
$$\text{发生的概率等于它们各自概率之乘积}$$

- 两个不妥:
  - "A_1 与 A_2 同时发生" 应当是这两个事件的交，但它们分别是两个样本空间 Ω_1, Ω_2 的子集，无法进行运算;
  - 两个概率空间有各自的概率 P_1, P_2, 但涉及两个度验，命题中 "同时发生的概率" 既不能用 P_1 也不能用 P_2 来度量.
- 解决方法：构造可以同时描述两个试验的新概率空间 $(\Omega, \mathcal{F}, P)$.

乘积空间的构造

- 样本乘积空间: $\Omega := \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1 且 \omega_2 \in \Omega_2\}$;
- 乘积 σ- 代数 $\mathcal{F}_1 \times \mathcal{F}_2$:
  - 可测矩形集类: $\mathcal{G} := \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$;
  - $\mathcal{F}_1 \times \mathcal{F}_2 := \sigma(\mathcal{G})$
- 乘积概率测度:
  - 对于每个可测矩形 $A_1 \times A_2 \in \mathcal{G}$ 定义如下集函数:

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \quad A_i \in \mathcal{F}_i, i = 1, 2. \tag{21}$$

  - 理论上可以证明如上定义在 $\mathcal{G}$ 上的集函数 P 可唯一地扩张为 $\mathcal{F}_1 \times \mathcal{F}_2$ 上的概率测度，称之为 P_1 与 P_2 的乘积 (概率) 测度.
- 在上述乘积测度下

$$\begin{aligned} &P(A_1 \times \Omega_2) = P_1(A_1), \quad P(\Omega_1 \times A_2) = P_2(A_2) \\ &P\big((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)\big) = P(A_1 \times A_2) \\ &= P_1(A_1)P_2(A_2) = P(A_1 \times \Omega_1)P(\Omega_1 \times A_2) \end{aligned}$$

- $(\Omega_i, \mathcal{F}_i, P_i)$ 的独立性取决于乘积样本空间 $\Omega_1 \times \Omega_2$ 上的概率是否取作由 (21) 所确定的乘积测度

n 个试验相互独立的定义

定义 2.32 设有 n 个随机试验，第 i 个试验的概率空间为 $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, \cdots, n$. 代表这 n 个试验的乘积样本空间 $\Omega = \Omega_1 \times \cdots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_n = \sigma(\mathcal{G})$, 其中 $\mathcal{G}$ 为形如 $B_1 \times \cdots \times B_n (B_i \in \mathcal{F}_i)$ 的可测矩形的全体。如果 $(\Omega, \mathcal{F})$ 上的概率测度 P 是 $P_1, \cdots, P_n$ 的乘积测度，即对任何 $B_1 \times \cdots \times B_n \in \mathcal{G}$ 满足

$$P(B_1 \times \cdots \times B_n) = P_1(B_1) \cdots P(B_n),$$

则称这 n 个度验独立. 如果现设

$$\Omega_i \equiv \Omega_0, \mathcal{F}_i \equiv \mathcal{F}_0, P_i \equiv P_0, i = 1, \cdots, n,$$

即 n 个试验有相同的概率空间，则称它们为 n 个 (重) 独立重复试验. 如果在 n 个独立重复实验中，每次试验的可能结果为两个:A 或 $\overline{A}$, 则称这种试验为 n 重伯努利试验.

彩票问题

例 2.78 某彩票每周开奖一次，每次提供十万分之一的中奖机会，且各周开奖是独立的。若你每周买一张彩票，坚持十年（每年按 52 周计算），试求未中奖的概率.

解: 依假设，每次中奖的概率为 10^{-5}, 于是每次不中奖的概率是 $1-10^{-5}$. 另外十年一共购买 520 次彩票，而每次开奖都是独立的，相当于进行了 520 次独立重复试验. 若记 A_i 为“第 i 次开奖不中奖”，则 $A_1, \cdots, A_{520}$ 相互独立，从而

$$P(A_1A_2\cdots A_{520}) = (1-10^{-5})^{520} = 0.9948$$

第九讲：随机变量定义及相关性质

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

随机变量的引入

- 赌徒输光：甲和乙初始资金分别为 $i, a-i$ 元，每一局甲赢的概率为 p
- 关注的问题
  - 甲最终获胜的概率
  - 甲乙两人在任意时刻的剩余资产:k 轮赌博后恰好剩下 j 元
  - k 轮赌博后甲乙两人资产的差额 Z
  - 赌博持续时间 R
- 表示方法:
  - $E :=\{$甲最终获胜 $\}$, $Q_i := P(\mathrm{E})$
  - $A_{jk} :=\{$甲在 k 轮赌博后恰好剩下 j 元 $\}$
  - $B_{jk} :=\{$乙在 k 轮赌博后恰好剩下 j 元 $\}$
  - k 轮赌博后甲乙两人资产的差额如何表示?
  - 赌博持续时间 R 如何表示?
  - 很难用事件来表示或者表示很复杂

随机变量的引入

- $X_k :=$ 甲在 k 轮赌博后的资产
  - 乙在 k 轮赌博后的资产 $Y_k = a - X_k$
  - 资产差额:$Z = X_k - Y_k = 2X_k - a$
  - 持续时间:$R = \min\{n : X_n = 0, \text{或} Y_n = 0\}$
- X_k 取值的特点
  - 依赖于前面 k 次赌博这一“随机试验”的结果
  - 在“随机试验”完成之前，X_k 取值不确定，因此具有不确定性
  - k 次赌博“随机试验”一旦完成，X_k 的值必然确定
  - 记 k 次赌博“随机试验”样本空间为 Ω, 则给定 $\omega \in \Omega$, 则 X_k 值确定
- 以 $k = 2$ 为例，看一下 X_2 的取值情况，设 $i \geq 2$
  - $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, 其中
    $\omega_1 = (\text{胜，胜}), \omega_2 = (\text{胜，败}), \omega_3 = (\text{败，胜}), \omega_4 = (\text{败，败})$
  - $X_2(\omega_1) = i + 2,\ X_2(\omega_2) = X_2(\omega_3) = i,\ X_2(\omega_4) = i - 2$
- X_k 可以看作定义在样本空间 Ω 上的函数，即
  $X_k : \Omega \to \{0, 1, \cdots, a\}$
- 一般的，随机变量可以看作从样本空间到实数的映射:$X : \Omega \to R$

随机变量的直观定义

定义 2.1 (直观定义) 称 X 为随机变量，如果 X 是从样本空间 Ω 到实数的映射，即 $X:\Omega \to R$.

一个随机变量的例子

例 2.1 考虑硬币抛掷两次的随机试验，其样本空间为

$$\Omega = \{HH, HT, TH, TT\}.$$

- 令 X 表示正面朝上的次数，则 X 是一个随机变量, 相应的映射如下

$$X(TT) = 0, X(HT) = X(TH) = 1, X(HH) = 2.$$

- 令 Y 表示反面朝上的次数，则 $Y = 2 - X$, 对于任意的 $\omega \in \Omega$, 有 $Y(\omega) = 2 - X(\omega)$.
- 设 I 是由第一次掷硬币的结果决定的随机变量：若第一次硬币正面朝上则 $I = 1$, 反之 $I = 0$, 即

$$I(HH) = I(HT) = 1, \quad I(TH) = I(TT) = 0.$$

- 若正面记 1, 反面记 0, 此时 $\Omega = \{(1,1), (1,0), (0,1), (0,0)\}$. 上述的 X, Y 和 I 可表示为:

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2, Y(\omega_1, \omega_2) = 2 - \omega_1 - \omega_2, I(\omega_1, \omega_2) = \omega_1.$$

与函数对比分析

- 数学分析中的函数 $f(x)$
  - $f(x): D \to R$, 其中 $D \subset R$
  - R 上可定义距离 $d(x,y) = |x-y|$
  - 可根据距离 d 定义函数的连续性
- 概率论中的随机变量 X
  - $X(\omega): \Omega \to R$
  - 定义域 Ω 没有距离的定义，但有事件域 $\mathcal{F}$ 及定义其上的概率 P, 即具有结构 $(\Omega, \mathcal{F}, P)$
  - 值域 R 有距离，但因定义域无距离，故无法考虑随机变量的连续性
  - 值域 R 还有 σ- 代数 $\mathcal{B} := \mathcal{B}(R)$, 有可测空间结构 $(R, \mathcal{B})$
- 是否需要对 $X(\omega): \Omega \to R$ 做一些额外的限定，以便更好的研究 X?
- 从赌徒输光问题可以看出，对随机变量 X, 我们会关注
  - $X(\omega) = x$ 的概率，$X(\omega) \le x, X(\omega) \in [b,c]$ 的概率
  - 更一般的，$\forall B \in \mathcal{B}(R), X(\omega) \in B$ 的概率
- 概率的定义域是 $\mathcal{F}$, 要想计算 $X(\omega) \in B$ 的概率，当且仅当

$$\forall B \in \mathcal{B}(R),\ X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

随机变量的严格数学定义

定义 2.2 (可测映射) 设 $(\Omega, \mathcal{F})$ 和 $(E, \mathcal{E})$ 为两个可测空间并令 X 为从样本空间 Ω 到 E 的映射，即 $X(\omega): \Omega \to E$. 若对任意的 $B \in \mathcal{E}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

则称 X 为 $(\Omega, \mathcal{F})$ 到 $(E, \mathcal{E})$ 的可测映射.

若将上述定义中的可测空间 $(E, \mathcal{E})$ 更换为 $(R, \mathcal{B})$，则

定义 2.3 (可测函数或随机变量) 设 $(\Omega, \mathcal{F})$ 是可测空间，X 为从样本空间 Ω 到实数集 R 的映射，即 $X(\omega): \Omega \to R$. 如果对 $\forall B \in \mathcal{B}$ 均有

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

则称 X 为可测空间 $(\Omega, \mathcal{F})$ 上的可测函数或随机变量.

定义 2.4 (随机变量的另一定义) 设 $(\Omega, \mathcal{F})$ 是可测空间，X 为从样本空间 Ω 到实数集 R 的映射，即 $X(\omega): \Omega \to R$. 如果对任意的 $x \in R$ 均有

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F} \text{ 或} \{\omega : X(\omega) < x\} \in \mathcal{F}$$

则称 $X(\omega)$ 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，简称随机变量.

随机变量两种定义的等价性

由定义 2.4 推定义 2.3: 仅需说明若定义 2.4 成立，则对任意 $B \in \mathcal{B}$ 均有

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

即只需说明以下集合包含关系成立即可

$$\mathcal{A} := \{A : X^{-1}(A) \in \mathcal{F}\} \supset \mathcal{B}$$

欲证上面包含关系成立，我们只需说明以下两点即可:

1. $\mathcal{A}$ 是 σ 代数;
2. $O_1 := \{(-\infty, x] : x \in R\} \subset \mathcal{A}$.

再由 $\mathcal{B} := \sigma(O_1)$ 知 $\mathcal{B} \subset \mathcal{A}$.

$\mathcal{A} := \{A : X^{-1}(A) \in \mathcal{F}\}$ 为 σ 代数

- $X^{-1}(R) = \{\omega : X(\omega) \in R\} = \Omega \in \mathcal{F}$, 故 $R \in \mathcal{A}$;
- 若 $A \in \mathcal{A}$, 即 $X^{-1}(A) \in \mathcal{F}$, 则

$$\begin{aligned} X^{-1}(\overline{A}) &= \{\omega : X(\omega) \in \overline{A}\} = \{\omega : X(\omega) \notin A\} \\ &= \overline{\{\omega : X(\omega) \in A\}} = \overline{X^{-1}(A)} \\ &\in \mathcal{F} \end{aligned}$$

- 对于 $A_j \in \mathcal{A}, j = 1, 2, \cdots$, 有 $X^{-1}(A_j) \in \mathcal{F}, j = 1, 2, \cdots$. 从而

$$\begin{aligned} X^{-1}(\cup_{j=1}^{\infty} A_j) &= \{\omega : X(\omega) \in \cup_{j=1}^{\infty} A_j\} = \cup_{j=1}^{\infty} \{\omega : X(\omega) \in A_j\} \\ &= \cup_{j=1}^{\infty} X^{-1}(A_j) \\ &\in \mathcal{F} \end{aligned}$$

随机变量的分类以及两个注记

- 随机变量 X 是从样本空间 Ω 到实数 R 的映射，故根据其值域集合可粗略的分为两大类
  - 离散型随机变量：其值域集合是有限点集或可数点集即

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\} = \{a_n\}_{n\geq 1}$$

  - 非离散型随机变量：其值域集合不是有限点集或可数点集

注 2.1 随机变量的两点注记:

- 首先, 随机变量是确定性函数, 自身并没有随机性. 给定样本空间上的样本点, 有唯一确定的实数值与之相对应, 这种对应关系并没有不确定性. 所有的不确定性都体现在样本点是否在试验结果中出现, 和随机变量本身没有关系。随机变量的引入, 更多地是为了数学处理上的方便.
- 其次, 随机变量并不是概率论中独有的概念. 若将前面可测映射中的 $(\Omega, \mathcal{F}), (E, \mathcal{E})$ 均取为 $(R, \mathcal{B}(\mathbb{R}))$, 则可测映射的定义便退化为实分析中的 "可测函数". 随机变量是一特殊的可测函数.

示性随机变量

例 2.2 设 Ω 是某随机试验的样本空间，$\mathcal{F}$ 为其事件域 (σ 代数)，则对于任意的 $A \in \mathcal{F}$，示性函数 $I_A(\omega) := \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases}$ 是随机变量.

解: 由示性函数的定义知:

$$\{\omega : I_A(\omega) \le x\} = \begin{cases} \emptyset, & x < 0, \\ \overline{A}, & x \in [0, 1), \\ \Omega, & x \ge 1. \end{cases}$$

显然，无论 x 取何值，均有 $\{\omega : I_A(\omega) \le x\} \in \mathcal{F}$

随机变量的性质

定理 2.1 若 $X, Y, \{X_n, n \geq 1\}$ 都为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，则

1. $|X|$，$aX + bY, (a, b \in R)$ 均为随机变量;
2. $X^+ := X \vee 0, X^- := (-X) \vee 0$ 均为随机变量;
3. XY 为随机变量;
4. 若 X/Y 处处有意义，则 X/Y 为随机变量;
5. $\inf_n X_n, \sup_n X_n, \liminf_{n\to\infty} X_n, \limsup_{n\to\infty} X_n$ 均为随机变量.

证明:

1.
   - $\{\omega : |X| < x\} = \{\omega : -x < X < x\} = \{\omega : X < x\} \cap \overline{\{\omega : X \leq -x\}} \in \mathcal{F}$;
   - $\{\omega : aX < x\} = \{\omega : X < \dfrac{x}{a}\} \in \mathcal{F}$, (当 $a > 0$ 时);
   - 设 Q 为有理数集，则

$$
\begin{aligned}
\{\omega : X + Y < x\} &= \{\omega : X < x - Y\} = \cup_{r\in Q} \{\omega : X < r < x - Y\} \\
&= \cup_{r\in Q} \{\omega : X < r, Y < x - r\} \\
&= \cup_{r\in Q} (\{\omega : X < r\} \cap \{\omega : Y < x - r\}) \\
&\in \mathcal{F}
\end{aligned}
$$

2. 注意到 $X^+ = \dfrac{|X|+X}{2}, X^- = \dfrac{|X|-X}{2}$，易得 X^+, X^- 均为随机变量;

3. 首先假定 X, Y 非负，则对任意的 $x > 0$ 有

$$\begin{aligned}\{XY < x\} &= \{X = 0\} \cup \{Y = 0\} \cup \left(\cup_{r \in Q_+} \left[\{X < r\} \cap \{Y < \frac{x}{r}\}\right]\right) \\ &\in \mathcal{F}.\end{aligned}$$

对一般的 X, Y, 由 X^+, X^-, Y^+, Y^- 为随机变量，可得

$$XY = (X^+ - X^-)(Y^+ - Y^-) = (X^+Y^+ + X^-Y^-) - (X^+Y^- + X^-Y^+)$$

为随机变量.

4. 设 $|Y| > 0$ 处处成立，易证 $\dfrac{1}{Y}$ 是随机变量，故 $\dfrac{X}{Y} = X \cdot \dfrac{1}{Y}$ 为随机变量.

5. 对任意的 $x \in R$, 我们有

$$\{\inf_n X_n < x\} = \cup_n \{X_n < x\}, \quad \{\sup_n X_n \le x\} = \cap_n \{X_n \le x\}$$

简单随机变量

例 2.3 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，$A_i \in \mathcal{F}, i = 1, \cdots, n$ 为 Ω 的一个分割，$a_i, i = 1, \cdots, n$ 为 n 个不同的实数，则

$$X(\omega) := \sum_{i=1}^{n} a_i I_{A_i}(\omega) \tag{22}$$

作为 n 个示性随机变量的线性组合，仍为随机变量。我们称形如 (22) 的 $X(\omega)$ 为简单随机变量.

随机变量的函数是否仍为随机变量？

定理 2.2 设 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，$g(x)$ 为 $(R, \mathcal{B}) \to (R, \mathcal{B})$ 上的 Borel 可测函数，证 $Y := g(X)$ 为 $(\Omega, \mathcal{F})$ 上的随机变量.

证明: 注意到，对任意的 $B \in \mathcal{B}$, $g^{-1}(B) \in \mathcal{B}$, 故

$$\begin{aligned} Y^{-1}(B) &= \{\omega : Y(\omega) \in B\} \\ &= \{\omega : g(X(\omega)) \in B\} \\ &= \{\omega : X(\omega) \in g^{-1}(B)\} \\ &= X^{-1}(g^{-1}(B)) \\ &\in \mathcal{F} \end{aligned}$$

随机变量的结构

对于 $(\Omega, \mathcal{F}, P)$ 上的任意非负随机变量 X 及自然数 n，我们可将 Ω 按 X 的取值进行分割。即令

$$\begin{aligned} A_k(\omega) &:= \{\omega : \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\}, k = 0, 1, \cdots, n2^n - 1 \\ A_{n2^n}(\omega) &:= \{\omega : X(\omega) \ge n\} \end{aligned}$$

则

$$X_n(\omega) := \sum_{k=0}^{n2^n} \frac{k}{2^n} I_{A_k}(\omega)$$

为简单随机变量且随机变量序列 $\{X_n, n \ge 1\}$ 满足

$$0 \le X_1(\omega) \le X_2(\omega) \le \cdots \le X_n(\omega) \to X(\omega)$$

$X_n(\omega)$ 的单调性

注意到对任意的 $k = 0, 1, \cdots, n2^n - 1$,

$$\begin{aligned} A_k(\omega) &= \{\omega : \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\} = \{\omega : \frac{2k}{2^{n+1}} \le X(\omega) < \frac{2(k+1)}{2^{n+1}}\} \\ &= \{\omega : \frac{2k}{2^{n+1}} \le X(\omega) < \frac{2k+1}{2^{n+1}}\} \cup \{\omega : \frac{2k+1}{2^{n+1}} \le X(\omega) < \frac{2k+2)}{2^{n+1}}\} \\ &= A_k^1(\omega) \cup A_k^2(\omega) \end{aligned}$$

故在集合 $A_k(\omega), k = 0, 1, \cdots, n2^n - 1$ 上,

$$X_{n+1}(\omega) = \begin{cases} \frac{2k}{2^{n+1}}, & \omega \in A_k^1(\omega) \\ \frac{2k+1}{2^{n+1}}, & \omega \in A_k^2(\omega) \end{cases} \quad (\ge \frac{k}{2^n} = X_n(\omega))$$

而在 $A_{n2^n}(\omega) := \{\omega : X(\omega) \ge n\} = \{\omega : X(\omega) \ge \frac{n2^{n+1}}{2^{n+1}}\}$ 上，显然有

$$X_{n+1}(\omega) \ge n = X_n(\omega).$$

$X_n(\omega)$ 的收敛性

注意到，对任意的 ω, 必定存在 k 使得

$$\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n},$$

从而

$$0 \leq X(\omega) - X_n(\omega) \leq \frac{1}{2^n}$$

显然有

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

随机变量的简单随机逼近

定理 2.3 对 $(\Omega, \mathcal{F}, P)$ 上的实值变量 $X(\omega)$ 为随机变量的充要条件是: 存在简单随机变量序列 $\{X_n(\omega), n \geq 1\}$ 使得

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega$$

而且当 X 非负时，还可选取 $\{X_n(\omega), n \geq 1\}$ 为非负单调不减的简单随机变量序列.

随机变量的生成 σ-代数

定义 2.5 (单个随机变量的生成 σ-代数) 设 X 为定义在 $(\Omega, \mathcal{F})$ 上的随机变量并令 $\mathcal{F}_X := \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$. 我们称 $\sigma(X) := \sigma(\mathcal{F}_X)$ 为随机变量 X 的生成 σ-代数.

定义 2.6 (有限个随机变量的生成 σ-代数) 设 $X_k, k = 1, 2, \cdots, n$ 为定义在 $(\Omega, \mathcal{F})$ 上的随机变量并令 $\mathcal{F}_{X_k} := \{X_k^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$. 我们称

$$\sigma(X_1, X_2, \cdots, X_n) := \sigma\left(\cup_{k=1}^n \mathcal{F}_{X_k}\right)$$

为随机变量 $X_1, X_2, \cdots, X_n$ 的生成 σ-代数.

- $\mathcal{F}_X, \mathcal{F}_{X_k}$ 均为 σ-代数, 但 $\cup_{k=1}^n \mathcal{F}_{X_k}$ 不一定是 σ-代数, 因此

$$\sigma(X) = \mathcal{F}_X, \quad \text{但} \quad \sigma(X_1, X_2, \cdots, X_n) \neq \cup_{k=1}^n \mathcal{F}_{X_k}.$$

- 若 X 是 $(\Omega, \mathcal{F})$ 上的随机变量, 则必有 $\sigma(X) := \mathcal{F}_X \subset \mathcal{F}$, 即 $\sigma(X)$ 是 Ω 上使得 X 成为随变量所需要的最小 σ-代数.
- 类似的, $\sigma(X_1, X_2, \cdots, X_n)$ 是 Ω 上使得 $X_1, X_2, \cdots, X_n$ 为随机变量所需要的最小 σ-代数.
- 随机变量 X 生成的 $\sigma-$ 代数 $\sigma(X)$ 集中体现了 X 的取值信息.

两个随机变量之间的关系

定义 2.7 考虑两个随机变量 X, Y，如果对任意的 $B \in \mathcal{B}(\mathbb{R})$ 均有

$$Y^{-1}(B) \in \sigma(X), \text{ 等价的 } \sigma(Y) \subset \sigma(X).$$

则称 Y 适应 (adaptive to) X.

定理 2.4 考虑可测空间 $(\Omega, \mathcal{F})$, X 和 Y 为定义其上的随机变量. 若 Y 适应 X, 则存在可测函数 $g : \mathbb{R} \to \mathbb{R}$, 使得

$$Y(\omega) = g(X(\omega)), \ \forall \omega \in \Omega.$$

注 2.2

- 随机变量的生成 σ-代数研究随机变量间关系起着重要作用.
- 直观地看,Y 适应 X, 意味着 Y 包含的信息被 X 包含的信息所涵盖. 换句话说,Y 和 X 间存在导出关系.

随机向量: 如何定义 $\mathcal{B}(\mathbb{R}^n)$

- 直观上来讲, 随机向量就是取值于 $\mathbb{R}^n$ 的随机变量.
- 如何定义 $\mathcal{B}(\mathbb{R}^n)$
  - 注意到
$$\mathbb{R}^n := \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} = \Pi_{k=1}^n \mathbb{R}$$
  - 我们希望
$$\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R}) = \Pi_{k=1}^n \mathcal{B}(\mathbb{R})$$
  - 但 $\Pi_{k=1}^n \mathcal{B}(\mathbb{R})$ 不是 σ-代数.
  - 因此, 定义
$$\mathcal{B}(\mathbb{R}^n) := \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})) = \sigma(\Pi_{k=1}^n \mathcal{B}(\mathbb{R}))$$
  - 我们也可以类比 $\mathcal{B}(\mathbb{R})$ 的定义方法, 先定义 $\mathbb{R}^n$ 上的立方体
$$I = I_1 \times I_2 \times \cdots \times I_n = \Pi_{k=1}^n I_k,\ I_k = (a_k, b_k].$$
    然后定义 $\mathcal{B}(\mathbb{R}^n)$ 为:
$$\mathcal{B}(\mathbb{R}^n) := \sigma(\{I : I\text{为}\mathbb{R}^n\text{上的立方体}\}).$$

随机向量

定义 2.8 (n 维随机向量) 考虑可测空间 $(\Omega, \mathcal{F})$, 若映射 $X: \Omega \to \mathbb{R}^n$ 使得对任意的 $B \in \mathcal{B}(\mathbb{R}^n)$ 均有

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

则称 X 为 n 维随机向量.

注 2.3 若记 $X = (X_1, X_2, \cdots, X_n)$, 则 X 为 n 维随机向量当且仅当 $X_k, k = 1, \cdots, n$ 为随机变量.

定义 2.9 (复随机变量) 考虑可测空间 $(\Omega, \mathcal{F})$ 以及映射 Z

$$Z(\omega) = X(\omega) + \mathrm{i}Y(\omega) \in \mathbb{C}$$

其中 $\mathrm{i} = \sqrt{-1}$ 是虚数单位. 如果 $(X(\omega), Y(\omega))$ 构成二维随机向量, 则称 Z 为复随机变量.

第十讲：随机变量的分布与性质

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　数学学院

分布与分布函数

定理 2.5 设 X 为 $(\Omega, \mathcal{F}, P)$ 上的随机变量，对于 Borel 集 B, 定义集函数 $\mathbf{F}(B)$ 如下:

$$\mathbf{F}(B) := P(X^{-1}(B)) = P \circ X^{-1}(B) = P(X \in B) \tag{23}$$

则 $\mathbf{F}(\cdot)$ 为 $(R, \mathcal{B})$ 上的概率，称之为随机变量 X 的诱导概率测度.

定义 2.10 称由 (23) 式定义在 $(R, \mathcal{B})$ 上的概率测度 $\mathbf{F}(\cdot)$ 为随机变量 X 的概率分布，简称分布.

- 给定概率空间 $(\Omega, \mathcal{F}, P)$，任给随机变量均可在 $(R, \mathcal{B})$ 上诱导出一个概率测度。由此可见，在同一个可测空间上可以定义不同的概率测度.
- 对于任意的 $B \in \mathcal{B}$，随机变量 X 落入 B 中的概率可通过 B 的概率测度 $\mathbf{F}(B)$ 得出。这也就是说，概率分布 $\mathbf{F}(\cdot)$ 完全刻画了随机变量 X 取值的概率规律.

随机变量的分布函数

如果我们将 $(R, \mathcal{B})$ 上的测度仅局限于集类 $\mathcal{P} := \{(-\infty, x], x \in R\}$ 上，由于 $\mathcal{P}$ 中的每条半直线被它的右端点 x 所决定，于是集函数 F 就化为 R 上的点函数.

定义 2.11 对于随机变量 X 而言，称 x 的函数

$$F(x) := \mathbf{F}((-\infty, x]) = P(X \leq x)$$

为 X 的概率分布函数或累积分布函数，简称分布函数并记作 $X \sim F(x)$，有时也以 $F_X(x)$ 表明是 X 的分布函数.

注 2.4 也有一些教材按如下方式定义分布函数:

$$F(x) := \mathbf{F}((-\infty, x)) = P(X < x)$$

分布函数的性质

定理 2.6 任一分布函数 $F(x)$ 都具有以下三条基本性质

1. 单调性非降性：$F(x)$ 是单调非减函数即对任意的 $x_1 < x_2$，有 $F(x_1) \leq F(x_2)$；
2. 右连续性：$F(x)$ 是 x 的右连续函数，即

$$F(x_0) = F(x_0+) := \lim_{x \to x_0+} F(x)$$

3. 规范性：对任意的 x 有，$0 \leq F(x) \leq 1$ 且

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$
$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1$$

注 2.5 在函数论中，我们一般不从随机变量出发来定义分布函数，而是直接定义满足上面三条性质的函数为分布函数.

分布函数性质的证明

1. 对任意的 $x < y$, $F(y) - F(x) = P(x < X \leq y) \geq 0$;
2. 因 $F(x)$ 是单调有界非降函数，所以其任一点 x_0 的右极限 $F(x_0+)$ 必存在，为证其连续性，只需证对单调上下降且收敛至 x_0 的数列 $\{x_n\}$ 有 $\lim_{n\to\infty} F(x_n) = F(x_0)$ 即可。注意到

$$\begin{aligned}\lim_{n\to\infty} F(x_n) &= \lim_{n\to\infty} P(X \leq x_n) = P(\cap_{n=1}^{\infty}\{X \leq x_n\}) = P(X \leq x_0) \\ &= F(x_0)\end{aligned}$$

3. 由 F 的单调性及概率的连续性可知

$$\begin{aligned}F(+\infty) &= \lim_{n\to+\infty} F(n) = \lim_{n\to+\infty} P(X \leq n) \\ &= P(\cup_{n=1}^{\infty}\{X \leq n\}) = P(X < \infty) = 1\end{aligned}$$

同理可证 $F(-\infty) = 0$.

事件概率的分布函数表示

- $P(X > b) = 1 - F(b)$;
- $P(a < X \leq b) = F(b) - F(a)$;
- $P(X < a) = F(a-)$;
- $P(X = a) = F(a) - F(a-)$;
- $P(X \geq b) = 1 - F(b-)$;
- $P(a \leq X < b) = F(b-) - F(a-)$;
- $P(a \leq X \leq b) = F(b) - F(a-)$;
- $P(a < X < b) = F(b-) - F(a)$;
- 对于不交区间并: $P(X \in \cup_{i=1}^{n}(a_i, b_i]) = \sum_{i=1}^{n}[F(b_i) - F(a_i)]$;
- 一般的，$P(X \in B) = \int_B dF(x)$.

随机变量与其分布 (函数) 之间的关系

- 分布函数是随机变量最基本的性质，也是对随机变量概率特性最全面的描述。为清晰起见，通常记随机变量 X 的分布函数为 $F_X(x)$.
- 分布函数和随机变量密不可分. 每个随机变量都有自己的分布函数; 反过来, 每个分布函数都能找到与之相对应的随机变量.
- 尽管随机变量与其分布函数密切关联，但是两者间并没有一一对应关系。一个随机变量只能有唯一的分布函数, 不同的随机变量却可能具有相同的分布函数.

例 2.4 (不同的随机变量有相同的分布) 考虑随机变量 X, 其分布满足

$$P(X=1)=P(X=-1)=\frac{1}{2}.$$

设随机变量 $Y=-X$, 那么 Y 的分布为

$$P(Y=1)=P(Y=-1)=\frac{1}{2}$$

显然, X 和 Y 的分布函数完全相同, 但明显是两个不同的随机变量.

分布函数表示与 Borel 域上概率测度之间的关系

定理 2.7 任给一个定义在 Borel 域 $\mathcal{B}(\mathbb{R})$ 上的概率 $\bar{P}$, 都存在如下定义的分布函数 F 与之相对应: $F(x)=\bar{P}((-\infty,x])$. 反过来, 任给一个定义于实数轴上的分布函数 F, 都存在定义在 Borel 域 $\mathcal{B}(\mathbb{R})$ 上的概率 $\bar{P}$ 与之相对应.

证明: 第一个结论可直接验证. 由概率的单调性, 若 $x_1 \leqslant x_2$, 则

$$\bar{P}\left((-\infty,x_1]\right) \leqslant \bar{P}\left((-\infty,x_2]\right) \Rightarrow F\left(x_1\right) \leqslant F\left(x_2\right),$$

从而得到 F 的单调不减. 再由概率的连续性,

$$\begin{aligned}\lim_{x_n\downarrow x}(-\infty,x_n]=(-\infty,x] \quad &\Rightarrow \quad \lim_{x_n\downarrow x}\bar{P}\left((-\infty,x_n]\right)=\bar{P}((-\infty,x]) \\ &\Rightarrow \quad \lim_{x_n\downarrow x}F\left(x_n\right)=F(x)\end{aligned}$$

定理的第二个结论本质上是概率 "扩张". 根据测度扩张定理, 由分布函数所确定的定义在 $\mathcal{P}$ 上的集函数 $\mathbf{F}((-\infty,x]):=F(x)$ 可以唯一的扩张到 $\mathcal{B}:=\sigma(\mathcal{P})$ 上, 成为 $\mathcal{B}$ 上的概率测度. 扩张后的概率测度称之为分布函数 $F(x)$ 所引出的勒贝格 - 斯蒂尔吉斯测度。实际上这个 $\mathbf{F}$ 正好是我们前面引进的概率分布.

离散型随机变量及分布

定义 2.12 (离散型随机变量) 如果随机变量 X 只取有限个值 $x_1, x_2, \cdots, x_n$ 或可列个值 $x_1, x_2, \cdots$，就称 X 为离散型随机变量，简称离散随机变量，其分布函数称之为离散型的.

定义 2.13 (离散型随机变量的分布列或概率质量函数) 对于离散型随机变量 X, 称 X 取值 x_k 的概率

$$p_k := p(x_k) = P(X = x_k), k = 1, 2, \cdots,$$

为 X 的概率分布列或简称为分布列，记 $X \sim \{p_k\}$. 分布列也常用下面的矩阵来表示

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_k, & \cdots \\ p_1, & p_2, & \cdots, & p_k, & \cdots \end{pmatrix}$$

容易验证，分布列有以下性质

1. 非负性：$p_k \geq 0, k = 1, 2, \cdots;$
2. 正则性：$\sum_k p_k = 1$

离散型随机变量的概率分布及其分布函数

- 由概率分布的定义，对任意的 $B \in \mathcal{B}$，我们有

$$\begin{aligned}\mathbf{F}(B) &= P(X \in B) = P(\cup_{k:x_k \in B}\{X = x_k\}) \\ &= \sum_{k:x_k \in B} P(X = x_k) = \sum_{k:x_k \in B} p_k\end{aligned}$$

- 由分布函数的定义知，

$$\begin{aligned}F(x) &= P(X \le x) = P(\cup_{k:x_k \le x}\{X = x_k\}) \\ &= \sum_{k:x_k \le x} P(X = x_k) = \sum_{k:x_k \le x} p_k \\ &= \sum_{k=1}^{\infty} p_k U(x - x_k)\end{aligned}$$

此处，$U(x) = \begin{cases} 1, & x \ge 0, \\ 0, & x < 0, \end{cases}$ 为阶跃函数.

- 易见离散型随机变量 X 的分布函数是一个纯跳跃函数：在 X 的每个可能取值 x_k 上有跃度 p_k，在每个不含 x_k 的区间上恒取常值.

退化随机变量

例 2.5 常数 c 可看作仅取一个值的随机变量 X, 即

$$P(X=c)=1$$

这个分布常称为 单点分布 或 退化分布，其分布函数为

$$F(x)=\begin{cases} 0, & x<c \\ 1, & x\geq c \end{cases}$$

例 2.6 计算例 2.1 中的所有随机变量的分布列或概率质量函数.

- X 表示正面朝上的次数，其概率质量函数 p_X 为:

$$p_X(0) = P(X = 0) = 1/4, \quad p_X(1) = P(X = 1) = 1/2,$$
$$p_X(2) = P(X = 2) = 1/4, \quad p_X(x) = P(X = x) = 0, x \neq 0, 1, 2.$$

- $Y = 2 - X$, 表示反面朝上的次数。注意到

$$P(Y = y) = P(2 - X = y) = P(X = 2 - y) = p_X(2 - y),$$

因此，随机变量 Y 的概率质量函数为

$$p_Y(0) = P(Y = 0) = 1/4, \quad p_Y(1) = P(Y = 1) = 1/2,$$
$$p_Y(2) = P(Y = 2) = 1/4, \quad p_Y(y) = P(Y = y) = 0, y \neq 0, 1, 2.$$

- I 表示第一次是否正面朝上的示性随机变量.

$$p_I(0) = P(I = 0) = 1/2, \quad p_I(1) = P(I = 1) = 1/2,$$
$$p_I(i) = P(I = i) = 0, i \neq 0, 1.$$

X, Y 和 I 的概率质量函数图

例 2.7 掷两颗骰子，其样本空间 Ω 含有 36 个等可能的样本点

$$\Omega = \{(x, y) : x, y = 1, 2, \cdots, 6\}$$

令 X 和 Y 表示每个骰子分别出现的点数。试求下面随机变量的分布列:

1. $T_1 := X + Y =$ 骰子点数之和;
2. $T_2 := 14 - (X + Y)$;
3. $T_3 :=$ 点数为 6 点的骰子的个数;
4. $T_4 := \max\{X, Y\} =$ 骰子的最大点数

- T_1, T_2 的概率分布列为

$$\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

图 3.4　两个骰子的点数之和的概率质量函数。

- T_3 的概率分布列为

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{25}{36} & \frac{10}{36} & \frac{1}{36} \end{pmatrix}$$

- T_4 的概率分布列为

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{36} & \frac{3}{36} & \frac{5}{36} & \frac{7}{36} & \frac{9}{36} & \frac{11}{36} \end{pmatrix}$$

连续型随机变量及分布

定义 2.14 (连续型随机变量) 设 X 为一随机变量，$F(x)$ 为随机变量 X 的分布函数，如果存在非负可积函数 $p(x)$ 使得

$$F(x) = \int_{-\infty}^{x} p(y)dy \tag{24}$$

则称 X 为连续型随机变量，其分布函数称之为连续型分布函数，函数 $p(x)$ 称为 X 的概率密度函数，简称密度函数或密度.

注 2.6

- 能够表为 (24) 式变上限积分的函数 $F(x)$ 在分析中称为绝对连续函数。绝对连续函数必为连续函数.
- 在若干个点上或零测集上改变密度函数 $p(x)$ 的值并不影响其积分的值，从而不影响分布函数 $F(x)$ 的值，这意味着连续分布的密度函数不唯一.

密度函数的性质

容易验证，随机变量 X 的密度函数有以下性质

1. 非负性：$p(x) \geq 0$;
2. 正则性：$\int_{-\infty}^{\infty} p(x)dx = 1$.

连续型随机变量分布的一些常见性质

- $p(x) = F'(x)$;
- $P(a < X \leq b) = F(b) - F(a) = \int_a^b p(x)dx$;
- $0 \leq P(X = a) \leq P(X \in (a - \epsilon, a)) = \int_{a-\epsilon}^{a} p(x)dx \xrightarrow{\epsilon \to 0} 0$, 故 $P(X = a) = 0$，即连续型随机变量取值单点的概率为 0;
- $P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X < b)$;
- 对任意的 Borel 集 B,

$$P(X \in B) = \int_B p(x)dx$$

- $P(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} p(y)dy = p(\xi)\Delta x \approx p(x)\Delta x$

一个连续型随机变量的例子

例 2.8 考虑概率空间 $(\Omega, \mathcal{F}, P)$, 其中 $\Omega := [0,1], \mathcal{F} := \mathcal{B}([0,1])$, $P(A) := m(A)$ 即 A 的 Lebesgue 测度. 定义

$$\theta(\omega) = \omega, \ \forall \omega \in [0,1].$$

显然, $\theta(\omega)$ 是随机变量, 且其分布函数为

$$F(x) = P(\theta(\omega) \leq x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

其分布密度为

$$p(x) = F'(x) = 1, x \in (0,1).$$

例 2.9 定义函数 $F(x)$ 如下

$$F(x)=\begin{cases}0, & x<0,\\ \dfrac{1+x}{2}, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

试说明

1. $F(x)$ 为分布函数;
2. $F(x)$ 既非离散型也非连续型分布;
3. $F(x)$ 可分解为

$$F(x)=\frac{1}{2}F_1(x)+\frac{1}{2}F_2(x)$$

其中

$$F_1(x)=\begin{cases}0, & x<0,\\ 1, & x\ge 0.\end{cases}\qquad F_2(x)=\begin{cases}0, & x<0,\\ x, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

勒贝格分解

定理 2.8 (勒贝格分解) 对任一分布函数 $F(x)$ 有如下分解

$$F(x) = c_1F_1(x) + c_2F_2(x) + c_3F_3(x),$$

其中常数 $c_1, c_2, c_3 \geq 0, c_1 + c_2 + c_3 = 1$, 而 $F_1(x), F_2(x), F_3(x)$ 都是分布函数，$F_1(x)$ 为纯跳跃离散型函数，$F_2(x)$ 为连续型分布函数，$F_3(x)$ 为奇异型分布函数 (见教材 3.6 节奇异连续型分布构造).

- 上述定理中奇异函数的含义及定理的证明可参见一般的实变函数论教科书，这里我们不再详述，仅指出几种特殊情况:
  - 在分解式中取 $c_1 = 1, c_2 = c_3 = 0$ 便得到我们所讨论的离散型分布函数;
  - 在分解式中取 $c_2 = 1, c_1 = c_3 = 0$ 便得到连续型分布函数;
  - 若取 $c_3 = 0, c_1 \neq 0, c_2 \neq 0, c_1 + c_2 = 1$ 便得到离散与连续混合分布
- 从上面分析可看出，随机变量除了离散型与连续型外还有很多其他类型.

分布函数与随机变量

$$\left.\begin{array}{r}(\Omega,\mathcal{F},P)\\ \\ X(\omega)\end{array}\right\}\Rightarrow P(X\le x)=:F(x)\Rightarrow\left\{\begin{array}{l}\text{单调非降性}\\ \text{右连续性}\\ \text{规范性}\end{array}\right.$$

$$\left.\begin{array}{r}\text{单调非降性}\\ \text{右连续性}\\ \text{规范性}\end{array}\right\}\text{的}F(x)\text{给定}\Rightarrow\text{存在}\left\{\begin{array}{l}(\Omega,\mathcal{F},P)\\ \\ X(\omega)\end{array}\right.\quad\text{使得}F(x)=P(X\le x)?$$

随机变量的存在性问题分析

定理 2.9 若 $F(x)$ 是右连续、单调非降函数，且 $F(-\infty)=0, F(+\infty)=1$, 则存在一个概率空间 $(\Omega, \mathcal{F}, P)$ 及其上的随机变量 $X(\omega)$, 使 $X(\omega)$ 的分布函数恰好是 $F(x)$.

分析:

- 对 $[0,1]$ 上的均匀分布随机变量 $\theta(\omega)=\omega$, 其分布函数为

$$P(\theta(\omega) \leq x) = P(\omega \in [0,x]) = x, \ \forall x \in [0,1];$$

- $F(x) \in [0.1]$, 若将上式中的 x 替换为 $F(x)$, 则有

$$P(\theta(\omega) \leq F(x)) = F(x);$$

- 若分布函数 $F(x)$ 可逆，则

$$P(F^{-1}(\theta(\omega)) \leq x) = P(\theta(\omega) \leq F(x)) = F(x);$$

- 考虑 $X(\omega) := F^{-1}(\theta(\omega))$, 则 $F_X(x) = P(F^{-1}(\theta(\omega)) \leq x) = F(x)$;

随机变量的存在性问题分析

- $F(x)$不一定可逆, 能否定义一个函数 G 使得

$$\{\theta(\omega) \le F(x)\} = \{G(\theta(\omega)) \le x\}?$$

- 上述等式也即寻求如下等价性

$$G(\theta) \le x \Leftrightarrow F(x) \ge \theta$$

- 因此，若函数 G 存在，则对任意给定的 θ 必须满足

$$G(\theta) \le x, \quad \forall x \in \{x : F(x) \ge \theta\}$$

- 故所寻找的函数 G 需有以下性质

$$G(\theta) \le \inf\{x : F(x) \ge \theta\}$$

- 上述不等式右侧的下确界也是 G 的一种选择，可以证明选此下确界作为函数 G 的定义的确具有我们所要求的性质.

单调逆 (一般逆) 的定义

定义 2.15 设 $F(x)$ 是右连续、单调非降函数，且 $F(-\infty)=0, F(+\infty)=1$. 对任意的 $p\in(0,1)$, 我们称

$$F^{-1}(p) := \inf\{x : F(x) \geq p\},$$

为函数 $F(x)$ 的单调逆或一般逆.

注 2.7 $F^{-1}(p)$ 在概率论中也称为函数 $F(x)$ 的 p 分位数函数或与其相对应的随机变量 X 的 p 分位数，通常用 x_p 或 ξ_p 来表示.

定理 2.10 设 $F(x), F^{-1}(p)$ 定义如上，则

$$F^{-1}(p) \leq x \Leftrightarrow F(x) \geq p$$

证明: 令 $A := \{y : F(y) \geq p\}$, 则

- $\Leftarrow$: $F(x) \geq p$ 蕴含 $F^{-1}(p) := \inf A \leq x$
- $\Rightarrow$: $F^{-1}(p) \leq x$ 即 $x \geq F^{-1}(p) := \inf A =: x_0$
  - 若 $x_0 \in A$, 则显然 $F(x) \geq F(x_0) \geq p$
  - 若 $x_0 \notin A$, 则存在 $\{x_n\}_{n\geq 1} \subset A$ 使得 $x_n > x_0$ 且 $\lim_{n\to\infty} x_n = x_0$,从而

$$F(x) \geq F(x_0) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) \geq p;$$

随机变量存在性的证明

定理 2.9 的证明:

- 取 $\Omega = [0,1]$, $\mathcal{F}$ 为 $[0,1]$ 上的 Borel 集全体，取 P 为直线上的 Lebesgue 测度 (是长度概念的推广，但对一切 Borel 集有定义);
- 定义: $\theta(\omega) = \omega$, 则 $\theta(\omega)$ 是 $(\Omega, \mathcal{F}, P)$ 上的随机变量，并且，对一切的 $0 \le x \le 1$,

$$P(\theta(\omega) \le x) = P(\omega \in [0,x]) = x;$$

- 考虑 $X(\omega) := F^{-1}(\theta(\omega))$, 则

$$F_X(x) = P(X \le x) = P(F^{-1}(\theta(\omega)) \le x) \xlongequal[p \le F(x)]{F^{-1}(p) \le x} P(\theta(\omega) \le F(x)) = F(x)$$

单调逆 (一般逆) 的性质

- $F^{-1}(F(x)) \le x, \forall x \in R$: 设 $x_0 \in R$ 任意给定且 $F(x_0) = p_0$, 则

$$\begin{aligned} F^{-1}(F(x_0)) = F^{-1}(p_0) &= \inf\{x : F(x) \ge p_0\} \\ &= \inf\{x : F(x) \ge F(x_0)\} \le x_0 \end{aligned}$$

- $F(F^{-1}(p)) \ge p, \forall p \in (0,1)$:设 $p_0 \in (0,1)$ 任意给定且

$$F^{-1}(p_0) = x_0 = \inf\{x : F(x) \ge p_0\} =: \inf A,$$

  - 若 $x_0 \in A$, 则显然有 $F(F^{-1}(p_0)) = F(x_0) \ge p_0$;
  - 若 $x_0 \notin A$, 则 存在 $\{x_n\}_{n\ge1} \subset A$ 使得 $x_n > x_0$ 且 $\lim_{n\to\infty} x_n = x_0$,从而由函数 $F(x)$ 的右连续性可知

$$F(F^{-1}(p_0)) = F(x_0) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) \ge p_0;$$

- $F^{-1}(p)$ 关于 p 单调非降: 集合 $\{x : F(x) \ge p\}$ 关于 p 单调不增，故其下确界也关于 p 单调非降

单调逆 (一般逆) 的性质

下面的几条性质

- $F^{-1}(p)$ 关于 p 是左连续的;
- $F^{-1}(p) := \inf\{x : F(x) \geq p\} = \sup\{x : F(x) < p\}$
- $x < F^{-1}(p) \Leftrightarrow F(x) < p$

思考

若 $F(x)$ 是左连续、单调非降函数，且 $F(-\infty)=0, F(+\infty)=1$, 是否存在一个概率空间 $(\Omega, \mathcal{F}, P)$ 及其上的随机变量 $X(\omega)$, 使得

$$F(x)=P(X<x).$$

第十一讲：常见的离散型随机变量

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

伯努利实验

- 只有两种可能结果的试验称为伯努利试验；例如抽检产品，可能是合格品，也可能是次品；掷两颗骰子，可能得到同点，也可能得到不同点，等等都是伯努利试验.
- 伯努利试验的样本空间 Ω 并不一定只含有两个样本点，有时只是把我们所关心的一部分样本点归结为一种结果 A, 同时把其余的样本点的集合看作另一种结果 $\bar{A}$;
- 在上述掷骰子的试验中，样本空间 Ω 共含有 36 个样本点，如果我们只关心同点是否发生，就可以把其中的 6 个样本点组成的事件 $A := \{(i,i) : i = 1, \cdots, 6\}$ 视为一种结果，而其余的 30 个样本点组成另一结果 $\bar{A} := \{\text{不同点}\}$;
- 此外我们不再关心由 Ω 的其他非空子集组成的事件，于是对于伯努利试验而言，事件 σ 代数应取为 $\mathcal{F} = \{\emptyset, A, \bar{A}, \Omega\}$;
- 通常把结果 A 称作" 成功", 而把 $\bar{A}$ 称作" 失败";
- 再取定成功失败的概率 $p = P(A), q = P(\bar{A})$ $(p > 0, q > 0$ 且 $p + q = 1)$, 则建立了一次伯努利试验的概率空间 $(\Omega, \mathcal{F}, P)$.

- 在概率论的理论与应用中，经常以一系列独立重复的伯努利试验作为概率模型;
- 所谓重复，粗略的说即各次试验的概率空间都是上述的 $(\Omega, \mathcal{F}, P)$;
- 而 n 个试验的独立性则是指各次试验的结果互不影响，即对于第 i 次试验的任何结果 $E_i (i = 1, \cdots, n)$ 均有

$$P(E_1 E_2 \cdots E_n) = P(E_1) P(E_2) \cdots P(E_n)$$

- 将一次伯努利试验独立重复 n 次，称作 n 重伯努利试验;
- 将一次伯努利试验独立地重复下去所得到的一系列试验，称为可列重伯努利试验.

二项分布:n 重伯努利试验中成功的次数 X

- X:n 重伯努利试验中成功 (事件 A 发生) 的次数;
- X 的所有可能取值为: $0, 1, \cdots, n$;
- 下面我们考虑 X 的分布列
  - n 重伯努利试验的样本空间:
    $\Omega := \{\omega = (\omega_1, \cdots, \omega_n) : \omega_i \text{或者为} A \text{或者为} \bar{A}\}$
  - 样本空间样本点的个数为 2^n 个;
  - $\{X = k\} = \{\omega = (\omega_1, \cdots, \omega_n) : \omega_1, \cdots, \omega_n \text{中有} k \text{个} A\}$, 共包含 C_n^k 个样本点;
  - 若任给样本点 $\omega = (\omega_1, \cdots, \omega_n) \in \{X = k\}$, 则意味着 $\omega_1, \cdots, \omega_n$ 中有 k 个 A, $n-k$ 个 $\bar{A}$, 故由独立性可知

$$P(\omega) = p^k (1-p)^{n-k}$$

  - 而事件 $\{X = k\}$ 中共有 C_n^k 个类似的 ω, 故

$$P(X = k) = C_n^k p^k (1-p)^{n-k} := b(k; n, p), k = 0, 1, \cdots, n$$

    这个分布常称为二项分布，记为 $X \sim B(n, p)$.
  - 容易验证，$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$

两点分布

- $n=1$ 时的二项分布 $B(1,p)$ 称为二点分布，或 $0-1$ 分布，或称伯努利分布，其分布列为

$$P(X=k)=p^k(1-p)^{1-k}, k=0,1$$

或记为

$$\begin{pmatrix} 0, & 1 \\ 1-p, & p \end{pmatrix}$$

- $B(1,p)$ 主要用于描述一次伯努利试验中成功的次数 (0 或 1);
- 若记 X_i 表示第 i 次伯努利试验中成功的次数，则 X_i 相互独立且有

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

即二项分布随机变量可写为 n 个独立同为两点分布随机变量的和.

二项分布的性质

- 对 $k \geq 1$, 考虑比值

$$\begin{aligned}\frac{b(k;n,p)}{b(k-1;n,p)} &= \frac{C_n^k p^k (1-p)^{n-k}}{C_n^{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)} \\ &= \frac{k(1-p)+(n+1)p-k}{k(1-p)} = 1 + \frac{(n+1)p-k}{k(1-p)}\end{aligned}$$

- 当 $(n+1)p > k$ 时，$b(k;n,p) > b(k-1;n,p)$;
- 当 $(n+1)p < k$ 时，$b(k;n,p) < b(k-1;n,p)$;
- 从而，对于固定的 $n,p,\{X=k\}$ 的概率 $b(k;n,p)$ 先随 k 增大而增大，再随 k 增大而减小，故必有最大值:
  - 如果 $m := (n+1)p$ 为整数，则 $b(m;n,p) = b(m-1;n,p)$ 同为 $b(k;n,p)$ 的最大值
  - 如果 $(n+1)p$ 不为整数，则 $b(k;n,p)$ 在 $m = [(n+1)p]$ 处取到最大值 (此处 $[a]$ 表示不超过 a 的最大整数)
- 我们称使得 $b(k;n,p)$ 取得最大值的 m 为二项分布随机变量的最可能值，或称为 n 重伯努利试验中最可能的成功次数.

二项分布的性质

定理 2.11 设 $X \sim B(n,p)$, 且 $q = 1-p$(通常用 q 表示伯努利试验失败的概率), 则有 $n - X \sim B(n,q)$.

- 借用二项分布的直观定义：将 X 为 n 次独立伯努利试验成功的次数，则 $n-X$ 为这些试验中失败的次数.
- 互相交换成功与失败的角色，可知 $n - X \sim B(n,q)$.
- 也可从分布列 (概率质量函数) 的角度出发得到 $n - X \sim B(n,q)$.
- 令 $Y = n - X$, 则 Y 的分布列 (概率质量函数) 为

$$
\begin{aligned}
P(Y=k) =& P(n-X=k) = P(X=n-k) \\
=& \binom{n}{n-k} p^{n-k} q^k = \binom{n}{k} q^k p^{n-k},
\end{aligned}
$$

二项分布的性质

定理 2.12 设 $X \sim B(n,p)$, 其中 n 为偶数，$p = 1/2$, 则 X 的分布关于 $n/2$ 对称，即对任意的非负整数 j, 均有

$$P(X = \frac{n}{2} + j) = P(X = \frac{n}{2} - j).$$

解: 由定理 2.11 可知，$n - X$ 同样服从 $B(n, 1/2)$. 因此对任意非负整数 k 均有

$$P(X = k) = P(n - X = k) = P(X = n - k).$$

令 $k = n/2 + j$, 即可得证.

例 2.10 设每台自动机床在运行过程中需要维修的概率均为 $p = 0.01$, 并且各机床是否需要维修相互独立。如果:

1. 每名维修工人负责看管 20 台机床;
2. 3 名维修工人负责看管 80 台机床;

求机床不能及维修的概率.

解: 1. 这是 $n = 20$ 重伯努利试验，参数 $p = 0.01$, 故需要维修的机床数 X 服从 $B(20, 0.01)$ 分布。故不能及时维修的概率为

$$\begin{aligned} P(X > 1) &= 1 - P(X \le 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - C_{20}^{0} 0.01^{0} (1 - 0.01)^{20} - C_{20}^{1} 0.01 (1 - 0.01)^{20-1} \approx 0.0169 \end{aligned}$$

2. 此时需要维修的机床数 X 服从 $B(80, 0.01)$ 分布，类似可得不能及时维修的概率为

$$P(X > 3) = 1 - \sum_{k=0}^{3} b(k; 80, 0.01) \approx 0.0087$$

小概率事件终将发生

例 2.11 在可列重伯努利试验中，求事件 $E := \{试验终将成功\}$ 的概率.

解: 考虑所求概率事件的反面即 $\overline{E} := \{试验永不成功\}$.若我们记

$$F_n := \{前n次试验均失败\},$$

则易知，$\{F_n\}$ 为单调下降事件序列，且

$$\lim_{n\to\infty} F_n = \cap_{n=1}^{\infty} F_n = \overline{E}$$

从而

$$P(\overline{E}) = P(\lim_{n\to\infty} F_n) = \lim_{n\to\infty} P(F_n) = \lim_{n\to\infty} C_n^0 p^0 (1-p)^n = 0$$

故

$$P(E) = 1 - P(\overline{E}) = 1 - 0 = 1$$

无论成功的概率有多小，但是试验最终成功的概率为 1, 也就是说小概率事件终将发生的概率为 1.

几何分布：可列重伯努利试验中首次成功的等待时间 X

- 记 X 为可列重伯努利试验中首次成功的等待时间即首次成功所需要试验的次数;
- $\{X=k\}=\{\underbrace{\overline{A}\cdots\overline{A}}_{k-1\text{个}}A\}$, 故

$$P(X=k)=(1-p)^{k-1}p:=g(k;p),k=1,\cdots,$$

这个分布常称为几何分布，记为 $X\sim G(p)$.

几何分布的性质

定理 2.13 取值自然数的随机变量 X 为几何分布当且仅当 X 有无记忆性:

$$P(X > m+n | X > m) = P(X > n), \text{对任意的} m, n \geq 1. \tag{25}$$

证明: 若 X 为几何分布，则

$$P(X > m+n | X > m) = \frac{P(X > m+n, X > m)}{P(X > m)} = \frac{P(X > m+n)}{P(X > m)}$$

而

$$P(X > n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = \frac{(1-p)^n p}{1-(1-p)} = (1-p)^n$$

故

$$P(X > m+n | X > m) = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n = P(X > n)$$

若 X 具有无记忆性，则由 ~ (25) 知

$$Q_n := P(X > n) > 0, \quad \text{对任意的} n \geq 1$$

并且有

$$Q_{m+n} = P(X > m+n) = P(X > m)P(X > m+n | X > m) = Q_m Q_n$$

从而

$$Q_m = Q_1^m$$

注意到 $Q_1 \in (0,1)$, 事实上,

- $Q_1 = P(X > 1) > 0$ 显然;
- 若 $Q_1 = 1$, 则对一切的 m 均有 $Q_m = P(X > m) = 1$, 这与 X 取自然数矛盾，故 $Q_1 \in (0,1)$.

故取 $p = 1 - Q_1 \in (0,1)$, 且对任意的 $k \geq 1$ 有

$$\begin{aligned} P(X = k) &= P(X > k-1) - P(X > k) = Q_{k-1} - Q_k \\ &= (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1} p \end{aligned}$$

无记忆性

$$P(X > m + n | X > m) = P(X > n)$$

- 上述的无记忆性表明：已知试验了 m 次未获得成功，再加做 n 次试验仍不成功的概率，等于从开始算起做 n 次试验都不成成功的概率.
- 换句话放，已做过的 m 次失败的试验被忘记了；
- 产生几何分布的这种无记忆性的根本原因在于，我们进行的是独立重复试验，这是不学习，不总结经验的试验，已经做过的试验当然不会留下记忆.

例 2.12 10 把外形相同的钥匙中只有一把能打开门。现一一试开，试对每次试毕放回与不放回两种情形，分别求事件
$E := \{$至多试3次能打开门$\}$ 的概率.

解: 1. 放回情形是独立重复试验，属伯努利概型。以 X 表示首次打开门的等待时间，则 X 服从几何分布 $G(0.1)$. 故所求概率为

$$P(E) = P(X \leq 3) = \sum_{k=1}^{3} (1-0.1)^{k-1} 0.1 = 0.271$$

2. 不放回情形不再是独立重复试验，适用于古典概型。样本点总数 $n(\Omega) = C_{10}^3$, 而 $n(\overline{E}) = C_9^3$. 故

$$P(E) = 1 - P(\overline{E}) = 1 - \frac{C_9^3}{C_{10}^3} = 0.3$$

或令 $A_i := \{$第i次取到能开门的钥匙$\}$, 则 A_1, A_2, A_3 互不相容，由可加性及抽签的公平性可得

$$P(E) = P(\cup_{i=1}^3 A_i) = \sum_{i=1}^{3} P(A_i) = 0.3$$

帕斯卡 (Pascal) 分布：第 r 次成功的等待时间

- 记 X_r 为可列重伯努利试验中第 r 次成功的等待时间即第 r 次成功所需要试验的次数;
- 易见 X 的所有可能取值为 $k = r, r+1, \cdots$，并且有

$$\begin{aligned}
\{X_r = k\} &= \{\text{前}k-1\text{次试验恰有}r-1\text{次成功且第}k\text{次成功}\} \\
P(X_r = k) &= P(\text{前}k-1\text{次试验恰有}r-1\text{次成功})P(\text{第}k\text{次成功}) \\
&= C_{k-1}^{r-1} p^{r-1}(1-p)^{k-1-(r-1)} p \\
&= C_{k-1}^{r-1} p^{r}(1-p)^{k-r} := f(k;r,p), \quad k = r, r+1, \cdots
\end{aligned}$$

这个分布常称为帕斯卡 (Pascal) 分布或负二项分布.

- $f(k;r,p), k = r, r+1, \cdots$，可以成为离散型分布的密度，事实上: $f(k;r,p) > 0$ 显然，其和

$$\begin{aligned}
\sum_{k=r}^{\infty} f(k;r,p) &= \sum_{k=r}^{\infty} C_{k-1}^{r-1} p^r (1-p)^{k-r} \xlongequal[q=1-p]{k-r=i} \sum_{i=0}^{\infty} C_{r+i-1}^{r-1} p^r q^{r+i-r} \\
&= \sum_{i=0}^{\infty} C_{r+i-1}^{i} p^r q^i = \sum_{i=0}^{\infty} C_{-r}^{i} p^r (-q)^i = p^r (1-q)^{-r} = 1
\end{aligned}$$

帕斯卡分布的性质

- 若帕斯卡分布中的 $r=1$, 则此时的帕斯卡分布即为几何分布;
- 如果记

$$\tau_1 = X_1, \quad \tau_n = X_n - X_{n-1}, \quad n > 1,$$

则随机变量 τ_n 是第 $n-1$ 次成功到第 n 次成功的间隔时间。显然有

$$X_r = \tau_1 + \tau_2 + \cdots + \tau_r$$

- 以后我们会看到: $\tau_1, \cdots, \tau_r$ 是 r 个相互独立的随机变量且每个 τ_k 均服从几何分布.

例 2.13 某人口袋中有两盒火柴，开始时每盒各装 n 根。每次他从口袋中任取一盒使用其中的一根火柴。求此人掏出一盒发现已空，而另一盒尚余 r 根的概率.

解: 记

$$E = \{\text{掏出甲盒已空而乙盒尚余}r\text{根}\}$$

则由对称性可知所求概率为 $2P(E)$. 若我们以取出甲盒为" 成功", 这便是一个成功率 $p = 1/2$ 的独立重复伯努利试验. 而

$$E = \{\text{第}n+1\text{次成功发生在第}2n-r+1\text{次试验}\}$$

故所求概率为

$$\begin{aligned} 2P(E) &= 2f(2n-r+1; n+1, 1/2) = 2C_{2n-r}^{n}(\frac{1}{2})^{n+1}(\frac{1}{2})^{2n-r-n} \\ &= C_{2n-r}^{n}2^{r-2n} \end{aligned}$$

分赌注问题

例 2.14 1654 年，当时的职业赌徒 DeMere 爵士向法国的大数学家 Pascal 提出如下问题：甲乙两人各下赌注 m 元，商定先胜三局者取得全部赌金。假定在每一局中二人获胜的机会相等，且各局胜负相互独立。如果当甲胜一局而乙胜零局时赌博被迫中止，问赌注如何分？

- 为解决这个问题，Pascal 与当时声望很高的数学家 Fermat 建立了通信联系。他们进行了卓有成效的讨论，不仅完满的回答了分赌注问题，而且为解决其他概率问题建立起了框架，极大的促进了概率论的建立与发展；
- Pascal 令人信服的指出，赌金的分法应当取决于若赌博能继续进行下去甲乙各自获胜的概率，这个概率即为在 $p = 0.5$ 的可列重伯努利试验中 2 次成功发生在 3 次失败之前的概率；
- 更一般的，下面我们求一下 n 次成功发生在 m 次失败之前的概率.

例 2.15 在可列重伯努利试验中，求下面事件的概率:

$$E = \{n\text{次成功发生在}m\text{次失败之前}\}$$

解: 记 $F_k = \{\text{第}n\text{次成功发生在第}k\text{次试验}\}$, 则

$$E = \cup_{k=n}^{n+m-1} F_k$$

从而由 F_k 的互不相容性可得

$$P(E) = \sum_{k=n}^{n+m-1} P(F_k) = \sum_{k=n}^{n+m-1} C_{k-1}^{n-1} p^n (1-p)^{k-n}$$

利用上面的公式可计算 $n = 2, m = 3, p = 1/2$ 时，其相应的概率为

$$P(\text{甲胜}) = P(E) \xlongequal{n=2,m=3,p=1/2} \frac{11}{16}$$

故赌注应以 $11:5$ 的比例分配给甲乙两人.

图 1: 泊松

- 西莫恩　德尼　泊松：法国数学家、几何学家和物理学家；
- 泊松的科学生涯开始于研究微分方程及其在摆的运动和声学理论中的应用；
- 对积分理论、行星运动理论、热物理、电磁理论、位势理论和概率论都有重要贡献；
- 19 世纪概率统计领域里的卓越人物，改进了概率论的运用方法，特别是用于统计方面的方法，建立了描述随机现象的一种概率分布　泊松分布；
- 推广了“大数定律”，并导出了在概率论与数理方程中有重要应用的泊松积分。

泊松 (Poisson) 分布

- Poisson 分布是概率论中一种重要的离散型分布，它在理论与实践中都有广泛的应用，常与单位时间 (面积) 内上的计数过程相联系；
  - 一天内到达某商场的顾客数;
  - 单位时间内，电路受外界电磁波的冲击次数;
  - 一定时期内，某放射性物质放射出来的粒子数等
- 在二项分布中，当参数 n 较大时，计算二项概率 $b(k;n,p)$ 会非常麻烦.

定理 2.14 设有一列二项分布 $\{b(k;n,p_n)\}$, 若其参数列 p_n 满足

$$\lim_{n\to\infty} np_n = \lambda > 0,$$

则对任何非负整数 k 有

$$\lim_{n\to\infty} b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

证明: 记 $\lambda_n := np_n$, 则

$$\begin{aligned} b(k;n,p_n) &= C_n^k p_n^k (1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}(\frac{\lambda_n}{n})^k(1-\frac{\lambda_n}{n})^{n-k} \\ &= \frac{\lambda_n^k}{k!}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})(1-\frac{\lambda_n}{n})^{n-k} \end{aligned}$$

注意到

$$\lim_{n\to\infty}\lambda_n^k = \lambda^k,$$

$$\lim_{n\to\infty}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n}) = 1$$

$$\lim_{n\to\infty}(1-\frac{\lambda_n}{n})^{n-k} = \lim_{n\to\infty} e^{(n-k)\ln(1-\frac{\lambda_n}{n})}$$

$$= e^{\lim_{n\to\infty}(n-k)\ln(1-\frac{\lambda_n}{n})} = e^{\lim_{n\to\infty}(n-k)(-\frac{\lambda_n}{n}+o(\frac{1}{n}))}$$

$$= e^{\lim_{n\to\infty}(-\lambda_n+\frac{k\lambda_n}{n}+(n-k)o(\frac{1}{n}))} = e^{-\lambda}$$

从而 $$\lim_{n\to\infty} b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0,1,2,\cdots$$

有了上述定理，当 n 很大而 p 很小时，可以用近似公式计算二项概率

$$b(k;n,p) \approx \frac{(np)^k}{k!}e^{-np}$$

这里我们要求 p 很小，为保证 n 很大时，乘积 np 有适度的大小.

定义 2.16 (泊松分布) 对参数 $\lambda > 0$, 记

$$p(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, \cdots .$$

易见 $p(k;\lambda) > 0$ 且

$$\sum_{k=0}^{\infty} p(k;\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda}e^{-\lambda} = 1$$

即 $\{p(k;\lambda)\}$ 可以看成离散型分布的密度，我们就把它称为以 λ 为参数的泊松分布，记作 $P(\lambda)$.

泊松分布的性质

- 泊松分布列 $p(k;\lambda)$ 随 k 变化情况与二项分布相似，事实上，考虑比值

$$\frac{p(k;\lambda)}{p(k-1;\lambda)} = \frac{\lambda^k (k-1)! e^{-\lambda}}{\lambda^{k-1} k! e^{-\lambda}} = \frac{\lambda}{k}, k \geq 1$$

  - 当 $k < \lambda$ 时有，$p(k;\lambda) > p(k-1;\lambda)$;
  - 当 $k > \lambda$ 时有，$p(k;\lambda) < p(k-1;\lambda)$;
  - 因此 $p(k;\lambda)$ 随 k 先升后降，在 $m = [\lambda]$ 处达到最大值，而 λ 为整数时，$p(k;\lambda)$ 在 $m = \lambda, \lambda - 1$ 处同时取到最大值.
- 泊松分布在随机选择下的不变性：假设某块放射性物质在单位时间内发射出的粒子数 X 服从 $P(\lambda)$ 分布。而每个粒子被记录下来的概率为 p 即粒子有 $1-p$ 的概率被计数器遗漏，如果各粒子是否被记录相互独立，试求记录下的粒子数 Y 的分布.

$$
\begin{aligned}
P(Y=k) &= \sum_{n=0}^{\infty} P(X=n, Y=k) = \sum_{n=0}^{\infty} P(X=n)P(Y=k|X=n) \\
&= \sum_{n=k}^{\infty} P(X=n)P(Y=k|X=n) = \sum_{n=k}^{\infty} p(n;\lambda)B(k;n,p) \\
&= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} C_n^k p^k (1-p)^{n-k} = \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} \frac{1}{k!} e^{-\lambda} (\lambda p)^k \\
&= \frac{1}{k!} e^{-\lambda p} (\lambda p)^k
\end{aligned}
$$

超几何分布

假设某个罐子中有 w 个白球和 b 个黑球，考虑如下两种取球方式

- 有放回的抓取 n 个球
  - $X := n$ 个球中白球的数量;
  - $X \sim B(n, \dfrac{w}{w+b})$
- 不放回的抓取 n 个球
  - $X := n$ 个球中白球的数量;
  - X 服从超几何分布，记为 $X \sim H(w, b, n)$;
  - 易知

$$P(X=k) = \frac{C_w^k C_b^{n-k}}{C_{w+b}^n} := h(k, w, b, n), k = 0, 1, \cdots, r,$$

  其中 $r := \min\{w, n\}$.
  - 若要验证以上给出的确实为一个概率分布列，只需注意到下面的组合等式成立

$$\sum_{k=0}^{r} C_w^k C_b^{n-k} = C_{w+b}^n$$

超几何分布：两个例子

- 不合格品抽检问题
  - 设有 N 件产品，其中有 M 件不合格品，从中不放回的抽取 n 件;
  - 令 $X := n$ 件抽取的产品中不合格品的件数;
  - $X \sim H(M, N-M, n)$
- 麋鹿的捕获 - 再捕获问题
  - 森林中共有 N 头麋鹿。某天，捕获了 M 头麋鹿，标记后将这 M 头麋鹿再放回野外.
  - 几天后，又重新随机地捕获 n 头麋鹿。假设重新捕获的麋鹿也同样可能是之前捕获的麋鹿.
  - 令 $X :=$ 再次被捕获的麋鹿的数量;
  - $X \sim H(M, N-M, n)$;
  - 第一次捕获的 M 头麋鹿相当于前面例子中白球的总数，第一次未捕获的 $N-M$ 头麋鹿相当于前面例子中黑球的总数，再次被捕获的 n 头麋鹿相当于前面例子中抽样的数量.

超几何分布的适用基础

- 除上面例子外，超几何分布还可出现在许多情况下;
- 超几何分布的适用基础是总体根据两套标签进行分类:
  - 在罐子的示例中，每个球不是白色就是黑色 (第一套标签);
  - 每个球要么是样本要么不是样本 (第二套标签);
- 两套标签中至少有一个是被完全随机分配的 (在罐子的例子中，球是随机抽样的);
- X 代表被两套标签都标记的数量：在罐子的例子中，关注的是既被抽样又是白色的球。
- X 服从超几何分布.

超几何分布之间的对称性

定理 2.15 $X \sim H(w, b, n)$, $Y \sim H(n, w+b-n, w)$, 则 X 和 Y 是同分布的.

证明:

- 考虑一个由 w 个白球和 b 个黑球充满的罐子，现在随机不放回地从罐子里抓取 n 个球;
- 白球或者黑球看作第一套标签，有没有被抽取看作第二套标签;
- X 表示抽取的样本球中白球的数量 $\sim H(w, b, n)$;
- 若将是否被抽取看作第一套标签，白球或者黑球看作第二套标签;
- Y 表示所有白球中被抽中的数量 $\sim H(n, w+b-n, w)$;
- 显然 X 和 Y 都是表示被抽取的白球数量，所以它们有相同的分布;
- 也可以用代数的方法来检查 X 和 Y 是否具有相同的概率质量函数，即验证 $P(X=k) = P(Y=k)$.

二项分布与超几何分布

定理 2.16 如果 $X \sim B(n,p), Y \sim B(m,p)$, 且 X 和 Y 相互独立，则当给定条件 $X+Y=r$ 时，X 的条件分布为超几何分布 $H(n,m,r)$.

定理 2.17 如果 $X \sim H(M,b,n)$, 且当 $N=M+b \to \infty$ 时 $p=\dfrac{M}{N}$ 保持不变，则 X 的分布列 (概率质量函数) 收敛到 $B(n,p)$ 的分布列.

超几何分布的二项近似

注意到

$$h(k;M,b,n)=\frac{C_M^kC_b^{n-k}}{C_{M+b}^n}=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$$

$$=\frac{\dfrac{M!}{k!(M-k)!}\dfrac{(N-M)!}{(n-k)!(N-M-(n-k))!}}{\dfrac{N!}{n!(N-n)!}}$$

$$=\frac{n!}{k!(n-k)!}\frac{M!}{(M-k)!}\frac{(N-M)!}{(N-M-(n-k))!}\frac{(N-n)!}{N!}$$

$$=\frac{n!}{k!(n-k)!}\frac{M(M-1)\cdots(M-(k-1))}{N(N-1)\cdots(N-(k-1))}\frac{(N-M)\cdots(N-M-(n-k)+1)}{(N-k)\cdots(N-n+1)}$$

$$=\frac{n!}{k!(n-k)!}\frac{\frac{M}{N}(\frac{M}{N}-\frac{1}{N})\cdots(\frac{M}{N}-\frac{(k-1)}{N})}{1(1-\frac{1}{N})\cdots(1-\frac{(k-1)}{N})}\frac{(1-\frac{M}{N})\cdots(1-\frac{M}{N}-\frac{(n-k)-1}{N})}{(1-\frac{k}{N})\cdots(1-\frac{n-1}{N})}$$

$$\to C_n^k\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}=b(k;n,p)$$

第十二讲：常见的连续型分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

均匀分布

定义 2.17 任给参数 $a < b$, 函数

$$p(x) = \frac{1}{b-a}, \quad a < x < b$$

满足密度函数的两个性质即:$p(x) \geq 0$ 且 $\int_{-\infty}^{\infty} p(x)dx = 1$. 我们称以上式中的 $p(x)$ 为密度的连续型分布为区间 (a, b) 上的均匀分布，记作 $U(a, b)$.

- 易见，均匀分布的分布函数为

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x < b \\ 1. & x \geq b \end{cases}$$

- 如果 X 服从 $U(a, b)$ 分布，则对任何 $a \leq x < y \leq b$ 有

$$P(x < X \leq y) = \int_x^y \frac{1}{b-a} dt = \frac{y-x}{b-a}$$

GU5672972S2
DEUTSCHE BUNDESBANK
Banknote
10
ZEHN DEUTSCHE MARK
1777–1855 Carl Friedr. Gauß
GU5672972S2
Deutsche Bundesbank
Frankfurt am Main
1. September 1999
Zehn
Deutsche Mark

图 2: 高斯

- 卡尔。弗里德里希。高斯 1777 年出生于布伦瑞克;
- 11 岁时就发现了二项式定理;
- 19 岁发现正十七边形的尺规作图法，解决了困扰数学家们两千多年的难题;
- 1809 年提出“最小二乘法”并在此基础上建立的正态分布方程，是概率统计中一个非常重要的工具，广泛应用于数学、物理学等领域;
- 一生发表了 155 篇论文，对数论、代数学、非欧几何、复变函数和微分几何等领域都做出了开创性的贡献，被誉为数学王子.

正态分布 (高斯分布)

考虑下述函数

$$p_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

下面我们验证 $p_{\mu,\sigma}(x)$ 满足密度函数的两个性质即非负性与正则性：非负性显然；正则性则需要计算积分 $I := \int_{-\infty}^{\infty} p_{\mu,\sigma}(x)dx$.

由于 $p_{\mu,\sigma}(x)$ 的原函数不是初等函数，我们无法通过微积分基本公式计算 I. 我们考虑计算 I^2:

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdrd\theta \\
&= 1
\end{aligned}
$$

正态分布的定义

定义 2.18 若随机变量 X 的密度函数为

$$p_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

则称 X 服从参数为 (μ, σ^2) 的正态分布，记作 $X \sim N(\mu, \sigma^2)$. 特别的称 $N(0,1)$ 分布为标准正态分布，并简写 $p_{0,1}(x)$ 为 $\varphi(x)$. 正态分布又称高斯分布.

正态分布的性质

- 分布密度关于参数 μ 对称，即有

$$p_{\mu,\sigma}(\mu - x) = p_{\mu,\sigma}(\mu + x), \quad \forall x \in R$$

特别的，$N(0,1)$ 分布的密度函数为偶函数.

- $p_{\mu,\sigma}(x)$ 在 $x = \mu$ 处取得最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$. 注意到密度曲线下的面积应保持等于 1, 并且密度函数在 x 处的值反映了此分布取值 x 附近的概率大小。故 σ^2 越小，密度的曲线越尖陡，分布取值越集中；σ^2 越大，密度曲线越平缓，分布取值越分散.
- 正态分布的分布函数为

$$\Phi_{\mu,\sigma}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(t-\mu)^2}{2\sigma^2}} dt$$

特别的记标准正态分布函数为 $\Phi(x)$. 其与 $\Phi_{\mu,\sigma}(x)$ 有以下关系:

$$\Phi_{\mu,\sigma}(x) = \Phi(\frac{x-\mu}{\sigma})$$

例 2.16 (3σ 原则) 设随机变量 X 服从 $N(\mu,\sigma^2)$ 分布，求 $P(|X-\mu|<3\sigma)$.

解: 注意到

$$\begin{aligned} P(|X-\mu|<3\sigma) &= P(\mu-3\sigma<X<\mu+3\sigma) \\ &= \Phi_{\mu,\sigma}(\mu+3\sigma)-\Phi_{\mu,\sigma}(\mu-3\sigma) \\ &= \Phi(3)-\Phi(-3)=\Phi(3)-(1-\Phi(3)) \\ &= 2\Phi(3)-1=0.9973 \end{aligned}$$

这表明，尽管正态分布的取值范围是 $(-\infty,+\infty)$, 但是其取值落在区间 $(\mu-3\sigma,\mu+3\sigma)$ 的概率高达 99.73%

标准正态分布与一般正态分布之间的关系

例 2.17 若 $X \sim N(\mu, \sigma^2)$, 则 $Y := \dfrac{X-\mu}{\sigma} \sim N(0,1)$; 反之，若 $X \sim N(0,1)$, 则 $Y := \sigma X + \mu \sim N(\mu, \sigma^2)$.

解: 我们仅给出第一种情况的证明。事实上，我们只需要说明 Y 的分布密度为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

即可. 事实上，考虑 Y 的分布函数

$$\begin{aligned} F_Y(x) &= P(Y \le x) = P(\frac{X-\mu}{\sigma} \le x) = P(X \le \sigma x + \mu) \\ &= \int_{-\infty}^{\sigma x+\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \end{aligned}$$

从而 Y 的分布密度为

$$p_Y(x) = F_Y'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x)$$

正态分布函数的简单性质

记 $\Phi(x)$ 为标准正态分布的分布函数，$U \sim N(0,1)$, $\Phi_{\mu,\sigma}(x)$ 为 $N(\mu,\sigma^2)$ 分布的分布函数，$X \sim N(\mu,\sigma^2)$, 则

- $\Phi(0) = \dfrac{1}{2}$;
- $\Phi_{\mu,\sigma}(\mu) = \dfrac{1}{2}$;
- $\Phi(-x) = 1 - \Phi(x)$;
- $P(|U| < x) = P(-x < U < x) = P(U < x) - P(U \le -x)$
  $= \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1$;
- $\Phi_{\mu,\sigma}(x) = P(X \le x) = P(\dfrac{X-\mu}{\sigma} \le \dfrac{x-\mu}{\sigma}) = \Phi(\dfrac{x-\mu}{\sigma})$
- $P(a < X \le b) = \Phi(\dfrac{b-\mu}{\sigma}) - \Phi(\dfrac{a-\mu}{\sigma})$.

伽玛函数 (Gamma 函数)

定义 2.19 称以下函数

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}e^{-x}dx, \quad \alpha > 0,$$

为 Gamma 函数. Gamma 函数有以下性质:

- $\Gamma(1) = \int_0^\infty x^{1-1}e^{-x}dx = 1$;
- $\Gamma(\frac{1}{2}) = \int_0^\infty x^{\frac{1}{2}-1}e^{-x}dx = 2\sqrt{\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-x}d\sqrt{2}x^{\frac{1}{2}}$
  $= 2\sqrt{\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt = \sqrt{\pi}$;
- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, 故 $\Gamma(n+1) = n\Gamma(n) = \cdots = n!$; 事实上,

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^\infty x^\alpha e^{-x}dx = \int_0^\infty -x^\alpha de^{-x} \\ &= -x^\alpha e^{-x}|_0^\infty + \int_0^\infty \alpha x^{\alpha-1}e^{-x}dx = \alpha\Gamma(\alpha)\end{aligned}$$

伽玛分布

定义 2.20 (Gamma 分布) 若随机变量 X 的密度函数为

$$p(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称 X 服从参数为 (α, λ) 的 Gamma 分布，记作 $X \sim \Gamma(\alpha, \lambda)$.

指数分布与 χ^2(卡方) 分布

定义 2.21 称 $\Gamma(1,\lambda)$ 分布为指数分布，其密度函数与分布函数分别为

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

定义 2.22 称 $\Gamma(\frac{n}{2},\frac{1}{2})$ 分布为 χ^2 分布，记作 $\chi^2(n)$, 其密度函数为

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

泊松分布与伽玛分布

例 2.18 假定有一个于随机时刻陆续到来的质点流，我们若以 $N(t)$ 表示在 $[0,t]$ 时间内到来的质点的个数，并假设其服从参数为 λt 的泊松分布。试证明第 n 个质点到达的时间 S_n 服从 $\Gamma(n,\lambda)$ 分布.

证明: 首先我们考虑一下 S_n 的分布函数

$$\begin{aligned} P(S_n \le t) &= P(N(t) \ge n) = \sum_{k=n}^{\infty} P(N(t)=k) \\ &= \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

故其密度 $p(t)$ 为

$$\begin{aligned} p(t) = F'(t) &= \sum_{k=n}^{\infty} \left[\frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} - \frac{\lambda(\lambda t)^k}{k!} e^{-\lambda t} \right] \\ &= \frac{\lambda(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \end{aligned}$$

第十三讲：多维概率分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

December 12, 2024

n 维随机向量的定义

定义 2.23 (n 维随机向量或随机变量) 如果 $X_1, \cdots, X_n$ 是定义在同一个概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 个随机变量，即

$$\{\omega : X_i(\omega) \in B_i\} \in \mathcal{F}, \quad 任意的 B_i \in \mathcal{B},$$

则称 $X(\omega) := (X_1(\omega), \cdots, X_n(\omega))$ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 维随机向量或 n 维随机变量.

注 2.8

- 如果 $(X_1, \cdots, X_n)$ 是一个 n 维随机向量，则对任何不超过 n 的正整数 k 和 $1 \leqslant j_1 < \cdots < j_k \leqslant n$, $\left(X_{j_1}, \cdots, X_{j_k}\right)$ 都是一个 k 维随机向量;
- 特别地，当 $k = 1$ 时就是随机变量.

n 维随机向量的定义

定理 2.18 $X(\omega) = (X_1(\omega), \cdots, X_n(\omega))$ 是定义在某个概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 维随机向量当且仅当对任意的 $(x_1, \cdots, x_n) \in \mathbb{R}^n$ 均有

$$\{\omega : X_1(\omega) \le x_1, \cdots, X_n(\omega) \le x_n\} = \{X_1 \le x_1, \cdots, X_n \le x_n\} \in \mathcal{F},$$

或等价的有对任意的 $B \in \mathcal{B}(\mathbb{R}^n)$

$$\{\omega : X(\omega) \in B\} = \{\omega : (X_1(\omega), \cdots, X_n(\omega)) \in B\} \in \mathcal{F}.$$

多维分布及联合分布函数

定义 2.24 (多维分布) 设 $X=(X_1,\cdots,X_n)$ 是定义在某个概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 维随机向量，则对任意的 $B\in\mathcal{B}(\mathbb{R}^n)$, 称

$$\mathbf{F}(B):=P(X\in B)=P((X_1,\cdots,X_n)\in B)$$

为 n 维随机向量 $X=(X_1,\cdots,X_n)$ 的分布.

定义 2.25 (分布函数或联合分布函数) 设 $X=(X_1,\cdots,X_n)$ 是定义在某个概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 维随机向量，称

$$F(x_1,\cdots,x_n)=P(X_1\leqslant x_1,\cdots,X_n\leqslant x_n),\quad \forall(x_1,\cdots,x_n)\in\mathbb{R}^n$$

为随机向量 X 的分布函数，也称为随机变量 $X_1,\cdots,X_n$ 的联合分布函数.

注 2.9

- n 维随机向量的分布函数 $F(x_1,\cdots,x_n)$ 是定义在 $\mathbb{R}^n$ 上的 n 元函数;
- 称一个 n 元函数 $F(x_1,\cdots,x_n)$ 为 n 维分布函数，如果存在某个随机向量以它作为分布函数.

n 维随机向量分布函数的性质

定理 2.19 n 维随机向量 $X=(X_1,\cdots,X_n)$ 的分布函数 $F(x_1,\cdots,x_n)$ 具有下述性质:

1. $F(x_1,\cdots,x_n)$ 对每个变元非降;
2. $F(x_1,\cdots,x_n)$ 对每个变元右连续;
3. 对任意的 $1\le j\le n$,

$$\lim_{x_j\to-\infty}F\left(x_1,\cdots,x_j,\cdots,x_n\right)=0,$$

$$\lim_{x_1\to\infty,\cdots,x_n\to\infty}F(x_1,\cdots,x_n)=1$$

4. $F(x_1,\cdots,x_n)$ 具有增量非负性: 对任意的 $1\le j\le n$ 及 $a_j\leqslant b_j$ 均有

$$\Delta_{(a_1,\cdots,a_n)}^{(b_1,\cdots,b_n)}F=\sum_{\boldsymbol{x}=(x_1,\cdots,x_n)\in\mathcal{O}}\operatorname{sgn}(\boldsymbol{x})F(\boldsymbol{x})\geqslant 0,$$

其中

$$\mathcal{O}:=\{(x_1,\cdots,x_n):x_j=a_j\text{或}b_j,1\le j\le n\},$$

$$\operatorname{sgn}(\boldsymbol{x})=\begin{cases}1, & |\{j:x_j=a_j\}|\text{为偶数},\\ -1, & |\{j:x_j=a_j\}|\text{为奇数}.\end{cases}$$

n 维随机向量分布函数的性质

特别的，当 $n=2$ 或 3 时，$\Delta_{(a_1,\cdots,a_n)}^{(b_1,\cdots,b_n)}F$ 有以下具体形式:

$$\Delta_{(a_1,a_2)}^{(b_1,b_2)}F = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2)$$

$$\begin{aligned}\Delta_{(a_1,a_2,a_3)}^{(b_1,b_2,b_3)}F = {} & F(b_1,b_2,b_3) - F(a_1,b_2,b_3) - F(b_1,a_2,b_3) \\ & - F(b_1,b_2,a_3) + F(a_1,a_2,b_3) + F(a_1,b_2,a_3) \\ & + F(b_1,a_2,a_3) - F(a_1,a_2,a_3)\end{aligned}$$

二维 (联合) 分布函数

定义 2.26 设 (X, Y) 为 $(\Omega, \mathcal{F}, P)$ 上的随机向量，称

$$F(x, y) := P(X \leq x, Y \leq y)$$

为 (X, Y) 的二维 (联合) 分布函数.

二维 (联合) 分布函数的性质

定理 2.20 $F(x,y)$ 具有性质

1. $F(x,y)$ 对每个自变量是单调不减的;
2. $F(x,y)$ 对每个自变量都是右连续的;
3. $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 1 - \lim_{x\to+\infty,y\to+\infty} F(x,y) = 0$;
4. $F(x,y)$ 在任一矩形 $(a_1,b_1]\times(a_2,b_2]$ 上具有非负增量，即有

$$\Delta F = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) \geq 0.$$

二维离散型分布

定义 2.27 若随机向量 (X, Y) 有至多可列对可能值 $(x_i, y_i), i, j = 1, 2, \cdots$,, 则称随机向量及其联合分布是离散型的，而

$$p_{ij} := P(X = x_i, Y = y_j), \quad i, j = 1, 2, \cdots,$$

称为 (X, Y) 的联合分布列 (律). 我们常用下表来表示 (X, Y) 的分布列

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

二维离散型分布列的基本性质

- 非负性: $p_{ij} \geq 0, i,j = 1, 2, \cdots$;
- 正则性: $\sum_i \sum_j p_{ij} = 1$;
- 与一维情形类似，对任意的 $B \in \mathcal{B}^2$,

$$P((X, Y) \in B) = \sum_{i,j:(x_i,y_j)\in B} p_{ij}.$$

计算联合分布列的方法

- 确定随机变量 (X, Y) 的所有取值数对;
- 计算取每个数值对的概率;
- 列出表格

例 2.19 从 $1, 2, 3, 4$ 中任取一数记为 X, 再从 $1, \cdots, X$ 中任取一数记作 Y, 求 (X, Y) 的联合分布列及 $P(X = Y)$.

二维连续型分布

定义 2.28 若存在二元非负可积函数 $p(x,y)$ 使得 (X,Y) 的二维 (联合) 分布函数

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv, \forall (x,y)\in\mathbb{R}^2$$

则称 (X,Y) 及其概率分布为连续型的，称 $p(x,y)$ 为其联合分布密度.

易知，联合分布密度具有以下两个基本性质:

1. 非负性: $p(x,y)\geq 0$;
2. 正则性: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)dxdy=1$.

联合分布密度与概率分布之间的关系

- $p(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$
- 对于连续型随机变量，对任意的 $B \in \mathcal{B}^2$ 有

$$P((X,Y) \in B) = \iint_B p(x,y)dxdy$$

例 2.20 设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases} Ae^{-2x-3y}, & x>0,y>0 \\ 0, & \text{其他} \end{cases}$$

试确定常数 A, 并求概率 $P(X<2,Y<1)$ 及 $P(2X+3Y<6)$.

解:

- 由 $\iint p(x,y)dxdy=1$ 可得 $A=6$;
- $P(X<2,Y<1)=\int_{-\infty}^{2}\int_{-\infty}^{1}p(x,y)dxdy=\int_{0}^{2}\int_{0}^{1}p(x,y)dxdy=(1-e^{-4})(1-e^{-3})$;

注意到如果我们令 $D := \{(x,y) : 2x+3y<6\}$, 则

$$
\begin{aligned}
P(2X+3Y<6) &= P((X,Y)\in D) = \iint_D p(x,y)dxdy \\
&= \int_0^3 \int_0^{\frac{1}{3}(6-2x)} 6e^{-(2x+3y)}dxdy \\
&= 6\int_0^3 e^{-2x}(-\frac{1}{3}e^{-3y})|_0^{\frac{1}{3}(6-2x)}dx = 1-7e^{-6}
\end{aligned}
$$

边缘分布：随机向量的分量各自的概率分布

定义 2.29 (边缘分布) 在 (X,Y) 的联合分布函数中，令 $y\uparrow+\infty$ 可得

$$\begin{aligned}\lim_{y\to+\infty}F(x,y) &= \lim_{y\to+\infty}P(X\le x,Y\le y)\\ &= P(X\le x,Y\le+\infty)=F(x,+\infty)\\ &= P(X\le x):=F_1(x).\end{aligned}$$

称 $F_1(x)$ 为 X 的边缘分布. 类似的，Y 的边缘分布为

$$F_2(y)=P(Y\le y)=\lim_{x\to+\infty}F(x,y)=F(+\infty,y)$$

从边缘分布的定义可知，若 (X,Y) 的联合分布为 $F(x,y)$, 则

$$\begin{aligned}X&\sim F_1(x)=F(x,+\infty)\\ Y&\sim F_2(y)=F(+\infty,y)\end{aligned}$$

边缘分布的离散情形

- 边缘一词源于离散型情形。在二维离散型概率分布 $\{p_{ij}\}$ 的列表中，将各行求和写在表的最右一列，再将各列求和写在表的最下一行.
- 若 (X, Y) 的联合分布列为 $\{p_{ij}\}$, 则

$$X\text{的分布列为}:P(X=x_i)=\sum_j P(X=x_i, Y=y_j)=\sum_j p_{ij}=p_{i\cdot};$$

$$Y\text{的分布列为}:P(Y=y_j)=\sum_i P(X=x_i, Y=y_j)=\sum_i p_{ij}=p_{\cdot j}$$

i,j	1	2	$\cdots$	j	$\cdots$	$p_{i\cdot}$
1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	

边缘分布的连续情形

- 注意到

$$
\begin{aligned}
F_1(x) &= P(X \le x) = F(x, +\infty) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} p(u,v)dudv \\
&= \int_{-\infty}^{x} \underbrace{\left[\int_{-\infty}^{+\infty} p(u,v)dv\right]}_{p_1(u)} du = \int_{-\infty}^{x} p_1(u)du
\end{aligned}
$$

类似的有，

$$
\begin{aligned}
F_2(y) &= P(Y \le y) = F(+\infty, y) = \int_{-\infty}^{y} \int_{-\infty}^{+\infty} p(u,v)dudv \\
&= \int_{-\infty}^{x} \underbrace{\left[\int_{-\infty}^{+\infty} p(u,v)du\right]}_{p_2(v)} dv = \int_{-\infty}^{y} p_2(v)dv
\end{aligned}
$$

- 对于连续型随机向量，其分量仍为连续型，相应的边缘密度分别为

$$
p_1(x) = \int_{-\infty}^{+\infty} p(x,y)dy, \quad p_2(y) = \int_{-\infty}^{+\infty} p(x,y)dx
$$

常见的多维离散型分布：多项分布

- 多项分布：进行 n 次独立重复试验，每次试验有 r 个互不相容的结果:$A_1, \cdots, A_r$ 之一发生；
- 每次试验中 A_i 发生的概率为 $p_i = P(A_i), i = 1, \cdots, r$ 且 $p_1 + \cdots + p_r = 1$;
- 记 X_i 为 n 次独立试验中 A_i 出现的次数，则 $(X_1, X_2, \cdots, X_r)$ 取值 $(n_1, n_2, \cdots, n_r)$ 的概率，即 $A_i, i = 1, \cdots, r$ 出现 n_i 次的概率为

$$P(X_1 = n_1, \cdots, X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

  其中 $n = n_1 + \cdots + n_r$;
- 这个联合分布列称为 r 项分布，又称多项分布，记作 $M(n, p_1, \cdots, p_r)$.
- 上述概率是多项式 $(p_1 + p_2 + \cdots + p_r)^n$ 的一项，故其和为 1.
- $r = 2$ 时即为二项分布.

常见的多维离散型分布：多维超几何分布

- 多维超几何分布：袋中有 N 个球，其中有 N_i 个 i 号球，$i = 1, 2, \cdots, r$, 且 $N_1 + N_2 + \cdots + N_r = N$;
- 从中任取 n 个球，若记 X_i 为取出的 n 个球中 i 号球的个数，$i = 1, 2, \cdots, r$, 则

$$P(X_1 = n_1, X_2 = n_2, \cdots, X_r = n_r) = \frac{C_{N_1}^{n_1} C_{N_2}^{n_2} \cdots C_{N_r}^{n_r}}{C_N^n}$$

  其中 $n_1 + n_2 + \cdots + n_r = n$;
- $r = 2$ 时即为超几何分布.

多维均匀分布

定义 2.30 设 D 为 $\mathbb{R}^n$ 中的一个有界区域，其度量为 S_D, 如果多维随机变量 $(X_1, \cdots, X_n)$ 的联合密度函数为

$$p(x_1, \cdots, x_n) = \begin{cases} \dfrac{1}{S_D}, & (x_1, x_2, \cdots, x_n) \in D \\ 0, & \text{其他} \end{cases}$$

则称 $(X_1, \cdots, X_n)$ 服从 D 上的多维均匀分布，记作 $(X_1, \cdots, X_n) \sim U(D)$.

二维均匀分布所描述的随机现象就是向平面区域 D 中随机投点，如果该点的坐标 (X, Y) 落在 D 的子区域 G 中的概率只与 G 的面积有关，而与 G 的位置无关，则

$$P((X, Y) \in G) = \iint_G p(x, y) dxdy = \iint_G \frac{1}{S_D} dxdy = \frac{S_G}{S_D}$$

例 2.21 设 (X,Y) 服从单位圆 $D=\{(x,y):x^2+y^2<1\}$ 上的均匀分布，试求其边缘密度函数 (非均匀分布).

(X,Y) 的联合密度为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2<1\\ 0, & \text{其他}\end{cases}$$

故当 $|x|\geq 1$, $p(x,y)=0$, 从而 $p_1(x)=0$; 当 $|x|<1$ 时，

$$\begin{aligned}p_1(x) &= \int_{-\infty}^{+\infty}p(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy\\ &= \frac{2}{\pi}\sqrt{1-x^2}\end{aligned}$$

类似可得 Y 的密度为

$$p_2(y)=\frac{2}{\pi}\sqrt{1-y^2}$$

二维正态分布

定义 2.31 设参数 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 满足 $\sigma_1 > 0, \sigma_2 > 0$ 及 $-1 < \rho < 1$, 称以

$$
\begin{aligned}
p(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\
&\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}
\end{aligned}
$$

为密度函数的连续型分布为二维正态分布，记作

$$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho).$$

后面我们会给出

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)dxdy = 1$$

的证明.

二维正态分布密度图像

0.0018
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
-20
-10
0
10
20
y
20
10
0
-10
-20
x

二维正态分布的边缘分布

定理 2.21 若 (X,Y) 为二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

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

证明: X 的边缘分布密度为

$$\begin{aligned}
p_1(x) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
&\quad \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}dy \\
&= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right\}dv \\
&= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left[\left(\frac{v-\rho u}{\sqrt{1-\rho^2}}\right)^2 + u^2\right]\right\}dv \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}
\end{aligned}$$

类似的可得 $Y \sim N(\mu_2,\sigma_2^2)$.

第十四讲：条件分布与随机变量的独立性

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

条件分布

- $(\Omega, \mathcal{F}, P)$ 为一概率空间，任意固定 $B \in \mathcal{F}$ 且 $P(B) > 0$, $P(\cdot|B)$ 仍为 $(\Omega, \mathcal{F})$ 上的概率;
- 对于 $(\Omega, \mathcal{F})$ 上的随机变量 X, 考虑其关于条件概率 $P(\cdot|B)$ 的分布即在概率空间 $(\Omega, \mathcal{F}, P(\cdot|B))$ 上考虑 X 的分布:

$$F(x|B) := P(X \leq x|B) = \frac{P(X \leq x, B)}{P(B)}, x \in R$$

- 现假定有另一随机变量 Y, 则当 $P(Y = y) > 0$ 时，$F(x|Y = y)$ 则为已知 $Y = y$ 时 X 的条件分布函数.

离散型条件分布

定义 2.32 若离散型随机向量 (X, Y) 的联合分布为 $\{p_{ij}\}$, 则当

$$P(Y = y_j) = p_{\cdot j} > 0$$

时，称

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

为已知 $Y = y_j$ 时 X 的条件分布. 类似的，当 $P(X = x_i) = p_{i\cdot} > 0$ 时，称

$$p_{j|i} := P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i\cdot}}$$

为已知 $X = x_i$ 时 Y 的条件分布.

离散型条件分布函数

根据上面的定义，给定 $Y=y_j$ 条件下 X 的条件分布函数为

$$F(x|y_j)=\sum_{i:x_i\le x}P(X=x_i|Y=y_j)=\sum_{i:x_i\le x}p_{i|j}$$

类似的给定 $X=x_i$ 条件下 Y 的条件分布函数为

$$F(y|x_i)=\sum_{j:y_j\le y}P(Y=y_j|X=x_i)=\sum_{j:y_j\le y}p_{j|i}$$

连续型条件分布

- 设二维连续型随机变量 (X,Y) 的联合密度函数为 $p(x,y)$, 边际密度函数分别为 $p_X(x), p_Y(y)$;
- 条件分布函数 $P(X \le x|Y=y)$: 因为 $P(Y=y)=0$, 故无法通过条件概率直接计算;
- $P(X \le x|Y=y) := \lim_{h\to 0} P(X \le x|Y \in [y, y+h])$;
- 根据如上定义:

$$
\begin{aligned}
F(x|y) \quad := \quad P(X \le x|Y=y) &= \lim_{h\to 0} \frac{P(X \le x, Y \in [y, y+h])}{P(Y \in [y, y+h])} \\
&= \lim_{h\to 0} \frac{\int_{-\infty}^{x} \int_{y}^{y+h} p(u,v)dvdu}{\int_{y}^{y+h} p_Y(v)dv} = \lim_{h\to 0} \frac{\int_{-\infty}^{x} \{\frac{1}{h}\int_{y}^{y+h} p(u,v)dv\}du}{\frac{1}{h}\int_{y}^{y+h} p_Y(v)dv} \\
&\xlongequal[\text{积分中值定理}]{p(x,y),p_Y(y)\text{在}y\text{连续}} \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)}du
\end{aligned}
$$

- 这表明给定 $Y=y$ 时 X 的条件分布仍为连续型，其密度函数为

$$p(x|y) = p(x,y)/p_Y(y).$$

连续随机变量的条件分布函数及条件密度函数

定义 2.33 对一切使 $p_Y(y) > 0$ 的 y, 给定 $Y = y$ 条件下 X 的条件分布函数和条件密度函数分别为

$$
\begin{aligned}
F(x|y) &= \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)} du, \\
p(x|y) &= \frac{p(x,y)}{p_Y(y)}.
\end{aligned}
$$

同理，对一切使 $p_X(x) > 0$ 的 x, 给定 $X = x$ 条件下 X 的条件分布函数和条件密度函数分别为

$$
\begin{aligned}
F(y|x) &= \int_{-\infty}^{y} \frac{p(x,v)}{p_X(x)} dv, \\
p(y|x) &= \frac{p(x,y)}{p_X(x)}.
\end{aligned}
$$

连续场合的全概率公式和贝叶斯公式

- 由条件密度函数的定义可得

$$p(x,y) = p_X(x)p(y|x), \quad p(x,y) = p_Y(y)p(x|y)$$

- 对 $p(x,y)$ 求边际密度函数可得

$$\begin{aligned}
p_X(x) &= \int_{-\infty}^{+\infty} p_Y(y)p(x|y)dy \\
P(X \le x) &= \int_{-\infty}^{x} \left[\int_{-\infty}^{+\infty} p_Y(y)p(u|y)dy \right] du \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{x} p(u|y)du \right] p_Y(y)dy \\
&= \int_{-\infty}^{+\infty} P(X \le x|Y = y)dP(Y \le y)
\end{aligned}$$

连续场合的全概率公式和贝叶斯公式

- 类似推导可得

$$
\begin{aligned}
p_Y(y) &= \int_{-\infty}^{+\infty} p_X(x)p(y|x)dx \\
P(Y \le y) &= \int_{-\infty}^{+\infty} P(Y \le y|X = x)dP(X \le x)
\end{aligned}
$$

- 贝叶斯公式的密度函数形式:

$$
\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)p(y|x)}{\int_{-\infty}^{+\infty} p_X(x)p(y|x)dx} \\
p(y|x) &= \frac{p(x,y)}{p_X(x)} = \frac{p_Y(y)p(x|y)}{\int_{-\infty}^{+\infty} p_Y(y)p(x|y)dy}
\end{aligned}
$$

例 2.22 若 (X,Y) 为二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 试求给定 $Y=y$ 时 X 的条件分布密度.

解: X 的条件分布密度为

$$
\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p_Y(y)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}} \\
&= \frac{\exp\left\{\frac{-1}{2\sigma_1^2(1-\rho^2)}\left[(x-\mu_1)^2-2\rho\frac{\sigma_1}{\sigma_2}(x-\mu_1)(y-\mu_2)+\rho^2\frac{\sigma_1^2}{\sigma_2^2}(y-\mu_2)^2\right]\right\}}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right]^2\right\} \\
&\sim N(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2)) \overset{\rho=0}{=\!=\!=} N(\mu_1,\sigma_1^2)
\end{aligned}
$$

类似可求得 Y 关于 $X=x$ 的条件分布密度.

例 2.23 设 (X,Y) 服从单位圆 $D=\{(x,y):x^2+y^2<1\}$ 上的均匀分布，试求条件分布密度函数 $p(y|x)$.

解: (X,Y) 的联合密度及 X 的边缘密度分别为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2<1,\\ 0, & \text{其他}\end{cases},\quad p_X(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2}, & |x|<1,\\ 0, & \text{其他}.\end{cases}$$

故

$$p(y|x)=\frac{p(x,y)}{p_X(x)}=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}}, & |y|\le\sqrt{1-x^2}\\ 0, & \text{其他}\end{cases}$$

随机变量的独立性

- 首先我们回顾一下事件的独立性:

$$P(AB) = P(A)P(B) \Leftrightarrow P(A|B) = P(A), \text{当} P(B) > 0;$$

- 二维离散型随机变量

$$p_{i|j} := \frac{p_{ij}}{p_{\cdot j}} = P(X = x_i | Y = y_j) \xlongequal[\text{不影响概率}]{\text{若条件事件}} P(X = x_i) = p_{i\cdot};$$

- 故如果对任意的事件 $\{X = x_i\}, \{Y = y_j\}, i,j = 1, 2, \cdots$, 均独立，我们则可称随机变量 X, Y 独立，即 X, Y 独立等价于

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

- 此时的联合分布函数 $F(x, y)$ 为

$$\begin{aligned} F(x,y) &= P(X \le x, Y \le y) = \sum_{x_i \le x} \sum_{y_j \le y} p_{ij} = \sum_{x_i \le x} p_{i\cdot} \sum_{y_j \le y} p_{\cdot j} \\ &= \sum_{x_i \le x} P(X = x_i) \sum_{y_j \le y} P(Y = y_j) = F_X(x) F_Y(y) \end{aligned}$$

随机变量的独立性

- 二维连续型随机变量

$$
\begin{aligned}
F(x|y) &= \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)} du = P(X \le x | Y = y) \xlongequal[\text{不影响概率}]{\text{若条件事件}} P(X \le x) \\
&= \int_{-\infty}^{x} p_X(u) du \\
&\Rightarrow \quad p(x,y) = p_X(x) p_Y(y)
\end{aligned}
$$

- 此时，其分布函数

$$
\begin{aligned}
F(x,y) &= P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(u,v) du dv \\
&= \int_{-\infty}^{x} p_X(u) du \int_{-\infty}^{y} p_Y(v) dv = F_X(x) F_Y(y)
\end{aligned}
$$

随机变量独立性的定义

定义 2.34 设 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的联合分布函数为 $F(x_1, \cdots, x_n)$, $F_i(x_i)$ 为 X_i 的边缘分布函数。如果对任意的 n 个实数 $x_1, x_2, \cdots, x_n$, 有

$$F(x_1, x_2, \cdots, x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n)$$

则称 $X_1, X_2, \cdots, X_n$ 相互独立。特别的,

1. 在离散随机变量场合，如果对任意 n 个取值 $x_1, x_2, \cdots, x_n$, 有

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_n = x_n),$$

   则称 $X_1, X_2, \cdots, X_n$ 相互独立.

2. 在连续随机变量场合，如果对任意的 n 个取值 $x_1, x_2, \cdots, x_n$, 有

$$p(x_1, x_2, \cdots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n)$$

   则称 $X_1, X_2, \cdots, X_n$ 相互独立.

第十五讲：随机变量函数的分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

一维离散型随机变量函数的分布

定理 2.22 设 $X \sim \{p_k\}$ 为一维离散型随机变量，若 $Z = f(X)$, 则对随机变量 Z 的所有可能值 z_k 有

$$P(Z = z_k) = \sum_{i:f(x_i)=z_k} p_i$$

证明: 注意到 $\{Z = z_k\} = \{f(X) = z_k\} = \cup_{i:f(x_i)=z_k}\{X = x_i\}$

故 $P(Z = z_k) = \sum_{i:f(x_i)=z_k} P(X = x_i) = \sum_{i:f(x_i)=z_k} p_i$.

例 2.24 若 $P(X = 1) = P(X = -1) = 0.25, P(X = 0) = 0.5$, 求 $Z = X^2$ 的分布.

解: $P(Z = 1) = \sum_{i:x_i^2=1} p_i = P(X = 1) + P(X = -1) = 0.5$

$$P(Z = 0) = P(X = 0) = 0.5$$

$Y=f(X)$: $f(x)$ 严格单调且 X 为连续型随机变量

定理 2.23 设 X 是连续随机变量，其密度函数为 $p_X(x)$. $Y=f(X)$ 是另一个随机变量。若 $y=f(x)$ 严格单调，其反函数 $h(y)$ 有连续导数，则 $Y=f(X)$ 的密度函数为

$$p_Y(y)=\begin{cases} p_X(h(y))|h'(y)|, & a\le y\le b \\ 0, & \text{其他} \end{cases}$$

其中 $a=\min\{f(-\infty),f(+\infty)\}, b=\max\{f(-\infty),f(+\infty)\}$.

证明: 由 $f(x)$ 为严格单调函数可知:

$$\begin{aligned} F_Y(y) &= P(Y\le y)=P(f(X)\le y) \\ &= \begin{cases} 0, & y<a \\ 1, & y>b \\ P(X\le h(y))=\int_{-\infty}^{h(y)} p_X(x)dx, & a\le y\le b\text{且}f(x)\text{严格递增} \\ P(X\ge h(y))=\int_{h(y)}^{+\infty} p_X(x)dx, & a\le y\le b\text{且}f(x)\text{严格递减} \end{cases} \end{aligned}$$

几个常见的函数形式 f

- $f(x) = ax + b$：当 X 为均匀分布，正态分布，伽玛分布时，求 $Y = f(X)$ 的分布密度;
- 当 $X \sim N(\mu, \sigma^2)$ 时，可以验证 $aX + b \sim N(a\mu + b, a^2\sigma^2)$;
- $f(x) = e^x$：当 X 为均匀分布，正态分布，伽玛分布时，求 $Y = f(X)$ 的分布密度;
- 对无单调性的函数 $f(x)$, 我们一般根据 X 的分布密度想法求出 Y 的分布函数 $F_Y(y)$,

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) = P(f(X) \le y) = P(X \in f^{-1}((-\infty, y])) \\
&= \int_{f^{-1}((-\infty,y])} p_X(x)dx
\end{aligned}
$$

- 然后对 $F_Y(y)$ 关于 y 求导即可得到其分布密度 $p_Y(y)$.

多维随机变量函数的分布：离散情形

定理 2.24 (离散卷积公式) 设 X 与 Y 为相互独立的非负整值随机变量，各有分布 $\{a_k\}$ 与 $\{b_k\}$. 则其和有分布

$$P(X+Y=n)=\sum_{k=0}^{n} a_k b_{n-k}, n=0,1,2,\cdots .$$

证明: 注意到 X 与 Y 相互独立，故对任何的 $n=0,1,2,\cdots$，有

$$\begin{aligned}
P(X+Y=n) &= \sum_{k=0}^{\infty} P(X+Y=n, X=k) = \sum_{k=0}^{n} P(X+Y=n, X=k) \\
&= \sum_{k=0}^{n} P(X=k)P(X+Y=n|X=k) = \sum_{k=0}^{n} P(X=k)P(Y=n-k|X=k) \\
&= \sum_{k=0}^{n} P(X=k)P(Y=n-k) = \sum_{k=0}^{n} a_k b_{n-k}
\end{aligned}$$

泊松分布的可加性

例 2.25 (泊松分布的可加性) 设随机变量 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$, 且 X, Y 相互独立，则 $Z = X + Y \sim P(\lambda_1 + \lambda_2)$

解: 由题意知

$$
\begin{aligned}
P(Z=k) &= P(X+Y=k) = \sum_{i=0}^{\infty} P(X=i, X+Y=k) \\
&= \sum_{i=0}^{\infty} P(X=i, Y=k-i) = \sum_{i=0}^{k} P(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} P(X=i)P(Y=k-i) = \sum_{i=0}^{n} \left(\frac{\lambda_1^i}{i!} e^{-\lambda_1}\right)\left(\frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2}\right) \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} = \frac{(\lambda_1+\lambda_2)^k}{k!} e^{-(\lambda_1+\lambda_2)}
\end{aligned}
$$

$k = 0, 1, \cdots,$

泊松分布的可加性

- 泊松分布的上述性质可以叙述为：泊松分布的卷积仍是泊松分布，并记为

$$P(\lambda_1) * P(\lambda_2) = P(\lambda_1 + \lambda_2);$$

- 这里的卷积是指" 寻求两具独立随机变量和的分布运算";
- 上述泊松的性质可推广至有限个独立泊松随机变量之和的分布上去，即:

$$P(\lambda_1) * P(\lambda_2) * \cdots * P(\lambda_n) = P(\lambda_1 + \lambda_2 + \cdots + \lambda_n);$$

- 特别的，当 $\lambda_1 = \lambda_2 = \cdots = \lambda_n$ 时有

$$P(\lambda) * P(\lambda) * \cdots * P(\lambda) = P(n\lambda);$$

二项分布的可加性

例 2.26 (二项分布的可加性) 设随机变量 $X \sim B(n,p), Y \sim B(m,p)$, 且 X, Y 独立，则 $X+Y \sim B(n+m,p)$. 特别的，若

$$X_i \sim B(1,p), i=1,2,\cdots,n, \text{则} X_1+X_2+\cdots+X_n \sim B(n,p).$$

多维随机变量函数的分布：连续情形的一般方法

若 $Z = f(X_1, X_2, \cdots, X_n)$, $(X_1, X_2, \cdots, X_n)$ 的密度函数为 $p(x_1, x_2, \cdots, x_n)$, 则 Y 的分布函数为

$$G(z) = P(Z \le z) = \int \cdots \int_{f(x_1, x_2, \cdots, x_n) \le z} p(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$

二维随机变量和的分布 (卷积公式): $Z = X + Y$

定理 2.25 若 $Z = X + Y$, 而 (X, Y) 的联合密度函数为 $p(x, y)$, 则

$$\begin{aligned} F(z) &= P(Z \le z) = \iint_{x+y\le z} p(x,y)dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} p(x,y)dydx \end{aligned}$$

特别的，若 X, Y 相互独立时有 $p(x, y) = p_1(x)p_2(y)$, 将其代入上式得

$$\begin{aligned} F(z) &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} p_1(x)p_2(y)dy \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z} p_1(x)p_2(u-x)du \\ &= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} p_1(x)p_2(u-x)dx\right] du \end{aligned}$$

故其密度函数为 $p(z) = \int_{-\infty}^{+\infty} p_1(x)p_2(z-x)dx = \int_{-\infty}^{+\infty} p_1(z-y)p_2(y)dy$

正态分布的可加性

例 2.27 设随机变量 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且 X, Y 独立，则 $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

证明: 显然 $Z = X+Y$ 仍在 $(-\infty, +\infty)$ 上取值。据卷积公式可得

$$
\begin{aligned}
p_z(z) &= \int_{-\infty}^{+\infty} p_1(z-y)p_2(y)dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[\frac{(z-y-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}dy \\
&\xlongequal[B=\frac{(z-\mu_1)}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}]{A=\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[A(y-\frac{B}{A})^2+\frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right]\right\}dy \\
&= \frac{\sqrt{2\pi}\sqrt{1/A}}{2\pi\sigma_1\sigma_2}e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sqrt{1/A}}\exp\left\{-\frac{(y-\frac{B}{A})^2}{2(\sqrt{1/A})^2}\right\}dy \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2\sqrt{A}}e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}
\end{aligned}
$$

正态分布的可加性

- 若 $X \sim N(\mu, \sigma^2)$, 则对任意非零常数 a 有，$aX \sim N(a\mu, a^2\sigma^2)$;
- 若 $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$ 为 n 个独立的正态随机变量，则其线性组合仍服从正态分布，即

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n \sim N(\mu_0, \sigma_0^2)$$

其中 $\mu_0 = \sum_{i=1}^n a_i\mu_i, \sigma_0^2 = \sum_{i=1}^n a_i^2\sigma_i^2$.

Gamma 分布的可加性

例 2.28 若 $X \sim \Gamma(\alpha_1, \lambda), Y \sim \Gamma(\alpha_2, \lambda)$, 且 X, Y 独立，则

$$Z = X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda).$$

例 2.29 若 $X_i \sim \exp(\lambda) = \Gamma(1, \lambda), i = 1, 2, \cdots, n$, 且相互独立独立，则

$$X_1 + X_2 + \cdots + X_n \sim \Gamma(n, \lambda).$$

例 2.30 若 $X_i \sim \chi^2(n_i) = \Gamma(\frac{n_i}{2}, \frac{1}{2}), i = 1, 2, \cdots, m$, 且相互独立独立，则

$$\begin{aligned} X_1 + X_2 + \cdots + X_m &\sim \Gamma(\frac{n_1 + n_2 + \cdots + n_m}{2}, \frac{1}{2}) \\ &= \chi^2(n_1 + n_2 + \cdots + n_m) \end{aligned}$$

二维随机变量商的分布: $Z = X/Y$

定理 2.26 (商的密度) 如果随机变量 (X, Y) 的联合密度函数为 $p(x, y)$, 则它们的商 $Z := X/Y$ 仍为连续型，且其密度为

$$p(z) = \int_{-\infty}^{+\infty} |y| p(zy, y) dy.$$

证明: 商的分布函数为

$$\begin{aligned}
F(z) &= P(X/Y \le z) = \iint_{x/y \le z} p(x, y) dx dy \\
&= \iint_{x/y \le z, y > 0} p(x, y) dx dy + \iint_{x/y \le z, y < 0} p(x, y) dx dy \\
&= \int_0^{+\infty} dy \int_{-\infty}^{yz} p(x, y) dx + \int_{-\infty}^{0} dy \int_{yz}^{+\infty} p(x, y) dx \\
&= \int_0^{+\infty} dy \int_{-\infty}^{z} y p(yu, y) du + \int_{-\infty}^{0} dy \int_{z}^{-\infty} y p(yu, y) du \\
&= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} |y| p(yu, y) dy \right] du
\end{aligned}$$

多维独立随机变量的最大 (小) 值分布

定理 2.27 若 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量，其各自的分布为 $F_k(x), k = 1, 2, \cdots, n$, 若记

$$\overline{X} := \max_{k=1}^{n} X_k, \quad \underline{X} = \min_{k=1}^{n} X_k$$

则

$$P(\overline{X} \le x) = \Pi_{k=1}^{n} F_k(x), \quad P(\underline{X} \le x) = 1 - \Pi_{k=1}^{n}(1 - F_k(x))$$

证明: $\overline{X}, \underline{X}$ 的分布函数为

$$\begin{aligned}
F_{\overline{X}}(x) &= P(\overline{X} \le x) = P(X_1 \le x, X_2 \le x, \cdots, X_n \le x) \\
&= \Pi_{k=1}^{n} P(X_k \le x) = \Pi_{k=1}^{n} F_k(x) \\
F_{\underline{X}}(x) &= 1 - P(\underline{X} > x) = 1 - P(X_1 > x, \cdots, X_n > x) \\
&= 1 - \Pi_{k=1}^{n} P(X_k > x) = 1 - \Pi_{k=1}^{n}(1 - F_k(x))
\end{aligned}$$

独立同分布情形下的最大最小值分布

推论 2.1 若上述定理中的 F_k 均相同为 $F(x)$, 并且假设 $F(x)$ 具有分布密度 $p(x)$, 则 $\overline{X}, \underline{X}$ 分布密度分别为

$$\begin{aligned} p_{\overline{X}}(x) &= n[F(x)]^{n-1}p(x), \\ p_{\underline{X}}(x) &= n[1-F(x)]^{n-1}p(x) \end{aligned}$$

思考: 计算 $(\overline{X}, \underline{X})$ 的联合分布函数.

随机向量函数的联合分布：一般情形

求连续型随机向量的函数的分布最一般的提法是:
设已知 $(X_1, X_2, \cdots, X_n)$ 的联合密度函数 $p(x_1, x_2, \cdots, x_n)$, 而

$$\begin{cases} Y_1 = f_1(X_1, X_2, \cdots, X_n) \\ Y_2 = f_2(X_1, X_2, \cdots, X_n) \\ \cdots \\ Y_m = f_m(X_1, X_2, \cdots, X_n) \end{cases} \tag{26}$$

求随机变量 $(Y_1, Y_2, \cdots, Y_m)$ 的分布.

- 保证 $(Y_1, Y_2, \cdots, Y_n)$ 是随机向量: 只需 $f_i, i = 1, 2, \cdots, m$ 是 Borel 函数即可;
- 我们需要利用 $(X_1, \cdots, X_n)$ 的分布密度计算 $(Y_1, \cdots, Y_n)$, 故我们需要从 (26) 中反解出 $(X_1, \cdots, X_n)$ 即要求方程组 (26) 有解:
  - 若 $m > n$, 既使对线性函数 f_i 也不能保证方程组有解;
  - 若 $m < n$, 我们可以增补 $Y_j = X_j, j = m+1, \cdots, n$ 化为 $m = n$ 的情形;
  - 故我们只考虑 $m = n$ 的情形.

随机向量函数的联合分布：一般情形

定理 2.28 设随机向量 $(X_1, X_2, \cdots, X_n)$ 的联合密度函数 $p(x_1, x_2, \cdots, x_n)$，而 $Y_k = f_k(X_1, \cdots, X_n), k = 1, 2, \cdots, n$. 若假设每个 f_k 均为 n 维 Borel 函数，并且对 $(Y_1, \cdots, Y_n)$ 的每一组可能值 $(y_1, \cdots, y_n)$，方程组

$$\begin{cases} y_1 = f_1(x_1, \cdots, x_n) \\ \cdots \\ y_n = f_n(x_1, \cdots, x_n) \end{cases} \tag{27}$$

有唯一解

$$\begin{cases} x_1 = h_1(y_1, \cdots, y_n) \\ \cdots \\ x_n = h_n(y_1, \cdots, y_n) \end{cases} \tag{28}$$

并且每个 $h_k(y_1, \cdots, y_n)$ 有连续的一阶偏导数。那么 $(Y_1, \cdots, Y_n)$ 是连续型随机向量，并且有联合密度函数

$$q(y_1, \cdots, y_n) = \begin{cases} p\big(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\big)|J| \\ 0, \text{若}(y_1, \cdots, y_n)\text{使}(27)\text{无解} \end{cases}$$

其中 J 为变换 (28) 的 Jacobi 行列式.

随机向量函数的联合分布：一般情形

证明: $(Y_1, \cdots, Y_n)$ 的联合分布函数为

$$
\begin{aligned}
F(u_1, \cdots, u_n) &= P(Y_1 \le u_1, \cdots, Y_n \le u_n) \\
&= \int_{f_1(x_1,\cdots,x_n)\le u_1} \cdots \int_{f_n(x_1,\cdots,x_n)\le u_n} p(x_1, \cdots, x_n) dx_1 \cdots dx_n \\
&= \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_n} p\big(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\big) |J| dy_1 \cdots dy_n
\end{aligned}
$$

故 $(Y_1, \cdots, Y_n)$ 是连续型随机向量，并且有联合密度函数

$$
q(y_1, \cdots, y_n) = \begin{cases} p\big(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\big) |J| \\ 0, 若(y_1, \cdots, y_n)使(27)无解 \end{cases}
$$

注: 若 Jacobi 行列式 J 不易计算时，可计算 J^{-1} 即变换 (27) 的 Jacobi 行列式.

随机变量积的分布

例 2.31 设随机向量 (X,Y) 的联合密度函数为 $p(x,y)$. 则 $U=XY$ 的密度函数为

$$p_U(u)=\int_{-\infty}^{+\infty}p(\frac{u}{v},v)\frac{1}{|v|}dv$$

证明: 增补变量 $V=Y$, 则 $\begin{cases}u=xy,\\ v=y,\end{cases}$ 的反函数为 $\begin{cases}x=\dfrac{u}{v},\\ y=v,\end{cases}$. 其 Jacobi 行列式为

$$J=\begin{vmatrix}\dfrac{1}{v} & -\dfrac{u}{v^2}\\ 0 & 1\end{vmatrix}=\frac{1}{v},$$

故 (U,V) 的联合密度函数为 $q(u,v)=p(\frac{u}{v},v)\frac{1}{|v|}$. 对 $q(u,v)$ 关于 v 积分便可得 $U=XY$ 的密度函数. (随机变量商的密度也可类似求出)

第十六讲：数学期望

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

平均值与加权平均值

有甲乙两名射手，其射击技术可用下表表出

甲射手

击中环数	8	9	10
概　率	0.3	0.1	0.6

乙射手

击中环数	8	9	10
概　率	0.2	0.5	0.3

试问哪一个射手技术较好?

- 显然这个问题的答案不是一眼就可以看出来的;
- 这也表明分布列虽然完整的描述了随机变量，但却不够集中的反应其变化情况;
- 有必要找一些量来更集中，更概括的描述随机变量;
- 所要找的量多是某种平均值.

平均值与加权平均值

- 最常见的平均值求法:$\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$;
- 没有考虑每个数据相对重要性，比如一小学生考试成绩为：语文 95 分，数学 85 分，常识 60 分，若按上述计算其平均成绩为 $\bar{x} = 80$;
- 上述计算没有考虑到三个科目的相对重要性，不太能反映学生的真正成绩：如果在这个年级中，每周有语文 10 节课，数学 8 节课，常识 2 节课，则利用下述的平均计算其平均成绩似科更合理一些:

$$\bar{x}_w = 95 * \frac{10}{20} + 85 * \frac{8}{20} + 60 * \frac{2}{20} = 87.5$$

- 加权平均：给定权 w_i 满足 $\sum_{i=1}^{n} w_i = 1$, 则

$$\bar{x}_w = \sum_{i=1}^{n} w_i x_i$$

称为加权平均值.

有甲乙两名射手，其射击技术可用下表表出

甲射手

击中环数	8	9	10
概　率	0.3	0.1	0.6

乙射手

击中环数	8	9	10
概　率	0.2	0.5	0.3

考虑以概率为权重的加权平均:

$$\bar{x}_{甲} = 8 * 0.3 + 9 * 0.1 + 10 * 0.6 = 9.3$$
$$\bar{x}_{乙} = 8 * 0.2 + 9 * 0.5 + 10 * 0.3 = 9.1$$

则平均起来，甲每枪射中 9.3 环，乙每枪射中 9.1 环，故甲的技术更好一些.

数学期望的定义

定义 3.1 设离散型随机变量的分布列为

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_k, & \cdots \\ p_1, & p_2, & \cdots, & p_k, & \cdots \end{pmatrix}$$

如果

$$\sum_{i=1}^{\infty} |x_i| p_i < \infty$$

则称

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i p_i$$

为随机变量 X 的数学期望，或称该分布的数学期望，简称期望或均值。若级数 $\sum_{i=1}^{\infty} |x_i| p_i$ 不收敛，则称 X 的期望不存在.

从离散到连续

- 假设连续型随机变量 X 的分布密度为 $p(x)$;
- 考虑随机变量 X 的如下近似：记 $A_i := \{\omega : X(\omega) \in (x_i, x_{i+1}]\}$

$$\tilde{X} := \sum_i x_i I_{A_i}(\omega)$$

其中 $I_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i \\ 0, & \omega \notin A_i \end{cases}$.

显然 $\tilde{X}$ 为离散型随机变量，且

$$P(\tilde{X} = x_i) = \int_{x_i}^{x_{i+1}} p(x)dx \approx p(x_i)(x_{i+1} - x_i);$$

- 故 $\tilde{X}$ 的期望为

$$E(\tilde{X}) \approx \sum_i x_i p(x_i)(x_{i+1} - x_i) \to \int xp(x)dx$$

连续型随机变量数学期望的定义

定义 3.2 设连续随机变量 X 的密度函数为 $p(x)$. 如果

$$\int_{-\infty}^{\infty} |x|p(x)dx < \infty,$$

则称

$$E(X) := \int_{-\infty}^{\infty} xp(x)dx$$

为随机变量 X 的数学期望，简称期望或均值。若 $\int_{-\infty}^{\infty} |x|p(x)dx$ 不收敛，则称 X 的期望不存在.

数学期望的一般情形

- 若随机变量 X 的分布函数为 $F(x)$;
- 类似于连续情形，考虑随机变量 X 的如下近似:

$$\tilde{X} := \sum_i x_i I_{A_i}(\omega)$$

其中 $A_i := \{\omega : X(\omega) \in (x_i, x_{i+1}]\}, I_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i \\ 0, & \omega \notin A_i \end{cases}$.
显然 $\tilde{X}$ 为离散型随机变量，且

$$P(\tilde{X} = x_i) = P(X \in (x_i, x_{i+1}]) = F(x_{i+1}) - F(x_i);$$

- 故 $\tilde{X}$ 的期望为

$$E(\tilde{X}) = \sum_i x_i \left[F(x_{i+1}) - F(x_i)\right] \to \int x dF(x)$$ 斯蒂尔切斯积分

数学期望的一般定义

定义 3.3 假设随机变量 X 的分布函数为 $F(x)$, 则定义

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

为随机变量 X 的数学期望。这里我们要求上述积分绝对收敛，否则称数学期望不存在.

斯蒂尔切斯 (Stieltjes) 积分的性质

$$I = \int_{-\infty}^{+\infty} g(x)dF(x)$$

- 当 $F(x)$ 为右连续的阶梯函数，在 $x_i(i = 1, 2, \cdots,)$ 具有跃度 p_i 时，上面的积分化为

$$I = \sum_i g(x_i)\Big(F(x_i) - F(x_{i-1})\Big) = \sum_i g(x_i)p_i$$

- 当 $F(x)$ 存在导数 $F'(x) = p(x)$ 时，上述积分化为普通积分

$$I = \int_{-\infty}^{+\infty} g(x)p(x)dx$$

- 线性性质

$$\int_{-\infty}^{+\infty} (ag_1(x) + bg_2(x)dF(x) = a\int_{-\infty}^{+\infty} g_1(x)dF(x) + b\int_{-\infty}^{+\infty} g_2(x)dF(x)$$

- 若 $g(x) \geq 0$, $F(x)$ 单调不减，$b > a$, 则 $\int_a^b g(x)dF(x) \geq 0$

几点注记

- 在引入上述相关的积分后，

$$
\begin{aligned}
F(x) &= P(X \le x) = \int_{(-\infty,x]} dF(x) \\
P(X \in B) &= \int_B dF(x)
\end{aligned}
$$

- 数学期望的另一计算公式：关于概率测度的积分

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x dF(x) \quad \text{注意到} F(x) = P(X(\omega) \le x) := P(X^{-1}(x)) \\
&\overset{x=X(\omega)}{=\!=\!=} \int_\Omega X(\omega) dP(X^{-1}(X(\omega))) = \int_\Omega X(\omega) dP(\omega)
\end{aligned}
$$

第十七讲：数学期望 (续) 与方差

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

随机变量关于概率测度的积分

- 若 $X(\omega)=\sum_{i=1}^{n} a_i I_{A_i}(\omega), A_1, \cdots, A_n \in \mathcal{F}$, 即 X 为简单随机变量时，定义

$$\int_{\Omega} X(\omega) P(d\omega) := \sum_{i=1}^{n} a_i P(A_i)$$

- 若 $X(\omega)$ 为非负随机变量，则由之前随机变量的结构易知:

$$\text{存在简单随机变量序列} X_n(\omega) \uparrow X(\omega)$$

故可定义

$$\int_{\Omega} X(\omega) P(d\omega) := \lim_{n\to\infty} \int_{\Omega} X_n(\omega) P(d\omega)$$

- 若 $X(\omega)$ 为任一随机变量，则由于 $X(\omega)=X^{+}(\omega)-X^{-}(\omega)$, 故定义

$$\int_{\Omega} X(\omega) P(d\omega) := \int_{\Omega} X^{+}(\omega) P(d\omega) - \int_{\Omega} X^{-}(\omega) P(d\omega)$$

随机变量可积与积分存在的定义

定义 3.4 称随机变量 $X(\omega)$ 关于概率测度 P 的积分存在，如果

$$\int_\Omega X^+(\omega)P(d\omega) \text{ 与 } \int_\Omega X^-(\omega)P(d\omega)\text{不同时为}\infty$$

称随机变量 $X(\omega)$ 关于概率测度 P 可积，如果 $\int_\Omega X(\omega)P(d\omega) < \infty$, 即

$$\int_\Omega X^+(\omega)P(d\omega) < \infty \text{ 且 } \int_\Omega X^-(\omega)P(d\omega) < \infty.$$

定义 3.5 若 $X(\omega)$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的可积随机变量，则称

$$E(X) := \int_\Omega X(\omega)P(d\omega)$$

为随机变量 $X(\omega)$ 的数学期望.

随机变量期望的性质

1. 若 c 为常数，则 $E(c) = c$;
2. 若 $X \geq 0$, 则 $E(X) \geq 0$
3. 对任意的常数 a, 有 $E(aX) = aE(X)$;
4. 对任意两个函数 $g_1(x), g_2(x)$ 有

$$E(g_1(X) \pm g_2(X)) = E(g_1(X)) \pm E(g_2(X))$$

5. 对任意两个随机变量 X, Y, 有 $E(X+Y) = E(X) + E(Y)$;
6. 一般的，对任意两个随机变量 X, Y 及任给两个常数 a, b, 有

$$E(aX + bY) = aE(X) + bE(Y);$$

7. 更一般的有

$$E(\sum_{i=1}^{n} a_i X_i + b) = \sum_{i=1}^{n} a_i E X_i + b$$

积分变换定理

定理 3.1 设 X 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，$\mathbf{F}$ 为随机变量 X 的分布，则对任意的 Borel 函数 $f(x)$ 有

$$\int_{\Omega} f(X(\omega))P(d\omega) = \int_{\mathbb{R}} f(x)\mathbf{F}(dx) \left(= \int_{\mathbb{R}} f(x)dF(x) \right).$$

特别的，若 $f(x) = x$, 则有

$$E(X) = \int_{\Omega} X(\omega)P(d\omega) = \int_{\mathbb{R}} x\mathbf{F}(dx) = \int_{\mathbb{R}} xdF(x)$$

随机变量的期望与独立性

定理 3.2 设 X, Y 为概率空间 $(\Omega, \mathcal{F}, P)$ 上两个相互独立的随机变量，且 X, Y 均可积，则乘积 XY 也可积，且

$$E(XY) = E(X)E(Y)$$

定理 3.3 概率空间 $(\Omega, \mathcal{F}, P)$ 上两个随机变量 X, Y 相互独立的充要条件是，对于使得 $f(X)$ 与 $g(Y)$ 均可积的任何 Borel 函数 f, g 均有

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

随机变量函数的期望

定理 3.4 若随机变量 X 的分布函数为 $F(x)$, 则 X 的某一函数 $g(X)$ 的期望为

$$
\begin{aligned}
E[g(X)] &= \int_{-\infty}^{+\infty} g(x)dF(x) \\
&= \begin{cases} \sum_i g(x_i)[F(x_i) - F(x_{i-1})] = \sum_i g(x_i)P(X = x_i), \\ \int_{-\infty}^{+\infty} g(x)p(x)dx. \end{cases}
\end{aligned}
$$

随机变量函数的期望：多维情形

定理 3.5 若随机向量 $(X_1, \cdots, X_n)$ 的分布函数为 $F(x_1, \cdots, x_n)$, $g(x_1, \cdots, x_n)$ 为 n 元 Borel 函数，则

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

特别的，

$$E(X_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i dF(x_1, \cdots, x_n) = \int_{-\infty}^{+\infty} x_i dF_i(x_i)$$

其中 $F_i(x_i)$ 为 X_i 的分布函数. 更进一步，如果 $(X_1, \cdots, X_n)$ 为连续型随机向量即具有联合分布密度 $p(x_1, \cdots, x_n)$, 则

$$E[g(X_1, \cdots, X_n)] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \cdots, x_n) p(x_1, \cdots, x_n) dx_1 \cdots dx_n$$

$$E(X_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i p(x_1, \cdots, x_n) dx_1 \cdots dx_n = \int_{-\infty}^{+\infty} x_i p_{X_i}(x_i) dx_i$$

随机向量的数学期望

定义 3.6 假设随机向量 $(X_1, \cdots, X_n)$ 的每个分量 X_i 的数学期望都存在，则称

$$E(X) = (EX_1, EX_2, \cdots, EX_n)$$

为随机向量 X 的数学期望向量，简称为 X 的数学期望。这里

$$E(X_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i dF(x_1, \cdots, x_n) = \int_{-\infty}^{+\infty} x_i dF_i(x_i)$$

这里 $F_i(x_i)$ 为 X_i 的分布函数.

期望的极限性质：单调收敛定理

定理 3.6 (单调收敛定理) 若随机变量序列 $\{X_n\}$ 满足条件

$$0 \le X_1(\omega) \le X_2(\omega) \le \cdots \le X_n(\omega) \uparrow X(\omega), \omega \in \Omega,$$

则

$$\begin{aligned} \int_\Omega \lim_{n\to\infty} X_n(\omega) P(d\omega) &= \lim_{n\to\infty} \int_\Omega X_n(\omega) P(d\omega), \\ \text{i.e.} \quad E(\lim_{n\to\infty} X_n) &= \lim_{n\to\infty} E(X_n). \end{aligned}$$

单调收敛定理的证明 I

- 设 $\{X_{nk}\}_{k\geqslant 1}$ 为定义 $E(X_n)$ 的简单随机变量列，且满足

$$0 \leqslant X_{11} \leqslant X_{12} \leqslant \cdots \leqslant X_{1k} \uparrow X_1$$

$$0 \leqslant X_{21} \leqslant X_{22} \leqslant \cdots \leqslant X_{2k} \uparrow X_2$$

$$\vdots$$

$$0 \leqslant X_{n1} \leqslant X_{n2} \leqslant \cdots \leqslant X_{nk} \uparrow X_n$$

$$\lim_{k\to\infty} E(X_{nk}) = E(X_n)$$

- 令 $\widetilde{X}_k = \max_{1\leqslant i\leqslant k} X_{ik}$ ，则 $\widetilde{X}_k$ 为非降简单随机变量序列，

$$X_{nk} \leqslant \widetilde{X}_k \leqslant X_k, \quad E(X_{nk}) \leq E\left(\widetilde{X}_k\right) \leq E(X_k)$$

单调收敛定理的证明 II

- 下证 $\lim_{k\to\infty}\widetilde{X}_k = X$, $\lim_{k\to\infty}E(\widetilde{X}_k) = \lim_{n\to\infty}E(X_n)$
  - 固定 n，令 $k\to+\infty$ 可得

$$X_n \le \lim_{k\to+\infty}\widetilde{X}_k \le X, \qquad E(X_n) \le \lim_{k\to+\infty}E\left(\widetilde{X}_k\right) \le \lim_{k\to+\infty}E(X_k)$$

  - 再令 $n\to+\infty$ 可得

$$\lim_{k\to\infty}\widetilde{X}_k = X, \qquad \lim_{k\to\infty}E(\widetilde{X}_k) = \lim_{n\to\infty}E(X_n)$$

- 结合 $E(\lim_{n\to\infty}X_n) = E(X) := \lim_{n\to\infty}E(\widetilde{X}_k)$ 可知定理得证.

期望的极限性质：Fatou 引理

定理 3.7 (Fatou 引理) 若 $\{X_n\}$ 是一随机变量序列，

- 如果存在可积随机变量 $\underline{Y}$ 使得 $X_n \geq \underline{Y}$, 则有

$$\begin{aligned} \int_\Omega \liminf_{n\to\infty} X_n(\omega)P(d\omega) &\leq \liminf_{n\to\infty} \int_\Omega X_n(\omega)P(d\omega), \\ \text{i.e.} \quad E(\liminf_{n\to\infty} X_n) &\leq \liminf_{n\to\infty} E(X_n). \end{aligned}$$

- 如果存在可积随机变量 $\overline{Y}$ 使得 $X_n \leq \overline{Y}$, 则有

$$\begin{aligned} \int_\Omega \limsup_{n\to\infty} X_n(\omega))P(d\omega) &\geq \limsup_{n\to\infty} \int_\Omega X_n(\omega))P(d\omega), \\ \text{i.e.} \quad E(\limsup_{n\to\infty} X_n) &\geq \limsup_{n\to\infty} E(X_n). \end{aligned}$$

Fatou 引理证明

- 如果 $\liminf_{n\to\infty} E(X_n) = +\infty$, 则不等式显然成立. 以下设 $\liminf_{n\to\infty} E(X_n)$ 有限.
- 令 $Y_n = \inf_{k\geq n}(X_k - \underline{Y}), n \geq 1$, 则 $\{Y_n\}$ 为非负随机变量的上升序列，且以 $\liminf_{n\to\infty}(X_n - \underline{Y})$ 为极限.
- 由单调收敛定理可得:

$$E\left[\liminf_{\to\infty}(X_n - \underline{Y})\right] = \lim_{n\to\infty} E\left[\inf_{k\geq n}(X_k - \underline{Y})\right] \leq \liminf_{n\to\infty} E(X_n - \underline{Y}).$$

- 消去有限的 $E(\underline{Y})$ 可证得结论
- 对 $\{-X_n\}$ 用下确界的结论可证得上确界的结论.

期望的极限性质：控制收敛定理

定理 3.8 (控制收敛定理) 设 $\{X_n\}$ 是一随机变量序列。如果存在可积随机变量 Y 使得 $|X_n| \le Y$, 且 $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, 则随机变量 $X(\omega)$ 可积，且有

$$\begin{aligned} \int_\Omega \lim_{n\to\infty} X_n(\omega)P(d\omega) &= \lim_{n\to\infty} \int_\Omega X_n(\omega)P(d\omega), \\ \text{i.e.} \quad E(X) = E(\lim_{n\to\infty} X_n) &= \lim_{n\to\infty} E(X_n). \end{aligned}$$

控制收敛定理的证明

- $|X_n| \leq Y$ 蕴含 $|X| \leq Y$, 故 X 可积.
- 取 $\underline{Y} = -Y$ 与 $\overline{Y} = Y$, 用 Fatou 引理便得到

$$E(X) \leq \liminf_{n\to\infty} E\left(X_n\right) \leq \limsup_{n\to\infty} E\left(X_n\right) \leq E(X)$$

- 故 $\lim_n E\left(X_n\right)$ 存在且等于 $E(X)$, 定理至此得证.

随机变量方差的定义

定义 3.7 若随机变量 X^2 的期望 $E(X^2)$ 存在，则称

$$D(X) := E\left[(X-EX)^2\right]$$

为随机变量 X 的方差。有时我们也用 $Var(X)$ 来表示 X 的方差。称方差 $D(X)$ 的正平方根 $\sqrt{D(X)}$ 为随机变量 X 的标准差，记为 $\sigma(X)$ 或 σ_X.

定理 3.9 假设随机变量 X 的方差存在，则

$$\begin{aligned} D(X) &= E[(X-EX)^2] = \int_{-\infty}^{+\infty}(x-EX)^2 dF(x) \\ &= \begin{cases} \sum_i (x_i-EX)^2[F(x_i)-F(x_{i-1})] = \sum_i (x_i-EX)^2 P(X=x_i) \\ \int_{-\infty}^{+\infty}(x-EX)^2 p(x)dx \end{cases} \end{aligned}$$

方差的性质

1. $D(X) = E[(X - EX)^2] \geq 0$, $D(X) = 0$ 当且仅当 $P(X = EX) = 1$;
2. 常数的方差为 0 即 $D(c) = 0$, 其中 c 为常数;
3. $D(X) = E(X^2) - (EX)^2$: 事实上，根据方差的定义及期望的线性性质，我们有

$$
\begin{aligned}
D(X) &= E[(X - EX)^2] = E[X^2 - 2X \cdot EX + (EX)^2] \\
&= E(X^2) - 2EX \cdot EX + (EX)^2 \\
&= E(X^2) - (EX)^2
\end{aligned}
$$

方差的性质 (续)

4. 若 $E(X^2)=0$, 则 $E(X)=0, D(X)=0$. 事实上

$$0 \le D(X) = E(X^2) - (EX)^2 = -(EX)^2 \le 0$$

5. 若 a, b 为常数，则 $D(aX+b) = a^2 D(X)$, 事实上，

$$\begin{aligned} D(aX+b) &= E(aX+b-E(aX+b))^2 = E(aX-aEX)^2 \\ &= E[a^2(X-EX)^2] = a^2 E[(X-EX)^2] = a^2 D(X) \end{aligned}$$

6. $f(c) := E(X-c)^2$ 当且仅当 $c = E(X)$ 时取到最小值:由

$$f(c) = E(X-EX+EX-c)^2 = E(X-EX)^2 + [EX-c]^2$$

可得.

切比雪夫 (Chebyshev) 不等式

定理 3.10 设随机变量 X 的期望与方差均存在，则对任意的 $\epsilon > 0$ 有

$$P(|X - EX| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$$

证明: 记 $a = EX$, 则

$$\begin{aligned} P(|X-a| \geq \epsilon) &= \int_{\{x:|x-a|\geq\epsilon\}} dF(x) \leq \int_{\{x:|x-a|\geq\epsilon\}} \frac{(x-a)^2}{\epsilon^2} dF(x) \\ &\leq \frac{1}{\epsilon^2}\int_{-\infty}^{+\infty}(x-a)^2 dF(x) = \frac{D(X)}{\epsilon^2} \end{aligned}$$

$D(X)=0 \Leftrightarrow P(X=EX)=1$

定理 3.11 若随机变量 X 的方差存在，则

$$D(X)=0 \Leftrightarrow P(X=EX)=1.$$

特别的，如果 $E(X^2)=0$, 则 $E(X)=0, D(X)=0$, 从而 $P(X=0)=1$.

证明: 充分性显然，下面证必要性。设 $D(X)=0$, 此时 EX 存在。注意到

$$\{|X-EX|>0\} = \cup_{n=1}^{\infty}\{|X-EX|\geq \frac{1}{n}\}$$

故

$$\begin{aligned} P(|X-EX|>0) &= P(\cup_{n=1}^{\infty}\{|X-EX|\geq \frac{1}{n}\}) \\ &\leq \sum_{n=1}^{\infty} P(|X-EX|\geq \frac{1}{n}) \leq \sum_{n=1}^{\infty}\frac{D(X)}{(1/n)^2}=0 \end{aligned}$$

从而

$$P(X=EX)=1-P(|X-EX|>0)=1-0=1$$

二项分布 $B(n,p)$ 的期望:$P(X=k)=C_n^k p^k(1-p)^{n-k}$

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} \frac{k\cdot n!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n\cdot (n-1)!}{(k-1)!(n-1-(k-1))!}pp^{k-1}(1-p)^{n-1-(k-1)} \\
&= np\sum_{k=1}^{n} C_{n-1}^{k-1}p^{k-1}(1-p)^{n-1-(k-1)} \\
&= np
\end{aligned}
$$

特别的，对于两点分布 $(B(1,p))$ 随机变量 X, 有 $E(X)=p$. 事实上对于两点分布其期望为: $E(X)=0\cdot P(X=0)+1\cdot P(X=1)=p$.

计算二项分布期望的另一方法

前面在讲二项分布时，我们知道:

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

其中 $X_k, k = 1, \cdots, n$ 为第 k 次伯努利试验成功的次数，显然服从两点分布。故

$$\begin{aligned} EX_i &= 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = p \\ EX_i^2 &= 0^2 \cdot P(X_i = 0) + 1^2 \cdot P(X_i = 1) = p \\ D(X_i) &= EX_i^2 - (EX_i)^2 = p - p^2 = p(1-p) \end{aligned}$$

再由期望的线性性知

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

二项分布的方差 $D(X)$

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 P(X=k) = \sum_{k=1}^{n} k(k-1+1) C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k(k-1) C_n^k p^k (1-p)^{n-k} + \sum_{k=1}^{n} k C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1) C_n^k p^k (1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{k=2}^{n} C_{n-2}^{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} + np \\
&= n(n-1)p^2 + np
\end{aligned}
$$

故

$$D(X) = E(X^2) - (EX)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

特别的，两点分布的方差为 $p(1-p)$.

泊松分布的期望与方差: $P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty} \lambda\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda \\
E(X^2) &= \sum_{k=0}^{\infty} k^2P(X=k) = \sum_{k=1}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty} [(k-1)+1]\frac{\lambda^k}{(k-1)!}e^{-\lambda} \\
&= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda^2+\lambda \\
D(X) &= E(X^2)-(EX)^2 = \lambda^2+\lambda-\lambda^2 = \lambda
\end{aligned}
$$

几何分布的期望与方差:$P(X=k)=(1-p)^{k-1}p$

$$
\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty}\sum_{n=1}^{k}(1-p)^{k-1} = p\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}(1-p)^{k-1} \\
&= p\sum_{n=1}^{\infty}(1-p)^{n-1}\frac{1}{1-(1-p)} = \sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{1}{p} \\
E(X^2) &= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\left[\sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1} - \sum_{k=1}^{\infty} k(1-p)^{k-1}\right] \\
&= 2p\sum_{k=1}^{\infty}\sum_{n=1}^{k} n(1-p)^{k-1} - p\sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= 2p\sum_{n=1}^{\infty} n\sum_{k=n}^{\infty}(1-p)^{k-1} - \frac{1}{p} = 2p\sum_{n=1}^{\infty} n(1-p)^{n-1}/p - \frac{1}{p} \\
&= \frac{2}{p^2} - \frac{1}{p} \\
D(X) &= E(X^2) - (EX)^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}
\end{aligned}
$$

负二项分布的期望与方差: $P(X=k)=C_{k-1}^{r-1}p^r(1-p)^{k-r}$

$$
\begin{aligned}
E(X) &= \sum_{k=r}^{\infty} kC_{k-1}^{r-1}p^r(1-p)^{k-r} = \sum_{k=r}^{\infty} \frac{r}{p}\frac{k\cdot(k-1)!}{r\cdot(r-1)!(k-r)!}p^{r+1}(1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} \frac{r}{p}C_{k+1-1}^{r+1-1}p^{r+1}(1-p)^{k-r} = \frac{r}{p}\sum_{l=r+1}^{\infty} C_{l-1}^{r+1-1}p^{r+1}(1-p)^{l-(r+1)} = \frac{r}{p} \\
E(X^2) &= \sum_{k=r}^{\infty} k^2C_{k-1}^{r-1}p^r(1-p)^{k-r} = \sum_{k=r}^{\infty} k(k+1-1)C_{k-1}^{r-1}p^r(1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} k(k+1)C_{k-1}^{r-1}p^r(1-p)^{k-r} - \sum_{k=r}^{\infty} kC_{k-1}^{r-1}p^r(1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} \frac{r(r+1)}{p^2}C_{k+2-1}^{r+2-1}p^{r+2}(1-p)^{(k+2)-(r+2)} - \frac{r}{p} \\
&= \frac{r(r+1)}{p^2} - \frac{r}{p} \\
D(X) &= E(X^2)-(EX)^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}
\end{aligned}
$$

均匀分布的期望与方差:$X \sim U(a,b)$

- 若 $X \sim U(a,b)$, 即 $p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,b) \\ 0, & \text{其他} \end{cases}$, 故

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xp(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{a+b}{2} \\
E(X^2) &= \int_{-\infty}^{+\infty} x^2p(x)dx = \int_a^b x^2\frac{1}{b-a}dx = \frac{a^2+ab+b^2}{3} \\
D(X) &= E(X^2) - (EX)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}
\end{aligned}
$$

标准正态分布的期望与方差:$X \sim N(0,1)$

- 若 $X \sim N(0,1)$ 即 $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 则

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x p(x) dx = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 \\
D(X) &= E[(X - EX)^2] = E(X^2) = \int_{-\infty}^{+\infty} x^2 p(x) dx \\
&= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x d(-e^{-\frac{x^2}{2}}) \\
&= \frac{1}{\sqrt{2\pi}} \left(-x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right) \\
&= 1
\end{aligned}
$$

正态分布的期望与方差:$X \sim N(\mu, \sigma^2)$

- 若 $X \sim N(\mu, \sigma^2)$, 则易知 $\dfrac{X-\mu}{\sigma} \sim N(0,1)$, 故

$$
\begin{aligned}
E(X) &= E(\sigma \cdot \frac{X-\mu}{\sigma} + \mu) = \sigma E(\frac{X-\mu}{\sigma}) + \mu = \mu \\
D(X) &= D(\sigma \cdot \frac{X-\mu}{\sigma} + \mu) = \sigma^2 D(\frac{X-\mu}{\sigma}) = \sigma^2
\end{aligned}
$$

Gamma 分布的期望

- 若 $X \sim \Gamma(\alpha, \lambda)$ 即 $p(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$，故

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x p(x) dx = \int_0^{+\infty} x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\
&= \int_0^{+\infty} \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{\alpha+1-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}; \\
E(X^2) &= \int_{-\infty}^{+\infty} x^2 p(x) dx = \int_0^{+\infty} x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\
&= \int_0^{+\infty} \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} \frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)} x^{\alpha+2-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\lambda^2} \\
D(X) &= E(X^2) - (EX)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}
\end{aligned}
$$

指数分布与 χ^2 分布的期望与方差

- 若 $X \sim \Gamma(1, \lambda)$ 分布即 X 服从指数分布，此时 $\alpha = 1$

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} = \frac{1}{\lambda} \\
E(X^2) &= \frac{\alpha(\alpha+1)}{\lambda^2} = \frac{2}{\lambda^2} \\
D(X) &= \frac{\alpha}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}
$$

- 若 $X \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ 即 X 服从 $\chi^2(n)$ 分布，此时 $\alpha = n/2, \lambda = 1/2$

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} = n \\
E(X^2) &= \frac{\alpha(\alpha+1)}{\lambda^2} = n(n+2) \\
D(X) &= \frac{\alpha}{\lambda^2} = 2n
\end{aligned}
$$

第十八讲：协方差与相关系数

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

协方差

定义 3.8 对于随机向量 $X=(X_1,\cdots,X_n)$, 类似于随机向量期望的定义，我们定义随机向量 X 的方差为 $D(X):=(D(X_1),D(X_2),\cdots,D(X_n))$.

定义 3.9 设随机向量 $(X_1,\cdots,X_n)$ 的每个分量 X_i 的方差均存在，称

$$cov(X_i,X_j):=E[(X_i-EX_i)(X_j-EX_j)],i,j=1,2,\cdots,n$$

为 X_i 与 X_j 的协方差. 而将协方差构成的 $n\times n$ 方阵

$$B=(b_{ij}),\quad b_{ij}=cov(X_i,X_j)$$

称为随机向量的协方差阵。一般来说，

- 若 $cov(X_i,X_j)>0$ 时，称 X_i 与 X_j 正相关;
- 若 $cov(X_i,X_j)<0$ 时，称 X_i 与 X_j 负相关;
- 若 $cov(X_i,X_j)=0$ 时，称 X_i 与 X_j 不相关或零相关.

协方差的性质:$cov(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)]$

- $cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$: 事实上

$$\begin{aligned} cov(X_i, X_j) &= E[(X_i - EX_i)(X_j - EX_j)] \\ &= E\Big[X_iX_j - X_iE(X_j) - X_jE(X_i) + E(X_i)E(X_j)\Big] \\ &= E(X_iX_j) - E(X_i)E(X_j) \end{aligned}$$

- 若 a 为常数，则 $cov(X_i, a) = 0$;
- 对任意常数 a, b, 有 $cov(aX_i, bX_j) = ab \cdot cov(X_i, X_j)$;
- $cov(X_i + X_j, X_k) = cov(X_i, X_k) + cov(X_j, X_k)$.

协方差与方差: $cov(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)]$

- $cov(X_i, X_i) = E(X_i - EX_i)^2 = D(X_i)$;
- $D(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^n D(X_i) + 2\sum_{1\le i<j\le n} cov(X_i, X_j)$: 事实上

$$
\begin{aligned}
D(\sum_{i=1}^n X_i) &= E(\sum_{i=1}^n X_i - E\sum_{i=1}^n X_i)^2 = E(\sum_{i=1}^n (X_i - EX_i))^2 \\
&= E\left[\sum_{i=1}^n (X_i - EX_i)^2 + 2\sum_{1\le i<j\le n} (X_i - EX_i)(X_j - EX_j)\right] \\
&= \sum_{i=1}^n D(X_i) + 2\sum_{1\le i<j\le n} cov(X_i, X_j)
\end{aligned}
$$

- 特别的,

$$D(X_i + X_j) = D(X_i) + D(X_j) + 2cov(X_i, X_j)$$

协方差矩阵的性质:$cov(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)]$

- 对称性: $cov(X_i, X_j) = cov(X_j, X_i)$, 由定义可直接推得，从而协方差阵 B 是对称矩阵;
- 协方差矩阵 B 是非负定矩阵，事实上，对任意向量 $y = (y_1, y_2, \cdots, y_n)$,

$$
\begin{aligned}
yBy' &= \sum_{i,j} y_i y_j b_{ij} = \sum_{i,j} y_i y_j E[(X_i - EX_i)(X_j - EX_j)] \\
&= E \sum_{i,j} y_i (X_i - EX_i) \cdot y_j (X_j - EX_j) \\
&= E[\sum_i y_i (X_i - EX_i)]^2 \geq 0
\end{aligned}
$$

例 3.1 设二维随机变量 (X,Y) 的联合密度为

$$p(x,y)=\begin{cases}3x, & 0<y<x<1,\\ 0, & \text{其他}\end{cases}$$

试求 $cov(X,Y)$.

解: 由 $cov(X,Y)=E(XY)-EXEY$ 知，我们只需计算 $E(XY),EX,EY$ 的值.

$$\begin{aligned}
EX &= \int_{-\infty}^{+\infty} xp_X(x)dx=\int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} p(x,y)dydx\\
&= \int_0^1 x\int_0^x 3xdydx=\int_0^1 3x^3dx=3/4\\
EY &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)dydx=\int_0^1\int_0^x 3xydydx=3/8\\
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp(x,y)dydx=\int_0^1\int_0^x xy\cdot 3xdydx=3/10\\
cov(X,Y) &= \frac{3}{10}-\frac{3}{4}\cdot\frac{3}{8}=\frac{3}{160}>0
\end{aligned}$$

例 3.2 设二维随机变量 (X,Y) 的联合密度为

$$p(x,y)=\begin{cases}\dfrac{x+y}{3}, & 0<x<1, 0<y<2,\\ 0, & \text{其他}\end{cases}$$

试求 $D(2X-3Y+8)$.

解: 注意到

$$\begin{aligned}D(2X-3Y+8) &= D(2X-3Y)=D(2X)+D(-3Y)+2cov(2X,-3Y)\\ &= 2^2D(X)+(-3)^2D(Y)+2\cdot 2\cdot(-3)cov(X,Y)\\ &= 4D(X)+9D(Y)-12cov(X,Y)\end{aligned}$$

所以我们需要计算:$E(X),E(X^2),E(Y),E(Y^2),E(XY)$.

$$\begin{aligned}E(X) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)dydx=\int_0^1\int_0^2 x\frac{x+y}{3}dydx=5/9\\ E(X^2) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2p(x,y)dydx=\int_0^1\int_0^2 x^2\frac{x+y}{3}dydx=7/18\end{aligned}$$

类似的可计算

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)dydx = \int_0^1\int_0^2 y\frac{x+y}{3}dydx = 11/9 \\
E(Y^2) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2p(x,y)dydx = \int_0^1\int_0^2 y^2\frac{x+y}{3}dydx = 16/9 \\
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp(x,y)dydx = \int_0^1\int_0^2 xy\frac{x+y}{3}dydx = 2/3
\end{aligned}
$$

从而

$$
\begin{aligned}
D(X) &= 7/18-(5/9)^2 = 13/162, \\
D(Y) &= 16/9-(11/9)^2 = 23/81, \\
cov(X,Y) &= 2/3-5/9\cdot 11/9 = -1/81 \\
D(2X-3Y+8) &= 4\cdot 13/162+9\cdot 23/81-12\cdot(-1/81) = 245/81
\end{aligned}
$$

二维正态分布的协方差矩阵

若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 即其联合分布密度为

$$
\begin{aligned}
p(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times \\
&\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}
\end{aligned}
$$

则由上一章的知识知:$X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2,\sigma_2^2)$, 故

$$
\begin{aligned}
&E(X)=\mu_1, \quad E(Y)=\mu_2 \\
&b_{11}=cov(X,X)=D(X)=\sigma_1^2, \\
&b_{22}=cov(Y,Y)=D(Y)=\sigma_2^2
\end{aligned}
$$

$$b_{12} = cov(X,Y) = E[(X-EX)(Y-EY)] = \iint (x-\mu_1)(y-\mu_2)p(x,y)dxdy$$

$$\xlongequal[v=(y-\mu_2)/\sigma_2]{u=(x-\mu_1)/\sigma_1} \iint \frac{\sigma_1\sigma_2 uv}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\}dudv$$

$$= \iint \frac{\sigma_1\sigma_2 uv}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}\left[(u-\rho v)^2+(1-\rho^2)v^2\right]\}dudv$$

$$\xlongequal[t=v]{s=(u-\rho v)/\sqrt{1-\rho^2}} \frac{\sigma_1\sigma_2}{2\pi} \iint (\sqrt{1-\rho^2}st+\rho t^2)\exp\{-\frac{s^2+t^2}{2}\}dsdt$$

$$= \frac{\sigma_1\sigma_2}{2\pi}\left(\sqrt{1-\rho^2}\int_{-\infty}^{+\infty} se^{-\frac{s^2}{2}}ds \int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}}dt + \rho\int_{-\infty}^{+\infty} e^{-\frac{s^2}{2}}ds \int_{-\infty}^{+\infty} t^2e^{-\frac{t^2}{2}}dt\right)$$

$$= \frac{\sigma_1\sigma_2}{2\pi}(\sqrt{1-\rho^2}\cdot 0 + \rho\sqrt{2\pi}\cdot\sqrt{2\pi}) = \rho\sigma_1\sigma_2$$

故二维正态分布的协方差矩阵为

$$B = \begin{pmatrix} \sigma_1^2, & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2, & \sigma_2^2 \end{pmatrix}$$

相关系数

定义 3.10 设 X, Y 为方差存在的两个随机变量，则称

$$r := \frac{cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为 X 与 Y 的相关系数.

相关系数的性质: $r := \dfrac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

- $r = 0 \Leftrightarrow cov(X,Y) = 0$ 即 X, Y 不相关;
- 若记 $X^* := \dfrac{X - EX}{\sqrt{D(X)}}, \quad Y^* := \dfrac{Y - EY}{\sqrt{D(Y)}}$, 则 $EX^* = EY^* = 0$, 从而

$$\begin{aligned} r &:= \frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{E[(X-EX)(Y-EY)]}{\sqrt{D(X)}\sqrt{D(Y)}} \\ &= E(X^*Y^*) = E[(X^* - EX^*)(Y^* - EY^*)] \\ &= cov(X^*, Y^*) \end{aligned}$$

- $|r| \le 1$, 且
  - $r = 1$ 当且仅当 $P(X^* = Y^*) = 1$;
  - $r = -1$ 当且仅当 $P(X^* = -Y^*) = 1$.

柯西 - 施瓦兹 (Cauchy-Schwarz) 不等式

定理 3.12 对任意的随机变量 X 与 Y 都有

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

等式成立当且仅当存在常数 t_0 使得

$$P(Y = t_0 X) = 1$$

Cauchy-Schwarz 不等式的证明:$[E(XY)]^2 \le E(X^2)E($

证明: 对任意的实数 t, 定义

$$u(t) := E(tX - Y)^2 = t^2E(X^2) - 2tE(XY) + E(Y^2).$$

显然，对一切的 $t \in R$, $u(t) \ge 0$, 故方程 $u(t) = 0$ 或者没有实根或者有重根，从而

$$\Delta = [2E(XY)]^2 - 4E(X^2)E(Y^2) \le 0$$

即

$$[E(XY)]^2 \le E(X^2)E(Y^2)$$

上述不等式等号成立，当且仅当 $\Delta = 0$ 即 $u(t) = 0$ 存在一个重根 t_0, 这时

$$u(t_0) = E(t_0X - Y)^2 = 0$$

从而由可知

$$P(t_0X - Y = 0) = 1 \Leftrightarrow P(Y = t_0X) = 1.$$

相关系数绝对值小于等于 1 的证明

- 首先由 $X^* = \dfrac{X-EX}{\sqrt{D(X)}}, Y^* = \dfrac{Y-EY}{\sqrt{D(Y)}}$ 知:

$$\begin{aligned} EX^* &= EY^* = 0 \\ D(X^*) &= D\left(\frac{X-EX}{\sqrt{D(X)}}\right) = \frac{1}{D(X)}D(X-EX) = \frac{1}{D(X)}D(X) = 1 \\ D(Y^*) &= D\left(\frac{Y-EY}{\sqrt{D(Y)}}\right) = \frac{1}{D(Y)}D(Y-EY) = \frac{1}{D(Y)}D(Y) = 1 \end{aligned}$$

- 其次，由 $r = cov(X^*, Y^*)$ 知

$$\begin{aligned} |r| &= |cov(X^*, Y^*)| = |E(X^*Y^*) - E(X^*)E(Y^*)| \le \sqrt{E[(X^*)^2]E[(Y^*)^2]} \\ &= \sqrt{D(X^*)D(Y^*)} = 1; \end{aligned}$$

- $|r| = 1$ 当且仅当存在 t_0 使得 $P(Y^* = t_0X^*) = 1$, 而此时有

$$r = E(X^*Y^*) = t_0E[(X^*)^2] = t_0$$

随机变量不相关时期望与方差的性质

定理 3.13 对于随机变量 X, Y, 下面四个事实是等价的:

1. X, Y 不相关;
2. $r = 0$ 即相关系数为 0;
3. $E(XY) = E(X)E(Y)$;
4. $D(X+Y) = D(X) + D(Y)$.

证明: (1) 与 (2) 等价是显然的，根据定义即可得. 由于

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

故 $cov(X, Y) = 0$ 当且仅当 $E(XY) = E(X)E(Y)$, 即 (1) 和 (3) 等价. 另外，又由于

$$D(X+Y) = D(X) + D(Y) + 2cov(X, Y)$$

故 $cov(X, Y) = 0$ 当且仅当 $D(X+Y) = D(X) + D(Y)$, 即 (1) 与 (4) 等价.

独立与不相关的联系：独立必定不相关

定理 3.14 若 X, Y 独立，则 X 与 Y 不相关.

证明: 我们仅对连续型随机变量给出证明。因为 X, Y 独立，故其联合分布密度 $p(x,y) = p_X(x)p_Y(y)$, 从而

$$\begin{aligned} E(XY) &= \iint xyp(x,y)dxdy = \iint xyp_X(x)p_Y(y)dxdy \\ &= \int xp_X(x)dx \int yp_Y(y)dy = E(X)E(Y) \end{aligned}$$

从而 X, Y 不相关.

定理 3.15 若 X, Y 独立，则

$$E(XY) = E(X)E(Y), \quad D(X+Y) = D(X) + D(Y)$$

更一般的，我们有若 $X_1, \cdots, X_n$ 为相互独立的随机变量，则.

$$\begin{aligned} E(X_1 \cdots X_n) &= E(X_1)E(X_2)\cdots E(X_n); \\ D(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} D(X_i) \end{aligned}$$

独立与不相关的联系：不相关未必独立

例 3.3 设随机变量 $X \sim N(0, \sigma^2)$, 并令 $Y = X^2$, 显然 X, Y 不独立，但是此时 X, Y 不相关。事实上，此时

$$cov(X, Y) = E(X \cdot X^2) - E(X)E(X^2) = 0.$$

例 3.4 设 θ 为 $[0,2\pi]$ 上的均匀分布，a 为一固定常数，令

$$X=\cos\theta,\quad Y=\cos(\theta+a).$$

则我们有

$$\begin{aligned}
E(X) &= \int_0^{2\pi}\cos t\cdot\frac{1}{2\pi}dt=0,\quad E(Y)=\int_0^{2\pi}\cos(t+a)\cdot\frac{1}{2\pi}dt=0\\
E(X^2) &= \int_0^{2\pi}\cos^2 t\cdot\frac{1}{2\pi}dt=\frac{1}{2},\quad E(Y^2)=\int_0^{2\pi}\cos^2(t+a)\cdot\frac{1}{2\pi}dt=\frac{1}{2}\\
E(XY) &= \int_0^{2\pi}\cos t\cos(t+a)\cdot\frac{1}{2\pi}dt=\frac{1}{2}\cos a
\end{aligned}$$

因此 $r=\cos a$. 故

- $a=0$ 时，$r=1, X=Y$;
- $a=\pi$ 时，$r=-1, X=-Y$;
- $a=\frac{\pi}{2}$或$\frac{3\pi}{2}$ 时，$r=0$, X,Y 不相关，但此时 $X^2+Y^2=1$, 因此不独立.

二维正态分布：不相关与独立等价

定理 3.16 对于二维正态分布 (X,Y),X 与 Y 不相关与 X,Y 独立等价.

证明: 因为独立必然不相关，故我们仅需证在不相关条件下，X,Y 独立即可. 对于二维正态分布，我们有

$$cov(X,Y)=\rho\sigma_1\sigma_2$$

从而 X,Y 不相关，当且仅当 $\rho=0$, 此时

$$\begin{aligned}
p(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times \\
&\quad \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}\exp\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\} \\
&= p_X(x)P_Y(y)
\end{aligned}$$

故 X,Y 独立，定理得证.

第十九讲：条件数学期望与母函数

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

条件数学期望

- 在第二章中对于任何有正概率的事件 B, 我们引入了条件概率 $P(\cdot|B)$ 及条件分布 $F(x|B) := P(X \leq x|B)$;
- 类似于概率与分布函数，对于上述的条件概率及条件分布，我们也可以定义相应的条件数学期望:

定义 3.11 如果下述积分绝对收敛，则称

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

为已知事件 B 发生后 X 的条件数学期望.

条件数学期望的两类特殊情形：离散型与连续型

定理 3.17 若 X, Y 均为离散型随机变量，并且条件数学期望定义中的 B 选为 $B := \{Y = y_j\}$ 的形式，则

$$E(X|Y = y_j) = \int_{-\infty}^{+\infty} x dF(x|Y = y_j) = \sum_i x_i P(X = x_i|Y = y_j)$$

定理 3.18 若 X, Y 均为连续型随机变量，并且条件数学期望定义中的 B 选为 $B := \{Y = y\}$ 的形式，则

$$E(X|Y = y) = \int_{-\infty}^{+\infty} x dF(x|Y = y) = \int_{-\infty}^{+\infty} x p(x|y) dx$$

随机变量函数的条件数学期望及条件方差

定理 3.19 设 $g(x)$ 为 Borel 函数，则 $g(X)$ 关于 $Y=y$ 的条件期望为

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

定义 3.12 假设 X 的方差存在，则称

$$\begin{aligned} D(X|Y=y) &:= E[(X-E(X|Y=y))^2|Y=y] \\ &= \int_{-\infty}^{+\infty}(x-E(X|Y=y))^2 dF(x|y) \end{aligned}$$

为给定 $Y=y$ 后 X 的条件方差.

例 3.5 若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则

$$E(X|Y=y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2).$$

证明: 在第二章中，我们知道给定 $Y=y$ 下，X 的条件分布服从

$$N(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2))$$

故

$$E(X|Y=y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$$

一般来说，给定 $Y=y$ 后 X 的条件数学期望是 Y 的可能值 y 的函数，记之为

$$\varphi(y):=E(X|Y=y)$$

如果再将 y 用 Y 代回，就得到一个随机变量 $\varphi(Y)$, 相应的，我们记

$$E(X|Y):=\varphi(Y)$$

并称随机变量 $E(X|Y)$ 为 X 关于 Y 的条件数学期望.

条件数学期望的性质：重期望

定理 3.20 设 (X,Y) 为二维随机向量，则对于 Borel 函数 $g(x)$ 有

$$E[E(g(X)|Y)] = E(g(X))$$

证明: 我们仅对二维连续型随机变量给出证明。令 $\varphi(y) := E(g(X)|Y=y)$, 则 $E(g(X)|Y) = \varphi(Y)$, 从而

$$\begin{aligned}
\varphi(y) &= E(g(X)|Y=y) = \int_{-\infty}^{+\infty} g(x)dF(x|y) \\
&= \int_{-\infty}^{+\infty} g(x)p(x|Y=y)dx = \int_{-\infty}^{+\infty} g(x)\frac{p(x,y)}{p_Y(y)}dx \\
E[E(g(X)|Y)] &= E(\varphi(Y)) = \int_{-\infty}^{+\infty} \varphi(y)p_Y(y)dy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)\frac{p(x,y)}{p_Y(y)}p_Y(y)dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)p(x,y)dxdy = E(g(X))
\end{aligned}$$

重期望的不同表达形式

- 上述定理的重期望公式可写成如下形式

$$
\begin{aligned}
E(g(X)) &= \int_{-\infty}^{+\infty} E[g(X)|Y=y]dF_Y(y) \\
&= \begin{cases} \sum_j E(g(X)|Y=y_j)P(Y=y_j) \\ \int_{-\infty}^{+\infty} E(g(X)|Y=y)p_Y(y)dy \end{cases}
\end{aligned}
$$

- 特别的，若 $g(X)=X$, 则有

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} E[X|Y=y]dF_Y(y) \\
&= \begin{cases} \sum_j E(X|Y=y_j)P(Y=y_j) \\ \int_{-\infty}^{+\infty} E(X|Y=y)p_Y(y)dy \end{cases}
\end{aligned}
$$

巴格达窃贼问题

例 3.6 一个窃贼被关在有 3 个门的地牢中。其中 1 号门通向自由，出 1 号门后走 3 个小时便回到地面；2 号门通向一个地道，在此地道走 5 个小时后返回地牢；3 号门通向一个更长的地道，沿这个地道走 7 个小时后回到地牢。如果窃贼每次选择 3 个门的可能性总相等，求他为获自由而奔走的平均时间.

解: 设窃贼需要走 X 小时到达地面，则 X 的所有可能取值为

$$3, 5+3, 7+3, 5+5+3, 5+7+3, 7+7+3, \cdots,$$

则显然要写出 X 的分布列是困难的，所以无法直接求 $E(X)$. 但是如果我们引入 Y 表示窃贼每次对 3 个门的选择，则

$$P(Y=1)=P(Y=2)=P(Y=3)=1/3$$

$$E(X|Y=1)=3, E(X|Y=2)=5+E(X), E(X|Y=3)=7+E(X)$$

从而由重期望公式可得

$$E(X)=\sum_{i=1}^{3} E(X|Y=i)P(Y=i)=5+\frac{2}{3}E(X) \Rightarrow E(X)=15$$

随机个随机变量和的期望

定理 3.21 设 $X_1, X_2, \cdots$，为一列独立同分布的随机变量，随机变量 N 只取正整数值，且 N 与 $\{X_n\}$ 独立，则

$$E\left(\sum_{i=1}^{N} X_i\right) = E(X_1)E(N)$$

证明: 由重期望公式可知

$$\begin{aligned}
E\left(\sum_{i=1}^{N} X_i\right) &= E\left[E\left(\sum_{i=1}^{N} X_i | N\right)\right] = \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{N} X_i | N=n\right) P(N=n) \\
&= \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{n} X_i | N=n\right) P(N=n) = \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{n} X_i\right) P(N=n) \\
&= \sum_{n=1}^{+\infty} nE(X_1)P(N=n) = E(X_1)E(N)
\end{aligned}$$

矩的概念

定义 3.13 如果 $E|X|^k < +\infty$, 则称 $m_k := E(X^k)$ 为随机变量 X(及其分布) 的 k 阶原点矩。而称 $c_k := E(X-EX)^k$ 为随机变量 (及其分布) 的 k 阶中心矩.

定理 3.22 当 $E|X|^k < +\infty$ 时有,

$$c_k = \sum_{i=0}^{k} C_k^i (-m_1)^{k-i} m_i, \quad m_k = \sum_{i=0}^{k} C_k^i c_{k-i} m_1^i.$$

证明: 注意到

$$\begin{aligned} (X-EX)^k &= \sum_{i=0}^{k} C_k^i X^i (-EX)^{k-i}, \\ X^k = (EX+X-EX)^k &= \sum_{i=0}^{k} C_k^i (EX)^i (X-EX)^{k-i} \end{aligned}$$

对上面两式取期望即可得证.

母函数的定义

定义 3.14 对任何实数列 $\{p_n\}$, 如果幂级数

$$G(s) = \sum_{n=0}^{\infty} p_n s^n \tag{29}$$

的收敛半径 $s_0 > 0$, 则称 $G(s)$ 为数列 $\{p_n\}$ 的母函数。特别当 $\{p_n\}$ 为某非负整值随机变量 X 的概率分布时，(29) 式至少在区间 $[-1, 1]$ 上绝对收敛且一致收敛，此时有

$$G(s) = E(s^X),$$

称此 $G(s)$ 为随机变量 X 或其概率分布 $\{p_n\}$ 的母函数.

母函数与分布之间的反演公式

- 已知 $G(s)$, 如何确定 $\{p_n\}$?
- 注意到

$$G(0) = p_0, \quad G^{(n)}(s) = n!p_n + \sum_{k=n+1}^{\infty} k(k-1)\cdots(k-n+1)p_k s^{k-n}$$

- 故

$$p_n = \frac{1}{n!}G^{(n)}(0)$$

- 非负整值概率分布 $\{p_n\}$ 与其母函数 $G(s)$ 是一一对应的，母函数可以作为描述这种分布的一种工具.

母函数与随机变量矩的关系

定理 3.23 设非负随机变量 X 的母函数为 $G(s)$, 如果 $E(X)$ 与 $E(X^2)$ 有限，则

$$G'(1) = E(X), \quad G''(1) = E(X^2) - E(X)$$

常见离散型分布的母函数

- Poisson 分布的母函数

$$G(s) = E(s^X) = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(s-1)}$$

- 几何分布的母函数

$$G(s) = E(s^X) = \sum_{k=1}^{\infty} s^k q^{k-1} p = \frac{ps}{1-qs}$$

独立和的母函数

定理 3.24 如果 X, Y 为相互独立的随机变量，它们分别有概率分布 $\{a_n\}, \{b_n\}$ 及对应的母函数 $A(s), B(s)$, 则它们的和 $X+Y$ 的母函数为

$$C(s) = A(s)B(s).$$

更进一步，如果 $X_1, \cdots, X_n$ 相互独立，其对应的母函数分别为 $A_1(s)$, $\cdots, A_n(s)$, 则 $X_1 + \cdots + X_n$ 的母函数为

$$C(s) = \Pi_{i=1}^{n} A_i(s)$$

证明: 注意到由 X, Y 相互独立可知 s^X, s^Y 也相互独立，从而

$$C(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X)E(s^Y) = A(s)B(s)$$

例 3.7 设 X 服从二项分布 $B(n,p)$, 则 X 的母函数为 $G(s) = (q+ps)^n$.

随机个非负整值随机量之和的母函数

定理 3.25 设 $\{X_k\}$ 为相互独立同分布的非负整值随机变量序列，其共同的母函数为 $G(s)$. 如果 N 为另一非负整值随机变量，其母函数为 $F(s)$. 则当 N 与每一个 X_k 均独立时，$X=\sum_{k=1}^{N} X_k$ 的母函数为

$$H(s)=F(G(s)).$$

证明: 由重期望公式可得

$$\begin{aligned}
H(s) &= E(s^X)=E(E(s^X|N))=\sum_{n=0}^{\infty} P(N=n)E(s^{\sum_{k=1}^{N} X_k}|N=n) \\
&= \sum_{n=0}^{\infty} P(N=n)E(s^{\sum_{k=1}^{n} X_k}|N=n)=\sum_{n=0}^{\infty} P(N=n)E(s^{\sum_{k=1}^{n} X_k}) \\
&= \sum_{n=0}^{\infty} P(N=n)[G(s)]^n=E(G(s)^N) \\
&= F(G(s))
\end{aligned}$$

随机个非负整值随机量之和的期望与方差

- 对 $H(s)$ 求导可得

$$H'(s) = F'(G(s))G'(s)$$

- 令 $s = 1$ 并注意到 $G(1) = 1$ 可得

$$E(X) = E(N)E(X_1)$$

- 再对 $H'(s)$ 求导可得

$$H''(s) = F''(G(s))[G'(s)]^2 + F'(G(s))G''(s)$$

- 令 $s = 1$ 可得

$$\begin{aligned} E(X^2) - E(X) &= [E(N^2) - E(N)][E(X_1)]^2 + E(N)[E(X_1^2) - E(X_1)] \\ &= E(N)D(X_1) + E(N^2)[E(X_1)]^2 - E(N)E(X_1) \end{aligned}$$

- 从而

$$D(X) = E(X^2) - [E(X)]^2 = E(N)D(X_1) + D(N)[E(X_1)]^2$$

第二十讲：特征函数

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

复随机变量及其期望

定义 3.15 若 $X(\omega), Y(\omega)$ 为定义在 Ω 上的实值随机变量，则称 $Z(\omega) = X(\omega) + iY(\omega)$ 为复随机变量；称 $\overline{Z} = X(\omega) - iY(\omega)$ 为 $Z(\omega)$ 的复共轭随机变量；称 $|Z| := \sqrt{X^2 + Y^2}$ 为复随机变量 Z 的模.

定义 3.16 若随机变量 X, Y 的期望 $E(X), E(Y)$ 都存在，则复随机变量 Z 的数学期望定义为 $E(Z) := E(X) + iE(Y)$.

- $Z_1 = X_1 + iY_1$ 与 $Z_2 = X_2 + iY_2$ 独立当且仅当 (X_1, Y_1) 与 (X_2, Y_2) 独立；
- $E(e^{iX}) = E(\cos X) + iE(\sin X)$;
- $|e^{iX}| = \sqrt{\cos^2 X + \sin^2 X} = 1$;
- 若 X, Y 独立，则 e^{iX} 与 e^{iY} 也独立.

特征函数的定义

定义 3.17 设 X 是一个随机变量，称

$$\begin{aligned}\varphi(t) &:= E(e^{itX}) = E[\cos(tX)] + iE[\sin(tX)] \\ &= \int_{-\infty}^{+\infty} \cos(tx)dF(x) + i\int_{-\infty}^{+\infty} \sin(tx)dF(x) \\ &= \int_{-\infty}^{+\infty} e^{itx}dF(x), \quad -\infty < t < +\infty\end{aligned}$$

为 X 的特征函数。特别的，

- 如果 X 为离散型随机变量，其分布列为 $p_k = P(X = x_k)$，则

$$\varphi(t) = \sum_{k=1}^{+\infty} e^{itx_k}p_k, \quad -\infty < t < +\infty;$$

- 如果 X 为连续型随机变量，其分布密度为 $p(x)$，则

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}p(x)dx, \quad -\infty < t < +\infty;$$

特征函数的性质 I

1. 由 $\varphi(t) = E[\cos(tX)] + iE[\sin(tX)]$ 知，特征函数总是存在的;
2. $\varphi(0) = 1, |\varphi(t)| \le 1$: $\varphi(0) = E(1) = 1$, 而

$$\begin{aligned} |\varphi(t)| &= |E[\cos(tX)] + iE[\sin(tX)]| = \left[(E[\cos(tX)])^2 + (E[\sin(tX)])^2\right]^{1/2} \\ &\le \left[E[\cos^2(tX)] + E[\sin^2(tX)]\right]^{1/2} = 1; \end{aligned}$$

3. $\varphi(-t) = \overline{\varphi(t)}$: $\varphi(-t) = E[\cos(tX)] - iE[\sin(tX)] = \overline{\varphi(t)}$;
4. 若 $Y = aX + b$, 其中 a, b 为常数，则 $\varphi_Y(t) = e^{ibt}\varphi_X(at)$, 事实上,

$$\begin{aligned} &\varphi_Y(t) = E[e^{it(aX+b)}] = E[\cos(t(aX+b))] + iE[\sin(t(aX+b))] \\ &= E[\cos(taX)\cos(tb) - \sin(taX)\sin(tb)] + iE[\sin(taX)\cos(tb) + \cos(taX)\sin(tb)] \\ &= \cos(tb)\big(E[\cos(taX)] + iE[\sin(taX)]\big) + i\sin(tb)\big(E[\cos(taX)] + iE[\sin(taX)]\big) \\ &= e^{itb}Ee^{itaX} = e^{itb}\varphi_X(at) \end{aligned}$$

特征函数的性质 II

5. 若 X, Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$, 事实上，

$$
\begin{aligned}
&\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX}e^{itY}) \\
&= E[(\cos(tX) + i\sin(tX))(\cos(tY) + i\sin(tY))] \\
&= E[(\cos(tX) + i\sin(tX))\cos(tY)] + iE[(\cos(tX) + i\sin(tX))\sin(tY)] \\
&= E[\cos(tY)]E[e^{itX}] + iE[\sin(tY)]E[e^{itX}] \\
&= E(e^{itX})E(e^{itY})
\end{aligned}
$$

6. 更一般的，若 $X_1, X_2, \cdots, X_n$ 相互独立，则对于 $Y = X_1 + \cdots + X_n$ 的特征函数有

$$\varphi_Y(t) = \Pi_{k=1}^n \varphi_{X_k}(t)$$

特征函数的性质 III

7. 如果 $E(X^k)$ 存在，则

$$\varphi^{(k)}(t) = i^k E[X^k e^{itX}], \quad \varphi^{(k)}(0) = i^k E[X^k]$$

事实上，

$$\frac{d^k}{dt^k}\varphi(t) = \frac{d^k}{dt^k}E[e^{itX}] = E[\frac{d^k}{dt^k}e^{itX}] = E[(iX)^k e^{itX}] = i^k E[X^k e^{itX}]$$

特征函数的性质 IV

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

$$
\begin{aligned}
&|\varphi(t+h)-\varphi(t)| = |E(e^{i(t+h)X})-E(e^{itX})| = |E[e^{itX}(e^{ihX}-1)]| \\
&\le E|e^{itX}(e^{ihX}-1)| = E|e^{ihX}-1| = \int_{-\infty}^{+\infty}|e^{ihx}-1|dF_X(x) \\
&= \int_{|x|>A}|e^{ihx}-1|dF_X(x)+\int_{-A}^{A}|e^{ihx}-1|dF_X(x) \\
&\le 2\int_{|x|>A}dF_X(x)+\int_{-A}^{A}|hx|dF_X(x) \\
&\le 2P(|X|>A)+hAP(|X|\le A) \le 2P(|X|>A)+hA
\end{aligned}
$$

选取 A 使得 $P(|X|>A)\le\epsilon/4$, $h\le\delta=\dfrac{\epsilon}{2A}$, 则只要 $|h|\le\delta$ 必有

$$|\varphi(t+h)-\varphi(t)|\le 2P(|X|>A)+hA\le\epsilon$$

特征函数的性质 V

9. $\varphi(t)$ 是非负定函数，即对任意的正整数 n 及 n 个实数 $t_1, \cdots, t_n$ 及 n 个复数 $z_1, \cdots, z_n$，均有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k - t_j)z_k\bar{z}_j \geq 0$$

证明： 由特征函数的定义可得

$$\begin{aligned}\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k - t_j)z_k\bar{z}_j &= \sum_{k=1}^{n}\sum_{j=1}^{n}z_k\bar{z}_jE[e^{i(t_k-t_j)X}] = \sum_{k=1}^{n}\sum_{j=1}^{n}E[z_ke^{it_kX}\bar{z}_j\overline{e^{it_jX}}] \\ &= E[\sum_{k=1}^{n}z_ke^{it_kX}\sum_{j=1}^{n}\overline{z_je^{it_jX}}] = E[|\sum_{k=1}^{n}z_ke^{it_kX}|^2] \geq 0\end{aligned}$$

常用分布的特征函数 I

- 单点分布 $P(X=a)=1$: $\varphi(t)=e^{ita}$;
- 两点分布 $P(X=x)=p^x(1-p)^{1-x}, x=0,1$: $\varphi(t)=pe^{it}+(1-p)$;
- 泊松分布:$\varphi(t)=\sum_{k=0}^{+\infty} e^{ikt}\dfrac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda e^{it}}e^{-\lambda}=e^{\lambda(e^{it}-1)}$;
- 均匀分布 $U(a,b)$:$\varphi(t)=\int_a^b \dfrac{e^{itx}}{b-a}dx=\dfrac{e^{ibt}-e^{iat}}{it(b-a)}$;
- 标准正态分布:

$$
\begin{aligned}
\varphi(t) &= \int_{-\infty}^{+\infty} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\
&= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{+\infty}\cos(tx)e^{-\frac{x^2}{2}}dx + i\int_{-\infty}^{+\infty}\sin(tx)e^{-\frac{x^2}{2}}dx\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(tx)e^{-\frac{x^2}{2}}dx
\end{aligned}
$$

常用分布的特征函数 II

故

$$\begin{aligned}\varphi'(t) &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x\sin(tx)e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sin(tx)de^{-\frac{x^2}{2}} \\ &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t\cos(tx)e^{-\frac{x^2}{2}}dx = -t\varphi(t)\end{aligned}$$

从而

$$\frac{d}{dt}[\varphi(t)e^{t^2/2}] = [\varphi'(t)+t\varphi(t)]e^{t^2/2} = 0$$

故 $\varphi(t)e^{t^2/2} = C$, 再利用 $C = \varphi(0) = 1$ 得

$$\varphi(t) = e^{-t^2/2}.$$

常用分布的特征函数 III

- 指数分布:

$$\varphi(t) = \int_0^\infty e^{itx}\lambda e^{-\lambda x}dx = \left(1 - \frac{it}{\lambda}\right)^{-1}$$

- 二项分布: $Y \sim B(n,p)$. 注意到 $Y = X_1 + \cdots + X_n$, 其中 X_i 为独立同分布的随机变量，且 $X_i \sim B(1,p)$, 故

$$\varphi_Y(t) = \Pi_{k=1}^n (pe^{it} + q) = (pe^{it} + q)^n, \quad q = 1 - p$$

- 一般正态分布: $Y \sim N(\mu, \sigma^2)$, 则由 $X := \dfrac{Y - \mu}{\sigma} \sim N(0,1)$ 及 $Y = \sigma X + \mu$ 可得

$$\varphi_Y(t) = e^{i\mu t}\varphi_X(\sigma t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$$

常用分布的特征函数 IV

- 伽玛分布:$Y \sim \Gamma(n, \lambda)$. 注意到 $Y = X_1 + \cdots + X_n$, 其中 X_i 独立同指数分布即 $X_i \sim \exp(\lambda)$, 故

$$\varphi_Y(t) = \Pi_{k=1}^{n}(1 - \frac{it}{\lambda})^{-1} = (1 - \frac{it}{\lambda})^{-n}$$

  更进一步，当 $Y \sim \Gamma(\alpha, \lambda)$, 我们有

$$\varphi_Y(t) = (1 - \frac{it}{\lambda})^{-\alpha}$$

- $\chi^2(n)$ 分布：因为 $\chi^2(n) = \Gamma(n/2, 1/2)$, 故其特征函数为

$$\varphi(t) = (1 - 2it)^{-n/2}$$

例 3.8 利用特征函数的方法求 $\Gamma(\alpha,\lambda)$ 分布的数学期望与方差.

解: Γ 分布的特征函数及其导数分别为

$$\begin{aligned}
\varphi(t) &= \left(1-\frac{it}{\lambda}\right)^{-\alpha} \\
\varphi'(t) &= \frac{i\alpha}{\lambda}\left(1-\frac{it}{\lambda}\right)^{-\alpha-1}, \quad \varphi'(0)=\frac{i\alpha}{\lambda} \\
\varphi''(t) &= \frac{i^2\alpha(\alpha+1)}{\lambda^2}\left(1-\frac{it}{\lambda}\right)^{-\alpha-2}, \quad \varphi''(0)=-\frac{\alpha(\alpha+1)}{\lambda^2}
\end{aligned}$$

故由 $\varphi^{(k)}(0)=i^kE(X^k)$ 知,

$$\begin{aligned}
E(X) &= \frac{\varphi'(0)}{i}=\frac{\alpha}{\lambda}, \quad E(X^2)=\frac{\varphi''(0)}{i^2}=\frac{\alpha(\alpha+1)}{\lambda^2} \\
D(X) &= E(X^2)-(EX)^2=\frac{\alpha(\alpha+1)}{\lambda^2}-\frac{\alpha^2}{\lambda^2}=\frac{\alpha}{\lambda^2}
\end{aligned}$$

特征函数与分布函数之间的关系：相互唯一决定

定理 3.26 (逆转公式) 设 $F(x)$ 与 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$ 有

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t) dt$$

定理 3.27 (唯一性定理) 随机变量的分布函数由其特征函数唯一决定.

证明: 在逆转公式中令 $x_2 = x, x_1 = y \to -\infty$ 可得

$$F(x) = \lim_{y\to-\infty} \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ity} - e^{-itx}}{it} \varphi(t) dt$$

而由分布函数本身是右连续可知，分布函数任一点的取值可由连续点上的值唯一决定.

狄利克雷积分 I

引理 3.1 (狄利克雷积分) 令

$$I(a,T)=\int_0^T \frac{\sin at}{t}dt, \quad \operatorname{sgn}\{a\}:=\begin{cases} -1, & a<0, \\ 0, & a=0, \\ 1, & a>0, \end{cases}$$

则

$$\lim_{T\to+\infty} I(a,T)=\lim_{T\to+\infty}\int_0^T \frac{\sin at}{t}dt=\frac{\pi}{2}\operatorname{sgn}\{a\}$$

证明: 注意到如果作积分换元 $y=at$ 可得

$$\lim_{T\to+\infty} I(a,T)=\operatorname{sgn}\{a\}\lim_{T\to+\infty} I(1,T)$$

狄利克雷积分 II

只需证明

$$\lim_{T\to+\infty} I(1,T) = \lim_{T\to+\infty}\int_0^T \frac{\sin t}{t}dt = \frac{\pi}{2}.$$

注意到 $\frac{1}{t} = \int_0^{+\infty} e^{-ut}du$, 故

$$\begin{aligned} I(1,T) &= \int_0^T\int_0^{+\infty} e^{-ut}\sin t du dt = \int_0^{+\infty}\left[\int_0^T e^{-ut}\sin t dt\right]du \\ &= \int_0^{+\infty}\left[\frac{1}{1+u^2} - \frac{1}{1+u^2}e^{-uT}(u\sin T + \cos T)\right]du \\ &= \frac{\pi}{2} - \int_0^{+\infty}\frac{u\sin T + \cos T}{1+u^2}e^{-uT}du \to \frac{\pi}{2} \end{aligned}$$

逆转公式的证明

证明: 考虑积分

$$
\begin{aligned}
J(T) &= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\varphi(t)dt \\
&= \frac{1}{2\pi}\int_{-T}^{T}\left[\int_{-\infty}^{+\infty}\frac{e^{-itx_1}-e^{-itx_2}}{it}e^{itx}dF(x)\right]dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\int_{-T}^{T}\frac{e^{it(x-x_1)}-e^{it(x-x_2)}}{it}dt\right]dF(x) \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\left[\int_{0}^{T}\frac{\sin t(x-x_1)}{t}dt-\int_{0}^{T}\frac{\sin t(x-x_2)}{t}dt\right]dF(x) \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\left[I(x-x_1,T)-I(x-x_2,T)\right]dF(x) \\
&\to \frac{1}{\pi}\frac{\pi}{2}\int_{-\infty}^{+\infty}\left[\mathrm{sgn}\{x-x_1\}-\mathrm{sgn}\{x-x_2\}\right]dF(x) \qquad T\to+\infty \\
&= \int_{x_1}^{x_2}dF(x)=F(x_2)-F(x_1)
\end{aligned}
$$

逆转公式的一般情形

定理 3.28 (逆转公式一般情形) 设 $F(x)$ 与 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对任意的 $x_1, x_2 \in R$ 有

$$\frac{F(x_2-0)+F(x_2)}{2}-\frac{F(x_1-0)+F(x_1)}{2}=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\varphi(t)dt$$

连续型分布情形下的逆转公式

定理 3.29 若 X 为连续型随机变量，其密度函数为 $p(x)$, 其特征函数为

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} p(x) dx,$$

如果 $\int_{-\infty}^{+\infty} |\varphi(t)| dt < +\infty$, 则

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt$$

证明: 由分布密度的定义可知

$$\begin{aligned} p(x) &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h\to 0} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-itx} - e^{-it(x+h)}}{it \cdot h} \varphi(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lim_{h\to 0} \frac{e^{-ith} - 1}{-it \cdot h} e^{-itx} \varphi(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt \end{aligned}$$

第二十一讲：随机变量的收敛性

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　数学学院

随机变量的几种收敛性

- 几乎必然收敛: 如果

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1,$$

则称随机变量列 $\{X_n\}$ 几乎必然收敛到 X, 记作 $X_n \xrightarrow{a.e.} X$, 或 $X_n \to X, a.s.$

- 依概率收敛: 如果对任意的 $\epsilon > 0$ 有

$$\lim_{n\to\infty} P(|X_n(\omega) - X(\omega)| > \epsilon) = 0,$$

则称 X_n 依概率收敛到 X, 记作 $X_n \xrightarrow{P} X$.

- r 阶平均收敛: 设 $r > 0$ 为常数，如果随机变量 X 与 $X_n(n \geq 1)$ 的 r 阶矩皆有限，并且有

$$\lim_{n\to\infty} E|X_n - X|^r = 0,$$

则称 $\{X_n\}$ 为 r 阶平均收敛到 X, 简称 r 阶收敛，记作 $X_n \xrightarrow{L^r} X$.

几乎必然收敛的刻划与性质

- 若记 $E := \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$, 则

$$E = \cap_{m=1}^{\infty} \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{m}\} \in \mathcal{F};$$

- $X_n \xrightarrow{a.s.} X$ 等价于 $P(E) = 1$ 或 $P(\bar{E}) = 0$;
- $X_n \xrightarrow{a.s.} X$ 的充分必要条件是对任意的 $\epsilon > 0$ 有

$$\lim_{k\to\infty} P(\cup_{n=k}^{\infty} |X_n - X| \geq \epsilon) = 0$$

- 若 $X_n \xrightarrow{a.s.} X$, 则 $X_n \xrightarrow{P} X$.

依概率收敛不能推出几乎必然收敛

例 4.1 取 $\Omega=[0,1),\mathcal{F}$ 为区间 $[0,1)$ 上的 Borel 集类，P 为 Lebesgue 测度。令

$$X_{11}(\omega)=1;\quad X_{21}(\omega)=\begin{cases}1, & \omega\in[0,\frac{1}{2}),\\ 0, & \omega\in[\frac{1}{2},1);\end{cases}\qquad X_{22}(\omega)=\begin{cases}1, & \omega\in[\frac{1}{2},1),\\ 0, & \omega\in[0,\frac{1}{2});\end{cases}$$

一般地，对每个自然数 k, 将 $[0,1)$ 等分为 k 个区间，并取

$$X_{ki}(\omega)=\begin{cases}1, & \omega\in[\frac{i-1}{k},\frac{i}{k}),\\ 0, & \text{其他},\end{cases}\quad i=1,\cdots,k$$

定义

$$Y_1=X_{11},Y_2=X_{21},Y_2=X_{22},Y_3=X_{31},\cdots$$

对于随机变量序列 Y_n, 对任意的 $\epsilon>0$ 有

$$P(|X_{ki}(\omega)|\geq\epsilon)\leq\frac{1}{k}\to 0,$$

但是收敛的点集 $E=\emptyset$, 即 Y_n 不几乎必然收敛.

r 阶收敛与依概率收敛

引理 4.1 (Markov 不等式) 设随机变量 X 的 r 阶矩有限，则对任意 $\epsilon > 0$ 有

$$P(|X| \geq \epsilon) \leq \frac{1}{\epsilon^r} E|X|^r$$

定理 4.1 若 $X_n \xrightarrow{L^r} X$, 则 $X_n \xrightarrow{P} X$. 反之不真.

例 4.2 仍取区间 $(0,1)$ 上的几何概率空间 $(\Omega, \mathcal{F}, P)$. 固定 $r > 0$, 令

$$X_n(\omega) = \begin{cases} n^{1/r}, & \omega \in (0, 1/n), \\ 0, & \omega \in [1/n, 1), \end{cases} \qquad X(\omega) = 0.$$

易知 $X_n \xrightarrow{a.s.} X$, 从而 $X_n \xrightarrow{P} X$, 但 $X_n \not\xrightarrow{L^r} X$.

依分布收敛 (弱收敛)

定义 4.1 设随机变量 $X, X_1, X_2, \cdots$, 的分布函数分别为 $F(x), F_1(x), F_2(x), \cdots$,, 如果对 $F(x)$ 的任一连续点 x 均有

$$\lim_{n\to} F_n(x) = F(x),$$

则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F(x)$. 此时也称 $\{X_n\}$ 按分布收敛于 X, 记作 $X_n \xrightarrow{W} X$.

注 4.1 后面我们用 C_F 表示函数 $F(x)$ 的连续点集，即

$$C_F := \{x : x\text{为}F(x)\text{的连续点}\}$$

为何要求分布函数在连续点收敛，而不是每一点收敛

例 4.3 考虑随机变量序列 $\{X_n\}$ 服从如下退化分布

$$P(X_n = \frac{1}{n}) = 1, n = 1, 2, \cdots,$$

它们的分布函数分别为 $F_n(x) = \begin{cases} 0, & x < \frac{1}{n} \\ 1, & x \geq \frac{1}{n} \end{cases}$

- 则在点点收敛要求下，$\{F_n(x)\}$ 的极限函数为

$$g(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

  显然上述极限不满足右连续性，即 $g(x)$ 不是一个分布函数.
- 因为 $F_n(x)$ 在 $x = 1/n$ 处有跳跃，故当 $n \to \infty$ 时，跳跃点的位置趋于 0,很自然的认为 $\{F_n(x)\}$ 应该收敛于点 $x = 0$ 的退化分布，即 $F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$. 但是对任意的 n, 有 $F_n(0) = 0$, 而 $F(0) = 1$, 所以 $\lim_{n\to\infty} F_n(0) = 0 \neq 1 = F(0)$, 而这个不收敛的点恰为 $F(x)$ 的不连续点.

依概率收敛与依分布收敛的关系

定理 4.2 若 $X_n \xrightarrow{P} X$, 则 $X_n \xrightarrow{W} X$.

证明: 设随机变量 $X, X_1, X_2, \cdots$, 的分布函数分别为 $F(x), F_1(x), F_2(x), \cdots$, 只需证在 $F(x)$ 的连续点上均有 $F_n(x) \to F(x)$ 即可. 而要证明这一点我们只需要说明, 对任意的 x 均有

$$F(x-0) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x+0)$$

注意到, 对任意的 $x' < x$,

$$\begin{aligned} \{X \le x'\} &= \{X_n \le x, X \le x'\} \cup \{X_n > x, X \le x'\} \\ &\subset \{X_n \le x\} \cup \{|X_n - X| \ge x - x'\} \end{aligned}$$

从而有 $F(x') \le F_n(x) + P(|X_n - X| \ge x - x')$. 再由 $X_n \xrightarrow{P} X$ 知

$$F(x') \le \liminf_{n\to\infty} F_n(x) \Rightarrow F(x-0) \le \liminf_{n\to\infty} F_n(x)$$

类似可证, 当 $x'' > x$ 时, $\limsup_{n\to\infty} F_n(x) \le F(x+0)$.

依分布收敛无法推出依概率收敛

例 4.4 设随机变量 X 的分布列为

$$P(X=-1)=P(X=1)=1/2$$

令 $X_n=-X$, 则 X_n 与 X 同分布，即 X_n 与 X 有相同的分布，故 $X_n \xrightarrow{W} X$, 但对任意的 $\epsilon \in (0,2)$,

$$P(|X_n-X| \geq \epsilon)=P(2|X| \geq \epsilon)=1 \nrightarrow 0$$

若 X_n 依分布收敛至常数 c，则 X_n 概率收敛至常数

定理 4.3 若 c 为常数，则 $X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{W} c$

只需证明充分性即可。记 X_n 的分布函数为 $F_n(x), n = 1, 2, \cdots$，常数 c 的分布函数为 $F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geq c. \end{cases}$ 故对任意的 $\epsilon > 0$,

$$\begin{aligned} P(|X_n - c| \geq \epsilon) &= P(X_n \geq c + \epsilon) + P(X_n \leq c - \epsilon) \\ &\leq P(X_n > c + \epsilon/2) + P(X_n \leq c - \epsilon) \\ &= 1 - F_n(c + \epsilon/2) + F_n(c - \epsilon) \\ &\to 1 - F(c + \epsilon/2) + F(c - \epsilon) = 0 \end{aligned}$$

从而 $X_n \xrightarrow{P} c$.

弱收敛的充分条件

定理 4.4 设 $F(x), F_n(x)(n \geq 1)$ 均为分布函数，如果对于 $F(x)$ 的连续点集 C_F 的某个稠密子集 D 上的一切 x 均有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则 $F_n \xrightarrow{W} F$.

Helly 第一定理

定理 4.5 (Helly 第一定理) 任意分布函数列 $F_n(x)$ 均存在子列 $F_{n(k)}(x)$ 以及一个有界、非降、右连续的函数 $F(x)$ 使得对任意的 $x \in C_F$, 均有

$$\lim_{k\to\infty} F_{n(k)}(x) = F(x).$$

注 4.2 极限函数 $F(x)$ 不一定为分布函数. 比如, 若 $a+b+c=1$, $G(x)$ 为分布函数, 令 $F_n(x) = a1_{(x\geq n)} + b1_{(x\geq -n)} + cG(x)$, 则易知

$$F_n(x) \to F(x) = b + cG(x),$$

但

$$\lim_{x\downarrow-\infty} F(x) = b \neq 0, \quad \lim_{x\uparrow\infty} F(x) = b + c = 1 - a \neq 1$$

Helly 第一定理证明

- 令 $q_1, q_2, \ldots$ 为有理数排列, $m_0(j) \equiv j$. 由于对每个 k,

$$F_m(q_k) \in [0,1], \quad m \geq 1.$$

因此, 存在 $m_{k-1}(j)$ 的子列 $m_k(i)$ 使得

$$\lim_{i\to\infty} F_{m_k(i)}(q_k) = G(q_k).$$

- 令 $F_{n(k)}(x) := F_{m_k(k)}(x)$, 则

$$\lim_{k\to\infty} F_{n(k)}(q) = G(q), \quad \forall q \in \mathbb{Q}.$$

- 令 $F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}$, 则易知 $F(x)$ 右连续. 事实上,

$$\begin{aligned}\lim_{x_n\downarrow x} F(x_n) &= \inf\left\{G(q) : q \in \mathbb{Q}, q > x_n \text{ 对某个} n\right\} \\ &= \inf\{G(q) : q \in \mathbb{Q}, q > x\} = F(x)\end{aligned}$$

Helly 第一定理证明 II

- 令 $x \in C_F$ 并取 $r_1, r_2, r_3 \in \mathbb{Q}$ 使得 $r_1 < r_2 < x < r_3$, 则

$$F(x) - \epsilon < F(r_1) \leq F(r_2) \leq F(x) \leq F(r_3) < F(x) + \epsilon$$

- 由于

$$F_{n(k)}(r_2) \to G(r_2) \geq F(r_1), \quad F_{n(k)}(r_3) \to G(r_3) \leq F(r_3),$$

因此, 当 k 充分大时,

$$F(x) - \epsilon < F_{n(k)}(r_2) \leq F_{n(k)}(x) \leq F_{n(k)}(r_3) < F(x) + \epsilon$$

- 综上可知

$$\lim_{k \to \infty} F_{n(k)}(x) = F(x), \quad x \in C_F.$$

胎紧 (tight)

定义 4.2 称分布函数列 $\{F_n(x)\}_{n\geq 1}$ 为胎紧的 (tight), 如果对任意的 $\epsilon > 0$, 存 M_ϵ 使得

$$\limsup_{n\to\infty}\left(1 - F_n\left(M_\epsilon\right) + F_n\left(-M_\epsilon\right)\right) = \limsup_{n\to\infty} P(|X_n| \geq M_\epsilon) \leq \epsilon$$

定理 4.6 分布函数列 $\{F_n(x)\}_{n\geq 1}$ 的子序列收敛到某个分布函数 $F(x)$ 当且仅当 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.

证明:

- 假设 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的, 且 $F_{n(k)}(x) \to F(x)$. 下证 $F(x)$ 为分布函数,即证: $F(+\infty) - F(-\infty) = 1$.
- 令 $r, s \in C_F$ 且 $r < -M_\epsilon, s > M_\epsilon$, 则由 $F_n(r) \to F(r)$, $F_n(s) \to F(s)$ 可得

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{k\to\infty}\left(1 - F_{n(k)}(s) + F_{n(k)}(r)\right) \\ &\leq \limsup_{n\to\infty}\left(1 - F_n\left(M_\epsilon\right) + F_n\left(-M_\epsilon\right)\right) \leq \epsilon \end{aligned}$$

- 令 $r \to -\infty, s \to \infty$ 后再令 $\epsilon \to 0$ 可得 $F(+\infty) - F(-\infty) = 1$

定理证明续

- 若极限函数是分布函数, 下面用反证法证明 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.
- 若 $\{F_n(x)\}_{n\geq 1}$ 不是胎紧的, 则存在 $\epsilon > 0$ 以及子列 $n(k) \to \infty$ 使得对任意的 $n(k)$ 均有

$$1 - F_{n(k)}(k) + F_{n(k)}(-k) \geq \epsilon,$$

- 由 Helly 第一定理可知, $\{F_{n(k)}(x)\}_{k\geq 1}$ 存在子列 $\{F_{n(k_j)}\}_{j\geq 1} \to F(x)$
- 令 $r, s \in C_F$, $r < 0 < s$, 则

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{j\to\infty} \left(1 - F_{n(k_j)}(s) + F_{n(k_j)}(r)\right) \\ &\geq \liminf_{j\to\infty} \left(1 - F_{n(k_j)}\left(k_j\right) + F_{n(k_j)}\left(-k_j\right)\right) \geq \epsilon \end{aligned}$$

- 令 $s \to \infty, r \to -\infty$ 可知 $F(x)$ 不是分布函数, 与假设矛盾.

Helly 第二定理

定理 4.7 (Helly 第二定理) 设分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 当且仅当对任何有界连续函数 $g(x)$ 有

$$\int_{\mathbb{R}} g(x)dF_n(x) \to \int_{\mathbb{R}} g(x)dF(x).$$

或等价的

$$E[g(X_n)] \to E[g(X)].$$

一个引理

定理 4.8 若分布函数列 $\{F_n(x)\}_{n\geq 1}$ 弱收敛于分布函数 $F(x)$, 则存在分布函数为 $F_n(x)$ 的随机变量序列 $\{Y_n\}_{n\geq 1}$ 及分布函数为 $F(x)$ 的随机变量 Y, 使得 $Y_n \to Y, a.s.$.

证明:

- 令 $\Omega = (0,1), \mathcal{F} = \mathcal{B}((0,1))$, P 为 $(0,1)$ 上的 Lebesgue 测度.
- 令

$$
\begin{aligned}
&Y_n(y) = F_n^{-1}(y) := \sup\{x : F_n(x) < y\} \\
&Y(y) = F^{-1}(y) := \sup\{x : F(x) < y\} \\
&a_y = \sup\{x : F(x) < y\}, \quad b_y = \inf\{x : F(x) > y\} \\
&\Omega_0 = \{y : (a_y, b_y) = \emptyset\}
\end{aligned}
$$

- 由于 (a_y, b_y) 互不相交且每个非空区间均包含不同的有理数, 因此可知 $\Omega - \Omega_0$ 中的元素个数可数, 从而 $P(\Omega - \Omega_0) = 0$.

引理证明 (续)

- 下证, 对任意的 $y \in \Omega_0$, 我们有 $F_n^{-1}(y) \to F^{-1}(y)$.
- 首先, 我们有 $\liminf_{n\to\infty} F_n^{-1}(y) \geq F^{-1}(y)$,
  - 令 $x \in C_F$ 使得 $x < F^{-1}(y)$, 则由 $y \in \Omega_0$ 知 $F(x) < y$
  - 由 $x \in C_F$ 以及 $F_n(x) \to F(x)$ 可知, 当 n 充分大时, $F_n(x) < y$, 也即 $F_n^{-1}(y) \geq x$
  - 令 $n \to \infty$ 取下极限后再令 $x \to F^{-1}(y)$ 可得本结果.
- 其次, 我们可证 $\limsup_{n\to\infty} F_n^{-1}(y) \leq F^{-1}(y)$
  - 令 $x \in C_F$ 使得 $x > F^{-1}(y)$, 则由 $y \in \Omega_0$ 知 $F(x) > y$
  - 由 $x \in C_F$ 以及 $F_n(x) \to F(x)$ 可知, 当 n 充分大时, $F_n(x) > y$, 也即 $F_n^{-1}(y) \leq x$
  - 令 $n \to \infty$ 取上极限后再令 $x \to F^{-1}(y)$ 可得本结果.

Helly 第二定理证明

- 必要性
  - 由引理可知, 存在 $\{Y_n\}_{n\geq 1}$ 及 Y 使得 $Y_n \to Y a.s.$
  - 由于 $g(x)$ 是连续函数, 从而 $g(Y_n) \to g(Y), a.s.$
  - 从而由控制收敛定理可知

$$E[g(X_n)] = E[g(Y_n)] \to E[g(Y)] = E[g(X)],$$

也即

$$\int_{\mathbb{R}} g(x)dF_n(x) \to \int_{\mathbb{R}} g(x)dF(x).$$

Helly 第二定理证明

- 充分性
  - 令

$$g_{x,\epsilon}(y) = \begin{cases} 1 & y \le x \\ 0 & y \ge x+\epsilon \\ \text{linear} & x \le y \le x+\epsilon \end{cases}$$

  - 由 $g_{x,\epsilon}$ 的定义及连续性可知

$$\limsup_{n\to\infty} P(X_n \le x) \le \limsup_{n\to\infty} Eg_{x,\epsilon}(X_n) = Eg_{x,\epsilon}(X) \le P(X \le x+\epsilon)$$

$$\liminf_{n\to\infty} P(X_n \le x) \ge \liminf_{n\to\infty} Eg_{x-\epsilon,\epsilon}(X_n) = Eg_{x-\epsilon,\epsilon}(X) \ge P(X \le x-\epsilon)$$

  - 令 $\epsilon \to 0$ 可得

$$\lim \sup_{n\to\infty} P(X_n \le x) \le P(X \le x),$$

$$\lim \inf_{n\to\infty} P(X_n \le x) \ge P(X < x) \overset{x\in C_F}{=} P(X \le x)$$

  - 由上可得, 对任意的 $x \in C_F$, 我们有

$$\lim_{n\to\infty} F_n(x) = F(x).$$

连续性定理及证明

定理 4.9 (连续性定理) 分布函数列 $\{F_n(x)\}$ 弱收敛到分布函数 $F(x)$ 的充分必要条件是: 相应的特征函数列 $\{\varphi_n(t)\}$ 逐点收敛到相应的特征函数 $\varphi(t)$.

证明:

- 必要性显然. 事实上, 由 $\cos(tx), \sin(tx)$ 为有界连续函数以及 Helly 第二定理可知

$$
\begin{aligned}
\varphi_n(t) &:= \int_{\mathbb{R}} e^{itx} dF_n(x) \\
&= \int_{\mathbb{R}} \cos(tx) dF_n(x) + i \int_{\mathbb{R}} \sin(tx) dF_n(x) \\
&\to \int_{\mathbb{R}} \cos(tx) dF(x) + i \int_{\mathbb{R}} \sin(tx) dF(x) \\
&= \int_{\mathbb{R}} e^{itx} dF(x) =: \varphi(t).
\end{aligned}
$$

连续性定理证明 (续)

下证充分性. 先证 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.

- 注意到

$$u^{-1}\int_{-u}^{u}(1-\varphi_n(t))\,dt = u^{-1}\int_{-u}^{u}\int_{\mathbb{R}}\left(1-e^{itx}\right)dF_n(x)dt$$

$$= u^{-1}\int_{\mathbb{R}}\int_{-u}^{u}\left(1-e^{itx}\right)dtdF_n(x) = 2\int_{\mathbb{R}}(1-\frac{\sin ux}{ux})dF_n(x)$$

$$\geq 2\int_{|x|\geq\frac{2}{u}}(1-\frac{1}{|ux|})dF_n(x) \geq \int_{|x|\geq\frac{2}{u}}dF_n(x) = P(|X_n|\geq\frac{2}{u})$$

- 另一方面,

$$\lim_{n\to\infty}u^{-1}\int_{-u}^{u}(1-\varphi_n(t))\,dt = u^{-1}\int_{-u}^{u}(1-\varphi(t))\,dt \xrightarrow{u\to 0} 0$$

- 从而

$$P(|X_n|\geq\frac{2}{u}) \leq u^{-1}\int_{-u}^{u}(1-\varphi_n(t))\,dt \leq 2\epsilon$$

连续性定理证明 (续)

其次, 证明对任意的有界连续函数 $g(x)$ 均有 $E[g(X_n)] \to E[g(X)]$.

- 首先注意到, 对于 $\{F_n(x)\}_{n\geq 1}$ 的任一子列 $\{F_{n(k)}(x)\}_{k\geq 1}$, 若 $F_{n(k)}(x) \xrightarrow{W} F(x)$, 则分布函数 $F(x)$ 的特征函数必为 $\varphi(t)$.
- 由 $\{F_n(x)\}_{n\geq 1}$ 的胎紧性可知, $\{F_n(x)\}_{n\geq 1}$ 的任一子列 $\{F_{n(k)}(x)\}_{k\geq 1}$ 均存在子列 $\{F_{n(k_j)}(x)\}_{j\geq 1}$ 弱收敛到某个分布函数 $F(x)$, 且 $F(x)$ 的特征函数为 $\varphi(t)$.
- 由于 $\{F_{n(k_j)}(x)\}_{j\geq 1}$ 弱收敛到某个分布函数 $F(x)$, 故由 Helly 第二定理可知, 对任意的有界连续函数 $g(x)$ 均有

$$\int g(x)dF_{n(k_j)}(x) \to \int g(x)dF(x)$$

- 故综上可知, $\int g(x)dF_n(x)$ 的任一子列 $\int g(x)dF_{n(k)}(x)$ 均存在子列 $\int g(x)dF_{n(k_j)}(x) \to \int g(x)dF(x)$. 从而, $\int g(x)dF_n(x) \to \int g(x)dF(x)$, 即

$$E[g(X_n)] \to E[g(X)].$$

第二十二讲：大数定律

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　数学学院

概率及频率

- 记 S_n 为 n 重伯努利试验中事件 A 出现次数，称 $\frac{S_n}{n}$ 为事件 A 出现的频率；
- 记 p 为一次试验中 A 发生的概率；
- 概率是频率的稳定值，何谓稳定值？似乎与频率的极限有关；
- 考虑 $|\frac{S_n}{n}-p|$, 如果能类似于函数列极限情形即对任意的 $\epsilon>0$, 存在 N 使得当 $n\geq N$ 时

$$|\frac{S_n}{n}-p|<\epsilon, \forall\omega\in\Omega.$$

- 但是上述情形无法做到，事实上，对任意的 $\epsilon<1-p$,

$$\begin{aligned}
&P(|\frac{S_n}{n}-p|<\epsilon)\leq P(\frac{S_n}{n}-p<\epsilon)\leq P(\frac{S_n}{n}-p<1-p)\\
&=P(\frac{S_n}{n}<1)=1-P(\frac{S_n}{n}=1)=1-P(X_1=1,\cdots,X_n=1)\\
&=1-p^n<1
\end{aligned}$$

伯努利大数定律

定理 4.10 设 S_n 为 n 重伯努利试验中事件 A 出现次数，p 为一次试验中 A 发生的概率，则对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - p| < \epsilon) = 1, \quad \text{或等价的} \quad \lim_{n\to\infty} P(|\frac{S_n}{n} - p| \geq \epsilon) = 0$$

证明: 因为 $S_n \sim B(n,p)$, 故 $S_n \sim B(n,p), E(\frac{S_n}{n}) = p, D(\frac{S_n}{n}) = \frac{p(1-p)}{n}$. 从而由切比雪夫不等式可得

$$1 \geq P(|\frac{S_n}{n} - p| < \epsilon) \geq 1 - \frac{D(\frac{S_n}{n})}{\epsilon^2} = 1 - \frac{p(1-p)}{n\epsilon^2} \to 1 \quad n \to \infty$$

故

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - p| < \epsilon) = 1$$

伯努利大数定律的含义

- 随着 n 的增大，事件 A 发生的频率 $\frac{S_n}{n}$ 与其概率 p 的偏差大于预先给定的精度 ϵ 的可能性越来越小，要多小有多小，这便是频率稳定于概率的含义;
- 事实上，由切比雪夫不等式可得

$$P(|\frac{S_n}{n}-p|\geq\epsilon)\leq\frac{D(\frac{S_n}{n})}{\epsilon^2}=\frac{p(1-p)}{\epsilon^2}\frac{1}{n}$$

- 伯努利大数定律提供了用频率来确定概率的理论依据;
- 比如要估计某种产品的不合格品率 p, 则可从该种产品中随机抽取 n 件，当 n 很大时，这 n 件产品中的不合格品的比例可作为不合格品率 p 的估计值.

伯努利大数定律的另一形式

- 注意到，在伯努利大数定律中，如果记 $X_i = \begin{cases} 1, & 第i试验A发生 \\ 0, & 第i试验A不发生 \end{cases}$，则 $S_n = \sum_{i=1}^n X_i$，并且

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i, p = E(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n E(X_i)$$

- 伯努利大数定律的结论为：对任意的 $\epsilon > 0$，有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^n [X_i - E(X_i)]| < \epsilon) = 1$$

也即

$$\frac{1}{n}\sum_{i=1}^n [X_i - E(X_i)] \xrightarrow{P} 0$$

(强) 大数定律的一般形式

定义 4.3 设有一随机变量序列 $\{X_n\}$, 如果对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)]| < \epsilon) = 1 \text{ 即 } \frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)] \xrightarrow{P} 0$$

则称该随机变量序列 $\{X_n\}$ 满足大数定律。如果进一步有

$$\frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)] \xrightarrow{a.s.} 0$$

则称 $\{X_n\}$ 满足强大数定律.

切比雪夫大数定律

定理 4.11 设 $\{X_n\}$ 为一列两两不相关的随机变量序列，若每个 X_i 的方差存在且有共同的上界 M, 即 $D(X_i) \le M, \forall i \in \mathbb{N}$, 则 $\{X_n\}$ 服从大数定律，即对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \epsilon) = 1$$

特别的，若 $\{X_n\}$ 独立同分布，且 $D(X_n) < \infty$, 则 $\{X_n\}$ 服从大数定律.

证明: 注意到

$$\begin{aligned} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| \ge \epsilon) &\le \frac{D(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\epsilon^2} \\ &= \frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \\ &\le \frac{1}{\epsilon^2}\frac{M}{n} \to 0 \end{aligned}$$

马尔科夫大数定律

定理 4.12 设随机变量序列 $\{X_n\}$ 满足如下马尔科夫条件

$$\frac{1}{n^2}D(\sum_{i=1}^{n}X_i) \to 0$$

则 $\{X_n\}$ 服从大数定律，即对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}E(X_i)| < \epsilon) = 1$$

辛钦大数定律

定理 4.13 设 $\{X_n\}$ 为一列独立同分布随机变量序列，若每个 X_i 的期望存在，则 $\{X_n\}$ 服从大数定律，即对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \epsilon) = 1.$$

特别的，如果 $E|X_i|^k, k = 1, 2, \cdots$, 存在，则 $\{X_n^k\}$ 服从大数定律.

证明: 设 $\{X_n\}$ 独立同分布，且 $E(X_i) = a, i = 1, \cdots$,, 则只需要证明

$$Y_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} a, \quad n \to \infty$$

由依概率收敛的性质，仅需证 $\varphi_{Y_n}(t) \to e^{iat}$ 即可。事实上，由

$$\begin{aligned}\varphi_{Y_n}(t) &= [\varphi(\frac{t}{n})]^n = [1 + \varphi'(0)\frac{t}{n} + o(\frac{t}{n})]^n = [1 + ia\frac{t}{n} + o(\frac{t}{n})]^n \\ &= \exp\{\frac{\ln(1 + ia\frac{t}{n} + o(\frac{t}{n}))}{1/n}\} \to e^{iat}\end{aligned}$$

Borel-Cantelli 引理

定理 4.14 对于事件列 $\{A_j\}$，有

- 若 $\sum_{n=1}^{\infty} P(A_n) < \infty$，则

$$P(\limsup_{n\to\infty} A_n) = \lim_{k\to\infty} P(\cup_{n=k}^{\infty} A_n) = 0$$

- 若 $\{A_n\}$ 相互独立，则 $\sum_{n=1}^{\infty} P(A_n) = \infty$ 当且仅当

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

证明: 1. 注意到 $\lim_{k\to\infty} P(\cup_{n=k}^{\infty} A_n) \le \lim_{k\to\infty} \sum_{n=k}^{\infty} P(A_n) = 0.$

2. 由于

$$\begin{aligned}
P(\cup_{n=k}^{\infty} A_n) &= \lim_{m\to\infty} P(\cup_{n=k}^{m} A_n) = \lim_{m\to\infty} (1 - P(\cap_{n=k}^{m} \overline{A}_n)) \\
P(\cap_{n=k}^{m} \overline{A}_n) &= \Pi_{n=k}^{m} P(\overline{A}_n) = \Pi_{n=k}^{m} (1 - P(A_n)) \\
&\le \Pi_{n=k}^{m} \exp(-P(A_n)) = \exp(-\sum_{n=k}^{m} P(A_n)) \xrightarrow{m\to\infty} 0
\end{aligned}$$

Borel 强大数定律律

定理 4.15 (Borel 强大数定律) 设 S_n 为 n 重伯努利试验中事件 A 出现次数，p 为一次试验中 A 发生的概率，则

$$\frac{S_n}{n} \xrightarrow{a.s.} p$$

分析: 记 $A_n = \{|S_n/n - p| \geq \varepsilon\}$, 则定理中的几乎必然收敛的充分必要条件是

$$P\left(\limsup_n A_n\right) = 0.$$

再 Borel-Cantelli 引理可知, 只需证明

$$\sum_{n=1}^{\infty} P(A_n) < +\infty.$$

若像证明弱大数定律那样用切贝谢夫不等式，我们只能得到

$$P(A_n) \leq \frac{1}{\varepsilon^2} D\left(\frac{S_n}{n}\right) = \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Borel 强大数定律证明

- 用 $r=4$ 时的马尔科夫不等式可得 $P(A_n) \le \frac{1}{\varepsilon^4} E\left|\frac{S_n}{n}-p\right|^4$.
- 将 S_n 写为 n 个独立伯努利分布随机变量之和 $S_n=\sum_{k=1}^n X_k$, 从而

$$E\left[\left|\frac{S_n}{n}-p\right|^4\right]=\frac{1}{n^4}\sum_{1\le i,j,k,l\le n} E\left[(X_i-p)(X_j-p)(X_k-p)(X_l-p)\right].$$

- 注意 $X_1,\cdots,X_n$ 相互独立，故上式和号中仅有

$$E\left[(X_i-p)^4\right]=pq\left(p^3+q^3\right)\ne 0,$$
$$E\left[(X_i-p)^2(X_j-p)^2\right]=p^2q^2\ne 0,(i\ne j).$$

- 注意到前者共 n 项，后者共 $C_n^2C_4^2=3n(n-1)$ 项，故我们有

$$E\left[\left|\frac{S_n}{n}-p\right|^4=\frac{pq}{n^4}\left[n\left(p^3+q^3\right)+3pq\left(n^2-n\right)\right]<\frac{1}{4n^2}.$$

- 代入马等科夫不等式右方可得 $P(A_n)\le\frac{1}{4n^2\varepsilon^4}$, 定理得证.

Borel 强大数定律的备注

注 4.3

- 上述定理证明，本质上只用到独立同分布序列的四阶中心矩 $E\left[(X_k-p)^4\right]$ 以及 $E\left[(X_k-p)^2(X_j-p)^2\right]$ 有限，并没有用到伯努利情形的其它特性. 所以，在四阶中心矩有限的前提下，其证明可以照搬到一般独立同分布情形.
- 在如下定理中，我们在明显减弱的条件下，得到一般独立同分布场合的强大数定律.
  - 证明中采用恰当选取子序列的方法，仅用切比雪夫不等式就得到了满足 Borel-Cantelli 引理要求的估计式.
  - 而且运用著名的“截尾法”，连切比雪夫不等式中“方差有限”的要求都可以绕过.

一般独立同分布情形下的强大数定律

定理 4.16 假定 $\{X_n\}$ 是相互独立同分布的随机变量序列，如果它们有有限的期望 $E(X_n) = a$, 则强大数定律成立，即 $n \to \infty$ 时有

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} a.$$

分析

- 记部分和 $S_n = \sum_{k=1}^n X_k$. 如非负情形成立, 则有

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^n X_k^+ - \frac{1}{n}\sum_{k=1}^n X_k^- \xrightarrow{a.s_i} E\left(X_1^+\right) - E\left(X_1^-\right) = a.$$

故不妨假定 $\{X_k\}$ 是非负随机变量列.

- 注意到

$$P\left\{\frac{1}{n}\left|S_n - E\left(S_n\right)\right| \geq \varepsilon\right\} \leq \frac{E[\sum_{i=1}^n (X_i - E(X_i))]^k}{\varepsilon^k n^k}, k = 2, 4$$

- 高阶矩不存在, 需要截断
  - 方法 1: $X_n^*(\omega) = \begin{cases} X_n(\omega), & X_n(\omega) \leq M \\ 0, & X_n(\omega) > M. \end{cases} = X_n(\omega) 1_{[0,M]}\left(X_n(\omega)\right).$
  - 方法 2: $X_n^*(\omega) = \begin{cases} X_n(\omega), & X_n(\omega) \leq n, \\ 0, & X_n(\omega) > n. \end{cases} = X_n(\omega) 1_{[0,n]}\left(X_n(\omega)\right).$
  - $S_n^* = \sum_{i=1}^n X_i^*$

分析 (续)

- 对截断方法 1:
  - 若 $P\{X_1 > M\} > 0$, 则
$$\sum_{n=1}^{\infty} P(X_n \neq X_{\mathrm{n}}^*) = \sum_{n=1}^{\infty} P(X_n > M) = \sum_{n=1}^{\infty} P(X_1 > M) = +\infty,$$
  - 无法推出 $X_n - X_n^* \to 0, a.s.$. 从而无法得到 $\dfrac{S_n - S_n^*}{n} \to 0, a.s.$
- 对截断方法 2:
  - $\sum P(X_n \neq X_{\mathrm{n}}^*) = \sum P(X_1 > n) \leq \int_0^{+\infty} P(X_1 > t)dt = E(X_1) < +\infty,$
  - 故 $X_n - X_n^* \to 0, a.s.$. 更进一步有, $\dfrac{S_n - S_n^*}{n} \to 0, a.s.$
  - 注意到
$$\frac{S_n}{n} = \frac{S_n - S_n^*}{n} + \frac{S_n^* - E(S_n^*)}{n} + \frac{E(S_n^*)}{n}$$
  - $E(X_n^*) = E[X_n(\omega)1_{[0,n]}(X_n(\omega))] = E[X_1(\omega)1_{[0,n]}(X_1(\omega))] \xrightarrow{n\to\infty} E(X_1)$
  - 从而 $\dfrac{E(S_n^*)}{n} = \dfrac{\sum_{i=1}^n E(X_i^*)}{n} \to E(X_1) = a$

分析 (续)

- 只需证

$$\frac{S_n^* - E(S_n^*)}{n} \xrightarrow{a.s.} 0$$

- 注意到

$$P\left(\frac{1}{n}\left|S_n^* - E\left(S_n^*\right)\right| \geq \varepsilon\right) \leq \frac{D(S_n^*)}{\varepsilon^2 n^2} = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n} D(X_i^*)$$

$$\leq \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\int_0^\infty x^2 1_{[0,i]}(x)dF(x) \leq \frac{1}{\varepsilon^2 n}\int_0^\infty x^2 1_{[0,n]}(x)dF(x)$$

- 从而

$$\sum_{n=1}^{\infty} P\left(\frac{1}{n}\left|S_n^* - E\left(S_n^*\right)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\int_0^\infty x^2 \sum_{n=1}^{\infty}\frac{1}{n}1_{[0,n]}(x)dF(x)$$

- 若 $\sum_{n=1}^{\infty}\frac{1}{n}1_{[0,n]}(x) < \frac{M}{x}$, 则由 X_i 期望存在可知上式小于 ∞, 问题得证.

分析 (续)

- 遗憾的是, 我们无法做到 $\sum_{n=1}^{\infty}\frac{1}{n}1_{[0,n]}(x) < \frac{M}{x}$.
- 但我们如果将 n 换成特定的子列 $u_n = [\theta^n] \geq \frac{\theta^n}{2}$, 并令 N 是使 $u_n \geq x$ 的最小自然数, 我们则有可能做到

$$\sum_{n=1}^{\infty}\frac{1}{u_n}1_{[0,u_n]}(x) = \sum_{n=N}^{\infty}\frac{1}{u_n} \leq 2\sum_{n=N}^{\infty}\theta^{-n} = M\theta^{-N} \leq \frac{M}{x}$$

- 综上, 我们可以证明, 对任意的 $\theta > 1$ 及相应的 $u_n = [\theta^n]$, 我们均有

$$\frac{S_{u_n}^* - E(S_{u_n}^*)}{n} \xrightarrow{a.s.} 0$$

- 从而可证

$$\frac{S_{u_n}}{u_n} = \frac{S_{u_n} - S_{u_n}^*}{u_n} + \frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n} + \frac{E(S_{u_n}^*)}{u_n} \xrightarrow{a.s.} a$$

分析 (续)

- 若 $u_n \le k \le u_{n+1}$, 则由 X_k 的非负性知

$$\frac{u_n}{u_{n+1}}\frac{S_{u_n}}{u_n} \le \frac{S_k}{k} \le \frac{u_{n+1}}{u_n}\frac{S_{u_{n+1}}}{u_{n+1}}$$

- 注意 $u_{n+1}/u_n \to \theta$, 故令 n 及相应的 k 趋于无穷可得, 以概率 1 有

$$\frac{1}{\theta}\cdot a \le \liminf_{k\to\infty}\frac{S_k}{k} \le \limsup_{k\to\infty}\frac{S_k}{k} \le \theta\cdot a.$$

- 上式对任意的 $\theta > 1$ 均成立, 故特别的取 $\theta = (\ell+1)/\ell$, 则以概率 1 有

$$\frac{\ell}{\ell+1}\cdot a \le \liminf_{k\to\infty}\frac{S_k}{k} \le \limsup_{k\to\infty}\frac{S_k}{k} \le \frac{\ell+1}{\ell}\cdot a.$$

- 对每个 $\ell \ge 1$, 记上式不成立的例外集为 E_ℓ, 且取 $E = \bigcup_{\ell=1}^{\infty} E_\ell$.
- 显然 $P(E) = 0$, 且对任意的 $\omega \notin E$, 上式对一切 $\ell \ge 1$ 成立.
- 令 $\ell \to \infty$ 便得证对任意的 $\omega \in \bar{E}$ 恒有

$$\lim_k \frac{S_k}{k} = \lim_n \frac{1}{n}\sum_{k=1}^{n} X_k = a$$

证明

- 不妨假定 $\{X_k\}$ 是非负随机变量列.
- 截断: $X_n^*(\omega) = \begin{cases} X_n(\omega), & X_n(\omega) \le n, \\ 0, & X_n(\omega) > n. \end{cases} = X_n(\omega) 1_{[0,n]}(X_n(\omega))$.
- 令 $S_n^* = \sum_{i=1}^n X_i^*$ 及 $u_n = [\theta^n]$
- 证明: $\dfrac{1}{u_n} S_{u_n} \xrightarrow{a.s.} a$.
  - 只需证明

$$\frac{S_n - S_n^*}{n} \xrightarrow{a.s.} 0, \quad \frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n} \xrightarrow{a.s.} 0, \quad \frac{E(S_n^*)}{n} \to a$$

- 由 θ 的任意性证明

$$\frac{S_n}{n} \xrightarrow{a.s.} a.$$

伯努利大数定律的应用：蒙特卡罗法计算定积分

问题 4.1 设 $0 \le f(x) \le 1$, 计算 $f(x)$ 在区间 $[0,1]$ 上的积分近似值

$$J = \int_0^1 f(x)dx$$

解: 注意到

$$J = \int_0^1 f(x)dx = \int_0^1 \int_0^{f(x)} dydx = P(Y \le f(X)) := p$$

其中 (X,Y) 为方形区域 $[0,1]\times[0,1]$ 上的均匀分布。此时我们有

- X,Y 分别服从 $[0,1]$ 上的均匀分布且 X,Y 独立;
- 故计算积分的值实际上就是计算事件 $A = \{Y \le f(X)\}$ 的概率 p;
- (X,Y) 为方形区域 $[0,1]\times[0,1]$ 均匀分布，因此可以看成在方形区域内的随机投点试验;
- 由伯努利大数定律知，我们可以用重复随机投点试验中事件 A 出现的频率：即随机点 (X,Y) 落入区域 $\{(x,y): y \le f(x)\}$ 中的频率作为 p 的估计值.

蒙特卡罗法计算定积分的具体步骤

1. 先用计算机产生 $(0,1)$ 均匀分布的 $2n$ 个随机数，组成 n 对随机数 $(x_i, y_i), i = 1, \cdots, n$, 这里的 n 可以很大;
2. 对 n 对数据 $(x_i, y_i), i = 1, \cdots, n$, 记录满足 $y_i \leq f(x_i)$ 的次数，即事件 A 发生的频数 S_n;
3. 根据频数 S_n, 计算频率 $\dfrac{S_n}{n}$, 即为积分 J 的近似值.

一般区间上的定积分的计算

问题 4.2 设 $c \le g(x) \le d, x \in [a,b]$, 计算 $g(x)$ 在区间 $[a,b]$ 上的积分近似值

$$J_{a,b} = \int_a^b g(x)dx$$

解:

$$\begin{aligned} J_{a,b} &= \int_a^b g(x)dx \xlongequal{x=a+(b-a)y} \int_0^1 g(a+(b-a)y)(b-a)dy \\ &= (b-a)(d-c)\int_0^1 \left(\frac{g(a+(b-a)y)-c}{d-c} + \frac{c}{d-c}\right)dy \\ &= S_0 \int_0^1 f(y)dy + c(b-a) \end{aligned}$$

其中

$$S_0 = (b-a)(d-c), \quad f(y) = \frac{g(a+(b-a)y)-c}{d-c}$$

独立情形下的强大数定律

定理 4.17 (Kolmogorov 极大不等式) 设随机变量 $X_1, \cdots, X_n$ 相互独立，它们都有零期望与有限方差。令 $S_n = \sum_{k=1}^{n} X_k$, 则对任意的 $\epsilon > 0$ 有

$$P(\max_{1 \le k \le n} |S_k| \ge \epsilon) \le \frac{1}{\epsilon^2} D(S_n)$$

独立情形下的强大数定律

定理 4.18 (Kolmogorov 强大数定律) 设 $\{X_n\}$ 是独立随机变量序列，满足

$$\sum_{k=1}^{\infty} \frac{D(X_k)}{k^2} < +\infty,$$

则强大数定律成立，即当 $n \to \infty$ 时有

$$\frac{1}{n} \sum_{k=1}^{n} [X_k - E(X_k)] \xrightarrow{a.s.} 0.$$

第二十三讲：中心极限定理

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　数学学院

中心极限定理：问题的由来

- 实际生活中，我们会经常会进行误差分析，而误差的产生是由大量微小的相互独立的随机因素叠加而成。举个例子来说：机床加工机械轴使其直径符合规定要求，加工后的机械轴的直径会受以下各种随机因素的影响:
  - 机床振动与转速的影响;
  - 刀具装配与磨损的影响;
  - 钢材成份与产地的影响;
  - 量具的误差，测量技术的影响;
  - 车间的温度，湿度，照明，工作电压的影响等等.
- 若记每个因素使每个加工轴的直径产生的误差为 $X_i, i = 1, \cdots, n$, 则最终对加工轴直径产生的误差为

$$Y_n = \sum_{i=1} X_i$$

- 显然在实际中，我们必然要关心当 $n \to \infty$ 时 Y_n 的分布是什么？
- 为什么我们不根据 X_i 的分布用卷积公式来计算 Y_n 的分布？

一个例子

例 4.5 设 $\{X_n\}$ 为独立同分布的随机变量序列，其共同的分布为区间 $(0,1)$ 上的均匀分布。记 $Y_n=\sum_{i=1}^n X_i$，试计算 Y_n 的分布.

图 4.4.1 均匀分布的卷积

问题：寻求 Y_n 的极限分布

- 从上面均匀分布的例子可知:

$$E(Y_n) = \sum_{i=1}^{n} E(X_i) = \frac{n}{2} \to \infty$$

$$D(Y_n) = \sum_{i=1}^{n} D(X_i) = \frac{n}{12} \to \infty$$

- 考虑 Y_n 的极限分布时，需对 Y_n 进行标准化:

$$Y_n^* := \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}}$$

- 在什么情况下

$$\lim_{n\to\infty} P(Y_n^* \le y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du$$

独立同分布情形下的中心极限定理

定理 4.19 (林德伯格 - 莱维 (Lindeberg-Lé vy) 中心极限定理) 设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $E(X_i) = \mu, D(X_i) = \sigma^2 > 0$ 存在，若记

$$Y_n^* = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}},$$

则对任意的 $y \in R$ 有

$$\lim_{n\to\infty} P(Y_n^* \le y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du$$

Lindeberg-Lévy 中心极限定理的证明

证明: 要证 Y_n^* 依分布收敛于标准正态分布，只需要说明其特征函数收敛于标准正态分布的特征函数即可. 注意到 $Y_n^* = \sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$. 故若记 $Z_i := X_i - \mu$ 的特征函数为 $\varphi(t)$, 则

$$
\begin{aligned}
\varphi(t) &= \varphi(0) + \varphi'(0)t + \varphi''(0)\frac{t^2}{2} + o(t^2) \\
&= 1 + iE(Z_i) + i^2E(Z_i^2)\frac{t^2}{2} + o(t^2) \\
&= 1 - \frac{\sigma^2 t^2}{2} + o(t^2) \\
\varphi_{Y_n^*} &= \left[\varphi(\frac{t}{\sigma\sqrt{n}})\right]^n = \left[1 - \frac{t^2}{2n} + o(\frac{t^2}{n\sigma^2})\right]^n \to e^{-t^2/2}
\end{aligned}
$$

标准正态随机数的产生

- 在随机模拟中经常会需要产生正态随机数，下面我们介绍一种利用中心极限定理通过 $(0,1)$ 上的均匀分布的随机数来产生正态分布 $N(\mu,\sigma^2)$ 随机数的方法;
- 中心极限定理理论分析：设 $X_1,\cdots,X_n,\cdots$ 为一列独立同 $(0,1)$ 上的均匀分布的随机变量序列，则 $E(X_i)=1/2, D(X_i)=1/12$, 故

$$\frac{\sum_{i=1}^n X_i - n/2}{\sqrt{n/12}} \to \sim N(0,1)$$

  特别的，若 $n=12$, 则 $X_1+X_2+\cdots+X_{12}-6$ 近似服从标准正态分布.
- 产生正态随机数的方法:
  - 从计算机中产生 12 个 $(0,1)$ 上均匀分布随机数，记为 $x_1,\cdots,x_{12}$;
  - 计算 $y=x_1+\cdots+x_{12}-6$, 由上面分析知可将 y 看成 $N(0,1)$ 的一个随机数;
  - 计算 $z=\mu+\sigma y$, 可将 z 看成 $N(\mu,\sigma^2)$ 分布的随机数;
  - 重复上面步骤 n 次，即可得到 n 个 $N(\mu,\sigma^2)$ 分布随机数.

二项分布的正态近似

定理 4.20 (棣莫弗 - 拉普拉斯中心极限定理) 设 n 重伯努利试验中，事件 A 在每次试验中出现的概率为 $p \in (0,1)$, 记 S_n 为 n 重伯努利试验中事件 A 出现的次数，则对任意的 $y \in R$ 均有

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le y\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du$$

在上述定理中，若记 $\Phi(y) = \beta, Y_n^* := \dfrac{S_n - np}{\sqrt{np(1-p)}}$, 则

$$P(Y_n^* \le y) \approx \Phi(y) = \beta$$

独立不同分布下的中心极限定理

- 设 $\{X_n\}$ 是一个相互独立的随机变量序列，其期望与方差有限分别为

$$E(X_i) = \mu_i, \quad D(X_i) = \sigma_i^2, \quad i = 1, 2, \cdots,$$

- 考虑随机变量的和 $Y_n = X_1 + \cdots + X_n$, 则

$$E(Y_n) = \mu_1 + \cdots + \mu_n, \quad \sigma(Y_n) := \sqrt{D(Y_n)} = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$$

- 考虑 Y_n 的标准化:$Y_n^* := \dfrac{Y_n - E(Y_n)}{\sigma(Y_n)} = \sum_{i=1}^n \dfrac{X_i - \mu_i}{\sigma(Y_n)}$
- 要求 Y_n^* 中的各项 $\dfrac{X_i - \mu_i}{\sigma(Y_n)}$ 均匀的小，即对任意的 $\tau > 0$, 要求事件

$$A_{ni} = \left\{ \frac{X_i - \mu_i}{\sigma(Y_n)} > \tau \right\} = \{|X_i - \mu_i| > \tau\sigma(Y_n)\}$$

发生的可能性小或直接要求其概率趋于 0.

- 为达目的，我们要求 $\lim_{n\to\infty} P(\max_{1\le i\le n} |X_i - \mu_i| > \tau\sigma(Y_n)) = 0$

$\lim_{n\to\infty} P(\max_{1\le i\le n} |X_i - \mu_i| > \tau\sigma(Y_n)) = 0$

注意到

$$
\begin{aligned}
&P(\max_{1\le i\le n} |X_i - \mu_i| > \tau\sigma(Y_n)) = P(\cup_{i=1}^n |X_i - \mu_i| > \tau\sigma(Y_n)) \\
&\le \sum_{i=1}^n P(|X_i - \mu_i| > \tau\sigma(Y_n)) = \sum_{i=1}^n \int_{|x-\mu_i|>\tau\sigma(Y_n)} dF_i(x) \\
&\le \frac{1}{\tau^2\sigma^2(Y_n)} \sum_{i=1}^n \int_{|x-\mu_i|>\tau\sigma(Y_n)} (x-\mu_i)^2 dF_i(x)
\end{aligned}
$$

故只要下面的 **林德伯格条件** 满足

$$
\lim_{n\to\infty} \frac{1}{\tau^2\sigma^2(Y_n)} \sum_{i=1}^n \int_{|x-\mu_i|>\tau\sigma(Y_n)} (x-\mu_i)^2 dF_i(x) = 0
$$

就可保证 Y_n^* 中各加项充分的小

林德伯格中心极限定理

定理 4.21 设独立随机变量序列 $\{X_n\}$ 满足林德伯格条件

$$\lim_{n\to\infty}\frac{1}{\tau^2\sigma^2(Y_n)}\sum_{i=1}^{n}\int_{|x-\mu_i|>\tau\sigma(Y_n)}(x-\mu_i)^2dF_i(x)=0$$

则对任意的 x, 有

$$\lim_{n\to\infty}P\left(\frac{1}{\sigma(Y_n)}\sum_{i=1}^{n}(X_i-\mu_i)\le x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du$$

注 4.4 假设 $\{X_n\}$ 独立同分布且方差有限，则容易验证上述的林德伯格条件满足，从而我们先前在独立同分布情形下的中心极限定理是林德伯格中收极限定理的特例.

推出 Lindeberg 条件的充分条件

定理 4.22 设 $\{X_n\}$ 为独立随机变量序列，如果存在常数列 $\{L_n\}$ 使得

$$\max_{1\le k\le n}|X_k|\le L_n,\ 且\lim_n\frac{L_n}{B_n}=0$$

则 Lindeberg 条件成立，从而 $\{X_n\}$ 满足中心极限定理.

李雅普诺夫中心极限定理

定理 4.23 设 $\{X_n\}$ 为独立随机变量序列，若存在 $\delta > 0$ 满足

$$\lim_{n\to\infty} \frac{1}{\sigma(Y_n)^{2+\delta}} \sum_{i=1}^{n} E(|X_i - \mu_i|^{2+\delta}) = 0$$

则对任意的 x, 有

$$\lim_{n\to\infty} P(\frac{1}{\sigma(Y_n)} \sum_{i=1}^{n} (X_i - \mu_i) \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$