西北工業大學

概率论与数理统计

数学与统计学院 **钱锋** 编

2023 年 12 月 26 日

不用翻译的汉译 Strik0r 学术垃圾译丛　丛书前言

这套《不用翻译的汉译 Strik0r 学术垃圾译丛》(Strik0r's Garbage-like Texts in Mathematics/Computer Science/Physics, 以下简称 Strik0r's GTM) 又臭又长, 你只能选择性地阅读. 本丛书是我在大学期间学习到的知识和工作中的见闻和经验的总结 (说人话就是所有东西都是抄来的), 内容横跨包括数学、物理学和计算机科学在内的多个学科及多个领域. 是一套纯粹的抄袭之作, 经典的盗版, 标准的垃圾.

但就笔者不多的学习经验来说, **坚持输出倒逼输入**是笔者在这短短几年的学习生涯中总结出来的一条关于学习的规律性认知, 这个认识与著名和 Feynmann 学习法、ZettelKasten 背后的原理和现代脑科学和认知科学的相关学术研究成果不谋而合. 实践证明, 不断地通过创造各种形式的输出来检验输入是否有效, 是确保学习成效和学习效率的关键因素和重要手段. 它有助于对知识点本身的理解, 也有助于知识网络和知识体系的搭建, 可以达到"既见树木, 又见森林"的效果, 打破"学而不思、思而不学"的怪圈. 因此这样一堆学术垃圾, 本质上是笔者学习、认识、实践和思考的副产物, 它们的诞生是有利于笔者本人的学习的.

如今, 借由西北工业大学数学与统计学院开展的"学霸直播间"朋辈学习辅导交流活动的平台, 笔者的这些学术垃圾得以面世, 在学院内部和网络上经过两学期的使用后取得了不错的反响, 故笔者将这些学术垃圾整理成册, 最终形成了这套资料. 这套资料知识体系庞大、内容丰富、深入浅出, 均代表了笔者对该领域的最高了解程度 (也就是说, 书里有的我都会, 书里没有的我听都没听过). 主要包括了以下方面的内容:

1. **数学与自然科学类**: 主要包括数学分析 (或微积分, 高等数学)、高等代数 (或线性代数)、概率论与数理统计、解析几何、数学物理方法 (复变函数、积分变换、常微分方程、偏微分方程)、离散数学和抽象代数等数学课程和大学物理 (力学、振动与波动、波动光学、热学、电磁学、近代物理学部分)、电路分析等物理类课程的内容.

2. **计算机科学类**: 主要包括程序设计基础 (Python、C、C++、Java、Kotlin)、软件技术 (MATLAB)、数据结构 (C、Python)、操作系统、计算机网络、计算机系统基础、计算机组成与设计等计算机科学与技术方向课程, 软件工程导论、面向对象分析与设计等软件工程方向课程, 以及机器学习、神经网络、深度学习等人工智能方向课程的内容.

3. **思想政治理论与军事类**: 主要包括马克思主义基本原理、中国近现代史纲要、思想道德与法治、毛泽东思想与中国特色社会主义理论体系概论、习近平新时代中国特色社会主义思想概论 5 门思政课程和军事理论、军事技能 2 门军事类课程的内容.

4. **审美与艺术类**: 主要包括美学原理、中国审美历程课程的内容. 后续考虑加入数学与音乐、基础乐理等音乐类的内容, 当前, 前提是我要进一步扩充我对有关方面的知识储备.

5. **外国语言文学类**: 有一本专门命名为 English 的资料, 是我在本科期间所学英语课程的究极缝合怪.
6. **运动与健康类**: 有一本运动与慢病防治, 说的就是这个内容.
7. **写作与沟通类**: 沟通表达技能是非常重要且非常关键的, 因此我撰写了一系列与写作和沟通有关的文档, 归在这一类当中了, 现在正在写的有学术文档写作 (LaTeX).

这套学术垃圾的编写原则

在编写这套学术垃圾的时候, 我们遵循了以下原则:

1. 注意材料的普适性

毕竟我只是个本科生, 没有任何一个高校会真的把我的资料当作教材来使用, 也不会有任何一个出版社愿意把我的资料出版 (虽然我并没有联系过). 因此在编写资料的时候, 我尽可能地让它们具有普适性, 能够满足大部分培养方案和教学大纲的学习要求, 以尽可能地让这些资料能够覆盖到更多的同学、帮助到更多的同学.

除此之外, 关于普适性的另一个阐述是同一份材料可以适应不同层次的课程教学需求, 例如《数学分析》可供数学院系同学参考, 也可供修读高等数学或微积分课程的理工类、经管类院系的同学参考,《程序设计基础》系列可供各种专业院系的同学参考. 为了满足不同教学层次的需求, 我特地在每一本资料的前言中都列了一张表格, 标注哪些内容是哪个层次的课程需要掌握的, 除此之外, 你也可以对照贵校的教学大纲以及相应课程的考试大纲来确定你的学习范围.

2. 注重系列材料的统一性和整体性

这个系列涉及的学科很多, 这个特点决定了我在编写这个系列的材料的时候会遇到很多来自不同领域的例子和解决方案, 因此你在阅读的时候会遇到的一个比较常见的操作是: 在《计算机系统基础》或者类似学科的资料中找到某一个理论, 然后在《抽象代数》中找到这个理论的分析或者证明 (放心, 我在行文的过程当中会给你指路的), 如此一来便可以极大程度地减小我们的例题数量, 转而通过更多我们更容易接触到的实例来向大家展示数学理论在工程实践和实际生活当中的应用. 相信对于兴趣广泛、求知欲强的同学来说, 一边阅读内容充实丰富的材料, 一边巩固学习到的专业知识, 一边接触学科交叉领域当中的实际应用, 是一件非常享受的事.

钱锋
西北工业大学软件学院
2023 年 12 月

第一版前言

概率论与数理统计是研究和揭示随机现象统计性规律的学科, 是数学学科的重要分支. 本书是根据西北工业大学数学与统计学院开设的《概率论与数理统计 II》课程的课程大纲, 结合期末考试的重要考察内容和笔者本人的科学见解, 并本着注重培养能力、塑造思维、拓宽知识面的原则编写而成的.

《概率论与数理统计》作为 Strik0r's GTM (Strik0r's Garbage-like Text in Mathematics/Computer Science/Physics, 即"不用翻译的汉译 Strik0r 学术垃圾译丛") 系列的第一部作品, 是我在广泛借鉴、疯狂抄袭, 以及恬不知耻的落款自己的名字而编写出来的. 我知道它很烂, 我知道你想喷, 但我比你更想喷, 所以我知道你很想喷, 但你先别喷. 在一切的一切开始以前, 我还是想先自吹自擂一下, 然后向各位读者介绍一下这门课程, 讲一下概率论与数理统计这门课程的特点, 并给出一些阅读本书的建议.

概率论与数理统计是怎么样的一门课?

概率论与数理统计是一门面向我国高等教育各个专业院系的课程, 在数学类、统计学类院系可能会分成概率论、数理统计两门课程来开设. 这门课程的主要内容是随机现象及其概率、随机变量的分布及其数字特征, 和以概率论为基础建立起来的, 以简单随机抽样、参数估计和假设检验为代表的数理统计的原理与方法.

概率论与数理统计这门课程的其中一个特点是内容多、知识量大. 在这门课程中我们每一章介绍一个模块的内容, 后面的模块与前面的模块紧密相扣, 并且在解题时需要将各个模块的知识综合起来使用 (这一点和高等代数/线性代数很像), 也就是说, 概率论与数理统计的内容知识量大, 而且具有较强的逻辑上的连贯性. 但纵览概率论与数理统计的各个章节, 你会发现这门课程的核心要义不过寥寥数条, 各个章节内部的"主线"也是非常明确的. 第一章在古典的等可能概型视角下研究随机事件发生的概率和不同随机事件的概率之间的关系 (条件概率、乘法公式、全概率公式、Bayes 公式和独立性); 第二章、第三章研究一维、二维离散型和连续性随机变量的概念及其分布, 进而给出了描述随机事件的一种新的方式——用随机变量在一定范围内的取值来描述随机事件, 进而利用随机变量的分布律或分布函数/概率密度函数来求解随机事件的概率, 在概率论中引入了分析学的方法; 第四章研究随机变量的两个重要的数字特征: 数学期望和方差, 刻画并给出了两个随机变量之间相关程度的协方差和相关系数; 第五章讨论在样本容量足够大时, 随机变量的概率分布会逼近于怎样的分布; 第六章介绍简单随机抽样的概念和三种重要的抽样分布, 并介绍了来自正态总体样本的统计量所服从的统计分布, 为第七章通过参数估计和假设检验来估计未知参数的值和做出统计推断提供支持.

第一版讲义的编写原则

早期版本应试优先, 在后续版本中再逐渐增加理论部分的内容. 应试是大学课程对每个同学的最低的要求, 准确地抓住重点、有的放矢是顺利通过大学期末考试的核心. 为此, 我们在讲义中搜罗了我校

近年来概率论与数理统计 II 课程的部分期末考试真题, 并且将它们按照章节和知识点顺序作为例题分散在全书中, 并在适当的思路分析后给出标准解答, 以便于同学们一边复习, 一边通过适当的例题巩固对知识点的理解.

在本书的第二版以及之后的版本中, 我们会逐渐的丰富本书的内容, 让它能够供更多学校的同学学习复习使用, 进而能为考研数学复习提供参考, 在这之后, 我们会考虑加入一些更理论性的内容来丰富和充实我们的内容框架, 并加入一些概率论与数理统计的应用案例来让我们学到的知识真正"活"起来.

如何使用本书?

首先强调一点, 本书并**不能代替正规的概率论与数理统计教材和课程**使用, 只能作为学习复习过程当中的参考资料, 这一点请各位读者知悉. 而对于确定要将本书作为概率论与数理统计课程参考书之一的同学而言, 我想向你说明的是这本书里的口水话很多, 因为我认为只有公式和逻辑符号, 没有文字描述的数学书读起来就像编码习惯不好还不写注释的屎山代码一样令人感到恶心, 因此我在梳理知识体系和讲解例题的时候加入了大量剖析思路的文字, 希望能够帮助到一些在平时学习课程中总是感到"啊嘞嘞, 这一步是怎么得出来的?"的、总是对数学题的解题思路感到不明觉厉的同学.

除此之外, **本书没有习题**, 我做过的大部分题目都被收录在了文档当中, 但我都写了答案, 但要想学好一门数学课, 练习是必不可少的, 因此各位读者需要自己寻找一些题目来练习解题. 记住: 对于应试而言, 一味地啃理论是没有意义的, 必须要保证一定的习题量才能保证足够好的学习成效. 建议各位读者一边学习, 一边就开始阅读课题组提供的《概率论与数理统计历年期末试题与试题解析》, 以更快熟悉题型、更早做好准备、提前进入状态, 避免陷入"考前破防, 考后狂浪、考前再破防"的恶性循环.

最后, 我将会在 2024 年的寒假录制一套基于这本资料的概率论与数理统计课程视频, 有学习需求的同学可以配合视频阅读本书以获得更好的学习体验. 在第二个大版本发布前, 本书将会在 QQ 频道「Strik0r 作业共享频」持续更新, 各位同学在学习时可以留意「常用资源共享」子频道内是否有新版的讲义发布.

最后的祝福

年关将近, 2023 年马上就要过去了. 在这一年里, 以 ChatGPT 为代表的 AI 大模型的出现对各个行业都造成了一定的冲击, 虽然这些人工智能领域的前沿动态并不在本书讨论的范畴内, 但我还是希望各位读者能从雨后春笋般涌现的语言大模型潮中看到我国作为 AI 大国尚且"大而不强"、在诸多关键环节、关键领域面临技术封锁和"卡脖子"的现状, 进而坚定我们学习理论知识、夯实理论基础的理想信念. 转眼又到一年年末, 不知今年的你收获了些什么, 也不知道在这一年里有什么样的失去才让你如梦中惊醒般后知后觉自己曾经拥有. 但无论如何, 都祝愿你能够在未来的生活中一切顺利, 学业有成. 让我们在美妙的数学旅程中永远坚持走向上的道路, 追求真理、正义、智慧和自由.

钱锋
西北工业大学 数学与统计学院
2023 年 12 月于启翔湖畔

目录

第一部分

概率论

第 1 章　概率论基础概念

在一定条件下必然发生, 条件可以完全决定结果的现象称为**确定性现象**, 在一定条件下可能发生也可能不发生, 条件不能完全决定结果, 个别试验结果呈现不确定性, 大量试验结果呈现**统计规律性**的现象称为**随机现象**, 概率论与数理统计是研究和揭示随机现象统计性规律的学科. 概率论的特点是先提出数学模型, 然后去研究其性质、特点和规律; 数理统计则是以概率论的理论为基础, 利用对随机现象的观测所得到的数据, 来研究数学模型并在此基础上给出预测与判断. 本章我们将要介绍许多概率论与数理统计中的基本概念, 这些概念将贯穿课程始终.

1.1　随机试验、样本空间、随机事件

本节介绍概率论与数理统计中研究随机现象的重要方法, 并介绍样本空间和随机事件的基本概念和实例.

1.1.1　随机试验、样本空间、随机事件的概念与实例

定义 1.1.1. (随机试验). **试验**是我们对自然世界随机现象观察或实验的一种抽象, 是一个含义广泛的术语. **随机试验**是我们研究随机现象的一种重要方法. 随机试验具有如下特征:

1. **可重复性**: 在相同的条件下, 随机试验可以重复进行.
2. **不唯一性**: 每次试验的可能结果不止一个, 并且能事先明确试验的所有可能结果.
3. **不可预测性**: 在进行一次随机试验之前不能确定本词随机试验的结果.

定义 1.1.2. (随机事件). 随机试验 E 所有可能的结果组成的集合称为随机试验 E 的**样本空间**, 一般记作 Ω; 样本空间中的元素即为随机试验 E 的可能结果, 称为**样本点**, 记作 ω; 样本空间中的子集为随机试验 E 的**随机事件**, 我们定义一个事件发生当且仅当该子集中的一个样本点出现; 样本空间 Ω 的单元素子集 $\{\omega_i\}$ 称为 E 的**基本事件**, 一个基本事件是相对于随机试验目的而言的不可再分的随机事件.

注. 事件的本质是集合.

样本空间是随机事件所有结果构成的集合, 样本空间的某个子集称为随机事件. 需要注意的是, 随机试验与样本空间并非一一对应的关系, 对于不同的试验, 其样本空间可以相同也可以不同 (一个样本空间可以概括许多内容大不相同的实际问题); 而对于同一随机试验, 对于不同的试验目的, 其样本空间也是不同的. 在具体问题的研究中, 描述随机现象的第一步就是建立样本空间.

显然 $\Omega \subset \Omega$, 因此样本空间 Ω 本身也构成随机试验 E 的一个事件, 对于随机试验 E 而言, Ω 中的样本点总是发生的, 因此 Ω 相对于 E 而言是一个**必然事件**; 空集 $\varnothing$ 中不包含任何样本点, 是不可能发生的, 称为**不可能事件**. 必然事件和不可能事件是随机试验 E 的两种特殊的事件, 称为**平凡的事件**.

刚才我们用集合论的语言描述了随机试验、样本空间和随机事件, 这样做的好处是便于使用集合的运算规则描述随机事件及其关系, 使得随机事件间也可以相互做运算. 我们将有关随机事件的几个基本概念与其在集合论中的对应概念罗列成下表, 供各位读者参考:

表 1.1 随机事件的基本概念及其在集合论中的表示

符号	概率论	集合论
Ω	样本空间, 必然事件	空间 (全集)
$\varnothing$	不可能事件	空集
ω	样本点	元素
$\{\omega\}$	基本事件	单元素子集
$A \subset \Omega$	事件	子集

例 1.1.1. 抛掷一枚硬币, 观察朝上的是正面还是反面, 是一个随机试验. 它有两种结果: 正面朝上或反面朝上. 因此该随机试验的样本空间 $\Omega = \{0, 1\}$, 其中 0 表示正面朝上, 1 表示反面朝上. 其中, 0,1 分别为该随机试验的两个基本事件. 这种的样本空间包含两个样本点的随机试验称为 $0-1$ 模型.

例 1.1.2. 一枚硬币连续抛掷 n 次, 观察正面出现的次数的样本空间 $\Omega = \{1, 2, \cdots, n\}$.

例 1.1.3. 记录某校车乘车点某一时刻的等车人数, 则样本空间 $\Omega = \mathbb{N}^+$.

例 1.1.4. 记录某地某时刻的气温, 是一个随机试验, 它的样本空间 Ω 是 $\mathbb{R}$ 上的一个区间 $[T_1, T_2]$. 该区间内的任何一个数值构成该随机试验的一个基本事件. 事件"气温在 $[t_1, t_2] \subset [T_1, T_2]$"(记作事件 A) 表述为

$$A = \{t | t_1 \leqslant t \leqslant t_2\} = [t_1, t_2].$$

例 1.1.5. 观测某服务器的一台网络交换机每分钟的数据传输量 (以 GB 计). 它的样本空间是 $\mathbb{R}^+$. 事件"数据传输量小于 10 GB"(记作事件 A) 表述为

$$A = \{x | 0 \leqslant x < 10\} = [0, 10).$$

1.1.2 事件间的关系与事件的运算

接下来我们用集合的运算法则来定义事件的运算, 用集合间的关系来定义事件间的关系. 用概率论的角度来解释和描述事件, 并利用集合论来叙述事件的关系和运算是本节学习的重点.

1. 事件的运算

定义 1.1.3. (和事件). 两个事件 A, B 中至少有一个事件发生, 记作

$$A \cup B \xlongequal{\text{def}} \{\omega | (\omega \in A) \vee (\omega \in B)\}.$$

定义 1.1.4. (积事件). 两个事件 A, B 同时发生, 记作

$$A \cap B = AB \xlongequal{\text{def}} \{\omega | (\omega \in A) \wedge (\omega \in B)\}.$$

我们可以将其推广到至多可列多个事件的情形, n 个事件 $A_1, A_2, \cdots, A_n$ 的和事件与积事件分别记作,

$$\bigcup_{i=1}^{n} A_i, \quad \bigcap_{i=1}^{n} A_i,$$

它们分别表示事件 $A_1, A_2, \cdots, A_n$ 中至少有一个发生和事件 $A_1, A_2, \cdots, A_n$ 同时发生.

定义 1.1.5. (差事件). 事件 A 发生但 B 不发生称为 A 与 B 的差, 记作

$$A - B = A \setminus B = A\overline{B} \xlongequal{\text{def}} \{\omega | (\omega \in A) \wedge (\omega \notin B)\}.$$

事实上, 差集的本质是从集合 A 中"排除"B 中的元素, 这也可以理解为从 A 中"排除"A 与 B 的公共元素 (因为 B 中不属于 A 的元素自然是不能被 A 排除的), 因此差事件也可以表示为

$$\boxed{A - B = A\overline{B} = A - AB}.$$

表 1.2 事件间的运算及其在集合论中的表示

运算	符号	概率论	集合论
和	$A \cup B$	事件 A 与 B 至少有一个发生	A 与 B 的并集
积	$A \cap B$ 或 AB	事件 A 与 B 同时发生	A 与 B 的交集
差	$A - B = A \setminus B = A\overline{B}$	事件 A 发生, 但事件 B 不发生	A 与 B 的差集

2. 事件间的关系

定义 1.1.6. (包含关系). 如果 事件 A 发生必然有事件 B 发生, 则称事件 B **包含** (contain) 事件 A, 记作 $B \subset A$.

性质 1. 如果 $A \subset B$, 则 $A \cup B = B, AB = A$.

根据集合的相关性质, 这个性质是显然的. 特别地, 由于 $\varnothing \subset A \subset \Omega$, 所以有

$$A \cup \varnothing = A, \qquad A\varnothing = \varnothing;$$
$$A \cup \Omega = \Omega, \quad A \cap \Omega = A.$$

定义 1.1.7. (相等关系). 事件 A, B 满足 $(A \subset B) \wedge (B \subset A)$, 则称 $A = B$. 即事件 A 生必然导致事件 B 发生, 且事件 B 发生必然导致事件 A 发生.

定义 1.1.8. (互斥关系). 若事件 A 发生必然导致事件 B 不发生, 事件 B 必然导致事件 A 不发生, 则称 A, B 互不相容, 记作 $A \cap B = \varnothing$.

当 $A \cap B = \varnothing$ 时, 可以将 $A \cup B$ 记为直和, 即 $A \cup B = A \oplus B$.

定义 1.1.9. (对立事件). $\overline{A} = \Omega - A$ 称为事件 A 的对立事件, 显然 $A, \overline{A}$ 必有一个发生. 且 A 对立事件 $\overline{A}$ 的对立事件即为 A 本身, 即

$$\overline{(\overline{A})} = A.$$

注意互斥事件与对立事件的关系, 如果 A, B 互斥, 则 $A \cap B = AB = \varnothing$, 若 A, B 对立, 则 $(A \cup B = \Omega) \wedge (A \cap B = AB = \varnothing)$, 因此对立一定互斥, 但互斥未必对立.

表 1.3 事件间的关系及其在集合论中的表示

关系	符号	概率论	集合论
包含	$A \subset B$	事件 A 发生则事件 B 必发生	A 是 B 的子集
等价	$A = B$ 或 AB	$(A \subset B) \wedge (B \subset A)$	集合 A 与 B 相等
互斥	$AB = \varnothing$	事件 A 与事件 B 不能同时发生	A 与 B 不相交
对立	$\overline{A}$	事件 A 的对立事件	A 在样本空间 Ω 中的余集 $A \cup \overline{A} = \Omega$ $A\overline{A} = \varnothing$

3. 事件间的运算定律

定理 1.1.1. (交换律). $A \cup B = B \cup A$, $AB = BA$.

定理 1.1.2. (结合率). $(A \cup B) \cup C = A \cup (B \cup C)$, $(AB)C = A(BC)$.

定理 1.1.3. (分配律). 事件之间的和运算与积运算满足以下分配律:

a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;

b) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$;

c) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;

d) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

定理 1.1.4. (De Morgan 对偶定理). 事件 A, B 至少有一个发生的对立事件是 A, B 均不发生, 即

$$\overline{A \cup B} = \overline{A} \cap \overline{B};$$

事件 A, B 均发生的对立事件是 A, B 至少有一个不发生, 即

$$\overline{A \cap B} = \overline{A} \cup \overline{B}.$$

将其推广到 n 个事件的情形:

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}, \quad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}.$$

1.2 随机事件的概率

1.2.1 概率的定义

定义 1.2.1. (频数与频率). 在相同的条件下, 进行了 n 次试验, 在这 n 次试验中, 事件 A 发生的次数 n_A 称为事件 A 发生的**频数**, 比值 $\dfrac{n_A}{n}$ 称为事件 A 发生的**频率**, 记为 $f_n(A)$.

当试验次数 n 不太大时, 频率 $f_n(A)$ 在 0,1 之间随机波动, 但是当 n 逐渐增大时, $f_n(A)$ 将会逐渐收敛于某个常数, 这种"频率的稳定性"就是通常而言的**统计规律性**.

定义 1.2.2. (概率). 设随机试验 E 的样本空间为 S, 对于每一个事件 $A \subset S$, 构造一个实值函数 $P(A)$,

$$\begin{aligned} P : S &\to \mathbb{R} \\ A &\mapsto P(A), \end{aligned}$$

如果函数 P 满足下列条件:

1) **非负性**: $\forall A \subset S (P(A) \geqslant 0)$.
2) **规范性**: $P(S) = 1$.
3) **可列可加性**: 设 $A_1, A_2, \cdots$ 两两互不相容, 即 $\forall i, j = 1, 2, \cdots (i \neq j \Longrightarrow A_i A_j = \varnothing)$, 则
$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

则称 $P(A)$ 为事件 A 的**概率** (probability).

1.2.2 概率的性质

性质 1. $P(\varnothing) = 0$.

性质 2. (有限可加性). 设 $A_1, A_2, \cdots, A_n$ 两两互不相同, 则
$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k).$$

性质 3. 设 $A, B \subset S$ 是两个事件, 若 $A \subset B$, 则
$$P(B - A) = P(B) - P(A), \quad P(B) \geqslant P(A).$$

推论 1. 设 $A, B \subset S$ 是两个事件, $P(A - B) = P\left(A\bar{B}\right) = P\left(A - AB\right) = P(A) - P(AB)$.

证明. 由于 $\forall A, B \subset S\, (A \subset B \Longrightarrow P(B - A) = P(B) - P(A))$, 而 $AB \subset A$ 显然成立. □

性质 4. $\forall A \subset S (P(A) \leqslant 1)$.

性质 5. (逆事件的概率). $\forall A \subset S \left(P\left(\bar{A}\right) = 1 - P(A)\right)$.

性质 6. (加法公式). $\forall A, B \subset S\, (P(A \cup B) = P(A) + P(B) - P(AB))$.

例 1.2.1. 设 A, B 为随机事件, 且 $P(A) = 0.4$, 又 $P(AB) = P\left(\bar{A}\bar{B}\right)$, 求 $P(B)$.

解. 别忘了我们在写 AB 的时候, 实际是在写 $A \cap B$, 同时根据 De Morgan 对偶律, $\bar{A}\bar{B} = \overline{A \cup B}$, 因此

$$P\left(\bar{A}\bar{B}\right) = 1 - P(A \cup B),$$

根据加法公式,

$$P(AB) = P(\bar{A}\bar{B}) = 1 - \left(P(A) + P(B) - P(AB)\right),$$

代入 $P(A) = 0.4$, 解这个一元一次方程, 即可求得 $P(B) = 0.6$.

1.3 等可能概型 (古典概型)

例 1.3.1. 某人有外形完全一样的 10 把钥匙, 现在随机抽取一把钥匙去打开一扇门, 试过的钥匙不需要再次尝试, 求他不超过 3 次就可以打开门的概率是多少?

解. 设 X 为尝试的次数, 所求概率为 p. 则

$$p = P\left\{X = 1\right\} + P\left\{X = 2\right\} + P\left\{X = 3\right\},$$

$X = k$ 隐含的条件是, 前面 $k - 1$ 把钥匙都是假钥匙, 只有最后一把才是真钥匙, 所以

$$P\left\{X = k\right\} = \frac{9}{10}\frac{8}{9}\cdots\frac{10-k+1}{10-k+2}\frac{1}{10-k+1},$$

于是

$$p = \frac{1}{10} + \frac{9}{10}\frac{1}{9} + \frac{9}{10}\frac{8}{9}\frac{1}{8} = \frac{3}{10},$$

因此 $p = \dfrac{3}{10}$ 即为所求.

注. 你会发现, 无论 k 取 $1, 2, \cdots, 10$ 中的哪一个数, $\dfrac{9}{10}\dfrac{8}{9}\cdots\dfrac{10-k+1}{10-k+2}\dfrac{1}{10-k+1} = \dfrac{1}{10}$, 所以你也可以把解题步骤简化为 $p = \dfrac{1}{10} + \dfrac{1}{10} + \dfrac{1}{10} = \dfrac{3}{10}$.

1.4 条件概率

1.4.1 条件概率

定义 1.4.1. (条件概率). 设 $A, B \subset S$ 是两个事件, 且 $P(A) > 0$, 称

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

为在事件 A 发生的条件下事件 B 发生的**条件概率**.

例 1.4.1. 假设你学习 10 分钟以上的概率为 80%, 学习 20 分钟以上的概率为 40%, 如果你已经学习了 10 分钟, 求你能学习 20 分钟以上的概率.

解. 记 A 为学习 10 分钟以上, B 为学习 20 分钟以上, 显然 $B \subset A$, 因此 $P(AB) = P(B = 0.4$, 根据条件概率的定义,

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.4}{0.8} = 0.5,$$

这个结论与心理学上的"开始 10 分钟"法的经验是相符合的, 也就是说, 完成一件你并不想做的事情的一个很好的办法是先尝试做一小段事件 (例如 10 分钟), 在这 10 分钟后, 你想要继续完成这件事情的概率就会上升, 积极性就会提高, 这会在一定程度上帮助你克服你的拖延症.

例 1.4.2. 设 A, B 为随机事件, $P(A) = 0.7$, $P(B) = 0.4$, $P(A - B) = 0.5$. 则 $P\left(A + B \,\middle|\, \bar{A}\right) =$ ____________.

解. 由于

$$P(A - B) = P(A - AB) = 0.7 - P(AB) = 0.5,$$

因此

$$P(AB) = 0.2,$$

而 $P\left(B\bar{A}\right) = P(B - BA) = P(B) - P(AB)$, 因此根据条件概率公式,

$$P\left(A + B \,\middle|\, \bar{A}\right) = \frac{P\left((A + B) \cap \bar{A}\right)}{P\left(\bar{A}\right)} = \frac{P(B - BA)}{1 - P(A)} = \frac{0.4 - 0.2}{0.3} = \frac{2}{3}.$$

1.4.2 乘法公式

定理 1.4.1. (乘法定理). 设 $A, B, C \subset S$ 是三个事件, 且 $P(A > 0)$, 则

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

若 $P(AB) > 0$ 则

$$\begin{aligned} P(ABC) &= P(C|AB)P(AB) \\ &= P(C|AB)P(B|A)P(A). \end{aligned}$$

1.4.3 全概率公式

定义 1.4.2. (样本空间的划分). 设试验 E 的样本空间为 S, $B_1, B_2, \cdots, B_n \subset S$ 是 E 的事件, 如果

1° $\displaystyle\bigcup_{k=1}^{n} B_k = B_1 \cup B_2 \cup \cdots \cup B_n = S$;

2° 当 $i \neq j$ 时 $B_iB_j = \varnothing$, 即 $B_1, B_2, \cdots, B_n$ 两两互斥,

则称 $\{B_k\}_{k=1}^{n}$ 为样本空间 S 的一个**划分** (split, 或称为完备事件组).

定理 1.4.2. (全概率公式). 设试验 E 的样本空间为 S, $\{B_k\}_{k=1}^{n}$ 为样本空间 S 的一个划分, 如果 $P(B_k) > 0 (k = 1, 2, \cdots, n)$, 则

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n) \\ &= \sum_{k=1}^{n} P(A|B_k)P(B_k). \end{aligned}$$

证明. 由于

$$A = AS = A\left(\bigcup_{k=1}^{n} B_k\right) = \bigcup_{k=1}^{n}(AB_k),$$

由于 $\{B_k\}_{k=1}^n$ 为样本空间 S 的一个划分, 因此 $(AB_i)(AB_j) = \varnothing$, 又 $P(B_k) > 0(k = 1, 2, \cdots, n)$, 所以有

$$P(A) = P(AB_1) + P(AB_2) + \cdots + P(AB_n) = \sum_{k=1}^{n} P(AB_k),$$

利用乘法定理,

$$P(A) = \sum_{k=1}^{n} P(A|B_k)P(B_k),$$

于是全概率公式得证. □

1.4.4 Bayes 公式

定理 1.4.3. (Bayes 公式). 设试验 E 的样本空间为 S, $A, B_1, B_2, \cdots, B_n$ 是 E 的事件, 且 $\{B_k\}_{k=1}^n$ 为样本空间 S 的一个划分, 如果 $P(A) > 0$, $P(B_k) > 0$, 则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum\limits_{k=1}^{n} P(A|B_k)P(B_k)}, \quad i = 1, 2, \cdots, n.$$

证明. 根据条件概率的定义,

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

而由于 $\{B_k\}_{k=1}^n$ 为样本空间 S 的一个划分, $P(A) > 0$, $P(B_k) > 0$, 因此根据全概率公式

$$P(A) = \sum_{k=1}^{n} P(A|B_k)P(B_k),$$

根据乘法定理,

$$P(B_iA) = P(A|B_i)P(B_i),$$

整理将 $P(A)$ 和 $P(B_iA)$ 分别代回原式中即为 Bayes 公式. □

例 1.4.3. 设袋子中共有 10 个球, 其中 4 个红球, 6 个白球, 从中取两次球, 每次取一球, 取后不放回. 求: (1) 第二次取到白球的概率? (2) 已知第二次取到白球, 则第一次也取到白球的概率?

解. 设 R_1, R_2, W_1, W_2 分别表示第 1 次取到红球, 第 2 次取到红球, 第 1 次取到白球, 第 2 次取到白球. 显然对于第一次取球的结果而言, R_1, W_1 构成样本空间 S_1 的一个划分, 而 $P(R_1) = \dfrac{4}{10}, P(W_1) = \dfrac{6}{10}$. 显然

$$W_2 = W_2S = W_2(R_1 + W_1) = W_2R_1 + W_2W_1,$$

则

$$P(W_2) = P(W_2R_1) + P(W_2W_1),$$

根据乘法定理,

$$P(W_2) = P(W_2|R_1)P(R_1) + P(W_2|W_1)P(W_1),$$

根据试验的实际情况不难得出 $P(W_2|R_1)=\dfrac{6}{9},P(W_2|W_1)=\dfrac{5}{9}$, 因此

$$P(W_2)=\frac{6}{9}\cdot\frac{4}{10}+\frac{5}{9}\cdot\frac{4}{10}=\frac{3}{5}.$$

而第二次取到白球, 第一次取到白球的概率

$$P(W_1|W_2)=\frac{P(W_1W_2)}{P(W_2)}=\frac{P(W_2|W_1)P(W_1)}{P(W_2)}=\frac{\dfrac{5}{9}\times\dfrac{6}{10}}{\dfrac{3}{5}}=\frac{5}{9}.$$

1.5　独立性

定义 1.5.1. (相互独立的随机事件). 设 $A,B\subset S$ 是两个事件, 如果满足等式

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

则称事件 A,B **相互独立**, 简称为 A,B 独立.

事件的独立性可以推广到多个事件, 设 $A,B,C\subset S$ 是三个事件, 如果满足等式

$$\begin{cases}P(AB)=P(A)P(B)\\P(BC)=P(B)P(C)\\P(AC)=P(A)P(C)\\P(ABC)=P(A)P(B)P(C)\end{cases},$$

则称 A,B,C 相互独立. 一般地, $n(n\geqslant 2)$ 个事件 $A_1,A_2,\cdots,A_n\subset S$ 相互独立的条件是, 如果对于任意的 $2,3,\cdots,n$ 个事件的积事件的概率都等于对应的各个事件概率之积, 则 $A_1,A_2,\cdots,A_n$ 相互独立.

推论 1. 若事件 $A_1,A_2,\cdots,A_n\subset S(n\geqslant 2)$ 相互独立, 则其中任意 $k(2\leqslant k\leqslant n)$ 个事件也是相互独立的.

推论 2. 若事件 $A_1,A_2,\cdots,A_n\subset S(n\geqslant 2)$ 相互独立, 则将其中任意多个事件换成它们的逆事件, 所得的 n 个事件仍然是相互独立的. 特别地, 若 A,B 独立, 则 $A,\overline{B}$ 独立, $\overline{A},B$ 独立, $\overline{A},\overline{B}$ 独立.

例 1.5.1. 设 $A,B,B\subset S$ 两两相互独立, 且 $ABC=\varnothing$, $P(A)=P(B)=P(C)<\dfrac{1}{2}$, 且 $P(A+B+C)=\dfrac{9}{16}$, 则 $P(A)=$ ____________.

解. 已知

$$P(A+B+C)=P(A)+P(B)+P(C)-P(Ab)-P(AC)-P(BC)+P(ABC),$$

而由于 $ABC=\varnothing$, 因此

$$P(ABC)=0,$$

又由于 A,B,C 两两相互独立, 因此

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

$$P(AC)=P(A)P(C),$$

$$P(BC)=P(B)P(C),$$

由于 $P(A)=P(B)=P(C)<\dfrac{1}{2}$, 不妨设 $p\xlongequal{\text{def}}P(A)=P(B)=P(C)<\dfrac{1}{2}$, 则根据 $P(A+B+C)=\dfrac{9}{16}$ 有一元二次方程

$$3p-3p^3=\frac{9}{16},$$

解之得

$$p=P(A)=\frac{1}{4}.$$

注. 用求根公式解上述一元二次方程时, 注意 $P(A)<\dfrac{1}{2}$ 的限制条件, 据此从两个解中选出符合要求的一个来.

例 1.5.2. A,B 两两独立, 且 A,B 都不发生的概率为 $\dfrac{1}{9}$, A 发生 B 不发生的概率与 A 不发生 B 发生的概率相等, 求 $P(A)=$ ____________.

解. 由于 A,B 独立, 因此 $\bar{A},\bar{B}$ 独立, $\bar{A},B$ 独立, $A,\bar{B}$ 也独立, 所以

$$P(\bar{A})P(\bar{B})=\frac{1}{9},$$

$$P(A\bar{B})=P(A)P(\bar{B})=P(\bar{A}B)=P(\bar{A})P(B),$$

即

$$P(A)(1-P(B))=P(B)(1-P(A)),$$

因此

$$P(A)=P(B),$$

不妨设 $p\xlongequal{\text{def}}P(A)$, 则有

$$(1-p)^2=\frac{1}{9},$$

解该一元二次方程, 得

$$P(A)=p=\frac{2}{3}.$$

注. 该一元二次方程有两个解 $\dfrac{4}{3}$ 和 $\dfrac{2}{3}$, 但由于概率 $p\leqslant 1$, 所以取 $p=\dfrac{2}{3}$.

例 1.5.3. 设随机事件 A 与 B 相互独立, 且 $0<P(A),P(B)<1$, 则 $P\left(A|B\right)+P(\bar{A}|\bar{B})=$____________.

解. 根据条件概率公式,

$$P\left(A|B\right)+P(\bar{A}|\bar{B})=\frac{P(AB)}{P(A)}+\frac{P(\bar{A}\bar{B})}{P(\bar{B})},$$

由于 A,B 相互独立, 所以 $A,\bar{B}$、$\bar{A},B$ 和 $\bar{A},\bar{B}$ 也相互独立, 因此

$$\text{上式}=\frac{P(A)P(B)}{P(B)}+\frac{P(\bar{A})P(\bar{B})}{P(\bar{B})},$$

显然 $\text{ans}=P(A)+P(\bar{A})=1$ 即为所求.

第 2 章　随机变量及其分布

2.1　随机变量

定义 2.1.1. (随机变量). 设随机试验的样本空间 $S=\{e\}$,

$$\begin{aligned} X: S &\to \mathbb{R}^1 \\ e &\mapsto X(e) \end{aligned}$$

为定义在样本空间 S 上的实值单值函数, 如果对于任意的实数 x, 事件 $\{e|X \leqslant x\}$ 有确定的概率, 则称 $X=X(e)$ 为**随机变量** (Random Variable).

对于样本空间 S 中任意的样本点 e, 都有唯一的实数 $X(e)$ 与之对应, 这里要注意的是, 随机变量所构造的函数关系未必是单射, 这就是说, 随机试验的一个结果仅有 1 个实数与之对应, 但不同随机试验的结果 (即样本空间中的不同元素) 可能对应的是同一个数值. X 的取值是随机的, 每个可能的取值都有确定的概率. 随机变量 X 在一定范围内的取值本质上就是随机事件.

2.2　离散型随机变量及其分布律

2.2.1　离散型随机变量

定义 2.2.1. (离散型随机变量的分布律). 可能取到的值是有限个或无限可列多个的随机变量称为**离散型随机变量**. 设离散型随机变量 X 可能取到的值为 $x_1, x_2, \cdots, X$ 取各个值的概率为 $P\{X=x_1\}, P\{X=x_2\}, \cdots$, 则

$$p_k \xlongequal{\text{def}} P\{X=x_k\}$$

称为 X 的分布律.

离散型随机变量的分布律可以用表格表示为

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	$\cdots$

性质 1. 根据概率的定义, p_k 满足条件:

a) $p_k \geqslant 0, k=1,2,\cdots$;

b) $\sum\limits_{k=1}^{\infty} p_k = 1$.

证明. 显然 $\bigcup\limits_{k=1}^{\infty}\{X=x_k\}$ 为必然事件, 且随机变量 X 为单值函数, 因此不存在 $e\in S$ 使得

$$(X(e)=x_k)\wedge(X(e)=x_j)\,,\quad (x_k\neq x_j),$$

因此对于任意的 $i,j\in\{1,2,\cdots\}^2$, 当 $k\neq j$ 时有 $\{X=x_j\}\cap\{X=x_k\}=\varnothing$. □

2.2.2 三种重要的离散型随机变量

1. $0-1$ 分布

随机变量 X 只可能取 $0,1$ 两个值, 它的分布律是

$$P\{X=k\}=p^k(1-p)^{1-k},k=0,1\quad(0<p<1)$$

则称 X 服从以 p 为参数的 $0-1$ 分布, $0-1$ 分布用于观察一件事情是否发生. 它的分布律也可以写成

X	0	1
p_k	$1-p$	p

2. Bernoulli 试验、二项分布

如果随机试验 E 只有两个可能的结果, A 与 $\overline{A}$, 则称 E 为 **Bernoulli 试验**. 设 $P(A)=p(0<p<1)$, 将 E 独立地重复进行 n 次, 则称这一串重复的独立试验为 n 重 Bernoulli 试验. 以 X 表示 n 重 Bernoulli 试验中事件 A 发生的次数, 则随机变量 X 的分布律为

$$P\{X=k\}=\mathrm{C}_n^kp^k(1-p)^{n-k},\quad k=0,1,2,\cdot,n,$$

则 X 服从参数为 n,p 的**二项分布**, 记作 $X\sim b(n,p)$.

当 $n=1$ 是二项分布退化为 $0-1$ 分布.

3. Poisson 分布

设随机变量 X d 饿所有可能取值为 $0,1,2,\cdots$, 它的分布律为

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

其中 $\lambda>0$ 是一个常数, 则称 X 服从以 λ 为参数的 **Poisson 分布**, 记为 $X\sim\pi(\lambda)$.

Poisson 分布常用语描述单位时间 (空间) 内随机事件发生的次数, 参数 λ 的实际意义是单位时间内随机事件发生的平均次数.

定理 2.2.1. 设 $\lambda>0$ 是一个常数, n 为任意正整数, 设 $np_n=\lambda$, 则对于任意固定的非负整数 k, 有

$$\lim_{n\to\infty}\mathrm{C}_n^kp_n^k(1-p_n)^{n-k}=\frac{\lambda^k\mathrm{e}^{-\lambda}}{k!},\tag{2.1}$$

即如果随机变量 $X\sim b(n,p)$, 当 n 很大 (一般 $n\geqslant 50$ 视为“很大”), p 很小时, X 近似服从参数为 $\lambda=np$ 的 Poisson 分布, 即

$$P\{X=k\}=\mathrm{C}_n^kp^k(1-p)^{n-k}\approx\frac{(np)^k\mathrm{e}^{-np}}{k!}.$$

2.3 随机变量的分布函数

定义 2.3.1. (随机变量的分布函数). 设 X 是一个随机变量, 称

$$F(x)=P\left\{X\leqslant x\right\},\quad -\infty<x<+\infty \tag{2.2}$$

为随机变量 X 的**分布函数**.

分布函数 $F(x)$ 实际表示随机变量 X 所有可能取值中, 小于等于 x 的所有取值的概率的累加, 即

$$F(x)=\sum_{X_i\leqslant x}P\left\{X=X_i\right\},$$

如果 X 是离散型随机变量, 则对于级数求和, 如果 X 是连续型随机变量, 则对应积分. 由分布函数的定义式 (2.2), 对于任意的 $a,b\in\mathbb{R}$, 不妨设 $a<b$,

$$P\left\{a<X\leqslant b\right\}=F(b)-F(a).$$

随机变量的引入使我们能用随机变量来描述各种现象, 并能利用分析方法对随机试验的结果进行深入广泛的研究和讨论. 研究随机变量及其分布函数是概率论的主要内容.

性质 1. $0\leqslant F(x)\leqslant 1, x\in\mathbb{R}$;

性质 2. $F(x)$ 单调不减的.

性质 3. $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0, F(+\inf)=\lim\limits_{x\to+\infty}F(x)=1.$

性质 4. $F(x)$ 右连续, 即 $\lim\limits_{x\to x_0^+}F(x)=F(x_0)$, 其中 $x\in\mathbb{R}$.

2.4 连续型随机变量及其概率密度

2.4.1 连续型随机变量

如果对于随机变量 X 的分布函数 $F(x)$, 存在非负可积函数 $f(x)$, 使得对于任意的实数 x,

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

则称 X 为**连续型随机变量**, 其中 $f(x)$ 称为 X 的**概率密度函数**, 简称概率密度.

性质 1. $f(x)\geqslant 0$.

性质 2. $\displaystyle\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1.$

性质 3. $\forall x_1,x_2\in\mathbb{R}\left(x_1<x_2\Longrightarrow P\left\{x_1<X<x_2\right\}=\displaystyle\int_{x_1}^{x_2}f(t)\mathrm{d}t.\right)$

证明. 显然 $P\left\{x_1<X<x_2\right\}=F(x_2)-F(x_1)$. □

性质 4. 如果 $f(x)$ 在点 x 处连续, 则 $F'(x)=f(x)$.

注. $\forall x_0 \in \mathbb{R}\,(P\{X = x_0\} = 0)$, 即对于任意给定的单点 x_0, 随机变量 X 取到 x_0 的概率为 0. 因此 X 在区间上的概率与区间端点是否取得到无关, 即 $P\{X \in (x_1, x_2)\} = P\{X \in [x_1, x_2)\} = P\{X \in (x_1, x_2]\} = P\{X \in [x_1, x_2]\}$.

例 2.4.1. 设连续型随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} kx+1, & x \in [0,2) \\ 0, & \text{else} \end{cases},$$

求: (1) 常数 k, (2) X 的分布函数 $F(x)$, (3) $P\{1 < X < 2\}$.

解. (1) 由归一性条件,

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \iff \int_0^2 kx + 1\mathrm{d}x = 1 \iff k = -\frac{1}{2},$$

(2) 根据概率密度函数的定义,

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \begin{cases} 0, & x \in (-\infty, 0) \\ -\frac{1}{4}x^2 + x, & x \in [0, 2) \\ 1, & x \in [2, +\infty) \end{cases}$$

(3) 根据概率密度函数或分布函数的定义,

$$P\{1 < X < 2\} = \int_1^2 \left(-\frac{x}{2} + 1\right)\mathrm{d}x = \frac{1}{4}.$$

2.4.2 常见的连续型随机变量的概率分布

1. 均匀分布

设连续型随机变量 X 具有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{else} \end{cases},$$

则称 X 在区间 (a, b) 上服从均匀分布, 记作 $X \sim U(a, b)$, 其分布函数为

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leqslant x < b \\ 1, & x \geqslant b \end{cases}.$$

2. 指数分布

若连续型随机变量 X 具有概率密度

$$f(x) = \begin{cases} \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), & x > 0, \\ 0, & \text{else}, \end{cases}$$

其中 $\theta > 0$ 为常数, 则称 X 服从参数为 θ 的**指数分布**, 记作 $X \sim E\left(\frac{1}{\theta}\right)$, 其分布函数为

$$F(x)=\begin{cases}1-\exp\left(-\frac{x}{\theta}\right), & x>0,\\ 0, & \text{else}.\end{cases}$$

服从指数分布的随机变量 X 具有无记忆性, 即

$$\forall s,t>0\,(P\{X>s+t|X>s\}=P\{X>t\}),$$

一个元件的使用寿命服从指数分布, 那么它已经工作了 s 小时, 还能连续工作 t 小时的概率, 就等于从最开始这个元件能工作 t 小时的概率, 只看时间间隔, 不看起始时间.

取 $\lambda=\frac{1}{\theta}$, 则指数分布可以表示为

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0\\ 0, & \text{else},\end{cases}\qquad F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x>0\\ 0, & \text{else}\end{cases}.$$

3. 正态分布

若连续型随机变量 X 的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{2.3}$$

其中 $\mu,\sigma(\sigma>0)$ 为常数, 则称 X 服从以 μ,σ 为参数的**正态分布** (normal distribution) 或 **Gauss 分布**, 记作 $X\sim N(\mu,\sigma^2)$.

正态分布的密度函数具有以下性质:

性质 1. 曲线关于 $x=\mu$ 对称, 当 $x=\mu$ 时 $f(x)$ 取得最大值 $\frac{1}{\sqrt{2\pi}\sigma}$.

性质 2. 曲线以 x 轴为渐近线.

性质 3. 曲线在 $x=\mu\pm\sigma$ 处有拐点.

性质 4. $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$.

证明. 显然 $f(x)\geqslant 0$, 令 $t=\frac{x-\mu}{t}$, 得到

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-p)^2}{2\sigma^2}\right)\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t.$$

记 $I=\int_{-\infty}^{+\infty}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t$, 则 $I^2=\int_{-\infty}^{+\infty}\exp\left(-\frac{(t^2+u^2)}{2}\right)\mathrm{d}t\mathrm{d}u$, 利用极坐标将其转化为累次积分, 得到

$$I^2=\int_0^{2\pi}\int_0^{\infty} r\exp\left(-\frac{r^2}{2}\right)\mathrm{d}r\mathrm{d}\theta=2\pi.$$

而 $I>0$, 因此有 $I=\sqrt{2\pi}$, 即有

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-p)^2}{2\sigma^2}\right) \mathrm{d}x = \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} = 1,$$

这说明正态分布概率密度曲线下方无穷区域的面积为 1. □

性质 5. 固定 σ, 改变 μ 的大小, $f(x)$ 形状不变, 沿着 x 轴平移; 固定 μ, 改变 σ 的大小, $f(x)$ 不发生平移, 以 $x=\mu$ 为中心拉伸, σ 越小, $f(x)$ 越"高、瘦", σ 越大, $f(x)$ 越"矮、胖".

取正态分布中的 $\mu=0,\sigma^2=1$, 则称为标准正态分布, 记作 $X\sim N(0,1)$, 此时密度函数

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

分布函数为

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x.$$

标准正态分布具有如下性质:

性质 1. $\varphi(x)$ 为偶函数, $\max\varphi(x)=\varphi(0)=\dfrac{1}{\sqrt{2\pi}}$.

性质 2. $\displaystyle\int_{-\infty}^{+\infty} \varphi(x)\mathrm{d}x = 1$.

性质 3. 分布函数满足 $\Phi(-x)=1-\Phi(x)$.

由于一般的正态分布在计算上不方便, 我们往往将一般的正态分布"标准化", 我们指出, 如果 $X\sim N(\mu,\sigma^2)$, 那么 $Z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$, 这时我们有:

1° $F(x)=P\{X\leqslant x\}=P\left\{\dfrac{X-\mu}{\sigma}\leqslant\dfrac{x-\mu}{\sigma}\right\}=P\left\{Z\leqslant\dfrac{x-\mu}{\sigma}\right\}=\Phi\left(\dfrac{x-\mu}{\sigma}\right)$.

2° $P\{x_1<X<x_2\}=P\left\{\dfrac{x_1-\mu}{\sigma}<\dfrac{X-\mu}{\sigma}<\dfrac{x_2-\mu}{\sigma}\right\}=\Phi\left(\dfrac{x_2-\mu}{\sigma}\right)-\Phi\left(\dfrac{x_1-\mu}{\sigma}\right)$.

3° $P\{X\geqslant x\}=1-P\{X<x\}=1-P\left\{\dfrac{X-\mu}{\sigma}<\dfrac{x-\mu}{\sigma}\right\}=1-\Phi\left(\dfrac{x-\mu}{\sigma}\right)$.

例 2.4.2. 设 $X\sim N(3,2^2)$, 求: (1) $P\{2<X\leqslant 5\}$; (2) 确定常数 c 使得 $P\{X>c\}=P\{X\leqslant c\}$.

解. 由于 $X\sim N(3,2^2)$, 故 $Z=\dfrac{X-3}{2}\sim N(0,1)$, 因此

$$\begin{aligned}
P\{2<X\leqslant 5\} &= P\left\{\frac{2-3}{2}<\frac{X-3}{2}\leqslant\frac{5-3}{2}\right\} \\
&= P\left\{-\frac{1}{2}<Z\leqslant 1\right\} \\
&= \Phi(1)-\Phi\left(-\frac{1}{2}\right) \\
&= \Phi(1)-\left(1-\Phi\left(\frac{1}{2}\right)\right) \\
&= 0.8413+0.6915-1 \\
&= 0.5328,
\end{aligned}$$

由题意知

$$P(X>c)=1-P\{X\leqslant c\}=P\{X\leqslant c\},$$

即 $P\{X\leqslant c\}=\dfrac{1}{2}$, 根据正态分布的对称性可知, $c=3$.

例 2.4.3. 设随机变量 $X\sim N(\mu,\sigma^2)$, 则随着 σ 的增大, 概率 $P\{|x-\mu|<\sigma\}$ 的取值保持不变.

证明. 将该正态分布标准化, 得

$$P\{|X-\mu|<\sigma\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|<1\right\}=P\left\{-1<\frac{X-\mu}{\sigma}<1\right\}=\Phi(1)-\Phi(-1),$$

而 $\Phi(1)-\Phi(-1)$ 是一个常数. □

注. $\forall a\in\mathbb{R}\,(\Phi(a)-\Phi(-a)=\Phi(a)-(1-\Phi(a))=2\Phi(a)-1)$, 这是一个比较常用的结论, 请各位考生加以记忆.

例 2.4.4. 一工厂生产的元件寿命 X (小时), 服从参数为 $\mu=160$, σ 未知的正态分布, 若想要保证元件寿命在 120 到 200 小时之间的概率不低于 0.8, 则 σ 最大能取多少?

解. 由题意知,

$$P\{120\leqslant X\leqslant 200\}\geqslant 0.8,$$

将正态分布标准化,

$$\begin{aligned}P\left\{\frac{120-160}{\sigma}\leqslant\frac{X-160}{\sigma}\leqslant\frac{200-160}{\sigma}\right\}&=\Phi\left(\frac{40}{\sigma}\right)-\Phi\left(-\frac{40}{\sigma}\right)\\&=2\Phi\left(\frac{40}{\sigma}\right)-1\geqslant 0.8\end{aligned}$$

即 $\Phi\left(\dfrac{40}{\sigma}\right)\geqslant 0.9$, 查表可知要使该式成立, 只要 $\dfrac{40}{\sigma}>1.29$, 即 $\sigma<31$.

2.5　随机变量的函数的分布

2.5.1　离散型随机变量函数的分布

例 2.5.1. 已知随机变量 X 的分布律如下表所示

X	0	1	2	3	4	5
P	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

求 $Y=(X-2)^2$ 的分布律.

解. 由已知得

$Y=(X-2)^2$	4	1	0	1	4	9
X	0	1	2	3	4	5
P	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

整理得到

$Y=(X-2)^2$	0	1	4	9
P	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{11}{36}$	$\frac{1}{9}$

例 2.5.2. 已知随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<-1,\\ 1/3, & -1\leqslant x<0,\\ 1/2, & 0\leqslant x<1,\\ 2/3, & 1\leqslant x<2,\\ 1, & x\geqslant 2,\end{cases}$$

求 $Y=\sin^2\left(\frac{\pi}{6}X\right)$ 的分布律.

解. 由已知, 得 X 的分布律为

X	-1	0	1	2
P	1/3	1/6	1/6	1/3

于是

$Y=\sin^2\left(\frac{\pi}{6}X\right)$	1/4	0	1/4	3/4
X	-1	0	1	2
P	1/3	1/6	1/6	1/3

整理得

$Y=\sin^2\left(\frac{\pi}{6}X\right)$	0	1/4	3/4
P	1/6	1/2	1/3

于是 Y 的分布函数为

$$F_Y(y)=\begin{cases}0, & y<0,\\ 1/6, & 0\leqslant y<1/4,\\ 2/3, & 1/4\leqslant y<3/4,\\ 1, & y\geqslant 3/4.\end{cases}$$

2.5.2 连续型随机变量函数的分布

设随机变量 Y 是随机变量 X 的函数 $Y=g(X)$, 其中 $g\in \mathrm{C}(-\infty,+\infty)$ 为连续函数, 设 X 的分布已知, 要求 $Y=g(X)$ 的分布, 则

(1) 求出 Y 的分布函数 $F_Y(y)=P\{Y\leqslant y\}$;

(2) 将 $F_Y(y)$ 关于 y 求导, 则 $\dfrac{\mathrm{d}}{\mathrm{d}y}F_Y(y)$ 就是 Y 的概率密度 $f_Y(y)$.

特别地, 当 g 是严格单调函数时, 有如下定理:

定理 2.5.1. 设随机变量 X 具有概率密度 $f_X(x)(x\in\mathbb{R})$, $g\in \mathrm{C}^1(-\infty,+\infty)$ 且 $\forall x\in\mathbb{R}(g'(x)>0\oplus g'(x)<0)$, 即 $g(x)$ 严格单调, 则 $Y=g(X)$ 是连续型随机变量, 且其概率密度为

$$f_Y(y)=\begin{cases}f_X[h(y)]\cdot|h'(y)|, & \alpha<y<\beta,\\ 0, & \text{else},\end{cases}$$

其中 $\alpha=\min\left(g(-\infty),g(+\infty)\right)$, $\beta=\max\left(g(-\infty),g(+\infty)\right)$, $h(y)$ 是 $g(x)$ 的反函数.
若 X 的概率密度为

$$\begin{cases}f_X(x), & a<x<b,\\ 0, & \text{else},\end{cases}$$

则 $\alpha=\min\left(g(a),g(b)\right)$, $\beta=\max\left(g(a),g(b)\right)$.

例 2.5.3. 设随机变量 X 的分布函数为 $F(x)$, 则随机变量 $Y=2X+1$ 的分布函数为 ____________.

解. 由分布函数的定义,

$$G(y)=P\{Y\leqslant y\}=P\{2X+1\leqslant y\}=P\left\{X\leqslant\frac{y-1}{2}\right\}=F\left(\frac{y}{2}-\frac{1}{2}\right).$$

例 2.5.4. 设随机变量 X 的概率密度函数为

$$f_X(x)=\begin{cases}\mathrm{e}^{-x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

求随机变量 $Y=\mathrm{e}^X$ 的概率密度 $f_Y(y)$.

解. 我们先求出 Y 的分布函数 $F_Y(y)$, 根据分布函数的定义,

$$F_Y(y)=P\{Y\leqslant y\}=P\left\{\mathrm{e}^X\leqslant y\right\}=P\{X\leqslant\ln y\}=\int_{-\infty}^{\ln y}f(x)\mathrm{d}x.$$

对 $\ln y$ 的正负值情况进行分类的讨论, 当 $\ln y \leqslant 0$, 即 $y \leqslant 1$ 时,

$$F_Y(y)=\int_{-\infty}^{\ln y} f(x)\mathrm{d}x=\int_{-\infty}^{\ln y} 0\mathrm{d}x=0,$$

当 $\ln y\ 0$, 即 $y>1$ 时,

$$F_Y(y)=\int_{-\infty}^{\ln y} f(x)\mathrm{d}x=\left(\int_{-\infty}^{0}+\int_{0}^{\ln y}\right) f(x)\mathrm{d}x=0+\int_{0}^{\ln y} \mathrm{e}^{-x}\mathrm{d}x=1-\frac{1}{y},$$

于是

$$F_Y(y)=\begin{cases}0, & y \leqslant 1, \\ 1-\dfrac{1}{y}, & y>1.\end{cases}$$

因此 Y 的概率密度函数

$$f_Y(y)=\begin{cases}0, & y \leqslant 1, \\ \dfrac{1}{y^2}, & y>1.\end{cases}$$

注. 由于 X 的概率密度函数是分段函数, 因此一定要注意对 Y 的分段处理, 从 $Y \leqslant y$ 解出 $X \leqslant h(y)$ 后, 一定要按照 $f_X(x)$ 的分段情况对 $h(y)$ 的取值进行分类讨论.

第 3 章　多维随机变量及其分布

3.1　二维随机变量

3.1.1　二维随机变量的联合分布

定义 3.1.1. (二维随机变量). 设随机试验 E 的样本空间是 $S = \{e\}$, 设 $X = X(e)$, $Y = Y(e)$ 是定义在样本空间 S 上的随机变量, 则由它们构成的向量 (X, Y) 称为**二维随机变量** (2 dimentional random variable) 或二位随机向量 (2 dimentioanl ramdom vector).

定义 3.1.2. (联合分布函数). 设 (X, Y) 是二维随机变量, x, y 是任意实数, 则二元函数

$$F(x,y) = P\left\{(x \leqslant x) \cap (Y \leqslant y)\right\} \xlongequal{\text{def}} P\left\{X \leqslant x, Y \leqslant y\right\}, \quad (x, y \in \mathbb{R})$$

称为二维随机变量 (X, Y) 的分布函数, 或称为随机变量 X 和 Y 的**联合分布函数**.

分布函数具有以下基本性质:

性质 1. $F(x, y)$ 是变量 x, y 的不减函数, 即对于任意固定的 y, 当 $x_2 > x_1$ 时有 $F(x_2, y) \geqslant F(x_1, y)$; 对于任意固定的 x, 当 $y_2 > y_1$ 时有 $F(x, y_2) \geqslant F(x, y_1)$. 这就是说:

$$\forall y \in \mathbb{R}(x_2 > x_1 \Longrightarrow F(x_2, y) \geqslant F(x_1, y)),$$

$$\forall x \in \mathbb{R}(y_2 > y_1 \Longrightarrow F(x, y_2) \geqslant F(x, y_1)).$$

性质 2. $0 \leqslant F(x, y) \leqslant 1$, 且对于任意固定的 y, $F(-\infty, y) = 0$, 对于任意固定的 x, $F(x, -\infty) = 0$, 且有

$$F(-\infty, -\infty) = 0, \quad F(\infty, \infty) = 1.$$

性质 3. $F(x + 0, y) = F(x, y)$, $F(x, y + 0) = F(x, y)$, 即 $F(x, y)$ 关于 x 和 y 是右连续的.

性质 4. 对于任意的 x_1, x_2, y_1, y_2, 如果 $x_1 < x_2, y_1 < y_2$ 则有

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \geqslant 0,$$

这个值表示 $(X, Y) \in (x_1, x_2) \times (y_1, y_2)$ 的概率.

3.1.2　二维离散型随机变量及其分布律

1. 二维离散型随机变量

若二维随机变量 (X, Y) 全部可能取到的不同的值是有限多对或可列对, 则称 (X, Y) 是**二维离散型随机变量** (descrete 2-dimentional ramdom vector), 设 (X, Y) 所有可能取的值为 $(x_i, y_j), i, j = 1, 2, \cdots,$

则称

$$P\{X=x_i, Y=y_j\}=p_{ij}, \quad i,j=1,2,\cdots,$$

为二维离散型随机变量 (X,Y) 的**联合分布律**.

关于离散型二维随机变量的联合分布, 我们做两点说明:

(1) 联合分布律可以用列表给出:

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nn}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$

(2) 有时为了使用矩阵代数, 也可以用矩阵来表示联合分布律:

$$\boldsymbol{p} \xlongequal{\text{def}} [p_{ij}]=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2n} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ p_{n1} & p_{n2} & \cdots & p_{nn} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

后面我们将看到, 用矩阵来表示离散型随机变量的联合分布会让许多统计量的形式变得非常简洁, 同时矩阵形式也为我们在关于随机变量的研究中使用线性代数理论提供了支持.

2. 二维离散型随机变量分布律和分布函数的性质

性质 1. $\forall (x_i,y_j)\in \mathrm{allPossibleValue}(X,Y)\,(p_{ij}\geqslant 0)$, 其中 $\mathrm{allPossibleValue}(X,Y) \xlongequal{\text{def}} \{(x_i,y_j)|i,j=1,2,\cdots\}$ 为二维离散型随机变量 (X,Y) 可能取到的值的集合.

性质 2. $\displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij}=1$.

性质 3. $\displaystyle\forall x,y\in\mathbb{R}\left(F(x,y)=\sum_{x_i\leqslant x}\sum_{y_i\leqslant y} p_{ij}\right)$.

3.1.3　二维连续型随机变量及其概率密度

对于二维随机变量 (X,Y) 的分布函数 $F(x,y)$, 如果存在非负二元函数 $f(x,y)$ 使得对于任意的 x,y 有

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

则称 (X,Y) 为**二维连续型随机变量** (continious 2-dimentioanl random vector), 函数 $f(x,y)$ 称为二维连续型随机变量 (X,Y) 的**联合概率密度**.

性质 1. $\forall x,y\in\mathbb{R}\,(f(x,y)\geqslant 0)$.

性质 2. $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y = F(-\infty,+\infty)=1$.

性质 3. $\forall G\subset\mathbb{R}^2\left(P\{(X,Y)\in G\}=\displaystyle\iint_G f(x,y)\mathrm{d}x\mathrm{d}y\right)$.

性质 4. *如果 $f(x,y)$ 在点 (x,y) 连续, 则* $\dfrac{\partial^2}{\partial x\partial y}F(x,y)=f(x,y)$.

注. 由于在考察二维连续型随机变量在一定范围内取值的概率时需要使用积分, 因此们在研究二维连续型随机变量时, 要养成先观察其联合概率密度非零区域的习惯.

3.2 边缘分布

定义 3.2.1. 二维随机变量 (X,Y) 作为一个整体具有分布函数 $F(x,y)$, 而 X 和 Y 都是随机变量, 设它们的分布函数分别为 $F_X(x)$ 和 $F_Y(y)$, 则有

$$F_X(x)=P\{X\leqslant x\}=\lim_{y\to+\infty}F(x,y),\quad F_Y(y)=P\{Y\leqslant y\}=\lim_{x\to+\infty}F(x,y). \tag{3.1}$$

$F_X(x)$ 和 $F_Y(y)$ 分别称为二维随机变量 (X,Y) 关于 X, 关于 Y 的**边缘分布函数**.

3.2.1 二维离散型随机变量的边缘分布

对于离散型随机变量 (X,Y), 设 (X,Y) 的分布律为

$$P\{X=x_i,Y=y_i\}=p_{ij},\quad i,j=1,2,\cdots,$$

则根据全概率公式, X,Y 的分布律分别为

$$P\{X=x_i\}=\sum_{j=1}^{\infty}p_{ij}\overset{\text{def}}{=\!=}p_{i\cdot},i=1,2,\cdots,$$
$$P\{Y=y_j\}=\sum_{i=1}^{\infty}p_{ij}\overset{\text{def}}{=\!=}p_{\cdot j},j=1,2,\cdots,$$

它们分别称为 (X,Y) 关于 X 和关于 Y 的**边缘分布律**. 用矩阵来表示联合分布律 $[p_{ij}]$ 时, 也可以记作

$$P\{X=x_i\}=\mathrm{sum}(\boldsymbol{p}[i,:]),$$

$$P\{Y=y_j\}=\mathrm{sum}(\boldsymbol{p}[:,j]).$$

如果用表格表示, 则边缘分布律刚好写在表格的“边缘”上, 即

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	$\cdots$	$X=x_i$
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nn}	$\cdots$	$p_{n\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
$Y=y_j$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot n}$	$\cdots$	1

我们把 X 的边缘分布律 $[p_{1\cdot}, p_{2\cdot}, \cdots, p_{n\cdot}]^{\mathrm{T}}$ 记作 $\boldsymbol{p}_X$, 类似地把 Y 的边缘分布律记作 $\boldsymbol{p}_Y$.

3.2.2　二维连续型随机变量的边缘分布

对于连续型随机变量 (X,Y), 设它的概率密度为 $f(x,y)$, 则 X,Y 的概率分布为

$$F_X(x) = F(x,+\infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y \right) \mathrm{d}x,$$

$$F_Y(y) = F(+\infty,y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x \right) \mathrm{d}y,$$

求它们的一阶导数, 便得到 X,Y 的概率密度分别为

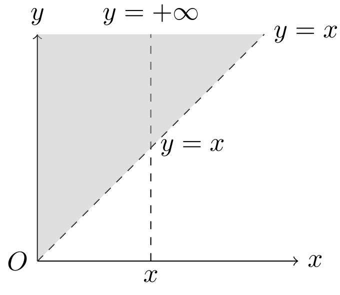
$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x,$$

它们分别称为 (X,Y) 关于 X 和关于 Y 的**边缘概率密度**.

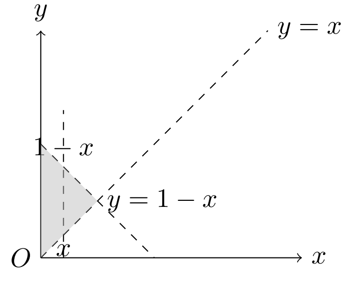
例 3.2.1. 设随机变量 (X,Y) 的概率密度为

$$f(x,y) = \begin{cases} \mathrm{e}^{-y}, & 0<x<y, \\ 0, & \text{else}, \end{cases}$$

求: (1) X 的概率密度; (2) $P\{X+Y \leqslant 1\}$.



(a)



(b)

解. 由已知得 $f(x,y)$ 的非零区域如上图 (a) 所示, 由 $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$,

a) 当 $x \leqslant 0$ 时, $f_X(x) = \int_{-\infty}^{+\infty} 0\mathrm{d}y = 0$.

b) 当 $x > 0$ 时, $f_X(x) = \int_{x}^{+\infty} \mathrm{e}^{-y}\mathrm{d}y = -\int_{x}^{+\infty} \mathrm{e}^{-y}\mathrm{d}(-y) = -\left.\mathrm{e}^{-y}\right|_x^{+\infty} = -\left(0 - \mathrm{e}^{-x}\right) = \mathrm{e}^{-x}$.

于是 X 的概率密度

$$f_X(x) = \begin{cases} \mathrm{e}^{-x}, & x > 0, \\ 0, & \text{else}. \end{cases}$$

接下来我们计算 $P\{X+Y \leqslant 1\}$, 对应的积分区域如上图 (b) 所示,

$$P\{X+Y \leqslant 1\} = \int_0^{1/2} \left(\int_x^{1-x} \mathrm{e}^{-y}\mathrm{d}y \right) \mathrm{d}x = \int_0^{1/2} \left(\mathrm{e}^{-x} - \mathrm{e}^{x-1}\right) \mathrm{d}x = 1 - \frac{2}{\mathrm{e}^{1/2}} + \frac{1}{\mathrm{e}}.$$

3.3 条件分布

3.3.1 二维离散型随机变量的条件分布律

设 (X, Y) 是二维离散型随机变量, 对于固定的 j, 如果 $P\{Y = y_j\} > 0$, 则根据条件概率公式,

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

为 $Y = y_j$ 条件下随机变量 X 的**条件分布律**. 同理, 对于固定的 i, 如果 $P\{X = x_i\} > 0$, 则称

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

为 $X = x_i$ 条件下随机变量 Y 的条件分布律.

当用矩阵 $\boldsymbol{p} = [p_{ij}]$ 来表示联合分布律时, 条件分布律

$$P\{X = x_k | Y = y_j\} = \frac{1}{\mathrm{sum}(\boldsymbol{p}[k, :])} \boldsymbol{p}[k, :] = \begin{bmatrix} p_{1j}/p_{\cdot j} \\ p_{2j}/p_{\cdot j} \\ \vdots \\ p_{nj}/p_{\cdot j} \\ \vdots \end{bmatrix},$$

$$P\{Y = y_k | X = x_i\} = \frac{1}{\mathrm{sum}(\boldsymbol{p}[k, :])} \boldsymbol{p}[k, :] = \begin{bmatrix} p_{i1}/p_{i\cdot} \\ p_{i2}/p_{i\cdot} \\ \vdots \\ p_{in}/p_{i\cdot} \\ \vdots \end{bmatrix}.$$

3.3.2 二维连续型随机变量的条件概率密度

设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y)$, (X,Y) 关于 Y 的边缘概率密度为 $f_Y(y)$. 若对于固定的 $y\in\mathbb{R}$, $f_Y(y)>0$ (即 $\forall y\in\mathbb{R}\,(f_Y(y)>0)$), 则称 $\dfrac{f(x,y)}{f_Y(y)}$ 为在 $Y=y$ 条件下 X 的**条件概率密度**, 记为

$$f_{X|Y}(x|y) \xlongequal{\text{def}} \frac{f(x,y)}{f_Y(y)}.$$

称 $\displaystyle\int_{-\infty}^{x} f_{X|Y}(x|y)\mathrm{d}x=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}$ 在 $Y=y$ 条件下 X 的**条件分布函数**, 记为

$$F_{X|Y}(x|y) \xlongequal{\text{def}} P\left\{X\leqslant x|Y=y\right\}=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\mathrm{d}x.$$

同理, 设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y)$, (X,Y) 关于 X 的边缘概率密度为 $f_X(x)$, 并且有 $\forall x\in\mathbb{R}\,(f_X(x)>0)$, 定义

$$f_{Y|X}(y|x) \xlongequal{\text{def}} \frac{f(x,y)}{f_X(x)},\quad F_{Y|X}(y|x) \xlongequal{\text{def}} P\left\{Y\leqslant y|X=x\right\}=\int_{-\infty}^{y}\frac{f(x,y)}{f_X(x)}\mathrm{d}y$$

为 $X=x$ 条件下 Y 的条件概率密度和条件分布函数.

3.4 相互独立的随机变量

定义 3.4.1. (相互独立的随机变量). 设 $F(x,y),F_X(x),F_Y(y)$ 分别为二维随机变量 (X,Y) 的分布函数及边缘分布函数, 如果对于任意的 $x,y\in\mathbb{R}$, 成立

$$P\left\{X\leqslant x,Y\leqslant y\right\}=P\left\{X\leqslant x\right\}P\left\{Y\leqslant y\right\},$$

即

$$F(x,y)=F_X(x)F_Y(y),$$

则称随机变量 X 和 Y **相互独立**.

3.4.1 二维离散型随机变量 (X,Y) 的相互独立性

定义 3.4.2. 如果对于二维离散型随机变量 (X,Y) 的所有可能取值 (x_i,y_i) 有

$$P\left\{X=x_i,Y=y_j\right\}=P\left\{X=x_i\right\}P\left\{Y=y_j\right\},$$

则称离散型随机变量 X,Y 相互独立.

注. 联合分布律 = 边缘分布律的乘积.

注. 如果 X 的可能取值有 n 个, Y 的可能取值有 M 个, 则 X,Y 相互独立要求 $n\times m$ 个式子同时成立. 相对的, 如果这些式子中至少有一个不成立, 则说明 X,Y 不相互独立.

注. 分布律使用矩阵表示时, (X,Y) 相互独立的条件是

$$p_{ij}=p_{i\cdot}p_{\cdot j}=\mathrm{sum}(\boldsymbol{p}[i,:])\mathrm{sum}(\boldsymbol{p}[:,j]),$$

即概率 p_{ij} 等于所在行的行和与所在列列和的乘积, 也就是

$$\boldsymbol{p} = \boldsymbol{p}_X \left(\boldsymbol{p}_Y\right)^{\mathrm{T}}.$$

例 3.4.1. 设随机变量 (X,Y) 具有分布律

$$P\left\{X=x,Y=y\right\} = p^2(1-p)^{x+y-2}, \quad 0<p<1, x,y\in\mathbb{N}^+,$$

验证 X,Y 是否相互独立.

证明. 可以计算

$$\begin{aligned}
P\left\{X=x\right\} &= \sum_{n=1}^{\infty} P\left\{X=x,Y=n\right\} = \sum_{n=1}^{\infty} p^2(1-p)^{x+n-2} \\
&= \sum_{n=1}^{\infty} p^2(1-p)^{x-2}(1-p)^n \xlongequal[\text{由 } 0<p<1 \text{ 易得 } |1-p|<1]{\text{等比级数求和}} \frac{p^2(1-p)^{x-2}(1-p)}{1-(1-p)} \\
&= p(1-p)^{x-1}.
\end{aligned}$$

同理, $P\left\{Y=y\right\} = p(1-p)^{y-1}$, 易得

$$P\left\{X=x,Y=y\right\} = p^2(1-p)^{x+y-2} = \left(p(1-p)^{x-1}\right)\left(p(1-p)^{y-1}\right) = P\left\{X=x\right\}P\left\{Y=y\right\},$$

因此 X,Y 相互独立. □

3.4.2 二维连续型随机变量 (X,Y) 的相互独立性

定义 3.4.3. 设连续型随机变量 (X,Y) 的联合概率密度为 $f(x,y)$, 边缘概率密度为 $f_X(x), f_Y(y)$, 如果

$$\forall x,y\in\mathbb{R}\left(f(x,y) = f_X(x)f_Y(y)\right),$$

即等式 $f(x,y)=f_X(x)f_Y(y)$ 在平面上几乎处处成立, 则称 X 与 Y 相互独立.

注. 验证连续型随机变量 X,Y 相互独立的方法:

(1) 求二维连续型随机变量的联合概率密度函数 $f(x,y)$;

(2) 根据联合概率密度函数求边缘概率密度函数 $f_X(x), f_Y(y)$;

(3) 验证等式 $f(x,y)=f_X(x)f_Y(y)$ 是否对任意的实数 x,y 成立.

显然这个定义与本节一开始引入的条件 $F(x,y)=F_X(x)F_Y(y)$ 是等价的. 于是我们可以给出另外一种验证连续型随机变量 X,Y 相互独立的方法:

(1) 求二维连续型随机变量的联合概率分布函数 $F(x,y)$;

(2) 根据联合概率密度函数求边缘概率分布函数 $F_X(x), F_Y(y)$;

(3) 验证等式 $F(x,y)=F_X(x)F_Y(y)$ 是否对任意的实数 x,y 成立.

例 3.4.2. 设二维随机变量 (X,Y) 具有概率密度

$$f(x,y)=\begin{cases}2\mathrm{e}^{-(2x+y)}, & x>0,y>0\\ 0, & \text{else}\end{cases},$$

证明 X 和 Y 相互独立.

证明. 根据边缘概率密度的计算公式, 知道

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\int_0^{+\infty}f(x,y)\mathrm{d}y, & x>0,\\ 0, & \text{else}\end{cases}=\begin{cases}\int_0^{+\infty}2\mathrm{e}^{-(2x+y)}\mathrm{d}y, & x>0\\ 0, & \text{else}\end{cases},$$

其中,

$$\int_0^{+\infty}2\mathrm{e}^{-(2x+y)}\mathrm{d}y=2\mathrm{e}^{-2x}\int_0^{+\infty}\mathrm{e}^{-y}\mathrm{d}y=2\mathrm{e}^{-2x},\quad (x>0)$$

因此

$$f_X(x)=\begin{cases}2\mathrm{e}^{-2x}, & x>0\\ 0, & \text{else}\end{cases},$$

同理可求得

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x=\begin{cases}\mathrm{e}^{-y}, & y>0\\ 0, & \text{else}\end{cases},$$

对于任意给定的 $x,y\in\mathbb{R}$, 有

$$f(x,y)=f_X(x)f_Y(y),$$

因此随机变量 X,Y 相互独立. □

例 3.4.3. 已知二维连续型随机变量 (X,Y) 的分布函数为

$$F(x,y)=\begin{cases}(1-\mathrm{e}^{-ax})y, & x>0,0\leqslant y\leqslant 1\\ 1-\mathrm{e}^{-ax}, & x\geqslant 0,y>1\\ 0, & \text{else}\end{cases},\quad (a>0)$$

求证 X,Y 相互独立.

证明. 根据边缘分布函数的计算公式, 得到

$$F_X(x)=\lim_{y\to+\infty}F(x,y)=\begin{cases}1-\mathrm{e}^{-ax}, & x\geqslant 0\\ 0 & \text{else}\end{cases},\quad F_Y(x)=\lim_{x\to+\infty}F(x,y)=\begin{cases}y, & 0\leqslant y\leqslant 1\\ 1, & y>1\\ 0, & \text{else}\end{cases},$$

显然对于任意给定的 $x,y\in\mathbb{R}$, 有 $F(x,y)=F_X(x)F_Y(y)$, 即 X,Y 相互独立. □

例 3.4.4. 设二维随机变量 (X,Y) 的联合概率密度为

$$f(x,y)=\begin{cases}kx^2y, & x^2<y<1,\\ 0, & \text{else},\end{cases}$$

(1) 求参数 k 的取值;
(2) 验证 X,Y 是否独立 (写出详细推导过程).

解. (1) 由归一化条件,

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

即

$$k\int_{-1}^{1}\left(\int_{x^2}^{1}x^2y\mathrm{d}y\right)\mathrm{d}x=1.$$

解之得到 $k=\dfrac{21}{4}$.
(2) 接下来我们求解 X,Y 的边缘概率密度, 当 $x\in(-1,1)$, 即 $x^2\in(0,1)$ 时,

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\frac{21}{8}x^2(10x^3),$$

于是

$$f_X(x)=\begin{cases}\dfrac{21}{8}x^2(1-x^4), & x\in(-1,1),\\ 0, & \text{else},\end{cases}$$

类似地,

$$f_Y(y)=\begin{cases}\dfrac{7}{2}y^{5/2}, & y\in(0,1),\\ 0, & \text{else},\end{cases}$$

显然 $f(x,y)=f_X(x)f_Y(y)$ 在 $\mathbb{R}^2$ 上并不处处成立, 故 X,Y 不独立.

3.5 二维随机变量的正态分布和均匀分布

3.5.1 二维正态分布

定义 3.5.1. (二维正态分布). 设二维随机变量 (X,Y) 的概率密度为

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

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 都是常数, 且 $\sigma_1,\sigma_2>0$, $-1<\rho<1$, 则称 (X,Y) 服从以 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 为参数的**二维正态分布**, 记为 $(X,Y)\sim N\left(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho\right)$. 特别地, 当 $\mu_1=\mu_2=0,\sigma_1=\sigma_2=1$ 时称 (X,Y) 服从二维标准正态分布 $N(0,1;0,1;\rho)$.

注. 回忆 $X\sim N(\mu,\sigma^2)\iff f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$.

注. 指数中的部分在形式上是与完全平方公式有一点类似的.

性质 1. $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 则 $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$. 即服从二维正态分布的二维随机变量 (X,Y) 的边缘分布是两个唯一确定的一维正态分布.

注. 给定两个服从正态分布的随机变量 $X\sim N(\mu_1,\sigma_1^2), Y\sim(\mu_2,\sigma_2^2)$, 不能确定 (X,Y) 是否服从二维正态分布.

注. 对于二维正态分布, 不同的 ρ 对应着不同的二维正态分布, 但其他参数不变, 仅 ρ 不同时, 边缘分布 $X\sim N(\mu_1,\sigma_1^2), Y\sim(\mu_2,\sigma_2^2)$ 是可以唯一确定的.

性质 2. 取 $\rho=0$, 则 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;0)$, 此时 $X\sim N(\mu_1,\sigma_1^2), Y\sim(\mu_2,\sigma_2^2)$, 并且有

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right),\quad f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right),$$

此时对于任意的 $x,y\in\mathbb{R}$, 成立

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right)=f_X(x)f_Y(y),$$

即对于二维正态随机变量 (X,Y), X,Y 相互独立的充分必要条件是 $\rho=0$.

3.5.2　二维均匀分布

定义 3.5.2. (二维均匀分布). 设 G 是平面上的有界区域, 其“面积”为 $A=m(G)$), 若二维随机变量 (X,Y) 具有概率密度

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in G\\ 0, & x\in\mathbb{R}^2\setminus G\end{cases},$$

则称 (X,Y) 在 G 上服从均匀分布.

性质 1. $\forall D\subset G\left(P\left\{(x,y)\in D\right\}=\dfrac{m(D)}{A}\right)$, 其中 $m(D)$ 为区域 D 的测度, 即一定意义下的“面积”.

注. 回忆随机变量 $X\sim U(a,b) \iff f(x)=\begin{cases}\dfrac{1}{b-a}, & x\in(a,b)\\ 0, & x\in\mathbb{R}\setminus(a,b)\end{cases}$, 并且一维均匀分布连续型随机变量具有重要性质: $\forall c\in(a,b)\forall d\in(c,b)\left(P\left\{c<X<d\right\}=\dfrac{d-c}{b-a}\right)$.

例 3.5.1. (X,Y) 在 $G=\left\{(x,y)\in\mathbb{R}^2|x^2+y^2\leqslant 1\right\}$ 上服从均匀分布, 则

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & (x,y)\in G\\ 0, & (x,y)\in\mathbb{R}^2\setminus G\end{cases}.$$

3.6　二维随机变量的函数的分布

3.6.1　二维离散型随机变量函数的分布

我们通常利用列表法来求解两个离散型随机变量函数的分布:

(X,Y)	$P\{X=x_i,Y=y_j\}$	$f_1(X,Y)$	$f_2(X,Y)$	$\cdots$
(x_1,y_1)	p_{11}	$f_1(x_1,y_1)$	$f_2(x_1,y_1)$	$\vdots$
(x_1,y_2)	p_{12}	$f_1(x_1,y_2)$	$f_2(x_1,y_2)$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
(x_2,y_1)	p_{21}	$f_1(x_2,y_1)$	$f_2(x_2,y_1)$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
(x_i,y_j)	p_{ij}	$f_1(x_i,y_j)$	$f_2(x_i,y_j)$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

完成列表后, 利用概率的可列可加性求解对应函数的分布律.

3.6.2 二维连续型随机变量函数的分布

一般地, 设二维连续型随机变量 (X,Y) 的概率密度函数为 $f(x,y)$, 则 $Z=g(X,Y)$ 的分布函数为

$$F_Z(z)P\{Z\leqslant z\}=\iint_{g(x,y)\leqslant z}f(x,y)\mathrm{d}x\mathrm{d}y,$$

其概率密度 $f_Z(z)=F_Z'(z)$.

1. $Z=X+Y$ 的分布

设 (X,Y) 的概率密度为 $f(x,y)$, 则 $Z=X+Y$ 的概率密度为

$$f_Z(z)=\int_{-\infty}^{+\infty}f(z-y,y)\mathrm{d}y$$

或

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)\mathrm{d}x.$$

特别地, 当 X,Y 相互独立时, 设它们的概率密度分别为 $f_X(x),f_Y(y)$, 则

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(z-y)f_Y(y)\mathrm{d}y,$$

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\mathrm{d}x.$$

它们称为**卷积公式** (convolution equation), 记作 f_X*f_Y, 即

$$f_X*f_Y=\int_{-\infty}^{+\infty}f_X(z-y)f_Y(y)\mathrm{d}y=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\mathrm{d}x.$$

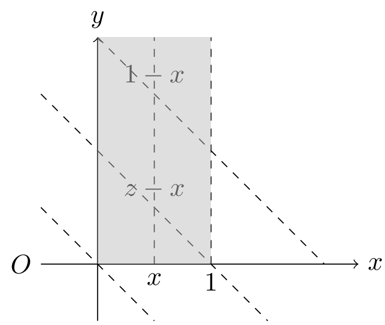
例 3.6.1. 设随机变量 X,Y 相互独立, 其概率密度分别为

$$f_X(x)=\begin{cases}1, & 0\leqslant x\leqslant 1,\\ 0, & \text{else},\end{cases},\quad f_Y(y)=\begin{cases}\mathrm{e}^{-y}, & y>0,\\ 0, & \text{else},\end{cases}$$

求 $Z=X+Y$ 的概率密度函数.

解. (分布函数法). 显然 (X,Y) 的分布函数为

$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}\mathrm{e}^{-y}, & 0\leqslant x\leqslant 1\wedge y>0,\\ 0, & \text{else},\end{cases}$$



概率密度非零区域如图所示.

$$F_Z(z)=P\{Z\leqslant z\}=P\{X+Y\leqslant z\}=\iint_{x+y\leqslant z}f(x,y)\mathrm{d}x\mathrm{d}y,$$

我们对 $z\in\mathbb{R}$ 的取值进行分类讨论:

a) $z<0$, $F_Z(z)=\iint_{x+y\leqslant z}\mathrm{d}x\mathrm{d}y=0$;

b) $0\leqslant z<1$, $F_Z(z)=\int_0^z\left(\int_0^{z-x}\mathrm{e}^{-y}\mathrm{d}y\right)\mathrm{d}x=z-1+\mathrm{e}^{-z}$;

c) $z\geqslant 1$, $F_Z(z)=\int_0^1\left(\int_0^{z-x}\mathrm{e}^{-y}\mathrm{d}y\right)\mathrm{d}x=1+\mathrm{e}^{-z}(1-\mathrm{e})$,

于是

$$F_Z(z)=\begin{cases}0, & z<0,\\ z-1+\mathrm{e}^{-z}, & 0\leqslant z<1,\\ 1+\mathrm{e}^{-z}(1-\mathrm{e}), & z\geqslant 1.\end{cases}$$

求其一阶导数, 则

$$f_Z(z)=\begin{cases}0, & z<0\\ 1-\mathrm{e}^{-z}, & 0\leqslant z<1,\\ (\mathrm{e}-1)\mathrm{e}^{-1}, & z\geqslant 1.\end{cases}$$

2. $Z=\dfrac{Y}{X},Z=XY$ 的分布

设 (X,Y) 是二维连续型随机变量, 它具有概率密度 $f(x,y)$, 则 $Z=\dfrac{Y}{X},Z=XY$ 仍为连续型随机变量, 其概率密度分别为

$$f_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|f(x,xz)\mathrm{d}x,$$

$$f_{XY}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f\left(x,\frac{z}{x}\right)\mathrm{d}x.$$

特别地, 若 X,Y 相互独立, 且具有边缘概率密度 $f_X(x),f_Y(y)$, 则

$$f_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|f_X(x)f_Y(xz)\mathrm{d}x,$$

$$f_{XY}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f_X(x)f_Y\left(\frac{z}{x}\right)\mathrm{d}x.$$

3. $M=\max(X,Y),m=\min(X,Y)$ 的分布

设 X,Y 是两个相互独立的随机变量, 它们的分布函数分别为 $F_X(x),F_Y(y)$, 则 $M=\max(X,Y),m=\min(X,Y)$ 的分布函数为

$$F_{\max}(z)=F_X(z)F_Y(z),$$

$$F_{\min}(z)=1-(1-F_X(z))(1-F_Y(z)).$$

第 4 章　随机变量的数字特征

4.1　数学期望

4.1.1　离散型随机变量的数学期望

设离散型随机变量 X 具有分布律

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

若级数 $\sum\limits_{k=1}^{\infty}x_kp_k$ 绝对收敛, 则称级数 $\sum\limits_{k=1}^{\infty}x_kp_k$ 的和为随机变量 X 的**数学期望** (expectation), 记为 $\mathbb{E}(X)$, 即

$$\mathbb{E}(X)=\sum_{k=1}^{\infty}x_kp_k.$$

注. 数学期望的本质是估值.

例 4.1.1. (2019 春, A 卷). 已知甲、乙两个袋子中装有同种产品, 其中甲袋中有 4 件合格品和 3 件次品, 乙袋中有 3 件合格品, 从甲袋中任意取 3 件袋子装入乙袋, 求: (1) 乙袋中次品件数 X 的数学期望 $\mathbb{E}(X)$; (2) 从乙袋中任取 1 件产品, 该品是次品的概率.

解. (1) 乙袋中所有的次品都是从甲袋中取得的, 因此 X 服从超几何分布,

$$P\{X=k\}=\frac{\mathrm{C}_3^k\cdot\mathrm{C}_4^{3-k}}{\mathrm{C}_7^3},\quad k=1,2,3,$$

则 X 的分布律为

X	0	1	2	3
P	4/35	18/35	12/35	1/35

因此数学期望

$$\mathbb{E}(X)=0\times\frac{4}{35}+1\times\frac{18}{35}+2\times\frac{12}{35}+3\times\frac{1}{35}=\frac{9}{7}.$$

(2) 设事件 B 表示从乙箱中取得 1 件产品为次品, 则

$$P(B)=\sum_{k=1}^{3}P(X=k)\cdot P(B|X=k)=\frac{4}{35}\cdot\frac{\mathrm{C}_0^1}{\mathrm{C}_6^1}+\frac{18}{35}\cdot\frac{\mathrm{C}_1^1}{\mathrm{C}_6^1}+\frac{12}{35}\cdot\frac{\mathrm{C}_2^1}{\mathrm{C}_6^1}+\frac{1}{35}\cdot\frac{\mathrm{C}_3^1}{\mathrm{C}_6^1}=\frac{3}{14}.$$

4.1.2 连续型随机变量的数学期望

设连续型随机变量 X 的概率密度为 $f(x)$, 若积分

$$\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$$

绝对收敛, 则称积分 $\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$ 的值为随机变量 X 的数学期望, 记为 $\mathbb{E}(X)$, 即

$$\mathbb{E}(x) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x.$$

例 4.1.2. (2022 年 II 级期末). 随机变量 X 服从均匀分布 $\mathrm{U}(0,1)$, 令随机变量 $Y = \mathrm{e}^X$.
(1) 求随机变量 Y 的概率密度函数;
(2) 求 Y 的数学期望.

解. (1) 由于 $X \sim \mathrm{U}(0,1)$, 因此 X 的概率密度函数为

$$p(x) = \begin{cases} 1, & x \in (0,1), \\ 0, & x \in \mathbb{R} \setminus (0,1), \end{cases}$$

根据分布函数的定义,

$$F_Y(y) = P\{Y \leqslant y\} = P\left\{\mathrm{e}^X \leqslant y\right\} = P\{X \leqslant \ln y\},$$

当 $y = 1$ 时 $\ln y = 0$, 当 $y = \mathrm{e}$ 时, $\ln y = 1$. 于是

i. $\forall y \leqslant 1\,(P\{X \leqslant \ln y\} = 0)$,
ii. $\forall y \in (1,\mathrm{e})\left(P\{X \leqslant \ln y\} = \int_{-\infty}^{+\infty} p(x)\mathrm{d}x = \int_0^{\ln y} \mathrm{d}x = \ln y\right)$,
iii. $\forall y \geqslant \mathrm{e}\,(P\{X \leqslant \ln y\} = 1)$.

所以 Y 的分布函数为

$$F_Y(y) = \begin{cases} 0, & y \leqslant 1, \\ \ln y, & y \in (1,\mathrm{e}), \\ 1, & y \geqslant \mathrm{e}, \end{cases}$$

所以

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Y(y) = \begin{cases} 1/y, & y \in (1,\mathrm{e}), \\ 0, & y \in \mathbb{R} \setminus (1,\mathrm{e}). \end{cases}$$

(2) 根据数学期望的定义,

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} yf_Y(y)\mathrm{d}y = \int_0^{\mathrm{e}} y\frac{1}{y}\mathrm{d}y = \mathrm{e} - 1.$$

4.1.3 随机变量的函数的数学期望

随机变量的函数仍然为随机变量, 因此要求随机变量的函数的数学期望, 只要在对应的计算公式中将随机变量的取值换成对应的函数值即可.

1. 一维随机变量的函数数学期望

若 Y 是随机变量 X 的函数 $Y=\varphi(X)$, 其中 φ 是连续函数. 设离散型随机变量 X 具有分布律

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

若级数 $\sum\limits_{k=1}^{\infty}\varphi(x_k)p_k$ 绝对收敛, 则 Y 的数学期望

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

设连续型随机变量 X 具有概率密度函数 $f(x)$ 且积分 $\int_{-\infty}^{+\infty}\varphi(x)f(x)\mathrm{d}x$ 绝对收敛, 则 Y 的数学期望

$$\mathbb{E}(Y)=\mathbb{E}(\varphi(X))=\int_{-\infty}^{+\infty}\boxed{\varphi(x)}f(x)\mathrm{d}x.$$

2. 二维随机变量的函数数学期望

设 W 是随机变量 X,Y 的函数 $W=\varphi(X,Y)$, 其中 φ 为连续函数. 设二维离散型随机变量 (X,Y) 具有分布律 $P\{X=x_i,Y=y_j\}=p_{ij}(i,j=1,2,\cdots)$, 且级数 $\sum\limits_{j=1}^{\infty}\sum\limits_{i=1}^{\infty}g(x_i,y_i)p_{ij}$ 绝对收敛, 则 W 的数学期望

$$\mathbb{E}(W)=\mathbb{E}(\varphi(X,Y))=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\boxed{g(x_i,y_i)}p_{ij};$$

设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y)$, 且积分 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\varphi(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$ 绝对收敛, 则 W 的数学期望

$$\mathbb{E}(W)=\mathbb{E}(\varphi(x,y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\boxed{\varphi(x,y)}\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\boxed{\varphi(x,y)}f(x,y)\mathrm{d}y\mathrm{d}x.$$

例 4.1.3. 设 (X,Y) 的联合分布律为

$$\begin{bmatrix}1/4 & 1/2\\ 1/8 & 1/8\end{bmatrix},$$

则 $Z=XY$ 的数学期望

$$\mathbb{E}(Z)=\mathbb{E}(XY)=0\times\left(\frac{1}{4}+\frac{1}{2}+\frac{1}{8}\right)+1\times\frac{1}{8}=\frac{1}{8}.$$

4.1.4 数学期望的性质

性质 1. 对于任意的常数 C, $\mathbb{E}(C) = C$.

性质 2. 设随机变量 $X_1, X_2, \cdots, X_n$ 的数学期望 $\mathbb{E}(X_1), \mathbb{E}(X_2), \cdots, \mathbb{E}(X_n)$ 存在, 则它们的线性组合的数学期望是它们数学期望的线性组合, 即

$$\mathbb{E}(k_1X_1 + k_2X_2 + \cdots + k_mX_m) = k_1\mathbb{E}(X_1) + k_2\mathbb{E}(X_2) + \cdots + k_n\mathbb{E}(X_n),$$

这就是说, 数学期望是随机变量的一个线性函数.

性质 3. 若 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量, 则

$$\mathbb{E}\left(\prod_{k=1}^{n} X_k\right) = \prod_{k=1}^{n} (\mathbb{E}(X_k)).$$

注. 需要说明的是, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ 不能推出 X, Y 相互独立.

4.2 方差

4.2.1 方差的定义

设 X 是一个随机变量, 若 $\mathbb{E}\left((X - \mathbb{E}(X))^2\right)$ 存在, 则称 $\mathbb{E}\left((X - \mathbb{E}(X))^2\right)$ 为 X 的**方差** (variance), 记作 $\mathbb{D}(X)$ 或 $\mathrm{var}(X)$, 即

$$\mathbb{D}(X) = \mathbb{D}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

注. 方差的本质就是随机变量 X 与其数学期望 $\mathbb{E}(X)$ 的偏离值的平方 $(X - \mathbb{E}(X))^2$ 的数学期望. 用于衡量数据点偏离其数学期望的平均程度, 刻画了一组数据的稳定性.

4.2.2 方差的计算公式

根据数学期望的性质,

$$\begin{aligned}
\mathbb{E}\left((X - \mathbb{E}(X))^2\right) &= \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}^2(X)\right) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}\left(X\mathbb{E}(X)\right) + \mathbb{E}(\mathbb{E}^2(X)) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}^2(X) + \mathbb{E}^2(X) \\
&= \mathbb{E}(X^2) - \mathbb{E}^2(X),
\end{aligned}$$

记得随机变量 X 的数学期望 $\mathbb{E}(X)$ 要么是一个级数的和, 要么是一个反常积分, 它们本质上都是一个常数, 是可以利用数学期望的性质从随机变量的函数中分离出来的, 于是我们实际上证明了

$$\boxed{\mathbb{D}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}^2(X)},$$

同时我们有 $\mathbb{E}(X^2) = \mathbb{D}(X) + \mathbb{E}^2(X)$.

例 4.2.1. 已知随机变量 X 服从以 λ 为参数的 Poisson 分布, 即 $X \sim P(\lambda)$, 且 $\mathbb{E}\left((X-1)(X-2)\right) = 1$, 求 λ.

解. 由 $X \sim P(\lambda)$, 知 $\mathbb{E}(X)=\lambda, \mathbb{D}(X)=\lambda$, 则

$$\mathbb{E}\left((X-1)(X-2)\right)=\mathbb{E}\left(X^2-3X+2\right)=\mathbb{E}(X^2)-3\mathbb{E}(X)+2,$$

其中 $\mathbb{E}(X^2)=\mathbb{D}(X)+\mathbb{E}^2(X)=\lambda^2+\lambda$, 因此

$$\mathbb{E}\left((X-1)(X-2)\right)=\lambda^2-2\lambda+2=1,$$

即

$$\lambda^2-2\lambda+1=0,$$

解之得 $(\lambda-1)^2=0$, 则 $\lambda=1$ 即为所求.

注. $X \sim P(\lambda) \iff \left(P\{X=k\}=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad \mathbb{E}(X)=\mathbb{D}(X)=\lambda\right)$.

4.2.3 方差的性质

性质 1. 设 C 为常数, 则 $\mathbb{D}(C)=0$.

性质 2. 设 X 为随机变量, C 为常数, 则

$$\mathbb{D}(CX)=C^2\mathbb{D}(X), \quad \mathbb{D}(X+C)=\mathbb{D}(X).$$

证明. 我们先感性地认识一下这两个公式, 数乘运算会把数据点距离其数学期望的偏差扩大 C 倍, 取平方, 就得到 C^2 倍了, 而加减一个常数不会改变数据集的平均偏离程度, 所以与常数做加法不改变数据集的方差. □

性质 3. 设 $c\in\mathbb{R}$, 则函数 $u(c)=\mathbb{E}((X-c)^2)$, 则该函数有最小值点 $c=\mathbb{E}(X)$, 且

$$\min_{c\in\mathbb{R}} u(c)=\mathbb{E}((X-c)^2)=\mathbb{E}((X-\mathbb{E}(X))^2)=\mathbb{D}(X),$$

即随机变量在其数学期望附近的偏离和波动是最小的.

性质 4. 设 X_1, X_2 是两个随机变量, $k_1, k_2\in\mathbb{R}$ 为任意常数, 则

$$\mathbb{D}(k_1X_1+k_2X_2)=k_1^2\mathbb{D}(X_1)+k_2^2\mathbb{D}(X_2)+2k_1k_2\mathrm{cov}(X_1,X_2).$$

特别地, 当 X_1, X_2 相互独立时有 $\mathrm{cov}(X_1,X_2)=0$, 此时

$$\mathbb{D}(k_1X_1+k_2X_2)=k_1^2\mathbb{D}(X_1)+k_2^2\mathbb{D}(X_2).$$

例 4.2.2. 随机变量 X, Y 相互独立, X 服从正态分布 $\mathrm{N}(0,2)$, Y 服从正态分布 $\mathrm{N}(1,2)$, 则随机变量 $X-2Y$ 服从的分布是 ____________.

解. 两个服从正态分布的随机变量的线性组合仍然服从正态分布. 由于 $X\sim\mathrm{N}(0,2)$, 因此 $\mathbb{E}(X)=0, \mathbb{D}(X)=2$, 而 $Y\sim\mathrm{N}(1,2)$, 因此 $\mathbb{E}(Y)=1, \mathbb{D}(X)=2$, 于是

$$\mathbb{E}(X-2Y)=\mathbb{E}(X)-2\mathbb{E}(Y)=-2,$$

$$\mathbb{D}(X-2Y)=\mathbb{D}(X)+(-2)^2\mathbb{D}(Y)=2+4\times 2=10,$$

于是 $X-2Y\sim\mathrm{N}(-2,10)$.

4.3　协方差及相关系数

4.3.1　协方差与相关系数的定义

设 (X,Y) 为二维随机变量, 如果 $\mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right)$ 存在, 则称它为随机变量 X,Y 的**协方差** (covariance), 记作 $\operatorname{cov}(X,Y)$, 即

$$\operatorname{cov}(X,Y) \xlongequal{\text{def}} \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right).$$

设 (X,Y) 为二维随机变量, 如果 $\mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right)$ 存在, 则称

$$\rho_{XY} \xlongequal{\text{def}} \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathbb{D}(X)}\sqrt{\mathbb{D}(Y)}}$$

为 X,Y 的**相关系数** (relevant coefficient).

4.3.2　协方差的计算

$$\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y).$$

4.3.3　协方差与相关系数的性质

1. 协方差的性质

性质 1. $\operatorname{cov}(X,X)=\mathbb{E}\left((X-\mathbb{E}(X))(X-\mathbb{E}(X))\right)=\mathbb{D}(X)$.

性质 2. 由于在协方差的定义中 X,Y 是对称的, 因此 $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$.

性质 3. 如果 $c\in\mathbb{R}$ 为常数, 则 $\operatorname{cov}(c,X)=0$.

性质 4. $\operatorname{cov}(k_1X_1,k_2X_2)=k_1k_2\operatorname{cov}(X_1,X_2)$.

性质 5. $\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cox}(X_2,Y)$.

性质 6. $\operatorname{cov}(k_1X_1+k_2X_2,Y)=k_1\operatorname{cov}(X_1,Y)+k_2\operatorname{cov}(X_2,Y)$.

性质 7. $\mathbb{D}(X+Y)=\mathbb{D}(X)+\mathbb{D}(Y)+2\operatorname{cov}(X,Y)$.

2. 相关系数的性质

性质 1. $|\rho_{XY}|\leqslant 1$, 且 $|\rho_{XY}|=1$ 的充分必要条件是存在常数 $a,b\in\mathbb{R}$ 使得 $P\left\{Y=a+bX\right\}=1$. 其中,

$$\rho_{XY}=-1 \iff \exists a<0\exists b\in\mathbb{R}\left(P\left\{Y=a+bX\right\}=1.\right)$$
$$\rho_{XY}=1 \iff \exists a>0\exists b\in\mathbb{R}\left(P\left\{Y=a+bX\right\}=1.\right)$$

性质 2. $\rho_{XY}=0 \iff \operatorname{cov}(X,Y)=0$, 此时称 X,Y 不相关.

性质 3. 相互独立的随机变量 X,Y 的协方差 $\operatorname{cov}(X,Y)=0$, 需要注意的是, 相互独立的随机变量一定不相关, 但不相关的随机变量未必相互独立.

例 4.3.1. 随机变量 X 服从正态分布 $\mathrm{N}(0,2)$, 随机变量 Y 服从指数分布 $\mathrm{Exp}\left(\frac{1}{2}\right)$, $U=2X$, $V=X-Y$, $\mathrm{cov}(U,V)=4$, 则 $\rho_{XY}=$ ____________.

解. 根据协方差的性质,

$$\begin{aligned}\mathrm{cov}(U,V)&=\mathrm{cov}(2X,X-Y)=2\mathrm{cov}(X,X)-2\mathrm{cov}(X,Y)\\&=2\mathbb{D}(X)-2\mathrm{cov}(X,Y)=4,\end{aligned}$$

由于 $X\sim\mathrm{N}(0,2^2)$, 因此 $\mathbb{E}(X)=0,\mathbb{D}(X)=2^2=4$, 而 $Y\sim\mathrm{Exp}\left(\frac{1}{2}\right)$, 因此 $\mathbb{E}(Y)=\frac{1}{1/2}=2,\mathbb{D}(Y)=\frac{1}{(1/2)^2}=4$, 于是

$$\mathrm{cov}(X,Y)=2,$$

而

$$\rho_{XY}=\frac{\mathrm{cov}(X,Y)}{\sqrt{\mathbb{D}(X)}\sqrt{\mathbb{D}(Y)}}=\frac{2}{\sqrt{4}\times\sqrt{\frac{1}{(1/2)^2}}}=\frac{1}{2}.$$

例 4.3.2. 已知二维连续型随机变量 (X,Y) 的联合概率密度函数为

$$f(x,y)=\begin{cases}\frac{1}{8}(x+y), & 0\leqslant x\leqslant 2,0\leqslant y\leqslant 2,\\ 0, & \text{else},\end{cases}$$

求 $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathrm{cov}(X,Y)$, ρ_{XY}, $\mathbb{D}(X+Y)$.

解. 根据数学期望、方差、协方差和相关系数的定义:

(1) $\mathbb{E}(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\mathrm{d}x\mathrm{d}y=\int_0^2\int_0^2\frac{1}{8}(x+y)\mathrm{d}x\mathrm{d}y=\frac{7}{6}$,

(2) $\mathbb{E}(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\mathrm{d}x\mathrm{d}y=\int_0^2\int_0^2\frac{1}{8}(x+y)\mathrm{d}x\mathrm{d}y=\frac{7}{6}$,

(3) $\mathbb{E}(X^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^2\int_0^2\frac{1}{8}(x+y)\mathrm{d}x\mathrm{d}y=\frac{5}{3}$,

(4) $\mathbb{E}(Y^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y^2f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^2\int_0^2\frac{1}{8}(x+y)\mathrm{d}x\mathrm{d}y=\frac{5}{3}$,

(5) $\mathbb{D}(X)=\mathbb{E}(X^2)-\mathbb{E}^2(X)=\frac{5}{3}-\left(\frac{7}{6}\right)^2=\frac{11}{36}$,

(6) $\mathbb{D}(Y)=\mathbb{E}(Y^2)-\mathbb{E}^2(Y)=\frac{5}{3}-\left(\frac{7}{6}\right)^2=\frac{11}{36}$,

(7) $\mathrm{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)=\frac{4}{3}-\frac{7}{6}\times\frac{7}{6}=-\frac{1}{36}$,

(8) $\rho_{XY}=\frac{\mathrm{cov(X,Y)}}{\sqrt{\mathbb{D}(X)}\sqrt{\mathbb{D}(Y)}}=\frac{-1/36}{\sqrt{11/36}\sqrt{11/36}}=-\frac{1}{11}$,

(9) $\mathbb{D}(X+Y)=\mathbb{D}(X)+\mathbb{D}(Y)+2\mathrm{cov}(X,Y)=\frac{11}{36}+\frac{11}{36}+2\times\left(-\frac{1}{36}\right)=\frac{5}{9}$.

第 5 章　　大数定律及中心极限定理

5.1　大数定律

5.1.1　随机变量序列的收敛性

定义 5.1.1. 设随机变量 $Y_n(n=1,2,\cdots)$ 和随机变量 Y 的分布函数分别为 $F_n(x)(n=1,2,\cdots)$ 和 $F(x)$, 若在 $F(x)$ 的所有连续点 x 上都有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称随机变量序列 $\{Y_n\}$ **依分布收敛于随机变量** Y (random variable sequence $\{Y_n\}$ convergence to random variable Y by distribution), 记作 $Y_n \xrightarrow{L} Y$.

依分布收敛表示, 当 n 充分大时, Y_n 的分布函数 $F_n(x)$ 收敛于 Y 的分布函数 $F(x)$, 它是概率论中的一种较弱的收敛性.

定义 5.1.2. 设又随机变量序列 $\{Y_n\}$ 和随机变量 Y, 若对于任意的正数 ε, 有

$$\lim_{n\to\infty} P\{|Y_n - Y| < \varepsilon\} = 1$$

或

$$\lim_{n\to\infty} P\{|Y_n - Y| \geqslant \varepsilon\} = 0,$$

则称随机变量序列 $\{Y_n\}$ **依概率收敛于随机变量** Y (convergence to random variable Y by probability), 记作 $Y_n \xrightarrow{P} Y$.

定理 5.1.1. (Chebyshev 不等式). 设随机变量 X 的数学期望 $\mathbb{E}(X)=\mu$, 方差 $\mathbb{D}(X)=\sigma^2$, 则对于任意给定的正数 ε, 成立

$$P\{|X-\mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}.$$

该定理的分析表述为:

$$\begin{aligned} &\forall \varepsilon > 0 \left(P\{|X-\mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2} \right), \\ \Longleftrightarrow\ &\forall \varepsilon > 0 \left(P\{|X-\mu| < \varepsilon\} \geqslant 1 - \frac{\sigma^2}{\varepsilon^2} \right). \end{aligned}$$

例 5.1.1. 设随机变量 X 的方差 $\mathbb{D}(X)=2$, 则根据 Chebyshev 不等式, 估计 $P\{|x-\mathbb{E}(X)| \geqslant 2\} \leqslant$ __________.

解. 根据 Chebyshev 不等式, 对于数学期望为 $\mathbb{E}(X)$, 方差为 $\mathbb{D}(X)$ 的随机变量 X,

$$\forall \varepsilon > 0 \left(P\left\{|X - \mathbb{E}(X)| \geqslant \varepsilon\right\} \leqslant \frac{D(x)}{\varepsilon^2} \right),$$

取 $\varepsilon = 2$, 则

$$P\left\{|x - \mathbb{E}(X)| \geqslant 2\right\} \leqslant \frac{D(X)}{2^2} = \frac{1}{2}.$$

例 5.1.2. 设 $X_1, X_2, \cdots, X_n, \cdots$ 为独立同分布的随机变量序列, $\mathbb{E}\left(X_k\right) = \mu, \mathbb{D}\left(X_k\right) = \sigma^2, k = 1, 2, \cdots$, 令

$$Y_n = \frac{2}{n(n+1)} \sum_{k=1}^{n} k X_k,$$

证明: 随机变量序列 $\{Y_n\}$ 依概率收敛于 μ.

证明. 根据数学期望和方差的相关性质,

$$\begin{aligned}
\mathbb{E}\left(Y_n\right) &= \mathbb{E}\left(\frac{2}{n(n+1)} \sum_{k=1}^{n} k X_k\right) = \frac{2}{n(n+1)} \mathbb{E}\left(\sum_{k=1}^{n} k X_k\right) \\
&= \frac{2}{n(n+1)} \mathbb{E}\left(X_k\right) \sum_{k=1}^{n} k = \frac{2}{n(n+1)} \cdot \mu \cdot \frac{n(n+1)}{2} = \mu,
\end{aligned}$$

$$\begin{aligned}
\mathbb{D}\left(Y_n\right) &= \mathbb{D}\left(\frac{2}{n(n+1)} \sum_{k=1}^{n} k X_k\right) = \left(\frac{2}{n(n+1)}\right)^2 \mathbb{D}(X_k) \sum_{k=1}^{n} k^2 \\
&= \frac{4}{n^2(n+1)^2} \cdot \sigma^2 \cdot \frac{n(n+1)(2n+1)}{6} = \frac{2(2n+1)\sigma^2}{3n(n+1)},
\end{aligned}$$

于是随机变量 Y_n 的数学期望 $\mathbb{E}(Y_n)$ 与方差 $\mathbb{D}(Y_n)$ 均存在, 根据 Chebyshev 不等式,

$$\forall \varepsilon > 0 \left(P\left\{\left|Y_n - \mathbb{E}\left(Y_n\right)\right| \geqslant \varepsilon\right\} \leqslant \frac{\mathbb{D}\left(Y_n\right)}{\varepsilon^2} \right),$$

这就是说,

$$0 \leqslant P\left\{|Y_n - \mu| \geqslant \varepsilon\right\} \leqslant \frac{2(2n+1)\sigma^2}{3n(n+1)} \frac{1}{\varepsilon^2},$$

而 $\lim\limits_{n\to\infty} \frac{2(2n+1)\sigma^2}{3n(n+1)} \frac{1}{\varepsilon^2} = 0$, 于是根据极限的迫敛性,

$$\lim_{n\to\infty} P\left\{|Y_n - \mu| \geqslant \varepsilon\right\} = 0,$$

即 $Y_n \xrightarrow{P} \mu$. □

5.1.2 大数定律

定理 5.1.2. (弱大数定律). 设 $X_1, X_2, \cdots$ 是相互独立, 服从同一分布的随机变量序列, 且具有数学期望 $\mathbb{E}(X_k) = \mu\,(k = 1, 2, \cdots)$, 作前 n 个变量的算术平均 $\frac{1}{n}\sum\limits_{k=1}^{n} X_k$, 则对于任意给定的正数 ε, 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1.$$

例 5.1.3. 设 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立且都服从参数为 2 的指数分布, 则当 $n \to \infty$ 时, $Y_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2$ 依概率收敛于 ________.

解. 对于随机变量序列 $X_1^2, X_2^2, \cdots, X_n^2, \cdots$, 由于 $\{X_i\}$ 独立同分布, 因此 $\left\{X_i^2\right\}$ 也独立同分布, 于是按照大数定律,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mathbb{E}(X_i^2),$$

而 $\mathbb{E}(X_i^2) = \mathbb{E}^2(X_i) + \mathbb{D}(X_i)$, 由于 $X_i \sim \mathrm{Exp}\,(2)$, 因此

$$\mathbb{E}(X_i) = \frac{1}{2}, \quad \mathbb{D}(X_i) = \frac{1}{2^2} = \frac{1}{4},$$

因此 $\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \xrightarrow{P} \dfrac{1}{2}$.

5.2 中心极限定理

5.2.1 独立同分布的中心极限定理

一组独立同分布的随机变量 $X_1, X_2, \cdots, X_n$ 之和 $\sum\limits_{k=1}^{n} X_i$ 在标准化后, 当 n 充分大时近似地服从标准正态分布, 即

$$\frac{\sum\limits_{i=1}^{n} X_i - \mathbb{E}\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{n\mathbb{D}\left(\sum\limits_{i=1}^{n} X_i\right)}} \sim \mathrm{AN}(0,1),$$

其中 $\mu = \mathbb{E}(X_i)$ 为随机变量序列的均值, $\sigma = \sqrt{\mathbb{D}(X_i)}$ 为随机变量序列的标准差, $\mathrm{AN}(0,1)$ 表示**近似标准正态分布** (approxmation of standard normal distribution). 那么此时由于随机变量序列独立同分布, $\mathbb{E}\left(\sum\limits_{i=1}^{n} X_i\right) = n\mu, \mathbb{D}\left(\sum\limits_{i=1}^{n} X_i\right) = n\sigma^2$, 于是上式可以写成

$$\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim \mathrm{AN}(0,1),$$

这就是**独立同分布的中心极限定理**.

例 5.2.1. 一工厂生产产品装箱后, 每箱重量随机, 假设 1 箱平均重量为 50kg, 标准差 5kg, 用载重为 5t 的卡车运送. 试用中心极限定理说明至多装多少箱才能使不超载的概率大于 0.977? ($\Phi(2) = 0.977$).

解. 对该问题进行建模, 我们做如下假设:

(1) X_i 表示第 i 箱的重量;

(2) n 表示运载箱数;

(3) $X \xlongequal{\text{def}} \sum\limits_{i=1}^{n} X_i$ 表示 n 箱的总重量,

根据中心极限定理,

$$\frac{\sum\limits_{i=1}^{n} X_i - \mathbb{E}\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{n\mathbb{D}\left(\sum\limits_{i=1}^{n} X_i\right)}} \sim \mathrm{AN}(0,1),$$

而 $\{X_i\}$ 独立同分布, 因此 $\mathbb{E}\left(\sum\limits_{i=1}^{n} X_i\right) = n\mathbb{E}(X_i) = 50n$, $\mathbb{D}\left(\sum\limits_{i=1}^{n} X_i\right) = \sum\limits_{i=1}^{n} \mathbb{D}(X_i) = 25n$, 所以

$$\begin{aligned} P\{X \leqslant 5000\} &= \left\{ \frac{\sum\limits_{i=1}^{n} X_i - \mathbb{E}\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{n\mathbb{D}\left(\sum\limits_{i=1}^{n} X_i\right)}} \leqslant \frac{5000 - \mathbb{E}\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{n\mathbb{D}\left(\sum\limits_{i=1}^{n} X_i\right)}} \right\} \\ &= P\left\{\frac{X-50n}{5\sqrt{n}} \leqslant \frac{5000-50n}{5\sqrt{n}}\right\} \\ &\approx \Phi\left(\frac{5000-50n}{5\sqrt{n}}\right) > 0.977, \end{aligned}$$

由于 $\Phi(2) = 0.977$, 所以要使上式成立, 只需

$$\frac{5000-50n}{5\sqrt{n}} > 2,$$

解得 $n < 98.0199$, 故最多装 $[98.0199] = 98$ 箱.

注. 要点在于把题目所要求的随机变量标准化 (在这一题中就是把 n 箱重量的总和标准化, 而且需要强调的是, 5000 也要标准化, 否则描述的就不是同一个随机事件了).

例 5.2.2. 一信号站同时收到 20 个新号 $V_k (k = 1, 2, \cdots, 20)$, 设这些信号 V_k 相互独立且分布于参数为 5 的 Poisson 分布, 则 $P\left\{\sum\limits_{k=1}^{20} V_k > 105\right\} =$ ____________. $(\Phi(0.5) = 0.692)$.

解. 由于 $V_k \sim \pi(5)$, 于是 $\mathbb{E}(V_k) = \mathbb{D}(V_k) = 5$, 因此

$$\mathbb{E}\left(\sum_{k=1}^{20} V_k\right) = \sum_{k=1}^{20} \mathbb{E}(V_k) = 100, \mathbb{D}\left(\sum_{k=1}^{20} V_k\right) = \sum_{k=1}^{20} \mathbb{D}(V_k) = 100,$$

根据题设条件,

$$P\left\{\sum_{k=1}^{20} V_k > 105\right\} = P\left\{\frac{\sum\limits_{k=1}^{20} V_k - \mathbb{E}\left(\sum\limits_{k=1}^{20} V_k\right)}{\sqrt{\mathbb{D}\left(\sum\limits_{k=1}^{20} V_k\right)}}\right\} = P\left\{\frac{\sum\limits_{k=1}^{20} V_k - 100}{\sqrt{100}} > \frac{105-100}{\sqrt{100}}\right\},$$

根据中心极限定理,

$$\text{上式} = P\left\{\frac{\sum_{k=1}^{20} V_k - 100}{\sqrt{100}} > 0.5\right\} = 1 - \Phi(0.5) = 0.308.$$

例 5.2.3. 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立同分布于均匀分布 $\mathrm{U}(0,1)$, 由中心极限定理, 当 n 充分大时, 随机变量 $Y = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 渐进服从 ____________.

解. 由题意知, $\mathbb{E}(X_i) = \frac{1-0}{2}, \mathbb{D}(X_i) = \int_0^1 \left(x^2 - \frac{1}{2}\right)^2 \mathrm{d}x = \frac{1}{12}$, 于是

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2\right) = \mathbb{D}\left(X_i\right) + \mathbb{E}^2\left(X_i\right) = \frac{1}{12} + \frac{1}{2^2} = \frac{1}{3}.$$

$$\begin{aligned}\mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) &= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n}\mathbb{D}\left(X_i^2\right) = \frac{1}{n}\int_0^1 \left(x^2 - \mathbb{E}\left(X_i^2\right)\right)^2 \mathrm{d}x \\ &= \frac{1}{n}\int_0^1 \left(x^2 - \left(\int_0^1 x^2 \mathrm{d}x\right)\right)^2 \mathrm{d}x = \frac{4}{45n}.\end{aligned}$$

因此 $Y \sim \mathrm{AN}\left(\frac{1}{3}, \frac{4}{45n}\right)$.

例 5.2.4. 设随机变量 $X_1, X_2, \cdots, X_{2n}$ 相互独立同分布, 且 $\mathbb{E}\left(X_i\right) = \mathbb{D}\left(X_i\right) = 1$, 当 $C =$____________ 时, 当 n 充分大时有 $Y_n = C\sum_{i=1}^{n}\frac{X_{2i} - X_{2i-1}}{\sqrt{n}}$ 近似服从标准正态分布.

解. 设 $Z_i = X_{2i} - X_{2i-1}$, 则 $\mathbb{E}\left(Z_i\right) = \mathbb{E}\left(X_{2i}\right) - \mathbb{E}\left(X_{2i-1}\right) = 0$, $\mathbb{D}\left(Z_i\right) = \mathbb{D}\left(X_{2i}\right) + \mathbb{D}\left(X_{2i-1}\right) = 2$, 于是

$$\mathbb{E}\left(\frac{C}{\sqrt{n}}\sum_{i=1}^{n} Z_i\right) = \frac{C}{\sqrt{n}} \cdot n \cdot 0 = 0,$$

$$\mathbb{D}\left(\frac{C}{\sqrt{n}}\sum_{i=1}^{n} Z_i\right) = \frac{C^2}{n} \cdot n \cdot 2 = 2C^2,$$

根据独立同分布的中心极限定理,

$$\frac{\frac{C}{\sqrt{n}}\sum_{i=1}^{n} Z_i}{\sqrt{2C^2}} = \frac{C}{\sqrt{n}}\sum_{i=1}^{n} Z_i \sim \mathrm{AN}(0,1),$$

得 $\sqrt{2C^2} = 1$, 解得 $C = \frac{\sqrt{2}}{2}$.

5.2.2 二项分布的中心极限定理

如果随机变量序列 X_k 相互独立且服从二项分布 $\mathbb{B}(n,p)$, 则随机变量序列的均值 $\mathbb{E}\left(X_k\right)=np$, 标准差 $\sqrt{\mathbb{D}\left(X_k\right)}=\sqrt{np(1-p)}$, 于是中心极限定理变为

$$\frac{\sum\limits_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \sim \mathrm{AN}(0,1),$$

这就是 **De Moivre-Laplace 中心极限定理**.

例 5.2.5. 某单位电话总机上有 200 条分机, 设每台分机有 5% 的时间需要使用外线, 且每台分机是否使用外线时相互独立的, 问: 总计需要设置多少条外线才能以 90% 的概率保证所有分机在需要使用外线时可供使用?

解. 设同时使用外线的分机总数为 X, 则 $X \sim \mathrm{B}(200, 0.05)$, 设外线总数为 N, 由已知, $\mathbb{E}(X)=200\times 0.05=10, \mathbb{D}(X)=200\times 0.05\times(1-0.05)=9.5$, 按照上述约定, "以 90% 的概率保证所有分机在需要使用外线时可供使用"就是说

$$P\left\{X \leqslant N\right\} \geqslant 0.9,$$

标准化,

$$P\left\{\frac{X-10}{\sqrt{9.5}} \leqslant \frac{N-10}{\sqrt{9.5}}\right\} \geqslant 0.9,$$

根据 De Moivre-Laplace 中心极限定理,

$$\frac{X-10}{\sqrt{9.5}} \sim \mathrm{AN}(0,1),$$

于是

$$P\left\{\frac{X-10}{\sqrt{9.5}} \leqslant \frac{N-10}{\sqrt{9.5}}\right\} \approx \Phi\left(\frac{N-10}{\sqrt{9.5}}\right) \geqslant 0.9 = \Phi(1.28),$$

即

$$\frac{N-10}{\sqrt{9.5}} \geqslant 1.28,$$

解之得 $N \geqslant 13.94$, 因此应该设置 ceiling$(13.94)=14$ 条外线.

第二部分

数理统计

第 6 章　样本及抽样分布

6.1　随机样本

6.1.1　总体和样本

数理统计中往往要研究有关对象的某一项数量指标, 对这项数量指标进行试验或观察, 将试验的全部可能的观察值称为总体, 每个可能的观察值称为个体, 总体的每一个个体是随机试验的一个观察值, 因此它是某一随机变量 X 的值. 这样以来, 一个总体对应一个随机变量, 总体所对应的随机变量的 X 的分布函数和数字特征就称为总体的分布函数和数字特征, 我们也用 X 来表示总体. 总体中包含的个体数量称为样本容量. 抽样调查指的是从总体中随机抽取一定数量的样本进行分析, 是数理统计的常用研究方法.

定义 6.1.1. (样本). 在相同的条件下, 对分布函数为 F 的总体 X 进行 n 次重复的、独立的观察, 得到 n 个结果 $X_1, X_2, \cdots, X_n$, 如果它们满足以下两条性质:

a) $X_1, X_2, \cdots, X_n$ 与总体 X 具有相同的分布;

b) $X_1, X_2, \cdots, X_n$ 相互独立,

则称随机变量 $\{X_k\}_{k=1}^n$ 为来自总体 X 或来自分布函数的样本容量为 n 的**简单随机样本**, 简称样本. 样本的观察值称为**样本值**, 这就是说, 设 $x_1, x_2, \cdots, x_n$ 分别为 $X_1, X_2, \cdots, X_n$ 的观察值, 则称它们为样本值.

6.1.2　统计量

对于一组样本观察值 $x_1, x_2, \cdots, x_n$, 我们往往可以得到一组数据, 对这些数据进行加工即可得到对总体 X 的统计推断. 加工处理这些数据的一种有效方法是, 构造样本的一些适当的、能反映样本的数字特征的函数. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, $g(X_1, X_2, \cdots, X_n)$ 是 $X_1, X_2, \cdots, X_n$ 的一个函数. 如果 g 中不包含任何未知参数, 则称 $g(X_1, X_2, \cdots, X_n)$ 为一个统计量. 如果 $x_1, x_2, \cdots, x_n$ 是对应于样本 $X_1, X_2, \cdots, X_n$ 的样本值, 则称 $g(x_1, x_2, \cdots, x_n)$ 为 $g(X_1, X_2, \cdots, X_n)$ 的观察值.

注. 统计量不包含任何未知参数.

我们把一些常用的统计量罗列在此:

1° 样本平均值: $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

2° 样本方差: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right)$,

注. 注意分母上是 $n-1$ 而不是 n.

定理 6.1.1. 总体 X 的均值为 μ, 方差为 σ^2, $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, $\bar{X}$ 为样本均值, S^2 为样本方差, 则总有

$$\mathbb{E}\left(\bar{X}\right)=\mu, \quad \mathbb{D}\left(\bar{X}\right)=\frac{\sigma^2}{n}, \quad \mathbb{E}(S^2)=\sigma^2.$$

证明. 根据样本的定义, $X_1, X_2, \cdots, X_n$ 相互独立且分布于总体 X 相同, 于是 $\mathbb{E}(X_i)=\mu, \mathbb{D}(X_i)=\sigma^2$,

$$\begin{aligned}\mathbb{E}\left(\bar{X}\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i\right) = \frac{n\mu}{n} = \mu.\end{aligned}$$

$$\begin{aligned}\mathbb{D}\left(\bar{X}\right) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\mathbb{D}\left(\sum_{i=1}^{n}X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}\left(X_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(S^2\right) &= \mathbb{E}\left(\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right)\right) = \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right) \\ &= \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}X_i^2\right) - n\mathbb{E}\left(\bar{X}^2\right),\end{aligned}$$

其中,

$$\mathbb{E}(X_i^2) = \mathbb{D}(X_i) + \mathbb{E}^2(X_i) = \sigma^2 + \mu^2,$$

$$\mathbb{E}\left(\bar{X}^2\right) = \mathbb{D}\left(\bar{X}\right) + \mathbb{E}^2\left(\bar{X}\right) = \frac{\sigma^2}{n} + \mu^2,$$

于是

$$\mathbb{E}(S^2) = \frac{1}{n-1}\left(n\left(\sigma^2+\mu^2\right) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) = \sigma^2.$$

不管服从什么分布, 只要均值和方差存在, 这三个式子总是成立的. □

6.2　抽样分布

统计量的分布称为**抽样分布**. 本节我们介绍来自正态总体的几个常用的统计量的分布.

6.2.1 χ^2 分布

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $\mathrm{N}(0,1)$ 的样本, 则称统计量

$$\chi^2 = \sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

服从自由度为 n 的 χ^2 **分布** (chi-square distribution), 记为 $\chi^2 \sim \chi^2(n)$.

性质 1. $\mathbb{E}\left(\chi^2\right) = n, \mathbb{D}\left(\chi^2\right) = 2n$.

性质 2. (可加性). $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$, 且 χ_1^2, χ_2^2 相互独立, 则

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2).$$

定义 6.2.1. (χ^2 分布的分位数). 对于给定的正数 $\alpha \in (0,1)$, 称满足条件

$$P\left\{\chi^2 > \chi_\alpha^2(n)\right\} = \int_{\chi_\alpha^2(n)}^{+\infty} f(y)\mathrm{d}y = \alpha$$

的数 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的**上 α 分位数**, 其中 $f(y)$ 为 $\chi^2(n)$ 分布的概率密度函数.

6.2.2 t 分布 (Student 分布)

设 $X \sim \mathrm{N}(0,1), Y \sim \chi^2(n)$, 且 X, Y 相互独立, 则称随机变量

$$t = \frac{X}{\sqrt{\dfrac{Y}{n}}}$$

服从自由度为 n 的 t **分布**, 记作 $t \sim t(n)$.

性质 1. 当 $n > 45$ 时, $T = \dfrac{X}{\sqrt{Y/n}} \sim \mathrm{AN}(0,1)$.

定义 6.2.2. (t 分布的分位数). 对于给定的正数 $\alpha \in (0,1)$, 称满足条件

$$P\left\{t > t_\alpha(n)\right\} = \int_{t_\alpha(n)}^{+\infty} h(y)\mathrm{d}y = \alpha$$

的数 $t_\alpha(n)$ 为 t 分布的**上 α 分位数**, 其中 $h(y)$ 为 t 分布的概率密度函数.

性质 2. $\forall \alpha \in (0,1)\, (t_{1-\alpha}(n) = -t_\alpha(n))$.

例 6.2.1. 设总体 $X \sim t(23)$, 其分布函数为 $F(x)$, 如果 $F(b) = \alpha$, 其中 $0 < \alpha < \dfrac{1}{2}$, 则 X 的上 α 分位数为 ____________.

A. b B. $-b$ C. $\dfrac{b}{2}$ D. $-\dfrac{b}{2}$

解. 根据 $F(b) = P\{X \leqslant b\} = \alpha \in (0, 1/2)$ 可知 $b < 0$, 且 $\displaystyle\int_{-\infty}^{b} h(y)\mathrm{d}y = \alpha$, 又根据上 α 分位数 $t_\alpha(23)$ 的定义可知, $\displaystyle\int_{t_\alpha(23)}^{+\infty} h(y)\mathrm{d}y = \alpha$, 根据 t 分布的对称性可知 $|b| = |\alpha|$, 因此 $\alpha = -b$.

6.2.3　F 分布

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, 且 U, V 相互独立, 则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

服从自由度为 n_1, n_2 的 F **分布**, 记为 $F \sim F(n_1, n_2)$.

性质 1. 如果 $F \sim F(n_1, n_2)$, 那么 $\dfrac{1}{F} \sim F(n_2, n_1)$.

定义 6.2.3. (F 分布的分位数). 对于给定的正数 $\alpha \in (0,1)$, 称满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1,n_2)}^{+\infty} \psi(y)\mathrm{d}y = \alpha$$

的数 $F_\alpha(n_1, n_2)$ 为 t 分布的**上 α 分位数**, 其中 $\psi(y)$ 为 F 分布的概率密度函数.

性质 2. $\forall \alpha \in (0,1) \left(F_{1-\alpha}(n_1, n_2) = \dfrac{1}{F_\alpha(n_2, n_1)} \right)$.

6.2.4　正态总体的样本均值与样本方差的分布

定理 6.2.1. 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $\mathrm{N}(\mu, \sigma^2)$ 的简单随机样本, $\bar{X}, S^2$ 分别为样本均值与样本方差, 则

1° $\bar{X} \sim \mathbb{N}\left(\mu, \dfrac{\sigma^2}{n}\right), \quad \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$;

2° $\dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;

3° $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$;

4° $\bar{X}$ 与 S^2 相互独立.

定理 6.2.2. 设 $X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别来自正态总体 $\mathbb{N}(\mu_1, \sigma_1^2)$、$\mathbb{N}(\mu_2, \sigma_2^2)$, 且这两个样本相互独立, 设 $\bar{X}, \bar{Y}$ 分别为它们的样本均值, S_1^2, S_2^2 分别为它们的样本方差, 则

1° $Z = \dfrac{\left(\bar{X}-\bar{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$;

2° $t = \dfrac{\left(\bar{X}-\bar{Y}\right) - (\mu_1 - \mu_2)}{S_W\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t((n_1-1)+(n_2-1))$, 其中 $S_W^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$;

3° $F = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$.

第 7 章　参数估计与假设检验

统计推断的两大主要工作分别是估计和假设检验, 参数估计的任务是, 已知总体 X 的分布函数的形式 $F(x;\theta)$ 的形式, 求解其中的未知参数 θ, $X_1, X_2, \cdots, X_n$ 为总体 X 的一个样本, $x_1, x_2, \cdots, x_n$ 为相应的样本值. 参数估计问题主要分为点估计和区间估计.

在总体的分布函数完全未知, 或只知道其形式但不知道其参数的情况下, 提出某些关于总体分布函数或关于其参数的假设, 然后抽取样本, 构造合适的统计量, 再根据样本对所提出的假设做出是接受还是拒绝的决策的问题, 称为**假设检验问题**. 假设检验问题的基本思想是, 小概率的事件一般不会发生, 因此我们往往基于所提出的假设对样本空间进行一个划分, 找出一个导致小概率事件发生的一个样本空间的子集来, 判断所构造的统计量的样本值是否落在这个子集当中, 来做出决策.

7.1　点估计

点估计的问题就是适当地选取一个统计量作为未知参数 θ 的估计, 称为估计量, 以估计量的观察值作为未知参数的近似值, 称为估计值, 估计量和估计值都记作 $\hat{\theta}$.

注. 注意估计量和估计值的区别, 估计量是一个样本的函数, 是一个统计量, 而估计值是基于一个具体的样本数据, 代入样本值得到的计算结果.

7.1.1　矩估计法

以样本矩作为相应的总体矩的估计量, 而以样本矩的连续函数作为相应的总体矩的连续函数的估计量的估计方法称为**矩估计**. 通过矩估计法得到的估计量称为**矩估计量**, 矩估计量的观察值称为**矩估计值**. 矩估计的核心是用样本均值 $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ 去估计总体均值 μ. 矩估计法的具体步骤如下:

(1) 求出总体矩 $\mu = \mathbb{E}(X) = g(\theta)$,

(2) 令 $\mu = g(\theta) = \bar{X}$, 反解得到矩估计量 $\hat{\theta} = h\left(\bar{X}\right)$,

(3) 将观察到的样本值 $\bar{x}$ (如果有的话) 代入 $\hat{\theta}$ 得到矩估计值.

7.1.2　最大似然估计法

最大似然估计法 (maximum likelihood estimarion method, MLE) 的思想是, 既然这组样本出现了, 那么我们有理由相信这组样本出现的概率是最大的, 我们构造这样的参数 θ 使得在这一个参数下, 这组样本出现的概率最大. 最大似然估计的具体做法是:

(1) 计算**似然函数** (likelihood): $L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)$, 其中 $\theta\in\Theta$, $f(x_i;\theta)$ 为 X 的概率密度函数 (连续型) 或分布律 (离散型).

(2) 取对数, 得到**对数似然函数** (log likelihood): $\ln L(\theta)=\sum_{i=1}^{n}\ln(f(x_i;\theta))$.

(3) 令 $\dfrac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=0$, 解得 $\hat{\theta}$.

7.2 估计量的评选标准

若估计量 $\hat{\theta}=\hat{\theta}\left(X_1,X_2,\cdots,X_n\right)$ 的数学期望 $\mathbb{E}\left(\hat{\theta}\right)$ 存在, 且

$$\forall\theta\in\Theta\left(\mathbb{E}(\hat{\theta})=\theta\right),$$

则称 $\hat{\theta}$ 是 θ 的**无偏估计量**.

例 7.2.1. 已知随机变量 X 的概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{2x}{theta^2}, & 0<x<\theta,\\ 0, & \text{else},\end{cases}$$

其中 θ 为未知参数, $X_1,X_2,\cdots,X_{2022}$ 为样本, 若 $C\sum_{i=1}^{2022}X_i^2$ 为 θ^2 的无偏估计, 则 $C=$____________.

解. 由题意知,

$$\mathbb{E}\left(C\sum_{i=1}^{2022}X_i^2\right)=\theta^2,$$

根据数学期望的性质,

$$\mathbb{E}\left(C\sum_{i=1}^{2022}X_i^2\right)=C\mathbb{E}\left(\sum_{i=1}^{2022}X_i^2\right)s=\theta^2,$$

由于 $X_1,X_2,\cdots,X_{2022}$ 为样本, 它们相互独立且与总体 X 同分布, 所以

$$\mathbb{E}\left(X_i^2\right)=\mathbb{E}(X^2)=\int_0^\theta x^2\cdot\frac{2x}{\theta^2}\mathrm{d}x=\frac{\theta^2}{2},$$

所以

$$\mathbb{E}\left(C\sum_{i=1}^{2022}X_i^2\right)=C\mathbb{E}\left(\sum_{i=1}^{2022}X_i^2\right)=2022C\frac{\theta^2}{2}=\theta^2,$$

解得 $C=\dfrac{1}{1011}$.

7.3 正态总体均值与方差的区间估计

区间估计的主要任务是, 对未知参数估计出一个范围, 它的主要思想是: 通过从总体抽取的样本, 根据一定的准确度要求构造出合适的区间作为未知参数真实值的估计区间.

7.3.1 参数的置信区间

设总体 X 的分布函数 $F(x;\theta)$ 含有一个未知参数 $\theta\in\Theta$, 其中 Θ 是 θ 可能取值的范围, 对于给定的 $\alpha\in(0,1)$, 若由来自总体 X 的样本 $X_1,X_2,\cdots,X_n$ 确定的两个统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$ 和 $\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$ 满足以下两个条件:

1° $\underline{\theta}<\overline{\theta}$;

2° $\forall\theta\in\Theta\left(P\left\{\underline{\theta}(X_1,X_2,\cdots,X_n)<\theta<\overline{\theta}(X_1,X_2,\cdots,X_n)\right\}\geqslant 1-\alpha\right)$,

则称随机区间 $\left(\underline{\theta},\overline{\theta}\right)$ 是 θ 的**置信水平** (或**置信度**) 为 $1-\alpha$ 的**置信区间**, $\underline{\theta},\overline{\theta}$ 分别称为置信水平为 $1-\alpha$ 的**置信下限**和**置信上限**.

7.3.2 单侧置信区间

设总体 X 的分布函数 $F(x;\theta)$ 含有一个未知参数 $\theta\in\Theta$, 其中 Θ 是 θ 可能取值的范围, 对于给定的 $\alpha\in(0,1)$, 若由来自总体 X 的样本 $X_1,X_2,\cdots,X_n$ 确定的统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$ 满足

$$\forall\theta\in\Theta\left(P\left\{\underline{\theta}(X_1,X_2,\cdots,X_n)<\theta\right\}\geqslant 1-\alpha\right),$$

则称随机区间 $(\underline{\theta},+\infty)$ 是 θ 的**置信水平** (或**置信度**) 为 $1-\alpha$ 的**置信区间**, $\underline{\theta}$ 称为置信水平为 $1-\alpha$ 的**置信下限**. 若 $\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$ 满足

$$\forall\theta\in\Theta\left(P\left\{\theta<\overline{\theta}(X_1,X_2,\cdots,X_n)\right\}\geqslant 1-\alpha\right),$$

则称随机区间 $\left(-\infty,\overline{\theta}\right)$ 是 θ 的**置信水平** (或**置信度**) 为 $1-\alpha$ 的**置信区间**, $\overline{\theta}$ 称为置信水平为 $1-\alpha$ **置信上限**.

7.3.3 求置信区间的方法

(1) 由已知数据求样本容量 n, 样本均值 $\bar{X}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i$ 与样本方差 $S^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i^2-n\bar{X}^2\right)$;

注. 在 CASIO fx-991CN X 函数计算器中, σ^2 对应的是样本方差, S^2 对应的是样本修正方差, 也就是我们一直以来使用的样本方差.

(2) 选择合适的枢轴量:

a. 单个正态总体: $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim \mathrm{N}(0,1)$、$t=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$、$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

b. 两个正态总体:

i. $Z=\dfrac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim \mathrm{N}(0,1)$;

ii. $t=\dfrac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{S_W\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t((n_1-1)+(n_2-1))$, 其中 $S_W^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$;

iii. $F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$.

(3) 根据枢轴量的分布函数图像, 做图分析置信区间的范围. 所谓置信区间, 就是未知参数落在这个范围内的概率很大, 因此可以利用分布函数图像的几何特征 (如果有对称性的话要注意对称性) 确定置信区间的范围, 它一般是枢轴量夹在两个上分位点的形式.

(4) 从上一步得到的不等式反解出未知参数的取值范围, 即为待求的置信区间.

7.4　假设检验问题

7.4.1　假设检验问题的基本概念

我们称需要着重考察的假设为**原假设**, 记作 H_0, 与原假设相对立的假设称为备择假设, 记作 H_1. 所构造出的用来观察和确定是否接受假设 H_0 的统计量称为**检验统计量**. 检验统计量基于假设, 将样本空间划分为两部分, 其中一部分对应 H_0 被拒绝的小概率事件, 称为**拒绝域**.

7.4.2　假设检验问题的求解步骤

(1) 由已知数据求样本容量 n, 样本均值 $\bar{X}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i$ 与样本方差 $S^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i^2-n\bar{X}^2\right)$;

注. 在 CASIO fx-991CN X 函数计算器中, σ^2 对应的是样本方差, S^2 对应的是样本修正方差, 也就是我们一直以来使用的样本方差.

(2) 选择合适的枢轴量 (枢轴量的选取与置信区间的求解完全一致):

a. 单个正态总体: $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim \mathrm{N}(0,1)$、$t=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$、$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

b. 两个正态总体:

i. $Z=\dfrac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim \mathrm{N}(0,1)$;

ii. $t=\dfrac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{S_W\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t((n_1-1)+(n_2-1))$, 其中 $S_W^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$;

iii. $F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$.

(3) 根据枢轴量的分布函数图像, 做图分析拒绝域的范围. 假设检验的基本思想是小概率事件几乎不会发生, 因此拒绝域是一个概率很小的区域, 它往往是从无穷远处 (或 0 点) 到某一个上分位点 (单侧检验的话只需要考虑一侧即可), 于是可以利用分布函数的几何特点确定拒绝域.

(4) 代入样本值判断检验统计量的观察值是否落在拒绝域内, 做出统计推断.

例 7.4.1. 对两批同类型的电子元件的电阻进行测试, 各抽取 6 件, 测得结果如下:

A 批	0.160	0.138	0.143	0.141	0.144	0.137
B 批	0.135	0.140	0.142	0.136	0.138	0.142

已知元件服从正态分布, $\bar{x} = 0.1405, \bar{y} = 0.1388, \sum_{i=1}^{6}(x_i - \bar{x})^2 = 3.75 \times 10^{-5}, \sum_{i=1}^{6}(y_i - \bar{y})^2 = 4.48 \times 10^{-5}$.
设 $\alpha = 0.05$, 问两批元件的电阻是否有显著差异.
参考数据: $F_{0.025}(5,5) = 7.15$, $t_{0.025}(10) = 2.2281$, $F_{0.05}(5,5) = 5.05$, $t_{0.05}(10) = 1.8125$.

解. 为了考察两批元件的电阻是否有显著差异, 我们从均值和方差两个角度出发来进行假设检验. 首先考虑方差, 假设 $H_0 : \sigma_1^2 = \sigma_2^2$, 则 $H_1 : \sigma_1^2 \neq \sigma_2^2$. 选取检验统计量

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \xlongequal{\sigma_1 = \sigma_2} \frac{\dfrac{1}{5}\sum_{i=1}^{6}(x_i - \bar{x})^2}{\dfrac{1}{5}\sum_{i=1}^{6}(y_i - \bar{y})^2} = \frac{3.75 \times 10^{-5}}{4.48 \times 10^{-5}} = 0.837,$$

拒绝域

$$(F \geqslant F_{0.025}(5,5) = 7.15) \vee \left(F \leqslant F_{1-0.025}(5,5) = \frac{1}{F_{0.05}(5,5)} = 0.13\right),$$

F 不在拒绝域内, 故接受原假设.
接下来我们考虑均值, 假设 $H_0 : \mu_1 = \mu_2$, 则 $H_1 : \mu_1 \neq \mu_2$. 选取检验统计量

$$t = \frac{\left|\bar{X} - \bar{Y}\right|}{S_W\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

其中

$$\begin{aligned} S_W &= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S^2}{(n_1 - 1) + (n_2 - 1)} = \frac{5 \times \dfrac{1}{5}\sum_{i=1}^{6}(x_i - \bar{x})^2 + 5 \times \dfrac{1}{5}\sum_{i=1}^{6}(y_i - \bar{y})^2}{10} \\ &= \frac{3.75 \times 10^{-5} + 4.48 \times 10^{-5}}{10} = 0.0029, \end{aligned}$$

于是 $t = 1.0153$. 拒绝域

$$|t| \geqslant t_{0.025}(10) = 2.2281,$$

t 不在拒绝域内, 故接受原假设. 因此, 可以认为两批元件的电阻没有显著差异.

附录 A　常用数据查询表

A.1　几种常用的概率分布表

分布	参数	分布律或概率密度	数学期望 $\mathbb{E}(X)$	方差 $\mathbb{D}(X)$
(0-1) 分布	$0<p<1$	$P\{X=k\}=p^k(1-p)^{1-k},\quad k=0,1$	p	$p(1-p)$
二项分布	$n\geqslant 1$ $0<p<1$	$P\{X=k\}=\mathrm{C}_n^k p^k(1-p)^{n-k}$ $k=0,1,\cdots,n$	np	$np(1-p)$
Poisson 分布	λ	$P\{X=k\}=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad k\in\mathbb{N}$	λ	λ
均匀分布	(a,b)	$f(x)=\begin{cases}\dfrac{1}{b-a}, & x\in(a,b),\\ 0, & x\in\mathbb{R}\setminus(a,b),\end{cases}$	$\dfrac{b-a}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布	λ	$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
正态分布	μ,σ	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2

后记

每一门课程的学习过程都是漫长的, 每一门课程的期末预习都是紧张、刺激而充实的. 无论你是一步一个脚印、扎实地学完了这门课, 还是像我撰写这本书时一样囫囵吞枣、72 小时速成, 不管怎样, 通过考试仅仅是一个开始, 概率论与数理统计在机器学习、人工智能、自然语言处理和传统的工程领域中都发挥着重要的作用, 它也是我们数学建模的重要理论支撑. 在未来的人生中, 你的任务就是尽量在概率统计的知识有用武之地的时候, 不遗余力地用它来解决你遇到的问题或是困难, 纸上得来终觉浅, 绝知此事要躬行, 唯有如此, 才不枉费你一朝一夕之间辛勤的学习.

你或许会忘掉这门课里的绝大多数知识, 但就像你不会记得自己曾经吃过多少碗饭一样, 这些知识都变成了你的养料, 等到有一天机会出现的时候, 这些知识将会在灵光一现中出现在你的大脑里供你调用, 为你解决问题提供有力的支持. 所以, 请不要认为对数学的学习只是徒劳, 因为真正美丽的花朵总是需要园丁辛勤的汗水的浇灌才能开放的.

心灵鸡汤就不再多讲了, 这些让人听了起鸡皮疙瘩的大道理人人都会讲, 而且大部分人讲的都比我好, 我就不在这里献丑了.

总之, 感谢我概率统计课的任课老师范艾利老师, 没有她的辛苦授课就不会有本书的诞生, 感谢我的辅导员程姣姣老师, 她对本书的诞生付出了辛勤的劳动. 感谢在我断更的这段时间继续支持我的新老粉丝朋友们, 希望这本书的表现能够对得起各位朋友的关注、信任、支持和期待. 再见, 朋友, 让我们把离别的惆怅和忧伤埋在心中, 就当作是播下了一颗再次相遇的种子, 带着彼此带给对方的祝福继续各自独一无二的人生吧. 概率论与数理统计的学习到这里就要告一段落了, 再见, 朋友, 希望本书对你的学习能有所帮助, 希望你能从中得到收获与启发. 再见, 朋友, 我们很快就会再见面的. 让我们在美妙的数学旅程中永远坚持走向上的道路, 追求真理、正义、智慧和自由.

钱锋
西北工业大学 数学与统计学院
2023 年 12 月于启翔湖畔

参考文献

[1] 盛骤, 谢式千, 潘承毅编, *概率论与数理统计第五版*, 北京: 高等教育出版社, 2019.12

[2] 盛骤, 谢式千, 潘承毅编, *概率论与数理统计附册学习辅导与习题选解: 浙大·五版*, 北京: 高等教育出版社, 2020.10

[3] 盛骤, 谢式千, 潘承毅编, *概率论与数理统计习题全解指南: 浙大·五版*, 北京: 高等教育出版社, 2020.10

I ♥ NPU

公诚勇毅　永矢毋忘
中华灿烂　工大无疆

2023 年　西北工业大学　中国西安